OBJECTIVE BAYESIAN ANALYSIS OF “ON/OFF” MEASUREMENTS

Diego Casadei
School of Engineering, FHNW, Bahnhofstrasse 6, 5210 Windisch, Switzerland; diego.casadei@fhnw.ch
Received 2014 July 25; accepted 2014 October 20; published 2014 December 9

ABSTRACT

In high-energy astrophysics, it is common practice to account for the background overlaid with counts from the source of interest with the help of auxiliary measurements carried out by pointing off-source. In this “on/off” measurement, one knows the number of photons detected while pointing toward the source, the number of photons collected while pointing away from the source, and how to estimate the background counts in the source region from the flux observed in the auxiliary measurements. For very faint sources, the number of photons detected is so low that the approximations that hold asymptotically are not valid. On the other hand, an analytical solution exists for the Bayesian statistical inference, which is valid at low and high counts. Here we illustrate the objective Bayesian solution based on the reference posterior and compare the result with the approach very recently proposed by Knoetig, and discuss its most delicate points. In addition, we propose to compute the significance of the excess with respect to the background-only expectation with a method that is able to account for any uncertainty on the background and is valid for any photon count. This method is compared to the widely used significance formula by Li & Ma, which is based on asymptotic properties.

Key words: gamma rays: general – methods: statistical

1. INTRODUCTION

In a counting experiment, the detector response to a trigger signal is saved, whenever at least one among (possibly many) different conditions is satisfied. The trigger requirements are defined in such a way to select interesting “events” and operate the detector in the most efficient way. Counting experiments are widespread in high-energy physics and astrophysics, and sometimes have to deal with very low event rates. This is the case, for example, when one tries to observe a very faint gamma-ray source with a space experiment, or when the goal is to detect the excess of counts corresponding to a new particle created by collisions produced by underground particle accelerators.

When only few events are collected, the asymptotic expressions that can be used with high count rates can no longer be adopted. Instead, it is of great importance to study the correct statistical model without simplifying assumptions that could invalidate the result. For counting experiments, it is commonly assumed that the integer number $n \geq 0$ of observed events follows the Poisson distribution:

$$\text{Poi}(n \mid a) = \frac{a^n}{n!} e^{-a},$$

where the real value $a \geq 0$ is the Poisson parameter, which coincides with the expected number of events and with the variance: $E[n] = V[n] = a$.

Realistic measurements always involve some degree of “background” counts, due to non-interesting events that satisfy (hopefully, but not always, with low probability) some trigger condition. We assume that the counts from the source of interest and those from the background are independent Poisson variables. A well-known property of the Poisson distribution is that the sum of two independent variables is again Poisson distributed, with the parameter given by the sum of the respective expectations:

$$P(n \mid s, b) = \text{Poi}(n \mid s + b) = \frac{(s + b)^n}{n!} e^{-(s+b)}, \quad (1)$$

where the expected number of events from the source $s \geq 0$ is the parameter of interest, and the background contribution $b \geq 0$ is the nuisance parameter ($n$ is an integer, whereas $s$ and $b$ are real numbers).

In high-energy astrophysics, it is common to estimate $b$ with the help of auxiliary measurements obtained by pointing the detector off the source. In this case, it is assumed that the source of interest does not contribute to the observed $k$ counts, such that one has a simple Poisson process:

$$\text{Poi}(k \mid B) = \frac{B^k}{k!} e^{-B}, \quad (2)$$

where $B \geq 0$ is the expected (background-only) photon count in the region off the source. By knowing the details (like the area of the sky and exposure time) of the source and off-source regions, it is possible to relate the expected counts from the background alone in the two regions: $b = \rho B$, where $\rho$ is a constant, assumed to be perfectly known (i.e., with negligible uncertainty compared to $B$).

In summary, the statistical inference about this “on/off” measurement makes use of the observed counts $n$ and $k$ in the source and off-source regions, of the known proportionality $\rho$ between the expected background fluxes in the two regions, and of the Poisson models (Equation (1) and Equation (2)) for the two measurements.

Recently, Knoetig (2014, MK2014 hereafter) summarized the previous approaches to the on/off inference problem and proposed an objective Bayesian solution that consists of two different steps. First, MK2014 checked whether the number $n$
of events in the source region is too high to be comfortably attributed to the background alone. If this is the case, one rejects the “null hypothesis” (background only) and claims a successful observation of the source. Next, the source intensity $s$ is estimated with the help of the auxiliary measurement. Note that MK2014 finds a (rather complicated) analytic solution to the Bayesian inference problem, in terms of special (Gamma and hypergeometric) functions, which are available in many libraries.

The procedure proposed by MK2014 is well motivated and it aims to achieve very desirable goals. However, it is not free from issues. The first source of possible problems is the proposed prior (Equation (15) of MK2014), which is obtained following Jeffreys’ rule in the bidimensional $(s, b)$ space (called $(\lambda_s, \lambda_{bg})$ in MK2014). Even though Jeffreys’ priors have a number of desirable properties for one-dimensional (1D) problems, it is well known that they behave badly in multidimensional problems (for a recent discussion, see Berger et al. 2013). Here we overcome this difficulty by reducing the problem to a 1D (marginal) model for which an objective prior is known.

One further complication in the procedure proposed by MK2014 is the comparison of the null hypothesis $H_0$ that $s = 0$ (called “background-only hypothesis” here) against the alternative hypothesis $H_1$ that $s > 0$ in the target region (the “source+background hypothesis”). In the absence of additional information, MK2014 assigns identical (prior) probabilities to the two hypotheses: $P(H_0) = P(H_1) = 1/2$. They are “nested” hypotheses, in the sense that $H_0$ assumes a single value, whereas $H_1$ allows the source strength to assume any other value in its domain. The fact that the measure of the allowed domain is zero for $H_0$ is problematic when using improper priors (see, for example, Bayarri et al. 2008 and references therein).

To overcome this problem, MK2014 fixes the ratio between the arbitrary scale factors of the source and background priors with a procedure relying on the assumption that when no counts are observed in both the target and off-source regions, the probabilities of $H_0$ and $H_1$ remain the same. This assumption is questionable, as counting zero events is not the same as performing no measurement. If the outcome is $k = 0$ in the auxiliary region, we learn that the background in the source region is very small (possibly negligible), and this may imply a better sensitivity (see Appendix A for more details). In addition, if we count $n = 0$ events in the source region, we know that there is no background contribution in this outcome and also that there is no count due to the source (which makes it easy to get an upper bound to its intensity). Thus, intuitively, one may think that the absence of counts in the target region decreases the probability that a source is actually there, hence increasing $P(H_0) = 1 - P(H_1)$, in contrast to MK2014. Although this way of thinking might also be criticized, what matters here is that the assumption that measuring zero counts in both regions does not update our degree of belief about the validity of both the background-only and the source+background hypotheses, is questionable and cannot be taken as a basis for the hypothesis test.

In this paper, we present the objective Bayesian approach based on the analytical solution of model (1) in the framework of the Bayesian reference analysis (Bernardo 2005a) obtained by Casadei (2012; DC2012 hereafter). This solution is based on the “reference prior” corresponding to the model given in Equation (1), which has a solid formal justification and does not suffer from the problems of multidimensional Jeffreys’ priors. In addition, frequentist coverage studies have been carried out by DC2012 and show a good average agreement between the posterior probability and the coverage (exact agreement is not possible, as this is a discrete problem).

The reference prior computed by DC2012 may be coded in any programming language (a C++ version is adopted here), and is also available in the Bayesian Analysis Toolkit (BAT; Caldwell et al. 2009). For the users of other analysis frameworks and/or programming languages, it may be useful to know that simple approximations also exist, which are even quicker to compute (Casadei 2014).

To illustrate the application of the objective Bayesian approach of DC2012 and compare it to MK2014, our solution will be applied in a simplified way (i.e., in a single-step procedure) to the same gamma-ray burst (GRB) data listed in Table 1 of MK2014. The marginal reference posterior probability density of the source strength $s$ will be estimated directly, without comparing $H_0$ with $H_1$. Similar to MK2014, the posterior distribution for $s$ will be summarized by providing its mode, i.e., the most likely value or peak position, together with highest posterior density (HPD) intervals, which are the narrowest intervals covering a predefined posterior probability. Whenever one of such intervals is limited by zero at the left, its right edge automatically provides an upper bound to the source intensity. In this case, it will be assumed that no source was detected, so that the alternative hypothesis of source+background is discarded.

In our conservative approach, when the posterior for $s$ suggests a non-negligible intensity, we check the statistical significance of the excess of counts with respect to the background-only hypothesis to decide whether or not to claim a successful source detection. This procedure provides results that are essentially equivalent to the two-step approach involving the comparison of two hypotheses, without complications arising from the latter in the presence of improper priors (as it is the case for the on/off problem).

2. METHODS

We have two Poisson models, characterizing a region where no source contribution is expected, Equation (2), and the target region where a source may produce counts in addition to those coming from the background alone, Equation (1). The available data are the number $n$ of photons detected when pointing to the source (called $N_{on}$ in MK2014), the $k$ off-source counts (called $N_{off}$ in MK2014), and the ratio $\rho = b/B$ (called $\alpha$ in MK2014) between the background fluxes in both regions. One first must estimate the background in the target region with the help of the auxiliary off-source measurement. Next, one considers the marginal model obtained by integrating over $b$ and finds the reference prior for $s$. Finally, one obtains the reference posterior for $s$ with the help of Bayes’ theorem.

The auxiliary measurement is analyzed first. The Poisson distribution (Equation (2)) allows us to estimate the off-source background intensity $B$ by means of the Bayes’ theorem:

$$p(B | k) \propto \text{Poi}(k | B) \pi(B)$$

A few lines of BAT instructions are sufficient to set up and solve the problem: a complete example is provided in the examples/advanced/referencecounting directory.

It is useful to recall the difference between an upper limit on the intensity that one expects to detect with a predefined probability, characterizing the sensitivity of the detection technique, and an upper bound on the intensity inferred from the actual measurement (Kashyap et al. 2010). Here we are only concerned with the latter. Appendix A provides details about the upper limits.
(we omit the proportionality constant, as the latter can be determined by imposing that the integral of \( p(B \mid k) \) be one). This equation expresses the posterior probability density function \( p(B \mid k) \) of the off-source background intensity \( B \) given the \( k \) observed counts in terms of the likelihood function (2) and of the prior density \( \pi(B) \).

If we have some prior estimate of the background flux in the off-source region, it is most convenient to represent it with a Gamma distribution (the conjugate prior of the Poisson model):

\[
\text{Ga}(x \mid S, R) = \frac{R^S}{\Gamma(S)} x^{S-1} e^{-Rx}, \quad (4)
\]

with shape parameter \( S > 0 \) and rate parameter \( R > 0 \). In this case, the posterior also belongs to the Gamma family, with a new shape and rate parameters \( S' = S + k \) and \( R' = R + 1 \), corresponding to \( k \) observed counts. For example, the prior parameters \( S, R \) can be fixed with the method of moments by imposing values for the prior expectation \( E[x] = S/R \) and variance \( V[x] = S/R^2 \).

In the absence of prior information, it is best to adopt an objective prior. The reference prior for the Poisson model coincides with Jeffreys’ prior, which is the limiting case of a Gamma function with a shape parameter \( S = 1/2 \) and a rate parameter \( R = 0 \). Hence, we use here the (improper) Jeffreys’ prior for the background contribution \( B > 0 \) and find the (normalized) density which represents the solution of Bayes’ formula (Equation (3)):

\[
p(B \mid k) = \text{Ga} \left( B \mid k + \frac{1}{2}, 1 \right). \quad (5)
\]

This is the reference posterior for the background in the off-source region, and the same solution is used in MK2014.5

Our goal is to estimate the source intensity \( s \) in the target region. We use the Bayes’ theorem to write the joint posterior density in the source region as

\[
p(s, b \mid n) \propto \text{Poi}(n \mid s + b) \pi(s) \pi(b). \quad (6)
\]

Later, we will integrate over \( b \) to find the marginal posterior density \( p(s \mid n) \). Hence, we need to find the prior \( \pi(b) \) by translating the background estimate in the control region into a background estimate in the target region. In order to do so, we use the posterior density \( p(B \mid k) \) from Equation (5) to determine the prior for the background contribution \( b \) in the source region.

With the change of variable \( b = \rho B \), we find

\[
\pi(b) = \text{Ga} \left( b \mid k + \frac{1}{2}, 1 \right). \quad (7)
\]

It is interesting to note that the expected background in the source region is \( E[b] = \rho(k + 1/2) \) with the most likely value \( \rho(k - 1/2) \) being the mode of \( \pi(b) \) when \( k \gg 1 \) (if \( k = 0 \) the prior peaks at zero). The commonly used maximum likelihood estimator \( \hat{b} = pk \) (Li & Ma 1983) is in between the peak value and the expected background. The background variance \( V[b] = \rho^2(k + 1/2) \) is also very similar (but not identical) to the commonly used value of \( \rho^2k \).

The next step is to write down the prior for the source strength \( s \). We assume no prior knowledge here, and hence adopt the reference prior calculated in DC2012. The starting point for determining \( \pi(s) \) is the marginal model

\[
P(n \mid s) = \int \text{Poi}(n \mid s + b) \pi(b),
\]

which in our case is

\[
P(n \mid s) = \left( \frac{1}{1 + \rho} \right)^{k+1/2} e^{-\rho s} f \left( s; n, k + \frac{1}{2}, 1, \frac{1}{\rho} \right). \quad (8)
\]

The polynomial

\[
f(x; n, c, d) = \sum_{m=0}^{n} \binom{c + m - 1}{m} (n - m)! (1 + d)^m \quad (9)
\]

is a function of the real variable \( x \geq 0 \) with integer parameter \( n \geq 0 \) and real parameters \( c, d > 0 \) whose properties are studied in DC2012.

From the marginal model, DC2012 finds the Fisher’s information, which in our case reads

\[
I(s) = \left( \frac{1}{1 + \rho} \right)^{k+1/2} e^{-s} \times \sum_{m=0}^{\infty} \left[ f \left( s; m, k + \frac{1}{2}, 1, \frac{1}{\rho} \right) \right]^2 - 1. \quad (10)
\]

The resulting reference prior is proportional to \( |I(s)|^{-1/2} \) and is improper. Hence, it is defined apart from a multiplicative constant. Making use of this degree of freedom, the expression proposed by DC2012 is

\[
\pi(s) = \sqrt{\frac{|I(s)|}{|I(0)|}} \quad (11)
\]

which is a monotonically decreasing function of \( s \) with a maximum at one for \( s = 0 \). In the limit of perfect prior background knowledge, \( \pi(b) = \delta(b - b_0) \), one gets Jeffreys’ prior for the offset variable \( s' = s + b_0 \):

\[
\pi(s) \rightarrow \sqrt{\frac{b_0}{s + b_0}} \equiv \pi_0(s), \quad (12)
\]

where \( b_0 = E[b] \) is a known value.

As shown in Appendix B, the limiting prior \( \pi_0(s) \) can often be used in place of the more complicated reference prior \( \pi(s) \) given in Equation (11). When this is not the case, a closed-form function provides an excellent approximation of the reference prior \( \pi(s) \) as described by Casadei (2014). This considerably simplifies the computation, although it is not used in this paper.

The marginal reference posterior for the source strength in the source region is finally

\[
p(s \mid n) \propto e^{-s} f \left( s; n, k + \frac{1}{2}, 1, \frac{1}{\rho} \right) \pi(s), \quad (13)
\]

5 The fact that the parameters assume new values well reflects the interpretation of the Bayes’ theorem as a way of updating our knowledge.

6 The constant at the numerator of Equation (14) in MK2014 is irrelevant, as an improper prior can have any scaling factor. Here, we took the latter to be one, for simplicity. What matters is that the posterior is a proper density, as is the case with Equation (5) above.

7 Incidentally, this makes it easy to compare it to the uniform prior, a very common (although mathematically ill-defined) choice with a long tradition.

8 \text{Ga}(x \mid c, d) \rightarrow \delta(x - b_0) \) in the limit \( c, d \rightarrow \infty \) while keeping \( b_0 = c/d \) constant.
obtained after having removed the inessential constant factor $(1 + \rho)^{-k-1/2}$ in front of the marginal model (Equation (8)), as the normalization of the marginal posterior is found by dividing by the integral from zero to infinity of the expression above (which is integrable for all possible values of the background shape and rate parameters).

The marginal reference posterior (Equation (13)) is the full solution of our inference problem. Very often, it will be summarized by providing only few figures of merit, like the most probable value or the posterior expectation, plus an interval enclosing the true value with some predefined posterior probability. To complement this solution, one can check whether the counts $n$ in the source region are compatible with the background-only hypothesis by comparing $n$ with the expectation $E[b] = \rho(k + (1/2))$ from a simple Poisson process in which no source is present.

The statistical significance $z$ quantifies the deviation between the observed counts and expected background in terms of the displacement from the peak of a normal distribution in units of standard deviations. An excess of counts for which $z = 3$ is commonly called “a 3σ excess” and considered a real effect (which is evidence for an additional contribution on top of the events expected from the background-only hypothesis), although more stringent requirements may be preferred, like the “5σ” excess traditionally required in high-energy physics to claim the discovery of a new particle, and required by MK2014 as well.9

A deviation from the expected background can occur in two directions: as an excess of counts when $n > E[b] = \rho(k + 1/2)$, or as a deficit when $n < E[b]$. The commonly used expression for the significance is Equation (17) of Li & Ma (1983, LM1983 hereafter). Such a formula always gives a positive value for $z$, while it is more appealing to differentiate between the excess and deficit of the observed events with respect to the expected background. In addition, strictly speaking, that formula is valid only asymptotically, although it has already been shown to behave well with moderately small values of $n$.

We compute the significance of the deviation from $E[b]$ with the inclusion of the uncertainty on the background in the source region, represented by the square root of the prior background variance $V[b]$, as described by Choudalakis & Casadei (2012, CC2012 in the following).10 A Poisson process with an uncertain parameter, whose probability density is represented by a Gamma distribution, is described by the Poisson–Gamma mixture,

$$P(n|c,d) = \int_0^{\infty} \text{Poi}(n | b) \Gamma a(n | c, d) \, db$$

$$= \frac{d^c}{\Gamma(c)} n! \left(1 + d\right)^{-c}. \quad (14)$$

In our case, we find $c, d$ with the method of moments, from $E[b] = c/d = \rho(k + 1/2)$ and $V[b] = c/d^2 = \rho^2(k + 1/2)$. As expected, the result is $c = k + 1/2$ and $d = 1/\rho$, the same as in Equation (7), but Equation (14) can also be used when there are other (e.g., systematic) contributions to the background uncertainty in the target region, by applying the method of moments with the corresponding (larger) variance.

For an excess, the probability $p$ to get a deviation that is not smaller than the observed one is given by the sum from $n$ to infinity of Equation (14). For a deficit, $p$ is given by the sum of the terms from 0 to $n$. Next, we compute the significance $z$ by imposing that the integral from $z$ to infinity (excess) or from minus infinity to $z$ (deficit) of a standard normal distribution is equal to $p$.

This definition of the $p$-value (hence of statistical significance $z$) is similar to the usual (frequentist) definition, with the exception that the Poisson–Gamma mixture (the marginal model given $H_0$) is used in place of the Poisson distribution, to which it reduces in the case of negligible background uncertainty. Hence, it represents a sort of hybrid computation of the probability that the data deviate from the expectation at least as much as in the actual observation. The result is valid for any value of $n$, even when $n = 0$, hence, it does not rely on asymptotic properties like the LM1983 significance. On the other hand, MK2014 defines the $p$-value as the probability of the model $H_0$ given the data (Equation (27) of MK2014). Hence, this significance (denoted by $S_0$) has a different meaning from the significance used here and by LM1983, despite the similarity of the numerical values.

3. RESULTS

We apply the methods described above on the same input data as MK2014 to make a detailed comparison with that solution. Table 1 shows GRB data collected by Fermi-LAT (Abdo et al. 2009) and VERITAS (Acciari et al. 2011). Such GRBs were selected by MK2014 because they have low counts (otherwise the difference with respect to asymptotic formulae is difficult to notice): at least one among $n$ and $k$ is not bigger than 15 counts.

The first step is to estimate the background $B$ in the off-source regions with the help of Equation (5). The reference posterior for $B$ only depends on the off-source counts $k$, hence it is the same for GRBs 080607 and 090418A (both with $k = 16$), and for GRBs 081024A and 090429B (both with $k = 7$). Figure 1 shows the reference posteriors for the background $B$ in the off-source regions. In addition to the counts in the off-source region, for each GRB the reference posterior mean and standard deviation of $B$ are reported, which may be useful summaries for back-of-the-envelope computations.

Next, one finds the background prior for $b = \rho B$ in the source region from Equation (7). Because the value of $\rho$ differs in each pair of GRBs with the same off-source background estimate, their background priors in the source regions are all distinct Gamma densities.

Once the reference prior from Equation (11) is computed, the final solution is provided by the (marginal) reference posterior for the source strength $s$ in the source region, Equation (13). It is worth noticing that in all cases considered here, the limiting prior $\pi_0(s)$ defined in Equation (12) works equally well. There would be no relevant change if $\pi_0(s)$ were used in place of the reference prior $\pi(s)$ defined in Equation (11), although the latter was used here.

Figure 2 shows the reference posteriors for $s$ for all GRBs listed in Table 1, where the posteriors are summarized by reporting the HPD credible intervals with 99%, 95%, 90%, and 68.3% posterior probability, the source intensity expectation ($E$), median ($M$), mode ($P$), plus variance ($V$), skewness ($S$), and excess kurtosis ($K$). Two decimal places are shown in the table, even though they do not bring any insight on the physics,
As noted by MK2014, the only clear detection is GRB 080825C, observed by Fermi-LAT (Abdo et al. 2009). The posterior HPD credible intervals are reported with 99%, 95%, 90%, and 68.3% posterior probability, together with source intensity expectation ($E$), median ($M$), mode ($P$), variance ($V$), skewness ($\Delta$), and excess kurtosis ($K$). The last two columns report the 99% upper bounds computed by MK2014 using his solution and the method by Rolke et al. (2005). GRB 080825C is the only clear detection: we obtain $s = 13.28^{+4.39}_{-2.92}$ with significance $\Delta = 6.26$, whereas MK2014 obtains $s = 13.28^{+4.39}_{-2.92}$ with significance $\Delta = 6.11$ computed by mapping the posterior probability of $H_0$ onto a Gaussian metric. Both Bayesian solutions find the same peak value for the source strength, which is only 3% weaker than the result of 13.7 units obtained by the Fermi-LAT collaboration, a difference 10 times smaller than the standard deviation computed here ($\sqrt{\Delta} = 3.95$), hence negligible. Our result is slightly more suggestive of higher source counts than MK2014, as the right asymmetry of our 68.3% credible interval is more pronounced. This implies that our source strength expectation (14.27 units) should be larger than MK2014 (where this value is not reported). However, this small difference is of little practical importance.

As noted by MK2014, the only clear detection is GRB 080825C, observed by Fermi-LAT (Abdo et al. 2009): we obtain $s = 13.28^{+4.39}_{-2.92}$ with significance $\Delta = 6.26$, whereas MK2014 obtains $s = 13.28^{+4.39}_{-2.92}$ with significance $\Delta = 6.11$ computed by mapping the posterior probability of $H_0$ onto a Gaussian metric. Both Bayesian solutions find the same peak value for the source strength, which is only 3% weaker than the result of 13.7 units obtained by the Fermi-LAT collaboration, a difference 10 times smaller than the standard deviation computed here ($\sqrt{\Delta} = 3.95$), hence negligible. Our result is slightly more suggestive of higher source counts than MK2014, as the right asymmetry of our 68.3% credible interval is more pronounced. This implies that our source strength expectation (14.27 units) should be larger than MK2014 (where this value is not reported). However, this small difference is of little practical importance.

Apart from 080825C, all other GRB data do not show any evidence for a detectable source, as the significance values reported in Figure 3 confirm. For GRBs 070521, 080310, 080330, 080604, and 081024A, the reference posterior is monotonically decreasing with maximum probability density at zero: the right edges of their HPD intervals are all upper bounds to the source strength. Although the posteriors of the other GRBs are not monotonically decreasing functions, their peaks are so close to zero that one has, in practice, upper bounds in these cases. The significance $\Delta$ of the deviation from the background-

![On/off measurements: background posteriors in the off-region](image)
On/off measurements: reference posteriors

Figure 2. Marginal posteriors for GRB data. The GRB name is followed by the counts in the off-source and source regions, and by the ratio of the background fluxes in these regions.

Figure 3. Observed counts (full dots) in the source region of each GRB are compared to the expected background and its uncertainty (red histogram with yellow bands representing fluctuations of one standard deviation in both directions). The plot at the bottom shows the significance of the deviation in each bin, computed with (cyan histogram) and without (red histogram) accounting for the uncertainty on the background. The black stars correspond to the significance computed according to Li & Ma (1983). The inset in the top left corner shows the pull distribution computed with background uncertainties, with a Gaussian fit (red line). A standard normal distribution (centered at zero with unit standard deviation) is also shown for comparison (dashed black line).
only expectation is 0.83 for 070419A, 0.93 for 070612B, 1.14 for 080607, 0.44 for 090418A, 0.87 for 090429B, and 0.33 for 0900515: clearly, $H_0$ cannot be rejected. In addition, their posterior 99% HPD intervals, when keeping a single decimal place, all start at zero. In conclusion, there is no clear evidence for some additional contribution in addition to the photon counts expected from the background alone.

It is interesting to look at the distribution of the significance $z$ of the deviations between observed counts and expected background, as this is a quick way of checking that the background-only hypothesis is valid for the entire set of measurements. By definition, under repeated sampling, the model $H_0$ will give values of $z$ that are normally distributed with center at zero and unit standard deviation.

A plot produced with all GRBs listed in Table 1 apart from GRB 080825C is shown in Figure 3. The values computed according to LM1983 are also shown for comparison. As mentioned above, they are always positive, which makes a difference when a deficit of events is observed. However, this case is less interesting than the observation of an excess of counts, for which the agreement is acceptable. The formula by LM1983 overestimates the significance when the latter is small (by 15% when $z \approx 1$, increasing when $z \to 0$ but decreasing when $z$ increases), which is not a big problem in practice. When the significance is high, it gives very similar results to our approach. For example, the detection of GRB 080825C by Fermi-LAT has a significance $z = 6.26$. The formula by LM1983 gives a value of 6.36, whereas MK2014 obtains 6.11; hence, the three methods agree that the result is very significant. The good agreement (1.6%) between the standard formula of LM1983 and our method is connected to their very similar background uncertainty estimates ($\rho \sqrt{k}$ for the standard method of LM1983 and $\rho \sqrt{k + 1/2}$ in our case). However, this uncertainty is purely statistical: if any additional contribution exists, then the formula by LM1983 is not able to account for it and the method by CC2012 should be used instead, as it gives a more general (the uncertainty is not assumed but is an input parameter). Another possibility is to compute the posterior probability of $H_0$ given the data and map it onto a Gaussian metric. This is what is done by MK2014, whose result is a bit smaller than, but still appreciably similar, to ours despite the difference in meaning.

By collecting all significance values, one creates the “pull distribution,” which, in the case of purely stochastic fluctuations, should follow a standard normal distribution (as it is the case when simulating a large number of pseudo-experiments). The inset at the top left corner of Figure 3 shows that the GRB measurements—with the exception of GRB 080825C, detected with more than 6σ statistical significance and not shown there—do not show any strong deviation from that distribution (dashed black curve), suggesting that $H_0$ does indeed hold for the entire sample. Although a Gaussian fit (red curve) actually confirms the preference for positive fluctuations which is visible in the bottom plot, it also shows that the results are more tightly clustered than expected. The shift of the barycenter is not significant, being about twice as large as the uncertainty on its position, confirming that the null hypothesis of pure background counts well describes the set of observations.

**4. SUMMARY AND DISCUSSION**

We have illustrated how the objective Bayesian solution to the inference problem for the Pois($s + b$) model can be applied to the on/off problem. Our solution is the marginal reference posterior probability density for the source strength $s$, given the measured counts $n$ in the source region, and the auxiliary measurement of background-only counts summarized by the off-source counts $k$ and the ratio $\rho$ between the background fluxes in the two regions. Based on the reference prior proposed by DC2012, this solution appears to be more conservative (higher upper bounds) when there is no clear detection of additional photons with respect to the background-only expectation in the source region, compared to the posterior proposed recently by MK2014 and to the frequentist method based on asymptotic properties of the profile likelihood test statistic by Rolke et al. (2005).

The approach by MK2014 also aims to provide an objective Bayesian result. Its most delicate point is the choice of the prior for the source region. The choice of Jeffreys’ prior in the $(s, b)$ space may result in problems that can be avoided if one considers the marginal model instead (obtained by integrating over the entire range of $b$, weighted by its prior). On the other hand, the marginal model is 1D and the corresponding reference prior is fixed. Reference priors, when available, are the recommended objective priors in the statistics literature, as they possess a number of desirable properties and are well “calibrated” from the frequentist point of view. In 1D problems, they usually coincide with Jeffreys’ priors, but this is not true in multidimensional problems.

Another potential source of problems is the hypothesis testing step in the method proposed by MK2014, as the improper priors used in both the off-source and source regions are not identical. MK2014 proposes an ad-hoc procedure to overcome this problem, which is based on the questionable assumption that measuring $k = 0$ and $n = 0$ does not change our degree of belief about the two alternative hypotheses. We propose to avoid the problem by omitting the comparison between $H_0$ and $H_1$. One can compute the statistical significance of the deviation between the observed counts and the background-only expectation on the basis of $H_0$ alone, as a conservative check that the probability of claiming a false detection is low enough. This complements the estimate of the source strength $s$ provided by the marginal reference posterior.

In case the hypothesis test is considered a fundamental step (which does not seem to be the case here), it might be worth noticing that Bernardo (2011) recommends basing our decision on the comparison between the reference posteriors of each model by means of an invariant information-based loss function (the “intrinsic discrepancy”). This promises to be the best way to achieve an universally applicable procedure that guarantees objective decisions. Unfortunately, this approach to the Bayesian hypothesis testing is not yet widespread in the scientific community, where most continue to look at the Bayes factors (ill defined in our case; see also the discussion in Bayarri et al. 2008).

The HPD intervals chosen by MK2014 cover 99% posterior probability. In other words, they are the shortest 99% credible intervals on the source strength $s$, given the measurements in the on/off regions. Even though they are not invariant under reparameterization, nor it is the most probable value, there is little or no discussion in the astrophysics community about the best choice for the parameter of interest: only $s$ is considered. Hence, here we also show the posterior mode with 99% credible.
HPD intervals, although one might consider more complicated ways of summarizing the result in an invariant way. The interested reader can find details about the “reference posterior intrinsic loss” function, which allows one to find the “intrinsic estimator” of the parameter of interest, together with “intrinsic 99% credible regions” for s in Bernardo (2005b, 2007).

Numerical comparison with the results obtained by MK2014 shows that the two methods are in decent agreement, although the use of Jeffreys’ prior leads to narrower posterior densities than the marginal reference posterior. This means that the upper bounds obtained by MK2014 are always tighter than those obtained here.

The only clear case of unambiguous GRB detection by is 080825C by Fermi-LAT (Abdo et al. 2009). For this GRB, MK2014 obtains \( s = 13.28^{+4.16}_{-3.40} \), whereas the result obtained here is \( s = 13.28^{+4.89}_{-2.92} \), even though the posterior peak is at the same position as in MK2014, our result is slightly more suggestive of a higher intensity, although it well overlaps with MK2014 within the uncertainties.

With all GRB data considered here, the very simple approximate reference prior \( \pi_0 \propto (s + E[b])^{-1/2} \) provides practically the same result as the (more complicated) reference prior. This is always true when the rate parameter describing the Gamma prior for \( b \) is large enough (in practice, it is sufficient to be larger than a few units), or when the shape parameter is large. In our case, the shape parameter of the background prior in the source region is \( S = k + 1/2 \), while the rate parameter is \( R = 1/\rho \). The approximate reference prior \( \pi_0(s) \) differs less than 1% from the reference prior when \( R > 4 \) or \( S > 40 \), plus a portion of the parameters space that does not satisfy any of these requirements (Appendix B).

The first condition is fulfilled by all GRBs in Table 1, apart from GRBs 080310 (\( S = 23.5, R = 7.8 \)), 081024A (\( S = 7.5, R = 7.04 \)), and 090515 (\( S = 24.5, R = 7.9 \)). However, these three GRBs have shape and rate parameters that fall in regions of the parameters space in which \( \pi_0(s) \) differs very little from the reference prior. This means that the approximate marginal reference posterior

\[
p_0(s \mid n) \propto Ga(s + E[b], n + \frac{1}{2}, 1)
\]

(Casadei 2014) could have been used in place of the marginal reference posterior (Equation (13)) with considerable simplification.

Finally, we have noticed that the significance \( z \) obtained with the standard asymptotic formula by Li & Ma (1983), when it is not too small, agrees well with the values calculated as suggested by CC2012, which are correct for any value of \( n \) (including the case \( n = 0 \)). By treating the background uncertainty as a free parameter, the more recent definition of \( z \) is more general than the standard formula and should be used whenever additional sources of uncertainty exist beyond the pure statistical fluctuations connected with the finite number of photons collected while pointing off-source.

5. CONCLUSION

The Bayesian approach proposed by MK2014 aims to provide an objective solution for the on/off measurement. Two possible sources of problems with this approach are connected with the use of Jeffreys’ prior in the bidimensional space \((s, b)\) of source and background intensities in the target region, and with the hypothesis test performed with improper priors.

Instead of using the Jeffreys’ prior, a different approach exists which provides an objective Bayesian solution as the marginal reference posterior for \( s \) obtained after reducing the problem to a 1D problem. This is done by integrating over the background intensity \( b \) in the target region, for which an informative prior exists. The reference prior for this 1D problem is known, which ensures that the solution possesses all the important features of the reference posteriors, including invariance under reparameterization and good frequentist properties. Thus, it is recommended to compute the reference posterior, possibly with the help of some approximation of the reference prior when the difference is small enough (which can be judged by checking the background parameters, as explained in Appendix B).

A formal two-step procedure, in which the comparison between background-only \( H_0 \) and source+background \( H_1 \) hypotheses is performed before estimating the source intensity in the target region, is not strictly necessary. A simpler approach is to estimate \( s \) directly, and to only check the significance of the deviation from the expectation from pure background counts in case the posterior for \( s \) suggests a non-negligible source intensity. As the significance is computed with \( H_0 \), one avoids the complications arising when comparing two nested hypotheses whose parameters have improper priors.

The significance \( S_0 \) defined by MK2014 is Bayesian in the sense that it corresponds to the probability that \( H_0 \) is true given the data. Thus, it is conceptually different from the usual definition of statistical significance \( z \) in terms of the probability of obtaining data with deviations at least as large as the observed one, given \( H_0 \). The latter is a frequentist concept, which is typically computed with the help of asymptotic formulae as in LM1983. Here we have adopted a hybrid approach that computes the probability of deviations given \( H_0 \) with the marginal model obtained after integrating over the background in the target region, given by a Poisson–Gamma mixture. This approach is superior to LM1983 for three reasons: it does not rely on asymptotics, it differentiates between excess and deficit, and it allows us to include any uncertainty on the background (while LM1983 only account for the statistical uncertainty from the auxiliary measurement). As expected, when the significance is high enough, the three definitions provide results that are numerically very similar. However, our approach behaves differently for mildly significant results, for which LM1983 tend to overestimate \( z \), and for all other cases (including deficits, which are assigned positive \( z \) values by LM1983).

### APPENDIX A

#### UPPER LIMITS ON DETECTABLE SOURCES

Kashyap et al. (2010) emphasize the important difference between the upper bound on the source intensity \( s \) in the target region, given the observed number \( n \) of counts and the prior knowledge about the background intensity \( b \) in the same region, and the upper limit on the detectable sources with the chosen detection technique. Upper bounds (typically at 90%, 95%, or 99% confidence level) are a way of summarizing the result of a measurement when there is no significant evidence of an excess with respect to the counts expected from the background alone. On the other hand, upper limits are connected to the detection technique and can/should be computed before looking at the experimental outcome. In other words, upper limits are connected to the sensitivity of the chosen technique and do not depend on the counts \( n \) in the source region. However,
they depend on the estimated background in the target region, which, in our case, means that they depend on the counts \( k \) in the off-source region, on the prior knowledge about the background rate \( B \) in this control region, and on the known ratio \( \rho = b/B \) between the background rates in both regions. In this paper, we only addressed upper bounds on the measured source intensity. Here we provide an algorithm to compute the upper limit according to Kashyap et al. (2010), whose approach is generally valid for any detection technique, for the case of on/off measurements.

The notation by Kashyap et al. (2010) is as follows. The source and background rates in the target region are \( \lambda_s, \lambda_B \), respectively, with corresponding exposure times \( \tau_s, \tau_B \). The ratio between the two regions is \( r \), and one performs two measurements whose results are the counts \( n_S \) and \( n_B \) in the source and off-source regions, with

\[
\begin{align*}
n_B(\lambda_B, \tau_B, r) & \sim \text{Poi}(r \tau_B \lambda_B) \\
n_S(\lambda_s, \lambda_B, \tau_s) & \sim \text{Poi}(\tau_s (\lambda_s + \lambda_B)).
\end{align*}
\]

The correspondence to our notation is as follows:

\[
\begin{align*}
n & = n_S, \quad s = \tau_s \lambda_s, \quad B = r \tau_B \lambda_B, \\
k & = n_B, \quad b = \tau_s \lambda_B, \quad \rho = \tau_s/(r \tau_B).
\end{align*}
\]

In addition to (1) the background prior \( \pi(b) \) in the source region, an upper limit also depends on (2) the “size of the test” \( \alpha \), which is the predefined maximum tolerable probability of false detection (or Type I error rate), and on (3) the predefined minimum “power of the test” \( \beta \), which is connected with Type II errors (\( \beta = 1 - \text{Type II error rate} \)) occurring when one makes a false exclusion (i.e., does not recognize that a source is present in the target region). Common values for \( \alpha \) are 0.05, 0.003 (a “3σ” threshold), and 0.001, whereas a “5σ” threshold would correspond to \( 5.7 \times 10^{-7} \). In addition, one typically sets \( \beta = 0.5 \) or \( \beta = 0.9 \) to require at least 50% or 90% detection probability when speaking about the expected sensitivity of the measurement process for a given source strength.

For the on/off measurement, we choose \( n \) as the test statistic and fix a threshold \( n_{\text{min}} \geq 0 \) such that if \( n > n_{\text{min}} \) we claim that a source has been detected, and if \( n \leq n_{\text{min}} \) we declare that the measurement is compatible with the background-only hypothesis. The threshold \( n_{\text{min}} \) is chosen such that the probability of false detection (i.e., of wrongly rejecting the null hypothesis \( H_0 \)) does not exceed \( \alpha \):

\[
\alpha \geq P(n > n_{\text{min}} | s = 0) = 1 - \sum_{n=0}^{n_{\text{min}}} \frac{\text{Poi}(n|b)\pi(b)}{db}
\]

\[
= 1 - \frac{1}{(1 + \rho)^{k+1/2} \Gamma(k + 1/2)} \sum_{n=0}^{n_{\text{min}}} \frac{\Gamma(n + k + 1/2)}{n! (1 + 1/\rho)^n}
\]

\[
(A3)
\]

where we used the Poisson–Gamma mixture from Equation (14) with \( c = k+1/2 \) and \( d = 1/\alpha \), i.e., we used the background prior \( \pi(b) \) from Equation (7). Because the threshold \( n_{\text{min}} \) changes at discrete steps, the actual Type I error rate will typically be smaller than \( \alpha \).

The expected detection probability when \( s > 0 \) is

\[
\beta(s|\alpha) = P(n > n_{\text{min}} | s = 1) = 1 - \sum_{n=0}^{n_{\text{min}}} \frac{\text{Poi}(n|s+b)\pi(b)}{db}
\]

\[
= 1 - \frac{e^{-s}}{(1 + \rho)^{k+1/2} \Gamma(k + 1/2)} \sum_{n=0}^{n_{\text{min}}} \frac{\Gamma(n + k + 1/2)}{n! (1 + 1/\rho)^n} f(s; n, k, 1/2, 1/\rho)
\]

\[
(A4)
\]

obtained after inserting the polynomial (Equation (9)) into the marginal model (Equation (8)). For \( s \to 0 \), Equation (A4) gives the last expression of (A3) because only the term with \( m = n \) survives in the second sum. Thus, \( \beta(0|\alpha) \leq \alpha \), which means that one must find a compromise while aiming at a low probability of claiming a fake source and a high probability to detect a true but faint source.

Now that all ingredients are available, we can formulate the algorithm for computing the upper limit. Once the size and power of the test have been chosen (say \( \alpha = 0.003 \) and \( \beta = 0.5 \)), one has to estimate the background in the target region where the prior \( \pi(b) \) shall be chosen as a Gamma density whose parameters are determined by the method of moments, starting from the background expectation and variance. In the on/off measurement, we use the control region to find the background in the target region, given by Equation (7). Next, we define a threshold for detecting the source in terms of the minimum number \( n_{\text{min}} \) of counts which ensures that our Type I error rate does not exceed \( \alpha \), following the inequality (Equation (A3)). Finally, one takes as the upper limit the smallest value of \( s \) for which \( \beta(s|\alpha) \geq 0.5 \), where \( \beta(s|\alpha) \) is computed in Equation (A4).

### APPENDIX B

#### APPROXIMATE FORMS FOR THE REFERENCE PRIOR

Here we report useful approximations to the reference prior \( \pi(s) \) defined in Equation (11) from Casadei (2014). An animation comparing the reference prior with these approximations and with the flat prior is available at https://www.youtube.com/watch?v=qUnRiwHcC, clearly showing when different approximations should be used.

The limiting form \( \pi_0(s) \) of the reference prior when there is certain knowledge of the background in the source region is given by Equation (12). The limit of perfect knowledge is approached by increasing values of the shape parameter, as the relative uncertainty on the background in the source region is \( \sqrt{\nu(b)/E(b)} = 1/\sqrt{5} \). However, it turns out that even at small values of \( s \) there are cases in which \( \pi_0(s) \) provides a very good approximation to \( \pi(s) \).

In order to quantify the deviation from \( \pi(s) \), their relative rms difference has been computed on the signal range \( 0 \leq s \leq 70 \), by dividing the distance

\[
d_2 = \left( \int_0^{70} [\pi_0(s) - \pi(s)]^2 ds \right)^{1/2}
\]

by the integral of the reference prior over the same range.

For most practical purposes, a relative rms difference below 1% is acceptable, as this is the order of magnitude of the maximum change in the posterior in the limit of very few or zero observed counts. For increasing \( n \), the changes of the posterior become smaller and smaller. Figure 4 shows that the \( \pi_0(s) \) is satisfactory (differing from \( \pi(s) \) by less than 1%) when the shape parameter is larger than 40 or the rate parameter is larger than 4, and in some cases even for lower values.
Relative RMS difference between $\pi(s)$ and $\pi_0(s)$ as a function of the shape and rate parameters of the background prior in the source region (Casadei 2014).

It should be emphasized that the threshold at 1% chosen here is arbitrary and quite conservative. In most applications, larger deviations can be acceptable, as the posteriors will quickly become indistinguishable for increasing number $n$ of observed counts. In addition, the common practice is to summarize the posterior by providing one value (e.g., the expectation or the mode) and some estimate of its uncertainty (e.g., the shortest interval covering 68.3% posterior probability), by rounding the values to the minimum meaningful number of digits. Often, this summary is quite robust compared to relative rms differences of several percent.

REFERENCES

Abdo, A. A., Ackermann, M., Asano, K., et al. 2009, ApJ, 707, 580
Acciari, V. A., Aliu, E., Arlen, T., et al. 2011, ApJ, 743, 62
Antcheva, I., Ballintijn, M., Bellenot, B., et al. 2009, CoPhC, 180, 2499
Bayarri, M. J., Berger, J. O., & Datta, G. S. 2008, IMS Collect., 3, 105
Berger, J. O., Bernardo, J. M., & Sun, D. 2013, in Overall Objective Priors, Tech. Rep., (Durham, NC: Duke Univ.), http://www.uv.es/~bernardo/BBS2013.pdf
Bernardo, J. M. 2005a, in Reference Analysis, Handbook of Statistics 25, ed. D. K. Dey & C. R. Rao (Amsterdam: Elsevier), 17
Bernardo, J. M. 2005b, Test, 14, 317
Bernardo, J. M. 2007, Sort, 31, 3
Bernardo, J. M. 2011, in Bayesian Statistics 9, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger et al. (Oxford: Oxford Univ. Press), 1
Caldwell, A., Kollar, D., & Kröninger, K. 2009, CoPhC, 180, 2197
Casadei, D. 2012, JINST, 7, P01012
Casadei, D. 2014, JINST, 9, T10006
Choudalakis, G., & Casadei, D. 2012, EPJP, 127, 25
Kashyap, V. L., van Dyk, D. A., Connors, A., et al. 2010, ApJ, 719, 900
Knoetig, M. L. 2014, ApJ, 790, 106
Li, T.-P., & Ma, Y.-Q. 1983, ApJ, 272, 317
Rolke, W. A., López, A. M., & Conrad, J. 2005, NIMPA, 551, 493