On Bayesian Statistics in Astronomical Investigation
Source Detection with Low Particle Counts

Mike West


1Institute of Statistics & Decision Sciences, Duke University, Durham, North Carolina 27706, United States of America
1. On Bayesian Statistics in Astronomical Investigation

1.1 On The Promise of Bayesian Inference for Astrophysics, by Thomas J. Loredo

It is no surprise (to a Bayesian) that inferential problems encountered using traditional statistical methods have a tendency to evaporate when the issues are addressed from formal Bayesian perspectives. This paper is largely concerned with cameo examples designed to communicate this precept to astronomers interested and involved in statistical work in their investigations. I like the paper, and have a great deal of respect for the entrepreneurial efforts of the author and his co-authors to shift the focus of statistical analysis in the field toward the Bayesian paradigm. That these efforts have been rewarded is clear from browsing some of the referenced articles, notably the works of Loredo and Lamb (referenced in text). Here we find advanced physical and statistical models subjected to formal and rigorous Bayesian analysis, some requiring high dimensional numerical integrations performed via Monte Carlo, that yield inferences in terms of posterior distributions for parameters of interest that clearly and unambiguously address detailed and substantive scientific issues. A reading of these works provides a clear picture of investigators led to adherence to the Bayesian paradigm on pragmatic and empirical grounds — the Bayesian approach gets the (right) job done where all others fail. The current paper, by comparison, focuses on elementary examples to clearly identify difficulties and inconsistencies, both conceptual and technical, inherent in traditional inferential paradigms, and zealously argues for the Bayesian approach as the preferred alternative. I agree with much of what Loredo has written here.

The Scientific Organising Committee exhorted invited discussants to delve deeply into some aspects of the material presented. If I have criticism of Loredo’s work here, it is whole-heartedly constructive and intended more as an addendum to the text of the paper. From my own Bayesian perspective, some basic comments in reading the paper are that (a) the Bayesian analysis of a specific problem does not exist, (b) priors always matter, but ever more so in problems when data are scarce, and (c) don’t neglect the pre- and post-data predictive opportunities inherent in the Bayesian paradigm. Let me elaborate on these general issues in the context of Section 5 of the paper, dealing with Poisson models for particle detection, specifically with respect to problems of low particle counts and source detection/estimation.

Consider the basic framework of Poisson counts arising from a possible source, such as in the ‘On/Off’ measurement paradigm of Section 5.2. In Loredo’s notation, \( b \) is the background intensity and \( s \) represents the intensity of the source, so that observed counts \( n \) are assumed to be Poisson distributed with mean \( b + s \). Assume here that \( b \) is known; the discussion below is affected only technically, and not in conceptual essence, by
admitting uncertainty in terms of a prior distribution for \((b|I)\). Also, expecting low counts implies \(b\) will be rather small (see also the paper by John Nousek, this volume, in discussion of low count radiation from SN1987A). Using a prior \(p(s|bI)\) that is uniform over a ‘large’ range (and does not depend on \(b\)), Loredo proceeds to summary inferences based on the posterior \(p(s|nbI)\) in his equation (5.13). Throughout the paper, such uniform priors are adopted as a routine on the basis of representing suitable forms of ‘ignorance’ about the quantity concerned. If any area of Bayesian inference has received too much attention during the last couple of decades it is surely the search for unique and ‘objective’ representation of ignorance – see [4] for a recent and partial review of the field. The maximum entropy school has been influential in physical sciences, as referenced by Loredo, and particularly predominant in expounding the view that a single prior may be found, in any given situation, to represent vagueness in the sense of maximum entropy subject to certain ‘plausible’ assumptions that typically stand for little more than mathematical convenience in determining a unique solution in the resulting MaxEnt framework. There is nothing unique, objective or otherwise scientifically persuasive about uniform priors for location parameters, or any of the plethora of vague, reference or indifference priors that abound. In investigations which admit an ‘objective’ (defined simply as consensus of informed observers) data model as here (ie. \(p(n|sI)\)), analysis should necessarily involve study of sensitivity to qualitative and quantitative aspects of the prior, including assessments of pre-data predictive validity of the \{data model}:\{prior\} combination, and post-data determination of the mapping from prior to posterior for ranges of scientifically plausible priors.

The issue of pre-data validity is addressed through the implied (prior) predictive distribution for the data, here \(p(n|bI) = \int p(n|sI)p(s|bI)ds\). When \(n\) is observed, the value of this density function provides the normalising constant in Bayes’ theorem \((C^{-1}\) in Loredo’s equation (5.6)). Prior to the data, however, this distribution describes the investigator’s view of experimental outcome. A uniform prior over a very large range translates essentially into a similar (though discrete) uniform \(p(n|bI)\), which most observers would be quite concerned about as a plausible and scientifically valid representation of expectations. The issue is particularly acute in problems of low counts and source detection when \(s\) (when non-zero) will be tend to be small — reasonable priors for \(s\), and thus predictions about \(n\), should surely reflect this. Competing ‘reference’ priors (and there are many – [4]), lead to posteriors that can differ markedly with low counts \(s\), though all such priors claim some form of ‘vagueness’ and ‘uniformity’ (on some scale). Scientific investigation must involve careful and thorough consideration of initial information, modes of incorporation of such information in summary inferences, and exploration of sensitivity to prior assumptions (which includes model and data assumptions and well as priors for model parameters – and sometimes the distinction is unclear and even irrelevant.
exp(− see [11]).

As a simple example, consider the class of exponential priors \( p(s|mbI) = \exp(-s/m)/m \), where the prior mean \( E(s|mbI) = m \) now appears in the conditioning. The predictive, or forecast, mean is now the expectation of low counts, and with \( b \) already small, scientifically plausible ranges of \( m \) should not include large values. For any chosen \( m \), the predictive density \( p(n|mbI) \) and the posterior \( p(s|mbI) \) are easily derived. Since concern lies with low count data, consider possible outcomes \( n = 0 \) and \( n = 1 \), and, for illustration, consider cameo examples based on \( b = 0.73 \) (from the SN1987A study in the Nousek paper, just referenced, section 1.2). We have predictive probabilities \( P(n = 0|mbI) = \exp(-b)/(1 + m) \) and \( P(n = 1|mbI) = \exp(-b)(m + b + mb)/(1 + m)^2 \). Each is appreciable for low integer values of \( m \), the chance of \( n = 0 \) falling below (that magic number) 0.05 at \( m = 9 \), and that for \( n = 1 \) at about \( m = 16 \). Priors with \( m \) in single digits provide reasonable degrees of support for conservatively large values of \( s \) while being predictively consistent with the expectations of low counts. Rejecting analysis on any single value of \( m \), we might add a prior over \( m \) or study sensitivity of inferences as \( m \) varies over an \( a \ priori \) plausible range. Consider, for instance, the key issue of source existence. Here we have a clear cut case for more traditional ‘testing’ of an hypothesis — we are interested in summarising experimental evidence for, or against, the hypothesis of no source, \( s = 0 \), with possible alternatives \( s > 0 \). Conditional on the model so far specified, the Bayesian summary of evidence is the Bayes’ factor (likelihood ratio, weight of evidence) \( B_n = p(n|(s = 0)I)/p(n|(s > 0)I) \). Given an initial probability \( \pi = P(s = 0|I) \) for the hypothesis, the Bayes’ factor maps \( \pi \) to \( \pi_n = P(s = 0|nbI) \) via Bayes’ theorem, viz. \( \pi_n = \pi B_n/(1 - \pi + \pi B_n) \).

In odds form, the final, data-based odds on \( s > 0 \) versus \( s > 0 \) are determined via \( \pi_n/(1 - \pi_n) = B_n \pi/(1 - \pi) \), or just the prior odds multiplied by the Bayes’ factor. Whatever the value of \( \pi \), \( B_n \) determines the increase/decrease of evidence, on the log-odds scale, due to the data. In our framework, and for any prior \( p(s|(s > 0)I) \) (not necessarily exponential), the Bayes’ factor may be evaluated via

\[
B_n^{-1} = \int_0^\infty (1 + s/b)^n e^{-s} p(s|(s > 0)I) ds
\]

for any observed number of counts \( n \). It is trivially deduced that

\[
B_n^{-1} \leq \max_{s>0}(1 + s/b)^ne^{-s} = (1 + s_n/b)^ne^{-s_n},
\]

where \( s_n = \max\{0, n - b\} \), just the maximum likelihood estimate of \( s \) given \( n \). For any datum \( n \), this gives a gross lower bound on \( B_n \) determining a first shot at a lower bound on \( \pi_n \), and hence a limit on the possible evidence in favour of source existence. Note that this lower bound is the maximised likelihood ratio traditionally used in non-Bayesian, likelihood based testing.
The fact that this is an absolute bound indicates that any Bayesian solution would be more conservative; the Bayes factor based on any prior for $s$ would support source existence less than the extreme produced by the likelihood approximation — in maximising, likelihoodists go overboard in the direction of the existence hypothesis. For possible priors in any specified class we might now explore bounds on $B_n$, paralleling such ideas in normal theory models ([3] section 4.3). Note first that $B_0$ does not depend on the background rate $b$ (this is obviously true for any prior $p(s|s > 0)I$) — so long as it does not depend on $b$, as assumed throughout) and always exceeds unity, indicating $n = 0$ always provides evidence against source existence. Under an exponential prior with mean $m$, for example, $B_0 = 1 + m$ and $B_1 = (1 + m)^2/(1 + m + m/b)$. So $B_0$ increases linearly as a function of the discrepancy between the prior: model predictions and the datum as measured by $|m| = |0 - m| = |n - m| = \text{observed-expected}$. For $n = 1$ counts, $B_1$ depends on $b$; if $b > 1$ then $B_1$ is an increasing function of $m$ and exceeds unity for all $m$, just like $B_0$. For background rates $b < 1$, $B_1 \geq 4b/(1 + b)^2$ taking this minimum value at $m = (1 - b)/(1 + b)$; $b$ must be less than 0.1 for this lower bound on $B_1$ to fall below 1/3, indicating a severe limit on the evidence in favour of source existence from observing just one count. Similar calculations may be performed for $n > 1$, as in my comments on John Nousek’s paper (this volume), where comparison is also made with the traditional significance testing approach — as is common in normal theory models, it is shown there that, by comparison with these bounds on Bayesian measures, the corresponding non-Bayesian significance levels heavily overstate the evidence for source existence.

This type of pre-data analysis, intended to provide insight into, and feedback on, the prior predictive validity of model:prior combinations, is clearly not specific to any model or prior form. If the exponential is replaced by Laredo’s uniform prior $p(s|abI) = a^{-1}$ for $0 < s < a$, (note the inclusion of the upper bound $a$ in the conditioning), similar arguments follow. It seems reasonable to me that a randomly chosen astronomer might have such a prior, at least approximately, but $a$ would have to be allowed to vary across investigators and to bear heavily on smaller values. What results here, for a Bayesian, is an elaborated model in which $p(s|bI)$ is drawn from a class by mixing over $a$,

$$p(s|bI) = \int_0^\infty p(s|abI) dF(a) = \int_s^\infty a^{-1} dF(a) \quad (3)$$

over $s > 0$ and where $F(a)$ is some prior distribution for $a > 0$. We can arrive at the same result from a quite different perspective (and one that will be, perhaps, more immediately acceptable to our non-Bayesian friends) as follows: suppose plausible priors for $s$ are deemed those that have decreasing density function — all such priors. I might suppose that my prior lies somewhere in this class. This is a very wide class of priors that qualitatively support smaller values of $s$ a priori, but that includes priors with close to
full support for values as large as you like — a comprehensive, ‘reference’ class of priors. Elementary probability theory (eg. [13] section 9) may be used to show that every prior in this class has a representation (1) for some function $F(a)$. At this point, we might be tempted into specifying $F(a)$, but this just translates the issue of choosing a prior for $s$ to that for $a$. For some inferences, we might alternatively consider sensitivity to priors in the class, or look at ranges of inferences, and bounds on inferences, implied by the class, following the so-called ‘robust’ Bayesian approach ([3] chapter 4). Further issues, such as ranges of possible inferences for other quantities of interest, may be developed in this, and similar, frameworks — space constraints here preclude further details, though this study of bounds on Bayes factors for the source existence hypothesis is continued, with numerical examples, in discussion of the paper by John Nousek, in this volume.

I share much of Loredo’s optimism, reflected in his title, for the future of statistics in astronomical and physical sciences. The tremendous payoffs of the Bayesian paradigm do not come for free, however. The study of impact of priors and robustness touched on above reflects the concern that careful and thorough consideration be given to prior specification on a case by case basis, and that we move (far) away from ‘automated’ analyses using arbitrary ‘vague’ priors. On possible future directions, it is clear that Bayesian developments during recent years have much to offer — I would identify prior modelling developments in *hierarchical models* ([3] section 4.6, [1], [9], [8]) as particularly noteworthy. Applications of such models have grown tremendously in biomedical and social sciences, but this has yet to be paralleled in physical sciences. Investigations involving repeat experimentation on similar, related systems provides the archetype logical structure for hierarchical modelling. In recent years, a major impediment to growth of application of Bayesian methods — simply that of performing the numerical integrations required to evaluate posterior quantities (eg. Loredo’s equation (4.3)) — has been dramatically alleviated with the development of, notably, Monte Carlo techniques of analysis, and some insight into the path of this ‘revolution’ may be gleaned from various of the papers in [5]. Issues of censored data analysis, density estimation and clustering, for example, of direct interest in astronomical investigations as evidenced by the activities at this conference, may now be addressed in computational Bayesian treatments of models otherwise inaccessible using traditional methods ([6], [7], [10]). There are clear opportunities for exploitation of these (and other) developments by astronomical investigators, and I look forward to Loredo’s predicted ‘Bayesian revolution .. in astrophysics’.

Acknowledgements: This work was supported in part by the National Science Foundation under grant DMS-8903842.
1. On Bayesian Statistics in Astronomical Investigation

1.2 On Source Existence and Parameter Fitting when Few Counts are Available, by John A. Nousek

This paper is concerned primarily with failure of traditional statistical methods in problems of inference in astronomical investigations characterised by the paucity of directly relevant experimental data. The author discusses issues concerning the suitability of statistical ‘procedures’ derived on the basis of gaussianity assumptions for the data, and reviews Bayesian and non-Bayesian attempts to address these failings in specific astronomical investigations.

Sections 1.2 and 1.3 concern inference about (possible) source rates in the presence of (possibly uncertain) background noise, moving onto inference about source existence in section 1.4. Much of my discussion of the paper of Loredo in this volume concerns just these issues, keyed out for Bayesian treatment in Loredo’s work. There I was concerned with issues of model and prior specification, while wholeheartedly endorsing the Bayesian view(s) of the problems advocated by Loredo; Nousek refers to similar Bayesian treatments [12], at least for the issues of inference about low Poisson count rates and the establishment of ‘confidence limits’ for such rates using formal Bayesian posterior distributions. Much of my discussion of Loredo’s paper concerning choices of priors and sensitivity analysis applies here.

On the issue of source existence, I am concerned that the approaches through confidence limits and approximate goodness of fit test described by Nousek obscure the issues. Source detection (as opposed to estimation) with low counts is sharply illuminated in the simplest paradigm of Loredo’s section 5. Adapting to Loredo’s notation, we are to observe \( n \) Poisson counts (Nousek’s \( N \)) arising from a possible radiating source with background rate \( b \) (Nousek’s \( B \)), and hypothesise a possible additional source of Poisson rate \( s \) (\( S \) of the paper [12] which rather thoroughly develops and illustrates the issues discussed in sections 1.2 and 1.3 of the current paper). We are interested in the issue of source existence: ie. \( s > 0 \)?

With respect to this simple Poisson setup, my discussion of Loredo’s paper provides the formal Bayesian basis for assessment of the hypothesis of source existence. Following Loredo, let \( I \) stand for relevant information assumed prior to experimentation, and write \( p(n|sbI) \) for the Poisson density for data \( n \) with a combined rate \( s + b \), so that \( p(n|(s = 0)bI) \) is the density if there is no source. Given any prior \( p(s|bI) \) for \( s \) under the source existence hypothesis, write \( p(n|(s > 0)bI) = \int p(n|sbI)p(s|bI)ds \) as the (prior) predictive density for the data under the general hypothesis of source existence, \( s > 0 \). The weight of experimental evidence against the existence hypothesis is measured by the Bayes’ factor \( B_n = p(n|(s = 0)bI)/p(n|(s > 0)bI) \) — just the likelihood ratio comparing the ‘null’ hypothesis of no source,
1. On Bayesian Statistics in Astronomical Investigation

$s = 0$, with the general alternative $s > 0$. Note the inclusion if the background rate $b$ in the conditioning of densities here — the discussion below is affected only technically, not conceptually, by introducing a prior for (the nuisance parameter) $b$ based on previous, source-free recordings. As in the current paper and [12], $b$ will be rather small in these problems of low counts. Also, in these references and the work of Loredo (this proceedings) $s$ is assumed independent of $b$ a priori, with $p(s|bI) = p(s|I)$. Following my discussion of Loredo, we have

$$B_0^{-1} = \int_0^\infty (1 + s/b)^n e^{-s} p(s|bI) ds,$$

for any observed number of counts $n$. Immediate deductions are that, since the prior $p(s|bI)$ does not depend on the background rate $b$, a record of $n = 0$ counts always supports the hypothesis of no source; this follows from (1) at $n = 0$ which gives $B_0^{-1}$ as the Laplace transform of $p(s|bI) = p(s|I)$, so that $B_0 \geq 1$ for any $p(s|I)$.

In related testing problems, ranges and bounds for Bayes’ factors (and other inferential quantities) have been usefully studied by Berger and co-authors ([3] section 4.3). As in discussion of Loredo, gross bounds are immediately available as $B_n \geq (1 + s_n/b)^{-n} e^{s_n}$, where $s_n = \max\{0, n - b\}$ is the maximum likelihood estimate of $s$ given $n$. This lower bound on $B_n$ is the maximised likelihood ratio traditionally used in non-Bayesian, likelihood based testing, so that any Bayesian solution would be more conservative and show less extreme support for the hypothesis of source existence. Just how conservative depends on the prior. Nousek’s discussion, and his more extensive work in [12], bears heavily on the use of ‘vague’ or ‘reference’ uniform priors — with improper priors, the posterior distribution $p(s|nbI)$ is perfectly well-defined though $p(n|(s > 0)bI)$ is also improper so the Bayes’ factor is undefined, a critical failing (one of many) of improper priors. A possible way out is to consider (proper) uniform priors over bounded ranges of the form $p(s|abI) = a^{-1}$ for $0 < s < a$, given some (arbitrary) upper bound $a$. Loredo uses such priors, which, as I noted in discussion there, might be defended in low count problems if ranges for $a$ postulated as scientifically plausible concentrate at small values. Immediate possibilities for further analysis now include exploring ranges of the Bayes’ factors that result as $a$ varies across suitable ranges. With such a prior, the Bayes’ factor

---

2Use of Bayes’ factors in simple testing situations has a very long history, and notably so in physical science applications. Indeed, Sir Harold Jeffreys, the eminent mathematical physicist and geophysicist who will be recognised by many physical scientists for his leading works in these areas, was an originator and leading proponent of Bayesian methods during the early part of this century, and Bayes’ factors as weights of evidence for or against hypotheses was central to much of his work [2].
On Bayesian Statistics in Astronomical Investigation

1. On Bayesian Statistics in Astronomical Investigation

Formula (1) is

\[ B_n^{-1} = a^{-1} \sum_{i=0}^{n} \binom{n}{i} b^{-i} \int_0^a s^i e^{-s} ds, \]

which may be easily evaluated for any specified \( a, b \) and number of counts \( n \). For all \( n > 0 \), \( B_n^{-1} \) has a unique maximum value greater than unity, a bound on the evidence for source existence. Take the value \( b = 0.73 \) in the SN1987A study in the Nousek papers (section 1.2 of the current paper) for illustration. Then, for \( n = 1, 2, 3, 4 \) and 5, the corresponding minimum values of \( B_n \) are about 0.97, 0.54, 0.18, 0.042 and 0.007, respectively. So, for any proper uniform (= ‘uninformative’) prior, the experimental evidence in favour of source existence does not exceed 0.18\(^{-1} \), representing an increase in odds of 11:2, if the number of counts is fewer than \( n = 4 \). Only with at least \( n = 4 \) counts does the bound indicate that evidence may strongly favour existence (and only then for some specific priors), with odds increasing by a factor of at most 0.042\(^{-1} \approx 24:1 \).

In addition to identifying the likelihood ratio test as overstating the case for source existence, we might make comparisons with the traditional significance testing approach. Under the null hypothesis of \( s = 0 \) the observed significance levels are simply tail areas under the Poisson distribution with rate \( b \), given by \( \sum_{x=n}^{\infty} b^x e^{-b}/x! \). At \( b = 0.73 \), this gives levels of 0.518, 0.166, 0.038 and 0.007 corresponding to \( n = 1, 2, 3 \) and 4, respectively. Thus a count of \( n = 3 \) is traditionally in favour of source existence at better than the 4% significance level, whereas the Bayes’ factor is bounded below by 0.18 in that case; the non-Bayes’ measure is unrealistically low, smaller than the absolute bound on the Bayes’ factor, for all \( n \).

These results indicate broadly applicable limits on the amount of evidence available from low count data in the presence of low background noise. Obviously, any specific prior may lead to much weaker evidence than indicated by these bounds, and the calculation should be performed case by case if unique priors are identified. Though apparently focused on the class of uniform priors, the result applies more generally using a simply derivable representation of all priors having decreasing density functions, mentioned in discussion of Loredo. There I claim the suitability of this class of priors as candidates for representation of realistic scientific opinion about the low count rate problem. This result can be applied in the spirit of [3] section 4.3 to determine bounds on inferential quantities – restricted here to just the Bayes’ factor – as follows. Any decreasing density function for \( s > 0 \) has the form

\[ p(s|bI) = \int_0^\infty p(s|abI) dF(a) = \int_s^\infty a^{-1} dF(a) \]

where \( F(a) \) is some prior distribution for \( a > 0 \). For any fixed \( a \), equation (2) defines the corresponding Bayes’ factor in favour of source existence, call this \( B_n^{-1}(a) \) here to explicitly denote the dependence on \( a \). Now the
1. On Bayesian Statistics in Astronomical Investigation

An overall measure is obtained by averaging with respect to the prior $F(a)$ — to obtain $B_n^{-1} = \int B_n^{-1}(a) dF(a)$. It follows that, whatever $F(a)$ may be (hence whatever decreasing density function $p(s|b)$ we choose), $B_n^{-1} \leq \max_{a=0}^{\infty} B_n^{-1}(a)$ so that the overall Bayes’ factor $B_n$ against existence still exceeds the lower bounds identified above by minimising over $a$ directly. Hence the earlier conclusions about limits to weights of evidence apply very widely indeed.

Of course, the Bayes’ factor is only part of the source existence question — given an initial probability $\pi = P(s = 0|bI)$ for the hypothesis of no source, the Bayes’ factor maps $\pi$, via Bayes’ theorem, to the posterior probability $P(s = 0|nbI) = \pi B_n/\{1 - \pi + \pi B_n\}$. Whatever the value of $\pi$, $B_n$ determines the increase/decrease of evidence, on the log-odds scale, due to the data, and the above discussion of $B_n$ as a weight of purely experimental evidence is relevant. However, the full picture must be borne in mind. A Bayes’ factor of $B_n = 0.3$, for example, indicates experimental evidence marginally in favour of source existence, but its eventual interpretation may vary widely depending on circumstances. I can envisage experiments designed to investigate potential sources whose existence is strongly suspected so that reasonable $\pi$ values are large, and others based on ‘scanning’ for sources in which reasonable $\pi$ values are small.

On more general issues, it it clear that investigations will often involve deeper physical modelling, introducing explanation of observed variation in detected counts through scientific description involving ‘regression’ type structures, though this simplistic paradigm captures the essence of the statistical issue and suffices for discussion here. The principles apply more generally; the technical and computational difficulties with more realistic, non-linear Poisson regression models are very much familiar to statisticians long schooled in modelling with non-gaussian and non-linear regressions; recent advances in high-dimensional numerical integration techniques permit Bayesian computations to be performed in complex models that are completely inaccessible to non-Bayesian approaches (see papers in [5], and references therein, for recent developments and applications). And certainly with respect to wider issues raised by Nousek’s paper, adherence to conditional (likelihood-based) inference procedures generally, and Bayesian methods more specifically, can often obviate the concerns of ‘breakdown’ of traditional methods.

Acknowledgements: This work was supported in part by the National Science Foundation under grant DMS-8903842.
1.3 References


