Predicting the future from observable sequences of events in astrophysics I. Maximum Likelihood Estimation for a fixed number of parameters

C. Röken, I. Lerche

Astrophys. Space Sci. Trans., 3, 1–13, 2007
www.astrophys-space-sci-trans.net/3/1/2007/

ASTRA
Astrophysics and **S**pace Sciences
Trans**a**ctions

Predicting the future from observable sequences of events in astrophysics I. Maximum Likelihood Estimation for a fixed number of parameters

C. Röken and I. Lerche

Institut für Theoretische Physik, Lehrstuhl IV: Weltraum und Astrophysik, Ruhr-Universität-Bochum, 44780 Bochum, Germany

Received: 8 November 2006 – Revised: 14 May 2007 – Accepted: 25 May 2007 – Published: 30 July 2007

Abstract. Using analytical and numerical methods, estimates are given of future predictions in astrophysics that can be gathered from a sequence of observed events, for example for γ-ray bursts. Some general probability considerations are provided and then a maximum likelihood estimation, together with an approximation for the large number limit of possible events. Illustrations are given of the numerical maximum likelihood estimation programs in the situations of both a large number and a finite number of events. The effects of data uncertainty are also considered. Numerical calculations and comparisons with theoretical expectations are presented too.

1 Introduction

In many fields of astrophysics one can observe a sequence of events, each of different intensity. Perhaps such is best exemplified by observations of γ-ray bursts which are short and intense pulses of soft γ-rays, where short means that the pulses can last from a fraction of a second to several hundred seconds. All known γ-ray bursts originate in extragalactic regions. Possible sources are for example core-collapsing rapidly rotating Wolf-Rayet stars or two colliding neutron stars orbiting in a binary. Today it is known that the form of a γ-ray burst is a narrow beam with corresponding energies around 10^{51} ergs (Frail et al., 2001; Panaitescu and Kumar, 2001; Piran et al., 2001). γ-ray bursts are detected in random directions, so they are an isotropic phenomenon. For example seven γ-ray bursts with photon energies that lie in an interval of 100 MeV to 18 GeV were measured with the high energy γ-ray detector Gamma-Ray Experiment Telescope (EGRET) at the Compton Gamma-Ray Observatory (Dingus and Catelli, 1998). For more detailed information about γ-ray bursts useful references are Piran (1999, 2000, 2004); Hurley (2002); Fishman and Meegan (1995); van Paradijs et al. (2000). Predictions of the energetics and the rate of such a sequence of γ-ray bursts with different intensities can be statistically discussed as shown in this paper.

One can formulate the questions of prediction in a more general way: From the observations of M events can one predict the most probable number, N, of all such observable events? Can one predict the fraction of all possible observable events that could lie less than (greater than) a given intensity? Can one predict whether the next event to arrive will lie above or below a given intensity? With what degree of confidence can one make any such predictions?

The purpose of this paper is to provide a set of procedures allowing one to address these questions under the following conditions:

1. Each event is statistically independent (no correlation) with any and all other events.
2. Each event arising from the remaining undiscovered population of events is random, with the probability of being measured proportional to a power of the arrival flux.

The first condition forbids the arrival of events from a periodic source, either due to dynamic repetitive motion of the source (such as a pulsar) or due to repetitive physical changes in the source (such as expansion and contraction). The second condition allows for the fact that arrival events may be strong but are "beamed" so that one sits on the "edge" of a beam and so sees a weak source. In either of these two extreme events, observability can be compromised. The basic questions are, then, within the framework of these two assumptions and given that one has measured M sequential events, can one estimate the residual total number of likely events, the power index of observability, the likely fraction of future events that will lie less than (greater than) a specific flux, and the probability that the next event in the sequence will have a flux less than (greater than) a specific value?

Correspondence to: C. Röken (cr@tp4.rub.de)

Published by Copernicus Publications on behalf of the Arbeitsgemeinschaft Extraterrestrische Forschung e.V.

2 Technical development

2.1 Probability considerations

Consider an ordered sequence of observed events, with event i of intensity I_i, and with M events measured. Suppose that a total of N events exists. One is interested in the probability that event $M+1$ will have an intensity I_{M+1} in excess of (or less than) a chosen value $I_{\rm cut}$, based on the information of the first M measured ordered events. For the N total events, let there be N_1 characterized by the average intensity $F_1\left(:=\left(\sum_{\rm low} I_i\right)/m_1\right)$, where the sum $\sum_{\rm low}$ extends over those events with intensities less than $I_{\rm cut}$ and N_2 characterized by the average intensity $F_2\left(:=\left(\sum_{i=1}^{M} I_i - \sum_{\rm low} I_i\right)/m_2\right)$, with $N_1+N_2=N$ and $m_1+m_2=M$. Note that this representation requires at least one observed event to lie above $I_{\rm cut}$. The probability that the first event lies in the class described by F_1 is

$$p_1 = \frac{F_1^a N_1}{N_1 F_1^a + N_2 F_2^a} \tag{1}$$

where a is the parameter of observability to be determined (see later) while the probability that the first event lies in the class described by F_2 is

$$p_2 = 1 - p_1 = \frac{F_2^a N_2}{N_1 F_1^a + N_2 F_2^a} \tag{2}$$

After m_1 measured events in class F_1 and m_2 measured events in class F_2, the probability the next event is in class $j\,(j=1,2)$ is

$$p_{\rm next}\,(j) = \frac{F_j^a\left(N_j - m_j\right)}{F_1^a\,(N_1-m_1) + F_2^a\,(N_2-m_2)} \tag{3}$$

Set the membership class $k_{ij}=1$ if the i^{th} event is in class j, and $k_{ij}=0$ if the i^{th} event is not in class j, with $k_{i1}+k_{i2}=1$. Then, after M events, ordered $i=1,\ldots,M$, the joint probability distribution is

$$p\,(M;\,N_1,\,N_2;\,a) = \prod_{i=1}^{M} \frac{\left[\sum_{j=1}^{2} k_{ij} F_j^a \left(N_j - \sum_{t=1}^{i-1} k_{tj}\right)\right]}{\left[\sum_{j=1}^{2} F_j^a \left(N_j - \sum_{t=1}^{i-1} k_{tj}\right)\right]} \tag{4}$$

With $F_1/F_2 = r$ and $N_2 = N - N_1$ Eq. (4) can be written:

$$p\,(M;\,N,\,N_1;\,a) \tag{5}$$
$$= \prod_{i=1}^{M} \frac{\left[k_{i1} r^a \left(N_1 - \sum_{t=1}^{i-1} k_{t1}\right) + k_{i2}\left(N - N_1 - \sum_{t=1}^{i-1} k_{t2}\right)\right]}{\left[r^a\left(N_1 - \sum_{t=1}^{i-1} k_{t1}\right) + N - N_1 - \sum_{t=1}^{i-1} k_{t2}\right]}$$

where the terms in the summations have to be omitted for $i=1$, the first event in the ordered sequence. Note that because only two classes are available (events with intensities below the cut and events with intensities above the cut) and because a given event must lie somewhere, then $k_{i1}=1-k_{i2}$. We use this fact in the next subsection. Unknown in Eq. (5) are the total number N of events that can occur, N_1 the number of events in the class characterized by F_1, as well as the parameter a. There are at least two procedures available to provide estimates of the parameters N, N_1 and a based on the M measured events. The procedures are: Maximum likelihood estimation and Bayesian updating (Aldrich, 1997; Fisher, 1925; Fisher , 1934; Arps et al., 1971; Jaynes, 1978).

In this paper we consider first the maximum likelihood estimation in general and then discuss the situation when the number of observed events to date, M, is taken to be small compared to the number, N, of available events.

2.2 Maximum Likelihood Estimation

The joint probability distribution given by Eq. (5) comes as close as it can to matching the observed sequence of events when Eq. (5) takes on its maximum value with respect to the parameters N, N_1 and a. It is convenient to change the variable N_1 to $N\sin^2\theta$ when $N_2 = N\cos^2\theta$ and to then use N, $\cos\theta\,(=\mu)$ and a as basic variables. The joint probability distribution has extremum values when

$$\frac{\partial p}{\partial N} = 0; \quad \frac{\partial p}{\partial \mu} = 0; \quad \frac{\partial p}{\partial a} = 0. \tag{6}$$

Note, however, that an extremum can be a maximum, minimum or point of inflection. Also note that the absence of an extremum does not imply the absence of a maximum. For instance $p=\mu^2\,(0\le\mu\le 1)$ has no extremum in $0\le\mu\le 1$, nevertheless it has a maximum at $\mu=1$. Further, the parameters N, μ^2 and a at an extremum can end up in physically unacceptable domains (e.g. $N<0$, $\mu^2<0$, $\mu^2>1$, $a<0$, etc.) in which case the extremum, even if a maximum, is not physically appropriate. We return to this point a little later.

From $\partial p/\partial N = 0$ one obtains the constraint

$$\sum_{i=1}^{M} \frac{k_{i1} r^a + \mu^2\,(k_{i2} - k_{i1} r^a)}{[N[k_{i1} r^a + \mu^2\,(k_{i2} - k_{i1} r^a)] - \alpha_i]}$$
$$= \sum_{i=1}^{M} \frac{r^a + \mu^2\,(1-r^a)}{[N[r^a + \mu^2\,(1-r^a)] - \beta_i]} \tag{7}$$

where

$$\alpha_i = k_{i1} r^a \sum_{t=1}^{i-1} k_{t1} + k_{i2} \sum_{t=1}^{i-1} k_{t2} \tag{8}$$

and

$$\beta_i = r^a \sum_{t=1}^{i-1} k_{t1} + \sum_{t=1}^{i-1} k_{t2} \tag{9}$$

with $\alpha_1 = \beta_1 = 0$.

From $\partial p/\partial \mu = 0$ one obtains the constraint

$$\sum_{i=1}^{M} \frac{k_{i2} - k_{i1} r^a}{N[k_{i1} r^a + \mu^2\,(k_{i2} - k_{i1} r^a)] - \alpha_i}$$
$$= \left(1 - r^a\right) \sum_{i=1}^{M} (N[r^a + \mu^2\left(1-r^a\right)] - \beta_i)^{-1} \tag{10}$$

while from $\partial p/\partial a = 0$ one obtains the constraint

$$\sum_{i=1}^{M} \frac{k_{i1}}{N[k_{i1}r^a + \mu^2\,(k_{i2} - k_{i1}r^a)] - \alpha_i} = \sum_{i=1}^{M}(N[r^a + \mu^2\,(1 - r^a)] - \beta_i)^{-1} \tag{11}$$

unless $r = 1$ in which case Eq. (11) is ignorable. For $r \neq 1$ the three constraints are not independent because one can immediately derive the third constraint from any two of Eqs. (7), (10) and (11). Thus, at best one can derive two of N, μ and a as functions of the third parameter. The constraint equations do not admit of simple analytic solutions for, say, N and μ as functions of a for a set M of observations, except in the somewhat simple case when $M = 2$.

Because a priori estimates for N and μ are needed in the Bayesian update procedure (see Appendix A), it is convenient to develop here the estimates of N and μ for the case $M = 2$. Then

$$N(M = 2) = r^{-a}[A - (1 - r^a)B] \tag{12}$$

$$\mu^2(M = 2) = \frac{Br^a}{A - (1 - r^a)B} \tag{13}$$

where A and B satisfy the pair of equations

$$2B(1 - 2k_{22})(1 - 2k_{11}) = -\alpha_1(1 - 2k_{11}) + A[(1 - 2k_{11})(1 - k_{22}) - k_{11}(1 - 2k_{22})] \tag{14}$$

$$A(A - \beta_1)(1 - k_{11} - k_{22}) = (1 - 2k_{22})(2A - \beta_1[Ak_{11} + (1 - 2k_{11})B] \tag{15}$$

Equations (14) and (15) provide linear algebraic equations for determining A and B. Two requirements on the solution are

1. $A > B(1 - r^a)$ and $A > 0$;
2. $A > B$

in order that the estimates $N(M = 2) > 0$ and $\mu^2(M = 2) > 0$. If $A > B$ and $A > 0$ then $A > B(1 - r^a)$ for all $a \geq 0$ in $r < 1$. Thus one only has to search for the solution of Eqs. (14) and (15) with $A > B$ and $A > 0$. From Eqs. (14) and (15) one has

$$A = \beta_1\alpha 1(1 - 2k_{11}) \cdot[2\alpha_1(1 - 2k_{11}) + \beta(k_{11} + k_{22} - 1)]^{-1} \tag{16}$$

$$B = \frac{\alpha_1}{1 - 2k_{22}} \frac{[\beta_1 k_{11}(2k_{22} - 1) - \alpha_1(1 - 2k_{11})]}{[\beta_1(k_{11} + k_{22} - 1) + 2\alpha_1(1 - 2k_{11}]} \tag{17}$$

For a given two event situation (where at least one event must have an intensity above $I_{\rm cut}$), it may happen that Eqs. (16) and (17) do not permit $N(M = 2) > 0$ and/or $0 \leq \mu^2(M = 2) \leq 1$ in which case the extremum constraints do not represent a physically allowable maximum. The same sense of requirement of physical allowability is true for any $N(M)$, $\mu^2(M)$ and $a(M)$, not just for $M = 2$. Thus, the constraint Eqs. (7) and (10) (which are impossible to solve analytically for $M \geq 5$) may not provide acceptable ($N > 0$, $0 \leq \mu^2 \leq 1$, $a \geq 0$) ranges for the desired parameters.

One way to avoid this extremum problem is to ignore the extremum constraint approach and deal directly with Eq. (4) in the form

$$p\,(M; N; \mu, a) = \prod_{i=1}^{M} \frac{\left[k_{i1}r^a\left(N\sin^2\theta - \gamma_i\right) + k_{i2}\left(N\cos^2\theta - \delta_i\right)\right]}{\left[r^a\left(N\sin^2\theta - \gamma_i\right) + N\cos^2\theta - \delta_i\right]} \tag{18}$$

with $\gamma_i = \sum_{t=1}^{i-1} k_{t1}$ and $\delta_i = \sum_{t=1}^{i-1} k_{t2}$.

Monte Carlo searching in $a \geq 0$, $N \geq M$, $0 \leq \mu^2 \leq 1$ then enables one to determine rapidly the parameter values yielding the maximum probability. In this situation one uses the physical acceptability requirements first so that one is guaranteed all values of p will lie in $0 \leq p \leq 1$. One catch is that there is no guarantee that the three parameters N, μ^2, a will be unique. It can happen that multiple domains in (N, μ^2, a) space will all produce an identical maximum probability. Thus a broad search is then required in order to identify all such domains and determine how to proceed further to narrow down, or eliminate, some of the domains.

One strategy that has proved somewhat successful in this regard is the so called Training Procedure (Lerche, 1997). Start with $M = 2$ and ensure that at least one of the first two events has an intensity above $I_{\rm cut}$. Run Monte Carlo simulations in $N \geq 2$, $0 \leq \mu^2 \leq 1$, $a > 0$ and record the largest value of $p(M = 2)$, say $p_{\max}(2)$, and the N, μ^2 and a values at which $p_{\max}(2)$ occurs. In the event multiple domains in (N, μ^2, a) space provide the same $p_{\max}(2)$ record all such domains. Now increase M from $M = 2$ to $M = 3$ and repeat the Monte Carlo procedure, again identifying all domains where $p_{\max}(3)$ occurs. Carry through this operational sequence at ever increasing M until all the observed events have been used. Then investigate the stability of the domains in (N, μ^2, a) space to determine these regions that provide a neighborhood within which (N, μ^2, a) stay and those domains that are highly unstable with respect to increases in M, i.e. these domains where (N, μ^2, a) for $p_{\max}(M)$ depart massively from the domains for $p_{\max}(M + 1)$. Then it is highly unlikely that such unstable domains represent appropriate values of the triad (N, μ^2, a), although they can not be categorically ruled out. Check the value of $p_{\max}(M)$ in relation to unity. If $p_{\max}(M)$ is close to unity then the parameter triad obtained represents the statement that the maximum likelihood estimation is close to the observed sequence (Cronquist, 1991).

While this basic strategy is somewhat successful in some cases in identifying and determining the relevant parameter domains from a given event sequence, it is less than satisfying for a variety of reasons. First the strategy requires too

much ad hoc user evaluation in the sense of being required to assess stable and unstable domains as M is increased. Second, the strategy requires a qualitative judgement on what constitutes close to unity for $p_{\max}(M)$. Third, the strategy is hardly unassailable as a procedure that will provide systematic improvements as the number of observed events is increased. Fourth, the strategy is extremely ineffective in assessing the number, N, of possible observable events when $N \gg M$ as we show in the next section of the paper. A better, and systematic, method is required for steady improvement as one increases the number of ordered observed events from $M = 2$ to the total observed, M. Precisely that concern is addressed by Bayesian updating. We discuss the Bayesian procedure in a later paper. However, the basic mathematical development for the Bayesian procedure is given in Appendix A. Equally, a systematic automatic procedure for assessing the relevant ranges of parameters is given in Appendix B (Lerche, 1997; Lumley, 1970).

Those mathematical developments are presented here both for the sake of completeness as well as so that only numerical procedures and their implementation and applications need be given in the second paper of this series.

2.3 The limit of large number of possible events, $N \gg M$

Expansion of Eq. (18) for $N \gg M$ yields, to first order in $1/N$, that

$$p = \left\{ \prod_{i=1}^{M} \frac{[k_{i1} r^a \sin^2\theta + k_{i2} \cos^2\theta]}{r^a \sin^2\theta + \cos^2\theta} \right\} \times \left\{ 1 - \frac{r^a}{N} \sum_{i=1}^{M} \cdot \frac{[\gamma_i \cos^2\theta \, (k_{i1} r^a - k_{i2}) + \delta_i \sin^2\theta \, (k_{i2} - k_{i1})]}{\left(k_{i1} r^a \sin^2\theta + k_{i2} \cos^2\theta\right) \left(\cos^2\theta + r^a \sin^2\theta\right)} \right\} \tag{19}$$

If one were to ignore the term factored by $1/N$ in Eq. (19) then the probability distribution is independent of N, indicating that one could perform a lowest order Monte Carlo search for the values of μ^2 and a that maximize

$$p_0 \equiv \prod_{i=1}^{M} \frac{k_{i1} r^a \sin^2\theta + k_{i2} \cos^2\theta}{r^a \sin^2\theta + \cos^2\theta} \tag{20}$$

The independence of p_0 from N is a significant improvement in computer time in that only two parameters have to be searched. The downside is that no information is made available on the value of N except, of course, that $N \gg M$. One strategy is to use the values of μ^2 and a, obtained by maximizing Eq. (20), as first order estimates in Eq. (19) and then search only the local neighborhood of these μ^2 and a values. Again, such a strategy calls for some external assessment of local neighborhood and can also suffer from the same disadvantage as the more general Training Procedure, producing multiple domains where p_0 is equally maximal (Lerche, 1997; McCray, 1969; Megill, 1971; Rose, 1987). To obviate all of these difficulties and drawbacks, it is appropriate to invoke Bayesian updating (see Appendix A).

Table 1. Input 1 ($N \gg M$) showing a systematic increase in intensity of sequential events and Input 2 ($N \gg M$) showing a more disordered set of sequential events.

Input 1	Input 2	μ	a
10	40	0.1	0.01
20	90	0.2	0.2
30	60	0.3	0.3
40	10	0.4	0.4
50	70	0.5	0.5
60	100	0.6	1.1
70	30	0.7	1.3
80	80	0.8	1.5
90	20	0.9	2.0
100	50		
Reference	39		

Another result for $N \gg M$, $N_1 \gg m_1$, and $N_2 \gg m_2$ is that Eq. (3) reduces to

$$p_{\text{next}}(1) = \frac{r^a \left(1 - \mu^2\right)}{r^a \left(1 - \mu^2\right) + \mu^2} \tag{21}$$

$$p_{\text{next}}(2) = \frac{\mu^2}{r^a \left(1 - \mu^2\right) + \mu^2} \tag{22}$$

3 Maximum Likelihood Estimation programs

To be able to address the questions asked in the Introduction one has to convert the mathematical formalism described in Sect. 2 into computer programs. In this section of the paper two programs for the maximum likelihood estimation procedure are introduced and the results for two different experiments (ordered sequences of observations) are discussed.

Of the two programs, one describes the situation including all parameters N, while the second program is appropriate for the special case $N \gg M$. Both programs are created via Open-Office using spreadsheets to achieve high user friendliness.

3.1 The special case, $N \gg M$

Starting with the description of the special case program for $N \gg M$. The mathematical basis for this algorithm is Eq. (20). Extrema occur when its partial derivations with respect to μ and a are zero as shown in Eqs. (6). In the program these equations are used for the variation of the parameters μ and a. The variation structure includes ten sections each for ten measured values. These values (ordered sequence of observations) are, in addition to the variation parameters μ

Table 2. This Table represents a small part of the calculation of the right side of Eq. (18). There are two sections each with 4 columns. In these sections one varies only one parameter (in this case the parameter μ; the columns in a section represent different μ values) while the other two parameters are constant (Note that one has different constant parameters in each section). The entries in the columns are the separate factors of Eq. (18). The last entries of the columns are the products of these factors.

	0.98	0.93	0.36	0.23
	0.98	0.93	0.35	0.23
	0.98	0.93	0.35	0.22
	0.02	0.07	0.66	0.78
	0	0.06	0.65	0.78
	-0.02	0.04	0.65	0.78
	-0.04	0.02	0.64	0.78
	-0.06	0	0.64	0.77
	-0.08	-0.02	0.63	0.77
	-0.11	-0.04	0.63	0.77
$p(\mu, a1, N1) =$	Not Valid	Not Valid	0.2	0.2
	0.99	0.94	0.42	0.28
	0.99	0.94	0.41	0.27
	0.99	0.94	0.41	0.27
	0.01	0.06	0.60	0.74
	0	0.04	0.59	0.74
	-0.02	0.03	0.59	0.73
	-0.03	0.02	0.58	0.73
	-0.05	0	0.58	0.73
	-0.06	-0.02	0.57	0.72
	-0.08	-0.03	0.57	0.72
$p(\mu, a2, N1) =$	Not Valid	Not Valid	0.16	0.22

and a, part of the user input as shown in Table 1 for two input sets. Both input sets are for the same measured values but in a permutated order compared to a reference value set to $I_{\rm cut} = 39$ (see Fig. 1). The difference that occurs when the order of the measured values is permuted is a change in the lowest number of observations for which the predictions make sense for a given $I_{\rm cut}$. For Input 1 this value is $M = 5$ and for Input 2 it is $M = 4$, because one value has always to lie above $I_{\rm cut}$. An example for a variation of a parameter is given in Table 2, showing the variation of μ for different, but fixed, a and N values for the 10 observed intensity values. The program determines the best μ and a values for the parameter interval chosen by the user (see Table 1 for the intervals and Fig. 2 for the best values for both inputs). These μ and a parameters fit best to Eqs. (6).

The problem that occurs in this approach is that the user has to vary the parameters μ and a manually. A great deal of time can be involved and for any of those chosen parameter intervals only the best values in this local area are obtained, which are not necessarily the globally best values. This deficiency can be repaired by an automatic search subroutine for calculating the best parameters, which is done in the second paper of this sequence (the mathematical procedure is presented in Appendix B). With the obtained best values for μ

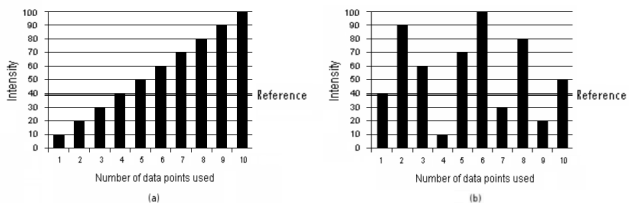

Fig. 1. Input 1(a) and Input 2(b) values in comparison with the reference intensity.

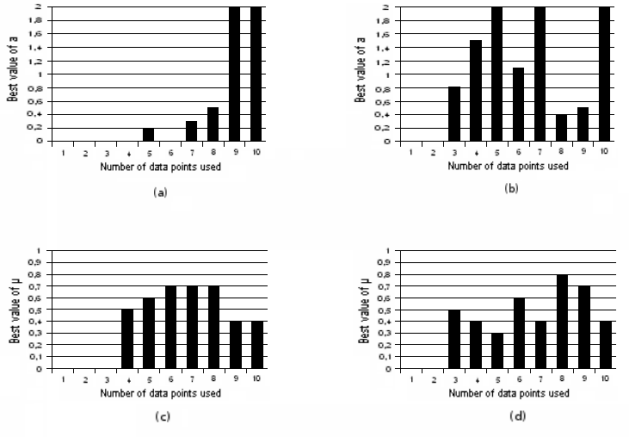

Fig. 2. Best values of a and μ (Input 1(a) and (c) and Input 2(b) and (d); $N \gg M$).

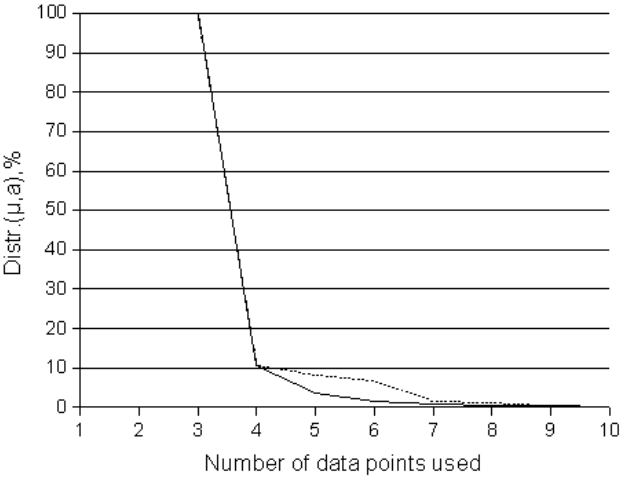

Fig. 3. Joint probability distributions (Input 1 (solid) and Input 2 (dot); $N \gg M$) estimated for all M - see text for further description.

and a the program can determine the joint probability distribution functions for the events (measured values).

These functions are shown in Fig. 3 for Input 1 (solid) and for Input 2 (dot). A slight difference can be observed in the tail of the functions ($M \geq 4$), caused by the permutation of the order of the measured values. The probability that the next event lies in class 1 (below $I_{\rm cut}$) or class 2 (above

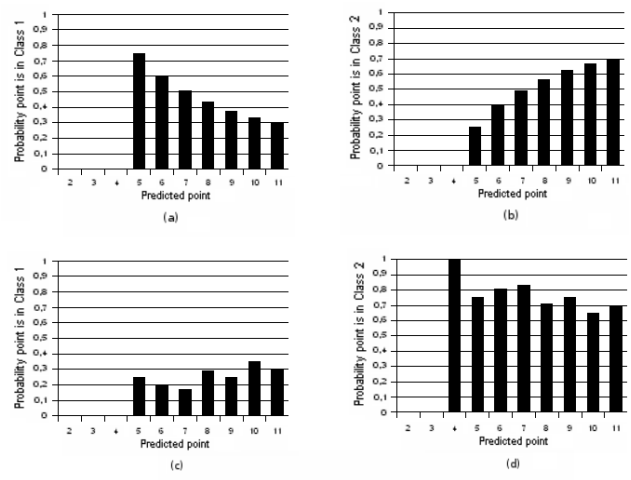

Fig. 4. Probability for the next event to lie in class 1 or class 2 (Input 1(a) and (b) and Input 2(c) and (d); $N \gg M$). (a) and (b): Note that the first three events have intensities $< I_{\rm cut}$, while the fourth has $I_4 > I_{\rm cut}$. Thus probability predictions can be given only for the fifth and higher events in the sequence. (c) and (d): Note that for the first event of Input 2 $I_1 > I_{\rm cut}$, so the range of validity for the probability for the next event to lie in class 1 or class 2 starts with a prediction for the fourth measured value.

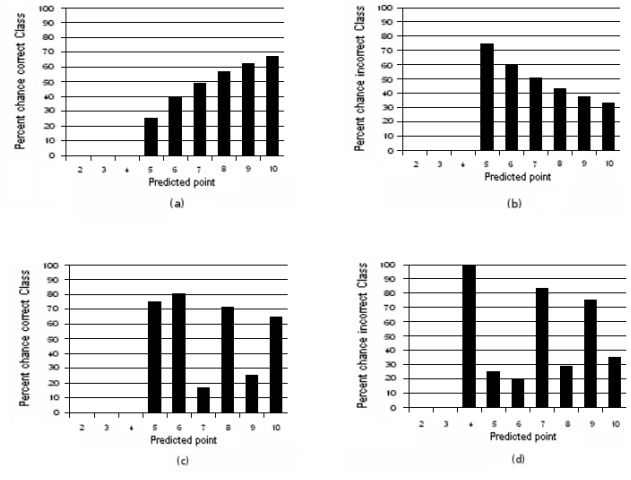

Fig. 5. Percent chance predicted point is in correct class (Input 1(a) and (b) and Input 2(c) and (d); $N \gg M$).

$I_{\rm cut}$), given by Eq. (3), is illustrated for both inputs in Fig. 4, respectively. With a systematic increase in the number of measured values the probability that the event lies in class 1 decreases while the probability that the event lies in class 2 increases (Figs. 4(a) and (b). This effect is caused by the linear growth of the ordered intensity values in Input 1. For Input 2 (see Figs. 4(c) and (d)) one has a more chaotic behavior of the data so that the decreasing/increasing effect of data Input 1 is now absent. With this program one is also able to calculate the probabilities that an event is either in the correct or incorrect class as shown in Fig. 5 for Inputs 1 and 2, respectively. Figures 5(a) and (b) illustrate the fact that if more measured values are incorporated into the calculations the higher is the probability to find an event to be in the correct class, which is again caused by the linear increase of the data values of Input 1. For data set 2 the probability to find an event in the correct class is between 15% and 80% (Figs. 5(c) and (d)). There is no specific pattern to be found.

Table 3. Input 1 and Input 2 ($N \geq M$).

Input 1	Input 2	μ	a	N
10	40	0.1	0.5	100
20	90	0.2	0.3	52
30	60	0.7	0.6	53
40	10	0.8	0.1	85
50	70			
60	100			
70	30			
80	80			
90	20			
100	50			
Reference	39			

3.2 The general case, $N \geq M$

In the program for the general case finite values of the parameter N are used, so Eq. (18) is the mathematical basis for the numerical variation calculations of the parameters N, μ and a. The analytical procedure utilized in this program differs from the one used in the special case program given above. Instead of finding the maxima of Eq. (18) by calculating its partial derivations one finds these by direct variation of the parameters so that p is as close to unity as possible for the prescribed values of the three input parameters. The three input boxes for the parameters μ, a and N are much smaller than for the special case program (see Table 3). The reason for this limitation of the input values is the rapid increase of the size of the program caused by every new parameter. Because $N \geq M$ there are restrictions for the minimum value of N for the calculations to make sense. Those do not occur in the special case $N \gg M$.

Here, the ordered sequences of observations (Inputs 1 and 2) and the reference intensity $I_{\rm cut}$ are the same as in the special case program. The general case program calculates the best μ, a and N values for a given set of data in a local parameter interval chosen by the user (see Fig. 6 for the best parameters of each of the two data sets). These μ, a and N values provide the highest value of p (see Eq. (18)). The joint probability distributions (see Eq. (4)) for these best values are shown in Fig. 7 for the data sets 1 (solid) and 2 (dot), respectively. The two joint probability distribution functions differ again only slightly in their tails. The probability that the next event lies in class 1 (below $I_{\rm cut}$) or class 2 (above

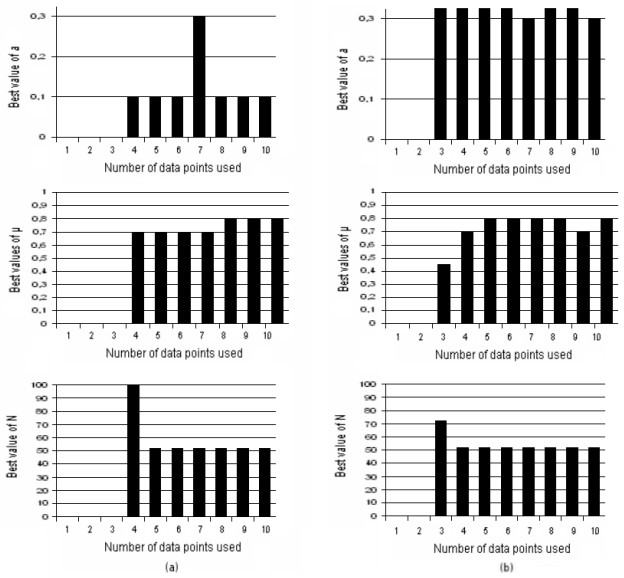

Fig. 6. Best values of a, μ and N (Input 1(a) and Input 2(b); $N \geq M$).

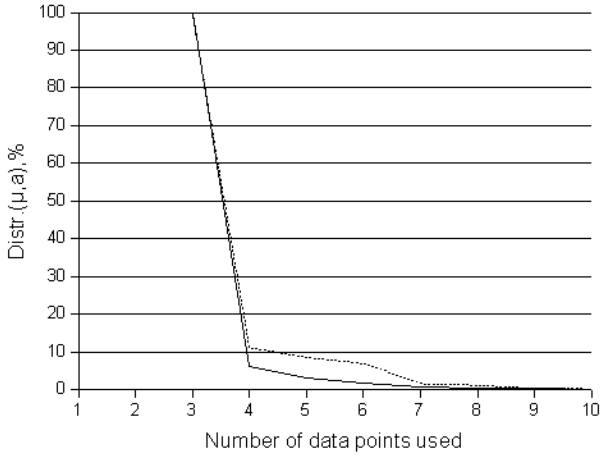

Fig. 7. Joint probability distribution (Input 1 (solid) and Input 2 (dot); $N \geq M$) estimated for all M - see text for further description.

$I_{\rm cut}$) is illustrated in Figs. 8(a) and (b) and Figs. 8(c) and (d) for the data sets 1 and 2, respectively. The behavior is nearly the same as in the special case program (see Fig. 4) with the difference that the increase of the probability in Figs. 8(a) and (b) is not as extreme as in Figs. 4(a) and (b). The probability that the events are in the correct or incorrect class is presented in Figs. 8(e) and (f) for Input 1 and Figs. 8(g) and (h) for Input 2. With a growing number of measured values the increase of the probability that the next data point is in the correct class is compared to the results from the special case program less powerful. In Fig. 8(g), no particular pattern can be found.

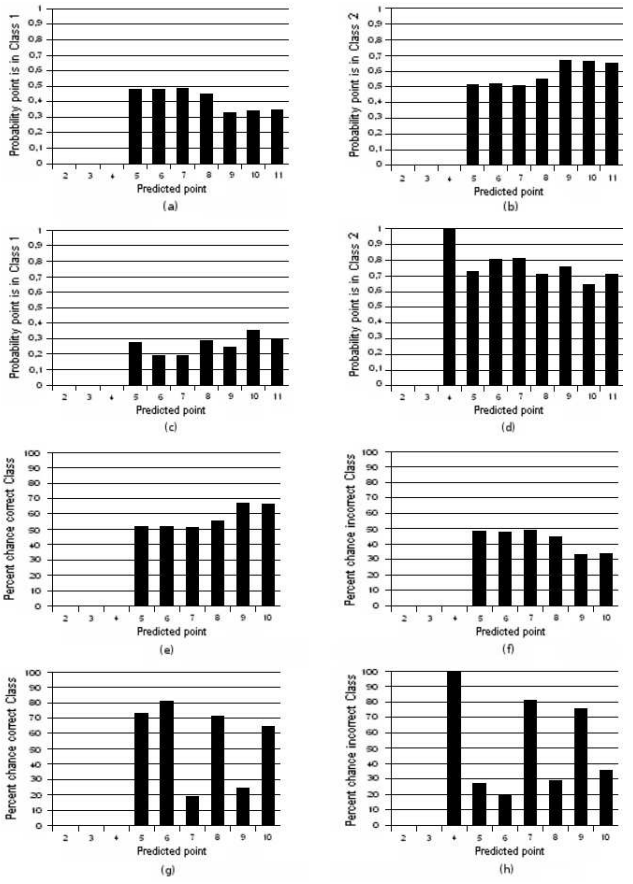

Fig. 8. Probability for the next event to lie in class 1 or class 2 (Input 1(a) and (b) and Input 2(c) and (d); $N \geq M$). (a) and (b): Note that the first three events have intensities $< I_{\rm cut}$, while the fourth has $I_4 > I_{\rm cut}$. Thus probability predictions can be given only for the fifth and higher events in the sequence. (c) and (d) Note that for the first event of Input 2 $I_1 > I_{\rm cut}$, so the range of validity for the probability for the next event to lie in class 1 or class 2 starts with a prediction for the fourth measured value. In (e) - (h) the percent chance that the predicted point is in the correct class is presented (Input 1(e) and (f) and Input 2(g) and (h); $N \geq M$).

As an aside, one has to be aware of numerical restrictions. On the one hand the number of the sequence of measured events is limited to a small number (here $M = 10$), because it is regulated by the size of the database as well as the CPU. On the other hand one has to pay attention that the conditions $N\left(1-\mu^2\right) \geq \gamma_i$, $N\mu^2 \geq \delta_i$ and $\mu_{best.} \neq 0$, $\mu_{best.} \neq 1$ are maintained. The central problem of the special case program is the ambiguity of the conditions $\partial p/\partial N = 0$, $\partial p/\partial \mu = 0$ and $\partial p/\partial a = 0$. One obtains not only the desired maxima, but also all general extremum values. It can not easily be determined if one has found a minimum, a point of inflection or a (right) maximum. The mathematical method used for the general program is less problematic, but a more extensive program has to be built. As mentioned before, the programs search for the best parameters in a local interval de-

Table 4. Summary of the Monte Carlo simulation variables used together with ranges and statistical characteristics.

Name	Maximum	Minimum	Mean	Variance	Std. Dev.	Dev./Mean in %
$p_{\mathrm{next},1.10}$	0.97	−0.41	0.36	0.05	0.23	64.62
$p_{\mathrm{next},1.4}$	76.87	−338.29	0.68	139.62	11.82	1747.18
$p_{\mathrm{next},1.5}$	36.23	−0.29	0.88	2.55	1.60	181.13
$p_{\mathrm{next},1.6}$	3.27	−25.11	0.63	0.73	0.85	136.09
$p_{\mathrm{next},1.7}$	0.97	−0.42	0.54	0.05	0.22	40.09
$p_{\mathrm{next},1.8}$	0.97	−0.41	0.48	0.05	0.22	46.90
$p_{\mathrm{next},1.9}$	0.98	−0.41	0.41	0.05	0.23	55.74
$p_{\mathrm{next},2.10}$	1.41	0.03	0.64	0.05	0.23	36.30
$p_{\mathrm{next},2.4}$	339.29	−75.87	0.32	139.62	11.82	3650.09
$p_{\mathrm{next},2.5}$	1.29	−35.23	0.12	2.55	1.60	1343.42
$p_{\mathrm{next},2.6}$	26.11	−2.27	0.37	0.73	0.85	228.61
$p_{\mathrm{next},2.7}$	1.42	0.03	0.46	0.05	0.22	47.54
$p_{\mathrm{next},2.8}$	1.41	0.03	0.52	0.05	0.22	43.04
$p_{\mathrm{next},2.9}$	1.41	0.02	0.59	0.05	0.23	39.11
a_1	0.20	0.01	0.10	0.00	0.06	53.32
a_2	0.30	0.01	0.15	0.01	0.08	54.53
a_3	0.40	0.10	0.25	0.01	0.09	34.45
a_4	0.90	0.01	0.45	0.07	0.26	57.52
a_5	0.99	0.20	0.59	0.05	0.22	38.35
a_6	1.50	0.20	0.85	0.14	0.37	44.08
a_7	3.00	0.90	1.93	0.36	0.60	31.28
a_8	19.99	0.01	10.39	33.41	5.78	55.64
a_9	99.98	10.16	55.01	666.86	25.82	46.94
$M1$	19.98	0.03	9.78	33.94	5.83	59.56
$M10$	149.73	60.01	104.80	662.74	25.74	24.57
$M2$	39.99	5.11	22.92	103.68	10.18	44.43
$M3$	35.00	15.02	25.00	33.25	5.77	23.06
$M4$	59.97	20.03	39.45	129.94	11.40	28.90
$M5$	74.99	30.07	52.68	174.80	13.22	25.10
$M6$	89.83	20.16	55.24	386.84	19.67	35.60
$M7$	79.98	65.01	72.65	18.84	4.34	5.97
$M8$	109.76	33.12	71.62	482.05	21.96	30.65
$M9$	122.94	67.26	95.13	250.01	15.81	16.62
μ_1	0.15	0.01	0.08	0.00	0.04	51.35
μ_2	0.29	0.11	0.20	0.00	0.05	26.03
μ_3	0.40	0.15	0.28	0.01	0.07	26.28
μ_4	0.55	0.22	0.38	0.01	0.09	24.61
μ_5	0.70	0.20	0.45	0.02	0.15	32.33
μ_6	0.99	0.33	0.66	0.04	0.19	28.95
μ_7	0.90	0.50	0.70	0.01	0.12	16.95
μ_8	1.20	0.40	0.79	0.05	0.23	28.58
μ_9	1.10	0.77	0.94	0.01	0.10	10.23

fined by the user and not by an automatic mode of operation. So one can not be certain in gaining the most appropriate parameter values.

4 Discussion and Conclusion

This paper has developed an analytical description and numerical methods for future predictions of events in astrophysics that can be garnered from a sequence of observed events. To determine the possible future behavior of occurring events one first investigates general probability considerations using a maximum likelihood estimation. The results received from the maximum likelihood estimation are further discussed for the approximation of the large number limit of possible events. The mathematical formalisms had to be transformed into computer codes, one, for a finite number system and one for a large number approximation as a special case of a finite number system.

The procedures developed so far make the intrinsic assumption that each intensity in the measured sequence of intensities is known with precision. In fact, any individual measured intensity has a degree of uncertainty, which

Table 5. Probability statistics for input $p_{\text{next},2.8}$ showing the chances of finding $p_{\text{next},2.8}$ in given ranges as well as the cumulative probability of obtaining a value for $p_{\text{next},2.8}$ less than a specified worth.

Variables	Correlation Coef.
μ_7	0.0112
$M1$	−0.0144
μ_6	0.0153
μ_3	−0.0181
a_4	0.0187
$M9$	−0.0193
a_6	0.0205
μ_5	−0.0214
a_5	0.0232
μ_1	0.0254
a_8	0.0261
a_1	−0.0262
a_2	−0.0269
μ_9	0.0285
$M3$	0.0304
a_9	0.0325
μ_2	−0.0534
μ_4	−0.0538
$M10$	0.0543
a_3	0.0610
$M2$	0.0642
$M7$	0.0657
a_7	−0.0682
$p_{\text{next},1.4}$	−0.0953
$p_{\text{next},2.4}$	0.0953
$p_{\text{next},1.5}$	−0.1144
$p_{\text{next},2.5}$	0.1144
$M8$	0.1223
$p_{\text{next},1.6}$	−0.1278
$p_{\text{next},2.6}$	0.1278
$M5$	0.1646
$M6$	0.2064
$M4$	0.2129
μ_8	0.2489
$p_{\text{next},1.10}$	−0.7059
$p_{\text{next},2.10}$	0.7059
$p_{\text{next},2.7}$	0.7648
$p_{\text{next},1.7}$	−0.7648
$p_{\text{next},2.9}$	0.8125
$p_{\text{next},1.9}$	−0.8125
$p_{\text{next},1.8}$	−1.0000

is why results are usually quoted to 1σ (occasionally 2σ). The cut intensity level, I_{cut}, is, of course, precisely known because one can choose the value. If a particular member of the measured sequence has a mean intensity I_i such that $\mid I_i - I_{\text{cut}} \mid \leq 1\sigma$ then, with confidence, one can ascribe that member to either the class less than or greater than the cut. However, when the uncertainty on I_i is sufficiently large that $\mid I_i - I_{\text{cut}} \mid \geq 1\sigma$ one does not know into which category the member is to be assigned. Accordingly, depending on the

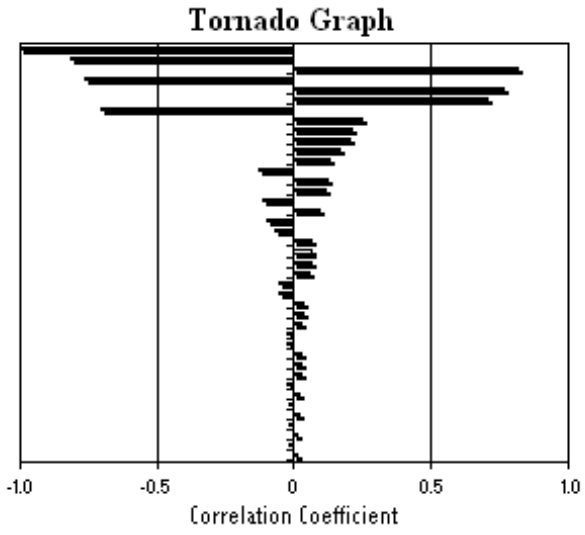

Fig. 9. Sensitivity of the output $p_{\text{next},2.8}$ in a Tornado Graph with the abscissa being the correlation coefficient range $(-1, 1)$. The largest values have the most influence on the sensitivity of $p_{\text{next},2.8}$ and are taken from Table 5.

value chosen for I_{cut}, there will be corresponding uncertainties in the parameters N, μ and a and also an uncertainty on assessing whether the $(M+1)^{st}$ intensity is predictable as being in the lower or upper classes. Fortunately, these problems can be addressed readily using Monte Carlo procedures as follows. Given that one knows the uncertainty distribution around the mean on each of the measured values then one chooses the same distribution for a suite of Monte Carlo runs from which one can compute the likely membership classes, the average values and uncertainties for N, μ and for a given sequence of measurements, and the uncertainty on the probability that the next value will lie above or below the I_{cut} value. A numerical illustration of this Monte Carlo procedure is now given using data set 1.

That simulation was performed with 1000 Monte Carlo iterations for the input set ("Value", μ, a) and the output set ($p_{\text{next},1}$, $p_{\text{next},2}$). Note that the output set starts with the fourth value of $p_{\text{next},1}$ and $p_{\text{next},2}$ because the first three values are less than I_{cut}. A continuously uniform distribution was used for the parameter μ and a and a discrete uniform distribution for the parameter "Value".

Table 4 provides a summary of the Monte Carlo simulation. The most important quantity is the mean value, because it represents the "best" value for the Monte Carlo output. The maximum and minimum values mark the edges of the simulation interval while the other quantities are standard statistical descriptors. Table 5 and Fig. 9 show, as an example, the sensitivity to the data (correlation coefficients) of the output $p_{\text{next},2.8}$ and the corresponding visualization in a Tornado Graph. Here $p_{\text{next},2.8}$ is the probability that the next event lies in class 2 using eight measured values. It is

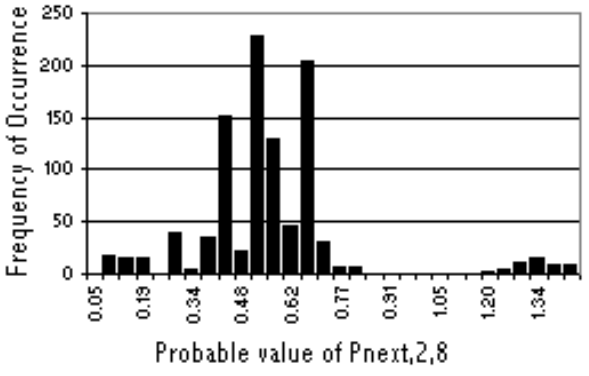

Fig. 10. Graph of the distribution for $p_{\text{next},2.8}$ due to uncertainties in parameters and input data.

most sensitive to changes in input μ_8 (best value of μ using eight measured values). The statistical data and the distribution plot for $p_{\text{next},2.8}$ are illustrated in Table 6 and Fig. 10. The most probable value for $p_{\text{next},2.8}$ is 0.52, but the range to half the peak value indicates $p_{\text{next},2.8} = 0.52\ \binom{+0.18}{-0.12}$, thereby providing some measure of the influence of uncertain data values and uncertain parameters on the predictability of the next event.

The programs as built include only 10 measurement value inputs. More data can be included in the programs but it was felt that 10 data points more than sufficed to illustrate the principles involved. In addition, in any future extension of the programs one should increase the number of classes for data points so that one can incorporate more information than that values lie either above or below I_{cut}.

In order to improve on the procedure itself for assessing the probability of predicting future events one must move from a user-intensive, and limited, manual input of parameter estimates to procedures that provide systematic updating based on priors. In the next paper we discuss the Bayesian update method and the systematic parameter determination procedure as relevant improvements of predictability for which basic mathematical developments are given in Appendix A and Appendix B.

The relevance of the prediction procedure for individual sources of high energy photons is that it enables one to sort out whether a source is truly weak or whether one is merely observing on the edge of a focused beam for a strong source. In addition, because the generation of pulses of high energy photons from individual objects appears to be somewhat random in time, the present procedure allows one to at least estimate from a given number of observed pulses a minimum number one is likely to observe.

One should be careful not to apply the procedure to all high energy photon sources simultaneously because then one has no idea of the mixture of long-lived and short-lived objects for which one is attempting to apply statistics. The assumptions of independence of pulses and non-recurring periodic nature from an individual source are the basic linchpins of the procedure and should not be violated. But within that framework, one has now available a set of procedures that enable one to provide assessments of future chances of observing events from a source and of determining what percentage are likely to be weak or strong intrinsically.

Table 6. Correlation coefficients (sensitivity data) for the probability that the next event is in class 2 with $M = 8$, $p_{\text{next},2.8}$, organized from lowest to highest in absolute value. The variables μ_i and a_i are the best μ and a parameters for i measured values M_i ($i = 1, ..., 10$).

Class Mark	Freq.	Ac. Freq.	Freq. in %	Ac.Freq. in %
0.05	1	1	0.10	0.10
0.10	17	18	1.70	1.80
0.15	15	33	1.50	3.30
0.19	15	48	1.50	4.80
0.24	1	49	0.10	4.90
0.29	39	88	3.90	8.80
0.34	4	92	0.40	9.20
0.39	36	128	3.60	12.80
0.43	152	280	15.20	28.00
0.48	22	302	2.20	30.20
0.53	227	529	22.70	52.90
0.58	129	658	12.90	65.80
0.62	47	705	4.70	70.50
0.67	203	908	20.30	90.80
0.72	30	938	3.00	93.80
0.77	6	944	0.60	94.40
0.82	6	950	0.60	95.00
0.86	0	950	0.00	95.00
0.91	0	950	0.00	95.00
0.96	0	950	0.00	95.00
1.01	0	950	0.00	95.00
1.05	0	950	0.00	95.00
1.10	0	950	0.00	95.00
1.15	0	950	0.00	95.00
1.20	2	952	0.20	95.20
1.25	4	956	0.40	95.60
1.29	12	968	1.20	96.80
1.34	15	983	1.50	98.30
1.39	8	991	0.80	99.10
1.44	9	1000	0.90	100.00

Appendix A Bayesian updating

Consider that one has used the first k observed events ($M > k \geq i \geq 1$) and constructed the probability distribution $p(k; N, \mu, a)$ using Eq. (18) with $M = k$. Now construct a Monte Carlo suite of runs on $N(\geq M)$, $\mu(0 \leq \mu \leq 1)$ and $a(> 0)$, recording the results $p(k; \mathbf{\Omega}_\alpha)$ where $\mathbf{\Omega}_\alpha$ is the vector (N, μ, a) for the α^{th} Monte Carlo run. One is free to choose the underlying distributions for each of N, μ and

a from which individual values are chosen for the Monte Carlo runs, subject only to $N \geq M$, $0 \leq \mu \leq 1$ and $a > 0$. Let the joint probability distribution for these underlying choices be $P_k(\mathbf{\Omega}_\alpha)$. Now suppose one is interested in adding the next observed event, $(k+1)$. Bayes theorem (Jaynes, 1978; Lerche, 1997; Harbaugh et al., 1977; Feller, 1968) then states that for the given set $\{\Omega_\alpha\}$ of Monte Carlo run parameters chosen the conditional probability $P_{k+1}(\mathbf{\Omega}_\alpha)$ that one knows event $(k+1)$ is provided through

$$P_{k+1}(\mathbf{\Omega}_\alpha) = \frac{P_k(\mathbf{\Omega}_\alpha)\, p(k+1;\Omega_\alpha)}{\sum_\alpha P_k(\Omega_\alpha)\, p(k+1;\Omega_\alpha)} \tag{A1}$$

Thus one has updated the probability distribution of the $\mathbf{\Omega}_\alpha$ so that one is progressively changing the probability that a given $\mathbf{\Omega}_\alpha$ triad $(N_\alpha, \mu_\alpha, a_\alpha)$ will more closely honor the observed events. Note from Eq. (18), that

$$\begin{aligned} p(k+1;\mathbf{\Omega}_\alpha) &= p(k;\mathbf{\Omega}_\alpha)\times \\ &\frac{[k_{k+1,1} r^a (N\sin^2\theta - \gamma_{k+1}) + k_{k+1,2} r^a (N\cos^2\theta - \delta_{k+1})]}{r^a (N\sin^2\theta - \gamma_{k+1}) + N\cos^2\theta - \delta_{k+1})} \\ &\equiv p(k;\mathbf{\Omega}_\alpha)\omega(k+1;\Omega_\alpha) \end{aligned} \tag{A2}$$

with

$$\gamma_{k+1} = \gamma_k + k_{k+1;1}, \quad \delta_{k+1} = \delta_k + k_{k+1;2} \tag{A3}$$

so that one merely has to calculate the Monte Carlo suite of values for the bracketed factor in Eq. (A2) once for each event. Bayes updating then proceeds iteratively by adding the next event, $k+2$, and so, in general, one obtains after all M events have been added

$$P_M(\Omega_\alpha) = \frac{P_{M-1}(\Omega_\alpha)\, p(M-1;\Omega_\alpha)}{\sum_\alpha P_{M-1}(\Omega_\alpha)\, p(M-1;\Omega_\alpha)} \tag{A4}$$

and so providing, for each triad $\mathbf{\Omega}_\alpha$ chosen, the probability $P_M(\Omega_\alpha)$ that the set of $\mathbf{\Omega}_\alpha$ comes as close as possible to honoring all the observed events. What the Bayes updating does not do is provide information on whether the set $\{\Omega_\alpha\}$ chosen represents the best possible fit. The point is that a finite suite of Monte Carlo calculations is involved. Thus one has chosen a finite number of values for $(N_\alpha, \mu_\alpha, a_\alpha)$.

The Bayes updating procedure indicates which of these finite number of values has the highest relative probability of honoring all the data but provides no information on non-chosen values of the parameters. Thus the absolute highest probability of honoring all the events may depend on other values than those chosen. One could iterate many times the whole Monte Carlo scheme, and associated Bayesian updating, with different random choices of the parameter triad (N, μ, a) from the underlying distributions of the parameters. In this way one would construct (eventually!) a dense set of parameter values and so identify almost surely the best parameter triad honoring most closely the observed events. However, such a procedure is not only computer intensive but may also be futile. The point is that in constructing the basis probability distribution functions for the Monte Carlo suite operations one has not only to ensure that one honors the physical requirements on $N(\geq M)$, $\mu(0 \leq \mu \leq 1)$ and $a(>0)$ but one also has to choose maximum values for N and a, say $N_{\max}$ and $a_{\max}$. It can happen that the values of N and a needed to satisfy the observed events lie greater than the chosen values $N_{\max}$ and /or $a_{\max}$.

What one needs is a procedure to supplement the Bayes updating that systematically and deterministically will obtain the parameter triad (N, μ, a) that will allow $p_M(M; N, \mu, a)$ to honor the ordered sequence of M observed events, starting with the triad determined from the Bayes updating procedure that has the highest relative probability, but allowing determination of parameter values not chosen in the original Monte Carlo suite of operations. In addition, any such systematic procedure must determine $N_{\max}$ and $a_{\max}$ so that the highest absolute probability for $p(M; N, \mu, a)$ is contained in $N_{\max} \geq N \geq M$ and $a_{\max} \geq a \geq 0$. This aspect of the problem is addressed in Appendix B.

Appendix B Systematic determination of parameter values

Let the triad value $\mathbf{\Omega}_0$ be that with highest relative probability $p(M; \mathbf{\Omega}_0)$ of satisfying all M observed events obtained from Bayesian updating. Now if the situation were to match perfectly with the observed M events then not only would $P(M; \mathbf{\Omega}_0)$ be unity but so, too, would each individual factor probability $p(k; \mathbf{\Omega}_0)$ $(k = 1, ..., M)$ (see Eq. (A2)). To the extent that there is not perfection, $p(k; \mathbf{\Omega}_0)$ will differ from unity. To determine the parameter triad that will match the M observations most closely define

$$\chi^2(\mathbf{\Omega}) = M^{-1} \sum_{k=1}^{M} [1 - p(k;\mathbf{\Omega})]^2 \tag{B1}$$

because, with perfection, $p(k; \mathbf{\Omega}_{\text{best}}) = 1$. Then one wishes to obtain a systematic procedure so that $\chi^2(\mathbf{\Omega})$ is minimized, starting with the Bayesian updated triad $\mathbf{\Omega}_0$, and allowing $N_{\max}$ and $a_{\max}$ to range outside of the values assigned in the Bayesian update method. Such a systematic procedure can be developed as follows. Because $N \geq M$, it is useful to write $N = M10^x$ where $x \geq 0$, and consider x as a basic variable. Then suppose, initially, one takes each component of the vector parameter $\boldsymbol{q} = (\chi, \mu, a)$ to lie in an initial chosen range $q^{(i)}_{\max} \geq q \geq q^{(i)}_{\min}$ where $i = 1, 2$ or 3 according as one handles with χ, μ or a, respectively. Then set

$$\sin^2\theta^{(i)} = \frac{q^{(i)}_{\max} - q^{(i)}}{q^{(i)}_{\max} - q^{(i)}_{\min}} \tag{B2}$$

with the initial value $\sin^2\theta^{(i)}_0 = (q^{(i)}_{\max} - q^{(i)}_0)/(q^{(i)}_{\max} - q^{(i)}_{\min})$ and $q^{(i)}_0$ determined from the Bayesian update procedure. An

iteration scheme (Lerche, 1997) designed to ensure convergence to the smallest value of χ^2 in Eq. (B1) is

$$\theta_{n+1}^{(i)} = \theta_n^{(i)} \exp\left(-\delta_n^{(i)} \frac{\partial\chi^2/\partial\theta_n^{(i)}}{\left|\partial\chi^2/\partial\theta_0^{(i)}\right|}\right) \tag{B3}$$

where $\delta_n^{(i)} = 1$ for $n = 0$ and

$$\delta_n^{(i)} = \frac{\left|\theta_n^{(i)} - \theta_{n-1}^{(i)}\right|}{\left[\tau^{-1}\sum_{j=1}^{\tau}\left|\theta_n^{(i)} - \theta_{n-1}^{(i)}\right|\right]} \quad \text{for} \quad n \neq 0\,;$$

and where the derivatives are given by the approximate evaluation

$$\frac{\partial\chi^2}{\partial\theta_n^{(i)}} = [\chi^2(\theta_n^{(1)}, ..., \theta_n^{(i)}(1+\beta), ...) - \chi^2(\theta_n^{(1)}, ..., \theta_n^{(i)}, ...)]/\beta\theta_n^{(i)} \tag{B4}$$

with $|\beta| \ll 1$. Note that the iteration scheme represented through Eq. (B3) has the following properties:

1. It guarantees that χ^2 will be either smaller or the same after each iteration;
2. It guarantees that all parameters will remain within the bounds $q_{\max}^{(i)} \geq q^{(i)} \geq q_{\min}^{(i)}$ for all iterations;
3. Because of the factor $\delta_n^{(i)}$, the iteration scheme treats first with those parameters that are causing the greatest change in χ^2, while minimizing the influence of other parameters until all parameters are causing essentially the same change in χ^2;
4. If $\sin^2\theta_n^{(i)} \to 1$ (0) then the iteration scheme is informing one that either the chosen lower (upper) value $q_{\min}^{(i)}(q_{\max}^{(i)})$ is too large ($\sin^2\theta_n^{(i)} \to 1$) or too small ($\sin^2\theta_n^{(i)} \to 0$) and must be decreased (increased), thereby providing the necessary information on which direction to change any initial chosen parameter ranges;
5. The scale value $\left|\partial\chi^2/\partial\theta_0^{(i)}\right|$ is best changed after a finite number, Q, of iterations by replacing $\theta_0^{(i)}$ by $\theta_Q^{(i)}$. Pragmatically it is superior in terms of convergence speed to perform Q iterations twice with the update of the scale value after the first Q iterations rather than perform $2Q$ iterations once with the original fixed scale value.

Coupled with the Bayesian update procedure, this systematic method then guarantees one obtains the values of N, μ and a most consistent (smallest χ^2) with the observed ordered event sequence. Once the values for N, μ and a are so determined then one can use Eq. (3) to evaluate the probability that the next event is in class j. As the intensity level, $I_{\rm cut}$, is progressively raised the probability that the next event will lie above $I_{\rm cut}$ is systematically lowered. But, at each level of $I_{\rm cut}$, one obtains estimates of the total number, N, of possible observable events together with the power index, a, as well as $N_1/(N_1+N_2) \equiv \mu^2$. The total number, N, of possible observable events, as well as the observability index a, should likely be independent of the chosen intensity level $I_{\rm cut}$ if the observed number of events, M, is representative of the total N. Numerical implementation of Appendices A and B is considered in the second paper of this series.

Acknowledgements. This work was supported by the DFG under SFB 591 and also by the award of a Mercator Professorship to Ian Lerche. We are most grateful for this support and also for the courtesies shown us by Prof. Reinhard Schlickeiser during our stay in Bochum.

Edited by: H. Fichtner
Reviewed by: two anonymous referees

References

Aldrich, J.: R. A. Fisher and the making of maximum likelihood 1912-1922, Stat. Science, 12, 162–167, 1997.

Arps, J. J., Smith, M. B., and Mortada, J.: Relationship between proved rese and exploratory effort, J. Pet. Technol., 23, 671–675 1971.

Cronquist, C.: Reserves and probabilities?synergism or anachronism?, J. Pet. Technol., 43, 1258–1264, 1991.

Dingus, B. L. and Catelli, J. R.: EGRET Detections of the Highest Energy Emission from Gamma-Ray Bursts in: Abstracts of the 19th Texas Symposium on Relativistic Astrophysics and Cosmology, 63, 1998.

Feller, J.: Elements of probability theory, McGraw-Hill, Englewood Cliffs, 1–2, 1968.

Fisher, R. A.: Theory of statistical estimation, Proc. Cambridge Philos. Soc., 22, 700–725, 1925.

Fisher, R. A.: Two new properties of mathematical likelihood, Proc. Roy. Soc., 144, 285–307, 1934.

Fishman, G. J. and Meegan, C. A.: Gamma-Ray Bursts, Annu. Rev. Astron. Astrophys., 33, 415–458, 1995.

Frail, D. A., Kulkarni, S. R., Sari, R., Djorgovski, S. G., Bloom, J. S., Galama, T. J., Reichart, D. E., Berger, E., Harrison, F. A., Price, P. A., Yost, S. A., and Diercks, A.: Beaming in Gamma-Ray Bursts: Evidence for a Standard Energy Reservoir, Astrophys. J. Lett., 562, L55, 2001.

Harbaugh, J. W., Doveton, J. H., and Davis, J. C.: Probability methods in oil exploration, Wiley, New York, 1977.

Hurley, K., Sari,R., and Djorgovski, S. G.: Cosmic Gamma-Ray Bursts, Their Afterglows, and Their Host Galaxies, astro-ph/0211620, 2002.

Jaynes, E. T.: Where do we stand on maximum entropy? in: The Maximum Entropy Formalism, MIT Press, Cambridge, MA, 15–118, 1978.

Lerche, I.: Geological risk and uncertainty in oil exploration, Academic Press, San Diego, 1997.

Lumley, J. L.: Stochastic tools in turbulence, Academic Press, New York, 1970.

McCray, A. W.: Evaluation of Exploratory Drilling Ventures By Statistical Decision Methods, J. Pet. Technol., 21, 1199–1209, 1969.

Megill, R. E.: Exploration economics, Petroleum Publishing Co., Tulsa, 1971.

Panaitescu, A. and Kumar, P.: Fundamental Physical Parameters of Collimated Gamma-Ray Burst Afterglows, Astrophys. J. Lett., 560, L49, 2001.

Piran, T., Kumar, P., Panaitescu, A., and Piro, L.: The Energy of Long-Duration Gamma-Ray Bursts, Astrophys. J. Lett., 560, 167–169, 2001.

Piran, T.: Gamma-Ray Bursts and the Fireball Model, Phys. Rep., 314, 575–667, 1999.

Piran, T.: Gamma-ray Bursts - A Puzzle Being Resolved, Phys. Rep., 333, 529–553, 2000.

Piran, T.: The physics of gamma-ray bursts, Rev. of modern physics, 76, 1143–1211, 2004.

van Paradijs, J., Kouveliotou, C., and Wijers, R. A. M. J.: Gamma-Ray Burst Afterglows, Annu. Rev. Astron. Astrophys., 38, 379–425, 2000.

Rose, P. R.: Dealing with risk and uncertainty: how can we improve?, AAPG Bull., 71, 1–16, 1987.