Studies in Astronomical Time Series Analysis: V. Bayesian Blocks, A New Method to Analyze Structure in Photon Counting Data

Jeffrey D. Scargle

Space Science Division

National Aeronautics and Space Administration

Ames Research Center

Subject headings: numerical methods – data analysis — models – X-ray astronomy — γ-ray astronomy

Received ________________________; accepted ________________________

Astrophysical Journal

ABSTRACT

I describe a new time-domain algorithm for detecting localized structures (bursts), revealing pulse shapes, and generally characterizing intensity variations. The input is raw counting data, in any of three forms: time-tagged photon events (TTE), binned counts, or time-to-spill (TTS) data. The output is the most probable segmentation of the observation into time intervals during which the photon arrival rate is perceptibly constant – *i.e.* has no statistically significant variations. The idea is not that the source is deemed to have this discontinuous, piecewise constant form, rather that such an approximate and generic model is often useful. Since the analysis is based on Bayesian statistics, I call the resulting structures Bayesian Blocks. Unlike most, this method does not stipulate time bins – instead the data determine a piecewise constant representation. Therefore the analysis procedure itself does not impose a lower limit to the time scale on which variability can be detected. Locations, amplitudes, and rise and decay times of pulses within a time series can be estimated, independent of any pulse-shape model – but only if they do not overlap too much, as deconvolution is not incorporated. The Bayesian Blocks method is demonstrated by analyzing pulse structure in BATSE γ-ray data. The MatLab scripts and sample data can be found on the World Wide Web at: `http://george.arc.nasa.gov/~scargle/papers.html`

Contents

1. The Problem: Structure in Photon Counting Data

Tracking a variable object's brightness changes, based on photon counting data, is a fundamental problem in astronomy. For example, the importance of activity of galactic and extragalactic objects on time scales at and below the millisecond range led NASA to design its X-ray and γ-ray observatories to detect individual photons with microsecond timing accuracy.

1.1. Difficulties

Existing methods do not fully and correctly extract the information in photon counts. The scientifically useful information, of course, is buried in the fluctuations inherent in the occurrence of discrete, independent events – *i.e.* photon detections. The shortest time scales are especially vulnerable to information losses. There are at least three reasons for this.

First are the *binning fallacies*. It is widely and incorrectly held that: (1) such data must be binned[1] in order to be analyzed at all, and (2) the bins must be large enough so that there are enough photons in each to provide a good statistical sample. The almost universal practice of binning event data throws away a considerable amount of information and introduces dependency of the results on the sizes and locations of the bins.

A second reason is that many analysts routinely use global methods, in essence averaging over the observation interval or subsegments of it that are sufficiently long to provide a good statistical sample. Power spectra, autocorrelation functions, and histograms are examples. While good for some problems, global methods dilute short bursts or other local signals.

---

[1] *I.e.*, one must divide the observation into equally spaced intervals and count photons within these *bins*.

Incorrect error models are the third source of information loss. It is usually assumed that observational errors are additive and normally distributed (as in χ^2 methods). Counting fluctuations are neither additive nor normal. Indeed, *the nearly ideal Poisson nature of photon detection provides the rare advantage of knowing statistical properties of the noise with great confidence, completeness, and precision.* (Typically the major way in which the data depart from this ideal is through lack of independence. In particular, detectors have a *dead-time* – arrival of a photon momentarily inhibits detection of subsequent photons.)

1.2. Approach

A single, simple idea sparked this development. *The probabilities of the elementary events – photon detection or nondetection – have such a simple but exact specification [equation (15) below] that it ought to be easy to derive an explicit statistical treatment of the total problem.* This led to a new algorithm, based on Bayesian principles, as described in §2 and demonstrated in §3. It exploits the full time resolution of the data, makes explicit use of the correct statistical distribution, avoids arbitrary binning, and operates in the time domain – focusing on local structures. It converts raw photon counts into the most probable piecewise constant representation of brightness as a function of time. This decomposition can provide simple estimates of the width, location, and amplitude of pulses – assuming their overlap is neglectable – and of the background level, without invoking parametric or other explicit pulse-shape models. An excellent overview of Bayesian methods, with an astronomical flavor, is Loredo (1992). Readers unfamiliar with Bayesian time series analysis might consult Sivia (1996), or the overview, with specific discussion of the changepoint problem, in Ò Ruanaidh and Fitzgerald (1996).

Before proceeding, a few comments on basic approach. As is common in astronomy, the following conceptual scheme underlies the data analysis. Some physical process in the astronomical object causes brightness variations. These fluctuations – modified by radiative transfer, viewing geometry, intervening matter, *etc.* – are modeled as an idealized signal, which in turn is compared with one or more physical models of the original dynamical process. Connection with the observed photon stream is made by interpreting the signal as determining a time-variable photon detection probability. Mathematical properties of this function (*e.g.* smoothness or differentiability), correspond to physical properties of the source – some of which are known but others of which are unknown.

In describing this kind of modeling, terms like *pulse*, *burst*, and *shot* have all been used, loosely, to mean more or less the same thing – namely a process that is in some sense *local*, as opposed to *global*, in time. I know of no generally accepted, rigorous definition of any of these terms, but the following notions may be useful. Consider a stochastic process with a continuous power spectrum of a simple functional form and extending over a broad range of time scales. Call this the *global* process. Self-similar or $\frac{1}{f}$ processes are examples [*cf.* Scargle, Steiman-Cameron, Young, Donoho, Crutchfield and Imamura (1993), Abry and Flandrin (1996), Young and Scargle (1996)]. A deterministic component with a line spectrum, such as a periodic signal, may also be present without materially changing the picture. *Bursts*, then, are non-periodic signals, localized in time, that are not part of the global process. That is, the spectrum of the total signal is altered by the presence of the bursts and is not of the simple form postulated for the global signal. Bursts can occur randomly, periodically, or in any other fashion. In this picture, whether or not a statistical ensemble of signal features is deemed to be bursts depends on the events' shapes, distribution, and relation to the global signal.

This distinction between global and local signals cannot always be made cleanly. For example, *intermittency* in a chaotic nonlinear dynamical system [*e.g.* Schuster (1988)] is in a sense localized, but is described by the same laws of motion which govern the chaotic behavior of the system. Furthermore it is obvious that, in the presence of noise, bursts can be detected only statistically.

The approach adopted here, using what statisticians call *change-point determination*, addresses part of this definitional problem head-on, as it is based directly on the statistical significance of putative local structure. On the other hand, distributions of the times, amplitudes, or shapes of pulses are not considered here; these would be concerns of a follow-on study, after Bayesian block analysis of the full time series.

1.3. Other Work

It has long been recognized that Bayesian methods are well-suited to finding changepoints [Smith (1975), Worsley (1986)]. A Bayesian analysis of Poisson data similar in spirit to the present work is Raftery and Akman (1986); see also Appendix C of Gregory and Loredo (1986). West and Ogden (1997) use methods similar to those described here to find changepoints in binned data, to an accuracy better than the bin size. (Their solutions are simultaneous maximum likelihood in the rates and changepoint location; the rate marginalization carried out here is probably preferable.) Sugiura and Ogden (1997) discuss detection of gradual, linear trends, rather than sudden changes.

Localized basis functions, such as wavelets, provide a partial solution to this problem [Abry, Goncalves and Flandrin (1995), Scargle (1997), Brillinger (1977)]. And the procedure described in Kolaczyk (1997) is somewhat related to the present approach; his segmentations are the standard dyadic intervals of wavelets, whereas here the intervals adapt themselves to the data and are therefore not generally evenly spaced. Donoho (1994) studied edge location in, and multi-segmented analysis of, time series. His methods, *segmentation pursuit* and *minimum entropy segmentation*, circumvent the fixed location of conventional wavelet methods, for a more general statistical model that that used here. Translation invariant wavelet transforms (Coifman and Donoho (1995)) also have potential for accurate location of changepoints.

Abry and Flandrin (1996) discuss the other side of the coin from the topic of the current paper, namely *long-range dependence* in point processes (the statistical term for event data, such as photon counting), using wavelet methods. Recent work has applied wavelets and wavelet denoising to the changepoint problem – see Ogden (1996), Ogden and Parzen (1996a), Ogden and Parzen (1997b).

I have recently become aware of the following work, closely related to this problem: Stark, Fitzgerald and Hladky (1997), Gustafsson (1998a), Gustafsson (1998b).

2. The Analysis Method: Bayesian Blocks

This section details a new algorithm implementing a Bayesian approach to the problem of detecting variability in photon counting data. A sketch of standard Bayesian model fitting will set notation and the context. We have some data D, and a model $\mathcal{M}$ containing a parameter θ. If there are several parameters, simply interpret θ as a vector. We want to estimate how probable it is that the model is correct, and we want to learn something about likely values of the parameter – all based on the data and any prior information that we might have.

The basic relation quantifying parameter inference is Bayes' Theorem, one form of which is

$$P(\theta|D,\mathcal{M})P(D|\mathcal{M}) = P(D|\theta,\mathcal{M})P(\theta|\mathcal{M}). \tag{1}$$

In order, the conventional names of the factors are the *posterior probability density* of θ, given the data, and the *prior predictive probability* for the data, on the left side; and the *likelihood* for the parameter, and the *prior* probability of the parameter, on the right side. These factors have other names to connote different emphasis; *e.g.* $P(D|\mathcal{M})$ is sometimes called the *global or marginal likelihood for the model.* Also, as described by Jaynes (1997), $P(D|\theta, \mathcal{M})$ is termed the likelihood when emphasizing its dependence on θ, but as the *sampling distribution* when emphasizing its dependence on D. All of the terms are to be interpreted *given the model*; this is the meaning of $\mathcal{M}$ behind the "|". The two sides of this equation are simply different ways of reckoning the probability of the same compound event, *i.e.* the model parameter having a specific value and the data being as observed. Standard practice is to write $P(D|\mathcal{M})$ as a divisor on the right-hand side, as this is the way Bayes' theorem is actually used: $P(\theta|D, \mathcal{M})$ is the probability distribution of the parameter and serves the role of quantifying the model's "goodness of fit" to the data.

2.1. Comparison of Alternative Models

A key tool is a procedure to decide which of two (or more) alternative models of a given chunk of data is more probable. This selection is based on those data plus any prior information on the relative likelihood of the models. *E.g.*, we might want to choose between the following two models of an astronomical light curve, based on observations over a time interval T [2]:

- $\mathcal{M}_1$: constant intensity over T
- $\mathcal{M}_2$: possibly different constant intensities in two sub-intervals, $T_1 + T_2 = T$

As will become apparent, this example is at the heart of the method proposed here.

[2]In what follows we use this symbol for both a time interval and its length; this should not cause confusion.

Consider a set of K models, say $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, ..., $\mathcal{M}_K$. (By $\mathcal{M}_k$ we mean the model without specification of any parameter values, so the terms model *class* or *structure* are better.) That we are limiting consideration to this set, plus all other relevant knowledge or assumptions, together comprise a background of information, conventionally denoted I. Bayes' theorem for model selection – as opposed to parameter estimation as in equation (1) – gives for the posterior probability of each model, given the data D and the background information I,

$$P(\mathcal{M}_k|D, I) = \frac{P(D|\mathcal{M}_k, I)P(\mathcal{M}_k|I)}{P(D|I)} \tag{2}$$

Since I does not change, and is therefore irrelevant in comparisons of the kind considered here, we often omit the symbol; its presence should be assumed in all equations derived from Bayes' theorem, including equation (1).

Equation (2) immediately gives a comparison of how well two models represent the data, in terms of the *odds ratio*

$$\frac{P(\mathcal{M}_k|D)}{P(\mathcal{M}_j|D)} = \frac{P(D|\mathcal{M}_k)P(\mathcal{M}_k)}{P(D|\mathcal{M}_j)P(\mathcal{M}_j)} \tag{3}$$

Note that $P(D|I)$ – the probability of observing the data, without regard to the model – is irrelevant to comparison of model classes and accordingly cancels out.

The quantity $P(D|\mathcal{M}_k)$, the probability of the data given the model, can be found by integrating equation (1) over θ, making use of the fact that $P(\theta_k|D, \mathcal{M}_k)$ is normalized:

$$P(D|\mathcal{M}_k) = \int P(D|\theta_k, \mathcal{M}_k)P(\theta_k|\mathcal{M}_k)d\theta_k, \tag{4}$$

The number and significance of the parameters may be different from model to model – hence the subscript on θ_k. Thus equation (3) becomes

$$\frac{P(\mathcal{M}_k|D)}{P(\mathcal{M}_j|D)} = \frac{\int P(D|\theta_k, \mathcal{M}_k)P(\theta_k|\mathcal{M}_k)d\theta_k}{\int P(D|\theta_j, \mathcal{M}_j)P(\theta_j|\mathcal{M}_j)d\theta_j} \; \frac{P(\mathcal{M}_k)}{P(\mathcal{M}_j)} \tag{5}$$

From this equation it is clear that

$$J(\mathcal{M}_k, D) \equiv P(\mathcal{M}_k) \int P(D|\theta_k, \mathcal{M}_k)P(\theta_k|\mathcal{M}_k)d\theta_k \tag{6}$$

is the fundamental quantity to be used in comparing models (J for joint probability for the model and the data). This factor includes prior information, and is independent of the number of, or values of, any model parameters. The model with the largest J value is the most likely to be correct. The integral on the right side of equations (4) and (6),

$$\mathcal{L}(\mathcal{M}_k, D) = \int P(D|\theta_k, \mathcal{M}_k) P(\theta_k|\mathcal{M}_k) d\theta_k \tag{7}$$

is often called the *global likelihood*, or sometimes the *marginal likelihood* or the *evidence* for the model.

It is the essence of the problems considered here that we are ignorant about the different model structures prior to analyzing the data. Accordingly the model priors $P(\mathcal{M}_k)$ (not to be confused with priors for the parameter) could all be taken equal and omitted from expressions for the global likelihood. However, for practical reasons it is useful to retain the *prior odds ratio*

$$\rho = \frac{P(\mathcal{M}_k)}{P(\mathcal{M}_j)} \tag{8}$$

as a scalar parameter of the computations. In the sample applications described in §3 below this quantity is used to suppress spurious blocks due to the statistical fluctuations.

Note that the complexities of the models, *e.g.* the number of parameters, are automatically accounted for in this comparison. Adding parameters to a model almost always increases its maximum likelihood (rigorously, never decreases it). But as is well known, the best model is not the most complex one. Some modeling techniques introduce a penalty factor that compensates for the added degrees of freedom represented by a more complex model. Here, as usual in Bayesian analysis, *this tradeoff between goodness of fit and model complexity is an automatic consequence of the integration over all model parameters* in equation (7). Sometimes in Bayesian analyses such a penalty factor is isolated and called the *Ockham factor*. Jaynes (1997) has a nice discussion of this issue; see also Chapter 4 of Sivia (1996).

2.2. Evidence for a Constant Poisson Rate Model

Now let's use equation (7) to compute the global likelihood for a source being of constant intensity during a given observation interval. The *Poisson process* is the mathematical model of such a source, with $\lambda \geq 0$ denoting the rate, here in photons per unit time, assumed to be constant over some time interval T. That is, the photon events are distributed identically, independently of each other, and with uniform probability over T at rate λ per unit time. Think of drawing a random integer from the Poisson counting distribution with mean λT and then throwing this number of darts randomly and uniformly across the interval. It is well known that this process has no memory or after effect in waiting-times: the arrival of a photon does not affect the probability of subsequent photon arrivals. This property implies that waiting times have an exponential distribution [Billingsley (1986), §23]. The Poisson model therefore has zero dead-time.

We actually use the discrete-time version, the *Poisson counting process* (PCP). That is, the observation interval is divided into a number of equal, fixed subintervals of length δt, and k – the number of counts in such an interval – is Poisson distributed:

$$P(k|PCP,\Lambda) = \frac{\Lambda^k e^{-\Lambda}}{k!}, \tag{9}$$

with parameter

$$\Lambda \equiv \lambda \delta t. \tag{10}$$

[Note: the count rate is expressed either per unit time, with λ (dimension is s^{-1}), or per interval with Λ (dimensionless).] Throughout, we assume that the arrival of a photon in any interval is independent of that in any other non overlapping interval; *i.e.*, the joint probability distribution of the random variables describing photon arrival in the two disjoint intervals is the product of the individual distributions. (Do not confuse the photon detection process with the possibly random process describing the source intensity as a function of time – which is typically correlated from one time to the next. See page 99 of Brillinger (1978) for a discussion of this issue, known as *doubly stochastic processes.*) We make considerable use of the fact that an event probability in an interval is the product of the probabilities in its subintervals.

2.2.1. Time-Tagged Event (TTE) Data

The recording mode called *event* or *time-tagged* data is common in X-ray and γ-ray astronomy, and capable of the highest time resolution. In this mode the detection times of individual photons are recorded. In principle, the raw data consist of a set of N *photon arrival times*

$$D_{TTE} : \{t_n, \quad n = 1, 2, 3, \ldots, N\} \tag{11}$$

over the range of times during which the instrument was active. See Brillinger (1978) for a discussion of this kind of process, consisting of discrete events – called *point processes* in the statistics literature.

In practice, of course, these times are recorded with small but finite resolution – the photons are assigned to narrow bins, as described in connection with equation (9). However, in most data systems there are two reasons for not thinking of these as ordinary bins: First, the time interval is very short (for BATSE $\delta t = 2\mu$-sec) compared to time scales of astrophysical interest. Second, the actual number of photons in the interval is not recorded – just whether one or more photons have arrived.[3] These considerations justify our thinking of this analysis as *bin free* and calling the intervals "ticks," by analogy to a digital clock, instead of bins.

[3]On the other hand, some systems (including BATSE and RXTE) have several detectors operating essentially independently and simultaneously, and photons from different detectors can be recorded with the same time stamp. I ignore these complications.

We introduce an integer time index m through

$$t_m = m \ \delta t, \tag{12}$$

where for an observation of duration $T = M\delta t, m = 1, 2, 3, \ldots, M$. The data consist of a set of N indices, one for each photon:

$$D_{TTE} : \{m_n, \quad n = 1, 2, 3, \ldots, N\}, \tag{13}$$

meaning that photon n is detected at time $m_n \ \delta t$.

A third way to represent these data, fully equivalent to the two above, is in terms of the observable X_m defined by

$$D_{TTE} : X_m = \begin{cases} 0 & \text{no photons during tick } m \\ 1 & \text{otherwise} \end{cases} \tag{14}$$

The probabilities of these values are

$$\begin{aligned} P(X_m = 0|\Lambda) &= p_0 \equiv e^{-\lambda\delta t} = e^{-\Lambda} \\ P(X_m = 1|\Lambda) &= p_1 \equiv 1 - p_0 \end{aligned} \tag{15}$$

Strictly speaking p_1, since it is the probability of one *or more* photons, is not proportional to the Poisson rate parameter. However, since this parameter is small – typically ≈ 0.01 counts per tick, or less – we have

$$p_1 = 1 - e^{-\lambda\delta t} \approx \lambda\delta t \equiv \Lambda \tag{16}$$

This approximation is useful at a few points, but the main analysis does not depend on it. Technically the above conditions define a *finite Bernoulli lattice process* Stoyan, Kendall, and Mecke (1995), since X takes on one of two possible values over a finite range of discrete times. Here I nevertheless follow common usage in referring to this as a Poisson process.

By the independence assumption discussed above, the joint probability of all the events X_m is just the product of the probabilities of the individual events. That is to say, defining $\mathcal{M}_1(\Lambda, T)$ as the Poisson process over interval T, with rate Λ per tick, we have

$$P[D_{TTE}|\mathcal{M}_1(\Lambda, T)] = \prod_{m=1}^{M} P(X_m|\Lambda) = p_1^N(1-p_1)^{M-N} \tag{17}$$

since N ticks contain a photon and the remaining $M - N$ do not. This probability is maximized at $p_1 = \frac{N}{M}$, and equation (15) gives as the most probable rate

$$\lambda = -\frac{1}{\delta t} log(1 - \frac{N}{M}), \tag{18}$$

which in the approximation of equation (16) reduces to

$$\lambda = \frac{1}{\delta t}\frac{N}{M}. \tag{19}$$

In view of the form of equation (17) we now switch from Λ to p_1 as the model parameter, to simplify the analysis. Furthermore, this change motivates selection of the following prior distribution:

$$P(p_1|\mathcal{M}_1) = \begin{cases} 1 & \text{for } 0 \leq p_1 \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

This normalized prior $[\int P(p_1|\mathcal{M}_1)dp_1 = 1]$ assigns probability uniformly to all physically realizable values. It is therefore less arbitrary than some priors adopted in Bayesian statistics, and we adopt it here in preference to alternatives, such as uniform in Λ or with cutoffs corresponding to some sort of *a priori* upper or lower limits on counting rates.

To evaluate the global likelihood in equation (7), multiply the likelihood in (17) by the above prior and integrate

$$\int P[D_{TTE}|\mathcal{M}_1(p_1)]P(p_1|\mathcal{M}_1)dp_1 = \int_0^1 p_1^N(1-p_1)^{M-N}dp_1 = B(N+1, M-N+1) \tag{21}$$

where the *beta function* B can be written in terms of the gamma function [Jeffrey (1995), §11.1.7]:

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} \tag{22}$$

In summary, the global likelihood for the single rate model is this simple function:

$$\mathcal{L}(\mathcal{M}_1|D_{TTE}) = \frac{\Gamma(N+1)\Gamma(M-N+1)}{\Gamma(M+2)} = \frac{N!(M-N)!}{(M+1)!}. \tag{23}$$

It may seem peculiar that this likelihood for a constant rate depends not at all on the distribution of the photon times within the interval – but on only the length of the interval and the number of photons in it. This quantity measures the likelihood of a single-rate model only when compared with the analogous quantity for another model class. This relationship is detailed in §2.4 where a single-rate, unsegmented model is compared with a two-rate, segmented model for the same data.

Note: had we used the probabilities from the truncated Poisson distribution – $e^{-\Lambda}$ and $\Lambda e^{-\Lambda}$ for zero and one photon, respectively – we would have arrived at

$$\mathcal{L}(\mathcal{M}_1|D_{TTE}) = \frac{\Gamma(N+1)}{(M+1)^{N+1}}, \tag{24}$$

a result obtained by Raftery and Akman (1986) – and applied to a study of the intervals between coal-mining disasters – but with a prior somewhat different from ours. In fact, equations (23) and (24) give very similar values, which may be taken as evidence that details of the prior do not matter very much. Equation (23) will be used here.

2.2.2. Binned Data

Sometimes the data are pre-binned into M evenly spaced intervals:

$$D_{BIN} : \{X_m, \quad m = 1, 2, \ldots, M\}, \tag{25}$$

where the integer X_m is the number of photons detected during the m-th such time interval. Taking the rate per bin to be constant, say Λ, the counts in a given bin obey Poisson statistics for this rate:

$$P(X_m|\Lambda) = \frac{\Lambda^{X_m} e^{-\Lambda}}{X_m!} \tag{26}$$

Independence of the counts X_m yields for the likelihood:

$$P[D_{BIN}|\mathcal{M}_1(\Lambda)] = \prod_{m=1}^{M} \frac{\Lambda^{X_m} e^{-\Lambda}}{X_m!} = \frac{\Lambda^N e^{-M\Lambda}}{\prod_{m=1}^{M} X_m!} \tag{27}$$

where $N = \sum_{m=1}^{M} X_m$ is the total number of photons. The maximum of this probability occurs at the same value given in equation (19).

Note that the denominator in equation (27) has the property that its value for an interval is just the product of its value for two or more subintervals. Hence this factor cancels out in a comparison of segmented *vs.* unsegmented versions of a given model, and we omit it. With Λ as the parameter, we adopt the nonuniform but normalized prior

$$P(\Lambda|\mathcal{M}_1) = \begin{cases} (1 - e^{-C})e^{-\Lambda} & 0 \leq \Lambda \leq C \\ 0 & \Lambda < 0 \text{ or } \Lambda > C \end{cases} \tag{28}$$

This prior, while nonuniform in Λ, corresponds to the same uniform p_0-distribution used in the TTE case. It is a special case of the Gamma distribution (power law times exponential) commonly used in Bayesian inference with the Poisson distribution [O'Hagan (1994), Lee (1997)]. This particular form reflects the prior belief that the rate is unlikely to exceed a specific, if approximate, value set by instrumental considerations (which in turn may be guided in the instrument design phase by the maximum expected source brightness). For example, C might be reckoned as roughly the bin interval divided by the instrument deadtime.

Integrating the above likelihood times this prior, and – absent a preferred value of C – taking the limit $C \to \infty$ (*i.e.* allowing bin counts to have any positive value) gives:

$$\mathcal{L}(\mathcal{M}_1|D_{BIN}) = \int_0^1 \Lambda^N e^{-(M+1)\Lambda} d\Lambda = \frac{\Gamma(N+1)}{(M+1)^{N+1}}, \tag{29}$$

curiously identical to equation (24).

2.2.3. Time-to-Spill Data

The last data mode considered is called *time-to-spill* (TTS). To reduce the telemetry data rate, only every S-th photon is recorded, where S is an integer[4] (typically 64 for the BATSE TTS mode):

$$D_{TTS} : \{\tau_n, \quad n = 1, 2, \ldots, N-1\}, \tag{30}$$

where τ_n is the interval between the n-th and the $n+1$-th spill events. It is well known that the distribution of such intervals is given by the gamma, or Erlang, distribution [Billingsley (1986), Haight (1967)]:

$$P(\tau_n|\Lambda) = \frac{\Lambda^S}{\Gamma(S)} \tau_n^{S-1} e^{-\Lambda \tau_n} \tag{31}$$

The usual independence assumption yields:

$$P[D_{TTS}|\mathcal{M}_1(\Lambda)] = [\frac{\Lambda^S}{\Gamma(S)}]^{N-1} (\prod_{n=1}^{N-1} \tau_n)^{S-1} e^{-\Lambda M}, \tag{32}$$

where $M = \sum_{n=1}^{N-1} \tau_n$ is the total length of the interval. As expected, this probability is maximum at

$$\Lambda = \frac{SN}{M}. \tag{33}$$

Equation (32), integrated with the same prior in equation (28), and again taking the limit $C \rightarrow \infty$, gives

$$\mathcal{L}(\mathcal{M}_1|D_{TTS}) = \frac{(\prod_{n=1}^{N-1} \tau_n)^{S-1}}{\Gamma(S)^{N-1}} \frac{\Gamma[S(N-1)+1]}{(M+1)^{S(N-1)+1}}. \tag{34}$$

Note that the interpretation of τ in terms of the true photon rate involves the same issue raised in the TTE case: because of detector dead-time, accumulation of S detector counts occurs at a slightly lower rate than does arrival of S photons. In practice the corresponding corrections can usually be ignored [*cf.* equations (18) and (19)].

---

[4]The data-descriptive constant S is not to be confused with a model parameter.

2.3. Evidence for a Segmented Poisson Rate Model

The previous section yielded estimates of the relative probabilities of the simplest model – namely the single constant-rate Poisson $\mathcal{M}_1(\Lambda)$ – for TTE, binned, and TTS data in equations (23), (29) and (34), respectively. These global likelihoods depend on only N and M, so we denote them as

$$\mathcal{L}(\mathcal{M}_1|D) = \phi_D(N, M) \tag{35}$$

where D here denotes the datatype, and where

$$\phi_{TTE}(N, M) = \frac{\Gamma(N+1)\Gamma(M-N+1)}{\Gamma(M+2)} \tag{36}$$

$$\phi_{BIN}(N, M) = \frac{\Gamma(N+1)}{(M+1)^{N+1}} \tag{37}$$

and

$$\phi_{TTS}(N, M) = \frac{(\prod_{n=1}^{N-1} \tau_n)^{S-1}}{\Gamma(S)^{N-1}} \frac{\Gamma[S(N-1)+1]}{M^{S(N-1)+1}} \tag{38}$$

These results will now be used to estimate the model in which the observation interval is broken into two sub-intervals over which the rates are assumed to be constant but different. (*Cf.* the example at the beginning of §2.1.) In the statistics literature, the point separating such segments is called a *changepoint* in the time series, because the underlying process changes abruptly there. Denote the two-segment model with constant Poisson rates $\mathcal{M}_2(\Lambda_1, \Lambda_2, t_{cp})$, where t_{cp} denotes the *changepoint* – *i.e.*, the time at which the rate switches from Λ_1 to Λ_2. In the notation of §2.1, the full interval T is partitioned into two intervals, T_1 and T_2, containing the times less than and greater than t_{cp}, respectively.

The probability of the compound model is, by the same independence assumption discussed above, just the product of the probabilities of the two segments considered separately:

$$P[D(T)|\mathcal{M}_2(\Lambda_1, \Lambda_2, t_{cp})] = P[D_1|\mathcal{M}_1(\Lambda_1, T_1)]P[D_2|\mathcal{M}_1(\Lambda_2, T_2)] \tag{39}$$

where D_1 is the data in interval 1, *etc.* Thus the global likelihood for the two-rate model is

$$\mathcal{L}(\mathcal{M}_2|D) = \int dt_{cp} \int d\Lambda_1 \int d\Lambda_2 P_{cp}(t_{cp}) P[D_1|\mathcal{M}_1(\Lambda_1, T_1)] P_\Lambda(\Lambda_1) P[D_2|\mathcal{M}_1(\Lambda_2, T_2)] P_\Lambda(\Lambda_2), \tag{40}$$

where P_Λ is the rate prior appropriate to the data type, and P_{cp} is the changepoint time prior. Note that the variables Λ_1, Λ_2, and t_{cp}, which are essentially nuisance parameters here, are integrated out, and the likelihood is therefore independent of them.

Consider now the TTE case. For actual data, time is discrete (*cf.* §2.2.1), so the integral in equation (40) is a sum and we denote the changepoint location by the integer m_{cp}. One could consider jumps at arbitrary clock times, m, but it simplifies the procedure to test for possible changepoints only at the arrival of an actual photon. Thus we parametrize the changepoint as

$$m_{cp} = m_{n_{cp}} \tag{41}$$

for some photon index n_{cp}. This simplification merely ignores the difference between points that identically divide the photons. Further, after carrying out the two Λ integrals, we can write the t_{cp}-integrand (or rather the corresponding discrete time summand) in equation (40) as a simple function of only the changepoint index n_{cp}, through the relations

$$N_1 = n_{cp}, \tag{42}$$

$$N_2 = N - N_1 = N - n_{cp}, \tag{43}$$

$$M_1 = m_{n_{cp}}, \tag{44}$$

and

$$M_2 = M - M_1 = M - m_{n_{cp}}, \tag{45}$$

From the expressions above, and with the definition

$$\psi(n_{cp}) = \phi(N_1, M_1)\phi(N_2, M_2), \tag{46}$$

we have

$$\mathcal{L}(\mathcal{M}_2|D) = \sum_{n_{cp}} \psi(n_{cp}) \Delta t_{n_{cp}} \tag{47}$$

where the factor $\Delta t_{n_{cp}}$ – defined as the time interval between successive photons – corresponds to a prior uniform in m, even though the sum in this equation is over n_{cp}, not m. In fact, the code in Appendix A omits this factor, because it appears to be a small correction in all the cases studied so far. The changepoint parameterization is slightly different for the other data modes; details are omitted.

2.4. Deciding Between Segmented and Unsegmented Model

The idea now is simple: compare the $J(\mathcal{M}_k, D)$ values from equation (6) of the unsegmented, single rate model $\mathcal{M}_1$ and the segmented, two-rate model $\mathcal{M}_2$, in terms of the odds ratio:

$$\mathcal{O}_{21} = \frac{\mathrm{J}(\mathcal{M}_2, D)}{\mathrm{J}(\mathcal{M}_1, D)} \tag{48}$$

This ratio, with the prior odds ratio equal to one, is often called the *Bayes factor*. If this ratio favors a segmented model, it is straightforward to compute from equation (46) the most probable changepoint location from among all possible changepoints. Finally and almost trivially, equation (18) or equation (19) determines the corresponding rates. The appendices contain computer code for all the necessary computations, and the procedure is demonstrated on real data in §3 below.

2.5. Multiple Change Points

As discussed earlier, the overall goal is to find the optimum block decomposition of the data – *i.e.* into a piecewise constant representation. The rigorously correct way to do this would be as follows. Let an arbitrary number, say N_{cp}, of changepoints divide the observation interval into $N_{cp} + 1$ subintervals. Compute the global likelihood, $\mathcal{L}(N_{cp})$ of this multiple changepoint decomposition; the value of N_{cp} which maximizes this quantity is the most probable number of changepoints. It is then a simple matter to find the most probable locations of the change points themselves, and the most probable values of the rates for each of the corresponding segments.

The case $N_{cp} = 2$ is relatively easy. In fact, the corresponding global likelihood – a function of the two changepoint location indices – can be computed with matrix operations that are quite efficient in MatLab. Some thinning of the data is necessary for the cases in which the number of photons is so large that the corresponding $N \times N$ matrix is too big. However, for more than 2 changepoints direct computation quickly becomes impractical. Therefore a simple iterative procedure was adopted as an attempt to approximate multiple changepoint determination. Start with the whole observation interval. Use the above method to decide between not segmenting this interval and segmenting it, with one changepoint, into two subintervals. If the latter is favored by the Bayes factor, apply the same procedure to both of the resulting subintervals. Continue in the same vein, applying the procedure to all new subintervals created at the previous step. That this method works approximately, but not exactly, is indicated by the fact that an algorithm that handles two simultaneous change-points (*i.e.* $N_{cp} = 2$) gave results similar, but not identical, to those obtained with iterative application of the single changepoint algorithm.

What stops this iteration? The obvious halting condition is that the odds ratios favor unsegmented models for all subintervals. Unfortunately this is too simple in practice. In the analysis of large data sets there are typically many computed odds ratios which are greater than 1 by only a small amount. Decisions based on these "coin flips" are wrong about half the time, subdividing many intervals that shouldn't be.

Since these cases tend to be short intervals containing only a few photons, much of the problem can be swept under the rug by imposing a minimum number of photons allowed in subintervals. A second approach is to impose a prior odds ratio that disfavors segmenting – *i.e.*, is biased toward keeping intervals unsegmented unless the odds ratio is strongly in favor of segmenting. (There is a simple argument in support of this second idea: an overall statistical assessment should take into account the number of roughly independent experiments carried out; this is on the order of the largest reasonable number of segmentation points, which in turn is determined by the resolution of the observation interval. This leads to a prior ratio in equation (8) of $\rho \approx \frac{\text{length of data interval}}{\text{desired time resolution}}$. It also has the advantage that it avoids the other idea's bias against short intervals. Unfortunately, this argument probably cannot be justified within the Bayesian formalism. Nevertheless, numerical experiments support the use of one or the other of these ideas.) The best approach may be to combine both, as was done by Bernaola-Galván, Román-Roldán, and Oliver (1996) in a similar segmentation algorithm, based on the *Jensen-Shannon divergence measure* in place of the likelihood, and applied to automatic detection of structure in DNA sequences. Gustafsson (1998b) uses a *stopping rule* based on somewhat different considerations. The code in Appendix B shows one way to carry out iterative segmentation and such a composite halting logic.

3. BATSE γ-ray burst data

This section demonstrates the method just described by applying it to γ-ray data from BATSE. The basic algorithm is employed to determine the detailed structure of pulses, such as are known to make up the time-profiles of many γ-ray bursts [Norris et al. (1996), Scargle, Norris and Bonnell (1997)].

Figure 1 depicts the logarithms of the odds ratios as a function of the position of the changepoint for BATSE data from the burst denoted Trigger 0551. The top panel shows for comparison the binned counts as a function of time (in microseconds). The raw data comprises about 29,000 photons. On the same time axis, the other panels show the logarithms of the odds ratio in equation (48), for TTE, binned, and TTS data, in order, as a function of the location of the changepoint. The binned and TTS data are derived directly from the TTE data. The spill data was constructed simply by sampling every 64-th photon from the TTE data.

Note several things: (1) The actual odds ratios are all astronomically large in favor of segmentation. (2) The most probable changepoint location is indicated with vertical dotted, dashed, and dot-dashed lines. If the actual odds ratios were plotted, this would be an extremely sharp maximum, indicating that there is very little uncertainty in the changepoint location. (3) The TTE and TTS changepoints are very nearly equal – suggesting that this method is rather efficient at extracting information from TTS data, and also that little information is lost in this mode. The fact that the value for the binned data is slightly different is consistent with the expected loss of time-resolution entailed by binning.

Figure 2 shows the result of iterating the segmentation procedure on the same TTE data. The Bayesian blocks are indicated with solid lines. The vertical dotted lines are the locations of pulses determined by a simple pulse finding routine that basically selects statistically significant local maxima; this algorithm will be described in Scargle, Norris and Bonnell (1997).

One can derive properties of the pulses from this block representation. In a separate paper [Scargle, Norris and Bonnell (1997)], this method will be used to determine peak times, amplitudes, and rise and fall times for gamma ray bursts. Specifically, we use the Bayesian Blocks technique to make crude estimates of the locations, amplitudes, and widths of the pulse structures within a burst, without a parametric pulse model and dealing with pulse overlap in a trivial way. The peak time and amplitude are taken as the center and height of the highest block in the pulse, and exponential rise and decay times are estimated using a simple quadrature of the corresponding halves of the burst profile. Then these crude pulses are used as initial guesses for a numerical routine that truly deconvolves overlapping pulses by fitting a parametric model. The initial guess is very important for the convergence of this fitting procedure to the (hopefully) global optimum; results with the Bayesian Blocks have proven very satisfactory. The lowest block provides a good estimate of the constant post-burst background, and will do so as long as the burst ends before the observation terminates.

4. Conclusions

The method developed here is applicable to all the photon event data modes common in high energy astrophysics: time-tagged events, binned counts, and time-to-spill data. The fundamental element of the method is a way to decide whether a single Poisson rate or two different rates is the better model for an interval. This decision is applied iteratively to build up a piecewise constant model of the data. This analysis method imposes no lower limit on the time scale; any such limits are set by the the data themselves.

The Bayesian Blocks method is designed to extract localized signals from counting data where statistical fluctuations are important. It is probably not useful in situations that require lots of time-averaging to extract coherent, global signals such as periodic or quasi-periodic variations.

Future work will include investigation of ways to determine multiple changepoints more rigorously. The principles behind a maximum likelihood determination of the number and location of changepoints is straightforward, and can surely be made computationally feasible. I have recently become aware of work by Chib (1996) developing a Markov chain Monte Carlo procedure for Bayesian estimation of multiple changepoint models that may be applicable to this problem. Phillips and Smith (1996) may also be of relevance.

In addition, it will be useful to extend the methods given here to include variable rates across the blocks, or other departures from the constant-rate model. I have explored both linear and exponentially varying rates. The approach in Sugiura and Ogden (1997) may be useful for this problem. I am pursuing extensions of the basic idea underlying Bayesian Blocks to higher dimensions; in particular spatial structure can be elucidated, and backgrounds removed, from two-dimensional photon counting data with generalizations of the one-dimensional algorithms given here.

It is also relatively easy to extend this methodology to a multivariate context – determination of block structure in pairs of time series in which it is assumed that the segmentation points occur at the same times in the two data series; of course, the rates are not in general the same. This will be particularly useful for BATSE gamma-ray burst data that consists of simultaneous photon counting in four broad energy channels. In this context, it will be useful to allow for, *e.g.* the known fact that there are time delays in the burst structures as a function of photon energy. Similarly, known gaps during which the instrument is not sensitive can be readily handled.

What to do with Bayesian Blocks? This depends on the context. For the pulse problem in γ-ray burst work (§3) we have indicated the use of the blocks to determine pulse attributes, at least in a crude way, without the need to adopt a specific model for pulse shapes. These attributes can in turn be used as starting guesses for further, parametric, nonlinear optimization, as discussed above. It is expected that many different uses can be made of Bayesian block decomposition.

Work is in progress in collaboration with Paul Hertz, Elliott Bloom, Jay Norris and Kent Wood, to use Bayesian Blocks to determine whether short-time-scale structure, or *bursts*, are present in Cygnus X-1. There is a long debate in the literature about the reality and meaning of short (millisecond) bursts in this accretion system. Almost certainly our approach will either detect or place upper limits on bursts, and has the possibility of detecting individual bursts at a high significance level. A different approach to this same problem, also using a Bayesian framework, was presented at a recent meeting of the High Energy Astrophysics Division of the American Astronomical Society [Marsden and Rothschild (1997)].

Note added in press: For TTE data, consider the time-scale transformation $\delta t \rightarrow \frac{1}{\alpha}\delta t, M \rightarrow \alpha M$, for α any integer > 1. This amounts to refining the clock ticks but leaving the photon times unchanged. Under this transformation the estimated block structure must be unaltered: the changepoint times and photon rates will stay fixed (although of course the rates *per tick* will decrease by a factor of α). By considering arbitrarily large α it follows that the asymptotic form (for $M \rightarrow \infty$) of equation (23) can be used without appreciable error. Details of this simplification will be posted on the World Wide Web site referenced in the Appendix and, together with a solution of the multiple changepoint problem using Markov Chain Monte Carlo methods, will be the subject of a future paper.

I am especially grateful to Tom Loredo for encouragement, numerous technical suggestions, and careful reading of several versions of the manuscript. Thanks also to Jay Norris and Jerry Bonnell for initiating the gamma-ray burst research that led to this work and for numerous useful suggestions, and to Eric Kolaczyk, Iain Johnstone, and Peter Cheeseman for statistical advice. I thank Bob Hogan and Mark Showalter, plus members of the SLAC Astrogravity group – Elliott Bloom, Chris Chaput, Daniel Engovatov, Gary Godfrey, Andrew Lee, and Ganya Shabad – for helpful comments and assistance. I am grateful to David Marsden and Rick Rothschild for an advance copy of their paper, and Bill Fitzgerald and Fredrik Gustafsson for helpful comments. This work is supported by grants from NASA's Astrophysics Data Program, the Compton Gamma-ray Observatory Guest Investigator Program, and the NASA-Ames Director's Discretionary Fund. The NASA data shown are from the BATSE instrument on the Compton Gamma Ray

Observatory.

5. Appendix

This appendix contains MatLab[5] code, an array-based data processing system. These MatLab scripts, and sample data allowing the reader to reproduce Figure 2 of this paper, can be found on the World Wide Web at: `http://george.arc.nasa.gov/~scargle/papers.html`. (See Buckheit and Donoho (1995) for a description of the philosophy of publishing scientific research in such a way the the reader can reproduce all results.) Much of this code is similar to that of IDL[6] and other similar software packages for data analysis, and can be considered as pseudocode for the procedure.

A few comments about the MatLab syntax is in order.

The function line at the beginning of each routine identifies the input and output variables. It will be seen that multiple input and output variables are possible; and the input and output variables are arrays (matrices, vectors, or scalars) in general.

The symbols $*$ and / specify matrix multiplication and division, respectively. Overriding the matrix operation in favor of the simple term-by-term operation is indicated by a dot (.) before $*$ or /. The statement `[ a_max, i_max ] = max( x )`, where `x` is a vector, returns both the value of the maximum of the array, and the index, `i_max` at which this maximum is achieved. The function `gammaln` is a built-in function that evaluates the natural logarithm of the gamma function of the argument array.

On any line, everything following the symbol `%` is treated as a comment and not processed. Three dots (...) at the end of a line indicates continuation onto the next line.

---

[5] ©the Mathworks, Inc.

[6] ©Research Systems, Inc.

The command `find` returns the indices of its argument that satisfy the condition specified in the argument. `isempty` is a logical function to determine whether the argument has been defined yet. `reverse` simply reflects an array, and `ceil` and `floor` are rounding of a real number to the next highest and lowest integer, respectively.

The expression `a'` means the matrix transpose of `a`.

5.1. Appendix A: Find A Change Point

This Appendix gives MatLab code for the procedure to find a single change point, as described in §2.4 of the text. The computation is particularly efficient because the evaluation of the global likelihoods can be carried out completely in terms of array operations on the vector containing all the candidate changepoints.

```
function [ change_point, odds_21, log_prob, log_prob_noseg ] = ...
    find_change( photon_times, t_0, t_n )
%
% Find most probable two-rate model for Poisson arrival time data,
% based on Bayesian analysis.
%
% Input:  photon_times -- photon arrival times
%                         (Note: These must be microseconds, not seconds,
%                         because the time values correspond to the
%                         clock rate at which the data are sampled.)
%         t_0          -- time just previous to photon_times(1)
%         t_n          -- time just after last time in photon_times
%
% Output: change_point -- index of "photon_times" which provides the maximum
%                         likelihood segmented model (that is, with one
%                         Poisson rate to the left of
%                               photon_times(change_point)
%                         and another to the right
%              odds_21 -- odds ratio: 2 unequal rates / 1 rate
%             log_prob -- log probability of segmented model, as a
%                         function of changepoint
%       log_prob_noseg -- log prob of nonsegmented model
```

```
%------------------------------------------------------------------------------------------

dt_half = 0.5 * diff( photon_times );
n_ph = length( photon_times ); % Number of photons

min_time = photon_times(    1 );
max_time = photon_times( n_ph );

t_left  = 0.5 * ( t_0 + min_time );
t_right = 0.5 * ( max_time + t_n );

% Number of microsecond "ticks" in the whole (extended) interval:
n_ticks = t_right - t_left + 1;

%------------------------------------------------------------------
%    Evaluate log-probability of the unsegmented model:
%------------------------------------------------------------------

log_prob_noseg = gammaln( n_ph + 1  ) + ...
                 gammaln( n_ticks - n_ph + 1 ) - ...
                 gammaln( n_ticks + 2);

%------------------------------------------------------------------
% Evaluate the log-probability of the segmented model as a
% function of changepoint; find optimum changepoint.
%------------------------------------------------------------------

% Array of trial changepoints:
n_1  = (1: n_ph - 1)';
n_2 = n_ph - n_1;

m_1 = photon_times( n_1 ) + dt_half( n_1 ) - t_left;
m_2 = n_ticks - m_1;

log_prob = - 1.e55 * ones( n_ph, 1 ); % mark all points as invalid
arg_1 = m_1 - n_1 + 1;
arg_2 = m_2 - n_2 + 1;
ii = find( arg_1 > 0 & arg_2 > 0 ); % indices of valid points
log_prob(ii) = gammaln( n_1(ii)+1 ) + gammaln( arg_1(ii) ) - gammaln( m_1(ii)+2 );
log_prob(ii) = log_prob(ii) + gammaln( n_2(ii)+1 ) + gammaln( arg_2(ii) ) - ...
              gammaln( m_2(ii) + 2 );
[ max_log, change_point ] = max( log_prob(ii) );
```

```
%-----------------------------------------------------------------
% Compute odds ratio: prob(seg) / prob(no_seg)
%-----------------------------------------------------------------
odds_21 = sum( exp( log_prob - log_prob_noseg ) );

if ~isfinite( odds_21 )
   odds_21 = 1000000;
end
```

5.2. Appendix B: Make Bayesian Blocks

This Appendix includes MatLab code for the iterative procedure to find a multiple change point, as described in §2.5 of the text.

```
function   [ n_seg_vec, xx_vec ] = make_segments( photon_times )
% function [ n_seg_vec, xx_vec ] = make_segments( photon_times )
%
% Input:  photon_times -- photon arrival times, in microseconds
%
% Output: n_seg_vec  -- array of changepoint times
%         xx_vec     -- count rates (c/usec) in the corresponding segments
%
% Note: t_seg = photon_times( n_seg_vec ) generates the changepoint times
%
% ------------------------------------------------------------------------

global prior_ratio min_photons

n_times = length( photon_times );
  min_time = photon_times(       1 );
  max_time = photon_times( n_times );
    delta_t  = ( max_time - min_time ) / ( n_times - 1 );
    min_tick = floor( min_time - 0.5 * delta_t );
    max_tick =  ceil( max_time + 0.5 * delta_t );
    n_ticks = max_tick - min_tick + 1; % Number of microsecond "ticks"

nseg_1_vec = [ 1 ];
nseg_2_vec = [ n_times ];
nosubs_vec = [ 0 ];
```

```
xx_vec = [ n_times / n_ticks ];
no_seg_flag = 0;

while no_seg_flag == 0

   num_segments = length( nseg_1_vec );
   no_seg_flag = 1; % set escape unless do a sub-segmentation

   nseg_1_work = [];
   nseg_2_work = [];
   nosubs_work = [];
   xx_work  = [];

   for seg_id = 1: num_segments

      do_it = 1; % do this one, unless ...

      % ... this one has been done before!
      if nosubs_vec( seg_id ) == 1
         do_it = 0;
      end

      nseg_1 = nseg_1_vec( seg_id );
      nseg_2 = nseg_2_vec( seg_id );
      x_seg  = xx_vec( seg_id );

      times_this = photon_times( nseg_1: nseg_2 );
      nt_this = length( times_this );

      if do_it > 0

         % Determine previous time
         time_this_1 = times_this(1);

         if time_this_1 == photon_times(1);

            % Special handling for first point in full array,
            %   or if it is the second point, but the first two
            %   (or more) times are equal:
            ii = find( times_this > time_this_1 );
            if isempty(ii)
               delt_t = 2; % Token value
            else
```

```
            delt_t = times_this(ii(1)) - time_this_1;
         end
         t_0 = time_this_1 - delt_t;

      else

         % t_0 is the time just previous to the sub-array
         t_0 = photon_times( nseg_1 - 1 );

      end

      % Determine subsequent time
      time_this_n = times_this(nt_this);

      if time_this_n == photon_times(n_times);

         % Special handling for last point in full array,
         %   or if it is the second-to-last point, but the
         %   last two (or more) times are equal:
         ii = find( times_this < time_this_n );
         if isempty(ii)
            delt_t = 2; % Token value
         else
            delt_t = time_this_n - times_this(ii(length(ii)));
         end
         t_n = time_this_n + delt_t;

      else

         % t_n is the time just after the sub-array
         t_n = photon_times( nseg_2 + 1 );

      end

      [ n_seg, odds_ratio, log_prob ] = find_change( times_this, t_0, t_n );

      % ... one of the proposed subsegments is too short:
      n_seg_right = nt_this - n_seg;
      if (n_seg <= min_photons) | (n_seg_right <= min_photons)
         do_it = 0;
      end

      % ... the significance criterion not met:
```

```
      if odds_ratio < prior_ratio
         do_it = 0;
      end

   end

   if do_it > 0

      % Subsegment this one; do not escape yet
      no_seg_flag = 0;

      ntimes_1_left  = nseg_1;
      ntimes_1_right = nseg_1 + n_seg - 1;

      ntimes_2_left  = nseg_1 + n_seg;
      ntimes_2_right = nseg_2;


      n_ticks_left  = times_this( n_seg   ) - times_this(     1 ) + 1;
      n_ticks_right = times_this( nt_this ) - times_this( n_seg ) + 1;

      nn_left  = n_seg;
      nn_right = nt_this - n_seg;

      x_seg_left  = nn_left  / n_ticks_left;
      x_seg_right = nn_right / n_ticks_right;

      nseg_1_work = [ nseg_1_work ntimes_1_left ntimes_1_right ];
      nseg_2_work = [ nseg_2_work ntimes_2_left ntimes_2_right ];
      xx_work     = [ xx_work  x_seg_left  x_seg_right  ];
      nosubs_work = [ nosubs_work 0 0 ];

   else

      % No sub-segmenting of this segment;
      % so just stuff in the beginning, end, mark
      % as "nosubs" so that it will not be done again
      nseg_1_work = [ nseg_1_work nseg_1 ];
      nseg_2_work = [ nseg_2_work nseg_2 ];
      xx_work  = [ xx_work x_seg ];
      nosubs_work = [ nosubs_work 1 ];

   end
```

```
   end

   % Post the segmentations just done:
   nseg_1_vec = nseg_1_work;
   nseg_2_vec = nseg_2_work;
   xx_vec  = xx_work;
   nosubs_vec = nosubs_work;

end

n_seg_vec = nseg_2_vec;
```

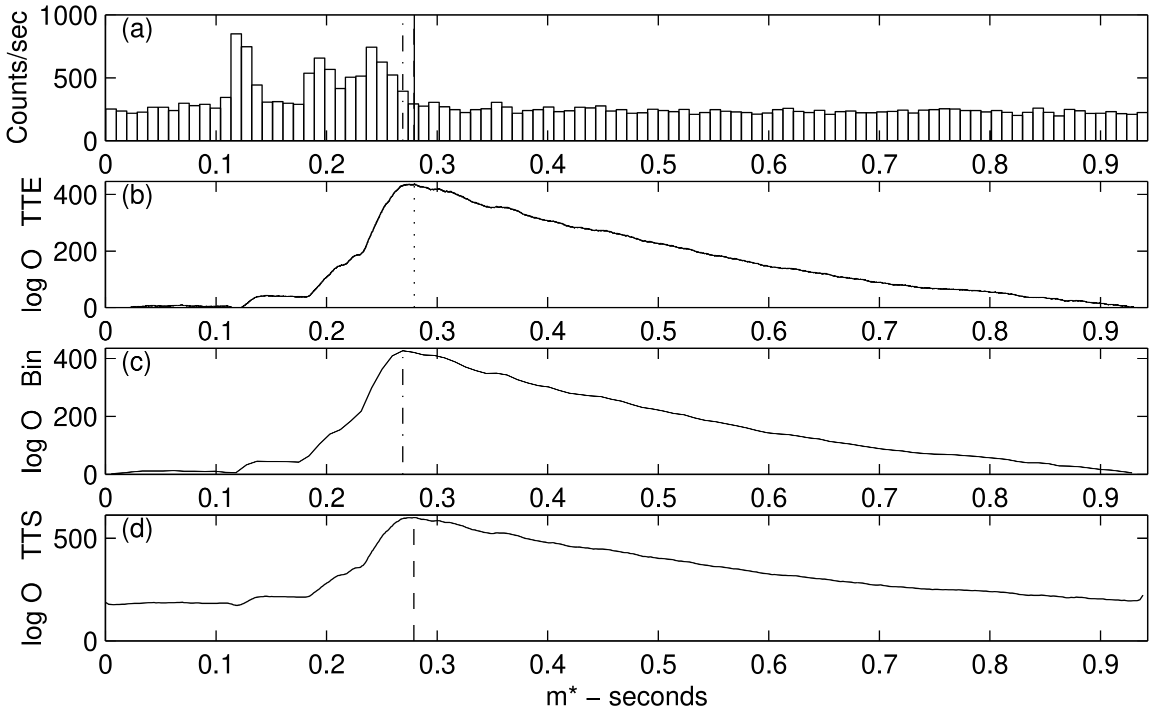

Fig. 1.— Changepoint location in BATSE data for Burst Trigger 0551. (a) Binned counts for comparison: 100 time bins, of width 9.42 ms. (b) For TTE data: log_{10} of the odds ratio in favor of segmentation, as a function of the changepoint location. (c) Same for binned data. (d) Same for TTS data. Vertical lines in all panels are at the maximum odds ratio; in (a) those for TTE and TTS modes are indistinguishable and appear as a solid line.

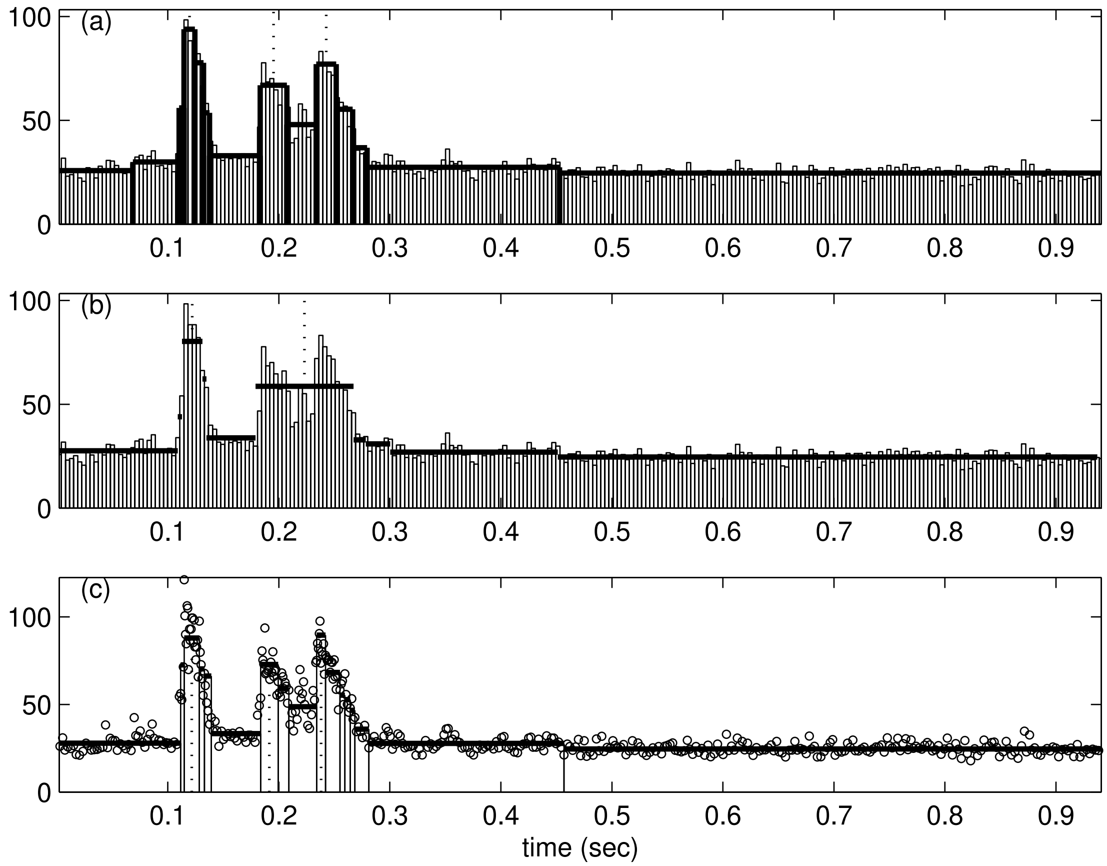

Fig. 2.— Bayesian Blocks for the same data as in Figure 1, determined as explained in the text. (a) TTE data; (b) TTS data; (c) binned data

REFERENCES

Abry, P., & Flandrin, P. 1996, "Point Processes, Long-Range Dependence and Wavelets," in Wavelets in Medicine and Biology, eds. A. Aldroubi & M. Unser, CRC Press, Boca Raton, 413

Abry, P., Goncalves, P., & Flandrin, P. 1995, "Wavelets, spectrum analysis and 1/f processes," in Lecture Notes in Statistics, No. 103, Wavelets and Statistics, eds. A. Antoniadis & G. Oppenheim, Springer-Verlag, 15

Bernaola-Galván, Pedro, Román-Roldán, Ramón, & Oliver, José L. 1996, Phys. Rev. E, 53, 5181

Billingsley, P., 1986, Probability and Measure, John Wiley & Sons, New York

Brillinger, David R., 1997, "Some Wavelet Analyses of Point Process Data," Proc. 31st Asilomar Conference on Signals, Systems and Computers. `http://stat-www.berkeley.edu/users/brill/papers.html`

Brillinger, David R., 1978, "Comparative Aspects of the Study of Ordinary Time Series and of Point Processes," pp. 33-133, Developments in Statistics, Vol. 1, Academic Press, Inc.

Buckheit, J., & Donoho, D., 1995, "WaveLab and Reproducible Research," in Wavelets and Statistics, eds. Antoniadis, A. & Oppenheim, G., Springer Lecture Notes in Statistics, No. 103; see also `http://stat.Stanford.EDU/reports/donoho/` and `http://sepwww.stanford.edu/research/redoc/`

Chib, S., 1996, "Estimation and Comparison of Multiple Change Point Models," Biometrika, in press

Coifman, R. R., & Donoho, D. L. 1995, "Translation-Invariant De-Noising," in Wavelets and Statistics, Anestis Antoniadia & Georges Oppenheim, eds., Springer-Verlag Lecture Notes

Donoho, David L. 1994, "On Minimum Entropy Segmentation," Wavelets: Theory, Algorithms, and Applications, ed. Chui, C.K., Montefusco, L., & Pucciio, L., Academic Press: New York, 233

Gregory, P. C., & Loredo, T. J. 1992, ApJ, 398, 146

Gustafsson, F. 1998a, "Segmentation of signals using piecewise constant linear regression," submitted to IEEE Transactions on Signal Processing.

Gustafsson, F. 1998b, "A Change Detection and Segmentation Toolbox for Matlab," Linköping University Technical Report LiTH-ISY-R-1669 `http://www.control.isy.liu.se/cgi-bin/reports?author~Gustafsson`

Haight, F. A., 1967, Handbook of the Poisson Distribution, John Wiley & Sons, New York

Jaynes, E. T., 1997, Probability Theory: The Logic of Science, available electronically at `http://bayes.wustl.edu`

Jeffrey, A. 1995, Handbook of Mathematical Formulas and Integrals, Academic Press: New York.

Kolaczyk, Eric D., 1997, ApJ, 483, 34

Lee, P., 1997, Bayesian Statistics: An Introduction, John Wiley & Sons

Loredo, T.J., 1992, "Promise of Bayesian Inference for Astrophysics," in Statistical Challenges in Modern Astronomy, ed. Feigelson and Babu, Springer-Verlag: New York, 275

Marsden, D., & Rothschild, R.E. 1997, Detection of Bursts in Time Series Data Using Bayesian Techniques, submitted to Ap. J.

Norris, J. P., Nemiroff, R. J., Bonnell, J. T., Scargle, J. D., Kouveliotou, C., Paciesas, W. S., Meegan, C. A., and Fishman, G. J. 1996, ApJ, 459, 393

Ogden, R. Todd, 1996, Wavelets in Bayesian Change-Point Analysis, American Statistical Association 1996 Proceedings of the Section on Bayesian Statistical Science, pp. 164-169.

Ogden, R. T. and Parzen, E. 1996a, Data Dependent Wavelet Thresholding in Nonparametric Regression with Change-point Applications, Computational Statistics and Data Analysis, 22, 53

Ogden, R. T. and Parzen, E. 1996b, Change-point Approach to Data Analytic Wavelet Thresholding, Statistics and Computing, 6, 93

O'Hagan, A., 1994, Kendall's Advanced Theory of Statistics: Bayesian Inference, Volume 2B, John Wiley & Sons

Ò Ruanaidh, J. J. & Fitzgerald, W. J., 1996, Numerical Bayesian Methods Applied to Signal Processing, Springer: New York

Phillips, D. B., & Smith, A, 1996, "Bayesian Model Comparison via Jump Diffusions" in Markov Chain Monte Carlo in Practice, ed. W. R. Gilks, S. Richardson, & D. J. Spiegelhalter, Chapman & Hall, London (1996), 215

Raftery, A. E., & Akman, V.E. 1986, Biometrika, 73, 85

Scargle, J., 1997, "Wavelet Methods in Astronomical Time Series Analysis," in Applications of Time Series Analysis in Astronomy and Metrology, Chapman & Hall, 226

Scargle, J. D., Norris, J, & Bonnell, J. T., 1997, "Attributes of Gamma-Ray Burst Pulses: I. Short Bursts Analyzed with BATSE Time-Tagged Event Data," in preparation

Scargle, J., Steiman-Cameron, T., Young, K., Donoho, D., Crutchfield, J., & Imamura, J. 1993, ApJ, 411, L91

Schuster, H. 1988, Deterministic Chaos (VCH, New York)

Sivia, D.S. 1996, Data Analysis: A Bayesian Tutorial, Clarendon Press: Oxford

Smith, A. F. M., 1975, Biometrika, 62, 407

Stark, J., Fitzgerald, W., & Hladky, S. 1997, "Multiple-order Markov Chain Monte Carlo Sampling Methods with Application to a Changepoint Model," Technical Report CUED/F-INFENG/TR. 302, http://www2.eng.cam.ac.uk/ jas/pubs.html

Stoyan, D., Kendall, W., and Mecke, J. (1995), Stochastic Geometry and its Applications, 2nd edition, John Wiley & Sons: New York

Sugiura, N. & Ogden, R. T., 1997, Testing Change-points with Linear Trend, Communications in Statistics B: Computation and Simulation, 23, 287

West, R. Webster, and Ogden, R. Todd, 1997, Continuous-time Estimation of a Change-point in a Poisson Process, Journal of Statistical Computation and Simulation, 56, 293

Worsley, K. J. 1986, "Confidence regions and tests for a change-point in a sequence of exponential family random variables," Biometrika, 73, 91

Young, K., & Scargle, J. D. 1996, ApJ, 468, 617

This manuscript was prepared with the AAS LaTeX macros v4.0.