New Horizons in Time-Domain Astronomy
Proceedings IAU Symposium No. 285, 2011
R.E.M. Griffin, R.J. Hanisch & R. Seaman, eds.

doi:10.1017/S1743921312000300

Sines, Steps and Droplets: Semi-parametric Bayesian Modelling of Arrival Time Series

Thomas J. Loredo

Department of Astronomy, Cornell University, Ithaca, New York, USA
email: loredo@astro.cornell.edu

Abstract. I describe ongoing work developing Bayesian methods for flexible modelling of arrival-time-series data without binning. The aim is to improve the detection and measurement of X-ray and gamma-ray pulsars and of pulses in gamma-ray bursts. The methods use parametric and semi-parametric Poisson point process models for the event rate, and by design have close connections to conventional frequentist methods that are currently used in time-domain astronomy.

Keywords. methods: statistical, methods: data analysis, pulsars: general, gamma rays: bursts

---

Measuring the arrival times, directions and energies of individual quanta—photons or particles—potentially provides the finest possible resolution of dynamical astronomical phenomena, particularly for high-energy sources that produce low detectable fluxes. The simplest methods for signal detection and measurement bin the data for statistical or computational convenience (for example, to allow the use of asymptotic Gaussian approximations, or to enable fast Fourier decomposition with an FFT). But methods that instead directly analyze the event data without binning can detect weaker signals and probe shorter time-scales than can ones that require binning.

The ongoing work I briefly describe here was motivated by studies of X-ray and gamma-ray pulsars, which produce periodic signals, and gamma-ray bursts, which produce chaotic signals that are typically comprised of multiple overlapping pulses. In the former case there may be less than one event per period (particularly in energy-resolved studies); in the latter, time-scales as short as milliseconds are relevant, and detected photons are sparse at high energies. Both phenomena motivate the development of data analysis techniques that can milk every hard-won event for what it is worth.

For simplicity we focus here just on the arrival-time data (also known as time-tagged event data), and presume that the events being analyzed have been selected to have directions consistent with an origin from a single source, and moreover that energy dependence of any putative signal is not significant (so the signal's temporal signature is not corrupted by ignoring event energies). We can represent the data as points on a time-line, as in Fig. 1. The dots denote events at times t_i that have been detected within small time intervals (δt), which represents the instrumental time-resolution. The empty intervals, denoted Δt_j, are informative; seeing no events in an observed interval provides a constraint on the signal, in contrast to simply not observing during the interval.

Conventional approaches to detecting signals in such data (binned or unbinned) adopt an approach of frequentist hypothesis testing: one devises a test statistic that measures departure of the data from the predictions of an uninteresting "null" model, and uses it to see if the null model may be safely rejected (implying that an interesting signal is present). No explicit signal model is needed to define such a test.

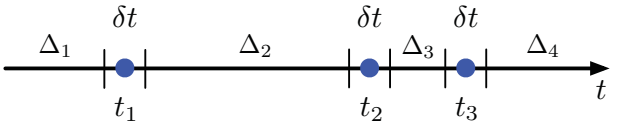

Figure 1. Arrival-time series depicted as points on a timeline, with detection and nondetection intervals noted.

Here I describe methods that were developed from the Bayesian approach, where one compares the null model to explicit alternative models that describe interesting signals. One motivation was to show how conventional "alternative free" test statistics arise in this framework (exactly or approximately). It illuminates implicit assumptions underlying conventional methods; more constructively, it provides a framework whereby generalizations of the implicit models may lead to new methods. Some other virtues of adopting a Bayesian approach to these problems, both pragmatic and conceptual, are outlined below; Loredo (2011) gives a more extensive, but still introductory, discussion.

We model the data with a non-homogeneous Poisson point process in time. A model, M, specifies an intensity function (event rate) $r(t;\mathcal{P})$ that depends on the model's parameters, $\mathcal{P}$. Parametric models have a parameter space of fixed dimension; for example, a periodic-signal model will typically have frequency, amplitude and phase parameters, and possibly additional parameters describing the light-curve shape. Non-parametric models have a parameter space whose (effective) dimension may grow with sample size, adapting to the data; it may formally be infinite-dimensional. Semi-parametric models have a parameter space with a fixed-dimension part (e.g. the frequency and phase of a periodic model), and a non-parametric part (e.g. an adaptive light-curve shape).

The data drive Bayesian inferences via the likelihood function, the probability for the data, $D = (\{t_i\}, \{\Delta t_j\})$, given values for the parameters. Referring to Fig. 1, we build the likelihood function by calculating the product of Poisson counting probabilities for zero counts in the empty intervals, and one count in each detection interval. The zero-count probabilities are of the form $\exp[-\int_{\Delta_i} dt\, r(t)]$, and the one-count probabilities are of the form $[r(t_i)\delta t]\exp[-r(t)\delta t]$ (presuming δt is small so that $r(t)\delta t \ll 1$). The likelihood function is thus:

$$\mathcal{L}_M(\mathcal{P}) \equiv p(D|\mathcal{P},M) = \exp\left[-\int_T dt\, r(t)\right] \prod_{i=1}^{N} r(t_i)\delta t, \tag{0.1}$$

where T denotes the full observing interval and N is the number of detected events. To go further, we must give specific rate models and priors for the model parameters. To fit a particular model, we use Bayes's theorem to calculate a posterior probability for the parameters, $p(\mathcal{P}|D,M) = p(\mathcal{P}|M)\mathcal{L}_M(\mathcal{P})/Z_M$, where Z_M is a normalization constant given by the integral of the product of the prior probability density, $p(\mathcal{P}|M)$, and the likelihood function. To detect a signal, we instead use Bayes's theorem on a hypothesis space including one or more models for interesting signals, and the null model (here, a constant-rate model with a single parameter, the amplitude, A, with $r(t) = A$). In this space, the normalization constant for a particular model, Z_M, plays the role of the likelihood for the model (as a whole); Z_M is often called the marginal likelihood for the model, where "marginal" refers to the integration over $\mathcal{P}$ used to calculate its value.

As a simple starting point, consider a model where the logarithm of the rate is proportional to a sinusoid plus a constant; the logarithm guarantees that the rate itself is non-negative. We may then write the rate as $r(t) = A\exp[\kappa\cos(\omega t - \phi)]/I_0(\kappa)$, where A is the time-averaged rate and $I_0(\cdot)$ denotes the modified Bessel function of order 0 (this normalizes the exponential factor so that A is the time-average). Light curves with that

https://doi.org/10.1017/S1743921312000300 Published online by Cambridge University Press

shape have a single peak per period; its width (equivalently, the duty cycle) is determined by the concentration parameter, κ, with large values corresponding to sharp peaks and $\kappa = 0$ corresponding to a constant rate. To estimate the frequency and concentration, we calculate the posterior for all four parameters, $p(A, \omega, \kappa, \phi|D, M)$, and marginalize (integrate) over A and ϕ. By adopting a flat prior for the phase, and nearly any prior for A that is independent of the other parameters, we find a marginal posterior probability density for frequency and concentration proportional to $I_0[\kappa R(\omega)]/[I_0(\kappa)]^N$, where $R(\omega)$ is the Rayleigh statistic given by

$$R^2(\omega) = \frac{1}{N}\left[\left(\sum_{i=1}^{N}\cos\omega t_i\right)^2 + \left(\sum_{i=1}^{N}\sin\omega t_i\right)^2\right]. \tag{0.2}$$

Estimation of ω alone, accounting for uncertainty in all other parameters, is found by further integrating over κ, which is easy to do numerically. Detection requires calculating the marginal likelihood, corresponding to a further integration over ω, and must be done numerically. It can be time-consuming for blind searches, but not significantly more so than the kind of frequency-grid searching employed by conventional tests.

The Rayleigh statistic was invented for the well-known (frequentist) Rayleigh test for periodic signals in arrival-time series; Lewis (1994) gives a review of the Rayleigh test and other frequentist period detection methods mentioned below. The quantity $2R^2(\omega)$ is called the Rayleigh power; it is the point process analogue of the periodogram or Fourier power spectral density. From a Bayesian point of view, the Rayleigh test implicitly assumes that periodic signals may be modelled well by log-sinusoid rate functions. Notably, there is no parameter corresponding to κ in the Rayleigh test; also, in practice it is known to work well only for smooth light curves with a single broad peak per period. Such light curves correspond to values of κ near unity—another implicit assumption of the Rayleigh test. These results indicate that one can implement Bayesian period searches using conventional computational tools already at hand for the Rayleigh test. They also indicate that explicit consideration of the κ parameter may lead to procedures which are more sensitive to sharply-peaked light curves than are the conventional Rayleigh tests.

Many pulsars have light curves with two or more peaks per period. That suggests generalizing the log-sinusoid model to a log-Fourier model, with the logarithm of the rate proportional to a sum of harmonic sinusoids. By adopting a finite sum of m harmonics with concentration parameters κ_k ($k = 1$ to m, with the fundamental corresponding to $k = 1$), we may proceed analogously to the above analysis. The larger number of phases and concentration parameters thwarts analytical integration; in an approximate treatment the posterior distribution for frequency and concentration is proportional to $\exp[S(\omega)]$, with $S(\omega) \equiv \sum_{k=1}^{m} \kappa_k R(k\omega)$. The frequentist test generalizing the Rayleigh test to multiple harmonics is the Z^2_m test, with $Z^2_m(\omega) = 2\sum_{k=1}^{m} R^2(k\omega)$, a sum of Rayleigh powers at harmonics. Notably, the Bayesian analysis uses the sum of $R(k\omega)$ values ("harmonic Rayleigh amplitudes") rather than powers. Note that, for $\kappa_k = 1$, $S^2(\omega) = Z^2_m(\omega)/2 + \sum_k \sum_{j\neq k} R(j\omega)R(k\omega)$; that is, S^2 contains information not in Z^2_m. Roughly speaking, Z^2_m corresponds to incoherently summing power in harmonics, but the quantity arising in the Bayesian treatment of a harmonic model instead sums amplitudes, accounting for phase information ignored by Z^2_m.

A popular frequentist period-detection method that aims to be sensitive to periodic signals of complex shape is χ^2 epoch folding (χ^2-EF). One folds the arrival times given by a trial period to produce a phase, θ_i, for each event, with θ_i in the interval $[0, 2\pi]$. For a constant signal, the phases should be uniformly distributed. The χ^2-EF method bins the phases into B bins and uses Pearson's χ^2 to test consistency of the binned phases with

https://doi.org/10.1017/S1743921312000300 Published online by Cambridge University Press

a uniform distribution. This motivates a Bayesian model with a piecewise-constant rate function and B steps per period. Gregory & Loredo (1992) (GL92) analyzed such a model, giving the rate in a bin as Af_k, where A is the average rate, and the shape parameters, f_k ($k = 1$ to B), specify the fraction of the rate attributed to each bin, with $\sum_k f_k = 1$. They assigned a constant prior to the shape parameters, from the intuition that that spreads probability across all possible shapes. After marginalizing over the shape parameters, the posterior for frequency and phase is inversely proportional to the multiplicity of the set of counts of events in phase bins. In a large-count limit, that is approximately $\exp(-\chi^2/2)$, providing a tie to χ^2-EF. The method has performed impressively, detecting an X-ray pulsar where the Rayleigh test failed, and performing well in a simulation study by Rots (1993) that compared it to other methods.

Despite those successes, there is room for improvement in the GL92 analysis, for a surprising reason. The constant prior adopted in GL92 does not in fact spread probability over all possible shapes. As the number of bins increases, the constant prior assigns ever larger probability to the neighborhood of flat models, making it harder than necessary to detect narrow peaks. The reason is a "curse of dimensionality" known as *concentration of measure*: a multi-dimensional distribution built out of the product of broad one-dimensional distributions with finite moments concentrates its probability in a decreasing volume of parameter space as dimension increases. Concentration can be avoided by letting the parameters of the one-dimensional component distribution vary with the target dimension. A theoretically appealing way to do that is to require *divisibility* of the prior; for example, the four-bin prior should reduce to the two-bin prior if we create a two-bin model out of the combination of bins 1 and 2, and 3 and 4. A divisible Dirichlet distribution prior accomplishes that, and improves the sensitivity to sharply peaked light curves so long as any constant background component is small (Loredo 2011).

Extending the construction to functions described with an infinite number of bins or points leads one to consider *infinitely divisible* priors for non-parametric functions: Gaussian process priors for curve fitting, Dirichlet process priors for modelling probability densities, and Lévy process priors for modelling Poisson intensities. With a team of statistician and astronomer colleagues, I am developing methods using priors built on Lévy processes for modelling pulses in gamma-ray bursts. This approach can quantify uncertainty even in a regime where pulses are highly overlapping. Its implementation involves compound Poisson processes, as arise in simple models of accumulation of rain, where drops with a distribution of sizes fall radomly over a region of space. And so Bayesian modelling of arrival-time series has led us from sines to steps to droplets.

Acknowledgements

Some of the research summarized here is currently supported by grant NNX09AK60G from NASA's Applied Information Systems Research Program. I am grateful to my collaborators on that grant, Mary Beth Broadbent, Carlo Graziani, Jon Hakkila, and Robert Wolpert, for their continuing contributions to this research.

References

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

Lewis, D. A. 1994, in: J. L. Stanford & S. B. Vardeman (eds.), *Statistical methods for physical science, Methods of Experimental Physics Vol. 28* (San Diego: Academic Press), p. 349

Loredo, T. J. 2011, in: J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, David Heckerman, A. F. M. Smith, & M. West (eds.), *Bayesian Statistics 9* (Oxford: Oxford Univ. Press), p. 361

Rots, A. H. 1993, in: P. J. Serlemitsos & S. J. Shrader (eds.), *BBXRT: a preview to astronomical X-ray spectroscopy in the 90's* (Maryland: NASA/GSFC), p. 363