Bayesian Inference With Log-Fourier Arrival Time Models and Event Location Data

Tom Loredo
May 1990; Sep–Nov 1993

1. INTRODUCTION

In these notes we describe Bayesian calculations with models that supply Bayesian counterparts to frequentist period searching with the Rayleigh and Z_n^2 statistics, and we outline how to analyze event arrival time data that includes a location for each event on a detector with a known point spread function. Many of these results are from my notebook of a few years ago, but I suppose nothing is true nowadays until it's in TEX! The reader is presumed to be familiar with the material in §§ 2 and 3 of Gregory & Loredo (1992; hereafter GL). In particular, our starting point is the arrival time likelihood function given by equation (3.5) of that paper. For the record, this likelihood function is not our invention; it is widely known in frequentist studies of inhomogeneous Poisson processes.

Before we start, let's summarize what is different about the Bayesian approach to arrival time series analysis. There are two key differences between Bayesian methods and their frequentist counterparts. First, Bayesian methods must assume a particular functional form (usually with several free parameters) for the possible periodic signal. Frequentist methods instead try to reject a constant model for the signal, without explicit reference to a specific family of periodic models. This would seem to be a serious drawback to Bayesian methods, since we usually don't know the shape of the periodic lightcurve we are trying to detect. However, from the Bayesian point of view the choice of statistic to be used in a frequentist test implicitly corresponds to a model choice. This is because different statistics correspond to different ideas about what departures from uniformity one wants to be most sensitive to, which is simply another way of thinking of a periodic model. This is recognized informally in frequentist literature in the acknowledgment that different choices of statistic are most sensitive to underlying signals of different shapes (e.g., the χ^2 statistic can detect signals with one or more narrow peaks well, while the Rayleigh statistic can detect smooth, sinusoidal signals well). It is recognized more formally in considerations of the *power* of a particular frequentist test: its ability to accurately detect a periodic signal *of specified shape* when such a signal is present. We have made something of an "industry" identifying particular models for which a Bayesian calculation leads to consideration of a function of the data directly related to a common frequentist statistic. This makes the implicit assumptions of a frequentist procedure explicit, so that its appropriateness can be more easily judged, and modifications pursued. These notes offer an example of this.

The second key distinction between Bayesian and frequentist methods is how a statistic is used to determine whether a periodic signal is present or not. In the frequentist approach, we evaluate the statistic at many frequencies, and note the largest observed value. We then calculate the chance probability for the statistic being as large as the maximum value or larger, assuming that the signal is constant; if this probability is small, the constant signal model is rejected. This calculation is significantly complicated by the need to examine many periods (and possibly many values of other parameters, such as the phase assigned to $t = 0$), and the number and location of the periods must be specified without consideration of the data. These aspects of a frequentist analysis introduce a troubling subjectivity in the results. In contrast, the Bayesian approach requires that one *integrate*, rather than maximize, a nonlinear function of the test statistic over the period range being searched (and over any other parameters in the model). As with the frequentist test, the range to search must be subjectively specified. However, the number and location of periods searched is irrelevant; as many periods should be used as is necessary to accurately compute the required integrals. The result is penalized for the unknown parameters automatically and objectively by the averaging performed by the integration process. Also, the result is conditional on the one observed data set; the frequentist probability is the probability, not only of the observed data, but of all worse data sets as well.

Another distinction worth mentioning is that the Bayesian approach forces us to clearly distinguish the tasks of signal detection and parameter estimation. In particular, the Bayesian approach identifies a nonlinear function of the test statistic whose integrals give straightforward probability statements about the allowed ranges ("credible regions") of unknown parameters, such as the frequency.

Okay, now that all (well, most) of the propaganda is out of the way, let's move on to the calculations!

2. THE LIKELIHOOD FUNCTION AND MODEL CRITERIA

We presume we have a parameterized model rate function, $r(t)$, which we will use to model data consisting of N arrival times, t_i, over some observing interval of total duration T. We use T to designate both the total "live-time" duration of the observations, and the set of intervals (possibly separated by gaps) that comprise the observations, so that $\int_T$ stands for an integral over all observing intervals (they need not be contiguous). For periodic models, the parameters for $r(t)$ will typically include an amplitude, A; a frequency, ω (or equivalently a period, $P = 2\pi/\omega$); a phase, ϕ; and possibly some other parameters, $\mathcal{S}$, that parameterize the lightcurve shape. These ingredients, together with Poisson assumptions, give the arrival time likelihood function in equation (3.5) of GL. We reproduce it here, ignoring the irrelevant Δt factors (they do not depend on the parameters, and cancel out in Bayes's theorem):

$$\mathcal{L}(A, \omega, \phi, \mathcal{S}) = \exp\left[-\int_T dt\, r(t)\right] \prod_{i=1}^{N} r(t_i). \tag{2.1}$$

Bayesian calculations require integrals over parameter space. Such integrals arise when we normalize distributions, find marginal distributions for interesting subsets of parameters, or compare rival models. Calculating these integrals can be *difficult*. In the context of searching for periodicities in arrival time data, any model will have at least four parameters: A, ω, ϕ, and at least one shape parameter. The likelihood function usually varies wildly with ω, so any integrals require evaluation of equation (2.1) at many frequencies. At *each* of these frequencies we have to integrate over the other parameters. This gets out of control pretty quickly! So the game we have to play in Bayesian modeling of arrival time data is to try to find physically useful models that allow at least some of the required integrations to be done analytically. In the next section, we show that the amplitude can always be integrated analytically. So our model choice should be guided by focussing on the other parameters. The GL model is cute in that it is a model with many $\mathcal{S}$ parameters—and thus capable of describing a diverse set of shapes—for which integrals over *all* of the $\mathcal{S}$ parameters can be done analytically. Only the ω and ϕ integrals need to be done numerically. Here we will take a complementary approach, identifying models for which ϕ integrals can be done analytically.

We note in passing that cases where the period is known or is well-constrained a priori (e.g. in searching for X-ray pulsations at a known radio pulsar period) can be significantly more tractable than cases where we must search for an unknown period. In the former case, it may be possible and interesting to consider more complicated models than one could otherwise consider because the ω dimension is eliminated, or at least small. But even in such cases it is useful to have computationally efficient models, such as those discussed here and in GL.

3. THE AMPLITUDE PARAMETER

3.1. *Normalized Models*

One parameter—the amplitude—is common to all models, and it turns out that we can take care of it once and for all, as we show in this section. We can always write a periodic model rate in the form,

$$r(t) = A\rho(\omega t - \phi), \tag{3.1}$$

where $\rho(\theta)$ is a periodic function with period 2π. Without loss of generality, we can also require that the amplitude parameter be the average rate,

$$A \equiv \frac{1}{P}\int_P dt\, r(t). \tag{3.2}$$

Together, these two equations impose a normalization constraint on ρ. To see this, change variables in equation (3.2) to $\theta = \omega t - \phi$, and use the fact that $\omega = 2\pi/P$. Then equation (3.2) becomes,

$$\int_0^{2\pi} d\theta\, \rho(\theta) = 2\pi, \tag{3.3}$$

or, equivalently,

$$\int_P dt\, \rho(t) = 1. \tag{3.4}$$

That is, ρ must be normalized as if $\rho/2\pi$ were a probability distribution in phase. We'll see in a minute why this is a convenient parameterization; roughly, it lets A control the expected number of events, and ρ describe the shape of the lightcurve. (Actually, our main conclusions will hold for any normalizing constant for ρ, as long as it does not depend on the model parameters.)

Now plug this form into the likelihood function. For a contiguous interval, T, the integral in the exponent of equation (2.1) is,

$$\int_0^T dt\, A\rho(\omega t - \phi) = \frac{AP}{2\pi}\int_{\phi}^{\omega T-\phi} d\theta\, \rho(\theta) \approx AT. \tag{3.5}$$

To get the last line, we used the fact that the integral divided by the 2π factor up front is just equal to the number of periods covered by T; multiplying by P thus gives T. It is approximate because T may not be an integral number of periods long, so the last part of the integral will include only part of one period. The integral of $\rho/2\pi$ over a fraction, f, of a period is not generally equal to f, hence the approximate result. But so long as $T \gg P$, the error we make in this final fractional period will be small compared to the total integral, because the total integral will contain many periods. So from now on we'll assume we are interested in periods such that many periods are contained in the data. This regime is very nice because the exponential factor in the likelihood function then depends only on A, and not on any other parameters.

For simplicity we've shown this for a contiguous interval; but it also holds for noncontiguous intervals of total duration $T \gg P$, so long as the "holes" in the observing time do not line up in phase (such alignment could happen for a strong periodic signal observed by a detector with bad dead time; but we probably don't need statistics to detect pulsations in such data!).

The likelihood function in the many-period regime is then,

$$\mathcal{L}(A, \omega, \phi, \mathcal{S}) = \left[A^N e^{-AT}\right] \prod_i \rho(\omega t_i - \phi). \tag{3.6}$$

Now we can see what this parameterization buys us: the likelihood factors into a part that depends only on A and a part that depends on all the other parameters (if ρ were not normalized, the right hand side of equation (3.5) would be a function of ω, ϕ, and $\mathcal{S}$ that would appear in the exponent in the likelihood). This means that so long as our prior similarly factors (that is, knowledge of A tells us nothing about the other parameters a priori, and vice versa), inferences about A will be independent of those about ω, ϕ, and whatever other parameters we have, and vice versa. Since this is true, let's take care of the amplitude inferences now, once and for all, so that in the rest of these notes we can focus on the other parameters.

3.2. *Inferring the Amplitude*

Let's presume the prior for the parameters factors in the way just stated, so that

$$p(A, \omega, \phi, \mathcal{S} \mid M) = p(A \mid M)\, p(\omega, \phi, \mathcal{S} \mid M), \tag{3.7}$$

where M is the background information specifying the model, its relation to the data (i.e. the Poisson assumptions we used to derive the likelihood function), and anything else we might know about the model (e.g. constraints on the period from other observations). To infer all of the parameters of a model, we use Bayes's theorem to calculate the full joint posterior for the parameters,

$$p(A, \omega, \phi, \mathcal{S} \mid D, M) = \frac{p(A \mid M)\, p(\omega, \phi, \mathcal{S} \mid M)}{p(D \mid M)} \mathcal{L}(A, \omega, \phi, \mathcal{S}). \tag{3.8}$$

Here $p(D \mid M)$ is the global likelihood for the model, given by,

$$\begin{aligned} p(D \mid M) &= \int dA \int d\omega \int d\phi \int d\mathcal{S}\; p(A, \omega, \phi, \mathcal{S} \mid M)\, \mathcal{L}(A, \omega, \phi, \mathcal{S}) \\ &= \int dA\, p(A \mid M) A^N e^{-AT} \int d\omega \int d\phi \int d\mathcal{S}\, p(\omega, \phi, \mathcal{S} \mid M) \prod_i \rho(\omega t_i - \phi). \end{aligned} \quad (3.9)$$

It is merely a normalization constant for purposes of inferring parameters; but it plays a much more important role in comparing alternative models (e.g. comparing a model with a periodic signal to one without). It is the hardest integral we'll eventually need to calculate.

Equation (3.8) gives the implications of the data for *all* of the model parameters. A summary of the implications *for the amplitude alone* is given by the marginal distribution for A, obtained by integrating the joint posterior over all the other parameters. Thanks to the factorization of the likelihood function in equation (3.6), we can calculate the marginal for A analytically for *any* model in normalized form as follows:

$$\begin{aligned} p(A \mid D, M) &= \int d\omega \int d\phi \int d\mathcal{S}\; p(A, \omega, \phi, \mathcal{S} \mid D, M) \\ &= \frac{1}{p(D \mid M)}\, p(A \mid M) A^N e^{-AT} \int d\omega \int d\phi \int d\mathcal{S}\, p(\omega, \phi, \mathcal{S} \mid M) \prod_i \rho(\omega t_i - \phi) \\ &= \frac{p(A \mid M) A^N e^{-AT}}{\int dA\, p(A \mid M) A^N e^{-AT}}. \end{aligned} \quad (3.10)$$

The factorization of the likelihood and prior lets us obtain this result without even requiring us to know how to integrate over $(\omega, \phi, \mathcal{S})$, because the required integral cancels.

The marginal posterior for A depends on the prior, as it must. However, the dependence is weak provided that the prior is nonzero near $A = N/T$ (where the likelihood peaks) and does not vary rapidly (i.e., on a scale $\sim \sqrt{N}/T$) there. Moreover, since the A parameter is common to all models (including nonperiodic ones), the prior for A is completely irrelevant for choosing among competing models (assuming it is the same for all models). We demonstrated these facts in Loredo (1992) and in GL using a flat prior with a sharp cutoff at some maximum rate, $A_{\max}$. Here, we repeat some of the calculations with an exponential prior, both to further demonstrate the insenstivity to the form of the prior, and because the algebra is particularly simple with this "conjugate" prior. Mike West (1992) has suggested this prior for similar Poisson problems.

So we presume our prior information includes a nonzero expectation value for A, which we denote A_0. Then the prior for A is

$$p(A \mid M) = \frac{1}{A_0} e^{-A/A_0}. \quad (3.11)$$

You can verify that this is normalized, and that $\langle A \rangle = A_0$. It will be more convenient to write this prior in terms of the timescale $\tau_0 = 1/A_0$, so that

$$p(A \mid M) = \tau_0 e^{-A\tau_0}. \quad (3.12)$$

Using this prior, the normalization constant in the denominator of equation (3.10) is,

$$\tau_0 \int_0^\infty dA\, A^N e^{-A(T+\tau_0)} = \frac{\tau_0 N!}{(T+\tau_0)^{N+1}}. \quad (3.13)$$

Thus the marginal posterior for A is,

$$p(A \mid D, M) = \left(1 + \frac{\tau_0}{T}\right)^{N+1} \frac{T(AT)^N}{N!} \exp\left[-AT\left(1 + \frac{\tau_0}{T}\right)\right]. \tag{3.14}$$

In the limit that $A_0 \to \infty$ (so $\tau_0 \to 0$), the prior flattens and becomes vanishingly small everywhere. But the posterior remains finite and perfectly well behaved, because the τ_0 factor in the prior also appears in the normalization constant, and thus cancels out. The $\tau_0 = 0$ limit is,

$$p(A \mid D, M) = \frac{T(AT)^N}{N!} e^{-AT}. \tag{3.15}$$

This is just of the form of a Poisson distribution N, multiplied by T so that it is normalized with respect to A rather than N. Considered as a function of A, it is a Gamma distribution. This limiting form is the same as that found in our earlier papers with a limiting flat prior, as it should be since as $A_0 \to \infty$, the exponential prior becomes flat.

The mode of equation (3.14) is the value of A that makes the derivative with respect to A vanish; it is,

$$\hat{A} = \frac{N}{T} \frac{1}{1 + \frac{\tau_0}{T}}. \tag{3.16}$$

The posterior expectation value for A is found by multiplying equation (3.14) by A and integrating; the result is,

$$\bar{A} = \frac{N+1}{T} \frac{1}{1 + \frac{\tau_0}{T}}. \tag{3.17}$$

This differs only slightly from the mode provided that $N \gg 1$. A measure of the width of the marginal posterior is provided by the posterior standard deviation, which is,

$$\sigma_A = \frac{\sqrt{N+1}}{T} \frac{1}{1 + \frac{\tau_0}{T}}. \tag{3.18}$$

All of these summaries of the marginal posterior differ from the infinite A_0 limit by factors of $(1 + \tau_0/T)$. Thus, so long as $\tau_0 \ll T$ (or $A_0 \gg 1/T$), our posterior estimates of A are insensitive to the precise prior information about A.

3.3. *Eliminating the Amplitude*

We have so far focused attention completely on A. But in practice, it is the other parameters that are usually of greater interest, particularly the period or frequency. In the remainder of these notes we will focus on the marginal posterior for all parameters other than A, which we call the *joint marginal distribution.* The joint marginal distribution is,

$$\begin{aligned} p(\omega, \phi, \mathcal{S} \mid D, M) &= \int dA\, p(A, \omega, \phi, \mathcal{S} \mid D, M) \\ &= C^{-1}\, p(\omega, \phi, \mathcal{S} \mid M) \prod_i \rho(\omega t_i - \phi), \end{aligned} \tag{3.19}$$

where the normalization constant is given by,

$$C = \int d\omega \int d\phi \int d\mathcal{S}\, p(\omega, \phi, \mathcal{S} \mid M) \prod_i \rho(\omega t_i - \phi). \tag{3.20}$$

To get equation (3.19), we cancelled the integral over A in the numerator with an identical integral factor in the denominator. This is possible only because of the factored form of the likelihood and prior, and it holds exactly for *any* independent prior for A.

Equation (3.19) will be the focus of attention in the remainder of these notes. We need to integrate it over ϕ and $\mathcal{S}$ to infer the frequency. We also need to evaluate C both to normalize interesting distributions, and to enable us to calculate the global likelihood, which we need to compare models. To see the relevance of C to the global likelihood, note that we can use equation (3.13) to compute one of the factors in the global likelihood given in equation (3.9), giving

$$\begin{aligned} p(D \mid M) &= \frac{\tau_0 N!}{(T+\tau_0)^{N+1}} \int d\omega \int d\phi \int d\mathcal{S}\, p(\omega, \phi, \mathcal{S} \mid M) \prod_i \rho(\omega t_i - \phi) \\ &= \frac{C \tau_0 N!}{(T+\tau_0)^{N+1}}. \end{aligned} \tag{3.21}$$

Thus if we can compute C for any model, we can trivially compute the global likelihood for that model.

In the following sections, we discuss models for which integrals over ϕ can be performed analytically, significantly simplifying the resulting numerical calculations. These models also bear close relationships with common frequentist statistics used for period detection in arrival time series.

3.4. *The Constant Model*

But before we discuss periodic models, we note that we have all the ingredients we need to completely treat the simplest nonperiodic model: a constant model. This model has only one parameter, the unknown "DC" amplitude, so that

$$r(t) = A, \tag{3.22}$$

corresponding to $\rho(t) = 1$. The likelihood function is simply $\mathcal{L}(A) = A^N \exp(-AT)$, and estimates of A and its uncertainty are just as were given above. Finally, the global likelihood for this model is the integral in the denominator of equation (3.10), which we evaluated in equation (3.13), so that,

$$p(D \mid M_0) = \frac{\tau_0 N!}{(T+\tau_0)^{N+1}}, \tag{3.23}$$

where M_0 denotes the information specifying the constant model. From equations (3.23) and (3.21), we see that the Bayes factor in favor of a particular periodic model over the constant model is simply,

$$B_{M,0} \equiv \frac{p(D \mid M)}{p(D \mid M_0)} = C. \tag{3.24}$$

Thus, up to the prior odds, the odds in favor of a periodic model is given by C. Note that all dependence on the prior for A has dropped out, because this parameter is common to both models.

4. THE LOG-SINUSOID MODEL

To detect and characterize a periodic signal, one's first impulse might be to consider a sinusoidal model rate, or a Fourier series of harmonic sinusoids. Such models work well for the analysis of time series consisting of samples of a signal contaminated with additive noise to which we assign a Gaussian probability distribution. Bretthorst (1988) has studied such models in depth with Bayesian methods, and shown how they are related to frequentist methods that rely on the DFT of the data.

Two aspects of arrival time series modeling conspire to argue against this choice. First, we must model an event rate, not a signal amplitude. An event rate must be nonnegative everywhere. Thus a simple sinusoid is not a valid model; we must instead consider a function like,

$$r(t) = A[1 + f\cos(\omega t - \phi)], \tag{4.1}$$

where a DC offset component is added (the pulsed fraction, f, is bound between 0 and 1). Second, in the Gaussian case, the likelihood function is an exponential of a *sum* of the signal model evaluated at the sample times. In our Poisson arrival time likelihood function, we have instead a *product* of rates evaluated at event times. As a result, models which are analytically tractable in the Gaussian case can be intractable in the Poisson case.

In fact, it is possible to use the model rate of equation (4.1), and analytically marginalize with respect to ϕ. However, the constant term necessary to make equation (4.1) everywhere nonnegative leads to a very large number of terms in the product of event rates, growing roughly like 3^N. Thus, analytical marginalization does not help us with this model. In fact, this model is more tractable if any required marginalizations over ϕ are performed numerically rather than analytically.

Since $r(t)$ must be everywhere nonnegative, and since we'd like products of $r(t)$ to have a simple form, it makes more sense to model the *logarithm* of the rate as a Fourier series. We begin in this section by considering a single sinusoid, taking

$$r(t) \propto e^{\kappa\cos(\omega t - \phi)}. \tag{4.2}$$

Our first task is to cast this model into normalized form, introducing a ρ function that is normalized according to equation (3.3). This task is simplified by noting that,

$$\int_0^{2\pi} d\theta e^{\kappa\cos\theta} = 2\pi I_0(\kappa), \tag{4.3}$$

where $I_n(\kappa)$ denotes the modified Bessel function of order n. Thus the normalized rate proportional to an exponentiated sinusoid is,

$$\rho(t) = \frac{1}{I_0(\kappa)}\, e^{\kappa\cos(\omega t - \phi)}. \tag{4.4}$$

This rate model has a single smooth pulse whose peak is at phase ϕ and whose shape is controlled by one shape parameter, κ, that jointly governs the "duty cycle" and the peak-to-background ratio. When $\kappa = 0$, the shape is flat; as $\kappa \to \infty$, the shape becomes a δ-function at phase ϕ. For large but finite κ, the lightcurve has a Gaussian shape near its peak with a standard deviation of $1/\sqrt{\kappa}$. The peak-to-background ratio is $e^{2\kappa}$. We restrict

κ to be ≥ 0 without loss of generality, since a change in sign of κ can always be accounted for by a change in ϕ of π.

We sometimes call this model the von Mises model, because $\rho/2\pi$ is a common distribution in the statistics of directional data on a circle known as the von Mises distribution. It is a circular generalization of the Gaussian distribution. But we will usually use a more descriptive name, calling equation (4.4) the log-sinusoid model. We will denote the information specifying this model by the symbol M_1.

To complete the specification of this (or any) model, we must specify priors for the model parameters. We will presume that the prior factors as the product of independent priors for ω, ϕ, and κ, so that

$$p(\omega, \phi, \kappa \mid M_1) = p(\omega \mid M_1)\, p(\phi \mid M_1)\, p(\kappa \mid M_1). \tag{4.5}$$

We assign a uniform prior for ϕ;

$$p(\phi \mid M_1) = \frac{1}{2\pi}. \tag{4.6}$$

This intuitively appealing flat prior can be formally justified if we insist that our conclusions be independent of the choice of the origin of time.

We use the same prior for ω that we used in GL:

$$p(\omega \mid M_1) = \frac{1}{\omega \log(\omega_{\rm hi}/\omega_{\rm lo})}. \tag{4.7}$$

The functional form of this prior comes from demanding independence of our conclusions with choice of time scale, and also is form-invariant if we change variables from ω to P (this is more an issue of principle than of practice; results will not change drastically in most cases if a flat prior is used). The prior range, however, is subjective. It will have little effect on our inferences about ω (provided any detected frequency lies in the prior range!) but it *will* affect the results of signal detection calculations. This cannot be helped, and actually is quite reasonable; even frequentist signal detection results depend on the search range.

Finally, we need to assign a prior for κ. For lack of anything better, we'll just use a flat prior up to some cutoff, $\kappa_{\rm max}$:

$$p(\kappa \mid M_1) = \frac{1}{\kappa_{\rm max}}. \tag{4.8}$$

If there is a periodic signal present, the value of $\kappa_{\rm max}$ will have negligible affect on our inferences about κ or ω. However, it will have an important effect on signal detection results, just as does the frequency range. The problem is: how do we choose $\kappa_{\rm max}$? It should come from our prior expectations for how small a duty cycle we would consider reasonable. This is a subjective aspect of the analysis—as is the choice of frequency range to search, and even the choice of model to consider. We'll skirt the issue here, and simply keep $\kappa_{\rm max}$ as an unspecified constant in the equations. But in an actual signal detection application, we'll have to worry explicitly about $\kappa_{\rm max}$. As a final note on this subject, we point out that the traditional Rayleigh test has no counterpart to κ, but that it is known to be poor at detecting signals that are not roughly sinusoidal. We conjecture that we could fix $\kappa = 1$ (i.e. use a δ-function prior) and duplicate this feature of the Rayleigh test; allowing κ to vary may improve the ability to detect "peakier" lightcurves than sinusoids. We'll find further evidence in support of this conjecture in the following section.

With the rate model and all the priors specified, we can now finally do some calculations!

4.1. *The Joint Marginal Distribution*

The marginal posterior for ω, ϕ, and κ is given by substituting equation (4.4) into equation (3.19):

$$p(\omega, \phi, \kappa \mid D, M_1) = C_1^{-1}\, p(\omega, \phi, \kappa \mid D, M_1)\, [I_0(\kappa)]^{-N} \exp\left[\kappa \sum_i \cos(\omega t_i - \phi)\right], \tag{4.9}$$

where C_1 is the normalization constant given by evaluating equation (3.20) for model M_1. The sum in the exponent can be simplified by using two trigonometric identities,

$$\begin{aligned} \sum_i \cos(\omega t_i - \phi) &= \cos(\phi) \sum_i \cos(\omega t_i) + \sin(\phi) \sum_i \sin(\omega t_i) \\ &= S(\omega)\, \cos(\phi - \phi'), \end{aligned} \tag{4.10}$$

where

$$S^2(\omega) = \left[\sum_i \cos(\omega t_i)\right]^2 + \left[\sum_i \sin(\omega t_i)\right]^2, \tag{4.11}$$

and

$$\tan(\phi') = \frac{\sum_i \sin(\omega t_i)}{\sum_i \cos(\omega t_i)}. \tag{4.12}$$

The joint marginal can thus be written,

$$p(\omega, \phi, \kappa \mid D, M_1) = C_1^{-1}\, p(\omega, \phi, \kappa \mid D, M_1)\, [I_0(\kappa)]^{-N} \exp\left[\kappa S(\omega) \cos(\phi - \phi')\right]. \tag{4.13}$$

The only parameter that S depends on is ω, and its functional form may look familiar: $S/\sqrt{N}$ is the Rayleigh statistic, and $2S^2/N$ is sometimes called the Rayleigh power or Fourier power of the time series. For future reference, we note two other ways of writing $S(\omega)$:

$$\begin{aligned} S^2(\omega) &= \left|\sum_j e^{i\omega t_j}\right|^2 \\ &= N + 2 \sum_i \sum_{j \neq i} \cos[\omega(t_j - t_i)]. \end{aligned} \tag{4.14}$$

It is clear from the first line that $S^2 \leq N^2$, with equality only when all events are separated from each other by exact integer multiples of a period. The second line shows that, roughly speaking, S^2 counts the number of pairs of events separated by an integer number of periods.

4.2. *Estimating and Eliminating the Phase*

The ϕ' constant defined by equation (4.12) is the most probable value for ϕ, conditional on particular values of ω and κ. To see this, note that the derivative of equation (4.13) with respect to ϕ is proportional to the product of $\sin(\phi - \phi')$ and the right hand side of this equation. The derivative thus vanishes at $\phi = \phi'$ and at $\phi = \phi' \pm \pi$. Since ϕ appears in the joint marginal only in the cosine in the exponential, multiplied by positive quantities, it is clear that the former root is a maximum, and the latter a minimum, proving our assertion. From equation (4.12), we see that ϕ' depends only on ω, and not on κ. Thus the location of the mode is independent of κ, although the width of the distribution about the mode depends on κ. Since $\cos(\phi - \phi') \approx 1 - (\phi - \phi')^2/2$ for ϕ near ϕ', equation (4.13) is approximately Gaussian as a function of ϕ near the mode, with a standard deviation of,

$$\sigma_\phi = \frac{1}{\sqrt{\kappa S(\omega)}}. \tag{4.15}$$

The Gaussian approximation will be good when $\kappa S \gg 1$.

Our next task is to eliminate ϕ from the joint posterior. Using equation (4.3), we can perform the necessary integral of equation (4.13) over ϕ analytically, giving us the marginal distribution for ω and κ:

$$\begin{aligned} p(\omega, \kappa \mid D, M_1) &= \int d\phi \, p(\omega, \phi, \kappa \mid D, M_1) \\ &= C_1^{-1} \, p(\omega \mid M_1) p(\kappa \mid M_1) \, \frac{I_0[\kappa S(\omega)]}{[I_0(\kappa)]^N}. \end{aligned} \tag{4.16}$$

This is one of our main results. It shows that the Rayleigh statistic is a sufficient statistic for estimating ω and κ in the context of the log-sinusoid model: the variation of S with ω tells us how to estimate the frequency; and the actual value of S at any given ω tells us how "clumped" the arrival times are, giving us an estimate of κ. The former fact is recognized in the Rayleigh test, but equation (4.16) tells us exactly how to "massage" the Rayleigh statistic mathematically in order to make probability statements about ω. There is no counterpart to κ in the formulation of the Rayleigh test, so the way that equation (4.16) extracts information about the "clumpiness" of the arrival times is unique to the Bayesian formulation.

4.3. *Estimating and Eliminating κ*

Equation (4.16) is the joint marginal for both ω and κ. Here we study the κ dependence of this distribution. It is of interest, both for estimating κ conditional on ω, and for integrating over κ. We need to integrate κ both to find the marginal for the most interesting parameter—the frequency—and to calculate the value of C_1, needed to normalize distributions and to compare the log-sinusoid model with competitors. We cannot go much further analytically, but we will close this section by developing some of the analytical background needed to deal with κ numerically. We discuss estimation of κ and integration of κ together, because these are not really separate tasks in practice: we need to know about the shape of the distribution in κ in order to intelligently integrate it, and this knowledge can be expressed as estimates of κ and its uncertainty.

We begin by examining the qualitative behavior of $p(\omega, \kappa \mid D, M_1)$ regarded as a function of κ, with ω fixed. The first thing to note is that the functional dependence on κ is completely determined by the magnitude of S, and does not depend on any other information about the distribution of the events in time. This compression of information into S is a computational advantage of this model. However, it may be a descriptive *dis*advantage, since there is clearly other information about the lightcurve shape in the distribution of arrival times. This is partly why we will consider generalizations of the log-sinusoid model in the following section.

To get a qualitative sense of the variation of equation (4.16) with κ, note that the modified Bessel functions have the following asymptotic behavior:

$$I_n(x) \approx \begin{cases} \left(\frac{x}{2}\right)^n \left[\frac{1}{n!} + \frac{x^2}{4\,(n+1)!}\right], & \text{as } x \to 0; \\ \frac{e^x}{\sqrt{2\pi x}} \left[1 - \frac{4n^2-1}{8x}\right], & \text{as } x \to \infty. \end{cases} \tag{4.17}$$

In particular, $I_0(x) \approx 1$ for $x \ll 1$. For small κ (more precisely, for $\kappa N \ll 1$),

$$p(\omega, \kappa \mid D, M_1) \approx C_1^{-1} \left[1 + \frac{\kappa^2}{4} (S^2(\omega) - N) \right]. \tag{4.18}$$

Thus $p(\omega, \kappa \mid D, M_1)$ is flat and nonzero at $\kappa = 0$, increases with κ if $S(\omega) > \sqrt{N}$, and decreases with κ if $S(\omega) < \sqrt{N}$. For large κ (more precisely, for $\kappa S \gg 1$),

$$p(\omega, \kappa \mid D, M_1) \approx \frac{p(\omega \mid M_1)}{C_1 \kappa_{\max} \sqrt{S(\omega)}} (2\pi\kappa)^{(N-1)/2} \exp\left[\kappa(S(\omega) - N)\right]. \tag{4.19}$$

Since $S \leq N$, the exponent is negative, so $p(\omega, \kappa \mid D, M_1)$ decreases exponentially for large κ (except for the pathological case where all event separations are integer numbers of periods, which corresponds to $\kappa = \infty$ anyway). These asymptotic behaviors imply that the most probable value of κ is $\hat{\kappa} = 0$ if $S(\omega) < \sqrt{N}$, corresponding to no periodicity; and that there is a nonzero mode only if $S(\omega) > \sqrt{N}$.

To locate the mode, we can calculate the partial derivative of equation (4.16) with respect to κ, using the fact that $I_0' = I_1$. It is a bit simpler to work with $\Lambda = \log p$. Denoting the partial derivative with respect to κ by ∂_κ, we find,

$$\partial_\kappa \Lambda = S \frac{I_1(\kappa S)}{I_0(\kappa S)} - N \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{4.20}$$

Since $I_1(0) = 0$, this always vanishes at $\kappa = 0$, as we anticipated from equation (4.18). This will be the mode if $S < \sqrt{N}$. But when $S > \sqrt{N}$, there will be another nonzero root, which we denote by $\hat{\kappa}$, because Λ must eventually fall at large κ. This root can be found by setting equation (4.20) equal to 0 and solving for $\hat{\kappa}$ numerically. A reasonable first guess for its location will help us find it. The asymptotic expressions are a place to start. For small κ, equation (4.17) gives,

$$\frac{I_1(x)}{I_0(x)} \approx \frac{x}{2} \left(1 - \frac{x^2}{8} \right). \tag{4.21}$$

Using this in equation (4.20), and setting the result equal to 0 gives,

$$\hat{\kappa}^2 \approx 8 \, \frac{S^2 - N}{S^4 - N}. \tag{4.22}$$

This expression will probably not be useful very often, as it applies only when $\hat{\kappa} N \ll 1$, which corresponds to *very* small values of κ when N is large. More useful is the large κ limit. We can see from equation (4.19) that the κ dependence of the ω–κ marginal is of the form of a Gamma distribution, $\propto \kappa^{(N-1)/2} \exp[\kappa(S - N)]$. The mode of this function gives an alternative estimate for $\hat{\kappa}$,

$$\hat{\kappa} \approx \frac{N - 1}{2[N - S(\omega)]}. \tag{4.23}$$

In practice, we should probably compute both estimates and use the one that is self consistent as our first guess. That is, we use equation (4.22) when that estimate satisfies $\hat{\kappa} N \ll 1$, and equation (4.23) when that estimate satisfies $\hat{\kappa} S \gg 1$.

Once $\hat{\kappa}$ is known, a measure of the accuracy of the estimate can be found by fitting a Gaussian to the peak. The mean of the Gaussian will be $\hat{\kappa}$, and its variance will be inversely proportional to the second derivative of Λ evaluated at $\hat{\kappa}$:

$$\sigma_\kappa^2 = -\frac{1}{\partial_\kappa^2 \Lambda(\hat{\kappa})}. \tag{4.24}$$

Using the result that $I_1'(x) = I_2(x) + I_1(x)/x$, the second derivative can be calculated from equation (4.20), giving,

$$\partial_\kappa^2 \Lambda = \frac{S^2}{I_0(\kappa S)}\left[I_2(\kappa S) + \frac{1}{\kappa S} I_1(\kappa S) - \frac{I_1^2(\kappa S)}{I_0(\kappa S)}\right] - \frac{N}{I_0(\kappa)}\left[I_2(\kappa) + \frac{1}{\kappa} I_1(\kappa) - \frac{I_1^2(\kappa)}{I_0(\kappa)}\right]. \tag{4.25}$$

To summarize, given a choice for ω, we can estimate κ as follows. If $S \leq \sqrt{N}$, $\hat{\kappa} = 0$ (i.e. the data favor no pulsations at this ω). Otherwise, we must find $\hat{\kappa}$ numerically by setting equation (4.20) equal to 0 and solving for $\hat{\kappa}$, say, using the Newton-Raphson method with equation (4.22) or equation (4.23) as a starting guess. A measure of the uncertainty in κ can be found by evaluating equation (4.25) at $\hat{\kappa}$ and using this to calculate σ_κ with equation (4.24).

More commonly, we will not have a particular ω in mind, and will instead need to integrate out κ in order to identify probable choices for ω (or whether *any* frequency is indicated at all). Two procedures suggest themselves. First, since $\kappa \geq 0$ and $p(\omega, \kappa \mid D, M_1)$ falls exponentially at large κ, Gauss-Laguerre quadrature will probably work efficiently, particularly when $S \lesssim \sqrt{N}$. If we have a code that gives abscissas, x_i, and weights, w_i, so that,

$$\int_0^\infty dx f(x) e^{-x} \approx \sum_i w_i f(x_i), \tag{4.26}$$

(e.g. the `gauleg` subroutine in *Numerical Recipes*), then a change of variables from κ to $x = \kappa(S - N)$ lets us write,

$$p(\omega \mid D, M_1) \approx \sum_i w_i e^{x_i} p(\omega, \kappa = x_i/(S-N) \mid D, M_1). \tag{4.27}$$

The change of variables matches the asymptotic behavior displayed in equation (4.19) to the Gauss-Laguerre formula, hopefully allowing us to use a small number of points to accurately estimate the integral.

If $S \gg \sqrt{N}$, there is likely to be a large peak at $\hat{\kappa}$ that will make Gauss-Laguerre quadrature inaccurate (unless we use an unreasonable number of abscissa values). Unfortunately, this is the most interesting case, since it corresponds to strong evidence for a periodic signal. One approach is to use Gauss-Hermite quadrature after a transformation to $x = (\kappa - \hat{\kappa})/\sigma_\kappa$. But I am suspect of this approach, since κ is bounded on one side of $\hat{\kappa}$ and its probability eventually falls exponentially, and not "Gaussianly", on the other side. A related alternative is to write,

$$p(\kappa) = f(\kappa) + p(\hat{\kappa}) G[(\kappa - \hat{\kappa})/\sigma_\kappa], \tag{4.28}$$

where $p(\kappa)$ denotes $p(\omega, \kappa \mid D, M_1)$ with ω fixed, $G(x) = \exp(-x^2/2)$, and $f(\kappa)$ is thus the difference between the function we want to integrate and a Gaussian fit to its peak. Then the desired integral of p is given by the integral of $p(\hat{\kappa})G$ and the integral of $f = p - p(\hat{\kappa})G$. The former integral is trivial; it is $p(\hat{\kappa})\sigma_\kappa\sqrt{2\pi}$. The f integral can be done in two parts. One part is for $\kappa < 0$, arising from the Gaussian being nonzero for negative κ; this integral is proportional to an error function. The other part is $p - p(\hat{\kappa})G$ for $\kappa \geq 0$, which can probably be evaluated accurately with Gauss-Laguerre quadrature, since the peak is subtracted off. Only some test calculations can reveal whether this procedure will work in practice.

Finally, we note that there will only be significant evidence for a periodic signal, and thus meaningful estimates for ω and κ, when S is large. Thus high accuracy in the κ integral is not required for values of ω that give small values of S , since either there is no evidence for a periodic signal at any ω, or there is strong evidence at some *other* ω, and integrals in that region of ω will dominate the total integral over all parameter space.

4.4. *Estimating the Frequency*

As with κ, we cannot do much analytically with regard to ω. Plotted as a function of ω, $I_0[\kappa S(\omega)]$ is usually very complicated. Even the peaks can have a complicated shape. However, we can get some idea of how a Bayesian frequency estimate will compare with a naive estimate based on the width of the Rayleigh peak as follows.

Let $\hat{\omega}$ denote the frequency that maximizes $S(\omega)$; this will be very near the most probable frequency value (it will not be precisely the mode because of the $/\omega$ dependence of the prior; it is the maximum likelihood value). Let $\hat{S} = S(\hat{\omega})$ be the value of S at $\hat{\omega}$. Near this peak, S will be approximately parabolic,

$$S(\omega) \approx \hat{S}\left[1 - \frac{(\omega - \hat{\omega})^2}{2\delta^2}\right], \tag{4.29}$$

where δ is the negative reciprocal of the second derivative of S with respect to ω at the peak; this is a simple measure of the width of the Rayleigh peak that would equal the half-width if S were exactly parabolic.

Let's consider the large $\hat{\kappa}\hat{S}$ limit, for which the marginal for ω and κ is given by equation (4.19), and for which $\hat{\kappa}$ is given by equation (4.23). In this limit, the ω dependence of the marginal is approximately Gaussian, with

$$p(\omega, \kappa = \hat{\kappa} \mid D, M_1) \propto \exp\left[-\hat{\kappa}\hat{S}\frac{(\omega - \hat{\omega})^2}{2\delta^2}\right]. \tag{4.30}$$

The standard deviation of this Gaussian is,

$$\begin{aligned} \sigma_\omega &= \frac{\delta}{\sqrt{\hat{\kappa}\hat{S}}} \\ &\approx \delta\sqrt{\frac{2(N - \hat{S})}{\hat{S}(N-1)}}. \end{aligned} \tag{4.31}$$

Since in this limit $\hat{\kappa}\hat{S} \gg 1$, from equation (4.23) we see that the factor multiplying δ is $\ll 1$. Thus the width of the posterior in ω can be *drastically* smaller than the width of the Rayleigh peak, because of the nonlinear "processing" done by the Bessel function to convert S into a probability density. Very similar conclusions were found by Bretthorst in his study of Gaussian spectrum analysis, where he often estimated frequencies with uncertainties orders of magnitude smaller than the width of the DFT peak (Bretthorst 1988).

4.5. *Signal Detection*

The final quantity we need to calculate is the global likelihood for M_1, $p(D \mid M_1)$. As shown in equation (3.21), if we can calculate C_1, we can calculate the global likelihood and any interesting odds ratios. The C_1 integral is given by equation (3.20), but for the log-sinusoid model we can perform the ϕ integral analytically, as in equation (4.16). Thus we have,

$$C_1 = \frac{1}{\kappa_{\max} \log(\omega_{\rm hi}/\omega_{\rm lo})} \int d\omega \int d\kappa \frac{I_0[\kappa S(\omega)]}{\omega [I_0(\kappa)]^N}. \tag{4.32}$$

The integrand is the same function (up to the constant C_1) whose numerical quadrature over κ we discussed in the previous subsection; all those comments apply here. In fact, we have to do the integrals in equation (4.32) first, because they determine C_1. So all we need to discuss is the quadrature over ω.

There is no particularly cute or intelligent way to code the ω integral. If you plot the result of the κ integral as a function of ω, you will see a function that varies wildly with ω. Only a brute force approach guarantees accurate estimation of the integral of such a complicated function. The one saving grace is that, if there is evidence for a signal, the value of the integral is typically *much* larger than it would be in the absence of a signal, and is completely dominated by the area under one or a few very narrow peaks. A good integration strategy is to do the integral with the trapezoid rule over the entire search range, with a step size chosen, not to guarantee accuracy of the integral over every bump and wiggle, but only to make sure each bump is sampled about twice. That is, make the step size about half the scale of variation of the function with ω. Along the way, note the location of the M largest peaks (with $M \approx 10$). Then go back, subtract off the contributions of the peaks to the original integral from the initial, crude grid, and redo the peaks with a much larger number of points per peak than was in the original grid.

Provided we can do this integral in a reasonable amount of time, we have everything we need in order to draw conclusions about the evidence for a log-sinusoid signal in the data. The Bayes factor in favor of M_1 over the constant model is simply $B_{1,0} = C_1$.

5. LOG-FOURIER MODELS

Without doing any numerical calculations with real (or simulated!) data, we can anticipate some weaknesses of the log-sinusoid model. Some of them were alluded to in the previous section. The basic problem is that we have only one shape parameter. Even if we suspect that the signal we are trying to detect has a single peak per period (as does the log-sinusoid lightcurve), a single parameter simply cannot describe the possible shapes we might expect. In particular, κ determines both the duty cycle and the peak-to-trough ratio of the lightcurve. But in general, there is no reason to suspect these characteristics to be linked. In particular, a source emitting *all* of its signal in a single peak (i.e., with a 100% duty cycle) could have almost any observed peak-to-trough ratio, depending on the background rate in the detector. Thus as a minimum we would like to have separate control of the background level and the pulse width.

We emphasize that these problems are not unique to the Bayesian treatment. The previous calculation elucidates the model assumptions implicit in frequentist use of the Rayleigh statistic, and weaknesses of the log-sinusoid model should also be present in frequentist analyses using the Rayleigh statistic. Indeed, it is widely known that the Rayleigh statistic is poor for detecting narrow pulses, although I am not aware of published studies elucidating how its sensitivity depends on background level as well as pulse width.

These problems encourage us to consider models with two or more shape parameters. One way to go is to add an explicit background term to the rate, writing

$$r(t) = B + \frac{A}{I_0(\kappa)} e^{\kappa \cos(\omega t - \phi)}, \tag{5.1}$$

where B is a parameter that partly determines the background rate. Unfortunately, adding a term *outside* of the exponent destroys the analytical simplicity of the log-sinusoid model. For example, integrals over ϕ no longer have a simple analytical form.

This leads us to consider adding terms *inside* the exponent. Here we will add additional sinusoidal terms, each with their own κ and ϕ, and with harmonic frequencies. This corresponds to modeling $\log(r)$ with a Fourier series. This is reasonable because $\log(r)$, unlike r itself, is not constrained to be positive.

In the rest of this section we work out some of the details arising in analysis with this model, which we call the *log-Fourier model*. We will find in the course of the analysis that, just as the Rayleigh statistic arose naturally as the sufficient statistic for the log-sinusoid model, so something like the Z_n^2 family of statistics (a harmonic sum of S values) will arise as approximate sufficient statistics for a subclass of log-Fourier models. The Bayesian calculation will thus elucidate the model assumptions implicit in the use of harmonic sums, and tell us how to convert them into probability statements about the existence of a signal and its frequency and shape. It will also suggest some generalizations to Z_n^2, although these may not be computationally feasible for many datasets.

But before going on, we might anticipate some drawbacks to this model. First, although adding additional κ coefficients greatly enlarges the accessible range of lightcurve shapes, there is no simple relationship between the coefficients and simple lightcurve characteristics (such as background level or pulse width). We have gained more control over the lightcurve shape, but it is *complicated* control. Second, as in the log-sinusoid case, integrations over the κ coefficients must be done numerically. Thus adding only one or two more terms can drastically increase the computational burden. A possible fix to this is to simply fix all κ values (say, at $\kappa = 1$), and let only the phases vary. Such a model is numerically tractable and describes a wide array of shapes, but may not describe the *physically relevant* array of shapes, particularly if the detector happens to have a significant background rate. We show below, however, that this constraint is implicitly assumed in tests that use simple harmonic sums, like the Z_n^2 statistic. So although we won't be able to go too far with this model analytically, we can at least treat a constrained version of it that may be useful as a Bayesian counterpart to Z_n^2.

These discouraging remarks are offered as motivation to the reader to come up with alternative models that combine numerical tractibility with physical relevance. There is a lot of potential for progress here, if only someone can come up with a clever enough model.

5.1. *The Normalized Log-Fourier Model*

The log-Fourier normalized rate with H harmonics is,

$$\rho(t) = \frac{1}{I(\boldsymbol{\kappa}, \boldsymbol{\phi})} \exp\left[\sum_{\alpha=1}^{H} \kappa_\alpha \cos(\alpha\omega t - \phi_\alpha)\right], \tag{5.2}$$

where $\boldsymbol{\kappa}$ denotes the set of κ_α coefficients; $\boldsymbol{\phi}$ denotes the set of phases, ϕ_α; and $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$ is a normalization constant given by,

$$I(\boldsymbol{\kappa}, \boldsymbol{\phi}) = \int_0^{2\pi} d\theta \, \exp\left[\sum_{\alpha=1}^{H} \kappa_\alpha \cos(\alpha\theta - \phi_\alpha)\right]. \tag{5.3}$$

Two serious weaknesses of this model are the lack of an analytical expression for $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$, and the explicit dependence of this integral on the phases. This forces us to do an extra numerical integral for every value of $\boldsymbol{\kappa}$ considered, and prevents rigorous analytical marginalization of the phases. We'd love to learn of any cute tricks that let us quickly (analytically?!) evaluate the necessary integrals!

For priors, we will use the same priors that were used in for the log-sinusoid model: flat priors for phases, a Jeffreys prior for ω (i.e., $\propto 1/\omega$), and independent flat priors for each of the κ_α. The complexity of the model makes the physical significance of the prior for $\boldsymbol{\kappa}$ unclear; we choose this prior simply for definiteness and simplicity.

We denote the information specifying a log-Fourier model with H harmonics by M_H. When $H = 1$, we recover the log-sinusoid model discussed previously.

5.2. *The Joint Marginal Distribution*

Plugging this model into equation (3.19), we find the joint marginal distribution for ω, all phases, and all κ_α:

$$p(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa} \mid D, M_H) = C_H p(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa} \mid M_H) \left[I(\boldsymbol{\kappa}, \boldsymbol{\phi})\right]^{-N} \exp\left[\sum_\alpha \kappa_\alpha \sum_i \cos(\alpha\omega t_i - \phi_\alpha)\right], \tag{5.4}$$

where C_H is the normalization constant found by integrating the other factors over all $2H + 1$ parameters. For each α term in the exponent, we can make the same trigonometric substitutions that were made in equations (4.10) through (4.12), so that the joint marginal can be cast into the form,

$$p(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa} \mid D, M_H) = C_H p(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa} \mid M_H) \left[I(\boldsymbol{\kappa}, \boldsymbol{\phi})\right]^{-N} \exp\left[\sum_\alpha \kappa_\alpha S(\alpha\omega) \cos(\phi_\alpha - \phi'_\alpha)\right]. \tag{5.5}$$

Here S is exactly the same function as in equation (4.11), only here it is evaluated at several harmonic frequencies. Also, the values of ϕ'_α are given by H independent versions of equation (4.12), one for each harmonic.

5.3. *An Approximate Treatment of a Constrained Model*

If $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$ did not depend explicitly on $\boldsymbol{\phi}$, we could straightforwardly find the most probable phases and the marginal for ω and $\boldsymbol{\kappa}$, using the same methods as in the previous section. The most probable phases would be the ϕ'_α values, and the marginal for $\boldsymbol{\kappa}$ and ω would be proportional to $\prod_\alpha I_0[\kappa_\alpha S(\alpha\omega)]$.

We will proceed by making a poorly justified assumption (acknowledging that there may be a yet-to-be-found reasonable justification). We will simply set $\phi_\alpha = \phi'_\alpha$, and study the joint marginal conditional on this assumption, which we call the *phase-conditioned marginal.* To the extent that $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$ varies weakly with $\boldsymbol{\phi}$, this corresponds to conditioning on the best-fit value of $\boldsymbol{\phi}$. I haven't been able to prove that $I(\boldsymbol{\kappa}, \boldsymbol{\phi})$ has weak enough dependence on $\boldsymbol{\phi}$ to justify this (I suspect it doesn't), but even if it does, this alone does not justify the approximation. We have to worry about possible correlations of other parameters with $\boldsymbol{\phi}$. In particular, it is clear that the width of the joint marginal in ω, and the location and the width in $\boldsymbol{\kappa}$, depend on $\boldsymbol{\phi}$. So we expect this "approximation" not to affect our best-fit frequency estimates, but to lead us to underestimate our uncertainty for ω. The approximation corrupts our $\boldsymbol{\kappa}$ estimates to an unknown degree. Finally, the approximation also corrupts our signal detection results by an unquantified amount.

Despite these drawbacks, we pursue this approximation for two reasons. First, in the Gaussian spectrum analysis case, we know that a similar assumption underlies use of the Lomb-Scargle periodogram. This is clear from the "least-squares" derivations in Lomb (1976) and Scargle (1982), where an unknown phase is simple set equal to its least squares value, with no account taken of the uncertainty of this phase. In fact, the equations are essentially the same as ours, so we are conditioning on least-squares phase estimates, which are often okay in the Gaussian case, but may not be okay in the Poisson case. (By the way, Bretthorst (1988) provides a Bayesian counterpart to the Lomb-Scargle work that explicitly and analytically accounts for the phase uncertainty.) Second, we will find that this approximation, combined with a constraint on $\boldsymbol{\kappa}$, leads to a conditioned marginal whose sufficient statistic is a harmonic sum similar to Z_n^2. Thus this approximation will give us insight into what is being implicitly assumed when we use Z_n^2.

The phase-conditioned marginal is,

$$p(\omega, \boldsymbol{\phi} = \boldsymbol{\phi}', \boldsymbol{\kappa} \mid D, M_H) = C_H p(\omega, \boldsymbol{\phi}, \boldsymbol{\kappa} \mid M_H) \, [I(\boldsymbol{\kappa}, \boldsymbol{\phi}')]^{-N} \exp\left[\sum_\alpha \kappa_\alpha S(\alpha\omega)\right]. \tag{5.6}$$

If we further constrain the model by setting $\kappa_\alpha = 1$, then the conditioned marginal for ω is given by,

$$p(\omega, \boldsymbol{\phi} = \boldsymbol{\phi}', \boldsymbol{\kappa} = 1 \mid D, M_H) \propto e^{z_H}, \tag{5.7}$$

where,

$$z_H = \sum_\alpha S(\alpha\omega). \tag{5.8}$$

This statistic is similar to Z_n^2 with $n = H$, which in our notation is given by,

$$Z_H^2 = \frac{2}{N} \sum_\alpha S^2(\alpha\omega). \tag{5.9}$$

We can clarify the relationship between the two of these statistics by noting that,

$$z_H^2 = \frac{N}{2} Z_H^2 + \sum_\alpha \sum_{\beta \neq \alpha} S(\alpha\omega) S(\beta\omega). \tag{5.10}$$

That is, z_H^2 has the quadratic S terms considered in Z_H^2, plus additional *bilinear* S terms. Recall that S^2 is a sum of squared cosines and sines of the data points, as shown in equation (4.11). Thus if we take our data set, cut it in half in time, and calculate Z_H^2 for each half, the value of Z_H^2 for the entire data set is given simply by summing the two values (up to a constant factor). On the other hand, the bilinear term in equation (5.10) indicates that z_H^2 for the whole data set is *not* proportional to the sum of z_H^2 over parts of the data set. Essentially, Z_H^2 computes an incoherent sum of Fourier power at harmonic frequencies, while z instead sums amplitudes rather than powers, and thus contains phase information not used by Z_H^2.

We might summarize what we have learned about Z_n^2 as follows. First, it is conditioned on particular values for the phases that are good estimates for the phases in the Gaussian (least squares) case, but that may not be good estimates in the Poisson case. This conditioning ignores the effects that phase uncertainty has on other parameters. Second, Z_n^2 may tacitly assume fixed harmonic ratios (but we have not found Bayesian estimates for κ_α, and these may lead to a similarly weighted sum). Finally, Z_n^2 does not use all of the phase information that is used in a Bayesian analysis with a log-Fourier model. We thus suspect that the Bayesian analysis will more easily detect real signals of various shapes than does Z_n^2; but further calculations are required to justify this suspicion. This motivates us to more fully treat the log-Fourier model, relaxing some of the assumptions made here. But this will have to wait for a later memo (and for some inspiration regarding the integrals!).

6. INCLUDING POSITION INFORMATION

We have presumed all along that the data consist only of the arrival times of the events. However, many instruments also provide positional information for each event, so that the data consist of triples, (x_i, y_i, t_i). If there is a background rate, the position information can help us distinguish possible source events from background events. Currently, this information is only crudely incorporated, by throwing out all data outside of a specified radius from a known or best-fit position. Something more sophisticated, which weights the points according to the point spread function, is likely to be better. We outline in broad-brush the Bayesian approach here. Similar (but not explicitly Bayesian) considerations are being pursued by Dieter Hartmann and Larry Brown at Clemson University. We don't go into too much detail regarding specific models, because a full analysis of such data gets computationally burdensome pretty quickly. It thus may not be practical for searching for possible signals of unknown frequency. However, if we are searching for a signal of *known* frequency (such as a radio pulsar counterpart), or if we have detected a signal and obtained a frequency estimate from a simpler method, the likelihood function described here may be quite usable.

With position information available, the relevant model rate must be a function of position as well as time; we denote it by $r(x, y, t)$. As already noted, the position information is important because it helps us distinguish a possible signal from background. Thus from the outset we will explicitly decompose r into background and signal parts. We will presume the background rate is constant in time (it is easy, at least in principle, to generalize to time-dependent cases) and that it is known as a function of position; we denote the background rate per unit time per unit x–y area by $b(x, y)$. It may have unknown parameters that we must infer from separate background measurements or from a prior assumptions about the signal+background data (e.g., that the background has a known spatial distribution, and that some annulus contains only background and thus sets its scale). But for simplicity we assume here that it is perfectly known.

We presume the point spread function for the instrument is known, so that that a source with flux $\Phi(t)$ in direction $\mathbf{n}$ produces a count rate at detector position (x, y) equal to $k(x, y \mid \mathbf{n})\Phi(t)$. In this notation, then, k has units of physical area per x–y area. Thus the total event rate is,

$$r(x, y, t) = b(x, y) + k(x, y \mid \mathbf{n})\Phi(t). \tag{6.1}$$

We will presume here that the source direction is known, though of course it could be a free parameter, or it could have a nontrivial prior from other imprecise observations.

Our task is to make inferences about models for the time-dependent flux, $\Phi(t)$. We might use a log-sinusoid or log-Fourier model for $\Phi(t)$; however, we will find that the presence of the $b(x, y)$ term in equation (6.1) largely ameliorates the benefits of these models. We point out that it may be relevant to explicitly introduce "background" (i.e., constant) terms in $\Phi(t)$, but that we must necessarily distinguish these from the instrumental background rate, $b(x, y)$.

Denote the information specifying a model for $\Phi(t)$ with parameters $\mathcal{P}$ by the symbol M. To make inferences, we need the likelihood function. We can calculate it from the Poisson distribution using a three-dimensional generalization of the derivation in § 3 of GL. We discretize (x, y, t) space into small cells of size $\delta x\delta y\delta t$ such that each cell contains either no event or one event. The data can then be described as a list of event detections, d_i, denoting the cells containing one event, and a list of nondetections, $\bar{d}_j$, denoting empty cells. The probability for the data conditioned on M and $\mathcal{P}$ is,

$$\mathcal{L}(\mathcal{P}) = \prod_{i=1}^{N} p(d_i \mid \mathcal{P}, M) \prod_j p(\bar{d}_j \mid \mathcal{P}, M) \tag{6.2}$$

From the Poisson distribution, the detection and nondetection probabilities are,

$$p(d_i \mid \mathcal{P}, M) = \exp\left[-\int_{\delta t_i} dt \int_{\delta x_i} dx \int_{\delta y_i} dy\, r(x, y, t)\right], \tag{6.3}$$

and,

$$p(\bar{d}_j \mid \mathcal{P}, M) = \left[\int_{\delta t_j} dt \int_{\delta x_j} dx \int_{\delta y_j} dy\, r(x, y, t)\right] \exp\left[-\int_{\delta t_j} dt \int_{\delta x_j} dx \int_{\delta y_j} dy\, r(x, y, t)\right], \tag{6.4}$$

where $\delta x_i \ldots$ denote the intervals associated with each detection and nondetection datum. When we insert these into equation (6.2), the arguments of the exponents add to give the integral of $r(x, y, t)$ over all (x, y, t) in the observations. If we denote the total background rate by,

$$B = \int dx \int dy\, b(x, y), \tag{6.5}$$

and the total detector area for flux from direction $\mathbf{n}$ by,

$$A_{\rm det}(\mathbf{n}) = \int dx \int dy\, k(x, y \mid \mathbf{n}), \tag{6.6}$$

then the total predicted event rate is,

$$\int dx \int dy \int dt\ r(x, y, t) = BT + A_{\rm det}(\mathbf{n}) \int dt\, \Phi(t), \tag{6.7}$$

where T is the total "live time" duration of the observations. Using this, and taking the discretized intervals to be small compared to the scale of variation of r, equation (6.2) takes the form,

$$\mathcal{L}(\mathcal{P}) = e^{-BT} e^{-A_{\rm det}(\mathbf{n}) \int dt \Phi(t)} \prod_i \left[b(x_i, y_i) + k(x_i, y_i \mid \mathbf{n})\Phi(t)\right] \delta x \delta y. \qquad (6.8)$$

This is the likelihood function for arrival time series with position data. If we consider $\mathbf{n}$ to be unknown, it is unchanged. We can see how this likelihood function weights events with position: an event is counted as part of the signal, Φ, according to the ratio of $k(x_i, y_i \mid \mathbf{n})$ to $b(x_i, y_i)$.

We'll leave it as an exercise for the future to look for $\Phi(t)$ models that let us do a significant part of our calculations analytically. We only note here that $\Phi(t)$ appears in a weighted sum with $b(x, y)$, so that log-Fourier models for $\Phi(t)$ do *not* lead to analytically tractable likelihood functions. We also note that we could have reformulated this likelihood so that the background arose as an integral of k times a background rate on the sky; but this formulation does not change the basic fact that sums of background and signal terms appear in the product. There may be no useful analytically tractable models. However, if the parameter space is not too large (in particular, if a signal has already been detected with a simpler approach, so that its frequency is known), a completely numerical treatment using equation (6.8) may be reasonable.

Finally, some detectors may have binned location information, so that only the pixel number for each event is recorded, rather than the precise event location. For such data, we replace $k(x, y \mid \mathbf{n})\delta x \delta y$ with a direction-dependent response function for pixel (x, y), denoted by $k_{xy}(\mathbf{n})$ (x and y are now integer valued), and simularly replace $b(x, y)\delta x \delta y$ with the background rate in a pixel, b_{xy}. The event rate for pixel (x, y) is then,

$$r_{xy}(t) = b_{xy} + k_{xy}(\mathbf{n})\Phi(t). \qquad (6.9)$$

To calculate the likelihood function, we again discretize t so that each time interval contains zero or one event. Nothing changes from the previous calculation except that k and b are discretized, and the δx and δy factors are absorbed into the definitions of k_{xy} and b_{xy}. If the time dimension is also binned, such that some bins contain more than one event, the likelihood function *does* take a different form, but it still follows simply from the Poisson distribution, leading to a likelihood that is the location generalization of that discussed in Gregory and Loredo (1993).

REFERENCES

Bretthorst, G. L. 1988, Bayesian Spectrum Analysis and Parameter Estimation (New York: Springer-Verlag)

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

Gregory, P. C., & Loredo, T. J. 1993, in Maximum Entropy and Bayesian Methods, Paris, France, 1992, ed. A. Mohammad-Djafari & G. Demoment (Dordrecht: Kluwer), in press

Loredo, T. J. 1992a, in Statistical Challenges in Modern Astronomy, ed. E. Feigelson & G. Babu (New York: Springer-Verlag), 275

West, M. 1992, in Statistical Challenges in Modern Astronomy, ed. E. Feigelson & G. Babu (New York: Springer-Verlag), 299