On Detecting Periodicity in Astronomical Point Processes

Peter Bickel[1], Bas Kleijn[2] & John Rice[1]

[1] *University of California, Berkeley*

[2] *Vrije Universiteit Amsterdam*

Abstract. We consider the problem of detecting periodicity in the rate function of a point process or a marked point process, motivated by the problem of detecting γ-ray pulsars. The detection problem poses both theoretical and computational challenges. On the theoretical side, there are no compelling optimality results that dictate the choice of a detection algorithm and the properties of detection procedures can be quite difficult to analyze. On the computational side, searching over a range of frequency and frequency drift can be a daunting task, even for a record consisting of only a thousand or so events. We discuss a class of detection procedures, weighted quadratic test statistics arising from likelihood expressions, whose properties we can understand and which do not impose excessive computational burdens. We show how knowledge of the point spread function associated with photon arrivals can be incorporated to improve power. We show that if a search over frequencies is conducted by discretizing a frequency band, the discretization must be very fine and we discuss the use of integration over frequency bands as an alternative. We also discuss the use of extreme value theory in conjunction with simulation in assessing statistical significance for such a search.

1. Introduction

Much of astronomical data analysis involves the attempt to detect and characterize a variable, possibly periodic, signal in the presence of a relatively intense background or with sparse measurements. Periodic sources are known with periods ranging from milliseconds for pulsars (rotating magnetized neutron stars) to months or longer for eclipsing binary star systems. This paper is concerned with the problem of detecting source periodicity from a sequence of arrival times.

Thus, consider an idealized detection problem: photon arrival times $0 < t_1 < \cdots < t_n < T$ are recorded. The underlying process is Poisson with rate function $\lambda(t)$ and we wish to test whether $\lambda(t)$ is periodic or constant. Let us even assume initially that the period $P << T$ is known; as will be shown below, the detection problem is then essentially equivalent to detecting deviations from uniformity of the phased arrival times $u_i = t_i \bmod P$. It can be thus viewed as testing goodness of fit of the model that the u_i are i.i.d. uniform on the interval $[0, 1]$ versus the alternative that they come from some other distribution, and we will assume that $\lambda(t)$ is not specified, so the alternative distribution is not specified.

This seems like an ancient and tightly specified problem, and it is natural to ask if there is an optimal detection procedure. The short answer is that there is not: a detection algorithm optimal for one function $\lambda(t)$ will not be optimal for another function. No matter how clever you are, no matter how rich the dictionary from which you adaptively compose a detection statistic, no matter how multilayered your hierarchical prior, your procedure will not be globally optimal. The long answer is interesting and subtle; for a good account see Chapter 14 of Lehmann and Romano (2005), which includes a precise statement and discussion of the following result (Janssen 2000): alternative probability densities have components in an infinite number of directions (as in a Fourier expansion) and *any* test can achieve high asymptotic power against local alternatives for at most a finite number of directions. In other words, associated with any particular test is a finite dimensional collection of targets and it is only for such targets that it is highly sensitive. Fortunately, the number of such directions grows with the strength of the signal and we expect that for directions highly correlated with ones at which the test is aimed there will be substantial power. Bickel et al. (2006) discuss the construction of tests designed to concentrate power in a number of orthogonal directions. As Jannsen's result suggests, these tests have low power in directions orthogonal to the ones chosen.

In the case of γ-ray pulsars the actual situation is substantially more complicated than this idealization, because of a variety of factors:

- The frequency may be unknown and be anywhere in the range of about 1 to 40 Hz.
- The frequency of a rotation-powered pulsar decreases because the pulsar spins down as it loses energy.
- The frequency may jump discontinuously, due presumably to stellar quakes.
- As mentioned above, the pulsars are seen against a bright celestial foreground and the angular resolution of the γ-ray telescope is relatively poor. The point spread function is energy dependent.
- The arrival times of the photons must be corrected for the location of the satellite and of the earth in its orbit about the sun, and for the delay due to the gravitational potential. These corrections depend on the location of the source, which may not be precisely known.
- Photons from a source are not recorded at all times, but only during certain viewing periods. EGRET viewing periods were typically one to two weeks in duration and during a viewing period, detections from any particular point source were modulated by occulations by the Earth. Also, photons may not be recorded because of "dead times"—for a short time after the detection of a photon, the detector is dead, so there is a minimum interval between detected events.

The challenges are daunting; a heroic search effort (Chandler et al. 2001) did not reveal any hitherto unknown γ-ray pulsars.

This paper is mainly concerned with a class of procedures which target specified directions. Some of these procedures have already been proposed in the literature, but we attempt to give a unified perspective and provide analysis which gives insights into their properties. The remainder of the paper is organized as follows: In Section 2 we describe the construction of a score test at a fixed frequency, relate it to well-known classical tests, and consider its power. In Section 3, we propose a method for integrating over frequency bands in a search as an alternative to a discretized search and show how an eigenfunction expansion can be used to computational advantage. In Section 4 we describe how the procedures described in previous sections are naturally extended to incorporate frequency drift. In Section 5 we discuss the possibility of using extreme value theory to assess statistical significance in a blind search over a wide frequency range. Section 6 consists of some concluding remarks.

2. Score test at a fixed frequency

We first consider the simple situation in which the frequency f is specified and there is no drift. Let $\nu_0(t)$ be a probability density on $[0,1]$, extended periodically with Fourier expansion

$$\nu_0(t) = \sum_n \alpha_n e^{2\pi i n t} \tag{1}$$

If $\nu_\tau(t) = \nu_0(t+\tau)$

$$\nu_\tau(t) = \sum_n \alpha_n e^{2\pi i n t + 2\pi i n \tau} \tag{2}$$

Here α_n is the amplitude of the n-th harmonic and $n\tau$ is its phase. We consider testing the hypothesis that the rate is constant versus an alternative target with shape specified by $\nu_0(t)$ and frequency f. Let $\phi(t) = ft$ and consider a Poisson process of photon arrivals. Let the possibly energy dependent point-spread function of the detector be $w(z|e)$, where z is the spatial variable and e denotes energy. Our model for the rate function of the Poisson process is

$$\lambda(t|\theta,\tau,\mu,f) = \mu[(1-\theta) + \theta\nu_\tau(\phi(t))], \quad 0 \le \theta \le 1 \tag{3}$$

a mixture of background and periodic source. This can be directly modified to reflect periods of time during which the detector was not active (due to occultations for example) by setting $\lambda(t) = 0$ during those time intervals.

The likelihood given (t_j, e_j, z_j), the arrival times, energies, and locations of the photons is

$$L(\mu,\theta,f,\tau) = \mu^n \prod_j [(1-\theta) + \theta w_j \nu_\tau(\phi(t_j))] \exp[-\int_0^T \lambda(t|\theta,\tau,\mu,f)dt] \tag{4}$$

where $w_j = w(z_j|e_j)$. A classical generalized likelihood ratio test would entail fitting the light curve under the null and alternative models. A score test, or Rao test, (Lehmann and Romano 2005), has the advantage of not requiring the latter step and has essentially the same asymptotic properties as does the likelihood ratio test. To construct a score test, differentiate the log likelihood with respect to θ and evaluate at $\theta = 0$. This gives

$$\ell'(\mu,f,\tau) = \int_0^T [\nu_\tau(\phi(t)) - 1]dW(t) - \mu\int_0^T [\nu_\tau(\phi(t)) - 1]dt \tag{5}$$

where $W(t)$ places weight w_j at t_j. We will neglect the second term on the grounds that T is very large compared to $1/f$. (If it were not ignored, we would estimate μ under the null by $N(T)/T$. This would be important to do if T was not large compared to $1/f$.) The statistic then only depends on (f, τ) through

$$\begin{aligned} S(f,\tau) &= \int_0^T \nu_\tau(\phi(t))dW(t) && (6) \\ &= \int_0^T \sum_n \alpha_n e^{2\pi i n f t + 2\pi i n \tau} dW(t) && (7) \\ &= \sum_n \alpha_n A_n e^{2\pi i n \tau} && (8) \end{aligned}$$

where

$$\begin{aligned} A_n &= \int_0^T e^{2\pi i n f t} dW(t) && (9) \\ &= \sum_j w_j e^{2\pi i n f t_j} && (10) \end{aligned}$$

To construct an invariant test, we square $S(f, \tau)$ and integrate from 0 to 1 with respect to τ. Observe that the Fourier coefficients of $S(f, \tau)$ as a function of τ are $\alpha_n A_n$, so that by Parseval's theorem

$$\begin{aligned} \int_0^1 |S(f,\tau)|^2 d\tau &= \sum_n |\alpha_n A_n|^2 && (11) \\ &= \sum_n |\alpha_n|^2 |A_n|^2 && (12) \end{aligned}$$

Let $Q(f) = \sum_n |\alpha_n|^2 |A_n(f)|^2$. It is noteworthy that this test statistic only depends upon the amplitudes of the Fourier coefficients, but not their phases.

One advantage of this construction is that it incorporates recorded energies and directions of photon arrivals in a principled and natural way, avoiding the necessity of making hard cuts. In empirical tests, we have found a resulting increase in power.

2.1. Relationship to some classical tests

In the case that the weights w_j are all equal to 1, various choices of $|\alpha_n|$ lead to different goodness of fit tests against uniformity for arrival times which have been folded onto the circle (observe that (6) only depends on the arrival times t_j through $s_j = \phi(t_j) \bmod 1$ since $\nu(\cdot)$ is periodic). The statistic $Q(f)$ was proposed in (Beran 1969) and shown to be locally most powerful invariant in the direction $\nu()$ at the frequency f. Truncating the expansion after $n = 1$ gives Rayleigh's test. Setting $|\alpha_n|^2 = \pi n^{-2}$ gives Watson's test (an invariant version of Cramer-von Mises). If $|\alpha_n| = 1$, $n \leq m$ and 0 otherwise, the Z_m^2 test (Buccheri et al. 1983) results. The classic text, Mardia (1972), describes many of these tests and others as well; see also Chapter 14 of Lehmann and Romano (2005), where they are referred to as weighted quadratic test statistics.

2.2. Relationship to tests based on density estimation

Consider the un-weighted case. Then $S(f,\tau) = \sum_j \nu_0(\tau + s_j)$. Since $\nu_0(t)$ is a probability density, this can be viewed as a kernel density estimate, $\hat{m}(\tau)$, say. A goodness of fit test against the uniform on $[0,1]$ is

$$\int_0^1 [\hat{m}(\tau) - 1]^2 d\tau = \int_0^1 \hat{m}^2(\tau) d\tau - 1 \tag{13}$$

and $\int_0^1 \hat{m}^2(\tau) d\tau = Q(f)$. Using a rectangular kernel with support on $[0,h]$ would amount to a continuous version of a chi-square goodness-of-fit test with h^{-1} bins. A convenient way to construct kernels and their Fourier series is to view the kernels as spectral densities and the Fourier series as the corresponding autocorrelation functions. For example, for an AR(1) process

$$\nu_0(t) = \frac{1-\beta^2}{1+\beta^2 - 2\beta\cos(2\pi t)} \tag{14}$$

$$\alpha_n = |\beta|^n \tag{15}$$

Kernel density estimation involves a choice of bandwidth, which tends to 0 as the sample size increases. For the AR(1), as $\beta \to 1$, the bandwidth tends to 0, the kernel becomes more peaked around 0, and its spectrum spreads out into higher frequencies. For the uniform kernel, the bandwidth is h. The choice of a small bandwidth, appropriate for density estimation, would amount to targeting the test in the direction of a sharply spiked $\nu_0(\cdot)$ with substantial high frequency content, which would not be appropriate for detection of low frequency objects (see remarks on power below). In DeJager et al. (1989) bandwidth selection methods were used to adaptively choose a truncation point for a trigonmetric expansion; it is difficult to analyze the properties of this modification and to see what price is paid for adaptation, but the analysis in the next section does indicate that an effective density estimate may not be especially effective for the purpose of detection.

2.3. Null distribution and power

Since the $n=0$ term contains no information about harmonic content in the signal, we consider the test statistic $Q_T = T^{-1}\sum_{n\neq 0} |\alpha_n|^2 |A_n|^2$ in the unweighted case with an hypothesized frequency f_0. To derive expressions related to power, we need to calculate

$$E|A_n|^2 = E\left|\int_0^T e^{2\pi i n f_0 t} dN(t)\right|^2 \tag{16}$$

$$= E\int_0^T \int_0^T e^{2\pi i n f_0 (t-s)} dN(t) dN(s) \tag{17}$$

using

$$E[dN(t)dN(s)] = \lambda(t)\lambda(s)dsdt + \lambda(t)\delta(s-t)dsdt \tag{18}$$

Under the null, $\lambda(t) = \mu$. Here and later we will use the identity

$$\int_0^T e^{i\omega t} dt = T e^{i\omega T/2} \frac{\sin(\omega T/2)}{\omega T/2} \tag{19}$$

$$= I_T(\omega) \tag{20}$$

The first term of (18) then gives

$$\mu^2 \left| \int_0^T e^{2\pi i n f_0 t} dt \right|^2 = \mu^2 \frac{\sin^2(\pi n f_0 T)}{(\pi n f_0)^2} \tag{21}$$

and the second term gives μT. Thus, under the null,

$$E(Q_T) \approx \mu \sum_{n \neq 0} |\alpha_n|^2 \tag{22}$$

For later use, we write the expression above as $E_0 = \sum_{n \neq 0} E_{0n}$. Also, under the null $2T^{-1}|A_n|^2$ is approximately distributed as a μ times a chi-squared random variable with two degrees of freedom, so $Var(T^{-1}|A_n|^2) \approx \mu^2$. From the orthogonality of the complex exponentials, $|A_n|^2$ is approximately independent of $|A_m|^2$, $n \neq m$. Then

$$Var(Q_T) \approx \mu^2 \sum_{n \neq 0} |\alpha_n|^4 \tag{23}$$

Now consider the behavior of the statistic under an alternative for which the expected total number of events is the same but which also contains a periodic component. Suppose that $\lambda(t) = \mu[(1-\theta) + \theta\gamma(t)]$ where $\gamma(t)$ is periodic with frequency f, with the actual light curve $\gamma(t)$ not necessarily equal to the target $\nu(t)$ and f not necessarily equal to the specified frequency f_0. Let $\gamma(t)$ have the Fourier expansion

$$\gamma(t) = \sum_k \gamma_k e^{2\pi k f_0 t} \tag{24}$$

Let $f_0 = f + \delta/T$. The second term in (18) again makes a contribution to $E(|A_n|^2)$ equal to E_0. For the first term, let

$$\omega_k = 2\pi n f + 2\pi k f + 2\pi k \delta/T \tag{25}$$

Then

$$E|A_n|^2 - E_{0n} = \mu^2 \left| (1-\theta) I_T(2\pi n f) + \theta \sum_k \gamma_k I_T(\omega_k) \right|^2 \tag{26}$$

The leading contribution is that from the term in the sum for which $k = -n$,

$$\begin{aligned} E|A_n|^2 - E_{0n} &\approx \mu^2\theta^2 \left| \gamma_{-n} e^{-\pi i n \delta} T \frac{\sin(\pi n \delta)}{\pi n \delta} \right|^2 && (27) \\ &= \mu^2\theta^2 T^2 |\gamma_n|^2 \left| \frac{\sin(\pi n \delta)}{\pi n \delta} \right|^2 && (28) \end{aligned}$$

Thus

$$E(Q_T) - E_0 \approx \mu^2\theta^2 T \sum_{n \neq 0} |\alpha_n|^2 |\gamma_n|^2 \left| \frac{\sin(\pi n \delta)}{\pi n \delta} \right|^2 \tag{29}$$

The n-th harmonic will only contribute to the power if $|\gamma_n|$ is substantial and if $n\delta$ is small. Asymptotically, power will be lost unless $\delta = o(T^{-1})$. Another way to understand the necessity of oversampling is to observe that unless the specified frequency is within $o(T^{-1})$ of the true frequency, the first and last photons will be out of phase.

Figure 1 shows $Q(f)$ (Rayleigh statistic) in units of standard deviations near the strong first harmonic of Geminga. If it were computed only at frequencies of the form k/T, the test statistic would be about 70 standard deviations, whereas the value at the peak is about 120 standard deviations. The signal is so overwhelmingly strong that it would not make a practical difference in this case, but for a weaker signal a factor of two could well be important in a blind search, so oversampling (relative to T^{-1} spacing) is necessary.

The power of the test is indicated by the ratio of (29) to the standard deviation under the null (23). It thus depends on the strength of the signal (μ, θ), the duration of observation (T), the correlation between the target shape $(\{|\alpha_n|^2\})$ and the signal shape $(\{|\gamma_n|^2\})$, and the frequency resolution δ. Inclusion of higher harmonics in the test statistic is thus only beneficial if the signal contains substantial power in those harmonics and if the sampling is fine. Otherwise, the cost in variance of including them may more than offset potential gains. Viewed from this perspective, tests based on density estimation with a small bandwidth are not attractive unless the light curve has substantial high frequency components and the target frequency is very close to the actual frequency.

We can tie this discussion to our comments on power in the introduction. A direction in the alternative corresponds to a specification of shape and frequency, f and $\{|\gamma_n|^2\}$, in relation to those at which the test is targeted, f_0 and $\{|\alpha_n|^2\}$. Our analysis shows that, as expected, we do obtain power for a shape and frequency close to those at which we have aimed. However, the analysis also shows that whereas convergence of $\{|\alpha_n|^2\}$ to $\{|\gamma_n|^2\}$ at a rate $r(T)$ implies that power as measured by signal to noise ratio increases at the same rate, $f - f_0$ must tend to 0 at rate $r(T)/T$ to obtain a power increase of order $r(T)$.

3. Searching over frequencies: integration as an alternative to discretization

We have seen above that when the frequency is not known precisely, it is desirable to search over a grid which is fine compared to the spacing of the natural Fourier frequencies. As an alternative to discretization, we propose integrating over a range of frequencies around the proposed frequency f_0. Integration should also be less sensitive to frequency glitches and drift. Thus consider integrating $S^2(f)$ using a symmetric density, $g(f)$, centered at f_0. We have to evaluate

$$\bar{A_n} = \int |A_n(f)|^2 g(f) df \tag{30}$$

$$= \int_0^T \int_0^T \int_f \exp(2\pi i n f(u-v)) g(f) df dW(u) dW(v) \tag{31}$$

$$= \int_0^T \int_0^T \hat{g}(2\pi n(u-v)) dW(u) dW(v) \tag{32}$$

$$= \sum_j \sum_k w_j w_k \hat{g}(2\pi n(t_j - t_k)) \tag{33}$$

where

$$\hat{g}(t) = \int_{-\infty}^{\infty} e^{itf} g(f) df \tag{34}$$

The drawback of the expression (33) is that its evaluation requires a number of operations that is proportional to the square of the number of photons, and this may impose a huge computational burden if many such integrations are to be carried out in a search over a very broad frequency range. This cost can be reduced by diagonalizng the quadratic form. The density $g(f)$ is a translation by f_0 of the density $g_0(f)$ which is centered at 0. The kernel $\hat{g}_0(2\pi n(u-v))$, $0 \le u, v \le T$ can be expanded in eigenfunctions as

$$\hat{g}_0(2\pi n(u-v)) = \sum_k \mu_{kn} \psi_{kn}(u) \psi_{kn}(v) \tag{35}$$

Since $\hat{g}(t) = e^{itf_0} \hat{g}_0(t)$,

$$\hat{g}(2\pi n(u-v)) = \sum_k \mu_{kn} e^{2\pi inuf_0} \psi_{kn}(u) e^{-2\pi invf_0} \psi_{kn}(v) \tag{36}$$

Using this representation in (33) gives

$$\bar{A}_n = \sum_k \mu_{kn} \left| \int_0^T e^{2\pi intf_0} \psi_{kn}(t) dW(t) \right|^2 \tag{37}$$

$$= \sum_k \mu_{kn} |V_{kn}|^2 \tag{38}$$

where

$$|V_{kn}|^2 = \left| \sum_j w_j e^{2\pi int_j f_0} \psi_{kn}(t_j) \right|^2 \tag{39}$$

The test statistic is thus

$$\bar{Q} = \sum_{n \neq 0} |\alpha_n|^2 \bar{A}_n \tag{40}$$

$$= \sum_{n \neq 0} |\alpha_n|^2 \sum_k \mu_{kn} |V_{kn}|^2 \tag{41}$$

Note that the evaluations of the eigenvectors do not depend on f_0.

We now consider the power of the test statistic $\bar{Q}_T = T^{-1}\bar{Q}$. Under the null, the expectation is thus the same as in the fixed frequency case above. Under the alternative, since $\delta = \delta(f) = T(f - f_0)$, from (29), and setting $g(f) = Tg_0(T(f - f_0))$

$$E\bar{Q}_T - E_0 = \mu^2 \theta^2 T \sum_{n \neq 0} |\alpha_n|^2 |\gamma_n|^2 \int \left| \frac{\sin(\pi nT(f - f_0))}{\pi nT(f - f_0)} \right|^2 g(f) df \tag{42}$$

$$= \mu^2 \theta^2 T \sum_{n \neq 0} |\alpha_n|^2 |\gamma_n|^2 \int \left| \frac{\sin(n\pi u)}{n\pi u} \right|^2 g_0(u) du \tag{43}$$

which should be compared to (29). Examination of this expression shows that power is still lost, especially for high frequencies. For example, if $g(f)$ is uniform on $f_0 \pm kT^{-1}$ the contribution from $|\gamma_n|^2$ is diminished by a factor of n^{-1} for large n, which can be seen using

$$\int_{-\infty}^{\infty} \left| \frac{\sin(\pi u)}{\pi u} \right|^2 du = 1 \tag{44}$$

Generally, as the support of $g(f)$ is increased, the power decreases.

3.1. Prolate spheroidal wave functions

The eigenfunctions arising from uniform weighting over a frequency interval are prolate spheroidal wave functions. They satisfy

$$\int_{-1}^{1} \frac{\sin(c(x-y))}{c(x-y)} \psi^S_{k,c}(y) dy = \frac{\pi}{c} \alpha_{k,c} \psi^S_{k,c}(x) \tag{45}$$

Suppose that $g_0(f)$ is uniform on $[-a, a]$. Then the characteristic function is

$$\hat{g}(t) = \frac{\sin(at)}{at} \tag{46}$$

and we wish to find the eigenfunctions that satisfy

$$\int_{0}^{T} \frac{\sin(2\pi na(u-v))}{2\pi na(u-v)} \psi_{kn}(v) dv = \mu_{kn} \psi_{kn}(u) \tag{47}$$

With the change of variables, $x = 2u/T - 1$, $y = 2v/T - 1$, this becomes

$$\int_{-1}^{1} \frac{sin(\pi naT(x-y))}{\pi naT(x-y)} \psi_{kn}(\frac{T}{2}(y+1)) dy = \frac{2}{T} \mu_k \psi_{kn}(\frac{T}{2}(x+1)) \tag{48}$$

Let $c = \pi naT$. Then we can relate to the spheroidal wave functions by observing that the equation above is satisfied if

$$\psi_{kn}(\frac{T}{2}(y+1)) = \psi^S_{k,c}(y) \tag{49}$$

$$\mu_{kn} = \frac{\pi}{c} \alpha_{k,c} \tag{50}$$

That is, the desired eigenfunctions are

$$\psi_{kn}(v) = \psi^S_{k,c}(\frac{2v}{T} - 1) \tag{51}$$

We are particularly interested in choices of the form $a = k/T$, $c = \pi nk$. For example, $k = 1$ results if we choose to integrate over a range of width $2/T$. Table 1 lists the eigenvalues for small values of k. They decay quite rapidly, so the quadratic form can be approximated well by only a few terms.

Table 1. Eigenvalues

$k=4$	$k=2$	$k=1$
0.2487	0.4883	0.7817
0.2482	0.3750	0.2065
0.2391	0.1234	0.0116
0.1815	0.0127	0.0002
0.0703	0.0006	0.0000
0.0112	0.0000	0.0000
0.0009	0.0000	0.0000
0.0000	0.0000	0.0000

4. Incorporating frequency drift

Suppose that $\phi(t) = ft + \frac{1}{2}\dot{f}t^2$. Modifying the notation above, we have the analogue of (12)

$$Q(f,\dot{f}) = \sum_{n\neq 0} |\alpha_n|^2 |A_n(f,\dot{f})|^2 \tag{52}$$

where

$$A_n(f,\dot{f}) = \int_0^T e^{2\pi i n f t + \pi i n \dot{f} t^2} dW(t) \tag{53}$$

Consider integrating with respect to densities g and h centered at f_0 and $\dot{f}_0$

$$\begin{aligned}
\bar{Q}_n &= \int_f \int_{\dot{f}} |A_n(f,\dot{f})|^2 g(f) h(\dot{f}) df d\dot{f} \\
&= \int_0^T \int_0^T \int_f \int_{\dot{f}} e^{2\pi i n f(u-v)} e^{\pi i n \dot{f}(u^2-v^2)} g(f) h(\dot{f}) df d\dot{f} dW(u) dW(v) \\
&= \int_0^T \int_0^T \hat{g}(2\pi n(u-v)) \hat{h}(\pi n(u^2-v^2)) dW(u) dW(v) \qquad (54)
\end{aligned}$$

We diagonalize the two kernels as before:

$$\hat{g}(2\pi n(u-v)) = \sum_k \mu_{kn} e^{2\pi i n f_0 u} e^{-2\pi i n f_0 v} \psi_{kn}(u)\psi_{kn}(v) \tag{55}$$

$$\hat{h}(\pi n(u^2-v^2)) = \sum_\ell \eta_{\ell n} e^{\pi i n \dot{f}_0 u^2} e^{-\pi i n \dot{f}_0 v^2} \varphi_{\ell n}(u)\varphi_{\ell n}(v) \tag{56}$$

Note that in the second equation, the eigenvalues and eigenfunctions are those of the operator whose kernel is $K(u,v) = \hat{h}_0(\pi n(u^2-v^2))$, $0 \le u, v \le T$. Using these representations (54) becomes

$$\bar{A}_n = \sum_k \sum_\ell \mu_{kn}\eta_{\ell n} \left| \int_0^T e^{2\pi i n f_0 t + \pi i n \dot{f}_0 t^2} \psi_{kn}(t)\varphi_{\ell n}(t) dW(t) \right|^2 \tag{57}$$

The test statistic is thus

$$\bar{Q} = \sum_{n \neq 0} |\alpha_n|^2 \bar{A}_n \tag{58}$$

$$= \sum_{k,\ell,n} \mu_{kn} \eta_{\ell n} |V_{kln}|^2 \tag{59}$$

where

$$V_{k\ell n} = \sum_j w_j e^{2\pi i n f_0 t_j} e^{\pi i n \dot{f}_0 t_j^2} \psi_{kn}(t_j) \varphi_{\ell n}(t_j) \tag{60}$$

In a search over a broad range one would tile the $(f, \dot{f})$ plane and calculate these statistics for each tile element, producing something akin to a two dimensional power spectrum. We illustrate the idea with data from the Vela pulsar for a single EGRET viewing period. We used all photons with energy greater than 25 MeV and within a conesize of 2 standard deviations. This resulted in 7208 events, which were weighted by an approximate energy dependent Gaussian psf. The pulse profile of Vela is very sharp, containing two peaks. One should be able to find it in a drunken blind search. The frequency is about 11.2 Hz and the drift is about -1.57×10^{-11} Hz/s. We tiled the $f \times \dot{f}$ plane in a narrow band around the known frequency , with the tiles being of size $df = 10^{-5} Hz$ and $d\dot{f} = 2.5 \times 10^{-12} Hz/s$. In each bin we calculated a test statistic which was the equally weighted sum of the contribution from the main frequency and the first harmonic, integrated using a uniform weight function. Results are displayed in Figure 2 below. The figure displays the values of the test statistic in units of standard deviations at locations for which the values were more than four standard deviations. The structure is interesting. At the true frequency, which is in the center of the frequency range, the test statistic is maximal near the true drift, as we would expect. But observe how the smaller values of the test statistic fan out into higher frequencies for large drift, apparently reflecting some kind of correlation structure among the statistics. The ridge corresponding to the true frequency is flanked by parallel ridges; consideration of the frequency offset reveals that these are caused by gaps in the records due to occultations by the Earth.

5. Assessing significance

If a search is made over only a narrow frequency band, statistical significance can be addressed in a straightforward way by simulation. Suppose that the observation period is composed of segments during which the detector is active (gaps between these segments can be due to instrumentation protocols and occultations by the Earth). Suppose that in the j-th segment which is of duration T_j, there are n_j arrivals. The null distribution of the test statistic can then be simulated by distributing events uniformly over the respective intervals (we are thus simulating the null distribution conditional on the number of arrivals in each segment). This can be done many times and the actual observed value of the test statistic can be compared to the empirical distribution of simulations under the null.

Such a simulation is however impractical if the search is over a broad frequency range and drift is allowed as well, since calculating the test statistic just once might take several days. Thus, some analytical approximation must be used, perhaps in conjunction with moderate simulation. For example, a standard approach is to approximate the distribution of the periodogram at frequency of the the form kT^{-1} by a chi-square distribution and to bound the maximum by the Bonferroni inequality—the result would be declared significant at level α if the p-value of the maximum were less than α/M, where M is the total number of frequencies. The potential weakness of this approach is that the limiting chi-square approximation may well be better in the center of the distribution than in the extreme tails where it is being used. In the case of the integrated statistic, the situation is even more complicated—the limiting distribution can be shown to be that of a weighted sum of chi-square random variables. Although an analytical expression can derived for the characteristic function, the resulting approximation may still be dubious in the extreme tails. It is thus worthwhile to consider alternative approximations. We have made some preliminary investigations of the use of classical extreme value theory in conjunction with affordable simulation and initial results are promising.

Consider $M_n = \max\{T_1, T_2, \dots, T_n\}$ where the T_i are iid random variables. Classical theory gives that if M_n has a limiting distribution it is of one of three types. In the case of weighted sums of chi-square random variables, the limiting distribution will be of Gumbel type. If F is the cdf of T, then there is a normalizing sequence (a_n, b_n) such that

$$F^n(a_n t + b_n) \to \exp[-\exp(-x)] \tag{61}$$

The Gumbel limit depends on the tail behavior of F. The existence of a limiting distribution implies that in a region of the tail around the expected maximum

$$\log[-\log F(t)] \sim -\lambda t + \beta \tag{62}$$

For large t, $F(t)$ is close to 1 and the approximation $\log[F(t)] \approx 1 - F(t)$ is good. Using this in (62) gives

$$P(T > t + s | T > t) \approx e^{-\lambda s} \tag{63}$$

This again holds in a region around the expected maximum. The conditional distribution is thus approximately exponential with mean $\alpha = 1/\lambda$. See Pickands (1975) for more detailed discussion.

We thus approximate tail probabilities by

$$P(T > t + s) \approx P_n(T > t) e^{-s/\alpha_n} \tag{64}$$

where P_n is the empirical measure based on simulation and α_n is the mean of the corresponding empirical conditional distribution. A related procedure is reported in Breiman et al. (1988).

To explore the feasibility of such a procedure and in particular to examine the sensitivity to the choice of t, we performed a modest empirical experiment using data from a single EGRET viewing period of the Geminga pulsar, for which the drift is so small that it does not need to be taken into account.

The test statistic was an equally weighted sum of the main and first harmonic integrated with a uniform weight over frequency bins of width $8T^{-1} = 6.6 \times 10^{-6}$ Hz spanning the range 1 to 40 Hz. There were thus about 5.7×10^6 values of the test statistic (i.e that many frequency bins). The test statistic was almost 50 standard deviations, leaving no room for reasonable doubt. The calculation took about one day on a single processor. The data were then scrambled once and the null distribution simulated as described above.

Figure 3 shows the empirical result: $\log[-\log F_n(t)]$ versus t for the simulated null distribution. Qualitatively, the simulated distribution was stable out to around six standard deviations, and the linear approximation looks reasonable. Table 2 shows results from the approximation (64) for varying cutoffs t. The results are reasonably consistent. We feel they would provide a plausible guide for assessing significance. For example, a short calculation shows that in order for a Bonferroni-corrected p-value to be less than .01, a test statistic of about 11 standard deviations or more would be required.

Table 2. Conditional means and tail probability approximations for various cuttoff values t.

t	α	$P(T > 10)$	$P(T > 15)$
3	.4852	3.0×10^{-9}	9.9×10^{-14}
4	.4548	1.3×10^{-9}	2.1×10^{-14}
5	.4438	9.4×10^{-10}	1.2×10^{-14}

6. Concluding remarks

In summary, we have investigated the theoretical properties of a class of detection tests that includes several commonly known special cases. We have shown that discretization of frequency finer than T^{-1} is necessary to attain good power, especially when the signal has substantial high frequency content. We have proposed integration over frequency bands as an alternative to fine discretization and have shown how this can be accomplished with a number of operations proportional to the number of events, via eigenfunction expansions. We have tentatively explored the use of extreme value theory in assessing significance of a search over a broad frequency range.

We have not yet fully implemented these results. One area that remains to be explored is that of developing a search strategy for covering a broad $(f, \dot{f})$ range. To speed computation, such a search might be implemented by first constructing a relatively coarse tiling of the $(f, \dot{f})$ plane and integrating over each tile. This coarsening would cause a loss of power, so the tiling would next be refined in regions for which the values of the test statistics were large. Alternatively, the observation period could be segmented in blocks with test statistics combined (incoherently) from the blocks. Coherent searches could then be carried out in neighborhoods of large test statistics.

We have not touched on Bayesian detection methods (Gregory and Loredo 1992). The natural Bayesian construction in the context above would be to place a prior distribution on shape via a prior on the Fourier coefficients of $\nu_0(t)$. For example, the coefficients could be modeled as independent mean zero Gaussians, with decaying variances. See Verdinelli and Wasserman (1998) for such a construction and discussion of the computation of the Bayes factor, and for an early proposal on Bayesian density estimation see Brunk (1978). The computations needed for a broad-band search are formidable indeed.

In nonparametric, or high dimensional problems, the choice of prior has important consequences, unlike in low dimensional parametric problems in which the prior can be taken to be uninformative and is typically overwhelmed by the likelihood. There is an evolving literature on frequentist robustness of nonparametric Bayes procedures, for example Ghosh and Ramamoorthi (2003), and many issues are not well understood. In nonparametric problems the prior is informative and performs crucial smoothing. Bayes factors are particularly sensitive to the choice of priors. The prior is a useful device, but in an infinite dimensional setting it does not represent prior opinion or a state of knowledge in the same sense in which it would for coin tossing, making interpretation difficult. To understand the performance of a Bayesian detection method, and to compare the effectiveness of different priors and frequentist detection procedures, such as the score tests we have presented, some frequentist calibration would be informative. For example, suppose we wished to compare the performance of two priors and two frequentist procedures. One way to do this would be to evaluate the probabilities that the Bayes factors were less than unity both in the presence and the absence of a signal and to relate these to the type I and type II error probabilities of the frequentist procedures. We are not aware of any such studies, either theoretical or empirical.

Both the score test and a Bayesian procedure involve choosing a target, via the choice of the Fourier coefficients, α_n, of the score test or in the choice of the variances of those coefficients for the Bayesian procedure. These choices should indeed be informed by prior knowledge about the possible range of light curve shapes. Even though the light curves of only a small number of γ-ray pulsars are known, using them to shape the target is presumably worthwhile.

Especially in light of the fact that theory alone can not make unequivocal conclusions about the comparative power of detection procedures, empirical experimentation is crucial. We understand that GLAST plans to make blind comparisons of a variety of procedures on synthetic signals. The results should be extremely interesting.

Acknowledgments. We thank Seth Digel, Patrick Nolan, and Tom Loredo for many helpful conversations.

References

R. J. Beran. Asymptotic theory for a class of tests for uniformity of a circular distribution. *The Annals of Mathematical Statistics*, 40:1196–1206, 1969.

P. Bickel, Y. Ritov, and T. Stoker. Tailor-made Tests for Goodness-of-Fit to Semiparametric Hypotheses. *The Annals of Statistics*, 34:721-741, 2006.

L. Breiman, C. Stone, and C. Kopperberg. Confidence bounds for extreme quantiles. *Journal of Statistical Computation and Simulation*, 127–149, 1988.

D. Brunk. Univariate density estimation by orthogonal series. *Biometrika*, 65: 521–528, 1978.

A. M. Chandler et al. A search for radio-quiet gamma-ray pulsars. *The Astrophysical Journal*, 556: 59–69, 2001.

O. C. De Jager et al. A powerful test for weak period signals with unknown light curve shape in sparse data. *Astronomy and Astrophysics*, 221:180–190, 1989.

R. Buccheri et al. Search for pulsed γ-ray emission from radio pulsars in the cos-b data. *Astronomy and Astrophysics*, 128:245, 1983.

J. K. Ghosh and R. V. Ramamoorthi. *Bayesian Nonparametrics*. Springer, 2003.

P. C. Gregory and Thomas J. Loredo. A new method for the detection of a periodic signal of unknown shape and period. *The Astrophysical Journal*, 398:146–168, 1992.

A. Janssen. Global power functions of goodness of fit tests. *The Annals of Statistics*, 28:239–253, 2000.

E. L. Lehmann and J. R. Romano. *Testing Statistical Hypotheses*. Springer, third edition, 2005.

K. V. Mardia. *Statistics of Directional Data*. Academic Press, 1972.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

I. Verdinelli and L. Wasserman. Bayesian goodness-of-fit testing using infinite dimensional exponential families. *The Annals of Statistics*, 26:1215–1241, 1998.

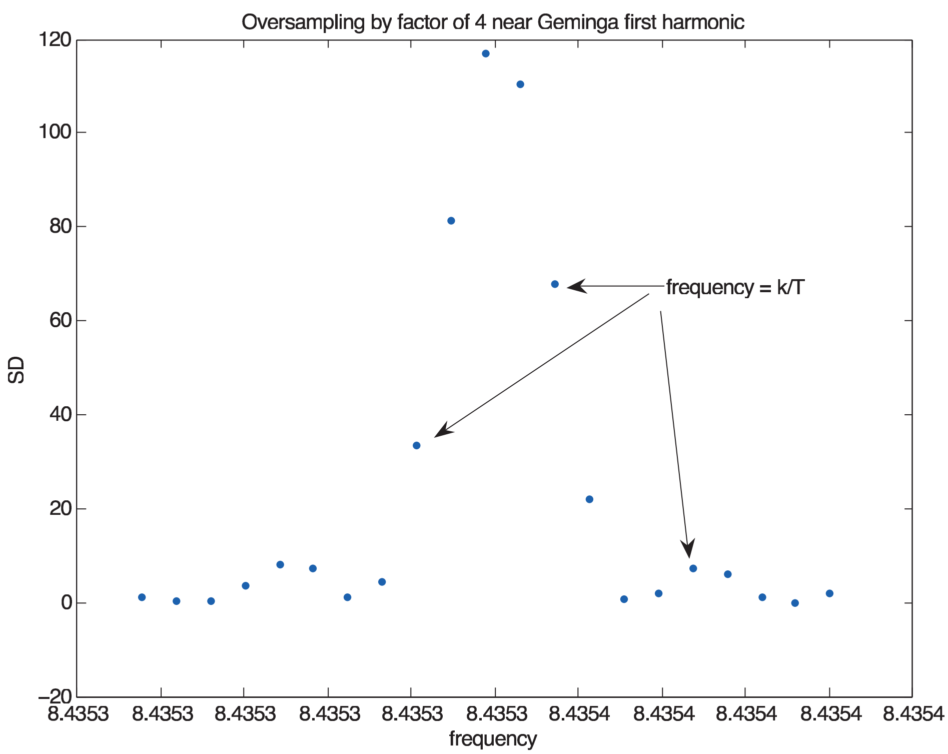

Figure 1. Test statistics for Geminga

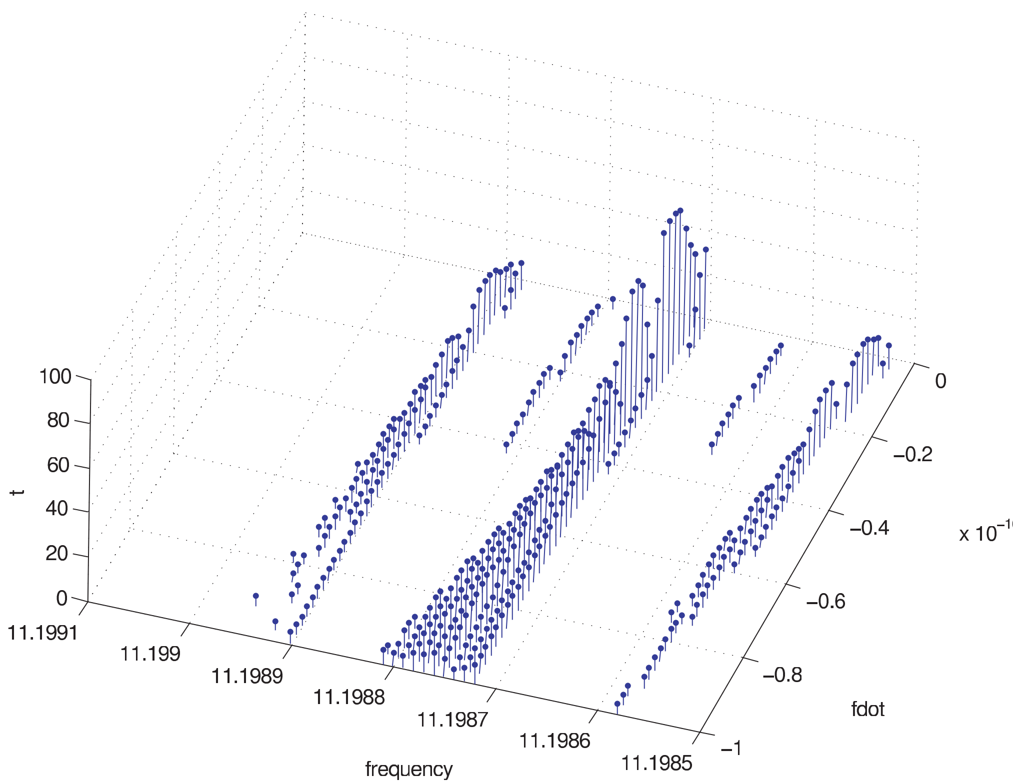

Figure 2. Test statistics (in units of standard deviations) for the Vela pulsar as a function of f and $\dot{f}$. Only those statistics greater than four standard deviations are plotted.

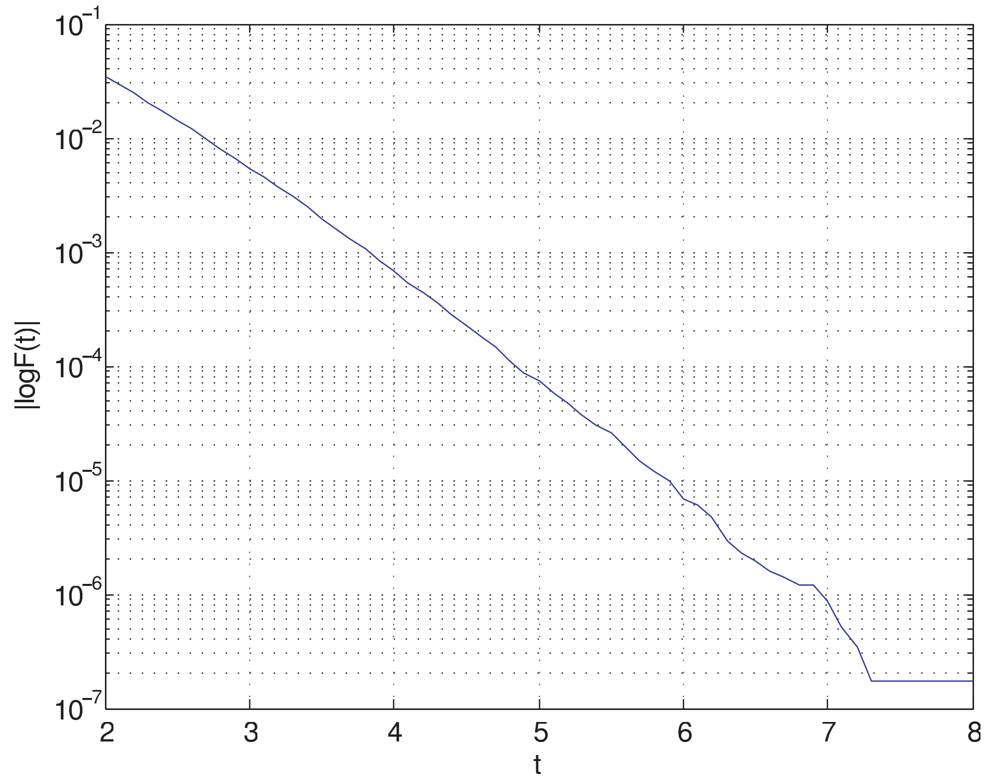

Figure 3. Comparison to extreme value theory: $\log[-\log F(t)]$ versus t, where t is in units of standard deviations.