Event Weighted Tests for Detecting Periodicity in Photon Arrival Times

John Rice
Statistics Department
University of California, Berkeley

Joint work with Peter Bickel and Bas Kleijn. Thanks to Seth Digel, Patrick Nolan, Tom Loredo, Charlotte Wickham, Jeremy Shen
Outline

- Motivation: EGRET sources and gamma-ray pulsars
- Detection as an hypothesis testing problem
- A score test and discussion of its properties
- Difficulties of a blind search
- Integration over frequency bands as an alternative to discretization
- Use of simulation and extreme value theory to assess significance
EGRET Sources
Many gamma-ray sources are unidentified and may be pulsars, but establishing that these sources are periodic is difficult. Might only collect ~1500 photons during a 10 day period.
Detection problem

Barycentric corrected arrival times \(0 < t_1 < \ldots < t_N < T\). For EGRET, \(N \sim 1000\), \(T \sim 10\) days

Energy and incidence angle of each photon. *How should this information be used?*

Some photons come from source and some from background.

*Question:* Is the source periodic, perhaps with decaying frequency? Pulse profile unknown.
Detection problem as hypothesis test

Unpleasant fact: There is no optimal test. Even if the frequency were known, a detection algorithm optimal for one pulse profile will not be optimal for another one. 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 pulse profile $\nu(t)$ is an infinite dimensional object. Any test can achieve high 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.

Consequence: You have to be a [closet] Bayesian and choose directions a priori.

Lehman & Romano. Testing Statistical Hypotheses. Chapt 14
Specifying a target

Consider testing against a template for the pulse profile, a probability density:

\[ \nu_0(t) = 1 + \eta \sum_{n \neq 0} \alpha_n e^{2\pi i nt} \]

\[ \nu_\tau(t) = 1 + \eta \sum_{n \neq 0} \alpha_n e^{2\pi i (t+\tau)} \]

Model the arrival times as a Poisson process with rate function:

\[ \lambda(t|\theta, \tau, \mu, f) = \mu c(t)[(1 - \theta) + \theta \nu_\tau(\phi(t))], \quad 0 \leq \theta \leq 1 \]

\[ \phi(t) = \phi_0 + ft \quad \text{or} \quad \phi(t) = \phi_0 + ft + \frac{1}{2}ft^2 \]

\[ H_0: \eta = 0 \]
Likelihood function and score test

Associated with each event is auxiliary information, the incidence angle and the measured energy; denote these variables by $z$. Let $f_B(z)$ denote the probability density function of $z$ for a background event and $f_S(z)$ the density function for a source event. The likelihood of the marked point process is

$$L = \mu^N \prod_{j=1}^{N} c(t_j)[(1 - \theta) f_B(z_j) + \theta f_S(z_j) \nu_\tau(\phi(t_j))] \times \exp(-\mu \int_0^T c(t)[(1 - \theta) + \theta \nu_\tau(\phi(t))] dt)$$

A score test is found by evaluating the derivative of the log likelihood with respect to $\eta$ at 0.
Score test

\[ S(\tau) = \sum_{i=1}^{N} \left( \frac{\theta f_S(z_j)}{(1-\theta)f_B(z_j) + \theta f_S(z_j)}(\nu_\tau(\phi(t) - 1)) \right) - \mu \theta \int_0^T c(t)[\nu_\tau(\phi(t)) - 1] dt \]

Posterior probability that the photon was from the source

Negligible for large T

To eliminate the dependence on \( \tau \), square and integrate, resulting in

\[ Q_T = \frac{1}{T} \sum_{n \neq 0} |\alpha_n|^2 |A_n|^2 \]

where

\[ A_n = \sum w_j \exp(2\pi in\phi(t_j)) \]
Relationship to classical tests in the un-weighted case

\[ Q_T = \frac{1}{T} \sum_{n \neq 0} |\alpha_n|^2 |A_n|^2 \quad A_n = \sum_j w_j \exp(2\pi in\phi(t_j)) \]

Rayleigh’s test (1919): \( \alpha_n = 0, \ n \geq 2 \)

Bucherri et al. (1983): \( Z_m^2: \alpha_n = 1, \ n \leq m \)

De Jager et al. (1989): choose \( m \) adaptively

Beran (1969) showed test to be locally most powerful among invariant tests for uniformity on the circle
Weight function

\[ w_j = \frac{\theta f_S(z_j)}{(1 - \theta) f_B(z_j) + \theta f_S(z_j)} \]

\[ z = (E, \varphi) \]

\[ f_B(z) = f_B(E) f_B(\varphi|E) \]

\[ f_S(z) = f_S(E) f_S(\varphi|E) \]

\[ w(E, \varphi) = \frac{\theta f_S(E) f_S(\varphi|E)}{\theta f_S(E) f_S(\varphi|E) + (1 - \theta) f_B(E) f_B(\varphi|E)} \]

Depends on spectra of source and background through their ratio
Weight function

The optimal weight function depends on the ratio of background and source spectra. If this is unknown, can use

$$w(E, \varphi) = \frac{\theta f_S(\varphi|E)}{\theta f_S(\varphi|E) + (1 - \theta)f_B(\varphi|E)}$$
Weight function

The weight function depends on $\theta$, the proportion of photons from the source. This may be known from previous studies. It might also be estimated from the log likelihood function under the null hypothesis (source is not periodic)

$$\ell(\theta) = N \log \mu + \sum_{j=1}^{N} \log c(t_j) +$$

$$\sum_{j=1}^{N} \log [(1 - \theta)f_B(z_j) + \theta f_S(z_j)] - \mu \int_0^T c(t) dt.$$

For a weak source (small $\theta$)

$$w \approx \frac{f_S(E)f_S(\varphi|E)}{f_B(E)f_B(\varphi|E)}$$
Detection sensitivity: power of the test

Let the pulse shape of the source be

\[ \gamma(t) = \sum_{n \neq 0} \gamma_n e^{2\pi int} \]

Then,

\[ \frac{E_K(Q_T) - E_H(Q_T)}{\sigma_H} \sim \theta^2 T \mu_0 \mathcal{E}(w) \frac{|\gamma_n|^2 |\alpha_n|^2}{\left[ \sum_{n \neq 0} |\alpha_n|^4 \right]^{1/2}} \]

where \( H \) denotes the null hypothesis and \( K \) the alternative hypothesis and for any weight function \( w(z) \)

\[ \mathcal{E}(w) = \frac{[E(W|Source)]^2}{E(W^2)} \]
Power

\[
\frac{E_K(Q_T) - E_H(Q_T)}{\sigma_H} \sim \theta^2 T \mu_0 \mathcal{E}(w) \frac{\sum_{n \neq 0} |\gamma_n|^2 |\alpha_n|^2}{[\sum_{n \neq 0} |\alpha_n|^4]^{1/2}}
\]

\( \theta^2 T \): threshold for detecting weak signal is \( \theta \sim T^{-1/2} \)

\( \sum_{n \neq 0} |\gamma_n|^2 |\alpha_n|^2 \): corr of actual lightcurve with template

\( \mathcal{E}(w) = \frac{[E(WW_{opt})]^2}{E(W^2)} \): correlation of weight function with optimal weight function
Power: effect of frequency misspecification

\[ f = f_0 + \Delta/T \quad \text{and} \quad \dot{f} = \dot{f}_0 + \Delta/T^2, \quad \Delta < 1 \]

\[ \sum_{n \neq 0} |\gamma_n|^2 |\alpha_n|^2 (1 + O((n\Delta)^2)) \]

High accuracy is required to gain power from higher harmonics
Example: template

Light Curves

Normalized Fourier Coefficients

Intensity

Proportion of phase

Size

Component

Pulsar
- Vela
- Geminga
- Crab

Pulsar
- V1
- G
- Cl
Example: template

Table 1: Relative efficiencies for truncated equal weights

<table>
<thead>
<tr>
<th>number of coefficients</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crab</td>
<td>28</td>
<td>65</td>
<td>89</td>
<td>83</td>
<td>88</td>
<td>84</td>
<td>80</td>
<td>78</td>
<td>74</td>
</tr>
<tr>
<td>Geminga</td>
<td>23</td>
<td>82</td>
<td>67</td>
<td>71</td>
<td>66</td>
<td>63</td>
<td>59</td>
<td>57</td>
<td>54</td>
</tr>
<tr>
<td>Vela</td>
<td>42</td>
<td>67</td>
<td>86</td>
<td>79</td>
<td>87</td>
<td>80</td>
<td>84</td>
<td>79</td>
<td>79</td>
</tr>
</tbody>
</table>

Table 2: Relative efficiencies obtained from using the first five and first ten average coefficients as the template.

<table>
<thead>
<tr>
<th>number of terms</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crab</td>
<td>96</td>
<td>97</td>
</tr>
<tr>
<td>Geminga</td>
<td>85</td>
<td>85</td>
</tr>
<tr>
<td>Vela</td>
<td>89</td>
<td>93</td>
</tr>
</tbody>
</table>
Example: weight function

Consider a source which emits photons at rate $\alpha$ and a background whose rate is $\rho$ per unit area and suppose that photons are collected in a disc of radius $R$ (rather than a spherical cap, for simplicity). Then

\[
\mu = \pi R^2 \rho + \alpha \\
\theta = \frac{\alpha}{\pi R^2 \rho + \alpha} \\
\]

\[
f_B(\varphi|E) = \frac{2\varphi}{R^2}, \quad 0 \leq \varphi \leq R
\]

The optimal weight function is then

\[
w_{opt}(E, \varphi) = \frac{f_S(E)f_S(\varphi|E)}{f_S(E)f_S(\varphi|E) + \beta \varphi f_B(E)}
\]

where $\beta = 2\pi \rho / \alpha$
Example: weight function

If the psf is bivariate circular Gaussian with standard deviation $\sigma(E)$, then $\varphi$, the distance to the origin, has the probability density function

$$f_S(\varphi|E) = \frac{\varphi}{\sigma(E)} \exp\left(-\frac{\varphi^2}{2\sigma^2(E)}\right)$$

(This assumes that $\sigma(E) \ll R$, otherwise the density truncated at $R$ has to be normalized to have unit area.) Then the optimal weight function is

$$w_{opt}(E, \varphi) = \frac{f_S(E)}{f_S(E) + \beta \sigma(E) \exp(\varphi^2/2\sigma^2(E))f_B(E)}$$
Example: weight function

If photons are not differentially weighted according to the ratio of the energy spectra, one has the weight function

\[ w(E, \varphi) = \frac{1}{1 + \beta \sigma(E) \exp(\varphi^2/2\sigma^2(E))} \]

The decay of the weight function depends on the parameter \( \xi = 2\pi \rho \sigma(E)/\alpha \).

\( \xi = 100, 10, 1, 0.1, 0.01 \)
Difficulties

- Frequency unknown
- Spin down
- Large search space
- Glitches
- Celestial foreground
- Barycentric time correction
- Pulse profile unknown

Computational demands for a blind search are very substantial. A heroic search using a 512 processor supercomputer did not find any previously unknown gamma-ray pulsars in EGRET data. (Chandler et al, 2001).
Fig. 1.—The $f_f^2$ phase space relevant to our search. The 472 pulsars in the Princeton catalog (Taylor et al. 1993) that have positive period derivatives are denoted by filled circles. The known EGRET pulsars are indicated by triangles. They are, in order of decreasing rotation frequency, Crab, PSR 1951 + 32, Vela, PSR 1706 – 44, Geminga, and PSR 1055 – 52.

From Chandler et al. (2001)
Search Space

Consider no drift. Good frequency resolution depends on matching phase of photons at beginning and end of the record. If true frequency is $f_0$, the number of cycles in time $T$ is $T/f_0$, so if the hypothesized frequency is $f = f_0 + \delta f$, $\delta f$ should be $o(T^{-1})$ in order for a photon at the end of the record to be in phase with one at the beginning. The phase error at the end of the record is $T\delta f$.

10 days = 864,000 sec, $\delta f = T^{-1}$. If a 40 Hz range has to be searched, a minimum of $40 \times 864000 = 34,560,000$ possible frequencies must be examined.

Similarly, drift must be resolved within $o(T^{-2})$. To search the interval of possible frequency derivatives at this resolution, about 400-500 values must be examined.

Consequence is that a test statistic must be evaluated $\sim 10^9$ values of frequency and its derivative.
Histograms and density estimates for phased and folded arrival times from Geminga $t^* = t \mod P$ ($P = 1/f$) for increments in $f$ of size $10^{-7}$ (.12/T)
Example: Vela
Crab
Power vs Computational Cost

Power $\propto \theta^2 T$

$f$ resolution $\propto T^{-1}$

$\dot{f}$ resolution $\propto T^{-2}$

Calculation of statistic for a single $(f, \dot{f}) \propto T$

FLOPS $\propto T^4$

Partition $T$ into $B$ blocks of length $\sigma$. Compute statistic in each block and average.

Power $\propto \frac{\theta^2 T}{\sqrt{B}} = \theta^2 \sigma^{1/2} T^{1/2}$

FLOPS $\propto \sigma T$
Blocking Vela and Crab

Vela: 318 blocks

Crab: 25 blocks
Integration versus discretization

Rather than fine discretization of frequency, consider integrating the test statistic over a frequency band using a symmetric probability density $g(f)$.

\[
\bar{Q}_n = \int |A_n(f)|^2 g(f) \, df
\]

\[
= \sum_j \sum_k w_j w_k \hat{g}(2\pi n (t_j - t_k))
\]

where

\[
\hat{g}(t) = \int_{-\infty}^{\infty} e^{itf} g(f) \, df
\]
\[ Q_n = \sum_j \sum_k w_j w_k \hat{g}(2\pi n(t_j - t_k)) \]

Requires a number of operations quadratic in the number of photons. However the quadratic form can be diagonalized in an eigenfunction expansion, resulting in a number of operations linear in the number of photons.

\[ Q_n = \sum_k \mu_{kn} \left| \int_0^T e^{2\pi i n f_0} \psi_{kn}(t) dW(t) \right|^2 \]

(In the case that \( g() \) is uniform, the eigenfunctions are the prolate spheroidal wave functions.) Then

\[
E \bar{Q}_T = \mu^2 \theta^2 T \sum_n |\alpha_n|^2 |\gamma_n|^2 \int \left| \frac{\sin(n\pi u)}{n\pi u} \right|^2 g_0(u) du \\
g(f) = T g_0(T(f - f_0))
\]

Power is still lost in high frequencies unless the support of \( g \) is small.

This procedure can be extended to integrate over tiles in the \((f, \dot{f})\) plane when

\[ \phi(t) = ft + \frac{1}{2} \dot{f} t^2 \]
Assessing significance

At a single frequency, significance can be assessed easily through simulation. In a broadband blind search this is not feasible and furthermore one may feel nervous in using the traditional chi-square approximations in the extreme tail (it can be shown that the limiting null distribution of the integrated test statistic is that of a weighted sum of chi-square random variables). We are thus investigating the use of classical extreme value theory in conjunction with affordable simulation.
Gumbel Approximation

Consider $M_n = \max\{T_1, T_2, \ldots, T_n\}$ 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 mixtures of chi-square random variables, the limiting distribution will be of Gumbel type.

$$F^n(a_n t + b_n) \rightarrow \exp[-\exp(-x)]$$

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

We thus approximate tail probabilities by

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

where $P_n$ is the empirical measure and $\alpha_n$ is the mean of the corresponding empirical conditional distribution.
Tail Approximations

Table 1: Conditional means and tail probability approximations for various cutoff values $t$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>$\alpha$</th>
<th>$P(T &gt; 10)$</th>
<th>$P(T &gt; 15)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>.4852</td>
<td>$3.0 \times 10^{-9}$</td>
<td>$9.9 \times 10^{-14}$</td>
</tr>
<tr>
<td>4</td>
<td>.4548</td>
<td>$1.3 \times 10^{-9}$</td>
<td>$2.1 \times 10^{-14}$</td>
</tr>
<tr>
<td>5</td>
<td>.4438</td>
<td>$9.4 \times 10^{-10}$</td>
<td>$1.2 \times 10^{-14}$</td>
</tr>
</tbody>
</table>

According to this approximation, in order for a Bonferroni corrected p-value to be less than 0.01, a test statistic of about 11 standard deviations or more would be required.
$\log[- \log F(t)]$ versus $t$
Thank you
Example: Geminga

Main frequency only, N=1

Main frequency plus equally weighted first harmonic
Oversampling by factor of 4 near Geminga first harmonic

frequency = k/T