Supplement: Jaynes-Bretthorst algorithm, and Bayesian periodograms for event data

Tom Loredo
Cornell Center for Astrophysics and Planetary Science
http://www.astro.cornell.edu/staff/loredo/bayes/

AAS237 — 8 Jan 2021

For event data details see Loredo (1993): "Bayesian Inference With Log-Fourier Arrival Time Models and Event Location Data" (unpublished report)

Agenda

1. **Jaynes-Bretthorst algorithm**

2. **Periodograms for event time series**

Agenda

1. **Jaynes-Bretthorst algorithm**

2. **Periodograms for event time series**

Bayesian Curve Fitting & Least Squares

Setup

Data $D = \{d_i\}$ are measurements of an underlying function $f(x;\theta)$ at N sample points $\{x_i\}$. Let $f_i(\theta) \equiv f(x_i;\theta)$:

$$d_i = f_i(\theta) + \epsilon_i, \qquad \epsilon_i \sim N(0, \sigma_i^2)$$

We seek learn θ, or to compare different functional forms (model choice, M).

Likelihood

$$\begin{aligned} p(D|\theta, M) &= \prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{d_i - f_i(\theta)}{\sigma_i}\right)^2\right] \\ &\propto \exp\left[-\frac{1}{2}\sum_i \left(\frac{d_i - f_i(\theta)}{\sigma_i}\right)^2\right] \\ &= \exp\left[-\frac{\chi^2(\theta)}{2}\right] \end{aligned}$$

Posterior

For prior density $\pi(\theta)$,

$$p(\theta|D, M) \propto \pi(\theta) \exp\left[-\frac{\chi^2(\theta)}{2}\right]$$

If you have a least-squares or χ^2 code:

- Think of $\chi^2(\theta)$ as $-2 \log \mathcal{L}(\theta)$.
- Bayesian inference amounts to exploration and numerical integration of $\pi(\theta)e^{-\chi^2(\theta)/2}$.
- If noise level is uncertain, keep the $1/\sigma_i$ factors (dropped above!) and include noise parameters in inference (e.g., scale all σ_i by a parameter, α)

Important Case: Separable Nonlinear Models

A (linearly) separable model has parameters $\theta = (A, \psi)$:

- Linear amplitudes $A = \{A_\alpha\}$
- Nonlinear parameters ψ

$f(x; \theta)$ is a linear superposition of M nonlinear components $g_\alpha(x; \psi)$:

$$
\begin{aligned}
d_i &= \sum_{\alpha=1}^{M} A_\alpha g_\alpha(x_i; \psi) + \epsilon_i \\
\text{or} \qquad \vec{d} &= \sum_\alpha A_\alpha \vec{g}_\alpha(\psi) + \vec{\epsilon}.
\end{aligned}
$$

Note: "linear/nonlinear" refers to how the predictions depend on the *parameters*, not how they depend on the sample location!

Examples

Polynomials (simple or orthogonal); $\psi = \emptyset$:

$$
\begin{aligned}
f(x) &= A_0 + A_1 x + A_2 x^2 + A_3 x^3 \\
&= A_0 + A_1 x + A_2'(2x^2 - 1) + A_3'(4x^3 - 3x), \quad x \in [-1, 1] \\
&= A_0 g_0(x) + A_1 g_1(x) + A_2' g_2(x) + A_3' g_3(x)
\end{aligned}
$$

Sinusoids; $\psi = \omega$:

$$
\begin{aligned}
f(x) &= A \cos(\omega x + \phi) \\
&= A_1 \cos \omega x + A_2 \sin \omega x \\
&= A_1 g_1(x, \omega) + A_2 g_2(x, \omega)
\end{aligned}
$$

Chirps; $\psi = (\omega, \alpha)$:

$$
\begin{aligned}
f(x) &= A \cos(\alpha x^2 + \omega x + \phi), \quad \text{inst. freq.} = \omega + 2\alpha x \\
&= A_1 \cos(\alpha x^2 + \omega x) + A_2 \sin(\alpha x^2 + \omega x)
\end{aligned}
$$

Spectral continuum + lines; multipoles/spherical harmonics...

The Jaynes-Bretthorst Algorithm

Why separable structure is important: You can marginalize over A *analytically* $\rightarrow$ *Jaynes-Bretthorst algorithm* ("Bayesian Spectrum Analysis & Param. Estimation" 1988)

Algorithm is closely related to linear least squares, diagonalization (eigenvectors/values), and SVD

Goals:

- Estimate the nonlinear parameters ψ
- Estimate amplitudes
- Compare rival models

The log-likelihood is a quadratic form in A_α,

$$
\begin{aligned}
\mathcal{L}(A,\psi) &\propto \frac{1}{\sigma^N}\exp\left[-\frac{Q(A,\psi)}{2\sigma^2}\right] \\
\text{with}\quad Q &= \left[\vec{d}-\sum_\alpha A_\alpha \vec{g}_\alpha\right]^2 \\
&= \left[\vec{d}-\sum_\alpha A_\alpha \vec{g}_\alpha\right]\cdot\left[\vec{d}-\sum_\beta A_\beta \vec{g}_\beta\right] \\
&= d^2 - 2\sum_\alpha A_\alpha \vec{d}\cdot\vec{g}_\alpha + \sum_{\alpha,\beta} A_\alpha A_\beta \eta_{\alpha\beta} \\
&\quad \text{where}\quad \eta_{\alpha\beta}(\psi) = \vec{g}_\alpha(\psi)\cdot\vec{g}_\beta(\psi)
\end{aligned}
$$

We seek to integrate out the amplitudes, but completing the square is complicated because of the nontrivial metric $\eta_{\alpha\beta}$

Change the basis for $\vec{f}$ from $\vec{g}_\alpha$ to an *orthonormal basis* $\vec{h}_\mu$:

$$\vec{g}_\alpha = \sum_\mu a_{\alpha\mu}\vec{h}_\mu \qquad \text{with } \vec{h}_\mu \cdot \vec{h}_\nu = \delta_{\mu\nu}$$

which implies $\vec{h}_\mu = \sum_\alpha (a^{-1})_{\mu\alpha}\vec{g}_\alpha$. Note $a = a(\psi)$.

Rewriting $\vec{f}$,

$$\vec{f}(\theta) = \sum_{\alpha=1}^{M} A_\alpha \vec{g}_\alpha(\psi) = \sum_{\mu=1}^{M} B_\mu(A,\psi)\vec{h}_\mu(\psi)$$

with orthonormal amplitudes $B_\mu(A,\psi) = \sum_\alpha A_\alpha a_{\alpha\mu}(\psi)$

Some linear algebra shows that $\eta = aa^T$, so we can get a from η via Cholesky/eigen/QR decomposition.

Now write the quadratic form in terms of the Bs instead of the As:

$$\begin{aligned} Q &= d^2 - 2\sum_\alpha A_\alpha \vec{d}\cdot\vec{g}_\alpha + \sum_{\alpha,\beta} A_\alpha A_\beta \eta_{\alpha\beta} \\ &= d^2 - 2\sum_\mu B_\mu \vec{d}\cdot\vec{h}_\mu + \sum_\mu B_\mu^2 \\ &= \sum_\mu \left[B_\mu - \hat{B}_\mu(\psi)\right]^2 + r^2(\psi) \end{aligned}$$

with $\hat{B}_\mu(\psi) \equiv \vec{d}\cdot\vec{h}_\mu(\psi)$ and the residual $\vec{r}(\psi) \equiv \vec{d} - \sum_\mu \hat{B}_\mu \vec{h}_\mu$

The posterior in terms of Bs is

$$p(B,\psi|D,I) \quad \propto \quad \frac{\pi(\psi)J(\psi)}{\sigma^N}\exp\left[-\frac{r^2(\psi)}{2\sigma^2}\right]\exp\left[\frac{-1}{2\sigma^2}\sum_\mu [B_\mu - \hat{B}_\mu(\psi)]^2\right]$$

$J(\psi) = (\det\eta)^{1/2}$ comes from changing variables from As to Bs

Marginalize B's analytically (*check the range!*):

$$p(\psi|D,I) \propto \frac{\pi(\psi)J(\psi)}{\sigma^{N-M}} \exp\left[-\frac{r^2(\psi)}{2\sigma^2}\right]$$

If σ unknown, marginalize using $p(\sigma|I) \propto \frac{1}{\sigma}$.

$$p(\psi|D,I) \propto \pi(\psi)J(\psi)\left[r^2(\psi)\right]^{\frac{M-N}{2}}$$

For given ψ, r^2 is just the residual sum of squares from a least squares fit to the basis functions. We can write

$$\begin{aligned} r^2(\psi) &= d^2 - \sum_\mu \hat{B}_\mu^2(\psi) \\ &= d^2 - S(\psi) \end{aligned}$$

with $S(\psi) = \sum_\mu [\vec{d} \cdot \vec{h}_\mu(\psi)]^2$, the sum of squared projections

Bayesian Spectrum Analysis

Adopt a sinusoid periodic signal model:

$$\begin{aligned} f(t) &= A\cos(\omega t - \phi) && \text{parameters } \omega, A, \phi \\ &= A_1 \cos\omega t + A_2 \sin\omega t && \text{parameters } \omega, A_1, A_2 \end{aligned}$$

$$d_i = f(t_i) + e_i \qquad \text{Gaussian error } pdf\text{s; rms} = \sigma$$

Estimate ω:

$$\begin{aligned} p(\omega|D) &\propto \int dA_1 \int dA_2\; p(\omega, A_1, A_2)\mathcal{L}(\omega, A_1, A_2) \\ &\propto p(\omega)J(\omega)\exp\left[\frac{S(\omega)}{\sigma^2}\right] \end{aligned}$$

- Equally-spaced samples: $S(\omega) \to P(\omega)$ for large N (when η is nearly diagonal)
- Unequally-spaced samples: $S(\omega) \approx$ Lomb-Scargle periodogram

Two Sinusoids

Adopt a model with two sinusoids at distinct frequencies:

$$\begin{aligned} f(t) &= A\cos(\omega_1 t - \phi_1) + B\cos(\omega_2 t - \phi_2) \\ &= A_1 \cos\omega_1 t + A_2 \sin\omega_1 t \\ &\quad + A_3 \cos\omega_2 t + A_4 \sin\omega_2 t \end{aligned}$$

Use JB algorithm to find $p(\omega_1, \omega_2|D)$ or $p(\omega_1, \delta\omega|D)$

Call $\log p$ the *doublet periodogram*

If the frequencies are well-separated, can show that

$$p(\omega_1, \omega_2|D) \propto p(\omega_1, \omega_2) J(\omega_1, \omega_2) \exp\left[\frac{S(\omega_1)}{\sigma^2}\right] \exp\left[\frac{S(\omega_2)}{\sigma^2}\right]$$

a product of two *independent* dist'ns, each determined by the periodogram

If the frequencies are close, there is no such simplification; the real and imaginary parts of the DTFT at each frequency (projections!) combine in a more complicated way

Exoplanets

Data are radial velocities; model as due to Keplerian motion.

For single planet:

$$V(t) = A_1 + A_2[e + \cos \upsilon(t)] + A_3 \sin \upsilon(t)$$

$$\upsilon(t) = f(t; \tau, e, M_0) \qquad \text{via Kepler's eqn}$$

Period τ
3 linear amplitudes (COM velocity, orbital velocity, arg. of periastron)
2 other nonlinear parameters (e, M_0)

Follow the recipe! For $e = 0 \to$ LSP, but for nonzero eccentricity it generalizes to a *Kepler periodogram*

For astrometry, 2D data require $x(t)$, $y(t)$.
Extra parameters: inclination, parallax, proper motion.

Terminology for Generalized Periodograms

Fourier models

$$\ln p(\omega|D,M) \quad \propto \quad \text{periodogram (Schuster, Lomb-Scargle)}$$

$$\ln p(\omega_1,\omega_2|D,M) \quad \propto \quad \text{doublet periodogram}$$

Chirp model

$$A_1\cos(\omega t+\alpha t^2)+A_2\sin(\omega t+\alpha t^2)$$
$$\ln p(\omega,\alpha|D,M) \propto \text{Chirpogram (Jaynes)}$$

Keplerian reflex motion model

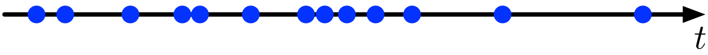
$$\ln p(\tau,e,M_p|D,M_p) = \text{(Radial) Keplerogram}$$
$$\log p(\tau|D) \propto \text{(Radial) Kepler periodogram}$$

Cf. Box and other periodograms. . .

Agenda

1. Jaynes-Bretthorst algorithm
2. Periodograms for event time series

Time-Tagged Event/Arrival Time Series



Data D: $\{t_i\}$ for $i = 1$ to N

Nonhomogeneous Poisson point process sampling distribution, rate $r(t;\theta)$:

$$p(D|\theta, M) = \exp\left[-\int_T dt\, r(t)\right] \prod_{i=1}^{N} r(t_i)$$

Parametric models: $r(t;\theta)$; θ fixed dimension, e.g. (A, ω, ϕ)

Semiparametric models: $r(t;\theta,\psi)$

- θ of fixed dimension
- ψ of variable dimension (e.g., adaptive shape, # of components...)

Conventional period detection approaches

Try to reject a "null" (constant rate) hypothesis with an omnibus test; report p-value.

- Rayleigh statistic:

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

- Z_n^2 statistic:

$$Z_n^2(\omega) = \sum_{j=1}^{n} R^2(j\omega)$$

- χ^2-Epoch folding:
  - Fold data with trial period $\rightarrow$ phases $\theta_i = \omega t_i + \phi \mod 2\pi$; bin $\rightarrow n_j$, $j = 1$ to M
  - Calculate Pearson's $\chi^2(\omega, \phi)$ vs. $n_j = N/M$; *average over phase*

Bayesian Setup

Rate $=$ avg. rate $A \times$ periodic shape $\rho(\theta)$, shape params $\mathcal{S}$:

$$r(t) = A\rho(\omega t - \phi; \mathcal{S})$$

Non-homogeneous point process likelihood (for $T \gg$ period):

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

Marginal likelihood for ω, ϕ, $\mathcal{S}$:

$$\mathcal{L}_m(\omega, \phi, \mathcal{S}) \propto \prod_i \rho(\omega t_i - \phi; \mathcal{S})$$

Goal: Find shapes where some Bayesian integrals are *analytical*

Bayesian Harmonic Modeling

Analytical phase marginalization

Log-sinusoid model

$$\rho(\theta) \propto e^{\kappa \cos\theta}, \qquad \text{Von Mises dist'n}$$

Can analytically marginalize ϕ:

$$\mathcal{L}_m(\omega) \propto I_0\left[\kappa N R(\omega)\right] / I_0^N(\kappa)$$

Rayleigh statistic "appears;" κ provides sensitivity to narrow pulses (vs. Rayleigh test)

Log-Fourier models

For multi-peak light curves, consider log-Fourier models:

$$\rho(\theta) \propto \exp\left[\sum_j \kappa_j \cos(j\theta)\right]$$

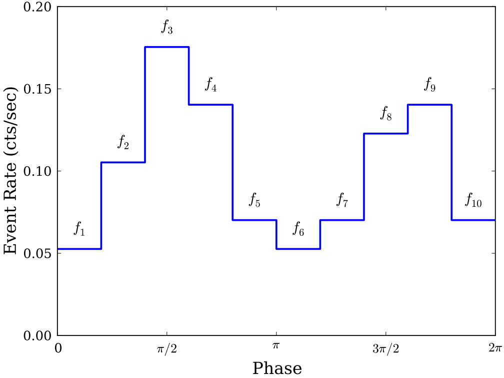
Issues: cannot analytically integrate ϕ; dimensionality of κ

Laplace approximation for ϕ:

$$\mathcal{L}_m(\omega, \boldsymbol{\kappa}) \approx \text{weighted } Z_n^2 + \text{interference terms}$$

Bayesian Stepwise Modeling

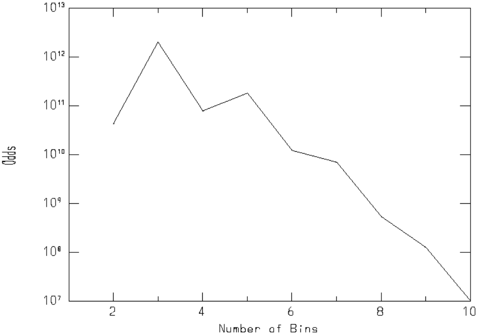
Analytical shape marginalization

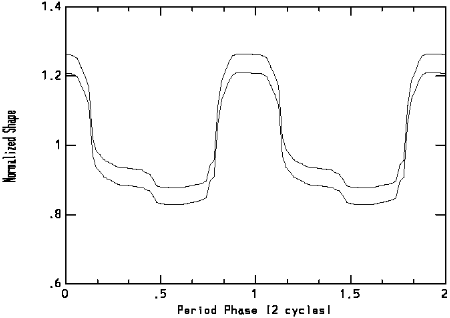


Flat prior on $\boldsymbol{f} \rightarrow$ Can marginalize $\boldsymbol{f}$; for M bins:

$$\mathcal{L}(\omega, \phi, M) \propto \frac{(M-1)!}{(N+M-1)!} \left[\frac{n_1!\; n_2!\; \ldots n_M!}{N!}\right] \sim \exp\left[\chi^2(\omega, \phi)\right]$$

X-Ray Pulsar PSR 0540-693 (Gregory & TL 1996)
3300 events over 10^5 s, many gaps, Rayleigh test fails



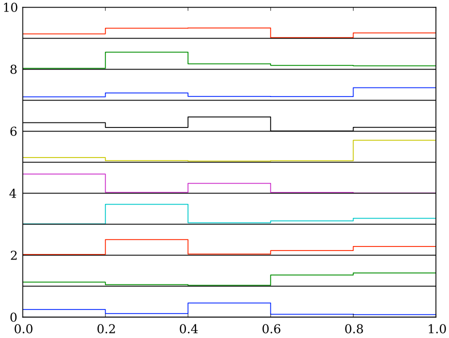


Other applications by Arnold Rots; see IAU 285 poster

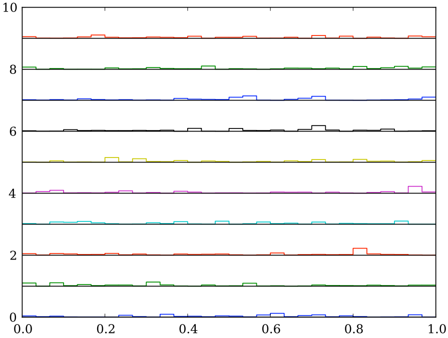
How Might We Do Better?

The flat stepwise shape prior is. . . *flat!*

Flat prior, *m*=5



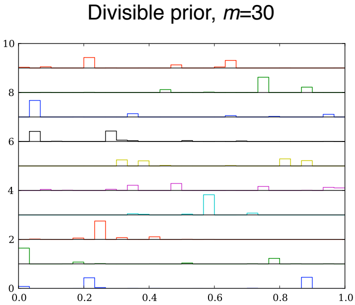
Flat prior, *m*=30

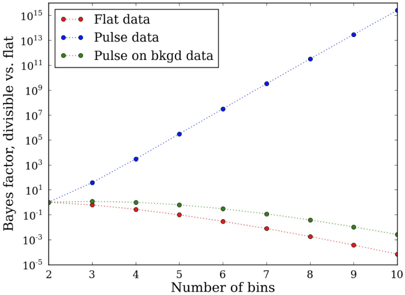


A simple example of a general phenomenon:
Concentration of measure

Promising direction

- Cross-model consistency requirement: 4-bin prior should become 2-bin prior when binned up, etc.
- $\to$ Form of prior should depend on dimension,
  $$p(f) \propto \prod_{i=1}^{M} f^{\alpha-1} \text{ with } \alpha = \text{Const}/M$$
- This is called *divisibility*





Best solutions combine divisibility & structure, e.g., mixtures