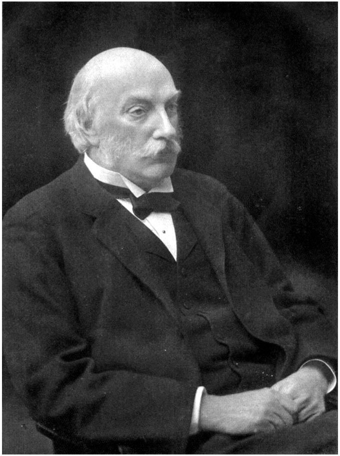
Understanding and using periodograms

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

AAS237 — 8 Jan 2021

Periodograms and spectral analysis of time series

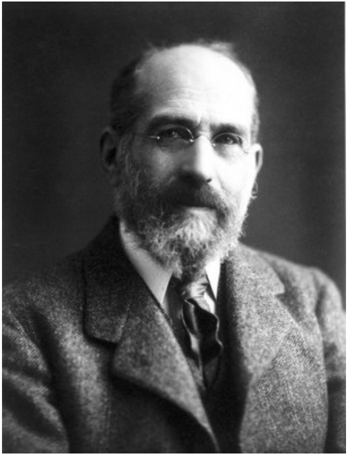
Lord Rayleigh
(John William Strutt)
1842–1919



Wikipedia

Cavendish Lab, University of Cambridge
(succeeded Maxwell);
Royal Inst. of Great Britain

Arthur Schuster
1851–1934



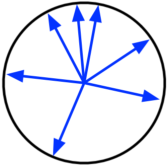
Wikipedia

Cavendish Lab (w. Maxwell, Rayleigh);
Owens College, Victoria University;
succeeded by Rutherford

Schuster (& Rayleigh) inventing periodograms

Motivation

- Find "*hidden periodicities*" (vs. "obvious periodicities" like tides, sunspot maxima)
- *Quantify uncertainty*: "[A]pply the theory of probability in such a way that we may be able to assign a definite number for the probability that the effects found by means of the usual methods are real, and not due to accident." *[sic! A p-value is NOT a FAP!]*



Two types of periodograms

- "Occurence periodogram" for event times (*Rayleigh statistic*): Add wrapped/folded unit vectors
- "Magnitude periodogram" for equally-spaced scalar measurements (*Schuster periodogram*): Add wrapped/folded vectors

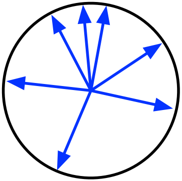
Null distributions (no periodic signal)

- Resultant vector magnitude, R: Rayleigh distribution, $p(R) \propto Re^{-R^2/\sigma^2}$
- Squared magnitude (power): Exponential distribution, $p(S) \propto e^{-S/\sigma^2}$ — a χ^2_2 dist'n

(Use $\sigma^2 \to N$ for occurence periodograms; beware normalizations!)

“Occurrence” periodogram (Rayleigh power):

$$\frac{n r_1}{2 a_0} = \left\{ (\cos k t_1 + \cos k t_2 + \quad . \quad . \quad . \quad \cos k t_n)^2 + (\sin k t_1 + \sin k t_2 + \quad . \quad . \quad . \quad \sin k t_n)^2 \right\}^{\frac{1}{2}}$$



“Magnitude” periodogram (Schuster p-gram):

$$\frac{1}{2} p r_1 = \left\{ (T_1 \cos \theta_1 + T_2 \cos 2\theta + \dots)^2 + (T_1 \sin \theta_1 + T_2 \sin 2\theta + \dots)^2 \right\}^{\frac{1}{2}}.$$

Schuster showed this amounts to fitting $A \cos kt + B \sin kt$ to the data, and computing $\hat{A}^2 + \hat{B}^2$ — the Fourier power for the fundamental frequency

Name & interpretation

- "*Fourier's analysis* here serves the same purpose as the prismatic analysis of a luminous disturbance. . . ."
- "It is convenient to have a word for some representation of a variable quantity which shall correspond to the 'spectrum' of a luminous radiation. I propose the word *periodogram*. . ."
- Envisioned application to *periodic* ("lines" from musical instruments), *quasiperiodic* ("bands" from sunspots) and *stochastic* ("broad band" from noise) variability


Wikipedia

"More lives have been lost looking at the raw periodogram than by any other action involving time series!"
— *John Tukey*

"More lives have been lost looking at the **raw** periodogram than by any other action involving time series!"
— *John Tukey*

It's how you *manipulate the periodogram* that makes it useful—for *multiple purposes*

Literature (vast!) relevant to understanding periodograms:

- Spectral analysis of time series
- Applied harmonic analysis
- Harmonic regression
- Trigonometric regression

Three uses for periodograms

- Nonparametric spectrum estimation
- Parametric period detection & estimation (least squares, Bayes)
- Semiparametric stochastic process (Gaussian process) modeling

Agenda

1. **Math background: Fourier transforms**

2. **Periodograms for sampled/probed time series**
   - Periodogram as a smooth spectrum estimator
   - Periodogram as a sinusoid period finder
   - Periodogram as a log likelihood for a GP

Agenda

1. **Math background: Fourier transforms**

2. **Periodograms for sampled/probed time series**
   - Periodogram as a smooth spectrum estimator
   - Periodogram as a sinusoid period finder
   - Periodogram as a log likelihood for a GP

The continuous Fourier transform

Describe periodicity via:

- Period τ (e.g., seconds, days)
- (Natural) frequency $\nu = 1/\tau$ (e.g., Hz, cycles/day)
- Angular frequency $\omega = 2\pi\nu$ (e.g., rad/s, rad/day)

Fourier transform:

$$\begin{aligned} \tilde{f}(\nu) &\equiv \int_{-\infty}^{\infty} dt\, f(t)\, e^{-i2\pi\nu t} \\ &= \int_{-\infty}^{\infty} dt\, f(t)\, [\cos(2\pi\nu t) - i\sin(2\pi\nu t)] \end{aligned}$$

Inverse Fourier transform:

$$f(t) \equiv \int_{-\infty}^{\infty} d\nu\, \tilde{f}(\nu) e^{+2\pi i\nu t}$$

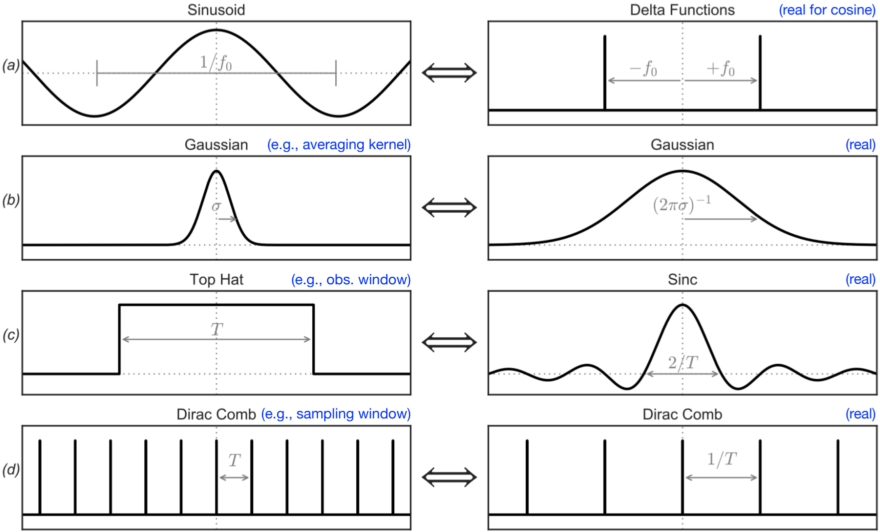
Notation:

$$\begin{aligned} \mathcal{F}\{f\} &\equiv \tilde{f}(\nu) \\ \mathcal{F}^{-1}\{\tilde{f}\} &\equiv f(t) \end{aligned}$$

Power spectrum or power spectral density (PSD):

$$\mathcal{P}_f(\nu) \equiv |\mathcal{F}\{f\}|^2$$

Fourier transform pairs



From Jake vanderPlas's notes on GitHub

The convolution theorem

Convolution of f and g:

$$\{f * g\}(t) \equiv \int du\, f(u)\, g(u - t)$$

Transform of a convolution is a product of transforms:

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\}$$

Transform of a product is a convolution of transforms:

$$\mathcal{F}\{f \cdot g\} = \mathcal{F}\{f\} * \mathcal{F}\{g\}$$

Wiener-Khinchin theorem relating PSD to the covariance function of a stationary Gaussian process:

$$C(u) = \mathrm{E}[f(t)f(t-u)] \qquad \text{(zero-mean case)}$$
$$\rightarrow \mathcal{P}_f = \mathcal{F}\{C(u)\}$$

Signal vs. data transforms

For N noiseless measurements,

$$\begin{aligned}
\{y_i\} &= \{f(t_i)\} && (1) \\
&= f(t) \times \text{Top Hat} \times \text{Comb} && (2)
\end{aligned}$$

The δ function spectrum of a sinusoid becomes a superposition of many smooth bumps

Noise introduces an additional additive corruption

Discretization & computation

Discrete time Fourier transform (DTFT)

$$\begin{aligned} \tilde{f}(\nu) &\equiv \sum_j f(t_j)\, e^{-i2\pi\nu t} \\ &= \sum_j f(t_j)\, [\cos(2\pi\nu t) - i\sin(2\pi\nu t)] \end{aligned}$$

This is a continuous function of ν

Discrete Fourier transform (DFT)

For equally-spaced $t_j = jT/N$, the DFT evaluates the DTFT at a specific set of N equally-spaced frequencies, from $\nu = 0$ to the Nyquist frequency: the *Fourier frequencies*

$$\tilde{f}_k \equiv \sum_{j=0}^{N-1} f(t_i) \exp\left(-i\frac{2\pi k}{N} T\right)$$

(For real-valued $f(t)$, half of the $\tilde{f}_k$ values are repeated.)

Agenda

1. **Math background: Fourier transforms**

2. **Periodograms for sampled/probed time series**
   - Periodogram as a smooth spectrum estimator
   - Periodogram as a sinusoid period finder
   - Periodogram as a log likelihood for a GP

Periodogram-based spectral analysis

Underlying idea:

- A periodic signal may be recognized by looking for a δ-function in the *signal's* power spectrum
- $\Rightarrow$ Devise an *estimator* for the signal's power spectrum and look for δ-functions in an estimate

Periodogram is the power spectrum of the *data*, $y_i = f(t_i) + \epsilon_i$:

$$P(\omega) = \frac{1}{N}\left[\left(\sum_i y_i \cos \omega t_i\right)^2 + \left(\sum_i y_i \sin \omega t_i\right)^2\right]$$

(Note: Normalization conventions differ, depending on noise properties)

- Evenly sampled data: Schuster periodogram
- Unevenly sampled data: Lomb-Scargle periodogram (LSP) (with an optimal choice of phase)

Bias and variability

Interpret data power spectrum as signal power spectrum corrupted by noise and sampling "window"

Under no-signal "null" hypotheses, $N/2$ values at the *Fourier frequencies* (as returned by the DFT) are *statistically independent*

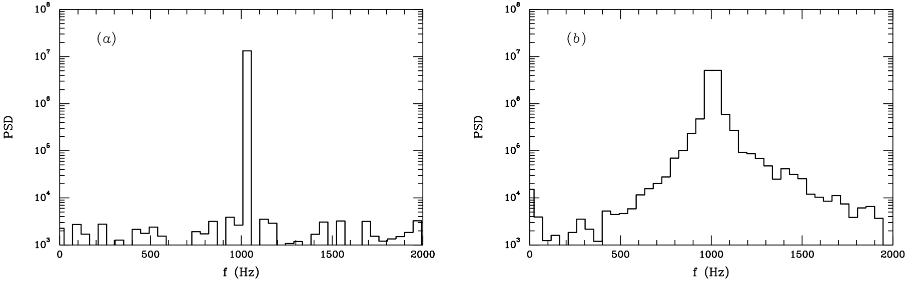
(This follows from orthogonality of sines/cosines on a uniform grid; LSP case is more complicated)

$\Rightarrow$ *focus on Fourier frequencies*

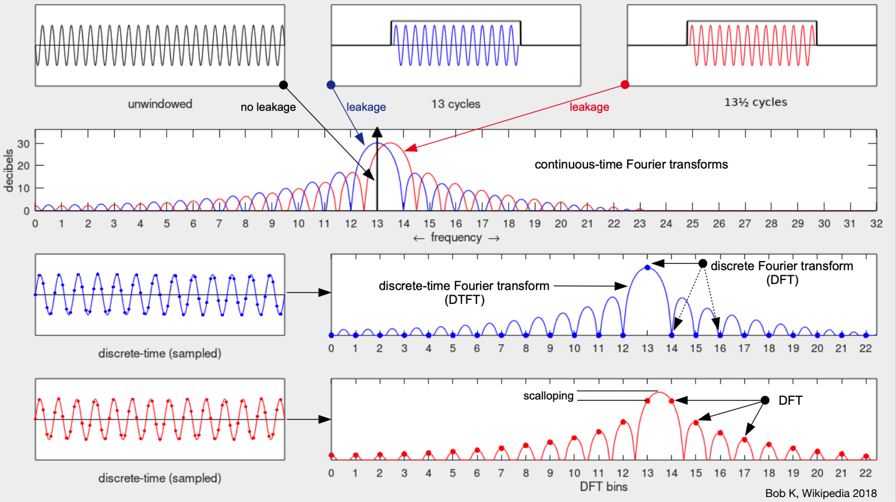
- Variability: The number of spectrum estimates (at Fourier frequencies) grows with N $\rightarrow$ periodogram value doesn't converge; it's an *inconsistent* estimate of the signal power spectrum
- Bias: The window function (finite duration, discrete sampling) biases the expectation value of the periodogram away from the true signal power—*spectral leakage*

Spectral Leakage

1024 samples at 48 kHz sampling rate
$S/N = 5$, white noise
$f = 1031.25$ Hz (Fourier frequency) and $f = 1008$ Hz



Spectral leakage caused by “windowing”



Variance and bias reduction

Bias reduction

Tapering (aka, apodizing, windowing): Multiply data by a smoothly varying function that reduces leakage
Multitapering averages results from a few special tapers tuned to avoid redundancy. See Scargle (1997).

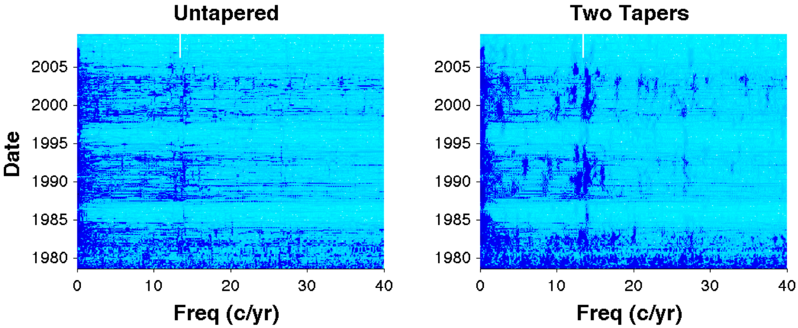
Variance reduction

Idea: Trade decreased frequency resolution (fewer frequencies than $N/2$) for reduced variability (by averaging estimates for kept frequencies)

- *Bartlett's method:* Average periodograms of non-overlapping chunks of data
- *Welch's method:* Average periodograms of *overlapping* chunks of data, but tapered to ameliorate effect of overlap

Time-frequency power spectrum via multitapering

MATLAB implementation by Scargle, applied to 33 y of solar Ca II K line data from Sacramento Peak National Solar Observatory (Keil & Worden 1984):



Tapering and averaging work well for estimating fairly smooth features *in the spectrum, but are suboptimal for measuring genuinely periodic signals*

Agenda

1. **Math background: Fourier transforms**

2. **Periodograms for sampled/probed time series**
   - Periodogram as a smooth spectrum estimator
   - Periodogram as a sinusoid period finder
   - Periodogram as a log likelihood for a GP

Periodogram as a period finder

Adopt a sinusoid periodic signal model (a δ-function spectrum!):

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

$$y_i = f(t_i) + \epsilon_i \qquad \text{Gaussian error } \textit{pdf}\text{s; rms} = \sigma$$

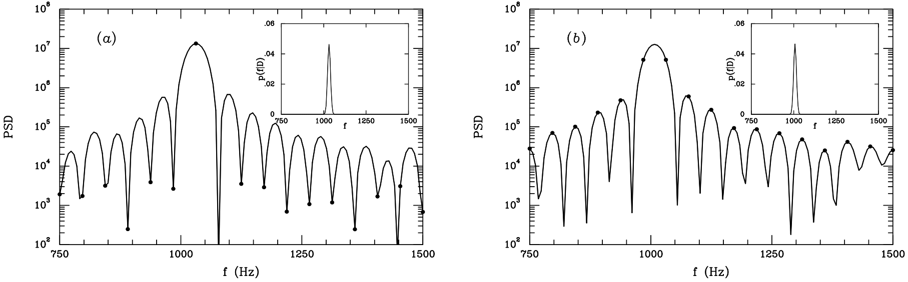
Estimate ω via profile likelihood, or Bayes (*Jaynes-Bretthorst alg.*):

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

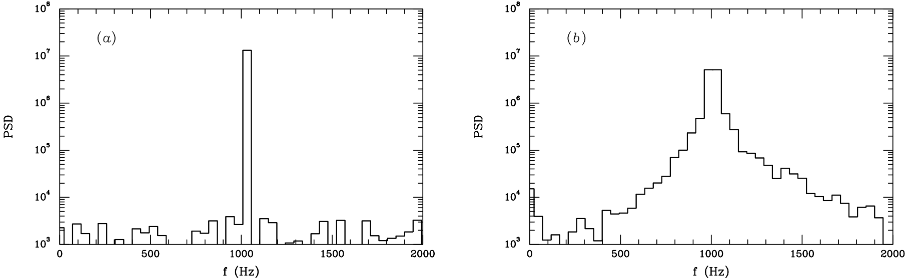
- Equally-spaced samples: $S(\omega) \to P(\omega)$ for large N
- Unequally-spaced samples: $S(\omega) \approx$ Lomb-Scargle periodogram

The posterior dist'n for ω requires computing a *continuous* version of the periodogram/power spectrum—the DTFT, not the DFT

1024 samples at 48 kHz sampling rate
$S/N = 5$
$f = 1031.25$ Hz (Fourier frequency) and $f = 1008$ Hz



Same functions, but here evaluated only at Fourier frequencies (whence "leakage" when true frequency $\neq$ Fourier frequency):



Leakage is just the shape of the actual likelihood function becoming partly apparent when the true frequency isn't a Fourier frequency, and you only look at Fourier frequencies

To produce accurate frequency estimates, you need to use the DTFT, not the DFT, and evaluate it much more finely than at Fourier frequencies (e.g., by direct calculation, or via algorithms like fractional FTs)

Contrast with spectrum estimation

- Periodogram here is not a power spectrum estimator, but the logarithm of the marginal *pdf* for ω
- No special role for Fourier frequencies
- No "leakage;" log *pdf* has similar structure for *all* signal frequencies, but sidelobes get *exponentiatally attenuated* (no need for smoothing) — *sidelobes quantify period uncertainty*
  See Jupyter notebook for an exercise exploring this
- Detection via Bayesian model comparison uses the *marginal likelihood* for the signal (vs. a pure noise marginal likelihood):

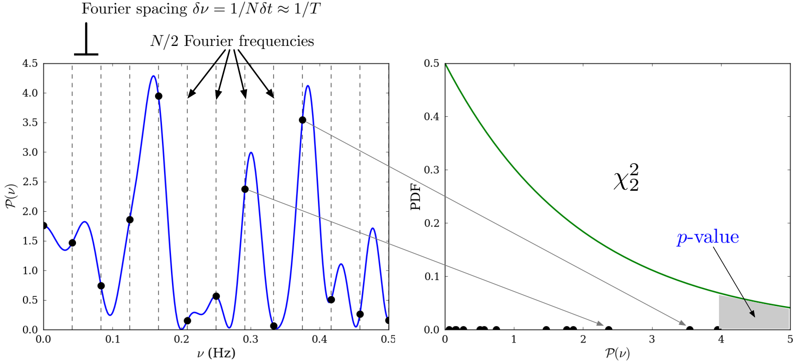
  $$\mathcal{L}(signal) \approx \exp\left[\frac{S_{\rm max}}{\sigma^2}\right] \times \frac{\text{peak width}}{\text{prior search range}}$$

  Handles period/phase/amplitude uncertainty via *marginalization* over the search space, *not optimization*
- Conventional periodograms optimal only for *single sinusoids*

Periodic signal detection: NHST vs. Bayes

Null hypothesis significance testing

Compute null dist'n for periodogram ordinates:



Find the maximum observed value of the periodogram

Report a *p-value*: Probability under the null (just noise) hypothesis for seeing a periodogram peak as big as observed, *or bigger*

This is a measure of how surprising the periodogram peak is, but statisticians still debate exactly how to interpret it

What a p-value really means

In the voice of Don LaFontaine or Lake Bell:

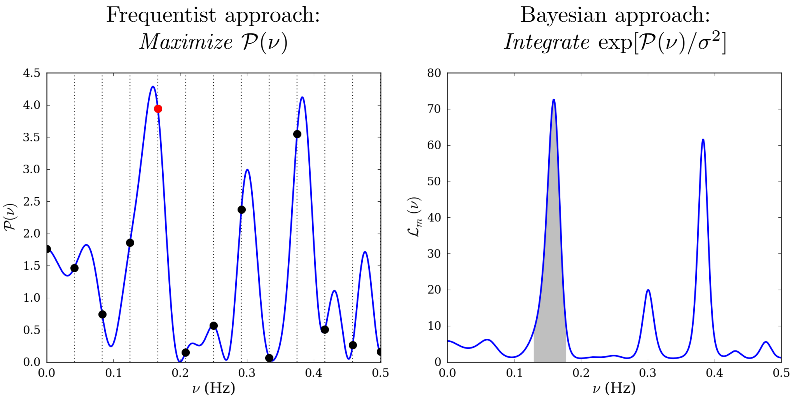
In a world. . . *with absolutely no periodic sources,*
with a threshold set so we wrongly claim to detect a source $100 \times p\%$ *of the time,*
this data would wrongly be considered a detection—and it would be the data providing the weakest *evidence for a periodic source in that world.*

Who wants to say *that*?!

Whence "p-value"—a measure of "surprisingness" under the null whose main virtue is that p is uniformly distributed under the null.

Frequentist vs. Bayesian detection

Optimize vs. marginalize



Frequentist p-value must adjust for # of frequencies examined;
$p_{\mathrm{mult}} \approx N p_1$

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

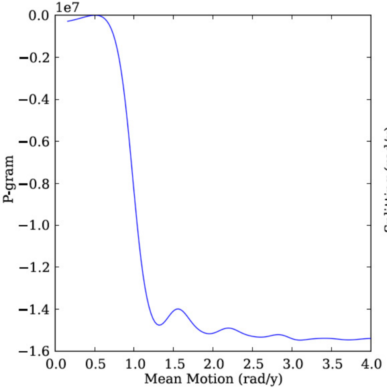
$$p(\omega_1, \omega_2|D) \propto p(\omega_1, \omega_2) J(\omega_1, \omega_2) \exp\left[\frac{S(\omega_1)}{\sigma^2}\right] \exp\left[\frac{S(\omega_2)}{\sigma^2}\right]$$

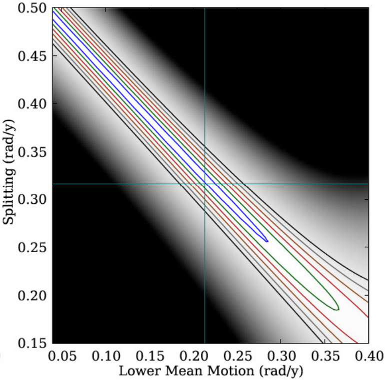
a product of two *independent* dist'ns, each determined by the periodogram

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

Two-planet example

Singlet periodogram (i.e., single-sinusoid, left) and *doublet periodogram* (right) for simulated *SIM* astrometry data from a Jupiter (11.9 y) + Saturn (29.5 y) system at 10 pc with $\approx 45^\circ$ inclination. The data span 10 y, with 0.86 μas errors. The true mean motions (shown by crosshair) are 0.53 and 0.21 rad/y.





Agenda

1. Math background: Fourier transforms

2. **Periodograms for sampled/probed time series**
   - Periodogram as a smooth spectrum estimator
   - Periodogram as a sinusoid period finder
   - Periodogram as a log likelihood for a GP

Periodogram as a Gaussian process log likelihood

Model *stochastic variability* via a *stationary Gaussian process with parametric covariance function* $C(u;\theta)$ for lag u and params θ:

$$C(u;\theta) = \mathrm{E}[f(t)f(t-u)] \qquad \text{(zero-mean case)}$$

This process has power spectrum

$$P(\omega;\theta) = \mathcal{F}\{C(u;\theta)\}$$

Then (under some conditions on C), for equally spaced samples and large N, the likelihood function for θ may be approximately calculated using the periodogram (*Whittle likelihood*, $O(N \log N)$ i/o $O(N^2)$; see Vaughan's `spritzer`:

$$\mathcal{L}(\theta) \propto \prod_{j=1}^{N/2} \frac{1}{P(\omega_j;\theta)} \exp\left[-\frac{S(\omega_j)}{P(\omega_j;\theta)}\right]$$

where $\{\omega_j\}$ are the Fourier frequencies (can be signif. biased)

Intuition: Recall that under the white-noise null, periodogram ordinates at Fourier frequencies follow independent χ^2_2 (i.e., exponential) distributions

Wikipedia

"More lives have been lost looking at the raw periodogram than by any other action involving time series!"
— *John Tukey*

Using periodograms w/o losing your life

Tailor manipulations to goals:

- Nonparametrically estimate a smooth spectrum:
  - Raw power null distribution $\propto e^{-S(\omega_i)/\sigma^2}$
  - Raw periodogram is inconsistent and biased (window, leakage) $\rightarrow$ smooth/taper it
  - Multitapering: Scargle's MATLAB code handling time-frequency and irregular sampling; NIPY's nitime package
- Detect a periodic signal (spectrum assumed to have δ-function):
  - Single sinusoid: $\mathcal{L}_m(\omega) \propto e^{S(\omega)/\sigma^2}$
  - Multiple sinusoids (harmonics?) $\rightarrow$ *multiplet periodograms*
  - Jaynes-Bretthorst algorithm for building generalized periodograms (Python package forthcoming)
- Whittle likelihood for semiparametric stationary GPs