Lecture 13
Spectral analysis revisited

Reading: Chapter 13 of Gregory “The Bayesian Revolution in Spectral Analysis”
Next few lectures

• Revisit spectral analysis
  • Issues with Fourier based estimators
    – 100% errors unless degrees of freedom increased
    – Nonuniform sampling: Lomb-Scargle method
      » Still suffers from large sidelobes
      » Can apply CLEAN algorithm
    – Poor performance with ‘red’ processes (leakage = bias)
  • Alternative approaches:
    – Bayesian method
    – Cholesky decomposition
    – Maximum entropy method
    – “Designer” methods for specific criteria (e.g. no bias)
    – Etc.
Spectral Analysis Methods

• Goals:
  – Power in a spectral line
  – Frequency of spectral line
  – Shape of a spectrum
    – continuum only,
    – continuum + line(s)
  – Signal detection
    – Null hypothesis = noise only (what type?)
    – Competing hypothesis: signal + noise
  – Testing null hypothesis if white noise
    – Tests for “whiteness” = test for statistical independence
    – Line fitting: consistent with zero slope?
    – Zero-crossing tests after subtracting mean
    – Etc.
A Simple Example

- Consider white noise that we calculate a spectral estimate for using the DFT.
  - N data points, \( \{x_j, j=1,N\} \)
  - what does the spectral estimate look like?

- Why would we use such an estimator?

- Suppose we know that the process is white noise. What would be a better spectral estimator than the DFT based one?

- In general, where the “true” spectrum could be anything, how do we incorporate prior information?
Consequences of Non-uniform Sampling for Fourier-based Spectral Analysis

- Gaps, non-uniformities = missing information
- Zero fill: fill missing values with zeros
- Spectrum = $|\text{DFT}|^2/N$:
  - Distortion of spectral lines
  - Increased sidelobes
  - Decreased S/N
  - Correlations between Fourier amplitudes
- Basis vectors no longer orthogonal
- Covariance matrix of Fourier amplitudes no longer diagonal
- An ad hoc fix: CLEAN algorithm
DFT Examples
Examples of Nonuniform Sampling

Cases shown: time series and power spectra

- Uniform sampling of sinusoids
- Gaps = large blocks of missing samples
- Irregular sampling: fraction of points missing at random locations
3 Sine waves  uniform sampling

Data

Spectrum = \|\text{FFT}\|^2
Data

3 Sine waves with gaps
3 Sine waves with gaps

Data

Spectrum = |FFT|^2
Data

3 Sine waves random sampling 50 per cent zeroed

Spectrum = $|\text{FFT}|^2$
Spectrum of random window
3 Sine waves with gaps
3 Sine waves with gaps

Data

Spectrum = |FFT|^2
3 Sine waves random sampling 95 per cent zeroed

Data

Spectrum = |FFT|^2
A Simple Example

• Consider white noise that we calculate a spectral estimate for using the DFT.
  – N data points, \( \{x_j, j=1,N\} \)
  – what does the spectral estimate look like?
• Why would we use such an estimator?
• Suppose we know that the process is white noise. What would be a better spectral estimator than the DFT based one?
• In general, where the “true” spectrum could be anything, how do we incorporate prior information?
Zero-crossing Test

- $X_n$, $n=1, \ldots, N$ values
- $\delta X_n = X_n - \langle X \rangle_{\text{est}}$ Subtract mean
- Expected number of zero crossings for white Gaussian noise
- $Z_W = ?$ What is your guess?
- Non-Gaussian noise (e.g. shot noise):
  - Can have far fewer zero crossings
Lecture Notes in Statistics

Edited by J. Berger, S. Fienberg, J. Gani, K. Krickeberg, and B. Singer

G. Larry Bretthorst

Bayesian Spectrum Analysis and Parameter Estimation

Springer-Verlag
A Bayesian Approach to Spectral Analysis

Chirped Signals

Chirped signals are oscillating signals with time variable frequencies, usually with a linear variation of frequency with time. E.g.

\[ f(t) = A \cos(\omega t + \alpha t^2 + \theta). \]

Examples:
- plasma wave diagnostic signals
- Signals propagated through dispersive media (seismic cases, plasmas)
- Gravitational waves from inspiraling binary stars
- Doppler-shifted signals over fractions of an orbit (e.g. acceleration of pulsar in its orbit)

Jaynes’ Approach to Spectral Analysis:


Briefly by Gregory in Chapter 13 of *Bayesian Logical Data Analysis for the Physical Sciences*

Result: Optimal processing is a **nonlinear** operation on the data without recourse to smoothing. However, the DFT-based spectrum (the “periodogram”) plays a key role in the estimation.
Start with Bayes’ theorem

\[
p(H/DI) = p(H/I) \frac{p(D/HI)}{p(D/I)}
\]

In this context, probabilities represent a simple mapping of degrees of belief onto real numbers.

Recall

\[ p(D/HI) \text{ vs. } D \text{ for fixed } H = \text{“sampling distribution”} \]

\[ p(D/HI) \text{ vs. } H \text{ for fixed } D = \text{“likelihood function”} \]

Read \( H \) as a statement that a **parameter vector** lies in a region of parameter space.
Measured Quantity:

\[ y(t) = f(t) + e(t) \]
\[ f(t) = A \cos(\omega t + \alpha t^2 + \theta) \]
\[ e(t) = \text{white gaussian noise}, \langle e \rangle = 0, \langle e^2 \rangle = \sigma^2 \]

Data Set:

\[ D = \{ y(t), \ |t| \leq T \}, \quad N = 2T + 1 \text{ data points}. \]
Data Probability: The probability of obtaining a data set of $N$ samples is

$$P(D|HI) = \sum_t P[y(t)] = \prod_{t=-T}^{T} (2\pi\sigma^2)^{-1/2} e^{-\left\{ \frac{1}{2\sigma^2} [y(t) - f(t)]^2 \right\}},$$  \hspace{1cm} (1)$$

which we can rewrite as a likelihood function once we acquire a data set and evaluate the probability for a specific $H$. Writing out the parameters explicitly, the likelihood function is

$$L(A, \omega, \alpha, \theta) \propto e^{-\left\{ \frac{1}{2\sigma^2} \sum_{t=-T}^{T} [y(t) - A\cos(\omega t + \alpha t^2 + \theta)]^2 \right\}}$$

For simplicity, assume that $\omega T \gg 1$ so that many cycles of oscillation are summed over.

Then

$$\sum_t \cos^2(\omega t + \alpha t^2 + \theta) = \sum_t \frac{1}{2} [1 + \cos 2(\omega t + \alpha t^2 + \theta)]$$

$$\approx \frac{2T + 1}{2}$$

$$\equiv \frac{N}{2}$$
Expand the argument of the exponential in the likelihood function, we have

\[
[y(t) - A \cos(\omega t + \alpha t^2 + \theta)]^2 = y^2(t) + A^2 \cos^2(\omega t + \alpha t^2 + \theta) - 2Ay(t) \cos(\omega t + \alpha t^2 + \theta)
\]

We care only about terms that are functions of the parameters, so we drop the \(y^2(t)\) term to get

\[
- \frac{1}{2\sigma^2} \sum_{t=-T}^{T} [y(t) - A \cos(\omega t + \alpha t^2 + \theta)]^2 \rightarrow - \frac{1}{2\sigma^2} \sum_{t} [A^2 \cos^2(\omega t + \alpha t^2 + \theta) - 2Ay(t) \cos(\omega t + \alpha t^2 + \theta) + \frac{NA^2}{4\sigma^2}]
\]

The likelihood function becomes

\[
L(A, \omega, \alpha, \theta) \propto e^{\left\{\frac{A}{\sigma^2} \sum_{t} y(t) \cos(\omega t + \alpha t^2 + \theta) - \frac{NA^2}{4\sigma^2}\right\}}
\]

**Integrating out the phase:**

In calculating a power spectrum [in this case, a chirped power spectrum (“chirpogram”)], we do not care about the phase of any sinusoid in the data. In Bayesian estimation, such a parameter is called a nuisance parameter.

Since we do not know anything about \(\theta\), we integrate over its prior distribution, a pdf that is
uniform over $[0, 2\pi]$:

$$f_\theta(\theta) = \begin{cases} \frac{1}{2\pi} & 0 \leq \theta \leq 2\pi \\ 0 & \text{otherwise.} \end{cases}$$

The marginalized likelihood function becomes

$$L(A, \omega, \alpha) \propto \frac{1}{2\pi} \int_0^{2\pi} d\theta \ L(A, \omega, \alpha, \theta)$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\theta \ \exp \left\{ \frac{A}{\sigma^2} \sum \ y(t) \cos(\omega t + \alpha t^2 + \theta) - \frac{NA^2}{4\sigma^2} \right\}$$

$$= \exp \left( \frac{NA^2}{4\sigma^2} \right) \left[ \frac{1}{2\pi} \int_0^{2\pi} d\theta \ \exp \left( \frac{A}{\sigma} \sum \ y(t) \cos(\omega t + \alpha t^2 + \theta) \right) \right]$$

Using the identity

$$\cos(\omega t + \alpha t^2 + \theta) = \cos(\omega t + \alpha t^2) \cos \theta - \sin(\omega t + \alpha t^2) \sin \theta$$

we have

$$\sum_{t} y(t) \cos(\omega t + \alpha t^2 + \theta) = \cos \theta \sum_{t} y(t) \cos(\omega t + \alpha t^2) - \sum_{t} y(t) \sin(\omega t + \alpha t^2)$$

$$\equiv P \cos \theta - Q \sin \theta$$

$$= \sqrt{P^2 + Q^2} \cos[\theta + \tan^{-1}(Q/P)].$$
This result may be used to evaluate the integral over $\theta$ in the marginalized likelihood function:

$$
\frac{1}{2\pi} \int_0^{2\pi} d\theta \exp\left(\frac{A}{\sigma} \sum_t y(t) \cos(\omega t + \alpha t^2 + \theta)\right) = \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{\frac{A}{\sigma^2} \sqrt{P^2 + Q^2} \cos[\theta + \tan^{-1}(Q/P)]}
$$

To evaluate the integral we use the identity,

$$
I_0(x) \equiv \frac{1}{2\pi} \int_0^{2\pi} d\theta e^{x \cos \theta} = \text{modified Bessel function}
$$

This yields

$$
\frac{1}{2\pi} \int_0^{2\pi} d\theta \exp\left(\frac{A}{\sigma} \sum_t y(t) \cos(\omega t + \alpha t^2 + \theta)\right) = I_0\left(\frac{A}{\sigma^2} \sqrt{P^2 + Q^2}\right)
$$

We now simplify $P^2 + Q^2$:

$$
P^2 + Q^2 = \left[\sum_t y(t) \cos(\omega t + \alpha t^2)\right]^2 + \left[\sum_t y_t \sin(\omega t + \alpha t^2)\right]^2
$$

$$
= \sum_t \sum_{t'} y(t)y(t') \left[\cos(\omega t + \alpha t^2) \cdot \cos(\omega t' + \alpha t'^2) + \frac{\sin(\omega t + \alpha t^2) \sin(\omega t' + \alpha t'^2)}{\cos[\omega(t - t') + \alpha(t - t')^2]}\right]
$$

$$
P^2 + Q^2 = \sum_t \sum_{t'} y(t)y(t') \cos[\omega(t - t') + \alpha(t - t')^2].
$$
Define

\[
C(\omega, \alpha) \equiv N^{-1}(P^2 + Q^2) = N^{-1} \sum_t \sum_{t'} y(t)y(t') \cos[\omega(t - t') + \alpha(t - t')^2],
\]

Then the integral over \( \theta \) gives

\[
\int_0^{2\pi} d\theta \ L(A, \omega, \alpha, \theta) \equiv I_0 \left( \frac{A\sqrt{NC(\omega, \alpha)}}{\sigma^2} \right)
\]

and the marginalized likelihood is

\[
L(A, \omega, \alpha) = e^{-\frac{NA^2}{4\sigma^2}} I_0 \left[ \frac{A\sqrt{NC(\omega, \alpha)}}{\sigma^2} \right].
\]
Notes:

(1) The data appear only in $C(\omega, \alpha)$.

(2) $C$ is a **sufficient** statistic, meaning that it contains all information from the data that are relevant to inference using the likelihood function.

(3) How do we read $L(A, \omega, \alpha)$? As the **probability distribution** of the parameters $A, \omega, \alpha$ in terms of the data dependent quantity $C(\omega, \alpha)$. (Note that $L$ is not normalized as a PDF). As such, $L$ is a quite different quantity from the Fourier-based power spectrum.

(4) What is the quantity

$$C(\omega, \alpha) \equiv N^{-1} \sum_t \sum_{t'} y(t)y(t') \cos[\omega(t-t') + \alpha(t-t')^2]?$$

For a given data set, $\omega, \alpha$ are variables. If we plot $C(\omega, \alpha)$, we expect to get a large value when $\omega = \omega_{\text{signal}}, \alpha = \alpha_{\text{signal}}$.

(5) For a non-chirped but oscillatory signal ($\omega \neq 0, \alpha = 0$), the quantity $C(\omega, \alpha)$ is nothing other than the periodogram (the squared magnitude of the Fourier transform of the data). We then see that, for this case, the likelihood function is a nonlinear function of the Fourier estimate for the power spectrum.
A Limiting Form:

For argument $x \gg 1$, the Bessel function

$$I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}}.$$ 

In this case the marginalized likelihood is

$$L(A, \omega, \alpha) \propto e^{-\frac{NA^2}{4\sigma^2}} I_0\left(\frac{A\sqrt{NC(\omega, \alpha)}}{\sigma^2}\right)$$

$$\propto e^{-\frac{NA^2}{4\sigma^2}} \times e^{\frac{A\sqrt{NC(\omega, \alpha)}}{\sigma^2}} \times \frac{A\sqrt{NC(\omega, \alpha)}}{(2\pi A\sqrt{NC(\omega, \alpha)}\sigma^2)^{1/2}}.$$ 

Since $C(\omega, \alpha)$ is large when $\omega$ and $\alpha$ match those of any true signal, we see that it is exponentiated as compared to appearing linearly in the periodogram.
Interpretation of the Bayesian and Fourier Approaches

We found the marginalized likelihood for the frequency and chirp rate to be

$$L(A, \omega, \alpha) = e^{-\frac{NA^2}{4\sigma^2}} I_0 \left[ \frac{A\sqrt{N}C(\omega, \alpha)}{\sigma^2} \right].$$

and the limiting form for the Bessel function’s argument $x \gg 1$ is

$$I_0(x) \sim \frac{e^x}{\sqrt{2\pi x}}.$$

In this case the marginalized likelihood is

$$L(A, \omega, \alpha) \propto e^{-\frac{NA^2}{4\sigma^2}} I_0 \left( \frac{A\sqrt{N}C(\omega, \alpha)}{\sigma^2} \right)$$

$$\propto e^{-\frac{NA^2}{4\sigma^2}} \times e^{-\frac{A\sqrt{N}C(\omega, \alpha)}{\sigma^2}} \times \left( \frac{2\pi A\sqrt{N}C(\omega, \alpha)}{\sigma^2} \right)^{1/2}.$$

Since $C(\omega, \alpha)$ is large when $\omega$ and $\alpha$ match those of any true signal, we see that it is exponentiated as compared to appearing linearly in the periodogram.
Now let’s consider the case with no chirp rate, $\alpha = 0$. Examples in the literature show that the width of the Bayesian PDF is much narrower than the periodogram, $C(\omega, 0)$. Does this mean that the uncertainty principle has been avoided?

The answer is no!

**Uncertainty Principle in the Periodogram:**

For a data set of length $T$, the frequency resolution implied by the spectral window function is

$$\delta \omega \sim 2\pi \delta f \sim \frac{2\pi}{T}.$$ 

**Width of the Bayesian PDF:**

When the argument of the Bessel function is “large” the exponentiation causes the PDF to be much narrower than the spectral window for the periodogram.
**Interpretation:**

The periodogram is the distribution of power (or variance) with frequency for the particular realization of data used to form the periodogram. The spectral window also depicts the distribution of variance for a pure sinusoid in the data (with infinite signal to noise ratio).

The Bayesian posterior is the PDF for the frequency of a sinusoid and therefore represents a very different quantity than the periodogram and are thus not directly comparable.
1. The Bayesian method addresses the question, “what is the PDF for the frequency of the sinusoid that is in the data?"

2. The periodogram is the distribution of variance in frequency.

3. If we use the periodogram to estimate the sinusoid’s frequency, we get a result that is more comparable:
   
   (a) First note that the width of the posterior PDF involves the signal to noise ratio (in the square root of the periodogram) $\sqrt{N}A/\sigma$ while the width of the periodogram’s spectral window is independent of the SNR.

   (b) General result: if a spectral line has width $\Delta \omega$, its centroid can be determined to an accuracy

   \[ \delta \omega \sim \frac{\Delta \omega}{\text{SNR}}. \]

   This result follows from matched filtering, which we will discuss later on.

   (c) Quantitatively, the periodogram yields the same information about the location of the spectral line as does the posterior PDF.

4. Problem: derive an estimate for the width of the posterior PDF that can be compared with the estimate for the periodogram.
Figure 1: Left: Time series of sinusoid + white noise with $A/\sigma = 1$ sampled $N = 500$ times over an interval of length $T = 500$. Right: Plot of the periodogram (red) and Bayesian PDF of the time series.
Figure 2: Left: Time series of sinusoid + white noise with $A/\sigma = 1/4$ sampled $N = 500$ times over an interval of length $T = 500$. Right: Plot of the periodogram (red) and Bayesian PDF of the time series.