Time Series Analysis and Classification

Tarek Medkour

- Part 1: Time Series Analysis
  - Introduction
  - Time Series Models
  - Spectral Analysis
  - Time-Frequency Representation
  - Multivariate Time Series
- Part 2: Time Series Classification
  - Pattern Recognition and Detection
  - Feature Extraction and Selection
  - Models and Representation Learning
  - Data Enhancement and Preprocessings
  - Change-Point and Anomaly Detection

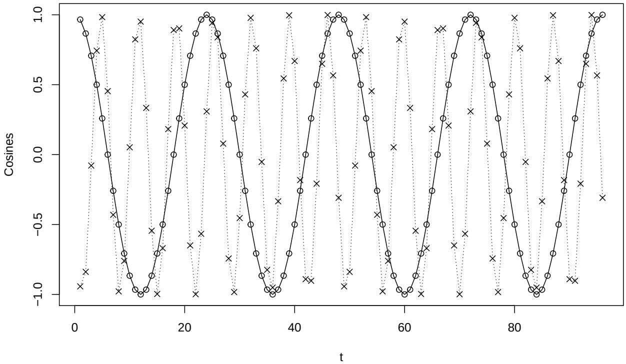
Introduction to Spectral Analysis

Idea:

Decompose a stationary time series $\{X_t\}$ into a combination of sinusoids, with random (and uncorrelated) coefficients. Just as in Fourier analysis, where we decompose (deterministic) functions into combinations of sinusoids. This is referred to as âspectral analysisâ or analysis in the âfrequency domain,â in contrast to the time domain approach we have considered so far. The frequency domain approach considers regression on sinusoids; the time domain approach considers regression on past values of the time series.

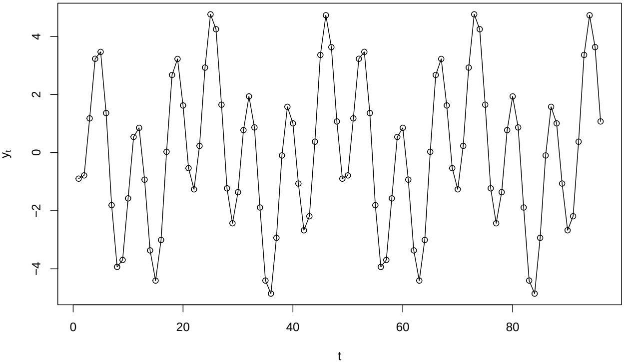
Example

We consider two discrete-time cosine curves with time running from 1 to 96. The frequencies are 4/96 and 14/96, respectively. The lower-frequency curve has a phase of zero, but the higher-frequency curve is shifted by a phase of 0.6π.



Linear combination of the two cosine curves with a multiplier of 2 on the low-frequency curve and a multiplier of 3 on the higher-frequency curve and a phase of 0.6π

$$Y(t) = 2\cos\left(2\pi\frac{4t}{96}\right) + 3\cos\left(2\pi\frac{14t}{96} + 0.3\right)$$



Recall

The cosine curve has the equation:

$$x(t) = R\cos(2\pi f t + \Phi) \tag{1}$$

where $R(> 0)$ is the amplitude, f the frequency, and Φ the phase of the curve.

The period of the cosine wave is $1/f$ time units, as the curve repeats itself exactly every $1/f$ time units.

Reparameterising the Cosine Curve

This expression is not convenient for estimation because the parameters R and Φ do not enter the expression linearly. Instead, we use a trigonometric identity to reparameterise the equation as:

$$R\cos(2\pi ft + \Phi) = A\cos(2\pi ft) + B\sin(2\pi ft) \tag{2}$$

where:

$$R = \sqrt{A^2 + B^2} \tag{3}$$

$$\Phi = \arctan(-B/A) \tag{4}$$

$$A = R\cos(\Phi) \tag{5}$$

$$B = -R\sin(\Phi) \tag{6}$$

Linear Combination of Cosine Curves

A general linear combination of m cosine curves with arbitrary amplitudes, frequencies, and phases can be written as:

$$Y(t) = A_0 + \sum_{j=1}^{m} [A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)] \tag{7}$$

Fourier Frequencies

If the frequencies f_j are the **Fourier frequencies**. That is of the form j/n, for $j = 1, \ldots, n$, the cosine and sine predictor variables are orthogonal

In this case, the least squares estimates are given by:

$$\hat{A}_0 = \bar{Y} \tag{8}$$

$$\hat{A}_j = \frac{2}{n}\sum_{t=1}^{n} Y_t \cos(2\pi t j/n) \tag{9}$$

$$\hat{B}_j = \frac{2}{n}\sum_{t=1}^{n} Y_t \sin(2\pi t j/n) \tag{10}$$

Example Fitting

If we apply these formulas to the series shown in the example, we would obtain perfect results. At frequency $f_4 = 4/96$, we would obtain $\hat{A}_4 = 2$ and $\hat{B}_4 = 0$, and at frequency $f_{14} = 14/96$, we would obtain $\hat{A}_{14} = -0.927051$ and $\hat{B}_{14} = -2.85317$. We would obtain estimates of zero for the regression coefficients at all other frequencies.

Note

Any series of any length n, whether deterministic or stochastic and with or without any true periodicities, can be fit perfectly by the combination of cosines model

$$Y(t) = A_0 + \sum_{j=1}^{m} [A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)] \qquad (11)$$

by choosing $m = n/2$ if n is even and $m = (n-1)/2$ if n is odd. There are then n parameters to adjust (estimate), in order to fit the series of length n.

The Periodogram

Odd Sample Size

For odd sample sizes with $n = 2k + 1$, the periodogram I at frequency $f = j/n$ for $j = 1, 2, \ldots, k$ is defined as:

$$I\left(\frac{j}{n}\right) = \frac{n}{2}(\hat{A}_j^2 + \hat{B}_j^2) \tag{12}$$

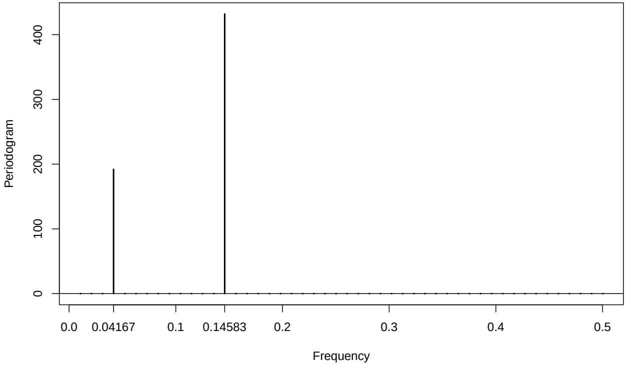
Even Sample Size

If the sample size is even and $n = 2k$, the periodogram is given by:

$$I\left(\frac{j}{n}\right) = \frac{n}{2}(\hat{A}_j^2 + \hat{B}_j^2) \text{ for } j = 1, 2, \ldots, k-1 \tag{13}$$

$$I\left(\frac{1}{2}\right) = n(\hat{A}_k)^2 \tag{14}$$

For our example, the heights show the presence and relative strengths of the two cosine-sine components quite clearly. Note also that the frequencies $4/96 \approx 0.04167$ and $14/96 \approx 0.14583$ have been marked on the frequency axis.



Interpretation of the Periodogram

Relative Strength of Frequencies

The height of the periodogram shows the relative strength of cosine-sine pairs at various frequencies in the overall behavior of the series.

Analysis of Variance

The periodogram $I(j/n)$ is the sum of squares with two degrees of freedom associated with the coefficient pair (A_j, B_j) at frequency j/n, so we have:

$$\sum_{j=1}^{n}(Y_j - \bar{Y})^2 = \sum_{j=1}^{k} I\left(\frac{j}{n}\right) \tag{15}$$

when $n = 2k + 1$ is odd. A similar result holds when n is even.

Efficient Computation

Fast Fourier Transform

For long series, the computation of a large number of regression coefficients might be intensive. Fortunately, quick, efficient numerical methods based on the fast Fourier transform (FFT) have been developed that make the computations feasible for very long time series.

Periodogram with Noisy Data

Does the periodogram work just as well when we do not know where or even if there are cosines in the series? What if the series contains additional ”noise”?

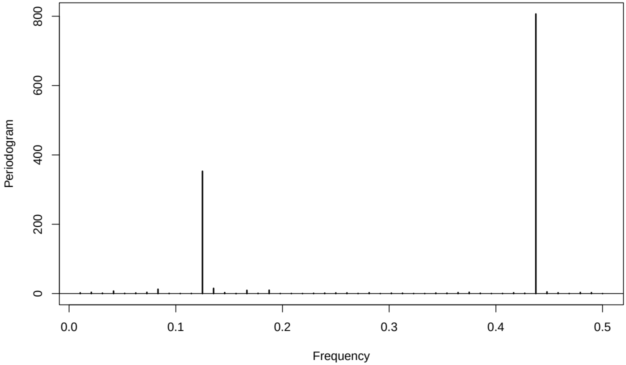
Simulated Example

To illustrate, we generate a time series using randomness to select the frequencies, amplitudes, and phases, and with additional additive white noise. The key elements are:

- The two frequencies are randomly chosen without replacement from among $1/96, 2/96, \ldots, 47/96$.
- The A's and B's are selected independently from normal distributions with means of zero and standard deviations of 2 for the first component and 3 for the second.
- A normal white noise series, $\{W_t\}$, with zero mean and standard deviation 1, is chosen independently of the A's and B's and added on.

$$
\begin{aligned}
Y_t = {} & A_1 \cos(2\pi f_1 t) + B_1 \sin(2\pi f_1 t) && (16)\\
& + A_2 \cos(2\pi f_2 t) + B_2 \sin(2\pi f_2 t) + W_t && (17)
\end{aligned}
$$

The periodogram clearly shows that the series contains two cosine-sine pairs at frequencies of about $12/96 = 0.125$ and $42/96 = 0.4375$ and that the higher-frequency component is much stronger.



Extending the Periodogram

Periodogram Definition

Although the Fourier frequencies are special, we extend the definition of the periodogram to all frequencies in the interval $[0, \frac{1}{2}]$:

$$I(f) = \frac{n}{2}(\hat{A}_f^2 + \hat{B}_f^2) \tag{18}$$

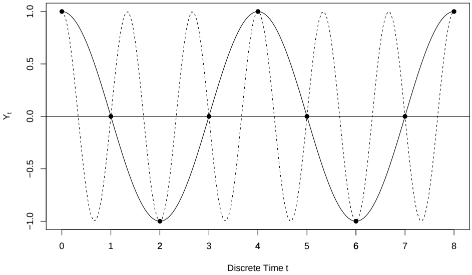
$$\hat{A}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \cos(2\pi t f) \tag{19}$$

$$\hat{B}_f = \frac{2}{n}\sum_{t=1}^{n} Y_t \sin(2\pi t f) \tag{20}$$

Frequency Interval

We restrict frequencies to the interval from 0 to $\frac{1}{2}$ because with discrete-time observations, we could never distinguish between frequencies f and $f + k(\frac{1}{2})$ for any positive integer k. These frequencies are said to be aliased with one another, and it suffices to limit attention to frequencies within the interval from 0 to $\frac{1}{2}$.

Here we have plotted two cosine curves, one with frequency $f = \frac{1}{4}$ and the other at frequency $f = \frac{3}{4}$.
If we only observe the series at the discrete-time points 0, 1, 2, 3,... the two series are identical. With discrete-time observations, we could never distinguish between these two curves. We say that the two frequencies $\frac{1}{4}$ and $\frac{3}{4}$ are aliased with one another.



Spectral Representation

Consider a time series represented as:

$$Y_t = \sum_{j=1}^{m} [A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)] \tag{21}$$

where the frequencies $0 < f_1 < f_2 < \cdots < f_m < \frac{1}{2}$ are fixed, and A_j and B_j are independent normal random variables with zero means and $\mathrm{Var}(A_j) = \mathrm{Var}(B_j) = \sigma_j^2$.

Covariance Structure

Then the time series $\{Y_t\}$ is stationary with mean zero and:

$$\gamma_k = \sum_{j=1}^{m} \sigma_j^2 \cos(2\pi k f_j) \tag{22}$$

- Thus, we can represent γ_k using a Fourier series. The coefficients are the variances of the sinusoidal components.

In particular, the process variance γ_0 is a sum of the variances due to each component at the various fixed frequencies:

$$\gamma_0 = \sum_{j=1}^{m} \sigma_j^2 \tag{23}$$

Spectral Density of a Time Series

If a time series $\{X_t\}$ has autocovariance γ satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$$

then we define its spectral density as

$$S(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h}$$

for $-\infty < \nu < \infty$.

Spectral Density: Some Facts

1. We have $\sum_{h=-\infty}^{\infty} \left|\gamma(h)e^{-2\pi i\nu h}\right| < \infty$.
2. S is periodic, with period 1. Thus, we can restrict the domain of S to $-1/2 \le \nu \le 1/2$.
3. S is even (that is, $S(\nu) = S(-\nu)$).
4. $S(\nu) \ge 0$.
5. $\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\nu h} S(\nu)\, d\nu$. The autocaviariance function and the spectral density are a **Fourier pair**

Example: White Noise

For white noise $\{W_t\}$, we have seen that $\gamma(0) = \sigma_w^2$ and $\gamma(h) = 0$ for $h \neq 0$.
Thus,

$$S(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h}$$
$$= \gamma(0) = \sigma_w^2.$$

That is, the spectral density is constant across all frequencies: each frequency in the spectrum contributes equally to the variance.
This is the origin of the name white noise: it is like white light, which is a uniform mixture of all frequencies in the visible spectrum.

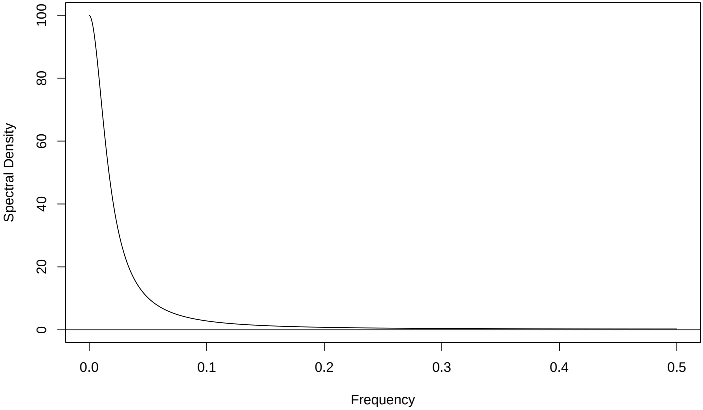
Example: AR(1)

For $X_t = \phi_1 X_{t-1} + W_t$, we have seen that $\gamma(h) = \sigma_w^2 \phi_1^{|h|}/(1-\phi_1^2)$. Thus,

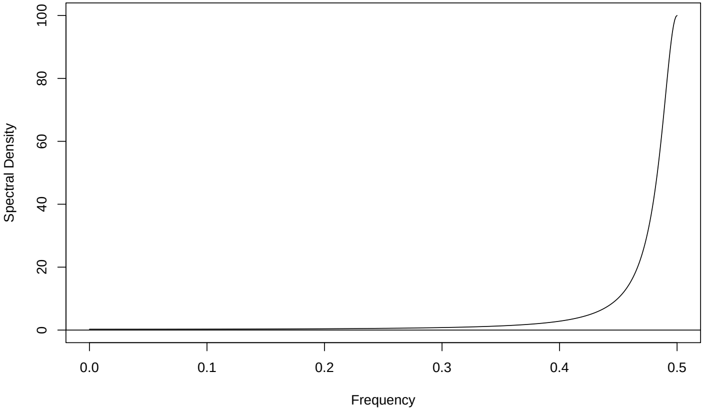
$$
\begin{aligned}
S(\nu) &= \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h} = \frac{\sigma_w^2}{1-\phi_1^2} \sum_{h=-\infty}^{\infty} \phi_1^{|h|} e^{-2\pi i \nu h} \\
&= \frac{\sigma_w^2}{1-\phi_1^2} \left(1 + \sum_{h=1}^{\infty} \phi_1^h \left(e^{-2\pi i \nu h} + e^{2\pi i \nu h} \right) \right) \\
&= \frac{\sigma_w^2}{1-\phi_1^2} \left(1 + \frac{\phi_1 e^{-2\pi i \nu}}{1-\phi_1 e^{-2\pi i \nu}} + \frac{\phi_1 e^{2\pi i \nu}}{1-\phi_1 e^{2\pi i \nu}} \right) \\
&= \frac{\sigma_w^2}{(1-\phi_1^2)} \frac{1-\phi_1 e^{-2\pi i \nu} \phi_1 e^{2\pi i \nu}}{(1-\phi_1 e^{-2\pi i \nu})(1-\phi_1 e^{2\pi i \nu})} \\
&= \frac{\sigma_w^2}{1-2\phi_1 \cos(2\pi\nu) + \phi_1^2}
\end{aligned}
$$

- If $\phi_1 > 0$ (positive autocorrelation), spectrum is dominated by low frequency components—smooth in the time domain.
- If $\phi_1 < 0$ (negative autocorrelation), spectrum is dominated by high frequency components—rough in the time domain.

Example: AR(1) Spectrum



Example: AR(1) Spectrum



Example: MA(1)

$$X_t = W_t + \theta_1 W_{t-1}.$$

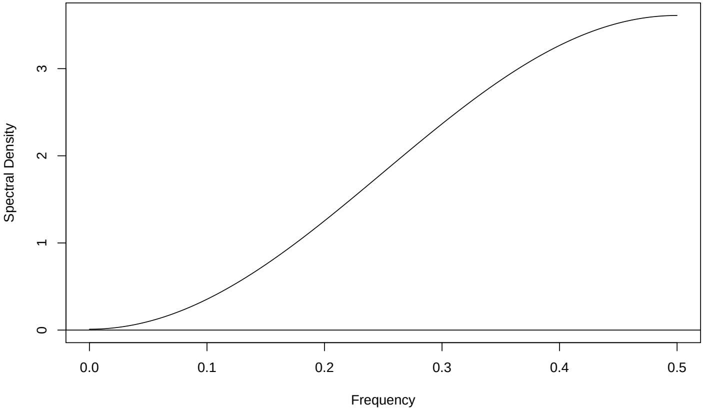
$$\gamma(h) = \begin{cases} \sigma_w^2(1+\theta_1^2) & \text{if } h = 0, \\ \sigma_w^2\theta_1 & \text{if } |h| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} S(\nu) &= \sum_{h=-1}^{1} \gamma(h) e^{-2\pi i \nu h} \\ &= \gamma(0) + 2\gamma(1)\cos(2\pi\nu) \\ &= \sigma_w^2(1 + \theta_1^2 + 2\theta_1 \cos(2\pi\nu)) \end{aligned}$$

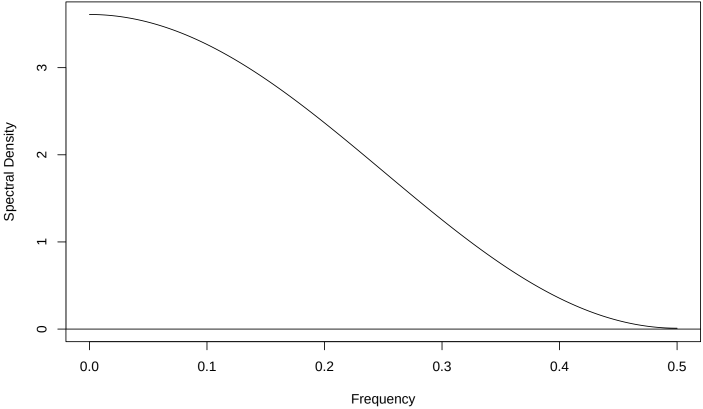
Example: MA(1)

- If $\theta_1 > 0$ (positive autocorrelation), spectrum is dominated by low frequency components—smooth in the time domain.
- If $\theta_1 < 0$ (negative autocorrelation), spectrum is dominated by high frequency components—rough in the time domain.

Example: MA(1) Spectrum



Example: MA(1) Spectrum



Discrete Spectral Distribution Function

For $X_t = A\sin(2\pi\lambda t) + B\cos(2\pi\lambda t)$, we have $\gamma(h) = \sigma^2\cos(2\pi\lambda h)$, and we can write

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i\nu h} dF(\nu),$$

where F is the discrete distribution

$$F(\nu) = \begin{cases} 0 & \text{if } \nu < -\lambda, \\ \frac{\sigma^2}{2} & \text{if } -\lambda \le \nu < \lambda, \\ \sigma^2 & \text{otherwise.} \end{cases}$$

The Spectral Distribution Function

For any stationary $\{X_t\}$ with autocovariance γ, we can write

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \nu h} dF(\nu),$$

where F is the spectral distribution function of $\{X_t\}$.
If γ is absolutely summable, F is continuous: $dF(\nu) = S(\nu)d\nu$, and we get

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \nu h} S(\nu) d\nu,$$

where $S(\nu)$ is the spectral density.

The Spectral Distribution Function

For $X_t = \sum_{j=1}^{k}(A_j \sin(2\pi\nu_j t) + B_j \cos(2\pi\nu_j t))$, the spectral distribution function is $F(\nu) = \sum_{j=1}^{k} \sigma_j^2 F_j(\nu)$, where

$$F_j(\nu) = \begin{cases} 0 & \text{if } \nu < -\nu_j, \\ \frac{1}{2} & \text{if } -\nu_j \le \nu < \nu_j, \\ 1 & \text{otherwise.} \end{cases}$$

The heights of the jumps in the spectral distribution give the variances associated with the various periodic components, and the positions of the jumps indicate the frequencies of the periodic components.
For $0 \le \nu_1 < \nu_2 \le \frac{1}{2}$, the integral $\int_{\nu_1}^{\nu_2} dF(\nu)$, gives the portion of the (total) process variance $F\left(\frac{1}{2}\right) = \gamma_0$ that is attributable to frequencies in the range ν_1 to ν_2.

Time-invariant linear filters

- A filter is an operator; given a time series $\{X_t\}$, it maps to a time series $\{Y_t\}$.
- We can think of a general linear process $X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}$ as the output of a causal linear filter with a white noise input.
- A time series $\{Y_t\}$ is the output of a linear filter $A = \{a_{t,j} : t, j \in \mathbb{Z}\}$ with input $\{X_t\}$ if

$$Y_t = \sum_{j=-\infty}^{\infty} a_{t,j} X_j.$$

- If $a_{t,t-j}$ is independent of t (i.e., $a_{t,t-j} = \psi_j$), then we say that the filter is **time-invariant**.
- If $\psi_j = 0$ for $j < 0$, we say the filter ψ is **causal**.
- We'll see that the name 'filter' arises from the frequency domain viewpoint.

Time-Invariant Linear Filters: Examples

1. $Y_t = X_{-t}$ is linear, but not time-invariant.
2. $Y_t = \frac{1}{3}(X_{t-1} + X_t + X_{t+1})$ is linear, time-invariant, but not causal:
$$\psi_j = \begin{cases} \frac{1}{3} & \text{if } |j| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$
3. For polynomials $\phi(B)$, $\theta(B)$ with roots outside the unit circle, $\psi(B) = \frac{\theta(B)}{\phi(B)}$ is a linear, time-invariant, causal filter.

Frequency response of a time-invariant linear filter

- The operation $\sum_{j=-\infty}^{\infty} \psi_j X_{t-j}$ is called the convolution of X with ψ.
- The sequence ψ is also called the impulse response.
- Suppose that $\{X_t\}$ has spectral density $S_X(\nu)$ and ψ is stable, that is, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$. Then

$$Y_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} = \psi(B) X_t$$

has spectral density

$$S_Y(\nu) = |\psi(e^{-2\pi i \nu})|^2 S_X(\nu).$$

Proof

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Y_{t-k}) &= \mathrm{Cov}\left(\sum_{j=-\infty}^{\infty} \psi_j X_{t-j}, \sum_{s=-\infty}^{\infty} \psi_s X_{t-k-s}\right) \\
&= \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \psi_j \psi_s \,\mathrm{Cov}(X_{t-j}, X_{t-k-s}) \\
&= \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \psi_j \psi_s \int_{-1/2}^{1/2} e^{2\pi i (s+k-j) f} S_X(f)\, df \\
&= \int_{-1/2}^{1/2} \left| \sum_{s=-\infty}^{\infty} \psi_s e^{-2\pi i s f} \right|^2 e^{2\pi i f k} S_X(f)\, df \\
&= \int_{-1/2}^{1/2} |\psi(e^{-2\pi i f})|^2 e^{2\pi i f k} S_X(f)\, df
\end{aligned}
$$

So,

$$\mathrm{Cov}(Y_t, Y_{t-k}) = \int_{-1/2}^{1/2} |\psi(e^{-2\pi i f})|^2 S_X(f) e^{2\pi i f k}\, df$$

But,

$$\mathrm{Cov}(Y_t, Y_{t-k}) = \int_{-1/2}^{1/2} S_Y(f) e^{2\pi i f k}\, df$$

Thus, we must have,

$$S_Y(f) = |\psi(e^{-2\pi i f})|^2 S_X(f)$$

Frequency response of a time-invariant linear filter

- The function $\nu \mapsto \psi(e^{-2\pi i\nu})$ (the polynomial $\psi(z)$ evaluated on the unit circle) is known as the **frequency response or transfer function** of the linear filter.
- The squared modulus, $\nu \mapsto |\psi(e^{-2\pi i\nu})|^2$, is known as the **power transfer function** of the filter.

Linear Processes and Filters

We have seen that a general linear process, $Y_t = \psi(B)e_t$, is a special case, since

$$S_Y(\nu) = |\psi(e^{-2\pi i\nu})|^2 S_e(\nu) = |\psi(e^{-2\pi i\nu})|^2 \sigma_e^2.$$

When we pass a time series $\{X_t\}$ through a linear filter, the spectral density is multiplied, frequency-by-frequency, by the squared modulus of the frequency response $\nu \mapsto |\psi(e^{-2\pi i\nu})|^2$. This is a version of the equality $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$, but the equality is true for the component of the variance at every frequency.
This is also the origin of the name 'filter.'

Frequency Response: Examples

Consider the moving average

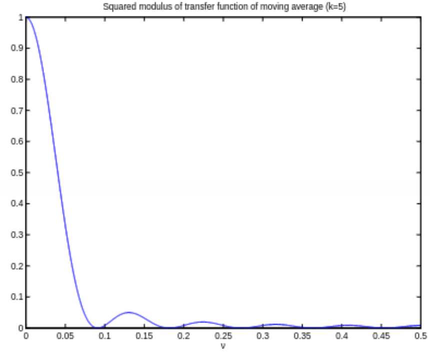
$$Y_t = \frac{1}{2k+1} \sum_{j=-k}^{k} X_{t-j}.$$

This is a time-invariant linear filter (but not causal). Its transfer function is the Dirichlet kernel

$$\psi(e^{-2\pi i\nu}) = D_k(2\pi\nu) = \frac{1}{2k+1} \sum_{j=-k}^{k} e^{-2\pi i j\nu} \tag{24}$$

$$= \begin{cases} 1 & \text{if } \nu = 0, \\ \frac{\sin(2\pi(k+1/2)\nu)}{(2k+1)\sin(\pi\nu)} & \text{otherwise.} \end{cases} \tag{25}$$

Example: Dirichlet Kernel



This is a **low-pass filter**: It preserves low frequencies and diminishes high frequencies. It is often used to estimate a monotonic trend component of a series.

Frequency Response: Examples (cont.)

Consider the first difference

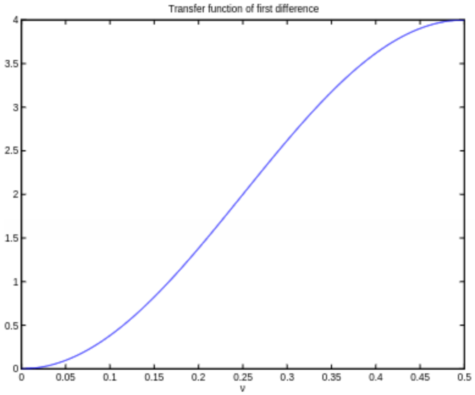
$$Y_t = (1 - B)X_t.$$

This is a time-invariant, causal, linear filter. Its transfer function is

$$\psi(e^{-2\pi i\nu}) = 1 - e^{-2\pi i\nu},$$

so $|\psi(e^{-2\pi i\nu})|^2 = 2(1 - \cos(2\pi\nu))$.

Example: First Difference Kernel



This is a **high-pass filter**: It preserves high frequencies and diminishes low frequencies. It is often used to eliminate a trend component of a series.

Spectral Density of a Linear Process

If X_t is a linear process, it can be written
$X_t = \sum_{i=0}^{\infty} \psi_i W_{t-i} = \psi(B) W_t$.
Then

$$f(\nu) = \sigma_w^2 |\psi(e^{-2\pi i \nu})|^2.$$

That is, the spectral density $S(\nu)$ of a linear process measures the modulus of the ψ (MA(∞)) polynomial at the point $e^{2\pi i \nu}$ on the unit circle.

Spectral Density of a Linear Process

For an ARMA(p, q), $\psi(B) = \frac{\theta(B)}{\phi(B)}$, so

$$\begin{aligned} f(\nu) &= \sigma_w^2 \frac{\theta(e^{-2\pi i\nu})\theta(e^{2\pi i\nu})}{\phi(e^{-2\pi i\nu})\phi(e^{2\pi i\nu})} \\ &= \sigma_w^2 \left| \frac{\theta(e^{-2\pi i\nu})}{\phi(e^{-2\pi i\nu})} \right|^2 . \end{aligned}$$

This is known as a rational spectrum.

Rational Spectra

Consider the factorization of θ and ϕ as

$$\theta(z) = \theta_q(z - z_1)(z - z_2)\cdots(z - z_q)$$
$$\phi(z) = \phi_p(z - p_1)(z - p_2)\cdots(z - p_p),$$

where $z_1, \ldots, z_q$ and $p_1, \ldots, p_p$ are called the zeros and poles.

$$S(\nu) = \sigma_w^2 \left| \frac{\theta_q \prod_{j=1}^{q}(e^{-2\pi i\nu} - z_j)}{\phi_p \prod_{j=1}^{p}(e^{-2\pi i\nu} - p_j)} \right|^2 = \sigma_w^2 \frac{\theta_q^2 \prod_{j=1}^{q} |e^{-2\pi i\nu} - z_j|^2}{\phi_p^2 \prod_{j=1}^{p} |e^{-2\pi i\nu} - p_j|^2}.$$

As ν varies from 0 to 1/2, $e^{-2\pi i\nu}$ moves clockwise around the unit circle from 1 to $e^{-\pi i} = -1$.
And the value of $S(\nu)$ goes up as this point moves closer to (further from) the poles p_j (zeros z_j).

Rational Spectra: Examples

- Recall AR(1): $\phi(z) = 1 - \phi_1 z$. The pole is at $1/\phi_1$.
  - If $\phi_1 > 0$, the pole is to the right of 1, so the spectral density decreases as ν moves away from 0.
  - If $\phi_1 < 0$, the pole is to the left of -1, so the spectral density is at its maximum when $\nu = 0.5$.
- Recall MA(1): $\theta(z) = 1 + \theta_1 z$. The zero is at $-1/\theta_1$.
  - If $\theta_1 > 0$, the zero is to the left of -1, so the spectral density decreases as ν moves towards -1.
  - If $\theta_1 < 0$, the zero is to the right of 1, so the spectral density is at its minimum when $\nu = 0$.

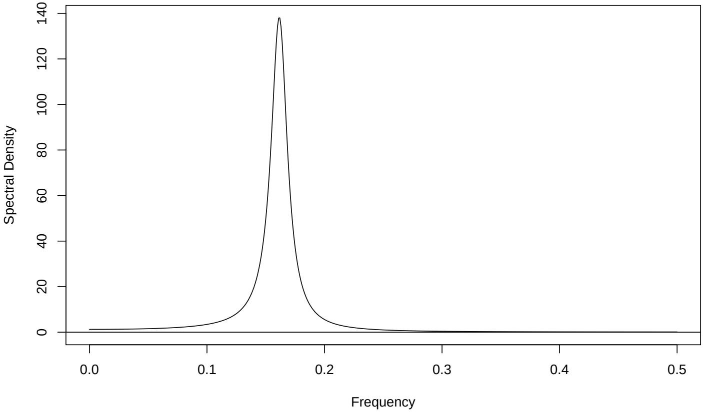
Example: AR(2)

Consider $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + e_t$ with $\phi_1 = 1$, $\phi_2 = -0.9$, and $\sigma_e^2 = 1$. In this case, the poles are at $p_1, p_2 \approx 0.5555 \pm i0.8958 \approx 1.054e^{\pm i1.01567} \approx 1.054e^{\pm 2\pi i 0.16165}$. Thus, we have

$$S(\nu) = \frac{\sigma_w^2}{\phi_2^2 |e^{-2\pi i\nu} - p_1|^2 |e^{-2\pi i\nu} - p_2|^2},$$

and this gets very peaked when $e^{-2\pi i\nu}$ passes near $1.054e^{-2\pi i 0.16165}$.

Example: AR(2) Spectrum



Estimating the Spectrum: Outline

- We have seen that the spectral density gives an alternative view of stationary time series.
- Given a realization $x_1, \ldots, x_n$ of a time series, how can we estimate the spectral density?
- One approach: replace $\gamma(\cdot)$ in the definition

$$S(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h},$$

  with the sample autocovariance $\hat{\gamma}(\cdot)$.
- Another approach, called the periodogram: compute $I(\nu)$, the squared modulus of the discrete Fourier transform (at frequencies $\nu = k/n$).

Estimating the Spectrum: Discussion

- These two approaches are identical at the Fourier frequencies $\nu = k/n$.
- The asymptotic expectation of the periodogram $I(\nu)$ is $S(\nu)$. We can derive some asymptotic properties, and hence do hypothesis testing.
- Unfortunately, the asymptotic variance of $I(\nu)$ is constant. It is not a consistent estimator of $S(\nu)$.
- We can reduce the variance by smoothing the periodogram -averaging over adjacent frequencies-. If we average over a narrower range as $n \to \infty$, we can obtain a consistent estimator of the spectral density.

Estimating the Spectrum: Sample Autocovariance

Idea: use the sample autocovariance $\hat{\gamma}(\cdot)$, defined by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}), \text{ for } -n < h < n,$$

as an estimate of the autocovariance $\gamma(\cdot)$, and then use a sample version of

$$S(\nu) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \nu h},$$

That is, for $-1/2 \leq \nu \leq 1/2$, estimate $f(\nu)$ with

$$\hat{S}(\nu) = \sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-2\pi i \nu h}.$$

Estimating the Spectrum: Periodogram

Discrete Fourier Transform: For a sequence $(x_1, \ldots, x_n)$, define the discrete Fourier transform (DFT) as $(X(\nu_0), X(\nu_1), \ldots, X(\nu_{n-1}))$, where

$$X(\nu_k) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi i \nu_k t},$$

and $\nu_k = k/n$ (for $k = 0, 1, \ldots, n-1$) are called the Fourier frequencies.

First, let's show that we can view the DFT as a representation of x in a different basis, the Fourier basis.

Periodogram

The periodogram is defined as

$$I(\nu_j) = |X(\nu_j)|^2 = \frac{1}{n} \left| \sum_{t=1}^{n} e^{-2\pi i t \nu_j} x_t \right|^2 = X_c^2(\nu_j) + X_s^2(\nu_j),$$

where

$$X_c(\nu_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \cos(2\pi t \nu_j) x_t,$$

$$X_s(\nu_j) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sin(2\pi t \nu_j) x_t.$$

Periodogram

Since $I(\nu_j) = |X(\nu_j)|^2$ for one of the Fourier frequencies $\nu_j = j/n$ (for $j = 0, 1, \ldots, n-1$), the orthonormality of the e_j implies that we can write

$$x^* x = \left(\sum_{j=0}^{n-1} X(\nu_j) e_j \right)^* \left(\sum_{j=0}^{n-1} X(\nu_j) e_j \right) = \sum_{j=0}^{n-1} |X(\nu_j)|^2 = \sum_{j=0}^{n-1} I(\nu_j).$$

For $\bar{x} = 0$, we can write this as

$$\hat{\sigma}_x^2 = \frac{1}{n} \sum_{t=1}^{n} x_t^2 = \frac{1}{n} \sum_{j=0}^{n-1} I(\nu_j).$$

Estimating the Spectrum: Periodogram

Why is the periodogram at a Fourier frequency (that is, $\nu = \nu_j$) the same as computing $S(\nu)$ from the sample autocovariance? Almost the same—they are not the same at $\nu_0 = 0$ when $\bar{x} \neq 0$. But if either $\bar{x} = 0$, or we consider a Fourier frequency ν_j with $j \in \{1, \ldots, n-1\}$, ...

Estimating the Spectrum: Periodogram

$$
\begin{aligned}
I(\nu_j) &= \frac{1}{n}\left|\sum_{t=1}^{n} e^{-2\pi i t\nu_j} x_t\right|^2 \\
&= \frac{1}{n}\left|\sum_{t=1}^{n} e^{-2\pi i t\nu_j}(x_t-\bar{x})\right|^2 \\
&= \frac{1}{n}\left(\sum_{t=1}^{n} e^{-2\pi i t\nu_j}(x_t-\bar{x})\right)\left(\sum_{t=1}^{n} e^{2\pi i t\nu_j}(x_t-\bar{x})\right) \\
&= \frac{1}{n}\sum_{s,t} e^{-2\pi i (s-t)\nu_j}(x_t-\bar{x})(x_s-\bar{x}) \\
&== \frac{1}{n}\sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-2\pi i h\nu_j}
\end{aligned}
$$

Asymptotic Properties of the Periodogram

Gaussian White Noise

If $X_1, \ldots, X_n$ are i.i.d. $N(0, \sigma^2)$ (Gaussian white noise; $f(\nu) = \sigma^2$), then the $X_c(\nu_j)$ and $X_s(\nu_j)$ are all i.i.d. $N(0, \sigma^2/2)$. Thus,

$$\frac{2}{\sigma^2} I(\nu_j) = \frac{2}{\sigma^2}(X_c^2(\nu_j) + X_s^2(\nu_j)) \sim \chi_2^2 \tag{26}$$

So for the case of Gaussian white noise, the periodogram has a chi-squared distribution that depends on the variance σ^2 (which, in this case, is the spectral density).

General Asymptotic Properties

Under more general conditions (e.g., normal $\{X_t\}$, or linear process $\{X_t\}$ with rapidly decaying ACF), the $X_c(\nu_j)$, $X_s(\nu_j)$ are all asymptotically independent and $N(0, S(\nu_j)/2)$.

Consider a frequency ν. For a given value of n, let $\hat{\nu}(n)$ be the closest Fourier frequency (that is, $\hat{\nu}(n) = j/n$ for a value of j that minimizes $|\nu - j/n|$). As n increases, $\hat{\nu}(n) \to \nu$, and (under the same conditions that ensure the asymptotic normality and independence of the sine/cosine transforms), $S(\hat{\nu}(n)) \to S(\nu)$.

In that case, we have:

$$\frac{2}{S(\nu)} I(\hat{\nu}(n)) = \frac{2}{S(\nu)} \left(X_c^2(\hat{\nu}(n)) + X_s^2(\hat{\nu}(n))\right) \xrightarrow{d} \chi_2^2 \qquad (27)$$

Parametric Spectral Estimation

In parametric spectral estimation, we consider the class of spectral densities corresponding to ARMA models.

Recall that, for a linear process $Y_t = \psi(B)W_t$, the spectral density is given by:

$$f_y(\nu) = \left|\psi(e^{2\pi i\nu})\right|^2 \sigma_w^2.$$

For an AR model, $\psi(B) = 1/\phi(B)$, so Y_t has the rational spectrum:

$$f_y(\nu) = \frac{\sigma_w^2}{\left|\phi(e^{-2\pi i\nu})\right|^2} = \frac{\sigma_w^2}{\phi^2} \prod_{j=1}^{p} \left|e^{-2\pi i\nu} - p_j\right|^2,$$

where p_j are the poles or roots of the polynomial ϕ.

Typical Approach to Parametric Spectral Estimation

The typical approach to parametric spectral estimation is to use the maximum likelihood parameter estimates $(\hat{\phi}_1, \ldots, \hat{\phi}_p, \hat{\sigma}_w^2)$ for the parameters of an AR(p) model for the process, and then compute the spectral density for this estimated AR model:

$$\hat{f}_y(\nu) = \hat{\sigma}_w^2 / \left| \hat{\phi}(e^{-2\pi i \nu}) \right|^2 .$$