STAT 5170: Applied Time Series

Course notes for part A of learning unit 6

In this learning unit, we shift our attention from the *time domain* perspective of time series analysis to the *frequency domain* perspective. Our study of this perspective has, in fact, already begun; its starting point was our discussion of seasonal regression variables and the scaled periodogram, which we discussed in the previous learning unit. The frequency domain perspective pushes the mathematical ideas that underly seasonal variables quite far, to the point of offering an entirely new way to conceive of a time-series model, and to think about time series analysis.

Section 6A.1: Spectral analysis.

We start our development of frequency domain perspective by revisiting the trigonometric setup that gives rise to seasonal variables. When doing so we make a substantial conceptual change in how we regard the variables, not as components of a deterministic trend, as they are treated in regression, but as random functions, which ultimately define a stationary time series. This change in our conception of seasonal variables from deterministic trend to stationary randomness is one of the hallmarks of the frequency domain perspective.

Similar to what we did before, let us consider a simple trigonometric function, but this time use it to define a time series (x_t) according to

$$x_t \quad = \quad A\cos(2\pi\omega t + \phi).$$

For this time series to be random in some sense, some variable to the right of the above equality would be a random variable. For present purposes, it is the amplitude, A, and phase, ϕ, that would be random, while the frequency, ω, is to remain fixed. Now recall that application of a trigonometric identity yields the equivalent formula

$$x_t \quad = \quad U_1\cos(2\pi\omega t) + U_2\sin(2\pi\omega t),$$

where

$$U_1 = A\cos(\phi) \quad \text{and} \quad U_2 = -A\sin(\phi).$$

This formula defines the time series as a linear function of two term, whose coefficients are random. The familiarity of this setup suggests that the variables U_1 and U_2 would be specified as independent, normally-distributed random variables.

Observe also that an inversion of the above transformation provides that

$$A = \sqrt{U_1^2 + U_2^2} \quad \text{and} \quad \phi = \tan^{-1}(-U_2/U_1).$$

These formulas would specify the distributions of A and ϕ, from those of U_1 and U_2.

Observe also a peculiar feature of this simple trigonometric time series, which is that conditioning on U_1 and U_2 makes (x_t) deterministic. In other words, if values for U_1 and U_2 were known, then forecasting any future value could be done with exact precision.

In contrast, consider a simple version of the model that would be used in the regression setup for seasonal variables. This would be understood through the decomposition $x_t = \mu_t + y_t$, a "signal-plus-noise" model, wherein μ_t is constant and y_t is random. Defining μ_t in terms of a simple trigonometric function yields

$$x_t \;=\; \beta_1 \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t) + y_t.$$

In this setup, knowing the linear coefficients β_1 and β_2 would define the mean function μ_t at all time points, t, but it would not allow forecasting with exact precision since the corresponding value x_t is a shift of μ_t by random noise, y_t.

Two interesting properties of the relationships discussed above are as follows:

Properties: Suppose the time series (x_t) is defined according to $x_t = U_1 \cos(2\pi\omega t) + U_2 \sin(2\pi\omega t)$, for random variables $U_1 = A\cos(\phi)$ and $U_2 = -A\sin(\phi)$. It follows that...

(i.) If U_1 and U_2 are independent, standard normal random variables ($i.e.$, $U_i \sim N(0,1)$), then $A^2 \sim \chi^2_2$ and, independently, $\phi \sim \text{unif}(-\pi, \pi)$. The converse of this statement also holds.

(ii.) If U_1 and U_2 are uncorrelated, mean-zero random variables, each with variance σ^2 ($i.e.$, $E[U_i] = 0$ and $Var[U_i] = \sigma^2$), then x_t is stationary with $E[x_t] = 0$ and $\gamma(h) = \sigma^2 \cos(2\pi\omega h)$.

□

Proofs of these properties are as follows.

Proof: Property (i) is an exercise in applying the two-dimensional transformation theorem.

Property (ii) is established as follows. Write $c_t = \cos(2\pi\omega t)$ and $s_t = \sin(2\pi\omega t)$. Then,

$$E(x_t) \;=\; E(U_1 c_t + U_2 s_t) = c_t E(U_1) + s_t E(U_2) = 0$$

and

$$\begin{aligned}
\gamma(h) &= Cov(x_{t+h}, x_t) \\
&= Cov(U_1 c_{t+h} + U_2 s_{t+h}, U_1 c_t + U_2 s_t) \\
&= Cov(U_1 c_{t+h}, U_1 c_t) + Cov(U_1 c_{t+h}, U_2 s_t) \\
&\quad + Cov(U_2 s_{t+h}, U_1 c_t) + Cov(U_2 s_{t+h}, U_2 s_t) \\
&= c_{t+h} c_t Cov(U_1, U_1) + c_{t+h} s_t Cov(U_1, U_2) \\
&\quad + s_{t+h} c_t Cov(U_2, U_1) + s_{t+h} s_t Cov(U_2, U_2) \\
&= \sigma^2 c_{t+h} c_t + 0 + 0 + \sigma^2 s_{t+h} s_t
\end{aligned}$$

The identity

$$\cos(\alpha - \beta) \;=\; \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)$$

provides that

$$
\begin{aligned}
\gamma(h) &= \sigma^2(c_{t+h}c_t + s_{t+h}s_t) \\
&= \sigma^2\{\cos(2\pi\omega(t+h))\cos(2\pi\omega t) + \sin(2\pi\omega(t+h))\sin(2\pi\omega t)\} \\
&= \sigma^2\cos(2\pi\omega(t+h) - 2\pi\omega t) \\
&= \sigma^2\cos(2\pi\omega h).
\end{aligned}
$$

$\square$

Property (*ii*) of the above discussion readily extends to establish the following property of a time series that is defined by aggregating simple trigonometric time series.

Property: Suppose

$$x_t = \sum_{k=1}^{q}\{U_{k1}\cos(2\pi\omega_k t) + U_{k2}\sin(2\pi\omega_k t)\},$$

for distinct ω_k, where U_{ki} are independent with $E(U_{ki}) = 0$ and $Var(U_{ki}) = \sigma_k^2$. Then $E(x_t) = 0$ and

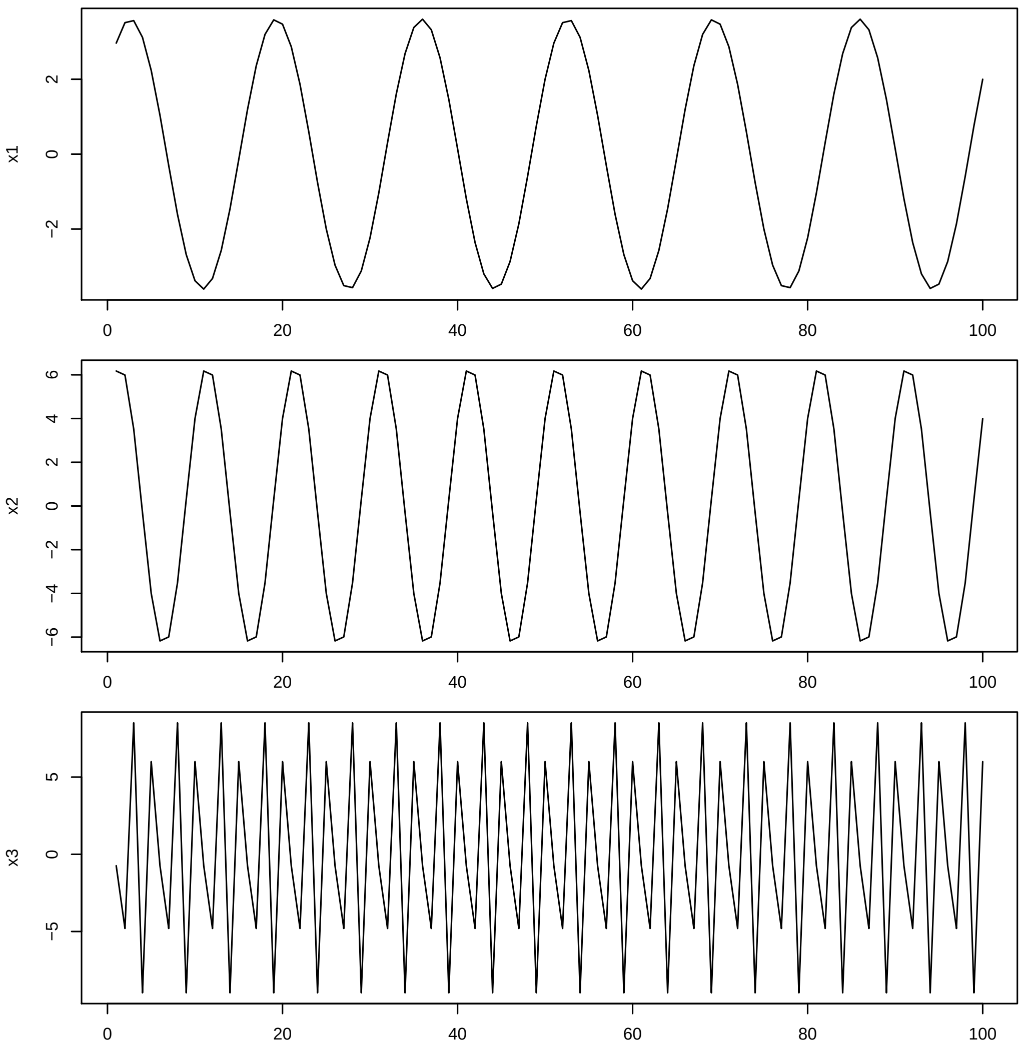
$$\gamma(h) = \sum_{k=1}^{q}\sigma_k^2\cos(2\pi\omega_k h).$$

$\square$

The concept of defining a time series by aggregating simple trigonometric time series is illustrated as follows.

Example: Aggregating simple trigonometric time series

The figure below plots three time series, (x_{1t}), (x_{2t}), and (x_{3t}), each a simple periodic function with the same frequency, across the time points $t = 1, \ldots, 100$.



The time series are defined by the following formulas:

$$
\begin{aligned}
x_{1t} &= 2\cos(2\pi \times 0.06t) + 3\sin(2\pi \times 0.06t) \\
x_{2t} &= 4\cos(2\pi \times 0.10t) + 5\sin(2\pi \times 0.10t) \\
x_{3t} &= 6\cos(2\pi \times 0.40t) + 7\sin(2\pi \times 0.40t)
\end{aligned}
$$

That is, each is a version of the formula

$$x_{kt} = U_{k1}\cos(2\pi\omega_k t) + U_{k2}\sin(2\pi\omega_k t),$$

for which

$$
\begin{array}{lll}
\omega_1 = 0.06, & U_{11} = 2, & U_{12} = 3, \\
\omega_2 = 0.10, & U_{21} = 4, & U_{22} = 5, \\
\omega_3 = 0.40, & U_{31} = 6, & \text{and } U_{32} = 7.
\end{array}
$$

With these values, one may, for example, calculate the amplitude of each series: The formula is

$$A_k = \sqrt{U_{k1}^2 + U_{k2}^2}$$

which evaluates to

$$\begin{aligned} A_1 &= \sqrt{U_{11}^2 + U_{12}^2} = \sqrt{2^2 + 3^2} = 3.6056 \\ A_2 &= \sqrt{U_{21}^2 + U_{22}^2} = \sqrt{4^2 + 5^2} = 6.4031 \\ A_3 &= \sqrt{U_{31}^2 + U_{32}^2} = \sqrt{6^2 + 7^2} = 9.2195. \end{aligned}$$

Observe on the horizontal axes how the amplitudes are reflected in each plot, as the maximum magnitude of the plotted values.

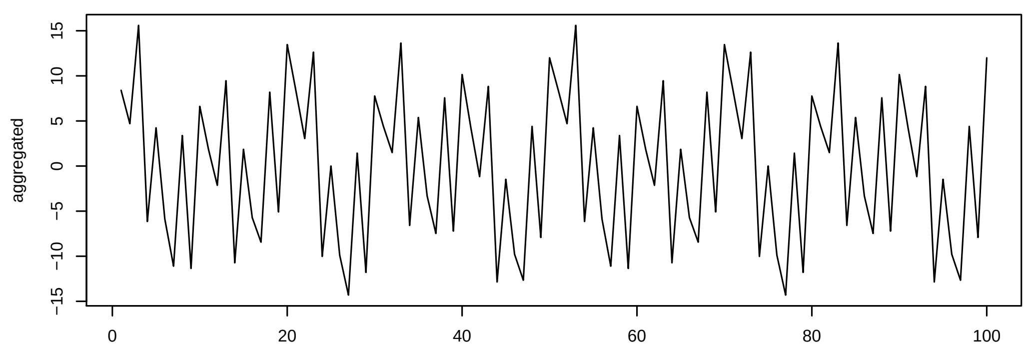
The aggregate of these times series is

$$x_t = x_{1t} + x_{2t} + x_{3t},$$

or, expressed as an aggregate of sinusoids,

$$x_t = \sum_{k=1}^{3} \{U_{k1} \cos(2\pi\omega_k t) + U_{k2} \sin(2\pi\omega_k t)\}.$$

A plot of the aggregated series is as follows.



Note how this example illustrates that it is possible to create complex patterns from aggregating simple periodic functions. We can start to see behavior in this time series that resembles patterns seen in time series generated from other models and processes. □

The previous discussion and example offers a stepping stone to the frequency domain approach to time-series analysis. In terms of modeling, the basic premise of this perspective is that any stationary time series may be understood as an aggregate of simple trigonometric time series, possibly a great many of them. In what follows, we define a mathematical concept called the *spectral density*, which keeps track of the strengths at which various trigonometric time series are emphasized in the aggregation. That is, it keeps track of the strengths at which the range frequencies are expressed in the patterns of the time series.

Definition: The *spectral density* of a stationary time series (x_t) is

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

for $-1/2 \leq \omega \leq 1/2$. $\square$

Note that the spectral density is defined from the autocovariance function of a time series. This is interesting in part because it makes a connection between the time domain and frequency domain perspectives of time series analysis. In the time domain, the focus of attention is on the autocovariance function, which becomes the mathematical basis of building $ARMA(p, q)$ models. In the frequency domain, the spectral density is the focus of mathematical attention, and any stationary time series is modeled as an aggregate of simple trigonometric time series. The spectral density is used to understand the components of such an aggregation.

Note also that the spectral density is defined using *complex numbers*, specifically complex exponentiation. The following illustration offers a review of such concepts.

<u>Illustration:</u> Properties of complex exponentiation A complex number is $z = a + ib$, where $i = \sqrt{-1}$ is the imaginary constant. Euler's formula defines the concept of complex exponentiation, and suggests intuition for complex numbers in terms of angles and magnitudes within a polar coordinate system. Euler's formula is

$$e^{i\alpha} = \cos\alpha + i \sin\alpha.$$

Observe, for instance, that the identities $\cos 0 = 1$ and $\sin 0 = 0$ imply

$$e^{0} = e^{i0} = \cos 0 + i \sin 0 = 1,$$

as would be expected from exponential notation.

The properties

$$\cos(-\alpha) = \cos\alpha \quad \text{and} \quad \sin(-\alpha) = -\sin\alpha$$

characterize cosine as an *even*, or *symmetric*, function, and sine as an *odd* function. From these properties, it is straightforward to deduce that

$$e^{i\alpha} + e^{-i\alpha} = 2\cos\alpha \quad \text{and} \quad e^{i\alpha} - e^{-i\alpha} = 2i \sin\alpha.$$

The magnitude of a complex number $z = a + ib$ is

$$|z| \;\; = \;\; \sqrt{a^2 + b^2}.$$

This may be computed through

$$|z|^2 \;\; = \;\; (a + ib)(a - ib) = a^2 - abi + abi + b^2 = a^2 + b^2.$$

If $z = \beta e^{i\alpha}$, a scaled exponential, its squared-magnitude is

$$|z|^2 \;\; = \;\; |\beta\cos\alpha + i\beta\sin\alpha|^2 = \beta^2\{(\cos\alpha)^2 + (\sin\alpha)^2\} = \beta^2,$$

which makes use of the trignometric identity $(\cos\alpha)^2 + (\sin\alpha)^2 = 1$.

If $z = 1 + \beta_1 e^{i\alpha} + \beta_2 e^{2i\alpha} + \cdots + \beta_p e^{pi\alpha}$, a linear combination of exponentials, a formula for its squared-magnitude is

$$|z|^2 \;=\; 1 + \sum_{j=1}^{p} \beta_j^2 + 2\sum_{j=1}^{p-1}\left(\beta_j + \sum_{k=1}^{p-j}\beta_k\beta_{j+k}\right)\cos(j\alpha) + 2\beta_p\cos(p\alpha)$$

assuming $p \geq 2$. To see this, first observe that

$$\begin{aligned}
\left|1+\sum_{j=1}^{p}\beta_j e^{ij\alpha}\right|^2 &= \left|1+\sum_{j=1}^{p}\beta_j\cos(j\alpha) + i\sum_{j=1}^{p}\beta_j\sin(j\alpha)\right|^2 \\
&= \left(1+\sum_{j=1}^{p}\beta_j\cos(j\alpha)+i\sum_{j=1}^{p}\beta_j\sin(j\alpha)\right)\left(1+\sum_{j=1}^{p}\beta_j\cos(j\alpha)-i\sum_{j=1}^{p}\beta_j\sin(j\alpha)\right) \\
&= \left(1+\sum_{j=1}^{p}\beta_j e^{ij\alpha}\right)\left(1+\sum_{j=1}^{p}\beta_j e^{-ij\alpha}\right).
\end{aligned}$$

(This property is itself notable and will be used later.) Then expand the product and apply a triangle argument according to

$$\begin{aligned}
\left|1+\sum_{j=1}^{p}\beta_j e^{ij\alpha}\right|^2 &= 1+\sum_{j=1}^{p}\beta_j(e^{ij\alpha}+e^{-ij\alpha}) + \sum_{j=1}^{p}\sum_{k=1}^{p}\beta_j\beta_k e^{-i(j-k)\alpha} \\
&= 1+\sum_{j=1}^{p}\beta_j^2 + \sum_{j=1}^{p}\beta_j(e^{ij\alpha}+e^{-ij\alpha}) + \sum_{j=1}^{p-1}\sum_{k=1}^{p-j}\beta_k\beta_{j+k}(e^{ij\alpha}+e^{-ij\alpha}).
\end{aligned}$$

□

Having reviewed complex exponentiation and some of its properties, several properties the spectral density are accessible to us. They are stated as follows without proof:

Properties:

(i.) When the spectral density exists, the autocovariance function may be recovered from the spectral density according to the formula

$$\gamma(h) \;=\; \int_{-1/2}^{1/2} e^{2\pi i\omega h} f(\omega)d\omega$$

In particular, $\gamma(0) = \int_{-1/2}^{1/2} f(\omega)d\omega$.

(ii.) The spectral density exists whenever

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

(*iii.*) The spectral density is always non-negative, $f(\omega) \geq 0$.

(*iv.*) The spectral density is symmetric, $f(\omega) = f(-\omega)$. This suggests that exploration of the spectral density need only focus on the region $0 \leq \omega \leq 1/2$.

(*v.*) If (x_t) is $ARMA(p, q)$ such that $\phi(B)x_t = \theta(B)w_t$ and $Var(w_t) = \sigma_w^2$, then

$$f(\omega) = \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2}\sigma_w^2.$$

This is equivalently stated

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

where $\psi(z) = \theta(z)/\phi(z)$.

□

The sequence of examples and illustrations below connect ideas about the spectral density to times series models that we have already seen.

Example: Spectral density of white noise

Suppose the time series (x_t) is white noise with variance $\sigma^2 = Var(x_t)$. Its autocovariance function is

$$\gamma(h) = \begin{cases} \sigma^2 & \text{if } h = 0 \\ 0 & \text{otherwise} \end{cases}$$

With this, the spectral density is quickly derived as

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)e^{-2\pi i\omega h} = \sigma^2 e^0 = \sigma^2$$

The form of the spectral density is uniform across the entire range of frequencies, $-1/2 \leq \omega \leq 1/2$. In term of the distribution of simple periodic components, this indicates that all frequencies are represented evenly. The term "white" in "white noise" refers to this property, an even mix of all frequencies. □

Example: Spectral density of $MA(1)$

Suppose the time series (x_t) is $MA(1)$, so that

$$x_t = w_t + \theta_1 w_{t-1}$$

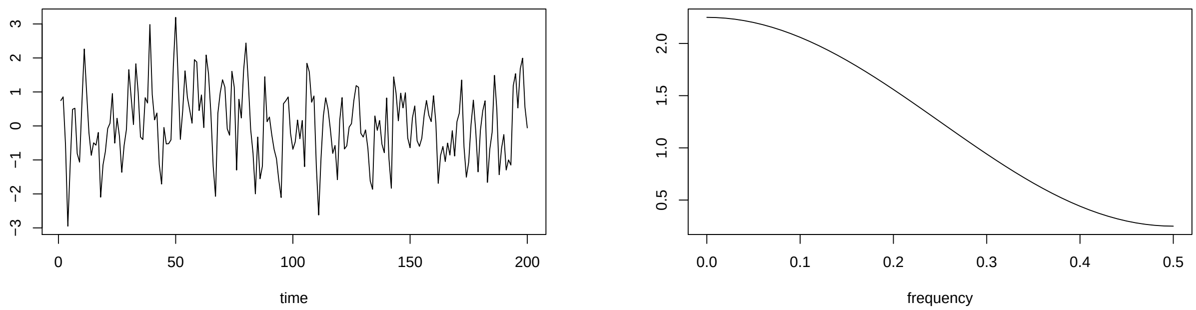
for the moving-average parameter θ_1 and white noise variance $\sigma_w^2 = Var(w_t)$. The corresponding autocovariance function is

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

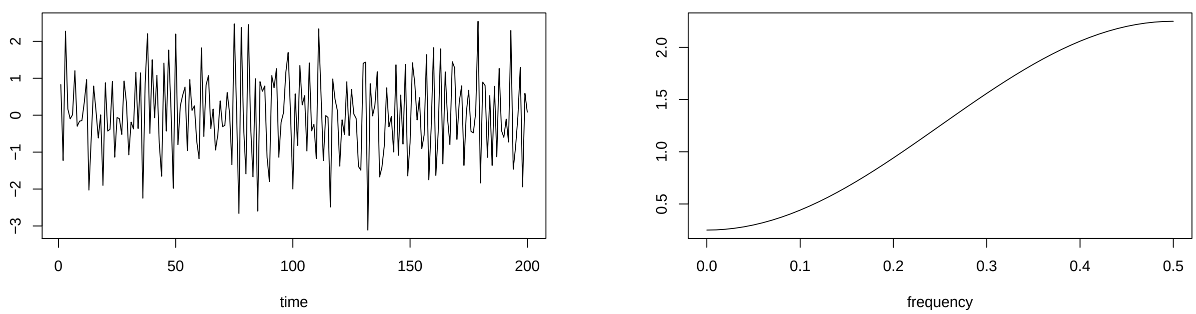
Its spectral density is derived as

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

Setting $\theta_1 = 0.5$, the spectral density under this model is displayed below, to the right of a time series plot of $n = 200$ simulated measurements.



Having now set the $\theta_1 = -0.5$, parallel figure appears below.



Observe how, under the setting $\theta_1 = 0.5$, the spectral density indicates a prevalence of low frequency oscillations, and how these are appear more strongly in the simulated measurements. $\square$

Example: Spectral density of $AR(2)$

Suppose the time series (x_t) is $AR(2)$, according to

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + w_t$$

for autoregressive parameters ϕ_1 and ϕ_2 and white noise variance $\sigma_w^2 = Var(w_t)$. The autoregressive relationship is written in operator notation according to

$$\phi(B)x_t = w_t$$

for the autoregressive polynomial

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2.$$

The spectral density is then provided by the formula for that of an ARMA time series, which reduces to

$$f(\omega) \;=\; \frac{1}{|\phi(e^{-2\pi i\omega})|^2}\sigma_w^2.$$

A previous formula provides that the squared-magnitude of the form of linear combination that appears in the denominator is

$$\begin{aligned}
|\phi(e^{-2\pi i\omega})|^2 &= |1-\phi_1 e^{-2\pi i\omega} - \phi_2 e^{-4\pi i\omega}|^2 \\
&= 1+\phi_1^2+\phi_2^2+2(-\phi_1+\phi_1\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega) \\
&= 1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega).
\end{aligned}$$

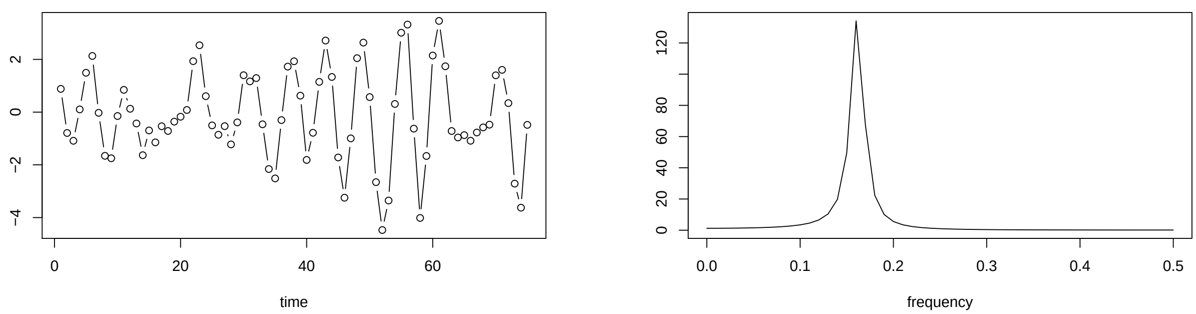
Alternatively, one might derive this directly according to

$$\begin{aligned}
|\phi(e^{-2\pi i\omega})|^2 &= (1-\phi_1 e^{-2\pi i\omega} - \phi_2 e^{-4\pi i\omega})(1-\phi_1 e^{2\pi i\omega} - \phi_2 e^{4\pi i\omega}) \\
&= 1+\phi_1^2+\phi_2^2-\phi_1(1-\phi_2)(e^{2\pi i\omega}+e^{-2\pi i\omega}) - \phi_2(e^{4\pi i\omega}+e^{-4\pi i\omega}). \\
&= 1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega).
\end{aligned}$$

The spectral density is therefore

$$f(\omega) \;=\; \frac{\sigma_w^2}{1+\phi_1^2+\phi_2^2-2\phi_1(1-\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega)}.$$

Setting $\phi_1 = 1$ and $\phi_2 = -0.9$, the spectral density under this model is displayed below, to the right of a time series plot of $n = 200$ simulated measurements.



Here, the spectral density indicates a prevalence of strong periodicities in a narrow range of frequencies around $\omega = 0.16$, which corresponds to a frequency of about $1/\omega = 6.25$. $\square$

<u>Example:</u> Roots and poles of a $ARMA(p,q)$ spectral density

For an $ARMA(p,q)$ time series, the spectral density is

$$f(\omega) \;=\; \sigma_w^2\frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2}.$$

By knowing the roots $z_1,\ldots,z_q$ of $\theta(z)$ and $p_1,\ldots,p_p$ of $\phi(z)$, it is possible to factor these polynomials according to

$$\begin{aligned}
\theta(z) &= \theta_q(z-z_1)\times\cdots\times(z-z_q) = \theta_q\prod_{j=1}^{q}(z-z_j) \\
\phi(z) &= \phi_p(z-p_1)\times\cdots\times(z-p_p) = \phi_p\prod_{j=1}^{p}(z-p_j)
\end{aligned}$$

hence write

$$f(\omega) \quad = \quad \sigma_w^2 \frac{\theta_q^2 \prod_{j=1}^q |e^{-2\pi i\omega} - z_j|^2}{\phi_p^2 \prod_{j=1}^p |e^{-2\pi i\omega} - p_j|^2}.$$

In this context, the roots $p_1, \ldots, p_p$ of $\phi(z)$ are called the "poles" of the spectral density, because $f(\omega)$ becomes large whenever ω is such that $e^{-2\pi i\omega}$ is near some p_j.

For instance, a previous example examined the spectral density of an $AR(2)$ time series such that

$$x_t \quad = \quad x_{t-1} - 0.9x_{t-2} + w_t.$$

Here, the autoregressive polynomial is $\phi(z) = 1 - z + 0.9z^2$. By the quadratic formula, its roots (*i.e.*, poles) are

$$\frac{1 \pm \sqrt{(-1)^2 - 4(1)(0.9)}}{2(0.9)} = 0.5556 \pm i0.8958 = 1.0541e^{\pm 2\pi i \times 0.1617}.$$

This explains the peak in $f(\omega)$ at $\omega = 0.1617$, for it is at that frequency that $e^{-2\pi i\omega}$ is near one of the poles. □

This next example introduces a concept that is sometimes useful for mathematical study of spectral densities.

Example: Autocovariance generating functions

Suppose (x_t) is a stationary process with autocovariance function $\gamma(h)$. Its *autocovariance generating function* is

$$G(z) \quad = \quad \sum_{h=-\infty}^{\infty} \gamma(h) z^h,$$

provided there is some $r > 0$ such that $G(z)$ for all z such that $1/r < |z| < r$, for $r > 0$.

The autocovariance generating function, $G(z)$, is sometimes very easy to derive. When it may be written down explicitly, the autocovariance function $\gamma(h)$ may be read off from the coefficients associated with z^h or z^{-h}.

Suppose the (x_t) is a linear process, such that

$$x_t \quad = \quad \sum_{j=-\infty}^{\infty} \psi_j w_{t-j},$$

where (w_t) is white noise with $\sigma_w^2 = Var(w_t)$. We know that

$$\gamma(h) \quad = \quad \sigma_w^2 \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h}.$$

The autocovariance generating function is

$$
\begin{aligned}
G(z) &= \sum_{h=-\infty}^{\infty} \gamma(h) z^h \\
&= \sigma_w^2 \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} z^h, \\
&= \sigma_w^2 \left\{ \sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{h=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+h} (z^h + z^{-h}) \right\} \\
&= \sigma_w^2 \left(\sum_{j=-\infty}^{\infty} \psi_j z^j \right) \left(\sum_{k=-\infty}^{\infty} \psi_k z^{-k} \right),
\end{aligned}
$$

in which the last step follows from a triangle argument.

By defining the infinite polynomial

$$
\psi(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j,
$$

the autocovariance generating function may be written more simply as

$$
G(z) = \sigma_w^2 \psi(z) \psi(z^{-1})
$$

in which the last step follows from a triangle argument.

This extends to any causal $ARMA(p,q)$ time series, (x_t), for which $\phi(B)x_t = \theta(B)w_t$. Such a time series has the causal representation $x_t = \psi(B)w_t$, defined by

$$
\psi(z) = \frac{\theta(z)}{\phi(z)},
$$

expressed as an infinite moving average polynomial

$$
\psi(z) = \phi_0 + \phi_1 z + \phi_2 z^2 + \cdots,
$$

where $\phi_0 = 1$. The autocovariance generating function is therefore

$$
G(z) = \sigma_w^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}.
$$

For example, if (x_t) is $MA(2)$, then $\theta(z) = 1 + \theta_1 z + \theta_2 z^2$ and

$$
\begin{aligned}
G(z) &= \sigma_w^2 \theta(z)\theta(z^{-1}) = (1 + \theta_1 z + \theta_2 z^2)(1 + \theta_1 z^{-1} + \theta_2 z^{-2}) \\
&= \sigma_w^2 \{(1 + \theta_1^2 + \theta_2^2) + (\theta_1 + \theta_1\theta_2)(z + z^{-1}) + \theta_2(z^2 + z^{-2})\}.
\end{aligned}
$$

Reading off the coefficients establishes that $\gamma(0) = \sigma_w^2(1 + \theta_1^2 + \theta_2^2)$, $\gamma(\pm 1) = \sigma_w^2(\theta_1 + \theta_1\theta_2)$, $\gamma(\pm 2) = \sigma_w^2\theta_2$, and $\gamma(\pm h) = 0$ for $h > 2$.

The spectral density is defined from the autocovariance generating function according to

$$f(\omega) \;=\; \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h} = G(e^{-2\pi i\omega}).$$

Thus, for an $ARMA(p,q)$ time series, the spectral density is

$$f(\omega) \;=\; \sigma_w^2 \frac{\theta(e^{-2\pi i\omega})\theta(e^{2\pi i\omega})}{\phi(e^{-2\pi i\omega})\phi(e^{2\pi i\omega})} = \sigma_w^2 \frac{|\theta(e^{-2\pi i\omega})|^2}{|\phi(e^{-2\pi i\omega})|^2},$$

which establishes a property stated in previous discussion. □

Section 6A.2: Periodograms.

The spectral density is a concept associated with a time series model, and would not be known when only measured data from a time series (x_t) is available. The following discussion develops a statistic called a *periodogram*, which serves an analogous role as the spectral density density in a time series model and may be regarded as an estimate of it. It is derived from similar ideas as those that lead our discussion of seasonal variables in regression to consider the scaled periodogram. Here, these ideas are discussed in the concept of a time series of aggregated simple trigonometric time series.

To start, let us suppose that $x_1, \ldots, x_n$ are the first n measurements of a stationary time series (x_t) that is defined as follows. The following property (stated without proof) indicates how these data may be understood *via* simple trigonometric time series.

Property: There are coefficients U_{k1}, and U_{k2} such that, if n is odd, then $x_1, \ldots, x_n$ has the representation

$$x_t \;=\; \mu + \sum_{k=1}^{(n-1)/2} \{U_{k1}\cos(2\pi kt/n) + U_{k2}\sin(2\pi kt/n)\};$$

and, if n is even,

$$x_t \;=\; \mu + \sum_{k=1}^{n/2-1} \{U_{k1}\cos(2\pi kt/n) + U_{k2}\sin(2\pi kt/n)\} + U_{n/2}\cos(\pi t).$$

□

This property defines $x_1, \ldots, x_n$ as an aggregate of simple periodic components

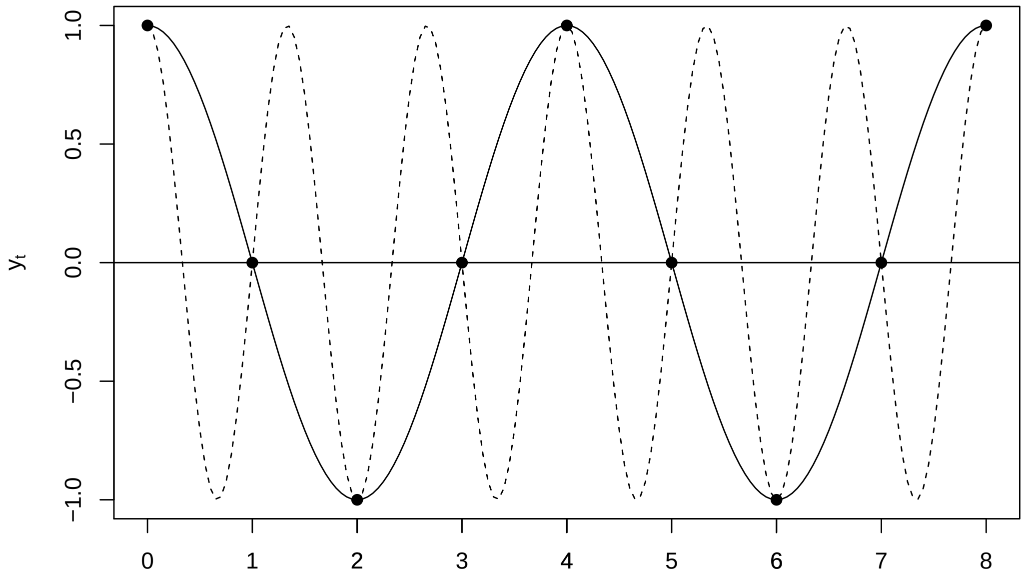
$$x_{kt} \;=\; U_{k1}\cos(2\pi kt/n) + U_{k2}\sin(2\pi kt/n).$$

The coefficients U_{k1}, and U_{k2} are called *Fourier coefficients*. The frequency associated with the k'th simple periodic component is $\omega_k = k/n$, which is defined for $k = 1, \ldots, (n-1)/2$ if n is odd and $k = 1, \ldots, n/2-1$ if n is even; it is called a *fundamental Fourier frequency*.

Fundamental Fourier frequencies are defined with reference to a measured data sequence, $x_1, \ldots, x_n$, not from an infinite or double-infinite, hypothetical time series (x_t). The following illustration explores a peculiar property associated with fundamental Fourier frequencies that reflects the situation of finite data.

<u>Illustration:</u> Aliasing The coarseness of discrete time measurements prevents high-order frequencies from being expressed in a finite data set of measurements at equally spaced time points, $t = 1, \ldots, n$.

To illustrate consider the two time series, overlaid in the plot below, that are defined from simple periodic functions and measured at the time points $t = 1, \ldots, 8$.



The corresponding formulas are

$$x_t = 2\cos(2\pi t \times 0.25) \quad \text{and} \quad y_t = 2\cos(2\pi t \times 0.75).$$

Notice in particular that the frequency associated with (x_t) is $\omega_x = 0.25$ and that associated with (y_t) is $\omega_y = 0.75$.

Examining the plot closely, one sees that the measurements collected at the discrete points $t = 1, \ldots, 8$ are identical between the two series, despite that the two functions that define them are vastly different when viewed across the continuous interval such that $0 < t \leq 8$.

In this way, the discrete measurements collected on a higher-frequency function appear as though they might have been generated from a lower-frequency periodic component. It is said that the higher-frequency, $\omega_y = 0.75$, is an "alias" of the lower-frequency $\omega_x = 0.25$.

For time series measured at discrete time points, any frequency $\tilde{\omega}$ with $\tilde{\omega} \geq 0.5$ is an alias of some frequency ω with $\omega < 0.5$, and will be expressed in the data at the lower frequency. There is no need, therefore, to examine frequencies higher than 0.5.

The frequency value demarcating these ranges, 0.5, is sometimes called the *folding frequency* for discrete sampling. □

In the frequency domain perspective, the time series model corresponding to representations as aggregated periodic components would treat the Fourier coefficients U_{k1}, and U_{k2} as independent, mean zero (*i.e.*, $E[U_{ki}] = 0$) random variables. Their associated variances are denoted $\sigma_k^2 = Var[U_{ki}]$. Under these assumptions, the expectations and variances of each simple periodic component $x_{k1}, \ldots, x_{kn}$ time series are readily deduced:

$$E[x_{kt}] = \cos(2\pi kt/n)E[U_{k1}] + \sin(2\pi kt/n)E[U_{k2}] = 0$$

and

$$\begin{aligned} Var[x_{kt}] &= \{\cos(2\pi kt/n)\}^2 Var[U_{k1}] + \{\sin(2\pi kt/n)\}^2 Var[U_{k2}] \\ &= [\{\cos(2\pi kt/n)\}^2 + \{\sin(2\pi kt/n)\}^2]\sigma_k^2 \\ &= \sigma_k^2, \end{aligned}$$

since $(\cos\alpha)^2 + (\sin\alpha)^2 = 1$.

Let us now set up ideas that will lead to a definition of the periodogram. Associated with the aggregated periodic-component representations of the data, $x_1, \ldots, x_n$, in terms of the Fourier coefficients, U_{k1} and U_{k2}, are a set of *inverse* formulas that represent the Fourier coefficients in terms of the data:

$$\begin{aligned} U_{k1} &= \frac{2}{n}\sum_{t=0}^{n-1} x_{t+1}\cos(2\pi kt/n) \\ U_{k2} &= \frac{2}{n}\sum_{t=0}^{n-1} x_{t+1}\sin(2\pi kt/n). \end{aligned}$$

With these ideas at hand, the periodogram is defined as follows:

Definition: Suppose $x_1, \ldots, x_n$ are the first n measurements of a time series (x_t). The *periodogram* at $\omega_k = k/n$, for $k = 0, 1, \ldots, n-1$, is

$$I(\omega_k) = \tfrac{n}{4}U_{k1}^2 + \tfrac{n}{4}U_{k2}^2;$$

the *scaled periodogram* at that frequency is

$$P(\omega_k) = U_{k1}^2 + U_{k2}^2.$$

It is clear that these two quantities are related according to $I(\omega_k) = \frac{n}{4}P(\omega_k)$. □

To interpret these quantities, note that $E[U_{k1}^2 + U_{k2}^2] = 2\sigma_k^2 = 2Var[x_{kt}]$. This indicates that the periodogram, $I(\omega_k)$, and scaled periodogram, $P(\omega_k)$, each describes the relative magnitude at which the simple periodic time series, x_{kt}, is expressed in $x_t = \sum_k x_{kt}$.

The periodogram and the underlying representations that motivate it are related to a broader mathematical framework whose central concept is the *discrete Fourier transform* (DFT). This is defined as

$$d(k/n) = \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} x_{t+1}e^{-2\pi ikt/n},$$

for $k = 0, 1, \ldots, n-1$. The inverse DFT is

$$x_{t+1} = \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} d(k/n) e^{2\pi i k t/n}.$$

Without going into details, an important computational idea that is associated with the DFT is the *fast Fourier transform*, which is an efficient algorithm for calculating the DFT.

The DFT makes a connection to the periodogram in the following way. First, observe that

$$\begin{aligned} d(k/n) &= \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x_{t+1} e^{-2\pi i k t/n}. \\ &= \frac{1}{\sqrt{n}} \left\{ \sum_{t=0}^{n-1} x_{t+1} \cos(2\pi k t/n) - i \sum_{t=0}^{n-1} x_{t+1} \sin(2\pi k t/n) \right\} \\ &= \tfrac{\sqrt{n}}{2} U_{k1} - i \tfrac{\sqrt{n}}{2} U_{k2}. \end{aligned}$$

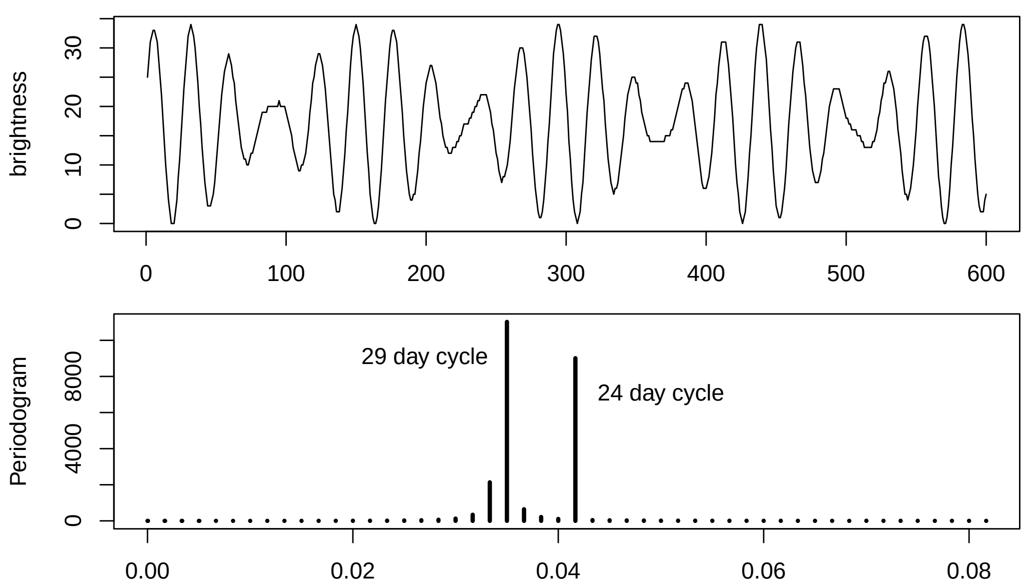
From this, it is quickly seen that

$$|d(k/n)|^2 = \left(\tfrac{\sqrt{n}}{2} U_{k1}\right)^2 + \left(\tfrac{\sqrt{n}}{2} U_{k2}\right)^2 = \tfrac{n}{4} U_{k1}^2 + \tfrac{n}{4} U_{k2}^2 = I(\omega_k).$$

The use of these ideas is illustrated in the following example.

Example: Star brightness

The time series shown in the upper panel of the figure below is of the brightness of a certain star measured at midnight on $n = 600$ consecutive days.



The lower panel is of the scaled periodogram calculated at frequencies $\omega_k = k/n$ for $k = 0, 1, \ldots, 48$; that is, $\omega_k = 0.0000, 0.0017, 0.0033, 0.0050, 0.0067, \ldots, 0.08$. Periodogram values calculated at higher frequencies are nearly zero. The periodogram clearly identifies two relatively strong periodic patterns in these data, one with a 29-day cycle and the other with a 24-day cycle. □

The remainder of our discussion sets up ideas that motivated techniques for classical statistical inference under the frequency domain perspective. These ideas derive from asymptotic sampling properties of the periodogram as the length, n, of the measured time series $x_1, \ldots, x_n$ grows large.

Insight into the sampling properties is found in an alternative formula for the periodogram, given by

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

Observe the parallels of this formula with that of the spectral density

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}.$$

Motivated by such parallels, the periodogram, $I(\omega_k)$, is sometimes called a *sample spectral density*.

Next, define the quantities

$$d_c(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x_{t+1} \cos(2\pi k t/n) \quad \text{and} \quad d_s(\omega_k) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} x_{t+1} \sin(2\pi k t/n)$$

so that the DFT and periodogram are

$$d(\omega_k) = d_c(\omega_k) - i d_s(\omega_k) \quad \text{and} \quad I(\omega_k) = d_c(\omega_k)^2 + d_s(\omega_k)^2.$$

With these quantities defined, the following large-sample, asymptotic properties (stated without proof) may be established.

Property: If n is large, then the following distributional properties hold approximately:

(*i.*) $d_c(\omega_k) \sim N(0, \frac{1}{2} f(\omega_k))$ and $d_s(\omega_k) \sim N(0, \frac{1}{2} f(\omega_k))$;

(*ii.*) If $j \neq k$, $d_c(\omega_j)$, $d_s(\omega_j)$, $d_c(\omega_k)$, and $d_s(\omega_k)$ are independent;

(*iii.*) $E[I(\omega_k)] = f(\omega_k)$ and $Var[I(\omega_k)] = f(\omega_k)^2$;

(*iv.*) and, $2I(\omega_k)/f(\omega_k) \sim \chi^2_2$.

Suppose $k_1, k_2, \ldots$ is such that $\omega_{k_n} \to \omega$ as $n \to \infty$. The property $E[I(\omega_k)] \approx f(\omega_k)$ in (*iii*) implies that when n is large the periodogram $I(\omega_{k_n})$ is approximately unbiased for the spectral density $f(\omega)$. Property (*iv*) provides an approximate confidence interval formula associated with this estimate: an approximate $(1-\alpha)100\%$ confidence interval for $f(\omega)$ has endpoints

$$\frac{2I(\omega_{k_n})}{\chi^2_{2,1-\alpha/2}} \quad \text{and} \quad \frac{2I(\omega_{k_n})}{\chi^2_{2,\alpha/2}}$$

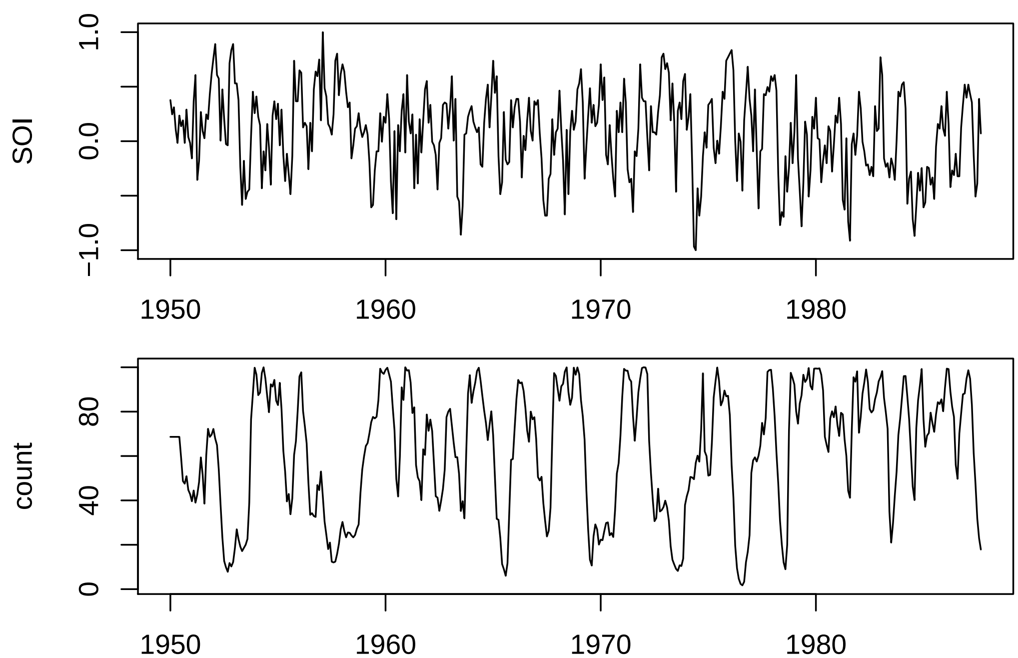
writing $\chi^2_{2,p}$ is the p'th quantile of the χ^2_2 distribution.

However, the property $Var[I(\omega_k)] \approx f(\omega_k)^2$ in (*iii*) contradicts the more desirable property $Var[I(\omega_k)] \to 0$ as $n \to \infty$, which would imply that the periodogram $I(\omega_{k_n})$ is a consistent estimator for the spectral density $f(\omega)$. In other words, $I(\omega_{k_n})$ does not become a more precise estimate of $f(\omega)$ as sample size increases. Reflecting this, the confidence interval for $f(\omega)$ does not narrow to a point as sample size increases! This is a troublesome property of the periodogram, which will be dealt with more carefully in Part B of the learning unit.

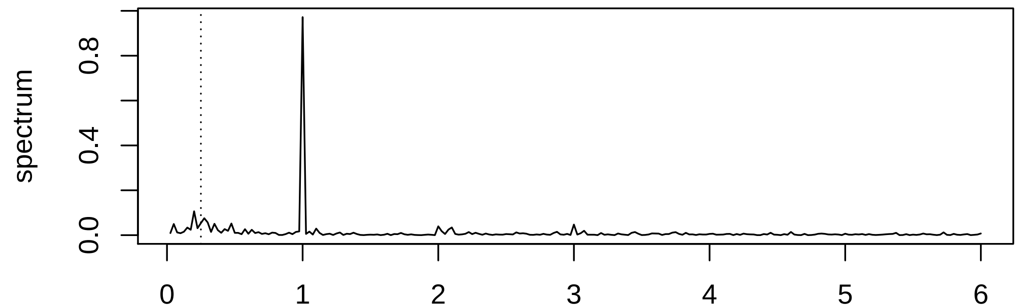
In the meantime, consider the following example which illustrates these ideas.

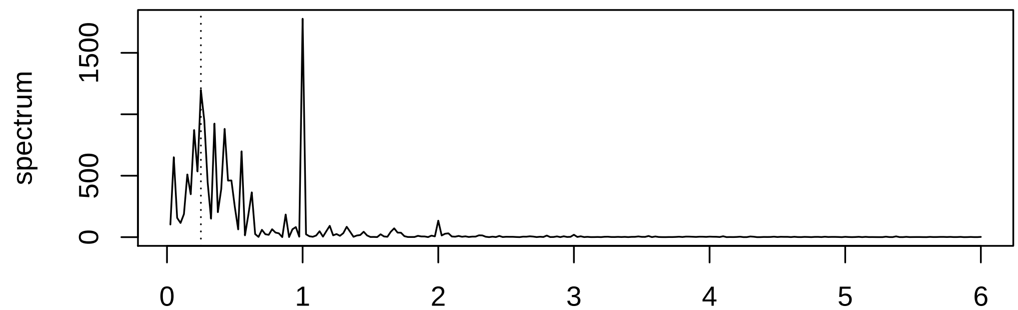
Example: Pacific Ocean time series

The figure below displays time series plots of two time series that we have see before. The top panel plots the Southern Oscillation Index (SOI) time series, which measures ocean surface temperature, and the bottom panel plots a times series of fish population counts. Both are measured in the Pacific Ocean at the same rate, and over the same time period, $n = 453$ monthly measurements during the years 1950-1987.



The figure below displays the periodograms of the two time series. The horizontal axis is labeled in units of 1/12. Thus, for instance, the frequency $\omega = 1/12$ (*i.e.*, a yearly cycle) is one unit; a frequency $\omega = 2/12$ (*i.e.*, a half-year cycle) is two units, *etc.*





From the periodograms, we see that both series exhibit a strong yearly cycle, which is to be expected. There are also indications of lower frequency cycles, particularly in the fish-count data set, particularly around the frequency $\omega = 1/48$ (*i.e.*, a four-year cycle), which is indicated at unit 1/4 on the horizontal axis. The cycle at this frequency is sometimes call the El Niño effect.

Confidence intervals for values of the spectral density at $\omega = 1/12$ and $\omega = 1/48$ (corresponding to yearly and four-year cycles, respectively) are calculated below. Each is for 95% confidence, hence uses the quantile-values $\chi^2_{2,0.025} = 0.0506$ and $\chi^2_{2,0.975} = 7.3778$ of the chi-square distribution with two degrees of freedom.

For the SOI time series at $\omega = 1/12$, the periodogram is $I(1/12) = 0.9722$. The endpoints of the confidence interval for the spectral density value $f(1/12)$ are.

$$\frac{2I(\omega_{k_n})}{\chi^2_{2,1-\alpha/2}} = \frac{2(0.9722)}{7.3778} = 0.2636 \quad \text{and} \quad \frac{2I(\omega_{k_n})}{\chi^2_{2,\alpha/2}} = \frac{2(0.9722)}{0.0506} = 38.4011.$$

It is perhaps more sensible to view this interval on a logarithmic scale. Taking natural logarithms, the confidence interval for $\log f(1/12)$, which is estimated at $\log I(1/12) = \log 0.9722 = -0.0282$, has endpoints

$$\log 0.2636 = -1.3335 \quad \text{and} \quad \log 38.4011 = 3.6481.$$

On either scale, the resulting interval is very, very wide, indicating an extraordinary lack of precision in the estimate. Nevertheless, even the lower endpoint of the interval for $f(1/12)$ is larger than any estimated value of the periodogram at any other frequency, which suggests (informally, not as a rigorous inference) the presence of a strong yearly cycle.

For the SOI time series at $\omega = 1/48$, the periodogram is $I(1/48) = 0.0537$. The endpoints of the confidence interval for the associated spectral density value $f(1/48)$ are

$$\frac{2I(\omega_{k_n})}{\chi^2_{2,1-\alpha/2}} = \frac{2(0.0537)}{7.3778} = 0.01457 \quad \text{and} \quad \frac{2I(\omega_{k_n})}{\chi^2_{2,\alpha/2}} = \frac{2(0.0537)}{0.0506} = 2.1222.$$

On a logarithmic scale, the estimate of $\log f(1/48)$ is $\log I(1/48) = \log 0.0537 = -2.9238$, and the endpoints of the corresponding confidence interval are

$$\log 0.01457 = -4.2291 \quad \text{and} \quad \log 2.1222 = 0.7525.$$

Again, we calculate intervals that are very wide. In this instance, there is very little that can be precisely inferred about this particular value of the spectral density.

Corresponding inferences for the spectral density of the fish population time series are as follows:

At $\omega = 1/12$, the estimate of $f(1/12)$ is $I(1/12) = 1777.75$, and the associated confidence interval is $[481.92, 70217.18]$. On a logarithmic scale, the estimate of $\log f(1/12)$ is $\log 1777.75 = 7.4831$ and the associated confidence interval has endpoints $\log 481.92 = 6.1778$ and $\log 70217.18 = 11.1593$.

At $\omega = 1/48$, the estimate of $f(1/48)$ is $I(1/48) = 1197.37$, and the associated confidence interval is $[324.59, 47293.54]$. On a logarithmic scale, the estimate of $\log f(1/48)$ is $\log 1197.37 = 7.0879$ and the associated confidence interval has endpoints $\log 324.59 = 5.7826$ and $\log 47293.54 = 10.7641$.

All of these calculations illustrate how inference on a particular value of the spectral density may be very imprecise. $\square$