Spectral Analysis/The Spectral Domain

March 11, 2019

The material in this set of notes is based on S&S Example 2.8 and S&S Chapter 4.

Theory

The general idea is that there's another nice way to write out models for time series data. So far, we have always talked about modeling each element of a time series as a linear function of past values and/or past errors. Instead, we might think of each element of a time series as a noisy realization of several cyclical processes:

$$x_t = \sum_{k=1}^{r} v_k \cos\left(2\pi\omega_k t\right) + u_k \sin\left(2\pi\omega_k t\right), \; v_k, u_k \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_k^2\right). \tag{1}$$

There is a very useful theorem called the **spectral representation theorem** that states that if x_t is a stationary process, we can find a **spectral representation** for x_t, i.e. we can find r underlying cyclical processes that vary at rates denoted by $\omega_1, \ldots, \omega_r$ and account for different amounts of the variability $\sigma_1^2, \ldots, \sigma_r^2$ such that (6) holds approximately. Note that it is possible r could be very big! This means that all of our stationary **ARMA**(p, q) models can approximately be written as in (6) certain choices of r, $\omega_1, \ldots, \omega_r$, and $\sigma_1^2, \ldots, \sigma_r^2$, although there may not be nice expressions for r, $\omega_1, \ldots, \omega_r$, and $\sigma_1^2, \ldots, \sigma_r^2$ in terms of our **ARMA**(p, q) parameters $\mu_x, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$, and σ_w^2.

A nice thing about the **spectral representation** of the time series is that it allows

straightforward computation of $\gamma_x(h)$ using some trigonometric identities. When $h = 0$, we have:

$$\begin{aligned}
\gamma_x(0) =& \mathbb{E}\left[x_t^2\right] \\
=& \mathbb{E}\left[\left(\sum_{k=1}^{r} v_k \cos(2\pi\omega_k t) + u_k \sin(2\pi\omega_k t)\right)^2\right] \\
=& \mathbb{E}\left[\sum_{k=1}^{r} \left(v_k \cos(2\pi\omega_k t) + u_k \sin(2\pi\omega_k t)\right)^2\right] + \\
& \mathbb{E}\left[\sum_{k=1}^{r}\sum_{l=1,l\neq k}^{r} \left(v_k \cos(2\pi\omega_k t) + u_k \sin(2\pi\omega_k t)\right)\left(v_l \cos(2\pi\omega_l t) + u_l \sin(2\pi\omega_l t)\right)\right] \\
=& \sum_{k=1}^{r} \mathbb{E}\left[v_k^2\right]\cos^2(2\pi\omega_k t) + \mathbb{E}\left[u_k^2\right]\sin^2(2\pi\omega_k t) \qquad \mathbb{E}[v_k u_k] = 0,\ \mathbb{E}[v_k v_l] = \mathbb{E}[u_k u_l] = 0,\ k \neq l \\
=& \sum_{k=1}^{r} \sigma_k^2. \qquad \sin^2(a) + \cos^2(a) = 1
\end{aligned}$$

More generally, for any $h > 0$ we can write

$$\gamma_x(h) = \sum_{k=1}^{r} \sigma_k^2 \cos(2\pi\omega_k h). \tag{2}$$

Independence of $v_1, \ldots, v_k, u_1, \ldots, u_k$ is the key to straightforward computation of $\gamma_x(h)$, and it is also useful interpretationally as it allows us to decompose a time series with possibly correlated elements into a set of independent random variables.

The actual **spectral representation theorem** is quite technical, and I won't expect you to have a deep understanding of it or ask you to prove it but I will show it and try to explain how it leads us to conclude that any stationary process can approximately be represented by (6) for certain choices of $\omega_1, \ldots, \omega_r$ and $\sigma_1^2, \ldots, \sigma_r^2$. First, we need to define the spectral distribution function $F(w)$.

The spectral distribution function $F(w)$ is closely related to the autocovariance function, $\gamma_x(h)$. It is a monotone non-decreasing function defined on the domain $[-0.5, 0.5]$ that

satisfies:

$$\gamma_x(h) = \int_{-0.5}^{0.5} \exp\{2\pi i\omega h\}\, dF(\omega). \tag{3}$$

By virtue of being monotone non-decreasing and satisfying (3), we can conclude that $F(\omega)$ must have various specific properties - we won't go through them all here, but you can find them in Appendix C of S&S. The important thing to understand is that the spectral distribution function $F(\omega)$ contains the same information as the autocovariance function $\gamma_x(h)$, in the way that the characteristic function of a random variable contains the same information as a random variable's probability density function.

The **spectral representation theorem** states that if x_t is a mean-zero stationary process with spectral distribution function $F(\omega)$ that satisfies (3), then there exists a complex-valued stochastic process $z(\omega)$ on the interval $\omega \in [-0.5, 0.5]$ having stationary uncorrelated increments $dz(\omega)$ such that

$$x_t = \int_{-0.5}^{0.5} \exp\{2\pi i t\omega\}\, dz(\omega), \tag{4}$$

where for $-0.5 \leq \omega_1 \leq \omega_2 \leq 0.5$, $\mathbb{V}[z(\omega_2) - z(\omega_1)] = F(\omega_2) - F(\omega_1)$.

Letting $\omega_1, \ldots, \omega_m$ be an increasing sequence of values from -0.5 to 0.5 and noting that if $z(\omega)$ is stationary, then (4) is equivalent to:

$$x_t = \lim_{r\to\infty} \sum_{k=1}^{r} \exp\{2\pi i t\omega_k\}\, y_k, \tag{5}$$

where the increments $y_k = z(\omega_k) - z(\omega_{k-1})$ are independent, stationary random variables.

We can go from (5) to (6) by applying a bunch of trigonometric identities, and by assuming that the infinite sum in (5) can be reasonably well represented by a finite sum. The technical details are very involved and well beyond the scope of this class, and I have not found a resource that gives a very clear big picture explanation.

Going back to

$$x_t = \sum_{k=1}^{r} v_k \cos\left(2\pi\omega_k t\right) + u_k \sin\left(2\pi\omega_k t\right), \; v_k, u_k \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_k^2\right). \tag{6}$$

The variances $\sigma_1^2, \ldots, \sigma_r^2$ are specific values of the **spectral density function** of x, which is defined as follows. When x_t is stationary and has a summable autocovariance function $\gamma_x(h)$, i.e.

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

then it can be written as

$$\gamma_x(h) = \int_{-0.5}^{0.5} \exp\{2\pi i \omega h\} f(\omega) \, d\omega,$$

where $h = \ldots, -2, -1, 0, 1, 2, \ldots$. This is the inverse transform of the **spectral density**, which in turn is given by

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

where $-0.5 \leq \omega \leq 0.5$. If you are familiar with the Fourier transform, you might recognize that the **spectral density** $f(x)$ as the Fourier transform of the autocovariance function $\gamma_x(h)$. If we refer back to our simple spectral representation for x_t given by (6), each σ_k^2 corresponds to the spectral density function evaluated at frequency ω_k, $\sigma_k^2 = f(\omega_k)$.

The spectral density of a stationary process x_t with summable autocovariances behaves a lot like a probability density function. It corresponds to the derivative of the spectral distribution function $F(\omega)$ and is always nonnegative. However, just as the spectral distribution function is unlike a cumulative distribution function in that $F(0.5) = \gamma_x(0)$ instead of 1, the spectral density function does not always integrate to 1 but rather integrates to $\gamma_x(0)$. The spectral density also has two other nice properties:

- It is symmetric about 0, i.e. $f(\omega) = f(-\omega)$.

- It has period one 1, i.e. $f(\omega + 1) = f(\omega)$.

As a result, we can always just examine $f(\omega)$ on the interval $[0, 0.5]$ without any loss of information - once we know the values of $f(\omega)$ for $0 \leq \omega \leq 0.5$, we know the values of $f(\omega)$ for $-0.5 \leq \omega \leq 0$ and $f(\omega_k + \omega_j)$, where $-0.5 \leq \omega_k \leq 0.5$ and ω_j is any integer.

We can think of the values of the function $f(\omega)$ as the unknown parameters that describe the behavior of a time series x_t when we are working in the **spectral domain**, whereas the values of the function $\gamma_x(h)$ are the unknown parameters that describe the behavior of a time series x_t when we are working in the **time domain**.

Estimation

In practice, we will be interested in estimating the spectral density function $f(\omega)$. It is easiest to think about estimating the spectral density as estimating $\sigma_1^2, \ldots, \sigma_r^2$ based on the model (6), but where the number of cyclical components r is unknown. This is a little odd - it is easiest to write down the spectral representation as an approximate finite sum as in (6), in which case the spectral density function is just a sequence of r values $\sigma_1^2, \ldots, \sigma_r^2$ corresponding to frequencies $\omega_1, \ldots, \omega_r$. However, we will tend to think of spectral density function $f(\omega)$ as a continuous function of ω that is based on the exact spectral representation given by (4).

In practice, what we'll do is use as many cyclical components as possible. Note that (6) looks like a regression problem where $\boldsymbol{x}$ is the response, $\cos(2\pi\omega_k \boldsymbol{t})$ and $\sin(2\pi\omega_k \boldsymbol{t})$ are the covariates with $\boldsymbol{t} = (1, \ldots, n)$, v_k and u_k are the regression coefficients with $\sigma_k^2 = \mathbb{E}\left[\left(v_k^2 + u_k^2\right)/2\right]$. If we want to include as many cyclical components as possible and have n observed response values, then we'll want to use a design matrix with exactly n covariates - the most we can have while still having a full rank design matrix:

$$x_t = \begin{cases} a_0 + \sum_{k=1}^{(n-1)/2} a_k \cos\left(2\pi\left(k/n\right)t\right) + b_k \sin\left(2\pi\left(k/n\right)t\right) & \text{if } n \text{ is odd} \\ a_0 + \left(\sum_{k=1}^{n/2-1} a_k \cos\left(2\pi\left(k/n\right)t\right) + b_k \sin\left(2\pi\left(k/n\right)t\right)\right) + a_{n/2}\cos\left(\pi t\right) & \text{if } n \text{ is even} \end{cases} \tag{7}$$

There is no error in (7) - (7) corresponds to a **saturated regression model**. If we compute all of the regression coefficients a_j and b_j using least squares, we can perfectly recover $\boldsymbol{x}$. This is because we have the same number of covariates as observations, *and* the intercept, sine, and cosine terms in (7) are not linearly dependent. For convenience, we'll introduce some notation for the design matrix that corresponds to (7). The $n \times n$ design matrix $\boldsymbol{Z}$ has an intercept as the first column $\boldsymbol{z}_1 = \mathbf{1}_n$ and remaining columns given by:

$$\boldsymbol{z}_{2(k-1)+2} = \cos\left(2\pi\left(k/n\right)\boldsymbol{t}\right)$$

$$\boldsymbol{z}_{2(k-1)+3} = \sin\left(2\pi\left(k/n\right)\boldsymbol{t}\right)$$

This is a special design matrix, insofar as its columns are nearly orthogonal, i.e. $\boldsymbol{z}_j'\boldsymbol{z}_k \approx 0$ if $j \neq k$. This makes computation especially easy because we can compute the values of a_j and b_j that satisfy (7) either by regressing $\boldsymbol{x}$ on all of the columns of $\boldsymbol{Z}$ simultanously, or by regressing out one column of $\boldsymbol{Z}$ at a time. We can think of each k/n as a frequency. The corresponding squared coefficients at each frequency

$$P\left(k/n\right) = \begin{cases} \hat{a}_k^2 & k = 0 \text{ or } k = n/2 \\ \hat{a}_k^2 + \hat{b}_k^2 & \text{otherwise} \end{cases} \tag{8}$$

is referred to as the **scaled periodogram**. Conveniently, because the scaled periodogram does not depend on the individual values of a_k and b_k it can be computed even more easily using the fast Fourier transform.

The scaled periodogram is almost an estimator of the values of the spectral density function evaluated at the frequencies k/n, but off by a constant. If we are just interested in the shape of the spectral density function, the scaled periodogram will suffice. This may often

be the case in practice, because we will tend to examine to be looking for evidence of very large values of the spectral density function $f(\omega)$, which indicate that certain frequencies are especially relevant for explaining variation of a certain time series.

If we want to actually try to estimate the values of the spectral density function, we need to rescale the **scaled periodogram**. We define the **periodogram** $I(k/n) = \left(\frac{n}{4}\right) P(k/n)$ to be an estimator of the values of the spectral density function evaluated at the frequencies k/n.

Unfortunately, the periodogram is a very noisy estimator of the spectral density, and it does not get less noisy as we observe a longer time series! Intuitively, this makes some sense because can think of our estimate of the periodogram as the result of a regression of n response values on n covariates - as we get more data, we are adding more terms to the model we are fitting.

We're going to need to introduce some new notation to describe how the the asymptotic distribution of elements of the periodogram as $n \to \infty$. Letting $m = n/2$ if n is even and $m = (n-1)/2$ if n is odd, as $n \to \infty$, we can think of our periodogram as estimating values of the spectral density function at a finer resolution $\{0, 1/n, \ldots, m/n\}$ on the interval $[0, 1]$. Let $\boldsymbol{s}$ be a subset of L consecutive frequencies from the $m+1$ frequency values $\{0, 1/n, \ldots, m/n\}$. As $n \to \infty$ and L stays fixed, all of the entries of $\boldsymbol{s}$ will converge to a constant value s, i.e. $\boldsymbol{s} \to s\mathbf{1}_L$.

If a time series $\boldsymbol{x}$ has a linear process representation with $x_t = \sum_{j=-\infty}^{\infty} \psi_j w_{t-j}$, where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$, $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$, and $\sum_{h=-\infty}^{\infty} |h| \, |\gamma_x(h)| < \infty$, then for any collection $\boldsymbol{s}$ of L frequencies that converge to $s\mathbf{1}_L$ as $n \to \infty$ with $f(s) > 0$,

$$\begin{pmatrix} \frac{2I(s_1)}{f(s)} \\ \vdots \\ \frac{2I(s_L)}{f(s)} \end{pmatrix} \xrightarrow{d} \begin{pmatrix} c_1 \\ \vdots \\ c_L \end{pmatrix}, \tag{9}$$

where $c_1, \ldots, c_L \overset{i.i.d.}{\sim} \chi_2^2$ as $n \to \infty$.

Approximate level-α confidence intervals can be obtained in the usual way. Note that (9) does *not* indicate that the limiting distribution does *not* have a decreasing variance as $n \to \infty$ - the periodogram is not getting less noisy! There are two main approaches to addressing this, smoothing the periodogram and parameteric estimation of the periodogram.

- **Smoothing** the periodogram. If the spectral density is a smooth function of ω, we can share information across consecutive periodogram values. Again, let $\boldsymbol{s}$ be a subset of L consecutive frequencies from the $m+1$ frequency values $\{0, 1/n, \ldots, m/n\}$ that converges to a constant s as $n \to \infty$. We can naively construct a lower variance estimate of the spectral density at s, $\bar{f}(\boldsymbol{s}) = \frac{1}{L}\sum_{i=1}^{L} I(s_i)$.

  If the conditions that ensure (9) hold, as $n \to \infty$,

  $$\frac{2L\bar{f}(\boldsymbol{s})}{f(s)} \xrightarrow{d} \chi^2_{2L}. \tag{10}$$

  In practice, (10) will apply as long as we choose L to be small relative to n and as long as $f(\omega)$ is not varying too much on the interval $(s_1, \ldots, s_L)$. We will often quantify the amount of smoothing imposed by referring to the bandwidth $B = L/n$, which gives the lengths of the intervals chosen for smoothing. Larger bandwidths correspond to more aggressive smoothing.

  A more sophisticated approach might not weight elements of $\boldsymbol{s}$ the same in the spectral density estimate. For weights $h_1, \ldots, h_L > 0$ which are symmetric about $h_{(L-1)/2}$ with $h_{(L-1)/2-k} = h_{(L-1)/2+k}$ and $\sum_{i=1}^{L} h_i = 1$, we can construct another estimate of the spectral density at s, $\hat{f}(\boldsymbol{s}) = \sum_{i=1}^{L} h_i I(s_i)$. The choice of weights $h_1, \ldots, h_L$ are determined according to a kernel function.

  If the conditions that ensure (9) hold and $\sum_{i=1}^{L} h_i^2 \to 0$ as $n \to \infty$ and $L \to \infty$ but $L/n \to 0$, then for any collection $\boldsymbol{s}$ of L frequencies that converge to $s\mathbf{1}_L$ as

$n \to \infty$ with $f(s) > 0$, then as $n \to \infty$,

$$\frac{2L_h \hat{f}(\boldsymbol{s})}{f(s)} \xrightarrow{d} \chi^2_{2L_h}, \tag{11}$$

where $L_h = \left(\sum_{i=1}^{L} h_i^2\right)^{-1}$.

Again, (11) will apply in practice as long as we choose L to be small relative to n and as long as $f(\omega)$ is not varying too much on the interval $(s_1, \ldots, s_L)$. To quantify the amount of smoothing imposed, we generalize the definition of bandwith to be $B = L_h/n$.

A note of caution - how much smoothing is imposed matters a lot! Smoothing too much can make it impossible to observe evidence of peaks of the spectral density, and smoothing too little can make it easy to mistake noise for peaks of the spectral density. There are principled approaches to deciding how to smooth and how much, but they are beyond the scope of this class.

- **Parametric estimation** of the periodogram. This is a pretty simple idea - we first find the AIC minimizing $\mathbf{AR}(p)$ model for a time series n, and then examine the spectral density assuming that the estimated $\mathbf{AR}(p)$ parameters are the truth. The spectral density is not a simple function of the $\mathbf{AR}(p)$ parameters, but the spectral density estimate constructed by plugging the $\mathbf{AR}(p)$ parameter estimates into the spectral density function of an $\mathbf{AR}(p)$ model does improve as $n \to \infty$ without additional smoothing.

  - This is justifiable in practice thanks to a very powerful theorem that tells us that if $f(\omega)$ is the spectral density of a stationary process, we can find a causal $\mathbf{AR}(p)$ process with spectral density $g(\omega)$ that approximates $f(\omega)$ arbitrarily well! However, we might need to choose a value of p that is prohibitively large to get a good enough approximation of the spectral density from an $\mathbf{AR}(p)$ model in practice.

– Unfortunately, the existing theory that gives a limiting distribution of the a spectral density estimate obtained in this way relies on assumptions that may be too strong, so the bootstrap is often used to get confidence intervals for the parametric estimate of the periodogram in practice.