TIME SERIES ANALYSIS

Chapter 7: Spectral analysis

The idea is to 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 the spectral analysis, or analysis in the frequency domain, 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.

1 The Autocovariances and the Spectral Representation

The autocovariance and autocorrelation functions describe the evolution of a process through time (time domain). Spectral density function describes the frequency properties of a time series, complementary to the acv.f. and ac.f. functions (frequency domain).

Suppose we suspect that a time series contains a periodic sinusoidal component with a known wavelength. Then a natural model is

$$X_t = R\cos(\omega t + \phi) + Z_t$$

where ω is called the frequency of the sinusoidal variation, R is called the amplitude of the variation, ϕ is called the phase, $2\pi/\omega$ is the period and Z_t denotes a stationary random series.

In practice, the variation in a time series may be caused by variation at several different frequencies. For example, sales figures may contain weekly, monthly, yearly, and other cyclic variation. In other words, data my show variation at high, medium and low frequency. It is natural therefore to generalize the above model to

$$X_t = \sum_{j=1}^{k} R_j \cos(\omega_j t + \phi_j) + Z_t.$$

Since $cos(\omega t + \phi) = cos\omega t \cos\phi - \sin\omega t \sin\phi$, the model becomes

$$X_t = \sum_{j=1}^{k} a_j \cos(\omega_j t) + b_j \sin(\omega_j t) + Z_t,$$

where $a_j = R_j \cos\phi_j$ and $b_j = -R_j \sin\phi_j$.

However we may ask why there should only be a finite number of frequencies involved in the above models. In fact, letting $k \to \infty$, it is shown that any discrete-time stationary process measured at unit intervals may be represented as

$$X_t = \int_0^{\pi} \cos(\omega t) du(\omega) + \int_0^{\pi} \sin(\omega t) dv(\omega) \tag{1.1}$$

where $u(\omega)$ and $v(\omega)$ are uncorrelated continuous processes with orthogonal increments. (**??**) is called **the spectral representation** of the process. Every frequency in the range $(0, \pi)$ may contribute to the variation of X_t.

The main point of introducing the spectral representation is to show that every frequency in the range $(0, \pi)$ may contribute to the variation of the process. However the processes $u(\omega)$ and $v(\omega)$ are of little direct practical interest. Instead we introduce a single function, $F(\omega)$, called the **(power) spectral distribution function**.

THEOREM 1 (Wiener-Khintchine Theorem) For any real valued stationary stochastic process with acvf γ_k there exists a non-decreasing function $F(\omega)$ such that

$$\gamma_k = \int_0^{\pi} \cos(\omega k) dF(\omega) \tag{1.2}$$

where $F(\omega)$ is the spectral distribution function. The above equation is called the spectral representation of the autocovariance function.

$F(\omega)$ is the contribution to the variance of X_t, which is accounted for big frequencies in the range $(0, \pi)$. It has the following properties.

- ▷ $F(\omega) = 0$ for $\omega < 0$.
- ▷ $\gamma_0 = \int_0^{\pi} \cos(0) dF(\omega) = F(\pi) = Var(X_t) = \sigma_X^2$.
- ▷ $F(\omega)$ is monotonically increasing between $\omega = 0$ and $\omega = \pi$.

```
> t=1:96; cos1=cos(2*pi*t*4/96); cos2=cos(2*pi*(t*14/96+.3))
> plot(t,cos1, type='o', ylab='Cosines')
> lines(t,cos2,lty='dotted',type='o',pch=4)
```

Exhibit 1 displays two discrete-time cosine curves with time running from 1 to 96. We would only see the discrete points, but the connecting line segments are added to help our eyes follow the pattern. 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.

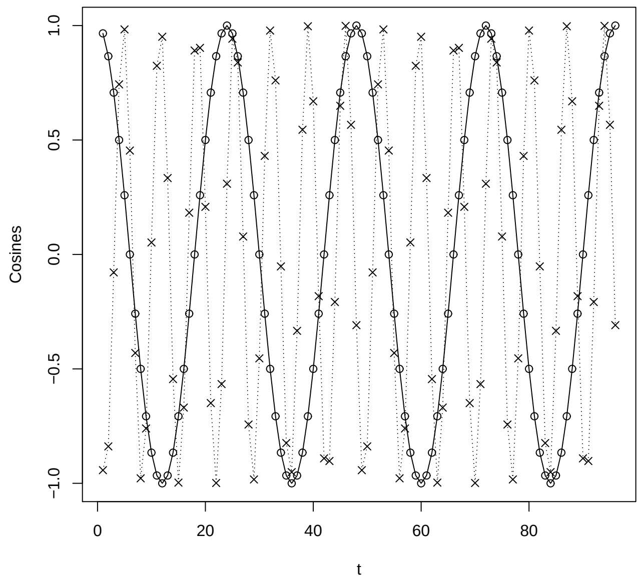

Figure 1: Cosine Curves with n = 96 and Two Frequencies and Phases

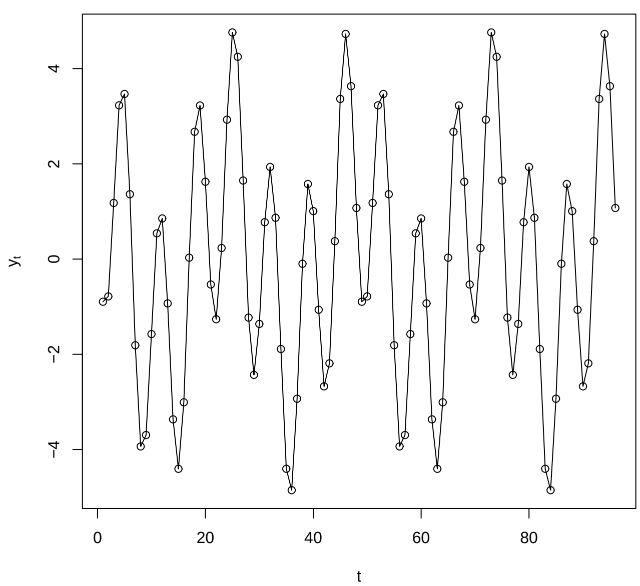

Figure 2: Linear Combination of Two Cosine Curves

```
> y=2*cos1+3*cos2
> plot(t,y, type='o',ylab=expression(y[t]))
```

Now the periodicity is somewhat hidden. Spectral analysis provides tools for discovering the "hidden" periodicities quite easily. Of course, there is nothing random in this time series.

2 The Spectral Density Function

Definition 1 The spectral density function, also called the spectrum, is given by

$$f(\omega) = \frac{dF(\omega)}{d\omega}. \tag{2.1}$$

In view of this, (**??**) becomes

$$\gamma(k) = \int_0^{\pi} \cos(\omega k) f(\omega) d\omega. \tag{2.2}$$

This yields

$$\gamma(0) = \int_0^{\pi} \cos(0) f(\omega) d\omega = F(\pi) = \sigma_X^2.$$

The physical interpretation of the spectrum is that $f(\omega)d\omega$ represents the contribution to variance of components with frequencies in the range $(\omega, \omega + d\omega)$. A peak in the spectrum indicates an important contribution to variance at frequencies near the value that corresponds to the peak.

It is important to realize that the autocovariance function (acvf) and the spectral density function are equivalent ways of describing a stationary process. Both functions contain the same information but express it in different ways.

It can be shown that the corresponding inverse relationship to (**??**)) is given by

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-ik\omega} \tag{2.3}$$

so that the spectrum is the **Fourier transform** of the acvf. Since $\gamma(k)$ is even in k and $e^{-i\alpha} = \cos\alpha + i\sin(\alpha)$ (**??**) is often written in the equivalent form

$$f(\omega) = \frac{1}{\pi}\Big[\gamma(0) + 2\sum_{k=1}^{\infty} \gamma(k)\cos(\omega k)\Big]. \tag{2.4}$$

Note that if we try to apply the above equation to a process containing a deterministic component at a particular frequency ω_0, then $\sum \gamma(k) cos(\omega k)$ will not converge when $\omega = \omega_0$.

The normalized form of $f(\omega)$ is

$$f^*(\omega) = \frac{f(\omega)}{\sigma_X^2} = \frac{1}{\pi}\Big[\gamma(0) + 2\sum_{k=1}^{\infty}\rho(k)\cos(\omega k)\Big],$$

where $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$. This means that $f^*(\omega)$ is the Fourier transform of acf and $f^*(\omega)d\omega$ is the proportion of variance on $(\omega, \omega + d\omega)$.

3 periodogram

Given a realization $x_1, ..., x_n$ of a time series, how can we estimate the spectral density ? One plausible approach may be to replace $\gamma(k)$ in the definition

$$f(\omega) = \frac{1}{\pi}\sum_{k=-\infty}^{k=+\infty}\gamma_k e^{-i\omega k}$$

with sample acvf $\hat{\gamma}(k)$.

The alternative is to use periodogram. Define $R_p = \sqrt{\hat{\alpha}_p^2 + \hat{\beta}_p^2}$.

DEFINITION 2 For odd sample sizes with n = 2k + 1, the periodogram $I(\omega_p)$ at frequency $\omega_p = 2\pi p/n$, is defined to be

$$I(\omega_p) = \frac{n}{4\pi}R_p^2.$$

The plot of $I(\omega_p)$ vs p is called the periodogram.

Note that the formula for R_p^2, and hence for $I(\omega_p)$, can be written in several equivalent ways that look quite different. For example it can be shown that

$$I(\omega_p) = \frac{1}{\pi n}\Big|\frac{1}{n}\sum_{t=1}^{n}e^{it\omega_p}x_t\Big|^2. \tag{3.1}$$

Note that $I(\omega_p)$ is the squared modulus of the discrete Fourier transform (at frequencies ω_p). Moreover using equation

$$\sum_{t=1}^{n}\cos(\omega_p t) = \sum_{t=1}^{n}\sin(\omega_p t) = 0$$

we may further write

$$I(\omega_p) = \frac{1}{\pi n}\Big|\frac{1}{n}\sum_{t=1}^{n}e^{it\omega_p}(x_t - \bar{x})\Big|^2$$

$$= \frac{1}{\pi n}\Big(\frac{1}{n}\sum_{t=1}^{n} e^{it\omega_p}(x_t - \bar{x})\Big)\Big(\frac{1}{n}\sum_{t=1}^{n} e^{-it\omega_p}(x_t - \bar{x})\Big)$$

$$= \frac{1}{\pi n}\Big(\sum_{t,s=1}^{n} e^{-i(s-t)\omega_p}(x_t - \bar{x})\Big)(x_s - \bar{x})\Big)$$

$$= \frac{1}{\pi}\sum_{k=-(n-1)}^{n-1} \hat{r}_k e^{-ik\omega_p}, \tag{3.2}$$

where

$$\hat{r}_k = \frac{1}{n}\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x}).$$

The formula (**??**) is called a discrete finite fourier transform (assuming that $\hat{r_k} = 0$ for $|k| > n$).

Figure 3 below displays a graph of the periodogram for the time series in Fig 2. The heights show the presence and relative strengths of the two cosine-sine components quite clearly. Note also that the frequencies 4/96 . 0.04167 and 14/96 . 0.14583 have been marked on the frequency axis.

```
>library(TSA)
>periodogram(y)
>abline(h=0)
> axis(1,at=c(0.04167,0.14583))
```

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"? To illustrate, we generate a time series using randomness to select the frequencies, amplitudes, and phases and with additional additive white noise. The two frequencies are randomly chosen without replacement from among $1/96, 2/96, \cdots, 47/96$. 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. Finally, a normal white noise series, Z_t, with zero mean and standard deviation 1, is chosen independently of the A's and B's and added on. The model is

$$y_t = A_1\cos(2\pi f_1 t) + B_1\sin(2\pi f_1 t) + A_2\cos(2\pi f_2 t) + B_2\cos(2\pi f_2 t) + Z_t.$$

Fig 4 displays a time series of length 96 simulated from this model. Once more, the periodicities are not obvious until we view the periodogram shown in fig 5.

```
> set.seed(134); t=1:96; integer=sample(48,2)
> freq1=integer[1]/96; freq2=integer[2]/96
> A1=rnorm(1,0,2); B1=rnorm(1,0,2)
> A2=rnorm(1,0,3); B2=rnorm(1,0,3); w=2*pi*t
> y=A1*cos(w*freq1)+B1*sin(w*freq1)+A2*cos(w*freq2)+
B2*sin(w*freq2)+rnorm(96,0,1)
> plot(t,y,type='o',ylab=expression(y[t]))
```

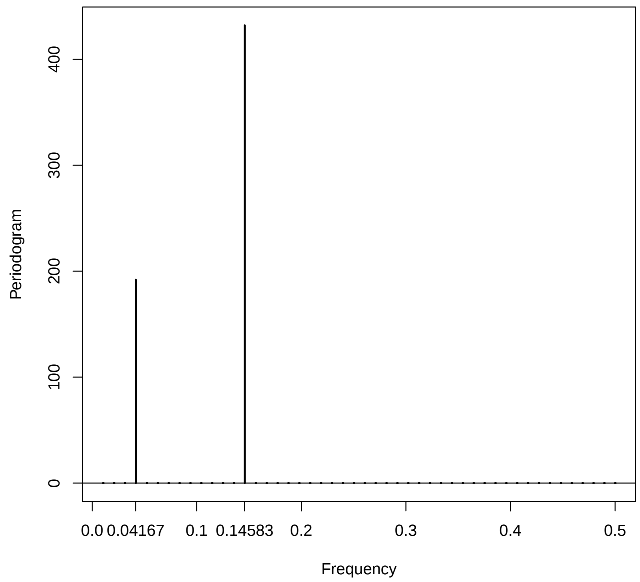

Figure 3: Periodogram of the Series in Fig 2

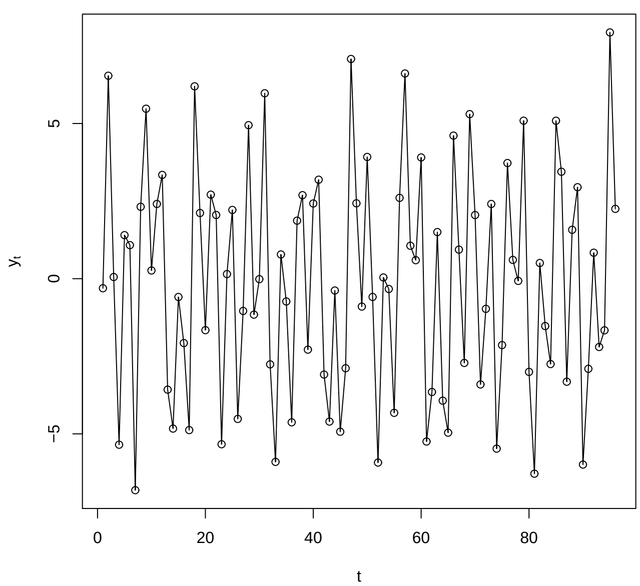

Figure 4: Time Series with "Hidden" Periodicities

The periodogram (fig 5) clearly shows that the series contains two cosine-sine pairs at frequencies of about 0.11 and 0.32 and that the higher-frequency component is much stronger. There are some other very small spikes in the periodogram, apparently caused by the additive white noise component. (When we checked the simulation in detail, we found that one frequency was chosen as 10/96 . 0.1042 and the other was selected as 30/96 = 0.3125.)

```
>periodogram(y);abline(h=0)
```

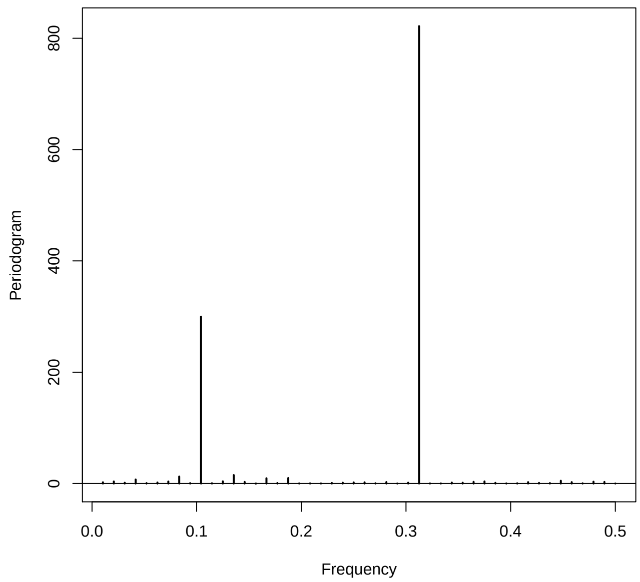

Figure 5: Periodogram of the Time Series Shown in Fig 4

The two sharp peaks suggest a model for this series with just two cosine-sine pairs with the appropriate frequencies or periods,

$$y_t = A_1 \cos(2\pi f_1 t) + B_1 \sin(2\pi f_1 t) + A_2 \cos(2\pi f_2 t) + B_2 \cos(2\pi f_2 t) + Z_t.$$

If we estimate this regression model, we obtain highly statistically significant regression coefficients for all five parameters and a multiple R-square value of 99.9%.

4 Derivation of some Spectral Density Functions

Example 1 A Purely Random Process Z_t is a sequence of independent random variables. Its acvf is given by

$$\gamma(k) = \begin{cases} \sigma^2, & k = 0 \\ 0, & otherwise \end{cases}.$$

It follows that the power spectral density function is

$$f(\omega) = \frac{1}{\pi}\Big[\gamma_0 + 2\sum \gamma_k \cos \omega k\Big] = \frac{1}{\pi}\Big[\gamma_0 + 0\Big] = \frac{\gamma_0}{\pi}.$$

Hence the spectrum is a constant on $(0, \pi)$. ■

Example 2 Moving Average Process MA(1):

$$X_t = Z_t + \beta Z_{t-1}.$$

Its acf is

$$\rho(k) = \begin{cases} 1, & k = 0 \\ \frac{\beta}{1+\beta^2} & k = \pm 1 \\ 0, & otherwise \end{cases}.$$

It follows that the normalized spectral density function is

$$f^\star(\omega) = \frac{1}{\pi}\Big[1 + 2\sum \rho_k \cos \omega k\Big] = \frac{1}{\pi}\Big[1 + \frac{2\beta\cos\omega}{1+\beta^2}\Big].$$

Note that the spectral density function is then

$$f(\omega) = \sigma_X^2 f^*(\omega),$$

with $\sigma_X^2 = (1 + \beta^2)\sigma_Z^2$.

The shape of the spectrum depends on the value of β_1. If $\beta_1 > 0$, the power concentrates at low frequencies, which is referred to as a low-frequency spectrum; If $\beta_1 < 0$, the power concentrates at high frequency which is referred to as a high-frequency spectrum.

Example 3 Autoregressive Process AR(1)

$$X_t = \alpha X_{t-1} + Z_t,$$

has an acf given by

$$\gamma_k = \sigma_X^2 \alpha^{|k|}, \quad k = 0, \pm 1, \pm 2, \cdots$$

and

$$\sigma_X^2 = \frac{\sigma_Z^2}{1 - \alpha^2}.$$

Note that

$$f(\omega) = \frac{1}{\pi}\sum_{k=-\infty}^{k=+\infty}\gamma_k e^{-i\omega k} = \frac{\sigma_X^2}{\pi}\left[\frac{1}{1-\alpha e^{i\omega}} + \frac{\alpha e^{i\omega}}{1-\alpha e^{i\omega}}\right]$$
$$= \frac{\sigma_X^2}{\pi}\frac{1-\alpha^2}{1-2\alpha\cos\omega+\alpha^2}.$$

The shape of the spectrum depends on the value of α. If $\alpha > 0$ $f(\omega)$ is large when ω is small—low frequency spectrum while if $\alpha < 0$ $f(\omega)$ is large when ω is large—high frequency spectrum. One should note that

- $\alpha > 0, \rho_k > 0, X_t$ tends to be similar
- $\alpha < 0, \rho_k < 0, X_t$ tends to be different, which implies that values of X_t will tend to change sign at every time point, and rapid oscillations like this correspond to high frequency variation.

Example 4 (Deterministic Sinusoidal Perturbations.) Suppose that

$$X_t = \cos(\omega_0 t + \phi)$$

where ω_0 is a constant in $(0, \pi)$ and ϕ has a uniform distribution on $(0, 2\pi)$. One can verify that

$$EX_t = 0, \quad Var(X_t) = \frac{1}{2}, \quad \gamma_k = \frac{1}{2}\cos\omega_0 k.$$

Note that γ_k does not tend to zero as k increases. Recall that

$$\gamma_k = \int_0^{\pi}\cos\omega k dF(\omega).$$

This implies that the power spectral distribution function is

$$F(\omega) = \begin{cases} 0 & \omega < \omega_0 \\ \frac{1}{2}, & \omega \geq \omega_0 \end{cases}$$

Because there is no derivative of $F(\omega)$ at ω_0, $f(\omega)$ is not defined at $\omega = \omega_0$. ■

Example 5 (A Mixture of Deterministic and Stochastic Components.) Suppose that

$$X_t = \cos(\omega_0 t + \phi) + Z_t$$

where ω_0, ϕ are defined as in the last example and Z_t is $N(0, \sigma_Z^2)$. Its acf is

$$\gamma_k = \begin{cases} \frac{1}{2} + \sigma_Z^2 & k = 0 \\ \frac{1}{2}\cos\omega_0 k, & k = \pm 1, \pm 2, \cdots. \end{cases}$$

For $\cos(\omega_0 t + \phi)$,

$$F_1(\omega) = \begin{cases} 0 & \omega < \omega_0 \\ \frac{1}{2}, & \omega \geq \omega_0 \end{cases}$$

For Z_t, $F_2(\omega) = \frac{\sigma_Z^2}{\pi}\omega$, $0 < \omega < \pi$. It follows that

$$F(\omega) = F_1(\omega) + F_2(\omega),$$

which is a step function and $f(\omega)$ is not defined at $\omega = \omega_0$. ■

In fact, one may show that if X_t and Y_t are independent stationary processes with spectral density function $f_X(\omega)$ and $f_Y(\omega)$ then $V_t = X_t + Y_t$ is also stationary with spectrum $f_V(\omega) = f_X(\omega) + f_Y(\omega)$.

5 Spectral Analysis

Suppose that a function $f(t)$ is defined on $(-\pi, \pi]$ and satisfies the do-called Dirichlet conditions. The approximation of a function by taking sum of sine and cosine terms, is called the Fourier series representation and, for a function f(t), is given by

$$f(t) = \frac{a_0}{2} + \sum_{r=1}^{k} a_r \cos(rt) + b_r \sin(rt),$$

where $a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)dt$,

$$a_r = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(rt)dt, \quad b_r = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(rt)dt, \quad r = 1, 2, \cdots .$$

Fourier series converges to $f(t)$ as $k \to \infty$ except at points of discontinuity, where it converges to $\frac{1}{2}(f(t+0) + f(t-0))$.

In order to apply Fourier analysis to discrete time series, we need to consider the Fourier series representation of $f(t)$ is defined only on the integers $1, 2, \cdots, N$. We below demonstrate that the required Fourier series emerges naturally by considering a simple sinusoidal model.

Consider the model

$$X_t = \mu + \sum_{j=1}^{m} \Big[\alpha_j cos(2\pi f_j t) + \beta_j sin(2\pi f_j t)\Big]. \tag{5.1}$$

Ordinary least squares regression can be used to fit α_j and β_j. Suppose that n is odd and write $n = 2k + 1$. Then the frequencies of the form $1/n, 2/n, \cdots, k/n$ (= 1/2 - 1/(2n)) are called the Fourier frequencies. The cosine and sine predictor variables at these frequencies (and at f = 0) are known to be orthogonal, and the least squares estimates are simply $\mu = \bar{x}$

$$\hat{\alpha}_j = \frac{2}{n}\sum_{t=1}^{n} x_t cos(2\pi tj/n), \quad \hat{\beta}_j = \frac{2}{n}\sum_{t=1}^{n} x_t sin(2\pi tj/n), j = 1, 2, \cdots$$

The overall effect of the Fourier analysis is to partition the variability of the series into components at frequencies $\frac{2\pi}{n}, \frac{4\pi}{n}, \cdots, \pi$.

5.1 Asymptotic properties of the periodogram

Recall from (**??**)

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{+\infty} \gamma(k) e^{-ik\omega}.$$

This, together with (**??**), implies that $I(\omega_p)$ seems a natural estimate of $f(\omega)$. However, although we find

$$\lim_{n\to\infty} E(I(\omega)) = f(\omega),$$

so that the periodogram is asymptotically unbiased we see below that the variance of $I(\omega)$ does not decrease as n increases. Thus, $I(\omega_p)$ is not a consistent estimator, it is a bad estimator.

Indeed, when $x_1, x_2, \cdots, x_n$ are i.i.d $N(0, \sigma^2)$ for model (**??**) one may verify that

$$Var(I(\omega_p) = \frac{\sigma^4}{\pi^2}.$$

As this variance is a constant it does not tend to zero as $n \to \infty$.

5.2 Estimating the spectral density: transforming the truncated autocovariance

One type of estimation procedure consists of taking a Fourier transform of the truncated weighted sample acvf. It is clear that the precision of the values of $\hat{r}_k$ decreases as k increases because the coefficients are based on fewer and fewer terms. Thus it would seem intuitively reasonable to give less weight to the values of $\hat{r}_k$ as k increases.

For a set of weights λ_k, and a number M < N, introduce

$$\hat{f}(\omega) = \frac{1}{\pi}\left[\lambda_0 \hat{r}_0 + 2\sum_{k=1}^{M} \lambda_k \hat{r}_k \cos \omega k\right],$$

where λ_k are called the lag window and M is called the truncation window. Usually $M = o(n)$.

Two best known lag windows are as follows.

Definition 3 Tukey window:

$$\lambda_k = \frac{1}{2}(1 + \cos\frac{\pi k}{M}), \quad k = 0, 1, \cdots, n.$$

DEFINITION 4 Parzen window

$$\lambda_k = \begin{cases} 1 - \frac{1}{6}(\frac{k}{M})^2 + 6(\frac{k}{M})^3 & k \leq M/2 \\ 2(1-k/M)^3 & M/2 \leq k \prec M \end{cases}.$$

Tukey is probably the most commonly used. Choice of M relies on the balance of "resolution" against "variance": the larger M - the rougher the result; commonly, M is selected to be $2\sqrt{n}$.

5.3 Estimating the spectral density: Smoothing the periodogram

The basic idea here is that most spectral densities will change very little over small intervals of frequencies. As such, we should be able to average the values of the sample spectral density over small intervals of frequencies to gain reduced variability. In doing so, we must keep in mind that we may introduce bias into the estimates if, in fact, the theoretical spectral density does change substantially over that interval. There will always be a trade-off between reducing variability and introducing bias. We will be required to use judgment to decide how much averaging to perform in a particular case.

Let ω be a Fourier frequency. Consider taking a simple average of the neighboring sample spectral density values centered on frequency f and extending m Fourier frequencies on either side of ω. We are averaging 2m + 1 values of the sample spectrum, and the smoothed sample spectral density is given by

$$\hat{f}(\omega) = \frac{1}{2m+1}\sum_{k=-m}^{+m} I(\omega + \frac{2\pi k}{n}).$$

When averaging for frequencies near the end points of 0 and π, we treat the periodogram as symmetric about 0 and π. Taking m to be odd with $m^* = (m-1)/2$ we have

$$\hat{f}(0) = I(0) + \frac{2}{m}\sum_{k=1}^{m*} I(\frac{2\pi k}{n})$$

and

$$\hat{f}(\pi) = I(\pi) + \frac{2}{m}\sum_{k=1}^{m*} I(\pi - \frac{2\pi k}{n}).$$

The larger the value of m the smaller will be the variance of the resulting estimate but the larger will be the bias.

A somewhat better option is to use a weighted average

$$\hat{f}(\omega) = \sum_{k=-m}^{+m} W_m(k) I(\omega + \frac{2\pi k}{n}),$$

where the weights satisfy

$$W_m(k) \geq 0, \quad W_m(k) = W_m(-k), \quad \sum_{k=-m}^{m} W_m(k) = 1.$$