Chapter 8

Spectral Methods

8.1 Introduction

Spectral methods are based on thinking of a time series as a superposition of sinusoidal fluctuations of various frequencies – the analogue for a random process of the Fourier representation of a deterministic function. Statistical analysis aims to make inferences about the importance of the different frequencies. This form of analysis is popular for engineering and physical sciences data and in some theoretical studies. The methods extend nicely to multiple time series and spatial processes.

Alternative name: frequency domain methods.

The aim of this short chapter is to outline basic ideas only, in a roughly historical order.

References for a complete treatment are:

P. Bloomfield (1973) *Fourier Analysis of Time Series: an Introduction*, Wiley.

M. B. Priestley (1981) *Spectral Analysis and Time Series*, Academic Press.

P. J. Brockwell & R. A. Davis (1993) *Time Series: Theory and Methods*, Springer-Verlag.

8.2 Harmonic Regression and the Periodogram

Suppose that a data set (eg the Beveridge Wheat Price series) is suspected of having a periodic behaviour. A simple model for variation around the mean might be

$$X_t = \alpha \cos \omega t + \beta \sin \omega t + \epsilon_t, \tag{8.1}$$

for $t = 1, \ldots, n$, where ω is some frequency and ϵ_t is white noise $\mathrm{WN}(0, \sigma^2)$.

(Recall that for $\cos \omega t$ and $\sin \omega t$:

wavelength	=	$2\pi/\omega$
cycles/obs.	=	$\omega/2\pi$)

We can take $\omega \in [0, \pi]$, because if originally $\omega \in [2j\pi, (2j+1)\pi)$ for a positive integer j, then for integer t

$$\cos \omega t = \cos(\omega - 2j\pi)t \qquad \text{and} \qquad \sin \omega t = \sin(\omega - 2j\pi)t,$$

and if $\omega \in [(2j+1)\pi, (2j+2)\pi)$ then

$$\cos \omega t = -\cos(\omega - (2j+1)\pi)t \qquad \text{and} \qquad \sin \omega t = -\sin(\omega - (2j+1)\pi)t,$$

so that the form of the model stays the same as if ω were in $[0, \pi]$.

Note that values of ω close to 0 give slowly changing waves, and values of ω close to π give an alternating sequence of positive and negative values, the fastest oscillation we can expect to see in discrete time.

We will concentrate on ω values of the form $\omega = \omega_p = 2\pi p/n$ for an integer p with $1 \leq p < n/2$. This is little restriction if the sample size n is not too small, and is highly convenient mathematically. The frequencies ω_p are called the *Fourier frequencies.*

The simple model (8.1) is a linear model

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} \cos \omega_p & \sin \omega_p \\ \cos 2\omega_p & \sin 2\omega_p \\ \vdots & \vdots \\ \vdots & \vdots \\ \cos n\omega_p & \sin n\omega_p \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \\ \epsilon_n \end{pmatrix},$$

that is,

$$\mathbf{X} = S\theta + \epsilon, \tag{8.2}$$

say, where S is the $n \times 2$ matrix of sines and cosines above and $\theta = (\alpha, \beta)'$.

The least squares estimates of α and β are therefore

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (S'S)^{-1}S'\mathbf{X},$$

which reduce to

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \frac{2}{n}\sum_{j=1}^{n} X_j \cos j\omega_p \\ \frac{2}{n}\sum_{j=1}^{n} X_j \sin j\omega_p \end{pmatrix}$$

through the first three of the trigonometric identities

$$\sum_{j=1}^{n} \cos^2 j\omega_p = \sum_{j=1}^{n} \sin^2 j\omega_p = n/2$$

$$\text{and} \qquad \sum_{j=1}^{n} \cos j\omega_p \sin j\omega_p = \sum_{j=1}^{n} \cos j\omega_p = \sum_{j=1}^{n} \sin j\omega_p = 0.$$

(The final two identities will be needed later.)

The sum of squares due to regression on the two covariates is $\mathbf{X}'S(S'S)^{-1}S'\mathbf{X}$, which reduces to

$$\begin{aligned} \text{Regression SS}(\omega_p) &= \frac{2}{n}\left[\left(\sum_{1}^{n} X_j \cos j\omega_p\right)^2 + \left(\sum_{1}^{n} X_j \sin j\omega_p\right)^2\right] \\ &= n(\hat{\alpha}^2 + \hat{\beta}^2)/2. \end{aligned}$$

Thus the regression sum of squares is proportional to the squared amplitude $\hat{\alpha}^2 + \hat{\beta}^2$ of the fitted sinusoid.

A more refined model would allow sinusoidal variation at different frequencies:

$$X_t = \sum_{k=1}^{m} \{\alpha_k \cos \omega_k t + \beta_k \sin \omega_k t\} + \epsilon_t, \tag{8.3}$$

for some $m < n/2$. When $p \neq q$:

$$\sum_{j=1}^{n} \cos j\omega_p \cos j\omega_q = \sum_{j=1}^{n} \cos j\omega_p \sin j\omega_q = \sum_{j=1}^{n} \sin j\omega_p \sin j\omega_q = 0,$$

and this implies that in the matrix representation of the model (8.3), $S'S$ is again $n/2$ times the unit matrix, so the estimates of α_k and β_k are still

$$\begin{pmatrix} \hat{\alpha}_k \\ \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \frac{2}{n} \sum_{j=1}^{n} X_j \cos j\omega_k \\ \frac{2}{n} \sum_{j=1}^{n} X_j \sin j\omega_k. \end{pmatrix}$$

The sum of squares due to regression on $\cos \omega_k t$ and $\sin \omega_k t$ is similarly still

$$\text{Regression SS}(\omega_k) = n(\hat{\alpha}_k^2 + \hat{\beta}_k^2)/2,$$

and the diagonal nature of $S'S$ implies that the overall regression sum of squares for the whole model (8.3) is the sum of the m individual regression sums of squares associated with the m sinusoidal components. As m is taken larger this procedure partitions the variability in the X_t sequence into amounts associated with all of the Fourier frequencies.

Periodogram

The **periodogram ordinate at frequency** ω_q is defined to be

$$\begin{aligned} I(\omega_q) &= n(\hat{\alpha}_q^2 + \hat{\beta}_q^2)/4 \\ &= \frac{1}{n}\left[\left(\sum_{1}^{n} X_j \cos j\omega_q\right)^2 + \left(\sum_{1}^{n} X_j \sin j\omega_q\right)^2\right] \end{aligned}$$

for $q = 0, \ldots, [n/2]$, (ie half the regression sum of squares).

A plot of $I(\omega_q)$ vs $\omega_q > 0$ is called the **periodogram** of $\{X_t\}$.
(The value for $\omega_q = 0$ is usually omitted since it corresponds to variation at frequency zero – that is, no variation at all.)

If model (8.1) really holds (that is, if variation is at exact frequency ω_p for a particular value of p) then:

$$E(X_t) = \alpha \cos t\omega_p + \beta \sin t\omega_p,$$

so that

$$E(\hat{\alpha}_p) = \alpha, \qquad E(\hat{\beta}_p) = \beta$$

and

$$E(\hat{\alpha}_q) = E(\hat{\beta}_q) = 0 \text{ for } q \neq p$$

whilst

$$\text{Var}(\hat{\alpha}_q) = \text{Var}(\hat{\beta}_q) = \frac{2\sigma^2}{n} \text{ for all } q$$

whether $q = p$ or not.

Thus

$$E(I(\omega_q)) = \begin{cases} \sigma^2 + n(\alpha^2 + \beta^2)/4 & \text{for } q = p \\ \sigma^2 & \text{for } q \neq p. \end{cases}$$

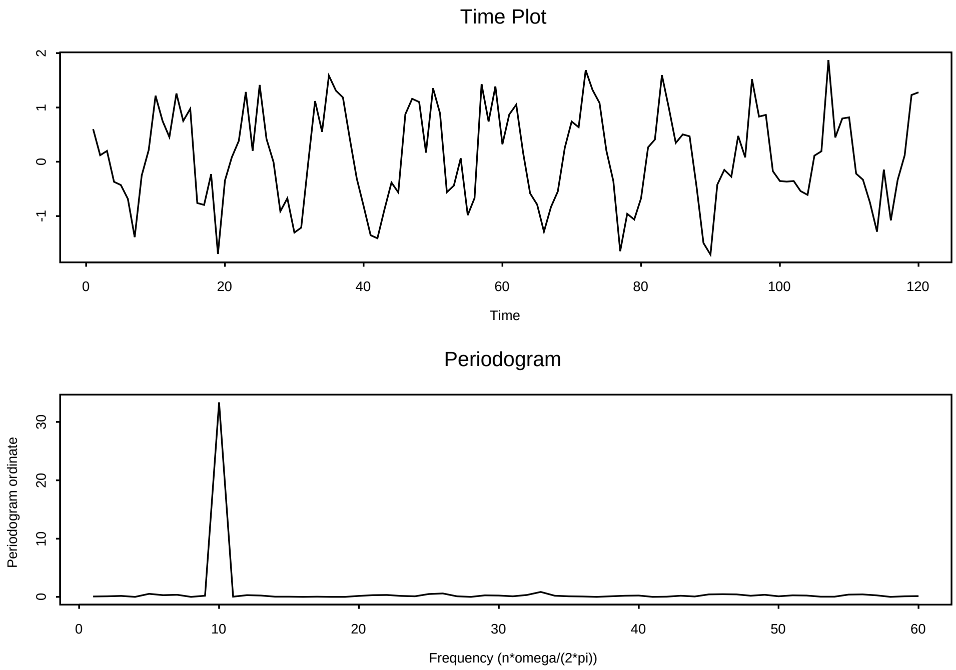
The periodogram is therefore expected to show a spike at ω_p.

Example 1: A Single Sinusoid with Noise

For 120 observations simulated from

$$X_t = \cos(2\pi t/12) + \epsilon_t$$

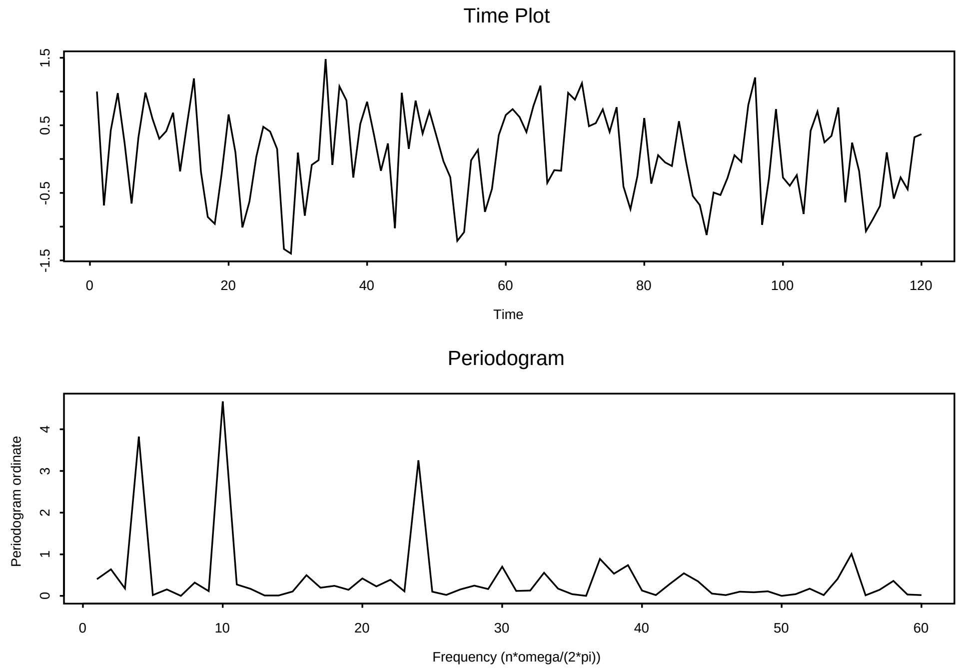
where ϵ_t is WN$(0, 0.25)$ we get the plots:



More generally, for a model with several sinusoidal components we would expect to see several spikes in the periodogram, one at each of the component frequencies.
This was the basis for methods dating from the early 1900's for detecting hidden periodicities in data.

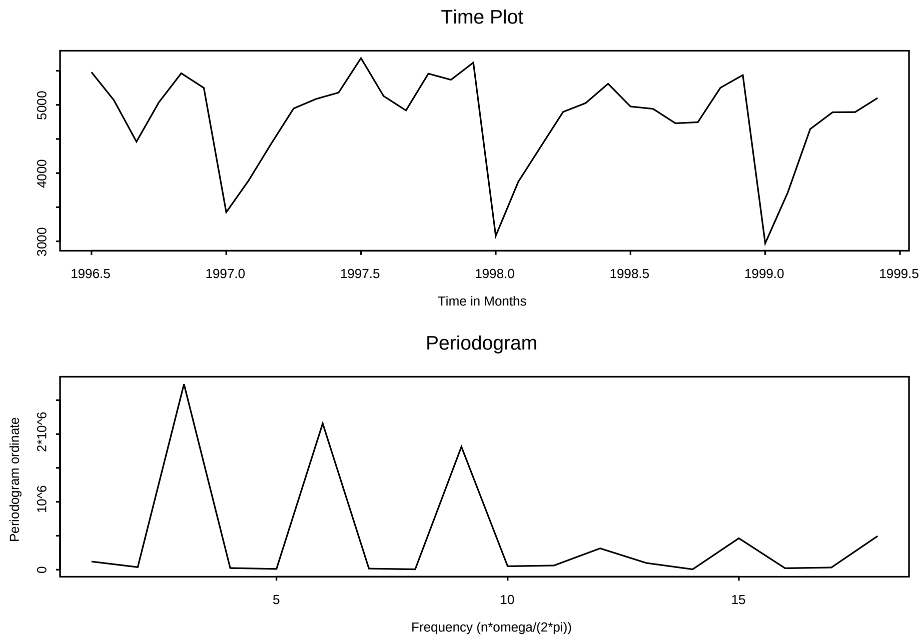
Example 2: A Triple Sinusoid with Noise

The plots below are for 120 observations simulated from a model similar to that in Example 1 except that now there are three trigonometric terms with periods 12, 5 and 30 and amplitudes 0.4, 0.3 and 0.3 respectively.



Example 3: UK Beer Consumption

UK beer consumption 1997–1999 (in thousands of hectolitres) shows clear seasonality:



Notes

(a) The periodogram has the property that the sum of periodogram ordinates from, say, ω_{p_1} to ω_{p_2} represents the amount of variation in the series $\{X_t\}$ associated with fluctuations of frequency between ω_{p_1} and ω_{p_2}.

(b) Tests for WN can be based on the periodogram. One is based on the *cumulative periodogram*, which is defined as follows. Let $C_k = \sum_{j=1}^{k} I(\omega_j)$ for $k = 1, \ldots, (n/2)$, where $(n/2)$ is the largest integer strictly less than $n/2$, and set $U_k = C_k/C_m$. A plot of U_k vs $k/((n/2)-1)$ is called the cumulative periodogram. If the data were actually a realization of a white noise sequence then the amplitudes of the sinusoidal components in (8.3) would all be zero, so the periodogram ordinates should differ only because of sampling fluctuation, and the cumulative periodogram should increase approximately linearly. A statistic measuring departure from linearity will therefore give a test of the WN hypothesis. One such is the maximum horizontal distance D of the U_k sequence from the 45^o line joining the origin to (1,1). Approximate 5% critical values of D, found by Stephens (1974), are given by

$$1.358/(n_1^{1/2} + 0.12 + 0.11 n_1^{-1/2})$$

where $n_1 = (n/2) - 1$.

(c) There is a close connection between the periodogram and the correlogram. In fact trigonometrical manipulations show that

$$I(\omega_p) \propto 1 + 2 \sum_{k=1}^{n-1} r_k \cos k\omega_p. \qquad (8.4)$$

This and (a) are reasons for the wider importance of I, going much beyond its use in tests for hidden periodicities.

8.3 The Spectral Representation of a Stationary Process

The regression models above are not stationary. Now consider probability models for a *weakly stationary* X_t which follow up the idea of representing the variability of X_t in terms of trigonometric variation of different frequencies.

Because of the relation between sine and cosine functions and complex exponentials, and the ease of manipulating the latter, it is mathematically convenient to discuss these models in terms of complex valued random variables $X_t = X_t^{(1)} + iX_t^{(2)}$. The definitions of E and Cov are extended for complex random variables to:

$$E(X) = E(X^{(1)}) + iE(X^{(2)})$$

and,

$$\begin{aligned} \text{Cov}(X, U) &= E(X - E(X))(\bar{U} - E(\bar{U})) \\ &= E\left\{X^{(1)} - EX^{(1)} + i(X^{(2)} - EX^{(2)})\right\}\left\{U^{(1)} - EU^{(1)} - i(U^{(2)} - EU^{(2)})\right\}, \end{aligned}$$

so in particular

$$\text{Var}(X) = \text{Cov}(X, X) = E|X|^2 \geq 0$$

is real.

If $\text{Cov}(X, U) = 0$, say X and U are *orthogonal.*

Some models.

Model 1: A Single Sinusoid

Let

$$X_t = Se^{i\omega t},$$

where $0 \leq \omega \leq \pi$ is a fixed frequency and S is a random variable with

$$E(S) = 0, \quad \text{Var}(S) = g.$$

It is easily verified that this is a weakly stationary process with auto-covariance function

$$\gamma_h = E(X_{t+h}X_t) = ge^{i\omega h},$$

and in particular variance $\gamma_0 = g$.

Model 2: Two Sinusoids

Let

$$X_t = S_1 e^{i\omega_1 t} + S_2 e^{i\omega_2 t},$$

where ω_1 and ω_2 are fixed frequencies between 0 and π, not connected with the Fourier frequencies, and S_1, S_2 are random variables with

$$E(S_i) = 0, \quad \text{Var}(S_i) = g_i, \quad i = 1, 2.$$

For this model to be weakly stationary it is necessary that

$$E(S_1 \bar{S_2}) = 0,$$

that is, S_1 and S_2 should be orthogonal. In that case

$$\text{Var}(X_t) = g_1 + g_2,$$

and

$$\gamma_h = g_1 e^{i\omega_1 h} + g_2 e^{i\omega_2 h}.$$

Model 3: Multiple Sinusoids

Let

$$X_t = \sum_j S_j e^{i\omega_j t}, \tag{8.5}$$

where again the ω_j are fixed frequencies between 0 and π, not connected to the Fourier frequencies, and the S_j have

$$E(S_j) = 0, \quad \text{Var}(S_j) = g_j.$$

For weak stationary it is found again that the S_j should be orthogonal, and in that case

$$\text{Var}(X_t) = \sum_j g_j,$$

and

$$\gamma_h = \sum_j g_j e^{i\omega_j h}.$$

If the ωs are arranged in increasing order, then the variability in X_t associated with frequencies $\in [\omega_k, \omega_{k+l}]$, that is, $\text{Var}(\sum_k^{k+l} S_j e^{i\omega_j t})$ is

$$\text{Var}(\sum_k^{k+l} S_j e^{i\omega_j t}) = \text{Var}(\sum_k^{k+l} S_j) = \sum_k^{k+l} g_j.$$

General Case

Imagine that the number of terms $\to \infty$ and the ω_js become more dense in $[0, \pi]$. The natural way to generalize (8.5) is to replace the sum by an integral and let

$$X_t = \int_0^\pi e^{i\omega t} dS(\omega). \tag{8.6}$$

The integral can be defined rigorously via the idea of convergence in quadratic mean: see references.

Interpretation of (8.6)

$\int_0^{\omega'} dS(\omega) = S(\omega') - S(0)$ corresponds to $S_1 + S_2 + \ldots + S_l$, if $\omega' = \omega_l$, so S is a stochastic process got by adding terms which are orthogonal – *an additive process with orthogonal increments.* Only changes in $S(\omega)$ matter in (8.6), so we can if we wish take $S(0) = 0$. As in *Model 3*

$$E(\int_{\omega_1}^{\omega_2} dS(\omega)) = 0;$$

and for weak stationarity we need also, as before,

$$\text{Cov}(\int_{\omega_1}^{\omega_2} dS(\omega), \int_{\omega_3}^{\omega_4} dS(\omega)) = 0,$$

when the intervals (ω_1, ω_2) and (ω_3, ω_4) don't overlap.
In this case

$$\text{Var}(\int_{\omega_1}^{\omega_2} dS(\omega)) = \int_{\omega_1}^{\omega_2} dG(\omega) = G(\omega_2) - G(\omega_1)$$

where G is a real increasing function corresponding to $g_1 + g_2 + \ldots + g_{\omega_l}$ in *Model 3.*

Thus the part of the variability in X_t resulting from sinusoids with frequencies in the range $(\omega_1, \omega_2]$ is given by $G(\omega_2) - G(\omega_1)$ and the overall variance of X_t is $\sigma_X^2 = \text{Var}(X_t) = G(\pi) - G(0) = G(\pi)$ if we set $G(0) = 0$.

(In concise notation:

$$\begin{aligned} E(dS(\omega)) &= 0, \\ E(dS(\omega)d\bar{S}(\omega')) &= 0, \quad \omega \neq \omega', \\ E(|dS(\omega)|^2) &= dG(\omega).) \end{aligned}$$

The autocovariance function of (8.6) is

$$\gamma_h = E(X_{t+h}\bar{X}_t) = \int_0^\pi e^{i\omega h} dG(\omega).$$

G is called the *power spectral distribution function* of X_t. It is a non-decreasing function on $[0, \pi]$ with $G(0) = 0, G(\pi) = \text{Var}(X) = \sigma_X^2$.
(The power of an alternating electrical current is proportional to the square of the current's amplitude. Thus G specifies the total power in X_t and how it is distributed over different frequencies. Power in this context is identified with variance.)

Define the *spectral distribution function* of X_t to be the function

$$F(\omega) = G(\omega)/\sigma_X^2.$$

Then F is a *probability* distribution function on $[0, \pi]$, and the acf of X may be written as

$$\rho_h = \int_0^\pi e^{i\omega h} dF(\omega), \quad h = 0, 1, \ldots. \tag{8.7}$$

Though the model (8.6) was introduced here as a heuristic extension of *Models 1–3*, the remarkable thing is that *every weakly stationary sequence* $\{X_t\}$ *has a representation of the form (8.6)*, in the sense that there exists a distribution function F_X on $[0, 1]$ and an additive process S with orthogonal increments having zero means and variances given by $\text{Var}(S(\omega_2) - S(\omega_1)) = \sigma^2(F_X(\omega_2) - F_X(\omega_1))$ for some constant σ, for which (8.6) holds. Moreover the autocorrelation function of X_t is given by (8.7).

The spectral distribution function may have a density. When it does the density is called the *spectral density function of* X_t, or simply *the spectrum of* X_t. If this density is denoted by f then (8.7) becomes

$$\rho_h = \int_0^\pi e^{i\omega h} f(\omega) d\omega, \quad h = 0, 1, \ldots$$

that is, for a real process,

$$\rho_h = \int_0^\pi \cos(h\omega) f(\omega) d\omega, \qquad h = 0, 1, \ldots. \tag{8.8}$$

showing that the acf $\{\rho_h\}$ is simply the Fourier transform of the spectrum f (cf the characteristic function of a probability distribution).

Since Fourier transforms may be inverted, it is simple in principle to pass from ρ_h to $f(\omega)$ too: the inversion of (8.8) is

$$f(\omega) = \frac{1}{\pi}\left\{1 + 2\sum_{k=1}^{\infty} \rho_k \cos(k\omega)\right\} \quad 0 \le \omega \le \pi. \tag{8.9}$$

Formula (8.9) is useful for finding the spectrum of specific models for weakly stationary processes.

Examples

1: White Noise

If X_t is WN$(0, \sigma^2)$ then $\rho_k = 1$ for $k = 0$ and 0 for $k > 0$, so from (8.9)

$$f(\omega) = 1/\pi, \qquad 0 \le \omega \le \pi.$$

2: AR(1) Process

When $X_t = \alpha X_{t-1} + \epsilon_t$ then $\rho_k = \alpha^k$ for $k \ge 0$, so

$$f(\omega) = \frac{1}{\pi}\left\{1 + 2\sum_{1}^{\infty} \alpha^k \cos(k\omega)\right\},$$

which after some manipulation reduces to

$$f(\omega) = \frac{1}{\pi}\,\frac{1-\alpha^2}{1 - 2\alpha\cos(\omega) + \alpha^2}, \qquad 0 \le \omega \le \pi.$$

8.4 Spectral Estimation

Comparison of (8.9) and the expression (8.4) for the periodogram in Note (c) of §8.2,

$$I(\omega_p) \propto 1 + 2\sum_{k=1}^{n-1} r_k \cos k\omega_p,$$

reveals a close similarity. In fact the periodogram is an empirical analogue of the spectrum. It is natural therefore to expect to be able to use I as an estimator of f. However it turns out that although I is unbiassed for f, it is not a consistent estimator because its variance does not converge to zero as sample size increases. Various forms of smoothing of I are needed to obtain practically useful spectral estimators. For details see references.