show that $\hat{x}(N, k)=\alpha^k x_N$ for $k=1, 2, \ldots$. Also show that the variance of the k-steps-ahead forecast error is given by $(1-\alpha^{2k})\sigma_Z^2/(1-\alpha^2)$.

For the AR(1) model given by

$$X_t-\mu=\alpha(X_{t-1}-\mu)+Z_t$$

show that $\hat{x}(N, k)=\mu+\alpha^k(x_N-\mu)$ for $k=1, 2, \ldots$. (In practice the least squares estimate of α would be substituted into the above formulae.)

5.3 Consider the SARIMA$(1, 0, 0)(0, 1, 1)_{12}$ model used as an example in Section 5.2.4. Show that

$$\hat{x}(N, 2)=x_{N-10}+\alpha^2(x_N-x_{N-12})+\theta\alpha z_{N-11}+\theta z_{N-10}$$

5.4 For the SARIMA$(0, 0, 1)(1, 1, 0)_{12}$ model, find forecasts at time N for up to 12 steps ahead in terms of observations and estimated residuals up to time N.

5 For the model $(1-B)(1-0.2B)X_t=(1-0.5B)Z_t$ in Exercise 3.12, find forecasts for one and two steps ahead, and show that a recursive expression for forecasts three or more steps ahead is given by

$$\hat{x}(N, k)=1.2\hat{x}(N, k-1)-0.2\hat{x}(N, k-2)$$

Find the variance of the one-, two- and three-steps-ahead forecast errors. If $z_N=1$, $x_N=4$, $x_{N-1}=3$ and $\sigma_Z^2=2$, show that $\hat{x}(N, 2)=3.64$ and that the standard error of the corresponding forecast error is 1.72.

6 Consider the ARIMA$(0, 1, 1)$ process

$$(1-B)X_t=(1-\theta B)Z_t$$

Show that $\hat{x}(N, 1)=x_N-\theta z_N$, and $\hat{x}(N, k)=\hat{x}(N, k-1)$ for $k\geqslant 2$. Express $\hat{x}(N, 1)$ in terms of x_N and $\hat{x}(N-1, 1)$ and show that this is equivalent to exponential smoothing. By considering the ψ weights of the process, show that the variance of the k-steps-ahead prediction error is $[1+(k-1)(1-\theta)^2]\sigma_Z^2$.

6

Stationary processes in the frequency domain

6.1 INTRODUCTION

In Chapter 3 we described several types of stationary stochastic process, placing emphasis on the autocovariance (or autocorrelation) function which is the natural tool for considering the evolution of a process through time. In this chapter we introduce a complementary function called the **spectral density function**, which is the natural tool for considering the frequency properties of a time series. Inference regarding the spectral density function is called an analysis in the **frequency domain**.

Some statisticians initially have difficulty in understanding the frequency approach, but the advantages of frequency methods are widely appreciated in such fields as electrical engineering, geophysics and meteorology. These advantages will become apparent in the next few chapters.

We shall confine ourselves to real-valued processes. Many authors consider the more general problem of complex-valued processes, and this results in some gain of mathematical conciseness. But, in my view, the reader is more likely to understand an approach restricted to real-valued processes. The vast majority of practical problems are covered by this approach.

6.2 THE SPECTRAL DISTRIBUTION FUNCTION

In order to introduce the idea of a spectral density function, we must first consider a function called the **spectral distribution function**. The approach adopted is heuristic and not mathematically rigorous, but will, hopefully, give the reader a better understanding of the subject than a more theoretical approach.

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

$$X_t = R\cos(\omega t + \theta) + Z_t \tag{6.1}$$

where ω is called the **frequency** of the sinusoidal variation, R is called the **amplitude** of the variation, θ is called the **phase**, and Z_t denotes some stationary

random series. Note that the angle $(\omega t + \theta)$ is usually measured in units called radians, where π radians $= 180°$. Since ω is the number of radians per unit time it is sometimes called the **angular** frequency, but in keeping with most authors we call ω the frequency. However some authors, notably Jenkins and Watts (1968), refer to frequency as $f = \omega/2\pi$, the number of cycles per unit time, and this form of frequency is much easier to interpret from a physical point of view. We usually use the angular frequency ω in mathematical formulae for conciseness, but will often use the frequency $f = \omega/2\pi$ for the interpretation of data. The period of a sinusoidal cycle, called the **wavelength**, is clearly $1/f$ or $2\pi/\omega$. An example of a sinusoidal function is shown in Figure 6.1. There $f = 1/6$ and the wavelength is 6.

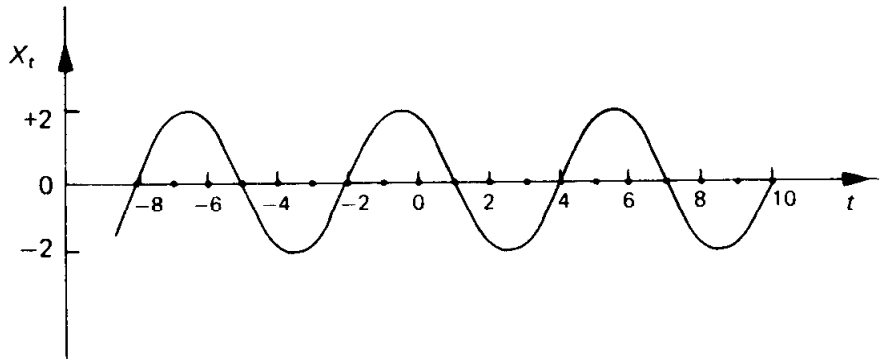

Figure 6.1 A graph of $R\cos(\omega t + \theta)$ with $R = 2$, $\omega = \pi/3$ and $\theta = \pi/6$.

Model (6.1) is a very simple model, but 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 cyclical variation. In other words the data show variation at high, medium and low frequencies. It is natural therefore to generalize (6.1) to

$$X_t = \sum_{j=1}^{k} R_j \cos(\omega_j t + \theta_j) + Z_t \tag{6.2}$$

where R_j is the amplitude at frequency ω_j.

The reader will notice that models (6.1) and (6.2) are **not** stationary if R, θ, $\{R_j\}$ and $\{\theta_j\}$ are fixed constants because $E(X_t)$ will change with time. In order to apply the theory of stationary processes to models like (6.2), it is customary to assume that $\{R_j\}$ are (uncorrelated) random variables with mean zero, or that $\{\theta_j\}$ are random variables with a uniform distribution on $(0, 2\pi)$, which are fixed for a single realization of the process (see Section 3.5 and Exercise 3.14). This is something of a 'mathematical trick', but it does enable us to treat time series containing one or more deterministic sinusoidal components as stationary series.

Since $\cos(\omega t+\theta)=\cos\omega t\cos\theta-\sin\omega t\sin\theta$, model (6.2) can be expressed as a sum of sine and cosine terms in the form

$$X_t=\sum_{j=1}^{k}(a_j\cos\omega_j t+b_j\sin\omega_j t)+Z_t \tag{6.3}$$

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

But we may now ask why there should only be a finite number of frequencies involved in model (6.2) or (6.3). In fact, letting $k\to\infty$, the work of Wiener and others has shown that any discrete stationary process measured at unit intervals may be represented in the form

$$X_t=\int_0^{\pi}\cos\omega t\,\mathrm{d}u(\omega)+\int_0^{\pi}\sin\omega t\,\mathrm{d}v(\omega) \tag{6.4}$$

where $u(\omega)$, $v(\omega)$ are uncorrelated continuous processes with orthogonal increments (see Section 3.4.8) which are defined for all ω in the range $(0, \pi)$. Equation (6.4) is called the **spectral representation** of the process; it involves stochastic integrals, which require considerable mathematical skill to handle properly. It is intuitively more helpful to ignore these mathematical problems and simply regard X_t as a linear combination of orthogonal sinusoidal terms. Thus the derivation of the spectral representation will not be considered here (see for example Cox and Miller, 1968, Chapter 8).

The reader may wonder why the upper limits of the integrals in (6.4) are π rather than ∞. For a continuous process the upper limits would indeed be ∞, but for a discrete process measured at unit intervals of time there is no loss of generality in restricting ω to the range $(0, \pi)$, since

$$\cos[(\omega+k\pi)t]=\begin{cases}\cos\omega t & k,\ t \text{ integers with } k \text{ even}\\ \cos(\pi-\omega)t & k,\ t \text{ integers with } k \text{ odd}\end{cases}$$

and so variation at frequencies higher than π cannot be distinguished from variation at a corresponding frequency in $(0, \pi)$. The frequency $\omega=\pi$ is called the **Nyquist frequency**. We will say more about this in Section 7.2.1. For a discrete process measured at equal intervals of time of length Δt, the Nyquist frequency is $\pi/\Delta t$. In the next two sections we consider discrete processes measured at unit intervals of time, but the arguments carry over to discrete processes measured at intervals Δt if we replace π by $\pi/\Delta t$.

The main point of introducing the spectral representation (6.4) 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 $v(\omega)$ in (6.4) are of little direct practical interest. Instead we introduce a function $F(\omega)$ called the (power) spectral distribution function, which arises from a theorem (e.g. Bartlett, 1966, Section 6.1), called the Wiener-Khintchine theorem, named after N. Wiener and A. Y. Khintchine. As applied to real-valued processes, this theorem says

that, for any stationary stochastic process with autocovariance function $\gamma(k)$, there exists a monotonically increasing function $F(\omega)$ such that

$$\gamma(k)=\int_0^{\pi} \cos \omega k \, \mathrm{d}F(\omega) \tag{6.5}$$

Equation (6.5) is called the spectral representation of the autocovariance function, and involves a type of integral (called Stieltjes) which may be unfamiliar to some readers. It can however be shown that the function $F(\omega)$ has a direct physical interpretation: it is the contribution to the variance of the series which is accounted for by frequencies in the range $(0, \omega)$. It is most important to understand this physical interpretation of $F(\omega)$. There is no variation at negative frequencies, so that

$$F(\omega)=0 \qquad \text{for } \omega<0$$

For a discrete process measured at unit intervals of time, the highest possible frequency is π and so all the variation is accounted for by frequencies less than π. Thus

$$F(\pi)=\mathrm{Var}(X_t)=\sigma_X^2$$

This last result also comes directly from (6.5) with $k=0$, when

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

In between $\omega=0$ and $\omega=\pi$, $F(\omega)$ is monotonically increasing.

If the process contains a deterministic sinusoidal component at frequency ω_0, say $R\cos(\omega_0 t+\theta)$ where R is a constant and θ is uniformly distributed on $(0, 2\pi)$, then there will be a step increase in $F(\omega)$ at ω_0 equal to $E[R^2\cos^2(\omega_0 t+\theta)]=\frac{1}{2}R^2$.

As $F(\omega)$ is monotonic, it can be decomposed into two functions, $F_1(\omega)$ and $F_2(\omega)$, such that

$$F(\omega)=F_1(\omega)+F_2(\omega) \tag{6.6}$$

where $F_1(\omega)$ is a non-decreasing continuous function and $F_2(\omega)$ is a non-decreasing step function. This decomposition usually corresponds to the Wold decomposition, with $F_1(\omega)$ relating to the purely indeterministic component of the process and $F_2(\omega)$ relating to the deterministic component. We shall be mainly concerned with purely indeterministic processes, where $F_2(\omega)\equiv 0$, so that $F(\omega)$ is a continuous function on $(0, \pi)$.

The adjective 'power', which is sometimes prefixed to 'spectral distribution function', derives from the engineer's use of the word in connection with the passage of an electric current through a resistance. For a sinusoidal input, the power is directly proportional to the squared amplitude of the oscillation. For a more general input, the power spectral distribution function describes how

the power is distributed with respect to frequency. In the case of a time series, the variance may be regarded as the total power.

Note that some authors use a normalized form of $F(\omega)$ given by

$$F^*(\omega) = F(\omega)/\sigma_X^2 \tag{6.7}$$

Thus $F^*(\omega)$ is the **proportion** of variance accounted for by frequencies in the range $(0, \omega)$. Since $F^*(\pi) = 1$, and $F^*(\omega)$ is monotonically increasing, $F^*(\omega)$ has similar properties to a cumulative distribution function.

6.3 THE SPECTRAL DENSITY FUNCTION

For a purely indeterministic discrete stationary process, the spectral distribution function is a continuous (monotone bounded) function in $(0, \pi)$, and may therefore be differentiated with respect to ω in $(0, \pi)$. (Strictly speaking, $F(\omega)$ may not be differentiable on a set of measure zero, but this is of no practical importance.) We will denote the derivative by $f(\omega)$, so that

$$f(\omega) = \frac{\mathrm{d}F(\omega)}{\mathrm{d}\omega} \tag{6.8}$$

This is the (power) spectral density function. The term 'spectral density function' is often shortened to **spectrum**, and the adjective 'power' is sometimes omitted.

When $f(\omega)$ exists, equation (6.5) can be expressed in the form

$$\gamma(k) = \int_0^\pi \cos \omega k f(\omega)\, \mathrm{d}\omega \tag{6.9}$$

This is an ordinary (Riemann) integral and therefore much easier to handle. Putting $k = 0$, we have

$$\gamma(0) = \sigma_X^2 = \int_0^\pi f(\omega)\, \mathrm{d}\omega = F(\pi) \tag{6.10}$$

The physical meaning of the spectrum is that $f(\omega)\, \mathrm{d}\omega$ represents the contribution to variance of components with frequencies in the range $(\omega, \omega + \mathrm{d}\omega)$. When the spectrum is drawn, equation (6.10) indicates that the total area underneath the curve is equal to the variance of the process. A peak in the spectrum indicates an important contribution to variance at frequencies in the appropriate interval. An example of a spectrum is shown in Figure 6.2, together with the corresponding normalized spectral distribution function.

It is important to realize that the autocovariance function (acv.f.) and the power spectral density function are equivalent ways of describing a stationary stochastic process. From a practical point of view, they are complementary to each other. Both functions contain the same information but express it in

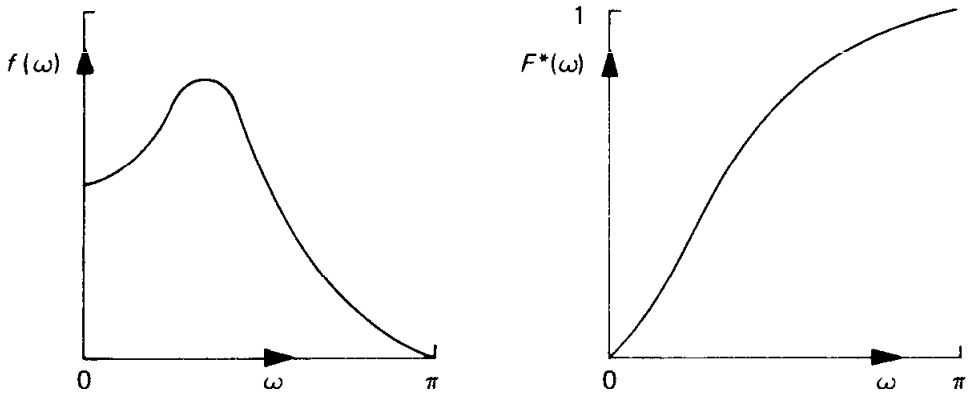

Figure 6.2 An example of a spectrum, together with the corresponding normalized spectral distribution function.

different ways. In some situations a time-domain approach based on the acv.f. is more useful, while in other situations a frequency-domain approach is preferable.

Equation (6.9) expresses $\gamma(k)$ in terms of $f(\omega)$ as a cosine transform. The inverse relationship (see Appendix A) is given by

$$f(\omega) = \frac{1}{\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} \tag{6.11}$$

so that the spectrum is the **Fourier transform** of the autocovariance function. Since $\gamma(k)$ is an even function, (6.11) is often written in the equivalent form

$$f(\omega) = \frac{1}{\pi}\left[\gamma(0) + 2\sum_{k=1}^{\infty} \gamma(k) \cos \omega k\right] \tag{6.12}$$

Note that if we try to apply (6.12) to a process containing a deterministic component at frequency ω_0, then $\Sigma\gamma(k) \cos \omega_0 k$ will not converge, since $F(\omega)$ is not differentiable at ω_0 and so $f(\omega_0)$ is not defined.

The reader should note that several other definitions of the spectrum are given in the literature, most of which differ from (6.12) by a constant multiple and by the range of definition of $f(\omega)$. The most popular approach is to define the spectrum in the range $(-\pi, \pi)$ by

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} \tag{6.13}$$

whose inverse relationship (see Appendix A) is

$$\gamma(k) = \int_{-\pi}^{\pi} e^{i\omega k} f(\omega)\, d\omega \tag{6.14}$$

Jenkins and Watts (1968) use these equations, except that they take $f = \omega/2\pi$ as

the frequency variable (see equations (A.3) and (A.4)). Equations (6.13) and (6.14), which form a Fourier transform pair, are the more usual form of the Wiener-Khintchine relations. The formulation is slightly more general in that it can be applied to complex-valued time series. But for real time series we find that $f(\omega)$ is an even function, and then we need only consider $f(\omega)$ for $\omega > 0$. In my experience the introduction of negative frequencies, while having certain mathematical advantages, serves only to confuse the student. As we are concerned only with real-valued processes, we prefer (6.11) defined on $(0, \pi)$.

It is sometimes useful to use a normalized form of the spectral density function, given by

$$f^*(\omega) = f(\omega)/\sigma_X^2 = \frac{\mathrm{d}F^*(\omega)}{\mathrm{d}\omega} \tag{6.15}$$

This is the derivative of the normalized spectral distribution function (see equation (6.7)). Then we find that $f^*(\omega)$ is the Fourier transform of the **autocorrelation** function, namely

$$f^*(\omega) = \frac{1}{\pi}\left[1 + 2\sum_{k=1}^{\infty} \rho(k)\cos \omega k\right] \tag{6.16}$$

and that $f^*(\omega)$ is the **proportion** of variance in the interval $(\omega, \omega + \mathrm{d}\omega)$. Kendall, Stuart and Ord (1983, equation 47.20) define the spectral density function in the range $(0, \pi)$ in terms of the autocorrelation function but omit the constant $1/\pi$ from equation (6.16). This makes it more difficult to give the function a physical interpretation. Instead they introduce an intensity function which corresponds to our power spectrum.

6.4 THE SPECTRUM OF A CONTINUOUS PROCESS

For a continuous purely indeterministic stationary process $X(t)$, the autocovariance function $\gamma(\tau)$ is defined for all τ and the (power) spectral density function $f(\omega)$ is defined for all positive ω. The relationship between these functions is very similar to that in the discrete case except that there is no upper bound to the frequency. We have

$$\begin{aligned} f(\omega) &= \frac{1}{\pi}\int_{-\infty}^{\infty} \gamma(\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\,\mathrm{d}\tau \\ &= \frac{2}{\pi}\int_{0}^{\infty} \gamma(\tau)\cos \omega\tau\,\mathrm{d}\tau \end{aligned} \tag{6.17}$$

for $0 < \omega < \infty$, with the inverse relationship

$$\gamma(\tau) = \int_0^{\infty} f(\omega)\cos \omega\tau\,\mathrm{d}\omega \tag{6.18}$$