1

Introduction

1.1 Periodic functions

We encounter periodic phenomena every day of our lives. Those of us who still use analogue clocks are acutely aware of the 60 second, 60 minute and 12 hour periods associated with the sweeps of the second, minute and hour hands. We are conscious of the fact that the Earth rotates on its axis roughly every 24 hours and that it completes a revolution of the Sun roughly every 365 days. These periodicities are reasonably accurate. The quantities we are interested in measuring are not precisely periodic and there will also be error associated with their measurement. Indeed, some phenomena only *seem* periodic. For example, some biological population sizes appear to fluctuate regularly over a long period of time, but it is hard to justify using common sense any periodicity other than that associated with the annual cycle. It has been argued in the past that some cycles occur because of predator-prey interaction, while in other cases there is no obvious reason. On the other hand, the sound associated with musical instruments can reasonably be thought of as periodic, locally in time, since musical notes are produced by regular vibration and propagated through the air via the regular compression and expansion of the air. The 'signal' will not be *exactly* periodic, since there are errors associated with the production of the sound, with its transmission through the air (since the air is not a uniform medium) and because the ear is not a perfect receiver. The word 'periodic' will therefore be associated with models for phenomena which are approximately periodic, or which would be periodic under perfect conditions. While the techniques discussed in this book can be applied to *any* time series, it should be understood that they have been derived for random sequences constructed from regularly sampled periodic functions and additive noise.

The simplest periodic function that models nature is the sinusoid. Consider a particle which rotates in a plane around some origin at a uniform speed of λ radians per second. The y-coordinate of the particle at time t (secs) is then of the form

$$y(t) = A\cos(\lambda t + \phi), \tag{1.1}$$

where A is the (constant) distance between the origin and the particle and ϕ is another constant which indicates the relative position of the cycle at time 0. The graph of $y(t)$ for $\lambda = 2\pi\frac{15}{512}$ and $\phi = 0$ is shown in Figure 1.1. Where a frequency is quoted, it will be understood from now on that it is measured in radians per unit time, unless otherwise stated.

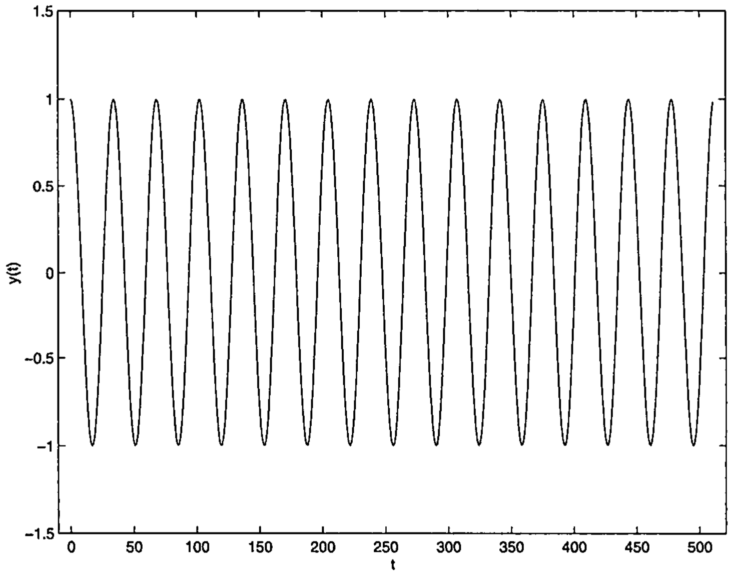

Fig. 1.1. Sine wave, frequency 0.1841

The period of $y(t)$ is the shortest time taken for $y(t)$ to repeat itself, and this is obviously $2\pi/\lambda$. This function is also the general solution of the differential equation which describes simple harmonic motion

$$\ddot{y}(t) = -\lambda^2 y(t).$$

In general, if a well-behaved function $y(t)$ is periodic, with period $2\pi/\lambda$, then it can be written in the form

$$y(t) = c_0 + \sum_{k=1}^{\infty} c_k \cos(k\lambda t) + \sum_{k=1}^{\infty} s_k \sin(k\lambda t),$$

which is called the Fourier expansion of $y(t)$. The coefficients may be calculated using the equations

$$\begin{aligned}\int_0^{2\pi/\lambda} \cos(j\lambda t)\, y(t)\, dt &= c_j \int_0^{2\pi/\lambda} \cos^2(j\lambda t)\, dt \\ &= \begin{cases} c_j \frac{\pi}{\lambda} & ; \quad j \geq 1 \\ c_0 \frac{2\pi}{\lambda} & ; \quad j = 0 \end{cases}\end{aligned}$$

and

$$\begin{aligned}\int_0^{2\pi/\lambda} \sin(j\lambda t)\, y(t)\, dt &= s_j \int_0^{2\pi/\lambda} \sin^2(j\lambda t)\, dt \\ &= s_j \frac{\pi}{\lambda}.\end{aligned}$$

Note that the range of integration may be replaced by any interval of length $2\pi/\lambda$.

The ‘square wave’ function, depicted in Figure 1.2, again for the case where $\lambda = 2\pi\frac{15}{512}$ and $\phi = 0$, is periodic and has a Fourier expansion, even though the latter has many discontinuities.

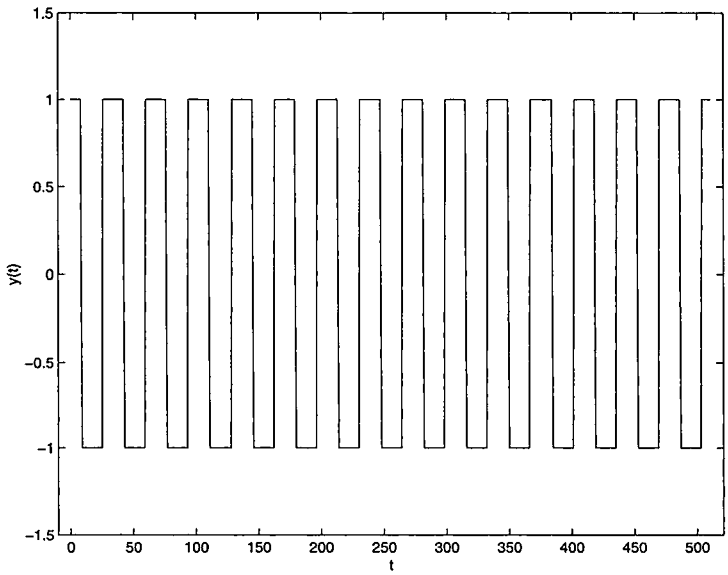

Fig. 1.2. Square wave, frequency 0.1841

Here

$$y(t) = \begin{cases} 1 & ; \quad -\pi/(2\lambda) < t \leq \pi/(2\lambda) \\ -1 & ; \quad \pi/(2\lambda) < t \leq 3\pi/(2\lambda), \end{cases}$$

and the coefficients in the Fourier expansion of $y(t)$ are

$$\begin{aligned} s_j &= \int_{-\pi/(2\lambda)}^{\pi/(2\lambda)} \sin(j\lambda t)\,dt - \int_{\pi/(2\lambda)}^{3\pi/(2\lambda)} \sin(j\lambda t)\,dt \\ &= \frac{\lambda}{j\pi\lambda}\left(\int_{-j\pi/2}^{j\pi/2} \sin x\,dx - \int_{j\pi/2}^{3j\pi/2} \sin x\,dx\right) \\ &= 0, \\ c_0 &= \frac{\lambda}{2\pi}\left(\int_{-\pi/(2\lambda)}^{\pi/(2\lambda)} dt - \int_{\pi/(2\lambda)}^{3\pi/(2\lambda)} dt\right) \\ &= 0, \\ c_{2j} &= \frac{\lambda}{\pi}\left\{\int_{-\pi/(2\lambda)}^{\pi/(2\lambda)} \cos(2j\lambda t)\,dt - \int_{\pi/(2\lambda)}^{3\pi/(2\lambda)} \cos(2j\lambda t)\,dt\right\} \\ &= \frac{\lambda}{2j\pi\lambda}\left(\int_{-j\pi}^{j\pi} \cos x\,dx - \int_{j\pi}^{3j\pi} \cos x\,dx\right) \\ &= 0,\ j \geq 1 \end{aligned}$$

and, for $j \geq 0$,

$$\begin{aligned} c_{2j+1} &= \frac{\lambda}{(2j+1)\pi\lambda}\left(\int_{-j\pi-\pi/2}^{j\pi+\pi/2} \cos x\,dx - \int_{j\pi+\pi/2}^{3j\pi+3\pi/2} \cos x\,dx\right) \\ &= \frac{4(-1)^j}{(2j+1)\pi}. \end{aligned}$$

Thus $y(t)$ may be written in the form

$$y(t) = \sum_{j=0}^{\infty} \frac{4(-1)^j}{(2j+1)\pi} \cos\{(2j+1)\lambda t\}.$$

The expansion, in this extreme case, will be valid only at the points of continuity of $y(t)$. This is most easily seen by noticing that, when $t = \pi/(2\lambda)$, the above expansion gives

$$y\left(\frac{\pi}{2\lambda}\right) = \sum_{j=0}^{\infty} \frac{4(-1)^j\, 0}{(2j+1)\pi} = 0,$$

which is midway between the true values just before and just after $\pi/(2\lambda)$.

Note that the expansion of $y(t)$ has no sine terms and that the even-indexed cosine terms are also missing.

The frequencies and periods of the sine wave and square wave can easily be determined by inspection of the above pictures. However, more often than not, when data is measured, it comes with noise, which can often be assumed to be additive. A model for the received data might therefore be

$$y(t) = A\cos(\lambda t + \phi) + x(t), \tag{1.2}$$

rather than equation (1.1), where $x(t)$ is the noise associated with the measurement at time t. It may then not be possible to tell from the data that $y(t)$ has a periodic deterministic component. For example, the Figures 1.3 and 1.4 depict the noisy sinusoid and square waves corresponding to the previous two figures, with $A = 1$ and noise standard deviation 2.

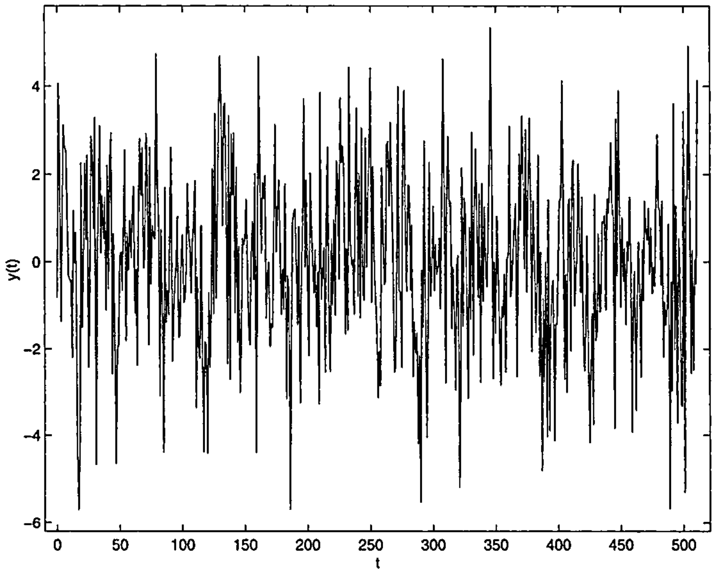

Fig. 1.3. Sine wave + noise, frequency 0.1841

If the deterministic function is more general, but is still periodic, we could consider the model

$$y(t) = \mu + \sum_{j=1}^{\infty} A_j \cos(j\lambda t + \phi_j) + x(t),$$

or a version with only a finite number of terms, since an infinite number

Cambridge University Press
978-0-521-80446-2 - The Estimation and Tracking of Frequency
B. G. Quinn and E. J. Hannan
Excerpt
More information

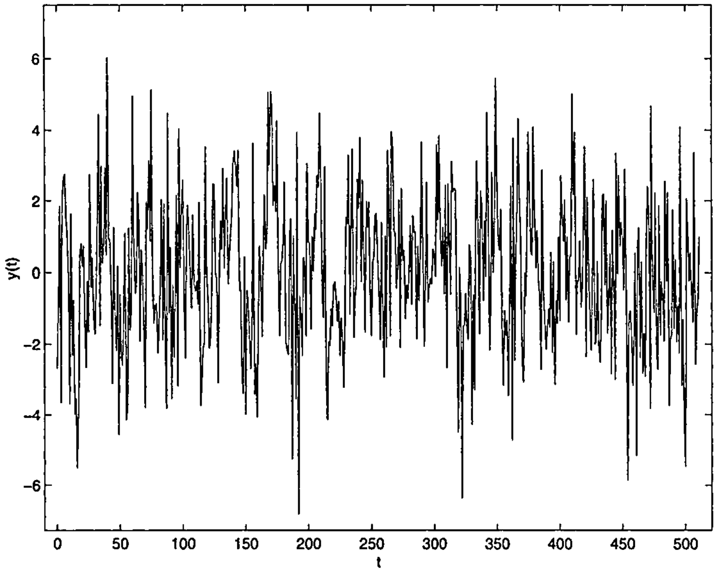

Fig. 1.4. Square wave + noise, frequency 0.1841

of parameters would in practice be impossible to estimate. More generally still, we shall consider in this book the model

$$y(t) = \mu + \sum_{j=1}^{r} A_j \cos(\lambda_j t + \phi_j) + x(t),$$

where r is the number of sinusoidal terms, and the frequencies

$$\{\lambda_j; j = 1, \ldots, r\}$$

may or may not be harmonically related. The term μ is called the overall mean or 'DC' term, and A_j, λ_j and ϕ_j are the amplitude, frequency and (initial) phase of the jth sinusoid. We shall be interested in estimating several (fundamental) frequencies and possibly their harmonics.

Even when there is no noise, it may be difficult to identify the frequencies. For example, if $r = 2$, and the two frequencies λ_1 and λ_2 are close together, the phenomenon known as 'beats' arises. Figure 1.5 was generated using $\lambda_1 = 2\pi 15/512 \sim 0.1841$ and $\lambda_2 = 0.2$. The amplitudes in both cases were 1 and the phases were 0. Although the data seem periodic, there appear to be a frequency near the two, and another much lower frequency associated with the evolution of the 'envelope'. This is because

$$\cos(\lambda_1 t) + \cos(\lambda_2 t) = 2\cos\left(\frac{\lambda_1+\lambda_2}{2}t\right)\cos\left(\frac{\lambda_1-\lambda_2}{2}t\right),$$

which is the product of two sine waves. We can think of the above either as an amplitude-modulated sinusoid with frequency $\frac{\lambda_1+\lambda_2}{2}$, or an amplitude-modulated sinusoid with frequency $\frac{\lambda_1-\lambda_2}{2}$.

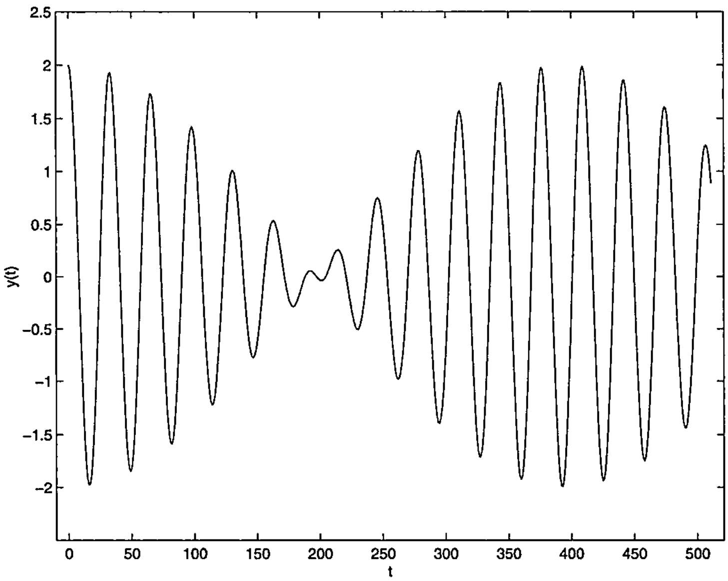

Fig. 1.5. Sum of two sine waves, frequencies 0.1841 and 0.2000

It is possible to estimate the two frequencies by observation, but, again, if there is noise, it may not be possible to tell from a graph that there are any periodicities present. Interestingly enough, it is possible to 'hear' the periodicities when listening to the noisy data depicted in Figure 1.6, even though the two sinusoids cannot be seen.

1.2 Sinusoidal regression and the periodogram

We consider now the simple sinusoidal model but with an additive constant term, and rewritten in the slightly different, but equivalent, form

$$y(t) = \mu + \alpha\cos(\lambda t) + \beta\sin(\lambda t) + x(t).$$

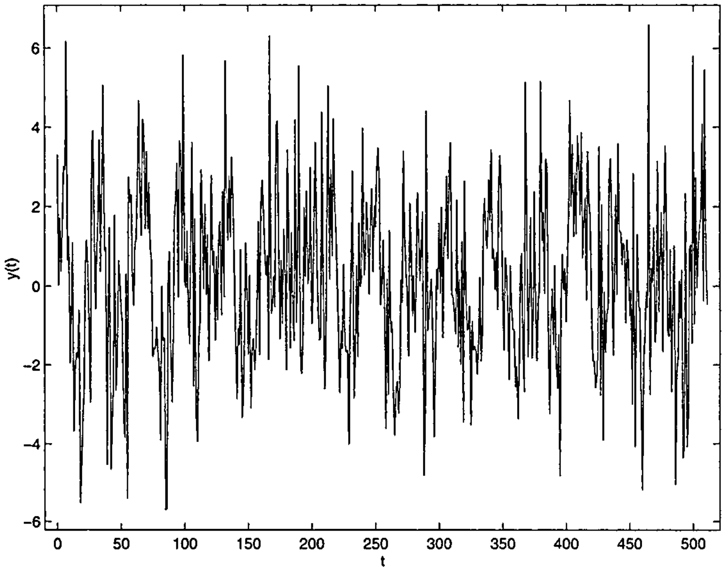

Fig. 1.6. Sum of two sine waves + noise, frequencies 0.1841 and 0.2000

As long as we sample at T equidistant time-points, we may assume without loss of generality that the data are observed at $t = 0, 1, 2, \ldots, T-1$. The model is just a regression model, and for λ fixed, is a *linear* regression model. The least-squares regression estimators of μ, α and β are given by

$$\begin{bmatrix} \widehat{\mu} \\ \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = D^{-1}(\lambda) E(\lambda),$$

where

$$D(\lambda) = \begin{bmatrix} T & \sum_{t=0}^{T-1} \cos(\lambda t) & \sum_{t=0}^{T-1} \sin(\lambda t) \\ \sum_{t=0}^{T-1} \cos(\lambda t) & \sum_{t=0}^{T-1} \cos^2(\lambda t) & \sum_{t=0}^{T-1} \sin(\lambda t)\cos(\lambda t) \\ \sum_{t=0}^{T-1} \sin(\lambda t) & \sum_{t=0}^{T-1} \sin(\lambda t)\cos(\lambda t) & \sum_{t=0}^{T-1} \sin^2(\lambda t) \end{bmatrix}$$

and

$$E(\lambda) = \begin{bmatrix} \sum_{t=0}^{T-1} y(t) \\ \sum_{t=0}^{T-1} y(t)\cos(\lambda t) \\ \sum_{t=0}^{T-1} y(t)\sin(\lambda t) \end{bmatrix}.$$

The residual sum of squares, for fixed λ, is then given by

$$\begin{aligned}\mathrm{SS}(\lambda) &= \sum_{t=0}^{T-1} y^2(t) - E'(\lambda) D^{-1}(\lambda) E(\lambda) \\ &= \sum_{t=0}^{T-1} \{y(t) - \overline{y}\}^2 - \left\{E'(\lambda) D^{-1}(\lambda) E(\lambda) - \overline{y}^2\right\},\end{aligned}$$

where $\overline{y} = T^{-1} \sum_{t=0}^{T-1} y(t)$. We may thus estimate the parameter λ, which makes the model a *nonlinear* regression model, by maximising with respect to λ the regression sum of squares

$$E'(\lambda) D^{-1}(\lambda) E(\lambda) - \overline{y}^2. \tag{1.3}$$

This expression is simple enough to compute exactly, but the computations will be onerous if T is large. However, many of the calculations can be simplified by using the fact that

$$\sum_{t=0}^{T-1} e^{i\lambda t} = \begin{cases} \frac{e^{i\lambda T}-1}{e^{i\lambda}-1} & ; \quad e^{i\lambda} \neq 1 \\ T & ; \quad e^{i\lambda} = 1 \end{cases}.$$

In fact, for large T, we have the approximations

$$D(\lambda) = \begin{bmatrix} T & O(1) & O(1) \\ O(1) & \frac{T}{2} + O(1) & O(1) \\ O(1) & O(1) & \frac{T}{2} + O(1) \end{bmatrix},$$

and

$$D^{-1}(\lambda) = T^{-1} \begin{bmatrix} 1 & o(1) & o(1) \\ o(1) & 2 + o(1) & o(1) \\ o(1) & o(1) & 2 + o(1) \end{bmatrix},$$

where $o(1)$ and $O(1)$ denote terms which converge to zero and are bounded, respectively. Hence, the regression sum of squares is, from (1.3), approximately

$$\begin{aligned} I_y(\lambda) &= \frac{2}{T} \left\{ \sum_{t=0}^{T-1} y(t) \cos(\lambda t) \right\}^2 + \frac{2}{T} \left\{ \sum_{t=0}^{T-1} y(t) \sin(\lambda t) \right\}^2 \\ &= \frac{2}{T} \left| \sum_{t=0}^{T-1} y(t) e^{i\lambda t} \right|^2, \end{aligned}$$

which is called the *periodogram* of $\{y(t); t = 0, 1, \ldots, T-1\}$. This leads us to consider a second method of estimating λ; namely, maximising $I_y(\lambda)$ with respect to λ. If the model is correct, the two methods lead to equivalent estimators.

Figures 1.7 and 1.8 depict the periodograms of the noiseless sine and square waves, measured at the times $t = 0, 1, \ldots, 511$. The periodograms have been calculated using the finite Fourier transform, at the so-called Fourier frequencies

$$\left\{2\pi\frac{j}{T}; j = 0, 1, \ldots, T-1\right\}.$$

The Fourier transform is calculated at these frequencies because the the 'fast' Fourier transform algorithm of Cooley and Tukey (1965), and later more general versions, which are available in packages such as MATLAB$^{\mathrm{TM}}$, enable the T complex coefficients, equispaced in frequency, to be calculated in $O(T \log T)$ operations rather than the expected $O(T^2)$. It will be seen later, furthermore, that the joint statistical properties of these Fourier coefficients are simpler than those at other frequencies.

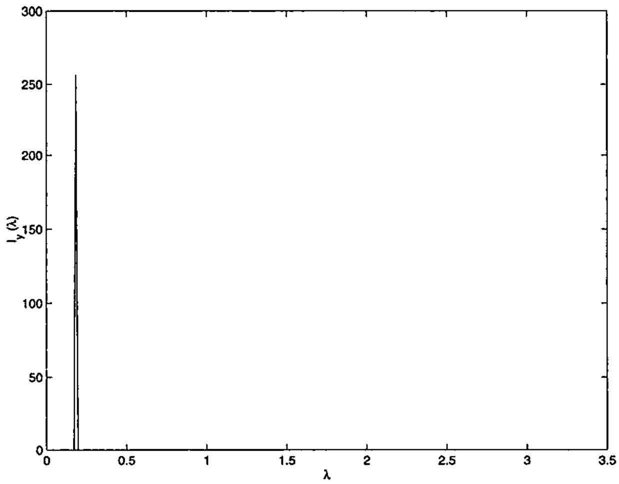

Fig. 1.7. Periodogram of sine wave, frequency 0.1841

Note that the periodogram of the sine wave has a very pronounced peak, near the true frequency, and that the periodogram of the square wave has pronounced peaks at the odd harmonics of the fundamental frequency. The latter might seem a little surprising, since periodogram maximisation, as a frequency estimation technique, has been motivated only by the single frequency case. It is shown in Chapter 3 that when there are r sinusoids, the regression sum of squares, as a function of the r frequencies, is approximated