# Part I

## Overview

The first six chapters of this book are an advertisement for time-series analysis. They are written to be read casually, to introduce some new vocabulary to the reader on primarily an intuitive level, with visual illustrations. Do not read these chapters expecting the technical terms to be defined precisely. Relax. Eventually, the text will cycle back over these ideas and add the necessary precision to this intuitively based introduction.

Cambridge University Press
978-0-521-10336-7 - Time-Series Analysis: A Comprehensive Introduction for Social Scientists
John M. Gottman
Excerpt
More information

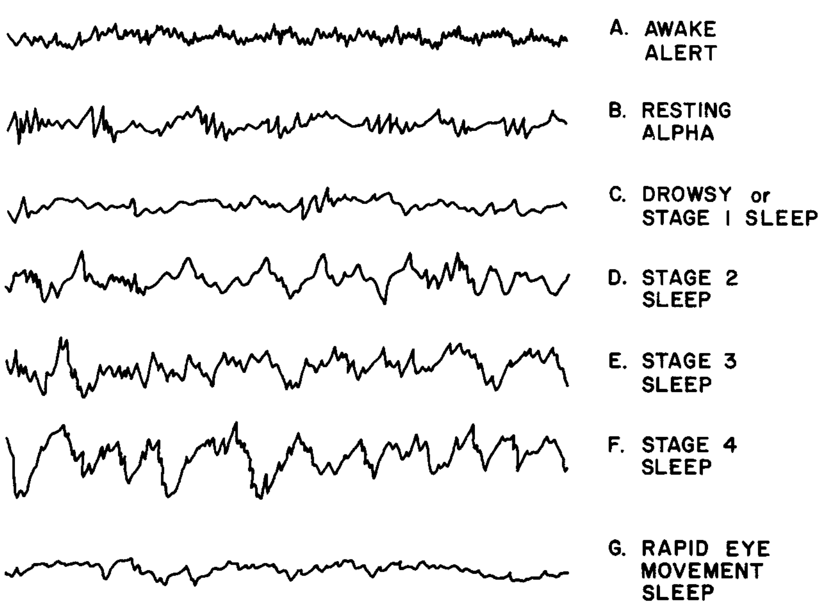
# 1
# The search for hidden structures

*This chapter introduces one concept, the spectral density function, which is capable of detecting the frequencies embedded in the time series even when the data are masked by noise. For the time being, we will discuss the case when there are true cycles in the data. Later, this notion will be generalized to the case when the oscillations are not perfectly cyclic.*

## Brain waves

When we record changes in electrical potential on the scalp of a subject who is either awake, drowsy, or in different stages of sleep, we obtain electroencephalographic (EEG) brain-wave patterns of the type illustrated in Figure 1.1. When a person is awake and active, the EEG record consists of rapid waves of low amplitude, as in part A. As a person becomes relaxed and eventually drowsy, the EEG is characterized by a wave of 8 to 12 hertz (Hz; cycles per second) called an *alpha wave* (B). As the person falls asleep the waves become slower and have larger amplitude (C to F). During

Figure 1.1. Electroencephalographic patterns characteristic of various states of wakefulness and sleep.



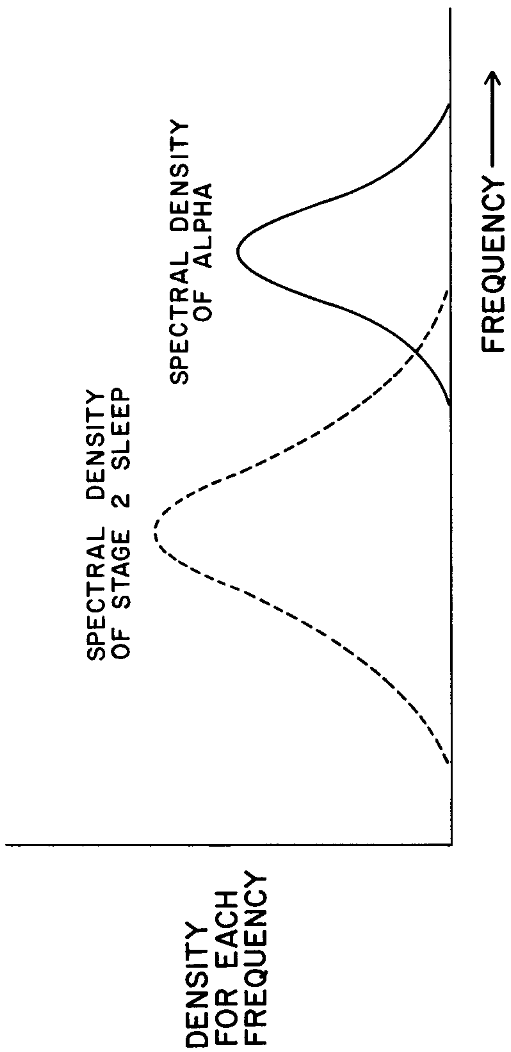



Figure 1.2. Hypothetical spectral density function for stage 2 and alpha EEG patterns.

rapid-eye-movement dreaming (G) the EEG pattern is much more like that of wakefulness shown in part A than those of the other sleeping waves.

We can examine these graphs visually, but is there some way of describing them systematically? The waking EEG is usually described as containing "high-frequency, low-amplitude, erratic cycles." What do the terms "frequency," "amplitude," and "cycles" mean? The term "frequency" in time-series analysis does not mean how often an event occurs, as it does in introductory probability and statistics. Instead, *frequency* means how rapidly things repeat themselves. For example, a stagecoach wagon wheel that is rotating slowly is rotating at a slow frequency; a particular spot on the wheel will be on top every full revolution of the wheel. This will occur slowly (slow frequency) if the wheel is rotating slowly, and rapidly (fast frequency) if the wheel is rotating rapidly. In terms of time-series data, if the data wave smoothly and slowly up to a peak and then smoothly and slowly down to a valley, the data are said to have a lower frequency than if they moved back and forth from peak to valley with every successive observation. *Amplitude* refers to how large the magnitude of the change is from peak to valley. In Figure 1.1 the graph for A, awake–alert, has less amplitude than the graph for F, stage 4 sleep. *Cycle* implies *regular* repetition over a fixed time, called the *period*, of this transition from peak to valley and back to peak. A rapid cycle refers to a high-frequency wave that moves from peak to peak in a short time, or short period.

In the EEG of a drowsy person, cycles called *alpha*, the rhythm of relaxed wakefulness, predominate. In the change from wakefulness to drowsiness, the EEG picks up a cycle. To describe this, a function called the *spectral density* would be used. The spectral density would tell us something like how much variance is accounted for by each frequency we can measure. The spectral density of the awake and relaxed state would thus show a peak in those frequency ranges where we expect to find alphas, and would not show a peak elsewhere. By using the spectral density, we could quantitatively assess the state of our subject. Figure 1.2 illustrates this concept of spectral density.

Why is the spectral density necessary? The answer is that the textbook example of Figure 1.1 is not what we usually observe in practice. In the next section we see how a pure cycle can be hidden by noise so that visual inspection alone is inadequate, but the spectral density can easily pick out our pure cycle.

## Hidden cycles

Figure 1.3 shows a pure sine wave. No one would have to compute a spectral density or do any statistics at all if our data were this regular. In this case,

Cambridge University Press
978-0-521-10336-7 - Time-Series Analysis: A Comprehensive Introduction for Social Scientists
John M. Gottman
Excerpt
More information

Figure 1.3. Example of a pure sine-wave process.

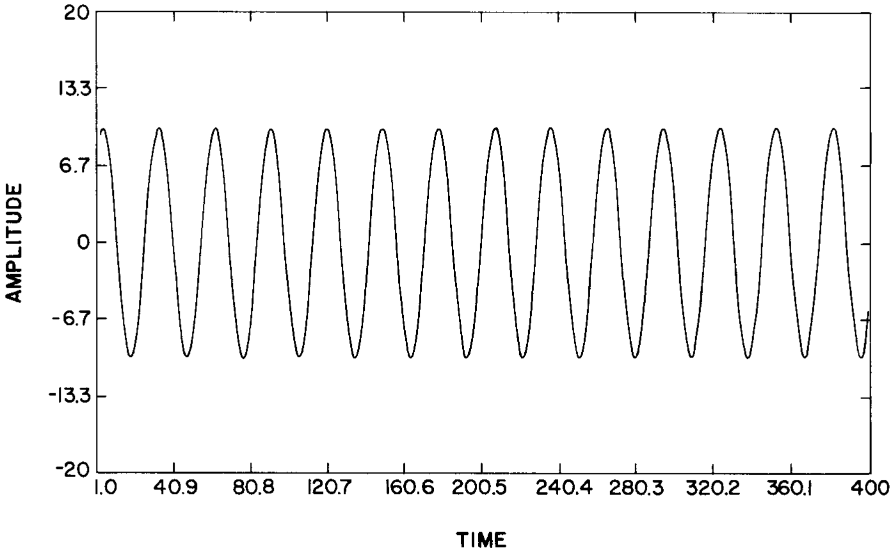
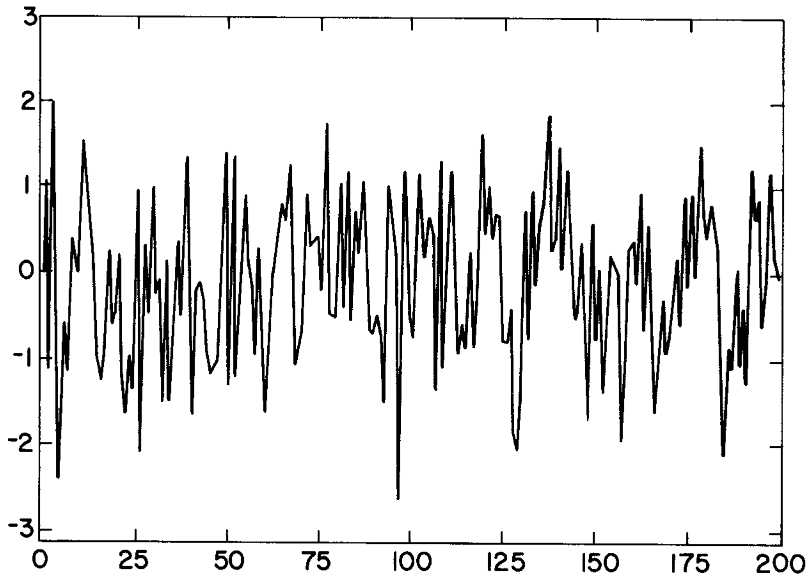



Figure 1.4. Example of a purely random ("white-noise") process.



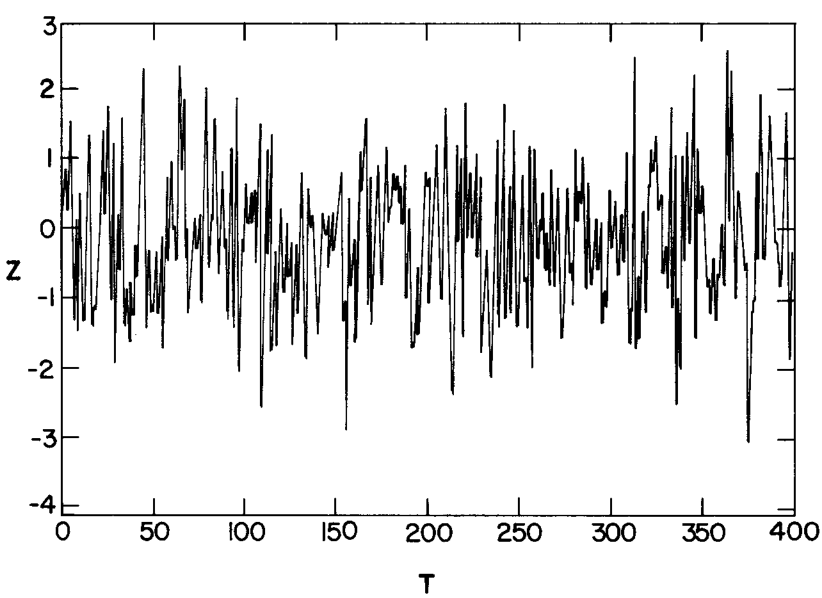
in fact, the spectral density function would be a spike at one frequency and zero everywhere else. Let us examine the opposite state of affairs, a set of completely independent random numbers, called *white noise*, depicted by Figure 1.4. Here there is absolutely no regularity, and all cycles are present with equal intensity, much as white light consists of light of all colors mixed with equal brightness. The spectral density of white noise should theoretically be a straight line with no peaks.

Now consider a pure tone masked by white noise, displayed in Figure 1.5. It is extremely difficult to detect any pure cycle in Figure 1.5a. But the spectral density of the data in Figure 1.5b shows a spike at exactly the right frequency to detect the hidden cycle. This example shows how a spectral analysis of the data can point out structure in the data not visible to the eye.

### Detecting hidden changes

In some cases the mean and variance of a time series may not change after some planned experiment, but some other facet of the data, such as its time dependence, does change; however, if we rely only on analysis of variance we might falsely conclude that no change had taken place. In Figure 1.6 we

Figure 1.5. (a) Example of a sine wave masked by white noise. (b) The spectral density of a sine wave masked by white noise reveals a spike at the correct frequency.




(a)

Figure 1.5. (*cont.*)

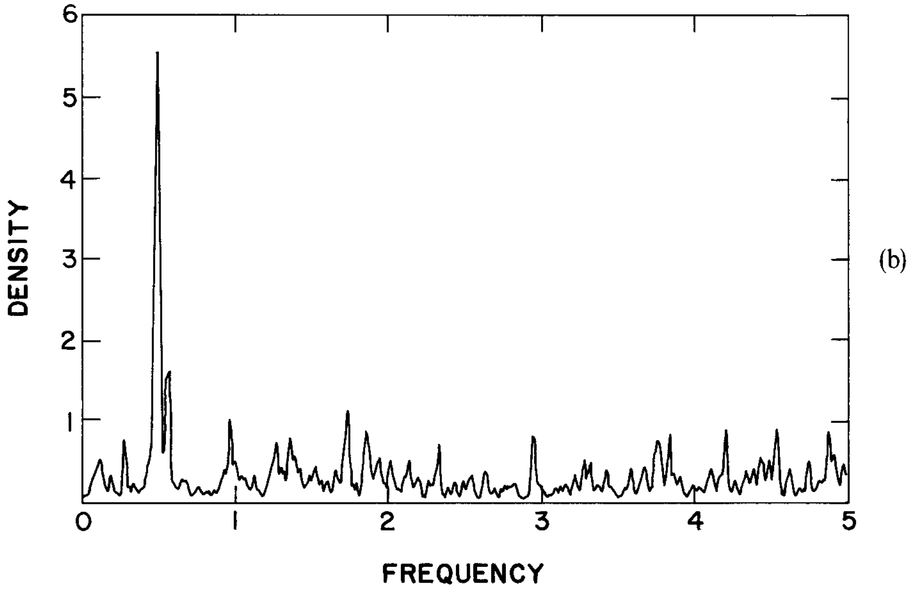
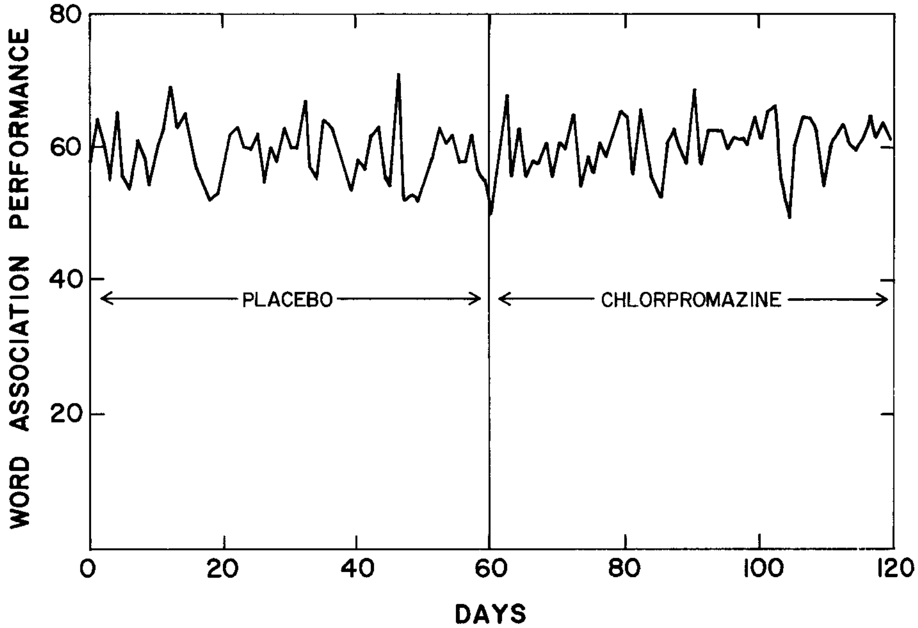



Figure 1.6. Word-association test scores of one schizophrenic patient in two drug conditions.



see the daily Word Association Relatedness scores for one schizophrenic patient for 120 days (Holtzman, 1963). This measure presumably assesses the normality of associations or thought processes, with lower scores indicative of more private, or "strange," associations. The means of the two halves of the series are 59.10 for placebo and 60.75 for drug, and the variances are 18.32 and 16.25, respectively. There really has been no change in mean level or variance in the behavior of this person from placebo to drug.

However, if we search for cyclicity in the placebo condition, we find (Figure 1.7) a peak at one cycle, with a frequency of .1. The period of the cycle is the time it takes for the behavior to repeat itself, and it is the reciprocal of the frequency, or $1/.1 = 10$ days. There is a dominant cycle, when the person is on the placebo, of 10 days.

What happens to this cycle on the drug? If we look at the spectral density of the 60 drug days we find that there is no longer a peak at .1, but that the peak has shifted over to the right, and is over a frequency of .2. This is a new period of $1/.2 = 5$ days (Figure 1.8). Therefore, the drug has shifted the cycle to a more rapid frequency.[1] The behavior of the schizophrenic has become more erratic and less smooth. At this point in our research we may not know if this change is "good" or "bad." This example illustrates how thinking in terms of means and variances may miss what could be valuable

Figure 1.7. Spectral density function of word-association data during the placebo condition, showing a 10-day cycle.

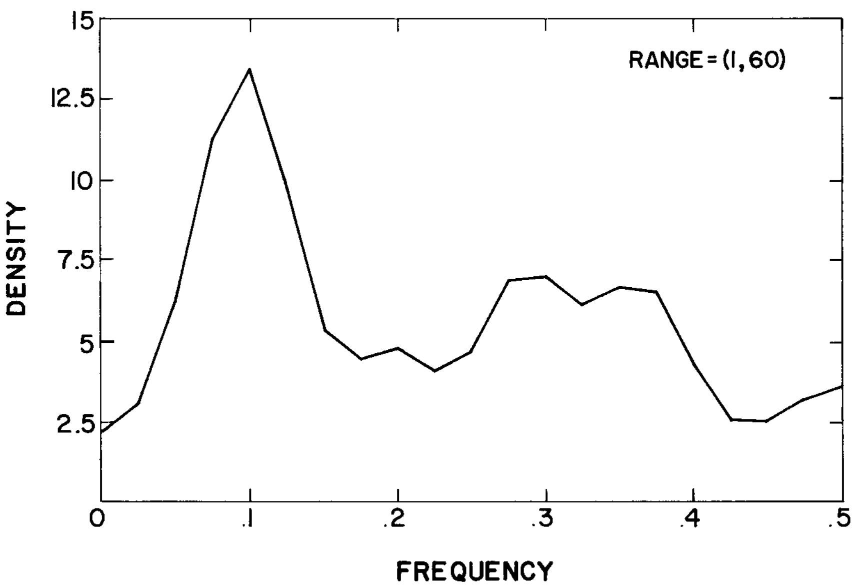
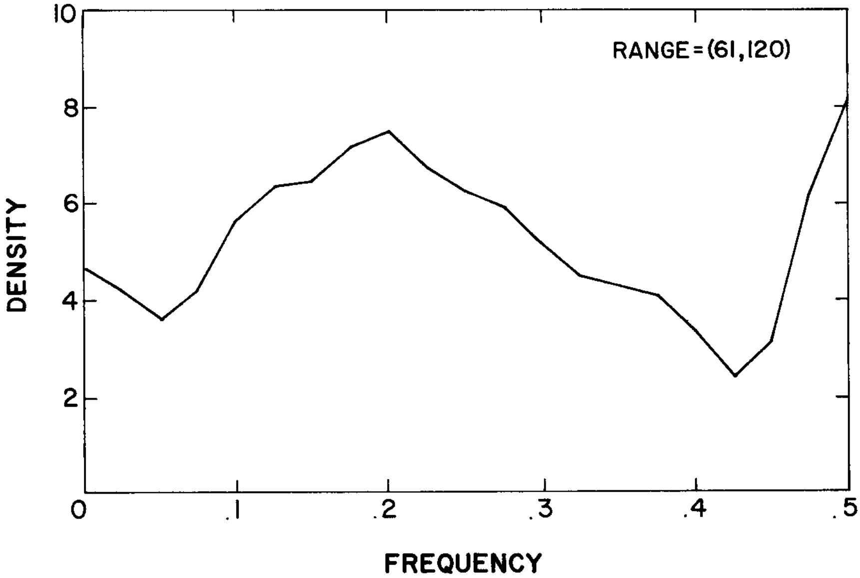


Figure 1.8. Spectral density function of the word-association data during the chloropromazine condition, showing a 5-day cycle.




information in the data. These shifts in frequency are statistically significant, and this book will discuss how to perform tests of significance on data such as these.

# 2
# The ubiquitous cycles

*This chapter continues the discussion of cyclicity and introduces the idea of spectral decomposition, in which a time series is represented as a sum of independent cycles of different frequency. Traffic fatalities data are analyzed visually using this concept of spectral decomposition. The idea of a sum of cycles as an approximation is discussed, which introduces the notion of a Fourier series.*

Most people are familiar with the rainbowlike spectrum into which sunlight is transformed when passed through a prism, and many have also seen the sharp line spectrum that results when monochromatic light is passed through a prism. Mixtures of monochromatic light give rise to lines corresponding to the separate components when passed through a prism, and chemical elements, when properly excited by flame or electric spark or some other means, exhibit characteristic line spectra or band spectra or some combination of the two types. The line spectra are caused by electrons changing energy states in their atoms, but the band spectra are caused by more complicated motions of atoms and molecules. The many frequencies present in sunlight result from the numerous energy transitions going on in the chaos of the thermonuclear reactions. Since any steady-state source of visible light will have a definite spectrum, we can characterize the source by measuring the line and band components. In all cases, broad-band spectra will be associated with more complicated and random atomic and molecular motions, which I will call the "stochastic" component.

A similar situation obtains if we feed an electrical signal composed of a sine wave plus a stochastic component (such as might be produced by vacuum-tube shot noise or an intermittent spark) into a radio receiver. If the radio tunes reasonably sharply and with a fairly flat response over frequencies surrounding that of our sine wave, we will hear the sine wave as a pure tone when we are tuned to its frequency, and we will hear the same tone but with diminishing volume as we tune away from that frequency. Actually, since the ear has a nonlinear response, there will also be a slight tone shift. However, the stochastic component will be heard at about the same volume through a wide tuning range. Anyone who has ever tried to listen to AM radio during a lightning storm and tried to tune out the accompanying static is aware of this phenomenon. Once again we have a case in which a pure tone is characterized by a narrow band of radio frequencies, or "spectral line," and the stochastic component by a broad band of radio frequencies.

In many areas of human activity, measurements of some quantity are made either continuously or at discrete points along the time axis. Furthermore, over long periods of time many of the phenomena being measured have a steady-state quality, or the data can be transformed to exhibit steady-state behavior. (Chapter 8 will explicate this concept of steady-state behavior by introducing the notion of "stationarity.")

In view of the preceding discussion, it is then natural to ask whether there exists a "mathematical prism" or "mathematical tuner" through which we can pass such data and pick out line spectra and band spectra that help characterize and classify the data. Return to the example of brain-wave data measured at various stages of sleep (Figure 1.1). The stochastic component will have a spectral density function that covers a band of frequencies and thus the data will be repetitive, but not as precisely as one would expect if the spectral density function were composed of precise lines. The frequencies of the bands of cycles shift with different stages of sleep. There may also be precise cycles, or line spectra, represented in a set of data, in which case the data