Chapter 6

SPECTRAL ESTIMATION

The previous chapters surveyed the general theory. In this chapter we will specialize the analysis to frequency and spectrum estimates. Our ultimate aim is to derive explicit Bayesian estimates of the power spectrum and other parameters when multiple nonstationary frequencies are present. We will do this by proceeding through several stages beginning with the simplest spectrum estimation problem. We do this because as was shown by Jaynes [12] when multiple well-separated frequencies are present [$|\omega_j - \omega_k| \gg 2\pi/N$], the spectrum estimation problem essentially separates into independent single-frequency problems. It is only when multiple frequencies are close together that we will need to use more general models.

In Chapters 3 and 4 we derived the posterior probability of the $\{\omega\}$ parameters independent of the amplitudes and noise variance and without assuming the sampling times t_i to be uniformly spaced. Much of the discussion in this Chapter will center around understanding the behavior of the posterior probability density for multiple frequencies. This discussion is, of course, simpler when the t_i are uniform, because then the sine and cosine terms are orthogonal in the sense discussed before. We will start by making this assumption; then, where appropriate, the results for nonuniform times will be given.

This should not be taken to imply that uniform time spacing is the "best" way to obtain data. In fact, nonuniform time intervals have some significant advantages over uniform intervals. We will discuss this issue shortly, and show that obtaining data at apparently random intervals will significantly improve the discrimination of high frequencies even with the same amount of data.

6.1 The Spectrum of a Single Frequency

In Chapter 2 we worked out an approximate Bayesian solution to the single stationary harmonic frequency problem. Then in Chapter 3 we worked out what amounts to the general solution to this problem. Because we have addressed this problem so thoroughly in other places we will investigate some other properties of the analysis that may be troubling the reader. In particular we would like to understand what happens to the ability to estimate parameters when one or more of our assumptions is violated. We would like to demonstrate that the estimates derived in Chapter 4 are accurate, and that when the assumptions are violated the estimated frequencies are still reasonably correct but the error estimates are larger, and therefore, more conservative.

6.1.1 The "Student t-Distribution"

We begin this chapter by demonstrating how to use the general formalism to derive the exact "Student t-distribution" for the single frequency problem on a uniform grid. For a uniformly sampled time series, the model equation is

$$f_l = B_1 \cos \omega l + B_2 \sin \omega l$$

where l is an index running over a symmetric time interval ($-T \leq l \leq T$) and ($2T + 1 = N$). The matrix g_{ij}, Eq. (3.4), becomes

$$g_{ij} = \begin{pmatrix} \sum_{l=-T}^{T} \cos^2 \omega l & \sum_{l=-T}^{T} \cos \omega l \sin \omega l \\ \sum_{l=-T}^{T} \cos \omega l \sin \omega l & \sum_{l=-T}^{T} \sin^2 \omega l \end{pmatrix}.$$

For uniform time sampling the off diagonal terms are zero and the diagonal term may be summed explicitly to obtain

$$g_{ij} = \begin{pmatrix} c & 0 \\ 0 & s \end{pmatrix}$$

where

$$c = \frac{N}{2} + \frac{\sin(N\omega)}{2\sin(\omega)}$$

$$s = \frac{N}{2} - \frac{\sin(N\omega)}{2\sin(\omega)}.$$

Then the orthonormal model functions may be written as

$$H_1(t) = \frac{\cos(\omega t)}{\sqrt{c}}$$

$$H_2(t) = \frac{\sin(\omega t)}{\sqrt{s}}.$$

The posterior probability of a frequency ω in a uniformly sampled data set is given by Eq. (3.17). Substituting these model functions gives

$$P(\omega|D,I) \propto \left[1 - \frac{R(\omega)^2/c + I(\omega)^2/s}{N\overline{d^2}}\right]^{\frac{2-N}{2}} \tag{6.1}$$

where $R(\omega)$ and $I(\omega)$ are the squares of the real and imaginary parts of the discrete Fourier transform (2.4, 2.5).

We see now why the discrete Fourier transform does poorly for small N or low frequencies: the constants c and s are normalization constants that usually reduce to $N/2$ for large N; however, these constants can vary significantly from $N/2$ for small N or low frequency. Thus the discrete Fourier transform is only an approximate result that must be replaced by (6.1) for small amounts of data or data sets which contain a low frequency. The general solution is represented by (3.17), and this equation may be applied even when the sampling is nonuniform.

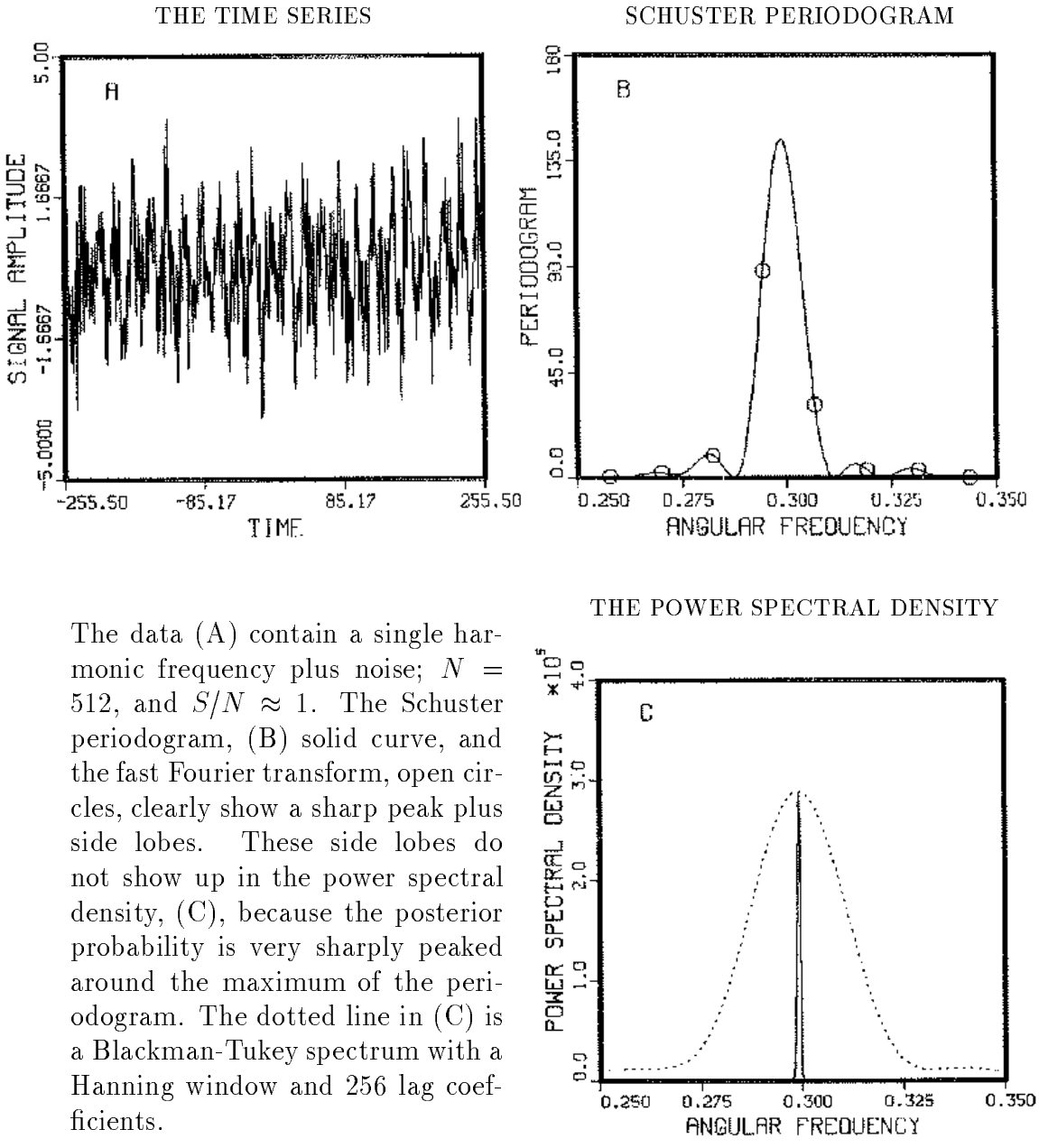
6.1.2 Example – Single Harmonic Frequency

To obtain a better understanding of the use of the power spectral density derived in Chapter 2 (2.16), we have prepared an example: the data consist of a single harmonic frequency plus Gaussian white noise, Fig. 6.1. We generated these data from the following equation

$$d_j = 0.001 + \cos(0.3j + 1) + e_j \tag{6.2}$$

where j is a simple index running over the symmetric interval $-T$ to T in integer steps ($2T+1 = 512$), and e_j is a random number with unit variance. After generating the time series we computed its average value and subtracted it from each data point: this ensures that the data have zero mean value. Figure 6.1(A) is a plot of this computer simulated time series, and Fig. 6.1(B) is a plot of the Schuster periodogram (continuous curve) with the fast Fourier transform marked with open circles. The

Figure 6.1: Single Frequency Estimation



The data (A) contain a single harmonic frequency plus noise; $N = 512$, and $S/N \approx 1$. The Schuster periodogram, (B) solid curve, and the fast Fourier transform, open circles, clearly show a sharp peak plus side lobes. These side lobes do not show up in the power spectral density, (C), because the posterior probability is very sharply peaked around the maximum of the periodogram. The dotted line in (C) is a Blackman-Tukey spectrum with a Hanning window and 256 lag coefficients.

periodogram and the fast Fourier transform have spurious side lobes, but these do not appear in the plot of the power spectral density estimate, Fig. 6.1(C), because as noted in Chapter 2, the processing in (4.15) will effectively suppress all but the very highest peak in the periodogram. This just illustrates numerically what we already knew analytically; it is only the very highest part of the periodogram that is important for estimation of a single frequency.

We have included a Blackman-Tukey spectrum using a Hanning window (dotted line) in Figure 6.1(C) for comparison. The Blackman-Tukey spectrum has removed the side lobes at the cost of half the resolution. The maximum lag was set at 256, i.e. over half the data. Had we used a lag of one-tenth as Tukey [13] advocates, the Blackman-Tukey spectrum would look nearly like a horizontal straight line on the scale of this plot.

Of course, the peak of the periodogram and the peak of the power spectral density occur at the same frequency. Indeed, for a simple harmonic signal the peak of the periodogram is the optimum frequency estimator. But in our problem (i.e. our model), the periodogram is not even approximately a valid estimator of the power spectrum, as we noted earlier. Consequently, even though these techniques give nearly the same frequency estimates, they give very different power spectral estimates and, from the discussion in Chapters 2 and 4, very different accuracy estimates.

Probably, one should explain the difference on the grounds that the two procedures are solving different problems. Unfortunately, we are unable to show this explicitly. We have shown above in detail that the Bayesian procedure yields the optimal solution to a well-formulated problem, by a well-defined criterion of optimality. One who wishes to solve a different problem, or to use a different optimality criterion, will naturally seek a different procedure. The Blackman-Tukey procedure has not, to the best of our knowledge, been so related to any specific problem, much less to any optimality criterion; it was introduced as an intuitive, *ad hoc* device. We know that Blackman and Tukey had in mind the case where the entire time series is considered a sample drawn from a "stationary Gaussian random process"; thus it has no mention of such notions as "signal" and "noise". But the "hanning window" smoothing procedure has no theoretical relation to that problem; and of course the Bayesian solution to it (given implicitly by Geisser and Cornfield [24] and Zellner [25] in their Bayesian estimates of the covariance matrix of a multivariate Gaussian sampling distribution) would be very different. It is still conceivable that the Blackman-Tukey procedure is the solution to some well-defined problem, but we do not know what that problem is.

6.1.3 The Sampling Distribution of the Estimates

We mentioned in Chapter 4 that we would illustrate numerically that the Bayesian estimates for the $\{\omega\}$ parameters were indeed accurate under the conditions supposed, even by sampling-theory criteria. For the example just given the true frequency was 0.3 while the estimated frequency from data with unity signal-to-noise ratio was

$$(\omega)_{\rm est} = 0.2997 \pm 0.0006$$

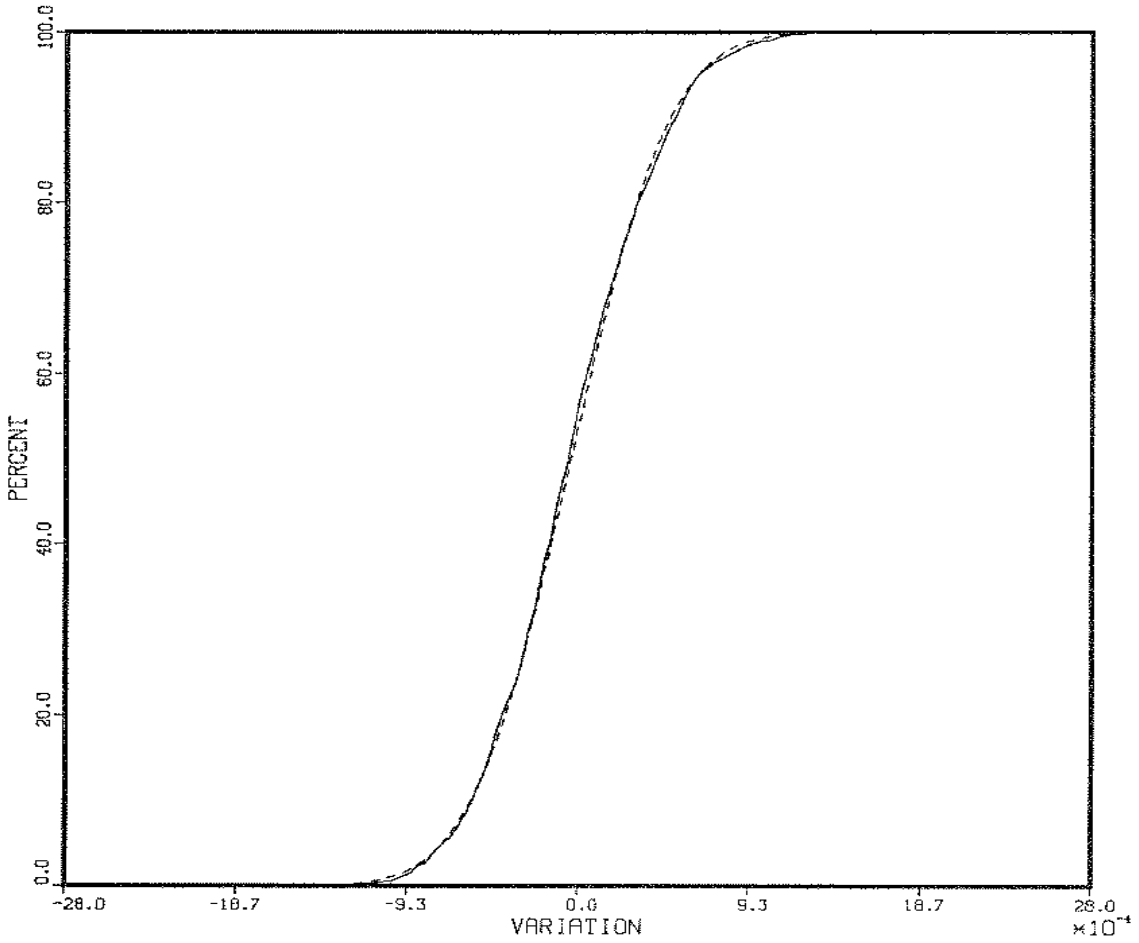
at two standard deviations, in dimensionless units. But one example in which the estimate is accurate is not a sufficient demonstration. Suppose we generate the signal (6.2) a number of times and allow the noise to be different in each of these. We can then compute a histogram of the number of times the frequency estimate was within one standard deviation, two standard deviations $\cdots$ etc. of the true value. We could then plot the histogram and compare this to a Gaussian, or we could integrate the histogram and compare the total percentage of estimates included in the interval $\hat{\omega} - \langle\omega\rangle$ to a Gaussian. This would tell us something about how accurately the results are reproducible over different noise samples. This is not the same thing as the accuracy with which ω is estimated from one given data set; but orthodox statistical theory takes no note of the distinction. Indeed, in "orthodox" statistical theory, this sampling distribution of the estimates is the sole criterion used in judging the merits of an estimation procedure.

We did this numerically by generating some 3000 samples of (6.2) and estimating the frequency ω from each one. We then computed the histogram of the estimates, integrated, and plotted the total percentage of estimates enclosed as a function of $\hat{\omega} - \langle\omega\rangle$, Fig. 6.2 (solid line). From the 3000 sample estimates we computed the mean and standard deviation. The dashed line is the equivalent plot for a Gaussian having this mean and standard deviation. With 3000 samples the empirical sampling distribution is effectively identical to this Gaussian, and its width corresponds closely to the Bayesian error estimate. However, as R. A. Fisher explained many years ago, this agreement need not hold when the estimator is not a sufficient statistic.

6.1.4 Violating the Assumptions – Robustness

We have said a number of times that the estimates we are making are the "most conservative" estimates of the parameters one can make. We would like to convey a

Figure 6.2: The Distribution of the Sample Estimates



We generated the single harmonic signal (6.2) some 3000 times and estimated the frequency. We computed the mean, standard deviation, and a histogram from these data, then totaled the number of estimates from left to right; this gives the total percentage of estimates enclosed as a function of $\hat{\omega} - \langle\omega\rangle$ (solid line). We have plotted an equivalent Gaussian (dashed line) having the same mean and standard deviation as the sample. Each tick mark on the plot corresponds to two standard deviations.

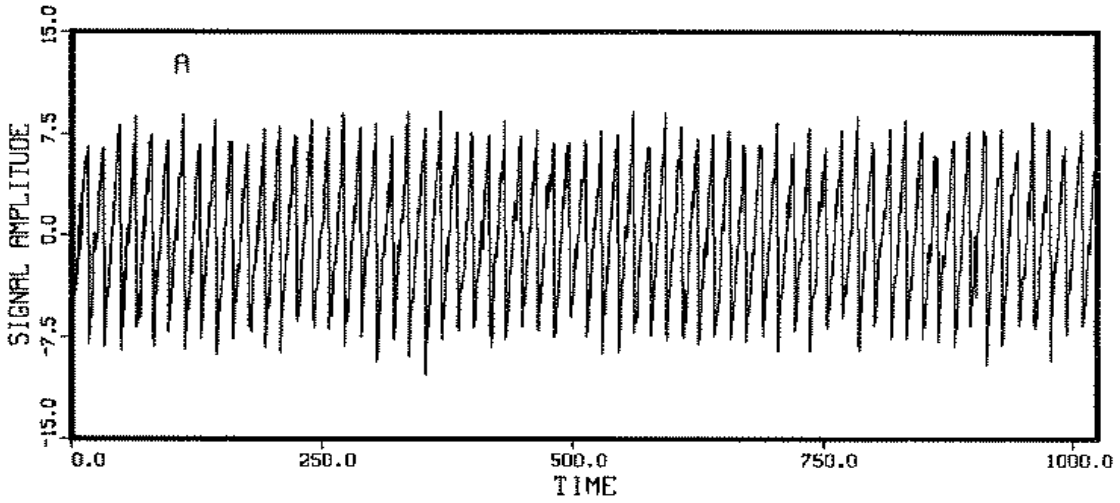
better understanding of that term now. General theorems guarantee [25], [26], [27], [28] that if all of the assumptions are met exactly, then the estimate we obtain will be the "best" estimate of the parameters that one can make from the data and the prior information. But in all cases where we had to put in prior information, we specifically assumed the least amount of information possible. This occurred when we chose a prior for the noise – we used maximum entropy to derive the most uninformative prior we could for a given second moment: the Gaussian. It occurred again when we assigned the priors for the amplitudes, and again when we assigned the prior for the $\{\omega\}$ parameters. This means that any estimate that takes into account additional information by using a more concentrated prior will always do better! But, further if the model assumptions are not met by the data (e.g. the noise is not white, the "true" signal is different from our model, etc.), then probability theory will necessarily make the accuracy estimates even wider because the models do not fit the data as well! These are bold claims, and we will demonstrate them for the single frequency model.

Periodic but Nonharmonic Signals

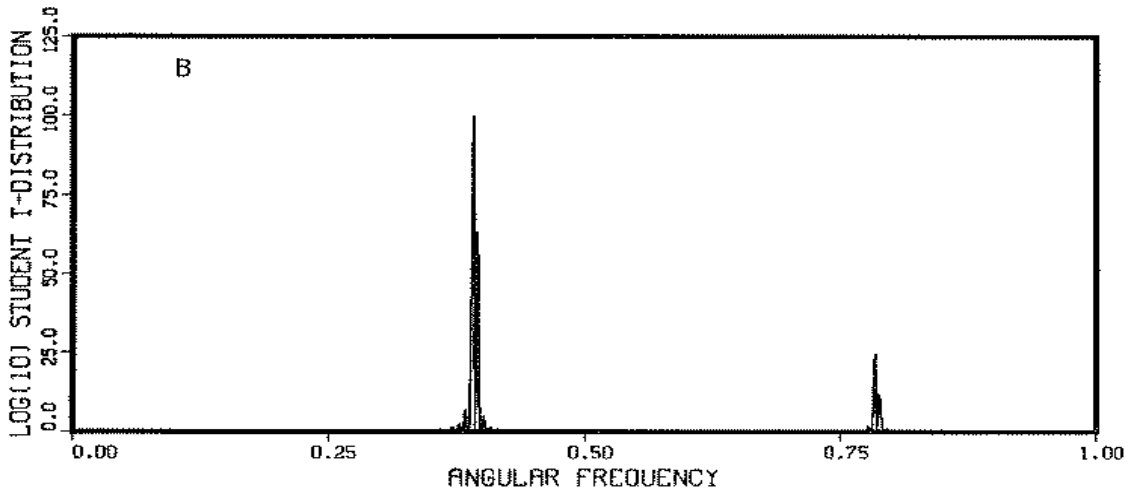
First let us investigate what will happen if the true signal in the data is different from that used in the model (i.e. it does not belong to the class of model functions assumed by the model). Consider the time series given in Fig. 6.3(A); this signal is a series of ramp functions. We generated the data with $N = 1024$ data points by simply running a counter from zero to 15, and repeated this process 64 times. The RMS is then $[1/16 \sum_{k=0}^{15}(k - 7.5)^2]^{\frac{1}{2}} = 4.61$. We then added a unit normal random number to the data, and last we computed the average value of the data and subtracted this from each data point.

This signal is periodic but not harmonic; nonetheless we propose to use the single harmonic frequency model on these data. Figure 6.3(B) is a plot of the $\log_{10}$ of the probability of a single harmonic frequency in these data: essentially this is the discrete Fourier transform of the data. We see in Fig. 6.3(B) the discrete Fourier transform has at least four peaks. But we have demonstrated that the discrete Fourier transform is an optimal frequency estimator for a single harmonic frequency: all of the structure, except the main peak, is a spurious artifact of not using the true model. The main peak in Fig. 6.3(B) is some 25 orders of magnitude above the second largest peak: probability theory is telling us all of that other structure is completely negligible. We then located this frequency as accurately as possible and computed the estimated

Figure 6.3: Periodic but Nonharmonic Time Signals



BASE 10 LOGARITHM OF THE PROBABILITY OF A
HARMONIC FREQUENCY IN NONSINUSOIDAL DATA



The data in (A) contain a periodic but nonharmonic frequency, with $N = 1024$, and $S/N \approx 4.6$. The Schuster periodogram, (B), clearly indicates a single sharp peak plus a number of other spurious features. Estimating the frequency from the peak of the periodogram gives 0.3927 ± 0.0003 while the true frequency is 0.3927.

error in the frequency using (4.14). This gives

$$(\omega)_{\text{est}} = 0.3927 \pm 0.0003$$

at two standard deviations. The true frequency is

$$(\omega)_{\text{true}} = 0.392699$$

while the "best" estimate possible for a sinusoidal signal with the same total number of data values and the same signal-to-noise would be given by

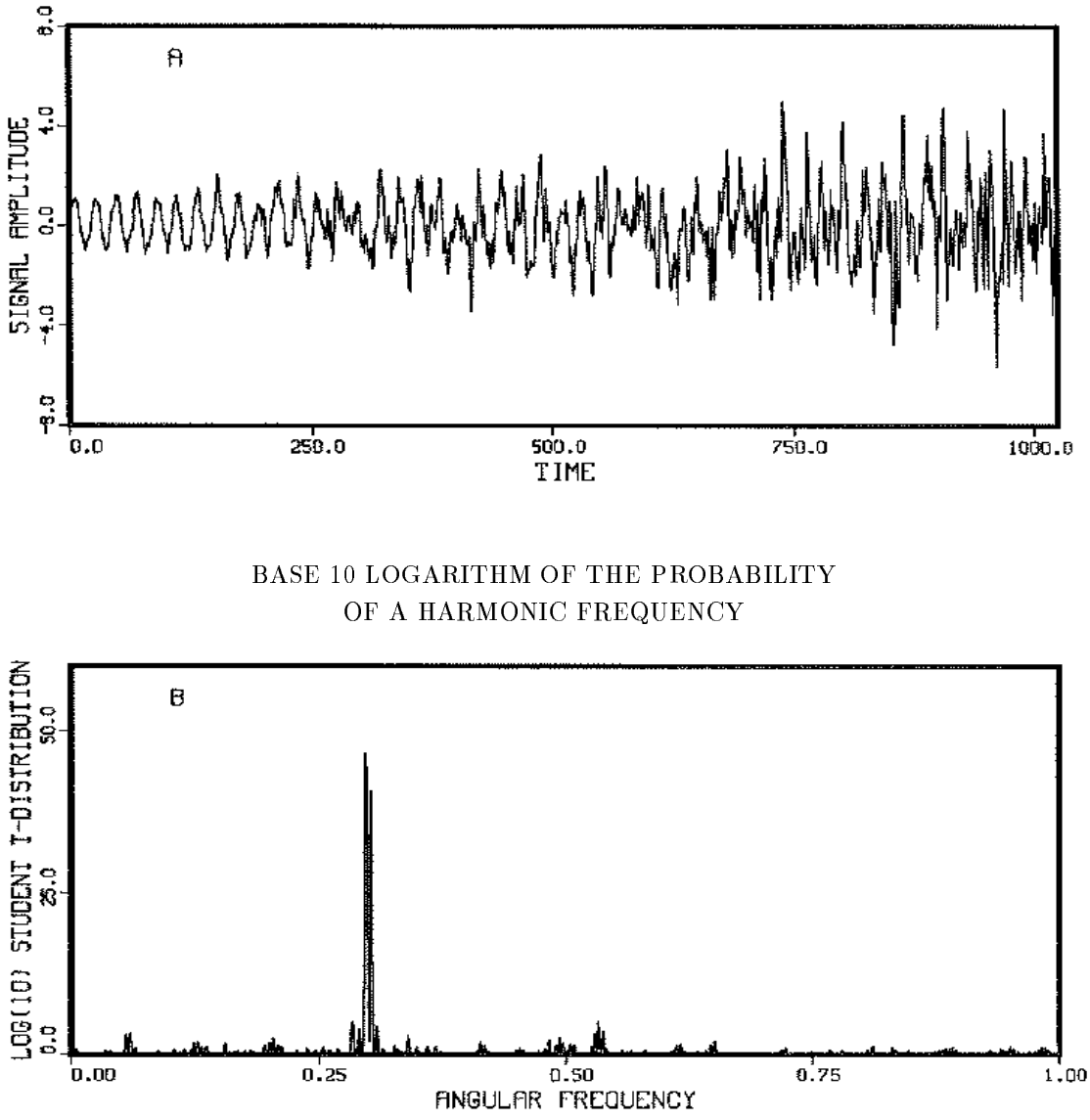
$$(\omega)_{\text{best}} = 0.392699 \pm 0.00003.$$

The estimate is a factor of ten worse than what could be obtained if the true signal met the assumptions of the model (i.e. was sinusoidal with the same signal-to-noise ratio). The major difference is in the estimated noise: the true signal-to-noise is 4.6, but the estimated signal-to-noise using the harmonic model is only 1.5.

The Effect of Nonstationary, Nonwhite Noise

What will be the effect of nonwhite noise on the ability to estimate a frequency? In preparing this test we used the same harmonic signal as in the simple harmonic frequency case (6.2). Although the noise is still Gaussian, we made it different from independent, identically distributed (iid) Gaussian in two ways: first, we made the noise increase linearly in time and second, we filtered the noise with a 1-2-1 filter. Thus the noise values not only increased in time; they were also correlated. The data for this example are shown in Fig. 6.4(A), and the $\log_{10}$ of the "Student t-distribution" is shown in Fig. 6.4(B). The data were prepared by first generating the simple harmonic frequency. We then prepared the noise using a Gaussian distributed random number generator, scaling linearly with increasing time, and filtering; finally, we added the noise to the data. The noise variance in these data ranges from 0.1 in the first data values to 2.1 in the last few data values – there are $N = 1000$ data values. We next computed the $\log_{10}$ probability of a single harmonic frequency in the data set, Fig. 6.4(B). There are two close peaks near 0.3 in dimensionless units. However, we now know that only the highest peak is important for frequency estimation. The highest peak is some 10 orders of magnitude above the second. Thus the second peak is completely negligible compared to the first. We estimated the frequency from this peak and found 0.297 ± 0.003; the correct value is 0.3. Thus one pays a penalty in

Figure 6.4: The Effect of Nonstationary, Nonwhite Noise



The data in Fig. 6.4(A) contain a periodic frequency but the noise is nonstationary and nonwhite, as described in the text. There are 1000 data points with $S/N < 0.5$, The Schuster periodogram, Fig. 6.4(B), clearly indicates a single sharp peak from which we estimated the frequency to be 0.297 ± 0.003; the correct value is 0.3.

accuracy; but the Bayesian conclusions still do not mislead us about this. Actually, the nonstationarity, which obscures part of the data, was much more serious than the nonwhiteness.

Amplitude modulation and other violations of the assumptions

It should be relatively clear by now what will happen when the amplitude of the signal is not constant. For the single stationary frequency problem the sufficient statistic is the Schuster periodogram, and we know from past experience that this statistic is at least usable on nonstationary series with Lorentzian or Gaussian decay. We can also say that when the amplitude modulation is completely unknown, the single largest peak in the discrete Fourier transform is the only indication of frequencies: all others are evidence but not proof. If one wishes to investigate these others one must include some information about the amplitude modulation.

It should be equally obvious that when the signal consists of several stationary sinusoids, the periodogram continues to work well as long as the frequencies are reasonably well separated. But any part of the data that does not fit the model is noise. In cases where we analyze data that contain multiple stationary frequencies using a one-frequency model, all of the frequencies except the one corresponding to the largest peak in the discrete Fourier transform are from the standpoint of probability theory just noise – and extremely correlated, non-Gaussian noise.

All of these effects, and why probability theory continues to work after a fashion in spite of them, are easily understood in terms of the intuitive picture given earlier on page 36. We are picking the frequency so that the dot product between the data and the model is as large as possible. In the case of the sawtooth function described earlier, it is obvious that the "best" fit will occur when the frequency matches that of the sawtooth, although the fit of a sawtooth to a sinusoid cannot be very good; so probability theory will estimate the noise to be large. The same is true for harmonic frequencies with decay; however, the estimated amplitude and phase of the signal will not be accurate. This interpretation should also warn you that when you try to fit a semiperiodic signal (like a Bessel oscillation) to a single sinusoidal model, the fit will be poor. Fundamentally, the spectrum of a nonsinusoidal signal does not have a sharp peak; and so the sharpness of the periodogram is no longer a criterion for how well its parameters can be determined.

6.1.5 Nonuniform Sampling

All of the analysis done in Chapters 2 through 5 is valid when the sampling intervals are nonuniformly spaced. But is anything to be gained by using such a sampling technique? Initially we might anticipate that the problem of aliasing will be significantly reduced. Additionally, the low frequency cutoff is a function of the length of time one samples. We will be using samples of the same duration, so we do not expect to see any significant change in the ability to detect and resolve low frequencies. But will the ability to detect any signal be changed? Will sampling at a nonuniform rate make it possible to estimate a frequency better? We will attempt to address all of these concerns. But most of this will be in the form of numerical demonstrations. No complete analytical theory exists on this subject.

Aliasing

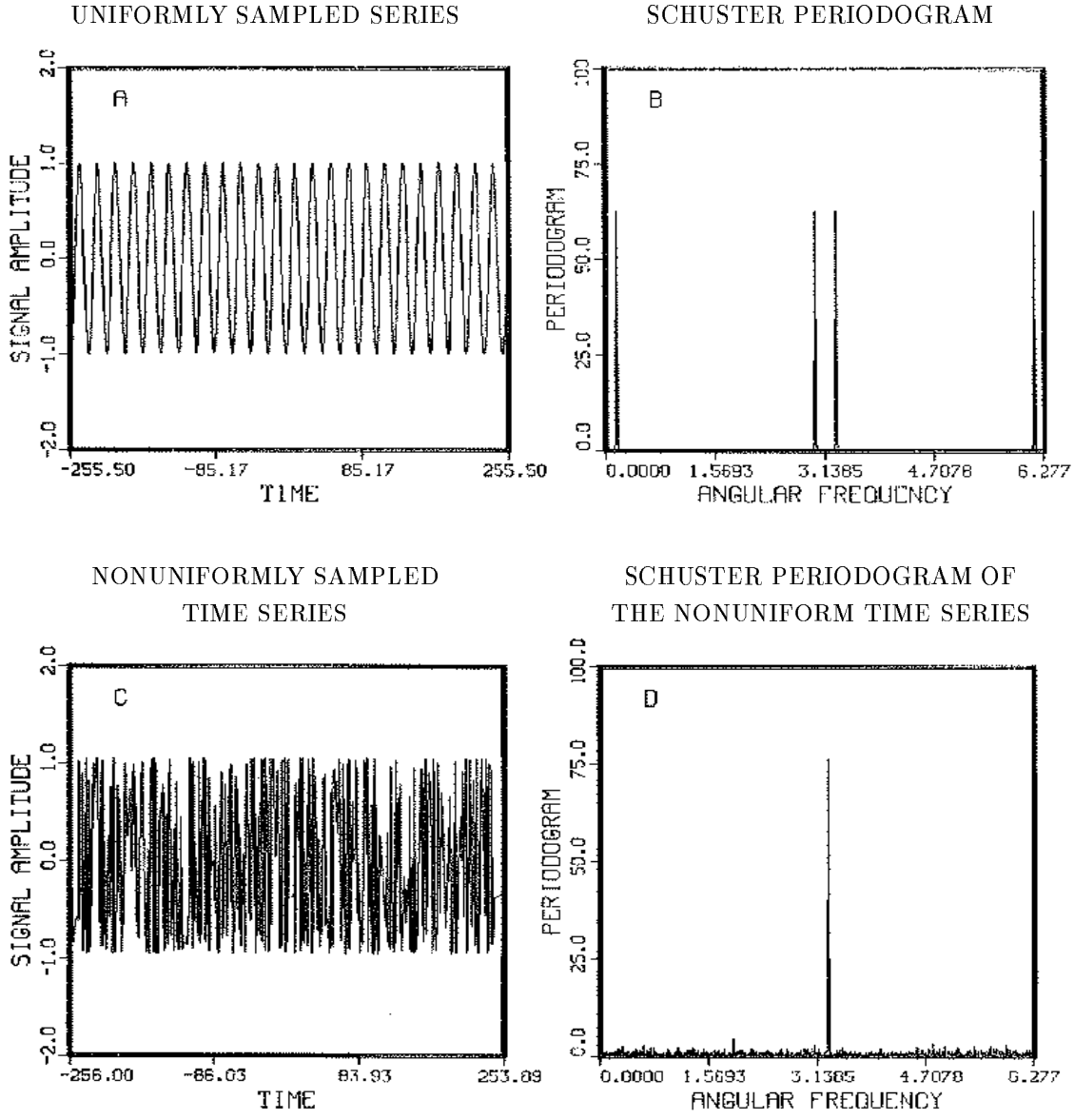
We will address the question of aliasing first. To make this test as clear as possible we have performed it without noise. The data were generated using

$$d_j = \cos([\pi + 0.3]t_j + 1).$$

For the uniform sampled data t_j is a simple index running from $-T$ to T by integer steps and $2T + 1 = 512$. Except for the lack of noise and the addition of π to the frequency this is just the example used in Fig. 6.1. Figure 6.5(A) is a plot of this uniformly sampled series. The true frequency is $0.3 + \pi$, but the plot has the appearance of a frequency of only 0.3 radians per unit step. In the terminology introduced by Tukey, this is an "alias" of the true frequency. The true frequency is oscillating more than one full cycle for each time step measured. The nonaliased frequencies that can be discriminated, with uniform time samples, have ranges from 0 to $\pi/2$. The periodogram of these data Fig. 6.5(B) has four peaks in the range $0 \leq \omega \leq 2\pi$: the true frequency at $0.3 + \pi$ and three aliases.

The nonuniform sampled time series Fig. 6.5(C) also has a time variable which takes on values from $-T$ to T. There are also 512 data points. The true frequency is unchanged. The time variable was randomly sampled. A random number generator with uniform distribution was used to generate 512 random numbers. These numbers were scaled onto the proper time intervals and the simulated signal was then evaluated at these points. No one particular region was intentionally sampled more than any

Figure 6.5: Why Aliases Exist



Aliasing is caused by uniform sampling of data. To demonstrate this we have prepared two sets of data: A uniformly sampled set (A), and a nonuniformly sampled set (C). The periodogram for the uniform signal, (B), contains a peak at the true frequency $(0.3 + \pi)$ plus three alias peaks. The periodogram of the nonuniformly sampled data, (D) has no aliases.

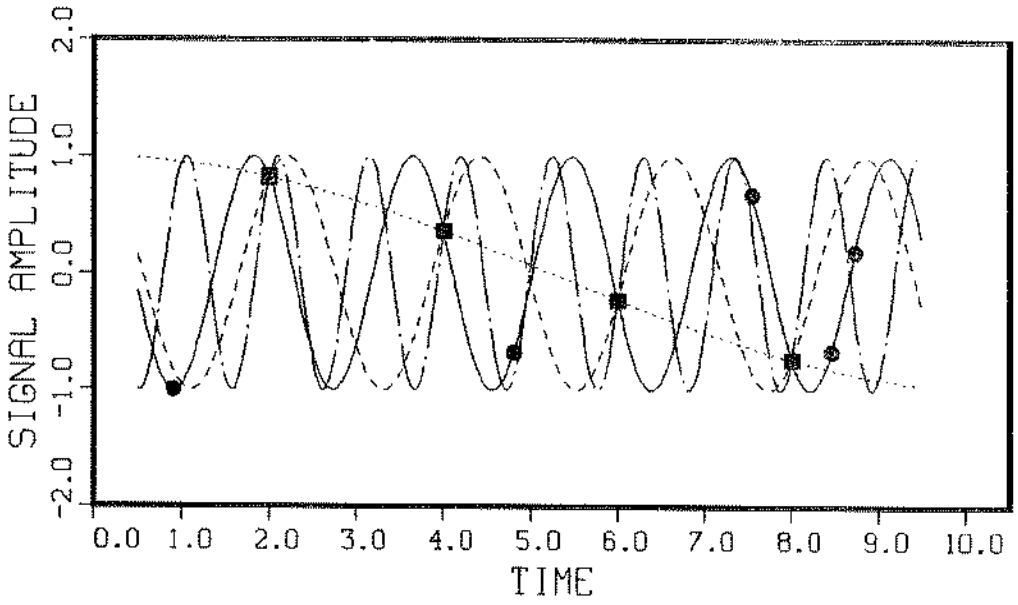
other. The time series, Fig. 6.5(C) looks very different from the uniformly sampled time series. By sampling at different intervals the presence of the high frequency becomes apparent.

We then computed the periodogram for these data and have displayed it in Fig. 6.5(D). The first striking feature is that the aliased frequencies (those corresponding to negative frequencies as well as those due to adding π to the frequency) have disappeared. The second feature which is apparent is that the two periodograms are nearly the same height. Sampling at a nonuniform rate does not significantly alter the precision of the frequency estimates, provided we have the same amount of data, and the same total sampling time. Third, the nonuniformly sampled time series has small features in the periodogram which look very much like noise, even though we know the signal has no noise. Small wiggles in the periodogram are not caused just by the noise; they can also be caused by the irregular sampling times (it should be remembered that these features are not relevant to the parameter estimation problem). The answer to the first question: "Will aliasing go away when one uses a nonuniformly sampled time series?" is yes.

Why aliasing is eliminated for a nonuniform time series is easily understood. Consider Fig. 6.6; here we have illustrated the true frequency (solid line) and the three alias frequencies from the previous example, Fig. 6.5. The squares mark the location of three uniform sample points, while the circles mark the location of the nonuniform points. Looking at Fig. 6.6 we now see aliasing in an entirely different light. Probability theory is indicating (quite rightly) that in the frequency region $0 \leq \omega \leq 2\pi$ there are four equally probable frequencies, Fig. 6.5(B), while for the nonuniformly sampled data, probability theory is indicating that there is only one frequency consistent with the data, Fig. 6.5(D).

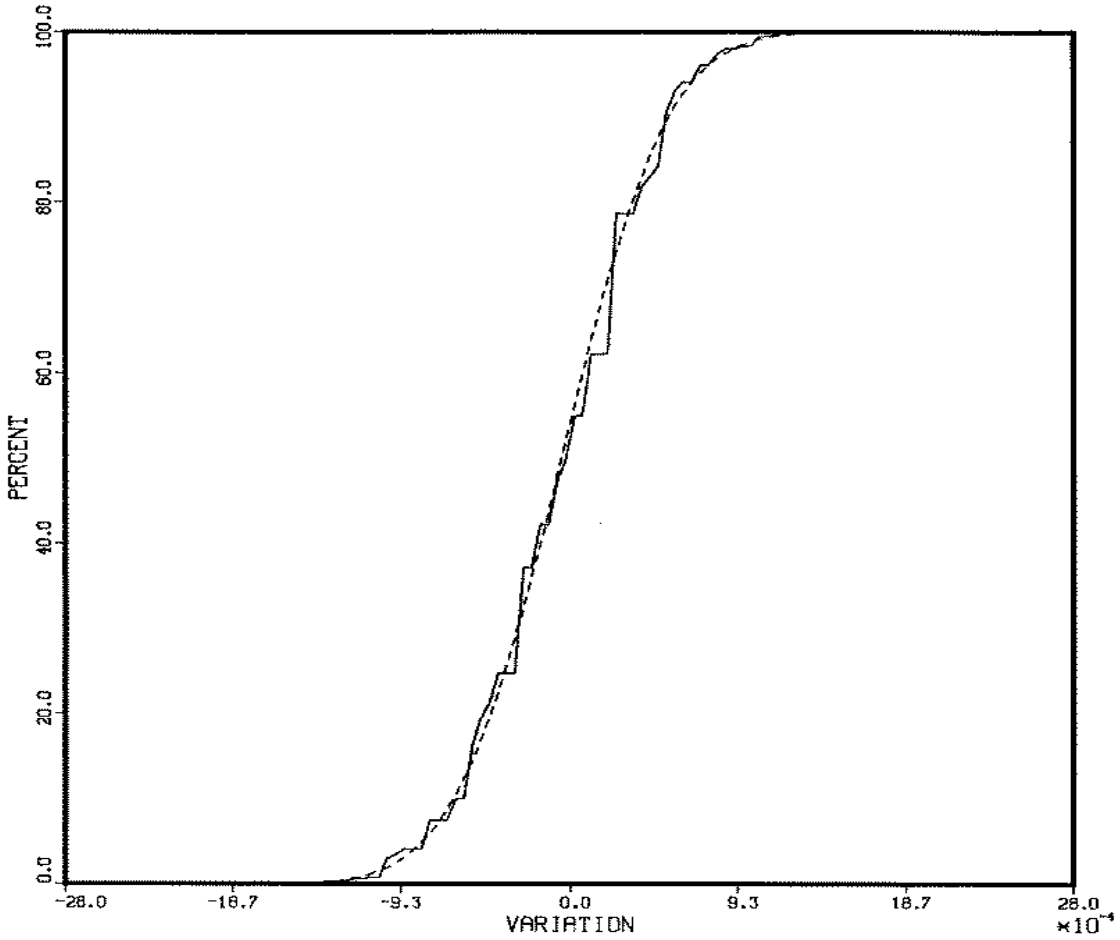
Of course it must be true that the aliasing phenomenon returns for some sufficiently high frequency. If the sampling times $\{t_i\}$, although nonuniform, are all integer multiples of some small interval δt, then frequencies differing by $2\pi/\delta t$ will still be aliased. We did one numerical test with a signal-to-noise ratio of one and the same true frequency. We then calculated the periodogram for higher frequencies. We continued increasing the frequency until we obtained the first alias. This occurred at a frequency around 60π, almost 1.8 orders of magnitude improvement in the frequency band free of aliasing. Even then this second large maximum was many orders of magnitude below that at the first "true" frequency.

Figure 6.6: Why Aliases Go Away for Nonuniformly Sampled Data



When aliases occur in a uniformly sampled time series, probability theory is still working correctly; indicating there is more than one frequency corresponding to the "best" estimate in the data. Suppose we have a signal $\cos(0.3 + \pi)t$ (solid line). For a uniformly sampled data (squares) there are four possible frequencies: $\hat{\omega} = 0.3$ (dotted line), $\hat{\omega} = \pi - 0.3$ (dashed line), $\hat{\omega} = 2\pi - 0.3$ (chain dot), and the true frequency (solid line) which pass through the uniform data (marked with squares). For nonuniformly sampled time series (marked with circles), aliases effectively do not occur because only the "true" (solid line) signal passes through the nonuniformly spaced points (circles).

Figure 6.7: Uniform Sampling Compared to Nonuniform Sampling



We generated some 3000 sets of data with nonuniform data samples, estimated the frequency, computed a histogram, and computed the cumulative number of estimates summing from left to right on this plot (solid line). The equivalent plot for uniformly sampled data is repeated here for easy reference (dotted line). Clearly there is no significant difference in these plots.

Nonuniform Sampling and the Frequency Estimates

The second question is "Will sampling at a nonuniform rate significantly change the frequency estimate?". To answer this question we have set up a second test, Fig. 6.7. The simulated signal is the same as that in Fig. 6.2, only now the samples are nonuniform. We generated some 3000 samples of the data and estimated the frequency from each. We then computed a histogram and integrated to obtain cumulative sampling distribution of the estimates, Fig. 6.7. If nonuniform sampling improves the frequency resolution then we would expect the cumulative distribution (solid line) to rise faster than for the uniformly sampled case (dotted line). As one can see from this plot, nonuniform sampling is clearly equivalent to uniform sampling when it comes to the accuracy of the parameter estimates; moreover, nonuniform sampling improves the high frequency resolution but does not change the frequency estimates otherwise.

Some might be disturbed by the irregular appearance of the solid line in Fig. 6.7. This irregular behavior is simply "digitization" error in the calculation. When we performed this calculation for the uniform case the first time, this same effect was present. We were unsure of the cause, so we repeated the calculation forcing our searching routines to find the maximum of the periodogram much more precisely. The irregular behavior was much reduced. We did not repeat this procedure on the nonuniformly sampled data, because it is very expensive computationally.

6.2 A Frequency with Lorentzian Decay

The simple harmonic frequency problem discussed in Chapter 2 may be generalized easily to include Lorentzian or Gaussian decay. We assume, for this discussion, that the decay is Lorentzian; the generalization to other types of decay will become more obvious as we proceed. For a uniformly sampled interval the model we are considering is

$$f(l) = [B_1 \cos(\omega l) + B_2 \sin(\omega l)]\mathrm{e}^{-\alpha l} \tag{6.3}$$

where l is restricted to values ($1 \leq l \leq N$). We now have four parameters to estimate: the amplitudes B_1, B_2; the frequency ω; and the decay rate α.

6.2.1 The "Student t-Distribution"

The solution to this problem is a straightforward application of the general procedures. The matrix g_{ij} (3.4) is given by

$$g_{ij} = \begin{pmatrix} \sum_{l=1}^{N} \cos^2(\omega l)\mathrm{e}^{-2\alpha l} & \sum_{l=1}^{N} \cos\omega l \sin\omega l \mathrm{e}^{-2\alpha l} \\ \sum_{l=1}^{N} \cos\omega l \sin\omega l \mathrm{e}^{-2\alpha l} & \sum_{l=1}^{N} \sin^2(\omega l)\mathrm{e}^{-2\alpha l} \end{pmatrix}.$$

This problem can be solved exactly. However, the exact solution is tedious, and not very informative. Fortunately an approximate solution is easily obtained which exhibits most of the important features of the full solution; and is valid in the same sense that a discrete Fourier transform is valid. We approximate g_{ij} as follows: First the sum over the sine squared and cosine squared terms may be approximated as

$$\begin{aligned} c &\equiv \sum_{l=1}^{N} \cos^2(\omega l)\mathrm{e}^{-2l\alpha} \approx \sum_{l=1}^{N} \sin^2(\omega l)\mathrm{e}^{-2l\alpha} \\ &= \frac{1}{2}\sum_{l=1}^{N}[1 \pm \cos(2\omega l)]\mathrm{e}^{-2l\alpha} \\ &\approx \frac{1}{2}\sum_{l=1}^{N}\mathrm{e}^{-2l\alpha} \\ &= \frac{1}{2}\left[\frac{1-\mathrm{e}^{-2N\alpha}}{\mathrm{e}^{2\alpha}-1}\right]. \end{aligned} \tag{6.4}$$

Second, the off diagonal terms are at most the same order as the ignored terms; these terms are therefore ignored. Thus the matrix g_{ij} can be approximated as

$$g_{ij} \approx \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}.$$

The orthonormal model functions may then be written as

$$H_1(l) = \cos(\omega l)\mathrm{e}^{-\alpha l}/\sqrt{c} \tag{6.5}$$

$$H_2(l) = \sin(\omega l)\mathrm{e}^{-\alpha l}/\sqrt{c} \tag{6.6}$$

The projections of the data onto the orthonormal model functions (3.13) are given by

$$h_1 \equiv \frac{R(\omega,\alpha)}{\sqrt{c}} = \frac{1}{\sqrt{c}}\sum_{l=1}^{N} d_l \cos(\omega l)\mathrm{e}^{-\alpha l}$$

$$h_2 \equiv \frac{I(\omega,\alpha)}{\sqrt{c}} = \frac{1}{\sqrt{c}}\sum_{l=1}^{N} d_l \sin(\omega l)\mathrm{e}^{-\alpha l}$$

and the joint posterior probability of a frequency ω and a decay rate α is given by

$$P(\omega,\alpha|D,I) \propto \left[1 - \frac{R(\omega,\alpha)^2 + I(\omega,\alpha)^2}{Nc\overline{d^2}}\right]^{\frac{2-N}{2}}. \tag{6.7}$$

This approximation is valid provided there are plenty of data, $N \gg 1$, and there is no evidence of a low frequency. There is no restriction on the range of α: if $\alpha > 0$ the signal is decaying with increasing time, if $\alpha < 0$ the signal is growing with increasing time, and if $\alpha = 0$ the signal is stationary. This equation is analogous to (2.8) and reduces to (2.8) in the limit $\alpha \to 0$.

6.2.2 Accuracy Estimates

We derived a general estimate for the $\{\omega\}$ parameters in Chapter 4, and we would like to use those estimates for comparison with the single stationary frequency problem. To do this we can approximate the probability distribution $P(\omega,\alpha|D,\sigma,I)$ by a Gaussian as was done in Chapter 4. This may be done readily by assuming a form of the data, and then applying Eqs. (4.9) through (4.14). From the second derivative we may obtain the desired (mean) $\pm$ (standard deviation) estimates. Approximate second derivatives, usable with real data, may be obtained analytically as follows. We take as the data

$$d(t) = \hat{B}\cos(\hat{\omega}t)\mathrm{e}^{-\hat{\alpha}t}, \tag{6.8}$$

where $\hat{\omega}$ is the true frequency of oscillation and $\hat{\alpha}$ is the true decay rate. We have assumed only a cosine component to effect some simplifications in the discussion. It will be obvious at the end of the calculation that the result for a signal of arbitrary phase and magnitude may be obtained by replacing the amplitude $\hat{B}^2$ by the squared magnitude $\hat{B}^2 \to \hat{B}_1^2 + \hat{B}_2^2$.

The projection of the data (6.8) onto the model functions (6.5), and (6.6) is:

$$h_1 = \frac{\hat{B}}{2\sqrt{c}}\left[\sum_{l=1}^{N}\cos(\omega-\hat{\omega})l\mathrm{e}^{-(\alpha+\hat{\alpha})l} + \sum_{l=1}^{N}\cos(\omega+\hat{\omega})l\mathrm{e}^{-(\alpha+\hat{\alpha})l}\right].$$

The second term is negligible compared to the first under the conditions we have in mind. Likewise, the projection h_2 is essentially zero compared to h_1 and we have

ignored it. These sums may be done explicitly using (6.4) to obtain

$$h_1 = \frac{\hat{B}}{4\sqrt{c}} \left[\frac{1 - \mathrm{e}^{-2Nv}}{\mathrm{e}^{2v} - 1} + \frac{1 - \mathrm{e}^{-2Nu}}{\mathrm{e}^{2u} - 1} \right]$$

where

$$v = \frac{\alpha + \hat{\alpha} - i(\omega - \hat{\omega})}{2} \qquad \text{and} \qquad u = \frac{\alpha + \hat{\alpha} + i(\omega - \hat{\omega})}{2},$$

and $i = \sqrt{-1}$ in the above equations. Then the sufficient statistic $\overline{h^2}$ is given by:

$$\overline{h^2} = \frac{\hat{B}^2}{16} \left[\frac{\mathrm{e}^{2\alpha} - 1}{1 - \mathrm{e}^{-2N\alpha}} \right] \left[\frac{1 - \mathrm{e}^{-2Nv}}{1 - \mathrm{e}^{2v}} + \frac{1 - \mathrm{e}^{-2Nu}}{1 - \mathrm{e}^{2u}} \right]^2$$

The region of the parameter space we are interested in is where the unitless decay rate is small compared to one, and $\exp(N\hat{\alpha})$ is large compared to one. In this region the true signal decays away in the observation time, but not before we obtain a good representative sample of it. We are not considering the case were the decay is so slow that the signal is nearly stationary, or so fast that the signal is gone within a small fraction of the observation time. Within these limits the sufficient statistic $\overline{h^2}$ is

$$\overline{h^2} \approx \frac{\hat{B}^2 \alpha}{2} \left[\frac{\alpha + \hat{\alpha}}{(\alpha + \hat{\alpha})^2 + (\omega - \hat{\omega})^2} \right]^2 .$$

The first derivatives of $\overline{h^2}$ evaluated at $\omega = \hat{\omega}$ and $\alpha = \hat{\alpha}$ are zero, as they should be. The mixed second partial derivative is also zero, indicating that the presence of decay does not (to second order) shift the location of a frequency, and this of course explains why the discrete Fourier transform works on problems with decay. This gives the second derivatives of $\overline{h^2}$ as

$$b_\alpha \equiv - \left(\frac{\partial^2 \overline{h^2}}{\partial \alpha^2} \right)_{\alpha = \hat{\alpha}} = \frac{\hat{B}^2}{16 \hat{\alpha}^3} \qquad \text{and} \qquad b_\omega \equiv - \left(\frac{\partial^2 \overline{h^2}}{\partial \omega^2} \right)_{\omega = \hat{\omega}} = \frac{\hat{B}^2}{2 \hat{\alpha}^3}.$$

From these derivatives we then make the (mean) $\pm$ (standard deviation) estimates of the frequency and decay rate to obtain

$$(\alpha)_{\mathrm{est}} = \hat{\alpha} \pm \frac{\sigma}{\sqrt{b_\alpha}} \qquad \text{and} \qquad (\omega)_{\mathrm{est}} = \hat{\omega} \pm \frac{\sigma}{\sqrt{b_\omega}}$$

where

$$\frac{\sigma}{\sqrt{b_\alpha}} \approx \frac{2.8 \sigma \hat{\alpha}^{\frac{3}{2}}}{|\hat{B}|} \qquad \text{and} \qquad \frac{\sigma}{\sqrt{b_\omega}} \approx \frac{\sigma \hat{\alpha}^{\frac{3}{2}}}{|\hat{B}|}. \tag{6.9}$$

Converting to physical units, if the sampling rate is Δt and $\hat{\alpha}$ is now the true decay rate in hertz, these accuracy estimates are

$$\frac{\sigma}{\sqrt{b_\alpha}} \approx \frac{2.8 \sigma}{|\hat{B}|} \sqrt{\hat{\alpha}^3 \Delta t} \quad \text{Hertz} \qquad \text{and} \qquad \frac{\sigma}{\sqrt{b_\omega}} \approx \frac{\sigma}{|\hat{B}|} \sqrt{\hat{\alpha}^3 \Delta t} \quad \text{Hertz.}$$

Just as with the single frequency problem the accuracy depends on the signal-to-noise ratio and on the amount of data. In the single frequency case the amount of data was represented by the factor of N. Here the amount of data depends on two factors: the true decay rate $\hat{\alpha}$, and the sampling time Δt. The only factor the experimenter can typically control is the sampling time Δt. With a decaying signal, to improve the accuracy of the parameter estimates one must take the data faster, thus ensuring that the data are sampled in the region where the signal is large, or one must improve the signal-to-noise ratio of the data.

How does this compare to the results obtained before for the simple harmonic frequency? For a signal with $N = 1000$, a decay rate of $\hat{\alpha} = 2$Hz, $\hat{B}/\sqrt{2}\sigma = 1$, and again taking data for 1 second gives the accuracy estimates for frequency and decay as

$$(\omega)_{\text{est}} = \hat{\omega} \pm 0.06 \text{ Hz} \qquad \text{and} \quad (\alpha)_{\text{est}} = \hat{\alpha} \pm 0.17 \text{ Hz}.$$

The uncertainty in ω is 0.13Hz compared to 0.025Hz for an equivalent stationary signal with the same signal-to-noise ratio. This is a factor of 2.4 times larger than for a stationary sinusoid, and since the error varies like $N^{-\frac{3}{2}}$ we have effectively lost all but one third of the data due to decay. When we have reached the unitless time of $t = 250$ the signal is down by a factor of 12 and has all but disappeared into the noise. Again, the results of probability theory correspond nicely to the indications of common sense – but they are quantitative where common sense is not.

6.2.3 Example – One Frequency with Decay

To illustrate some of these points we been making, we have prepared two more examples: first we will investigate the use of this probability density when the decay mode is known and second when it is unknown.

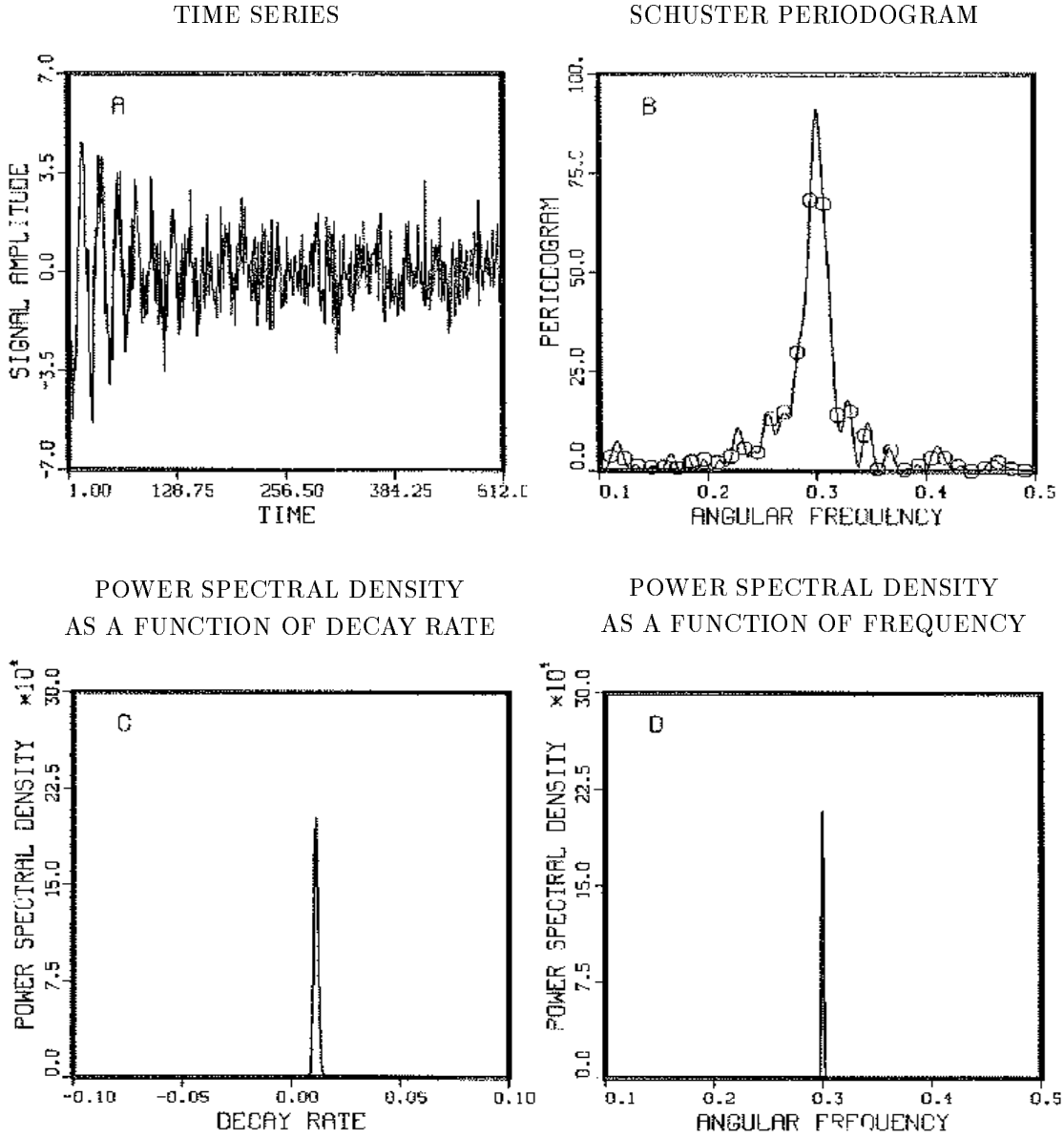
Known method of decay

Figure 6.8 is an example of the use of the posterior probability (6.7) when the signal is known to be harmonic with Lorentzian decay. This time series was prepared from the following equation

$$d_j = 0.001 + \cos(0.3j + 1)\mathrm{e}^{-0.01j} + e_j. \tag{6.10}$$

The $N = 512$ data samples were prepared in the following manner: first, we generated the data without the noise; we then computed the average of the data, and subtracted

Figure 6.8: Single Frequency with Lorentzian Decay



The data (A) contain a simple frequency with a Lorentzian decay plus noise. In (B) the noise has significantly distorted the periodogram (continuous curve) and the fast Fourier transform (open circles). The power spectral density may be computed as a function of decay rate α by integrating over the frequency (C), or as a function of frequency ω by integrating over the decay (D).

it from each data point, to ensure that the average of the data is zero; we then repeated this process on the Gaussian white noise; next, we scaled the computer generated signal by the appropriate ratio to make the signal-to-noise ratio of the data analyzed exactly one. The time series clearly shows a small signal which rapidly decays away, Fig. 6.8(A). Figure 6.8(B), the periodogram (continuous curve) and the fast Fourier transform (open circles) clearly show the Lorentzian line shape. The noise is now significantly affecting the periodogram: the periodogram is no longer an optimum frequency estimator.

Figures 6.8(C) and 6.8(D) contain plots of the power spectral density (4.16). In Fig. 6.8(C) we have treated the frequency as a nuisance parameter and have integrated it out numerically; as was emphasized earlier this is essentially the posterior probability distribution for α normalized to a power level rather than to unity. In Fig. 6.8(D) we have treated the decay as the nuisance parameter and have integrated it out. This gives the power spectral estimate as a function of frequency.

The width of these curves is a measure of the uncertainty in the determination of the parameters. We have determined full-width at half maximum (numerically) for each of these and have compared these to the theoretical "best" estimates (6.9) and find

$$(\omega)_{\rm est} = 0.2998 \pm 5 \times 10^{-4} \qquad \text{and} \qquad (\omega)_{\rm best} = 0.3000 \pm 6 \times 10^{-4},$$

$$(\alpha)_{\rm est} = 0.0109 \pm 1.6 \times 10^{-3} \qquad \text{and} \qquad (\alpha)_{\rm best} = 0.0100 \pm 1.6 \times 10^{-3}.$$

The theoretical estimates and those calculated from these data are effectively identical.

Unknown Method of Decay

Now what effect does not knowing the true model have on the estimated accuracy of these parameters? To test this we have analyzed the signal from Fig. 6.8 using four different models and have summarized the results in Table . There are several significant observations about the accuracy estimates; including a decay mode does not significantly affect the frequency estimates; however it does improve the accuracy estimates for the frequency as well as the estimated standard deviation of the noise σ, but not very much.

As we had expected, the Gaussian decay does not fit the data well: it decays away too fast, and the accuracy estimates are a little poorer. As with the single

Table 6.1: The Effect of Not Knowing the Decay Mode

Description	model	frequency ω	σ	$P(f_j\|D,I)$
Stationary:	$B\cos(\omega t+\theta)$	$0.3001 \pm 6\times10^{-4}$	1.260	8.3×10^{-33}
Gaussian in time:	$B\cos(\omega t+\theta)\mathrm{e}^{-\alpha t^2}$	$0.2991 \pm 7\times10^{-4}$	0.993	6.5×10^{-4}
Lorentzian in time:	$\dfrac{B\cos(\omega t+\theta)}{1+\alpha t^2}$	$0.2998 \pm 5\times10^{-4}$	0.978	0.0027
Lorentzian in frequency:	$B\cos(\omega t+\theta)\mathrm{e}^{-\alpha t}$	$0.2998 \pm 5\times10^{-4}$	0.979	0.9972

We analyzed the single frequency plus decay data (6.10) using four different decay models: stationary harmonic frequency, Gaussian decay, Lorentzian in time, and last Lorentzian in frequency. The stationary harmonic frequency model (first row) gives a poor estimate of the standard deviation of the noise, and consequently the estimated uncertainty of the frequency is larger. The probability of this model is so small that one would not even consider this as a possible model of the data. The second model is a single frequency with Gaussian decay. Here the estimated standard deviation of the noise is accurate, but the model fits the data poorly; thus the relative probability of this model effectively eliminates it from consideration. The third model is a single frequency with a Lorentzian decay in time. The relative probability of this model is also small indicating that although it is better than the two previous models, it is not nearly as good as the last model. The last model is a single frequency with Lorentzian decay. The relative probability of the model is effectively one, within the class of models considered.

harmonic frequency problem when we were demonstrating the effects of violating the assumptions, nothing startling happens here and maybe that is the most startling thing of all. Because it means that we do not have to know the exact models to make significant progress on analyzing the data. All we need are models which are reasonable for the data; i.e. models which take on most of the characteristics of the data.

The last column in this table is the relative probability of the various models (5.1). The relative probability of the single harmonic frequency model, 8.3×10^{-33} completely rules this model out as a possible explanation of these data. This is again not surprising: one can look at the data and see that it is decaying away. This small probability is just a quantitative way of stating a conclusion that we draw so easily without any probability theory. The Gaussian model fits the data much better, 6.5×10^{-4}, but not as well as the two Lorentzian models. The Lorentzian model in time has only about one chance in 500 of being "right" (i.e. of providing a better description of future data than the Lorentzian in frequency). Thus probability theory can rank various models according to how well they fit the data, and discriminates easily between models which predict only slightly different data.

6.3 Two Harmonic Frequencies

We now turn our attention to the slightly more general problem of analyzing a data set which we postulate contains two distinct harmonic frequencies. The "Student t-distribution" represented by (3.17) is, of course, the general solution to this problem. Unfortunately, that equation does not lend itself readily to understanding intuitively what is in the probability distribution. In particular we would like to know the behavior of these equations in three different limits: first, when the frequencies are well separated; second, when they are close but distinct; and third, when they are so close as to be, for all practical purposes, identical. To investigate these we will solve, approximately, the two stationary frequency problem.

6.3.1 The "Student t-Distribution"

The model equation for the two-frequency problem is a simple generalization of

the single-harmonic problem:

$$f(t) = B_1 \cos(\omega_1 t) + B_2 \cos(\omega_2 t) + B_3 \sin(\omega_1 t) + B_4 \sin(\omega_2 t).$$

The model functions can then be used to construct the g_{jk} matrix. On a uniform grid this is given by

$$g_{jk} = \begin{pmatrix} c_{11} & c_{12} & 0 & 0 \\ c_{12} & c_{22} & 0 & 0 \\ 0 & 0 & s_{11} & s_{12} \\ 0 & 0 & s_{12} & s_{22} \end{pmatrix}$$

where

$$c_{jk} = \sum_{l=-T}^{T} \cos(\omega_j l)\cos(\omega_k l) = \frac{\sin(\frac{1}{2}N\omega_+)}{2\sin(\frac{1}{2}\omega_+)} + \frac{\sin(\frac{1}{2}N\omega_-)}{2\sin(\frac{1}{2}\omega_-)} \tag{6.11}$$

$$s_{jk} = \sum_{l=-T}^{T} \sin(\omega_j l)\sin(\omega_k l) = \frac{\sin(\frac{1}{2}N\omega_-)}{2\sin(\frac{1}{2}\omega_-)} - \frac{\sin(\frac{1}{2}N\omega_+)}{2\sin(\frac{1}{2}\omega_+)} \tag{6.12}$$

$$\begin{aligned} \omega_+ &= \omega_j + \omega_k, \qquad (j,k = 1 \text{ or } 2) \\ \omega_- &= \omega_j - \omega_k. \end{aligned}$$

The eigenvalue and eigenvector problem for g_{jk} splits into two separate problems, each involving 2×2 matrices. The eigenvalues are:

$$\lambda_1 = \frac{c_{11}+c_{22}}{2} + \sqrt{(c_{11}-c_{22})^2 + 4c_{12}^2}, \qquad \lambda_2 = \frac{c_{11}+c_{22}}{2} - \sqrt{(c_{11}-c_{22})^2 + 4c_{12}^2},$$

$$\lambda_3 = \frac{s_{11}+s_{22}}{2} + \sqrt{(s_{11}-s_{22})^2 + 4s_{12}^2}, \quad \text{and} \quad \lambda_4 = \frac{s_{11}+s_{22}}{2} - \sqrt{(s_{11}-s_{22})^2 + 4s_{12}^2}.$$

Well Separated Frequencies

When the frequencies are well separated $|\omega_1 - \omega_2| \gg 2\pi/N$, the eigenvalues reduce to $\lambda = N/2$. That is, g_{jk} goes into $N/2$ times the unit matrix. Then the model functions are effectively orthogonal and the sufficient statistic $\overline{h^2}$ reduces to

$$\overline{h^2} = \frac{2}{N}\left[C(\omega_1) + C(\omega_2)\right].$$

The joint posterior probability, when the variance is known, is given by

$$P(\omega_1, \omega_2 | D, \sigma, I) \propto \exp\left[\frac{C(\omega_1) + C(\omega_2)}{\sigma^2}\right]. \tag{6.13}$$

The problem has separated: one can estimate each of the frequencies separately. The maximum of the two-frequency posterior probability density will be located at the two greatest peaks in the periodogram, in agreement with the common sense usage of the discrete Fourier transform.

Two Very Close Frequencies

The labels ω_1, ω_2, etc. for the frequencies in the model are arbitrary, and accordingly their joint probability density is invariant under permutations. That means, for the two-frequency problem, there is an axis of symmetry running along the line $\omega_1 = \omega_2$. We do not know from (6.13) what is happening along that line. This is easily investigated: when $\omega_1 = \omega_2 \equiv \omega$ the eigenvalues become

$$\lambda_1 = N, \qquad \lambda_2 = 0, \qquad \lambda_3 = N, \qquad \lambda_4 = 0.$$

The matrix g_{jk} has two redundant eigenvalues, and the probability distribution becomes

$$P(\omega|D,\sigma,I) \propto \exp\left\{\frac{C(\omega)}{\sigma^2}\right\}. \tag{6.14}$$

The probability density goes smoothly into the single frequency probability distribution along this axis of symmetry. Given that the two frequencies are equal, our estimate of them will be identical, in value and accuracy, to those of the one frequency case. In the exact solution, the factor of two that we would have if we attempted to use (6.13) where it is not valid, is just cancelled out.

Close But Distinct Frequencies

We have not yet addressed the posterior probability density when there are two close but distinct frequencies. To understand this aspect of the problem we could readily diagonalize the matrix g_{jk} and obtain the exact solution. However, just like the single frequency case with Lorentzian decay, this would be extremely tedious and not very productive. Instead we derive an approximate solution which is simpler and valid nearly everywhere if N is large. To obtain this approximate solution one needs only to examine the matrix g_{jk} and notice that the elements of this matrix consist of the diagonal elements given by:

$$c_{11} = \frac{N}{2} + \frac{\sin(N\omega_1)}{2\sin(\omega_1)} \approx \frac{N}{2},$$

$$c_{22} = \frac{N}{2} + \frac{\sin(N\omega_2)}{2\sin(\omega_2)} \approx \frac{N}{2},$$

$$s_{11} = \frac{N}{2} - \frac{\sin(N\omega_1)}{2\sin(\omega_1)} \approx \frac{N}{2},$$

$$s_{22} = \frac{N}{2} - \frac{\sin(N\omega_2)}{2\sin(\omega_2)} \approx \frac{N}{2},$$

and the off-diagonal elements. The off-diagonal terms are small compared to N unless the frequencies are specifically in the region of $\omega_1 \approx \omega_2$; then only the terms involving the difference $(\omega_1 - \omega_2)$ are large. We can approximate the off diagonal terms as:

$$c_{12} \approx s_{12} \approx \frac{1}{2}\sum_{l=-T}^{T} \cos\frac{1}{2}(\omega_1 - \omega_2)l = \frac{1}{2}\frac{\sin\frac{1}{2}N(\omega_1 - \omega_2)}{\sin\frac{1}{2}(\omega_1 - \omega_2)} \equiv \frac{B}{2}. \tag{6.15}$$

When the two frequencies are well separated, (6.15) is of order one and is small compared to the diagonal elements. When the two frequencies are nearly equal, then the off-diagonal terms are large and are given accurately by (6.15). So the approximation is valid for all values of ω_1 and ω_2 that are not extremely close to zero and π.

With this approximation for g_{jk} it is now possible to write a simplified solution for the two-frequency problem. The matrix g_{jk} is given approximately by

$$g_{jk} = \frac{1}{2}\begin{pmatrix} N & B & 0 & 0 \\ B & N & 0 & 0 \\ 0 & 0 & N & B \\ 0 & 0 & B & N \end{pmatrix}.$$

The orthonormal model functions (3.5) may now be constructed:

$$H_1(t) = \frac{1}{\sqrt{N+B}}\left[\cos(\omega_1 t) + \cos(\omega_2 t)\right], \tag{6.16}$$

$$H_2(t) = \frac{1}{\sqrt{N-B}}\left[\cos(\omega_1 t) - \cos(\omega_2 t)\right],$$

$$H_3(t) = \frac{1}{\sqrt{N+B}}\left[\sin(\omega_1 t) + \sin(\omega_2 t)\right],$$

$$H_4(t) = \frac{1}{\sqrt{N-B}}\left[\sin(\omega_1 t) - \sin(\omega_2 t)\right].$$

We can write the sufficient statistic $\overline{h^2}$ in terms of these orthonormal model functions to obtain

$$\overline{h^2} = h_+^2 + h_-^2,$$

$$h_+^2 \equiv \frac{1}{4(N+B)}\left\{[R(\omega_1) + R(\omega_2)]^2 + [I(\omega_1) + I(\omega_2)]^2\right\},$$

$$h_-^2 \equiv \frac{1}{4(N-B)}\left\{[R(\omega_1) - R(\omega_2)]^2 + [I(\omega_1) - I(\omega_2)]^2\right\},$$

where R and I are the sine and cosine transforms of the data as functions of the appropriate frequency. The factor of 4 comes about because for this problem there are $m = 4$ model functions. Using (3.15), the posterior probability that two distinct frequencies are present given the noise variance σ^2 is

$$P(\omega_1, \omega_2 | D, \sigma, I) \propto \exp\left\{\frac{2\overline{h^2}}{\sigma^2}\right\}. \tag{6.17}$$

A quick check on the asymptotic forms of this will verify that when the frequencies are well separated one has $\overline{h^2} = \frac{1}{2}[C(\omega_1) + C(\omega_2)]$, and it has reduced to (6.13). Likewise, when the frequencies are the same the second term goes smoothly to zero, and the first term goes into $\frac{1}{2}C(\omega)$, to reduce to (6.14) as expected.

6.3.2 Accuracy Estimates

When the frequencies are very close or far apart we can apply the results obtained by Jaynes [12] concerning the accuracy of the frequency estimates:

$$(\omega)_{\text{est}} = \hat{\omega} \pm \frac{\sigma}{\hat{B}}\sqrt{48/N^3}. \tag{6.18}$$

In the region where the frequencies are close but distinct, (6.17) appears very different. We would like to understand what is happening in this region, in particular we would like to know just how well two close frequencies can be estimated. To understand this we will construct a Gaussian approximation similar to what was done for the case with Lorentzian decay. We Taylor expand the $\overline{h^2}$ in (6.17) to obtain

$$P(\omega_1, \omega_2 | D, \sigma, I) \approx \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{2}\sum_{k=1}^{2} b_{jk}(\omega_j - \hat{\omega}_j)(\omega_k - \hat{\omega}_k)\right\}$$

where

$$b_{11} = -2\frac{\partial^2\overline{h^2}}{\partial\omega_1^2}\Big|_{\substack{\omega_1=\hat{\omega}_1 \\ \omega_2=\hat{\omega}_2}}$$

$$b_{22} = -2\frac{\partial^2\overline{h^2}}{\partial\omega_2^2}\Big|_{\substack{\omega_1=\hat{\omega}_1 \\ \omega_2=\hat{\omega}_2}}$$

$$b_{12} = -2\frac{\partial^2\overline{h^2}}{\partial\omega_1\partial\omega_2}\Big|_{\substack{\omega_1=\hat{\omega}_1 \\ \omega_2=\hat{\omega}_2}}$$

where $\hat{\omega}_1$, $\hat{\omega}_2$ are the locations of the maxima of (6.17). If we have uniformly sampled data of the form

$$d_l = \hat{A}_1 \cos(\hat{\omega}_1 l) + \hat{A}_2 \cos(\hat{\omega}_2 l) + \hat{A}_3 \sin(\hat{\omega}_1 l) + \hat{A}_4 \sin(\hat{\omega}_2 l) \tag{6.19}$$

where $-T \leq l \leq T$, $2T+1 = N$, $\hat{A}_1$, $\hat{A}_2$, $\hat{A}_3$, $\hat{A}_4$ are the true amplitudes, and $\hat{\omega}_1$, $\hat{\omega}_2$ are the true frequencies, then h_j is given by the projection of H_j (6.16) onto the data (6.19) to obtain

$$h_j = \frac{1}{\sqrt{N + B(\omega_1, \omega_2)}} \sum_{l=-T}^{T} H_j(t_l) d_l$$

where

$$\frac{B(\omega_1, \omega_2)}{2} \equiv \frac{1}{2} \sum_{l=-T}^{T} \cos(\omega_1 - \omega_2) l = \frac{1}{2} \frac{\sin \frac{1}{2} N(\omega_1 - \omega_2)}{\sin \frac{1}{2}(\omega_1 - \omega_2)}. \tag{6.20}$$

For a uniform time series these h_j may be summed explicitly using (6.20) to obtain

$$\begin{aligned}
h_1 = \frac{1}{2\sqrt{N + B(\omega_1, \omega_2)}} \quad \times \quad & \Big\{ \hat{A}_1 [B(\hat{\omega}_1, \omega_1) + B(\hat{\omega}_1, \omega_2)] \\
+ \quad & \hat{A}_2 [B(\hat{\omega}_2, \omega_1) + B(\hat{\omega}_2, \omega_2)] \Big\} \\
h_2 = \frac{1}{2\sqrt{N - B(\omega_1, \omega_2)}} \quad \times \quad & \Big\{ \hat{A}_1 [B(\hat{\omega}_1, \omega_1) - B(\hat{\omega}_1, \omega_2)] \\
+ \quad & \hat{A}_2 [B(\hat{\omega}_2, \omega_1) - B(\hat{\omega}_2, \omega_2)] \Big\} \\
h_3 = \frac{1}{2\sqrt{N + B(\omega_1, \omega_2)}} \quad \times \quad & \Big\{ \hat{A}_3 [B(\hat{\omega}_1, \omega_1) + B(\hat{\omega}_1, \omega_2)] \\
+ \quad & \hat{A}_4 [B(\hat{\omega}_2, \omega_1) + B(\hat{\omega}_2, \omega_2)] \Big\} \\
h_4 = \frac{1}{2\sqrt{N - B(\omega_1, \omega_2)}} \quad \times \quad & \Big\{ \hat{A}_3 [B(\hat{\omega}_1, \omega_1) - B(\hat{\omega}_1, \omega_2)] \\
+ \quad & \hat{A}_4 [B(\hat{\omega}_2, \omega_1) - B(\hat{\omega}_2, \omega_2)] \Big\}.
\end{aligned}$$

We have kept terms corresponding to the differences in the frequencies. When the frequencies are close together it is only these terms which are important: the approximation is consistent with the others made.

The sufficient statistic $\overline{h^2}$ is then given by

$$\overline{h^2} = \frac{1}{4}(h_1^2 + h_2^2 + h_3^2 + h_4^2). \tag{6.21}$$

To obtain a Gaussian approximation for (6.17) one must calculate the second derivative of (6.21) with respect to ω_1 and ω_2. The problem is simple in principle but tedious in practice. To get these partial derivatives, we Taylor expand (6.21) around the maximum located at $\hat{\omega}_1$ and $\hat{\omega}_2$ and then take the derivative. The intermediate steps are of little concern and were carried out using an algebra manipulation package. Terms of order one compared to N were again ignored, and we have assumed the frequencies are close but distinct, also we used the small angle approximations for the sine and cosine at the end of the calculation. The local variable δ [defined as $(\hat{\omega}_2 - \hat{\omega}_1)/2 \equiv \delta/N$] measures the distance between two adjacent frequencies. If δ is π then the frequencies are separated by one step in the discrete Fourier transform. The second partial derivatives of $\overline{h^2}$ evaluated at the maximum are given by:

$$b_{11} \approx (\hat{A}_1^2 + \hat{A}_3^2)N^3 \left(\frac{3\sin^2\delta - 6\delta\cos\delta\sin\delta + \delta^2[\sin^2\delta + 3\cos\delta] - \delta^4}{24\delta^3[\sin\delta - \delta][\sin\delta + \delta]} \right)$$

$$b_{22} \approx (\hat{A}_2^2 + \hat{A}_4^2)N^3 \left(\frac{3\sin^2\delta - 6\delta\cos\delta\sin\delta + \delta^2[\sin^2\delta + 3\cos\delta] - \delta^4}{24\delta^3[\sin\delta - \delta][\sin\delta + \delta]} \right)$$

$$b_{12} \approx (\hat{A}_1\hat{A}_2 + \hat{A}_3\hat{A}_4)N^3 \left(\frac{\delta^4\sin\delta + 2\delta^3\cos\delta - 3\delta^2\sin\delta + \sin^3\delta}{8\delta^3[\sin\delta - \delta][\sin\delta + \delta]} \right).$$

If the true frequencies $\hat{\omega}_1$ and $\hat{\omega}_2$ are separated by two steps in the discrete Fourier transform, $\delta = 2\pi$, we may reasonably ignore all but the δ^4 term to obtain

$$b_{11} \approx \frac{(\hat{A}_1^2 + \hat{A}_3^2)N^3}{24}$$

$$b_{22} \approx \frac{(\hat{A}_2^2 + \hat{A}_4^2)N^3}{24}$$

$$b_{12} \approx \frac{(\hat{A}_1\hat{A}_2 + \hat{A}_3\hat{A}_4)N^3\sin(\delta)}{8\delta}.$$

Having the mixed partial derivatives we may now apply the general formalism (4.14) to obtain

$$(\omega_1)_{\text{est}} = \hat{\omega}_1 \pm \sqrt{\frac{48\sigma^2}{N^3(\hat{A}_1^2 + \hat{A}_3^2)\left\{1 - \frac{9(\hat{A}_1\hat{A}_2 + \hat{A}_3\hat{A}_4)^2\sin^2(\delta)/\delta^2}{4(\hat{A}_1^2 + \hat{A}_3^2)(\hat{A}_2^2 + \hat{A}_4^2)}\right\}}}$$

$$(\omega_2)_{\text{est}} = \hat{\omega}_2 \pm \sqrt{\frac{48\sigma^2}{N^3(\hat{A}_2^2 + \hat{A}_4^2)\left\{1 - \frac{9(\hat{A}_1\hat{A}_2 + \hat{A}_3\hat{A}_4)^2\sin^2(\delta)/\delta^2}{4(\hat{A}_1^2 + \hat{A}_3^2)(\hat{A}_2^2 + \hat{A}_4^2)}\right\}}}$$

The accuracy estimates reduce to (6.18) when the frequencies are well separated. When the frequencies have approximately the same amplitudes and δ is order of 2π (the frequencies are separated by two steps in the fast Fourier transform) the interaction term is down by approximately 1/36; and one expects the estimates to be nearly the same as those for a single frequency. Probability theory indicates that two frequencies which are as close together as two steps in a discrete Fourier transform do not interfere with each other in any significant way. Also note the appearance of the sinc function in the above estimates. When the frequencies are separated by a Nyquist step ($|\omega_1 - \omega_2| = 2\pi/N$) the frequencies cannot interfere with each other. Although this is a little surprising at first sight, a moment's thought will convince one that when the frequencies are separated by $2\pi/N$ the sampled vectors are exactly orthogonal to each other and because we are effectively taking dot products between the model and the data, of course they cannot interfere with each other.

6.3.3 More Accuracy Estimates

To better understand the maximum theoretical accuracy with which two frequencies can be estimated we have prepared Table 6.2. To make these estimates comparable to those obtained in Chapter 2 we have again assumed $N = 1000$ data points and $\sigma = 1$. There are three regions of interest: when the frequency separation is small compared to a single step in the discrete Fourier transform; when the separation is of order one step; and when the separation is large. Additionally we would like to understand the behavior when the signals are of the same amplitude, when one signal is slightly larger than the other, and when one signal is much larger than the other. When we prepared this table we used the joint posterior probability of two frequencies (3.16) assuming the variance σ^2 known. The estimates obtained are the "best" in the sense that in a real data set with $\sigma = 1$, and $N = 1000$ data points the accuracy estimates one obtains will be, nearly always, slightly worse than those contained in table 6.2.

The three values of $(\omega_1 - \omega_2)$ examined correspond to $\delta = 1/4$, $\delta = 4$, and $\delta = 16$: roughly these correspond to frequency separations of 0.07, 0.3, and 5.1 Hz. We held the squared magnitude of one signal constant, and the second is either 1, 4 or 128 times larger.

When the separation frequency is 0.07 Hz the frequencies are indistinguishable. The smaller component cannot be estimated accurately. As the magnitude of the

Table 6.2: Two Frequency Accuracy Estimates

$\frac{\sqrt{B_2^2+B_4^2}}{\sqrt{B_1^2+B_3^2}}$	$\Delta f = 0.07$ Hz		$\Delta f = 0.3$ Hz		$\Delta f = 5.1$ Hz	
	$\delta\hat{f}_1$ Hz	$\delta\hat{f}_2$ Hz	$\delta\hat{f}_1$ Hz	$\delta\hat{f}_2$ Hz	$\delta\hat{f}_1$ Hz	$\delta\hat{f}_2$ Hz
1	± 0.091	± 0.091	± 0.027	± 0.027	± 0.025	± 0.025
4	± 0.091	± 0.088	± 0.027	± 0.013	± 0.025	± 0.012
128	± 0.091	± 0.034	± 0.025	± 0.0024	± 0.025	± 0.0022

We ran a number of simulations to determine how well two frequencies could be determined. In column 1 the two frequencies are separated by only 0.07 Hz and cannot be resolved. In column 2 the separation frequency is now 0.3 Hz and the resolution is approximately 0.0025 Hz for each of the three amplitudes tested. We would have to move one of the frequencies by 11 standard deviations before they would overlap each other. In column 3 the frequencies are separated by 5.1 Hz and we would have to move one of the frequencies by 200 standard deviations before they overlapped.

second signal increases, the estimated accuracy of the second signal becomes better as one's intuition would suppose it should (the signal looks more and more like one frequency). But even at 128:1 probability theory still senses that all is not quite right for a single frequency, and gives an accuracy estimate wider than for a true one frequency signal. However, for very close frequencies the true resolving power is conveyed only by the two-dimensional plot like Fig. 6.10 below; not by the numbers in Table 6.2.

When the separation frequency is 0.3 Hz or about one step in the discrete Fourier transform, the accuracy estimates indicate that the two frequencies are well resolved. By this we mean one of the frequencies would have to be moved by 11 standard deviations before it would be confounded with the other (two parameters are said to be confounded when probability theory cannot distinguish their separate values). This is true for all sample signals in the table; it does, however, improve with increasing amplitude. According to probability theory, when two frequencies are as close together as one Nyquist step in the discrete Fourier transform, those frequencies are clearly resolvable by many standard deviations even at $S/N = 1$; the Rayleigh criterion is far surpassed.

When the separation frequency is 5.1Hz, the accuracy estimates determine both frequencies slightly better. Additionally, the accuracy estimates for the smaller frequency are essentially 0.025Hz which is the same as the estimate for a single harmonic

frequency that we found previously (2.12). Examining Table 6.2 more carefully, we see that when the frequencies are separated by even a single step in the discrete Fourier transform, the accuracy estimates are essentially those for the single harmonic frequencies. The ability to estimate two close frequencies accurately is essentially independent of the separation frequency, as long as it is greater than or approximately equal to one step in the discrete Fourier transform!

6.3.4 The Power Spectral Density

The power spectral density (4.16) specifically assumed there were no confounded parameters. The exchange symmetry, in the two-harmonic frequency problem, ensures there are two equally probable maxima. We must generalize (4.16) to account for these. The generalization is straightforward. We have from (4.16)

$$\hat{p}(\{\omega\}) \approx 4\overline{h^2}\frac{P(\{\omega\}|D,\langle\sigma^2\rangle,I)}{\int d\{\omega\}P(\{\omega\}|D,\langle\sigma^2\rangle,I)}. \tag{6.22}$$

The generalization is in the approximating of $P(\{\omega\}|D,\langle\sigma^2\rangle,I)$. Suppose for simplicity that the two frequencies are well separated and the variance σ^2 is known; then the matrix b_{jk} becomes

$$b_{jk} = -\frac{m}{2}\frac{\partial^2\overline{h^2}}{\partial\omega_j^2}\delta_{jk}.$$

Which gives

$$\begin{aligned}\frac{P(\{\omega\}|D,\langle\sigma^2\rangle,I)}{\int d\{\omega\}P(\{\omega\}|D,\langle\sigma^2\rangle,I)} &\approx \left(\frac{b_{11}}{2\pi\langle\sigma^2\rangle}\frac{b_{22}}{2\pi\langle\sigma^2\rangle}\right)^{\frac{1}{2}} \\ &\times \exp\left\{-\frac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega_1)^2-\frac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega_2)^2\right\}\end{aligned}$$

when expanded around $\omega_1 \approx \hat{\omega}_1$, $\omega_2 \approx \hat{\omega}_2$, and

$$\begin{aligned}\frac{P(\{\omega\}|D,\langle\sigma^2\rangle,I)}{\int d\{\omega\}P(\{\omega\}|D,\langle\sigma^2\rangle,I)} &\approx \left(\frac{b_{11}}{2\pi\langle\sigma^2\rangle}\frac{b_{22}}{2\pi\langle\sigma^2\rangle}\right)^{\frac{1}{2}} \\ &\times \exp\left\{-\frac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega_2)^2-\frac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega_1)^2\right\}\end{aligned}$$

when we expand around the other maximum. But to be consistent we must retain the same symmetries in the approximation to the probability density as it originally

possessed: the approximation which is valid everywhere is

$$\begin{aligned}\frac{P(\{\omega\}|D,\langle\sigma^2\rangle,I)}{\int d\{\omega\}P(\{\omega\}|D,\langle\sigma^2\rangle,I)} &\approx \frac{1}{2}\left(\frac{b_{11}}{2\pi\langle\sigma^2\rangle}\frac{b_{22}}{2\pi\langle\sigma^2\rangle}\right)^{\frac{1}{2}} \\ &\times \left[\exp\left\{-\frac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega_1)^2-\frac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega_2)^2\right\}\right. \\ &+ \left.\exp\left\{-\frac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega_2)^2-\frac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega_1)^2\right\}\right]\end{aligned}$$

The factor of $1/2$ comes about because there are two equally probable maxima. The power spectral density is a function of both ω_1 and ω_2, but we wish to plot it as a function of only one variable ω. We can do this by integrating out the nuisance parameter (in this case one of the two frequencies). From symmetry, it cannot matter which frequency we choose to eliminate. We choose to integrate out ω_1 and to relabel ω_2 as ω. Performing this integration we obtain

$$\begin{aligned}\hat{p}(\omega) &\approx 2\overline{h^2}(\hat{\omega}_2,\omega)\sqrt{\frac{b_{11}}{2\pi\langle\sigma^2\rangle}}\exp\left\{-\tfrac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega)^2\right\} \\ &+ 2\overline{h^2}(\hat{\omega}_1,\omega)\sqrt{\frac{b_{22}}{2\pi\langle\sigma^2\rangle}}\exp\left\{-\tfrac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega)^2\right\}\end{aligned}$$

and using the fact that

$$\overline{h^2}(\hat{\omega}_1,\hat{\omega}_2)=\overline{h^2}(\hat{\omega}_2,\hat{\omega}_1)=C(\hat{\omega}_1)+C(\hat{\omega}_2)$$

we have

$$\begin{aligned}\hat{p}(\omega) &\approx 2[C(\hat{\omega}_1)+C(\hat{\omega}_2)]\left[\sqrt{\frac{b_{11}}{2\pi\langle\sigma^2\rangle}}\exp\left\{-\tfrac{b_{11}}{2\langle\sigma^2\rangle}(\hat{\omega}_1-\omega)^2\right\}\right. \\ &+ \left.\sqrt{\frac{b_{22}}{2\pi\langle\sigma^2\rangle}}\exp\left\{-\tfrac{b_{22}}{2\langle\sigma^2\rangle}(\hat{\omega}_2-\omega)^2\right\}\right].\end{aligned}$$

We see now just what that exchange symmetry is doing: The power spectral density conveys information about the total energy carried by the signal, and about the accuracy of each line, but the two terms have equal areas; it contains no information about how much energy is carried in each line. That is not too surprising; after all we defined the power spectral density as the total energy carried by the signal per unit $\{\omega\}$. That is typically what one is interested in for an arbitrary model function.

However, the multiple frequency problem is unique in that one is typically interested in the power carried by each line; not the total power carried by the signal. This is not really a well defined problem in the sense that as two lines become closer and closer together the frequencies are no longer orthogonal and power is shared between them. The problem becomes even worse when one considers nonstationary frequencies. We will later define a line spectral density which will give information about the power carried by one line when there are multiple well separated lines in the spectrum.

6.3.5 Example – Two Harmonic Frequencies

To illustrate some of the points we have been making about the two-frequency probability density (6.17) we prepared a simple example, Fig. 6.9. This example was prepared from the following equation

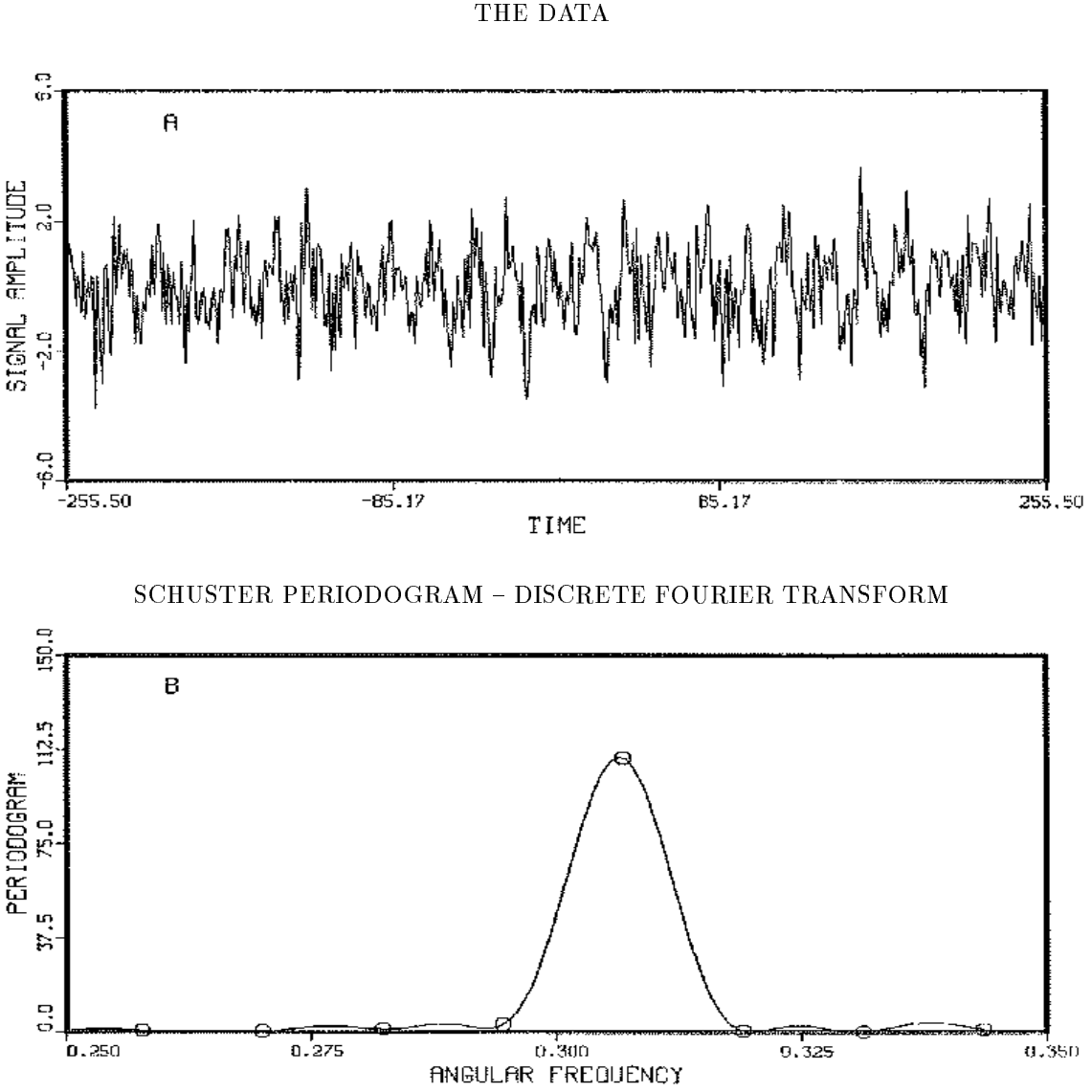
$$d_i = \cos(0.3i + 1) + \cos(0.307i + 2) + e_i$$

where e_i has variance one and the index runs over the symmetric time interval $(-255.5 \leq i \leq 255.5)$ by unit steps. This time series, Fig. 6.9(A), has two simple harmonic frequencies separated by approximately 0.6 steps in the discrete Fourier transform. One step corresponds to $|\hat{\omega}_1 - \hat{\omega}_2| \approx 2\pi/512 = 0.012$.

From looking at the raw time series one might just barely guess that there is more going on than a simple harmonic frequency plus noise, because the oscillation amplitude seems to vary slightly. If we were to guess that there are two close frequencies, then by examining the data one can guess that the difference between these two frequencies is not more than one cycle over the entire time interval. If the frequencies were separated by more than this we would expect to see beats in the data. If there are two frequencies, the second frequency must be within 0.012 of the first (in dimensionless units). This is in the region where the frequency estimates are almost but not quite confounded.

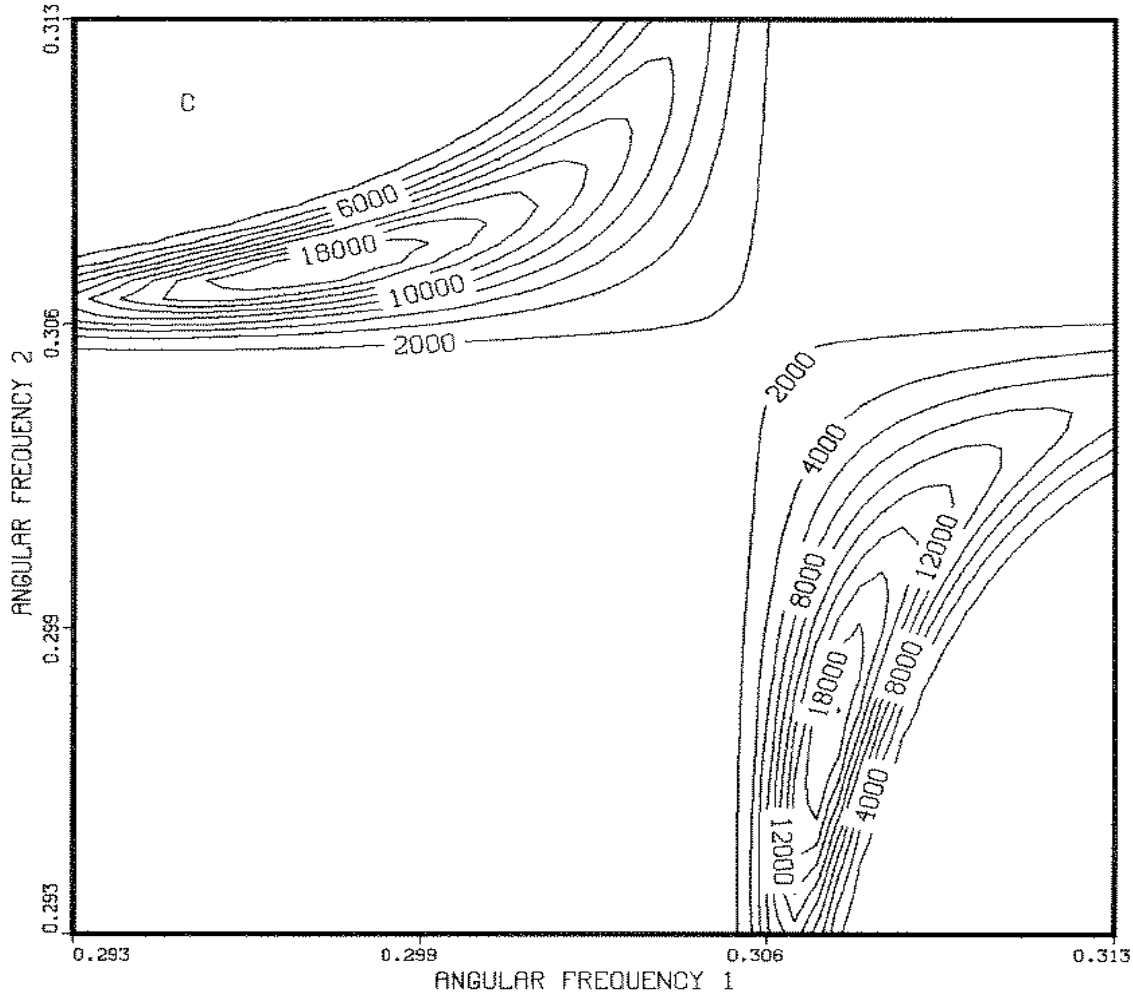
Now Fig. 6.9(B) the periodogram (continuous curve) and the fast Fourier transform (open circles) show only a single peak. The single frequency model has estimated a frequency which is essentially the average of the two. Yet the two frequency posterior probability density Fig. 6.10 shows two well resolved, symmetrical maxima. Thus the inclusion of just this one simple additional fact – that the signal may have two frequencies – has greatly enhanced our ability to detect the two signals. Prior information, even when it is only qualitative, can have a major effect on the quantitative

Figure 6.9: Two Harmonic Frequencies – The Data



The data (A) contain two frequencies. They are separated from each other by approximately a single step in the discrete Fourier transform. The periodogram (B) shows only a single peak located between the two frequencies.

Figure 6.10: Posterior Probability density of Two Harmonic Frequencies



This is a fully normalized posterior probability density of two harmonic frequencies in the data, Fig. 6.9. The two-frequency probability density clearly indicates the presence of two frequencies. The posterior odds ratio prefers the two-frequency model by 10^7 to 1.

conclusions we are able to draw from a given data set.

This plot illustrates numerically some of the points we have been making. First, in the two harmonic frequency probability density there are three discrete Fourier transforms: one along each axis, and a third along $\omega_1 = \omega_2$. The two transforms along the axes form ridges. If the frequencies are very close and have the same amplitude the ridges are located at the average of the two frequencies: $0.5(0.3 + 0.307) = 0.335$. The discrete Fourier transform along the line of symmetry $\omega_1 = \omega_2$ can almost be imagined. As we approach the true frequencies, $\omega_1 \approx 0.307$ and $\omega_2 \approx 0.3$, these ridges have a slight bend away from the value indicated by the discrete Fourier transform: these very close frequencies are not orthogonal. When the true frequencies are well separated, these ridges intersect at right angles (the cross derivatives are zero) and the frequencies do not interfere with each other. Even now, two very close frequencies do not interfere greatly.

According to probability theory, the odds in favor of the two-frequency model compared to the one-frequency model are 10^7 to 1. Now that we know the data contain two partially resolved frequencies, we could proceed to obtain data over a longer time span and resolve the frequencies still better. Regardless, it is now clear that what one can detect clearly depends on what question one asks, and thus on what prior information we have to suggest the best questions.

6.4 Estimation of Multiple Stationary Frequencies

The problem of estimating multiple stationary harmonic frequencies can now be addressed. The answer to this problem is, of course, given by the "Student t-distribution" Eq. (3.17) using

$$f(t) = \sum_{j=1}^{r} B_j \cos \omega_j t + \sum_{j=1}^{r} B_{r+j} \sin \omega_j t \tag{6.23}$$

as a model. No exact analytic solution to this problem exists for more than a few frequencies. However, a number of interesting things can be learned by studying this problem.

6.5 The "Student t-Distribution"

We begin this process by calculating the g_{ij} matrix explicitly. For a uniformly sampled time series this is given by

$$g_{jk} = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1r} & 0 & \cdots & \cdots & 0 \\ c_{21} & c_{22} & \cdots & c_{2r} & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{r1} & c_{r2} & \cdots & c_{rr} & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & s_{11} & s_{12} & \cdots & s_{1r} \\ \vdots & \vdots & \vdots & \vdots & s_{21} & s_{22} & \cdots & s_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & s_{r1} & s_{r2} & \cdots & s_{rr} \end{pmatrix}$$

where c_{jk} and s_{jk} were defined earlier (6.11, 6.12). To investigate the full solution we first make the same large N approximations we made in the two-frequency problem.

When the frequencies are well separated, $|\omega_j - \omega_k| \gg 2\pi/N$, the diagonal elements are again replaced by $N/2$ and the off diagonal elements are given by $B(\omega_j, \omega_k)/2$, using the notation $B_{jk} \equiv B(\omega_j, \omega_k)$ defined earlier by Eq. (6.20). This simplifies the g_{jk} matrix somewhat:

$$g_{jk} = \frac{1}{2}\begin{pmatrix} N & B_{12} & \cdots & B_{1r} & 0 & \cdots & \cdots & 0 \\ B_{21} & N & \cdots & B_{2r} & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{r1} & B_{r2} & \cdots & N & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & N & B_{12} & \cdots & B_{1r} \\ \vdots & \vdots & \vdots & \vdots & B_{21} & N & \cdots & B_{2r} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & B_{r1} & B_{r2} & \cdots & N \end{pmatrix}$$

The problem separates into finding the eigenvalues and eigenvectors of an $r \times r$ matrix.

Multiple Well-Separated Frequencies

For convenience, assume the frequencies are ordered: $\omega_1 < \omega_2 < \cdots < \omega_r$. The exchange symmetries in this problem ensure that we can always do this. Now assume that $|\omega_j - \omega_k| \gg 2\pi/N$. The g_{jk} matrix will simplify significantly, because all of the off-diagonal elements are essentially zero:

$$g_{jk} \approx \frac{N}{2}\delta_{jk}.$$

The problem has separated completely, and the joint posterior probability of multiple harmonic frequencies when the variance is known is given by

$$P(\omega_1, \cdots, \omega_r | D\sigma, I) \propto \exp\left\{\sum_{j=1}^{r} \frac{C(\omega_j)}{\sigma^2}\right\}.$$

This result was first found by Jaynes [12]. The estimated frequencies are the r largest peaks in the discrete Fourier transform, again in agreement with common sense. Of course, the accuracy estimates of the frequencies are those obtained from the single harmonic frequency problem.

If one were to estimate the accuracy from multiple frequency data using a single-frequency model, the answers would not be the same; the estimated noise variance σ^2 would be far greater, because multiple-frequency data will not fit a single-frequency model. Thus in realistic cases where the noise variance σ^2 must be estimated from the data, it is essential to use the estimated variance from the multiple-frequency model even when the frequencies are well separated.

The results from this section let us see the discrete Fourier transform in yet another way: the discrete Fourier transform is a sufficient statistic for the estimation of multiple-well separated harmonic frequencies. P. Whittle [29] derived the periodogram from the principle of maximum likelihood in 1954 and stated that "... in practice the periodogram presents a wildly irregular appearance, suggesting little or nothing to the eye." It now appears that this depends on the condition of the brain behind that eye; after a little Bayesian education, a periodogram suggests a great deal to the eye because one knows where to look. When the frequencies are well separated, it is only the very largest peaks in the periodogram that are important for frequency estimation. The common practice of taking the log of the periodogram is just about the worst thing one could do, because it accents the noise and suppresses information about the frequencies.

Two Close Frequencies

Now assume that the first two frequencies are close together: $|\omega_1 - \omega_2| \approx 2\pi/N$. Then the off diagonal term B_{12} is not small. But by assumption all the remaining off-diagonal terms are negligible. The problem separates into a two-frequency problem for the close frequencies and $r - 1$ one-frequency problems. A feasible procedure for estimating multiple harmonic frequencies is now clear. We calculate the probability

of a single harmonic frequency in the data. We take the single largest peak from the data and we examine it with a two-frequency model. If there is any evidence of two frequencies, we will obtain a better fit; if not the frequencies will confound with each other. Now generate the best model function from the estimated parameters (either a one or two-frequency model) and subtract it from the data. The difference is the residual signal which must be analyzed further.

What we are contemplating here is, in spirit, what an economist would call detrending (i.e. estimating a trend and then subtracting it from the data). Normally this is a bad thing to do, because the trend and the signal of interest are not orthogonal. We can do this here because the orthogonality properties of multiple harmonic model functions ensures that the error is small. But, we stress, it is only the special properties of the sine and cosine functions that make this possible.

Next we examine the residual signal using the same procedure. We compute the posterior probability of a frequency in the residuals and examine the largest peak for two frequencies. We repeat the entire procedure until we have reduced the residuals to noise (i.e. until they exhibit no visible regularity). Determining the stopping place is not generally a problem, but if there are many small signals present it will be necessary to use the procedures developed in Chapter 5 to determine the total number of frequencies present. We stress again that this procedure is only applicable to the multiple stationary frequency problem and then only because of the special properties of the sine and cosine functions. Even here, if there is evidence of multiple close frequencies it will be necessary to use the estimates obtained from this procedure as initial estimates for a full multiple-frequency analysis on the data.

6.5.1 Example – Multiple Stationary Frequencies

To illustrate some of the points we have been making we have prepared a simple example of a stationary signal with multiple harmonic frequencies. This simple example was prepared from

$$\begin{aligned} f(t) &= \cos(0.1i+1) + 2\cos(0.15i+2) + 5\cos(0.3i+3) \\ &+ 2\cos(0.31i+4) + 3\cos(1i+5) + e_i \end{aligned}$$

and shown in Fig. 6.11(A), where e_i has unit variance and there are $N = 512$ data points. The periodogram Fig. 6.11(B) resolves the four well-separated frequencies and then hints that the frequency near 0.3 could be two frequencies. To estimate these frequencies we simply postulated a five-frequency model and used the estimates from

the periodogram as initial estimates of the frequencies. We located the maximum of the five-dimensional posterior probability density and determined the accuracy estimate using the procedure given in (4.14).

The estimated frequencies and amplitudes from these data are

frequency	amplitude
0.0998±0.0001	0.9±0.08
0.1498±0.0002	2.08±0.08
0.3001±0.0002	4.97±0.08
0.3102±0.0001	1.95±0.08
0.9999±0.0001	2.92±0.08

These are in excellent agreement with the true values. The estimated noise variance for these data is 0.98 and the true variance is 1.0. For this data set with $N = 512$ data values, the "best" estimate for the well-separated frequencies is given by (2.10)

$$\omega_{\text{est}} \approx \hat{\omega} \pm \sqrt{48\sigma^2/N^3(B_1^2 + B_2^2)}$$

which gives

frequency	amplitude
0.1000±0.0006	1.0±0.08
0.1500±0.0002	2.0±0.08
0.3000±0.0001	5.0±0.08
0.3100±0.0002	2.0±0.08
1.0000±0.0002	3.0±0.08

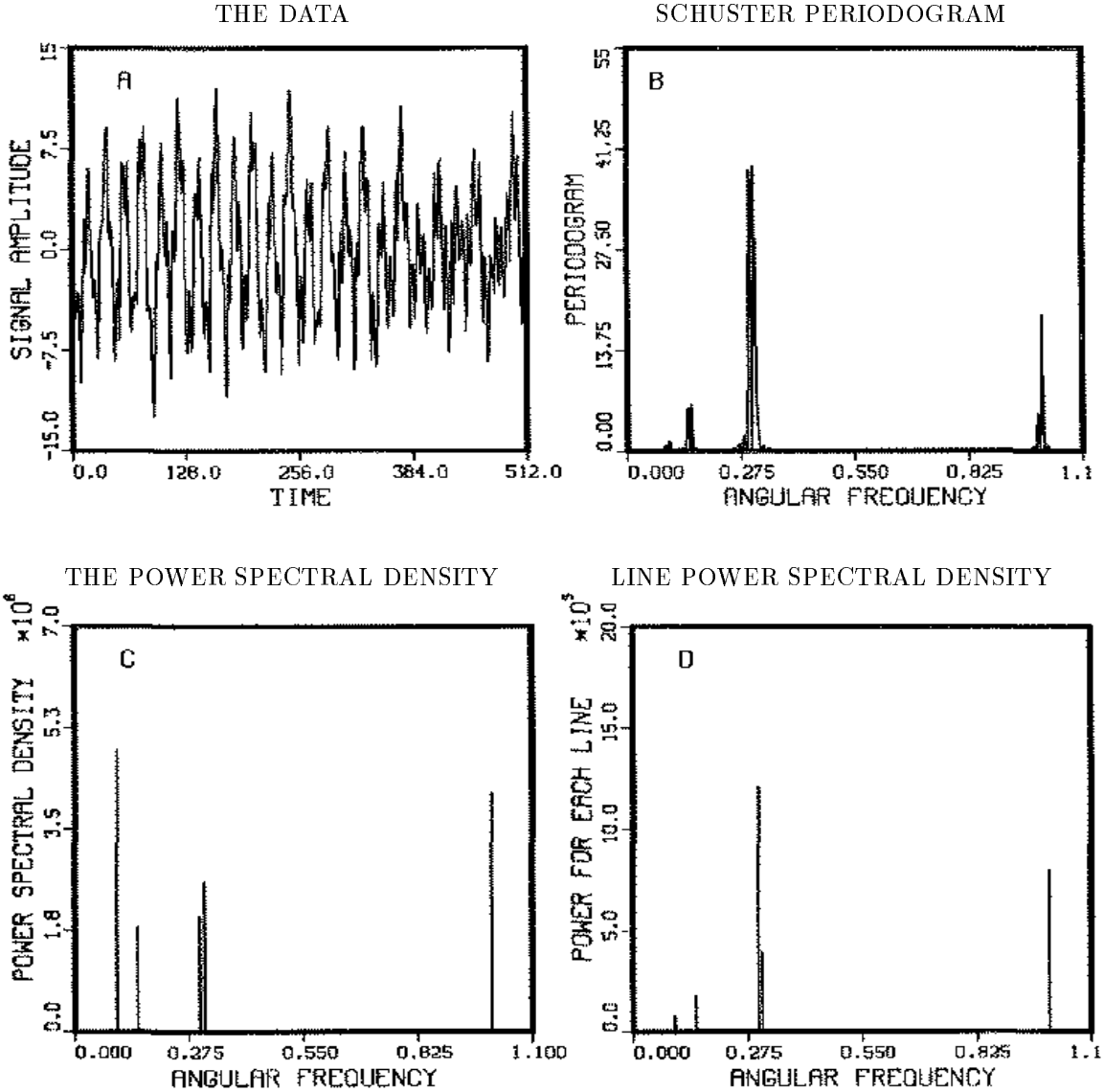
and the actual values we obtained are all comparable to these: in some cases a little better and others a little worse.

6.5.2 The Power Spectral Density

We saw in the two-frequency problem that the power spectral density $\hat{p}(\omega)$ is telling us something about the energy density of the signal and about the accuracy of the line. We have not generalized that function to account for the symmetry properties of the multiple frequency problem, and we do that now. The generalization is straightforward and we simply give the result for well-separated frequencies here.

When the frequencies are well separated the problem essentially splits into a series of one-frequency problems: all we must do is to maintain the symmetries of the original probability density. Maintaining those symmetries and integrate out all but

Figure 6.11: Multiple Harmonic Frequencies



The data (A) contain five frequencies. Three of the five are well separated. The Schuster periodogram (B) resolves the three well-separated frequencies, but one cannot tell if the peak near $\omega = 0.3$ is one or two frequencies. The power spectral density $\hat{p}(\omega)$ (C) clearly separates all five frequencies while the height is indicative of the resolution. The height of the line power spectral density (D) is indicative of the energy carried by the line.

one frequency, the power spectral density may be approximated as

$$\hat{p}(\omega) \approx 2\left[\sigma^2 + \sum_{j=1}^{r} C(\hat{\omega}_j)\right] \sum_{k=1}^{r} \left[\frac{b_{kk}}{2\pi\sigma^2}\right]^{\frac{1}{2}} \exp\left\{-\frac{b_{kk}(\hat{\omega}_k - \omega)^2}{2\sigma^2}\right\}.$$

As was stressed earlier this function expresses information about the total energy carried by the signal and about the accuracy of each line, but nothing about the power carried by one line. After the fact this is not too surprising: after all, we asked a question about the total energy carried by the signal, and not a question about the power carried by one line. If we wish information about the power carried by one line, we must ask a question about one line, and we do that in the next subsection. First we illustrate the generalized power spectral density with a simple example. In addition to determining the frequencies in the previous example we have plotted the power spectral density in Fig. 6.11(C). We see from (C) that the five frequencies have been well resolved by the "Student t-distribution": the widths of the lines from (C) are indications of how well the lines have been determined from the data while the integral over all lines is the total energy carried by the signal in the observation time.

6.5.3 The Line Power Spectral Density

We would like to plot a power spectral density that is an indication of the power carried by the individual spectral lines. This is easily done simply by defining the appropriate spectral density. Here we define a line power spectral density $\hat{S}(\omega)$ as the posterior expected value of the energy carried by one sinusoidal component of the signal in the frequency range $d\omega$. This is given by

$$\begin{aligned}
\hat{S}(\omega) &= \frac{N}{2}\int (B_1^2 + B_{1+r}^2) P(\{B\}, \{\omega\}|\sigma, D, I) dB_1 \cdots dB_m \, d\omega_2 \cdots d\omega_r \\
&= \frac{2}{r}\left[\sigma^2 + C(\omega)\right] \sum_{k=1}^{r} \left[\frac{b_{kk}}{2\pi\sigma^2}\right]^{\frac{1}{2}} \exp\left\{-\frac{b_{kk}(\hat{\omega}_k - \omega)^2}{2\sigma^2}\right\}
\end{aligned}$$

where we have performed the integrals over all but the first frequency. We have relabeled ω_1 as ω. When we computed this expectation value we used the amplitudes for frequency ω_1 however, the exchange symmetries in this problem ensure we will obtain the same result whichever one we chose to leave behind. This is essentially just the marginal posterior probability density normalized to the power carried by one spectral line. The integral over ω will give the total energy carried by all of the spectral lines, and in this approximation each line contributes its total energy to the

integral. We have included an example of this in the multiple frequency example given earlier see Fig. 6.11(D). In this figure the lines are normalized to a power level, the heights are indications of the total energy carried by a line, and the width is an indication of the accuracy of the estimates.

6.6 Multiple Nonstationary Frequency Estimation

The problem of multiple nonstationary frequencies is easily addressed using the "Student t-distribution" (3.17). As with the multiple stationary frequency estimation problem, an analytic solution is not feasible for more than a few frequencies. However, we already know that this problem separates. If it did not, the discrete Fourier transform would not be useful on this problem; and it is.

The way to handle this problem is to apply the "Student t-distribution" numerically. One can apply the single-frequency-plus-decay model when the nonstationary frequencies are well separated, and then use more complex models where needed. The numerical procedure to use is to calculate the discrete Fourier transform of the data, and from it compute the logarithm of the probability of a single harmonic frequency. Then set up a nonstationary frequency model using the single best frequency from the discrete Fourier transform. Locate the maximum of the probability density and then compute the residuals. These residuals are essentially what probability theory is calling the noise. Repeat the Fourier transform step on the residuals. If there are additional frequencies in the data, repeat the process using two, three, $\cdots$, frequencies model until all frequencies have been accounted for. Of course, one can save time here by starting with an initial model that has the same number of well-separated peaks as are in the Fourier transform of the data. But care must be taken; if these signals are decaying, one must supply reasonable estimates for the decay rates and this can be very difficult.

When applying this procedure, there is no need to check to see if any of the peaks have multiple frequencies. Later passes through the procedure will resolve double structure. If any of the peaks has multiple frequencies, then when one fits the main peak not all of the signal will be removed, and on some later cycle through the procedure the second frequency will be the largest remaining effect in the data and the procedure will pick it out. The procedure works so well and the effects are so striking, that an example is needed. We give this example in the next chapter.