3.1. INTRODUCTION

In the introductory chapter we pointed out the usefulness of covariance functions and spectral representations as ways of describing continuous data that are mixtures of signal and noise. These two ways of representing continuous dynamic processes lead to powerful methods of signal analysis. However, we indicated that the analysis procedures are generally performed not upon specimen functions of the original continuous processes, processes that are essentially infinite in duration, but upon finite segments of their sampled versions. The results and conclusions drawn from these analyses are then used to draw inferences about the original processes: the waveform of a response, its spectrum, its correlation with another response, its dependence upon a stimulus parameter, etc. The question is, how good are these inferences? Although we made some effort to point out the legitimacy of the procedures under many circumstances of practical interest, it is important that we establish their validity somewhat more securely. Once this is done we can examine specific applications of covariance functions and spectral analysis in more depth and detail. This will permit us in addition to move to related methods of signal analysis, such as coherence functions, which also have four applicability in studying the relationships between pairs of processes. Finally, these methods are applicable not only to the study of continuous processes such as the EEG but also form the basis for the analysis of certain aspects of single and multiple unit activity. Thus an understanding of how continuous processes are analyzed forms a basis for studying unit activity.
3.2. DISCRETE FOURIER REPRESENTATIONS OF CONTINUOUS PROCESSES

At the outset it is important to state a basic attribute of band-limited signals that is of fundamental importance: Whether periodic or not, such signals must be infinite in duration. This fact follows directly from the properties of the Fourier transform for continuous signals. On the other hand, the properties of the Fourier transform also guarantee that the spectrum of a finite duration signal, such as a segment of an infinite duration signal, cannot be band-limited even when the infinite duration signal is. This means that there is an inherent contradiction built into our procedures for analyzing infinite duration signals from their finite segments. The contradiction is only resolved when the infinite duration signal is truly periodic. In all other cases we are forced to settle for errors of estimation. The sampling procedure does not alleviate these errors but introduces problems of its own, the kinds of problems we deal with here.

Although it may be a fiction, we have assumed that the processes we are studying are stationary mixtures of signals and noise with at least the noise being a random process. The analysis procedures are by necessity performed upon their finite duration segments. And here we invoke the next assumption, a true fiction. This is that the finite duration segment is a single period of a periodic specimen function. As objectionable as this might seem at first, it does no real harm since we have no knowledge of the specimen function's behavior outside this observed interval. Because of stationarity, the statistical behavior of the specimen function outside this observed interval is not likely to be much different. Thus we are not disregarding any information that we have concerning the specimen's behavior. In the first chapter we assumed that the repetition period was equal to the time of observation T. Other periodicity assumptions are also possible. If we want, we can consider the repetition period to be longer than T, T' say, by padding out the observed segment with a zero amplitude data segment lasting T' - T sec. In a sense this is falsifying the data, but we know exactly how we have falsified it and we can take this into consideration in the subsequent analyses in order to avoid arriving at erroneous conclusions. Padding out the data with zero amplitude segments is a routine procedure when dealing with the estimation of the covariance functions. In this case, as we shall see, it is convenient to make T' = 2T. For the moment, however, let the repetition period be T. Let us keep in mind, then, the fact that we have forced periodicity upon the process and that for practical purposes we can make this periodicity length T or longer. Later on we shall use a 2T repetition period to deal with autocovariance function estimation.

In Chapter 1 we introduced the Fourier series representation for a T-continuous periodic signal and showed that if the signal were band-limited, its waveform could be completely represented by a finite number of parameters. Specifically, if the period of the signal is T and its bandwidth is F, then N = 2FT terms are involved in either the real or complex Fourier series representation. It was also demonstrated that the signal could be completely represented by N consecutive sample amplitudes spaced Δ sec apart where Δ = 1/2F sec. The Fourier and time sample representations are closely related, the relationship between the two involving what is called the discrete Fourier transform. We introduce it here and show that in the bandwidth-limited situation, it leads to the same Fourier coefficients as would be obtained from a Fourier series representation of the original T-continuous signal.

We start with a single T sec segment of a band-limited signal which we consider to have period T. We obtain N samples of this signal at times Δ sec apart starting at the beginning of the segment, t = 0. Using these samples we can partially reconstruct the original signal by means of weighted sinc functions. The partial representation is

\[ x(t) = \sum_{t^* = 0}^{N-1} x(t^*Δ) \frac{\sin[n(t - t^*Δ)/Δ]}{n(t - t^*Δ)/Δ} \]  \hspace{1cm} (3.1)
The reason for the reconstruction being partial is that we have ignored the tails of the weighted sinc functions outside the \( T \) sec segments. We can, however, insert them because of the assumed periodicity of \( x(t) \). The complete reconstruction takes in all the weighted sinc functions throughout time:

\[
x(t) = \sum_{t'=-\infty}^{\infty} x(t') \frac{\sin[\pi(t - t')/\Delta]}{\pi(t - t')/\Delta} \int_{-\infty}^{\infty} x(t') \frac{\sin[\pi(t - t')/\Delta]}{\pi(t - t')/\Delta} e^{-j2\pi nt/\Delta} dt
\]

This now holds for all \( t \). Now let us take the complex Fourier series representation of a single period from 0 to \( T = N\Delta \):

\[
X_T(n) = \frac{1}{T} \int_0^T x(t) \frac{\sin[\pi(t - t')/\Delta]}{\pi(t - t')/\Delta} e^{-j2\pi nt/\Delta} dt
\]

\[
= \frac{1}{T} \sum_{t'=0}^{\infty} x(t') \int_0^T \frac{\sin[\pi(t - t')/\Delta]}{\pi(t - t')/\Delta} e^{-j2\pi nt/\Delta} dt
\]

This is actually a simple equation to deal with, given the periodicity of \( x(t) \). Because of periodicity we have \( x((N + t')\Delta) = x(t') \). When this fact is taken into account for values of \( t' \) outside the range 0 to \( N - 1 \), Eq. (3.3) simplifies, after some elementary substitutions, to

\[
X_T(n) = \frac{N}{T} \sum_{t'=0}^{N-1} x(t') \exp(-j2\pi nt') \int_{-\infty}^{\infty} \sin \frac{\pi x}{\pi} \cos \frac{2\pi nx}{N} dx
\]

As long as \( n/N < 1 \), which is true for the band-limited signal, this further reduces to

\[
X_T(n) = \frac{N}{T} \sum_{t'=0}^{N-1} x(t') \exp(-j2\pi nt')
\]

The original integration operation upon \( x(t) \) has thus been modified into a summation operation upon \( x(t') \).

This is an opportune time to reconsider the steps that led us to Eq. (3.5). Our data specimen was a \( T \) sec segment of an ongoing process band-limited to real frequencies between 0 and \( 1/2\Delta \). We assumed, solely for the purpose of analysis, that this segment was one period of a periodic process. We then sampled the segment at intervals \( \Delta \) sec apart to obtain an \( N \) sample representation of it. Because of the bandwidth limitation and the periodicity assumption, we need only \( N \) Fourier components at complex frequencies spaced equally from \(-1/2\Delta\) to \( 1/2\Delta\) to represent the data completely. Now, in the majority of situations, the data do not arise from a periodic process but are specimens of an aperiodic process with power distributed at all frequencies up to \( 1/2\Delta \) (or at all the complex frequencies between \(-1/2\Delta\) and \( 1/2\Delta\)). Hence, our periodicity assumption has in a sense falsified the data. It has produced a representation of the signal requiring only \( N \) Fourier components. This is not a serious falsification, however. What it amounts to is saying that all the frequency components in the narrow frequency band between \((n - 1/2)/N\Delta\) and \((n + 1/2)/N\Delta\), a band \( 1/N\Delta \) wide, are considered to be concentrated at the single frequency \( n/N\Delta \), and represented by \( X_T(n) \). \( X_T(n) \) is therefore essentially the product of the frequency density of the Fourier representation times the incremental bandwidth \( 1/N\Delta \). The density in that frequency region can then be obtained by dividing \( X_T(n) \) by \( 1/N\Delta \). This gives

\[
N\Delta X_T(n) = \sum_{t'=0}^{N-1} x(t') \exp(-j2\pi nt'/\Delta)
\]

\( \Delta \) is a constant independent of the duration of the specimen and plays only a minor role in the reconstruction of \( x(t) \) from the Fourier representation. For this reason we define the discrete Fourier transform (DFT) of \( x(t) \) as \( X_N(n) \):

\[
X_N(n) = \sum_{t'=0}^{N-1} x(t') \exp(-j2\pi nt'/N)
\]

The elimination of the factor \( N \) that appeared in Eq. (3.5) means that in order to recover \( x(t') \) from \( X_N(n) \), we must define the inverse DFT as
To see this, we multiply both sides of Eq. (3.7) by
\[ \exp(j2\pi nu^*/N) \]
and sum over all the values of \( n \) between \(-N/2\) and \((N/2) - 1\), the range of the complex Fourier expansion. We obtain
\[
\frac{N/2-1}{n=-N/2} X_n(n) \exp(j2\pi nu^*/N) \]
\[
= \frac{N/2-1}{n=-N/2} \left[ \frac{N-1}{t^*=1} x(t^*\Delta) \exp(-j2\pi n\Delta t^*/N) \right] \exp(j2\pi nu^*/N) \quad (3.9)
\]
We then interchange the order of the two summations on the right-hand side and consider the summation with respect to \( n \). This is
\[
\frac{N/2-1}{n=-N/2} \exp[j2\pi (u^* - t^*)/N]
\]
For any value of \( t^* \) different from \( u^* \), this summation is zero (as can be seen by using the summation formula for a geometric series). But when \( t^* = u^* \), the summation is \( N \). Thus, Eq. (3.9) reduces to Eq. (3.8) which is what we wished to show.

Equations (3.7) and (3.8) are a discrete Fourier transform pair and have been justified on a heuristic basis. Later in the chapter we shall establish the validity of the relation somewhat more carefully, paying closer attention to the properties of continuous processes. It is also worth noting that the definition of the DFT varies from author to author according to the handling of the factor \( N \). The definition adopted here seems to be the most common one.

The cosine and sine versions of the DFT are given by
\[
A^*_n(n) = \frac{N-1}{t^*=0} x(t^*\Delta) \cos(2\pi n t^*/N) \\
B^*_n(n) = \frac{N-1}{t^*=0} x(t^*\Delta) \sin(2\pi n t^*/N)
\]
(3.9a)
These are associated with the complex relations \( A^*_n(n) = X^*_n(n) + X^*_n(-n) \) and \( B^*_n(n) = j[X_n(n) - X^*_n(-n)] \). It is worthwhile pointing again that \( X_n(n) \), the direct DFT, is a periodic function of \( n \), period \( N \), and its inverse \( x(t^*\Delta) \) is a periodic function of time. That is, \( X_n(-(N + n)) = X_n(n + N) \), etc. and \( x((N + t^*)\Delta) = x((N + t^*)\Delta) \), etc. In the previous chapters we considered the index for the direct DFT to run from \(-N/2\) to \((N/2) - 1\). It is clear now that because of the periodicity it is equally satisfactory to consider \( n \) to range from 0 to \( N - 1 \).

The periodicity of the direct and inverse DFT emphasizes the fact that when the DFT is applied to an \( N \) sample sequence of data points, it is done under the assumption that the data arise from a periodic process, period \( N \). Sometimes the period can be considered to be greater than \( N \) by appending or "padding" a sequence of zero amplitude samples, \( N' - N \) of them so that the overall length of the resulting sequence is \( N' \). This padding with zeros is a technique commonly employed in digital filtering and in the estimation of the acvf and spectrum of a specimen function, as we shall see later. The resulting sequence of sample values can be considered to arise from a periodic band-limited signal \( \tilde{x}(t) \), period \( N' \), which is zero at \( L = N' - N \) consecutive sample times. The DFT of this signal is
\[
\tilde{X}_N(m) = \sum_{t^*=0}^{N'-1} \tilde{x}(t^*\Delta) \exp(-j2\pi mt^*/N')
\]
(3.10)
Because of the fact that \( \tilde{x}(t) = x(t) \) for values of \( t^* \) ranging from 0 to \( N - 1 \) and is zero for values of \( t^* \) ranging from \( N \) to \( N' - 1 \), we have
\[
\tilde{X}(m) = \sum_{t^*=0}^{N-1} \tilde{x}(t^*\Delta) \exp(-j2\pi mt^*/N')
\]
(3.11)
An especially important case is \( N' = 2N \). Here we have

\[
\tilde{X}_{2N}(m) = \sum_{t^n=0}^{N-1} x(t^n) \exp(-j2\pi mt^n/2N)
\]

Because of the \( 2N \) periodicity of \( \tilde{x}(t) \), the values of \( n \) range from \(-N\) to \( N - 1 \) instead of from \(-N/2\) to \((N/2) - 1\). If we examine Eqs. (3.7) and (3.10), we see that when \( m = 2n \), i.e., it is an even number or zero,

\[
\tilde{X}_{2N}(2n) = \sum_{t^n=0}^{N-1} x(t^n) \exp(-j2\pi nt^n/2N) = X_N(n)
\]

This shows that the even index terms for \( \tilde{X}_{2N}(n) \) are completely determined by the values of the \( X_N(n) \). But what about the odd index terms? Some reflection on this reveals that these terms arise solely because of the padding procedure. They are necessary to force \( \tilde{x}(t) \) to be zero at the sample times between \( N \) and \( N' - 1 \). They provide no additional information about \( x(t) \), but, interestingly enough, are an essential ingredient for obtaining an estimate of the acvf from the estimated spectrum. This point will be discussed later. Finally, it is easy to see that similar results would be obtained if \( N' \) were any other multiple value of \( N \).

### 3.3. Aliasing

As we discussed in Chapter 1, the necessity for sampling a signal at a rate compatible with its bandwidth, the Nyquist rate, is vital to a meaningful interpretation of a spectral analysis. Here we wish to establish this point somewhat more securely and show in what way improper sampling, sampling at too low a rate for a given bandwidth, obscures and falsifies spectral analysis.

Let us begin by considering the continuous signal \( x(t) \) to be periodic \( T \), and to have an unlimited bandwidth. The Fourier series representation for such a signal is given by

\[
x(t) = \sum_{n=-\infty}^{\infty} X_T(n) \exp(j2\pi nt/T)
\]

where

\[
X_T(n) = \frac{1}{T} \int_{0}^{T} x(t) \exp(-j2\pi nt/T) \, dt
\]

We wish to deal with the sampled representation \( x(t^\Delta) \) and so we sample \( x(t) \) every \( \Delta \) sec, obtaining \( N \) samples such that \( T = N\Delta \). We then blindly take the DFT,

\[
X_T(n) = \sum_{t^n=0}^{N-1} x(t^n) \exp(-j2\pi nt^n/N)
\]

We have used the dagger symbol to indicate our suspicion that something may be amiss in this representation, i.e., that \( X_T(n) \) may not be the same as \( X_T(\hat{n}) \). That such is the case may be seen by substituting for each sample value its Fourier series expansion as given by Eq. (3.14):

\[
X_T(n) = \sum_{t^n=0}^{N-1} \left[ \sum_{m=-\infty}^{\infty} X_T(m) \exp(j2\pi mt^n/T) \exp(-j2\pi nt^n/N) \right]
\]

The exponential term here has the important property that when \( m - n = 0 \) or some integer multiple of \( N \), the summation over \( t^n \) is equal to \( N \); otherwise it is identically 0. That is, for fixed \( m, \)

\[
\sum_{t^n=0}^{N-1} \exp[j2\pi (m - n)t^n/N] = \begin{cases} N, & m = kN + n \\ 0, & m \neq kN + n \end{cases}
\]

where \( k \) is an integer. Using this fact in Eq. (3.17), it can be seen that

\[
X_T(n) = \sum_{k=-\infty}^{\infty} X_T(kN + n)
\]
This means that each term in the DFT of \( x(t) \) is the sum of a possibly infinite set of Fourier coefficients associated with the higher frequency components in \( x(t) \). The higher frequency components are those corresponding to frequencies that are greater than \( N \) by an amount \( kN \). If \( x(t) \) has no Fourier series components for values of \( n \) equal to or greater than \( N/2 \) (corresponding to frequencies \( 1/2\Delta \) or greater), \( X_N^+(n) = NX_T(n) = X_N^0(n) \); otherwise, \( X_N^+(n) \neq NX_T(n) \). This means that the DFT for \( x(t) \) yields correct results only if \( x(t) \) is band-limited to frequencies below \( 1/2\Delta \).

When \( x(t) \) has a greater bandwidth, the high frequency components add to the low frequency ones, an effect that is called aliasing because the high frequency components are misrepresented or misinterpreted as low frequency ones. Once aliasing occurs, there is no way to properly sort out the \( X_T^0(n) \) components from the \( X_T^+(n) \). This is why the cutoff frequency \( F \) of the analog prefilter must be matched to the sampling rate such that \( F \leq 1/2\Delta \). It is essential to the proper analysis of continuous data by sampling techniques.

The numerical value of \( n \) corresponding to the highest frequency representable by the sampling procedure is \( N/2 \). As shown previously, it is determined by the relation \( n/T = 1/2\Delta \). To see the effect of aliasing more clearly, consider Fig. 3.1 which shows a cosine wave of frequency \( F = 1/2\Delta \) being sampled at the negative and positive peaks. If the frequency of the wave increases a little above \( F \) to \( F + a \) (dotted line), sine waves of frequency \( F + a \) and \( F - a \) can be drawn through the sampling points equally well. This gives us reason to suspect that a wave of real frequency \( F + a \) will, after sampling, be confused with a wave of real frequency \( F - a \). With this in mind, let us examine Eq. (3.19) when \( n \) has a value of \( (N/2) - i \). Then all the \( X_T^0(kN + n) \) such that 

\[
kN + n = kN + (N/2) - i = (k + 1/2)N - i
\]

will contribute to the terms \( X_T^0[(N/2) - i] \). A real frequency term at \( (N/2) - i \) corresponds to complex frequency terms \( X_T^0[(N/2) - i] \) and \( X_T^0[(-N/2) + i] \). The aliases of \( X_T^0[(N/2) - i] \) are at frequencies \( (-3N/2) - i, (-N/2) - i, (3N/2) - i, \ldots \) while the aliases of \( X_T^0[(-N/2) + i] \) are at frequencies \( (-3N/2) + i, (N/2) + i, (3N/2) + i, \ldots \). If we group these aliasing terms in pairs, one term from each sequence, we find that \( X_T^0[(-N/2) - i] \) pairs with \( X_T^0[(N/2) + i] \) to give a real frequency term at \( (N/2) - i \). Similarly, there are real frequency terms at \( (3N/2) + i, (3N/2) - i, (5N/2) + i, (5N/2) - i, \ldots \). Thus a real frequency data component at \( (N/2) - i \) will have alias contributions from whichever of these higher frequency terms that are present in the data input to the ADC. In effect the original Fourier representation of \( x(t) \) has been folded in accordion fashion about frequencies that are multiples of \( 1/2\Delta \) and collapsed into the frequency region extending from 0 to \( 1/2\Delta \) which is also called the folding frequency. (Fig. 3.2)

It is of some interest that aliasing effects can also enter into sampled representations of data that are band-limited to the Nyquist frequency. We have seen previously how the discrete Fourier transform is a completely adequate representation of a continuous periodic band-limited signal as long as the signal samples are taken frequently enough to eliminate the possibility of aliasing. But in actuality, few of the data one analyzes are
Fig. 3.2. The accordion-like folding of the frequency (or n) axis due to sampling of a continuous signal. Frequency components of the original signal marked with x's on the f axis are interpreted in the sampled version as belonging to the lowest frequency, an encircled x.

periodic or band-limited, although the latter condition can be approached as closely as desired by analog prefiltering prior to sampling. Periodicity is another matter. Even when periodic stimulation is employed and the response or signal component of the data is periodic, the remainder, the noise, is not. Periodicity is then lacking in the data. What the data analysis procedure does in this situation is to effectively create periodic data from the T sec data segment we have available to study. That is, we analyze the T sec segment as though it originated from a process with period T or greater. This introduces some complications which we need to consider. The "periodicized" process created from a T sec segment of data (1) is generally not band-limited even if the original data are, (2) can contain frequency components, apart from aliases, that are not present in the original data. Let us deal with these complications in order, using as an illustration a signal that is both band-limited and periodic, a cosine wave whose period is 3T/8, T being the period of its observation. The periodicized version of this signal is shown in Fig. 3.3. It is clear that there are discontinuities in the periodicized signal which guarantee that it will not be band-limited. In fact, it may be stated that unless the original signal has rather special properties, i.e., that its amplitude and time derivatives at t = 0 are the same as those at t = T, there will be discontinuities in the periodic waveform and its derivatives that guarantee that the periodicized signal will not be band-limited. We know that if we sample this process, every Δ sec such that T = NW, we are sure to encounter aliasing, its
severity depending upon the sampling rate. If we apply the DFT to the samples and treat the resulting Fourier coefficients as though there were no aliasing involved, we effectively consider the data as having arisen from a periodic band-limited process, i.e., one that has no discontinuities of any kind at the ends of the interval. This recreated signal is also shown in Fig. 3.3 for \( N = 16, \delta = T/16 \). This means that the sampling has distorted the original data, primarily at the ends of the interval. The high frequency components associated with the discontinuities at 0 and \( T \) have been aliased into the spectral representation. The numeric results obtained from the DFT show the results of this aliasing. Both covariance and spectral analysis of the data can be affected. Fortunately, the larger \( N \) is, the smaller the end effects tend to become. They also diminish as the severity of the discontinuities diminishes.

### 3.4. Leakage

#### A. Fourier Series

Besides the aliasing that is introduced into the DFT representation of a time-limited segment of a nonperiodic signal, we must deal with another form of signal misrepresentation, referred to as spectral leakage. It occurs with all aperiodic data and even with periodic band-limited data whose period is not integrally related to the time of observation. In the Fourier analysis procedures, the frequency composition of the data is computed to be a set of frequency constituents harmonically related to \( 1/T \), the fundamental of the time of observation. The frequency components that are closest to the original frequencies in the data contribute most to the analysis, but more remote frequencies may also be interpreted as being present when in fact they are not. To see a specific example of this, consider the signal to be the cosine wave whose period is \( 3T/8 \) (Fig. 3.3). We compute the Fourier series representation of this signal first because it avoids all aliasing effects. The Fourier series coefficients are given by

\[
X_T(n) = \frac{1}{T} \int_0^T \cos\left(\frac{2\pi n}{3T} t\right) \exp\left(-j2\pi T t\right) dt
\]

for \(- (N - 1)/2 \leq n \leq (N - 1)/2 \).

\[
A_T(n) = \frac{2}{T} \int_0^T \cos\left(\frac{2\pi n}{3T} t\right) \cos\left(\frac{2\pi T}{3T} t\right) dt
\]

\[
B_T(n) = \frac{2}{T} \int_0^T \cos\left(\frac{2\pi n}{3T} t\right) \sin\left(\frac{2\pi T}{3T} t\right) dt
\]

for \( 0 \leq n \leq (N - 1)/2 \). The values for \( A_T(n) \) and \( B_T(n) \) are obtained by standard integration formulas and are tabulated in Table 3.1 for \( n = 1, 2, \ldots, 8 \).

#### Table 3.1

FOURIER SERIES AND DFT COEFFICIENTS FOR COS(2\(\pi 8T/3T\))

<table>
<thead>
<tr>
<th>( n )</th>
<th>( A_T(n) )</th>
<th>( A_N(n) )</th>
<th>( B_T(n) )</th>
<th>( B_N(n) )</th>
<th>( X_T(n) )</th>
<th>( X_N(n) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>16</td>
<td>256</td>
<td>16</td>
<td>256</td>
<td>16</td>
<td>256</td>
</tr>
<tr>
<td>1</td>
<td>-0.120</td>
<td>-0.017</td>
<td>-0.114</td>
<td>-0.078</td>
<td>-0.085</td>
<td>-0.078</td>
</tr>
<tr>
<td>2</td>
<td>-0.236</td>
<td>-0.133</td>
<td>-0.230</td>
<td>-0.307</td>
<td>-0.320</td>
<td>-0.307</td>
</tr>
<tr>
<td>3</td>
<td>0.389</td>
<td>0.493</td>
<td>0.395</td>
<td>0.758</td>
<td>0.738</td>
<td>0.758</td>
</tr>
<tr>
<td>4</td>
<td>0.083</td>
<td>0.188</td>
<td>0.089</td>
<td>0.215</td>
<td>0.187</td>
<td>0.215</td>
</tr>
<tr>
<td>5</td>
<td>0.041</td>
<td>0.147</td>
<td>0.047</td>
<td>0.134</td>
<td>0.098</td>
<td>0.133</td>
</tr>
<tr>
<td>6</td>
<td>0.026</td>
<td>0.133</td>
<td>0.031</td>
<td>0.099</td>
<td>0.055</td>
<td>0.099</td>
</tr>
<tr>
<td>7</td>
<td>0.018</td>
<td>0.127</td>
<td>0.024</td>
<td>0.080</td>
<td>0.025</td>
<td>0.080</td>
</tr>
<tr>
<td>8</td>
<td>0.013</td>
<td>0.125</td>
<td>0.019</td>
<td>0.067</td>
<td>0.000</td>
<td>0.067</td>
</tr>
</tbody>
</table>
Inspection of the Fourier components as determined by Eq. (3.20) reveals that the analysis has decomposed the original cosine wave into frequency components at all values of $n$. None of these corresponds to the frequency of the original signal which lies slightly below $n = 3$, but the coefficients are largest at $n = 3$ and next largest at $n = 2$. There is a gradual diminution of component amplitudes as $n$ departs from these values. What has happened is that the power of the original signal has been dispersed or "leaked" out from the original signal frequency into the neighboring frequencies of the Fourier analysis. No spurious power is added by the analysis, for if all the $A_T(n)$ and $B_T(n)$ were squared and summed, their total contribution would equal that of the original signal in the $T$ sec interval. The net effect, however, is a rather serious misrepresentation of the original signal whose spectrum is a single real frequency component at $\theta/3T$. The cause of the misrepresentation is that only a finite length of the signal segment has been used for the analysis. It is possible to show that the Fourier representation of a $T$ sec segment of data results from a convolution of the spectrum of the original, infinite duration signal with the sinc function $\sin(\pi nT)/(\pi nT)$. To see how this comes about, we refer back to the expression for $X_T(n)$ in Eq. (3.20) where we replace the illustrative frequency $\theta/3T$ by the general frequency $f$.

$x(t) = \cos 2\pi ft$. We can calculate the $A_T(n)$ and $B_T(n)$ for this signal and find them to be

$$A_T(n) = \frac{1}{T} \left[ \sin \frac{2\pi T(f - (n/T))}{2\pi(f - (n/T))} + \sin \frac{2\pi T(f + (n/T))}{2\pi(f + (n/T))} \right]$$

$$B_T(n) = -\frac{1}{T} \left[ \cos \frac{2\pi T(f - (n/T))}{2\pi(f - (n/T))} - 1 - \cos \frac{2\pi T(f + (n/T))}{2\pi(f + (n/T))} - 1 \right]$$

(3.21)

The terms containing $f - n/T$ and $f + n/T$ are a manifestation of the fact that cosine and sine waves consist of positive and negative complex frequency terms. We are considering real (positive) frequency data and so both $f$ and $n$ are greater than 0. In most cases $f$ will be sufficiently greater than 0 to make the second term of Eq. (3.21) negligible compared to the first. This results in the approximation

$$A_T(n) = \frac{1}{T} \left[ \sin \frac{2\pi T(f - (n/T))}{2\pi(f - (n/T))} \right]$$

$$B_T(n) = -\frac{1}{T} \left[ \cos \frac{2\pi T(f - (n/T))}{2\pi(f - (n/T))} - 1 \right]$$

(3.22)

From this we obtain the spectral power at real frequency $n/T$:

$$|X_T(n)|^2 + |X_T(-n)|^2 = \frac{1}{2} \left[ |A_T(n)|^2 + |B_T(n)|^2 \right]$$

$$= \frac{1}{2} \left[ \frac{\sin \pi T(f - (n/T))}{\pi T(f - (n/T))} \right]^2$$

(3.23)

The total power of $x(t) = \cos 2\pi ft$ is $1/2$ and is concentrated solely at frequency $f$. The Fourier analysis has in effect dispersed or leaked this power out into neighboring frequencies that are harmonically related to $1/T$. This also means that if one is interested in estimating the spectral component of the data at a particular frequency, there will be included in the estimate a contribution from nearby spectral components that have had their power leaked into the frequency where the estimate is being made. The weighting factor for these extraneous contributions is that given by the bracketed term in Eq. (3.23). It shows that the larger $T$ becomes, the smaller is the frequency range over which leakage is a significant factor.

Leakage may also magnify the undesirable effects of 60 Hz or other single frequency artifacts in the data. These may arise from a variety of causes: ineffective electrical shielding, stray coupling of stimulus frequencies into the responses, and so on. An important attribute of a signal with a line spectrum, one expressed by delta functions in the spectrum, is that a rather substantial amount of power is confined to an infinitesimally narrow
frequency band rather than being spread out over a broader range of frequencies. It is this concentration of power that can be so potent in producing leakage into the estimates of power density in the neighboring regions of the spectrum. The leakage occurs, as Eq. (3.23) indicates, if the line component is not exactly located at a harmonic of the fundamental analysis interval. To see this, suppose a spurious line component is located midway between adjacent harmonic frequencies of the analysis interval and that the rms strength of the line is \( \sigma_a \). The leakage of this component into the neighboring frequency terms is well approximated by Eq. (3.23) as long as the line is reasonably far from 0 frequency. It can be seen that the larger \( N \) is, the narrower will be the frequency range over which significant amounts of leakage occur. Because of the side lobes of the sinc function, leakage effects can occur between rather widely spaced frequencies when \( \sigma_a \) is large.

It is also true that the closer the frequency of a line component is to a harmonic of the analysis interval, the smaller is the leakage effect. The most generally useful way of minimizing leakage is by means of spectral "windowing" techniques of which more will be said later. These techniques, which are another form of linear filtering, have the effect of estimating the spectrum in a way that greatly minimizes the side lobe contributions to the spectral estimate.

B. DISCRETE FOURIER TRANSFORMS

Leakage is not alleviated by resort to the DFT. Rather, the situation persists and is also overlaid with aliasing effects so that the resulting data representation contains both, inextricably combined. To see this we refer again to the signal \( x(t) = \cos 2\pi ft/3T \) and represent it by its DFT as given by Eqs. (3.7) and (3.9), rewritten here for \( N = 16 \):

\[
X_N(n) = \sum_{t=0}^{15} \cos(2\pi t^6/6) \exp(-j2\pi nt/16)
\]

\[
A_N(n) = \sum_{t=0}^{15} \cos(2\pi t^6/6) \cos(2\pi nt/16)
\]

\[
B_N(n) = \sum_{t=0}^{15} \cos(2\pi t^6/6) \sin(2\pi nt/16)
\]

In Table 3.1, we show the DFT coefficients for \( n \) ranging from 1 to 8 when there are two different sample intervals, the first being \( T/16 \) with \( N = 16 \), and the second \( T/256 \) with \( N = 256 \). The discrepancy between the tabulated values for either situation and those obtained from the continuous Fourier series expansion arises from the aliasing introduced by sampling. As the sampling interval becomes shorter, the discrepancy diminishes and what remains is the pure leakage effect. Again, what causes it is the finite length of the signal segment, \( N \) samples in duration. If we let \( x(t) = \cos 2\pi ft \) and perform a calculation similar to that just done for the Fourier series, we find that power has leaked from frequency \( f \) into frequency \( n/N \). The amount that has leaked is given by:

\[
|X_N(n)|^2 + |X_N(-n)|^2 = \frac{1}{2} \left[ |A_N(n)|^2 + |B_N(n)|^2 \right]
\]

\[
= \frac{1}{2} \left[ \frac{\sin \pi N(fN - n/N)}{N \sin \pi (fN - n/N)} \right]^2
\]

The approximation arises as before because of the fact that we have ignored the usually small terms involving \( f + (n/N) \). From Eq. (3.25) we see that the leakage from frequency \( f \) into the \( n \)th component of the DFT has very nearly the same behavior as it had for the Fourier series representation. Thus leakage in the two cases is comparable although the leakage in the DFT tends to be the larger of the two because the denominator of Eq. (3.25) is smaller than that of Eq. (3.22).
Another aspect of leakage is associated with the presence of a constant dc component in the data. If only the spectrum of the data is of interest, leakage is not a factor because the steady component shows up only in the \( n = 0 \) term of the Fourier representation. But when one uses the spectrum as an intermediary step for obtaining an estimate of the acvf (or ccvf) of the data, then leakage becomes a factor. Such a procedure is quite common when one employs the fast Fourier transform to first obtain the spectral estimate and then the acvf from it. The reason that leakage becomes a factor is that in this procedure it is necessary to pad out the original sequence of \( N \) data points with a sequence of zero amplitude samples, \( L \) of them if one wishes to estimate the acvf for lags up to \( L \). This means that the DFT that one works with is

\[
X_N(t) = \sum_{t=0}^{N-1} x(t)e^{-j2\pi nt/N} 
\]

The upper limit is \( N - 1 \) rather than \( N' - 1 \), \( (N' = L + N) \), because the last \( L \) values of \( x(t) \) are taken to be 0. When \( x(t) \) has an average value \( a \), the contribution of this to \( X_N(n) \) is

\[
[X_N(n)]_{dc} = \frac{aL}{N} \exp(-j2\pi nt'/N') - \frac{a(1 - \exp(-j2\pi nt'/N'))}{1 - \exp(-j2\pi nt'/N')}
\]

The contribution to the raw spectral estimate \( [C_{xx}(n)]_{dc} = \frac{|X_N(n)|_{dc}}{2} \) follows directly. It is

\[
|X_N(n)|_{dc}^2 = \frac{a^2}{4} \left[ \frac{\sin(\pi n/N)}{\sin(\pi n/N')} \right]^2
\]

For \( n = 0 \), the result is \( (aN)^2 \) as is to be expected. If a data record of length \( N = 1000 \) were padded with 10 zeros to permit estimation of the acvf out to \( 100 \), the dc leakage at \( h = 1 \) would be \( 100a^2 \). If the record were padded with 100 zeros, the dc leakage at \( n = 1 \) would be \( 1.053 \times 10^4a^2 \). The effect obviously depends upon the strength of the dc term. In the second case, if \( a \) is 5 times the amplitude of the real component at \( n = 1 \), one could expect an error in the spectral estimate amounting to about 24%, a rather serious matter. To eliminate leakage, a good procedure is to first remove the average value from the data before padding it with zeros.

3.5. TREND

Another effect that we need to be aware of is one that is brought about by the presence of very low frequency components in the data, frequencies that are less than that of the fundamental frequency of the analysis interval. Such components are referred to as producing trends in the data. These are progressive changes in the short term mean of the data, a mean that is calculated over a relatively small segment of the data. Trends may also be found in other properties of the data such as the variance and covariance functions, but here we are concerned only with trends in the mean and, more specifically, linear trends, i.e., those trends that can be described by data having the form \( x(t) = bt + v(t) \), \( bt \) being the trend component and \( v(t) \) the component one normally considers in a trend-free situation. It is also possible to take into account trends which are not linear (Otines and Enochson, 1974) but here we are only interested in seeing how linear trends affect a spectrum analysis. When a linear trend is present in an \( N \) sample sequence of data, it will contribute to the DFT according to

\[
[X_N(n)]_{trend} = \frac{1}{N} \sum_{t=0}^{N-1} bt^n e^{-j2\pi n/N} 
\]

The expression can be summed without difficulty. When \( n \) is small compared to \( N \), we find that

\[
[C_{xx}(n)]_{trend} = \left[ \frac{X_N(n)}{2} \right]^2 = \left( \frac{bn/2\pi n}{2} \right)^2
\]

In effect the trend leaks into the nearby low frequency components in a manner that is inversely proportional to \( n^2 \). Note that \( bn \) is
the total trend in the data from the beginning to the end of the
sequence. To eliminate contamination of the spectral estimates by
trends, the trends should be estimated and removed before a spec-
trum analysis. Procedures for doing this are given in Otnes and

3.6. THE POWER SPECTRUM, GENERAL CONSIDERATIONS

When investigating the properties of samples of random vari-
ables, it is useful to characterize them by population statistics.
In the case of a simple, univariate random variable, the mean is
a measure of its location (from zero), and the variance is a mea-
sure of its dispersion about the mean. These two statistics are
also of use when investigating random signals. The mean specifies
a baseline about which the signal fluctuates. If the physical
signal is an electrical one, then the mean corresponds to the dc
level of the signal. The variance provides a measure of the mag-
nitude of the signal's fluctuation about its mean. For electrical
signals, the variance corresponds to the power of the ac component
of the signal. While the mean and variance are useful and readily
computed statistics, they provide no information concerning the
temporal character of the fluctuation of a random signal. We
cannot infer from them whether the signal's fluctuations are slow
or rapid or whether they possess some rhythmicity or a high degree
of irregularity. However, as we noted in Chapter 1, if the signal
is wide sense stationary, such information can be provided by the
power spectrum of the signal.

The power spectrum provides a statement of the average dis-
tribution of power of a signal with respect to frequency. If the
signal varies slowly, then its power will be concentrated at low
frequencies; if the signal tends to be rhythmic, then its power
will be concentrated at the fundamental frequency of the rhythm,
perhaps at its harmonic frequencies; if the signal lacks rhythmi-
city, then its power will be distributed over a broad range of
frequencies.

A way of obtaining an estimate of the power spectrum of a
signal at a given frequency is to pass the signal through a narrow
band linear filter centered at the frequency of interest, and then
to compute the variance (power) of the filter output. This opera-
tion can be performed at any frequency of interest. The variance
of the output of the filter will be proportional to the amount of
power in the signal at frequencies close to the filter center fre-
quency. The variance can then be plotted as a function of the
filter's center frequency and the resulting graph will be an approx-
imate indication of the frequency distribution of the signal's
power. This filtering approach was the traditional way of analyz-
ing spectra before the advent of high speed digital computers. It
is still useful conceptually although the mechanization of the
filtering techniques has been changed drastically by the computer.

The concept of a power spectrum applies to both T-continuous
and T-discrete signals. Because we are usually interested in con-
tinuous signals, we will begin with a discussion of the power spec-
tra of wide sense stationary continuous signals. Then we move to
consider more fully the computation and interpretation of power
spectra from wide sense stationary sampled data. This is the rep-
resentation of continuous signals that digital computers usually
operate upon.

Illustrations of how a power spectrum characterizes the
temporal behavior of a signal are provided in the following
examples. First, consider an EEG recording from a subject in
deep sleep (Fig. 3.4a). In such a case the EEG consists primarily
of slowly fluctuating, high amplitude delta wave activity. Conse-
sequently, most of the power is concentrated at low frequencies and
so the spectrum will be relatively large at those frequencies, and
small elsewhere (Fig. 3.4b). As a second example, consider the
EEG of an awake but resting subject. In this case the EEG may
consist of primarily rhythmic, quasinusoidal alpha wave activity
in the 9 to 12 Hz frequency range (Fig. 3.5a). The associated
power spectrum will have a peak in the 9 to 12 Hz range and be
Fig. 3.4. (a) A hypothetical example of a low frequency EEG waveform recorded from an individual in deep sleep. (b) Power spectrum corresponding to the low frequency EEG process. relatively small elsewhere (Fig. 3.5b). In the third example, consider the EEG of an alert subject. Here the EEG tends to consist of low amplitude waves with rapid, irregular fluctuations (Fig. 3.6a). No predominant rhythms or slow fluctuations are apparent. The corresponding power spectrum will tend to be broadly distributed over the frequency range of the EEG (Fig. 3.6b), a range which extends to an upper frequency of about 30 to 50 Hz. 

Fig. 3.5. (a) A hypothetical example of EEG alpha activity. (b) Power spectrum corresponding to an EEG alpha process with pronounced alpha activity.

The three foregoing examples illustrate how the power spectrum provides a characterization of the "average" temporal behavior of a random signal. But it does not uniquely specify the signal it is derived from. One cannot reconstruct the signal given only its power spectrum because the power spectrum does not preserve the phase information in the signal. In effect, the spectrum specifies
Fig. 3.6. (a) A hypothetical example of rapid, irregularly fluctuating EEG recorded from an alert individual. (b) Power spectrum corresponding to the rapid, irregularly fluctuating EEG.

The average strength of a signal at each frequency. The average strength at a given frequency reflects both the amount of time during which there is activity at that frequency and the strength of that activity. For example, consider Fig. 3.7 which illustrates both a persistent, relatively low amplitude rhythmic random signal (a), and a signal in which relatively high amplitude bursts of rhythmic activity occur irregularly (b). The magnitudes of the power spectra corresponding to the two signals may be the same near the frequency of the rhythm. Although the signal in Fig. 3.7b has higher amplitudes during the bursts of rhythmic activity, the average power near the frequency of the rhythm is no greater than that of the signal in Fig. 3.7a because the duration of the rhythmic activity in (b) is less than in (a).
3.7. POWER SPECTRUM OF CONTINUOUS RANDOM SIGNALS

In the above discussion we presented the concept of the power spectrum from an empirical point of view. We held that the variance of the output signal of a narrow band linear filter provides a measure of the power of the components of the input signal whose frequencies are in the pass band of the filter. We now examine this statement more closely, taking a mathematical point of view. Consider Fig. 3.8. \( x(t) \) is a wide sense stationary random signal whose power spectrum is of interest to us. For simplicity, we assume that the mean value of \( x(t) \) is zero. \( H(f) \) is the transfer function and \( h(t) \) the corresponding impulse response (weighting function) of the linear filter used to obtain a spectral estimate of \( x(t) \). We shall compute the variance of the filter's output \( x_h(t) \) and relate it to \( x(t) \) as well as to \( h(t) \) and \( H(f) \) and to the power spectrum of \( x(t) \), \( C_{xx}(f) \).

We first state the output signal in terms of the convolution relation between output and input, established in Chapter 2:

\[
x_h(t) = \int_{-\infty}^{\infty} h(\tau)x(t - \tau) \, d\tau
\]  

(3.31)

The variance of the output can be expressed in terms of the variance of the input signal and the filter impulse response function, as given above. Since the output, like the input, has zero mean,

\[
\text{var}\left[x_h(t)\right] = E\left[\left|x_h(t)\right|^2\right] = \left\{E\left[\left(\int_{-\infty}^{\infty} h(\tau)x(t - \tau) \, d\tau\right)^2\right]\right\}
\]

(3.32)

Now the square of an integral can be expressed as the product of two identical integrals, differing only in the symbols used to denote the variable over which the integration is performed. Then we have

\[
\text{var}\left[x_h(t)\right] = \mathbb{E}\left[\left(\int_{-\infty}^{\infty} h(\tau)x(t - \tau) \, d\tau\right)^2\right]
\]

(3.33)

Since the averaging operation is with respect to the random variable \( x(t) \), Eq. (3.33) can be rearranged so that the averaging operation is performed prior to integration over \( \tau \) and \( u \).

\[
\text{var}\left[x_h(t)\right] = \int_{-\infty}^{\infty} h(t) \, dt \int_{-\infty}^{\infty} h(u) \, du \, \mathbb{E}[x(t - \tau)x(t - u)] \, du
\]

(3.34)

\( \mathbb{E}[x(t - \tau)x(t - u)] \) is the autocovariance function (acvf) of \( x(t) \). Since \( x(t) \) is wide sense stationary, the acvf is a function only of the difference between \( \tau \) and \( u \). Denote the acvf by \( C_{xx}(\tau - u) \) and substitute \( C_{xx}(\tau - u) \) into Eq. (3.34). This gives

\[
\text{var}\left[x_h(t)\right] = \int_{-\infty}^{\infty} h(t) \, dt \int_{-\infty}^{\infty} h(u) C_{xx}(\tau - u) \, du
\]

(3.35)

Equation (3.35) indicates that the variance of \( x_h(t) \), the filter output, is determined solely by the filter characteristics and the second-order statistics (acvf) of the input signal. However, Eq. (3.35) does not show clearly just how the filter's action upon the input signal determines the variance of \( x_h(t) \). This can be brought out if Eq. (3.35) is expressed in terms of the frequency response of the filter and the spectrum of the signal as we shall do in the next step. But some comments upon this step are first in order. Up to this point we have used a deductive argument to arrive at Eq. (3.35). We have assumed nothing about the nature or even the existence of the power spectrum. We have only assumed that the input process is stationary and that it has the acvf \( C_{xx}(\tau) \). We now make use of the fact, first mentioned in Chapter 1,
that the power spectrum and the acvf of a wide sense stationary process constitute a Fourier transform pair. The power spectrum is the direct Fourier transform of the acvf, and the acvf is the inverse Fourier transform of the power spectrum. The latter is indicated below, with the power spectrum of \( x(t) \) denoted by \( C_{xx}(f) \).

\[
C_{xx}(f) = \int_{-\infty}^{\infty} C_{xx}(f) \exp(j2\pi ft) \, df
\] (3.36)

Substitution of Eq. (3.36) into Eq. (3.35) yields an expression which relates the variance of the filter output to the power spectrum of the input signal:

\[
\text{var} x_h(t) = \int_{-\infty}^{\infty} h(\tau) \, d\tau \int_{-\infty}^{\infty} h(u) \, du \int_{-\infty}^{\infty} C_{xx}(f) \exp[j2\pi f(\tau - u)] \, df
\] (3.37)

Equation (3.37) can be further simplified by changing the order of integration, as follows:

\[
\text{var} x_h(t) = \int_{-\infty}^{\infty} C_{xx}(f) \, df \int_{-\infty}^{\infty} h(\tau) \exp(j2\pi ft) \, d\tau \int_{-\infty}^{\infty} h(u) \exp(-j2\pi fu) \, du
\] (3.38)

The two right-most integrals in Eq. (3.38) are Fourier transforms of the filter impulse response, and hence may be stated in terms of the filter's transfer function:

\[
\int_{-\infty}^{\infty} h(u) \exp(-j2\pi fu) \, du = H(f)
\] (3.39a)

\[
\int_{-\infty}^{\infty} h(\tau) \exp(j2\pi ft) \, d\tau = H(-f) = H^*(f)
\] (3.39b)

Note that \( H^*(f) \) is the complex conjugate of \( H(f) \). Since the product of a complex quantity and its conjugate equals the squared magnitude of the quantity, substitution of Eqs. (3.39a and b) into Eq. (3.38) yields

\[
\text{var} x_h(t) = \int_{-\infty}^{\infty} C_{xx}(f) |H(f)|^2 \, df
\] (3.40)

This can be seen to specify the variance of the filter's output in terms of both the power spectrum of the input signal and the squared magnitude of the filter's transfer function.

Equation (3.40) indicates that the power spectrum of a random signal is the density of average power at a given frequency. The units are power per Heriz. To see this, suppose that the filter transfer function is unity over a narrow band \( b \) of frequencies centered at frequency \( f_c \) and zero elsewhere. Then,

\[
|H(f)| = \begin{cases} 
1, & f_c - \frac{b}{2} \leq f \leq f_c + \frac{b}{2} \\
0, & \text{elsewhere}
\end{cases}
\] (3.41)

Substitution of Eq. (3.41) into Eq. (3.40) yields

\[
\text{var} x_h(t) = \int_{f_c - \frac{b}{2}}^{f_c + \frac{b}{2}} C_{xx}(f) \, df
\] (3.42)

Since \( b \) is small, the integral in Eq. (3.42) can be approximated by

\[
\text{var} x_h(t) = b C_{xx}(f_c)
\] (3.43)

Rearranging Eq. (3.43), and taking the limit as \( b \) becomes infinitesimally small, yields

\[
C_{xx}(f_c) = \lim_{b \to 0} \frac{\text{var} x_h(t)}{b}
\] (3.44)

Note that \( \text{var} x_h(t) \) represents the total average power of the random process in the narrow pass band of the filter: \( f_c - (b/2) \) to \( f_c + (b/2) \). Thus, from Eq. (3.44) it can be seen that the power spectrum is a density function.
The integral of $C_{xx}(f)$ over all frequencies equals the total power of the random process. This can be inferred from Eq. (3.40) by setting $|H(f)|^2 = 1$ for all $f$. Passing a signal through a filter with a transfer function of unity magnitude in no way alters the amount or the frequency distribution of the average power of a signal. Hence, for this case Eq. (3.40) reduces to

$$\text{var}[x(t)] = \int_{-\infty}^{\infty} C_{xx}(f) \, df$$

(3.45)

It is useful here to reconsider two important properties of the power spectrum previously discussed in Chapter 1.

1. As Eqs. (3.40) and (3.42) indicate, $C_{xx}(f)$ is non-negative at all frequencies.
2. It is an even function of frequency.

With regard to the first property, if negative values could occur, then by suitable filtering one could obtain an output signal with negative power. However, this is impossible since the power of a signal is the signal's variance, and variance, being the average of a squared quantity, can never be negative. The second property can be inferred from the Fourier transform relationship between the power spectrum and acvf, as follows.

$$C_{xx}(f) = \int_{-\infty}^{\infty} c_{xx}(t) \exp(-j2\pi ft) \, dt$$

(3.46)

Replacing the exponential in Eq. (3.46) with its Euler identity yields

$$C_{xx}(f) = \int_{-\infty}^{\infty} c_{xx}(t) (\cos 2\pi ft - j\sin 2\pi ft) \, dt$$

(3.47)

Since $c_{xx}(t)$ is an even function of $t$ and $\sin 2\pi ft$ is an odd function of $t$, the integral of the product of the acvf with the sinusoid will be zero. Hence

$$C_{xx}(f) = \int_{-\infty}^{\infty} c_{xx}(t) \cos 2\pi ft \, dt$$

(3.48)

Changing $f$ to $-f$ in Eq. (3.48) does not alter the cosine and therefore does not alter the integral. Consequently, $C_{xx}(f)$ must be a real, even function of $f$.

### 3.8. THE POWER SPECTRUM OF T-DISCRETE RANDOM SIGNALS

Use of a digital computer for power spectrum computations requires that the continuous signal be sampled. It is important that aliasing errors be avoided if an accurate estimate of the power spectrum is to be obtained. When the signal is band-limited, sampling at the Nyquist rate or faster will insure that aliasing will not occur. If the signal is not band-limited or cannot be sampled at twice its upper band-limit, then it should be low-pass filtered prior to sampling, so that activity at frequencies above one-half the sampling frequency will be effectively eliminated. A power spectrum estimate that is free of aliasing errors can then be obtained for frequencies below one-half the sampling frequency. However, information concerning activity at higher frequencies will necessarily be lost. Although the power spectrum properties of T-discrete signals are closely related to those of the original continuous signals, there are important differences which it is most useful to discuss.

The two approaches commonly used to estimate power spectra via digital computation are:

1. The estimation first of the acvf and from it the power spectrum by the use of the discrete Fourier transform (DFT).
2. The computation of the periodogram, the "raw" spectrum estimate, by applying the DFT to a finite $N$ sample segment of the signal.

With the advent of the fast Fourier transform algorithm (Oppenheim and Schafer, 1975), the periodogram approach is usually the more rapid one. Once the periodogram has been obtained, further steps are necessary to improve the goodness of the spectral
estimate. We will discuss these after paying initial attention to the properties of the periodogram.

3.9. THE FOURIER TRANSFORM FOR T-DISCRETE SIGNALS

The Fourier transform relationship between the power spectrum and the acvf for T-continuous signals has been developed and discussed in Chapter I. The Fourier transform pair is restated here.

\[ C_{xx}(f) = \int_{-\infty}^{\infty} c_{xx}(t) \exp(-j2\pi ft) \, dt \]  
\[ c_{xx}(t) = \int_{-\infty}^{\infty} C_{xx}(f) \exp(j2\pi ft) \, df \]  

(3.49)  
(3.50)

An analogous relationship can be shown to hold for T-discrete signals. If the period between samples is \( \Delta \) sec and the upper band-limit of the signal is less than or equal to \( 1/2\Delta \), then Eq. (3.50) becomes

\[ C_{xx}(t^o\Delta) = \int_{-1/2\Delta}^{1/2\Delta} C_{xx}(f) \exp(j2\pi ft^o\Delta) \, df \]  

(3.51)

The acvf is defined only at the discrete times of \( t^o\Delta \), where \( t^o \) is an integer that can range from minus to plus infinity. However, \( C_{xx}(f) \) is a continuous function of frequency. Note that Eq. (3.51) is obtained from Eq. (3.50) by direct substitution of \( t^o\Delta \) for \( t \) and setting the limits of integration to correspond to one-half the Nyquist frequency.

The discrete analog of Eq. (3.49) is a summation over the discrete set of acvf values:

\[ C_{xx}(f) = \Delta \sum_{t^o=-\infty}^{\infty} c_{xx}(t^o\Delta) \exp(-j2\pi ft^o\Delta) \]  

(3.52)

When Eq. (3.52) is compared with Eq. (3.49), we see that \( t^o\Delta \) replaces \( t \), a summation replaces the integral, and the finite time increment \( \Delta \) replaces the infinitesimal \( dt \). The correspondence between Eq. (3.52) and (3.49) has been given here by making some intuitively reasonable changes in the original T-continuous transform pair. We will now demonstrate that the relationship is a mathematically valid one. This is done by substituting for \( C_{xx}(f) \) in Eq. (3.51) the right side of Eq. (3.52).

\[ c_{xx}(t^o\Delta) = \int_{-1/2\Delta}^{1/2\Delta} \Delta \sum_{\tau=-\infty}^{\infty} c_{xx}(\tau^o\Delta) \exp(-j2\pi f\tau^o\Delta) \exp(j2\pi f\tau^o\Delta) \, df \]  

(3.53)

Interchange of the order of integration and summation yields

\[ c_{xx}(t^o\Delta) = \Delta \sum_{\tau=-\infty}^{\infty} c_{xx}(\tau^o\Delta) \int_{-1/2\Delta}^{1/2\Delta} \exp[j2\pi f(t^o - \tau^o)\Delta] \, df \]  

(3.54)

The integral on the right side is easily shown to be

\[ \frac{1}{\Delta} \sin \frac{\pi(t^o - \tau^o)}{\Delta} = \int_{-1/2\Delta}^{1/2\Delta} \exp[j2\pi f(t^o - \tau^o)\Delta] \, df \]  

(3.55)

When both \( t^o \) and \( \tau^o \) are integers, the above integral is zero except for \( t^o = \tau^o \), for which case the integral equals \( 1/\Delta \). Hence, substitution of Eq. (3.55) into Eq. (3.54) results in the elimination of all terms in the summation over \( \tau^o \), except the \( \tau^o = t^o \) term. The \( \Delta \) and \( 1/\Delta \) factors cancel. What is left is an identity proving the equality of Eq. (3.54) and demonstrating the validity of the Fourier transform pair for T-discrete signals, Eqs. (3.51) and (3.52).

We noted above that \( C_{xx}(f) \) is a continuous function. Examination of Eq. (3.52) also indicates that \( C_{xx}(f) \) is a periodic function of frequency, since all the complex exponentials in the summation are periodic with the fundamental frequency being \( 1/\Delta \). This property was to be expected in view of the discussion of aliasing in Section 3.3. Note that only the frequency components between \(-1/2\Delta\) and \( 1/2\Delta \) are needed to describe the signal.
3.10. THE PERIODOGRAM

The intention of this section is to show that the power spectrum of a stationary random process can be estimated through use of the periodogram without having first to estimate the acvf. We will show that the periodogram is equivalent to a Fourier transform of the acvf. To do this we first discuss (1) the properties of an estimated acvf which is based upon a finite segment of a T-discrete waveform; and (2) the properties of an estimated power spectrum which is based upon the Fourier transform of such a specimen acvf.

An estimate of the acvf of a stationary random process can be computed from a T sec segment of the process. A set of N consecutive samples spaced Δ sec apart is used as follows:

\[ \hat{C}_{xx}(\tau) = \frac{1}{N} \sum_{t=0}^{N-1} x(t) x(t+\tau) \Delta, \quad |\tau| \leq N - 1 \]  

(3.56)

Note that the upper limit of the summation is a function of \( \tau \).

This is because there are only a finite number of sample products available. For example, in the \( \tau = 0 \) case, all N points can be used to compute the cross products \( x(t) x(t) \). In the \( \tau = 1 \) case, only \( N - 1 \) points can be used to compute the cross products since, when \( t = N - 1 \), the cross product becomes \( x[(N - 1)\Delta] x(N\Delta) \).

The only data samples available are for the time points at 0 through \( (N - 1)\Delta \). There is no NA time sample available unless, as noted in Section 3.2, the data are periodicized. This will be discussed further in Section 3.18. Thus the summation over the cross products must be limited to the range of \( \tau = 0 \) to \( \tau = N - 2 \) when \( \tau = 1 \). Similar reasoning is applicable to larger magnitudes of \( \tau \), in which case still fewer sample cross products are available. The expected value of \( \hat{C}_{xx}(\tau) \) is

\[ \mathbb{E}[\hat{C}_{xx}(\tau)] = \frac{1}{N} \sum_{t=0}^{N-1} \mathbb{E}(x(t) x(t+\tau)) \Delta \]

(3.57)

Since the summation consists of \( N - |\tau| \) terms, all equal to unity, the sum equals \( N - |\tau| \) and so Eq. (3.57) becomes

\[ \mathbb{E}[\hat{C}_{xx}(\tau)] = \left( 1 - \frac{1}{N} \right) \hat{C}_{xx}(\tau) \]

(3.58)

Equation (3.58) indicates that Eq. (3.56) is a biased estimator of the acvf, and that as the number of sample times \( N \) becomes large with respect to \( |\tau| \), the bias becomes small.

An estimate of the power spectrum can then be obtained by using Eq. (3.52) to compute the Fourier transform of the acvf estimate, Eq. (3.56). The summation index \( \tau \) is confined to the range \(- (N - 1)\) to \( (N - 1)\) since only \( N \) time points are available in the original sampled data segment and positive and negative values of \( \tau \) are permitted up to \( N - 1 \). We then have for the estimate of the power spectrum,

\[ \hat{C}_{xx}(f) = \frac{1}{N} \sum_{\tau=-(N-1)}^{N-1} \mathbb{E}(x(t) x((t + \tau)\Delta) \exp(-j2\pi ft) \Delta) \]

(3.59)

This equation forms a basis for estimating the power spectrum although, as will be shown, some modifications are needed so as to obtain statistically acceptable results.

From a practical point of view, evaluation of Eq. (3.59) can entail relatively large amounts of computer time when \( N \) is large. For this reason it may be advantageous to estimate the power spectrum directly by means of the periodogram of the waveform specimen as expressed by the following equation:
\[ P_{xx}(n) = \frac{\Delta}{N} \left| \sum_{t^o=0}^{N-1} x(t^o\Delta) \exp(-j2\pi nt^o/N) \right|^2 \] (3.60)

\[ P_{xx}(f) = \frac{1}{T} \left| X(f) \right|^2 \] (3.61)

When \( x(t) \) is band-limited, we can resort to the sampled representation and the Fourier transform,

\[ P_{xx}(f) = \frac{\Delta}{N} \left| \sum_{t^o=0}^{N-1} x(t^o\Delta) \exp(-j2\pi f t^o\Delta) \right|^2 \] (3.62)

Usually only the harmonic frequencies \( f_n = n/T = n/N\Delta \) are of interest to us (by periodicizing the original data), and we can obtain the periodogram from the DFT:

\[ P_{xx}(f_n) = \frac{\Delta}{N} \left| X_N(n) \right|^2 \]

\[ = \frac{\Delta}{N} \left| \sum_{t^o=0}^{N-1} x(t^o\Delta) \exp(-j2\pi n t^o\Delta) \right|^2 \] (3.63)

On occasion we shall write \( P_{xx}(f_n) \) as \( P_{xx}(n) \) so that the two notations are equivalent.

We will now show that \( P_{xx}(f) \) is equal to the \( \hat{C}_{xx}(f) \) defined in Eq. (3.59). As a first step, we note that the square of the magnitude of a complex quantity is equal to the product of that quantity and its complex conjugate. Hence,

\[ P_{xx}(f_n) = \frac{\Delta}{N} \sum_{t^o=0}^{N-1} x(t^o\Delta) \exp(-j2\pi f_n t^o\Delta) \sum_{u^o=0}^{N-1} x(u^o\Delta) \exp(j2\pi f_n u^o\Delta) \]

\[ = \frac{\Delta}{N} \sum_{t^o=0}^{N-1} \sum_{u^o=0}^{N-1} x(t^o\Delta) x(u^o\Delta) \exp(-j2\pi f_n (t^o - u^o)\Delta) \] (3.64)

We now make a change of variables, substituting \( \tau^o \) for \( t^o - u^o \).

Since both \( t^o \) and \( u^o \) range from 0 to \( N - 1 \), the range will be

\[ - (N - 1) \text{ to } (N - 1) \]. Hence, Eq. (3.64) becomes

\[ P_{xx}(f) = \frac{\Delta}{N} \sum_{\tau^o=-(N-1)}^{N-1} \sum_{u^o=0}^{N-1} x(u^o\Delta) x[(u^o + \tau^o)\Delta] \exp(-j2\pi f \tau^o\Delta) \] (3.65)

Note that the upper limit of the summation over \( u^o \) has been reduced by \( |\tau^o| \). The reasons are the same as for the summation in Eq. (3.56). Comparison of Eq. (3.59) with Eq. (3.65) indicates that the periodogram \( P_{xx}(f) \) is identical with the spectral estimate \( \hat{C}_{xx}(f) \) obtained by means of the Fourier transform of the sample acvf. The reason for preferring the periodogram as the vehicle for spectral estimation is that it can be computed more rapidly, provided that a fast Fourier transform algorithm is used.

3.11. STATISTICAL ERRORS

OF THE PERIODOGRAM--BIAS

We previously indicated that the specimen or sample acvf, which is used explicitly in Eq. (3.59) and implicitly in Eq. (3.60), provides a biased estimate of the acvf. Consequently, the periodogram will provide a biased estimate of the power spectrum. The expected value of the periodogram can be obtained by substituting Eq. (3.58), the expected value of the sample acvf, into Eq. (3.65).

\[ E[P_{xx}(f_n)] = \frac{\Delta}{N} \sum_{\tau^o=-(N-1)}^{N-1} \left( 1 - \frac{1}{N} \right) c_{xx}(\tau^o\Delta) \exp(-j2\pi f_n \tau^o\Delta) \] (3.66)

Comparison of this equation with Eq. (3.52), which defines \( c_{xx}(f) \) as the Fourier transform of \( c_{xx}(\tau^o\Delta) \), yields

\[ E[P_{xx}(f_n)] = C_{xx}(f_n) - \sum_{\tau^o=-(N-1)}^{N-1} c_{xx}(\tau^o\Delta) \exp(-j2\pi f_n \tau^o\Delta) \]

\[ - \sum_{\tau^o=-(N-1)}^{N-1} c_{xx}(\tau^o\Delta) \exp(-j2\pi f_n \tau^o\Delta) \]

\[ = \frac{\Delta}{N} \sum_{\tau^o=-(N-1)}^{N-1} \left| \tau^o \right| c_{xx}(\tau^o\Delta) \exp(-j2\pi f_n \tau^o\Delta) \] (3.67)
The three right-most terms in Eq. (3.67) constitute the bias. Assuming that \( x(t) \) is a zero mean random process, the bias will tend toward zero as \( N \) becomes large.

To examine the nature of the bias in the frequency domain, we can rewrite Eq. (3.66) in a somewhat more general form, as follows:

\[
E[P_{xx}(f)] = \frac{\Delta}{N} \sum_{\tau = -N}^{N} w_B(\tau \Delta) c_{xx}(\tau \Delta) \exp(-j2\pi f \tau \Delta) \tag{3.68}
\]

where

\[
w_B(\tau \Delta) = \begin{cases} 
1 - \frac{|\tau \Delta|}{N}, & |\tau \Delta| < N \\
0, & |\tau \Delta| \geq N
\end{cases} \tag{3.69}
\]

The function \( w_B(\tau \Delta) \) can be thought of as a "lag window" function which multiplies or weights the set of acvf terms, and, since it is different from unity, "causes" the periodogram to be a biased estimate of the power spectrum. Since we showed in Chapter 1 that multiplication in the time domain is the equivalent of convolution in the frequency domain, Eq. (3.68) can be stated in the frequency domain as

\[
E[P_{xx}(f)] = \frac{\Delta}{N} \int_{-1/2\Delta}^{1/2\Delta} c_{xx}(f) w_B(f_n - f) \, df \tag{3.70}
\]

where \( w_B(f) \) is the Fourier transform of \( w(t) \). It can be shown that

\[
w_B(f) = \frac{N-1}{N} \sum_{\tau = -N}^{N-1} \exp(-j2\pi \tau f \Delta)
\]

\[
= \frac{1}{N} \left( \frac{\sin \pi \Delta f}{\sin \pi f} \right)^2 \tag{3.71}
\]

\( w_B(f) \) can be thought of as a "frequency window" function. Substituting Eq. (3.71) into Eq. (3.70) gives

\[
E[P_{xx}(f)] = \frac{\Delta}{N} \int_{-1/2\Delta}^{1/2\Delta} \frac{1}{N} \left( \frac{\sin \pi \Delta(f_n - f)}{\sin \pi \Delta(f_n - f)} \right)^2 c_{xx}(f) \, df \tag{3.72}
\]

Fig. 3.9. (a) Plot of the window function, \( W_B(f) = (\sin \pi \Delta f)^2 / (\sin \pi f)^2 \), for \( N = 10 \). (b) An illustration of how the product of \( c_{xx}(f) \) and \( w_B(f_n - f) \) determines the expected value of the estimate of the power spectrum.

The frequency band of interest, -1/2\( \Delta \) to 1/2\( \Delta \), \( W_B(f_n - f) \) is near zero except at \( f = f_n \). Hence, only that portion of \( c_{xx}(f) \) which is near frequency \( f_n \) will contribute much to the periodogram estimate of the power spectrum. This is illustrated in Fig. 3.9b.

Note that \( E[P_{xx}(f_n)] \) is equal to an area determined by the product of \( c_{xx}(f) \) and \( W_B(f_n - f) \). As \( N \) becomes large, the area becomes
more closely confined to frequencies that are near to \( f_n \). This means that by increasing \( N \) we can obtain a high resolution, small bias estimate whose expected value is close to \( C_{xx}(f_n) \). If \( N \) is small, the expected value of the estimate will contain spectral components covering a broad range of frequencies. In this circumstance only a low resolution, large bias estimate can be obtained, and nuances such as sharp peaks in the spectrum may not be detected.

A rough guide to the size of \( N \) necessary for the bias to become negligible can be obtained from consideration of Eq. (3.72), Fig. 3.9b, and the fact that

\[
\Delta \int_{-1/2\Delta}^{1/2\Delta} \omega_B(f - f) \, df = \int_{-1/2\Delta}^{1/2\Delta} \frac{\sin \pi \Delta(f - f)}{\pi \Delta(f - f)} \, df = 1.0
\]

This means that if \( C_{xx}(f) \) is relatively constant over the band of frequencies where the frequency window function is markedly greater than zero, then the right side of Eq. 3.72 is approximately equal to \( C_{xx}(f_n) \). Inspection of Fig. 3.9b suggests that \( N \) should be such that \( C_{xx}(f) \) does not vary significantly over a frequency bandwidth of about \( 4/\pi \Delta \) Hz.

The concept of leakage that was discussed in Section 3.4 is simply another way of describing bias or resolution. Examination of Fig. 3.9b indicates that the low resolution, large bias situation is one in which activity at frequencies other than the one of interest contributes to (i.e., leaks into) the estimate of \( C_{xx}(f_n) \).

3.12. STATISTICAL ERRORS OF THE PERIODOGRAM--VARIANCE

Since \( P_{xx}(f_n) \) is a function of the set of \( N \) random variables, the \( x(t^n) \), \( P_{xx}(f_n) \) is also a random variable. We already know that as \( N \) becomes large, the mean of \( P_{xx}(f) \) approaches \( C_{xx}(f) \), the power spectrum of \( x(t) \). We now arrive at a troublesome property of the periodogram, namely, that its variance does not become small as \( N \) increases. Instead, the estimation errors contained in \( P_{xx}(f_n) \) will be of the same order of magnitude as the \( P_{xx}(f) \) itself, regardless of \( N \). Consequently, the raw periodogram is not a consistent estimate of the power spectrum \( C_{xx}(f) \). For \( P_{xx}(f) \) to be a consistent estimate in the statistical sense, its mean must approach the true spectrum and its variance must become small as \( N \) becomes large. The periodogram meets the former but not the latter criterion. However, by application of suitable averaging procedures the latter criterion can also be satisfied. We will now discuss the basis of such averaging procedures.

First, to gain insight into the nature of the variance of the periodogram, let us consider the case of a zero mean, white Gaussian process. In this case, the samples \( x(t^n) \) are independent of one another and the power of the process is uniformly distributed over the frequency band from \(-1/2\Delta\) to \(1/2\Delta\). The acvf, \( c_{xx}(\tau^n) \), is zero for all \( \tau^n \) except \( \tau^n = 0 \), at which point it has the value \( \sigma_x^2 \). The spectrum \( C_{xx}(f) \) of such a process is easily seen to be equal to \( \sigma_x^2 \) for all \( f \).

We find the mean and variance of the periodogram at zero frequency by setting \( f = 0 \) in Eq. (3.60). This yields

\[
P_{xx}(0) = \Delta \sum_{n=0}^{N-1} x(t^n)^2
\]

In Section 3.11 it was pointed out that the distribution of the sum of \( N \) identically distributed normal \((\mu, \sigma)\) random variables is normal \((N\mu, \sqrt{N}\sigma)\). Hence, the distribution of \( \sum_{n=0}^{N-1} x(t^n)^2 \) is Gaussian with a mean of zero and variance equal to \( N\sigma_x^2 \). It was further shown in Section 3.11 that the square of a zero mean, unit-standard-deviation Gaussian random variable has a chi-squared distribution with one degree of freedom. Thus inspection of Eq. (3.73) indicates that it can be expressed as the product of a constant times a chi-squared random variate, as follows:

\[
P_{xx}(0) = \Delta \sigma_x^2 \sum_{n=0}^{N-1} x(t^n)^2
\]
The square of the quantity within the brackets has a chi-squared distribution with one degree of freedom. Two other results from Section 1.13 are useful here. The first is that the mean of a chi-squared variable with \( m \) degrees of freedom equals \( m \) and its variance equals \( 2m \). The second is that the product of a constant \( a \) times a chi-squared random variate with \( m \) degrees of freedom has a mean equal to \( am \) and a variance equal to \( 2a^2 m \). Applying these to Eq. (3.74), we have

\[
\begin{align*}
\mathbb{E}[P_{xx}(0)] &= \sigma_x^2 \\
\text{var}[P_{xx}(0)] &= 2\sigma_x^4
\end{align*}
\]

Hence, the standard deviation of \( P_{xx}(0) \) is \( \sqrt{2} \sigma_x \). Since \( C_{xx}(0) \) equals \( \sigma_x^2 \), the standard deviation of \( P_{xx}(0) \) equals \( \sqrt{2} C_{xx}(0) \). This shows that although the expected value of \( P_{xx}(0) \) equals \( C_{xx}(0) \), the variance of \( P_{xx}(0) \) is independent of \( N \), the number of time samples. The coefficient of variation of \( P_{xx}(0) \) is \( \sqrt{2} \).

The above result applies only to the estimate at zero frequency. A basically similar but computationally more tedious development can be made for the value of a periodogram of a white, Gaussian process at any arbitrary frequency. The mean and variance of the periodogram are (Oppenheim and Schafer, 1975)

\[
\begin{align*}
\mathbb{E}[P_{xx}(f)] &= \sigma_x^2 \\
\text{var}[P_{xx}(f)] &= \sigma_x^4 + \frac{\sin^2 2\pi f\Delta}{N \sin 2\pi f\Delta}
\end{align*}
\]

Equation (3.78) reduces to Eq. (3.76) when \( f \) equals zero or \( 1/2\Delta \). Since \( [\sin 2\pi f\Delta/N \sin(2\pi f\Delta)]^2 \) ranges between zero and one, we find that \( \text{var}[P_{xx}(f)] \) does not become small, viz.,

\[
\sigma_x^4 \leq \text{var}[P_{xx}(f)] \leq 2\sigma_x^4
\]

In practice it is only necessary to compute the periodogram at the discrete set of frequencies \( f = n/N\Delta \), where \( n \) is an integer. This makes it possible to use a fast Fourier transform algorithm. At these frequencies, \( \text{var}[P_{xx}(n/N\Delta)] \) equals \( \sigma_x^4 \). Hence for a zero mean, white Gaussian process the expected value of the periodogram equals \( C_{xx}(f) \) while its standard deviation is also approximately equal to \( C_{xx}(f) \). Increasing \( N \) will not reduce the standard deviation.

For a nonwhite random process the results are quite similar. The periodogram again provides a biased estimate of \( C_{xx}(f) \), as indicated by Eqs. (3.67), (3.68) and (3.72). An approximate expression for the variance of the periodogram is (Oppenheim and Schafer, 1975)

\[
\text{var}[P_{xx}(f)] = C_{xx}^2(f) \left[ 1 + \left( \frac{\sin 2\pi f\Delta}{N \sin 2\pi f\Delta} \right)^2 \right]
\]

Thus, as in the case of a white Gaussian process, the standard deviation of the periodogram is equal to \( C_{xx}(f) \) at frequencies \( n/N\Delta \) and is slightly larger at other frequencies. While increasing \( N \) will decrease the bias, as indicated by Eq. (3.72), it will not effectively decrease the standard deviation of the periodogram estimate.

### 3.13. AVERAGING THE PERIODGRAM--THE BARTLETT ESTIMATOR

It is apparent from the preceding discussion that since the periodogram is not a consistent estimator of the power spectrum, a procedure is required that will attenuate the random fluctuations associated with the periodogram and produce a useful spectrum estimate. One such procedure is to divide the signal specimen into a series of subsegments, compute a periodogram for each of them and then average the periodograms. This approach, first suggested by Bartlett (Oppenheim and Schafer, 1975), also gives one the opportunity of testing for stationarity. It is implemented as follows. Let the signal segment be divided into \( N \) subsegments, each \( \Delta \) sec long. Denote the signal in the \( m \)th subsegment by
The corresponding periodogram for the \( m \)th subsegment is
\[
P^{(m)}_{xx}(f) = \frac{1}{M} \sum_{m=1}^{M} \left| \sum_{n=0}^{N-1} x^{(m)}(t^n \Delta) \exp(-j2\pi f t^n \Delta) \right|^2
\]
(3.81)

Thus, using Bartlett's method, the estimate of the power spectrum of \( x(t) \) is
\[
B_{xx}(f) = \frac{1}{M} \sum_{m=1}^{M} P^{(m)}_{xx}(f)
\]
(3.82)

The expected value of the Bartlett estimator at frequency \( f_n \) is
\[
E[B_{xx}(f_n)] = \frac{1}{M} \sum_{m=1}^{M} E[P^{(m)}_{xx}(f_n)]
\]
(3.83)

The expected value of the periodogram, \( P_{xx}(f_n) \), is the same for all \( m \) and is given by Eq. (3.72). Hence, the expected value of the Bartlett estimator is the same as the expected value of the individual periodograms and is given by
\[
E[B_{xx}(f_n)] = \frac{1}{M} \int_{-1/2\Delta}^{1/2\Delta} \left[ \sin \pi \Delta (f_n - f) \right]^2 C_{xx}(f) \, df
\]
(3.84)

The bias leakage properties of the Bartlett estimator are also the same as that of the individual periodograms so that the remarks following Eq. (3.72) concerning bias and leakage of raw periodograms apply here as well. What is most important is that the variance of the Bartlett estimator is less than that of the periodogram, as we shall show in the next section. The argument is based upon the assumption that there is a total of \( N \) data points available, \( N \) being the number of data points in each of the \( M \) subsegments.

3.14. VARIANCE OF THE BARTLETT ESTIMATOR

If \( N_m \), the number of time points in a subsegment, is sufficiently large so that \( c_{xx}(t^n \Delta) \) is small for \( t^n > N_m \), then the various subsegment periodograms, \( P^{(m)}_{xx}(f) \), will tend to be statistically independent of one another. This means that the variance of the average of the \( M \) periodograms will be approximately equal to the variance of the individual periodograms divided by \( M \). (See Sections 1.13 and 4.1.) Using this and Eq. (3.79), it follows that the variance of the Bartlett estimator is approximately
\[
\begin{align*}
\text{var}[B_{xx}(f)] &= \left( \frac{1}{M} \sum_{m=1}^{M} \text{var}[P^{(m)}_{xx}(f)] \right) / M \\
&= \frac{c_{xx}^2(f)}{M} \left[ 1 + \left( \frac{\sin 2\pi f \Delta}{M \sin 2\pi f \Delta} \right)^2 \right]
\end{align*}
\]
(3.85)

Thus, for \( f_n = n/N \Delta \) and unequal to zero or \( 1/2\Delta \),
\[
\text{var}[B_{xx}(f_n)] = \frac{c_{xx}^2(f_n)}{M}
\]
(3.86a)
and when \( f_n = 0 \) or \( 1/2\Delta \),
\[
\begin{align*}
\text{var}[B_{xx}(0)] &= 2c_{xx}^2(0)/M \\
\text{var}[B_{xx}(1/2\Delta)] &= 2c_{xx}^2(1/2\Delta)/M
\end{align*}
\]
(3.86b, 3.86c)

This means that Bartlett's method is a consistent estimator of the power spectrum since, as the total number of data points \( N = M \) increases, both the bias and variance of the estimate become small. The bias, as given by Eq. (3.84), is determined solely by the length of the subsegments \( N_m \), and diminishes as \( N_m \) increases. The variance is determined by the number of subsegments \( M \) to which it is inversely proportional.

Since only a fixed number of time samples \( N \) is available for estimation of the power spectrum, however it is done, there is a trade-off between the size of the variance and the resolution of the Bartlett estimator. Variance is reduced by dividing the data segment into as many subsegments as possible, thereby increasing \( M \). But by so doing, one shortens the length of the sub-
segments $N_m$, and hence increases the bias and decreases the resolution. Thus, the size of variance and bias are inversely related to one another: as one increases, the other decreases. Variance itself is related closely to spectral resolution, the ability to detect fine structure in the spectrum. Decreasing the variance of an estimate is brought about by decreasing the length $N_m$ of a data subsegment. This means that the periodograms have fewer frequency components in them (smaller $N_m$) so that the frequency resolution decreases. Reduced resolution is therefore concomitant with reduced variance. Later we shall show this is another way by speaking of frequency resolution in terms of bandwidth.

3.15. THE FAST FOURIER TRANSFORM AND POWER SPECTRUM ESTIMATION

We mentioned in Section 3.8 that the main reason for using the periodogram approach to power spectrum estimation is that it can be carried out more rapidly than by computing the acvf and then taking its Fourier transform. The savings in time come about by use of the fast Fourier transform algorithm (Bergland, 1969; Oppenheim and Schafer, 1975) to compute the Fourier transforms of the original data, as specified by Eqs. (3.60) and (3.81). In order to take advantage of the fast Fourier transform or FFT, we must confine the frequencies for which the spectral estimate is computed to the discrete set of $f_n = n/N_{\Delta}$, $n = 0, \ldots, N - 1$, where $N_{\Delta}$ is the duration of the segment. This is no restriction since the periodogram of a band-limited process is completely represented by its sample values at frequencies $n/N_{\Delta}$. We must emphasize the fact that the value of the spectral estimate at each frequency does not depend upon whether the FFT or some other algorithm is used. Neither are the bias and variance of the estimate affected by the choice of the algorithm. The only difference may be in computational round-off error, which may be smaller with the FFT, since the FFT entails fewer steps.

3.16. SMOOTHING OF SPECTRAL ESTIMATES BY WINDOWING

We have shown above that although the periodogram itself is not a consistent estimator of the power spectrum, a way of obtaining one is to average across a set of sequentially obtained periodograms. Here we shall develop a different approach to smoothing of the periodogram which also yields a consistent spectral estimate. Our argument will apply mainly to estimates obtained at the discrete set of frequencies $f_n = n/N_{\Delta}$.

Rather than dividing the data into numerous time sequential subsegments and averaging across time, the periodogram can be smoothed by averaging over narrow bands of frequency. One important property of periodogram estimates that we make use of here is that $P_{xx}(f_n)$ for a white Gaussian process is the sum of the square of two identical and independent Gaussian random variables (Jenkins and Watts, 1968), except when $f_n = 0, 1/2N_{\Delta}$. This property is also approximately valid when the Gaussian restriction is eliminated and any peak in the spectrum is broad compared to $1/N_{\Delta}$. This means that in most situations of interest, $P_{xx}(f_n)$ is proportional to a chi-squared random variable with two degrees of freedom. Since

$$E[P_{xx}(f_n)] = C_{xx}(f_n)$$

and

$$\text{var}[P_{xx}(f_n)] = C_{xx}^2(f_n)$$

$2P_{xx}(f_n)/C_{xx}(f_n)$ is a $\chi^2_2$ random variable. A second important property of periodogram estimates is that for a white Gaussian process $\text{cov}[P_{xx}(f_n), P_{xx}(f_m)] = 0$ when $n \neq m$. This property is also approximately valid for nonwhite and some non-Gaussian processes. Thus one can treat values of the periodogram at integer multiples of $1/N_{\Delta}$ as uncorrelated random variables. For more details, see Jenkins and Watts (1968).
Let us now consider a spectral estimate made up of a weighted sum of periodogram values:

\[
\hat{C}_{XX} (f_n) = \sum_{k=n-K}^{n+K} \hat{p}_{XX} (f_k) \hat{W}(f_n - f_k) \tag{3.87}
\]

The \( \hat{W}(f_k) \) are the weights of a spectral smoothing filter which weights and sums the periodogram estimates from \( f_{n-K} \) to \( f_{n+K} \). \( \hat{C}_{XX} (f_n) \) is a new random variable, and when the process is Gaussian, its mean and variance are given by

\[
E[\hat{C}_{XX} (f_n)] = \sum_{k=n-K}^{n+K} E[\hat{p}_{XX} (f_k)] \hat{W}(f_n - f_k) \tag{3.88a}
\]

\[
\text{var}[\hat{C}_{XX} (f_n)] = \sum_{k=n-K}^{n+K} \text{var}[\hat{p}_{XX} (f_k)] \hat{W}^2(f_n - f_k) \tag{3.88b}
\]

Since frequency averaging is usually applied to periodograms obtained from long data segments, the results of Section 3.12 indicate that Eqs. (3.88a and b) can be approximated by

\[
E[\hat{C}_{XX} (f_n)] = \sum_{k=n-K}^{n+K} \hat{C}_{XX} (f_k) \hat{W}(f_n - f_k) \tag{3.89a}
\]

\[
\text{var}[\hat{C}_{XX} (f_n)] = \sum_{k=n-K}^{n+K} \hat{C}_{XX} (f_k) \hat{W}^2(f_n - f_k) \tag{3.89b}
\]

These equations can be further simplified when the process is a white one (even if only in the range of frequencies covered by the summation), in which case its mean and variance are given by

\[
E[\hat{C}_{XX} (f_n)] = \hat{C}_{XX} (f_n) \sum_{k=n-K}^{n+K} \hat{W}(f_k) \tag{3.90a}
\]

\[
\text{var}[\hat{C}_{XX} (f_n)] = \hat{C}_{XX} (f_n) \sum_{k=n-K}^{n+K} \hat{W}^2(f_k) \tag{3.90b}
\]
the degrees of freedom and the smaller the variance. We may assign to the smoothing filter a generalized bandwidth parameter. This is the bandwidth (or the number of frequency components averaged over) that a uniformly weighted filter would have in order to yield an estimator with the same variance as the actual smoothing filter. This assumes the data have a flat spectrum over the range of the smoothing filter. The bandwidth and variance are inversely related so that their product is a constant. This can be readily seen for a white noise process being smoothed by a uniformly weighted filter. The bandwidth = 2K + 1 and the variance = var[xx(n)]/(2K + 1). This means that there is always a trade-off between variance and bandwidth. Small variance is obtained at the cost of large bandwidth (or low resolution) and vice versa.

The trade-offs between variance and resolution are much the same whether a Bartlett estimator, Eq. (3.82), or a more general windowing approach, Eq. (3.87), is used. However, there are differences in details. Inspection of Eq. (3.84) indicates that the Bartlett estimator is the equivalent of using a frequency smoothing filter of the form \( (\sin \pi n f_0 / N \sin \pi f_0)^2 \), referred to as the Bartlett window. While the Bartlett window has been widely used and provides a reasonable balance between variance and resolution, in some instances other window shapes may be more desirable. An advantage of the averaging over the frequency approach is that a wide variety of window functions can be devised according to the particular spectral smoothing problem at hand. Details such as the precise width of the window function can be controlled by direct specification of the \( W(f_k) \) terms.

Although there is some latitude in selecting a spectral window function \( W(f_k) \) for a given application, there are practical constraints that should be evaluated. Thus, while a window that extends over a broad frequency range will yield a low variance estimate, it is associated with leakage from frequencies that are far from the one at which the spectrum is being estimated. If these distant spectral components are large, the window width should be narrowed to reduce the leakage. Another consideration has to do with the values of the \( W(f_k) \). There are relatively common window functions that have negative values for some of the \( W(f_k) \). Such windows must be used with caution since they can lead to negative spectrum estimates.

From Eqs. (3.90b) and (3.91a) it can be seen that the magnitude of \( \sum_{k=-K}^{K} W(f_k) \) is crucial in determining the variance of the spectrum estimate. The smaller the sum of the squares, the smaller the variance. Given the constraint that \( \sum_{k=-K}^{K} W(f_k) = 1 \), it can be shown that the sum of the \( W^2(f_k) \) terms will be smallest when all \( W(f_k) = 1/(2K + 1) \), in which case the sum of the squares equals 1/(2K + 1).

Spectral windowing can also be implemented in the time domain by dealing with the acvf. Since convolution in the frequency domain is equivalent to multiplication in the time domain, the time domain equivalent of Eq. (3.87) is

\[
C_{\hat{xx}}(f) = \Delta \sum_{t=-(N-1)}^{N-1} \hat{w}(t\Delta) \hat{C}_{\hat{xx}}(t\Delta) \exp(-j2\pi ft\Delta) \tag{3.92}
\]

where \( \hat{C}_{\hat{xx}}(t\Delta) \) is given by Eq. (3.56), and \( \hat{w}(t\Delta) \), commonly referred to as a "lag window," is in effect the Fourier transform of \( W(f) \). Prior to the late 1960s, when the FFT became widely known, windowing was usually implemented in the time domain, via Eq. (3.92). It is the Fourier transform of these lag windows that sometimes yields negative \( W(f_k) \). A detailed discussion of the properties of spectral and lag window functions and their implementation can be found in Jenkins and Watts (1968), Otnes and Enochson (1972), and Welch (1967).

3.17. THE CROSS SPECTRUM

In Chapter 1 we discussed the concept of the cross covariance function (ccvf). The Fourier transform of the ccvf is referred to as the cross spectrum. The cross spectrum provides a statement of how common activity between two processes is distributed across
frequency. The cross spectrum is the Fourier transform of the ccvf, as indicated by Eq. (1.69). As an example, consider two processes each of which consists of a quasiperiodic signal embedded in wide band noise processes. Suppose the quasiperiodic signals are due to a common phenomenon so that they are closely related. The wide band noise processes, on the other hand, are due to random fluctuations that are unique to each process and so are unrelated. The cross spectrum of the two processes would be relatively large in the frequency band of the shared, quasiperiodic signal and small at other frequencies, since the wide band noise processes are independent and not shared activity.

To some extent the cross spectrum can provide insight into the relationships between a pair of random processes. Further insight can be obtained from the coherence function, which is derived from the power spectra and cross spectrum of the pair of random processes. The coherence function will be discussed in Section 3.19.

The procedures and problems in estimating cross spectra are similar to those described in the preceding discussion of the power spectra. It can be computed by Fourier transform of the sample ccvf. However, with the availability of the FFT algorithm, a periodogram approach in some instances may be preferable. The bias-resolution and variance properties of the cross spectrum are the same for both approaches and are similar to those of the power spectrum.

For example, consider the ccvf and cross spectrum for two wide sense stationary random signals, \( x(t) \) and \( y(t) \). The sample ccvf may be computed in the same manner as an acvf [see Eq. (3.56)], as follows,

\[
\hat{c}_{xy}(\tau^\Delta) = \frac{1}{N} \sum_{\tau^\Delta=0}^{N-1} x(t^\Delta)y[(t^\Delta + \tau^\Delta)\Delta], \quad |\tau^\Delta| \leq N - 1 \tag{3.93}
\]

The sample cross spectrum can be obtained in the same manner as the sample power spectrum [see Eq. (3.59)], as follows,

\[
\hat{C}_{xy}(f) = \frac{1}{N} \sum_{\tau^\Delta=-(N-1)}^{N-1} \hat{c}_{xy}(\tau^\Delta) \exp(-j2\pi f \tau^\Delta)
\]

The expected value of the above cross-spectrum estimate can be found by the same steps used to arrive at Eq. (3.72), the expected value of the periodogram estimate. The result is

\[ E[\hat{C}_{xy}(f)] = \Delta \int_{-\frac{1}{2}\Delta}^{\frac{1}{2}\Delta} \frac{1}{N} \left( \sin \pi NT(f_n - f) \right) \frac{1}{\sin \pi T(f_n - f)} C_{xy}(f) \, df \tag{3.95} \]

Eq. (3.95) is directly comparable to Eq. (3.72), the expression for the expected value of the periodogram estimate of the power spectrum. As in the case of the periodogram, increasing the length of the epoch segment \( N \) will decrease the bias of the cross-spectral estimate but its variance will not be effectively decreased. Consequently, averaging and/or windowing techniques, as described earlier for estimation of the power spectrum, must also be employed when estimating the cross power spectrum. Further details about cross-spectral estimates may be found in Chapters 8 and 9 of Jenkins and Watts (1968).

### 3.18. COVARIANCE FUNCTIONS

The auto- and cross covariance functions were introduced in Chapter 1 and shown to be a way of representing the temporal relationships within an individual dynamic process and also between different dynamic processes. The Fourier relationship between the cvfs and power spectra was also established for continuous stationary processes and for \( T \) sec realizations of them. To do this for the power spectra we resorted to the artifice of considering a \( T \) sec segment of data to be one period of a periodic process. This provided us with an estimator for the cvf and the spectrum of the continuous aperiodic process. The properties of the spectral estimators have been discussed in the preceding section. Now we move to a more detailed consideration of the covariance function, pointing out some essential features of its estimation and how
PRINCIPLES OF NEUROBIOLOGICAL SIGNAL ANALYSIS

The representation of a T sec segment of data as one period of a periodicized specimen function \( x(t) \) means that the estimated acvf is given by

\[
\hat{c}_{xx,N}(\tau^*) = \frac{1}{N} \sum_{\tau^*=0}^{N-1} \hat{x}(\tau^*) \hat{x}^*(\tau^* + \tau^*)
\]  

(3.96)

and is itself periodic, \( N \). We use the tilde to denote that the acvf has arisen from periodicized data \( \hat{x}(t) \). The subscript \( N \) indicates the periodicity. (Throughout this discussion we will assume the original specimen function to be band limited, \( F = 1/2 \), and sampled at the Nyquist rate so that \( \Delta = 1 \).) Whenever \( \tau^* + \tau^* \) exceeds \( N - 1 \), \( \tau^* + \tau^* \) is to be considered as having its value taken "modulo \( N \)." That means, in this instance, that if \( \tau^* + \tau^* = 117 \) and \( N = 100 \), the value taken for \( \tau^* + \tau^* \) is 17 and \( \hat{x}(117) = \hat{x}(17) \). This follows from the periodicity of \( \hat{x}(\tau^*) \). The estimated acvf that results from the use of Eq. (3,96) is sometimes referred to as a circular covariance function because of this method of computation—the data are in effect considered to be wrapped around a cylinder whose circumference is \( T = N \). The circular covariance function estimator has a serious deficiency that limits its usefulness. The nature of this deficiency can be seen by representing it as two summations:

\[
\hat{c}_{xx}(\tau^*) = \frac{1}{N} \left[ \sum_{\tau^*=0}^{N-1-|\tau^*|} \hat{x}(\tau^*) \hat{x}^*(\tau^* + \tau^*) + \sum_{\tau^*=-N}^{N-1-|\tau^*|} \hat{x}(\tau^*) \hat{x}^*(\tau^* + \tau^* - N) \right]
\]  

(3.97)

The absolute value sign serves to make the equation applicable to both positive and negative delays, though from the symmetry of the acvf about \( \tau = 0 \), only positive values need be considered. Using this fact, it can be seen that the above equation simplifies to

\[
\hat{c}_{xx,N}(\tau^*) = \left( \frac{N - |\tau^*|}{N} \right) \hat{c}_{xx}(\tau^*) + \frac{|\tau^*|}{N} \hat{c}(N - |\tau^*|)
\]  

(3.98)

\( \hat{c}_{xx}(\tau^*) \), of course, is just the average of products of the form \( \hat{x}(\tau^*) \hat{x}^*(\tau^* + \tau^*) \). This means that the circular acvf estimator is a combination of two estimators of the acvf, one for \( \tau^* \) and the other for \( N - |\tau^*| \). These two are inseparable from one another in this method of estimation. Interpretation of the estimated acvf can therefore be a problem. Of course, this is of no consequence when the data really do arise from a process with period \( T \). However, this is not usually the case. Consequently, it is desirable to look for acvf estimation procedures that are free of this problem. We need not seek far for one. All we need to do is adopt another periodicity artifice, one that begins by padding out the original sequence of \( N \) samples with a sequence of samples of 0 amplitude, let us say \( L \) of them. Then the data may be considered to arise from a specimen of a periodic process whose period is \( N' = N + L \). We consider this for the simplest situation, when \( L = N \) and \( N' = 2N \).

In our new sequence, \( \hat{x}(\tau^*) \) of length \( 2N \), data samples \( \hat{x}(N) \) through \( \hat{x}(2N - 1) \) are 0. Because of this, at each time lag \( \tau^* \) there can be only \( N - |\tau^*| \) nonzero products in the acvf estimate formed from the sequence. The acvf is then estimated as the average of these products with, however, the averaging factor being taken as \( 1/|\tau^*| \), \( N \) being the number of nonzero products when \( \tau^* = 0 \) rather than \( 1/(N - |\tau^*|) \). The reason for using the former is that the variance of the resulting estimator turns out to be smaller at larger values of \( \tau^* \) than when using the factor \( 1/(N - |\tau^*|) \). (See Jenkins and Watts, 1968.) This gives for the estimator
The tilde over the data samples is unnecessary. We have also re-
turned to the circumflex notation for the acvf estimate because
circularity has been eliminated in the computation even though we
have arrived at \( \hat{c}_{xx}(\tau^s) \) by an argument involving a periodicity
of \( 2N \). This estimate does not have the difficulty exhibited by the
circular acvf estimate with period \( N \) as given in Eq. (3.96). It
is therefore to be preferred to \( \hat{c}_{xx}(\tau^s) \) in most instances.

The statistical properties of \( \hat{c}_{xx}(\tau^s) \) are of interest. When
the data \( x(t) \) arise from a specimen function of random process \( x \),
we have

\[
E[\hat{c}_{xx}(\tau^s)] = \frac{1}{N} \sum_{t=0}^{N-1} E[x(t^s)x^*(t^s + \tau^s)]
\]

This means that \( \hat{c}_{xx}(\tau^s) \) is a biased estimate of \( c_{xx}(\tau^s) \) because,
as shown earlier, \( E[\hat{c}_{xx}(\tau^s)] - c_{xx}(\tau^s) = -|\tau^s|c_{xx}(\tau^s)/N \). Use of
the averaging factor \( 1/(N - |\tau^s|) \) would eliminate this problem,
but only, as noted above, at the expense of increasing the variance
of the estimate as \( \tau^s \) becomes large. This is generally
thought to be undesirable.

The variance of \( \hat{c}_{xx}(\tau^s) \) may be calculated from its defini-
tion in Eq. (3.99). The result depends upon the statistical prop-
ties of the process. In the Gaussian case, the one of most gen-
eral interest, it can be shown (Jenkins and Watts, 1968) that

\[
\text{var} [\hat{c}_{xx}(\tau^s)] = \frac{1}{N} \sum_{k=0}^{N-1} \left[ c_{xx}(k^s) + c_{xx}(k^s + \tau^s)c_{xx}(k^s - \tau^s) \right]
\]

This means that the variance of the acvf estimate of a Gaussian
process depends upon the acvf itself, something we generally do
not know beforehand. For the particular situation in which the
process is white noise with variance \( \sigma_x^2 \), \( c_{xx}(\tau^s) = \sigma_x^2 \delta(\tau^s) \) and
\( \text{var}[\hat{c}_{xx}(\tau^s)] = \sigma_x^4/N \) for all \( \tau^s \) except \( \tau^s = 0 \), in which case the
variance is \( 2\sigma_x^4/N \). Note that when \( x \) is an aperiodic process with
no dc component, \( c_{xx}(\tau^s) \) becomes small as \( \tau^s \) becomes large. This
means that the summation on the right-hand side of Eq. (3.101)
will be finite so that when we divide it by \( N \) to obtain the var-
iance of \( \hat{c}_{xx}(\tau^s) \), the result becomes small as \( N \) increases, indicat-
ing the estimator to be a consistent one. This also can be shown
to hold when the process is non-Gaussian. Further scrutiny of Eq.
(3.101) seems to indicate that difficulties are encountered when
\( x(t) \) has a periodic component in it, which can occur when there is
residual interference from 60 Hz power lines. In this case,
\( \hat{c}_{xx}(\tau^s) \) does not become small as \( \tau^s \) increases and the summation
becomes infinite. Does this mean that the variance of the esti-
mate is infinite regardless of \( N \)? The answer is no. The diffi-
culty arises in the formulation leading to Eq. (3.101). When
proper account is taken of the pure frequency component in \( x(t) \),
the variance of the estimate turns out to be the same as before.

The statistical relationship between estimates of the acvf
made at neighboring time points is also of some interest. This
refers to the fluctuations of the estimate about the estimated
mean of the acvf. What we are in effect discussing is the covari-
ance of the estimation errors. The problem is a thorny one, but
some results exist for the Gaussian stationary process. In partic-
ular, the covariance between acvf estimates at \( \tau_1^s \) and \( \tau_2^s \) is given by (Jenkins and Watts, 1968)

\[
\text{cov}[\hat{c}_{xx}(\tau_1^s), \hat{c}_{xx}(\tau_2^s)] = \frac{1}{N} \sum_{r=\text{max}}^{N-1} \left[ c_{xx}(r^s)c_{xx}(r^s + \tau_1^s - \tau_2^s) + c_{xx}(r^s + \tau_2^s)c_{xx}(r^s - \tau_2^s) \right]
\]

This equation, from which the previous one was derived, points out
some useful features of the acvf estimate. First, the estimates
are uncorrelated only when the \( x \) process is a white noise with \( c_{xx}(\tau^o) = \sigma^2 \delta(\tau^o) \). Second, for any process which has an acvf with non-zero values extending over \( K \) successive intervals, there will be a non-zero covariance between acvf estimates that are closer than \( 2K \) apart, that is, for which \( |\tau^o_1 - \tau^o_2| < 2K \). Narrow band processes have covariance functions of this type. The covariance between estimates becomes smaller as \( |\tau^o_1 - \tau^o_2| \) approaches \( 2K \). But the major fact is that when the process is a narrow band one, a larger \( N \) is required to obtain an acvf estimate in which the covariance between estimates is to be kept beneath a given maximum. This can be of importance in dealing with acvf estimates of the EEG. An EEG with a marked alpha component will, for a fixed \( N \), have a greater amount of covariance between acvf estimates than will an estimate of the covariance function obtained when the alpha component is small or lacking. Another aspect of the covariance function of narrow band processes is that there is little, if anything, to be gained by smoothing the acvf estimates because this does not reduce the covariance between neighboring estimates.

**B. ESTIMATION OF THE ACVF**

The functional form of the estimator in Eq. (3.99) suggests the obvious "brute force" way of calculating the estimates: averaging for each value of \( \tau^o \) the \( 0 \leq |\tau^o| < N \) products obtained from the \( N \) samples sequence. Computationally, the procedure is a lengthy one since complete evaluation of \( \hat{c}_{xx}(\tau^o) \) requires that there be \( N(N + 1)/2 \) multiplications and \( N(N - 1)/2 \) additions, a total of \( N^2 \) arithmetic operations. When \( N \) is large, the time required to complete this task becomes excessive. While some short cuts have been found for these time domain procedures, the net time savings has not been impressive. What has brought about a significant reduction in computation time has been the fast Fourier transform algorithm. Its use makes it possible to obtain estimates of the acvf by first estimating the periodogram of the data and then taking the inverse discrete Fourier transform. Since there are about \( N \log_2 N \) operations involved in estimating the periodogram and another \( 2N \log_2(2N) \) in taking the inverse DFT, the great computational savings are apparent. For example, when \( N = 1000 \), the method of Eq. (3.99) requires about \( 10^7 \) operations, while the DFT method requires about \( 4 \times 10^4 \) operations. The reduction in the number of operations is by a factor of over 25, a factor that increases as \( N \) increases. Because the DFT is such an efficient approach to acvf estimation when \( N \) is large, we shall describe it further.

We have already noted that the acvf estimate of Eq. (3.99) can be considered to arise from a periodicized process whose initial \( N \) samples are the \( x(t^o \delta) \) and whose final \( N \) samples are all zeros. To guard against spectral leakage effects of the dc component, we subtract out the average value of the \( N \) samples before padding the sequence with zeros. We may also de-trend the data if that seems warranted. The resulting sequence of \( 2N \) points then possesses the acvf we are interested in. An alternative way of arriving at this acvf is to first obtain the periodogram of the padded sequence. The periodogram of an unpadded sequence of \( N \) data points has been given in Section 3.10, Eq. (3.60). When the sequence is padded to length \( N' \) by adding \( L \) consecutive zeros such that \( N' = N + L \), the periodogram of the padded sequence is

\[
P_{ xx, N'}(n) = \frac{1}{N'} \left| X_{ N'}(n) \right|^2 = \frac{1}{N'} \left| \sum_{t^o=0}^{N-1} x(t^o) \exp(-j2\pi nt^o/N') \right|^2 \tag{3.103}
\]

The upper limit in the summation is \( N - 1 \) rather than \( N' - 1 \) because the last \( L \) values of the sequence are zero. When \( N' = 2N \), we have

\[
P_{ xx, 2N}(n) = \frac{1}{2N} \left| \sum_{t^o=0}^{N-1} x(t^o) \exp(-j2\pi nt^o/2N) \right|^2 \tag{3.104}
\]
Notice that because the fundamental interval is $2N$ rather than $N$ in length, there are twice as many frequencies present in the $2N$ periodogram. These additional frequency components are required to express the fact that the second half of the sample sequence is constrained to be zero. They afford no additional information about the original data but only serve as a computational vehicle to arrive at the acvf estimate. Note also that the presence of $2N$ rather than $N$ in the denominator does not increase the number of operations involved in the computation.

Having once obtained $P_{xx,2N}(n)$, its inverse DFT can be taken and it yields the estimated acvf:

$$
\hat{C}_{xx}(\tau^*) = \frac{1}{N} \sum_{n=-(N-1)}^{N-1} P_{xx,2N}(n) \exp(j2\pi n\tau^*/2N) \quad (3.105)
$$

Use of the factor $1/N$ rather than $1/2N$ in the above equation might, at first glance, appear to be an error. It can be verified to be correct by taking the DFT of the padded sequence $\hat{x}(\tau^*)$ and substituting this into Eq. (3.99, top). After carrying out the summations and using Eq. (3.103), we arrive at Eq. (3.105). Furthermore, because $C_{xx}(\tau^*)$ is an even function, the computation need only be carried out for positive values of $\tau^*$. Another way of writing Eq. (3.105) takes advantage of the fact that $P_{xx,2N}(n)$ is real. Using this, we have

$$
\hat{C}_{xx}(\tau^*) = \frac{P_{xx,2N}(0)}{N} + \frac{2}{N} \sum_{n=1}^{N-1} P_{xx,2N}(n) \cos(2\pi n\tau^*/2N) \quad (3.106)
$$

The derivation of the estimated acvf from the periodogram has just been shown to be valid for all values of $\tau^*$ up to $N$. In practice, there is usually little need to carry this out to such large lag values. Usually, lags that are less than 10% of $N$ are only of interest. Because of this there are further savings to be obtained in the use of the DFT. Let us assume that the acvf is of interest up to a lag of $L < N$. Then when we pad the original sequences of $N$ data samples, we need to add $L$ zeros to get an overall sequence of length $N' = N + L$. This guarantees that any estimation of $C_{xx}(\tau^*)$ at values of $\tau^* < L$ will be free from wraparound or overlap effects with the next period of the periodicized data. The effect of padding the data with $L$ zeros is shown in Fig. 3.10 when the lag is $L$. It can be seen that there are $N - L$ products which are nonzero and $L$ which are forced to be zero, and that none of the nonzero products arises from the overlap of one period with the next. The $N'$ periodogram of the padded data is given (after average values and possible trends have been removed) by Eq. (3.104) rewritten here

$$
P_{xx,N'}(n) = \frac{1}{N'} \sum_{\tau^*=0}^{N-1} x(\tau^*) \exp(-j2\pi n\tau^*/N') \quad (3.107)
$$

where $N' = N + L$. As before, we have a larger range of $n$ to deal with, but the additional frequency terms in the periodogram only serve to take the padding with zeros into account. The inverse DFT then yields our estimate of the acvf:

$$
\hat{C}_{xx}(\tau^*) = \frac{1}{N'} \sum_{n=-(N'-1)}^{N'-1} P_{xx,N'}(n) \exp(-j2\pi n\tau^*/N'), \quad 0 \leq |\tau^*| \leq L \quad (3.108)
$$
Because $L$ is usually small compared to $N$, the inverse transform involves not many more operations than does the computation of the periodogram.

C. CROSS COVARIANCE FUNCTION ESTIMATION

The computation of the ccvf for two $N$ length data sequences $x(t^*$) and $y(t^*)$ follows the same principles that hold for the acvf estimate. Again, we assume $\Lambda = 1$. The use of the direct and inverse DFT facilitates these computations when $N$ is large. If we are interested in estimating the ccvf for lags up to $L$, then padding the $x$ and $y$ sequences with $L$ zeros each eliminates the possibility of an overlap in the computation. The procedure to be used, therefore, after padding the sequences, is to obtain their respective DFTs, $X_N(n)$ and $Y_N(n)$. From them we obtain the raw cross-spectrum estimate $P_{xy,N}(n) = \frac{1}{N} X_N(n) Y_N^*(n)$ and then the estimated ccvf:

$$
\hat{c}_{xy}(\tau^*) = \frac{1}{N} \sum_{n=-N/2}^{N/2-1} P_{xy,N}(n) \exp(j2\pi\tau^*/N), \quad -L \leq \tau^* \leq L
$$

(3.109)

It will be remembered that $c_{xy}(\tau^*)$ is not an even function of $\tau^*$ and so its ccvf is to be estimated at both positive and negative values of $\tau^*$. This means a doubling of the length of the last step of the computation, but when $N$ is large, the FFT still produces a substantially shorter computation than the brute force method.

The statistical properties of the ccvf are close enough to those of the acvf so that a full development of them would in the main be repetitious. Consequently, we bring out only the highlights of the development and move quickly to the results. The most common form of the ccvf estimator is the biased version

$$
\hat{c}_{xy}(\tau^*) = \frac{1}{N} \sum_{t^*=0}^{N-1} x(t^*)y^*(t^* + \tau^*)
$$

(3.110)

$\hat{c}_{xy}(\tau^*)$ can be considered to be one period of a $2N$ periodic function, and, as already shown, this is especially important when it is obtained by Fourier methods. The biased version of the estimator is preferred for the same reason as is the biased version of the acvf, that it tends to yield a smaller variance in the estimate when $\tau^*$ becomes large. The variance of the ccvf estimator is derivable from its definition. When both processes are Gaussian, it is given by (Jenkins and Watts, 1968),

$$
\text{var}[\hat{c}_{xy}(\tau^*)] = \frac{1}{N} \sum_{\tau^*=0}^{2\pi} [c_{xx}(\tau^*)c_{yy}(\tau^*) + c_{xy}(\tau^* + \tau^*)c_{yx}(\tau^* - \tau^*)]
$$

(3.111)

This shows that the variance is calculable only when we know what the ccvf and both acvfs are. If both processes are white and uncorrelated, the second term drops out and we have

$$
\text{var}[\hat{c}_{xy}(\tau^*)] = \frac{\sigma_x^2\sigma_y^2}{N}
$$

(3.112)

The principal fact about the ccvf estimator is that it is a consistent one. Also in common with the acvf estimator, the covariance between estimates at two different lag times depends upon the difference between the lags and the covariance properties of the processes. The covariance of the estimator is a generalization of Eq. (3.111) which we show here for the special case when $x$ and $y$ are uncorrelated:

$$
\text{cov}[\hat{c}_{xy}(\tau^*), \hat{c}_{xy}(\tau^*_2)] = \frac{1}{N} \sum_{\tau^*=-\infty}^{\infty} c_{xx}(\tau^*)c_{yy}(\tau^* + \tau^*_2 - \tau^*)
$$

(3.113)

Equation (3.113) can be seen to be a discrete convolution of the two acvfs, the separation variable being $\tau^*_2 - \tau^*_1$. Among other things, this means that when $X$ and $Y$ are uncorrelated narrow band (nearly sinusoidal or pacemakerlike) processes centered at about the same frequency, the covariance between estimates can rise and fall cyclically over an extensive range of time separations. This
in turn can lead to spurious indications of covariance between processes unless special measures are taken, beyond merely increasing $N$, to reduce the magnitude of the estimated covariance between estimates. One such measure is prefiltering the $X$ and $Y$ data to individually "whiten" them before the covariance testing is carried out. The details of such a procedure are beyond the scope of this presentation and may be found in Jenkins and Watts (1968). However, the net import is that the use of the ccvf estimator as a means for measuring dependency between processes is beset with difficulties. These should be carefully assessed before experimentation designed to exploit ccvf estimation is entered into. There is a distinct danger of arriving at erroneous conclusions, especially in the case of pacemakerlike processes.

3.19. COHERENCE FUNCTIONS

The difficulties associated with ccvf estimation have brought about the development of an alternative method for evaluating the relationship between continuous processes, the coherence function. The coherence function is a measure based upon the auto- and cross-spectral properties of the processes, not upon their cvfs. It closely resembles the square of a correlation coefficient between the spectral components of the processes at a particular frequency $f$. Thus the coherence function, or squared coherence, is defined as

$$\kappa^2_{xy}(f) = \frac{|C_{xy}(f)|^2}{C_{xx}(f)C_{yy}(f)} \quad (3.114)$$

Because the $|C_{xy}(f)|^2$ ranges in absolute value from 0 to $C_{xx}(f)C_{yy}(f)$, $\kappa^2_{xy}(f)$ can be seen to be a normalization of the square of the cross spectrum by the product of the autospectra. The normalization is important because it compensates for large values in the cross spectrum that may have been brought about not by an increase in the coupling between the processes at frequency $f$ but by an inherently large concentration of power at that frequency in either the $X$ or $Y$ process. If the $X$ and $Y$ processes are identical, then $C_{xy}(f) = C_{yx}(f) = C_{xx}(f)$ and $\kappa^2_{xy}(f) = 1$ at all frequencies. At the opposite extreme, if $X$ and $Y$ are independent processes, $C_{xy}(f) = 0$ and $\kappa^2_{xy}(f) = 0$ at all frequencies. Between these two extremes there lies a wealth of possible relationships between the processes that can often be measured usefully by the coherence function. It may be, for example, that $X$ and $Y$ are closely related but only over a limited range of frequencies. This would be the case if $X$ and $Y$ each represented a noisy "locked in" response to a sinusoidal signal of frequency $f_0$. In this case the coherency would be nearly unity at $f_0$ and zero elsewhere. Similar situations may exist when the processes are not driven ones. They may be highly coherent over certain ranges of frequency and incoherent elsewhere. Note should be taken here of the fact that the coherence function suppresses any phase information concerning the two processes—it considers their relationship only in terms of power at a given frequency. Later in the chapter we discuss the use of phase measures to detect process interrelationships. It is also worth noting that when one of the processes is a well-defined stimulus, coherency measures are inferior to average response or cross-correlation techniques. Coherency measures find their major application when the processes are substantially random ones.

The coherence function exemplifies a change in emphasis from temporal to frequency measures. It can bring a certain amount of clarification to interprocess relationships. In this regard the estimator of the coherence function has properties that seem to be superior to those of the ccvf estimator. It is these properties which we consider now. The estimator $\kappa^2_{xy}(f)$ for the coherence function needs to be defined carefully. A meaningful estimate cannot be obtained directly from the raw auto- and cross-spectra of the processes. To see this, it is only necessary to examine what would happen if this were the case, viz.,
Clearly, this is a useless quantity. To be useful, a coherence function estimator must be formed from smoothed spectral estimates of the processes. The smoothing operations, however, necessitate consideration of the same issues that were dealt with in the estimation of auto- and cross-spectra, resolution, and bias. Their effect on the coherence function estimator is more difficult to determine, simply because of the way the coherence function has been defined. Though formal solutions for the bias and covariance of coherence function estimates have not been obtained for all the situations of interest involving (a) different kinds of processes, (b) different spectra, and (c) different smoothed spectral estimators, it has been possible by the use of simulation techniques to develop useful relationships for the bias and variance in many situations of interest. A property of major interest is that the coherence function estimator obtained from smoothed spectral estimates appears to be a robust one. That is, it is insensitive to whether the processes are Gaussian or not. This means that one can employ coherence function estimation without having to be particularly concerned about whether the results of the analysis are sensitive to the amplitude distributions of the particular processes involved.

As a rule, it is the small values of coherence that are especially important to deal with. They are the ones that are normally encountered in dealing with the EEG, for example. Electrode sites that are not close usually produce data in which clear correlations are not obvious. And if they were, there would be little reason to perform a coherence function analysis. To see how large the coherence function might be in a not too unreal situation, let us consider a simple model in which the data sources X and Y consist of a common signal process S embedded in independent noise processes N_1 and N_2. The temporal representation of this situation is

\[ x(t) = n_1(t) + s(t) \]
\[ y(t) = n_2(t) + s(t) \]

The power spectrum representation of this situation is

\[ C_{xx}(f) = C_{n_1 n_1}(f) + C_{ss}(f) \]
\[ C_{yy}(f) = C_{n_2 n_2}(f) + C_{ss}(f) \]
\[ C_{xy}(f) = C_{ss}(f) \]

The last relationship follows from the Fourier transform of the covf between X and Y. We must have \( C_{xy}(\tau) = C_{ss}(\tau) \) because the only correlation between X and Y is that caused by the presence of S in both. The coherence function is then

\[ \kappa_{xy}^2(f) = \frac{C_{ss}(f)}{[C_{n_1 n_1}(f) + C_{ss}(f)][C_{n_2 n_2}(f) + C_{ss}(f)]] \]

If we assume \( n_1 \) and \( n_2 \) to have identical spectra, this can be simplified to

\[ \kappa_{xy}^2(f) = \frac{1}{[1 + C_{nn}(f)/C_{ss}(f)]^2} \]

Let us now consider the signal process to have strength equal to the noise processes at frequency f. Then \( \kappa_{xy}^2(f) = 1/4 \), a rather small coherence. A signal-to-noise ratio of the order of unity tends to be large in comparison to that encountered in a number of interesting neurological situations, and so our major concern insofar as coherence function estimation is concerned must be with the behavior of \( \kappa_{xy}^2(f) \) when coherency is low.

The behavior of the coherence function estimator is best known when it is derived from smoothed spectral estimates having
20 or more degrees of freedom. This means, for example, smoothing over 10 neighboring frequencies with a rectangular spectral window or using 10 data sequences when Bartlett smoothing is employed.

Under these circumstances it has been found (Enochson and Goodman, 1965) that when the squared coherence is between 0.3 and 0.98, its estimator \( \hat{\kappa} \), expressed in terms of the Fisher z variable, has a nearly Gaussian distribution. \( \hat{z} \) is given by

\[
\hat{z} = \tanh^{-1} \kappa_{xy} = \frac{1}{2} \log \frac{1 + \kappa_{xy}}{1 - \kappa_{xy}}
\]

(3.120)

The mean and variance of \( \hat{z} \) are given by

\[
\mu_z = \tanh^{-1} \kappa_{xy} + \frac{1}{\text{d.f.} - 2}
\]

\[
\sigma_z^2 = \frac{1}{\text{d.f.} - 2}
\]

(3.121)

d.f. is the degrees of freedom associated with the spectral smoothing window and has been discussed previously. A rectangular window covering 10 neighboring frequencies has 20 degrees of freedom. The second term in the mean is a bias which becomes small as the degree of smoothing increases. The variance of the estimate also becomes small as the width of the spectral window increases, but obviously one does not wish to widen the window too much and thereby lose spectral resolution. One may surmise, however, that the covariance of coherence function estimates at nearby frequencies increases with the degree of smoothing. When the squared coherence is less than 0.3, one can continue to deal with the z transformed version of \( \hat{\kappa}_{xy} \), but the bias and the variance of the estimator need to be modified. Benignus (1969) has shown by using simulation techniques that a better estimate for \( \kappa_{xy}^2 \), small or large, is

\[
\hat{\kappa}_{xy}^2 = \hat{\kappa}^2 - \frac{2}{\text{d.f.}} \left( 1 - \hat{\kappa}_{xy}^2 \right)
\]

(3.122)

The same techniques also show that a better estimate of the variance of \( \hat{z} \) is given by

\[
\sigma_z^2 = \sigma_z^2 \left[ 1 - 0.014 (1.6 \hat{\kappa}_{xy}^2 + 0.22) \right]
\]

(3.123)

Further refinements to the estimator have been made by Lopes da Silva et al. (1974). Confidence limits for \( \hat{\kappa}_{xy}^2 \) may be constructed using these results. They are shown in Fig. (3.11). N is the number of segments used in Bartlett smoothing, and therefore is twice the number of degrees of freedom of the spectral estimate.

The discontinuities in the upper bounds result from the method of computation and are of no special significance. The thick and thin curves are instructive. Suppose we perform Bartlett smoothing with 16 segments of data. Only when \( \hat{\kappa}_{xy}^2 > 0.23 \) can we then say with about
95% confidence that the two processes have some coherence at the frequency tested. The expected value of the squared coherence is 0.23 but the confidence limits are 0 and 0.46. The figure clearly shows that rather large estimation errors will be the rule rather than the exception when the squared coherence is low. In view of these considerations, it is not surprising that nearly all who discuss the use of the coherence function recommend extreme caution in its use. Even large coherence function estimates may not justify the interpretation that there is dependency between the processes.

Several interesting applications of the coherence function to the study of the EEG have been made. We mention only two. Lopes da Silva et al. (1973) used the coherence function to study the relationship between cortical alpha rhythms and thalamic generators. They found instances of significant coherence between the two regions as well as cortico-cortical coherences which were high over large regions of the cortex. Another interesting application of the coherence function has been given by Gersch and Goddard (1970). They used it to test for the location of an epileptic focus in terms of its nearness to one of a number of electrode sites within the brain. This involved dealing with the coherence function for pairs of data sources (electrode sites) before and after possible coherence with a third site had been taken into account. By showing that the activity of two sites was coherent in the important frequency range of 4-12 Hz when the effects of a third site were present and then became incoherent when the effects of that site were computationally removed, they were able to infer that the third site was near the epileptic focus.

### 3.20. PHASE ESTIMATION

Another method for determining the existence of correlation between two processes is to use the information in the phase of the two processes rather than their power. The phase spectrum is derived from the cross spectrum by the relationship

$$\hat{F}_{xy}(f) = \arctan[-Q_{xy}(f)/L_{xy}(f)]$$  \hspace{1cm} (3.124)

The denominator is the real part of $C_{xy}(f)$ and the numerator the imaginary part. If the processes $X$ and $Y$ are uncorrelated, then no particular phase relationship is to be expected at any frequency. $F_{xy}(f)$ will be a random variable with mean 0 uniformly distributed over the range $-\pi/2$ to $\pi/2$. On the other hand, if there is a correlation between the two processes, this will show up in the phase spectrum in the form of a preferred phase angle related to frequency. For example, when $X$ and $Y$ both contain a signal process $S$ as in Eq. (3.116), the phase spectrum will be 0 for all $f$. If $X$ contains $S$, and $Y$ contains a linearly filtered version of $S$, $F_{xy}(f)$ can take on any real value. The estimator $\hat{F}_{xy}(f)$ of the phase spectrum is a random variable defined by

$$\hat{F}_{xy}(f) = \arctan[-\hat{Q}_{xy}(f)/\hat{L}_{xy}(f)]$$  \hspace{1cm} (3.125)

where $L_{xy}(f)$ and $Q_{xy}(f)$ are, respectively, the real and imaginary parts of $X(f)Y^*(f)$. The phase estimator, like the squared coherence estimator, is useful only when it is preceded by smoothing of the cross spectrum. Under these circumstances the variance of the estimate decreases with increasing squared coherence and the number of degrees of freedom of the smoothed spectral estimate. The relationship is

$$\text{var}[\hat{F}_{xy}(f)] = \frac{1}{\text{d.f.}} \left( \frac{1}{\hat{\kappa}_{xy}^2} - 1 \right)$$  \hspace{1cm} (3.126)

Decreasing the variance of the phase estimator obtained from a fixed length sample by increasing the degrees of freedom brings about, as before, a decrease in the spectral resolution and a lessened ability to detect correlations that may exist only over narrow frequency bands. Discussion of further properties of the
phase estimator may be found in Jenkins and Watts (1968). Thus far it has not been widely applied to the study of EEG activity.

REFERENCES