Report 013376-7-T

SPECTRAL ESTIMATION OF ALL-POLE PROCESSES USING BAYESIAN TECHNIQUES

R. J. Carpinella

COOLEY ELECTRONICS LABORATORY
Department of Electrical and Computer Engineering
The University of Michigan
Ann Arbor, Michigan 48109

June 1978

Technical Report No. 230

Prepared for
OFFICE OF NAVAL RESEARCH
Department of the Navy
Arlington, Virginia 22217

FOREWORD

A major problem in passive observation, one that goes to the very foundations of signal detection theory, is knowing the power spectrum of the background noise. Although the background noise is of no interest to the ultimate user, it determines the design and performance of detection and information extraction equipment.

In underwater acoustics the background noise is not well known because it changes continually, being made up of shipping noise, biological activity, rain noise, and wind-generated noise associated with surface activity. In recent years the noise from oil-drilling rigs, and explosive noise from oil prospecting has added considerably to the background. When these noise makers are somewhat distant from the receiving site, their sounds merge together, and the analyst and the equipment designer would like to lump them all together as one gross "random process."

This conglomeration of acoustic sources is inherently "non-stationary." Unfortunately, the mathematics developed to date does not provide practical structure or even useful guides for dealing with such a nonstationary process. Even the mathematical definition of "noise power spectral density" requires that the process be "wide sense stationary." Fortunately, the background is often sufficiently stable for long enough periods of time (tens of minutes, even hours sometimes) so that results based on stationary assumptions are useful when applied with care.

The research on noise spectrum measurement at Cooley Electronics

Laboratory has the long term goal of providing rigorous mathematics and practical measurement theory for the design and evaluation of detection and information extraction hardware in nonstationary noise backgrounds. The short term goal is to use the available theory to determine the "real essence" of spectral measurement, to find the essentials that can be carried forward in pursuit of the long term goal. The tests for "essential" are that they must be mathematically rigorous and at the same time be practical guides to measurement.

In 1974 J.G. Gobien completed a thesis on spectral estimation of an "all pole" process. The work was very mathematical. It differed from previous work in several ways. The foundation of the work was signal detection theory, not contemporary estimation theory. (The background is not being estimated for its own sake, but only to help equipment.) The measurement procedure uncovered as both necessary and sufficient differed from all current techniques, being neither an FFT (or DFT), nor a Maximum Entropy method, nor a variation on these. It prescribed power measurements of the reception and its several derivatives.

Gobien had to make numerous assumptions, many quite palatable, but two which were not. He had to assume that the "order" of the process and the very high frequency behavior of the process were known. He was forced into these assumptions because the mathematics of continuous processes specifies that these are "singularly estimatable," that is, they can be learned in an arbitrarily short time. (His realization of this, though not totally original, was

a major contribution as it explained why so many research attempts have aborted because they failed to realize it.) These unpalatable assumptions are where the math has failed us -- it says we know the process order and the high frequency behavior, and we know that in practice we do not know these.

The current work of Carpinella reported here removes these two assumptions, and shows how the order and the high frequency behavior can be learned quite rapidly. He attacked the same "all pole" process as Gobien (and as the maximum entropy method), but considers the measurement to be made using analog approximations to differentiators, followed by digital processing for integration and calculations. He allows the digital sampling to be very rapid, so the observation of the input was essentially continuous (in the engineering sense, but not in the mathematical sense). The quantization noise introduced by the digital processing forces an upper bound on the sampling rate which eliminates the "singularity" of the estimation; yet the maximum sampling rate is very high compared to current practice in FFT or MEM work.

Together the work of Gobien and Carpinella rigorously establish the basis for spectral estimation of a wide-sense-stationary all-pole Gaussian random process. Because this work is based on modern detection theory it contains all the "hooks" built into it to convert it from a learning machine to a tracking machine, an essential for attacking the real nonstationary problem. The all-pole assumption may be a mathematical fiction - it was the math that demanded it - but it should prove to be reasonable as the MEM work

makes the same assumption (author's opinion), and MEM work has had some excellent results in other fields. The Gaussian assumption is false, as usual, but it currently is the convenient way to force the math to stick to power measurements; it too has proven to be a useful and robust assumption.

The next step in this work depends on the talents of the next bright researcher to take up the work. So far the work has been very mathematical. A good experimentalist, with some feel for the math, might build a small unit (perhaps of order 5 or so), and after some lab testing have the nerve to attach it to a hydrophone. I hope so.

Theodore G. Birdsall

ACKNOWLEDGEMENTS

I would like to thank the members of my doctoral committee, for their time and energy in guiding this effort. I am especially indebted to Professor T. G. Birdsall, my committee chairman, who is a tribute to the Socratic ideal.

Appreciation is due to the Office of Naval Research Acoustic Program, which furnished support towards the production of this work.

Thanks are due also, to my wife for typing and editing of this work, and to Nell Perraut for her help with the typing of this manuscript.

The author's doctoral studies were made possible through the sponsorship of the Civilian Institutes Division of the United States Air Force Institute of Technology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I

STATEMENT OF PROBLEM AND INTRODUCTION

The problem to be considered is the spectral estimation of an Mth order, zero mean, stationary Gaussian process. The processes of interest are limited to those which possess an all pole spectral density of the form

$$S(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M} (\omega^2 + p_i^2)} \quad , \quad p_i \neq p_j . \tag{1.1}$$

The spectral estimation problem is cast as one in parameter estimation, which through the application of Bayes Rule provides probability density functions for the parameters in equation (1.1). This problem has been motivated by the work of Gobein (1974) who, using the results of Hajek (1962), showed the existence of a set of finite dimensional sufficient statistics relevant to this estimation problem. The solutions presented in this study are limited to the M=1 and M=2 cases, but the results can be extended to higher order finite dimensional processes. Because the M=2 problem exhibits the characteristics of higher dimensional processes, the emphasis of the simulation study is placed here. The main concern in the simulation study is the use of the simple measurements described by the sufficient statistics to develop the final estimate, and the circumstances under which this approach can be used.

Throughout, the discussion is expository and as such provides

a cursory background for the understanding of the spectral estimation problem. Chapter II introduces a more common approach to this problem: the periodogram estimate. These estimates are known to rely on empirical judgements, and form the basis for the approach which, in this study, is termed the "non-model" approach. The main issue in this study is placing the spectral estimation problem into a framework suitable for parameter estimation. This framework is established by the use of a time domain representation for the processes under investigation. Chapter III introduces the common models used for random processes, and the manner in which second order properties are related to model specifications. The latter part of Chapter III deals with the manner in which the observations model is obtained. This forms the basis for the estimation technique used in the simulation study.

The Maximum Entropy Method (MEM) is discussed in Chapter IV. It is one of the more successful approaches to spectral estimation based on the model approach. MEM uses the properties of an autoregressive process model to develop a series of approximations to the spectral density of the observations. These approximations are obtained by allowing the order of the model to be a free variable, and by providing a sequence of approximating models. The parameters in the models are obtained by a constrained optimization procedure, where the constraints on the range of the parameters are specified by the second order properties of the model. The MEM technique, in using a model approach, provides a contrast to the method developed in this study. This contrast is obtained through the prior assumptions that are made about the observations process so that the MEM approach applies.

Chapter V presents an introductory exposition of Bayesian

Parameter Estimation, and illustrates the basic techniques used in developing a post probability density function for the parameters of interest. The concepts of sufficient statistics and reproducing priors are introduced to facilitate the solution of the parameter estimation problem. These concepts are applied to the estimation of the parameters of an M=1, and M=2 autoregressive model and subsequently the parameters of the associated spectral representation. In this case the mean values of the post densities obtained are directly related to the model parameters specified by the Durbin-Levinson procedure discussed in Chapter III.

The continuous spectral estimation problem is addressed in Chapter VI. This problem is the one initially specified where estimates of the parameters p_i and σ^2, in equation (1.1), are sought. As before, these estimates are sought through the application of Bayes Rule. Initially, in addition to the assumption of an all pole structure, the order of the process (M) is assumed to be known. The order assumption is removed later in the investigation. Using the results of Chapter IV, an observations probability density function is obtained by modeling the observations as being densely sampled and representable by an autoregressive process which under dense sampling has the same second order properties as the process under investigation. In this densely sampled environment, the parameter σ^2 , equation (1.1), is shown to be learned faster than the others in the estimation problem. Finally, in Chapter VII, empirical verification of the estimation procedure is presented. The inclusion of indepenent quantizing noise is analyzed and empirically investigated. Through this vehicle, the trade-offs that are made between sampling

rate and resolution are illustrated.

CHAPTER II

CONVENTIONAL SPECTRAL ESTIMATION

2.1 Introduction

Power spectral estimation, as a prime tool in filter design, detection theory, and estimation of parameters of a linear system, finds its basis in the general study of estimation theory. The division of the techniques for spectral estimation can be based upon the prior assumptions that one makes about the process under investigation. The division proposed in this study is specified as "non-model" (conventional) and "model" techniques. The "model" techniques provide a way for the experimenter to inject varying degrees of prior knowledge about the source generating the observations, the measurement process, the range of parameters for the model, and the quality of the estimate desired. The "non-model" techniques use less prior information, and rely on a strong empirical basis. The major portion of this study is concerned with the methods that allow prior information to be used. To provide some degree of contrast, this chapter will present a cursory review of the common "non-model" techniques and their properties. A more complete description of these techniques can be found in Jenkins and Watts (1968), Blackman and Tukey (1959), and Brillinger (1975),which forms the foundation of the discussion to follow.

2.2 Correlation Estimates and the Periodogram

The conventional spectral estimates are traditionally based on the Fourier transform relationships between the correlation function and spectral density. These relationships are

$$S(\omega) = \sum_{n=-\infty}^{\infty} R(n)\ e^{-j\omega n} ,$$

and

$$R(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S(\omega)\ e^{j\omega n} d\omega ,$$

where $S(\cdot)$ is the spectral density and $R(n)$ is the correlation function.

The basis for the correlation estimator is the connection between ensemble and time averages. For a stationary process for which one desires to estimate the correlation, if it is possible to obtain a number of independent realizations (waveforms), then the average over these realizations will converge to the true value as the number of waveforms increases. In a practical situation, however, it is difficult or impossible to obtain many realizations, and hence it would be preferable to deal with a single realization. Under the conditions of ergodicity, when time averages and ensemble averages are equal to the same constant for almost-all realizations, it is possible to estimate the correlation from a single waveform.

In the process of forming these estimates, it is necessary to have some measure of their quality. The properties of an estimator that are used for this purpose are bias and variance. The bias of an estimator is defined as the true value of the parameter minus the expected value of the estimate. Formally, this definition is expressed as

$$\text{Bias } (B) \triangleq \gamma - E(\hat{\gamma}) \quad ,$$

where γ is the true value of the parameter and $\hat{\gamma}$ is the estimate. For a symmetrical probability density function, an unbiased estimator (B=0) gives an estimate whose value is at the center of the density function. The variance of the estimate indicates the width of the probability density function. A small variance implies that the probability density function is concentrated about its mean, which, if the estimate is unbiased, is the true value of the parameter. If as the number of observations increases, the bias and the variance tend toward zero, then the estimate is termed consistent. The ideal is then the existence of a consistent estimator.

One method of forming an estimate of the correlation function is from the relationship

$$C'_x(m) = \frac{1}{N - |m|} \sum_{n=0}^{N-|m|-1} x(n)x(n + m) \quad , \tag{2.1}$$

where

$$|m| < N \quad ,$$

N is the record length, m is the largest lag. It is evident that this estimate is unbiased since $E\big(x(n)x(n + m)\big) = R(m)$. The variance of this estimate is (Jenkins and Watts, 1968) approximately

$$\text{Var}(C'_x(m)) \simeq \frac{N}{\big[N - |m|\big]^2} \sum_{r=-\infty}^{\infty} R_x^2(r) + R_x(r + m)R_x(r - m) \tag{2.2}$$

Since this estimate is unbiased and

$$\lim_{N \to \infty} \{\text{Var}(C'_x(m))\} \to 0 \quad ,$$

it is a consistent estimate of $R_x(m)$. Another estimate for the auto-

correlation sequence is found from

$$C_x(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x(n)x(n+m). \tag{2.3}$$

Comparing this estimate to $C'_x(m)$, equation (2.1), it is apparent that

$$E(C_x(m)) = \frac{N - |m|}{N} C'_x(m)$$

$$= \frac{N - |m|}{N} R_x(m) . \tag{2.4}$$

Hence, $C_x(m)$ is asypmtotically unbiased, with variance

$$Var(C_x(m)) = \left(\frac{N - |m|}{N} \right)^2 Var(C'_x(m)) . \tag{2.5}$$

For N much larger than m this expression collapses to

$$Var(C_x(m)) = \frac{1}{N} \sum_{-\infty}^{\infty} \{R_x^2(r) + R_x(r+m)R_x(r+m)\} . \tag{2.6}$$

Both of these estimates are consistent under the condition that m is fixed and N grows large. At this point one might conclude that the Fourier transform of these estimates would provide a useful estimate of the spectral density, but this is not true.

An estimate of the power spectral density based on the Fourier transform of the estimate $C_x(m)$, equation (2.3), is called the periodogram. The periodogram is defined as

$$I_N(\omega) = \sum_{m=-(N-1)}^{(N-1)} C_x(m) e^{-j\omega m} , \tag{2.7}$$

and since the Fourier transform of the sequence x(n), $0 \leq n \leq N-1$, is

$$\underline{X}(e^{j\omega}) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \quad , \tag{2.8}$$

the periodogram can be shown to be also represented as

$$I_N(\omega) = \frac{1}{N} |\underline{X}(e^{j\omega})|^2 \quad . \tag{2.9}$$

The direct relationship between the input sequence and the periodogram, equation (2.9) accounts for the interest in such an estimate. To explore the feasibility of using such an estimate, the bias and variance are computed. The expected value of the periodogram is

$$E(I_N(\omega)) = \sum_{m=-(N-1)}^{(N-1)} E(C_x(m))e^{-j\omega m}$$

$$= \sum_{m=-(N-1)}^{(N-1)} \left(\frac{N-|m|}{N}\right) R_x(m)e^{-j\omega m} . \tag{2.10}$$

Because of the finite limits and the factor multiplying $R_x(m)$ in equation (2.10), the mean of the periodogram is not the transform of the correlation sequence and hence it is a biased estimate of the spectral density. An alternate view of the effects of the bias can be obtained from the complex convolution theorem which relates the transform of a product to a convolution integral. Specifically, this relationship is

$$K(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h(n)g(n)e^{-j\omega n} = \frac{1}{2\pi} \int_{-\pi}^{\pi} H(e^{j\omega})G(e^{j(\omega-\theta)})d\theta,$$

where $H(\cdot)$ and $G(\cdot)$ are the transforms of the time sequences $h(\cdot)$ and $g(\cdot)$. Applying this result to equation (2.10), the expected value of the periodogram is

$$E(I_N(\omega)) = \frac{1}{2\pi N} \int_{-\pi}^{\pi} S_x(\theta)G(e^{j(\omega-\theta)})d\theta \quad , \tag{2.11}$$

where $S_x(\theta)$ is the true spectral density and

$$G(e^{j\omega}) = \left[\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right]^2$$

is the transform of the factor weighting $R_x(m)$ in equation (2.10). From equation (2.11), the bias in the periodogram estimate shows up in a smearing of the spectral density $S_x(\theta)$ due to the convolution with the factor $G(\cdot)$. The variance of the periodogram is obtained by assuming that the sequence of observations is taken from a zero mean Gaussian process. Under this assumption, the variance of the periodogram (Jenkins and Watts, 1968) is

$$\mathrm{Var}(I_N(\omega)) = S_x^2(\omega)\left\{1 + \left[\frac{\sin(\omega N)}{N \sin \omega}\right]^2\right\}$$

$$= \begin{cases} \begin{cases} 2\, S_x^2(0) & \text{for } \omega = 0 \\ 2\, S_x^2(\pi) & \text{for } \omega = \pi \end{cases} \\ \\ S_x^2(\omega) \quad \text{else} \end{cases} , \tag{2.12}$$

where $S_x(\omega)$ is the spectral density of the observations. It is apparent that the variance of this estimate does not decrease as N becomes large. As a result of this, the estimate can be expected to have a rather erratic behavior. This behavior can be corrected by variance reduction techniques called smoothing.

2.3 Smoothing of the Periodogram

One approach to reducing the variance of the periodogram, proposed by Bartlett (1950), is to break the sequence of observations $x(n)$, $0 \leq n \leq N - 1$ into K segments of M samples so that $N = K \cdot M$. The K, M point periodograms are then averaged. An M point periodogram is formed as

$$I_M^i(\omega) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x^i(n)e^{-j\omega n}\right|^2 \qquad 1 \leq i \leq K \quad , \tag{2.13}$$

and then the estimate

$$\hat{S}_x(\omega) = \frac{1}{K}\sum_{i=1}^{K} I_M^i(\omega) \quad , \tag{2.14}$$

is formed. The expected value of this spectral estimate is

$$E\{\hat{S}_x(\omega)\} = \frac{1}{K}\sum_{i=1}^{K} E\{I_M^i(\omega)\}$$

$$= E\{I_M^i(\omega)\} \quad , \tag{2.15}$$

where from equation (2.10)

$$E\{I_M^i(\omega)\} = \sum_{-(M-1)}^{(M-1)}\left(\frac{M - |m|}{M}\right)R_x(m)e^{-j\omega m} \quad . \tag{2.16}$$

Again, based on the finite limits and the factor modifying $R_x(m)$, this estimate is also biased. Interpreting equation (2.16) as the transform of the product of the window function

$$w(m) = \begin{cases} \dfrac{M - |m|}{M} & |m| < M \\ 0 \quad \text{else} & \end{cases} \quad , \tag{2.17}$$

and the correlation function, an alternate form of equation (2.16) is

$$E\{I_M^i(\omega)\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(\theta)w(e^{j(\omega-\theta)})d\theta$$

$$= \frac{1}{2\pi M}\int S_x(\theta)\left(\frac{\sin((\omega-\theta)M/2)}{\sin((\omega-\theta)/2)}\right)^2 \quad . \tag{2.18}$$

The only difference between the mean of this estimate and that of the periodogram estimate, equation (2.11), is that N has been replaced by M. If the record length (N) is fixed and M=N/K, it is apparent that as the number (K) of M point periodograms increases the bias of the

Bartlett estimate increases. To see the value of this estimate, the variance is determined. Under the assumption that the K periodograms are independent, the variance is

$$\begin{aligned} \mathrm{Var}\{\hat{S}_x(\omega)\} &= \frac{1}{K}\{\mathrm{Var}\ I_M^i(\omega)\} \\ &\simeq \frac{1}{K} S_x^2(\omega) \left\{ 1 + \left[\frac{\sin(\omega M)}{M \sin \omega} \right]^2 \right\} . \end{aligned} \tag{2.19}$$

From this expression, it is apparent that as the number (K) of periodograms increases, the variance decreases. Thus, in using the Bartlett procedure, a trade off between variance and bias is made to obtain the spectral estimate.

An alternate method of obtaining the estimated spectral density is to apply a weight function (Blackman and Tukey, 1959) to the estimated correlation function, and thus transform the smoothed covariance function. The weighting function is chosen to weight the initial correlations heavily, and taper the weighting on latter covariances, so that the weights go to zero for large lags (m). The window functions are chosen to insure that the resulting spectral estimate will be a real, even, nonnegative function of frequency. Typically, the window function (g(m)) is chosen to be an even sequence of finite length, so that if

$$g(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\omega}) e^{j\omega m} d\omega , \tag{2.20}$$

for $-(M-1) < m < M-1$, and m = the number of lags such that $|m| < N$ (the record length), then

$$\hat{S}_x(\omega) = \sum_{-(M-1)}^{(M-1)} C_x(m) G(m) e^{-j\omega m} , \tag{2.21}$$

and the estimate will be real and even. It can be shown (Blackman and Tukey, 1959) that approximate relations for the mean and variance of this spectral estimate are

$$E(\hat{S}_x(\omega)) \simeq \frac{1}{2\pi} S_x(\omega) \cdot \int_{-\pi}^{\pi} G(e^{j\omega}) d\omega \ , \tag{2.22}$$

$$Var(\hat{S}_x(\omega)) \simeq \frac{1}{N} S_x^2(\omega) \cdot \int_{-\pi}^{\pi} G^2(e^{j\omega}) d\omega \ . \tag{2.23}$$

To obtain as asymptotically unbiased estimate, the window function is chosen so that

$$g(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(e^{j\omega}) d\omega = 1 \ . \tag{2.24}$$

The variance reduction afforded by this procedure is insured by satisfying the condition

$$\frac{1}{N} \sum_{m=-(M-1)}^{(M-1)} g^2(n) = \frac{1}{2\pi N} \int_{-\pi}^{\pi} G^2(e^{j\omega}) d\omega < 1 \ . \tag{2.25}$$

The distinction between this procedure and the periodogram estimates is the use of the correlation sequence, rather than the data sequence, in forming the estimate.

A method which combines windowing in the time domain, with averaging of periodograms is attributed to Welch (1970). The significant feature of the estimates produced by this approach is that, without regard to the window functions used, the spectral estimate will always satisfy the nonnegative requirement of a spectral density. This approach is well suited to the application of Fast Fourier Transform (FFT), thus providing a significant reduction in computational burden. This approach segments the data, as in equation (2.13), but a window function is applied to the data prior to forming

the periodogram. One begins by segmenting the data into K = N/M segments and then defining the K windowed periodograms

$$H_x^i(\omega) = \frac{1}{MV}\left|\sum_{n=0}^{M-1} x^i(n)g(n)e^{-j\omega n}\right|^2 , \tag{2.26}$$

for

$$i = 1,2,\ldots K ,$$

where

$$V = \frac{1}{M}\sum_{n=0}^{M-1} g^2(n) . \tag{2.27}$$

Now the spectral estimate is

$$\hat{S}_x(\omega) = \frac{1}{K}\sum_{i=1}^{K} H_x^i(\omega) . \tag{2.28}$$

The mean and variance for this estimate is (Welch, 1970)

$$E(\hat{S}_x(\omega)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(\theta)G(e^{j(\omega-\theta)})d\theta , \tag{2.29}$$

where

$$G(e^{j\omega}) = \frac{1}{MV}\left|\sum_{n=0}^{M-1} g(n)e^{-j\omega n}\right|^2 ,$$

and

$$\mathrm{Var}(\hat{S}_x(\omega)) \approx \frac{1}{K} S_x^2(\omega) . \tag{2.30}$$

As before, the variance reduction is achieved by introducing an arbitrary window function which, while providing a stable estimate, decreases the resolving power of the final estimate.

From the previous discussion, it is clear that little is assumed about the structure of the observations used to form the estimate. Most of the effort is placed on determining a suitable

window function to reduce the variability of the estimate. One of the major objections to the previously described estimates is just this point; that is, the window function is arbitrarily chosen, independent of any known structure within the observations. The goal of the "model" approach is to circumvent such arbitrary decisions and to provide a vehicle for injecting prior knowledge into the estimation process.

CHAPTER III

MODELS FOR RANDOM PROCESSES

3.1 Introduction

The framework for the estimation of a spectral density as a parameter estimation problem, is established by the Spectral Factorization and Representation Theorems (Astrom, 1970). These theorems indicate that a stationary random process with a rational spectral density can be characterized by operations on a white noise input. The Spectral Factorization Theorem guarantees that for a stationary random process with a rational spectral density $\phi(\cdot)$ there exists a rational transfer function $H(\cdot)$, of a stable system, such that the spectral density $\phi(\cdot)$ can be written as the frequency domain product $H(\cdot)H^*(\cdot)$. The Representation Theorem quarantees that when this stable system, with weighting function $h(\cdot)$, is excited by a process with orthogonal increments, the output process is stationary with spectral density $\phi(\cdot)$. With these two statements as guides, attention can be directed toward the methods of characterizing the random processes. Discrete random processes can be modeled in terms of a finite weighted sum of past values (autoregressive model), a weighted sum of uncorrelated random variables (moving average model) or a combination of both (autoregressive-moving average model). To insure the existence of such representations, specific conditions are placed on the character of the weights associated with the representations.

Continuous random processes have representations which are related to their discrete counter parts. The representation of a random process as the result of a convolution with white noise, is a counterpart to the moving average representation. This convolutional interpretation is justified through the use of a process with orthogonal increments and a bounded impulse response (Wong, 1971). Representing the random processes as the solution to a linear stochastic differential equation provides an alternative view which is the counter part of the auto-regressive model.

3.2 Autoregressive Models

One of the simplest methods of characterizing a discrete stationary random process is as a stochastic difference equation (auto-regressive model). In this characterization there is a sequence of random variables $\{y_t,\ t=1,2,\ldots\}$ which satisfies the equation

$$y_t + \beta_1 y_{t-1} + \ldots \beta_p y_{t-p} = e_t \ , \qquad (3.1)$$

where the sequence $\{e_t\}$ consists of uncorrelated random variables. The autoregressive (AR) model has several features which make it an attractive analytical model. The most attractive feature is the finite parameter description of the model and the relationship between the parameters and the second order properties. To illustrate this feature, consider the first order (p=1) AR model

$$y_t = -\beta_1 y_{t-1} + e_t \ , \qquad (3.2)$$

with $E(e_t^2) = \sigma^2$. The correlation function is determined from a few simple manipulations. Forming the product $(y_t y_{t-h})$ and taking its expectation, provides the result

$$E(y_t y_{t-h}) = -\beta_1 E(y_{t-1} y_{t-h}) + E(e_t y_{t-h}), \quad h > 0 \ . \qquad (3.3)$$

Since the model, equation (3.2), shows that the observation (y_t) is independent of future noise (e_t), the above result is further reduced to

$$R_h(h) = -\beta_1 R_y(h-1) \quad , h > 0 \quad . \tag{3.4}$$

Continuing on with this evaluation, the relationship for h=0 is

$$\begin{aligned} E(y_t y_t) &= -\beta_1 E(y_{t-1} y_t) + E(e_t y_t), \\ R(0) &= -\beta_1 R(1) + E\left(e_t(-\beta_1 y_{t-1} + e_t)\right) \\ R(0) &= -\beta_1 R(1) + \sigma^2 \quad . \end{aligned} \tag{3.5}$$

From equation (3.5) it is apparent that the zero lag correlation, R(0) is specified by the parameters σ^2 and β_1. This is illustrated by writing

$$R(1) = -\beta_1 R(0) \quad , \tag{3.6}$$

and then using this result, such that

$$\begin{aligned} R(0) &= -\beta R(1) + \sigma^2 \\ &= \beta_1^2 R(0) + \sigma^2 \\ R(0) &= \sigma^2/(1-\beta_1^2) \quad . \end{aligned} \tag{3.7}$$

With R(0) specified, the correlation at any lag (h) is found from

$$\begin{aligned} R(h) &= -\beta_1 R(h-1) \\ &= (-\beta_1)^h R(0) \\ &= (-\beta_1)^h \left[\sigma^2/(1-\beta_1^2)\right] \\ &= \left[\frac{R(1)}{R(0)}\right]^h \left[\frac{\sigma^2}{1-\beta_1^2}\right] \quad . \end{aligned} \tag{3.8}$$

Since $|\beta_1| < 1$, it is apparent that the correlation function is geometrically decreasing and as such is characteristic of a finite order AR process.

Another characteristic of an AR representation, which is particularly useful, is the Markov dependence of the observations.

This dependence most readily facilitates a probabilistic description of the observations (y_t). Under the condition that the forcing sequence e_t is zero mean and Gaussian, it becomes a simple matter to construct the density function $f(y_t/y_{t-1})$. Since this density is Gaussian, it is characterized by a mean and variance of the form

$$E(y_t/y_{t-1}) = -\beta y_{t-1} \tag{3.9}$$

$$Var(y_t/y_{t-1}) = \sigma^2 \quad . \tag{3.10}$$

With these factors determined, the one step conditional density of the observations is

$$f(y_t/y_{t-1}) = K \exp -\frac{(y_t + \beta y_{t-1})^2}{2\sigma^2} \tag{3.11}$$

The second order properties of higher order AR models are easily obtained from the p^{th} order representation

$$y_t + \beta_1 y_{t-1} + \ldots \beta_p y_{t-p} + e_t \quad , \tag{3.12}$$

where e_t is a white sequence of variance σ^2. Multiplying equation (3.12) by y_{t-h} and taking the expectation

$$E(y_t y_{t-h}) = -\beta_1 E(y_{t-1}y_{t-h}) \ldots - \beta_p E(y_{t-p}y_{t-h}) + e_t y_{t-h} \ ,$$

results in

$$R_y(h) = -\beta_1 R_y(h-1) - \beta_2 R_y(h-2) \ldots - \beta_p R_y(h-p), \ h > 0, \tag{3.13}$$

and

$$R_y(0) = -\beta_1 R_y(1) - \beta_2 R_y(2) \ldots - \beta_p R_y(p) + \sigma^2, h = 0 \ . \tag{3.14}$$

In general, for an AR process with roots of its characteristic equation

$$\lambda^p + \beta_1 \lambda^{p-1} + \ldots \beta_{p-1}\lambda + \beta_p = 0 \quad , \tag{3.15}$$

all less than one in magnitude, the correlation function satisfies equations (3.13) and (3.14). The converse of this statement has been the basis of several approaches to spectral estimation. That is, if the correlation function of a time series (y_t) satisfies

$$R(h) + \sum_{j=1}^{p} B_j R(h-j) \quad , \quad h = 1,2,\ldots p \tag{3.16}$$

and the roots of

$$\lambda^p + \sum_{j=1}^{p} B_j \lambda^{p-j} = 0 \tag{3.17}$$

have a magnitude of less than one, then the time series satisfies the difference equation

$$y_t + \beta_1 y_{t-1} \cdots + \beta_p y_{t-p} + e_t \tag{3.18}$$

It is apparent that if a set of correlations satisfies the requirements of equations (3.16) and (3.17), then the process can be modeled as autoregressive. Implicit in writing the difference equation for the model, is the evaluation of the set of coefficients $\{\beta_i\}$ from equation (3.16), which obviously requires a matrix inversion. This can be avoided through the use of the recursive procedure of Levinson (1947) and Durbin (1960). Essentially, the procedure states that the coefficients $\{\beta_{i,p}\}$ found from

$$\sum_{j=1}^{p} - \beta_{j,p} R(h-j) = R(h) \quad , \quad h = 1,2,\ldots p \quad , \tag{3.19}$$

for the p^{th} order system are related to the coefficients of the $p-1^{st}$ order system by

$$\beta_{j,p} = \beta_{j,p-1} + \beta_{p,p} \beta_{p-j,p-1} \quad , \quad j = 1, \ldots, p-1 \quad . \tag{3.20}$$

In addition, the noise scale factors are related by

$$\sigma_p^2 = \sigma_{p-1}^2 (1 - |\beta_{p,p}|^2) \quad . \tag{3.21}$$

An induction procedure (McDonough, 1974) can be used to demonstrate the validity of these statements. The recursive process starts with building a first order model (p=1) by using

$$\beta_{1,1} = -R_1/R_0 \quad , \tag{3.22}$$

and

$$\sigma_1^2 = R_0(1-|\beta_{1,1}|^2) \quad . \tag{3.23}$$

Using the relationships

$$\beta_{p,p} = \frac{-(\sum_{j=1}^{p-1} \beta_{j,p-1} R_{p-j} + R_p)}{(\sum_{j=1}^{p-1} \beta_{j,p-1} R_j + R_0)} \quad , \tag{3.24}$$

$$\beta_{j,p} = \beta_{j,p-1} + \beta_{p,p} \beta_{p,j,p-1} \quad , \quad j=1, \ \ldots \ ,p-1, \tag{3.25}$$

$$\sigma_p^2 = \sigma_{p-1}^2(1-|\beta_{pp}|^2) \quad , \tag{3.26}$$

the coefficient for the progressively higher order models, up to and including p can be determined.

If the time series being modeled, has been generated by an AR model of order p , then the order of the representation will be unique in the sense that coefficients of higher order models will collapse to those of order p . If the data has not been generated by an AR model, then the process of modeling becomes an approximation in which the order (p) results from some best fit.

3.3 Moving Average Representation

A moving average (MA) representation of a process is characterized by a weighted sum of uncorrelated random variables. Classification of this representation is in terms of the limit of the sum (finite, Infinite) and the range of the summation index (two-sided, one-sided). The most basic MA representation is one of finite order, which takes the form

$$x_t = \sum_{j=0}^{M} \alpha_j e_{t-j} \quad , \tag{3.27}$$

where e_{t-j} is a sequence of zero mean, uncorrelated random variables with variance σ^2. The correlation function of a MA representation is unique in that it is zero for all lags (h), $|h| > M$. That is, the correlation function is of the form

$$R_x(h) = E(x_t x_{t+h}) = \begin{cases} \sum_{j=0}^{M-|h|} \sigma^2 \alpha_j \alpha_{j+|h|} , & |h| \leq M \\ 0 , & h > M . \end{cases} \tag{3.28}$$

From the characteristics of the correlation functions associated with an MA representation, it could be conjectured that a finite order AR representation could be placed into an infinite order MA format, and maintain the same second order properties. This statement can be formalized (Anderson,1971) to provide the requirements under which such a transformation can be made. That is, assume that the AR representation is of the form

$$x_t + \alpha_1 x_{t-1} + \ldots \alpha_p x_{t-p} = e_t \tag{3.29}$$

with e_t a sequence of uncorrelated $(0,\sigma^2)$ random variables, and the roots of the characteristic equation

$$\lambda^p + \alpha_1 \lambda^{p-1} + \ldots \lambda_p = 0 , \tag{3.30}$$

all have the magnitude less than one. Under this condition, an MA representation of the form

$$x_t = \sum_{j=0}^{\infty} w_j e_{t-j} \tag{3.31}$$

can be found, where x_t is the mean square limit. The set of coefficients $\{w_j\}$ is found by the solution of the difference equation

$$w_j + \alpha_1 w_{j-1} + \ldots \alpha_p w_{j-p} = 0 \qquad j = p, p+1, \ldots . \tag{3.32}$$

subject to the boundary conditions

$$w_0 = 1$$

$$w_1 = -\alpha_1 w_0 \quad ,$$

$$w_2 = -\alpha_1 w_1 - \alpha_2 w_0 \quad ,$$

$$\vdots$$

$$w_{p-1} = -\alpha_1 w_{p-2} - \alpha_2 w_{p-3} \cdots \alpha_{p-1} w_0 \quad . \tag{3.33}$$

An example of converting an M=1 AR representation into its infinite oreder MA counterpart is presented next.

Considering the AR process (difference equation) that is modeled by

$$x_t - (1 - p\delta)x_{t-1} = e_t \quad . \tag{3.34}$$

Since the root (r) of the characteristic equation associated with this difference equation satisifes

$$|r| < 1 \tag{3.35}$$

then there exists an MA representation of the form

$$x_t = \sum_{j=0}^{\infty} w_j e_{t-j} \quad , \tag{3.36}$$

where the process x_t is the mean square limit of the sum, and the sequence of coefficients $\{w_j\}$ is found from the difference equation

$$w_j + \alpha_1 w_{j-1} = 0 \quad , \qquad j = 1,2,.... \tag{3.37}$$

subject to the boundary conditions

$$w_0 = 1$$

$$w_1 = -\alpha_1 w_0$$

with $\alpha_1 = -(1-p\delta)$. (3.38)

The MA representation that results is

$$x_t = \sum_{j=0}^{\infty} (-\alpha_1)^j e_{t-j} = \sum_{j=0}^{\infty} (1-p\delta)^j e_{t-j} \quad . \tag{3.39}$$

The duality between the MA and the AR representation can be continued in the finite order MA case; that is, corresponding to a finite order MA representation, there is an infinite order AR representation. The conditions under which this equivalence can be established, are given by Anderson (1971).

One other model of a discrete process that is possible, is the combined autoregressive-moving average model (ARMA). The ARMA representation is represented by

$$x_t + x_{t-1} \cdots + \alpha_p x_{t-p} = e_t + b_1 e_{t-1} + \ldots b_q e_{t-q} \quad , \tag{3.40}$$

where $\alpha_p \neq 0$, $b_q \neq 0$, and e_t is a sequence of uncorrelated $(0,\sigma^2)$ random variables. The ARMA representation can be converted into an infinite order MA or AR representation, under the condition that the roots of the characteristic equations have magnitude less that one (Anderson, 1971). Models of this form are used in practice to reduce the order of an approximating AR or MA representation of an observed process.

3.4 Spectral Representation of Discrete Random Processes

Any stationary random process can be characterized by its mean value and covariance (correlation) function. The use of the Fourier transform relationship on the covariance function, does not provide anything extra in characterization, but it does provide us with a different physical interpretation of the stationary random process.

The Fourier Integral relationship for discrete random processes

$$S(\omega) = \frac{1}{2\pi} \sum_{-\infty}^{\infty} R(h) e^{-ih\omega} \quad , \tag{3.41}$$

and

$$R(h) = \int_{-\pi}^{\pi} S(\omega) e^{-ih\omega} \quad , \tag{3.42}$$

are guaranteed to exist and are unique if the covariance R(h) is absolutely summable. Under this condition, the spectral density $S(\omega)$ is a continuous even nonnegative function in the frequency domain. By the uniqueness of the Fourier Transform, one would expect that the different characterization would have unique forms of the spectral density.

The spectral density of the MA representation

$$x_t = \sum_{j=0}^{p} \alpha_j e_{t-j} \quad , \tag{3.43}$$

where e_t is a sequence of uncorrelated $(0,\sigma^2)$ random variables, and the sequence α_j is absolutely summable, is

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{p} \alpha_j e^{-ij\omega} \right|^2 \quad . \tag{3.44}$$

An alternative form of the spectral representation can be found by rewriting the sum

$$\sum_{j=0}^{p} \alpha_j e^{-ij\omega} \quad , \quad \alpha_0 = 1 \tag{3.45}$$

as a polynomial in $e^{-i\omega}$, where

$$\sum_{j=0}^{p} \alpha_j e^{-ij\omega} = e^{-i\omega p} \sum_{j=0}^{p} \alpha_j e^{-i\omega(p-j)} = e^{-i\omega p} \prod_{j=1}^{p} (e^{i\omega} - m_j), \tag{3.46}$$

and the m_j are roots of the characteristic equation

$$m^p + \alpha_1 m^{p+1} + \ldots + \alpha_p = 0 \quad . \tag{3.47}$$

It now follows that the spectral representation is

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \left| \prod_{j=1}^{p} (e^{-i\omega} - m_j) \right|^2 , \tag{3.48}$$

which is an all zero spectral density.

The spectral density of an AR representaion can be found in a similar manner. With the process modeled as

$$\sum_{j=0}^{p} \alpha_j x_{t-j} = e_t \quad , \tag{3.49}$$

where $\alpha_0 = 1$, $\alpha_p \neq 0$, e_t a sequence of $(0,\sigma^2)$ uncorrelated random variables, and the roots of the characteristic equation have the magnitude less than one, the spectral density is given by

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{p} \alpha_j e^{-i\omega j} \right|^{-2} \quad , \text{ where } \alpha_0 = 1 \quad . \tag{3.50}$$

As before, rewriting the summation in terms of a polynomial in $e^{-i\omega}$, provides the alternative representation

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \left| \prod_{j=1}^{p} (e^{i\omega} - m_j) \right|^{-2} , \tag{3.51}$$

where the m_j are the roots of the characteristic equation. The representation in equation (3.51) is characterized as an all pole spectral density.

The spectral representation of an ARMA representation, equation (3.40), as might be conjectured, is a combination of the two previous results. The ARMA spectral density, which takes the form

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \frac{\left| \sum_{k=0}^{q} \beta_k e^{-i\omega k} \right|^2}{\left| \sum_{j=0}^{p} \alpha_j e^{-i\omega j} \right|^2}, \tag{3.52}$$

or

$$S_x(\omega) = \frac{\sigma^2}{2\pi} \frac{\left| \prod_{k=0}^{q} (e^{-i\omega} - r_k) \right|^2}{\left| \prod_{j=0}^{p} (e^{-i\omega} - m_j) \right|^2}, \tag{3.53}$$

where the r_k and m_j are the roots of the respective characteristic equations, is termed a rational spectral density.

It is apparent that these representations (AR, MA, ARMA) have the convenient feature that knowledge of the set of coefficients for the model provides the complete picture in terms of correlation properties and spectral behavior. The finite order AR representation provided a simple means of characterizing a process whose correlation function is of infinite extent. For an AR representation, only a finite set of coefficients are involved, and the probabilistic structure of the observations exhibited a Markov dependence. With the knowledge that a continuous periodic function can be approximated arbitrarily well by a trigonometric polynomial, it becomes apparent that a sampled continuous process may be modeled by an AR model if the order of the model is high enough (Jenkins and Watts, 1968). The simplicity of the AR models, coupled with approximation theory, has spurred some of the more successful approaches to spectral estimation.

To connect this to the continuous problem, where the Fourier relationships are

$$R(T) = \int_{-\infty}^{\infty} e^{-i\omega T} S(\omega)\, d\omega \quad , \tag{3.54}$$

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(T) e^{-i\omega T} dT \quad , \tag{3.55}$$

one has to find some approximation to $S(\omega)$, given a finite collection of observation $\{x_t,\ t=0,1,2,....\}$. Most often, the process $X(t)$ is assumed to be band limited. Then the sampling theorem can be applied to yield

$$R_k = R(k\Delta) = E(X_t X_{t+k\Delta}) \quad ,$$

$$R(T) = \sum_{k=-\infty}^{\infty} R_k \frac{\sin \pi(T - k\Delta)}{\pi(T - k\Delta)} \quad , \text{ for } \Delta \leq \Pi/\omega_H \ . \tag{3.56}$$

Then, as a result of this assumption

$$S(\omega) = \Delta \sum_{h=-\infty}^{\infty} R_k e^{-i\omega k\Delta} \quad , \tag{3.57}$$

where Δ is the sampling interval. Now, if K is sufficiently large, then the model

$$x_t + \sum_{j=1}^{K} \alpha_j x_{t-j} = e_t \quad , \tag{3.58}$$

represents an arbitrarily good approximation. Viewing the above relationship as an input/output description, one finds

$$S_x(\omega) = \left(|H(j\omega)|^2\right)^{-1} S_e(\omega) \quad , \tag{3.59}$$

with

$$H(j\omega) = \sum_{j=0}^{K} \alpha_j e^{-j\omega k\Delta} \quad , \quad \alpha_0 = 1 \ , \tag{3.60}$$

$$S_e(\omega) = \Delta \cdot \sigma_K^2 \quad , \tag{3.61}$$

$$S_x(\omega) = \frac{\Delta \cdot \sigma_K^2}{\left| \sum_{j=0}^{K} \alpha_j e^{-j\omega k\Delta} \right|^2} \ . \tag{3.62}$$

So again, with a known Δ and the values of the correlation at K lags, one can determine the K coefficients $\{\alpha_j\}_{j=0}^{K}$ and σ_K^2 from the Durbin-Levinson procedure, equations (3.22) and (3.26), and have a represenation for the process.

3.5 Representation of Continuous Time Processes

The continuous time analogue to a moving average representation would be a representation that is a weighted average of a white noise process. Forming this analogue leads to integrals of the form

$$\int f(t)n(t-T)dT \quad , \tag{3.63}$$

with $n(\cdot)$ a white noise process. Such an integral can be given meaning by rewriting it in the form (Doob, 1956)

$$\int f(t-s)dW(s) \quad , \tag{3.64}$$

where $W(s)$ is a process with orthogonal increments. In this way the process formed by the integrals

$$\int_{-\infty}^{t} f(t-T)dW(T) \quad , \tag{3.65}$$

and

$$\int_{-\infty}^{\infty} f(t-T)dW(T) \quad , \tag{3.66}$$

are analagous to the moving average representations

$$\sum_{k=0}^{\infty} \alpha_k \zeta(t-k) \quad , \tag{3.67}$$

$$\sum_{k=-\infty}^{\infty} \alpha_k \zeta(t-k) \quad . \tag{3.67}$$

The requirements (Doob, 1956) placed on the functions $f(t)$ for the integrals to be well defined are that

$$\int_0^\infty |f(t)|^2 \, dt < \infty \quad , \tag{3.68}$$

$$\int_{-\infty}^\infty |f(t)|^2 \, dt < \infty \quad . \tag{3.69}$$

Further, replacing the white noise $n(\cdot)$, in equation (3.63), with a process $W(\cdot)$ that has the property

$$E\{dW(t)dW(s)\} = K \cdot dt, \; t=s, \tag{3.70}$$

provides the justification for manipulation of integrals involving white noise. The significance of such integrals lie in the ability to represent a random process with a rational spectral density by such as integral. It can be shown (Astrom, 1970) that if the spectral density of a process can be represented as

$$S_y(\omega) = |H(\omega)|^2 \quad , \tag{3.71}$$

with all poles of $H(\omega)$ in LHP and $H^*(\omega)$ in RHP, then the output process can be modeled as

$$y(t) = \int_{-\infty}^t h(t-s)dW(s) = \int_0^\infty h(s)d\omega(t-s) \quad , \tag{3.72}$$

where $W(\cdot)$ is a process with orthogonal increments, and formally $dW(\cdot)$ is a white noise process.

Closely related to the stochastic integral representation, is the representation as a solution of a linear stochastic differential equation. This representation is written as

$$d\underline{x} = A(t)\underline{x}dt + d\underline{v} \quad , \tag{3.73}$$

where $\underline{v}(\cdot)$ is a wiener process with incremental covariance $R_1 dt$, and $x(t_0)$ is $N(0,R_0)$. The solution of the stochastic differential equation is a Gaussian process with zero mean and covariance

$$R(s,t) = \begin{cases} \phi(s,t)P(t), & s \geq t \\ P(s)\phi(t,s), & s \leq t \end{cases} \tag{3.74}$$

where $P(t)$ is the solution to

$$\dot{P} = AP + PA^T + R_1 \quad , \tag{3.75}$$

with

$$P(0) = R_0 \quad , \tag{3.76}$$

and

$$\dot{\phi}(t,t_0) = A(t)\phi(t,t_0) \quad . \tag{3.77}$$

If the coefficient matrix is constant, with eigenvalues that have negative real parts, then the solution is a stationary Gaussian process with

$$R(t-s) = \begin{cases} e^{A(t-s)} \cdot P_\infty & t \geq s \\ P_\infty \, e^{A(s-t)} & s \leq t \end{cases} \quad , \tag{3.78}$$

where P_∞ is the steady state solution to the differential equation (3.75). Now the process of interest can be interpreted as having a rational spectral density

$$S(\omega) = \int e^{-i\omega\tau} R(\tau) \, d\tau \quad , \tag{3.79}$$

or equivalently, the covariance function can be interpreted as

$$R(\tau) = \sum_i \alpha_i p_i(\tau) e^{-\lambda_i \tau} \quad , \tag{3.80}$$

where p_i are finite polynomials, and the real parts of λ_i are strictly positive.

The study of continuous time models can be approached via the theory of stochastic integrals, similar to the deterministic situation where the solution process is written as an integral equation, or by way of the limit of a difference equation. The difference equations of interest have the form

$$x^\Delta_{k+1} = f_\Delta(x^\Delta_k) + \sigma(x^\Delta_k)\zeta^\Delta_k \tag{3.81}$$

where $x^\Delta_k = x(k\Delta)$,

and Δ is the sampling interval. The forcing function is an independent increment process which is independent of x_0 and defined by

$$\zeta^{\Delta} = Z\left[(k+1)\Delta\right] - Z\left[k\Delta\right] ,$$

where $Z(\cdot)$ is a Wiener process. To show that such a representation provides a unique solution to the stochastic differential, which is approached in the limit as $\Delta \to 0$, requires a significant amount of detailed arguments. These arguments have been carried out by Kushner (1971). The important requirement on the difference equation, so that it will make sense in the limit, is the scaling of the parameters. Consider the linear difference equation

$$x_{k+1}^{\Delta} = (I + A_{\Delta})X_k^{\Delta} + \zeta_k^{\Delta} , \tag{3.82}$$

with

$$E\{\zeta_k^{\Delta}\ \zeta_k^{\Delta}\} = G_{\Delta} .$$

If one fixes the observation interval $T=N\Delta$ and intends to investigate the limiting behavior of the difference equation, the scaling on A_{Δ} and G_{Δ} must be of the form

$$G_{\Delta} = \Delta \cdot G ,$$

$$A_{\Delta} = \Delta \cdot A ,$$

so that

$$(I + A_{\Delta})^n = (I + \Delta A)^{t/\Delta} \xrightarrow[\substack{\Delta \to 0 \\ n \to \infty}]{} e^{At} . \tag{3.83}$$

In the general case, this scaling is reflected as

$$f_{\Delta}(x) = \Delta f(x) ,$$

$$\zeta_k^{\Delta} = Z\left((k+1)\Delta\right) - Z(k\Delta)$$

where the ζ_k^{Δ} are Gaussian with variance Δ. With this scaling, the general difference equation becomes

$$X^n_{k+1} = X^n_k + \Delta_n f(X^n_k) + \sigma(X^n_k)(\zeta^{\Delta n}_k), \tag{3.84}$$

or in summation form

$$X^n(k\Delta_n) = X_0 + \sum_{i=0}^{k-1} \Delta_n f\left(X^n(i\Delta_n)\right) + \sum_{i=0}^{k-1} \sigma\left(X^n(i\Delta_n)\right)\zeta^{\Delta n} \quad . \tag{3.85}$$

Since it can be proven that $X^n(\cdot)$ does converge to a continuous function, the first sum in equation (3.85) can be replaced by an integral,

$$\sum_{i=1}^{k-1} \Delta_n f\left(X^n(i\Delta_n)\right) \xrightarrow{\Delta \to 0} \int_0^t f^n(X^n_s)\, ds \quad . \tag{3.86}$$

The second sum in equation (3.85) has a limit which is a stochastic integral, the limit of an approximating sequence of non-anticipating step functions. The sum can be replaced in the limit by

$$\sum_{i=0}^{k-1} \sigma\left(X^n(i\Delta^n)\right)\left(Z\left((i+1)\Delta_n\right) - Z(i\Delta_n)\right) \xrightarrow{\Delta \to 0} \int_0^t \sigma^n(X^n_s)\, dZ_s \tag{3.87}$$

With these limiting forms defined, the difference equation in its summation form is

$$X^n_t = X_0 + \int_0^t f^n(X^n_s)ds + \int_0^t \sigma^n(X^n_s)dZ_s \quad . \tag{3.88}$$

It is shown (Kushner, 1971) that this equation holds with the limits X_t, $f(X_s)$, $\sigma(X_s)$ substituted, and the solution is unique.

An alternate method of obtaining a difference equation is through sampling the differential equation (Astrom, 1970)

$$d\underline{x} = A\underline{x}\, dt + d\underline{v} \quad , \tag{3.89}$$

where $\underline{v}(t)$ is a Wiener process with incremental covariance $R_1 dt$. Integration of equation (3.89), gives the difference equation

$$\underline{X}(t_{n+1}) = \phi(t_{n+1},t_n)\underline{X}(t_n) + \underline{\tilde{v}}(t_n) \quad , \tag{3.90}$$

where $\phi(\cdot\cdot)$ is defined by

$$\dot{\phi}(t,t_i) = A\phi(t,t_i) \quad , \qquad t_i \leq t \leq t_{i+1} \tag{3.91}$$

$$\phi(t_i,t_i) = I \tag{3.92}$$

and $\underline{\tilde{v}}(t_n)$ is defined by

$$\underline{\tilde{v}}(t_n) = \int_{t_n}^{t_{n+1}} A\phi(t_{n+1},t)\, d\underline{v}(t) \quad . \tag{3.92}$$

The properties of $\underline{\tilde{v}}(\cdot)$ are determined to be

$$E\left(\underline{\tilde{v}}(t_n)\right) = E\left\{\int_{t_n}^{t_{n+1}} \phi(t_{n+1},t)\, d\underline{v}(t)\right\} = 0 \quad , \tag{3.93}$$

$$E\left(\underline{\tilde{v}}(t_i)\underline{\tilde{v}}(t_j)\right) = E\left\{\int_{t_i}^{t_{i+1}}\int_{t_j}^{t_{j+1}} \phi(t_{i+1},t)d\underline{v}(t)d\underline{v}^T(s)\phi^T(t_{j+1},s)\right\}$$

$$= \begin{cases} \displaystyle\int_{t_i}^{t_{i+1}} \phi(t_{i+1},t)R_1\phi^T(t_{i+1},t)\, dt & t_i = t_j \\ 0 & t_i = t_j. \end{cases} \tag{3.94}$$

Now, at the sampling instants t_i, the outputs of the differential equation (3.89), and the difference equation (3.90), are identical.

3.6 Difference Equation For Sampled M=2 Process

To see the application of these ideas, assume that one desires to model the process that has as output correlation of the form

$$R_x(\tau) = \sigma^2 \sum_{i=1}^{M} \left\{ \frac{e^{-p_i|\tau|}}{2p_i \prod_{k=1}^{M} (p_k^2 - p_i^2)} \right\}, \quad p_i \neq p_k \; . \tag{3.95}$$

and spectral density of the form

$$S(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M} (\omega^2 + p_i^2)} \; . \tag{3.96}$$

As is well known, this behavior can be obtained by exciting an M dimensional linear dynamic system with transfer function

$$H(s) = \frac{\sigma}{\prod_{i=1}^{M} (s + p_i)}, \tag{3.97}$$

by white Gaussian noise.

Letting M=2 in the correlation equation, the system that generates the desired output statistics can be represented as

$$d\underline{x} = \begin{bmatrix} 0 & 1 \\ -p_1 p_2 & -(p_1 + p_2) \end{bmatrix} \underline{x}dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dv \quad , \tag{3.98}$$

$$y = [1, 0]\, \underline{x} \quad , \tag{3.99}$$

with

$$E\left(dv(t)\ dv(s)\right) = \sigma^2 dt$$

For this representation, the transition matrix $\phi(t,t_0)$ is

$$\frac{1}{p_2 - p_1} \begin{bmatrix} p_2 e^{-p_1(t-t_0)} - p_1 e^{-p_2(t-t_0)} & e^{-p_1(t-t_0)} - e^{-p_2(t-t_0)} \\ -p_1 p_2 \left(e^{-p_1(t-t_0)} - e^{-p_2(t-t_0)}\right) & p_2 e^{-p_2(t-t_0)} - p_1 e^{-p_1(t-t_0)} \end{bmatrix} \tag{3.100}$$

The output statistics $R_y(\tau)$ are found from

$$\begin{aligned} R_y(t-t_0) &= \phi[1,0]\ (t-t_0)\ R_x(t_0)\,[1,0]^T \\ &= [1,0] R_x(t-t_0)[1,0]^T \quad . \end{aligned} \tag{3.101}$$

For a stationary output, the correlation $R_{\underline{x}}(t_0=0)$, is the steady state solution to the equation

$$\dot{P}(t) = 0 = AP + PA^T + R$$

where

A is the coefficient matrix of equation (3.98),

P is the state covariance matrix,

R is the noise covariance.

The steady state solution to this equation (P_∞) is

$$P_\infty = R_{\underline{x}}(0) = \begin{bmatrix} \frac{\sigma^2}{2(p_1 + p_2)} & 0 \\ 0 & \frac{\sigma^2}{2p_1p_2(p_2 + p_1)} \end{bmatrix} . \qquad (3.102)$$

Now the output covariance function is

$$R_y(\tau) = [\underline{1},0]\phi(\tau)\, R_{\underline{x}}(0)\, [1,0]^T$$

$$= \frac{\sigma^2}{2p_1p_2(p_2^2 - p_1^2)} \left(p_2e^{-p_1|\tau|} - p_1e^{-p_2|\tau|} \right) . \qquad (3.103)$$

To obtain a discrete version of this model, equation (3.98), that will be equivalent at the sampling instants, the transition matrix

$$\phi(t,0) = \frac{1}{p_2 - p_1} \begin{bmatrix} p_2e^{-p_1t} - p_1e^{-p_2t} & e^{-p_1t} - e^{-p_2t} \\ -p_1p_2(e^{-p_1t} - e^{-p_2t}) & p_2e^{-p_2t} - p_1e^{-p_1t} \end{bmatrix} . \qquad (3.104)$$

will be sampled so that the constituent parts of the difference equation (3.90), repeated below for convenience,

$$\underline{x}(t_{i+1}) = \phi(t_{i+1},t_i)\, \underline{x}(t_i) + \tilde{\underline{v}}(t_i) , \qquad (3.105)$$

can be found. The first computation involves the determination of the covariance matrix for v(t), using equation (3.94), and results in

$$E\left(\tilde{\underline{v}}(t_i)\, \tilde{\underline{v}}(t_i)\right) = \int_{t_i}^{t_{i+1}} \phi(t_{i+1},s)\, R_v\phi^T(t_{i+1},s)\, ds$$

$$= \begin{bmatrix} R\tilde{v}_1\tilde{v}_1 & R\tilde{v}_1\tilde{v}_2 \\ R\tilde{v}_2\tilde{v}_1 & R\tilde{v}_2\tilde{v}_2 \end{bmatrix} , \qquad (3.106)$$

where the elements of equation (3.106) are

$$R\tilde{v}_1\tilde{v}_1 = \frac{\sigma^2}{(p_2 - p_1)^2}(p_1p_2)^2\left[\frac{1}{2p_1}(1-e^{-p_1\Delta}) - \frac{2}{p_1+p_2}\left(1-e^{-(p_1+p_2)\Delta}\right)\frac{1}{2p_2}(1-e^{-2p_2\Delta})\right], \quad (3.107)$$

$$R\tilde{v}_2\tilde{v}_1 = R\tilde{v}_1\tilde{v}_2 = \frac{\sigma^2}{(p_2 - p_1)^2}\left[\frac{p^2}{p_1+p_2}\left(1-e^{-(p_1+p_2)\Delta}\right) - \frac{p_1}{2p_1}(1-e^{-2p_1\Delta}) - \frac{p_2}{2p_2}(1-e^{-2p_1\Delta}) + \frac{p_1}{p_1+p_2}\left(1-e^{-(p_1+p_2)\Delta}\right)\right] . \quad (3.108)$$

and

$$R\tilde{v}_2\tilde{v}_2 = \frac{\sigma^2}{(p_2 - p_1)^2}\left[\frac{p_2}{2}(1-e^{-2p_2\Delta}) - \frac{2p_2p_1}{p_1+p_2}(1-e^{-(p_1+p_2)\Delta}) + \frac{p_1}{2}(1-e^{-2p_1\Delta})\right] . \quad (3.109)$$

Under fast sampling conditions $\left((p_1 + p_2)\Delta \ll 1\right)$ the quantities $R\tilde{v}_1\tilde{v}_1$, $R\tilde{v}_1\tilde{v}_2$ and $R\tilde{v}_2\tilde{v}_2$ reduce to

$$\Delta^{R\tilde{v}_2\tilde{v}_2} = \frac{\sigma^2}{(p_2 - p_1)^2}(p_2^2 - 2p_2p_1 + p_1^2)\Delta = \sigma^2\Delta , \quad (3.110)$$

$$\Delta^{R\tilde{v}_1\tilde{v}_1} = \frac{\sigma^2}{(p_2^2 - p_1^2)}(p_1p_2)^2(\Delta - 2\Delta + \Delta) = 0 , \quad (3.111)$$

$$\Delta^{R\tilde{v}_1\tilde{v}_2} = 0 . \quad (3.112)$$

So, it can be concluded that under these sample rate conditions the covariance matrix is

$$\Delta^{\underline{R}_v} = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2\Delta \end{bmatrix} .$$

Under the same conditions, the state transition matrix for the sampled

system becomes

$$\phi(\Delta) = \begin{bmatrix} 1 & \Delta \\ -p_1 p_2 \Delta & 1-(p_2+ p_1)\Delta \end{bmatrix} . \qquad (3.113)$$

Using equations (3.113) and (3.112), in equation (3.105), the difference equation being sought is

$$\underline{x}(t_{i+1}) = \left\{ I + \begin{bmatrix} 0 & 1 \\ -p_1 p_2 & -(p_2+ p_1) \end{bmatrix} \cdot \Delta \right\} \underline{x}(t_i) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} {}_{\Delta}\tilde{v}(t_i) , \qquad (3.114)$$

$$y(t_i) = [1,0]\, \underline{x}(t_i) ,$$

where I is the identity matrix and $E\left({}_{\Delta}\tilde{v}(t_i)\, {}_{\Delta}\tilde{v}(t_i)\right)=\sigma^2\Delta$.

The interesting consequence of this result is that it agrees with the assumptions on scaling required by equation (3.83), and further, it gives as estimate on the size of the sampling interval (Δ) for the difference equation to hold. Expanding equation (3.114) results in

$$x_1(t_{i+1}) = x_1(t_i) + \Delta \left\{ [1-\Delta(p_1+ p_2)]\, x_2(t_{i-1}) - \Delta p_1 p_2\, x_1(t_{i-1}) + {}_{\Delta}\tilde{v}(t_{i-1}) \right\} . \qquad (3.115)$$

Using the relationship

$$\Delta x_2(t_{i-1}) = x_1(t_i) - x_1(t_{i-1}) \qquad (3.116)$$

in equation (3.115), the result is

$$x_1(t_{i+1}) = \left(2-\Delta(p_1+ p_2)\right) x_1(t_i) - \left(1-\Delta(p_1+ p_2) + \Delta^2 p_1 p_2\right) x_1(t_{i-1}) + \Delta\, {}_{\Delta}\tilde{v}(t_{i-1}) ,$$

and since $y(t_i) = x_1(t_i)$, the final result is

$$y(t_{i+1}) = \left(2-\Delta(p_1+ p_2)\right) y(t_i) - \left(1-\Delta(p_1+ p_2) + \Delta^2 p_1 p_2\right) y(t_{i-1}) + \Delta \cdot {}_{\Delta}\tilde{v}(t_{i-1}) .$$

stability

$$|\alpha_{22}| = |(1-\Delta(p_1+ p_2) + \Delta^2(p_1p_2))|<1 ,$$

$$|\alpha_{22}| = |2-\Delta(p_1+ p_2)|<2 ,$$

under the fast sampling assumption

$$\Delta p_1 < 1 \quad , \Delta p_2 < 1 \quad , \quad \Delta(p_1+ p_2) < 1 \quad .$$

The same arguments and manipulations can be carried out for other finite dimensional structures. The model structure for the first order (M=1) model is determined in Appendix A. The most important aspect of the difference equation model is the output correlation function obtained. That is, there must be some assurance that it is sufficiently close to the correlation function of the processes under consideration. The next section demonstrates that the limiting form of the model does possess the desired properties.

3.7 Output Correlation Function of Sampled Differential Equation

The discrete model that resulted from sampling the differential equation (equation 3.82), is

$$\underline{X}^{\Delta}_{k+1} = (I + \Delta\cdot A)\ \underline{X}^{\Delta}_{k} + \underline{\zeta}^{\Delta}_{k} \tag{3.118}$$

with

$$E(\underline{\zeta}^{\Delta}_{k}\ \underline{\zeta}^{\Delta T}_{k}) = \Delta\cdot R_1 \quad , \quad E(\underline{X}_0\underline{X}^{T}_0) = R_0 \quad , \text{ and } E(\underline{\zeta}^{\Delta}_{k}\underline{X}^{T}_0) = 0 .$$

To see the limiting form of the covariance generated by this model, calculation of the state covariance of equation (3.118) yields

$$\text{Cov}\ \underline{X}^{\Delta}_{k} = \sum_{i=1}^{k} (I + \Delta\cdot A)^{k-i}(\Delta\cdot R_1)(I + \Delta\cdot A^{T})^{k-i} + (I + \Delta A)^{k}R_0(I + \Delta A^{T})^{k} \quad , \tag{3.119}$$

Now, fix the observation interval ($t=k\Delta$), so that as $k \to \infty$

$$(I + \Delta A)^{k} = (I + \Delta A)^{t/\Delta} \longrightarrow e^{At} .$$

With this condition established, the summation in equation (3.119) can be shown to be (Kushner, 1971) a Reimann sum approximation and

$$\lim_{\Delta\to 0} \text{Cov } \underline{X}_k^{\Delta} = \int_0^t e^{A(t-s)} R_1 e^{A^T(t-s)} ds + e^{At} R_0 e^{A^T t}. \qquad (3.120)$$

To determine the correlation function, fix the lag interval $\tau = m\Delta$, for $\tau \geq 0$, then

$$\lim_{\substack{m\to\infty \\ n\to\infty \\ \Delta\to 0}} E(\underline{X}_k^{\Delta}\, \underline{X}_{k+m}^{T\Delta}) \longrightarrow \int_0^t e^{A(t-s)} R_1 e^{A^T(t-s)} ds + e^{At} R_0 e^{A^T t}\, e^{A^T \tau}. \qquad (3.121)$$

If A has eigenvalues with negative real parts, then allowing the observation interval $t \to \infty$, provides the result

$$E(\underline{X}_k^{\Delta}\, \underline{X}_{k+m}^{T\Delta}) \longrightarrow \int_0^{\infty} e^{At} R_1 e^{A^T t}\, dt\; e^{A^T \tau} \quad . \qquad (3.122)$$

The same basic manipulations can be carried out for $\tau \leq 0$. Considering the linear, constant coefficient, vector differential equation

$$d\underline{X} = A\underline{X}\, dt + F\, d\underline{Z} \quad , \qquad (3.123)$$

where $\underline{Z}$ is a vector of independent Wiener processes, the solution is

$$\underline{X}(t) = e^{At}\underline{X}_0 + \int_0^t e^{A(t-s)} F\, d\underline{Z}(s) \quad . \qquad (3.162)$$

The statistical properties of the solution are

$$E(\underline{X}_t) = e^{At}\, E(\underline{X}_0) \quad , \qquad (3.124)$$

and for $\tau \geq 0$

$$R_{\underline{X}}(\tau) = \left(e^{At} \text{Cov } X_0\, e^{A^T t} + \int_0 e^{A(t-s)} FF^T e^{A^T(t-s)}\right) e^{A^T \tau}. \qquad (3.125)$$

The result obtained in equation (3.125) is the same as that obtained from the limit of the difference equation (3.118), if the substitutions Cov $X_0 = R_0$ and $R_1 = FF^T$, are made. Thus it can be concluded that taking limits of covariances corresponding to equation (3.118), the

covariances which are of interest can be generated by this model.

The approach to finding a difference equation, as discussed in this section, will be the basis for the solution of the estimation problem to be treated in Chapter VI. That is, the observations will be modeled as being generated from this difference equation, under the assumption that the sampling rate $(1/\Delta)$ is fast.Then the limit $(\Delta \to 0)$ will be taken to obtain the model with the desired properties. The next chapter provides a contrast to the problem of interest, in that it discusses an estimation technique that is based on observations generated by an autoregressive model, of arbitrary order without constraints on the sampling rate.

CHAPTER IV

THE MAXIMUM ENTROPY METHOD

This chapter introduces a technique for spectral estimation which is based on modeling the process under investigation as an autoregressive process. The approach taken in this technique relies on the correlation properties of this model, and the fact that as the order of this representation grows, an autoregressive model will arbitrarily well approximate a sampled continuous process. The assumptions placed on the application of the model will provide a contrast to the problem under investigation in this study. The MEM technique assumes no prior knowledge of the structure of the random process that has been sampled.

Burg (1967) introduced the MEM technique to circumvent the method used in conventional spectrum estimation techniques for handling data beyond the last available lag. Conventional techniques use window functions that have the common feature of multiplying unmeasured correlations by zero. In particular, Burg asserted that the direct average lag product procedure

$$R(i) = \frac{1}{N} \sum_{j=1}^{N-i} x_j x_{j+i} \tag{4.1}$$

implicitly assumes that the rest of the data is zero, and the use of the Fast Fourier Transforms treats the data as if it repeats itself periodically.

The MEM technique retains all known correlations without modification, and estimates the values of the unknown correlation through the use of the properties of the correlation matrix. That is, the set of values {R(0).....R(N)}, are the first N + 1 values of a correlation function if the matrix

$$\Gamma_{N+1} = \begin{bmatrix} R(0) & R(1) & . & . & . & R(N) \\ \vdots & & & & & \vdots \\ R(N) & . & . & . & & R(0) \end{bmatrix} \tag{4.2}$$

is positive definite, and hence its determinant is non-negative. If the first N values {R(0) . . . .R(N-1)}are known, then R(N) must have a value which makes the determinant non-negative, to insure that Γ_{N+1} is a correlation matrix. Burg (1967) indicated that this is equivalent to asking for the spectrum that maximizes the entropy of a Gaussian process , given the first N values of the correlation function. This statement is explicitly justified by Smylie, et al (1973), who demonstrate this equivalence through a theorem due to Grenander and Szego (1958). The demonstration begins with realizing that the entropy of a zero mean stationary Gaussian process is

$$H = \frac{1}{2} \log \{\text{Det } \Gamma_N\} \tag{4.3}$$

and under this condition, the matrix Γ_N is Toeplitz (equidiagonal). Since Γ_N is Toeplitz

$$\text{Det } \Gamma_N = \lambda_0\lambda_1\lambda_2 \cdot \ldots \cdot \lambda_{N-1} , \tag{4.4}$$

where λ_i are the eigenvalues of Γ_N,

Szego's theorem establishes the relationship between the determinant of Γ_N and the spectral density as

$$\lim_{N\to\infty} \frac{\log \lambda_0 + \log \lambda_1 \ldots \log \lambda_{N-1}}{N} = \lim_{N\to\infty} \log \left[(\lambda_0 \ldots \lambda_{N-1})^{1/N}\right]$$

$$= \lim_{N\to\infty} \log (\mathrm{Det}\, \Gamma_N)^{1/N}$$

$$= \Delta \int_{-1/2\Delta}^{1/2\Delta} \log(s(f)/\Delta)\, df \qquad (4.5)$$

Maximizing the determinant of the correlation matrix (Γ_N), subject to the constraints of the known correlations, is equivalent to maximizing the integral of the log-spectral density.

Edwards and Fitelson (1973) found the MEM spectrum as the solution of a variational problem. The variational problem is found by maximizing the integral

$$\int \log S(f)\, df \,, \qquad (4.6)$$

subject to the constraints imposed by the known correlations. A more intuitive approach, dealing directly with the correlation matrix shows the MEM technique to be equivalent to seeking a least square autoregressive model (Van den Bos, 1971). This equivalence is demonstrated by assuming that the first N values of the correlation function are known. The MEM procedure requires that the unknown correlation values R(N),R(N+1).... are chosen so that the entropy is maximized at each stage. The value of R(N) is determined by maximizing

$$\mathrm{Det}\ \Gamma_{N+1} = \mathrm{Det}\begin{bmatrix} R(0) & R(1) & \cdots & R(N) \\ \vdots & & & \vdots \\ R(N) & \cdots & & R(0) \end{bmatrix} \qquad (4.7)$$

with respect to R(N). The determinant of Γ_{N+1} is quadratic in R(N), and since the matrix Γ_{N+1} is positive definite, a single maximum exists. Applying the product rule of differentiation, the solution

to the maximization of equation (4.7), over R(N) is found to be the value of R(N) that satisfies

$$\text{Det}\begin{bmatrix} R(1) & R(0) & \cdots & R(N-2) \\ R(2) & & & \vdots \\ \vdots & & & \vdots \\ R(N) & \cdots & & R(1) \end{bmatrix} = 0 \quad . \tag{4.8}$$

Once R(N) has been found, R(N+1) can be determined by the equation

$$\text{Det}\begin{bmatrix} R(1) & R(0) & R(1) & \cdots & R(N-1) \\ R(2) & & & & \vdots \\ \vdots & & & & \vdots \\ R(N+1) & R(N) & \cdots & & R(1) \end{bmatrix} = 0 \quad . \tag{4.9}$$

Now consider the autoregressive model (of order N-1)

$$x(n) + \beta_1 x(n-1) + \cdots\cdots \beta_{N-1} x\left[n-(N-1)\right] = e^1(n)$$

In chapter III, it was demonstrated that the correlation function of a process generated by this model satisfied the equation

$$\begin{bmatrix} R(1) \\ \vdots \\ R(N-1) \end{bmatrix} + \begin{bmatrix} R(0) & \cdots & R(N-2) \\ \vdots & & \vdots \\ R(N-2) & \cdots & R(0) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_{N-1} \end{bmatrix} = 0. \tag{4.10}$$

If the first N values, {R(0) ... R(N-1)}, of any correlation function are available, substitution into equation (4.10) will yield the set of coefficients $\{\beta_1, \ldots \beta_{N-1}\}$. The value of R(N) for a model with these coefficients is uniquely determined by

$$\begin{bmatrix} R(1) & R(0) & \cdots & R(N-2) \\ \vdots & & & \vdots \\ R(N) & R(N-1) & \cdots & R(1) \end{bmatrix} \begin{bmatrix} 1 \\ \beta_1 \\ \vdots \\ \beta_{N-1} \end{bmatrix} = 0. \tag{4.11}$$

As a result, R(N) is selected so that

$$\text{Det}\begin{bmatrix} R(1) & R(0) & \cdots & R(N-2) \\ \vdots & & & \vdots \\ R(N) & R(N-1) & \cdots & R(1) \end{bmatrix} = 0. \tag{4.12}$$

This is precisely the same condition as required in equation (4.8), to find R(N) if the entropy is maximized. It can be concluded that the MEM technique is equivalent to finding an autoregressive model, given the correlation data.

In practice, only a finite record of the data is available, and the correlation function cannot be determined exactly, so the evaluation of the model coefficients proceeds along a different line. A method of determining the coefficients in this situation was proposed by Burg (1968). This method estimates the parameters $\{\beta_i\}$ in such a way that no assumptions outside the record length are made, and the positive definite character of the correlation matrix is maintained. To begin, an error function is defined as

$$e(n) = x(n) - \hat{x}(n) , \tag{4.13}$$

where $\hat{x}(n)$ is the predicted observation determined from past samples to be

$$\hat{x}(n) = -\sum_{i=1}^{M} \beta_i x(n-i) , \tag{4.14}$$

and M is the order of the model. Minimization of the squared error proceeds recursively for M=1... N-1, where N-1 is the record length. The minimization however, is performed so that the correlation matrix retains its positive definite structure. The error minimizing estimate for β_{11} (coefficient for M=1 model) is found from

$$\min_{\beta_{11}} P_1 = \min_{\beta_{11}} \frac{1}{2N-1} \sum_{n=1}^{N-1} e^2(n) \tag{4.15}$$

$$\min_{\beta_{11}} P_1 = \min_{\beta_{11}} \frac{1}{2N-1} \sum_{n=1}^{N-1} \{(x_{n+1}+\beta_{11}x_n)^2 + (x_n+\beta_{11}x_{n-1})^2\}.$$

This estimate is

$$\beta_{11} = - \frac{2(x_1x_2 + x_2x_3 \ldots x_Nx_{N-1})}{x_1^2 + 2x_2^2 + 2x_3^2 + \ldots 2x_{N-1}^2 + x_N^2} . \qquad (4.16)$$

The next step in the procedure is to minimize

$$\min_{\beta_{i,2}} P_2 = \frac{1}{2(N-2)} \sum_{n=1}^{N-2} \{(x_{n-2}+\beta_{12}x_{n-1}+\beta_{22}x_n)^2 + (x_n+\beta_{12}x_{n-1}+ \\ + (x_n+\beta_{12}x_{n-1}+\beta_{22}x_{n-2})^2\}, \qquad (4.17)$$

so that the coefficients for the M=2 model can be determined. The coefficients β_{12} and β_{22}, that minimize P_2, would normally be the estimated ones. However, because of the structure of the correlation matrix, the coefficients β_{11} and β_{12} are related by

$$\beta_{12} = \beta_{11}(1 + \beta_{22}) . \qquad (4.18)$$

This relationship follows from the Durbin - Levinson procedure, (Chapter III). As a consequence of equation (4.18), P_2 can only be minimized with respect to β_{22}. Proceeding in this manner, the coefficients for higher order models are determined, subject to the constraint

$$\beta_{i,M} = \beta_{i,M-1} + \beta_{MM}\beta_{M-i,M-1}, \quad i=1, \ldots M-1. \qquad (4.19)$$

This process can continue until the order of the model (M) is equal to the length of the record (N-1) or until some terminating condition is satisfied. An appropriate termination condition is the stage-to-stage stationarity of the power measure P_M. This criterion comes from the relationship

$$P_M = P_{M-1}(1 - \beta_{MM}^2) \qquad (4.20)$$

which relates the error power at stage M-1 to the error power at stage

M. For a stable difference equation ($|\beta_{MM}| < 1$), this power is monotone decreasing and hence a decision concerning the order of the model can be made based on the change in error power from stage M. The details of an iterative procedure for the coefficient determination are implemented in a form suitable for a computer program, in an article by Anderson (1974).

Once the estimated coefficients $\{\beta_{i,M}\}$ and the power P_M have been determined, the correlation function and the spectral density can immediately be specified. The MEM spectral density is

$$S(f) = \frac{P_M \Delta t}{\left|1 + \sum_{n=1}^{M} \beta_{n,M} e^{-j2\pi nf\Delta t}\right|^2}, \tag{4.21}$$

where M is the order of the model, and Δt is the sampling interval.

Estimates of the correlation function are made based on the properties of the autoregressive models specified by the coefficients. The correlation estimation procedure begins by forming

$$R(0) = \frac{1}{N} \sum_{i=1}^{N} x_i^2, \tag{4.22}$$

and then proceeds recursively using the estimated coefficients $\{\beta_{i,j}\}$ to form

$$R(T) = -\sum_{k=1}^{T} R(T-k)\beta_{k,T} \tag{4.23}$$

for T=1,2.....M, where M is the order of the model. Once the order of the model (M) has been decided on through the use of some appropriate criterion, the unknown correlations may be determined from

$$R(M + k) = -\sum_{m=1}^{M} R(M - m + k)\, \beta_{m,M} \tag{4.24}$$

for $k = 1,2,.....\infty$. To illustrate this point, assume that along with determining the coefficients $\{\beta_{i,j}\}$ at each stage, equation (4.23) is used to form the correlations

$$R(1) = -\beta_{11}R(0) , \tag{4.25}$$

$$R(2) = -\beta_{12}R(1) - \beta_{22}R(0) , \tag{4.26}$$

$$R(3) = -\beta_{13}R(2) - \beta_{23}R(1) - \beta_{33}R(0) . \tag{4.27}$$

Under the assumption that M=3, estimates of the unknown correlations are now determined from

$$R(3 + k) = -\sum_{m=1}^{3} R(3 - m + k)\beta_{M,3} , \tag{4.28}$$

for $k > 0$.

The asymptotic properties of the MEM spectral estimate have been investigated by Kromer (1970). Kromer has indicated that the MEM estimate is asymptotically unbiased,

$$E\left(\hat{S}(f)\right) = S(f) , \tag{4.29}$$

and asymptotically normal, with variance

$$\text{Var } \hat{S}(f) = \frac{2M}{N}\left(S^2(f)\right), \tag{4.30}$$

where N is the record length and M is the order of the model.

The problem to be investigated in this study can be contrasted with the MEM technique through the assumptions imposed. The problem under study will presuppose a structure, that is, the process under consideration will have an all pole spectral density. It will be assumed that the sampling rate is under the control of the experimenter and as such, will be fast. The coefficients in the models of the random process will be expressed in terms of probability density functions established via the application of Bayes Rule. It is the function of the next chapter to provide the requisite background in

Bayesian parameter estimation, so that the solution process can be accomplished.

CHAPTER V

BAYESIAN PARAMETER ESTIMATION

5.1 Introduction

The general parameter estimation problem is structured around the situation which is formed when an experimenter has a set of observations which have been assumed to come from a distribution, e.g. $N(\theta,\sigma^2)$. In a parameter estimation problem, one has the observations at hand and wishes to make some inferences about the unknown parameter (θ) of the distribution. In the Bayesian approach to this problem, a model for a prior density $p(\theta)$ is assumed, which expresses the knowledge about the parameter θ before observations are made. With the prior density $p(\theta)$, the observations model $p(\underline{y}/\theta)$ and observations $\underline{y}$. Bayes Rule can be applied to calculate the posterior density $p(\theta/\underline{y})$ and make inferences about the parameter θ. Therefore, the Bayesian approach is basically the application of Bayes Rule

$$p(\theta/\underline{y}) = \frac{p(\underline{y}/\theta)p(\theta)}{p(\underline{y})} ,$$

where

$$p(y) = E_\theta p(\underline{y}/\theta) = \begin{cases} \int_\theta p(\underline{y}/\theta)p(\theta)\, d\theta, & \theta \text{ continuous} \\ \sum_\theta p(\underline{y}/\theta)p(\theta) , & \theta \text{ discrete} . \end{cases} \qquad (5.1)$$

Given the observations, the denominator term in Bayes Rule is a normalizing constant which insures that the posterior density integrates to one. In this same circumstance, the density $p(\underline{y}/\theta)$ may be regarded

as a function of θ. When interpreted this way, it is called the likelihood function of θ, and is denoted as $\ell(\theta/\underline{y})$. With the likelihood function defined, Bayes Rule can be written as

$$p(\theta/y) = \frac{\ell(\theta/\underline{y})p(\theta)}{\int \ell(\theta/\underline{y})p(\theta)d\theta}$$

$$= K\ell(\theta/\underline{y})p(\theta) \quad , \tag{5.2}$$

Written this way, the role of the likelihood function $(\ell(\theta/\underline{y}))$ is demonstrated to be the manner in which the data affects the prior knowledge on the parameter θ.

The iterative application of Bayes Rule illustrates how knowledge is updated by the combination of what is learned from the observations, and from prior information. To see this, consider the post density formed from n observations,

$$p(\theta/\underline{Y}_n) = \frac{\prod_{i=1}^{n} p(y_i/\underline{Y}_{i-1},\theta)p(\theta)}{\int \prod_{i=1}^{n} p(y_i/\underline{Y}_{i-1},\theta)p(\theta)d\theta}$$

$$= \frac{\prod_{i=1}^{n} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)}{K(\underline{Y}_n)} \quad , \tag{5.3}$$

where $\underline{Y}_n$ is the observation vector denoted as

$$\underline{Y}_n = (y_1, y_2, \ldots y_n) \ .$$

Writing the post density formed from n-1 observations find

$$p(\theta/\underline{Y}_{n-1}) = \frac{\prod_{i=1}^{n-1} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)}{\int \prod_{i=1}^{n-1} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)d\theta}$$

$$= \frac{\prod_{i=1}^{n-1} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)}{K(\underline{Y}_{n-1})} \quad . \tag{5.4}$$

Now, interpreting the post density for n observations based on n-1 observations, yields

$$p(\theta/\underline{Y}_n) = \frac{p(y_n/\underline{Y}_{n-1},\theta)\cdot\prod_{i=1}^{n-1} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)}{\int p(y_n/\underline{Y}_{n-1},\theta)\cdot\prod_{i=1}^{n-1} p(y_i/\underline{Y}_{i-1}\theta)p(\theta)d\theta}$$

$$= \frac{p(y_n/\underline{Y}_{n-1},\theta)\cdot p(\theta/\underline{Y}_{n-1})\cdot N(\underline{Y}_{n-1})}{\int p(y_n/Y_{n-1}\theta)\ p(\theta/\underline{Y}_{n-1})\ N(\underline{Y}_{n-1})d\theta}$$

$$= \frac{p(y_n/\underline{Y}_{n-1},\theta)p(\theta/\underline{Y}_{n-1})}{\int p(y_n/\underline{Y}_{n-1},\theta)p(\theta/\underline{Y}_{n-1})d\theta}$$

$$= \frac{p(y_n/\underline{Y}_{n-1}\theta)p(\theta/\underline{Y}_{n-1})}{K(\underline{Y}_n)}\ . \qquad (5.5)$$

From this form, the sequential nature of Bayes Rule is evident, as the post from n-1 observations $\left(p(\theta/\underline{Y}_{n-1})\right)$ becomes the prior, which determines the post for n observations $\left(p(\theta/\underline{Y}_n)\right)$. It is now evident that Bayes Rule describes the process of learning from observations and how prior knowledge is affected as new observations are made. An example will illustrate the effect that priors have on the post and the learning that takes place.

Let the observation ($\underline{y}$) be taken from a distribution specified as N(85,16). Now, based on these samples and the two assumed priors $p_A(\theta) \sim N(90,4)$ and $p_B(\theta) \sim N(80,64)$, it is desired to estimate the mean (θ). If a single observation is made, the likelihood function $\ell(\theta/y)$ is represented by a normal distribution centered at the observation (y) with standard deviation 4. Bayes Rule can be applied to illustrate how the prior densities $p_A(\theta)$ and $p_B(\theta)$ effect the informa-

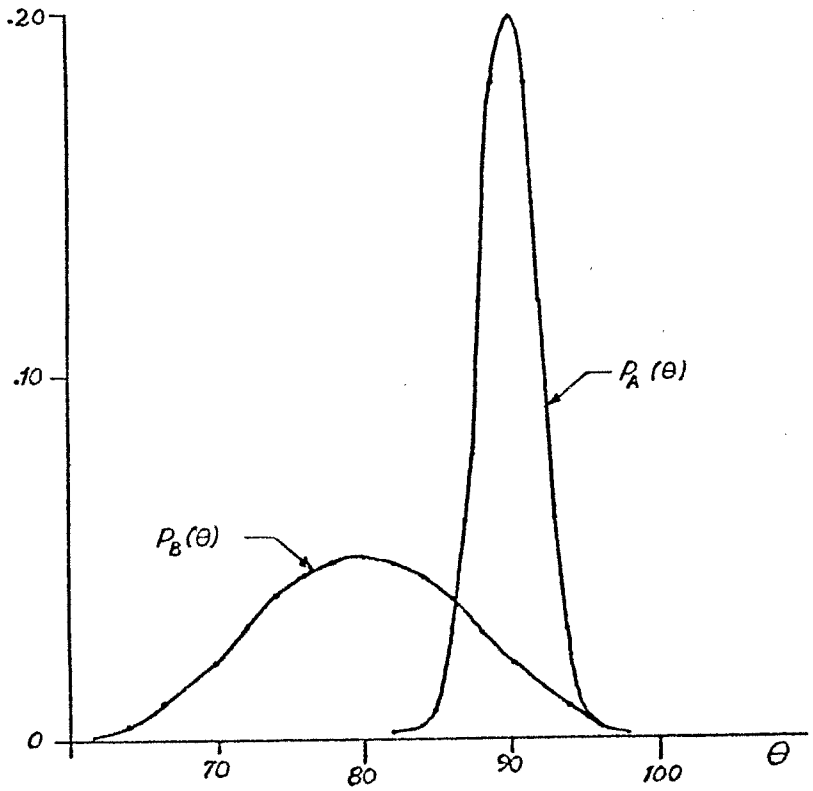

Fig. 1. Prior Density Functions for the Parameter θ.

tion presented by the observation. It is shown in Appendix B that the post density of θ is normal $N(\tilde{\theta},\tilde{\sigma}^2)$ where the post mean and variance are

$$\tilde{\theta} = \frac{N\alpha^2}{N\alpha^2 + \sigma^2}\left(\frac{1}{N}\sum y_i\right) + \frac{\sigma^2\mu}{N\alpha^2 + \sigma^2}, \tag{5.6}$$

$$\tilde{\sigma}^2 = \frac{\sigma^2\alpha^2}{N\alpha^2 + \sigma^2}, \tag{5.7}$$

Where N is the number of observations,

μ is the prior mean,

α^2 is the prior variance

σ^2 is the variance of the observations.

Now let the observation be y = 85, then the likelihood function takes the form shown in Figure 2. The effect of this observation is shown in Figure 3, where the post density based on the prior $p_A(\theta)$ is N(89,3.2), and the post density based on the prior $p_B(\theta)$ is N(84,13). Continuing with the example, assume that 10 observations have been made and as a consequnce, the sample mean $(1/N\sum y_i)$ is 87. The effects of these observations on the likelihood ratio are illustrated in Figure 4. The post density based on the prior $p_A(\theta)$ is N(87.1,1.15), and the post density based on the prior $p_B(\theta)$ is N(87,1.6). These post densities are illustrated in Figure 5.

So, even though the example started off with different prior information, $(p_B(\theta)$ represented more uncertainty than $p_A(\theta))$ the final answers were in agreement. After one observation, there was a wide spread in the answers. This was due to the likelihood function not being strongly peaked after one observation and hence the priors dominated the likelihood function. When the likelihood function became very peaked (after 10 observations) both priors now are dominated by

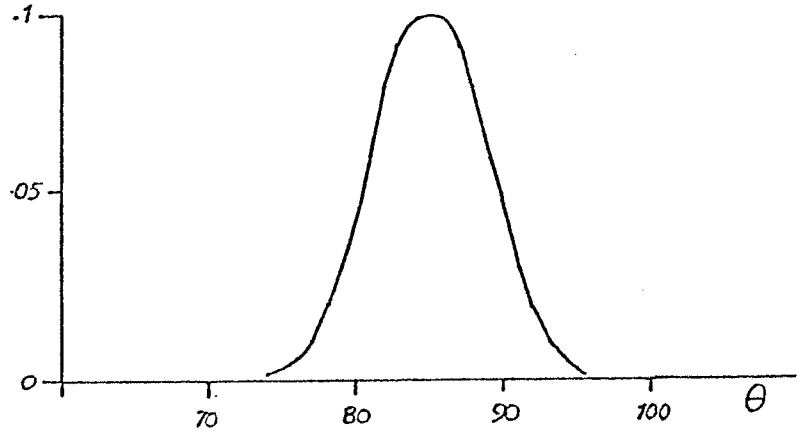

Fig. 2. The Likelihood Function for an Observation y = 85.

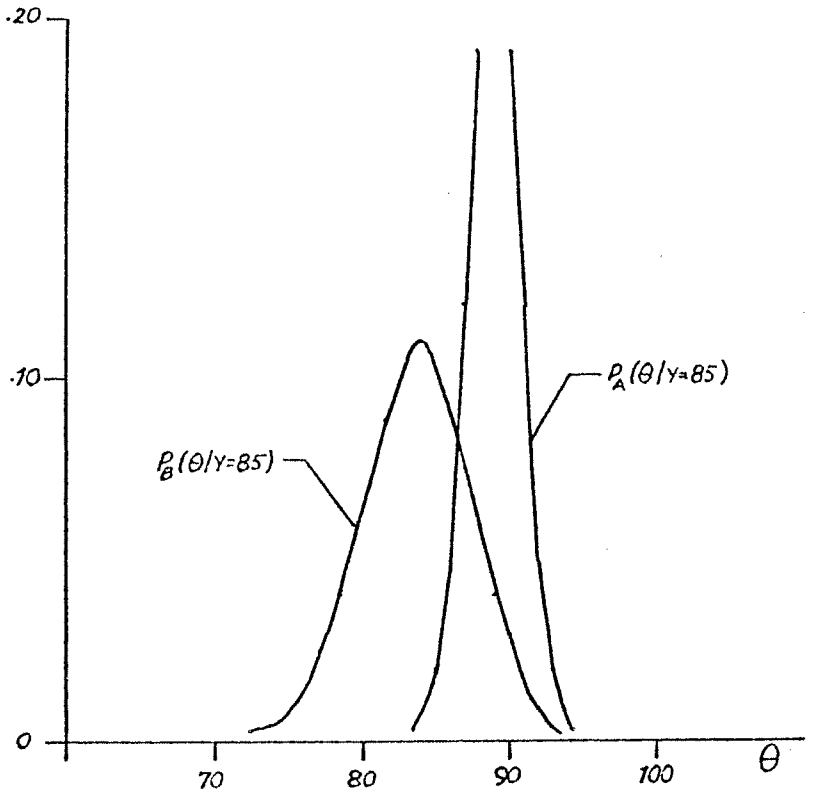

Fig. 3. Post Densities Resulting from the Observation y = 85.

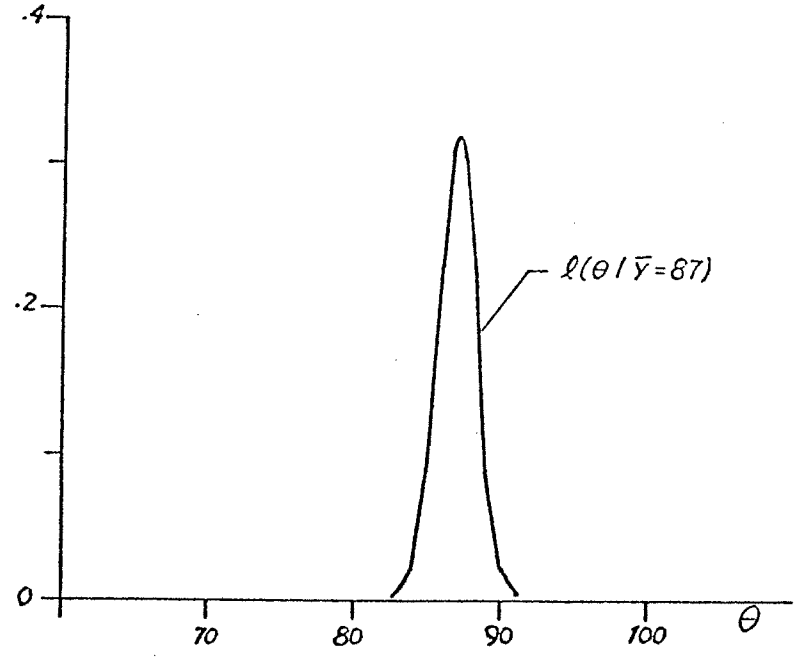

Fig. 4. The Likelihood Function for the Sample Mean $\bar{y} = 87$.

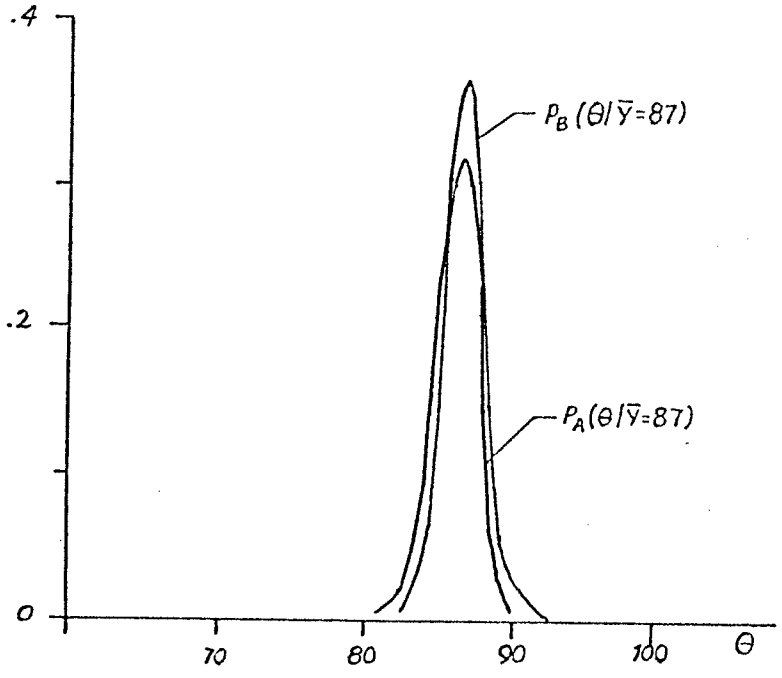

Fig. 5. Post Densities Resulting from $\bar{y} = 87$.

the likelihood function, and their effect on the post density is minimal.

5.2 Bayes Rule in the Presence of Structure

In general, if one is given the set of observations ($\underline{Y}_N$) where

$$\underline{Y}_N = (y_1, y_2, \ldots y_N),$$

the application of Bayes Rule to form a post density requires storage of a growing number of observations. For the problem of estimating the parameter θ, the post density is found from

$$f(\theta/\underline{Y}_N) = \frac{\prod_{i=1}^{N} f(y_i/\underline{Y}_{i-1},\theta) f(\theta)}{\int (\text{numerator}) d\theta}$$

where

$$f(\underline{Y}_N/\theta) = f(y_N/y_{N-1} \cdots y_1,\theta) f(y_{N-1}/y_{N-2} \cdots y_1,\theta) \cdot f(y_2/y_1\theta) f(y_1/\theta). \tag{5.8}$$

The evaluation of the post density requires storing the observation vector at each stage. If the observations are such that there is a structure imposed, evaluation of the post density function is simplified.

The class of observation processes of interest are those which are stationary and Markov. In dealing with stationary sequences (processes) it is assumed that the observations have been taken from $i = -\infty$. For the Mth order Markov processes, this fact is accounted for by a random initial condition vector

$$\underline{y}_0 = (y_0 \cdot \cdot \cdot y_{-(M-1)}),$$

where the random vector $\underline{y}_0$ is selected to insure this stationarity. With this definition in mind, the probability density function for the Mth order Markov observations is

$$f(y_N, y_{N-1} \cdots y_1, y_0 \cdots y_{-(M-1)}) = f(\underline{y}_0) \cdot \prod_{i=1}^{N} f(y_i / y_{i-1} \cdots y_{i-M}) . \tag{5.9}$$

Notationally, this equation can be made more convenient if the definition

$$\underline{y}_i = (y_i \cdot \cdot \cdot y_{i-(M-1)}) ,$$

is used to produce the result

$$f(\underline{Y}_N, \underline{y}_0) = f(\underline{y}_0) \cdot \prod_{i=1}^{N} f(y_i / \underline{y}_{i-1}) . \tag{5.10}$$

For an M=2 process, the observations density is

$$f(\underline{Y}_N, \underline{y}_0) = f(y_0, y_{-1}) \; f(y_N / y_{N-1}, y_{N-2}) f(y_{N-1} / y_{N-2}, y_{N-3}) \cdots \cdot f(y_1 / y_0, y_{-1}) \quad . \tag{5.11}$$

Comparison of this equation with equation (5.8) shows the simplification that results from assuming a structure in the observations. Conditioning the observations on the initial condition produces a further simplification which is

$$f(\underline{Y}_N / y_0) = \prod_{i=1}^{N} f(y_i / \underline{y}_{i-1}) \quad . \tag{5.12}$$

The result of assuming this structure is the reduction in storage at each stage, to an M vector which is of constant dimension from stage-to stage.

In simulating these observations, the initial condition ($\underline{y}_0$) is usually accounted for by running the model for a sufficient number of samples so that the transient effects have died out. These transient affects are due to assuming some arbitrary initial conditions other than the initial condition with the correct probabilistic properties to insure stationarity. If the number of observations is large, then such effects are negligible. From this point on, when dealing with

the observations generated from a Markov type model, the parameter estimate will be conditioned on the $\underline{y}_0$ vector, although it may not be explicitly stated.

5.3 Choice of Priors

Although the choice of priors has not been mentioned up to this point, it is one of the main aspects of the Bayesian approach. The freedom of choice allowed, provides a method of placing subjective knowledge into the problem, or using subjective knowledge to make the problem analytically tractable. In a decision problem, the prior information may be so strong that it would dominate what could be learned from taking observations. This is precisely what is desired in this type of problem. However, in some estimation problems the prior is vague, and hence the prior should be as broad as possible.

In selecting a prior to express vague opinions, it is important to realize that the use of a broad distribution is not an expression of uncertainty (Raiffa and Schlaifer, 1961). Rather, vagueness is expressed relative to the likelihood function (observations density) by using a prior which can be easily modified by a few observations (the likelihood function dominates the prior). Similarly, strong prior opinions are expressed by a prior which is more peaked than the likelihood function. One then considers vague (non-informative) priors as those which are broad over the range of values in which the likelihood function is strongly defined. Often, these priors give rise to improper distributions in the sense that the integral over the parameter space of the density function is infinite. For example, consider f(x)

to be uniform over $(-\infty,\infty)$, then the integral

$$\int_{-\infty}^{\infty} f(x)\ dx \ = \ c \int_{-\infty}^{\infty} dx \tag{5.13}$$

does not exist no matter how small c is. However, if one used such a distribution to represent the local behavior, where the likelihood is peaked, but not over the entire range of the parameter, and over this remaining range it has tails that trail to zero, the distribution is then proper. Often, an improper distribution is assumed as a prior to develop a post density which is proper once observations have been made. This is shown by carrying out the calculation in Bayes Theorem with a uniform prior. If $\ell(\theta/\underline{x})$ is the likelihood function of some set of observed values, and if after some observations

$$\int_{-\infty}^{\infty} \ell(\theta/\underline{x})\ d\theta \ < \ \infty \ ,$$

then the post density $f(\theta/\underline{x})$ will be

$$f(\theta/\underline{x}) \propto \ell(\theta/x) \tag{5.14}$$

and

$$\int f(\theta/\underline{x})\ d\theta < \infty \ .$$

The Principle of Stable Estimation (DeGroot, 1970) provides conditions under which a post density derived from an improper uniform density will closely approximate a post density that would be found from a more precise prior density. In fact, there is a theorem (Edwards, et al., 1963) that quantifies the measure of closeness of the approximate post density to the actual density that is achieved. In essence, this result shows that if the likelihood is peaked enough, then use of a uniform prior will provide a good approximation to the answer.

Another approach to choosing a prior distribution when information is vague, is through the use of reproducing (conjugate) densities. The use of a member of such a class guarantees that the post density is a member of the same class, which eases the burden of determining the normalizing constant in Bayes Rule. In general, one picks this member indexed by a parameter, and then studies the affect on the post distribution as the parameter approaches some limit. However, if the likelihood function is rather peaked, and a large number of observations are involved, major changes in the prior will have little effect on the post density.

5.4 Non-informative Priors

The reason for seeking non-informative priors is to obtain a distribution that would be representative of an uninformed observer. The selection of such a prior is guided by the requirement that this prior provides comparatively little information in reference to what can be expected from taking observations. For a single parameter, it is shown (Box and Tiao, 1973) that if the likelihood $\ell(\theta/\underline{Y})$ can be expressed as

$$\ell(\theta/\underline{Y}) = g\left[\phi(\theta) - f(\underline{Y})\right], \tag{5.15}$$

where $g\left[\cdot\right]$ is a known function independent of the data and $f(\underline{Y})$ is a function of $\underline{Y}$, then a locally uniform prior for θ can be found. This locally uniform prior is proportional to $|d\phi/d\theta|$, if $\phi(\theta)$ is a one to one transformation. For the case of normal observations with unknown mean, $\phi(\theta) = \theta$, and $f(\underline{Y}) = \sum y_i/N$, the non-informative prior is uniform. For the case of normal observations with unknown standard deviation (σ), $\phi(\theta) = \log \sigma$, and

$$f(\underline{Y}) = \log\left(\frac{\sqrt{\sum (y_i - \mu)^2}}{N}\right) , \tag{5.16}$$

hence, the non-informative prior is uniform. If more than one parameter is involved, the choice of non-informative priors is extremely complex, and to pursue this further would be of no immediate benefit save for the multiparameter Gaussian mean case. For the latter situation, the non-informative prior is uniform . Now, some examples illustrating the principles of selecting non-informative priors will be presented. To begin the illustration assume that a model with a Markov structure, specifically M=1, generates the observations. Hence, let the model be

$$y_i - \hat{a}y_{i-1} = \zeta_i , \tag{5.17}$$

where

$$\zeta_i \sim N(0,\sigma^2)$$

$$E(\zeta_i y_j) = 0 \qquad i \neq j .$$

Initially, let $\hat{a}$ be a random parameter and σ^2 be known. The likelihood function is then (for N observations)

$$\begin{aligned} \ell(\hat{a}/\underline{Y}_N, y_0) &\propto \prod_{i=1}^{N} \exp\left(-\frac{1}{2\sigma^2}(y_i - \hat{a}y_{i-1})^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(y_i - \hat{a}y_{i-1})\right) . \end{aligned} \tag{5.18}$$

For this example, a locally uniform prior can be represented as

$$f(\hat{a}) \propto K ,$$

over the range of $\hat{a}$, where $\ell(\hat{a}/\underline{Y}_N, y_0)$ is dominant. Continuing with the example, the post density is

$$f(\hat{a}/\underline{Y}_N, y_0) \propto \ell(\hat{a}/\underline{Y}_N, y_0) f(\hat{a})$$

$$f(\hat{a}/\underline{Y}_N, y_0) \propto K \exp - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \hat{a}y_{i-1})^2 \quad . \tag{5.19}$$

The post density has the form of a Gaussian density, except for the constant of integration. So, upon completing the square in $\hat{a}$, we find that

$$f(\hat{a}/\underline{Y}_N, y_0) = K \exp \sum_{i=1}^{N} - \frac{1}{2\sigma^2} (y_i^2 - 2\hat{a}y_i y_{i-1} + \hat{a}^2 y_{i-1}^2)$$

$$= K \exp \sum_{i=1}^{N} - \frac{1}{2\sigma^2} (\hat{a}^2 y_{i-1}^2 - 2\hat{a}y_i y_{i-1} + y_i^2)$$

$$= K \exp - \frac{1}{2\sigma^2} \left(\hat{a}^2 \sum y_{i-1}^2 - 2\hat{a} \sum y_i y_{i-1} + \sum y_i^2 \right)$$

$$= \tilde{K} \exp - \frac{1}{2\sigma^2} \sum y_{i-1}^2 \left(\hat{a} - \frac{\sum y_i y_{i-1}}{\sum y_{i-1}^2} \right)^2$$

$$= \frac{\sqrt{\Sigma\, y_{i-1}^2}}{\sqrt{2\pi\sigma^2}} \exp - \frac{1}{2\left(\frac{\sigma^2}{\sum y_{i-1}^2}\right)} \cdot \left(\hat{a} - \frac{\sum y_i y_{i-1}}{\sum y_{i-1}^2} \right)^2 \tag{5.20}$$

Defining the quantities

$$\hat{R}_0 = \frac{1}{N} \sum_{i=1} y_i^2 \quad , \tag{5.21}$$

$$\hat{R}_1 = \frac{1}{N} \sum y_i y_{i-1} \tag{5.22}$$

and replacing the corresponding quantities in the density function, in equation (5.20), the result is

$$f(\hat{a}/\underline{Y}_N, y_0) = \frac{1}{\sqrt{2\pi}} \quad \frac{\sqrt{N\hat{R}_0}}{\sigma} \cdot \exp - \frac{1}{2}\left(\frac{N\hat{R}_0}{\sigma^2}\right)\cdot\left(\hat{a} - \frac{\hat{R}_1}{\hat{R}_0}\right)^2$$

$$\hat{a} \sim N\left(\frac{\hat{R}_1}{\hat{R}_0}, \frac{\sigma^2}{N\hat{R}_0}\right) \quad .$$

From the relations in the model, the correlation function satisfies the normal equation for M=1, which is

$$R_1 - \hat{a}R_0 = 0 \ ,$$

where $\hat{a}$ is the root of the characteristic equation. For a stable solution to the difference equation

$$y_i - \hat{a}y_{i-1} = \zeta_i \ ,$$

it is required that $|\hat{a}| < 1$. The post density function provides this type of result since the mean is

$$\bar{\hat{a}} = \frac{\hat{R}_1}{\hat{R}} \ ,$$

and from the properties of the correlation function, it can be concluded that $|\hat{a}| < 1$.

Considering the same model, now it is assumed that the parameter $\hat{a}$ is known, but the parameter σ^2 is now the random variable to be estimated. That is, the likelihood function is

$$\ell(\sigma^2/\underline{Y}_N) \sim \sigma^{-N} \prod_{i=1}^{N} \exp - \frac{\sigma^{-2}}{2}(y_i - \hat{a}y_{i-1})^2$$

$$\sim \sigma^{-N} \exp - \frac{\sigma^{-2}}{2} \sum_{i=1}^{N} (y_i - \hat{a}y_{i-1})^2 \qquad (5.24)$$

Defining the quantity Ns^2 as

$$Ns^2 = \sum_{i=1}^{N} (y_i - \hat{a}y_{i-1})^2 ,$$

then the likelihood function is

$$\ell(\sigma^2/\underline{Y}_N) \propto \sigma^{-N} \exp - \frac{\sigma^{-2}}{2} \cdot Ns^2 . \tag{5.25}$$

From the previous discussion, the locally uniform prior for this likelihood function is

$$f(\sigma^2) = \sigma^{-2} .$$

Using this prior, the post density then becomes

$$f(\sigma^2/\underline{Y}_N) = K \cdot \ell(\sigma^2/\underline{Y}_N) \cdot \sigma^{-2}$$

$$= K \cdot \sigma^{-(N+2)} \exp \left\{ - \frac{Ns^2}{2\sigma^2} \right\} , \tag{5.26}$$

where

$$Ns^2 \triangleq \sum (y_i - \hat{a}y_{i-1})^2 .$$

All that is now necessary is the evaluation of the integration constant which makes the post density proper. This integration is possible through the use of the relations (Abramovitz and Stequn, 1970)

$$B^{-\alpha}\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-Bu} du = \int_0^\infty x^{-(\alpha+1)} e^{-B/x} dx . \tag{5.27}$$

The post density now takes the form

$$f(\sigma^2/\underline{Y}) = \begin{cases} \dfrac{\left(\dfrac{Ns^2}{2}\right)^{N/2}}{\Gamma(N/2)} \sigma^{-(N/2+1)} \exp - \dfrac{Ns^2}{2\sigma^2} & , \sigma^2 > 0 \\ 0 & , \text{else} \end{cases} \tag{5.28}$$

where $\Gamma(\alpha)$ is the gamma function defined by

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt .$$

This result makes sense since the mean and variance are given by

$$\overline{\sigma^2} = s^2 = \frac{1}{N} \sum (y_i - \hat{a}y_{i-1})^2,$$

$$\text{Var } (\sigma^2) = \frac{2s^4}{N},$$

where s^2 is the sample variance of the set of independent observations.

5.5 Sufficient Statistics and Reproducing Densities

In seeking an estimate of a parameter, it would simplify matters significantly if there was some summarizing measurement that contained as much information as the observations themselves. Such a measurement is called a sufficient statistic. A sufficient statistic is defined (Raiffa and Schlaiffer, 1961) as any mapping (g) from the observations space (Z) into Euclidean space (y). This mapping is a sufficient statistic when, for any prior density $f(\theta)$ and any observation (z), the post densities $\left(f(\theta/\cdot)\right)$ formed from the application of Bayes Rule satisfy

$$f(\theta/Z=z_1) = f(\theta/Z=z_2),$$

when the mapping satisfies

$$g(z_1) = g(z_2).$$

Although the above statements provided a formal definition of a sufficient statistic, it did not provide a methodology to determine one. The sufficient statistic is found from the likelihood function $\ell(\theta/Z)$, by the application of the Factorization Criterion. This criterion asserts that if the likelihood $\ell(\theta/Z)$ can be factored as

$$\ell(\theta/Z) = n\left(g(Z), \theta\right)p(Z), \tag{5.29}$$

then g(Z) is a sufficient statistic. Applying this concept in the application of Bayes Rule provides the result

$$f(\theta/Z) = \ell(\theta/Z)f(\theta) \;/\; \int(\text{numerator})\,d\theta$$
$$= n\big(g(Z),\theta\big)f(\theta) \;/\; \int n\big(g(Z),\theta\big)f(\theta)\,d\theta \quad . \tag{5.30}$$

It is evident that this minimal description for the likelihood function provides the information needed to form the post density. In seeking minimal representations, it would be appealing if the prior $\big(f(\theta)\big)$ could also be represented in this way. Such a description for the prior is termed a kernel of the density function. A kernel for $f(\theta)$ is defined as another function $r(\theta)$ that makes the ratio $r(\theta)/f(\theta)$ a constant relative to θ. This means that the density function and a kernel are related by

$$f(\theta) = r(\theta) \;/\; \int r(\theta)\,d\theta \quad . \tag{5.31}$$

Using a kernel of $f(\theta)$ in equation (5.30) provides the result

$$f(\theta/Z) = n\big(g(Z),\theta\big)\; r(\theta) \Big/ \int n\big(g(Z),\theta\big)r(\theta)\,d\theta \quad . \tag{5.32}$$

The significance here is that the post density is completely described in terms of a kernel of the likelihood function and the prior, and the functional forms of $n(\cdot,\cdot)$ and $r(\cdot)$ determine the form of the post density. The post density can be written as

$$f(\theta/Z) = K\; n\big(g(Z),\theta\big)r(\theta) \quad . \tag{5.33}$$

If $n(g,\theta)$ is such that the form of the post density $\big(f(\theta/z)\big)$ is the same as $r(\theta)$ then it leads to what is called a reproducing density.

This reproducing property is a very appealing aspect of using a Bayesian approach to estimation. When the prior distributions have this reproducing property, the computation of the post density is a simple process of computing new values for the parameters characterizing the distributions.

Following the work of Gobein (1974), this reproducing property

is ascribed to a family of density functions $\{W(\theta,p)\}$ indexed by a finite dimensional parameter (p). This family of density functions reproduces under the observations density $f(\underline{Y}_n/\theta)$ if for any n and any prior $f(\theta) = w(\theta,p_0)$, there is a parameter $p_n = p_n(p_0,\underline{Y}_n)$ such that when Bayes Rule is applied, the post density that results is $f(\theta/Y_n) = w(\theta,p_n)$. It turns out that this property is a direct result of the existence of a fixed dimensional sufficient statistic and is solely dependent on its functional form. To establish the reproducing type from the likelihood function, assume that $\ell(\theta/\underline{Y}_n)$ has a sufficient statistic of fixed dimension, so that

$$\ell(\theta/\underline{Y}_n) = g\{t(\underline{Y}_n),\theta\}h(\underline{Y}_n) \quad . \tag{5.34}$$

Then a reproducing density $p(\theta,t)$ can be found from

$$p(\theta,t_n) = g\{t(\underline{y}_n),\theta\} \Big/ \int g\{t(\underline{y}_n)\theta\}\,d\theta \quad , \tag{5.35}$$

if the integral above exists.

To demonstrate that a prior selected by use of equation (5.35) reproduces, the sequential form of Bayes Rule,

$$f(\theta/\underline{Y}_n) = \frac{f(y_n/\underline{Y}_{n-1},\theta)f(\underline{Y}_{n-1}/\theta)}{\int(\text{numerator})\,d\theta}$$

$$= \frac{f(\underline{Y}_n/\theta)f(\theta)}{\int(\text{numerator})\,d\theta} \tag{5.36}$$

is applied. Assuming that a sufficient statistic of fixed dimension exists, the density of $f(\underline{Y}_n/\theta)$ can be written as

$$f(\underline{Y}_n/\theta) = g\{t(\underline{Y}_n),\theta\}h(\underline{Y}_n)$$

$$= f(y_n/\underline{Y}_{n-1},\theta)f(\underline{Y}_{n-1}/\theta) \quad , \tag{5.37}$$

and similarly the density $f(\underline{Y}_{n-1}/\theta)$ can be written as

$$f(\underline{Y}_{n-1}/\theta) = g\{t(\underline{Y}_{n-1}),\theta\}h(\underline{Y}_{n-1}) \quad . \tag{5.38}$$

The priors are specified by

$$p(t_n,\theta) = g(t_n,\theta) \Big/ \int g(t_n,\theta)d\theta \ , \tag{5.39}$$

$$p(t_{n-1},\theta) = g(t_{n-1},\theta) \Big/ \int g(t_{n-1},\theta) \ . \tag{5.40}$$

Rewriting the observations density to reflect the stage wise dependence, and using the sufficient statistic results in

$$f(\underline{Y}_n/\theta) = f(y_n/\underline{Y}_{n-1},\theta)f(\underline{Y}_{n-1}/\theta) = g(t_n,\theta)h(\underline{Y}_n), \tag{5.41}$$

so that equation (5.41) can be written as

$$g(t_n,\theta)h(\underline{Y}_n) = f(y_n/\underline{Y}_{n-1},\theta)g(t_{n-1},\theta)h(\underline{Y}_{n-1}) \ . \tag{5.42}$$

Using the relationship

$$g(t_i,\theta) = p(t_i\cdot\theta)\int g(t_i,\theta)d\theta \ ,$$

between the priors and the kernel of the observations density in equation (5.37), the result is

$$p(t_n,\theta) = f(y_n/\underline{Y}_{n-1}\theta)p(t_{n-1},\theta)\left\{\frac{h(\underline{Y}_{n-1})}{h(\underline{Y}_n)}\cdot\int \frac{g(t_{n+1},\theta)}{g(t_n,\theta)}\,d\theta\right\}$$

$$= K\ p(t_{n-1},\theta|f(y_n/\underline{Y}_{n-1},\theta) \ . \tag{5.43}$$

From this, the post density based on n observations is represented as

$$p(t_n,\theta) = \frac{f(y_n/\underline{Y}_{n-1}\theta)p(t_{n-1},\theta)}{\int (\text{numerator})d\theta} \ ,$$

which is the sequential form of Bayes Rule with $p(t_{n-1},\theta)$ as a prior and $p(t_n,\theta)$ as a post density. This demonstrates that a prior selected to satisfy the conditions of equation (5.35) reproduces.

As an illustration of finding a reproducing prior, consider again the model for the M=1 process

$$y_i - \hat{a}y_{i-1} = \zeta_i \ ,$$

where ζ_i is $N(0,\sigma^2)$ and the parameter $\hat{a}$ is the random variable to be estimated. The likelihood function is

$$\ell(\hat{a}/\underline{Y}_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{a}y_{i-1}) \quad . \tag{5.44}$$

Using the "Factorization Theorem", equation (5.29), to find the kernel of the likelihood function, the result is

$$\begin{aligned}\ell(\hat{a}/\underline{Y}_n) &= \exp\cdot - \frac{1}{2\sigma^2}\sum(\hat{a}^2 y_{i-1}^2 - 2\hat{a}y_i y_{i-1}) \\ &\quad \cdot\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp - \frac{1}{2\sigma^2}\sum y_i^2 \\ &= g\left(t(\underline{Y}_n), a\right)\cdot h(\underline{Y}_n) \quad ,\end{aligned} \tag{5.45}$$

where

$$t(\underline{Y}_n) = \left[\sum y_{i-1}^2, \sum y_i y_{i-1}\right] = (t_{1n}, t_{2n}) \quad .$$

Selecting the prior so that

$$p(t_n, \hat{a}) = \frac{g(t_n, \hat{a})}{\int_{\Omega\hat{a}} g(t_n, \hat{a})\, d\hat{a}} \quad ,$$

then

$$p(t_n, \hat{a}) = K^{-1}\cdot\exp - \frac{1}{2\sigma^2} t_{1n}\left[\hat{a} + \left(\frac{t_{2n}}{t_{1n}}\right)\right]^2, \tag{5.46}$$

where

$$K^{-1} = \int_{-\infty}^{\infty} \exp - \frac{1}{2\sigma^2} t_{1n}\left[\hat{a} + \frac{t_{2n}}{t_{1n}}\right]^2 d\hat{a} = \frac{1}{\sqrt{2\pi\sigma^2/t_{1n}}} , \tag{5.47}$$

and upon letting

$$\frac{\sigma^2}{t_{1n}} = \zeta_{1n} \quad \text{and} \quad \frac{t_{2n}}{t_{1n}} = \zeta_{2n} \quad ,$$

the reproducing family has the form

$$p(\zeta_n, \hat{a}) = \frac{1}{\sqrt{2\pi\zeta_{1n}}} \exp - \frac{(\hat{a} + \zeta_{2n})^2}{2\zeta_{1n}} \quad . \tag{5.48}$$

Recalling that we are trying to model a stationary process, there is

one difficulty in taking $p(\zeta_0,\hat{a})$ as a prior. This prior is defined over the whole real line, and it is known, a priori, that the parameter $\hat{a}$ must have magnitude less than one. The incorporation of such a constraint is unique to Bayesian estimation. Injecting this constraint onto a prior which reproduces does not change this property. It should be evident that the reproducing property depends only on the kernel of the likelihood $g(t_n,\hat{a})$, and that constraining the allowable range of the random parameter will only effect the normalizing constant (integral) that makes the density $p(t_n,\hat{a})$ proper. That this is true is demonstrated by a theorem due to Spraggins (1965), and generalized by Gobein (1974), which shows that if the priors $f_1(\hat{a})$ and $f_2(\hat{a})$ are related by

$$f_2(\hat{a}) = r(\hat{a})f_1(\hat{a}) \quad , \tag{5.49}$$

where $r(\hat{a})$ is a positive function and $f_1(\hat{a})$ is a reproducing prior, then $f_2(\hat{a})$ reproduces.

To incorporate a constraint, using this theorem, let Ω_c be the range over which the parameter $\hat{a}$ is to be constrained. Next, define

$$f_2(\hat{a}) = r(\hat{a})f_1(\hat{a}) \quad ,$$

where

$$r(\hat{a}) = \begin{cases} K \ , & \hat{a} \in \Omega_c \\ o \ , & \hat{a} \notin \Omega_c \end{cases} \tag{5.50}$$

and

$$K = \left(\int_{\Omega_c} f_1(\hat{a})d\hat{a} \right)^{-1} \quad . \tag{5.51}$$

Let $f_1(\hat{a}|\underline{Y}_k)$ be the post density developed from the unconstrained reproducing prior $f_1(\hat{a})$, with

$$f_1(\hat{a}|\underline{Y}_n) = \frac{f(\underline{Y}_n|\hat{a})\ f_1(\hat{a})}{\int f(\underline{Y}_n|\hat{a})f_1(\hat{a})d\hat{a}} \tag{5.52}$$

$$f_1(\hat{a}|\underline{Y}_n) = \frac{f(\underline{Y}_n|\hat{a})f_1(\hat{a})}{f_1(\underline{Y}_n)} \quad . \tag{5.53}$$

Let $f_2(\hat{a}|\underline{Y}_n)$ represent the post density from the constrained prior $f_2(\hat{a})$, where

$$f_2(\hat{a}|\underline{Y}_n) = \frac{f(\underline{Y}_n|\hat{a})f_2(\hat{a})}{\int f(\underline{Y}_n|\hat{a})f_2(\hat{a})d\hat{a}} \tag{5.54}$$

$$= \frac{f(\underline{Y}_n|\hat{a})f_2(\hat{a})}{f_2(\underline{Y}_n)} \quad . \tag{5.55}$$

Using equation (5.50) in equation (5.54), the constrained post density is

$$f_2(\hat{a}|\underline{Y}_n) = \begin{cases} \dfrac{f(\underline{Y}_n|\hat{a})Kf_1(\hat{a})}{\int_{\Omega_c} (\text{numerator})d\hat{a}} & , \quad \hat{a} \in \Omega_c \qquad (5.56) \\ 0 & \hat{a} \notin \Omega_c \qquad (5.57) \end{cases}$$

Substituting the relationship

$$f(\underline{Y}_n|\hat{a})f_1(\hat{a}) = f_1(\hat{a}|\underline{Y}_n)f_1(\underline{Y}_n) \quad ,$$

into equation (5.56), produces the result

$$f_2(\hat{a}|\underline{Y}_n) = \frac{f_1(\hat{a}|\underline{Y}_n)f_1(\underline{Y}_n)\cdot K}{\int_{\Omega_c} f_1(\hat{a}|\underline{Y}_n)f_1(\underline{Y}_n)\cdot K \, d\hat{a}}$$

$$= f_1(\hat{a}|\underline{Y}_n) \cdot \frac{1}{\int_{\Omega_c} f_1(\hat{a}|\underline{Y}_n)d\hat{a}}$$

$$= f_1(\hat{a}|\underline{Y}_n) \cdot \frac{1}{\Pr\{(\hat{a}\varepsilon\Omega_c)|\underline{Y}_n\}} \quad . \tag{5.58}$$

This result shows that the structure of the post density is not changed by the constraint, only multiplied by a constant. This result suggests a useful procedure for solving constrained problems. Begin-

ning the problem by seeking the solution to the unconstrained problem, we are then able to find the unconstrained post density $\left(f_1(\hat{a}|\underline{Y}_n)\right)$. To find the constrained solution, all that is necessary is the evaluation of the scale factor

$$S(\hat{a}) = \begin{cases} \left(\Pr(\hat{a}\varepsilon\Omega_c)\,|\underline{Y}_n\right)^{-1}, & \hat{a} \in \Omega_c \\ 0 \quad , & \hat{a} \notin \Omega_c \end{cases} .$$

The preference for this approach is based on the reduction in complexity, and the fact that for large μ/σ ratios on the post $f_1(\hat{a}|\underline{Y}_n)$ the modifying constant will be close to one.

Precisely the same approach can be taken when dealing with non-informative priors. That is, solve the unconstrained problem with a non-informative prior, then find the modifying scale factor. This can be verified by viewing the sequential behavior of the post density, beginning with a uniform prior. With a locally uniform prior, sequentially applying Bayes Rule will result in a post density whose structure is determined (except for a proportionality constant) by the kernel of the likelihood function. After a few observations, the prior will result in a post density that is proper, and a member of a family determined by the kernel. This means that this post density will become the new prior, in the sequential formulation of Bayes Rule, and now equation (5.49) can be used. The following paragraphs will apply these concepts (reproducing densities) to obtain the post densities of the parameters associated with an M=1 and M=2 autoregressive process.

Consider the random process generated by

$$y_i - \hat{a}y_{i-1} = \zeta(i) \quad ,$$

where $\zeta(i) \sim N(0,\sigma^2)$, with $\hat{a}$ as a random parameter, and $|\hat{a}| < 1$. From

the previous example, the family of reproducing densities is $N(\zeta_{2n}, \zeta_{1n})$ where

$$\zeta_{2n} = \frac{t_{2n}}{t_{1n}} = \frac{\sum y_i y_{i-1}}{\sum y^2_{i-1}}, \quad \text{and } \zeta_{1n} = \frac{\sigma^2}{\sum y^2_{i-1}} .$$

Assume that a prior for the parameter a is

$$\hat{a} \sim N(\mu_a, \sigma_a^2) .$$

Then the post density after n observations is

$$f(\hat{a}|\underline{Y}_n) = K \exp -\frac{1}{2}\left(\frac{\sigma^2 + \sigma_a^2 \sum y^2_{i-1}}{\sigma^2\sigma_a^2}\right)\left(\hat{a} - \frac{\mu_a\sigma^2 + \sigma_a^2\sum y_i y_{i-1}}{\sigma^2 + \sigma_a^2\sum y^2_{i-1}}\right)^2, \tag{5.59}$$

where

$$K = \frac{1}{2\pi\left(\frac{\sigma^2\sigma_a^2}{\sigma^2 + \sigma_a^2 \sum y^2_{i-1}}\right)} . \tag{5.60}$$

The constrained post density is

$$f_c(\hat{a}|\underline{Y}_n) = f(\hat{a}|\underline{Y}_n) \cdot \frac{1}{\Pr\{|\hat{a}| < 1|\underline{Y}\}} , \tag{5.61}$$

where

$$\Pr\{(|\hat{a}| < 1)|\underline{Y}_n\} = K \int_{-1}^{1} f(\hat{a}|\underline{Y}_n)\, da .$$

Asymptotically, the result is $a_{post} \sim N(\overset{*}{\mu}_a, \overset{*}{\sigma}^2_a)$, where

$$\overset{*}{\mu}_a = \frac{\mu_a + \frac{\sigma_a^2}{\sigma^2}\sum y_i y_{i-1}}{1 + \frac{\sigma_a^2}{\sigma^2}\sum y^2_{i-1}} \xrightarrow{\text{N large}} \frac{\sum y_i y_{i-1}}{\sum y^2_{i-1}} = \frac{\hat{R}_1}{\hat{R}_0} , \tag{5.62}$$

$$\overset{*}{\sigma}^2_a = \frac{\sigma^2\sigma_a^2}{\sigma^2 + \sigma_a^2 \sum y^2_{i-1}} \xrightarrow{\text{N large}} \frac{\sigma^2}{\sum y^2_{i-1}} = \frac{\sigma^2}{N\hat{R}_0} . \tag{5.63}$$

Under the assumption that the $\overset{*}{\mu}_a/\overset{*}{\sigma}_a$ is large, the modifying factor in

equation (5.51) is close to unity and the parameters of the constrained post are close to the parameters of the unconstrained post density.

Again, consider the model

$$y_i - \hat{a}y_{i-1} = \zeta_i \ , \ \zeta_i \sim N(o,\sigma^2) \ ,$$

but now, both $\hat{a}$ and σ^2 are random parameters. Using the ideas of reproducing densities, determines the post densities for $\hat{a}$ and σ^2, where the likelihood function (observations density) is

$$\begin{aligned} f(\underline{Y}_N|y_0) &= \prod_{i=1}^{N} f(y_i|\underline{Y}_{i-1}) \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \prod_{i=1}^{N} \exp \ -\frac{1}{2\sigma^2}(y_i - \hat{a}y_{i-1})^2 \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \cdot \exp \ -\frac{1}{2\sigma^2}\sum (y_i - \hat{a}y_{i-1})^2 \ . \end{aligned} \tag{5.64}$$

Since σ^2 and $\hat{a}$ are random parameters, the likelihood is

$$f(\underline{Y}_N|y_0) \propto (r)^{N/2} \ \exp\{-\frac{r}{2}\sum (y_i - \hat{a}y_{i-1})^2\} \ ,$$

where

$$r = \frac{1}{\sigma^2} \ .$$

Now, if the priors are selected as

$$f(r) \propto r^{w-1}e^{-\ell r} \ , \tag{5.65}$$

$$f(\hat{a}/r) \propto r^{1/2} \exp\{-\frac{rT}{2}(\hat{a} - \mu)^2\} \tag{5.66}$$

with T a known constant, then these priors will reproduce under the given likelihood. To see this, use Bayes Rule and write the post density as

$$f(\hat{a},r|\underline{Y}_N,y_0) \propto f(\underline{Y}_N|y_0,\hat{a},r)\cdot f(\hat{a}|r)f(r)$$

$$f(\hat{a},r|\underline{Y}_N,y_0)\propto(r^{N/2})\ \exp\{\sum_{i=1}^{N}-\frac{r}{2}(y_i-\hat{a}y_{i-1})^2\}\cdot$$

$$(r^{1/2})\ \exp\{-\frac{rT}{2}(\hat{a}-\mu)^2\}\cdot$$

$$(r^{w-1})\ \exp\{-\ell r\} \quad . \tag{5.67}$$

Upon collecting terms and completing the square in the parameter $\hat{a}$, the post density is

$$f(\hat{a},r|\underline{Y}_N,y_0)\ \propto\ r^{w'-1}e^{-\ell' r}r^{1/2}\exp\left\{-\frac{rT'}{2}(\hat{a}-\mu')^2\right\} \quad , \tag{5.68}$$

where

$$\mu' = (\mu T+\textstyle\sum y_i y_{i-1})\Big/(T+\sum y_{i-1}^2) \quad ,$$

$$T' = T+\textstyle\sum y_{i-1}^2 \quad ,$$

$$w' = w+\frac{N}{2} \quad ,$$

$$\ell' = \ell+\frac{1}{2}(\mu^2 T+\textstyle\sum y_i^2 - T'\mu'^2) \quad .$$

The conditional density for the parameter $\hat{a}$ is

$$f(\hat{a}\ r\ \underline{Y}_N,y_0)\ \sim\ N(\mu',\sigma'^2) \quad ,$$

where

$$\mu' = \frac{\mu T+\sum y_i y_{i-1}}{T+\sum y_{i-1}^2} \quad , \quad \sigma'^2 = \frac{\sigma^2}{T+\sum y_{i-1}^2} \quad .$$

The density for r is

$$f(r)\ \propto\ r^{w'-1}e^{-\ell' r} \quad , \tag{5.69}$$

which is gamma with parameters ℓ', w'. More meaning is attached to these results if a few definitions are made, that is let

$$N\hat{R}_0 = \sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} y_{i-1}^2 - y_0^2 + y_N^2 \quad ,$$

$$N\hat{R}_1 = \sum_{i=1}^{N} y_i y_{i-1} \quad .$$

Using these definitions in μ' and σ'^2 , the resulting forms are

$$\mu_a' = \frac{\mu T + N\hat{R}_1}{T + N\hat{R}_0 + y_0^2 - y_N^2} \xrightarrow{N \text{ large}} \frac{\hat{R}_1}{\hat{R}_0} \quad , \tag{5.70}$$

$$\sigma_a'^2 = \frac{\sigma^2}{T + \sum y_{i-1}^2} = \frac{\sigma^2}{T + N\hat{R}_0 + y_0^2 - y_N^2} \xrightarrow{N \text{ large}} \frac{\sigma^2}{N\hat{R}_0} \to 0, \tag{5.71}$$

where μ_a' and $\sigma_a'^2$ are the parameters in the Gaussian post density of the random variable $\hat{a}$. Using these same definitions in the parameters for the gamma density, the parameters for the post density of r are

$$w' = w + \frac{N}{2} \quad , \tag{5.72}$$

$$\ell' = \ell + \frac{1}{2}\left\{\mu^2 T + \sum y_i^2 - \frac{(\mu T + \sum y_i y_{i-1})^2}{T + \sum y_{i-1}^2}\right\}$$

$$= \ell + \frac{1}{2}\left\{\mu^2 T + N\hat{R}_0 - \frac{(\mu T + N\hat{R}_1)^2}{T + N\hat{R}_0 + y_0^2 - y_N^2}\right\} \quad . \tag{5.73}$$

After sufficiently large numbers of observations, these parameters reduce to

$$w' \simeq \frac{N}{2} \quad ,$$

$$\ell' \simeq \frac{1}{2} N\hat{R}_0 - \frac{(N\hat{R}_1)^2}{N\hat{R}_0} = \frac{N}{2}\frac{(\hat{R}_0^2 - \hat{R}_1^2)}{R_0} \quad .$$

The mean and variance for r and σ^2, where $r = 1/\sigma^2$, are

$$E(r) = \frac{w'}{\ell'} \quad , \quad \text{var } (r) = \frac{w'}{\ell'^2}$$

$$E(\sigma^2) = \frac{\ell'}{w'} = \frac{\hat{R}_0 - \hat{R}_1}{\hat{R}_0} \quad ,$$

and

$$\text{Var}(\sigma^2) = \frac{\ell'^2}{w'^3} = \frac{\hat{R}_0 - \hat{R}_1}{N\hat{R}_0} \quad .$$

The mean values for the parameters $\hat{a}$ and σ^2 with the interpretation provided are precisely the results that would be obtained for these parameters from the use of the Durbin-Levinson procedure mentioned in

Chapter III. The parameter estimation problem for the M=2 autoregressive process is carried out in Appendix C.

5.6 Likelihood Ratio Functions

The next chapter will deal with situations where the number of observations (N) becomes arbitrarily large. In this case, probability density functions do not in general, exist, either because the integral required to find the normalizing constant does not exist, or because the density function is not defined over the entire range of the random variable. This can be illustrated by the example presented below. In this example, the observations vector $\underline{Y} = (y_1, y_2, \ldots y_N)$, is taken from a jointly Gaussian distribution,

$$y_i \sim N(0, \lambda_i^2)$$

with

$$\lambda_i^2 = 3V \cdot 4^{-i} \quad , \; V > 0 \tag{5.74}$$

and

$$E(\underline{Y}^T \underline{Y}) = E\left(\sum_{i=1}^{N} \lambda_i^2 \right) \quad ,$$

$$= 3V \sum_{i=1}^{N} 4^{-i} \quad . \tag{5.75}$$

Evaluating the series in equation (5.75), find

$$\lim_{N \to \infty} \sum_{i=1}^{N} 3V \cdot^{-i} = V \quad ,$$

which shows the random vector to have finite power. Upon forming the N dimensional density function, the result is

$$f(\underline{Y}_N) = \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\Pi}} \lambda_i^{-1} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N} \frac{y_i^2}{\lambda_i^2} \right\}$$

$$f(\underline{Y}_N) = \left(\frac{1}{2\Pi 3V}\right)^{N/2} \cdot \prod_{i=1}^{N} 2^i \cdot \exp\ -\frac{1}{6V} \sum_{i=1}^{N} 4^i y_i^2$$

$$= \left(\frac{1}{2\Pi 3V}\right)^{N/2} \cdot 2^{\{N(N+1)\}/2} \cdot \exp\ -\frac{1}{6V} \sum_{i=1}^{N} 4^i y_i^2$$

$$= \left(\frac{2^N}{3\Pi V}\right)^{N/2} \exp\ -\frac{1}{6V} \sum_{i=1}^{N} 4^i y_i^2 \qquad (5.76)$$

Now, let $\underline{Y}_N = \underline{0}$ and evaluate the limit

$$\lim_{N\to\infty} f(\underline{0}_N) = \lim_{N\to\infty} \left(\frac{2^N}{3\Pi V}\right)^{N/2} \qquad (5.77)$$

Eventually, the condition $2^N > 3V$ is satisfied and the limit does not exist, hence the density function is not defined at the origin. So then, rather than dealing with the density functions, a generalization (the likelihood ratio function) of the density function is used. The likelihood ration function $\left(\Lambda(\underline{Y})\right)$ is defined as the ratio of two probability measures

$$\Lambda(\underline{Y}_N) = \frac{P(\underline{Y}_N|*)}{P(\underline{Y}_N|\cdot)} = \frac{\mu_1}{\mu_0} \quad . \qquad (5.78)$$

If the measure $P(\underline{Y}|\cdot)$ is zero for some value of $\underline{Y}$ common to μ_1 and μ_0, with $\mu_1 > 0$, then

$$\Lambda(\underline{Y}_N) \triangleq \infty \quad .$$

However, if μ_1 and μ_0 are zero, but the density function $f(\underline{Y}|\cdot)$ is non zero, then the likelihood ratio function is

$$\Lambda(\underline{Y}_N) = \frac{f(\underline{Y}_N|*)}{f(\underline{Y}_N|\cdot)} = \frac{d\mu_1}{d\mu_0} \quad . \qquad (5.79)$$

If the density function $f(\underline{Y}_N|\cdot)$ is zero for some values common to $d\mu_1$ and $d\mu_0$, then

$$\Lambda(\underline{Y}_N) \triangleq \infty \quad .$$

With the likelihood ratio function defined, there is now a vehicle for dealing with the infinite dimensional observation vector. This vehicle consists of forming the finite dimensional likelihood ratio $\left(\Lambda(\underline{Y}_N)\right)$ and passing the limit as N goes to infinity. To perform this limiting operation and obtain a useful result, there must be some quarantee that this limit will exist. The Martingale Convergence Theorem (Wong, 1971) provides a necessary and sufficient condition for the limit

$$\lim_{N\to\infty} \Lambda(\underline{Y}_N) = \lambda\left(y(t)\right) \tag{5.80}$$

to exist. For Gaussian measures, this limit will exist if the two measures μ_0 and μ_1 can be shown to be equivalent. Before the conditions necessary for this equivalence are established, it is important to see how the likelihood ratio affects Bayes Rule.

Since the likelihood ratio is a generalization of the probability density function, it is suggested that with likelihood ratios a more general specification of Bayes Rule is possible. This generalization starts with N dimensional representation

$$\Lambda(\underline{Y}_N|\theta) = \frac{f(\underline{Y}_N|\theta)}{f(\underline{Y}_N|\theta_0)} \quad , \tag{5.81}$$

where θ_0 is any fixed value of the permissible range of the parameter θ such that a likelihood ratio exists. Using the likelihood ratio in Bayes Rule provides the result

$$f(\theta|\underline{Y}_N) = \frac{f(\underline{Y}_N|\theta)f(\theta)}{\int f(\underline{Y}_N|\theta)f(\theta)d\theta}$$

$$f(\theta|\underline{Y}_N) = \frac{\Lambda(\underline{Y}_N|\theta)f(\underline{Y}_N|\theta_0)f(\theta)}{\int\Lambda(\underline{Y}_N|\theta)f(\underline{Y}_N|\theta_0)f(\theta)d\theta}$$

$$= \frac{\Lambda(\underline{Y}_N|\theta)f(\theta)}{\int\Lambda(\underline{Y}_N|\theta)f(\theta)d\theta} \qquad (5.82)$$

If the limiting form of the likelihood ratio exists, Bayes Rule takes the form

$$f\big(\theta|Y(t)\big) = \lambda\big(y(t)|\theta\big)f(\theta) \Big/ \int\lambda\big(y(t)|\theta\big)f(\theta)d\theta \quad . \qquad (5.83)$$

The concepts of sufficient statistics and reproducing priors to form a post density, are still applicable when dealing with the likelihood ratio $\lambda\big(y(t)|\theta\big)$, since all that is required is the application of Bayes Rule. Generalizing the Factorization Criterion, a mapping $K(\cdot)$ is a sufficient statistic if and only if

$$\lambda\big(y(t)|\theta\big) = g\Big[K\big(y(t),\theta\big)\Big]G\big(y(t)\big) \quad , \qquad (5.84)$$

where $g\Big[k,\theta\Big]$ is of fixed dimension. Placing this into the context of Bayes Rule the post density is

$$f\big(\theta|y(t)\big) = \frac{g\Big[K,\theta\Big]\,G\big(y(t)\big)f(\theta)}{\int g\Big[K,\theta\Big]G\big(y(t)\big)f(\theta)d\theta}$$

$$= \frac{g\Big[K,\theta\Big]f(\theta)}{\int(\text{numerator})\ d\theta} \quad . \qquad (5.85)$$

The reproducing priors property can be demonstrated just as in the finite dimensional situation, where the prior is chosen to satisfy

$$f(K=k,\theta) = g\Big[K=k,\theta\Big] \Big/ \int g\Big[K=k,\theta\Big]d\theta \qquad (5.86)$$

After the prior is chosen in this manner, the sequential form of Bayes Rule is applied to show the reproducing property.

5.7 Equivalence of Gaussian Measures

In seeking the limit of the finite dimensional likelihood

ratio, the main issue is ensuring that the two measures under consideration are equivalent. Since, this study will be concerned with Gaussian random processes, the question of equivalence can be answered in terms of the mean value and covariance functions. Demonstrating the equivalence of two Gaussian measures is the same as showing that the Gaussian noise-in-noise detection problem is nonsingular (error free detection is not possible). The Gaussian noise-in-noise detection problem (Kailath, 1970) with the hypothesis

$$H_0 : x(t) = n(t) \quad , \tag{5.87}$$

$$H_1 : x(t) = z(t) + n(t) \quad , \tag{5.88}$$

where $n(t)$ is a zero mean, white Guassian noise (WGN) and $z(t)$ is a Gaussian process, is non singular if

$$E\left[n(t)z(s)\right] = 0 \quad , \; t > s \tag{5.89}$$

and

$$\int_0^T E\left[z^2(t)\right]dt < \infty \quad . \tag{5.90}$$

A more general statement of equivalence (Shepp, 1966) is possible by restating the problem as one with the hypotheses

$$H_0 : x(t) \text{ is a zero mean WGN process with covariance } \delta(t-s) \quad , \tag{5.91}$$

$$H_1 : x(t) \text{ is Gaussian with mean function } m(t) \text{ and covariance } R(t,s) \quad . \tag{5.92}$$

This problem is nonsingular if and only if the covariance $\left[R(t,s)\right]$ can be written as

$$R(t,s) = \delta(t-s) + K(t,s) \quad , \tag{5.93}$$

where $K(t,s)$ satisfies

$$\iint_0^T K^2(t,s)dtds < \infty \quad . \tag{5.94}$$

When the function K(t,s) is a covariance function, this problem reduces to the problem previously specified. When K(t,s) is not a covariance function, it can be shown (Kailath, 1970) that the problem can be restated in signal plus noise terms.

As an example of determining the equivalence of two Gaussian measures, the following problem is considered. The hypotheses of the problem are

$$H_1 : x(t) \text{ is a zero mean Gaussian process with}$$
$$R_{H_1}(t,u) = \frac{\sigma^2}{2p} e^{-p|t-u|} , \tag{5.95}$$

$$H_0 : x(t) \text{ is a zero mean Gaussian process with}$$
$$R_{H_0}(t,u) = \frac{\alpha^2}{2b} e^{-b|t-u|} \tag{5.96}$$

To apply the test for nonsingularity specified by equation (5.93), a filter is found such that on hypothesis H_0

$$\delta(t-u) = \iint h_{w_0}(t,x) R_{H_0}(s,t) h_{w_0}(u,y)\, dx dy . \tag{5.97}$$

The impulse response (h_{w_0}) of this whitening filter is

$$h_{w_0} = \frac{1}{\alpha}\left(\dot{\delta}(t-u) + b\,\delta(t-u)\right) , \tag{5.98}$$

and the transfer function is

$$H_{w_0}(j\omega) = \frac{j\omega + b}{\alpha} , \tag{5.99}$$

which has been determined from

$$|H_{w_0}(j\omega)|^2 = \frac{1}{S_{H_0}(\omega)} = \frac{\omega^2 + b^2}{\alpha^2} . \tag{5.100}$$

The response under H_1 of this filter is given by

$$R^*_{H_1}(t,u) = \iint h_{w_0}(t,x) R_{H_1}(x,y) h_{w_0}(u,y)\, dx\, dy$$

$$R^*_{H_1}(t,u) = \frac{1}{\alpha^2}\iint \left(\dot{\delta}(t-x) + b\,\delta(t-x)\right)\frac{\sigma^2}{2p}\,e^{-p|x-y|}\left(\dot{\delta}(u-y) + b\,\delta(u-y)\right)dx\,dy$$

$$= \frac{1}{\alpha^2}\int \left(\dot{\delta}(t-x) + b\,\delta(t-x)\right)\left(\frac{\partial}{\partial u}R_{H_1}(x,u) + bR_{H_1}(x,u)\right)dx$$

$$= \frac{1}{\alpha^2}\left[\frac{\partial^2}{\partial t\,\partial u}R_{H_1}(t,u) + b\,\frac{\partial}{\partial u}R_{H_1}(t,u) + b\,\frac{\partial}{\partial t}R_{H_1}(t,u) + b^2R_{H_1}(t,u)\right]. \quad (5.101)$$

As a result of this filtering operation, equivalence of the two measures is insured if $R^*_{H_1}(t,u)$ can be written as

$$R^*_{H_1} = \delta(t-u) + K^*_{H_1}(t,u)$$

where $K^*_{H_1}(t,s)$ is a square integrable function. The only non-square integrable part of $R^*_{H_1}(t,u)$ occurs in the first term, which is

$$\frac{1}{\alpha^2}\frac{\partial^2}{\partial t\,du}R^*_{H_1}(t,u) = \frac{1}{\alpha^2}\left[\sigma^2\,\delta(t-u) - \frac{\sigma^2 p}{2}\,e^{-p|t-u|}\right]. \quad (5.102)$$

It is apparent that if $\alpha^2 = \sigma^2$ then the condition for non-singularity is satisfied. Since the problem is symmetric, the problem is repeated by whitening on H_1. The result of this operation provides the same condition ($\alpha^2=\sigma^2$) to insure non-singularity.

Another, more easily applied test for equivalence of two measures is one given by Feldman (1958). Under the condition that the Gaussian measures μ_0 and μ_1 have covariances which admit rational spectral densities (degree of numerator polynomial strictly less than the degree of the denominator polynomial) then a necessary and sufficient condition for the equivalence of the measures is

$$\lim_{\omega \to \infty} \frac{S_{\mu_1}(\omega^2)}{S_{\mu_0}(\omega^2)} = 1 \quad , \tag{5.103}$$

where

$$S_{\mu_1}(\omega^2) = \int R_{\mu_1}(T) e^{-j\omega T} \, d\omega \quad , \tag{5.104}$$

$$S_{\mu_0}(\omega^2) = \int R_{\mu_0}(T) e^{-j\omega T} \, d\omega \quad , \tag{5.105}$$

Application of this test to the previous problem provides the same condition for equivalence ($\alpha^2 = \sigma^2$). Slepian (1958) using a theorem due to Baxter (1956) provided a sufficient condition for non-singularity of the problem just considered. That is, assume that under both hypotheses, one has a Gaussian process with mean function of bounded variation and covariances R_0 and R_1 that have bounded second derivatives $\partial^2/\partial t \partial s\left(R_i(t,s)\right)$ on $(0,T)X(0,T)$ except at $t = s$. Under these conditions, define

$$R_i' = \frac{\partial}{\partial s} R_1(t,s) \quad , \tag{5.106}$$

$$\sigma_i^2(t) = R_i'(t,s^-) - R_i'(t,s^+) \quad , \tag{5.107}$$

then a sufficient condition for non-singularity is

$$\sigma_0^2(t) = \sigma_1^2(t) \quad . \tag{5.108}$$

In summary, the problem of non-singularity (equivalence of measures), where WGN is present, can be resolved by insuring that the same intensity of WGN is present on both hypotheses, and the signal process has finite power. Without WGN, equivalence is assured if the high frequency behavior under both hypotheses is the same.

CHAPTER VI

THE CONTINUOUS ESTIMATION PROBLEM

6.1 Introduction

In dealing with discrete process representations, it became evident that an intimate relationship exists between the parameters of the difference equation and the coefficients in its spectral density. It was clear that specification of a finite set of parameters was sufficient to characterize either representation. Further, the discrete time representation allowed specification of a probability density function for the parameters and observations. The other aspect of note was that with a difference equation model, a Markov dependence existed which facilitated the use of Bayes Rule.

This chapter concentrates on obtaining the spectral estimate of a continuous random process. The estimate will be obtained through a limiting procedure performed on an equivalent discrete time representation of the process. The parameters that specify the spectrum will be characterized by a probability density function, rather than a point estimate form. This is a major departure from other approaches to this problem, since there is no "goodness" criterion which is to be satisfied by the solution process. The problem will be attempted in a manner that "holds off" specification of a goodness criterion, and instead seeks a probability density function which in turn could be applied to any arbitrary criterion.

The processes under consideration are a finite part of a zero mean stationary process with a spectral density of the form

$$S(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M}(\omega^2 + p_i^2)} \quad , \; p_i \neq p_j \quad . \tag{6.1}$$

Processes with this spectral density can be modeled as the solution of a linear Mth order stochastic differential equation excited by white Gaussian noise. The process so modeled is Mth order, Markov, and Gaussian, hence the set of parameters $\{\sigma^2, p_i|_{i=1}^{M}\}$, provide a complete description. The sample paths of this process has rather nice properties. The sample paths are almost everywhere uniformly continuous, bounded and M-1 times differentiable.

The estimation problem consists of finding the set of parameters $\{p_i,\sigma^2\}$ in the spectral representation, given that the observations have an all pole spectral density. Initially, the order of the process (M) will be assumed to be known, but this assumption will be removed in the next chapter. The parameters $\{p_i,\sigma^2\}$ are considered to be random variables, and a post density for these parameters will be sought via the application of Bayes Rule. The post density will be found through the use of a limiting procedure performed on a discrete representation which has the same second order properties. As mentioned in the last chapter, there are several technical difficulties involved in performing this procedure analytically.

6.2 Preliminary Requirements of the Estimation Problem

The first difficulty that arises is the evaluation of the normalizing factor (integral) in Bayes Rule. That is, in evaluating a post density for a parameter, say θ, by use of

$$f(\theta|\underline{X}_k) = \frac{f(x_k|\underline{X}_{k-1}\theta)\ f(\theta|\underline{X}_{k-1})}{\int f(x_k|\underline{X}_{k-1}\theta)\ f(\theta|\underline{X}_{k-1})d\theta} \tag{6.2}$$

the attainment of the result of $\left(f(\theta|\underline{X}_k)\right)$ analytically depends on being able to evaluate the integral in the denominator. This problem is alleviated if one chooses the prior density on the parameter (θ) out of the class of priors that reproduces. In so doing, the post density is a member of the same class, and the integral evaluation reduces to identifying the normalizing constant for that class.

A second difficulty arises when attempting to take the limit of the probability density function as the samples become dense in the observation interval. It was pointed out in the last chapter, that such a limit does not, in general, exist. This difficulty is circumvented by the introduction of the Likelihood Ratio function (LR) where the LR is defined as

$$\Lambda(\underline{X}_k,\theta) \triangleq f(\underline{X}_k|\theta) \ / \ f(\underline{X}_k|\theta_0), \tag{6.3}$$

where θ_0 is any member of the space Ω_θ such that the LR exists. With the LR function defined, Bayes Rule takes the form

$$f(\theta|\underline{X}_k) = \frac{\Lambda(\underline{X}_k,\theta)\ f(\theta)}{\int \Lambda(\underline{X}_k,\theta)\ f(\theta)d\theta}\ , \tag{6.4}$$

and now attention is directed towards the conditions required for the convergence of

$$\lim_{k\to\infty} \Lambda(\underline{X}_k,\theta) = \lim_{k\to\infty} \frac{f(\underline{X}_k|\theta)}{f(\underline{X}_k|\theta_0)} \to \lambda(x_t,\theta)\ . \tag{6.5}$$

The convergence of this limit can be tested via the application of equation (5.70). That is, the process under investigation is assumed to be an Mth order Gaussian process with a spectral density of the form

$$S_x(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M} (\omega^2 + p_i^2)} \quad , \; p_i \neq p_j \tag{6.6}$$

a correlation function

$$R_x(T) = \sigma^2 \sum_{i=1}^{M} \left\{ \frac{e^{-p_i|T|}}{2p_i \prod^{M}(p_k^2 - p_i^2)} \right\} \quad , \; p_i \neq p_k \quad , \tag{6.7}$$

and measure μ. To determine the LR function, it is necessary to assume the existence of another measure μ_0 with

$$S_x^0(\omega) = \frac{b^2}{\prod_{i=1}^{N} (\omega^2 + q_i^2)} \quad , \quad q_i \neq q_j \; , \tag{6.8}$$

$$R_x^0(T) = b^2 \sum_{i=1}^{N} \left\{ \frac{e^{-q_i(T)}}{2q_i \prod_{i=1}^{N} (q_k^2 - q_i^2)} \right\} \quad , \; q_i \neq q_k \; . \tag{6.9}$$

For the existence of the limit

$$\lim_{k\to\infty} \Lambda(\underline{X}, \cdot) = \lim_{k\to\infty} \frac{f(\underline{X}_k|\cdot)}{f_0(\underline{X}_k|\cdot)} = \lim_{k\to\infty} \frac{d\mu}{d\mu_0} \tag{6.10}$$

the two measures μ and μ_0 must be equivalent. The necessary and sufficient conditions for this equivalence are found from the satisfaction of

$$\lim_{\omega\to\infty} \frac{S_x(\omega)}{S_x^0(\omega)} = 1 \; .$$

The conditions that satisfy this requirement are observed to be M=N, and $b^2 = \sigma^2$. This means that the model that generates the observations process $(\underline{X}_k)$ must be of the same order (M=N) under both measures and have the same intensity driving function ($b^2 = \sigma^2$).

6.3 The M=1 Estimation Problem

The second order properties of the M=1 process are characterized by

$$R_x(\tau) = \frac{\sigma^2}{2p} e^{-p|\tau|} , \tag{6.11}$$

$$S_x(\omega) = \frac{\sigma^2}{\omega^2 + p^2} . \tag{6.12}$$

In consonance with the discussion in Chapter III, the equivalent discrete model (same second order properties) is given by

$$x_k = (1 - \Delta p)x_{k-1} + \zeta_k , \tag{6.13}$$

where Δ is selected so that

$$(1 - \Delta p)^{t/N} \to e^{-pt} , \quad t=N\Delta \quad . \qquad \begin{matrix} N\to\infty \\ \Delta\to 0 \end{matrix} \tag{6.14}$$

The driving sequence ζ_k is white zero mean, Gaussian with variance $\sigma^2\Delta$ and uncorrelated with x_0. The parameters (p,σ^2) are considered to be random variables, and a post density function will be sought to describe their behavior. To begin, the conditional density function for the observations is determined. Since the observations are Gaussian, the one step transition density is specified once the conditional mean and variance are determined. From equation (6.13) and the knowledge that

$$E(\zeta_k\zeta_n) = \begin{cases} \sigma^2\Delta & k = n \\ 0 & k \neq n \end{cases} \tag{6.15}$$

and $E(\zeta_n x_k) = 0$, for $n > k$, the conditional mean and variance are

$$E(x_k|x_{k-1}) = x_{k-1}(1-\Delta p), \tag{6.16}$$

$$\text{Var}\,(x_k|x_{k-1}) = \sigma^2\Delta . \tag{6.17}$$

Now, the one step density is

$$f(x_k|x_{k-1}) \propto \exp\{ -\frac{1}{2\sigma^2\Delta} (x_k-(1-\Delta p)x_{k-1})^2\} . \quad (6.18)$$

Using the properties of the M=1 process under consideration, the N step density is

$$f(\underline{X}_N|x_0,\sigma^2,p) \propto \exp -\frac{1}{2\sigma^2\Delta} \sum_{i=1}^{N} (x_i-(1-\Delta p)x_{i-1})^2 . \quad (6.19)$$

The formation of the LR function will be complete if it is now assumed that there is another density function

$$f(\underline{X}_N|x_0,n^2,p_0) \propto \exp -\frac{1}{2n^2\Delta} \sum_{i=1}^{N} (x_i-(1-\Delta p_0)x_{i-1})^2 , \quad (6.20)$$

that corresponds to the correlation function

$$R_x^o(T) = \frac{n^2}{2p_0} e^{-p_0|T|} , \quad (6.21)$$

and spectral density

$$S_x^0(\omega) = \frac{n^2}{\omega^2 + p_0^2} . \quad (6.22)$$

The Likelihood Ratio function (LR) is now formed as

$$\Lambda(\underline{X}_N|x_0,p,\sigma^2) = \frac{f(\underline{X}_N|x_0,\sigma^2,p)}{f(\underline{X}_N|x_0,n^2,p_0)} , \quad (6.23)$$

which provides the result

$$\Lambda(\underline{X}_N|x_0,p,\sigma^2) = \frac{\exp -\frac{1}{2\sigma^2\Delta} \sum_{i=1}^{N} (x_i-(1-\Delta p)x_{i-1})^2}{\exp -\frac{1}{2n^2\Delta} \sum_{i=1}^{N} (x_i-(1-\Delta p_0)x_{i-1})^2} , \quad (6.24)$$

where p_0 is any fixed element of the parameter space, so that the LR is defined. The estimation procedure now turns to investigate the behavior as $N \to \infty$ ($\Delta \to 0$). To insure the existence of this limit, the requirement left to be satisfied is that $\sigma^2 = n^2$. Incorporating this

into equation (6.24), and letting $p_0 = 0$, the LR function takes the form

$$\Lambda(\underline{X}_N|x_0,p,\sigma^2)= \exp -\frac{1}{2\sigma^2\Delta} \sum_{i=1}^{N} \left(x_i^2-2(1-\Delta p)x_i x_{i-1} +(1-\Delta p)^2 x_{i-1}^2\right)$$

$$-\left(x_i^2-2x_i x_{i-1} + x_{i-1}^2\right). \qquad (6.25)$$

Performing the reduction that is indicated, equation (6.25) is now

$$\Lambda(\underline{X}_N|x_0,\sigma^2,p)= \exp -\frac{1}{2\sigma^2\Delta} \left\{ \sum_{i=1}^{N} \Delta^2 p^2 x_{i-1}^2 + \sum_{i=1}^{N} 2\Delta p x_i x_{i-1} \right.$$

$$\left. - \sum_{i=1}^{N} 2\Delta p x_{i-1}^2 \right\} . \qquad (6.26)$$

At this point, it is convenient to expand certain factors, before taking the limit $N \to \infty$ $(\Delta \to 0)$. That is, expand the cross product term

$$2 \sum_{i=1}^{N} x_i x_{i-1} = \sum_{i=1}^{N} (x_i^2 + x_{i-1}^2) - \sum_{i=1}^{N} (x_i - x_{i-1})^2$$

$$= 2 \sum_{i=1}^{N} x_{i-1}^2 - \sum_{i=1}^{N} (x_i - x_{i-1})^2 - (x_0^2 - x_N^2) \quad .(6.27)$$

The LR function with this expansion inserted is now of the form

$$\Lambda(\underline{X}_N|x_0,\sigma^2,p) = \exp\left\{ - \frac{1}{2\sigma^2}\left[p^2 \sum_{i=1}^{N} x_{i-1}^2 \Delta - p \sum_{i=1}^{N} (x_i - x_{i-1})^2 \right.\right.$$

$$\left.\left. - p(x_0^2 - x_N^2)\right]\right\} \quad .(6.28)$$

At this point, the limit $N \to \infty$ will be taken to obtain the continuous LR function. The limit will be defined for a growing number of observations (N) in a fixed observation interval (T) where $T = N\Delta$. Using the sample path properties of the M=1 model, the observations are almost surely sample path continuous, and hence, the first sum in equation (6.28) converges to an integral. The second sum in equation

(6.28) under the stated limiting conditions will also converge to a known quantity which will be justified in the next section. For now, the limit of this sum will be defined as

$$\lim_{\substack{N\to\infty \\ \Delta\to 0}} \sum_{i=1}^{N} (x_i - x_{i-1})^2 \to \sigma^2 T \quad . \tag{6.29}$$

With these considerations in mind, the limit of the likelihood ratio function is

$$\lambda(x_t|x_0,p,\sigma^2) = \exp\left\{-\frac{1}{2\sigma^2}\left[p^2\int_0^T x^2(t)dt - p(x_0^2 - x_T^2) - p\sigma^2 T\right]\right\} . \tag{6.30}$$

Having found the limiting form of the LR function, Bayes Rule can be applied with the conjugate prior for the parameter p. The sufficient condition for the existence of such a prior was shown to be the existence of a finite dimensional sufficient statistic. Application of the Factorization Criterion to equation (6.30) shows the sufficient statistic to be the set of measurements

$$\int_0^T x^2(t)dt \ , \ x_0^2 \ , \ x_T^2 \quad .$$

Choosing the prior for p as $N(\mu,n^2)$, the application of Bayes Rule provides the result

$$f(p|x_t,x_0,\sigma^2) \propto \exp^{-\frac{1}{2\sigma^2}} \left\{p^2\int_0^T x^2(t)dt - p\left[\sigma^2 T+(x_0^2 - x_T^2)\right]\right\} \cdot \exp\left\{-\frac{1}{2n^2}(p-\mu)^2\right\} \quad . \tag{6.31}$$

Rearranging the exponential factors, and completing the square in p, the result is

$$f(p|x_t,x_0,\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\left(\int_0^T x^2dt + \frac{\sigma^2}{n^2}\right)\right.$$

$$\left.\left[p - \frac{\left(\sigma^2T + (x_0^2 - x_T^2) + \frac{2\mu\sigma^2}{n^2}\right)}{2\int_0^T x^2dt + \frac{\sigma^2}{n^2}}\right]^2\right\}. \quad (6.32)$$

From the form of the right hand side of equation (6.32), the post density is recognized as Gaussian. Hence, the normalizing constant can be determined by inspection. The post density for p is

$$N \sim \left(\frac{\sigma^2T + (x_0^2-x_T^2) + \frac{2\mu\sigma^2}{n^2}}{2\left(\int_0^T x^2(t)dt + \frac{\sigma^2}{n^2}\right)}, \quad \frac{\sigma^2}{\int_0^T x^2(t)dt + \frac{\sigma^2}{n^2}}\right),$$

where μ is the prior mean and n^2 is the prior variance. Under the assumption that $n^2 >> \sigma^2$, the post density becomes

$$N \sim \left(\frac{\sigma^2T + (x_0^2 - x_T^2)}{2\int_0^T x^2(t)dt}, \quad \frac{\sigma^2}{\int_0^T x^2(t)dt}\right),$$

which is precisely the result obtained if a locally uniform prior has been chosen.

The result just achieved, analytically, was based on a Gaussian prior. A real Bayesian approach would consist of using a Gaussian prior truncated to the positive side of the real line. This choice is made by virtue of the requirement that the quantity in equation (6.14) approaches e^{-pt}. Although this truncated distribution would reproduce, the result achieved is difficult to handle analytically and provides little insight. Rather than handling the constraint at the beginning of the problem, it is more efficacious to

solve the unconstrained problem first and then to handle the constraint by determining the modifying factor, equation (6.14), to yield the constrained solution.

6.4 The M=2 Estimation Problem

The case of most interest is the M=2 process with spectral density and correlation function of the form

$$R_x(T) = \frac{\sigma^2}{2p_1p_2(p_2^2 - p_1^2)} \left\{ p_2 e^{-p_1|T|} - p_1 e^{-p_2|T|} \right\} , \tag{6.33}$$

$$S_x(\omega) = \frac{\sigma^2}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)} . \tag{6.34}$$

The model that has the specified second order properties is characterized by the difference equation

$$x_1(n+1) = \left(2 - \Delta(p_1 + p_2)\right)x_1(n) - \left(1 - \Delta(p_1 + p_2) + \Delta^2 p_1 p_2\right)x_1(n-1) + \Delta \cdot \zeta(n-1) , \tag{6.35}$$

where $\zeta(n)$ is a zero mean white Gaussian sequence, that is independent of the initial condition vector $\underline{X}_0^T = (x_0, x_{-1})$. For this difference equation to be representative, the sampling rate must satisfy the condition $(p_1 + p_2)\Delta << 1$, so that

$$1 - e^{-(p_1 + p_2)\Delta} \simeq (p_1 + p_2)\Delta .$$

The procedure for finding the limiting form of the likelihood ratio function is the same as the M=1 case, with more details. This procedure is carried out in Appendix D, only the results of this procedure will be illustrated here. The conditional density for the observations generated by the difference equation model is

$$f(\underline{X}_N|x_0,x_{-1},q_1,\ q_2,\ \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2\Delta^3}\sum_{i=1}^{N}\left[x_i-(2-\Delta q_2)x_{i-1}\right.\right.$$
$$\left.\left.+(1-\Delta q_2+\Delta^2 q_1)x_{i-2}\right]^2\right\}, \quad (6.36)$$

where $q_2 = (p_1 + p_2)$, $q_1 = p_1 p_2$ and $\Delta = T/N$. Forming the likelihood ratio and performing the limiting operation, the result is

$$\lambda(x_t|\underline{X}_0,\sigma^2,q) = \exp\left\{-\frac{1}{2\sigma^2}\left[q_1^2\int x^2(t)dt + q_1q_2(x_T^2-x_0^2)\right.\right.$$
$$+q_2^2\int\dot{x}^2(t)dt - q_2(\sigma^2T + \dot{x}_0^2 - \dot{x}_T^2)$$
$$\left.\left.-2q_1\int\dot{x}^2(t)dt - 2q_1(x_T\dot{x}_T - x_0\dot{x}_0)\right]\right\} \quad (6.37)$$

Noting the form of the likelihood ratio function, it can be readily discerned that the finite dimensional sufficient statistics are the set of measurements

$$\int x^2(t)dt,\quad \int\dot{x}^2(t)dt,\ x_0,\ \dot{x}_0,\ x_T,\ \dot{x}_T \quad .$$

Using a locally uniform prior $\underline{q}$, and assuming that the end conditions have a negligible effect, the post density found from the application of Bayes Rule is

$$f(\underline{q}|x_t,\underline{x}_0,\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\left[\int_0^T x^2(t)dt\left(q_1-\frac{\int\dot{x}^2(t)dt}{\int x^2(t)dt}\right)^2\right]\right\} \quad (6.38)$$
$$\cdot\exp\left\{-\frac{1}{2\sigma^2}\left[\int_0^T\dot{x}^2(t)dt\left(q_2-\frac{\sigma^2T}{2\int\dot{x}^2(t)dt}\right)^2\right]\right\}.$$

From this formulation, the joint post density is seen to be normal, independent with parameters

$$\bar{q}_1 = \frac{\int\dot{x}^2(t)dt}{\int x^2(t)dt} \quad ; \quad \text{Var } q_1 = \frac{\sigma^2}{\int x^2(t)dt} \quad (6.39)$$

$$\bar{q}_2 = \frac{\sigma^2T}{2\int\dot{x}^2(t)dt} \quad ; \quad \text{Var } q_2 = \frac{\sigma^2}{\int\dot{x}^2(t)dt} \quad (6.40)$$

An interpretation for the parameters of the density is available upon use of the ergodicity of the stationary Gaussian process under consideration. That is interpreting the integrals as

$$\underset{T \to \infty}{\text{l.i.m.}}\ T^{-1}\int_0^T \dot{x}^2(t)dt \to R_{\dot{x}}(0)$$

$$\underset{T \to \infty}{\text{l.i.m.}}\ T^{-1}\int_0^T x^2(t)dt \to R_x(0)$$

Then the mean and variance take the form

$$\bar{q}_1 = \lim_{T \to \infty} \frac{TR_x(0)}{TR_{\dot{x}}(0)} \to p_1 p_2 = q_1$$

$$\bar{q}_2 = \lim_{T \to \infty} \frac{\sigma^2 T}{2TR_{\dot{x}}(0)} \to (p_1 + p_2) = q_2$$

$$\text{Var } q_1 = \lim_{T \to \infty} \frac{\sigma^2}{TR_x(0)} \to 0$$

$$\text{Var } q_2 = \lim_{T \to \infty} \frac{\sigma^2}{TR_{\dot{x}}(0)} \to 0$$

From these interpretations, the results are recognized as being unbiased and consistent. Applying these same considerations (envoking ergodicity) to the M=1 result shows that it has the same properties. the result achieved for the M=1, and M=2 estimation problem are in consonance with the work of Hajek (1962), which proceeds from a different starting point than presented here.

6.5 Estimation of the Random Parameter σ^2

To this point, the post densities presented as the final result were conditioned on the parameter σ^2, which is to be estimated or else known a priori. It is the purpose here to remove this conditioning

and to provide a means for determining this parameter. The vehicle used to perform this estimation is a theorem due to Baxter (1956). If the random process $x(t)$, $t\varepsilon[0,T]$ is Gaussian with a mean function that has a bounded first derivative and a continuous covariance function $R(t,s)$, with a bounded second derivative except at $t=s$, then

$$\lim_{n\to\infty} B_N = \lim_{n\to\infty} \sum_{k=1}^{2^n} \left(X(\frac{k}{2^n}) - X(\frac{k-1}{2^n})\right)^2 \underset{wp1}{\to} \int_0^T f(t)dt \ , \qquad (6.41)$$

where $f(t)$ is the jump in the first derivative of $R(t,s)$ defined as

$$f(t) = \lim_{s\to t^-} \frac{R(t,t) - R(s,t)}{t-s} - \lim_{s\to t^+} \frac{R(t,t) - R(s,t)}{t-s} \ . \qquad (6.42)$$

The existence of the first derivative at $t=s$ is not assumed, in fact it is not wanted, since this would make $f(t) \equiv 0$. Equation (6.41) indicates that as the sampling becomes dense in the observation interval ($\Delta = T/2^n$), the random variable B_N approaches its mean (the integral) almost surely. As is stands now, equation (6.41) is not rich in its meaning to the problems under consideration. To provide meaningfulness to these statements, attention will be focused on the M=1, and M=2 process being considered.

Begin by considering the M=1 process with correlation function

$$R_x(T) = \frac{\sigma^2}{2p} e^{-p|T|} \ , \qquad (6.43)$$

and complete the mean and variance of the random variable B_N. Through the application of the Chebyshev Inequality

$$Pr\{(B_N - \overline{B}_N) \geq \varepsilon\} \leq \frac{Var(B_N)}{\varepsilon^2} \ , \qquad (6.44)$$

and the Borel Cantelli lemma, it can be shown that

$$B_N \underset{N\to\infty}{\to} \overline{B}_N \qquad (6.45)$$

almost surely.

The significance of equation (6.45) is demonstrated through the computation of $\overline{B}_N$. Let $\underline{\Delta}$ be the first difference operator, and let δ be used for the time step Δ to avoid confusion with $\underline{\Delta}$, then

$$\underline{\Delta}x_k = x(\frac{k}{2^n}) - x(\frac{k-1}{2^n}) ,$$

$$a_{jk} = E(\underline{\Delta}x_j, \underline{\Delta}x_k) \text{ , for } j,k \in I^+ \text{with } 1 \leq j, k \leq 2^n ,$$

$$B_N = \sum_{k=1}^{2^n} \underline{\Delta}x_k^2 ,$$

and re-define the sampling interval to be $\delta = T/2^n$ and hence $t_k = kT/2^n$ for $k = 1,2, \ldots 2^n$. The mean of B_N is

$$\begin{aligned} E(B_N) &= \sum_k E(x_k - x_{k-1})^2 \\ &= \sum_k E(\underline{\Delta}x_k)^2 = \sum_k a_{kk} \\ &= 2\sum_k \left(R_x(0) - R_x(\delta)\right) . \end{aligned} \tag{6.46}$$

Under the fast sampling condition $p\delta << 1$, the mean becomes

$$\begin{aligned} E(B_N) &= 2 \sum_{k=1}^{2^n} \frac{\sigma^2}{2p}(1-e^{-p\delta}) \\ &= 2 \sum_{k=1}^{2^n} \frac{\sigma^2}{2p}(p\delta) + 0(\delta^2) \\ &= \sigma^2 \sum_{k=1}^{2^n} \delta + 0(\delta^2) , \end{aligned} \tag{6.47}$$

so that in the limit

$$\lim_{\substack{n \to \infty \\ \delta \to 0}} E(B_N) = \lim_{\substack{n \to \infty \\ \delta \to 0}} \sum_{k=1}^{2^n} \sigma^2\delta \to \int_0^T \sigma^2 dt = \sigma^2 T . \tag{6.48}$$

The significance of this result is that the measurement described by

the random variable B_N, has as its mean a scaled version of the random variable being estimated (σ^2). So now, motivation for showing that the random variable converges to its mean, has been provided. With this motivation, the computation of the variance is begun. The variance of B_N is computed as

$$E(B_N^2) = \sum_k \underline{\Delta}x_k^2 \cdot \sum_j \underline{\Delta}x_j^2 = \sum\sum_{kj} E(\underline{\Delta}x_k^2 \underline{\Delta}x_j^2)$$

$$= \sum_{j=k} E(\underline{\Delta}x_k^4) + 2\sum\sum_{j>k} E(\underline{\Delta}x_k^2 \underline{\Delta}x_j^2) \quad . \tag{6.49}$$

Since Δx_k is Gaussian, then

$$E(\underline{\Delta}x_k^4) = 3E(\underline{\Delta}x_k^2)^2 \triangleq 3a_{kk}^2 \quad ,$$

and

$$E(\underline{\Delta}x_k^2 \underline{\Delta}x_j^2) = E(\underline{\Delta}x_k^2)E(\underline{\Delta}x_j^2) + 2E(\underline{\Delta}x_k \underline{\Delta}x_j)$$

$$= a_{kk}\, a_{jj} + 2a_{jk}^2 \quad ,$$

so that

$$E(B_N^2) = \sum_{j=k} 3a_{kk}^2 + 2\sum\sum_{j>k} a_{kk}a_{jj} + 2a_{jk}^2 \quad . \tag{6.50}$$

The variance of B_N is then

$$\mathrm{Var}(B_N) = \sum_k 3a_{kk}^2 + 2\sum\sum_{j>k} a_{kk}a_{jj} + 2a_{jk}^2 - \sum_k a_{kk}^2 - 2\sum\sum_{j>k} a_{kk}a_{jj}$$

$$= 2\sum_{j=k} a_{kk}^2 + 4\sum\sum_{j>k} a_{jk}^2 \tag{6.51}$$

$$= 2\{\sum\sum_{jk} a_{jk}^2\} \quad . \tag{6.52}$$

Interpreting the constituent parts of equation (6.51) in terms of the given correlation provides the result

$$a_{kk}^2 = E(\underline{\Delta}x^2)^2 = 2\left(R_x(0) - R_x(\delta)\right)^2$$

$$= \sigma^4\delta^2 + \left(0(\delta^2)\right)^2 \quad , \tag{6.53}$$

$$a_{jk} = E(\underline{\Delta x}_k \underline{\Delta x}_j) = 2R_x(j-k) - R_x(j-k-1) - R_x(j-k+1), \quad (6.54)$$

where

$$R(j-k) = R_x\left((j-k)\delta\right) = \frac{\sigma^2}{2p} e^{-p(j-k)\delta}, \quad \text{for } j > k. \quad (6.55)$$

Using equation (6.55), under the condition $p\delta \ll 1$, in equation (6.54),

$$a_{jk} = E(\underline{\Delta X}_k \underline{\Delta X}_j) = \frac{\sigma^2}{2p}\left[2e^{-p\delta(j-k)} - e^{-p\delta(j-k+1)} - e^{-p\delta(j-k-1)}\right]$$

$$= \frac{\sigma^2}{2p}\left[e^{-p\delta(j-k)}(2 - 2\cosh p\delta)\right]$$

$$\cong \frac{\sigma^2}{2p}\left[e^{-p\delta(j-k)}(-p^2\delta^2)\right] = \frac{-\sigma^2 p\delta^2}{2}\left[e^{-p\delta(j-k)}\right] \quad (6.56)$$

$$a_{jk}^2 \leq \frac{\sigma^4\delta^4 p^2}{4} \quad (6.57)$$

The variance with the above reduction inserted is now in the form

$$\mathrm{Var}(B_N) \leq 2\sum_{j=k=1}^{2^n} \sigma^4\delta^2 + 2\sum_{j>k}^{2^n}\sum^{2^n} \frac{\sigma^4\delta^4 p^2}{4}, \quad (6.58)$$

$$\mathrm{Var}(B_N) \leq 2\sigma^4 T^2 \left\{ \frac{1}{2^n} + \frac{p^2}{2}\frac{T^2}{2^{2n}} \right\}. \quad (6.59)$$

Now, apply Chebyshev's Inequality to find

$$\Pr\left\{ |B_N - \overline{B}_N| \geq \varepsilon \right\} \leq \frac{\mathrm{Var}(B_N)}{\varepsilon^2}, \quad (6.60)$$

let $\varepsilon = n/\sqrt{2^n}$ then

$$\Pr\left\{ |B_N - \overline{B}_N| \geq \frac{n}{\sqrt{2^n}} \right\} \leq \frac{2^n\sigma^4 T^2}{n^2}\left\{\frac{1}{2^n} + \frac{p^2T^2}{2^{2n+1}}\right\}. \quad (6.61)$$

With the observation interval (T) fixed, the quantity on the right hand side of equation (6.61) is bounded. Defining the event A_n as

$$A_n \triangleq \{\omega: |B_n - \overline{B}_n| \geq \varepsilon\}$$

Then since

$$\sum_{n=1}^{\infty} p(A_n) \leq \sum_{n=1}^{\infty} \frac{M}{n^2} < \infty,$$

application of the Borel Cantelli lemma gives the result that the random variable B_N converges almost surely to its mean ($\overline{B}_N=\sigma^2 T$), as the sampling becomes dense in the observation interval.

Another interesting aspect of this result is the behavior of the increments ($\underline{\Delta}x_k$) as the sampling rate becomes high. From equation (6.56) the increment to increment correlation is

$$E(\underline{\Delta}x_k \underline{\Delta}x_j) = a_{jk} = -\frac{\sigma^2 p\, \delta^2}{2} .$$

As the sampling interval (δ) becomes smaller, the increments become independent, since the $\underline{\Delta}x_k$ are Gaussian. This suggests that the random variable B_N, since it is the sum of squared Gaussian random variables, could be modeled as chi - squared (χ^2).

To extend the result to processes of an order higher than M=1, an adaptation of Baxter's Theorem, equation (6.41),due to Slepian (1958) is invoked. This extension is valid when the correlation function R(t,s) has derivatives of an order less than 2M-1 which are continuous at t=s and the process x_t is M-1 times differentiable, with $R^{2M-2}(t,s)$ as the correlation function of the process $x^{(M-1)}(t)$. Under these conditions, the condition equivalent to equation (6.41) is

$$\lim_{n\to\infty} {}^{M}B_N = \lim_{\substack{n\to\infty \\ \delta\to 0}} \sum_{k=1}^{2^n} \left[X^{(M-1)}\left(\frac{k}{2^n}\right) - X^{(M-1)}\left(\frac{k-1}{2^n}\right) \right]^2 \longrightarrow \int_0^T \sigma^2_{2M-1}(t)\,dt , \tag{6.63}$$

where σ^2_{2M-1} is the jump in the 2M-1 st derivative of the correlation function at t=s, with

$$\lim_{n\to\infty} {}^{p}B_N = \lim_{\substack{n\to\infty \\ \delta\to 0}} \sum_{k=1}^{2^n} \left[X^{(p-1)}\left(\frac{k}{2^n}\right) - X^{(p-1)}\left(\frac{k-1}{2^n}\right) \right]^2 \longrightarrow 0 \tag{6.64}$$

for derivatives of the order $p < M-1$.

Now this result is applied to the M=2 process with the correlation function

$$R_x(T) = \frac{\sigma^2}{2p_1p_2(p_2^2 - p_1^2)} \left(p_2 e^{-p_1|T|} - p_1 e^{-p_2|T|} \right) , \tag{6.65}$$

which under the conditions of fast sampling $\left((p_1 + p_2)\delta << 1\right)$ reduces to (for small m values)

$$R_x(m\,\delta) = \frac{\sigma^2}{2p_1p_2(p_2^2 - p_1^2)} \left(p_2(1-p_1 m\,\delta) - p_1(1-p_2 m\,\delta)\right) . \tag{6.66}$$

Using equation (6.64) with M=2, the implication is that

$$\lim_{n \to \infty} {}_1B_N = \lim_{n \to \infty} \sum_{k=1}^{2^n} \left(x(\frac{k}{2^n}) - x(\frac{k-1}{2^n})\right) \longrightarrow 0$$

To see this, the same procedure will be carried out as was used for the M=1 process. Using the fast sampling representation for the correlation function , equation (6.66), the mean of ${}_1B_N$ is

$$E({}_1B_N) = \sum_{k=1}^{2^n} E\left[x(\frac{k}{2^n}) - x(\frac{k-1}{2^n})\right]^2 = 2\sum_{k=1}^{2^n} \left(R_x(0) - R_x(\delta)\right),$$

$$E({}_1B_N) = 2\sum_{k=1}^{2^n} \frac{\sigma^2}{2p_1p_2(p_2^2-p_1^2)} (p_1-p_2) - \left((p_2-p_1) + (p_1p_2-p_1p_2)\delta\right)$$

$$E({}_1B_N) = 0 \tag{6.67}$$

The variance computation for the random variable ${}_1B_N$, provides the result

$$Var({}_1B_N) = 2\sum_{j=k=1}^{2^n} E(\underline{\Delta}x_k^2)^2 + 4\sum\sum_{j>k}^{2^n} E^2(\underline{\Delta}x_k \underline{\Delta}x_j),$$

where

$$E(\underline{\Delta}x_k \underline{\Delta}x_j) = 2R(j-k) - R(j-k-1) - R(j-k+1) = \frac{\sigma^4}{(p_1+p_2)^2}\,\delta^4 ,$$

$$E(\underline{\Delta}x_k)^2 = 0 ,$$

so that

$$\mathrm{Var}({}_1B_N) \leq \frac{\sigma^2 T^2}{(p_1 + p_2)^2}\,\delta^2 \quad , \tag{6.68}$$

with the observation interval (T) fixed. Continuing with the M=2 process calculations, let

$${}_2B_N = \sum_{k=1}^{2^n} \left(\dot{x}(\tfrac{k}{2^n}) - \dot{x}(\tfrac{k-1}{2^n})\right)^2 \quad .$$

The correlation function for the derivative process is

$$R_{\dot{x}}(T) = \frac{\sigma^2}{2(p_2^2 - p_1^2)} \left\{ p_2 e^{-p_2|T|} - p_1 e^{-p_1|T|} \right\} , \tag{6.69}$$

and under fast sampling conditions, this becomes

$$R_{\dot{x}}(m\,\delta) = \frac{\sigma^2}{2(p_2^2 - p_1^2)} \left\{ p_2(1-p_2 m\,\delta) - p_1(1-p_1 m\,\delta) \right\} \quad . \tag{6.70}$$

Computation of the mean of ${}_2B_N$ provides the result

$$\begin{aligned} E({}_2B_N) &= \sum_{k=1}^{2^n} E\left(\dot{x}(\tfrac{k}{2^n}) - \dot{x}(\tfrac{k-1}{2^n})\right)^2 \\ &= 2\sum_{k=1}^{2^n} R_{\dot{x}}(0) - R_{\dot{x}}(\delta) \\ &= \sigma^2 \sum_{k=1}^{2^n} \delta \\ &= \sigma^2 T \quad . \end{aligned} \tag{6.71}$$

The variance of the random variable ${}_2B_N$ is

$$\mathrm{Var}({}_2B_N) = 2\sum_{j=k=1}^{2^n} E^2(\underline{\Delta}x_k^2) + 4\sum_{j>k}^{2^n}\sum E^2(\underline{\Delta}x_k \underline{\Delta}x_j) \tag{6.72}$$

$$\leq \frac{\sigma^2 T^4}{2^n} \left\{ 2 + \frac{T^2}{2^n} \frac{(p_2^2 + p_1 p_2 + p_1^2)}{(p_2 + p_1)^2} \right\} \quad .$$

As in the M=1 calculations, the application of Chebyshev's Inequality

and the Borel Cantelli lemma shows that for the M=2 process

$$ {}_iB_N \xrightarrow[\substack{N \to \infty \\ \delta \to 0}]{} {}_i\overline{B_N} \quad , \text{ for } i = 1,2 $$

with probability 1, where ${}_1\overline{B_N} = 0$, and ${}_2\overline{B_N} = \sigma^2 T$. The consequence of all this is that through the application of Baxter's Theorem, equations (6.63) and (6.64), there is a measurement that can provide an estimate of the random parameter σ^2. The accuracy of this estimate can be bounded through the use of Chebyshev's Inequality.

6.6 An Engineering Approach to Baxter's Theorem

An interpretation, in terms of linear filtering of a random process is possible for the operations indicated by Baxter's Theorem, equations (6.63) and (6.64). Through this interpretation, insight as to the fundamental properties being employed can be obtained. To begin with, interpret the relationship

$$ B_N = \sum_{k=1}^{N=2^n} \left(x\left(\frac{k}{2^n}\right) - x\left(\frac{k-1}{2^n}\right) \right)^2 \longrightarrow \overline{B}_N \quad , \tag{6.73} $$

as a filtering operation followed by a measurement, where the filter output is denoted by

$$ d(k) = x\left(\frac{k}{N}\right) - x\left(\frac{k-1}{N}\right) \quad . \tag{6.74} $$

The measurement is inferred from

$$ B_N = \sum_k d^2(k) = N R_d(0) \quad , \tag{6.75} $$

with $R_d(0)$ as the output power of the filter and $N = T/\delta$. The transfer function of the filter described by equation (6.74) has the squared magnitude

$$ |H(\omega)|^2 = 4 \sin^2(\omega \delta / 2) \quad , \tag{6.76} $$

and the output spectral density is

$$S_d(\omega) = 4 \sin^2(\omega \delta/2)\, S_x(\omega) \quad . \tag{6.77}$$

For values of ω which are small compared to $1/\delta$, the output spectral density has the form

$$S_d(\omega) \cong \omega^2 \delta^2 S_x(\omega) \ , \tag{6.78}$$

which shows the filter to be a differentiator for low frequency inputs. Specific illustration of this behavior is presented in Figure 6 , where one period of the filter transfer function is plotted for various rates (δ). As an illustration of this filtering interpretation, a second order process with spectral densities

$$S_x(\omega) = \frac{\sigma^2}{(\omega^2 + 1)(\omega^2 + 100)} \ , \tag{6.79}$$

and

$$S_{\dot{x}}(\omega) = \omega^2 S_x(\omega) \ ,$$

is processed by the filtering operation described by equation (6.77), for various sampling rates (δ). Figure 7 illustrates the response of the filter to the observations process, and Figure 8 illustrates the response to the derivative process of the M=2 model. For the observations process, as the sampling rate becomes smaller, the response compresses both in frequency span and magnitude. As a consequence of this compression, the output power $\left(R_d(0)\right)$ approaches zero rapidly. This is in consonance with the result predicted by Baxter's Theorem, since for a second order process, the mean of the random variable ${}_1B_N$ approaches zero under fast sampling conditions. The result of passing the derivative process through the same filter shows a flattening and spreading of the spectral response, from which it can be inferred that the output power is settling to a non

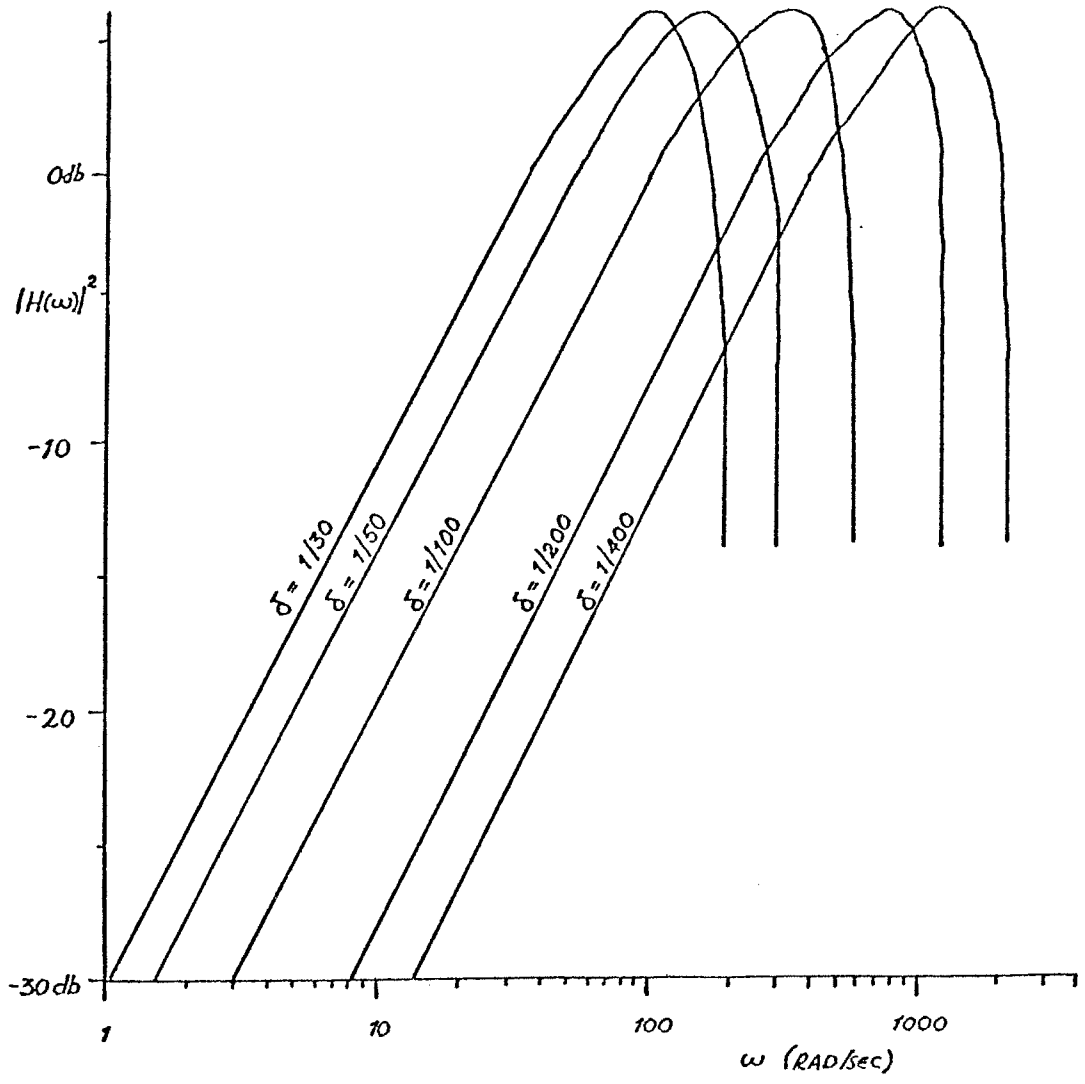

Figure 6. Squared Magnitude Filter Response for Various Sampling Rates (δ)

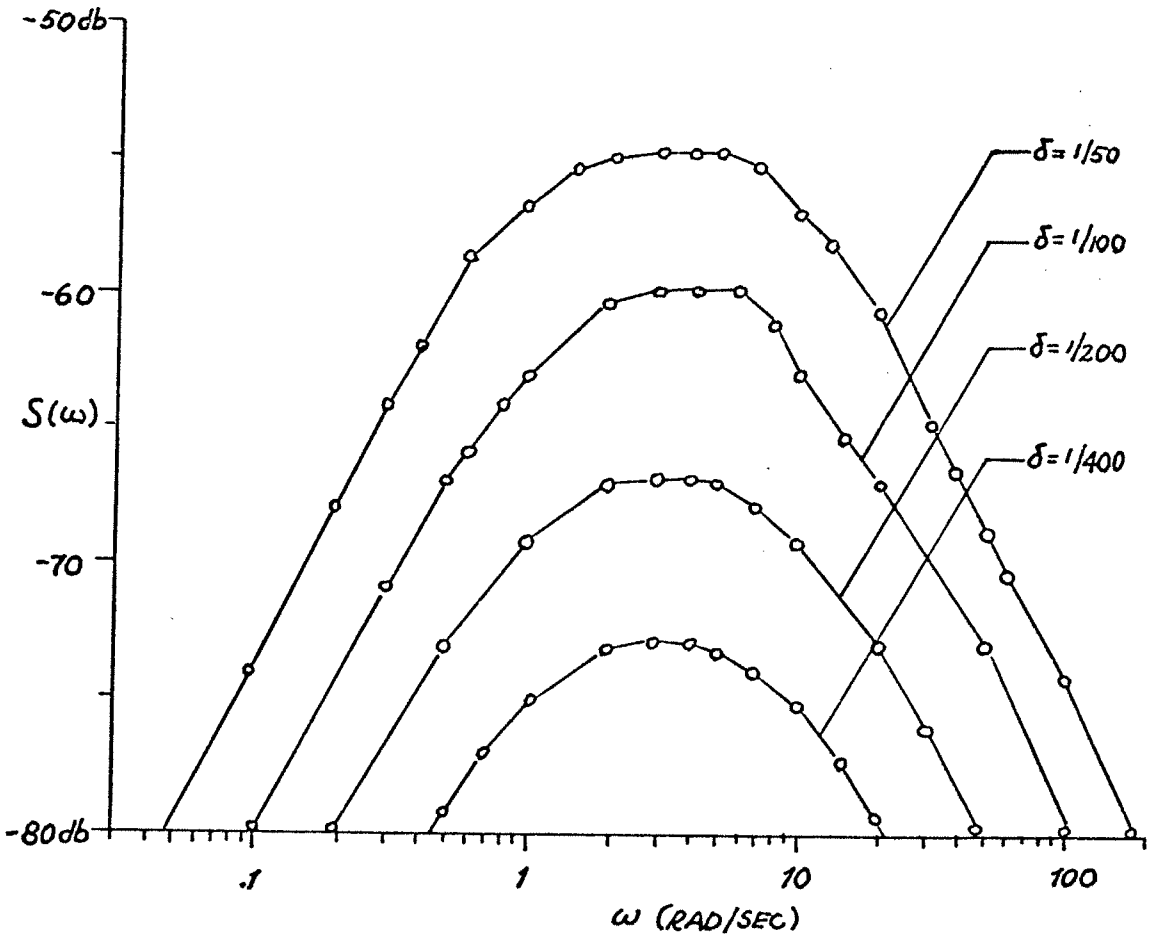

Figure 7. Response of the Observations Process to the Filter Specified by Sampling Rate (δ).

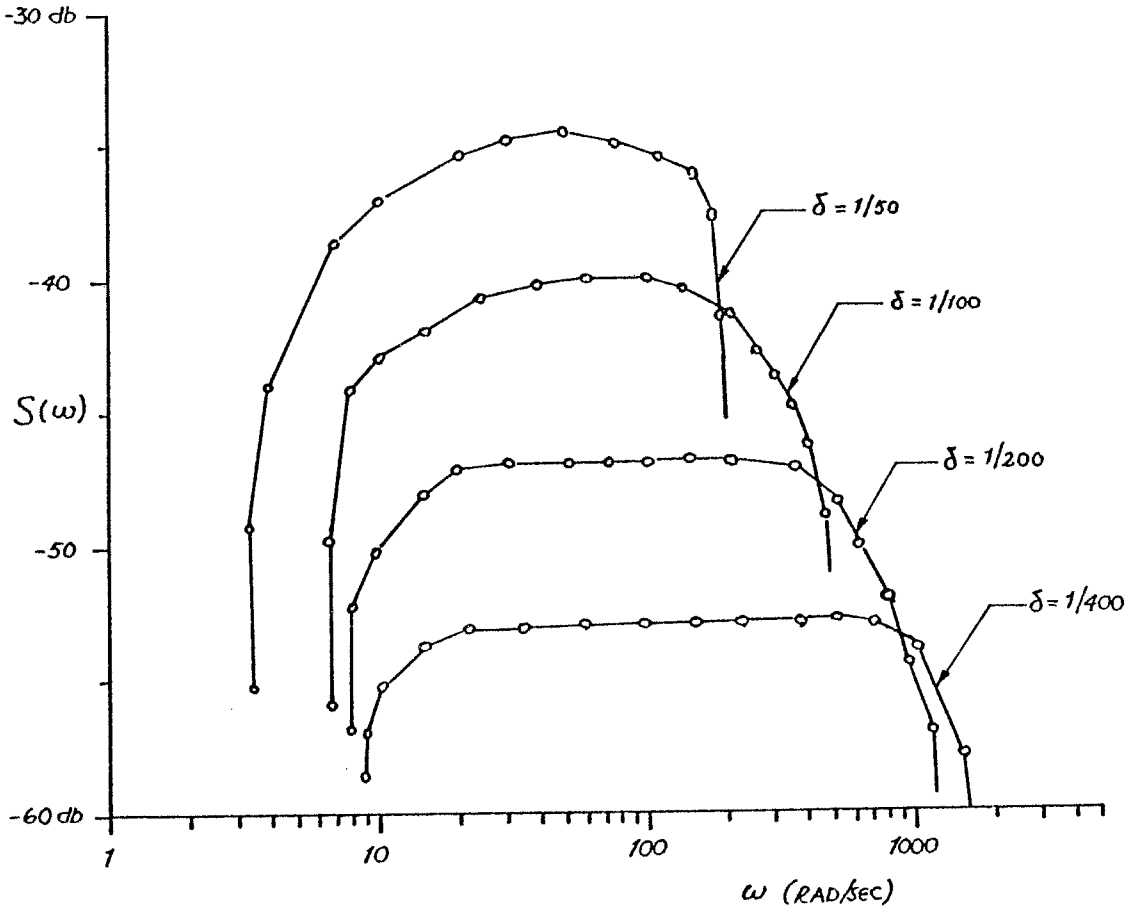

Figure 8. The Response of the Derivative Process to the Filter Specified by Sampling Rate (δ).

zero constant, which is the result predicted by the application of Baxter's Theorem to the derivative process.

In performing a simulation of the process just described, a discrete model is used to generate the observations process and an approximate derivative is formed from a difference quotient of the observations. To insure that the properties given by equation (6.73) and its adaptations, are still valid, the following analysis, in terms of filtering operations, is carried out. Analysis of the filtering operation is carried out through the use of Z transforms, and by specifying the order of the model. For the M=1 discrete process, the correlation function is of the form

$$R(n\,\delta) = \frac{\sigma^2}{2p} e^{-p|n\,\delta|} \tag{6.80}$$

Taking the Z transform of the correlation function, the result is

$$Z\{R(n\,\delta)\} = \frac{\sigma^2}{2p} \frac{1 - e^{-2p\,\delta}}{|1 - e^{-p\,\delta} Z|^2} \tag{6.81}$$

where

$$Z\{f(n\,\delta)\} \triangleq \sum_{-\infty}^{0} f(n\,\delta)Z^n + \sum_{0}^{\infty} f(n\,\delta)Z^{-n} - f(0) \quad .$$

Performing the filtering on the random sequence, the output spectral density is

$$\begin{aligned} S_d(Z) &= |H(Z)|^2 S_x(Z) \\ &= (1 - Z^{-1})(1 - Z^{1}) \frac{\sigma^2}{2p} (1 - e^{-2p\,\delta}) \\ &\quad \cdot \frac{1}{(1 - e^{-p\delta} Z)(1 - e^{-p\delta} Z^{-1})} \end{aligned} \tag{6.82}$$

Application of the inverse Z transform, to find the output correlation, provides the result

$$\oint_{Z=1} S_d(Z)\, Z^{n-1} dZ = R_d(n\delta) \quad , \tag{6.83}$$

and

$$R_d(0) = \oint_{Z=1} S_d(Z)\, Z^{-1} dZ$$

$$R_d(0) = \oint_{Z=1} \frac{(1-Z)(1-Z^{-1})\, Z^{-1}}{(1-e^{-p\delta}Z)(1-e^{-p\delta}Z^{-1})} \quad . \tag{6.84}$$

After performing the contour integration, the output power is

$$R_d(0) = \frac{\sigma^2}{2p}\, 2(1 - e^{-p\delta}) \quad . \tag{6.85}$$

Under the assumption that $p\delta \ll 1$, the expression for the output power reduces to

$$R_y(0) = \frac{\sigma^2}{2p}\, 2(1 - e^{-p\delta}) = \frac{\sigma^2}{2p}\, 2p\delta = \sigma^2\delta \quad , \tag{6.86}$$

where δ is the sampling interval. After scaling by the number of observations (N), this is precisely the result desired. Extending these thoughts to the second order process, the filtering will be performed on the observations and derivative processes. For the second order process, the model

$$x(n) = a_1 x(n-1) - a_2 x(n-2) + \zeta(n) \quad , \tag{6.87}$$

with

$$\zeta(n) = N(0, \sigma^2\delta^3)$$

$$a_1 = (2 - \delta(p_1 + p_2)$$

$$a_1 = \left(1 - \delta(p_1 + p_2) + \sigma^2 p_1 p_2\right) \quad ,$$

has correlation functions represented by

$$R_x(k\delta) = \frac{\sigma^2\delta^3}{(m_1-m_2)(1-m_1m_2)}\left\{\frac{m_1}{1-m_1^2}m_1^k - \frac{m_2}{1-m_2^2}m_2^k\right\}, \quad (6.88)$$

$$R_{\dot{x}}(k\delta) = \frac{\sigma^2\delta}{(m_1-m_2)(1-m_1m_2)}\left\{\frac{1-m_2}{(1+m_2)}m_2^k - \frac{1-m_1}{1+m_1}m_1^k\right\}, \quad (6.89)$$

where m_1, m_2 are related to the coefficients a_1, a_2 through the characteristic equation. The spectral density representations for the observation and the derivative processes, found through use of the two-sided Z transform, are

$$S_x(Z) = \frac{\sigma^2\delta^3}{|(Z-m_1)(Z-m_2)|^2}, \quad (6.90)$$

$$S_{\dot{x}}(Z) = \frac{\sigma^2\delta}{(m_1-m_2)(1-m_1m_2)}\left[\frac{(1-m_2)^2}{|Z-m_2|^2} - \frac{(1-m_1)^2}{|Z-m_1|^2}\right]. \quad (6.91)$$

Processing the observations with the filter specified by

$$|H(Z)|^2 = (1-Z^{-1})(1-Z), \quad (6.92)$$

the output spectral density is

$$S_d(Z) = \frac{(1-Z^{-1})(1-Z)}{|(Z-m_1)(Z-m_2)|^2}\sigma^2\delta^3, \quad (6.93)$$

and using the inverse transform, the output power is found to be

$$R_d(0) = \oint_{Z=1} \frac{\sigma^2\delta^3(1-Z^{-1})(1-Z)Z^{-1}}{|(Z-m_1)(Z-m_2)|^2}\,dZ$$

$$R_d(0) = \frac{2\sigma^2\delta^3}{(1-m_1m_2)(1+m_2)(1+m_1)} \quad (6.94)$$

$$= \frac{2\sigma^2\delta^3}{(1-m_1m_2)(1+m_1+m_2+m_1m_2)}.$$

To reduce this to a recognizable form, the definition of the roots

m_1, m_2 will be recalled. From the denominator of the spectral density, the characteristic equation

$$Z^2 - (m_1 + m_2)Z + (m_1 m_2) = 0 \quad . \tag{6.95}$$

The homgeneous difference equation associated with this characteristic equation is

$$x(n+2) = (m_1 + m_2)x(n+1) - m_1 m_2 x(n) \quad . \tag{6.96}$$

By association with equation (6.87), the coefficients (a_i) are related to the roots (m_i) through the equations

$$m_1 + m_2 = a_1 = \left(2 - \delta(p_1 + p_2)\right) = (1 - p_1\delta) + (1 - p_2\delta), \tag{6.97}$$

$$m_1 m_2 = a_2 = \left(1 - (p_1 + p_2)\delta + p_1 p_2 \delta^2\right) = (1 - p_1\delta)(1 - p_2\delta) \tag{6.98}$$

Using these definitions in the filter output correlation function

$$R_d(0) = \frac{2\sigma^2 \delta^3}{(1-m_1 m_2)\left(1 + (m_1 + m_2) + m_1 m_2\right)} \quad , \tag{6.99}$$

results in the form

$$R_d(0) = \frac{2\sigma^2 \delta^3}{\left(2 - 2(p_1 + p_2)\delta + p_1 p_2 \delta^2\right)\left((p_1 + p_2)\delta - p_1 p_2 \delta^2\right)} . \tag{6.100}$$

Upon scaling by N, and letting δ become small, the result is

$$\lim_{\delta \to 0} {}^1B_N = \lim_{\delta \to 0} NR_d(0) = \lim_{\delta \to 0} \frac{\sigma^2 T}{(p_1 + p_2)} \cdot \delta \to 0 \quad , \tag{6.101}$$

which is the condition predicted, through the application of Baxter's Theorem to the observations of an M=2 process.

Performing the same filtering operation on the approximate derivative process, the output spectral density is

$$S_d(Z) = |H(Z)|^2 S_{\dot{x}}(Z)$$

$$S_d(Z) = \frac{\sigma^2 \delta}{(m_1 - m_2)(1 - m_1 m_2)} \left\{ \frac{(1 - m_2)^2}{|Z - m_2|^2} - \frac{(1 - m_1)^2}{|Z - m_1|^2} \right\} \left\{ (1 - Z^{-1})(1 - Z^1) \right\} . \tag{6.102}$$

The output correlation at k=0 is found to be

$$R_{\dot{y}}(0) = \oint_{Z=1} S_{\dot{y}}(Z)\, Z^{-1} dZ$$

$$= \frac{2\sigma^2\delta}{(m_1-m_2)(1-m_1m_2)} \quad \frac{(1-m_2)^2(1+m_1) - (1-m_1)^2(1+m_2)}{(1+m_2)(1+m_1)} \,. \tag{6.103}$$

Using the definitions for m_1 and m_2, the result is further reduced to

$$R_{\dot{y}}(0) = \frac{\sigma^2\,\delta}{\left(1 - \left(\frac{p_1p_2}{p_1+p_2} + \frac{p_1+p_2}{2} - \frac{p_1p_2}{2}\right)\delta + 0(\delta^2)\right)} \tag{6.104}$$

Again, scaling by the number of observations and letting the sampling rate become small, the result is

$$\lim_{\delta \to 0} 2^{B_N} = \lim_{\delta \to 0} NR_d(0) = \lim_{\delta \to 0} \sigma^2 N\,\delta = \sigma^2 T \,. \tag{6.105}$$

In addition to the intuitive sense developed by the filtering interpretation, there is confidence that use of a discrete observation model will provide the result predicted by Baxter's Theorem. The conclusion which can be reached from this exposition is that through the use of dense sampling, one of the random variables in the spectral estimation problem can be determined significantly faster than the others. Further conditions under which this result can be used, and an empirical justification of this statement, is the concern of the simulation studies to be presented.

6.7 Structure of the Simulation Process

Primary emphasis in the simulation studies is placed on the M=2 process, because it exhibits the salient characteristics of higher order processes and the presentation of results is more amenable to graphic portrayal. The model used to generate the observations pro-

process is the basic second order difference equation

$$x(n) = ax(n-1) + bx(n-2) + \zeta(n)$$

where

$$a = \left(2 - \delta(p_1 + p_2)\right)$$

$$b = -\left(1 - \delta(p_1 + p_2) + \delta^2 p_1 p_2\right)$$

and $\zeta(n)$ is a zero mean white Gaussian sequence with variance $\sigma^2\,\delta^3$ with δ representing the sampling interval.

The processing of the observations generated to produce the post density for the parameter $\{p_i\ i=1,2\}$ is described by the modified set of sufficient statistics for the M=2 process,

$$\left\{\int_0^T x^2(t)dt\ ,\ \int_0^T \dot{x}^2(t)dt\right\}. \tag{6.106}$$

In the simulation study, these measures are realized by the set

$$\left\{\sum_{n=1}^{N} x^2(n)\cdot\delta\ ,\ \sum_{n=1}^{N} \dot{x}^2(n)\cdot\delta\right\}, \tag{6.107}$$

where an approximate derivative is formed from the difference quotient. Once this derivative is formed, the processing necessary to determine the random variable σ^2 is described as

$$B_N = \sum_{n=1}^{N} \left(\dot{x}(n) - \dot{x}(n-1)\right)^2 = \sigma^2 T \tag{6.108}$$

where T is the n-1 observations interval. Central to the estimation procedure is the choice of the sampling interval (δ). It is under the assumption that the sampling rate is fast that these measurements have meaning. The criterion, for the M=2 process, which till now has been necessary to insure the results achieved, is described by $(p_1 + p_2) \ll 1$. It might be argued that since p_1 and p_2 are random variables, that it would be difficult to insure that such a condition

is satisfied. In most circumstances p_1 and p_2 are sufficiently separated so that the fast sampling condition is nothing more than a Nyquist criterion, and hence requires the same magnitude of prior knowledge as would the application of the Nyquist criterion. Further, such a condition is a salient feature of the model approach to spectral estimation, where prior knowledge can be used to set the sampling rate.

CHAPTER VII

RESULTS OF SIMULATION STUDY

7.1 Introductory Comments

The significance of using an all pole spectral representation was that the properties of the correlation functions, related to this process, provided a vehicle for invoking the consequences of Baxter's Theorem and its extensions. Succinctly, this theorem provides justification for the inference that the random parameter σ^2 (white noise intensity driving the difference equation) converges significantly faster than the other parameters (p_i) in the estimation problem. Further, this theorem provides the statistic necessary to learn the parameter σ^2 under fast sampling conditions. It was demonstrated that for an M=2 process, the random variable

$$ {}_2B_N = \sum_{i=1}^{N} (\dot{x}_i - \dot{x}_{i-1})^2 \xrightarrow[\substack{h \to 0 \\ N \to \infty}]{a.s.} E({}_2B_N) = \sigma^2 T, \qquad (7.1) $$

where T is the observation interval, converges to its mean. In addition, this random variable was shown to be the sum of independent, identically distributed, squared Gaussian random variables, and under fast sampling conditions, the variance is given by

$$ \mathrm{Var}({}_2B_N) = 2(\sigma^2 T)^2 / N \ . \qquad (7.2) $$

The details leading to equation (7.2) are provided in Appendix E . Using the mean to standard deviation ratio (μ/σ) as an indication of the estimate's quality, where

$$\frac{\mu}{\sigma}\,(_2B_N) \;=\; \sqrt{\frac{N}{2}} \;=\; \sqrt{\frac{T}{2h}} \qquad (7.3)$$

it becomes apparent that for fixed observation interval (T), a smaller sample spacing (h) provides a better estimate (larger μ/σ). Figure 9 illustrates the results of the simulation studies, where the empirical μ/σ ratio is plotted versus the observation interval for various sampling rates. As predicted by the analysis, if the observation interval (T) is fixed, the use of smaller sampling intervals (h) produces a larger μ/σ ratio. From these results, it appears that unless otherwise restricted, there is a variety of T,h choices available to obtain the desired accuracy in the estimate of the random variable σ^2. Hence, at this point, the only limitiation on accuracy is based on the ability to sample fast enough. This picture will be altered later in this chapter, when quantization effects are included, but for now, the view presented will be pursued. Tables 1 and 2 illustrate the μ/σ ratio, for the random variable $_2B_N$, when different pole combinations are used. The point of these illustrations is that fast sampling rates render the behavior of the μ/σ ratio insensitive to the pole locations.

At this point it is instructive to establish the computational method of determining the constituent parts of the μ/σ ratio. The empirical averages of the random variable $_2B_N$ are formed by considering a fixed number of observations (η) where

$$\eta = M \cdot N \quad , \qquad (7.4)$$

with M representing the number of non-overlapping segments, and N representing the number of observations in each segment. The average of $_2B_N$ is then found from

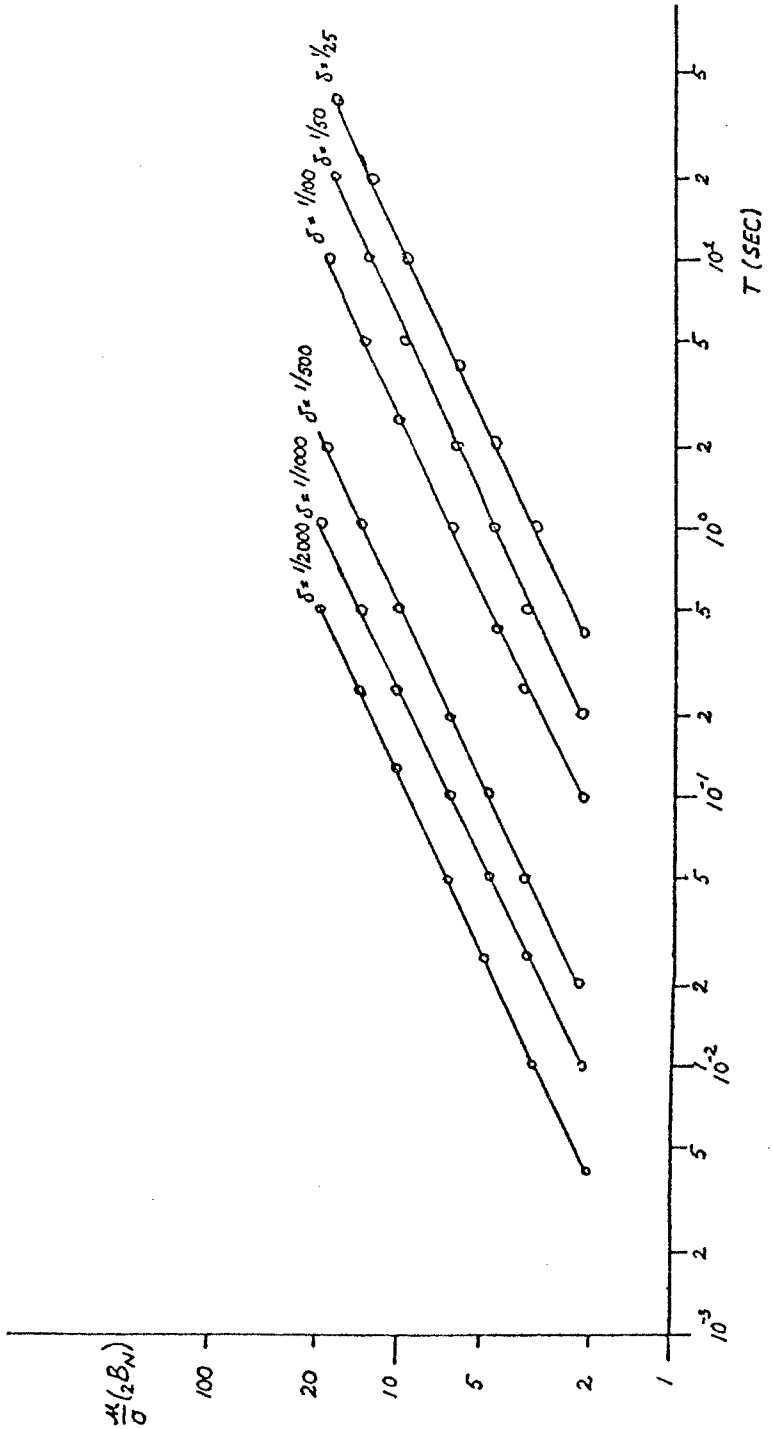

Figure 9. The Mean to Standard Deviation Ratio of the Random Variable $2B_N$ Versus Observation Interval (t) for the Various Sampling Intervals (δ).

TABLE 1

MEAN TO STANDARD DEVIATION RATIO OF $_2B_N$ FOR VARIOUS (p_1,p_2) AT SAMPLING RATE $h = 1\times10^{-2}$

$(p_1=1,p_2=5)$		$(p_1=1,p_2=10)$		$(p_1=2,p_2=10)$	
T(sec)	μ/σ	T(sec)	μ/σ	T(sec)	μ/σ
10.0	18.64	10.0	18.2	10.0	18.3
5.0	13.73	5.0	13.6	5.0	13.56
2.5	10.15	2.5	10.1	2.5	10.06
1.0	6.612	1.0	6.58	1.0	6.58
0.5	4.75	0.5	4.7	0.5	4.7
0.25	3.50	0.25	3.46	0.25	3.46
0.10	2.21	0.10	2.19	0.10	2.19

TABLE 2

MEAN TO STANDARD DEVIATION RATIO OF $_2B_N$ FOR VARIOUS (p_1,p_2) AT SAMPLING RATE $h = 1\times10^{-3}$

$(p_1=1,p_2=5)$		$(p_1=1,p_2=10)$		$(p_1=2,p_2=10)$	
T(sec)	μ/σ	T(sec)	μ/σ	T(sec)	μ/σ
1.0	18.65	1.0	18.7	1.0	18.6
0.5	13.83	0.5	13.77	0.5	13.7
0.25	10.3	0.25	10.2	0.25	10.2
0.1	6.65	0.1	6.62	0.1	6.61
0.05	4.78	0.05	4.77	0.05	4.76
0.025	3.54	0.025	3.53	0.025	3.53
0.01	2.22	0.01	2.22	0.01	2.22

$$\overline{{}_2B_N}^{M} = \frac{1}{M}\sum_{i=1}^{M} B_{N_i} \quad . \tag{7.5}$$

The procedure for determining the average from equation (7.4) and equation (7.5) can best be explained via an illustration. Consider a sequence of values for N, that is, $N_1=1$, $N_2=2,\ldots..N_k$, where the N_i are related to the total number of points (η), by

$$\eta = N_1M_1 = N_2M_2 \ldots = N_kM_k \quad . \tag{7.6}$$

The averages formed by the sequence are represented by

$$\overline{{}_2B_1}^{M_1} = \frac{1}{M_1}\sum(\underline{\Delta}\dot{x}_1^2) + (\underline{\Delta}\dot{x}_2^2) + \ldots (\underline{\Delta}\dot{x}^2_{M_1N_1}) \quad , \tag{7.7}$$

$$\overline{{}_2B_2}^{M_2} = \frac{1}{M_2}\sum(\underline{\Delta}\dot{x}_1^2+\underline{\Delta}\dot{x}_2^2)+ (\underline{\Delta}\dot{x}_3^2+\underline{\Delta}\dot{x}_4^2) + \ldots(\underline{\Delta}\dot{x}^2_{M_2N_2-1}+\underline{\Delta}\dot{x}^2_{M_2N_2}), \tag{7.8}$$

$$\overline{{}_2B_{N_k}}^{M_k} = \frac{1}{M_k}\sum_{k=1}^{M_k} {}_2B_{N_k} \quad , \tag{7.9}$$

where

$$B_{N_k} = \sum_{i=1}^{N_k}(\dot{x}_i - \dot{x}_{i-1})^2 \quad . \tag{7.10}$$

Recalling that the mean of B_N is

$$E({}_2B_N) = N\sigma^2h \quad , \tag{7.11}$$

a relationship among the various empirical averages above, can be established to facilitate computation. That is, since

$$\overline{{}_2B_{N_1}}^{M_1} = N_1\sigma^2h \quad , \tag{7.12}$$

and

$$\overline{{}_2B_{N_2}}^{M_2} = N_2\sigma^2h \quad , \tag{7.13}$$

then

$$\overline{{}_2B_{N_2}}^{M_2} = \frac{N_2}{N_1}\overline{{}_2B_N}^{M_1} \quad . \tag{7.14}$$

If, in addition, N =1, then the mean computation for the other N_i is $\overline{B_N}^M$ scaled by N_i.

Computation of the second moment presents a different story. Again, the best way to understand the computation is through illustration. Using a fixed number of points η the averages will be computed over M partitions of N points, so that η=M N. The computation of the second moment for a sequence of values N=1,2,....k is represented by

$$\overline{{}_2B_1^2}^{M_1} = \frac{1}{M_1}\,(\underline{\Delta}\dot{x}_1^2) + (\underline{\Delta}\dot{x}_2^2) + \ldots\ (\underline{\Delta}\dot{x}^2_{N_1M_1})^2 \quad , \tag{7.15}$$

$$\overline{{}_2B_2^2}^{M_2} = \frac{1}{M_2}\,(\underline{\Delta}\dot{x}_1^2 + \underline{\Delta}\dot{x}_2^2)^2 + \ldots\ (\underline{\Delta}\dot{x}^2_{M_2N_2-1} + \underline{\Delta}\dot{x}^2_{M_2N_2})^2 \quad , \tag{7.16}$$

$$\overline{2B^2_{N_k}}^{M_k} = \frac{1}{M}\sum\,(B_{N_M})^2 \quad , \tag{7.17}$$

where

$$B^2_{N_k} = \Big(\sum_{i=1}^{N_k}(\dot{x}_i - \dot{x}_{i-1})^2\Big)^2 \quad . \tag{7.18}$$

The averages used to construct the results previously presented were based on a value of η = 20,000. The largest value of N considered was N = 1000, and corresponding to this N, the minimum number of segments is M = 20. This smaller number of segments (M), used for averaging, accounts for the deviation of the computed μ/σ from the expected value at larger T. At the lower sampling rates (h=1/50, 1/25), the μ/σ ratio shows an increased departure from the value predicted by equation (7.3). This increased departure is a result of non satisfaction of the sampling requirement

$$(p_1 + p_2)h << 1 \quad ,$$

so that

$$e^{-(p_1+p_2)h} \cong 1 - (p_1+p_2)h \quad .$$

7.2 Discussion of Simulation Results

It was previously asserted that under fast sampling conditions the random parameter σ^2 converged significantly faster than the other parameters (q_i) in the estimation problem. It is the object here to provide a heuristic demonstration of this statement. The requirements of Baxter's Theorem demand sampling fast enough so that (for M=2) the correlation function for the derivative process can be closely approximated by a straight line. This means that the sample interval (h) is picked such that the approximations

$$e^{-p_1 h} \cong 1 - p_1 h$$

$$e^{-(p_1 + p_2)h} \cong \left(1 - (p_1 + p_2)h\right)$$

are valid. Using Figure 11$\left(R_{\dot{x}}(T) \text{ vs } T\right)$ to establish the range of the h acceptable, it is apparent that h is picked where the sample to sample correlation of the derivatives is high. Projecting this requirement on the correlation function for the observations, Figure 10 $\left(R_x(T) \text{ vs } T\right)$, the sample to sample correlation is even higher. The effect this constraint has on the convergence of the parameters q_1 and q_2 is apparent through the relationship between the sufficient statistics and the distribution parameters. The mean and variance relations are

$$\overline{q_1} = \frac{\sum \dot{x}_i^2 \cdot h}{\sum x_i^2 \cdot h}, \quad (7.19) \quad \overline{q_2} = \frac{\sigma^2 T}{2 \cdot \sum x_i^2 \cdot h} \quad , \qquad (7.20)$$

$$\text{Var } q_1 = \frac{\sigma^2}{\sum x_i^2 \cdot h}, \quad (7.21) \quad \text{Var } q_2 = \frac{\sigma^2}{\sum \dot{x}_i^2 \cdot h} \quad , \qquad (7.22)$$

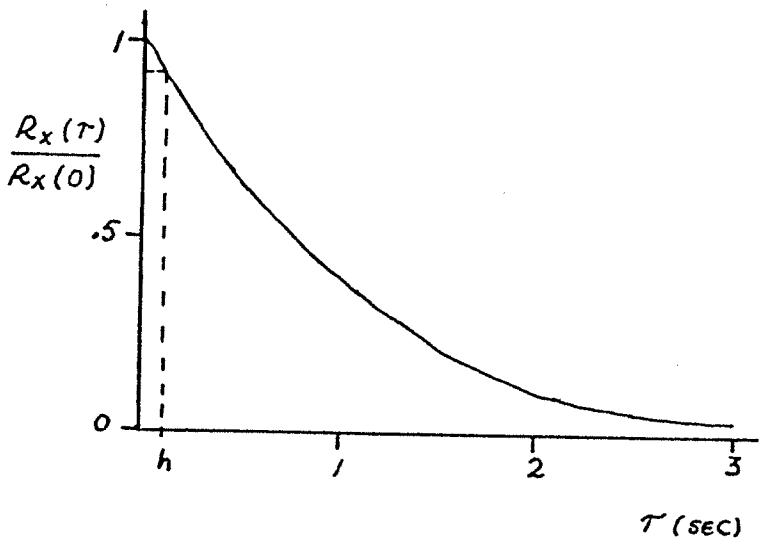

Figure 10. Normalized Covariance of Observations Process Versus lag (T).

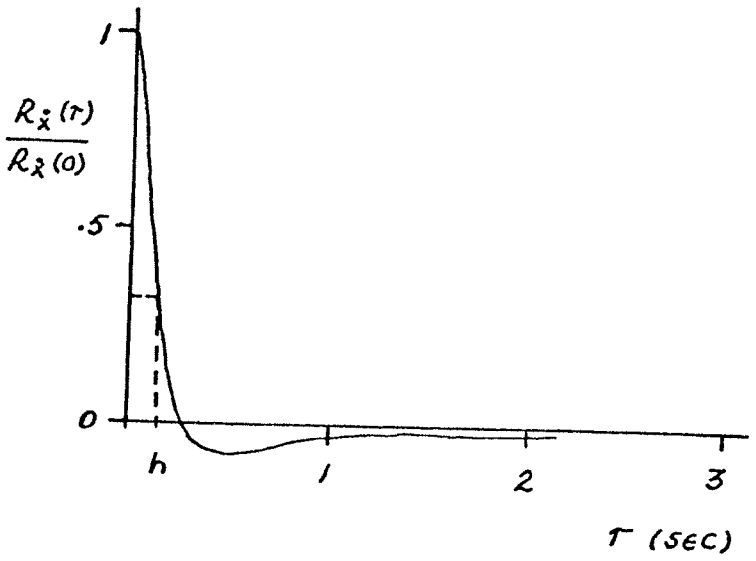

Figure 11. Normalized Covariance of Derivative Process Versus lag (T).

where $\sum x_i^2 \cdot h$ and $\sum x_i^2 \cdot h$ are the sufficient statistics. Interpreting the sufficient statistics as the sample variance for the observations and derivative process determined by periodic sampling of the random process, an understanding of the requirement for convergence can be obtained. The discussion below will be limited to interpretations placed on the sample variance for the observations. The same argument appropriately modified, can be applied to the sample variance for the derivative process. The sample variance is defined as

$$\hat{R}_x(0) = \frac{1}{Nh} \sum_{i=1}^{N} x_i^2 \cdot h \quad , \tag{7.23}$$

with

$$E\left(\hat{R}_x(0)\right) = \frac{1}{N} \sum_{i} E(x_i^2) = R_x(0) \quad , \tag{7.24}$$

and

$$\mathrm{Var}\left(\hat{R}_x(0)\right) = \frac{1}{N^2} \sum_{i}^{N} \sum_{j}^{N} R_{x^2}\left((i-j)h\right) - E^2\left(\hat{R}_x(0)\right) \tag{7.25}$$

where

h = sampling rate

N = Number of observations

T = Nh = observations interval.

Rearranging the double sum and using the Gaussian properties of the observations, the expression for the variance is

$$\mathrm{Var}\left(\hat{R}_x(0)\right) = \frac{2}{N}\left(R_x(0)\right)^2 + \frac{2}{N} \sum_{k=1}^{N-1} (1 - k/N)\, 2\left(R_x(k)\right)^2 \quad . \tag{7.26}$$

Next, to determine the effect of increasing the number of observations (N) while keeping the observation interval (T) fixed, rewrite equation (7.26) in the form

$$\mathrm{Var}\left(\hat{R}_x(0)\right) = \frac{2}{N}\left(R_x(0)\right)^2 + \frac{2}{T}\sum_{k=1}^{N-1}(1 - kh/T)2\left(R_x(kh)\right)^2 \cdot h. \tag{7.27}$$

Now, if T is held fixed and $N \to \infty$ ($h \to 0$) in a way such that $Nh=T$, the summation becomes an integral and the variance expression becomes

$$\lim_{N \to \infty} \mathrm{Var}\left(\hat{R}_x(0)\right) = \frac{2}{T}\int_0^T (1 - \frac{T}{T})2\left(R_x(T)\right)^2 dT \quad . \tag{7.28}$$

As a result, it is apparent that the variance of this estimate will not become arbitrarily small as $N \to \infty$ unless T also increases without limit. This is precisely the same requirement previously indicated, so that the estimates of the parameters q_1 and q_2 would be consistent. To achieve faster convergence for the estimates of the power, it would be required to space the samples so that the sample to sample correlation is small and hence the off diagonal terms in equation (7.26) provide a negligible contribution. This requirement contradicts the requirement imposed by using Baxter's Theorem, and hence, it can be concluded that under fast sampling the parameter σ^2 converges much faster than the parameters q_1 and q_2. The consequence of this result is that in using Bayes Rule to form a post density from

$$f(\underline{q},\sigma^2|x_t,\underline{x}_0) = \frac{f(x_t|\underline{q},\sigma^2,\underline{x}_0)\cdot f(\underline{q}|\sigma^2\underline{x}_0)f(\sigma^2)}{\int f(x_t|\underline{q},\sigma^2,x_0)\cdot f(\underline{q}|\sigma^2\underline{x}_0)f(\sigma^2)d\underline{q}d\sigma^2} \quad , \tag{7.29}$$

the random parameter σ^2 can be considered known and

$$f(q,\sigma^2|x_t,x_0) = f(q|x_t,x_0,\sigma^2)\Big|_{\sigma^2 = \overset{*}{\sigma}{}^2} \quad , \tag{7.30}$$

where $\overset{*}{\sigma}{}^2$ is the true value. That this in fact happens, will be borne out by the simulation results presented.

7.3 Presentation of the Results

Before presenting the results of the estimation problem from

the simulation studies, the method of presentation must be clarified. Since it has been shown that the post densities being dealt with are Gaussian, it is rather easy to specify the results by the mean and variance. However, the mean and variance being specified is not that of the parameters $\{p_i\}$, rather, they are a function of these parameters. Specifically, for the M=2 process, the random variables are $q_2;q_1$ where

$$q_2 = p_2 + p_1 \quad , \tag{7.31}$$

$$q_1 = p_1 \cdot p_2 \quad . \tag{7.32}$$

The relationship between the parameters $\{q_i\}$ and $\{p_i\}$ is given meaning by finding the complex spectral representation. For M=2, given that

$$S(\omega) = \frac{\sigma^2}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)} \quad , \tag{7.33}$$

let $s=j\omega$, and then

$$S(s) = \frac{\sigma^2}{|s^2 - (p_1 + p_2)s + p_1 p_2|^2} = \frac{\sigma^2}{|s^2 - q_2 s + q_1|^2} \quad . \tag{7.34}$$

The q_i are the coefficients of the characteristic equation, in the denominator of equation (7.34). Through the use of the characteristic equation, it is apparent that although q_1 and q_2 are real, the $\{p_i\}$ may be real or complex. For the M=2 process, the condition to be satisfied for real poles (p_1,p_2) is $q_2^2 \geq 4q_1$. Based on the model written for the difference equation, both p_1 and p_2 have positive real parts, and since the q_i are real numbers, their allowable range is depicted in Figure 12. All results of the estimation procedure will be presented in the q_1,q_2 plane, with the real/complex pole pair dividing line superimposed.

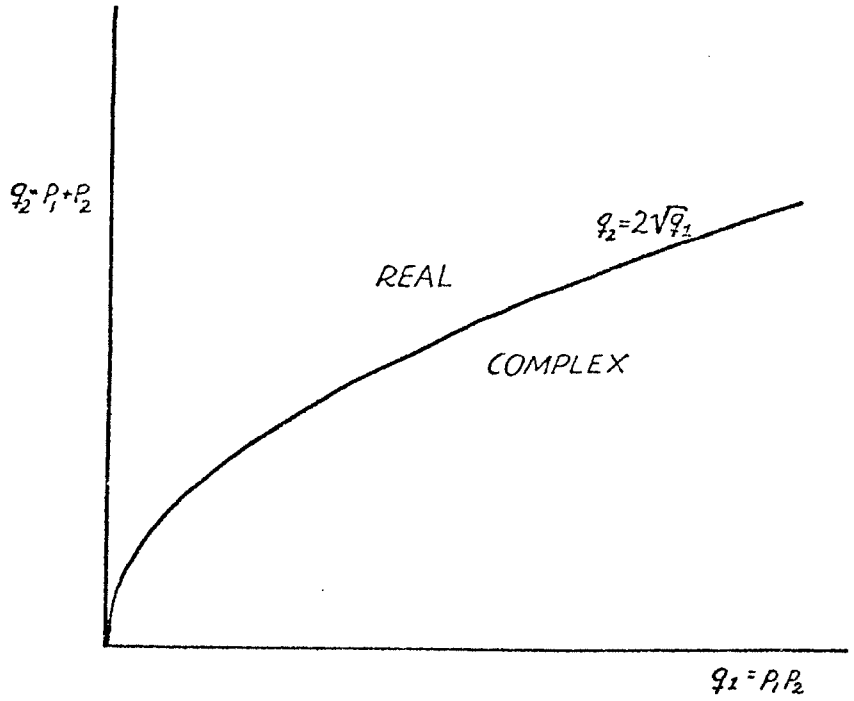

Figure 12. Allowable Range of Parameters for the M=2 Process.

In presenting the results of the estimation problem with a post density as an answer (M=2), the most direct way is through the use of constant probability density function contours. That is, a probability can be specifiec so that a contour in the q_1, q_2 plane is defined, and the parameters to be estimated fall within this contour with this probability. The expression for these contours(ellipses) are determined from the joint probability density function for a set of N Gaussian random variables, which is

$$f(\underline{a}) = \left(|2\pi|^{N/2} |\sum|^{1/2}\right)^{-1} \exp\ -\frac{1}{2}\, \underline{a}^t \sum^{-1} \underline{a} \quad . \tag{7.35}$$

From the above representation, one can observe that the equal height contours are given by the relation

$$\underline{a}^t \sum^{-1} \underline{a} = c^2 \quad , \tag{7.36}$$

which is an equation for an ellipse when N=2. The ellipses move out monotonically with c, and have the property that the probability of being inside the ellipse is

$$p = 1 - \exp(-c^2/2) \quad . \tag{7.37}$$

To compute the concentration ellipse for the M=2 process, represent the equation for the ellipse as

$$\frac{(q_2 - \mu_2)^2}{\sigma_2^2} + \frac{(q_1 - \mu_1)^2}{\sigma_1^2} = c^2 \quad . \tag{7.38}$$

Solving this equation for q_2 in terms of q_1 results in

$$q_2 = \mu_2 \pm \frac{\sigma_2}{\sigma_1} \sqrt{\sigma_1^2\, c^2 - (q_1 - \mu_1)^2} \quad . \tag{7.39}$$

If the parameter q_1 is determined in terms of increments of its standard deviation by

$$q_1 = \mu_1 \pm k\sigma_1 \quad , \tag{7.40}$$

then the expression for q_2 is

$$q_2 = \mu_2 \pm \sigma^2 \sqrt{c^2 - k^2} \tag{7.41}$$

where

$$c^2 = -2 \log_e(1 - p) \quad . \tag{7.42}$$

For all the results to be presented, the probability (p) was arbitrarily chosen to be p = .95, and the equations (7.40) and (7.41), were used with numerically determined means and variances to generate the ellipses.

Typical results of simulation studies are presented in Figures 13 through 15, where the 95% confidence contours are illustrated in the q_1 ($p_1 p_2$), q_2 (p_1+p_2) plane for various true values. The relatively rapid compression in the q_2 direction illustrated in Figure 13 is typical of large separations in the p_1, p_2 pair, and as previously explained, as the observation interval increases the contours become small. As the pole pairs get closer, the compression in the q_2, q_1 directions equalize. Again, this can be explained as a result of the fast sampling required to invoke Baxter's Theorem, which in turn slows down the convergence of the parameters q_2, q_1. Table 3 illustrates the difference in rate of convergence for the parameters σ^2, q_2, q_1, when the true values are p_1=1, p_2=10 (q_2=11, q_1=10). Comparison of the results presented in Table 3 illustrates the statement that the random variable σ^2 is learned faster than the other parameters. Hence, relative to the estimation of the parameters (q_1, q_2) the parameter σ^2 is a constant. The effect of the sampling interval on the parameters q_2 and q_1 is illustrated in Table 4. As can be seen from comparing the results in Table 3 to Table 4, the increase in sampling interval (h) slows the convergence on the parameters σ^2 and q_2 while increasing the

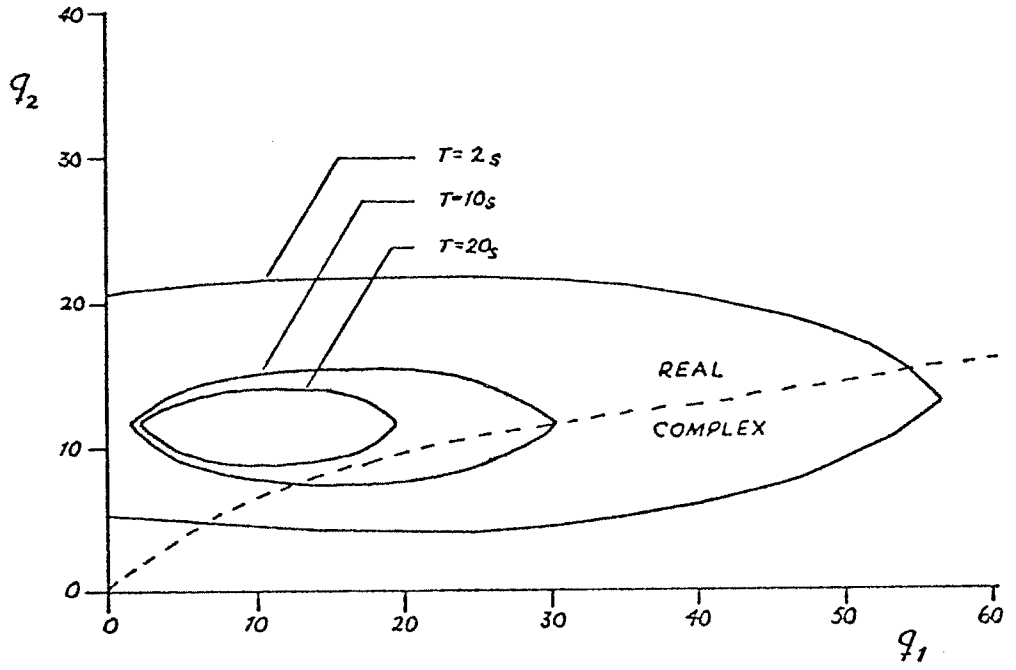

Figure 13. Pr = .95 Contours for True Values q_1=10, q_2=11, at Sampling Interval h = $2x10^{-3}$ for Various Observation Intervals (T).

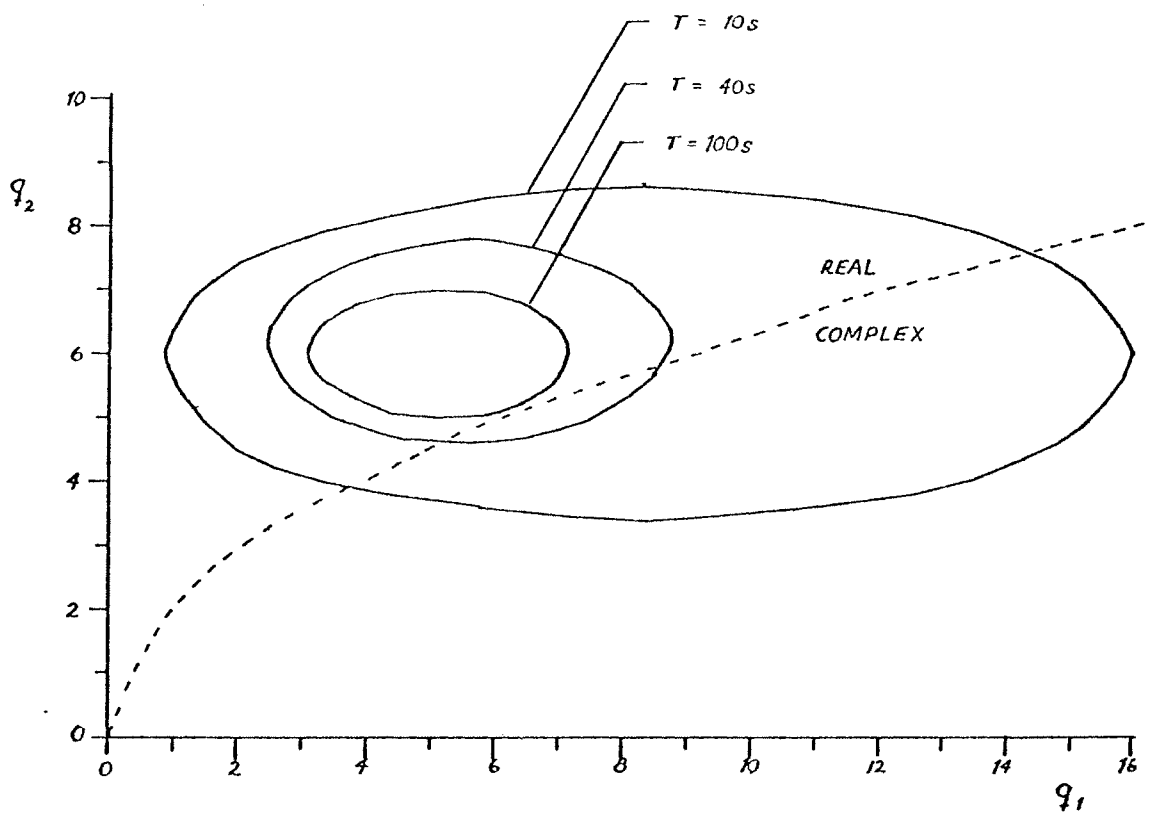

Figure 14. Pr = .95 Contours for True Values q_1=5, q_2=6, at Sampling Interval h = $1x10^{-2}$ for Various Observation Intervals (T)

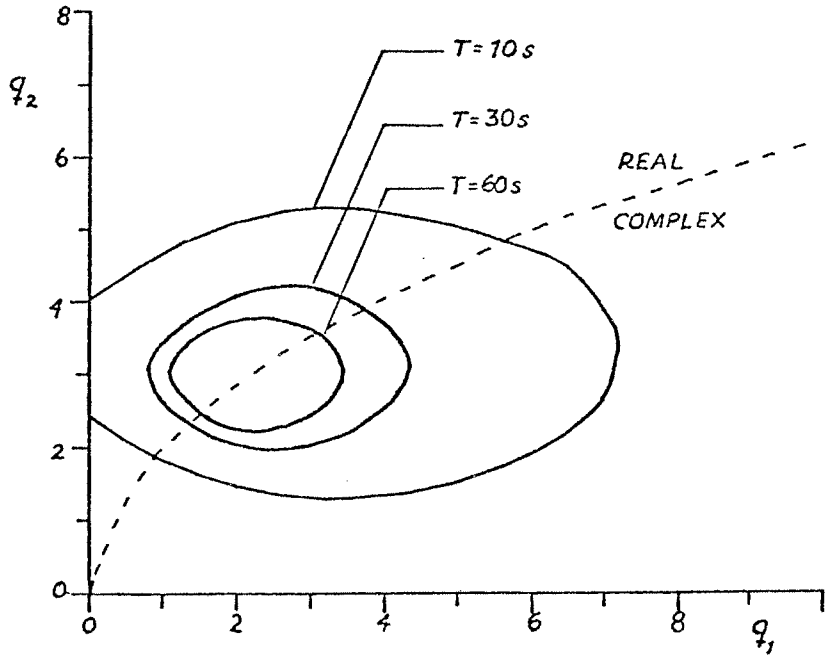

Figure 15. Pr = .95 Contours for True Values q_1=2, q_2=3, at Sampling Interval h = $1x10^{-2}$ for Various Observation Intervals (T).

TABLE 3

MEAN TO STANDARD DEVIATION RATIO AS A FUNCTION OF TIME FOR TRUE VALUES $q_1=10$, $q_2=11$, AT SAMPLING RATE $h = 2\times10^{-3}$.

T(sec)	$(\mu/\sigma)_{2^{B_N}}$	$(\mu/\sigma)_{q_1}$	$(\mu/\sigma)_{q_2}$
2	19	1.2	3.6
10	42	2.7	7.5
20	58	3.1	10.8

TABLE 4

MEAN TO STANDARD DEVIATION RATIO AS A FUNCTION OF TIME FOR TRUE VALUES $q_1=10$, $q_2=11$, AT SAMPLING RATE $h = 1\times10^{-2}$.

T(sec)	$(\mu/\sigma)_{2^{B_N}}$	$(\mu/\sigma)_{q_1}$	$(\mu/\sigma)_{q_2}$
10	19	2.9	7.4
20	26	3,6	10.3
30	33	4.4	12.5

TABLE 5

MEAN TO STANDARD DEVIATION RATIO AS A FUNCTION OF TIME FOR TRUE VALUES $q_1=2$, $q_2=3$, AT SAMPLING RATE $h = 1\times10^{-2}$.

T(sec)	$(\mu/\sigma)_{2^{B_N}}$	$(\mu/\sigma)_{q_1}$	$(\mu/\sigma)_{q_2}$
10	19	2.3	4.1
30	32	3.6	6.7
60	44	4.8	9.4

convergence on q_1. The results presented in Table 5 illustrate the rate of convergence for the parameters when a different structure is assumed. Predictably, the parameter σ^2 ($_2B_N$) converges rapidly as in the other results illustrated, but due to the high sampling rate, the other parameters converge more slowly since there is a greater low frequency content in the process.

7.4 An Alternative Method of Presenting the Results

One of the basic difficulties associated with presenting the results of the estimation problem in terms of confidence contours, is the absence of a direct method of relating the accuracy of the parameters $\{\sigma^2, p_i\}$ to the accuracy of the spectral representation constructed from these parameters. One approach to this problem is to present the estimated spectral representation by evaluation of the parametric form

$$S(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M} (\omega^2 + p_i^2)} \quad , \tag{7.43}$$

with the mean values and specified extremes substituted. For the M=1 and M=2 processes under consideration, this is a relatively straight forward, but tedious approach. For higher order processes, the mapping from the parameter space into the two dimensional spectral domain via this alternative is not readily amenable to analysis.

Another alternative is possible through the use of an equivalent low pass representation. This equivalent representation is described by an effective gain (G_E) and bandwidth (B_E) where

$$G_E = \max_{\omega} S(\omega) = S(0) \tag{7.44}$$

and

$$2B_E = \frac{1}{2\pi}\left[\frac{\int S(\omega)\,d\omega}{G_E}\right] = \frac{R(0)}{G_E} \quad . \tag{7.45}$$

For the M=2 process described by

$$S(\omega) = \frac{\sigma^2}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)} \quad , \tag{7.46}$$

the effective gain and bandwidth are (provided Re $p_k \geq |\text{Im } p_k|$)

$$G_E = \frac{\sigma^2}{(p_1 p_2)^2} = \frac{\sigma^2}{q_1^2} \quad , \tag{7.47}$$

$$2B_E = \frac{p_1 p_2}{2(p_1+p_2)} = \frac{q_1}{2q_2} \quad . \tag{7.48}$$

Introduction of this representation provides a means of obtaining an equivalent Time-Bandwidth product (TB_E). Blackman and Tukey (1959) show that a good description of the stability of a positive estimate is the equivalent number of degrees of freedom of a Chi-squared distribution, where

$$2TB_E = \text{degrees of freedom} \tag{7.49}$$

This product then provides a measure of the length of the observation interval (T) necessary to obtain a specified quality of the spectral estimate over the frequency range of interest. The basic assumption in using the Chi-Squared distribution is the independence of the observations. The independence assumption is not valid here, however, the use of a TB_E can still provide a guide to planning the length of the observation interval, so that a specified quality of the estimate $\hat{S}(\omega)$ is obtained. The use of the Time-Bandwidth product also provides a means of comparing the variability of the overall estimate without regard to structural distinction. For smaller B_E, larger observation intervals must be used to obtain a prespecified quality of

the estimate.

To see how the TB_E product relates to the accuracy of the parameters being estimated, the mean and variance relationships, equations (6.39 and (6.40), are employed. Interpreting the equivalent bandwidth as

$$2B_E = \frac{\overline{q_1}}{2\overline{q_2}} = \frac{(\overline{p_1 p_2})}{2(\overline{p_1+p_2})}$$

$$= \frac{\int \dot{x}^2(t)dt}{2\int x^2(t)dt} \cdot \frac{2\int \dot{x}^2(t)dt}{\sigma^2 T} , \qquad (7.50)$$

then

$$2TB_E = \frac{\int \dot{x}^2(t)dt}{\sigma^2 \int x^2(t)dt} . \qquad (7.51)$$

Recognizing the factors in equation (7.51) that relate to the variance of the parameters, the TB_E product is

$$2TB_E = \frac{Var(q_1)}{(Var(q_2))^2} . \qquad (7.52)$$

From this relationship, it is apparent that the TB_E product does provide a measure of accuracy, related to the parameter accuracy. How this measure relates to the overall spectral accuracy will be demonstrated empirically. The results presented will provide a more appealing demonstration of the estimation technique, than those obtained through the presentation of confidence contours in the abstract (q_1,q_2) plane.

The empirical presentation consists of an envelope of possible spectral representations obtained from the 95% contours at different observations intervals. This envelope results from the asymptotes of the spectral density specified by the pole locations, which are

determined by the relationships

$$q_1 = p_1 p_2 \quad ,$$

$$q_2 = p_1 + p_2 \quad .$$

The spectral densities prescribed, are obtained from the data contained in Tables 6 and 7 , which in some cases have not been illustrated in a contour format. The contours are readily determined through the application of equations (7.40) and (7.41).

Figures 16 through 21 illustrate the spectral envelope for a sequence of Time-Bandwidth products. A cursory comparison of these results reveals that the greatest spread of the asymptotes occurs over the low frequency range. The behavior of the asymptotes reflects the general characteristic of the estimation procedure. The high frequency asymptotes compress rapidly, whereas the reduction in low frequency variability requires longer observation intervals. From this observation it is apparent that the length of the observation interval controls the low frequency behavior. Comparison of the different structures (pole pairs) presented, with a fixed TB_E product, illustrates that the variability of the estimate is directly related to this product.

7.5 Solving the Constrained Problem

Up to this point, the results that have been presented represent the contours of the unconstrained solutions. Recall that the model being used required that the parameters $q_2(p_1 + p_2)$ and $q_1(p_1 p_2)$ be restricted to positive values. As mentioned in Chapter V, a direct attempt at solving the constrained problem analytically is difficult and provides little insight as to the behavior of the estimation

TABLE 6

PERTINENT DATA FOR SPECTRAL DENSITY COMPUTATION FOR THE TRUE VALUES $q_1=10$, $q_2=11$, AND SAMPLING RATE $h = 1/100$.

T(sec)	$2TB_E$	$\bar{q}_1$	Var q_1	$\bar{q}_2$	Var q_2
30	$13.\overline{63}$	13.1	9.5	10.9	.73
60	$27.\overline{27}$	11.4	4.3	11.1	.37
90	$40.\overline{90}$	10.97	2.6	10.9	.24

TABLE 7

PERTINENT DATA FOR SPECTRAL DENSITY COMPUTATION FOR TRUE VALUES $q_1=5$, $q_2=6$, AND SAMPLING RATE $h = 1/100$.

T(sec)	$2TB_E$	$\bar{q}_1$	Var q_1	$\bar{q}_2$	Var q_2
30	$12.\overline{5}$	6.51	2.6	6.0	.40
70	$29.1\overline{6}$	5.3	.96	6.4	.18
100	$41.\overline{6}$	5.1	.62	6.1	.12

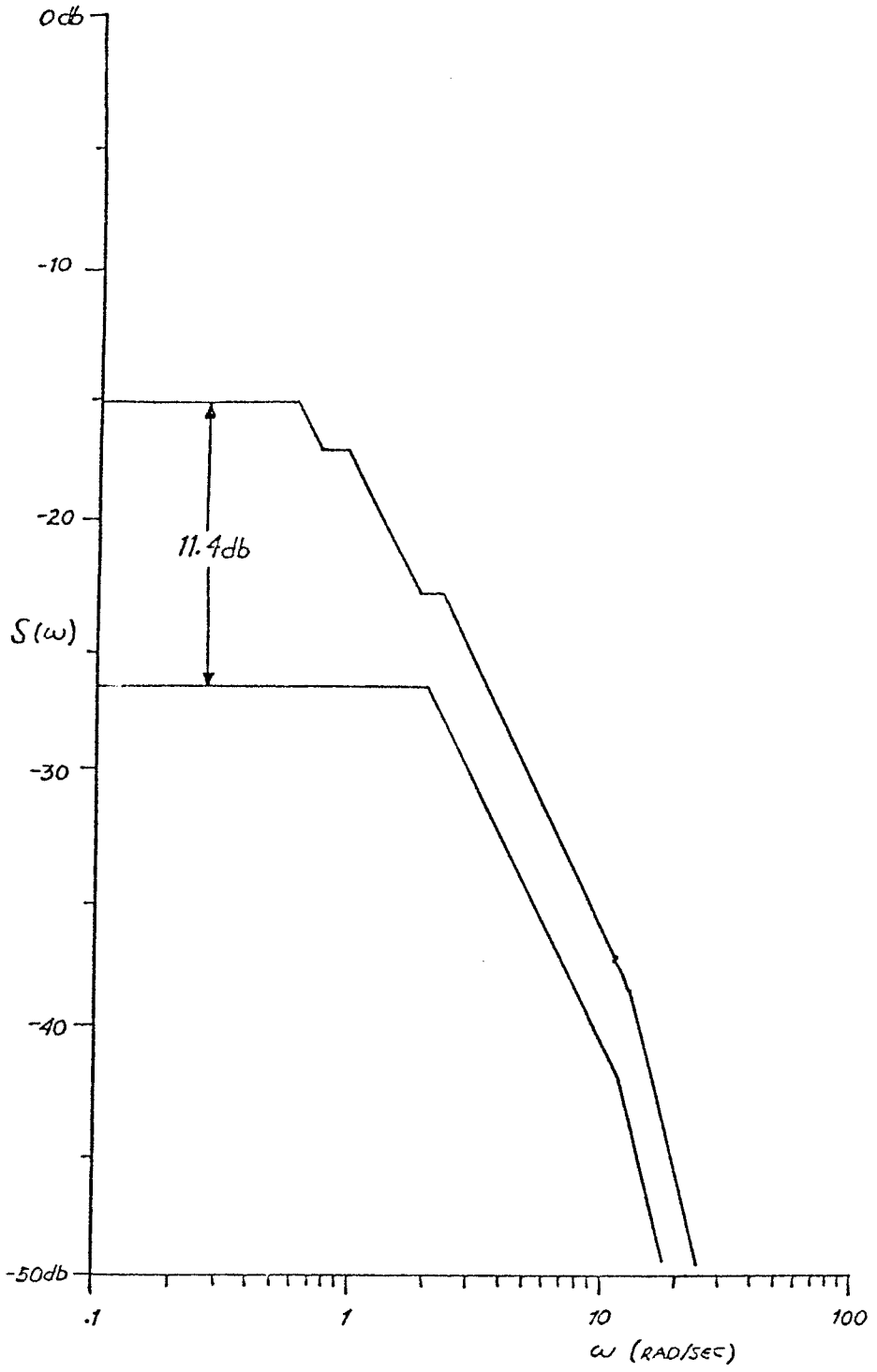

Figure 16. Spectral Envelope for True Values q_1=10, q_2=11, at $2TB_E$ = 13.63

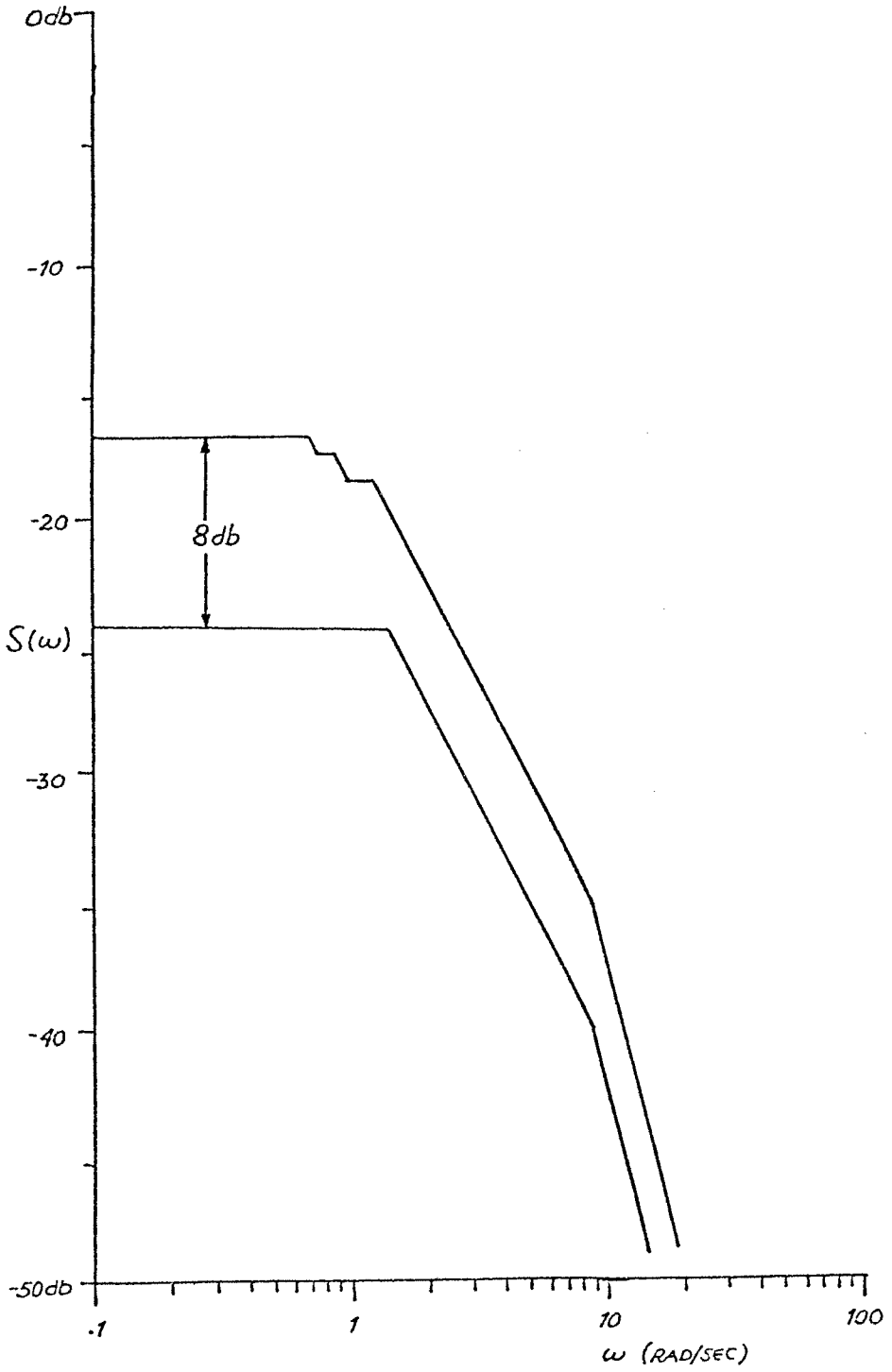

Figure 17. Spectral Envelope for True Values q_1=10, q_2=11 at $2TB_E$ = 27.

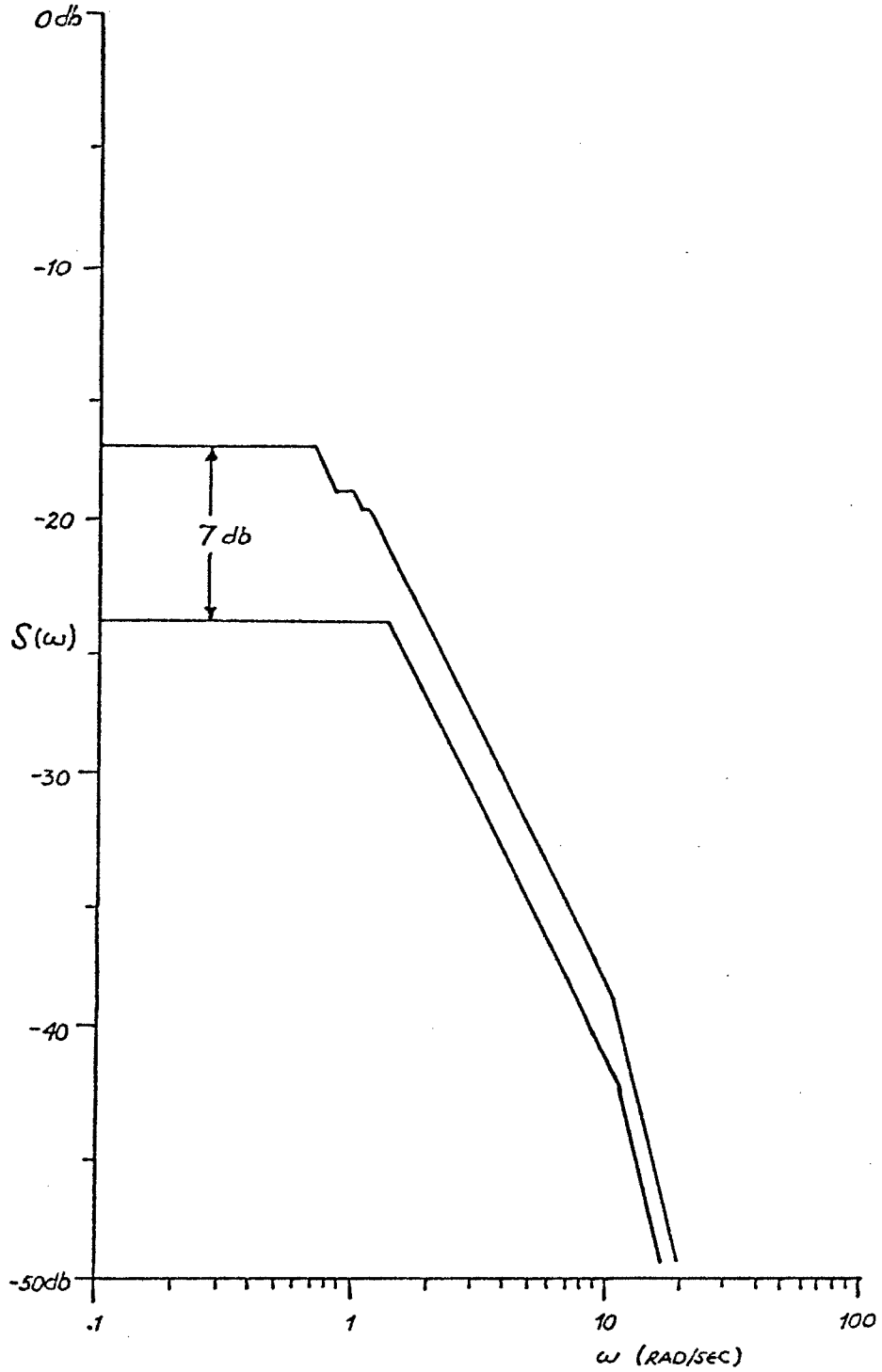

Figure 18. Spectral Envelope for True Values q_1=10, q_2=11, at $2TB_E = 41$.

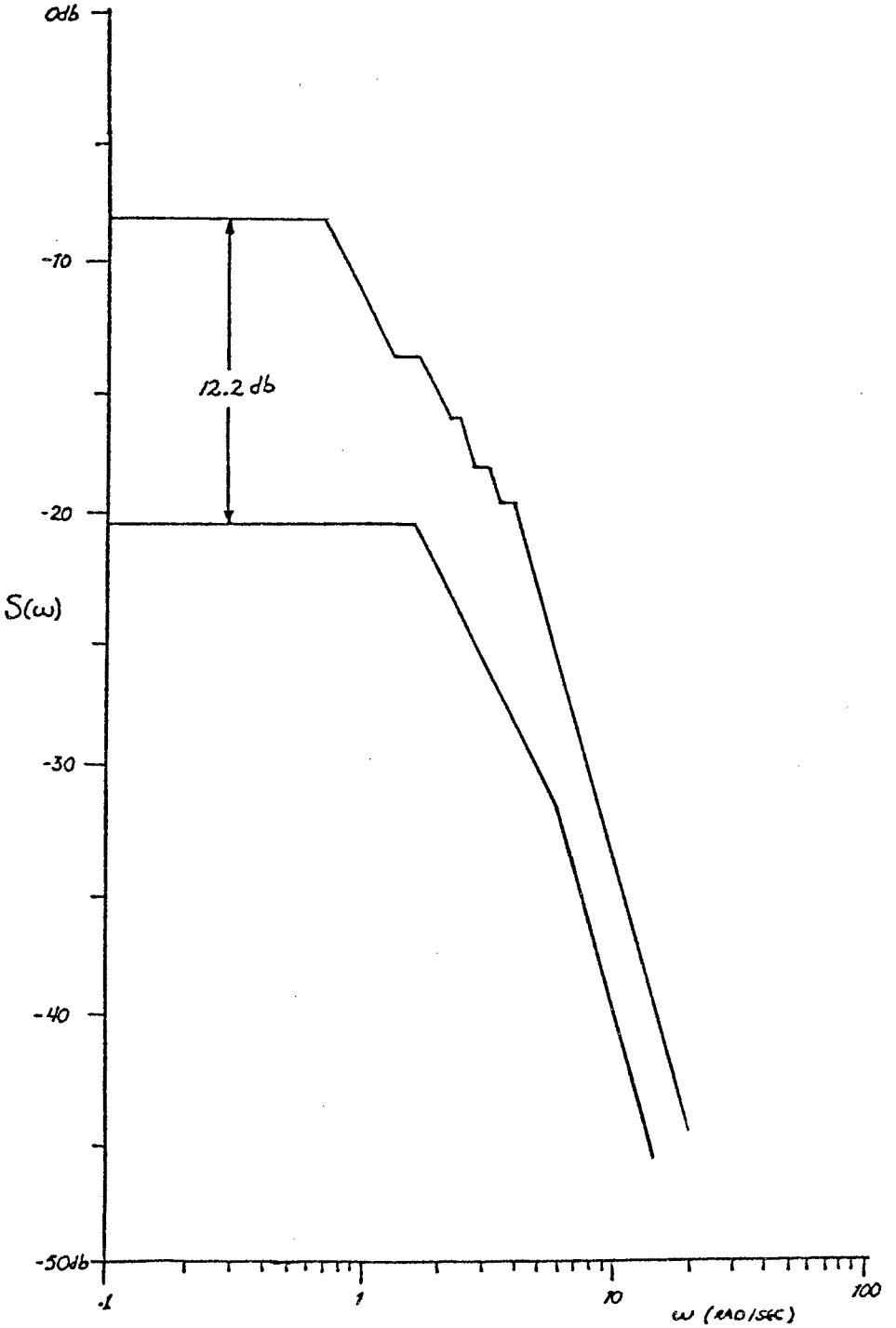

Figure 19. Spectral Envelope for True Values q_1=5, q_2=6, at $2TB_E$ = 12.5

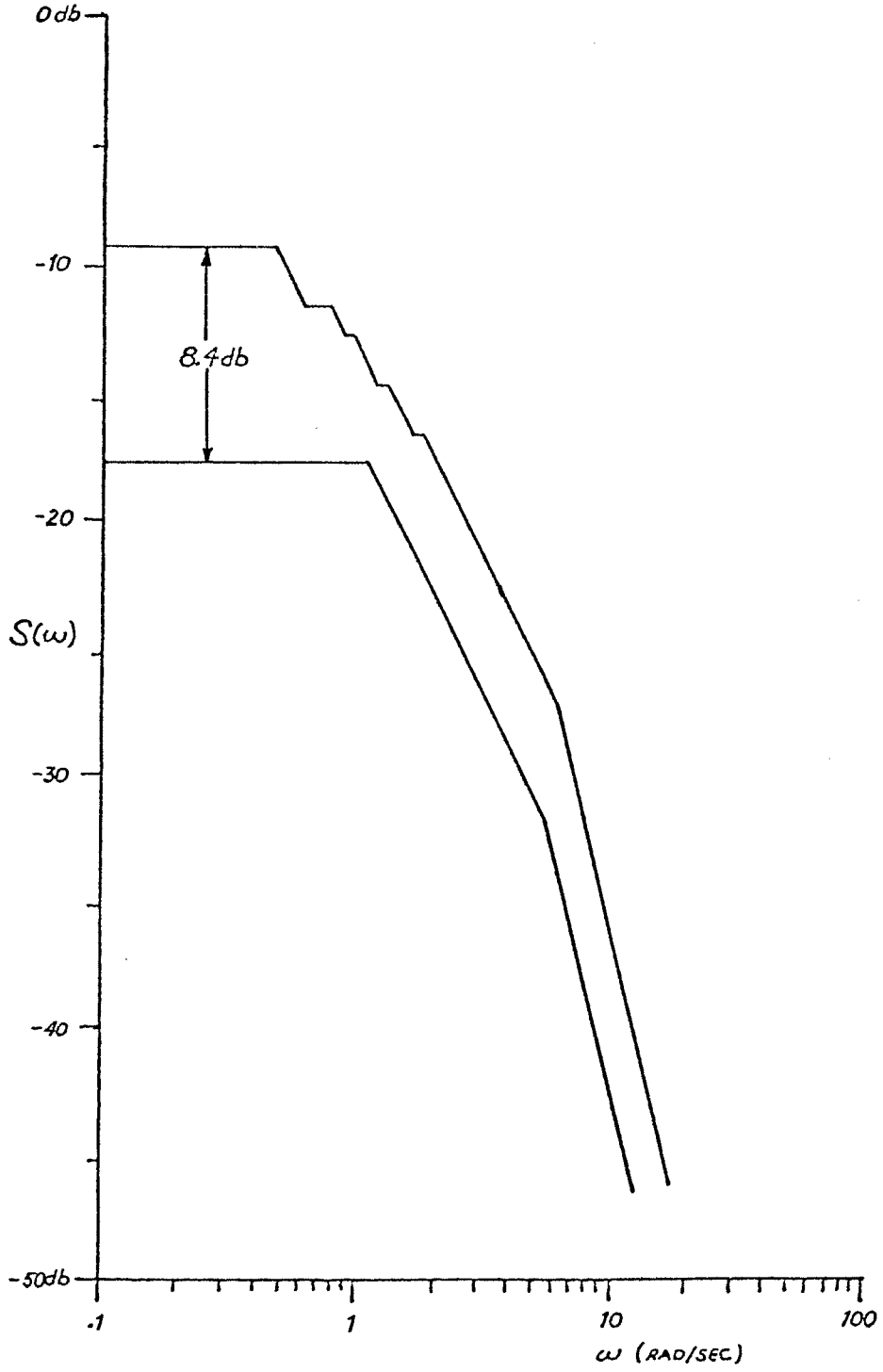

Figure 20. Spectral Envelope for True Values q_1=5, q_2=6, at $2TB_E = 29$

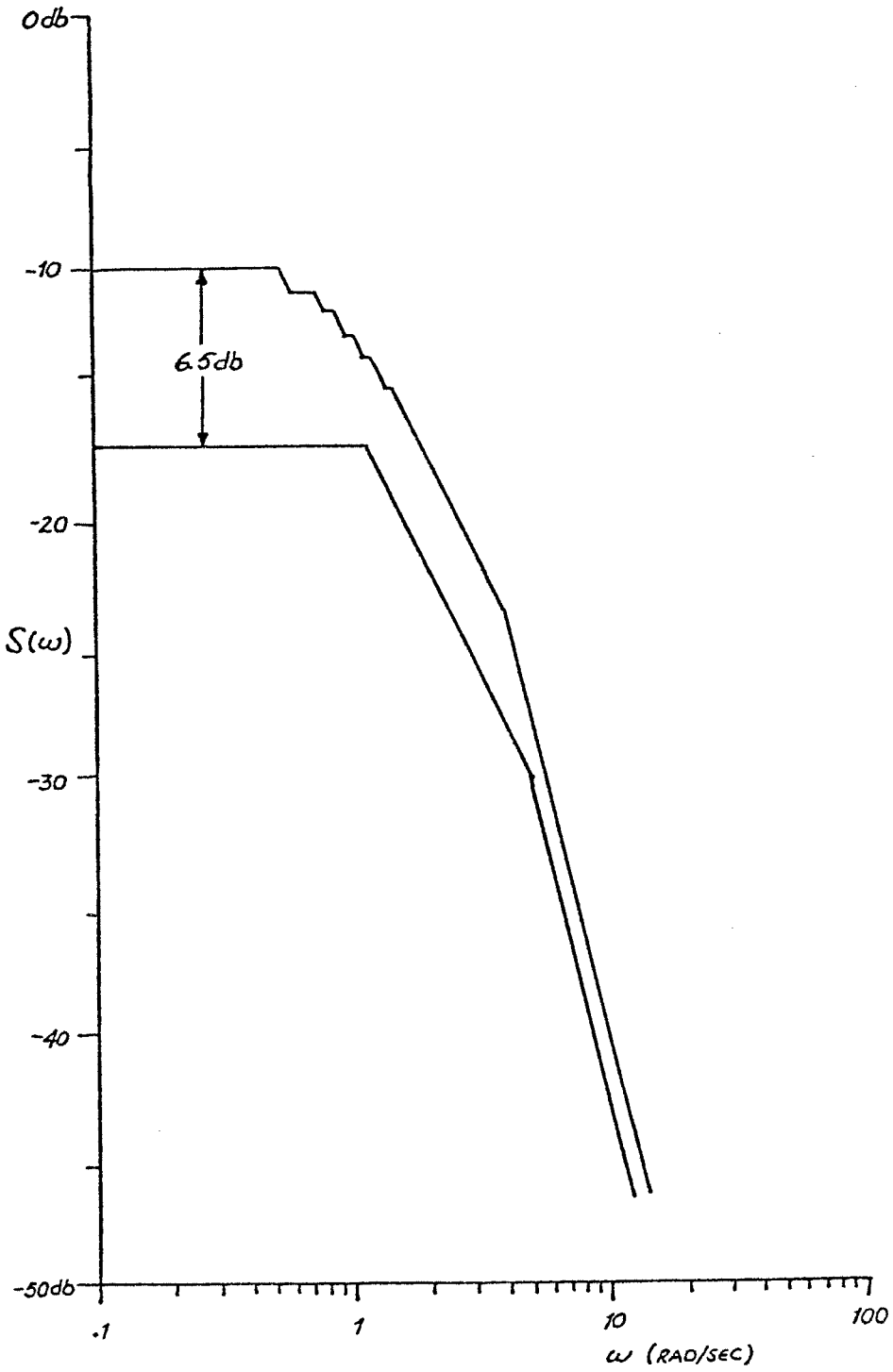

Figure 21. Spectral Envelope for True Values q_1=5, q_2=6, at $2TB_E$ = 42

results. It was proposed that an alternate approach, which is the intent of this solution to demonstrate, be followed. First, solve the unconstrained problem and hence find the respective post probability density functions, and then modify this solution to reflect the constraint. This approach will be illustrated for the results presented in Figure 13. Reference to this figure shows that for short observations intervals (T=2) the unconstrained post density has a significant portion of its 95% contour in the negative q_1 direction. As the observation interval becomes longer, the unconstrained solution becomes more concentrated in the positive q_1, q_2 directions. We find that the constrained and unconstrained solutions are the same. Recall that the constrained and unconstrained densities are related by

$$f_c(\underline{q}|y) = f_u(\underline{q}|y) \cdot S(\underline{q}) \quad ,$$

where $f_c(\cdot|\cdot)$ and $f_u(\cdot|\cdot)$ are the constrained and unconstrained post densities. The modifying factor $S(\underline{q})$ is defined as

$$S(\underline{q}) = \begin{cases} \left(\Pr(\underline{q} \geq 0|y)\right)^{-1} & \underline{q} \in \Omega_c \\ 0 & \underline{q} \notin \Omega_c \end{cases} ,$$

where Ω_c is the set of possible constrained values. The problem is somewhat simplified since the parameters q_1 and q_2 are independent, and therefore, the problem of determining the constrained solution is reduced to one of determining two separate constrained solutions. For the particular case under consideration, the modifying factor $S(q_i)$ is found from

$$\left(S(q_i)\right)^{-1} = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\mu_i}{\sigma_i}}^{\infty} e^{-t^2/2} dt = \Pr(q_i \geq 0|y) \quad .$$

As can be seen from the above relation, the modifying factor is close

to unit value for large μ_i/σ_i. When the modifying factor is close to one, the truncation of the density has a negligible effect and it can be concluded that the constrained and unconstrained solutions are practically equivalent. A summary of the parameters required to construct the constrained post density for the data being considered is presented in Table 8. The modifying factors for both densities are seen to approach unit value as the observation interval increases, so it is apparent that for $T \geq 10$ sec the contours presented in Figure 13 represent the constrained solution.

7.6 Effects of Independent Quantizing Noise

The introduction of independent quantizing noise is a means of providing a realistic constraint that would be encountered with a hardware realization of the processing previously described. The noise will be considered to be uniformly distributed with zero mean, and variance σ_e^2, where the variance is related to the quantizing step size (Q) by

$$\sigma_e^2 = Q^2/12 \quad . \tag{7.53}$$

The most pronounced effect of the noise will be felt by the processing used to determine the random variable σ^2. The effect on the M=1 processing is determined by evaluating the following expressions

$$E(\hat{B}_N) = E\left[\sum_{k=1}^{N} \underline{\Delta}\hat{x}_k^2\right] = E\sum_{k=1}^{N}\left[(x_k + \varepsilon_k)-(x_{k-1} + \varepsilon_{k-1})\right]^2 \tag{7.54}$$

$$Var(\hat{B}_N) = Var\left[\sum_{k=1}^{N} \underline{\Delta}\hat{x}_k^2\right] , \tag{7.55}$$

where ε_j represent the noise that produces the corrupted increment $\underline{\Delta}x_k$. Under the assumption of independence, the mean and variance

TABLE 8

POST DENSITY PARAMETER AND MODIFYING FACTORS FOR TRUE VALUES q_1=10, q_2=11, AT SAMPLING h = 2×10^{-3}.

T(sec)	$E(q_1)$	$\sqrt{\text{Var } q_1}$	$S(q_1)$	$E(q_2)$	$\sqrt{\text{Var } q_2}$	$S(q_2)$
2	18.2	15.3	1.13	12.9	3.6	1.006
10	16.1	6.1	1.003	11.4	1.5	1
20	10.8	3.5	1.001	11.3	1.05	1

calculations presented in Appendix F, result in

$$E(\hat{B}_N) = E(B_N) + 2N\sigma_e^2 \tag{7.56}$$

$$\text{Var } (\hat{B}_N) = \text{Var } (B_N) + 8\sigma_e^2 E(B_N) + 5.6\ N\sigma_e^4 \tag{7.57}$$

To provide an interpretation of these calculations, assume that the sampling of the observations is sufficiently fast ($p\delta \ll 1$), so that

$$\begin{aligned} E(\hat{B}_N) &= E(B_N) + 2N\sigma_e^2 \\ &= \sigma^2 T + 2N\sigma_e^2 \\ &= \sigma^2 T\left(1 + (2\sigma_e^2/h\sigma^2)\right) \\ &= \sigma^2 T\left(1 + 2F_s\sigma_e^2/\sigma^2)\right) \\ &= E(B_N)(1 + 2F_s\sigma_e^2/\sigma^2) \quad , \end{aligned} \tag{7.58}$$

where $F_s = 1/\delta$ is the sampling frequency, and

$$\begin{aligned} \mathrm{Var}(\hat{B}_N) &= \mathrm{Var}(B_N) + 8\sigma_e^2 E(B_N) + 5.6N\sigma_e^4 \\ &= 2\,\frac{(\sigma^2 T)^2}{N}\left[1 + \frac{4N\sigma_e^2}{\sigma^2 T} + \frac{5.6N^2\sigma_e^4}{2\sigma^4 T^2}\right] \\ &= 2\,\frac{(\sigma^2 T)^2}{N}\left[1 + \frac{4\sigma_e^2}{\sigma^2}F_s + \frac{5.6}{2}\left[\frac{\sigma_e^2}{\sigma^2}\right]^2 F_s^2\right] \quad . \end{aligned} \tag{7.59}$$

From these expressions it is evident that there is an upper limit on the sampling frequency that limits the recovery of the random parameter σ^2. Further insight into these relationships can be gained by seeking the sampling frequency (F_s) that minimizes $\mathrm{Var}(\hat{B}_N)$. Rewriting equation (7.59) in the form

$$\mathrm{Var}(\hat{B}_N) = \frac{2\sigma^4 T}{F_s^2}\left[1 + \frac{5.6}{2}\left[\frac{\sigma_e^2}{\sigma^2}\right]F_s^2 + 4\left[\frac{\sigma_e^2}{\sigma^2}\right]F_s\right] ,$$

and solving

$$\frac{d}{dF_s}\mathrm{Var}(\hat{B}_N) = \frac{2\sigma^4 T}{F_s^2}\left[5.6\left[\frac{\sigma_e^2}{\sigma^2}\right]F_s^2 + 4\left[\frac{\sigma_e^2}{\sigma^2}\right] - 1 + \frac{5.6}{2}F_s^2\left[\frac{\sigma_e^2}{\sigma^2}\right]^2 + 4\left[\frac{\sigma_e^2}{\sigma^2}\right]F_s\right] = 0,$$

for F_s, the result is

$$F_{s_{min}} = \frac{\sqrt{2}}{\sqrt{5.6}}\,\frac{\sigma^2}{\sigma_e^2} \quad .$$

By specifying the method of setting the quantizing step size, a more meaningful relationship for the minimizing sampling frequency ($F_{s_{min}}$) can be obtained. The quantizer step size is set to three times the expected rms voltage, as a result, the step size (Q) is related to the power of the observations (R_O) by

$$Q = \frac{3\sqrt{R_O}}{L/2} \quad , \tag{7.60}$$

where L is the number of quantization levels. As a result of this specification, the noise power is

$$\sigma_e^2 = \frac{1}{12}\left(\frac{3\sqrt{R_o}}{L/2}\right)^2$$

$$= \frac{3R_o}{L^2} \quad . \tag{7.61}$$

If for the power R_o, the expression for the first order process is used, the noise power is

$$\sigma_e^2 = \frac{3}{2pL^2}\sigma^2 \quad , \tag{7.62}$$

where

$$R_o = \frac{\sigma^2}{2p} e^{-p|\tau|}\Big|_{\tau=0} \quad ,$$

has been used in equation (7.61). With this interpretation, the minimizing sampling frequency takes the form

$$F_{s_{min}} = \sqrt{\frac{2}{5.6}}\left(\frac{\sigma^2}{\sigma_e^2}\right)$$

$$= \sqrt{\frac{2}{5.6}}\left(\frac{2}{3}pL^2\right) \quad . \tag{7.63}$$

For an 8 bit (L=256) quantizer the frequency $F_s$min would be on the order of $2\cdot 10^4 p$. The variance expression, repeated here for convenience,

$$\text{Var}\,(\hat{B}_N) = \frac{2\sigma^4 T}{F_s}\cdot\left\{1+\frac{5.6}{2}\left(\frac{\sigma e^2}{\sigma^2}\right)^2 F_s^2 + 4\left(\frac{\sigma e^2}{\sigma^2}\right)F_s\right\}$$

$$= \frac{2\sigma^4 T^2}{N}\left\{1+\frac{5.6}{2}\left(\frac{\sigma e^2}{\sigma^2}\right)F_s^2 + 4\left(\frac{\sigma e^2}{\sigma^2}\right)F_s\right\}$$

is suggestive of the qualitative representation in Figure 22. In this figure f_1 represents the frequency which satisfies the condition $F_s >> p$ $(p\delta << 1)$ and f_2 represents the frequency F_{sMIN},

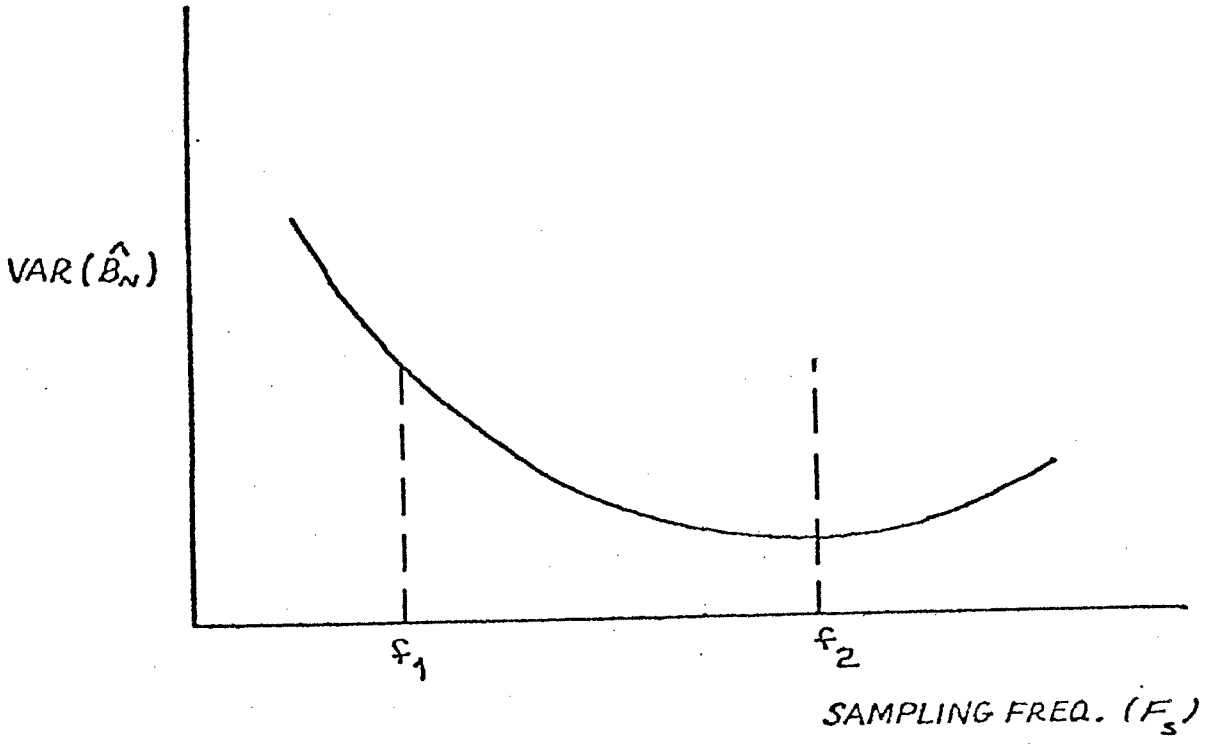

Figure 22. Variance of $\hat{B}_N$ Versus Sampling Frequency.

where the variance takes the value

$$\text{Var}\ (\hat{B}_N)\Big|_{F_s = F_{sMIN}} = \text{Var}\ (B_N)\ \{4.4\} \tag{7.64}$$

At this frequency the mean of the random variable B_N is

$$E(\hat{B}_N)\Big|_{F_s = F_{sMIN}} = E(B_N)\ (2.19).$$

So the random variable $\hat{B}_N$ as it converges to its mean provides a biased estimate of the random variable σ^2 .

For observations processes of an order higher than M=1, the derivative process of the observation is needed. The formation of this derivative can be achieved through either analog or digital means. The distinction between these two methods is reflected in the point at which quantization error is modeled . For digital differentiation, the noise is injected prior to the formation of the derivative, as illustrated in Figure 23. In analog differentiation, the quantizing effects can be considered to be independent of the observations and differentiation. With this in mind, the representation for analog differentiation takes the form illustrated in Figure 24.

Analysis of the digital differentiation for M=2, follows the same pattern as the analysis for the observations in the M=1 model, which allows the elimination of duplication. Define the following

$$ {}_1\hat{B}_N = \sum_{i=1}^{N} (\underline{\Delta}\hat{x}_k)^2 \quad , $$

$$ {}_2\hat{B}_N = \sum_{i=1}^{N} (\underline{\Delta}\dot{\hat{x}}_k)^2 \quad , $$

where

$$ \hat{x}_k = x_k + \varepsilon_k \quad , $$

$$ \dot{\hat{x}}_k = \dot{x}_k + (\varepsilon_k - \varepsilon_{k-1}) / h \quad , $$

h = sampling interval ,

$\sigma^2_{e_1}$ = quantizing noise power (N_1) .

Injection of the quantizing noise into the observations produces the result

$$ E({}_1\hat{B}_N) = E({}_1B_N) + 2N\sigma^2_{e_1} \quad , \tag{7.65} $$

$$ Var({}_1\hat{B}_N) = Var({}_1B_N) + 8\sigma^2_{e_1}E({}_1B_N) + N(5.6)\sigma_{e_1}{}^4 \quad , \tag{7.66} $$

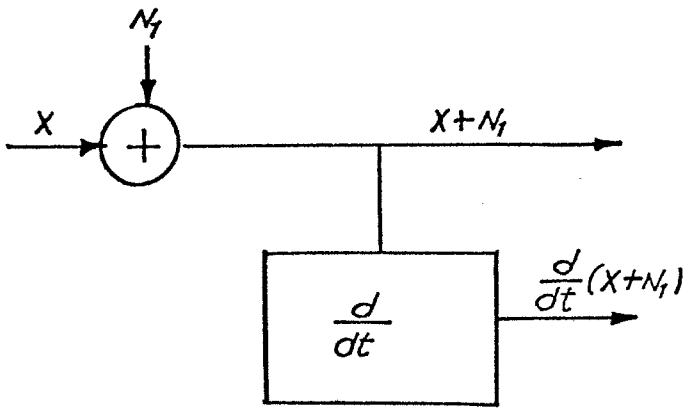

Figure 23. Pictorial Representation of Digital Differentiation.

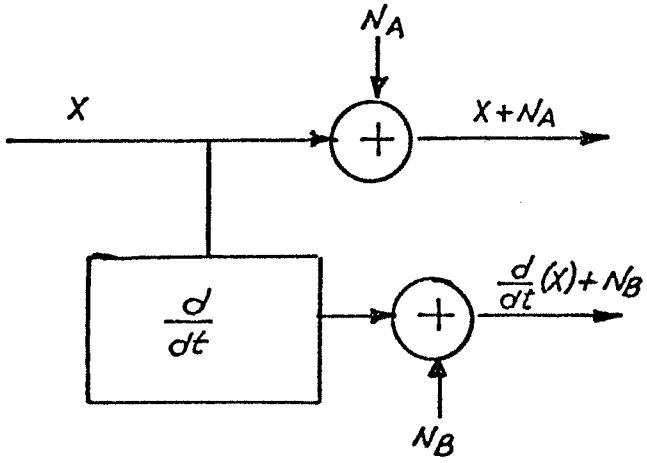

Figure 24. Pictorial Representation of Analog Differentiation.

where under fast sampling conditions $E({}_1B_N) \simeq 0$. The same noise source is now involved in the formation of the derivative process. The result of the digital differentiation yields

$$E({}_2\hat{B}_N) = E({}_2B_N) + 6(T/h^3)\sigma_{e_1}^2$$
$$= E({}_2B_N)\left(1 + 6F_s^3(\sigma_{e_1}^2/\sigma^2)\right) , \tag{7.67}$$

$$\mathrm{Var}({}_2\hat{B}_N) = \mathrm{Var}({}_2B_N) + 24\left(\frac{\sigma_{e_1}^2}{h^2}\right)E({}_2B_N) + 18\left(\frac{N}{h^4}\right)\sigma_{e_1}^4 \quad 2.8$$

$$= \mathrm{Var}({}_2B_N)\left\{1 + 12F_s^3\left(\frac{\sigma_{e_1}^2}{\sigma^2}\right) + 9\left(\frac{\sigma_{e_1}^2}{\sigma^2}\right)^2 2.8\frac{F_s^5}{T}\right\}$$

(7.68)

As is readily observable in the equation for $\mathrm{Var}({}_N)$, digital differentiation places a very stringent condition on the sampling frequency. It is evident that injection of the quantizing noise prior to forming the derivative makes the recovery of the quadratic variation impossible. On the other hand, analog differentiation will ease the recovery process, as will be demonstrated next. For the observations, define

$${}_1\hat{B}_N = \sum(\Delta\hat{x}_k)^2$$

where

$$\hat{x}_k = x_k + \varepsilon_{k_A} ,$$
$$\mathrm{Var}(\varepsilon_{k_A}) = \sigma_{e_A}^2 .$$

For the derivative process, define

$${}_2B_N = \sum(\Delta\hat{\dot{x}}_k)^2$$

where

$$\hat{\dot{x}}_k = \hat{\dot{x}}_k + \varepsilon_{k_B} ,$$
$$\mathrm{Var}(\varepsilon_{k_B}) = \sigma_{e_B}^2 .$$

Carrying out the manipulations to find the mean and variance of these newly defined terms, results in

$$E({}_1\hat{B}_N) = E({}_1B_N) + 2N\sigma^2_{e_A} \simeq 2N\sigma^2_{e_A} , \tag{7.69}$$

$$Var({}_1\hat{B}_N) = Var({}_1B_N) + 8\sigma^2_{e_A}E({}_1B_N)+5.6N\sigma^4_{e_A} \tag{7.70}$$

For the derivative process, the result is

$$\begin{aligned} E({}_2\hat{B}_N) &= E({}_2B_N) + 2N\sigma^2_{e_B} \\ &= E({}_2B_N)\left(1 + 2F_S(\sigma^2_{e_B}/\sigma^2)\right) \\ &= E({}_2B_N)\left(1 + 2(\sigma^2_{e_B}/\sigma^2 h)\right) , \end{aligned} \tag{7.71}$$

$$\begin{aligned} Var({}_2\hat{B}_N) &= Var({}_2B_N) + 8\,\sigma_{e_B}{}^2\,E({}_2B_N) + 5.6\,N\,\sigma_{e_B}{}^4 \\ &= Var({}_2B_N)\left\{1 + \frac{5.6}{2}\left(\frac{\sigma_{e_B}{}^2}{\sigma^2}\right)F_s{}^2 + 4\left(\frac{\sigma_{e_B}{}^2}{\sigma^2}\right)F_s\right\} \end{aligned} \tag{7.72}$$

Recalling that the random variable σ^2 is to be estimated (for M=2), from processing of ${}_2\hat{B}_N = \sum(\Delta\hat{\dot{x}}_k)^2$. The quantizing error is dependent on the square of the sampling frequency. Hence the same relationship as in Figure 22 is suggested. With analog differentiation there is a range of sampling frequencies for which the variance $\left(Var({}_2B_N)\right)$ is independent of second order effects, and quantization error. A qualitative interpretation of when the quantizing effects are irrelevant is available from equation (7.72)above. Viewing the term $(\sigma^2_e/\sigma^2 h)$ as a ratio of the quantizing power to desired power out of the first difference filter it is clear that as long as the quantizing noise power is an order of magnitude less than the desired power out $\sigma^2 h$, the quantizing noise has a negligible effect.

Another, more useful, interpretation can be provided when the

particular method used for setting the quantization step size is considered. As previously mentioned, the step size for the simulation study was based on a power measurement which, when incorporated into the noise variance computation, yields the result

$$\sigma^2_{e_B} = \frac{Q^2}{12} = \frac{3R_{\dot{x}}(0)}{L^2}$$

where

$$R_{\dot{x}}(0) = \frac{\sigma^2}{2(p_2^2 - p_1^2)} \left[p_2 e^{-p_2|T|} - p_1 e^{-p_1|T|}\right]\Big|_{T=0} = \frac{\sigma^2}{2(p_2+p_1)}$$

Using the true value for the power in the derivative process $R_{\dot{x}}(0)$, substituted into the mean and variance expressions, equations (7.71) and (7.72), yields

$$E(_2\hat{B}_N) = E(_2B_N)\left(1 + \frac{6}{L^2}\frac{R_{\dot{x}}(0)}{\sigma^2} F_s\right)$$

$$= E(_2B_N)\left(1 + \frac{F_s}{L^2} \cdot \frac{3}{(p_2 + p_1)}\right) \qquad (7.73)$$

$$Var(_2\hat{B}_N) \cong Var(_2B_N)\left\{1+5.6F_s^2\left[\frac{6R_{\dot{x}}(0)}{L^2\sigma^2}\right]^2\right\}$$

$$\approx Var(_2B_N)\left\{1+5.6F_s^2\left[\frac{3}{L^2(p_2 + p_1)}\right]^2\right\} . \qquad (7.74)$$

A simulation using poles at p_1=1, p_2= 10 rad/sec; ie. q_1=10, q_2=11 rad/sec and sampling rates of F_s= 100 and 1000 hertz was conducted. The results of the simulation studies of quantizing effects are presented in Figures 25 and 26 , where the mean to standard deviation ratio $\left(\frac{\mu_{2B_N}}{\sigma_{2B_N}}\right)$ is depicted for various numbers of quantizing levels (L) versus the number of observations (N). The distinction between these figure is the sampling rate (frequency). Both figures display the μ/σ ratio based on sample averages, but Figure 26

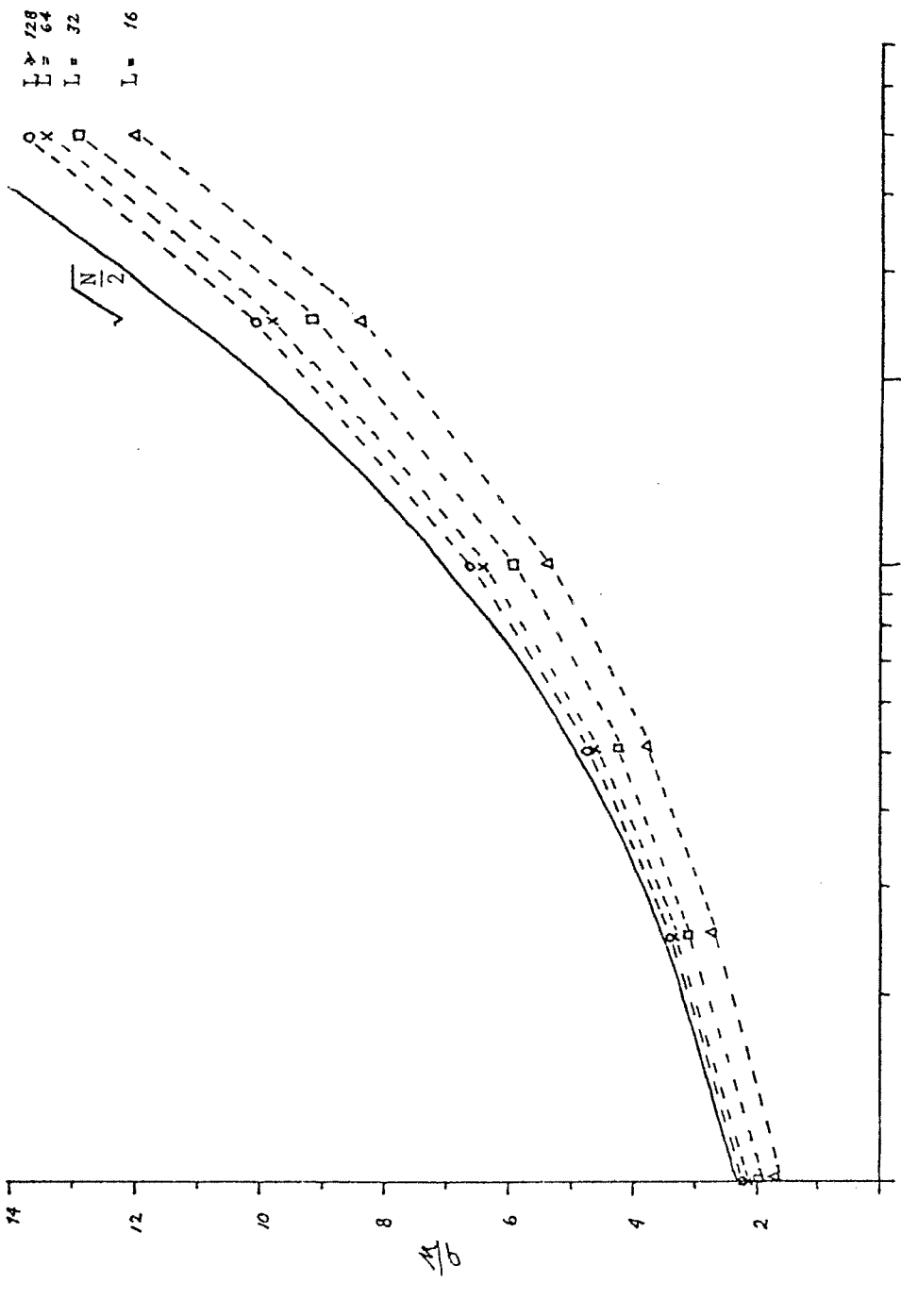

Figure 25. Mean to Standard Deviation Ratio of $_2B_N$ in the Presence of Quantizing Noise for Various Number of Levels (L) at $h = 1x10^{-2}$, $F_s = 100$

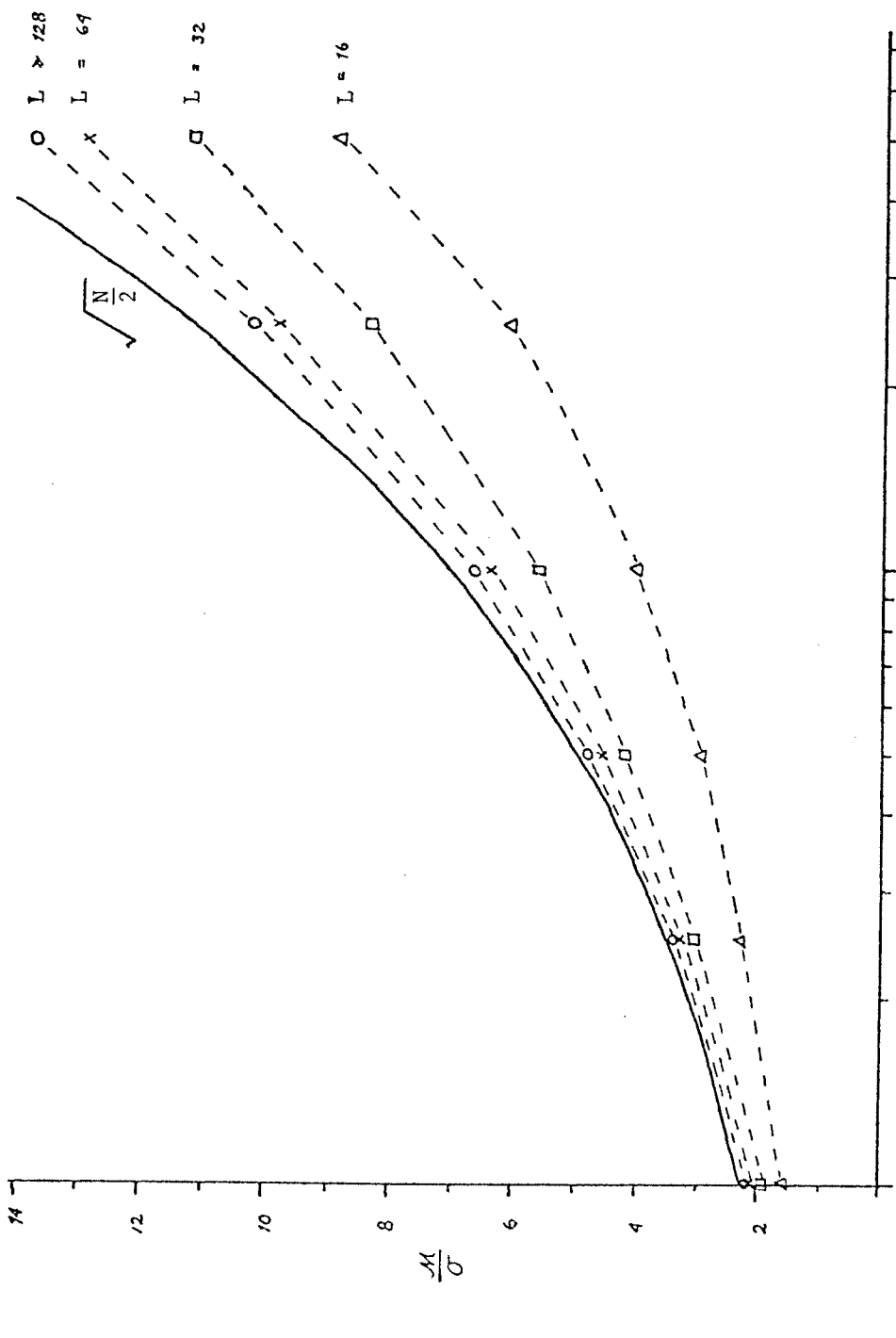

Figure 26. Mean to Standard Deviation Ration of $\hat{2^{B_N}}$ in the Presence of Quantizing Noise for Various Number of Levels (L) at $h = 1^{x}10^{-3}$, $F_s = 1000$

is the result of sampling an order of magnitude faster than the result presented in Figure 25. The results of the simulation studies are in consonance with the analysis, in that the higher frequency (smaller sampling interval) demonstrates a greater degradation due to quantization effects.

Digital processing, with its inherent quantization, has a restricted freedom of choice possible to achieve a desired accuracy in the result for the random variable σ^2. Without quantization (or other equipment noise) one could choose any number of combinations of sampling rate and number of observations, which would be compatible with equipment limitations and desired accuracy. The faster one samples, the better the estimate is in a fixed time. It is the realization of inherent equipment error that bursts this dream of singular estimation and frames the real engineering optimization.

The effect of quantizing error on the total result is illustrated in the next two figures. Figure 27 displays the 95% contours for the random variables $q_1(p_1 + p_2)$ and $q_2(p_1 p_2)$ for various observation intervals. Figure 28 represents the same results except that quantization effects are included. The quantization effects are demonstrated by a general upward motion and spreading of the contours, as illustrated in Figure 28 . That this is what should be expected, is evident from the consideration of the parameters that specify the contours. Recall that the mean and variance relationships are

$$\bar{q}_2 = \overline{(p_1 + p_2)} = \frac{\sigma^2}{2R_{\dot{x}}(0)} \quad , \quad \text{Var } q_2 = \frac{\sigma^2}{T R_{\dot{x}}(0)}$$

$$\bar{q}_1 = \overline{(p_1 p_2)} = \frac{R_x^{\bullet}(0)}{R_x(0)} \quad , \quad \text{Var } q_1 = \frac{\sigma^2}{T R_x(0)}$$

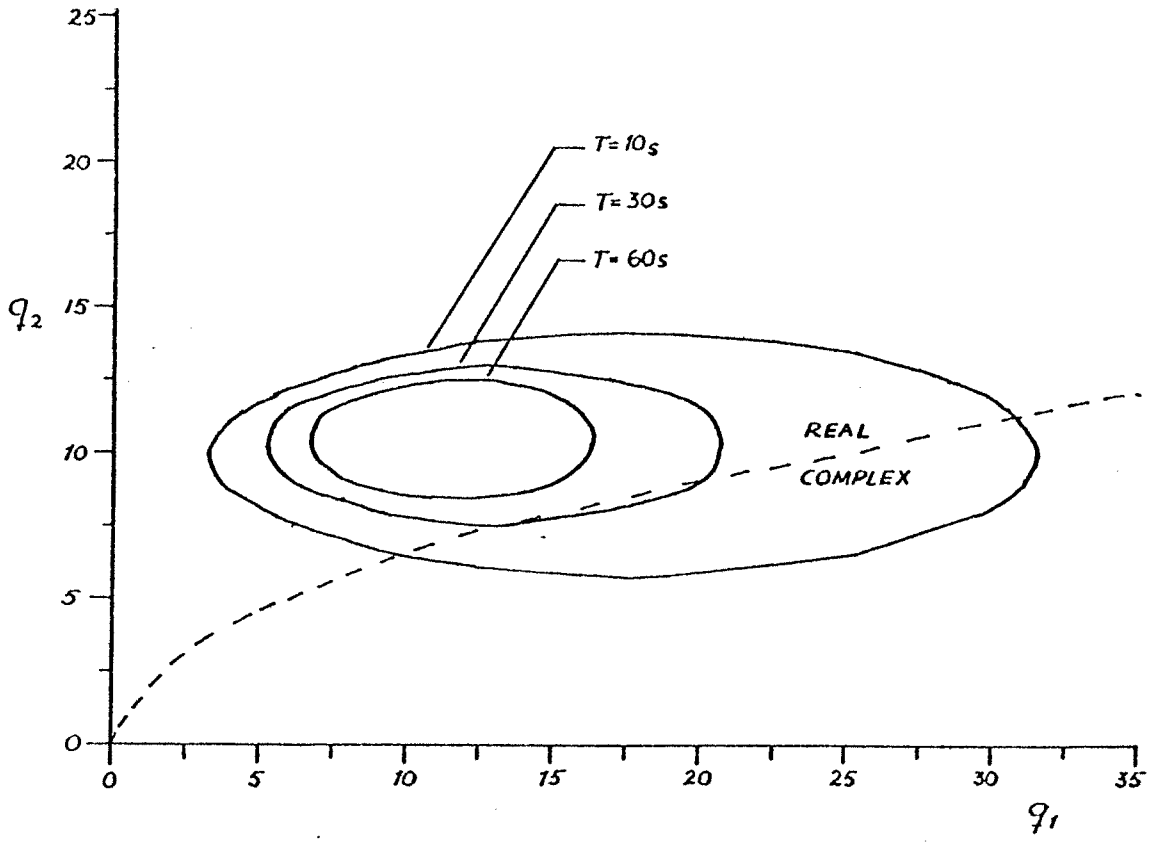

Figure 27. Pr = .95 Contour for True Values q_1=10, q_2=11, at Sampling Interval h = 10^{-2}.

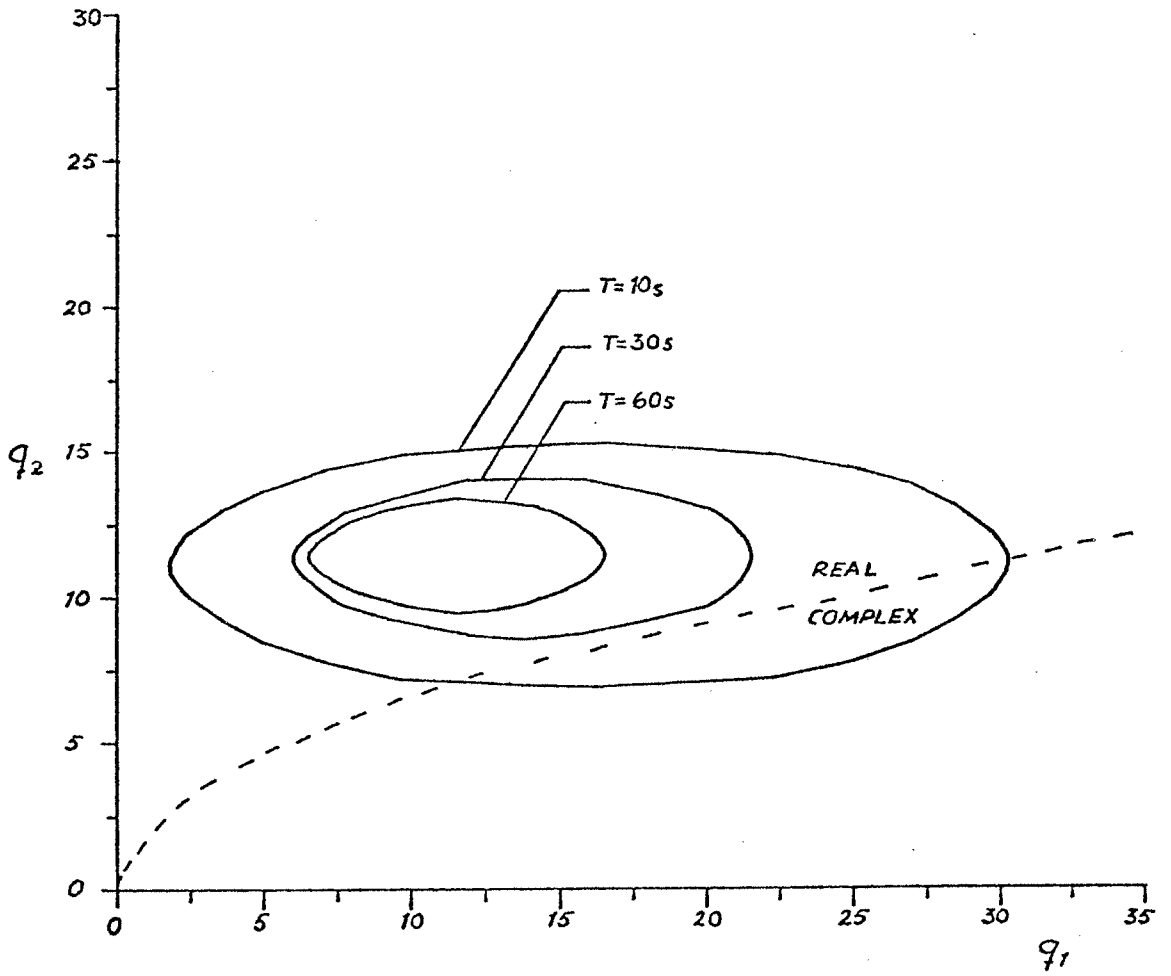

Figure 28. Pr = .95 Contour for True Values q_1=10, q_2=11, at Sampling Interval h = 10^{-2} with L=32 levels.

Introduction of the uniform quantizing noise caused an increase (positive bias) in the mean of σ^2. Specifically, the mean is

$$E({}_2\hat{B}_N) = E({}_2B_N)\left(1 + \frac{F_s}{L^2} \cdot \frac{3}{p_2 + p_1}\right),$$

where

$$E({}_2B_N) = \sigma^2 T$$

so that

$$E(\hat{\sigma}^2) = \sigma^2\left(1 + \frac{F_s}{L^2} \cdot \frac{3}{p_2 + p_1}\right) \quad . \tag{7.75}$$

The other measurements that are necessary for the mean and variance computation, the power measurements, are also biased upward as is demonstrated below. To incorporate the quantizing effects in the power measurements, define

$$NR_{\hat{x}}(0) = \sum_{i=1}^{N} \hat{x}_i^2 = \sum_{i=1}^{N} (x_i + \varepsilon_{i_A})^2$$

$$NR_{\hat{\dot{x}}}(0) = \sum \hat{\dot{x}}_i^2 = \sum_{i=1}^{N} (\hat{\dot{x}}_i + \varepsilon_{i_B})^2$$

then

$$E\left(R_{\hat{x}}(0)\right) = \frac{1}{N} E\left\{\sum x_i^2 + \sum \varepsilon_{i_A}^2\right\} = R_x(0) + \sigma_{\varepsilon_A}^2$$

$$E\left(R_{\hat{\dot{x}}}(0)\right) = \frac{1}{N} E\left\{\sum \dot{x}_i + \sum \varepsilon_{i_A}^2\right\} = R_x(0) + \sigma_{\varepsilon_B}^2 \quad .$$

Interpreting these results so that specific relationships for the quantizing noise can be used, the result is

$$E\, R_{\hat{x}}(0) = R_x(0) + \sigma_{\varepsilon_A}^2$$
$$= R_x(0) + \frac{3}{L^2} R_x(0) \tag{7.76}$$

$$E\, R_{\hat{\dot{x}}}(0) = R_{\dot{x}}(0) + \sigma_{\varepsilon_B}^2$$
$$= R_{\dot{x}}(0) + \frac{3}{L^2} R_{\dot{x}}(0) \quad . \tag{7.77}$$

Now, using these results in the mean and variance expressions, the

effects of quantization are observed to be

$$\overline{q_2} = \frac{\sigma^2\left(1 + \frac{F_s}{L^2} \cdot \frac{3}{p_2 + p_1}\right)}{2R_x(0)(1 + 3/L^2)} , \tag{7.78}$$

$$\text{Var } q_2 = \frac{\sigma^2\left(1 + \frac{F_s}{L2} \frac{3}{p_2 + p_1}\right)}{TR_x(0)(1 + 3/L^2)} , \tag{7.79}$$

$$\overline{q_1} = \frac{R_x(0)\ (1 + 3/_L{}^2)}{R\ (0)\ (1 + 3/_L{}^2)} , \tag{7.80}$$

$$\text{Var } q_1 = \frac{\sigma^2\left(1 + \frac{F_s}{L^2} \frac{3}{p_2 + p_1}\right)}{TR_x(0)\ (1 + 3/_L{}^2)} . \tag{7.81}$$

From these relationships, it becomes apparent that the power measurements are not significantly affected by the quantizing effects. The major contribution of quantization is reflected in the mean of the random variable σ^2. As a consequence of this alone both q_i variances are increased, hence the spreading of the contours. The mean of the random variable q_1 is not affected, but the mean $\overline{q_2}$ is biased upward due to the bias in σ^2.

It has been assumed that the quantization noise was "white" and independent of the data. These assumptions are reasonable if the quantization is sufficiently fine, and the waveform changes enough to traverse several quantum levels during the sampling period. It is intuitively clear when the quantizing is fine enough that

the samples are almost uniformly distributed within each quantum level. Bennett (1948) has calculated the noise correlation as a function of the correlation between successive signal samples, and has shown the noise to be practically uncorrelated unless the signal samples are almost completely correlated. With 8 bit quantizing L=256, the noise correlation is less than 0.0001, unless the signal samples have correlation greater than 0.9999.

Questions about the signal-noise correlation have been answered by Widrow (1956) and Requicha (1973). Using a sampling theory approach on the characteristic function of a quantized random variable it is shown that the joint moments can be calculated as though the random variables are independent when a Nyquist criterion is satisfied. This criterion requires that the sampled characteristic function (resulting from quantization) have negligible aliasing. For a Gaussian random variable, Widrow (1965) has shown that the use of a quantum size of the order of the standard deviation (3 bit quantizing, L=8) leads to negligible aliasing.

It is apparent that the parameter most sensitive to quantizing noise is the estimate of the random variable σ^2. The specification of a very high quality requirement (μ/σ ratio) on this parameter may well place the achievement of this quality beyond practical limits. That is, the required sampling rate (for a given μ/σ ratio) combined with a given resolution capability, may not be possible to satisfy. It is then important to observe the effect of different quality estimates on the spectral representation. The demonstration of the effect will be made through using different tolerances on the knowledge of the random parameter σ^2. The tolerances on the

of $\sigma^2(\pm k\%)$ are determined from the μ/σ ratio by

$$\mu \pm 3\sigma = (1 \pm k/100)\mu \quad , \tag{7.82}$$

so that

$$\mu/\sigma = 300/k \quad . \tag{7.83}$$

The tolerances under consideration k=10, k=20, correspond to the mean to standard deviation ratios of the random variable σ^2, $\mu/\sigma = 30$, $\mu/\sigma = 15$, respectively. The effect of these tolerances will be illustrated through the behavior of the mean values

$$\bar{q}_1 = \overline{(p_1 p_2)} = \frac{R_{\dot{x}}(0)}{R_x(0)} \quad , \tag{7.84}$$

$$\bar{q}_2 = \overline{(p_1 + p_2)} = \frac{\sigma^2(1 \pm k/100)}{2R_{\dot{x}}(0)} \quad , \tag{7.85}$$

when the numerical results are inserted into the parametric form of the spectral density

$$S(\omega) = \frac{\sigma^2(1 \pm k/100)}{(\omega^2 + p_1^2)(\omega^2 + p_2^2)} \quad . \tag{7.86}$$

The results presented assumed that the true values for the respective power measurements have been achieved. In this way the effects of the length of the observation interval on the power measurements are eliminated and the variability of the spectral representation will be due solely to the quality of knowledge about the random variable σ^2. The results of these computations for several pole pair combinations is presented in terms of spectral representations in Figures 29 through 31. As is evident from these figures, the high frequency asymptotes are rather closely packed about the true value (k=0). The largest variation of the spectral asymptotes occurs over the mid-frequency region and for processes with larger pole separations, the variation

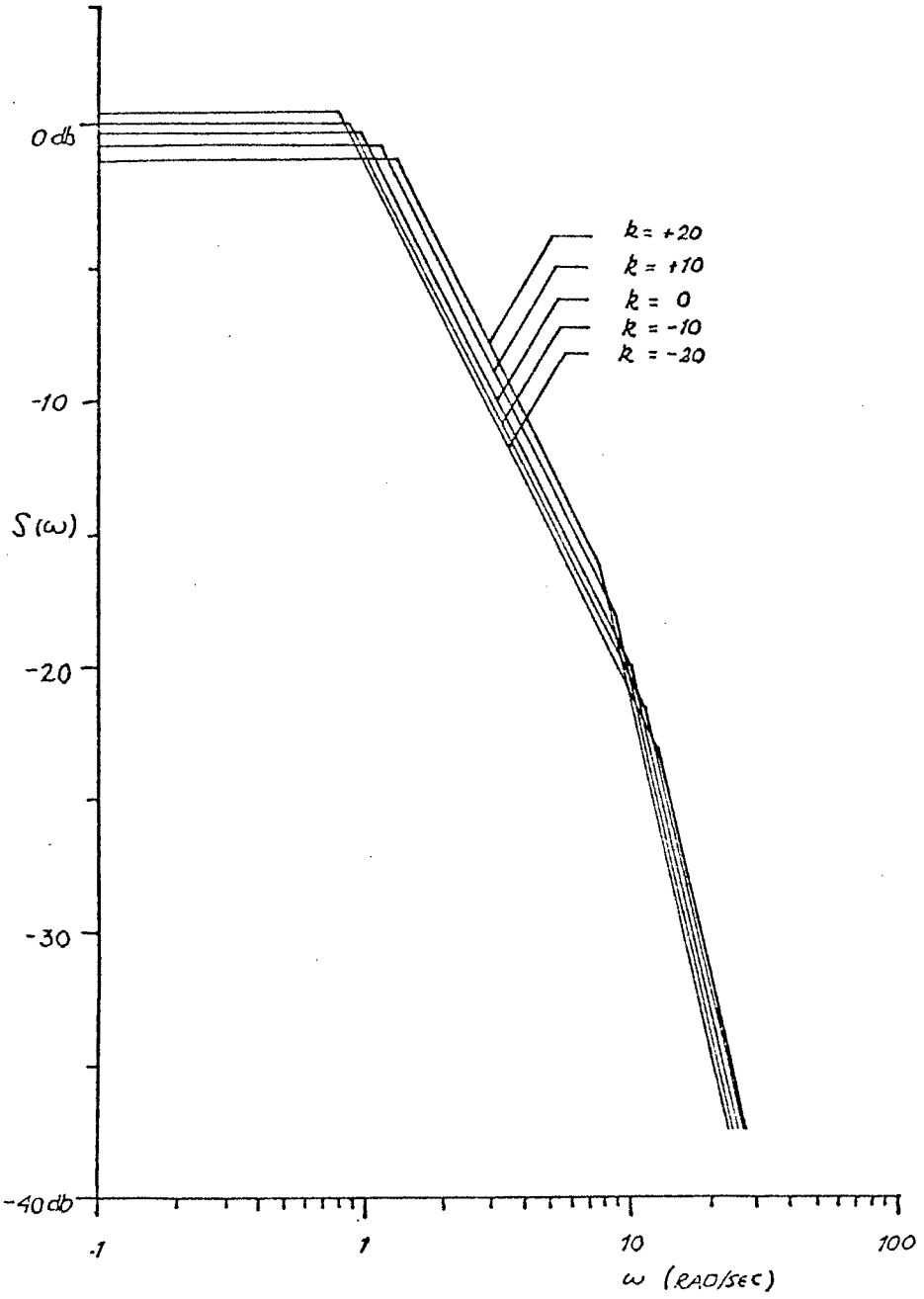

Figure 29. Spectral Asymptotes for True Values q_1=10, q_2=11, as a Function of Variability (k).

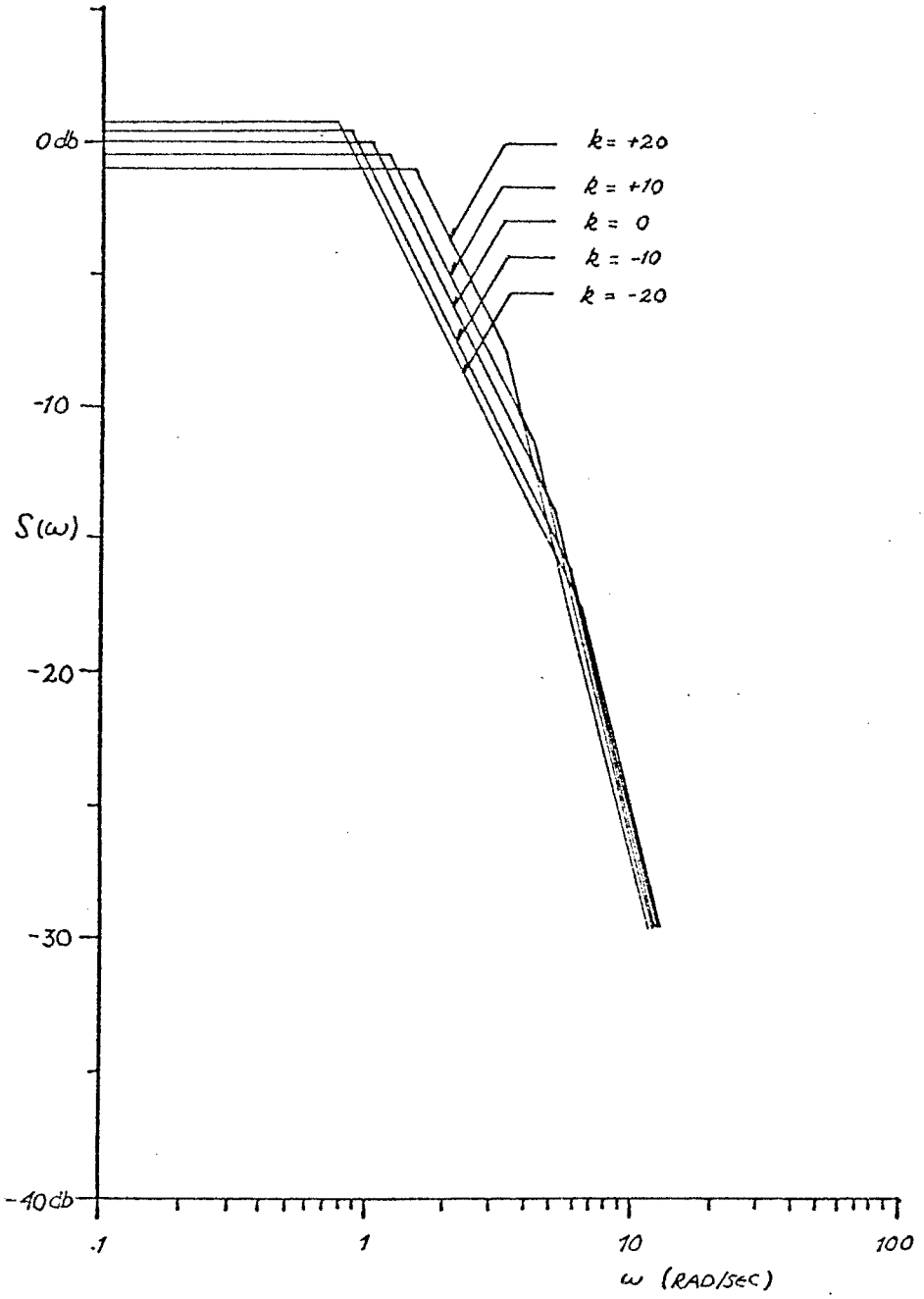

Figure 30. Spectral Asymptotes for True Values q_1=5, q_2=6, as a Function of Variability (k).

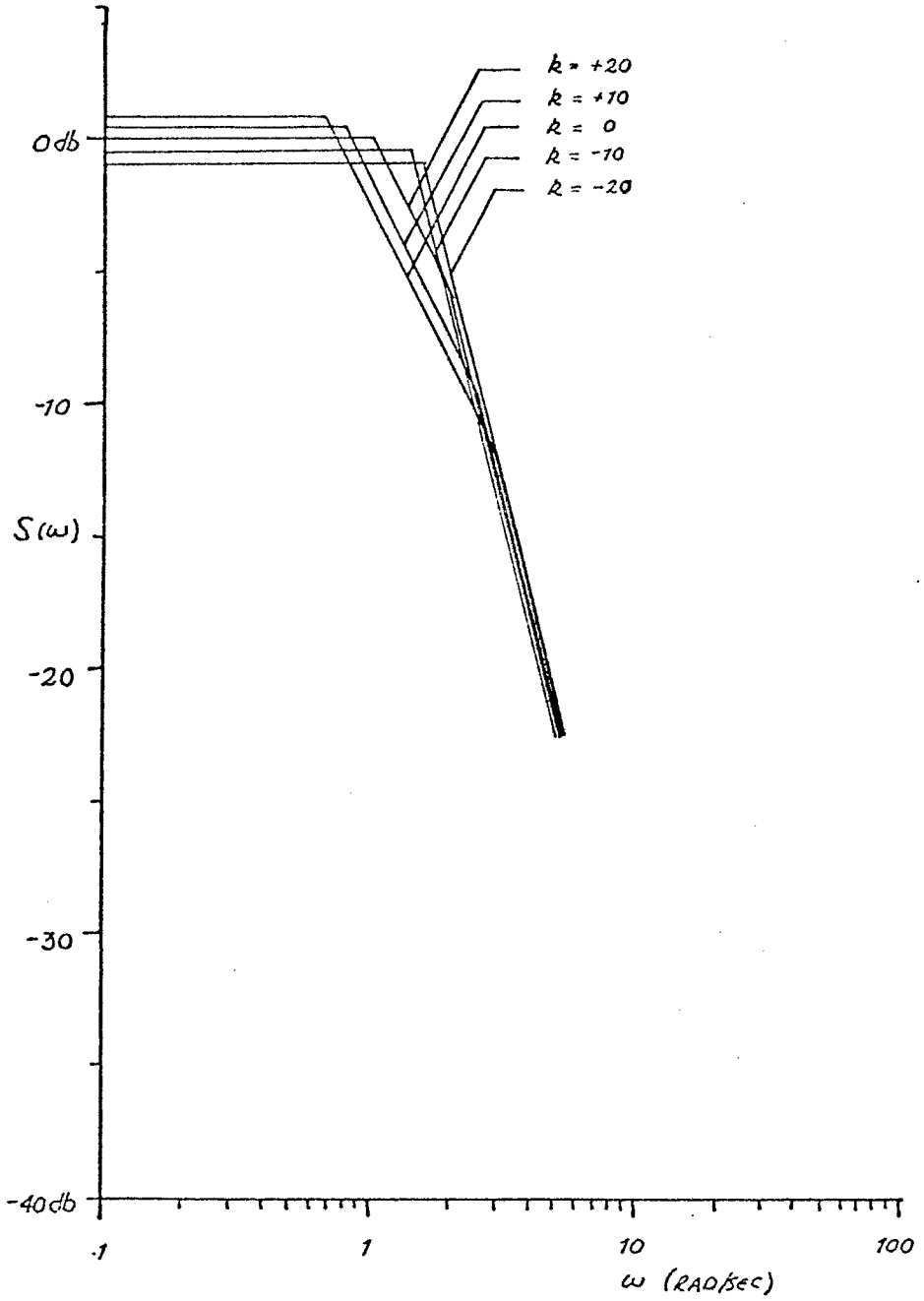

Figure 31. Spectral Asymptotes for True Values $q_1=2$, $q_2=3$, as a Function of Variability (k).

becomes minor. The large mid-frequency variation in Figure 31 is due to the complex conjugate poles that result when k = -10%, -20%. From these interpretations, it appears that μ/σ ratios of the order of μ/σ =30, are sufficient to minimize the variability of the spectral representation when there is a large high frequency content. On the other hand, larger μ/σ ratios are necessary when the poles are closely spaced.

7.7 Learning the Order of the Process (M)

Prior to this point, the discussion of the estimation problem has assumed that the order of the proces was known a priori. This knowledge and a fast sampling procedure provided the vehicle for learning the random variable σ^2 faster than the $\underline{q}$ parameters in the estimation problem. For the M=2 problem it was shown that as a consequence of Baxter's Theorem, the random variable

$$ {}_2B_N = \sum_i (\dot{x}_i - \dot{x}_{i-1})^2 \xrightarrow[a.s.]{} \sigma^2 T \quad , $$

provided the tool for determining σ^2. There is another part to this theorem that indicates that for an M=2 process, the random variable

$$ {}_1B_N = \sum_i (x_i - x_{i-1})^2 \xrightarrow[a.s.]{} 0 \quad . $$

This statement plus the knowledge that for an M=1 process the random variable ${}_1B_N$ behaves such that

$$ {}_1B_N = \sum_i (x_i - x_{i-1})^2 \xrightarrow[a.s.]{} \sigma^2 T \quad , $$

will establish the method of deciding the order (M). To place this in the context of a binary decision problem, the choices for M will be limited to M=1 or M=2. Extension to a higher decision problem is obvious. The decision problem is structured by specifying the two

hypotheses, H_M, given the decision variable ${}_1B_N$ Hypothesis (H_2) is that M=2, and the other hypothesis (H_1) is that M=1. To specify the decision process, the statistics of the decision variable are determined as proportional to a Chi-square with N degrees of freedom.

$$H_2 : {}_1B_N,$$

$$\text{mean} \quad \mu_2 = \frac{\sigma^2 T \cdot h}{2(p_2+p_1)} \quad ,$$

$$\text{variance} \quad \sigma_2 = 2\mu_2 2/N \quad , \tag{7.87}$$

$$H_1 : {}_1B_N,$$

$$\text{mean} \quad \mu_1 = \sigma^2 T \quad ,$$

$$\text{variance} \quad \sigma_1^2 = 2\mu_1 2/N \quad . \tag{7.88}$$

In reality, the sampling interval (h) will not be zero, so the mean μ_2 will be small but non zero. The μ/σ ratio $\sqrt{N/2}$ can be used as a measure of the quality of one's knowledge of ${}_1B_N$.

Under both hypotheses, the random variable converges at the same rate, to values which depend on the hypothesis in effect. To evaluate the performance of the decision problem as it stands would require the evaluation of a χ^2 ROC (receiver operating characteristic). Instead, to provide analytic tractability, the Central Limit Theorem is invoked and the distributions are assumed to be Gaussian, with parameters as specified. A pictorial representation of the situation is presented in Figure 32.

Select the decision cut C to yield equal probabilities of each type of error.

$$\int_C^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp - \frac{(x-\mu_2)^2}{2\sigma_2^2} \, dx = \int_{-\infty}^{C} \frac{1}{\sqrt{2\pi}\,\sigma_2} \exp - \frac{(x-\mu_1)^2}{2\sigma_1^2} \, dx \tag{7.89}$$

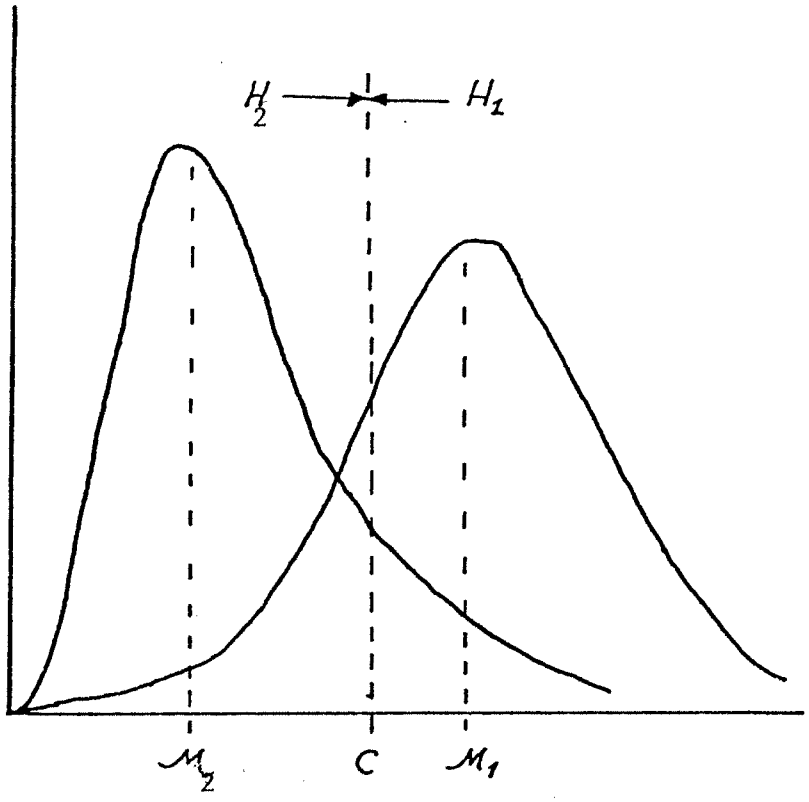

Figure 32. Decision Problem Representation.

The integrals are equal by selecting C such that the relations

$$k = (\mu_1 - C) / \sigma_1 \tag{7.90}$$

$$k = (C - \mu_2) / \sigma_2 \tag{7.91}$$

are satisfied. Once this is done, the probability of error is found from the equation

$$\Phi(k) = \int_{-\infty}^{k} \frac{1}{\sqrt{2\pi}} e^{-(y^2/2)} dy \quad . \tag{7.92}$$

Using (7.90) and (7.91) and $\mu/\sigma = \sqrt{N/2}$, the decision cut (C) is the harmonic mean of μ_1 and μ_2 .

$$C = \frac{2\mu_1\mu_2}{\mu_1+\mu_2} \tag{7.93}$$

The cut value for the means given is

$$C = \frac{\sigma^2_T \, h}{(p_1+p_2) + (h/2)} \approx \frac{\sigma^2_T \, h}{(p_1+p_2)} \tag{7.94}$$

k is determined from the given parameters to be

$$k = \frac{\mu_1-C}{\sigma_1} = \sqrt{\frac{N}{2}} \, \frac{\mu_1-\mu_2}{\mu_1+\mu_2}$$

$$= \sqrt{\frac{N}{2}} \, \frac{p_1+p_2 - h/2}{p_1+p_2 + h/2} \approx \sqrt{\frac{N}{2}} \tag{7.95}$$

To obtain a feel for the speed of the decision process, let k=2.32 then the probability of error is

$$\Phi(-2.32) = \int_{-\infty}^{-2.32} \frac{1}{\sqrt{2\pi}} e^{-(y^2/2)} dy = .01 \qquad (7.96)$$

The number of observations necessary to achieve 1% decision error is

$$N = 2k^2 \cong 11$$

The decision can be made on the order of the model within a relatively few observations. However a greater number of observations is necessary to insure the required accuracy for the q-vector estimation problem.

7.8 Concluding Remarks and Summary

The problem under investigation is the estimation of the set of parameters that describe the all pole spectral density

$$S(\omega) = \frac{\sigma^2}{\prod_{i=1}^{M} (\omega^2 + p_i)} \quad , \quad i \neq j \quad .$$

The observations process with this spectral representation is assumed to be zero mean and Gaussian. The parameters $\{\sigma^2, p_i\}$ are considered to be random variables and the results of the estimation problem are expressed as probability density funtions. Bayes' Rule was applied to the observations density that described a fast sampling environment, with the result that a summarizing set of measurements (sufficient statistics) provided the information necessary for the construction of the probability density functions. The sufficients statistics are shown to be, for an M^{th} order process, the measurements

$$\left\{\int_0^T x^2(t)dt, \quad \int_0^T \dot{x}^2(t)dt.... \right.$$

$$\left. \int_0^{T_{M-1}} x^2(t)dt, \; x(T), \; x(0) \ldots {}^{M-1}x(T), {}^{M-1}x(0) \right\}. \quad (7.97)$$

Although the analysis is carried out for only the M=1 and M=2 processes, this conclusion is valid for other finite dimensional processes, as demonstrated by Hajek (1962).

The fast sampling assumption coupled with the all pole structure of the observation process, provided a vehicle for the estimation of the parameter σ^2. The statistic

$$_pB_N = \sum_{k=1}^{N} ({}^{p-1}x_k - {}^{p-1}x_{k-1})^2 \qquad p = 1,2, \ldots M$$

was shown to converge almost surely to its mean, where depending on the order of the process (M), the mean is either zero or a constant (for a fixed observation interval, T). For the investigation here, the order of the process is limited to M=1 and M=2, and for the M=2 process, the means are respectively $_1\overline{B_N} = 0$ and $_2\overline{B_N} = \sigma^2 T$. So the use of the statistic $_2B_N$, provides a rapidly converging estimate of the parameter σ^2. Using the mean to standard deviation ratio as a measure of this convergence, it became evident that for the random variable $_MB_N$, this measure is

$$\frac{\mu}{\sigma}(_MB_N) = \sqrt{N/2} \qquad (7.98)$$

where the sampling becomes dense in a fixed observation interval T. Taking advantage of this property, by imposing fast sampling conditions, the parameter σ^2 is learned faster than the other parameters in the estimation problem. Another consequence of the fast

sampling condition is the relative rate of convergence of the low and mid frequency portions of the spectral representation. For an M=2 process, the mid frequency behavior is characterized by the parameter $q_2(p_1 + p_2)$ which is shown to converge more quickly than the parameter $q_2(p_1 p_2)$ which characterizes the low frequency behavior. This behavior is most easily discerned when the pole positions are widely separated. In this situation, it is clear that increasing the length of the observation interval is necessary for the settling of the low frequency portion of the spectral density.

The success of this estimation procedure is strongly dependent on the formation of the required derivatives, as indicated in equations (6.107) and (6.108), from the observation process. Quantization noise introduced severe error in digital differentiation and it was demonstrated that analog differentiation was necessary to avoid the overwhelming effects of quantizing noise. Analog differentiation was analyzed and simulated by allowing separate quantization schemes to be implemented for the observations and derivative processes. As a consequence, for the derivative process, the quantizing noise was injected after the formation of the derivative. In this circumstance, there is a trade-off that must be made between the desire to sample fast, and the resolution of the analog to digital converters. There is a minimum sampling rate required to perform the estimation through the use of the statistic $_2B_N$, and for a fixed quantizing scheme, there is an upper limit on the sampling rate beyond which quantization effects terminate the estimate. In general, for the numerical examples simulated in this study, acceptable results were obtained for sampling rates of 200 to 3000 times the Nyquist rate, and a

seven bit uniform quantizing scheme. From a practical standpoint, inherent equipment noise, examplified by quantization noise in digital processing, places a bound on sampling frequency and delimits the possibility of singular estimation promised by more abstract theory.

APPENDIX A

Difference Equation for the Sampled M=1 Process

The purpose of this Appendix is to illustrate the development of the difference equation that results from sampling the differential equation representation of an M=1 process. The M=1 process of interest has second order properties that are represented by

$$R_y(\tau) = \frac{\sigma^2}{2p} e^{-p|\tau|}, \tag{A.1}$$

$$S_x(\omega) = \frac{\sigma^2}{\omega^2+p^2} . \tag{A.2}$$

The differential equation that produces the desired output statistic can be represented as

$$dx = -pxdt + dv, \tag{A.3}$$

$$y = x$$

with $E(dv(t)\ dv(s)) = \sigma^2 dt$, for t=s. For this representation the transition matrix is

$$\Phi(t,t_o) = e^{-p(t-t_o)} \tag{A.4}$$

The output statistics (Ry(τ)) are found to be

$$R_y(t-t_o) = \Phi(t-t_o)\ R_x(t_o) ,$$

$$R_y(\tau) = e^{-p|\tau|}\ R_x(0) . \tag{A.5}$$

For a stationary process the correlation $R_x(0)$ is chosen as the steady state solution to equation (3.137), which is $R_x(0) = \frac{\sigma^2}{2p}$. The discrete version of equation (A.3), equivalent at the sampling instance, is obtained by determining the constituent

parts of

$$x(t_{i+1}) = \Phi(t_{i+1},t_i)\ x(t_i) + \tilde{v}(t_i)$$
$$y(t_i) = x(t_i) \qquad (A.6)$$

The statistical properties of the driving sequence $\tilde{v}(t_i)$ are found to be

$$E(\tilde{v}(t_i)\ \tilde{v}(t_i)) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)\ \sigma^2 dt\ \Phi^T(t_{i+1},s)ds$$
$$= \frac{\sigma^2}{2p}\ (1-e^{-2p(t_{i+1}-t_i)}) \quad , \qquad (A.7)$$

$$E(\tilde{v}(t_i)\ \tilde{v}(t_j)) = 0 \qquad (A.8)$$

Under fast sampling conditions, $(p(t_{i+1}- t_i) <1)$ these results reduce to

$$E(\tilde{v}(t_i)\ \tilde{v}(t_i)) = {}_{\Delta}R_{\tilde{v}}(0) = \sigma^2\Delta \quad , \qquad (A.9)$$

$$E(\tilde{v}(t_i)\tilde{v}(t_j)) = 0 \quad , \qquad (A.10)$$

where $\Delta = t_{i+1}-t_i$. Similarly, under fast sampling conditions the factor $\Phi(t_{i+1}, t_i)$ reduces to

$$\Phi(t_{i+1}\ ,\ t_i) = e^{-p\ (t_{i+1}-t_i)}$$
$$= 1-p\Delta \quad , \qquad (A.11)$$

so that the equivalent difference equation is of the form

$$y(t_{i+1}) = (1-p\Delta)\ y(t_i) + \tilde{v}(t_i) \qquad (A.12)$$

with $E(\tilde{v}(t_i)\ \tilde{v}(t_j)) = \sigma^2\ \Delta\ \delta_{ij}$. And as shown in Chapter III under fast sampling conditions this difference equation provides the desired second order properties at the sampling instants.

APPENDIX B

Bayesian Estimation of the Mean of a Normal Distribution

The purpose of this appendix is to illustrate the calculations used in developing the results presented in equations (5.7) and (5.8). The problem is to find a post density for the parameter θ (the mean of a normal pdf), given that the observations $\underline{Y}_N$ are described by the pdf

$$f(\underline{Y}_N / \theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{N/2} \exp \left\{ - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - \theta)^2 \right\} , \quad (B.1)$$

and the prior is described by

$$f(\theta) = \frac{1}{\sqrt{2\pi\,\alpha^2}} \exp \left\{ - \frac{1}{2\alpha^2} (\theta - \mu)^2 \right\} \cdot \quad (B.2)$$

Using Bayes Rule, the post density is written as

$$f(\theta/\underline{Y}_N) \propto \exp \left\{ - \frac{1}{2\,\sigma^2} \left(\sum_i (Y_i - \theta)^2 \right) \right\} \cdot \exp \left\{ - \frac{(\theta-\mu)^2}{2\,\alpha^2} \right\}$$

$$\propto \exp \left\{ - \frac{1}{2\sigma^2\alpha^2} \left[\theta^2 (N\alpha^2 + \sigma^2) - 2\theta (\alpha^2 \sum_i Y_i + 2\mu\sigma^2) \right. \right.$$

$$\left. \left. + \alpha^2 \sum_i Y_i^2 + \sigma^2\mu^2 \right] \right\} \cdot \quad (B.3)$$

Given the observations $(\underline{Y}_N)$, the terms not involving the parameter θ are constants which can be lumped into a constant of integration. Then the post density is represented as

$$f(\theta/Y_N) \propto K \exp \left\{ - \frac{1}{2\sigma^2\alpha^2} \left[\theta^2 (N\alpha^2 + \sigma^2) - 2\theta (\alpha^2 \sum_i Y_i + 2\mu\sigma^2) \right] \right\} \cdot \quad (B.4)$$

Completing the square in the exponential term above, the post

density is

$$f(\theta/\underline{Y}_N) = \frac{1}{\sqrt{2\pi\,\tilde{\sigma}^2}} \exp\left\{-\frac{1}{2\,\tilde{\sigma}^2}(\theta-\tilde{\mu})^2\right\}, \qquad \text{(B.5)}$$

where $\tilde{\sigma}^2 = \dfrac{\sigma^2\,\alpha^2}{N\alpha^2+\sigma^2}$,

$$\tilde{\mu} = \frac{N\alpha^2}{N\alpha^2+\sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N} Y_i\right) + \frac{\sigma^2}{N\alpha^2+\sigma^2}\,\mu \quad .$$

APPENDIX C

Bayesian Solution to M=2 Autoregressive Process

The second order Autoregressive Process is described by the difference equation

$$y(k+2) = -\beta_1 y(k+1) - \beta_2 y(k) + e(k), \qquad (C.1)$$

where $e(k) \sim N(0,\alpha^2)$.

The attempt here is to find the post densities on the parameters $\{\beta_1, \beta_2, a^2\}$ using the ideas of reproducing priors. To begin, the observations density is

$$f(\underline{Y}_N | \beta_1, \beta_2, a, \underline{Y}_o) \alpha \exp \sum_{i=1}^{N} \frac{-a}{2} (Y_i + \underline{\beta}^T \underline{Y}_{i-1})^2 , \qquad (C.2)$$

where $a = 1/\alpha^2$.

Bayes rule applied to the evaluation of the post density on the parameters takes the form

$$f(\beta | a, \underline{Y}_N)\, f(a | \underline{Y}_N) = \frac{f(\underline{Y}_N | a\underline{\beta}) \cdot f(a|\underline{\beta})\, f(\underline{\beta})}{\int f(\underline{Y}_N | a, \underline{\beta})\, f(a|\underline{\beta})\, f(\underline{\beta})\, da d\underline{\beta}} . \qquad (C.3)$$

It can be shown (Raiffa and Schlaiffer, 1961) that if the priors are chosen as

$$f(a) \propto a^{w-1} e^{-\ell a} \qquad a > 0 \quad \text{and,} \quad \ell > 0, \qquad (C.4)$$

$$f(\underline{\beta}/a) \propto a^{\frac{1}{2}} \exp\{ -\frac{a}{2} (\underline{\beta}-\underline{\mu})^T T (\underline{\beta}-\underline{\mu}) \} , \qquad (C.5)$$

where $f(a)$ is a gamma density function and $f(\underline{\beta}/a)$ is a gaussian density function, reproduce under the observations specified by

equation (C.2). The post density is now found to be

$$f(\underline{\beta}/a)\ f(a) \propto \exp\left\{-\frac{a}{2}(\underline{\beta}-\underline{\mu}')^T T'(\underline{\beta}-\underline{\mu}')\right\} \cdot a^{w'-1}e^{-\ell' a}, \quad (C.6)$$

where asymptotually

$$\underline{\mu}' = \begin{bmatrix} \mu' \\ 1 \\ \mu_2' \end{bmatrix} = \frac{1}{R_o^2 - R_1^2} \begin{bmatrix} R_2R_1 - R_1R_o \\ R_1^2 - R_2R_o \end{bmatrix}, \quad (C.7)$$

with $$R_1 = \frac{1}{N}\sum x_i x_{i-1},$$

$$R_o = \frac{1}{N}\sum x_i^2,$$

$$R_2 = \frac{1}{N}\sum x_i x_{i-2},$$

and the precision matrix for the gaussian post density on $\underline{\beta}$ is

$$T' = N \begin{bmatrix} R_o & R_1 \\ R_1 & R_o \end{bmatrix} = \Gamma^{-1}, \quad (C.8)$$

where Γ is the covariance matrix and

$$\Gamma = \frac{1}{N(R_o^2 - R_1^2)} \begin{bmatrix} R_o - R_1 \\ -R_1 \quad R_o \end{bmatrix}. \quad (C.9)$$

The parameters for the gamma post density f(a) are

$$w' = \frac{N}{2}, \quad (C.10)$$

and $$\ell' = \frac{N}{2(R_o^2 - R_1^2)} (R_o^3 - 2R_1^2R_o + 2R_1^2R_2 - R_2^2R_o). \quad (C.11)$$

Recalling the definition of the parameter a, the statistics of the post density can be represented as

$$E(\alpha^2) = E\left(\frac{1}{a}\right) = \frac{\ell'}{w'} = \frac{(R_o^3 - 2R_1^2 R_o + 2R_1^2 R_2 - R_2^2 R_o)}{(R_o^2 - R_1^2)} \quad . \tag{C.12}$$

Using the asymptotic mean of the parameter β_2 as the true value and placing this into equation (C.12) the result is

$$E(\alpha^2) = R_o \frac{(1-R_1^2)}{R_o^2}(1-\beta_2^2) \quad . \tag{C.13}$$

The variance is

$$Var(\alpha^2) = \frac{\ell'^2}{w'^3} = \frac{2}{N}\left(R_o\left(\frac{1-R_1^2}{R_o^2} \right)\left(1-\beta_2^2\right)\right)^2 \xrightarrow[N\to\infty]{} 0 \quad . \tag{C.14}$$

Before the post density can be claimed to be gamma one must find the conditions necessary to have the parameter ℓ' strictly positive. Using the given representation then

$$\ell' = \frac{N}{2} R_o \frac{(1-R_1^2)}{R_o^2}(1-\beta_2^2) > 0 \quad ,$$

if $\quad R_o^2 \neq R_1^2 \quad$ or $\quad \beta_2 < 1 \quad .$

With this restriction on β_2, and since β_1 can be written as

$$\beta_1 = -\frac{R_1}{R_o}(1+\beta_2) = \frac{R_1 R_2 - R_1 R_o}{R_o^2 - R_1^2} \quad ,$$

the constraint on β_1 is of the form

$$|\beta_1| < 2 \quad .$$

Hence, if the range on the parameters β_1, β_2 are restricted as above, then the existence of the density on the parameter α^2

can be guaranteed. This is reasonable since these statements are precisely the same as the requirements for the existence of a stable solution to the difference equation that served as the model. One further point that can be illustrated from this example is that the means for the parameters $\{\underline{\beta},\alpha^2\}$, when using the asymptotic results, are precisely the relationships given by the Durbin-Levinson (Chapter 3) for an M=2 process.

APPENDIX D

Details of the M=2 Estimation Problem

The purpose of this Appendix is to illustrate the calculations that lead up to equation (6.38), the post density for the parameters of the M=2 process. The M=2 process is characterized by a spectral density of the form

$$S_x(w) = \frac{\sigma^2}{(w^2+p_1^2)(w^2+p_2^2)} \tag{D.1}$$

and correlation function

$$R_x(\tau) = \frac{\sigma^2}{2p_1p_2(p_2^2-p_1^2)} \{ p_2 e^{-p_1|\tau|} - p_1 e^{-p_2|\tau|} \}. \tag{D.2}$$

The difference equation which, as the sampling interval (h) becomes dense, has the specified second order properties is

$$\underline{Y}(n) = \begin{bmatrix} 1 & h \\ -hp_1p_2 & 1-h(p_1+p_2) \end{bmatrix} \underline{Y}(n-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \zeta(n), \tag{D.3}$$

$$x(n) = y_1(n) \quad , \tag{D.4}$$

where

$$\zeta(\cdot) \sim N(0,\sigma^2 h),$$

$$E\{\zeta(k)\zeta(j)\} = \begin{cases} \sigma^2 h, & j=k \\ 0, & j\neq k \end{cases},$$

$$E\{y(n)\,\zeta(k)\} = 0, \quad k \geq n \quad .$$

Writing the difference equation in terms of the observation x(n), the equation

$$x(n)= x(n-1) - h^2 p_1 p_2\, x(n-1) +(1-h(p_1+p_2))(x(n-1)-x(n-2)) + h\zeta(n-1), \tag{D.5}$$

is achieved. To obtain the single step observation density ($f(\underline{X}_N/\underline{X}_o)$), all that is necessary is the evaluation of the conditional mean and variance. The conditional mean is

$$E(x(n)/x(n-1), x(n-2))= \left(2-h(p_1+p_2)\right)x(n-1) - \left(1-h(p_1+p_2) + h^2 p_1 p_2\right)x(n-2), \tag{D.6}$$

and the variance is

$$\mathrm{Var}(x(n)/x(n-1),x(n-2)) = \sigma^2 h^3 \quad . \tag{D.7}$$

With the mean and variance determined the one step density is

$$f(x(n)/x(n-1),x(n-2))\alpha\left\{\exp - \frac{1}{2\sigma^2 h^3}\left\{x_n-(2-h(p_1+p_2))x_{n-1} + (1-h(p_1+p_2) + h^2 p_1 p_2)x_{n-2}\right)^2\right\} \quad . \tag{D.8}$$

Using the properties of the M=2 process the N step observation's density is

$$f(\underline{x}_N/\underline{x}_o,\underline{q},\ \sigma^2) \propto \left\{\exp - \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} (x_i-(2-hq_2)x_{i-1} + (1-hq_2+h^2 q_1)\ x_{i-2})^2\right\} \tag{D.9}$$

where $\underline{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} p_1\ p_2 \\ p_1+p_2 \end{bmatrix}$,

and $\underline{x}_o = \begin{bmatrix} x_o \\ x_{-1} \end{bmatrix}$.

At this point it is convenient to introduce additional definitions to reduce the complexity in expanding the quadratic form in the exponential of equation (D.9). The expansion of this term is

necessary to group and recognize factors that converge to known quantities when the limit of the likelihood function is taken. Defining the quantity

$$h\dot{x}(n-2) = x(n-1) - x(n-2) \tag{D.10}$$

the quadratic form in (D.9) is written as

$$QF = \frac{1}{2\sigma^2 h^3} \sum_i (x_i - x_{i-1} + h^2 q_1 x_{i-2} - (1-hq_2)\dot{x}_{i-2})^2 \cdot \tag{D.11}$$

Expanding this form and grouping terms in like powers of q_1 and q_2 results in

$$\begin{aligned} QF = \frac{1}{2\sigma^2 h^3} \sum_i \{ & (x_i^2 - 2x_i x_{i-1} + x^2_{i-1}) \\ & + 2h^2 q_1 (x_{i-2}x_i - x_{i-2}x_{i-1}) \\ & + 2(1-hq_2)\, h(\dot{x}_{i-2}x_{i-1} - \dot{x}_{i-2}x_i) \\ & - 2h^2 q_1 (1-hq_2)\, h\dot{x}_{i-2}x_{i-2} \\ & + h^4 q_1^2 x^2_{i-2} \\ & + (1-hq_2)^2\, h^2 \dot{x}^2_{i-2} \} \cdot \end{aligned} \tag{D.12}$$

Grouping the factors of QF into two parts, those factors involving q and those which do not provides the result,

$$\begin{aligned} QF = & \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} (x_i^2 - 2x_i x_{i-1} + x^2_{i-1} + 2h\dot{x}_{i-2} \\ & (x_{i-1} - x_i) + h\dot{x}^2_{i-2}) \\ & + \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} (2h^2 q_1 x_{i-2} (x_i - x_{i-1}) + 2h^2 q_2^2 \dot{x}_{i-2}(x_i - x_{i-1}) \\ & - 2h^2 q_1 \dot{x}_{i-2}x_{i-2} - 2hq_2 h^2 \dot{x}^2_{i-2} \\ & + 2h^2 q_1 q_2 h^2 \dot{x}_{i-2}x_{i-2} + h^4 q_1^2 x^2_{i-2} + h^4 q_2^2 \dot{x}^2_{i-2}) \cdot \end{aligned} \tag{D.13}$$

The portion of Equation (D.13) that does not involve the parameters q_1, q_2 will be eliminated when forming the Likelihood Ratio $(f(\underline{X}_n/\underline{X}_o, \underline{q}_1\sigma^2) \;/\; f(\underline{X}_N/\underline{X}_o, \underline{q}_o, \sigma^2)$, where $\underline{q}_o = \underline{0})$ so attention can be directed towards reducing the terms that involve q_1 and q_2. Combining the terms in

$$A \triangleq \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} (2h^2 q_1 x_{i-2}(x_i - x_{i-1}) + 2h^2 q_2 \dot{x}_{i-2}(x_i - x_{i-1}) - 2h^2 q_1 \dot{x}_{i-2} x_{i-2} - 2hq_2 h^2 \dot{x}_{i-2}^2) \quad ,$$

results in

$$A = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2(q_1 x_{i-2} + q_2 \dot{x}_{i-2})(\dot{x}_{i-1} - \dot{x}_{i-2}) \quad . \tag{D.14}$$

Reducing the factors in

$$B \triangleq \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} 2h^2 q_1 q_2 h^2 \dot{x}_{i-2} x_{i-2}$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_1 q_2 x_{i-2}(x_{i-1} - x_{i-2}) \quad ,$$

Using the identity

$$2x_i x_{i-1} = x_i^2 + x_{i-1}^2 - (x_i - x_{i-1})^2 \quad ,$$

results in

$$B = \frac{1}{2\sigma^2} \{ q_1 q_2 \sum_{i=1}^{N} 2x_{i-1} x_{i-2} - 2q_1 q_2 \sum_{i=1}^{N} x_{i-2}^2 \}$$

$$= \frac{1}{2\sigma^2} \{ q_1 q_2 \sum_{i=1}^{N} x_{i-1}^2 + x_{i-2}^2 - (x_{i-1} - x_{i-2})^2 - 2q_1 q_2 \sum_{i=1}^{N} x_{i-2}^2 \} \quad . \tag{D.15}$$

Reducing the summation indices to common base provides the result

$$B = \frac{1}{2\sigma^2} q_1 q_2 \{ \sum_{i=1}^{N} (x_i - x_{i-1})^2 + x_o^2 - x_N^2 - 2x_o x_1 + 2x_N x_{N-1} \} \tag{D.16}$$

The term under the summation can be recognized as the quadratic variation and, for the M=2 process, goes to zero as the sampling becomes dense. Then in the limit ($h \to 0$), equation (D.16) will be of the form

$$B = \frac{1}{2\sigma^2} \; q_1 q_2 \; \{x_T^2 - x_o^2\} \quad , \tag{D.17}$$

by use of the a.s. sample path continuity of the x process.

The next candidate for reduction is the term

$$C \triangleq \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} h^4 q_1^2 x_{i-2}^2$$

$$= \frac{q_1^2}{2\sigma^2} \left\{ \sum_{i=1}^{N} x_i^2 \cdot h + h[x_o^2 + x_1^2 - x_N^2 - x_{N-1}^2] \right\} , \tag{D.18}$$

as the sampling interval becomes dense ($h \to 0$), because of the a.s. sample continuity and the finite average power

$$C = \frac{q_1^2}{2\sigma^2} \sum_{i=1}^{N} x_i^2 h \longrightarrow \frac{q_1^2}{2\sigma^2} \int_o^T x_i^2(t)dt \; . \tag{D.19}$$

The last factor to be reduced is

$$D \triangleq \frac{1}{2\sigma^2 h^3} \sum_{i=1}^{N} h^4 q_2^2 \dot{x}_{i-2}$$

$$= \frac{q_2^2}{2\sigma^2} \sum_{i=1}^{N} \dot{x}_{i-2} \cdot h$$

$$= \frac{q_2^2}{2\sigma^2} \sum_{i=1}^{N} \dot{x}_i^2 \cdot h + h(\dot{x}_o^2 + \dot{x}_1^2 - \dot{x}_N^2 - \dot{x}_{N-1}^2) \quad , \tag{D.20}$$

where under dense sampling, and since this is an M=2 process

$$D = \frac{q_2^2}{2\sigma^2} \sum_{i=1}^{N} \dot{x}_i^2 \cdot h \longrightarrow \frac{q_2^2}{2\sigma^2} \int_o^T \dot{x}^2(t)dt \; . \tag{D.21}$$

Returning to the factor denoted by A (D.14) where

$$A = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2(q_1 x_{i-2} + q_2 \dot{x}_{i-2})(\dot{x}_{i-1} - \dot{x}_{i-2})$$

and reducing the product to

$$A = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_i x_{i-2}(\dot{x}_{i-1} - \dot{x}_{i-2}) + \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_2 \dot{x}_{i-2}(\dot{x}_{i-1} - \dot{x}_{i-2}) \tag{D.22}$$

so that the identity

$$2\dot{x}_{i-2}\dot{x}_{i-1} = (\dot{x}_{i-1}^2 + \dot{x}_{i-2}^2) - (\dot{x}_{i-1} - \dot{x}_{i-2})^2 \quad ,$$

can be applied to the last sum in (D.22). The reduction afforded by this identity is reflected as

$$A = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_1 (\dot{x}_{i-1} - \dot{x}_{i-2}) x_{i-2}$$

$$+ \frac{1}{2\sigma^2} \sum_{i=1}^{N} q_2 (\dot{x}_{i-1}^2 - \dot{x}_{i-2}^2) - 2\dot{x}_{i-2}^2 - (\dot{x}_{i-1} - \dot{x}_{i-2})^2 \quad .$$

Further, reducing the indexed quantities to the same base results in

$$A = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_i x_{i-2}(\dot{x}_{i-1} \dot{x}_{i-2})$$

$$+ \frac{1}{2\sigma^2} \sum_{i=1}^{N} (-q_2)(\dot{x}_i - \dot{x}_{i-1})^2 + q_2(\dot{x}_N^2 - \dot{x}_o^2) \quad . \tag{D.23}$$

As the sampling becomes dense the last summation on (D.23) is recognized to be a constant ($\sigma^2 T$) via the application of Baxters Theorem. Now, the term left to be reduced is the

first summation in (D.23)

$$Al = \frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_i x_{i-2}(\dot{x}_{i-1} - \dot{x}_{i-2}) \quad . \tag{D.24}$$

To reduce this to a recognizable form, the summation by parts rule

$$\sum_{i=1}^{N} x_i \nabla_2 x_i = x_i \nabla_1 x_i \Big]_1^{N+1} - \sum_{i=1}^{N} \nabla_1 x_{i+1} \nabla_1 x_i \quad , \tag{D.25}$$

where ∇_1, and ∇_2 denote the first and second differences, is applied to (D.24) with the result

$$\frac{1}{2\sigma^2} \sum_{i=1}^{N} 2q_1 x_{i-2}(\dot{x}_{i-1} - \dot{x}_{i-2}) = \frac{1}{2\sigma^2 h} \sum_{i=1}^{N} 2q_1 x_{i-2} \nabla_2 x_{i-2} \quad ,$$

$$\frac{1}{2\sigma^2 h} \sum_{i=1}^{N} 2q_1 x_{i-2} \nabla_2 x_{i-2} = \frac{2q_1}{2\sigma^2 h} \Big\{ x_{i-2} \nabla_1 x_{i-2} \Big]_1^{N+1}$$

$$- \sum_{i=1}^{N} \nabla_1 x_{i-1} \nabla_1 x_{i-2} \Big\} \quad . \tag{D.26}$$

Rearranging the indices of the summand in (D.26) provides the result

$$- \frac{1}{2\sigma^2 h} \sum_{i=1}^{N} 2q_i \nabla_1 x_{i-1} \nabla_1 x_{i-2} = - \frac{1}{2\sigma^2 h} \sum_{i=1}^{N} 2q_1$$

$$[\nabla x_1 \nabla x_{i-1} + \nabla_1 x_o \nabla_1 x_{-1} - \nabla_1 x_N \nabla_1 x_{N-1}] \quad . \tag{D.27}$$

Using the identity,

$$2\nabla x_i \nabla x_{i-1} = \nabla_1^2 x_{i-1} + \nabla_1^2 x_i - (\nabla_1 x_1 - \nabla_1 x_{i-1})^2$$

in equation (D.27) yields

$$- \frac{1}{2\sigma^2 h} \sum_{i=1}^{N} 2q_i \nabla_1 x_{i-1} \nabla_1 x_{i-2} = - \frac{1}{2\sigma^2 h} \{ 2q_1 [\nabla_1 x_o \nabla_1 x_{-1} - \nabla_1 x_N \nabla_1 x_{N-1}]$$

$$+ \sum_{i=1}^{N} q_i [\nabla_1^2 x_i + \nabla_1^2 x_{i-1} - (\nabla_1 x_i - \nabla_1 x_{i-1})^2] \cdot \qquad (D.28)$$

Collecting terms the result is

$$Al = \frac{2q_i}{2\sigma^2 h} [x_N \nabla_1 x_N - x_o \nabla_1 x_o - \nabla_1 x_o x_{-1} + \nabla_1 x_N \nabla_1 x_{N-1}]$$

$$- \frac{1}{2\sigma^2} \left[\sum_{i=1}^{N} 2q_1 \frac{\nabla_1^2 x_i}{h^2} \cdot h + 2q_i h \left(\frac{\nabla_1^2 x_o - \nabla_1^2 x_N}{h^2} \right) \right.$$

$$\left. - \sum_{i=1}^{N} h 2q_1 \left(\frac{\nabla x_i - \nabla x_{i-1}}{h} \right)^2 \right] \cdot \qquad (D.29)$$

Realizing that for an M=2 process, the derivative exists and that the quadratic variation of the derivative approaches a bounded constant as the sampling becomes dense ($h \to o$), the limiting form of equation (D.29) is

$$\lim_{\substack{h \to 0 \\ N \to \infty}} Al = \frac{2q_1}{2\sigma^2} \lim_{h \to 0} \left[x_N \frac{\nabla_1 x_N}{h} - \frac{x_o \nabla_1 x_o}{h} - \frac{h \nabla_1 x_o \nabla_1 x_{-1}}{h^2} \right.$$

$$\left. + \frac{h \nabla_1 x_N \nabla_1 x_{N-1}}{h^2} \right]$$

$$- \frac{1}{2\sigma^2} \lim_{h \to 0} \left\{ \sum_{i=1}^{N} 2q_i \left(\frac{\nabla_1 x_i}{h} \right)^2 \cdot h + 2q_1 h \left[\left(\frac{\nabla_1 x_o}{h} \right)^2 - \left(\frac{\nabla_1 x_N}{h} \right)^2 \right] \right.$$

$$\left. - \sum_{i=1}^{N} 2q_1 \cdot h \left(\frac{\nabla x_i}{h} - \frac{\nabla x_{i-1}}{h} \right)^2 \right.$$

$$= \frac{2q_1}{2\sigma^2} [x_T \dot{x}_T - x_o \dot{x}_o] - \frac{1}{2\sigma^2} \int_o^T 2q_1 \dot{x}_o^2 dt \cdot \qquad (D.30)$$

Collecting all the reduced factors to form the likelihood ratio, the result is

$$\lambda(x(t)/\sigma^2 \underline{x}_o, q) \propto \exp \{ - \frac{1}{2\sigma^2} [q_1^2 \int_o^T x^2(t)dt$$

$$+q_1 q_2 (x_T^2 - x_o^2) + q_2^2 \int_o^T \dot{x}^2(t)dt - q_2(\sigma^2 T + \dot{x}_o^2 - \dot{x}_T^2)$$

$$-2q_i \int_o^T \dot{x}^2(t)dt - 2q_1 (x_T \dot{x}_T - x_o \dot{x}_o) \} \qquad \text{(D.31)}$$

APPENDIX E

Calculation of the Variance of the Random Variable $_2B_N$

In Chapter 6, the variance calculation (equation 6.72) centered around assuming that exponential terms are adequately well represented by

$$e^{-p\delta} = 1-p\delta \quad .$$

The intent here is to postpone such considerations, and carry out the calculation using the complete form. The variance is calculated from

$$\mathrm{Var}(_2B_N) = 2\left\{\sum_{j=k=1}^{N} E^2(\underline{\Delta}\dot{x}_k^2) + 2\sum_{j>k=1}^{N} E^2(\underline{\Delta}\dot{x}_k\underline{\Delta}\dot{x}_j)\right\} \quad . \quad (E.1)$$

To facilitate the calculation each sum is considered separately. The first sum is determined to be

$$\sum_{j=1}^{N} E^2(\underline{\Delta}\dot{x}_j^2) = \sum_{j=1}^{N} 2\left[R_x^{\cdot}(0)-R_x^{\cdot}(\delta)\right]^2$$

$$= N\left\{\frac{\sigma^4}{(p_2^2-p_1^2)}\left[p_2(1-e^{-p_2\delta}) - p_1(1-e^{-p_1\delta})\right]^2\right\} \quad (E.2)$$

where $R_x^{\cdot}(n\delta)$ is defined as

$$R_x^{\cdot}(n\delta) = \frac{\delta^2}{2(p_2^2-p_1^2)}\left\{p_2e^{-p_2n\delta} - p_1e^{-p_1n\delta}\right\} \quad . \quad (E.3)$$

The second sum is determined to be

$$2\sum_{j>k=1}^{N}\sum E^2(\dot{\Delta x}_k\dot{\Delta x}_j) = 2\Big(\sum_{j>k=1}^{N}\sum 2R_{\dot{x}}(j-k)-R_{\dot{x}}(j-k-1)-R_{\dot{x}}(j-k+1)\Big)^2$$

$$= 2\sum_{j>k=1}^{N}\sum K^2[p_2^2 e^{-2(j-k)p_2\delta}(1-\cosh p_2\delta)^2$$

$$+ 2p_1p_2e^{-(j-k)(p_2+p_1)}(1-\cosh p_1\delta)(1-\cosh p_2\delta)$$

$$+ p_1^2e^{-2(j-k)p_1\delta}(1-\cosh p_1\delta)^2] \quad , \quad (E.4)$$

where $K=\dfrac{\delta^2}{(p_2^2-p_1^2)}$. Rewriting the double sum in the form of a single sum the result is

$$2\sum_{j>k=1}^{N}\sum E^2(\dot{\Delta x}_k\dot{\Delta x}_j) = 2\sum_{j-k=m=1}^{N-1}K^2\left(1-\frac{m}{N}\right)$$

$$[p_2^2e^{-2(j-k)p_2\delta}(1-\cosh p_2\delta)^2 - 2p_1p_2e^{-(j-k)(p_2+p_1)}$$

$$(1-\cosh_1\delta)(1-\cosh p_2\delta) + p_1^2e^{-2(j-k)p_1\delta}(1-\cosh p_1\delta)^2] \cdot \quad (E.5)$$

Again there are two sums to be considered. The first sum is

$$A = \sum_{m=1}^{N-1}[p_2^2e^{-2mp_2\delta}(1-\cosh p_2\delta)^2 - 2p_1p_2e^{-m(p_1+p_2)\delta}$$

$$(1-\cosh p_1\delta)(1-\cosh p_2\delta) + p_1^2e^{-2mp_1\delta}(1-\cosh p_1\delta)^2]$$

$$= \frac{p_2^2(1-\cosh p_2h\delta)^2\, e^{-p_2\delta}(1-e^{-2p_2T})}{2\sinh(p_2\delta)}$$

$$- \frac{2p_1p_2(1-\cosh p_1\delta)(1-\cosh p_2\delta)e^{-\frac{(p_1p_2)\delta}{2}}(1-e^{-(p_1+p_2)T}}{2\sinh\left(\frac{p_1+p_2}{2}\right)\delta)}$$

$$+ \frac{p_1^2(1-\cosh p_1\delta)^2\, e^{-p_1\delta}(1-e^{-p_1T})}{2\sinh(p_1\delta)} \quad , \tag{E.6}$$

where the substitution $T=N\,\delta$ has been made and T is fixed.

The second sum is

$$B = (-1)\sum_{m=1}^{N-1} \frac{m}{N}\left[p_2^2 e^{-2mp_2\delta}(1-\cosh p_2\delta)^2 - 2p_1p_2 e^{-m(p_1+p_2)\delta}\right.$$

$$\left.(1-\cosh p_1\delta)(1-\cosh p_2\delta) + p_1^2 e^{-2mp_1\delta}(1-\cosh p_1\delta)^2\right] \cdot \tag{E.7}$$

To facilitate manipulations, let

$$B_1 = p_2^2(1-\cosh p_2\delta)\sum_{m=1}^{N-1} m\, e^{-2mp_2\delta} \tag{E.8}$$

$$B_2 = -2p_1p_2(1-\cosh p\delta)(1-\cosh p_2\delta)\sum_{m=1}^{N-1} m\, e^{-m(p_1+p_2)\delta}$$

$$B_3 = p_1^2(1-\cosh p_1\delta)^2\sum_{m=1}^{N-1} m\, e^{-2mp_1\delta} \quad , \tag{E.10}$$

so that

$$B = \frac{(-1)}{N}[B_1+B_2+B_3] \quad , \tag{E.11}$$

$$= \frac{(-1)\delta}{T}\sum_i B_i \quad \cdot$$

Evaluation of the sums associated with the B_i factors provides the result

$$B_1 = p_2^2\left\{(1-\cosh p_2\delta)^2(1-e^{-p_2T})\, e^{-p_2\delta}(\sinh p^2\delta-\cosh p_2\delta)\right\}$$

$$/4\sinh^2 p_2\delta \quad , \tag{E.12}$$

$$B_2 = -2p_1p_2\{(1-\cosh p_1\delta)(1-\cosh p_2\delta)\, e^{\frac{-(p_1+p_2)\delta}{2}}(1-e^{\frac{-(p_1+p_2)T}{2}}) \left(\sinh \frac{(p_1+p_2)\delta}{2} - \cosh \frac{(p_1+p_2)\delta}{2}\right)\} / 4 \sinh^2 \frac{(p_1+p_2)\delta}{2} \quad , \tag{E.13}$$

$$B_3 = p_1^2\{(1-\cosh p_1\delta)^2(1-e^{-p_1T})e^{-p_1\delta}(\sinh p_1\delta - \cosh p_2\delta)\} / 4\sinh^2 p_1\delta \quad . \tag{E.14}$$

With the constituent parts determined the variance expression is

$$\mathrm{Var}({}_2B_N) = 2\left\{\frac{N\sigma^4}{(p_2^2-p_1^2)^2}\left[p_2(1-e^{-p_2\delta})-p_1(1-e^{-p_1\delta})\right]^2 + \frac{2\sigma^4}{(p_2^2-p_1^2)^2}\left[A - \frac{1}{N}B\right]\right\} \quad , \tag{E.15}$$

where A and B have been defined in equations (E.6) and (E.11). Rather than attempting to simplify this expression, the assumption used throughout this presentation will be invoked. That is, assume that the sampling rate is such that the approximation

$$e^{-p\delta} = 1-p\delta,$$

is valid. Under this condition, the factors A and B are zero and the variance expression reduces to

$$\mathrm{Var}({}_2B_N) = 2\sigma^4 N\delta^2 = \frac{2\sigma^4 T^2}{N} \quad , \tag{E.16}$$

where T is a fixed observation interval.

APPENDIX F

Moment Calculations of the Noise Corrupted Random Variable $(\hat{B}_N)$

The purpose of this appendix is to illustrate the steps leading to the mean and variance computation of the noise corrupted measurement $(\hat{B}_N)$, for the M=1 and M=2 processes.

For the M=1 process let

$$\hat{B}_N = \sum_k \underline{\Delta}\hat{x}_k^2 \tag{F.1}$$

where $$\underline{\Delta}\,\hat{x}_k = x_k + e_k - (x_{k-1} - e_{k-1}) = \underline{\Delta}x_k + \underline{\Delta}\,e_k \tag{F.2}$$

and e_k represents the zero mean, white and data independent quantizing noise with variance σ_e^2. The mean is

$$\begin{aligned} E(\hat{B}_N) &= \sum_k E(\underline{\Delta}\,\hat{x}_k^2) \\ &= \sum_k E(\underline{\Delta}\,x_k^2 + 2\underline{\Delta}e_k\,\underline{\Delta}x_k + \underline{\Delta}\,e_k^2) \\ &= \sum_k E(\underline{\Delta}\,x_k^2) + 2E(\underline{\Delta}e_k^2) \\ &= E(B_N) + 2N\sigma_e^2 \end{aligned} \tag{F.3}$$

Begin the variance computation by defining

$$E(\hat{B}_N 2) = \sum_k \sum_j E(\underline{\Delta}\,x_k^2\,\underline{\Delta}x_j^2) \tag{F.4}$$

$$\begin{aligned} E(\hat{B}_N^2) = \sum_j \sum_k [\, & E(\underline{\Delta}\,x_k^2\,\underline{\Delta}\,x_j^2) + 2E(\underline{\Delta}e_k\,\underline{\Delta}x_k\underline{\Delta}x_j^2) \\ & + E(\underline{\Delta}\,e_k^2\,\underline{\Delta}x_j^2) + 2E(\underline{\Delta}e_j\underline{\Delta}\,x_j\,\underline{\Delta}x_k^2) \\ & + 4(\underline{\Delta}e_j\,\underline{\Delta}\,e_k\underline{\Delta}x_j\,\underline{\Delta}_k) + 2E(\underline{\Delta}e_k^2\,\underline{\Delta}e_j\,\underline{\Delta}x_j) \end{aligned}$$

$$+ E\,(\underline{\Delta}e_j^{\,2}\,\underline{\Delta}x_k^{\,2}) + 2E\,(\underline{\Delta}e_j^{\,2}\,\underline{\Delta}e_k\,\underline{\Delta}x_k)$$

$$+ E(\underline{\Delta}\,e_k^{\,2}\,\underline{\Delta}e_j^{\,2})] \quad . \tag{F.5}$$

Using the properties of the quantizing noise then

$$E(\underline{\Delta}e_k\,\underline{\Delta}x_k\,\underline{\Delta}x_j^{\,2}) = E[(e_k - e_{k-1})(\underline{\Delta}x_k\,\underline{\Delta}x_j^{\,2})]$$

$$= E(e_k\,\underline{\Delta}x_k\,\underline{\Delta}x_j^{\,2} - e_{k-1}\underline{\Delta}x_k\underline{\Delta}x_j^{\,2})$$

$$= E(e_k)\,E(\underline{\Delta}x_k\underline{\Delta}x_j^{\,2}) - E(e_{k-1})\,E(\underline{\Delta}\,x_k\,\underline{\Delta}x_j^{\,2}) = 0, \tag{F.6}$$

$$E(\underline{\Delta}\,e_k^{\,2}\,\underline{\Delta}x_j^{\,2}) = E(\underline{\Delta}e_k^{\,2})\,E(\underline{\Delta}x_j^{\,2})$$

$$= 2\sigma_e^{\,2}\,E(\underline{\Delta}x_j^{\,2}) \quad , \tag{F.7}$$

$$E(\underline{\Delta}\,e_j\,\underline{\Delta}x_j\,\underline{\Delta}x_k^{\,2}) = 0 \quad , \tag{F.8}$$

$$E(\underline{\Delta}e_j\underline{\Delta}e_k\underline{\Delta}x_j\,\underline{\Delta}x_k) = \begin{cases} E(\underline{\Delta}e_k^{\,2})\,E(\underline{\Delta}x_k^{\,2}) = 2\sigma_e^{\,2}\,E(\underline{\Delta}x_k^{\,2}), & j=k \\ E(\underline{\Delta}e_j\,\underline{\Delta}e_k)E(\underline{\Delta}x_k\,\underline{\Delta}x_j) = 0, & j\neq k \end{cases} \tag{F.9}$$

$$E(\underline{\Delta}\,e_j^{\,2}\underline{\Delta}x_k^{\,2}) = E(\underline{\Delta}e_j^{\,2})\,E(\underline{\Delta}x_k^{\,2}) = 2\sigma_e^{\,2}\,E(\underline{\Delta}x_k^{\,2}), \tag{F.10}$$

$$E(\underline{\Delta}\,e_k^{\,2}\underline{\Delta}e_j^{\,2}) = E(\underline{\Delta}e_k^{\,2})\,E(\underline{\Delta}e_j^{\,2}) = \begin{cases} (2\sigma_e^{\,2})^2, & j\neq k \\ E(\underline{\Delta}e_k^{\,4}) & j=k, \end{cases} \tag{F.11}$$

Using these terms

$$E(\hat{B}_N^{\,2}) = \sum_j \sum_k E(\underline{\Delta}\hat{x}_k^{\,2}\,\underline{\Delta}\,\hat{x}_j^{\,2})$$

$$= \sum_{j=k} E(\underline{\Delta}x_k^{\,4}) + 12\sigma_e^{\,2}\,E(\underline{\Delta}x_k^{\,2}) + E(\underline{\Delta}e_k^{\,4})$$

$$+ 2\sum_{j>k} \sum E(\underline{\Delta}x_k^{\,2}\underline{\Delta}x_j^{\,2}) + \sigma_e^{\,2}\,[2E(\underline{\Delta}x_j^{\,2}) + 2E(\underline{\Delta}x_k^{\,2}) + 4\,\sigma_e^{\,2}] \tag{F.12}$$

Computing the variance

$$\text{Var}\,(\hat{B}_N) = \sum_{j=k} E(\underline{\Delta}x_k^4) + 2\sum_{j>k}\sum E(\underline{\Delta}x_k^2\,\underline{\Delta}x_j^2) \quad -E^2(B_N)$$
$$+ \sum_{j=k} 12\,\sigma_e^2\,E(\underline{\Delta}x_k^2) + 2E(e_k^4) + 6(\sigma_e^2)^2$$
$$+ 2\sum_{j>k}\sum \sigma_e^2[\,2E(\underline{\Delta}x_k^2) + 2E(\underline{\Delta}x_j^2) + 4\sigma_e^2]$$
$$- 4N\sigma_e^2\,E(B_N) - 4N^2(\sigma_e^2)^2 \qquad \text{(F.13)}$$

$$\text{Var}(\hat{B}_N) = \text{Var}(B_N) + 12N\sigma_e^2\,E(\underline{\Delta}x_k^2) + 2N\,E(e_k^4)$$
$$+ 6N(\sigma_e^2)^2 + 4N(N-1)\,\sigma_e^2\,E(\underline{\Delta}x_k^2)$$
$$+4N(N-1)(\sigma_e^2)^2 - 4N\sigma_e^2\,E(B_N) \quad -4N^2(\sigma_e^2)^2, \qquad \text{(F.14)}$$
$$= \text{Var}\,(B_N) + 8\sigma_e^2\,E(B_N) + \quad 2N(\,\sigma_e^4 + E(e_k^4))$$

Expressing the quantity $E(e_k^4)$ in terms of the second moment $E(e_k^2)$, $E(e_k^4) = 1.8\,\sigma_e^4$, then

$$\text{Var}(\hat{B}_N) = \text{Var}(B_N) + 8\sigma_e^2\,E(B_N) + 5.6N\,\sigma_e^4 \qquad \text{(F.15)}$$

In dealing with the second order process it will be necessary to compute the moments of the quantities

$$1\hat{B}_N = \sum_k \underline{\Delta}\,\hat{x}_k^2, \qquad \text{(F.16)}$$
$$2\hat{B}_N = \sum_k \underline{\Delta}\,\dot{\hat{x}}_k^2, \qquad \text{(F.17)}$$

where for analog differentiation

$$\underline{\Delta}\hat{x}_k = \underline{\Delta}x_k + \underline{\Delta}e_{k_A} \qquad \text{(F.18)}$$
$$\underline{\Delta}\dot{\hat{x}}_k = \underline{\Delta}\dot{x}_k + \underline{\Delta}e_{k_B}\,. \qquad \text{(F.19)}$$

In this situation two different independent noise sources have been modeled where their respective variances are $\text{Var}(e_{k_A}) = \sigma_{e_A}^2$,

$$\mathrm{Var}(e_{k_B}) = \sigma^2_{e_B} \quad .$$

Following the development for the M=1 process

$$E(\hat{{}_1B_N}) = \sum_k E(\underline{\Delta}\, \hat{x}_k^2)$$

$$= E({}_1B_N) + 2N\sigma_{e_A}^2 \tag{F.20}$$

$$\mathrm{Var}(\hat{{}_1B_N}) = \mathrm{Var}\,({}_1B_N) + 12N\sigma_{e_A}^2\, E(\underline{\Delta}\, x_k^2)$$
$$+2N\, E(e_{k_A}^4) + 6N\,(\sigma_{e_A}^2)^2$$
$$+4N(N-1)(\sigma_{e_A}^2)^2 - 4N\sigma_{e_A}^2\, E({}_1B_N)$$
$$-4N^2(\sigma_{e_A}^2)^2 \quad . \tag{F.21}$$

$$\mathrm{Var}(\hat{{}_1B_N}) = \mathrm{Var}({}_1B_N) + 8\sigma_{e_A}^2\, E({}_1B_N) + 5.6N\,\sigma_{e_A}^4 \tag{F.22}$$

$$E(\hat{{}_2B_N}) = \sum_k E(\underline{\Delta}\, \hat{\dot{x}}_k^2)$$

$$= E({}_2B_N) + 2N\,\sigma_{e_B}^2 \tag{F.23}$$

$$\mathrm{Var}(\hat{{}_2B_N}) = \mathrm{Var}\,({}_2B_N) + 8\sigma_{e_B}^2\, E({}_2B_N) + 5.6\, N\sigma_{e_B}^4 \tag{F.24}$$

In dealing with the M=2 process construction of the derivative process can also be modeled by digital differentiation. In this situation the noise corrupted measurements are

$$\hat{x}_k = x_k + e_k \ ,$$
$$\hat{\dot{x}}_k = \dot{x}_k + (e_k - e_{k-1})/h$$

where h is the sampling interval.

The noise source is assummed to be independent with variance $\sigma_{e_1}^2$. Following the previous developments

$$E(\hat{{}_1B_N}) = \sum_k E(\underline{\Delta}\,\hat{x}_k^2)$$

$$= E({}_1B_N) + 2N\sigma_{e_1}^2 \qquad (F.25)$$

$$Var(\hat{{}_1B_N}) = Var({}_1B_N) + 8\sigma_{e_1}^2 E({}_1B_N) + 5.6N\sigma_{e_1}^4 \qquad (F.26)$$

For the measurement $\hat{{}_2B_N}$ define

$$\underline{\Delta}\,\hat{\dot{x}}_k = \frac{1}{h}\{(x_k - x_{k-1}) - (x_{k-1} - x_{k-2}) + (e_k - e_{k-1}) - (e_{k-1} - e_{k-2})\}$$

$$= \underline{\Delta}\dot{x}_k + (\underline{\Delta}\,e_k - \underline{\Delta}e_{k-1})/h \qquad (F.27)$$

The the mean of $\hat{{}_2B_N}$ is

$$E(\hat{{}_2B_N}) = \sum_k E(\underline{\Delta}\,\hat{\dot{x}}_k^2)$$

$$= \sum_k E(\underline{\Delta}\dot{x}_k^2) + \frac{2}{h}E(\underline{\Delta}e_k - \underline{\Delta}e_{k-1})\,\underline{\Delta}\dot{x}_k$$

$$+ E\,\frac{(\underline{\Delta}e_k - \underline{\Delta}e_{k-1})^2}{h^2}$$

$$= E({}_2B_N) + 6N\,\frac{\sigma_e^2}{h^2}$$

$$= E({}_2B_N)\left(1 + \frac{6\sigma_e^2}{\sigma^2 h^3}\right) \qquad (F.28)$$

The variance computation begins with

$$E(\hat{{}_2B_N^2}) = E\{\sum \underline{\Delta}\,\hat{\dot{x}}_j^2\,\underline{\Delta}\hat{\dot{x}}_k^2\}$$

$$= E\{\sum(\underline{\Delta}\dot{x}_k^2\,\underline{\Delta}\dot{x}_j^2 + \frac{2}{h}(\underline{\Delta}e_k - \underline{\Delta}\,e_{k-1})\,\underline{\Delta}\,\dot{x}_k\,\underline{\Delta}\dot{x}_j^2$$

$$+ \frac{1}{h^2} (\underline{\Delta}e_k - \underline{\Delta} e_{k-1})^2 \underline{\Delta}\dot{x}_j^2 + \frac{2}{h}(\underline{\Delta}e_j - \underline{\Delta}e_{j-1}) \underline{\Delta}\dot{x}_j \underline{\Delta}\dot{x}_k^2$$

$$+ \frac{4}{h^2} (\underline{\Delta}e_k - \underline{\Delta}e_{k-1}) \underline{\Delta}\dot{x}_j \underline{\Delta}\dot{x}_k$$

$$+ \frac{2}{h^3} (\underline{\Delta}e_j - \underline{\Delta}e_{j-1})(\underline{\Delta}e_k - \underline{\Delta}e_{k-1})^2 \underline{\Delta}\dot{x}_j$$

$$+ \frac{\underline{\Delta}\dot{x}_k^2}{h^2} (\underline{\Delta}e_j - \underline{\Delta}e_{j-1})^2$$

$$+ \frac{2}{h^3} (\underline{\Delta}e_k - \underline{\Delta}e_{k-1})(\underline{\Delta}e_j - \underline{\Delta}e_{j-1})^2 \underline{\Delta}\dot{x}_k$$

$$+ \frac{(\underline{\Delta}e_k - \underline{\Delta}e_{k-1})^2}{h^2} \frac{(\underline{\Delta}e_j - \underline{\Delta}e_{j-1})^2}{h^2} \Big\} \quad . \tag{F.29}$$

$$E(\hat{{}_2B_N^2}) = \sum_{j=k} \Big\{ E(\underline{\Delta}\dot{x}_k^4) + \frac{6\sigma_{e_1}^2}{h^2} E(\underline{\Delta}\dot{x}_j^2)$$

$$+ \frac{24\sigma_{e_1}^2}{h^2} E(\underline{\Delta}\dot{x}_j^2) + \frac{6\sigma_{e_1}^2}{h^2} E(\underline{\Delta}\dot{x}_k^2)$$

$$+ \frac{54}{h^4} (\sigma_{e_1}^2)^2 + 18\, E(e_k^4) \Big\}$$

$$+ 2 \sum_{j>k} \Big\{ E(\underline{\Delta}\dot{x}_k^2 \underline{\Delta}\dot{x}_j^2) + \frac{6\sigma_{e_1}^2}{h^2} E(\underline{\Delta}\dot{x}_j^2)$$

$$+ \frac{6\sigma_{e_1}^2}{h^2} E(\underline{\Delta}\dot{x}_k^2) + \frac{36}{h^4} (\sigma_{e_1}^2)^2 \quad . \tag{F.30}$$

$$E^2(\hat{{}_2B_N}) = E^2({}_2B_N) + 12 N\sigma_{e_1} E({}_2B_N) + 36 N^2 (\sigma_{e_1}^2)^2 \frac{}{h^2} \tag{F.31}$$

$$Var(\hat{{}_2B_N}) = Var({}_2B_N) + 36 N\sigma_{e_1}^2 \frac{}{h^2} E(\underline{\Delta}\dot{x}_k^2) + 54N (\sigma_{e_1}^2)^2 \frac{}{h^2}$$

$$+18\ N\ E(e_k^4) + N(N-1)\ 12\ \frac{\sigma_{e_1}^2}{h^2}\ E(\underline{\dot{\Delta}}\dot{x}_k^2)$$

$$+36\ N(N-1)\ \frac{(\sigma_{e_1}^2)^2}{h^2} - 12N\ \frac{\sigma_{e_1}^2}{h^2}\ E(\underline{\dot{\Delta}}\dot{x}_k^2)\ . \qquad \text{(F.32)}$$

$$\text{Var}\ (\hat{{}_2B_N}) = \text{Var}\ ({}_2B_N) + \frac{24\sigma_{e_1}^2}{h^2}\ E({}_2B_N) + 18\ N(2.8)\ \frac{\sigma_e^4}{h^4} \qquad \text{(F.33)}$$

BIBLIOGRAPHY

Abramowitz, M. and I. Stegun, Handbook of Mathematical Functions, Dover Publications, Inc., New York, 1970.

Andersen, N., "On the Calculation of Filter Coefficients for Maximum Entropy Spectral Analysis," Geophysics, Vol. 39, 1974, pp. 69-72.

Anderson, T. W., The Statistical Analysis of Time Series, John Wiley, New York, 1971.

Astrom, K. J., Introduction to Stochastic Control Theory, Academic Press, New York, 1970.

Bartlett, M. S., "Periodogram Analysis and Continuous Spectra," Biometrika, 37, 1950, pp.1-16.

Baxter, G., "A Strong Theorem for Gaussian Processes," Proc. Am. Math. Soc., 7, 1956, pp. 522-528.

Bennett, W. R., "Spectra of Quantized Signals," BSTJ, Vol. 27, 1948, pp. 446-472.

Birdsall, T. G., CICE 521, 1975, Lecture Notes

Blackman, R. B., and J.W. Tukey, The Measurement of Power Spectra, Dover Publications, Inc., New York, 1959.

Box, G. and G. Tiao, Bayesian Inference in Statistical Analysis, Addison-Wesley Publishing Co., Reading, Mass., 1973.

Brilinger, D. R., Time Series: Data Analysis on Theory, Holt, Rinehart, and Winston, New York, 1975.

Burg, J., "Maximum Entropy Analysis," 1967, presented at the 37th annual meeting Soc. Explor. Geophysics, Oklahoma City, Oklahoma.

Burg, J., "A New Analysis Technique for Time Series Data," 1968, presented at Advanced Study Institute on Signal Processing, NATO, Enschede, Netherlands.

DeGroot, M. H., Optimal Statistical Decisions, McGraw-Hill Book Company, New York, 1970.

Doob, J. L., Stochastic Processes, John Wiley, New York, 1953.

Durbin, J., "The Fitting of Time Series Models," Revue Inst. de Stat., 28, 1960, pp. 233-243.

Edwards, J. and M. Fitelson, "Notes on Maximum Entropy Processing," IEEE Trans. on Info. Thy., IT-19, 1973, pp. 232-234.

Edwards, W., H. Lindman, and L. Sarage, "Bayesian Statistical Inference for Psychological Research," Psychol. Rev., 70, 1963, pp. 193-242.

Feldman, J., "Equivalence and Perpendicularity of Gaussian Measures," Pacific J. Math., 8, 1958, pp. 699-708.

Fuller, W.A., Introduction to Statistical Time Series, John Wiley and Sons, New York, 1976.

Gikhman, I.I., "On Theory of Differential Equations of Stochastic Processes I-II," Am. Math Society Trans., Vol 1, Series 2, 1951, pp. 111-161.

Gobien, J.G., "Simultaneous Detection and Estimation: the Use of Sufficient Statistics and Reproducing Densities," Ph.D. Thesis, CICE Program, Univ. of Michigan, Ann Arbor, Michigan, 1974.

Grenander, N. and G. Szego, Toeplitz Forms and Their Applications, University of Calif. Press, Berkeley, 1958.

Hajek, J., "On Linear Statistical Problems in Stochastic Processes," Czech. Math. Journal, 1962, Vol. 12.

Halmns, P.R., Measure Theory, Van Nostrand, Princeton, N.J.,1950.

Jenkins, G.M. and D.G. Watts, Spectral Analysis and its Application, Holden-Day, San Francisco, 1968.

Kailath, T. "Innovation's Approach to Detection and Estimation," Proc. IEEE, 58, 1970, pp. 680-695.

Kromer, R., Asymptotic Properties of the Autoregressive Spectral Estimator, Ph.D. Thesis, Dept. of Statist., Stanford Univ., Stanford, California, 1970.

Kushner, H., Introduction to Stochastic Control, Holt, Rinehart and Winston Inc., New York, 1971.

Levinson, N., "The Wiener RMS Error Criterion in Filter Design," J. Math Physics, 25, 1947, pp. 261-278.

McDonough, R., "Maximum Entropy Spatial Processing of Array Data," Geophysics, 39, 1974, pp. 843-851.

Raiffa, H. and R. Schlaiffer, Applied Statistical Decision Theory, The Mit Press, Cambridge, Mass., 1961.

Reguicha, A.A., "Expected Values of Functions of Quantitized Random Variables," IEEE Trans on Comm. Thy., Vol 21.,1921, pp. 850-854.

Shepp, L.A., "Radon-Nikodyn Derivatives of Gaussian Processes," Ann. Math. Stat., 37, 1966, pp. 321-354.

Slepian, D., "Some Comments on the Detection of Gaussian Signals in Gaussian Noise," IRE Trans. on Info. Thy., 4, 1958, pp. 65-68.

Soong, T.T., Random Differential Equations, Academic Press, New York, 1973.

Spraggins, J., "A Note on the Iterative Application of Bayes Rule," IEEE Trans. on Info. Thy., IT-11, 1965, pp. 544-549.

Symlie, D.E., K.C. Clarke, and T.J. Ulrych, "Analysis of Irregularities in the Earth's Rotation," Methods in Computational Physics, 13, 1973, pp. 391-430, Academic Press, New York.

Van Den Bos, A., "Alternative Interpretation of Maximum Entropy Spectral Analysis," IEEE Trans. on Info. Thy., IT-17, 1971, pp. 493-494.

Van Trees, H.L., Detection, Estimation and Modulation Theory, Vol-L, John Wiley and Sons, Inc., New York, 1968.

Velichkin, A.I., "Correlation Function and Spectral Density of a Quantized Process," Telecomm. and Radio Engng, Part 2, Vol 17.,1962, pp. 70-77.

Welch, P.D., "The Use of Fast Fourier Transform for the Estimation of Power Spectra," IEEE Trans. Audio Electracoust. AV 15, 1970, pp. 70-73.

Widrow, B., "A Study of Rough Amplitude Quantization by Means of Nyguist Sampling Theory, " IRE Trans. Circuit Theory, Vol. CT-3, 1956, pp. 266-276.

Widrow, B., "Statistical Analysis of Amplitude-Quantized Sampled Data Systems," AIEE Trans. (Appl. Ind.), Vol 79.,1961, pp. 555-568.

Wong, E., Stochastic Processes in Information and Dynamical Systems, McGraw-Hill Book Co., New York, 1971.

DISTRIBUTION LIST

	No. of Copies
Office of Naval Research Department of the Navy Arlington, Virginia 22217 Attn: Code 222	2
102-OS	1
480	1
Director Naval Research Laboratory Technical Information Division Washington, D.C. 20375	6
Director Office of Naval Research Branch Office 1030 East Green Street Pasadena, California 91106	1
Office of Naval Research San Francisco Area Office 760 Market Street - Room 447 San Francisco, California 94102	1
Director Office of Naval Research Branch Office 495 Summer Street Boston, Massachusetts 02210	1
Office of Naval Research New York Area Office 207 West 24th Street New York, New York 10011	1
Director Office of Naval Research Branch Office 536 South Clark Street Chicago, Illinois 60605	1
Commander Naval Surface Weapons Center Acoustics Division White Oak Silver Springs, Maryland 20910 Attn: Dr. Zaka Slawsky	1
Commanding Officer Office of Naval Research Branch Office Box 39 FPO New York 09510	1

Officer in Charge
Naval Ship Research & Dev. Center
Annapolis Laboratory
Annapolis, Maryland 21402 — 1

Commander
Naval Sea Systems Command
Department of the Navy
Washington, D.C. 20362
Attn: SEA 037 — 1
Carey Smith, 06HI — 1
David F. Bolka, 06H2

Commanding Officer
Fleet Numerical Weather Central
Monterey, California 93940 — 1

Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314 — 12

Director of Navy Laboratories
Chief of Naval Material
2211 Jefferson Davis Highway
Crystal Plaza #5
Arlington, Virginia 20360
Attn: Dr. James Probus NAVMAT 03L — 1

Commander
Naval Electronic Systems Command
2511 Jefferson Davis Highway
Arlington, Virginia 20360
Attn: CDR A.R. Miller NAVELEX 320 — 1

Commander Naval Ship Research & Development Center
Department of the Navy
Bethesda, Maryland 20084
Attn: Mr. Craig Olson — 1
Unclassified Library — 1

Chief of Naval Operations
Room 4D518, Pentagon
Washington, D.C. 20350
Attn: CAPT A.H. Gilmore — 1

Commander
Naval Ocean Systems Center
Department of the Navy
San Diego, California 92132
Attn: Dr. Dan Andrews — 1
Mr. Henry Aurand — 1
Dr. Dean Hanna — 1

Superintendent
Naval Research Laboratory
Underwater Sound Reference Division
P.O. Box 8337
Orlando, Florida 32806 1

Commanding Officer
Naval Underwater Systems Center
New London Laboratory
New London, Connecticut 06320
Attn: Dr. A. Nuttall 1
Mr. A. Ellinthrope 1
Dr. D. M. Viccione 1

Commander
Naval Air Development Center
Department of the Navy
Warminster, Pennsylvania 18974
Attn: Unclassified Library 1

Superintendent
Naval Postgraduate School
Monterey, California 93940
Attn: Unclassified Library 1

Commanding Officer
Naval Coastal Systems Laboratory
Panama City, Florida 32401
Attn: Unclassified Library 1

Commanding Officer
Naval Underwater Systems Center
Newport Laboratory
Newport, Rhode Island 02840
Attn: Unclassified Library 1

Superintendent
U.S. Naval Academy
Annapolis, Maryland 21402
Attn: Library 1

Commanding Officer
Naval Intelligence Support Center
4310 Suitland Road
Suitland, Maryland 20390
Attn: Dr. Johann Martinek 1
Mr. E. Bissett 1

Office of the Assistant Secretary of the Navy
for Research, Engineering & Systems
Room 4E732, Pentagon
Washington, D.C. 20350
Attn: Mr. Gerald Cann 1

Commander
Naval Sea Systems Command
Washington, D.C. 20362
Attn: Unclassified Library, SEA 03E 1

Dr. Melvin J. Jacobson
Rensselaer Polytechnic Institute
Troy, New York 12181 1

Dr. Alan Winder
MSB Systems, Inc.
25 Sylvan Road South
West Point, Connecticut 06880 1

Dr. Charles Stutt
General Electric Company
P.O. Box 1088
Schenectady, New York 12301 1

Dr. T.G. Birdsall
Cooley Electronics Laboratory
University of Michigan
Ann Arbor, Michigan 48109 1

Dr. Harry DeFerrari
University of Miami
Rosenstiel School of Marine &
Atmospheric Sciences
4600 Rickenbacker Causeway
Miami, Florida 33149 1

Mr. Robert Cunningham
Bendix Electronics Center
15825 Roxford Street
Sylmar, California 91342 1

Dr. Stephen Wolff
John Hopkins University
Baltimore, Maryland 21218 1

Dr. M. A. Basin
S.D.P., Inc.
15250 Ventura Boulevard
Suite 518
Sherman Oaks, California 91403 1

Dr. Walter Duing
University of Miami
Rosenstiel School of Marine &
Atmospheric Sciences
4600 Rickenbacker Causeway
Miami, Florida 33149 1

Dr. David Middleton
127 East 91st Street
New York, New York 10028 — 1

Dr. Donald W. Tufts
University of Rhode Island
Kingston, Rhode Island 02881 — 1

Dr. Loren Nolte
Duke University
Department of Electrical Engineering
Durham, North Carolina 27706 — 1

Mr. S. W. Autrey
Hughes Aircraft Company
P.O. Box 3310
Fullerton, California 92634 — 1

Dr. Thomas W. Ellis
Texas Instruments, Inc.
13500 North Central Expressway
Dallas, Texas 75231 — 1

Dr. Terry Ewart
Applied Physics Laboratory
University of Washington
1013 Northeast Fortieth Street
Seattle, Washington 98195 — 1

Institute for Acoustical Research
Miami Division of the Palisades
Geophysical Institute
615 S. W. 2nd Avenue
Miami, Florida 33130 — 2
Attn: Mr. M. Kronengold
Dr. J. Clark
Dr. W. Jobst
Dr. S. Adams

Mr. Carl Hartdegen
Palisades Geophysical Institute
Sofar Station
FPO New York 09560 — 1

Mr. Charles Loda
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202 — 1

Special Assistant for ASW
Office of the Assistant Secretary of the Navy
for Research, Engineering & Systems
Washington, D.C. 20350 — 1
Attn: Dr. D. Hyde

Mr. Beaumont Buck
Polar Research Laboratory
123 Santa Barbara Avenue
Santa Barbara, California 93101 — 1

Dr. M. Weinstein
Underwater Systems, Inc.
8121 Georgia Avenue
Silver Spring, Maryland 20910 — 1

Mr. Thomas G. Kincaid
General Electric Company
P.O. Box 1088
Schenectady, New York 12301 — 1

Applied Research Laboratories
University of Texas at Austin
P.O. Box 8029
10000 FM Road 1325
Austin, Texas 78712 — 2
Attn: Dr. Lloyd Hampton
Dr. Charles Wood

Woods Hole Oceanographic Institution
Woods Hole, Massachusetts 02543 — 1
Attn: Dr. Paul McElroy
Dr. R. Porter
Dr. R. Spindel

Mr. John Bouyoucos
Hydroacoustics, Inc.
321 Northland Avenue
P.O. Box 3813
Rochester, New York 14610 — 1

Ronald J. Carpinella (Major)
Wright-Patterson AFB
Dayton, Ohio 45433 — 1

Atlantic Oceanographic &
Meteorological Laboratories
15 Rickenbacker Causeway
Miami, Florida 33149 — 1
Attn: Dr. John Proni

Dr. C. N. K. Mooers
University of Delaware
Newark, Delaware 19711 — 1

Systems Control, Inc
260 Sheridan Ave
Palo Alto, California 94306 — 1
Attn: Mr. Robert Baron

Westinghouse Electric Corporation
Advanced Development Program
Marketing Department - MS 227
P.O. Box 746
Baltimore, Maryland 21203
Attn: F. J. Frissyn 1

Oak Ridge National Laboratory
Union Carbide Corporation
Nuclear Division
P.O. Box X
Oak Ridge, Tennessee 37830 1

Professor Neil Bershad
University of California, Irvine
School of Engineering
Irvine, California 92664 1

Stuart Schwartz
Princeton University
Department of Electrical Engineering
Brackett Hall, Engineering Quad.
Princeton, New Jersey 08540 1

Ray Veenkant
Texas Instruments Inc.
North Central Expressway
MS 208
Dallas, Texas 75222 1

Dr. C. V. Kimball
University of Miami
Rosenstiel School of Marine &
Atmospheric Sciences
10 Rickenbacker Causeway
Miami, Florida 33149 1

Cooley Electronics Laboratory
University of Michigan
Ann Arbor, Michigan 48109 25