THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MINIMUM VARIANCE MEAN ESTIMATION FOR STOCHASTIC PROCESSES WITH NORMAL POWER SPECTRAL DENSITY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

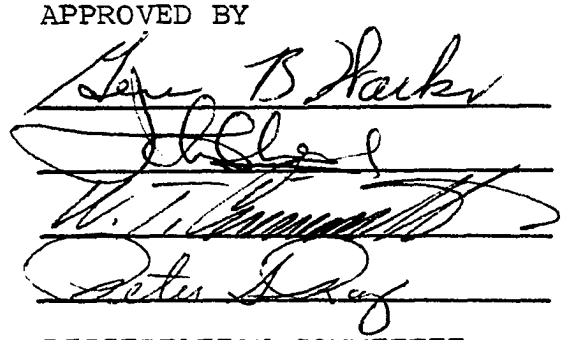
ARTHUR WAYNE MEEKS, JR.

Norman, Oklahoma

1979

MINIMUM VARIANCE MEAN ESTIMATION FOR STOCHASTIC PROCESSES WITH NORMAL POWER SPECTRAL DENSITY

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

Sincere appreciation is expressed:

To Dr. Gene B. Walker, major professor and dissertation committee chairman, for his valuable guidance and encouragement during the research and throughout my graduate education.

To Dr. J. Y. Cheung, Dr. W. T. Cronenwett, and Dr. P. S. Ray for serving on the committee and offering helpful advice and recommendations.

To Dr. W. L. Kuriger for reading and critiquing the dissertation manuscript.

To my employer, Fleet Analysis Center, for financial support through an educational fellowship during this period of graduate study.

To the School of Electrical Engineering at the University of Oklahoma for providing computer time for some of the research.

And to my wife, Donna, for several years of unwavering kindness and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

MINIMUM VARIANCE MEAN ESTIMATION FOR STOCHASTIC PROCESSES WITH NORMAL POWER SPECTRAL DENSITY

CHAPTER I

INTRODUCTION

The research documented in this dissertation concerns optimum estimation of the mean value for a class of stochastic processes. The class includes real-valued, wide-sense stationary processes with normally distributed power spectra. It will be assumed throughout that the processes discussed are ergodic for the mean value and autocovariance function so that statistical and temporal averages can be interchanged.

This research is motivated by its application to certain problems from radar theory. The stochastic processes described are frequently used to model signals caused by reflections from atmospheric precipitation or the surface of the sea. In the case of atmospheric precipitation, drop size spectra and precipitation rates can be measured from the intensity of the reflections. Estimation of the signal intensity can be mathematically equivalent to the previously stated estimation problem. Improvements in the performance

of meteorological radars, either in the accuracy of the measurements or in their time or spatial resolution, often require the development of estimation techniques which extract more information from the available observations. Optimum estimation techniques, such as studied herein, are usually desired when they can be found.

The primary emphasis in this dissertation is on discrete-time estimation; here, the observed data are a finite number of equally-spaced sample values from the stochastic process. The form of the estimator is restricted to linear, unbiased combinations of the observed data, and the minimization of estimator variance is the criterion for optimality. Under these assumptions, analytic expressions for the inverse of the covariance matrix, the optimum estimator, and its variance are developed. The variance of the optimum estimator is compared to the variance of a commonly used estimator, the sample mean. The sample mean is obtained by simple averaging of the observed data. The results from a numerical simulation are also presented.

A brief discussion of continuous-time estimation is included. The continuous-time estimator is based on observing the stochastic process for a continuous but finite interval and forming a linear, unbiased functional of the observed data. Minimization of estimator variance is again the criterion for optimality.

CHAPTER II

PROBLEM DEVELOPMENT

Definitions and Notation

The following definitions and notation are introduced to establish a common basis for discussing stochastic processes as they pertain to the stated problem. For the most part, the development follows Papoulis [1], although some changes will be introduced. The notation

$$X(t)$$

will represent a stochastic process which varies with time, denoted by t. The value of the process at some specific time, t_i, is the random variable

$$X(t_i) \text{ or } X_i$$

when no confusion will result. The mean of the stochastic process X(t) is

$$\mu(t) = E\{X(t)\} \tag{2-1}$$

where $E\{\cdot\}$ denotes the expectation operation. The autocorrelation of X(t) is given by

$$R(t_1,t_2) = E\{X(t_1)X(t_2)\}. \qquad (2\text{-}2)$$

Closely related to the autocorrelation is the autocovariance which is defined by

$$C(t_1,t_2) = E\{(X(t_1)-\mu(t_1))(X(t_2)-\mu(t_2))\}. \qquad (2\text{-}3)$$

For wide-sense stationary stochastic processes, the mean does not vary with time so that (2-1) becomes

$$\mu(t) = \mu. \qquad (2\text{-}4)$$

Also, the autocorrelation and autocovariance depend only on the absolute value of the difference

$$\tau = t_2 - t_1.$$

Thus, (2-2) and (2-3) become, respectively,

$$R(t_1,t_2) = R(\tau) = R(-\tau) \qquad (2\text{-}5)$$

and

$$C(t_1,t_2) = C(\tau) = C(-\tau) = R(\tau)-\mu^2. \qquad (2\text{-}6)$$

Henceforth, all stochastic processes mentioned will be wide-sense stationary unless otherwise stated.

At times, it will be useful to have the normalized autocovariance defined by

$$\rho(\tau) = C(\tau)/C(0). \qquad (2\text{-}7)$$

The power spectral density of a process is the

Fourier transform of its autocorrelation, denoted by

$$S(\omega) = \int_{-\infty}^{+\infty} R(\tau)\exp(-j\omega\tau)d\tau, \qquad (2\text{-}8)$$

where $j = \sqrt{-1}$ and ω is angular frequency.

Later discussions will refer to rational power spectral density, meaning that $S(\omega)$ can be expressed as the ratio of two finite and nonzero degree polynomials in ω^2. The degree of the numerator polynomial will be no greater than that of the denominator, and numerator and denominator will contain no common roots.

The variance of a random variable, X_i, is

$$\text{var}(X_i) = E\{(X_i - E\{X_i\})^2\}.$$

When X_i is a sample from $X(t)$ at $t = t_i$,

$$\text{var}(X_i) = C(0).$$

Thus a measure of the variance of the stochastic process can be derived from $C(\tau)$ at $\tau = 0$.

An estimate of the mean μ will be designated by

$$\hat{\mu},$$

and the bias of an estimate is defined by

$$\text{bias} = E\{\hat{\mu}\} - \mu. \qquad (2\text{-}9)$$

An estimate is said to be unbiased if the bias is identically zero. The variance of the estimate is

$$\mathrm{var}(\hat{\mu}) = E\{(\hat{\mu}-E\{\hat{\mu}\})^2\}$$

which can be summed with the square of the bias to obtain

$$\mathrm{var}(\hat{\mu})+\mathrm{bias}^2 = E\{(\hat{\mu}-\mu)^2\}.$$

The right-hand side of this expression is the mean-squared error (mse).

Problem Characterization

Examination of (2-8) shows that, if a stochastic process has nonzero mean, its power spectral density will contain an impulse $\delta(\tau)$, or Dirac delta function, at the origin of the frequency axis. At times this will be a mathematical inconvenience that can be circumvented by considering the original process as the sum of a zero-mean process and a constant. The zero-mean process has the same power spectral density as that of the original process, minus the impulse. This is equivalent to defining the power spectral density as the Fourier transform of the autocovariance.

The normal power spectral density for processes to be discussed will be taken to be

$$S(\omega) = C(0)(\sqrt{\pi}/\alpha)\exp(-\omega^2/4\alpha^2), \qquad (2\text{-}11)$$

where $C(0)$ and α are positive constants. The corresponding autocovariance is the inverse Fourier transform of (2-11), and can be calculated as done in [2]. The result of the calculation is

$$C(\tau) = C(0)\exp(-\alpha^2\tau^2). \qquad (2\text{-}12)$$

The normalized autocovariance is

$$\rho(\tau) = \exp(-\alpha^2\tau^2). \qquad (2\text{-}13)$$

Hence the autocovariance also has the shape of a normal density.

It is appropriate to note that a stochastic process with normal power spectral density is not, in general, the better known normal, or Gaussian, stochastic process. The latter process is characterized by the requirement that the N random variables

$$X(t_1),\ X(t_2),\ \ldots,\ X(t_N)$$

have jointly normal, or Gaussian, probability distributions for all N and all

$$t_1,\ t_2,\ \ldots,\ t_N.$$

In contrast, the requirement that a process have normal power spectral density leaves the form of the probability distributions open.

The power spectral density can be expressed using the Maclaurin series as

$$S(\omega) = C(0)(\sqrt{\pi}/\alpha)\Sigma_{k=0}^{\infty}\ (-\omega^2/4\alpha^2)^k/k! \qquad (2\text{-}14)$$

which, according to the previous definition, is irrational.

As shown later, knowledge of the autocovariance is

sufficient for the formulation of the linear, minimum-variance, unbiased estimator. This can be advantageous as experimental determination of the power spectral density and hence, autocovariance, is a more tractable engineering effort than determination of probability distributions. Also advantageous is that, for unbiased estimation, minimization of variance is tantamount to minimization of mse. This criterion often results in more tractable mathematics than other measures of estimator performance.

Previous Work

This section discusses the major historical developments of estimation theory as related to the stated problem. No attempt is made to be inclusive due to the great number of contributions to the theory. There are frequent references to rational power spectral densities for the following reason: the output of a linear, time-invariant physical system excited by white noise is a stochastic process with rational power spectral density. White noise refers to a process with power spectral density of the form

$$S(\omega) = k$$

where k is a positive constant.

The first treatments of estimation of stochastic processes were given independently by Wiener [4] and by Kolmogorov [5]. This theory considers the estimation of a process, called the signal, when observations are corrupted

by the addition of a second process, the noise. The interval of observation extends to the infinite past, and both signal and noise must be stationary. No deterministic components are permitted. The optimum estimator is found by solving a linear integral equation of particular form, called the Wiener-Hopf equation. Wiener's method of solution, spectral factorization, necessitates the restriction to processes with rational power spectral densities. A later derivation of the Wiener theory by Bode and Shannon [6] is based on electric circuit rather than probabilistic concepts but arrives at the same results as Wiener's.

The work of Zadeh and Ragazzini [7] extends the theory to include signals composed of a stochastic process summed with a polynomial of known maximum degree and unknown coefficients. The extension to finite observation time is also covered. However, the restriction to stochastic processes with rational power spectral densities is retained to permit solution of a modified Wiener-Hopf equation.

Lees [8], Johnson [9], and Blum [10] all solve problems similar to that of Zadeh and Ragazzini but for the discrete time estimator. Consequently, the optimum estimators are found as solutions to systems of linear equations rather than integral equations; therefore, the restriction to rational power spectral densities is not immediately applicable. However, the solutions are then in the form of inverses of covariance matrices for which analytic expres-

sions cannot generally be found.

The work of Kalman [11] and Kalman and Bucy [12] extends the basic Wiener theory to multivariate problems in discrete or continuous time. The signal is modeled as the output of a linear, possibly time-varying, physical system excited by white noise. The measurement noise is assumed to be white. The latter restriction is reduced by Bryson and Johansen [13] and Bucy [14] to requiring that the measurement noise be modeled as the signal.

For summarizing the relationship of the preceding to the problem at hand, one of two characterizations applies. Either the estimation method excludes noise models of the desired form, or the solution for the estimator is given as the inverse of a covariance matrix.

There have been several notable efforts at expressing the inverse of a covariance matrix explicitly. Siddiqui [15] and Whittle [16] have developed methods based on the process having an autoregressive representation, a requirement more restrictive than having rational power spectral density. Janos [17] and Skidmore [18] have generalized the method to rational power spectral densities.

These four methods all rely on solving a system of linear equations of dimension equal to the degree of the denominator of $S(\omega)$. When this dimension is considerably smaller than the dimension of the covariance matrix, a worthwhile computational advantage can result. These

methods do not, however, apply to normal power spectral densities.

Applications

After reviewing the preponderance of literature devoted to stochastic processes with rational power spectral densities, some question might naturally arise as to the significance of irrational power spectral densities. As noted by Whittle [16], " . . . almost invariably processes which involve an infinite number of variates (such as a random field in space) will have an s.d.f. [spectral density function] which is, indeed, transcendental." An example process is the meteorological radar signal, the result of radar energy reflected by precipitation in the atmosphere. This section considers estimating the average power of this signal, which is of practical interest because it is indicative of precipitation rates or drop size spectra.

Both Atlas [19] and Battan [20] contain comprehensive developments of meteorological radar signal characteristics, some of which are summarized here. The probability distributions of the signal amplitudes are jointly normal, so that the signal is a normal stochastic process. This fact is useful for developing the example although not necessary for formulation of the estimator.

The power spectral density of the signal closely fits a normal curve with variance σ_f^2 and mean μ_f. The variance results primarily from air turbulence and wind shear

within the sample volume defined by the antenna beamwidth and transmitted pulse length. Variance is related to the wavelength of the transmitted signal by

$$\sigma_f^2 = 4\sigma_v^2/\lambda^2$$

where λ is wavelength and σ_v^2, the variance of the radial velocity distribution, typically ranges from 1 to 10 meter^2/second^2. The mean, μ_f, is the sum of the transmitted frequency and the doppler shift caused by the mean radial velocity of the precipitation within the sample volume. The mean doppler shift, μ_d, and the mean radial velocity, μ_v, are related by

$$\mu_d = 2\mu_v/\lambda .$$

Typical values of μ_v range from -30 to 30 meter/second, and positive values of μ_v correspond to decreasing radial ranges. For most weather radars, λ lies between 1 and 10 cm, so that the bandwidth of the signal is quite small compared to the transmitting frequency.

In conventional radars the spectral density is linearly translated to a lower, intermediate frequency (IF) for amplification. The mean, μ_f, may then be considered to be the sum of the IF and the doppler shift without loss of generality. Compared to common IFs, about 30 MHz, the bandwidth of the signal is still small.

IF amplification is usually followed by signal

detection and low-pass filtering. The input-output relation is generally taken to be linear, square-law, or logarithmic. Only the square-law and linear responses will be considered here.

Davenport and Root [21] show that the output power spectral density of the square-law detector is the input spectral density convolved with itself when the input process is normal. For the input spectral density described above, the output spectral density consists of normal curves centered about $\pm 2\mu_f$, a normal curve centered about the origin, and a delta function at the origin. Low-pass filtering after detection removes the components at $\pm 2\mu_f$ without affecting the other components. What remains is a stochastic process with normal spectral density centered about the origin and nonzero mean as indicated by the delta function. The mean is equal to the average power of the input process, which is the quantity to be estimated.

The linear response case, as developed in [21], is similar to the preceding except that the mean of the output process is proportional to the square root of the input process average power. Consequently, the estimate of the output process mean must be squared and scaled to give an unbiased estimate of the input process average power.

In view of the preceding, we conclude that estimation of the meteorological radar signal average power is equivalent to estimation of the mean of a stochastic process

with normal power spectral density when the radar has either square-law or linear IF response. Two qualifications to this conclusion are necessary, however. The preceding discussion has not considered the effect of radar system noise, which may not be serious when the meteorological signal has relatively much greater power. Second, the estimation method to be developed requires a priori knowledge of σ_f^2, which is not generally available. Methods of estimating σ_f^2 are available; see, for example, Miller and Rochwarger [22].

CHAPTER III

DISCRETE-TIME ESTIMATION

Formulation

In this chapter, the problem of estimating, from a set of discrete samples, the mean of a stochastic process with normal power spectral density is formulated and solved. It is assumed that the number of samples is finite, and that the samples are equally spaced in time. It is also assumed that the normalized autocovariance, given by (2-13), is known.

Although the equations that the linear, minimum-variance, unbiased estimator, μ, must satisfy are known, they are derived here so that they will be readily available in the desired form. Given are the N sample values,

$$X(t),\ X(t+T_s),\ \ldots,\ X(t+(N-1)T_s),$$

of the process X(t) taken at intervals of T_s. The estimator, $\hat{\mu}$, is a linear combination of the samples so that

$$\hat{\mu} = \Sigma_{i=1}^{N} W_i X_i, \qquad (3\text{-}1)$$

where $X_i = X(t+(i-1)T_s)$ and the W_i's are unknown weights.

Since $\hat{\mu}$ is to be unbiased, (2-9) must be identically zero and

$$\mu = E\{\hat{\mu}\} = E\{\Sigma_{i=1}^{N} W_i X_i\} = \Sigma_{i=1}^{N} W_i E\{X_i\} = \mu\Sigma_{i=1}^{N} W_i.$$

Hence, for $\hat{\mu}$ to be unbiased, it must be true that

$$\Sigma_{i=1}^{N} W_i = 1. \tag{3-2}$$

By defining $\underline{X}$ as the column N-vector whose elements are the sample values, X_i, and $\underline{W}$ as the column N-vector whose elements are the weights, W_i, (3-1) can be written as

$$\mu = \underline{W}^T\underline{X} \tag{3-3}$$

where $\underline{W}^T$ denotes the transpose of $\underline{W}$. By defining $\underline{U}$ as the N-vector whose elements, U_i, all satisfy

$$U_i = 1,$$

(3-2) may be written as

$$\underline{W}^T\underline{U} = 1. \tag{3-4}$$

The variance of $\hat{\mu}$ which is to be minimized is given by

$$\text{var}(\hat{\mu}) = E\{(\hat{\mu}-E\{\hat{\mu}\})^2\} = E\{(\underline{W}^T\underline{X}-\underline{W}^T\underline{U}\mu)^2\}$$

$$= \underline{W}^T E\{(\underline{X}-\underline{U}\mu)(\underline{X}-\underline{U}\mu)^T\}\underline{W}. \tag{3-5}$$

The expression $(\underline{X}-\underline{U}\mu)(\underline{X}-\underline{U}\mu)^T$ is a square matrix of dimension

N. An element of row i and column j is $(X_i-\mu)(X_j-\mu)$. From (2-3),

$$E\{(X_i-\mu)(X_j-\mu)\} = C((i-j)T_s).$$

Therefore, the covariance matrix [C] can be defined by

$$[C] = E\{(\underline{X}-\underline{U}\mu)(\underline{X}-\underline{U}\mu)^T\}$$

with the elements of [C] denoted by $C_{ij}=C((i-j)T_s)$. Now, (3-5) may be expressed as

$$\text{var}(\hat{\mu}) = \underline{W}^T[C]\underline{W}. \qquad (3\text{-}6)$$

Rather than minimize the variance of μ, it is more convenient to use the normalized variance defined by

$$\sigma^2 = \text{var}(\hat{\mu})/C(0).$$

By defining the normalized covariance matrix $[\rho]$ with elements $\rho_{ij} = \rho((i-j)T_s)$, σ^2 can be written as

$$\sigma^2 = \underline{W}^T[\rho]\underline{W}. \qquad (3\text{-}7)$$

Since C(0) is a positive constant, minimizing σ^2 is equivalent to minimizing var(μ).

The minimization of σ^2 is straightforward and procedes by first defining the objective function, J, as

$$J(\underline{W},\gamma) = \underline{W}^T[\rho]\underline{W}+2\gamma(1-\underline{W}^T\underline{U})$$

where γ is a Lagrange multiplier. Next are required values

of $\underline{W}$ and γ which satisfy the conditions

$$\frac{\partial J}{\partial \underline{W}} = \underline{0}^T = 2\underline{W}^T[\rho] - 2\gamma \underline{U}^T, \tag{3-8}$$

$$\frac{\partial J}{\partial \gamma} = 0 = 1 - \underline{W}^T \underline{U}. \tag{3-9}$$

The notation $\underline{0}$ is used for the N element zero vector. With some rearrangement, (3-8) and (3-9) become

$$[\rho]\underline{W} = \gamma \underline{U}, \tag{3-10}$$

$$\underline{U}^T \underline{W} = 1. \tag{3-11}$$

Simultaneous solution of these equations is effected by multiplying (3-10) by $\underline{U}^T[\rho]^{-1}$ and substituting (3-11) to get

$$\underline{U}^T[\rho]^{-1}[\rho]\underline{W} = 1 = \gamma \underline{U}^T[\rho]^{-1}\underline{U}$$

which leads to

$$\gamma = (\underline{U}^T[\rho]^{-1}\underline{U})^{-1}. \tag{3-12}$$

Multiplying (3-10) by $[\rho]^{-1}$ and substituting (3-12) yields

$$\underline{W} = (\underline{U}^T[\rho]^{-1}\underline{U})^{-1}[\rho]^{-1}\underline{U}. \tag{3-13}$$

Substituting (3-13) into (3-7) and simplifying results in

$$\sigma^2 = (\underline{U}^T[\rho]^{-1}\underline{U})^{-1}. \tag{3-14}$$

In summary, the optimum estimator is determined by the weight vector, $\underline{W}$. In turn, $\underline{W}$ and the minimum value of

the normalized variance are determined by the inverse of the normalized covariance matrix through (3-13) and (3-14). It will be shown later that, for any finite number of samples from a process with normal power spectral density, the determinant of $[\rho]$ is always positive. This is sufficient to insure that the inverse of $[\rho]$ always exists. Furthermore, the determinants of the normalized covariance matrices for 1, 2, . . . , and N-1 samples are the principal minors of $[\rho]$ for N samples. Since these determinants are all positive, the N eigenvalues of $[\rho]$ are positive, which insures that the solution given by (3-13) is a true minimum. Were the eigenvalues of $[\rho]$ not all positive, the solution of (3-13) might have maximized the variance of $\hat{\mu}$.

Three and Four Sample Estimates

For estimates based on three or four samples, the weight vector, $\underline{W}$, can be found without specifying the form of $\rho(\tau)$ by manually inverting the normalized covariance matrix. The algebra is lengthy and tedious, so only the results will be given. For three samples,

$$\underline{W} = (3-4\rho_{12}+\rho_{13})^{-1} \begin{bmatrix} 1-\rho_{12} \\ 1-2\rho_{12}+\rho_{13} \\ 1-\rho_{12} \end{bmatrix} \qquad (3\text{-}15)$$

and for four samples,

$$\underline{W} = (4-2\rho_{12}-4\rho_{13}+2\rho_{14})^{-1} \begin{bmatrix} 1-\rho_{13} \\ 1-\rho_{12}-\rho_{13}+\rho_{14} \\ 1-\rho_{12}-\rho_{13}+\rho_{14} \\ 1-\rho_{13} \end{bmatrix} \qquad (3\text{-}16)$$

The validity of (3-15) and (3-16) may be verified by substitution into (3-13). The uniqueness of (3-15) and (3-16) follows directly from the uniqueness of $[\rho]^{-1}$.

These analytic expressions for $\underline{W}$ allow the demonstration of a fundamental difference between stochastic processes with rational versus normal power spectral densities. For three sample estimates and processes with rational power spectral density, it can be shown that

$$\lim_{T_s \to 0} \sigma^2 = 1.$$

by the following manipulation. First, substitute (3-15) into (3-7) and simplify to show that

$$\sigma^2 = \underline{W}^T[\rho]\underline{W} = \frac{1-2\rho_{12}^2+\rho_{13}}{3-4\rho_{12}+\rho_{13}}. \qquad (3\text{-}17)$$

For processes with rational power spectral density, Doob [23] shows that

$$\rho(\tau) = \Sigma_{i=1}^{2M} K_i \exp(\beta_i|\tau|) \qquad (3\text{-}18)$$

where the β_i's are the zeros of the denominator of $S(\omega^2)$,

assumed here to be of degree M. The K_i's are constants determined by the partial fraction expansion of $S(\omega^2)$. Since $\rho(0) = 1$, the K_i's must satisfy

$$\Sigma_{i=1}^{2M} K_i = 1.$$

By writing $\exp(\beta_i |\tau|)$ as a MacLaurin series, (3-18) can be rearranged to read

$$\rho(\tau) = 1+K\tau+0(\tau^2)$$

where $0(\tau^2)$ denotes terms of second or higher degree of τ and K is a constant given by

$$K = \Sigma_{i=1}^{2M} K_i\beta_i.$$

Hence,

$$\rho_{12} = 1+KT_s+0(T_s{}^2)$$

$$\rho_{13} = 1+2KT_s+0(T_s{}^2).$$

Substituting these expressions into (3-17) results in

$$\sigma^2 = \frac{-2KT_s+0(T_s{}^2)}{-2KT_s+0(T_s{}^2)} = \frac{1+0(T_s)}{1+0(T_s)}$$

Since all terms in $0(T_s)$ contain T_s to some positive power,

$$\lim_{T_s \to 0} 0(T_s) = 0,$$

and, finally,

$$\lim_{T_s \to 0} \sigma^2 = 1.$$

For processes with normal power spectral density, the corresponding limit is not 1 but 2/3. This can be shown by first defining

$$A = \exp(-\alpha^2 T_s^2). \tag{3-19}$$

Then, the elements of $[\rho]$ are given by

$$\rho_{ij} = \exp(-(i-j)^2 \alpha^2 T_s^2) = A^{(i-j)^2}. \tag{3-20}$$

Applying (3-20) to (3-17) yields

$$\sigma^2 = \frac{1-2A^2+A^4}{3-4A+A^4}. \tag{3-21}$$

The limit can not be taken yet as

$$\lim_{T_s \to 0} \sigma^2 = \frac{0}{0}.$$

Application of l'Hopital's rule will eliminate the indeterminant form; alternately, observe that (3-21) can be factored into

$$\sigma^2 = \frac{(1-A)^2}{(1-A)^2} \frac{(1+A)^2}{(3+2A+A^2)} = \frac{1+2A+A^2}{3+2A+A^2}. \tag{3-22}$$

The desired limit can now be calculated as

$$\lim_{T_s \to 0} \sigma^2 = \frac{2}{3} . \qquad (3\text{-}23)$$

A similar pair of calculations for four sample estimates results in the same numerical values for the two respective limits. The result of (3-23) may at first suggest the possibility of obtaining something for nothing by combining three samples so closely spaced in time that they have the "same" value and yet achieving a reduction in the variance of the estimate. Such is definitely not the case. Examination of the behavior of $\underline{W}$ for vanishingly small values of T_s shows that

$$\lim_{T_s \to 0} \underline{W} = \begin{bmatrix} +\infty \\ -\infty \\ +\infty \end{bmatrix} .$$

Consequently, the possibility suggested is physically unrealizable.

Inverse of the Covariance Matrix

Manual inversion of the covariance matrix, as used in the previous section, becomes prohibitively complex for more than four samples. Numerical inversion is possible by several known algorithms; however, for stochastic processes with normal power spectral density and large numbers of

samples, these algorithms become increasingly inaccurate due to round-off errors and limited dynamic range of digital computers. Also, numerical inversion does not lend itself to functional analysis of the inverse. For these reasons, an algorithm for generating a closed-form expression for the inverse seems attractive. The remainder of this section is devoted to deriving such an algorithm.

Because the covariance matrix and normalized covariance matrix are related by

$$[C] = C(0)[\rho] ,$$

their inverses are related by

$$[\rho]^{-1} = C(0)[C]^{-1} .$$

Therefore, it will suffice to invert either $[C]$ or $[\rho]$. To indicate that $[\cdot]$ is an $N \times N$ matrix, the notation

$$_N[\cdot]$$

will be adopted. For simplicity, the conventions of (3-19) and (3-20) will be followed:

$$A = \exp(-\alpha^2 T_s^2) ,$$

$$\rho_{ij} = A^{(i-j)^2} .$$

The normalized covariance matrix can be expressed as

$$ {}_N[\rho] = \begin{bmatrix} A^0 & A^1 & A^4 & \cdots & A^{(N-1)^2} \\ A^1 & A^0 & A^1 & \cdots & A^{(N-2)^2} \\ A^4 & A^1 & A^0 & \cdots & A^{(N-3)^2} \\ \vdots & \vdots & \vdots & & \vdots \\ A^{(N-1)^2} & A^{(N-2)^2} & A^{(N-3)^2} & \cdots & A^0 \end{bmatrix} \qquad (3\text{-}24) $$

The inverse of ${}_N[\rho]$ will be given in terms of the polynomial array, ${}_N[M]$. Each element of ${}_N[M]$ is a polynomial in B, where B is defined by

$$ B = A^2 . \qquad (3\text{-}25) $$

For N = 1, the array has one element and is given by

$$ {}_1[M] = [1] . \qquad (3\text{-}26) $$

For N > 1, the elements of ${}_N[M]$, denoted by ${}_N M_{ij}$, are found from the following recursion:

$$ \begin{aligned} {}_N M_{ij} = {} & (1-\delta_{i1})(1-\delta_{j1})\ {}_{N-1}M_{i-1,j-1} \\ & + (1-\delta_{iN})(1-\delta_{jN})B^{i-1}\ {}_{N-1}M_{ij} \\ & + (1-\delta_{i1})(1-\delta_{jN})B^{j-1+u(j-i)}\ {}_{N-1}M_{i-1,j} \\ & + (1-\delta_{iN})(1-\delta_{j1})B^{N-1+i-j+u(i-j)}\ {}_{N-1}M_{i,j-1} \end{aligned} \qquad (3\text{-}27) $$

where δ_{ij} is the Kronecker delta defined by

$$\delta_{ij} = 1 \text{ for } i = j$$

$$\delta_{ij} = 0 \text{ for } i \neq j ,$$

and u(n) is the unit step function defined by

$$u(n) = 1 \text{ for } n \geq 0$$

$$u(n) = 0 \text{ for } n < 0 .$$

The inverse of $[\rho]$ can now be stated in the form of a theorem using the notation

$$[G] = [\rho]^{-1} .$$

Theorem 1. For ${}_N[\rho]$ of the form given by (3-24), ${}_N[M]$ as defined in (3-26) and (3-27), and $N \geq 2$, the elements of ${}_N[G]$, the inverse of ${}_N[\rho]$, are defined by

$${}_N G_{ij} = (-A)^{|i-j|} \, {}_N M_{ij} \, \pi_{n=1}^{N-1} (1-B^n)^{-1} . \qquad (3\text{-}28)$$

The proofs of both Theorem 1 and the following theorem for the determinant of ${}_N[\rho]$ are deferred to Appendix A due to their lengths.

Theorem 2. For ${}_N[\rho]$ of the form given by (3-24) and $N \geq 2$, the determinant of ${}_N[\rho]$ is defined by

$$\det({}_N[\rho]) = \pi_{n=1}^{N-1} (1-B^n)^{N-n} . \qquad (3\text{-}29)$$

As noted earlier, when α and T_s are positive, $\rho(\alpha T_s) = A$ is always less than 1; therefore, B is also less than 1.

Examination of (3-29) shows that if B is not 1 the determinant of ${}_N[\rho]$ is never zero. Thus the existence of the inverse of ${}_N[\rho]$ is guaranteed.

Computation of the inverse of ${}_N[\rho]$, $\underline{W}$, and σ^2 using the above formulas is straightforward but becomes too complex for hand calculation for large values of N. When N = 6, for example, the expressions for $\underline{W}$ and σ^2 contain polynomials in A of degree 30. This point is further illustrated in Appendix C, which lists all of the polynomials, ${}_N M_{ij}$, for values of N from 2 to 6.

It seems quite feasible to program a digital computer to perform these computations, and, since they involve only integer values, numerical accuracy should not be a problem. Because the results of these computations are general expressions, the computations need be performed only once for each value of N. After the general expressions are determined, they could be truncated, retaining only as many terms as required for a specified degree of accuracy. In this way, accuracy of the solution could be controlled, rather than dictated by algorithmic or computer hardware limitations. However, this aspect of the problem will not be pursued further as part of this research.

Comparison of Estimator Variances

Having developed the defining equations for the minimum-variance, unbiased estimator and methods of solution, examination of the performance of this estimator is

now of interest. This is done by computing the normalized variance, σ^2, and comparing it with the normalized variance of a more common estimator, the sample mean. For N samples, the sample mean, $\bar{\mu}$, is defined by

$$\bar{\mu} = N^{-1} \Sigma_{i=1}^{N} X_i$$

or, in terms of previous notation

$$\bar{\mu} = N^{-1} \underline{U}^T \underline{X} . \tag{3-30}$$

The normalized variance, s^2, of the sample mean is derived by Costas [24] as

$$s^2 = N^{-1} + N^{-2} \Sigma_{i=1}^{N-1} 2(N-i) \rho(iT_s) ,$$

which can be equivalently expressed as

$$s^2 = N^{-2} \underline{U}^T [\rho] \underline{U} . \tag{3-31}$$

The normalized variances, σ^2 and s^2, were computed from (3-14) and (3-31), respectively, for a range of values of αT_s. Figures 1 through 5 display the results graphically for N = 3, 4, 8, 16, and 32, respectively. The curves on the graphs were drawn by connecting the data points with straight lines. All calculations were done using the IBM 370/158 computer at the University of Oklahoma.

In each figure, both σ^2 and s^2 asymtotically approach N^{-1} for large values of αT_s. For values of αT_s

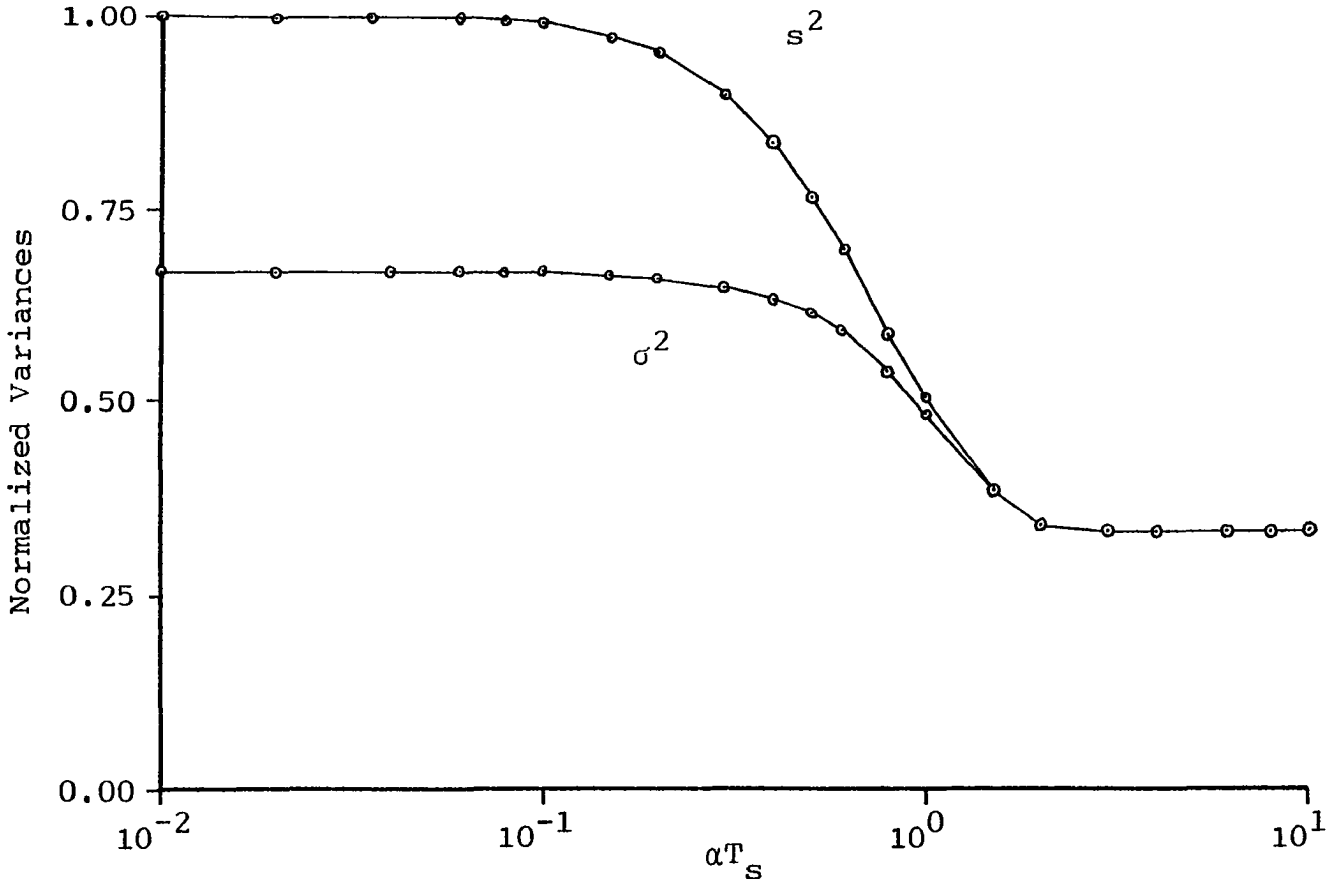


FIGURE 1. Normalized Variances, σ^2 and s^2, versus αT_s for 3 Samples

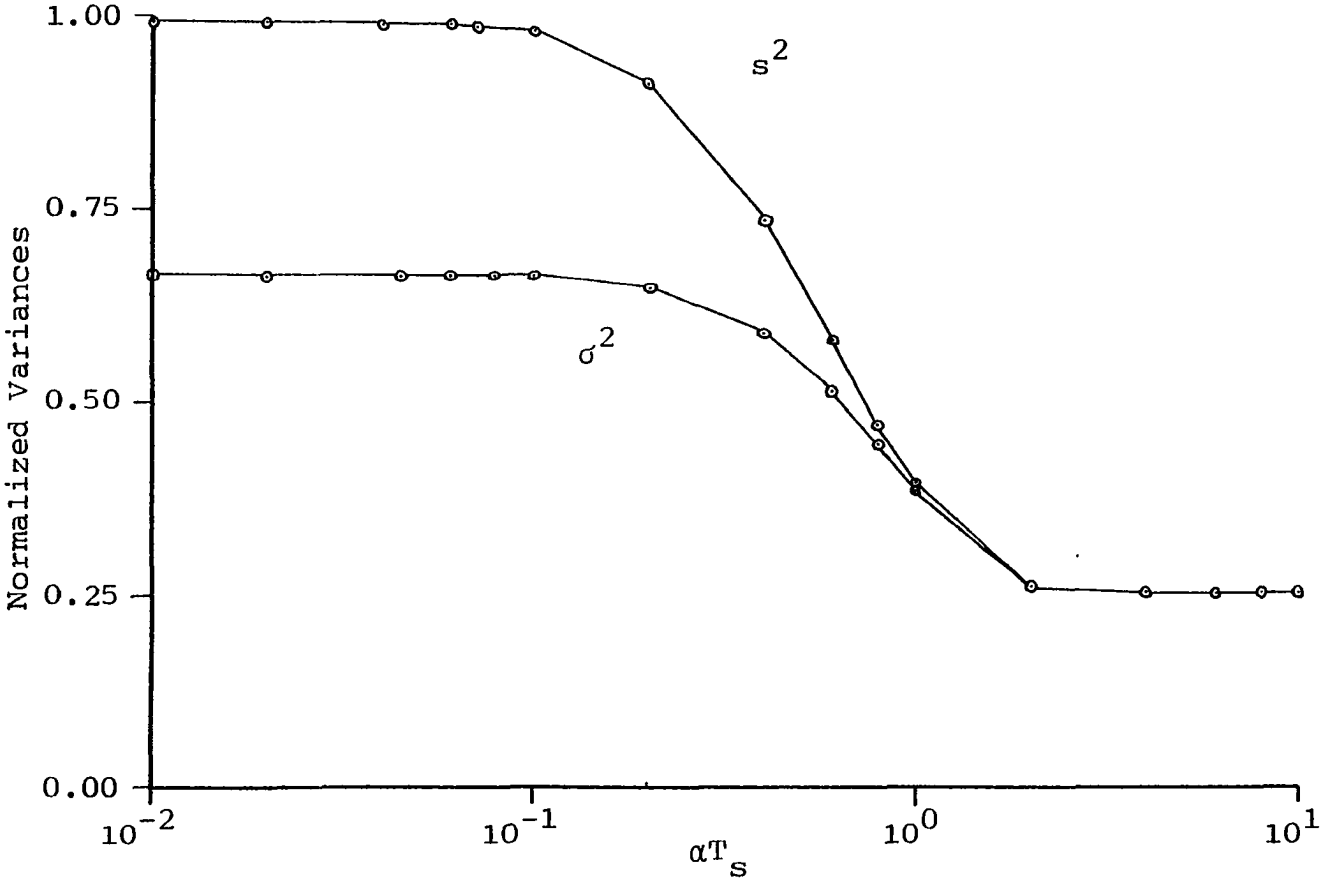


FIGURE 2. Normalized Variances, σ^2 and s^2, versus αT_s for 4 Samples

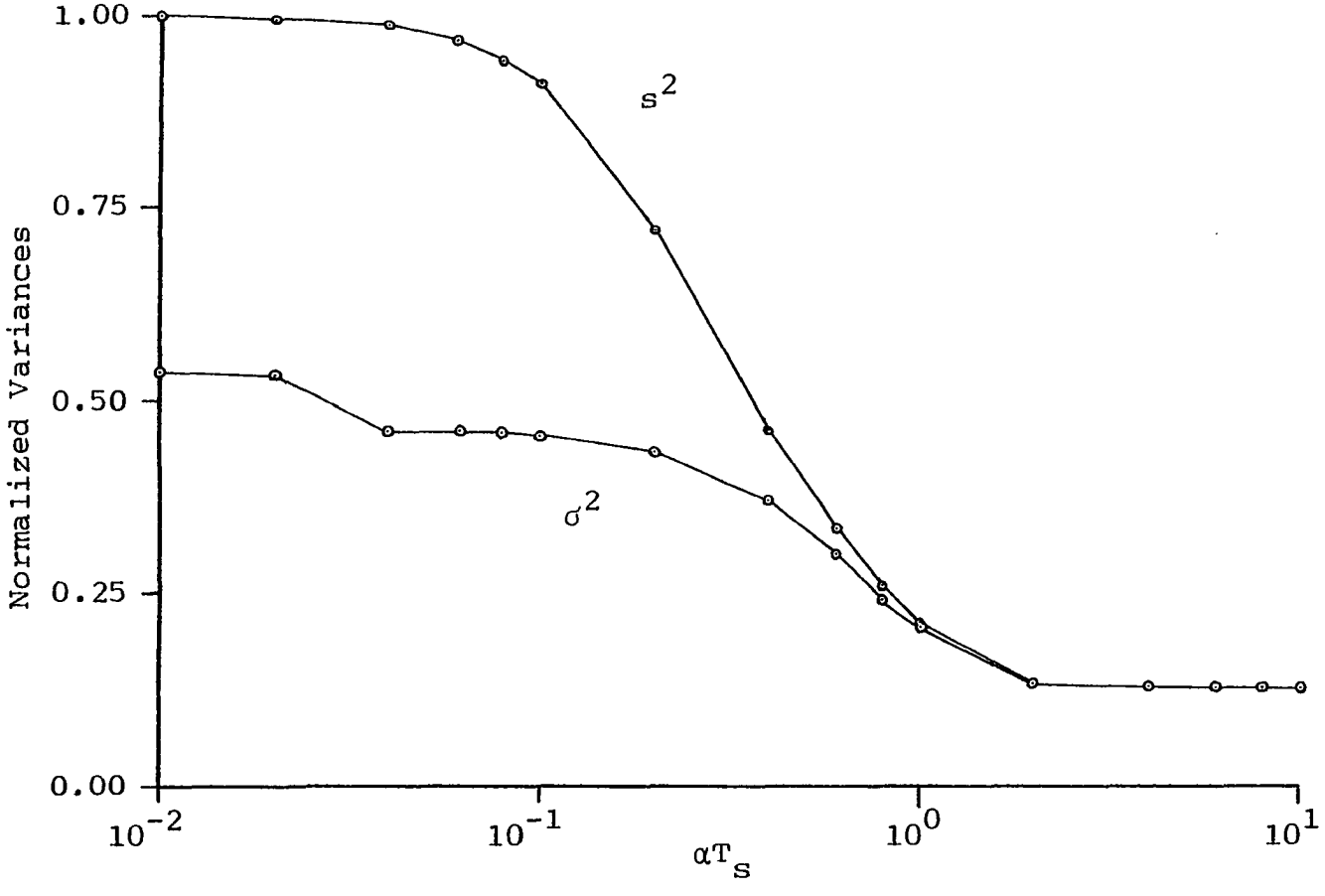


FIGURE 3. Normalized Variances, σ^2 and s^2, versus αT_s for 8 Samples

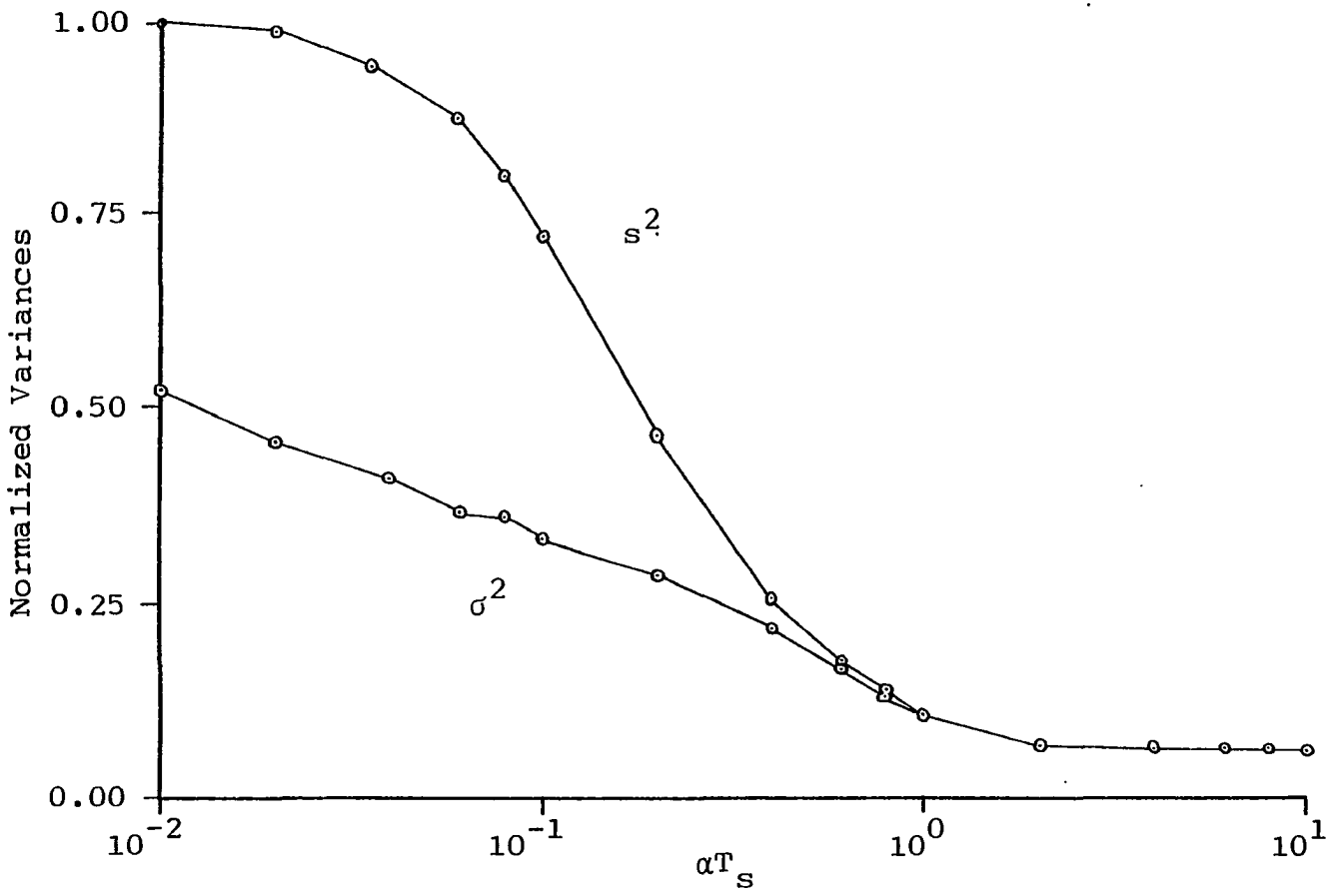


FIGURE 4. Normalized Variances, σ^2 and s^2, versus αT_s for 16 Samples

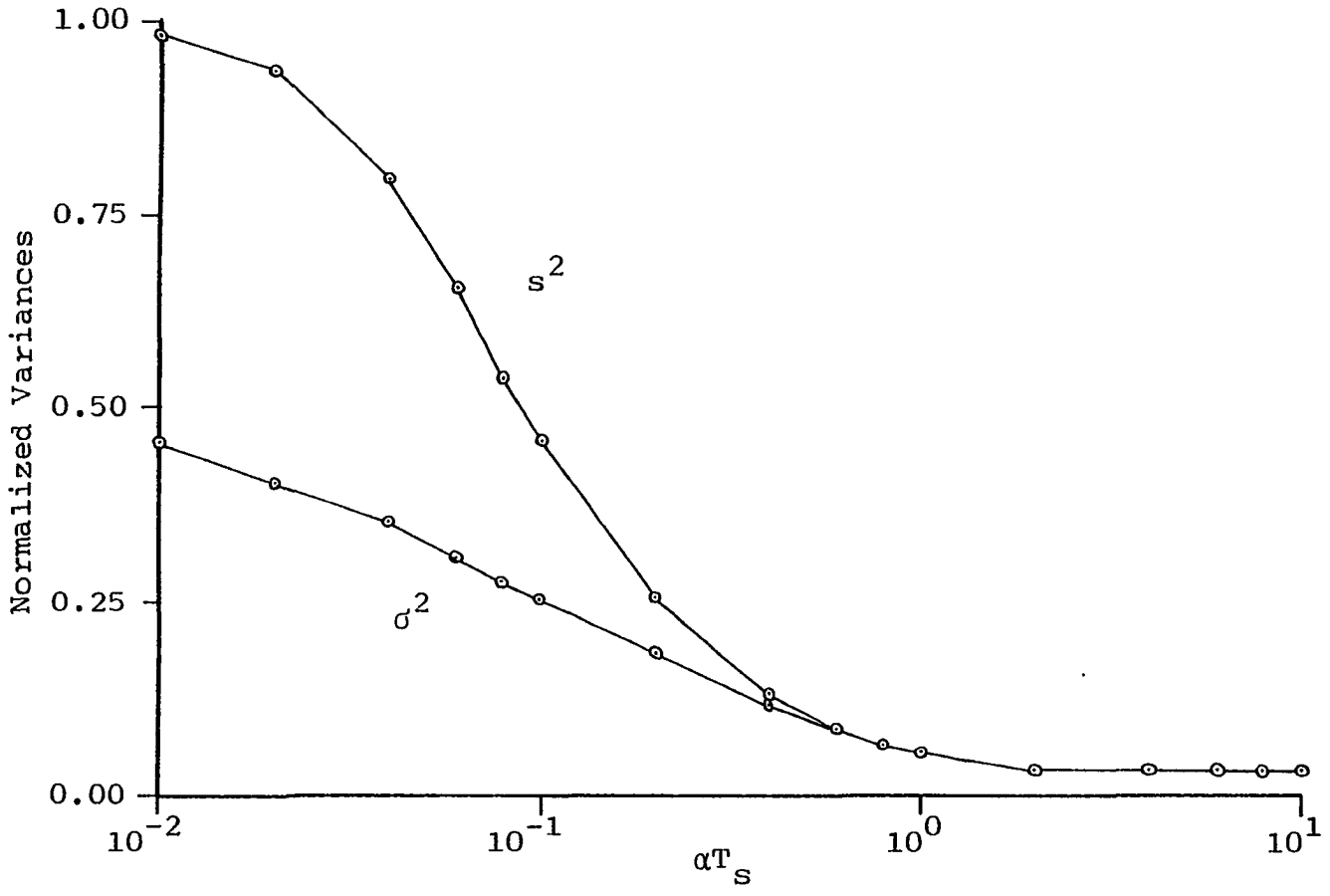


FIGURE 5. Normalized Variances, σ^2 and s^2, versus αT_s for 32 Samples

approaching zero, s^2 approaches 1. In Figure 1 for N = 3 and Figure 2 for N = 4, σ^2 approaches 2/3 as αT_s goes to zero. This behavior is consistent with analytic results given earlier. For N = 8, 16, and 32, σ^2 approaches a different limit as αT_s goes to zero. The limit appears to be equal to 8/15 although this has not been verified analytically. The ratio of optimum estimator variance to the sample mean variance is never greater than 1 (otherwise, the optimum estimator would not have minimum variance) and may be as small as about 1/2.

Numerical Simulation

The optimum estimator was applied to a simulated stochastic process to verify its predicted performance. The simulated stochastic process was generated using the University of Oklahoma IBM 370/158 computer and programs designed from descriptions by Sirmans and Bumgarner [25]. Their method uses pseudo-random numbers with exponential and uniform probability densities as the squared magnitudes and phases, respectively, for the Fourier coefficients of a sequence of sample values. With appropriate scaling, the squared magnitudes approximate the desired normal power spectral density. After conversion from polar to rectangular form, the Fourier coefficients were transformed by a Fast Fourier Transform algorithm to a set of sample values in the time domain.

After the sample values were generated, they were scaled and summed according to (3-3) to form the optimum estimate. The entire procedure was repeated many times, and the sample variance of the resulting set of estimates was computed. The sample variance was compared to the derived expression for estimator variance for verification of the estimator's performance.

The result of the simulation effort did not verify the performance of the optimum estimator, but instead revealed a sensitivity to inaccuracy in the covariance of the synthesized sample values. To illustrate the magnitude of the sensitivity, consider the effect of adding white noise to each of the sample values, represented by

$$Y_i = X_i + k e_i ,$$

where the e_i's are zero-mean random variables satisfying

$$E\{e_i e_j\} = \delta_{ij} ,$$

$$E\{e_i X_j\} = 0 .$$

The value of the constant, k, determines the relative amount of noise; the signal-to-noise power ratio is k^{-2}.

The covariance matrix of a sequence of Y_i's is defined by

$$[\phi] = [\rho] + k^2 [I]$$

where [I] is the identity matrix. When $[\phi]$ is substituted

for $[\rho]$ in (3-7), the variance of the estimator becomes

$$\sigma^2+k^2\underline{W}^T\underline{W} = \sigma^2+k^2\left|\underline{W}\right|^2 .$$

Because k^2 and $\left|\underline{W}\right|^2$ are both positive, the variance of the estimator is increased by the addition of noise to the samples.

The simulation described above used eight sample estimates and the value of αT_s was 0.1. The values of the components of $\underline{W}$ corresponding to these parameters are given in Table 1. The resulting value of $\left|\underline{W}\right|^2$ is 9.21×10^7. In this instance, a signal-to-noise ratio of 83 db will cause the variance of the optimum estimator to double.

The sensitivity to noise appears to decrease as the sample values become less correlated, or as αT_s increases. However, as αT_s increases, the improvement offered by the optimum estimator over that of the sample mean diminishes.

TABLE 1

Values of $\underline{W}$ for $N = 8$, $\alpha T_s = 0.1$

i	W_i
1	6.74×10^2
2	-3.10×10^3
3	5.28×10^3
4	-2.85×10^3
5	-2.85×10^3
6	5.28×10^3
7	-3.10×10^3
8	6.74×10^2

CHAPTER IV

CONTINUOUS-TIME ESTIMATION

The continuous-time estimator is restricted to be an unbiased, linear functional of a continuous, but finite, observation of the stochastic process. In the form derived here, the minimum-variance estimator is physically unrealizable but does estimate the mean value with zero variance.

The estimator is formulated by

$$\hat{\mu} = \int_0^T W(\tau)X(\tau)d\tau \qquad (4\text{-}1)$$

where $W(\tau)$ is a weight function and $X(\tau)$ is, as before, the stochastic process. The duration of the observation interval is T. The unbiasedness of the estimator is assured by requiring that

$$1 = \int_0^T W(\tau)d\tau \, . \qquad (4\text{-}2)$$

The variance of the estimator, denoted by $\text{var}(\hat{\mu})$, is

$$\text{var}(\hat{\mu}) = \int_0^T \int_0^T W(\tau)W(\nu)C(\tau-\nu)d\tau \, d\nu \, , \qquad (4\text{-}3)$$

where $C(t)$ is the autocovariance of the process.

Zadeh and Raggazini [7] show through classical variational techniques that minimization of (4-3) necessitates satisfying the condition,

$$\int_0^T W(\tau)C(\tau-\nu)d\tau = \gamma \quad \text{for } 0 \leq \nu \leq T \qquad (4\text{-}4)$$

where γ is a constant. The method of solving (4-4) given in [7] is restricted to stochastic processes with rational power spectral densities. Specifically, if

$$S(\omega^2) = (\Sigma_{i=0}^{n} a_i \omega^{2i})^{-1}, \qquad (4\text{-}5)$$

then the solution for $W(\tau)$ is of the form,

$$W(\tau) = k + \Sigma_{i=0}^{n-1} c_i \delta^{(i)}(\tau) + d_i \delta^{(i)}(T-\tau). \qquad (4\text{-}6)$$

Here the c_i's, d_i's, and k are all constants. When (4-6) is substituted into (4-1), the form of the optimum estimator becomes

$$\hat{\mu} = k\int_0^T X(\tau)d\tau + \Sigma_{i=0}^{n-1} c_i X^{(i)}(0) + d_i X^{(i)}(T). \qquad (4\text{-}7)$$

The normal power spectral density, as given by (2-14), may be uniformly approximated by finite series of the form in (4-5) for increasingly greater values of n. In light of (4-7), the optimum estimator should thus require derivatives of the stochastic process of all orders. An estimator using derivatives of all orders which has zero

variance can be constructed using the following theorem from Bartlett [26] or Papoulis [1].

Theorem 3. If $C(\tau)$ has derivatives of any order and

$$C(\tau) = \Sigma_{i=0}^{\infty} C^{(i)}(0)\tau^{i}(i!)^{-1}$$

then,

$$X(t+\tau) = \Sigma_{i=0}^{\infty} X^{(i)}(t)\tau^{i}(i!)^{-1} \qquad (4\text{-}8)$$

for all values of t and τ.

Since the requirement of Theorem 3 on $C(\tau)$ is satisfied by stochastic processes with normal power spectral density, then X(t) can be predicted without error for all values of t from knowledge at only one point in time of the process and its derivatives of all orders.

For stochastic processes which are ergodic in the mean, Papoulis [1] proves that

$$\mu = \lim_{T\to\infty}(2T)^{-1}\int_{-T}^{T} X(\tau)d\tau \, . \qquad (4\text{-}9)$$

By setting t equal to zero in (4-8) and substituting into (4-9), we arrive at the result,

$$\mu = \lim_{T\to\infty}(2T)^{-1}\int_{-T}^{T} \Sigma_{i=0}^{\infty} X^{(i)}(0)\tau^{i}(i!)^{-1}d\tau \, . \qquad (4\text{-}10)$$

The implication of this expression is that the mean can be estimated with variance of zero, at least conceptually.

No attention has been given to the question of convergence for (4-10) due to the physically unrealizability of the expression. However, considering the behavior of the analogous discrete-time estimators for large numbers of samples, convergence seems unlikely.

CHAPTER V

CONCLUSIONS

For discrete-time, minimum-variance estimation of the mean value of stochastic processes, the estimators and their variances depend solely on the inverse of the covariance matrix. One result of this research is a closed-form expression for the inverse of the covariance matrix when the stochastic process has normal power spectral density. The inverse is found from a sequence of recursively-generated polynomial arrays, which were also developed during the research.

The variance of the optimum estimator is shown to decrease relative to that of the sample mean as the covariance between sucessive sample values increases. Depending on the number of samples used in the estimate, the variance of the optimum estimator can be as small as half that of the sample mean. As the covariance between sucessive sample values decreases, the variance of the optimum estimator approaches that of the sample mean.

The indicated reduction in variance for the optimum estimator is attainable only if the power spectral density

is exactly normal. The addition of relatively small amounts of white noise causes a rapid increase in estimator variance unless the covariance between sucessive sample values is small. The inclusion of additive white noise is representative of thermal noise present in physical circuits and of quantization noise generated by analog-to-digital conversion. In simulations using digital computers, truncation errors and finite precision number representations also generate such noise. Since these noise sources are almost always present in practical applications, the viability of the estimation technique in its current state of development is questionable.

The estimator displays this sensitivity because, with the addition of noise, the appropriate covariance matrix is no longer $[\rho]$ as for the noise-free case, but $[\rho] + k^2[I]$, where k^2 is the reciprocal of the signal-to-noise ratio. Consequently, additional research to find a closed-form inverse for $[\rho] + k^2[I]$ would, if successful, be clearly beneficial. This approach would require knowledge of the signal-to-noise ratio, in contrast to the original implicit assumption that there is no noise present (k^2 is equal to zero). In many practical situations, the signal-to-noise ratio can be reliably estimated or measured. The sensitivity of estimator variance to deviations in k from its assumed value appears to be greatest when k is assumed to be zero. Thus, if more realistic values for k could be

used, the sensitivity may be sufficiently reduced to yield an effective estimation method.

The effort required to find an inverse for $[\rho] + k^2[I]$ will probably be at least as great as that required for the inverse of $[\rho]$. The problem appears to be untreated in the literature and could be a worthy dissertation topic. One approach to the problem would be to modify the polynomial arrays used for the inverse of $[\rho]$ to account for the addition of the noise term. However, preliminary studies of this approach did not show great promise.

The closed-form inverse developed for the covariance matrix, $[\rho]$, may find applications in other problems of prediction, filtering, or parameter estimation for processes with normal power spectral density. Applications could include estimator formulation or analysis.

The exploration of the continuous-time estimator does not lead to a practical estimator but does serve to demonstrate the zero variance estimate conceptually attainable.

LITERATURE CITED

1. Papoulis, A. Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill, 1965.

2. Humphreys, D. S. The Analysis, Design, and Synthesis of Electric Filters. Englewood Cliffs: Prentice-Hall, 1970.

3. Mood, A. M.; Graybill, F. A.; and Boes, D. C. Introduction to the Theory of Statistics. New York: McGraw-Hill, 1974.

4. Wiener, N. The Interpolation, Extrapolation, and Smoothing of Stationary Time Series. New York: John Wiley and Sons, 1949.

5. Kolomogorov, A. N. "Interpolation und Extrapolation von Stationären Zufällegen Folgen." Bull. Acad. Sci. (URSS), 1941; Sér. Math. 5, pp. 3-14.

6. Bode, H. W., and Shannon, C. E. "A Simplified Derivation of Linear Least Square Smoothing and Prediction." Proc. of the IRE, April 1950; vol. 38, no. 4, pp. 417-425.

7. Zedah, L. A., and Ragazzini, J. R. "An Extension of Wiener's Theory of Prediction." Journal of Applied Physics, July 1950; vol. 21, no. 7, pp. 645-655.

8. Lees, A. B. "Interpolation and Extrapolation of Sampled Data." IRE Transactions on Information Theory, March 1956; vol. IT-2, no. 1, pp. 12-17.

9. Johnson, K. "Optimum, Linear, Discrete Filtering of

Signals Containing a Nonrandom Component." IRE Transactions on Information Theory, June 1956; vol. IT-2, no. 2, pp. 49-55.

10. Blum, M. "An Extension of the Minimum Mean Square Prediction Theory for Sampled Input Signals." IRE Transactions on Information Theory, Sept. 1956; vol. IT-2, no. 3, pp. 176-184.

11. Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems." Transactions of the ASME, Series D - Journal of Basic Engineering, March 1960; vol. 82, no. 1, pp. 35-45.

12. Kalman, R. E., and Bucy, R. S. "New Results in Linear Filtering and Prediction Theory." Transactions of the ASME, Series D - Journal of Basic Engineering, March 1961; vol. 83, no. 1, pp. 95-108.

13. Bryson, A. E., Jr., and Johansen, D. E. "Linear Filtering for Time-Varying Systems Using Measurements Containing Colored Noise." IEEE Transactions on Automatic Control, Jan. 1965; vol. AC-10, no. 1, pp. 4-10.

14. Bucy, R. S. "Optimal Filtering for Correlated Noise." Journal of Mathematical Analysis and Applications, 1967; vol. 20, pp. 1-8.

15. Siddiqui, M. M. "The Inversion of the Sample Covariance Matrix in a Stationary Autoregressive Process." Annals of Math. Statistics, June 1958; vol. 129, pp. 585-588.

16. Whittle, P. Prediction and Regulation by Linear Least-Squares Methods. London: English Universities Press, 1963.

17. Janos, W. A. "Analytic Inversion of a Class of Covariance Matricies." IRE Transactions on Information

Theory, Sept. 1960; vol. IT-6, no. 4, pp. 477-484.

18. Skidmore, L. J. "Discrete and Continuous Estimation in Correlated Noise with Finite Observation Time." Ph.D. dissertation, California Institute of Technology, 1964.

19. Atlas, D. "Advances in Radar Meteorology." Advances in Geophysics 10. New York: Academic Press, 1964.

20. Battan, L. J. Radar Observation of the Atmosphere. Chicago: The University of Chicago Press, 1973.

21. Davenport, W. B., Jr., and Root, W. L. An Introduction to the Theory of Random Signals and Noise. New York: McGraw-Hill, 1958.

22. Miller, K. S., and Rochwarger, M. M. "A Covariance Approach to Spectral Moment Estimation." IEEE Transactions on Information Theory, Sept. 1972; vol. IT-18, no. 5, pp. 588-596.

23. Doob, J. L. Stochastic Processes. New York: John Wiley and Sons, 1953.

24. Costas, J. P. Periodic Sampling of Stationary Time Series. Technical Report No. 156. Cambridge: Research Laboratory of Electronics, Massachusetts Institute of Technology, [1950].

25. Sirmans, D., and Bumgarner, B. "Numerical Comparison of Five Mean Frequency Estimators." Journal of Applied Meteorology, Sept. 1975; vol. 14, no. 6, pp. 991-1003.

26. Bartlett, M. S. An Introduction to Stochastic Processes with Special Reference to Methods and Applications. Cambridge: University Press, 1966.

APPENDIX A

PROOF OF THEOREM 1

Theorem 1. For ${}_N[\rho]$ of the form given by (3-24), ${}_N[M]$ as defined in (3-26) and (3-27), and $N \geq 2$, the elements of the inverse, ${}_N[G]$, of ${}_N[\rho]$ are defined by

$$ {}_N G_{ij} = (-A)^{|i-j|} {}_N M_{ij} \prod_{n=1}^{N-1} (1-B^n)^{-1} . \qquad (A-1) $$

Proof. The proof of Theorem 1 is by induction. First use the recursion of (3-27) on (3-26) to show that ${}_2[M]$ is given by

$$ {}_2[M] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} . $$

Then, from Theorem 1, ${}_2[G]$ is given by

$$ {}_2[G] = (1-A^2)^{-1} \begin{bmatrix} 1 & -A \\ -A & 1 \end{bmatrix} . $$

That ${}_2[G]$ is the inverse of ${}_2[\rho]$ can be verified by showing that their product is the identity matrix (a necessary and sufficient condition),

$$ {}_2[G]\ {}_2[\rho] = (1-A^2)^{-1}\begin{bmatrix} 1 & -A \\ -A & 1 \end{bmatrix}\begin{bmatrix} 1 & A \\ A & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. $$

Therefore, Theorem 1 is true for N = 2. Next it is necessary to prove that, if Theorem 1 is true for N-1, then it is true for N. We will again use the fact that ${}_N[G]$ is the inverse of ${}_N[\rho]$ if and only if their product is the N x N identity matrix. This is equivalent to requiring that

$$ \Sigma_{j=1}^{N}\ {}_N G_{ij}\ {}_N\rho_{jk} = \delta_{ik} \qquad \text{(A-2)} $$

for all values of i and k from 1 to N. If Theorem 1 is assumed to be true for N-1, then

$$ \Sigma_{j=1}^{N-1}\ {}_{N-1}G_{ij}\ {}_{N-1}\rho_{jk} = \delta_{ik} \qquad \text{(A-3)} $$

can be taken as an identity for all values of i and k from 1 to N-1. It remains to be shown that if (A-3) is true, then (A-2) is true. Substitution of (3-20) and (3-27) into (A-2) results in

$$ \begin{aligned} \delta_{ik} = {} & \pi_{n-1}^{N-1}\ (1-A^{2n})^{-1}\ \Sigma_{j=1}^{N}\ (-A)^{|i-j|} A^{(k-j)^2} \\ & \cdot\{(1-\delta_{i1})(1-\delta_{j1})\ {}_{N-1}M_{i-1,j-1} \\ & +(1-\delta_{iN})(1-\delta_{jN})\ A^{2(i-1)}\ {}_{N-1}M_{ij} \\ & +(1-\delta_{i1})(1-\delta_{jN})\ A^{2(j-1+u(j-i))}\ {}_{N-1}M_{i-1,j} \\ & +(1-\delta_{iN})(1-\delta_{j1})\ A^{2(N-1+i-j+u(i-j))}\ {}_{N-1}M_{i,j-1}\} \end{aligned} $$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} \Sigma_{j=2}^{N} (-A)^{|i-j|} A^{(k-j)^2}$$

$$\cdot\{(1-\delta_{i1})\ {}_{N-1}M_{i-1,j-1}$$

$$+(1-\delta_{iN})\ A^{2(N+i-j-1+u(i-j))}\ {}_{N-1}M_{i,j-1}\}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} \Sigma_{j=1}^{N-1} (-A)^{|i-j|} A^{(k-j)^2}$$

$$\cdot\{(1-\delta_{iN})\ A^{2(i-1)}\ {}_{N-1}M_{i,j}$$

$$+(1-\delta_{i1})\ A^{2(j-1+u(j-i))}\ {}_{N-1}M_{i-1,j}\}$$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} \Sigma_{j=1}^{N-1} (-A)^{|i-j-1|} A^{(k-j-1)^2}$$

$$\cdot\{(1-\delta_{i1})\ {}_{N-1}M_{i-1,j}$$

$$(1-\delta_{iN})\ A^{2(N+i-j-2+u(i-j-1))}\ {}_{N-1}M_{i,j}\}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} \Sigma_{j=1}^{N-1} (-A)^{|i-j|} A^{(k-j)^2}$$

$$\cdot\{(1-\delta_{iN})\ A^{2(i-1)}\ {}_{N-1}M_{i,j}$$

$$(1-\delta_{i1})\ A^{2(j-1+u(j-i))}\ {}_{N-1}M_{i-1,j}\}$$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{i1}) \cdot$$

$$\{ \Sigma_{j=1}^{i-1} (-A)^{i-j-1} A^{(k-j-1)^2} {}_{N-1}M_{i-1,j}$$

$$+ (1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i+1} A^{(k-j-1)^2} {}_{N-1}M_{i-1,j} \}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{iN}) \cdot$$

$$\{ (1-\delta_{i1}) \Sigma_{j=1}^{i-1} (-A)^{i-j-1} A^{(k-j-1)^2} A^{2(N+i-j-2+u(i-j-1))} {}_{N-1}M_{i,j}$$

$$+ \Sigma_{j=i}^{N-1} (-A)^{j-i+1} A^{(k-j-1)^2} A^{2(N+i-j-2+u(i-j-1))} {}_{N-1}M_{i,j} \}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{iN}) \cdot$$

$$\{ (1-\delta_{i1}) \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2(i-1)} {}_{N-1}M_{i,j}$$

$$+ \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2(i-1)} {}_{N-1}M_{i,j} \}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{i1}) \cdot$$

$$\{ \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2(j-1+u(j-i))} {}_{N-1}M_{i-1,j}$$

$$+ (1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2(j-1+u(j-i))} {}_{N-1}M_{i-1,j} \}$$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} .$$

$$\{(1-\delta_{i1}) \Sigma_{j=1}^{i-1} (-A)^{i-j-1} A^{(k-j-1)^2} \; {}_{N-1}M_{i-1,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2(j-k+1)} (-1) \; {}_{N-1}M_{i-1,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2(N+i-k-1)} (-1) \; {}_{N-1}M_{i,j}$$

$$+(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2(N+i-k-1)} (-1) \; {}_{N-1}M_{i,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2(i-1)} \; {}_{N-1}M_{i,j}$$

$$+(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2(i-1)} \; {}_{N-1}M_{i,j}$$

$$+(1-\delta_{i1}) \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2(j-1)} \; {}_{N-1}M_{i-1,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2j} \; {}_{N-1}M_{i-1,j}\}$$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} \cdot$$

$$(1-\delta_{i1}) \Sigma_{j=1}^{i-1} (-A)^{i-j-1} A^{(k-j)^2+2j-2k+1} (1-A^{2k-2}) \; {}_{N-1}M_{i-1,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i+1} A^{(k-j)^2+2j-2k+1} (1-A^{2k-2}) \; {}_{N-1}M_{i-1,j}$$

$$+(1-\delta_{i1})(1-\delta_{iN}) \Sigma_{j=1}^{i-1} (-A)^{i-j} A^{(k-j)^2} A^{2i-2} (1-A^{2N-2k}) \; {}_{N-1}M_{i,j}$$

$$+(1-\delta_{iN}) \Sigma_{j=i}^{N-1} (-A)^{j-i} A^{(k-j)^2} A^{2i-2} (1-A^{2N-2k}) \; {}_{N-1}M_{i,j}\}$$

$$= \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{i1}) \cdot$$

$$\Sigma_{j=1}^{N-1} (-A)^{|i-j-1|} A^{(k-j-1)^2} (1-A^{2k-2}) \; {}_{N-1}M_{i-1,j}$$

$$+ \pi_{n=1}^{N-1} (1-A^{2n})^{-1} (1-\delta_{iN}) \cdot$$

$$\Sigma_{j=1}^{N-1} (-A)^{|i-j|} A^{(k-j)^2} (1-A^{2N-2k}) A^{2i-2} \; {}_{N-1}M_{i,j}$$

$$=(1-A^{2N-2})^{-1} (1-\delta_{i1})(1-A^{2k-2}) \Sigma_{j=1}^{N-1} \; {}_{N-1}G_{i-1,j} \; {}_{N-1}\rho_{j,k-1}$$

$$+(1-A^{2N-2})^{-1} (1-\delta_{iN})(A^{2i-2})(1-A^{2N-2k}) \Sigma_{j=1}^{N-1} \; {}_{N-1}G_{i,j} \; {}_{N-1}\rho_{j,k}$$

$$= (1-A^{2N-2})^{-1} (1-A^{2k-2})(1-\delta_{i1}) \delta_{ik}$$

$$+(1-A^{2N-2})^{-1} (A^{2i-2})(1-A^{2N-2k})(1-\delta_{iN})\delta_{ik}$$

$$= (1-A^{2N-2})^{-1} \delta_{ik}$$

$$\cdot\{(1-\delta_{i1})(1-A^{2i-2}) + (1-\delta_{iN})(A^{2i-2})(1-A^{2N-2i})\}$$

$$= \delta_{ik} (1-A^{2N-2})^{-1}$$

$$\cdot\{\delta_{i1}(1-A^{2N-2}) + \delta_{iN}(1-A^{2N-2})$$

$$+(1-\delta_{i1})(1-\delta_{iN})(1-A^{2i-2}+A^{2i-2}-A^{2i-2+2N-2i})\}$$

$$= \delta_{ik} (1-A^{2N-2})^{-1} (1-A^{2N-2}) \{\delta_{i1} + \delta_{iN} + (1-\delta_{i1})(1-\delta_{iN})\}$$

$$= \delta_{ik} .$$

Theorem 1 has now been proven for N = 2 and for N, given that it is true for N-1. Therefore, by induction, Theorem 1 must be true for all $N \geq 2$.

APPENDIX B

PROOF OF THEOREM 2

Theorem 2. For ${}_N[\rho]$ of the form given by (3-24) and $N \geq 2$, the determinant of ${}_N[\rho]$ is defined by

$$\det({}_N[\rho]) = \pi_{n=1}^{N-1} (1-B^n)^{N-n} . \qquad \text{(B-1)}$$

Proof. The proof of Theorem 2 is by induction. For $N = 2$, (B-1) can be solved to show that

$$\det({}_2[\rho]) = (1-B) = (1-A^2)$$

which is, by inspection, the correct expression for the determinant. Therefore, the theorem is satisfied for $N = 2$.

From (3-26), it is apparent that ${}_1M_{11} = 1$. By setting both i and j equal to 1 in (3-27), we see that

$${}_NM_{11} = {}_{N-1}M_{11}$$

for $N \geq 2$. By induction, it must be true that

$${}_NM_{11} = 1 \qquad \text{(B-2)}$$

for $N \geq 1$.

Substitution of (B-2) into (3-28) leads to the conclusion that, for $N \geq 1$,

$$ {}_{N}G_{11} = \pi_{n=1}^{N-1} (1-B^{n})^{-1} \quad . \tag{B-3} $$

As the inverse of ${}_{N}[\rho]$, ${}_{N}[G]$ can alternately be defined by

$$ {}_{N}[G] = (\det({}_{N}[\rho]))^{-1} \operatorname{adj}({}_{N}[\rho]) = (\det({}_{N}[\rho]))^{-1} \, {}_{N}[\Omega] $$

where $\operatorname{adj}({}_{N}[\rho]) = {}_{N}[\Omega]$ is the adjoint matrix of ${}_{N}[\rho]$.

Now, ${}_{N}\Omega_{11}$ is equal to the cofactor of ${}_{N}\rho_{11}$, and inspection of (3-24) shows that the cofactor of ${}_{N}\rho_{11}$ is the determinant of ${}_{N-1}[\rho]$. Hence,

$$ {}_{N}G_{11} = \pi_{n=1}^{N-1} (1-B^{n})^{-1} = (\det({}_{N}[\rho]))^{-1} \, (\det({}_{N-1}[\rho])) $$

or

$$ \det({}_{N}[\rho]) = (\det({}_{N-1}[\rho])) \, \pi_{n=1}^{N-1} (1-B^{n}) \quad . \tag{B-4} $$

Finally, if Theorem 2 is assumed true for N-1, then

$$ \det({}_{N-1}[\rho]) = \pi_{k=1}^{N-2} (1-B^{k})^{N-k-1} \quad . \tag{B-5} $$

Substitution of (B-5) into (B-4) results in

$$ \det({}_{N}[\rho]) = \pi_{n=1}^{N-1} (1-B^{n}) \, \pi_{k=1}^{N-2} (1-B^{k})^{N-1-k} $$

which can be simplified for the final result,

$$\det({}_N[\rho]) = \pi_{n=1}^{N-1} (1-B^n)^{N-n} \ .$$

Theorem 2 has now been verified for N, given that it is true for N-1; it was also verified for N = 2. Therefore, by induction, Theorem 2 must be true for $N \geq 2$.

APPENDIX C

TABLES OF POLYNOMIAL FORMS OF $_{N}M_{i,j}$

TABLE 2

POLYNOMIAL FORM OF $_{N}M_{i,j}$ FOR N = 2

i,j	$_{2}M_{i,j}$
1,1	1
1,2	1
2,1	1
2,2	1

TABLE 3

POLYNOMIAL FORM OF ${}_{N}M_{i,j}$ FOR N = 3

i,j	${}_{3}M_{i,j}$
1,1	1
1,2	$1 + B$
1,3	1
2,1	$1 + B$
2,2	$1 + B + B^2 + B^3$
2,3	$1 + B$
3,1	1
3,2	$1 + B$
3,3	1

TABLE 4

POLYNOMIAL FORM OF $_{N}M_{i,j}$ FOR N = 4

i,j	$_{4}M_{i,j}$
1,1	1
1,2	$1 + B + B^2$
1,3	$1 + B + B^2$
1,4	1
2,1	$1 + B + B^2$
2,2	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5$
2,3	$1 + 2B + 2B^2 + 2B^3 + B^4 + B^5$
2,4	$1 + B + B^2$
3,1	$1 + B + B^2$
3,2	$1 + 2B + 2B^2 + 2B^3 + B^4 + B^5$
3,3	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5$
3,4	$1 + B + B^2$
4,1	1
4,2	$1 + B + B^2$
4,3	$1 + B + B^2$
4,4	1

TABLE 5

POLYNOMIAL FORM OF ${}_{N}M_{i,j}$ FOR N = 5

i,j	${}_{5}M_{i,j}$
1,1	1
1,2	$1 + B + B^2 + B^3$
1,3	$1 + B + 2B^2 + B^3 + B^4$
1,4	$1 + B + B^2 + B^3$
1,5	1
2,1	$1 + B + B^2 + B^3$
2,2	$1 + B + 2B^2 + 3B^3 + 3B^4 + 3B^5 + 2B^6 + B^7$
2,3	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8$
2,4	$1 + 2B + 3B^2 + 3B^3 + 3B^4 + 2B^5 + B^6 + B^7$
2,5	$1 + B + B^2 + B^3$
3,1	$1 + B + 2B^2 + B^3 + B^4$
3,2	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8$
3,3	$1 + B + 3B^2 + 4B^3 + 6B^4 + 6B^5 + 6B^6 + 4B^7 + 3B^8 + B^9 + B^{10}$
3,4	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8$
3,5	$1 + B + 2B^2 + B^3 + B^4$

TABLE 5-Continued

i,j	$5^{M}_{i,j}$
4,1	$1 + B + B^2 + B^3$
4,2	$1 + 2B + 3B^2 + 3B^3 + 3B^4 + 2B^5 + B^6 + B^7$
4,3	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8$
4,4	$1 + B + 2B^2 + 3B^3 + 3B^4 + 3B^5 + 2B^6 + B^7$
4,5	$1 + B + B^2 + B^3$
5,1	1
5,2	$1 + B + B^2 + B^3$
5,3	$1 + B + 2B^2 + B^3 + B^4$
5,4	$1 + B + B^2 + B^3$
5,5	1

TABLE 6

POLYNOMIAL FORM OF $_{N}M_{i,j}$ FOR N = 6

i,j	$_{6}M_{i,j}$
1,1	1
1,2	$1 + B + B^2 + B^3 + B^4$
1,3	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
1,4	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
1,5	$1 + B + B^2 + B^3 + B^4$
1,6	1
2,1	$1 + B + B^2 + B^3 + B^4$
2,2	$1 + B + 2B^2 + 3B^3 + 4B^4 + 4B^5 + 4B^6 + 3B^7 + 2B^8 + B^9$
2,3	$1 + 2B + 3B^2 + 5B^3 + 6B^4 + 7B^5 + 7B^6 + 7B^7 + 5B^8 + 4B^9 + 2B^{10} + B^{11}$
2,4	$1 + 2B + 4B^2 + 5B^3 + 7B^4 + 7B^5 + 7B^6 + 6B^7 + 5B^8 + 3B^9 + 2B^{10} + B^{11}$
2,5	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8 + B^9$
2,6	$1 + B + B^2 + B^3 + B^4$
3,1	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
3,2	$1 + 2B + 3B^2 + 5B^3 + 6B^4 + 7B^5 + 7B^6 + 7B^7 + 5B^8 + 4B^9 + 2B^{10} + B^{11}$

TABLE 6-Continued

i,j	$_6M_{i,j}$
3,3	$1 + B + 3B^2 + 5B^3 + 8B^4 + 10B^5 + 13B^6 + 13B^7 + 13B^8 + 11B^9 + 9B^{10} + 6B^{11} + 4B^{12} + 2B^{13} + B^{14}$
3,4	$1 + 2B + 4B^2 + 6B^3 + 9B^4 + 11B^5 + 13B^6 + 13B^7 + 13B^8 + 10B^9 + 8B^{10} + 5B^{11} + 3B^{12} + B^{13} + B^{14}$
3,5	$1 + 2B + 4B^2 + 5B^3 + 7B^4 + 7B^5 + 7B^6 + 6B^7 + 5B^8 + 3B^9 + 2B^{10} + B^{11}$
3,6	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
4,1	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
4,2	$1 + 2B + 4B^2 + 5B^3 + 7B^4 + 7B^5 + 7B^6 + 6B^7 + 5B^8 + 3B^9 + 2B^{10} + B^{11}$
4,3	$1 + 2B + 4B^2 + 6B^3 + 9B^4 + 11B^5 + 13B^6 + 13B^7 + 13B^8 + 10B^9 + 8B^{10} + 5B^{11} + 3B^{12} + B^{13} + B^{14}$
4,4	$1 + B + 3B^2 + 5B^3 + 8B^4 + 10B^5 + 13B^6 + 13B^7 + 13B^8 + 11B^9 + 9B^{10} + 6B^{11} + 4B^{12} + 2B^{13} + B^{14}$
4,5	$1 + 2B + 3B^2 + 5B^3 + 6B^4 + 7B^5 + 7B^6 + 7B^7 + 5B^8 + 4B^9 + 2B^{10} + B^{11}$
4,6	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
5,1	$1 + B + B^2 + B^3 + B^4$
5,2	$1 + 2B + 3B^2 + 4B^3 + 4B^4 + 4B^5 + 3B^6 + 2B^7 + B^8 + B^9$

TABLE 6-Continued

i,j	$_6M_{i,j}$
5,3	$1 + 2B + 4B^2 + 5B^3 + 7B^4 + 7B^5 + 7B^6 + 6B^7 + 5B^8 + 3B^9 + 2B^{10} + B^{11}$
5,4	$1 + 2B + 3B^2 + 5B^3 + 6B^4 + 7B^5 + 7B^6 + 7B^7 + 5B^8 + 4B^9 + 2B^{10} + B^{11}$
5,5	$1 + B + 2B^2 + 3B^3 + 4B^4 + 4B^5 + 4B^6 + 3B^7 + 2B^8 + B^9$
5,6	$1 + B + B^2 + B^3 + B^4$
6,1	1
6,2	$1 + B + B^2 + B^3 + B^4$
6,3	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
6,4	$1 + B + 2B^2 + 2B^3 + 2B^4 + B^5 + B^6$
6,5	$1 + B + B^2 + B^3 + B^4$
6,6	1