LARGE-SAMPLE ESTIMATION OF AN UNKNOWN DISCRETE WAVEFORM WHICH IS RANDOMLY REPEATING IN GAUSSIAN NOISE[1]

By Melvin Hinich

Carnegie Institute of Technology

1. Summary and introduction. Suppose we have an input $X(t)$ made up of an unknown waveform $\theta(t)$ of known length, which is repeated randomly, and is imbedded in Gaussian noise with a known covariance function. The rate of recurrence of the waveform is a known small constant. In addition, the signal-to-noise ratio of the input $X(t)$ is quite low. We wish to estimate the waveform $\theta(t)$ and its autocorrelation $\psi(\tau) = \int \theta(t + \tau)\theta(t)\, dt$. Restricting ourselves to discrete-time observations on $X(t)$, we shall derive an optimal estimator of the discrete version of $\psi(\tau)$. This estimator is a weighted average of the sample autocorrelation and the square of a linear estimator of the time average (the zero-frequency or DC value) of the waveform. For the estimation of θ, the problem is more complicated. The optimality concept (asymptotic efficiency) used in this work is based upon large-sample theory and the Cramér-Rao Inequality (Chernoff [2] and Cramér [4]).

The problem stated above was motivated by a problem of electronic surveillance of an enemy communication system based upon pulse position modulation, PPM. To illustrate this system suppose station A is sending a message, which we wish to intercept and decode, to station B by PPM over a certain FM bandwidth. A continues to repeat a fixed pulse-type waveform $\theta(t) = \sum_{i=1}^{n} \theta_i H[t - (i - 1)T]$ where

$$H(t) = 1 \quad \text{if} \quad 0 \leqq t < T,$$
$$= 0 \quad \text{otherwise.}$$

The vector $\theta' = (\theta_1, \cdots, \theta_n)$, the parameter n, and T—the pulse width—are known to both A and B. Notice that nT is the time duration (length) of $\theta(t)$.

Since θ, n, and T are known to the receiver B, one may ask what the coding scheme is for the information that A is sending. The answer is that the length of time between successive recurrences of the waveform, is the variable which contains the information.

In many applications the average length of time between successive occurrences of $\theta(t)$ is around 10^2 times nT. While the intervals between repetitions are fundamental in the transmission of information, someone who does not yet know the waveform may regard the recurrences to be purely random in time.

This modulation technique has the effect of spreading the power in the FM

Received 15 February 1964.

[1] This research was sponsored by the Army Research Office, Office of Naval Research, and Air Force Office of Scientific Research.

bandwidth over a wider swath of the frequency scale. This makes surveillance more difficult because it requires that we somehow determine the actual bandwidth being used. Moreover, the spreading of the power makes jamming of the channel difficult. To further complicate matters, A transmits the pulses with low power so that B picks up an input $X(t)$ with low signal-to-noise ratio.

Since B knows $\theta(t)$, he uses matched filtering to detect the times of occurrence of the θ's. If the noise is assumed to be additive Gaussian noise with known covariance, then matched filtering is optimal in a decision theoretic sense (Wainstein and Zubakov [8]).

But suppose that we are listening in on this channel without knowing θ and we wish to find out what A is saying. First we must determine the frequency band of the channel which A is using. Then we must detect the times of occurrence of the θ's, although we do not know θ. However, let us assume that we have already determined n and T, although it will turn out that n is not a vital parameter in the estimator developed in this paper.

Jakowatz, Shuey, and White [6] present a special discrete-time (sampled data) system, called the Adaptive Filter, which estimates an unknown waveform which is repeating in additive noise. The system uses a complicated stochastic iterative procedure. The Filter obtains a crude estimate of the waveform from the initial input and uses the discrete cross-correlation between this estimate and the input to detect the times of occurrence of the waveform. When it decides that a waveform is present in the input, it refines the estimate by averaging it with the section of input where the waveform is thought to be present. Provided the autocorrelation of the waveform—$\psi(\tau)$—has $\psi(0)$, its maximum, a good deal larger than the relative maxima of ψ, and provided the noise is well behaved, then this iterative procedure results in an asymptotically stable estimate for the waveform. This stable estimate is then used as the matching element in matched filter detection of the waveform. Thus we could call the Adaptive Filter an adaptive matched filter. A partial analysis of this system is given by Hinich [5].

The estimate of the discrete autocorrelation of the waveform is helpful to the analysis of systems which are based upon discrete-time cross-correlation such as the Adaptive Filter, since the autocorrelation is a basic parameter in the distributions of the random variables (correlations) which arise in the operation of these systems.

Suppose we can obtain an expression for θ in terms of ψ, $\theta = f(\psi)$. Once we have obtained an asymptotically efficient estimator of ψ, call it $\hat{\psi}$, then $\hat{\theta} = f(\hat{\psi})$ would be an asymptotically efficient estimator of θ. Unfortunately, there is a multiplicity of θ's which have ψ as their autocorrelation. In Section 6 we will present a method for obtaining the correct θ (the one which appears in the input) from ψ by using the observations on $X(t)$. Unfortunately this method does not seem to be sufficiently practical.

To conclude, let us outline the rest of this paper. In Section 2 we give a formal statement and description of the problem posed above. In Section 3 we state the Cramér-Rao theorem and derive the information matrices relevant to the estimation of θ and ψ, as well as their inverses. In Section 4 we present the optimal estimator of ψ. We also show that while the normalized sample correlation is an unbiased estimator of ψ, it is not efficient. In Section 5 we discuss three examples. These are the general case white noise, the case of white noise when the discrete waveform has only two components θ_1, θ_2, and the case of Gaussian Markov noise ($EN(t + \tau)N(t) = \rho^{|\tau|}, 0 < \rho < 1$). In Section 6 we discuss the problem of estimating θ after the autocorrelation ψ has been estimated.

2. Statement of the problem. We shall develop a formal statement of the problem. First let us discuss it informally.

We observe a process $X(t)$ which consists of a randomly occurring unknown waveform plus noise. The noise process $N(t)$ is assumed to be stationary and Gaussian with mean zero and known covariance. Without loss of generality we may normalize so that the noise has variance $\sigma_{N(t)}^2 = EN^2(t) = 1$.

The waveform $\theta(t)$ has known specified length and can be represented as a step function as follows:

$$\begin{aligned} \theta(t) &= \textstyle\sum_{i=1}^n \theta_i H[t - (i - 1)T], && n,\ T \quad \text{known}, \\ H(t) &= 1 && \text{if} \quad 0 \leqq t < T, \\ &= 0 && \text{otherwise}. \end{aligned}$$

The parameter T represents the width of the steps of the step function, i.e., the pulse width. The number of equally spaced steps corresponding to the waveform $\theta(t)$ is given by n.

We assume that the time intervals between repetitions are large compared to the length of the waveform. Let γ be the rate of repetitions measured in units of T seconds (pulse width). Then $\gamma^{-1}T$ seconds is the average time between waveforms, and thus, γ is small.

The ultimate objective is to estimate $\theta' = (\theta_1, \cdots, \theta_n)$. However, the discrete autocorrelation $\psi' = (\psi_1, \cdots, \psi_n)$ defined by:

$$\begin{aligned} \psi_1 &= \tfrac{1}{2}(\theta_1^2 + \theta_2^2 + \cdots + \theta_n^2), \\ \psi_2 &= \theta_1\theta_2 + \theta_2\theta_3 + \cdots + \theta_{n-1}\theta_n, \\ \psi_3 &= \theta_1\theta_3 + \theta_2\theta_4 + \cdots + \theta_{n-2}\theta_n, \\ &\vdots \\ \psi_n &= \theta_1\theta_n, \end{aligned} \tag{1}$$

is an especially useful function of θ. Also useful is the discrete time-average

$$\psi_0 = \textstyle\sum_{i=1}^n \theta_i \qquad (\text{DC value of} \quad \theta). \tag{2}$$

We assume that $\|\theta\| = (\sum_{i=1}^n \theta_i^2)^{\frac{1}{2}}$ is small compared to the variance of the noise. We can state this in terms of R_θ, the signal-to-noise ratio of $X(t)$. By definition, $R_\theta = n^{-1} \sum_{i=1}^n \theta_i^2 / \int_{-\infty}^{\infty} S_N(f)\, df$ where $S_N(f)$ is the spectral density of the noise $N(t)$. But $\int_{-\infty}^{\infty} S_N(f)\, df = EN^2(t) = \sigma_{N(t)}^2 = 1$ by the normalization of N. Thus, $R_\theta = n^{-1} \sum_{i=1}^n \theta_i^2 = \|\theta\|^2/n$ and we assume that R_θ is small.

Now we will discuss the sampling procedure. For large m and a fixed integer w, we will take m groups of w successive discrete observations on $X(t)$; the successive observations being T seconds apart. That is for each t_i such that $t_1 < t_2 < \cdots < t_m$, we observe $X(t_i + T), X(t_i + 2T), \cdots, X(t_i + wT)$. The intervals between the t_i's $(t_{i+1} - t_i)$ are all substantially greater than wT and nT (the time duration of the waveform). Let $X^{(i)} = (X(t_i + T), \cdots, X(t_i + wT))$, $i = 1, \cdots, m$. We thus have a sample of m vector observations on a w-dimensional random variable.

This sampling scheme may be regarded as opening a sequence of windows of width wT seconds, through which we observe the process at m different stages of time.

There are several different possibilities which can occur when a window is opened. There may be no part of the waveform present during the wT second "look" at $X(t)$. In that case we observe only noise. However, the window may open just as the front part of the waveform is "visible". In that case only the head of the vector θ plus noise is observed. Similarly, we might observe only the tail of θ plus noise, or perhaps the middle of θ plus noise. Incidentally let us suppose that $wT \ll \gamma^{-1}T$ since we wish to exclude the possibility of catching two successive waveforms in the window.

Since the window has w components and θ has n components, there are $n + w - 1$ ways of catching part of θ along with the noise. We can represent these $n + w - 1$ possibilities for θ in the window by defining

$$(S_j\theta)' = (\theta_{j+1}, \cdots, \theta_{j+w}), \tag{3}$$

where $\theta_k = 0$ if $k \leqq 0$ or $k \geqq n + 1$.

For example $(S_{n-1}\theta)' = (\theta_n, 0, \cdots, 0)$ and $(S_{-w+1}\theta)' = (0, 0, \cdots, \theta_1)$. There is no way of knowing in advance which case is occurring. Each of the $n + w - 1$ possibilities may be regarded as equally likely. Figure 1 gives an example for $w = 3$, $n = 2$ where the noise has been removed.

We will catch some part of θ in the ith window if and only if $\theta(t)$ commences at time $t_0 : -nT + t_i + T < t_0 < t_i + wT$. Thus, the probability of this event is

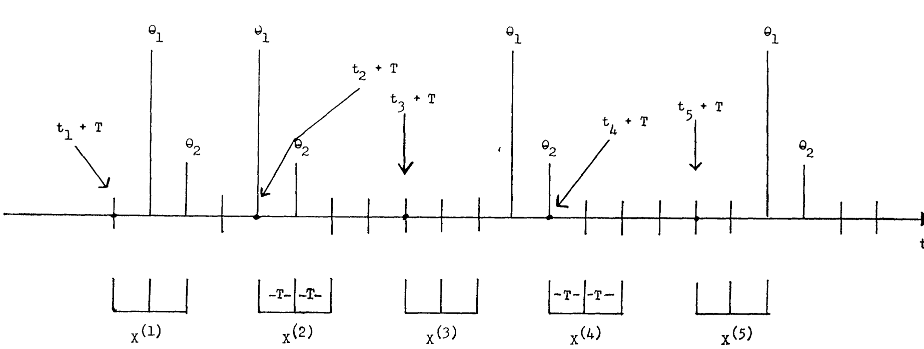

FIG. 1. Example of sampling system with noise removed and $n = 2$, $w = 3$

approximately $(n + w - 1)\gamma$ where γ is the recurrence rate. The probability for observing a specific one of the $n + w - 1$ possibilities is simply γ.

Since the distances between windows are greater than nT, a single waveform cannot appear in two successive windows. Moreover, suppose that for some τ_0, $EN(t + \tau)N(t) = 0$ for $\tau > \tau_0$. Then if we take the windows further apart than τ_0 seconds, the $X^{(i)}$'s are independent. The above restraint on the covariance of the noise holds approximately for many colored noise processes which occur in applications. Of course, for white noise, $\tau_0 = 0$.

We then can sum up this discussion with a formal statistical statement of an idealized version of the problem:

We have m independent w-dimensional random vectors $X^{(1)}, X^{(2)}, \cdots, X^{(m)}$, each identically distributed as X where

$$
\begin{aligned}
X &= N + S_{-w+1}\theta && \text{with probability } \gamma,\\
&= N + S_{-w+2}\theta && \text{with probability } \gamma,\\
&\;\;\vdots\\
&= N + S_{n-2}\theta && \text{with probability } \gamma,\\
&= N + S_{n-1}\theta && \text{with probability } \gamma,\\
&= N && \text{with probability } 1 - (n + w - 1)\gamma
\end{aligned}
\tag{4}
$$

and $S_i\theta$ is defined in (3).

The vector random variable N has a w-dimensional multivariate normal distribution with mean zero and known, non-singular covariance matrix Σ. We express this by $\mathcal{L}\{N\} = \mathfrak{N}(0, \Sigma)$ where $\mathcal{L}\{N\}$ is the distribution function of the random variable N.

We desire to estimate the n-dimensional vectors θ and ψ, ($\psi_1 = \frac{1}{2}\sum_1^n \theta_i^2$, $\psi_2 = \sum_1^{n-1}\theta_i\theta_{i+1}, \cdots, \psi_n = \theta_1\theta_n$), where it is assumed that γ is small, $\|\theta\| = [\sum_{i=1}^n \theta_i^2]^{\frac{1}{2}}$ is small and the sample size m is large.

3. Information matrix. The Cramér-Rao theorem gives a bound for the asymptotic variance of estimators in terms of the inverse of the information matrix (defined below). In this section we will compute the information matrix and its inverse for the given problem.

Given a parameter vector $\varphi' = (\varphi_1, \cdots, \varphi_n)$ and a random variable X with density function $p(x \mid \varphi)$, the information matrix for p and φ, $I(\varphi)$ is defined by:

$$I(\varphi) = E_\varphi[(\partial/\partial\varphi) \log p(X \mid \varphi)][(\partial/\partial\varphi) \log p(X \mid \varphi)]' : n \times n, \tag{5}$$

where

$$[(\partial/\partial\varphi) \log p(x \mid \varphi)]' = [(\partial/\partial\varphi_1) \log p(x \mid \varphi), \cdots, (\partial/\partial\varphi_n) \log p(x \mid \varphi)]$$

and E_φ represents expectation with respect to $p(x \mid \varphi)$.

Consider a random vector $T(X)' = (T_1(X), \cdots, T_n(X))$. $T(X)$ is called an unbiased estimator of φ if $E_\varphi T(X) = \varphi$. Furthermore we call $K_T =$

$E_\varphi[T(X) - E_\varphi T(X)][T(X) - E_\varphi T(X)]'$ the covariance of T, thus if T is an unbiased estimator of φ, $K_T = E_\varphi[T(X) - \varphi][T(X) - \varphi]'$.

The Cramér-Rao inequality states that if T is an unbiased estimator of φ, $K_T \geqq I^{-1}(\varphi)$ where the inequality between these matrices means that for any vector a, $a'K_Ta \geqq a'I^{-1}(\varphi)a$.

But suppose we have a random sample $X_1, \cdots, X_m$ of m independent observations from a population with density $p(x \mid \varphi)$. Let $\{T_m(X_1, \cdots, X_m): m = 1, 2, \cdots\}$ be a sequence of estimators of φ such that $\mathcal{L}\{m^{\frac{1}{2}}(T_m - \varphi)\} \to \mathfrak{N}(0, K)$ as $m \to \infty$. That is the distribution functions of $m^{\frac{1}{2}}[T_m(X_1, \cdots, X_m)]$ converge to an n-dimensional multivariate normal distribution with covariance matrix K. Then K is called the asymptotic covariance matrix of $\{T_m\}$.

Paraphrasing Stein's generalization of the Cramér-Rao theorem (see Theorem 1 in Chernoff [2]), there does not exist any estimator of φ which has an asymptotic covariance matrix K such that for some vector a, $a'Ka < a'I^{-1}(\varphi)a$, for all φ in some open set S. *Thus $I^{-1}(\varphi)$ is "essentially" the lower bound for the asymptotic covariance of estimators of φ.* Thus we shall call the sequence $\{T_m\}$ asymptotically efficient if $K = I^{-1}(\varphi)$. Cramér [4] shows that under mild conditions, the maximum likelihood estimator is asymptotically efficient.

We will now deal with the information matrices of interest in this work. However, to facilitate the algebra we make the one-to-one transformation,

$$Z = \Sigma^{-1}X, \tag{6}$$

where Σ^{-1} is the inverse of Σ, the covariance matrix of the Gaussian noise vector N. From (4) we have

$$\begin{aligned} Z &= N^* + \Sigma^{-1}S_j\theta \qquad \text{with probability} \quad \gamma \quad \text{for each} \\ &\qquad\qquad\qquad\qquad j = -w+1, \cdots, n-1, \\ &= N^* \qquad\qquad\qquad \text{with probability} \quad 1-(n+w-1)\gamma \end{aligned} \tag{7}$$

where $\mathcal{L}\{N^*\} = \mathfrak{N}(0, \Sigma^{-1})$.

Let $f(z \mid \theta)$ be the density of Z, parameterized by the waveform vector θ. Therefore from (7) we have

$$f(z \mid \theta) = [1-(n+w-1)\gamma]n(z \mid 0, \Sigma^{-1}) + \gamma \textstyle\sum_{j=-w+1}^{n-1} n(z \mid \Sigma^{-1}S_j\theta, \Sigma^{-1}), \tag{8}$$

where

$$n(z \mid \varphi, C) = (2\pi)^{-w/2}|C|^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(z-\varphi)'C^{-1}(z-\varphi)\right]$$

is a w-dimensional normal density with mean φ and covariance matrix C.

Notice that $f(z \mid \theta)$ is a convex combination of multivariate normal densities, but it is not in general multivariate normal itself. We shall handle it by making Taylor series approximations with θ in the neighborhood of zero.

Let $I(\theta)$ be the information matrix for $f(z \mid \theta)$ and θ. As in (5),

$$I(\theta) = E_\theta[(\partial/\partial\theta) \log f(Z \mid \theta)][(\partial/\partial\theta) \log f(Z \mid \theta)]'. \tag{9}$$

But we also wish to deal with the information matrix for the autocorrelation ψ given in (1). We see that $\psi(\theta)$ is not a one-to-one transformation for all θ. Let J be the Jacobian matrix of the transformation, $J_{ij} = \partial\psi_j/\partial\theta_i$ with $i, j = 1, \cdots, n$ It is easy to see that

$$(10) \qquad J = \begin{pmatrix} \theta_1 & \theta_0 + \theta_2 & \cdots & \theta_{2-n} + \theta_n \\ \theta_2 & \theta_1 + \theta_3 & \cdots & \theta_{3-n} + \theta_{n+1} \\ \vdots & \vdots & & \vdots \\ \theta_i & \theta_{i-1} + \theta_{i+1} & \cdots & \theta_{i-n+1} + \theta_{i+n-1} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_n & \theta_{n-1} + \theta_{n+1} & \cdots & \theta_1 + \theta_{2n-1} \end{pmatrix},$$

where we recall that $\theta_k = 0$ if $k \leqq 0$ or $k \geqq n + 1$. It is not difficult to see that

$$(11) \qquad |J| = [\textstyle\sum_{i=1}^n \theta_i][\sum_{i=1}^n (-1)^i\theta_i]P_{n-2}(\theta),$$

where $P_{n-2}(\theta)$ is a homogeneous $n - 2$ degree polynomial in the θ_i's. Therefore the set of θ's such that $|J| = 0$ form surfaces which divide the n-dimensional Euclidean space into at most 2^n regions where J does not vanish. We will restrict ourselves to θ vectors such that J does not vanish. Thus J^{-1} exists for these θ's. Moreover given a θ_0 such that $|J| \neq 0$, there exists a neighborhood of θ_0 in which $\psi(\theta)$ is one-to-one.

Now we will deal with the approximations of $f(z \mid \theta)$. From (8) we have

$$(12) \qquad \frac{f(z \mid \theta)}{f(z \mid 0)} = 1 - (n + w - 1)\gamma + \gamma \sum_{j=-w+1}^{n-1} \frac{n(z \mid \Sigma^{-1}S_j\,\theta, \Sigma^{-1})}{n(z \mid 0, \Sigma^{-1})},$$

since $f(z \mid 0) = n(z \mid 0, \Sigma^{-1})$. By expanding in Taylor's series about $\varphi = 0$ we have

$$(13) \qquad \begin{aligned} n(z \mid \Sigma^{-1}\varphi, \Sigma^{-1})/n(z \mid 0, \Sigma^{-1}) &= 1 + z'\varphi + \tfrac{1}{2}\varphi'(zz' - \Sigma^{-1})\varphi \\ &+ \textstyle\sum_{i,j,k=1}^w G_3(z \mid i, j, k)c_3(i, j, k)\varphi_i\varphi_j\varphi_k \\ &+ \textstyle\sum_{i,j,k,l=1}^w G_4(z \mid i, j, k, l)c_4(i, j, k, l)\varphi_i\varphi_j\varphi_k\varphi_l + \|\varphi\|^5K^*(z, \varphi), \end{aligned}$$

where the c's are constants and $|K^*(z, \varphi)| \leqq d \exp t_1^*|z_1| + \cdots + t_w^*|z_w|$ for some $t_i^* > 0$ and $d > 0$. Thus $E_0[K^*(Z, \varphi)]^r$ exists and is bounded by some number independent of φ for each $r \geqq 0$. With the notation $\Sigma^{-1} = (\sigma^{ij})$,

$$(14a) \qquad G_3(z \mid i, j, k) = z_iz_jz_k - \sigma^{ij}z_k - \sigma^{ik}z_j - \sigma^{jk}z_i,$$

$$(14b) \qquad \begin{aligned} G_4(z \mid i, j, k, l) &= z_iz_jz_kz_l - 2\sigma^{ij}z_kz_l - 2\sigma^{ik}z_jz_l \\ &- 2\sigma^{il}z_jz_k + \sigma^{ij}\sigma^{kl} + \sigma^{ik}\sigma^{jl} + \sigma^{il}\sigma^{jk}. \end{aligned}$$

From page 39 of Anderson [1], $E_0Z_iZ_jZ_kZ_l = \sigma^{ij}\sigma^{kl} + \sigma^{ik}\sigma^{jl} + \sigma^{il}\sigma^{jk}$, $E_0Z_iZ_j = \sigma^{ij}$. Moreover all odd moments of the Z_i's are zero. We then have the following orthogonality relationships:

$$(15) \qquad E_0Z_i(Z_jZ_k - \sigma^{ik}) = 0 \qquad \text{for all } i, j, k,$$

$$E_0Z_iG_3(Z \mid j, k, l) = 0 \qquad \text{for all } i, j, k, l,$$
$$E_0Z_iG_4(Z \mid j, k, l, m) = 0 \qquad \text{for all } i, j, k, l, m,$$
$$E_0(Z_iZ_j - \sigma^{ij})G_3(Z \mid k, l, m) = 0 \qquad \text{for all } i, j, k, l, m.$$

Putting $\varphi = S_j\theta$ in (13) and summing, we have from (12)

$$\begin{aligned} f(z \mid \theta)/f(z \mid 0) = 1 &+ \gamma[z'(\textstyle\sum_j S_j\theta) + \frac{1}{2}\sum_j (S_j\theta)'(zz' - \Sigma^{-1})(S_j\theta) \\ &+ \textstyle\sum_{i,j,k} G_3(z \mid i, j, k)O(\|\theta\|^3) + \sum_{i,j,k,l} G_4(z \mid i, j, k, l)O(\|\theta\|^4) \\ &+ K(z, \theta)O(\|\theta\|^5)]. \end{aligned} \tag{16}$$

$E_0[K(Z, \theta)]^r$ exists and is bounded by some number independent of θ for each $r \geqq 0$. Moreover the $O(\|\theta\|^3)$ and $O(\|\theta\|^4)$ terms are functions of θ which do not involve z.

Applying (1), (2), and (3) we have

$$\textstyle\sum_{j=-w+1}^{n-1} S_j\theta = \psi_0 1_w , \tag{17}$$

where $1_w' = (1, \cdots, 1)$ is a vector of w ones, and for any symmetric $w \times w$ matrix $A = (a_{ij})$

$$\begin{aligned} \tfrac{1}{2}\textstyle\sum_{j=-w+1}^{n-1} (S_j\theta)'A(S_j\theta) &= (\textstyle\sum_{i=1}^{w} a_{ii})\psi_1 + (\sum_{i=1}^{w-1} a_{i,i+1})\psi_2 \\ &\quad + \cdots + (\textstyle\sum_{i=1}^{w-n+1} a_{i,i+n-1})\psi_n , \end{aligned} \tag{18}$$

where a_{ij} is understood to be zero when i or j is greater than w or less than 1. Thus, if $w < n$, the coefficients of $\psi_{w+1}, \cdots, \psi_n$ vanish and $\frac{1}{2}\sum_j (S_j\theta)'A(S_j\theta) = (\sum_{i=1}^{w} a_{ii})\psi_1 + \cdots + a_{1w}\psi_w$.

Applying (17) and (18) to (16) we have the following result.

LEMMA 1. *Define* $Y(z)' = (Y_1(z), \cdots, Y_n(z))$ *by*

$$Y_k(z) = \textstyle\sum_{i=1}^{w-k+1} (z_iz_{i+k-1} - \sigma^{i,i+k-1}), \qquad k = 1, \cdots, n. \tag{19}$$

Then

$$\begin{aligned} f(z \mid \theta)/f(z \mid 0) = 1 &+ \gamma[(z'1_w)\psi_0 + Y(z)'\psi + \textstyle\sum_{i,j,k} G_3(z \mid i, j, k)O(\|\theta\|^3) \\ &+ \textstyle\sum_{i,j,k,l} G_4(z \mid i, j, k, l)O(\|\theta\|^4) + K(z, \theta)O(\|\theta\|^5)]. \end{aligned} \tag{20}$$

It is understood that $z_i = 0$ if $i \leqq 0$ or $i \geqq w + 1$, and thus if $w < n$, $Y_{w+1}(z) = \cdots = Y_n(z) \equiv 0$. We then define for $w < n$, $Y(z \mid w)' = (Y_1(z), \cdots, Y_w(z))$.

We now introduce the $n \times n$ matrix.

$$D = E_0Y(Z)Y(Z)' = (d_{ij}). \tag{21a}$$

If $w < n$ we can write

$$D = \begin{pmatrix} D_w & 0 \\ 0 & 0 \end{pmatrix},$$

where $D_w = E_0Y(Z \mid w)Y(Z \mid w)'$ is a $w \times w$ matrix. From (19) we can show for $i, j = 1, 2, \cdots, n$ that

(21b) $$d_{ij} = \sum_{k=1}^{w-i+1} \sum_{l=1}^{w-j+1} (\sigma^{kl}\sigma^{k+i-1,l+j-1} + \sigma^{k,l+j-1}\sigma^{k+i-1,l}),$$

where it is understoon that σ^{kl} is zero if k or l is greater than w or less than 1.

LEMMA 2. *For $w \geqq n$, D is positive definite. For $w < n$, D_w is positive definite.*

PROOF. Let $w \geqq n$ and notice that $E_0Y(Z) = 0$. Thus D is the covariance matrix of $Y(Z)$ and is therefore non-negative definite. Suppose D is singular. Then there exists a non-zero vector a such that $a'Y(Z) = 0$ with probability one, which implies that,

$$2a_1U_1 + a_2U_2 + \cdots + a_nU_n = \text{constant},$$

where

$$\begin{aligned} U_1 &= \tfrac{1}{2}\textstyle\sum_{k=1}^{w} Z_k^2, \\ U_2 &= \textstyle\sum_{k=1}^{w-1} Z_kZ_{k+1}, \\ &\vdots \\ U_n &= \textstyle\sum_{k=1}^{w-n+1} Z_kZ_{k+n-1}, \\ &\vdots \\ U_w &= Z_1Z_w. \end{aligned}$$

We see that $U(Z)' = (U_1, \cdots, U_w)$ results from an autocorrelation operation on the random vector Z. Therefore the Jacobian matrix of the transformation $U(Z)$ is given by (10) with n replaced by w and θ_k by Z_k.

As in the case of the transformation from θ to ψ, there exists a point z_0 in a neighborhood of which (i) the Jacobian is non-zero, and (ii) restricted to this neighborhood U is a one-to-one transformation mapping w-dimensional open sets onto w-dimensional open sets.

Now Z is a non-degenerate w-dimensional multivariate normal random variable and every region in the Z space has positive probability. Hence there is a neighborhood in the U space such that every w-dimensional cube in that neighborhood has positive probability. But then it is impossible that U be on the hyperplane $2a_1U_1 + a_2U_2 + \cdots + a_nU_n = \text{constant}$, with probability one. Hence D is non-singular.

For $w < n$, a similar argument holds for the hyperplane $2a_1U_1 + a_2U_2 + \cdots + a_wU_w = \text{constant}$.

We will now approximate $I(\theta)$. From (5) we rearrange terms and write

(22) $$I(\theta) = A(\theta) - B(\theta) + C(\theta),$$

where

$$A(\theta) = E_0\{[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta\}\{[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta\}',$$

(23) $$\begin{aligned} B(\theta) &= E_0\{[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta\}\{[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta\}' \\ &\qquad \cdot[f(Z \mid \theta)/f(Z \mid 0) - 1], \\ C(\theta) &= E_0\{[[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta][[f(Z \mid 0)]^{-1}\partial f(Z \mid \theta)/\partial\theta]' \\ &\qquad \cdot[f(Z \mid \theta)/f(Z \mid 0) - 1]^2 f(Z \mid 0)/f(Z \mid \theta)\} \end{aligned}$$

and

$$[\partial f(Z \mid \theta)/\partial\theta]' = (\partial f/\partial\theta_1, \cdots, \partial f/\partial\theta_n).$$

The Taylor expansion argument which gave (20) in Lemma 1 also yields

$$\begin{aligned}[f(z \mid 0)]^{-1}\partial f(z \mid \theta)/\partial\theta \\ = \gamma\, [(1_w'z)1_n + JY(z) + \textstyle\sum_{i,j,k} G_3(z \mid i, j, k)O(\|\theta\|^2) \\ + \textstyle\sum_{i,j,k,l} G_4(z \mid i, j, k, l)O(\|\theta\|^3) + K(z, \theta)O(\|\theta\|^4)],\end{aligned}$$

where $J = (\partial\psi_j/\partial\theta_i)$ is the Jacobian given by (10). Using (24) in (23) and noting that $f(z \mid 0)/f(z \mid \theta) \leqq [1 - (n + w - 1)\gamma]^{-1}$, we have

$$A(\theta) = \gamma^2[(1_w'\Sigma^{-1}1_w)1_n1_n' + JDJ' + O(\|\theta\|^4)],$$

$$\text{(25a)}\quad B(\theta) = \gamma^3\{[E_0(1_w'Z)^2Y(Z)']\psi(1_n1_n') + \psi_0[JE_0(1_w'Z)^2Y(Z)1_n' + 1_nE_0(1_w'Z)^2Y(Z)'J'] + O(\|\theta\|^2)\},$$

$$\text{(25b)}\quad B(\theta) = \gamma^3O(\|\theta\|^2),$$

$$\text{(25c)}\quad C(\theta) = \gamma^4O(\|\theta\|^2),$$

since the G's and K have bounded moments with respect to $n(z \mid 0, \Sigma^{-1})$. Hence from (22),

$$\text{(26)}\qquad I(\theta) = \gamma^2[(1_w'\Sigma^{-1}1_w)1_n1_n' + JDJ' + O(\gamma\|\theta\|^2) + O(\|\theta\|^4)].$$

Applying the chain rule (of differentiation with respect to θ) to $\psi_0 = (2\sum_1^n \psi_i)^{1/2} = \sum_1^n \theta_i$ we have $\psi_0^{-1}J1_n = 1_n$ and consequently $J^{-1}1_n = \psi_0^{-1}1_n$. Combining this with (26) we have

LEMMA 3.

$$I(\theta) = \gamma^2J[(1_w'\Sigma^{-1}1_w/\psi_0^2)1_n1_n' + D + O(\gamma) + O(\|\theta\|^2)]J'.$$

If $I(\theta)$ is non-singular, then $I^{-1}(\theta)$ is the lower bound for the asymptotic covariance of estimators of θ. To calculate $I^{-1}(\theta)$ we shall make use of the following lemma which we state without proof.

LEMMA 4. *Let $M = \alpha^{-1}11' + R$ where R is positive definite. Then,*

$$\begin{aligned}M^{-1} &= R^{-1} - [R^{-1}11'R^{-1}/(\alpha + 1'R^{-1}1)] \\ &= R^{-1} - [R^{-1}11'R^{-1}/1'R^{-1}1] + [R^{-1}11'R^{-1}/(1'R^{-1}1)^2]\alpha + O(\alpha^2)\end{aligned}$$

as $\alpha \to 0$.

Now let $\Sigma_{(\psi,\psi_0)}$ denote the Cramér-Rao lower bound for the asymptotic covariance of estimators of the $n + 1$ parameters $\psi_1, \psi_2, \cdots, \psi_n, \psi_0$.

In order to obtain $\Sigma_{(\psi,\psi_0)}$ from $I^{-1}(\theta)$, let us now state a result which was originally derived by a procedure called the delta method, (Mann and Wald [7]).

Let $\{T_m\}$ be a sequence of estimators of $\varphi^{*\prime} = (\varphi_1^*, \cdots, \varphi_r^*)$ such that $\mathcal{L}\{m^{\frac{1}{2}}(T_m - \varphi^*)\} \to \mathfrak{N}(0, K)$ as $m \to \infty$. Let g be a vector-valued function of φ continuously differentiable at φ^*. Then

(27) $$\mathcal{L}\{m^{\frac{1}{2}}[g(T_m) - g(\varphi^*)]\} \to \mathfrak{N}(0, [\partial g/\partial\varphi]'K[\partial g/\partial\varphi]),$$

where $[\partial g/\partial\varphi]$ is a matrix whose (i, j)th element is given by $(\partial g_j/\partial\varphi_i)_{\varphi=\varphi^*}$. If we let Σ_g denote the lower bound for the asymptotic covariance of estimators of $g(\theta)$, then if $I(\varphi)$ is non-singular

(28) $$\Sigma_g = [\partial g/\partial\varphi]'I^{-1}(\varphi)[\partial g/\partial\varphi].$$

We then have from (28) if $I(\theta)$ is non-singular,

(29) $$\Sigma_{(\psi,\psi_0)} = [\partial(\psi, \psi_0)/\partial\theta]'I^{-1}(\theta)[\partial(\psi, \psi_0)/\partial\theta],$$

where $\partial(\psi, \psi_0)/\partial\theta$ is the Jacobian matrix with elements $\partial\psi_j/\partial\theta_i$ for $j = 1, 2, \cdots, n, 0$ and $i = 1, \cdots, n$.

Thus we can write

(30) $$\partial(\psi, \psi_0)/\partial\theta = (J, 1_n).$$

Applying (29), (30), and Lemmas 2, 3, and 4 we have

THEOREM 1. *If $w \geqq n$, then $\Sigma_{(\psi,\psi_0)} = \gamma^{-2}K[I + O(\gamma) + O(\|\theta\|^2)]$, where I is the identity matrix and $K = (k_{ij})$ is a $(n + 1) \times (n + 1)$ symmetric matrix such that given the notation that $e_i' = (0, \cdots, 0, 1, 0, \cdots, 0)$ is a vector of zeros* with a one only in the ith position,

$$k_{ij} = e_i'D^{-1}e_j - (1_n'D^{-1}e_i)(1_n'D^{-1}e_j)/(1_n'D^{-1}1_n)$$

for $i, j = 1, \cdots, n$ and

$$k_{(n+1)i} = \psi_0(1_n'D^{-1}e_i)/(1_w'\Sigma^{-1}1_w)(1_n'D^{-1}1_n)$$

$$k_{(n+1)(n+1)} = (1_w'\Sigma^{-1}1_w)^{-1}.$$

COROLLARY 1.1. *Let $\sigma^2_{\psi_0}$ denote the Cramér-Rao lower bound for the asymptotic variance of estimators of ψ_0. Then*

$$\sigma^2_{\psi_0} = \gamma^{-2}[(1_w'\Sigma^{-1}1_w)^{-1} + O(\gamma) + O(\|\theta\|^2)].$$

COROLLARY 1.2. *Let Σ_ψ denote the Cramér-Rao lower bound for the asymptotic covariance of estimators of ψ. Then*

$$\Sigma_\psi = \gamma^{-2}[D^{-1} - (D^{-1}1_n1_n'D^{-1}/1_n'D^{-1}1_n) + O(\gamma) + O(\|\theta\|^2)].$$

If $w < n$ the main part of $I(\theta)$, as represented in Lemma 3, is a singular matrix. From (20) we see that for θ in the neighborhood of zero, $f(z \mid \theta)$ has $(\psi_1, \cdots, \psi_w)$ as natural parameters. We will obtain the lower bound for the asymptotic covariance of estimators of the n parameters $\psi_1, \cdots, \psi_{n-1}, \psi_0$, which will then give as a corollary the lower bound for the asymptotic covariance of estimators of $\psi_0, \psi_1, \cdots, \psi_w$.

If we let $\tau = (\psi_1, \cdots, \psi_{n-1}, \psi_0)$ denote the parameter vector, $f^*(z \mid \tau)$ denote the density of Z parameterized by τ, and $I^*(\tau)$ be the information matrix with respect to f^* and τ, then from Lemma 3 it can be shown that when $w < n$, I^* is singular.

However, let us define the pseudo-inverse with respect to P by

$$[I_p^*(\tau)]^{-1} = \lim_{\lambda\to 0+}[I^*(\tau) + \lambda P]^{-1}, \tag{31}$$

where P is a symmetric matrix such that $I^* + \lambda P$ is positive definite for $\lambda > 0$. Chernoff [3] shows that the (i, j)th element of I_P^{*-1} is independent of the choice of P if the ith and jth diagonal elements are finite. The submatrix of these invariant elements is the lower bound for the asymptotic covariance for estimators of the ψ_i corresponding to these finite diagonal elements.

Using Lemmas 2, 3, and 4 and by use of matrix inequalities, we have

THEOREM 2. *If $w < n$,*

$$I_P^{*-1} \geqq \gamma^{-2}\begin{bmatrix} D_w^{-1} & R_1 & 0 \\ R_1' & R_2 & R_3 \\ 0 & R_3' & (1_w'\Sigma^{-1}1_w)^{-1} \end{bmatrix}[I + O(\gamma) + O(\|\theta\|^2)],$$

were we leave the matrices R_1, R_2, and R_3 unspecified.

COROLLARY 2.1. *If $w < n$,*

$$\sigma^2_{\psi_0} \geqq \gamma^{-2}[(1_w'\Sigma^{-1}1_w)^{-1} + O(\gamma) + O(\|\theta\|^2)].$$

COROLLARY 2.2. *Let $\Sigma_{\psi(w)}$ be the Cramér-Rao lower bound for estimators of the w-dimensional vector $\psi(w)' = (\psi_1, \cdots, \psi_w)$. Then*

$$\Sigma_{\psi(w)} \geqq \gamma^{-2}[D_w^{-1} + O(\gamma) + O(\|\theta\|^2)].$$

If $w = n - 1$, we can determine ψ_n from τ because

$$\psi_n = \tfrac{1}{2}\psi_0^2 - \textstyle\sum_{j=1}^{n-1} \psi_j. \tag{32}$$

Let $\sigma^2_{\psi_n}$ be the Cramér-Rao bound for estimators of ψ_n. Then

$$\sigma^2_{\psi_n} = (\partial\psi_n/\partial\tau)'I_P^{*-1}(\partial\psi_n/\partial\tau). \tag{33}$$

Thus we have from (33) and Theorem 2,

COROLLARY 2.3. *If $w = n - 1$, $\sigma^2_{\psi_n} \geqq \gamma^{-2}[1_{n-1}'D_w^{-1}1_{n-1} + O(\gamma) + O(\|\theta\|^2)]$.*

4. Estimation of ψ and ψ_0. The random vector $Y(Z)$ defined in (19) as $Y_k(Z) = \sum_{i=1}^{w-k+1}(Z_iZ_{i+k-1} - \sigma^{i,i+k-1})$ could be called the autocorrelation statistic of $Z = \Sigma^{-1}X$. An important example is the case where $\Sigma = I$, the identity matrix. Then

$$\begin{aligned} Y_1(X) &= (X_1^2 - 1) + \cdots + (X_w^2 - 1), \\ Y_2(X) &= X_1X_2 + \cdots + X_{w-1}X_w, \\ &\vdots \\ Y_n(X) &= X_1X_n + \cdots + X_{w-n+1}X_w. \end{aligned} \tag{34}$$

In this section we will give asymptotically efficient estimators of ψ and ψ_0. These estimators are functions of $Y(Z)$ and $Z'1_w$.

In order to demonstrate efficiency, we will need the mean and covariance matrix of the $n+1$-dimensional random vector $(Y(Z)', Z'1_w)$.

LEMMA 5.

$$E_\theta \begin{pmatrix} Y(Z) \\ Z'1_w \end{pmatrix} = \gamma \begin{pmatrix} D\psi \\ 1_w'\Sigma^{-1}1_w\psi_0 \end{pmatrix}.$$

PROOF. For any φ,

$$\int (zz' - \Sigma^{-1}) n(z \mid \Sigma^{-1}\varphi, \Sigma^{-1})\, dz = (\Sigma^{-1}\varphi)(\Sigma^{-1}\varphi)' = \Sigma^{-1}\varphi\varphi'\Sigma^{-1}.$$

Thus from (7),

$$E_\theta(ZZ' - \Sigma^{-1}) = \gamma\Sigma^{-1}[\textstyle\sum_{j=-w+1}^{n-1} (S_j\theta)(S_j\theta)']\Sigma^{-1}. \tag{35}$$

The (r, s) element of $\gamma^{-1}E_\theta(ZZ' - \Sigma^{-1})$ is then,

$$E_\theta(Z_rZ_s - \sigma^{rs}) = \textstyle\sum_{j=1}^{n} \psi_j \sum_{l=1}^{w-j+1} (\sigma^{rl}\sigma^{s,l+j-1} + \sigma^{r,l+j-1}\sigma^{sl}). \tag{36}$$

Thus from (19), (21b), and (36) it follows that $E_\theta Y_k(Z) = \gamma \sum_{j=1}^{n} d_{kj}\psi_j$ and therefore

$$E_\theta Y(Z) = \gamma D\psi. \tag{37}$$

Now $\int z\, n(z \mid \Sigma^{-1}\varphi, \Sigma^{-1})\, dz = \Sigma^{-1}\varphi$ for any φ. Thus from (7),

$$\begin{aligned} E_\theta Z &= \gamma\Sigma^{-1}(\textstyle\sum_{j=-w+1}^{n-1} S_j\theta) \\ &= \gamma(\Sigma^{-1}1_w)\psi_0 . \end{aligned}$$

Thus $E_\theta Z'1_w = \gamma(1_w'\Sigma^{-1}1_w)\psi_0$.

LEMMA 6. *Let* $\mathbf{C}$ *be the covariance matrix of* $(Y(Z)', Z'1_w)$. *Then*

$$\mathbf{C} = \begin{pmatrix} D & \gamma O(\psi_0) \\ \gamma O(\psi_0) & 1_w'\Sigma^{-1}1_w \end{pmatrix} + \gamma O(\|\theta\|^2).$$

PROOF. $E_\theta[Y(Z) - \gamma D\psi][Y(Z)' - \gamma\psi'D] = E_\theta Y(Z)Y(Z)' - \gamma^2 D\psi\psi'D$. From (20) and (21a), we have

$$\begin{aligned} E_\theta Y(Z)Y(Z)' &= E_0 Y(Z)Y(Z)' + \gamma O(\|\theta\|^2) \\ &= D + \gamma O(\|\theta\|^2). \end{aligned}$$

From (36) it is easy to show that

$$E_\theta(Z'1_w)^2 = 1_w'\Sigma^{-1}1_w + \gamma 1_w'\mathbf{H}\psi, \tag{38}$$

where $\mathbf{H}^*$ is a $w \times n$ matrix with elements

$$\mathbf{h}_{ij} = \textstyle\sum_{s=1}^{w} \sum_{l=1}^{w-j+1} (\sigma^{il}\sigma^{s,l+j-1} + \sigma^{i,l+j-1}\sigma^{sl}).$$

We then have the indicated variance term for $Z'1_w$.

$E_\theta[Y(Z) - \gamma D\psi][Z'1_w - \gamma(1_w'\Sigma^{-1}1_w)\psi_0]$

$$= E_\theta Y(Z)(Z'1_w) - \gamma^2(1_w'\Sigma^{-1}1_w)D\psi\psi_0 = \gamma[E_0(Z'1_w)^2 Y(Z)\psi_0 + O(\|\theta\|^2)]$$

from (20). We thus have the desired result.

In the beginning of Section 2 we introduced the concept of asymptotic efficiency. However, in this section we will use a somewhat weaker concept of efficiency. A sequence of estimators $\{T_m\}$ will be called efficient near zero if for every vector a, $\lim_{\gamma,\varphi\to 0} (a'I^{-1}(\varphi)a/a'Ka) = 1$.

Let $^*[Y(Z)] = m^{-1}\sum_{k=1}^m Y(Z^{(k)})$ where $Z^{(k)} = \Sigma^{-1}X^{(k)}$. Then we have

THEOREM 3. *If $w \geqq n$, $\tilde{\psi} = \gamma^{-1}D^{-1}\,{}^*[Y(Z)]$ is an unbiased estimator of ψ but it is not asymptotically efficient near zero.*

PROOF. From the central limit theorem and Lemmas 5 and 6,

$$\mathfrak{L}\{m^{\frac{1}{2}}[{}^*[Y(Z)] - \gamma D\psi]\} \to \mathfrak{N}(0, K_Y),$$

where $K_Y = D + \gamma O(\|\theta\|^2)$ is the covariance of Y. Hence $\mathfrak{L}\{m^{\frac{1}{2}}(\tilde{\psi} - \psi)\} \to \mathfrak{N}(0, K_\psi)$ as $m \to \infty$, where $K_\psi = \gamma^{-2}D^{-1}K_Y D^{-1} = \gamma^{-2}[D^{-1} + \gamma O(\|\theta\|^2)]$ is the asymptotic covariance matrix of $\tilde{\psi}$. From Corollary 1.2, for any vector a,

$$a'\Sigma_\psi a = \gamma^{-2}[a'D^{-1}a - [(a'D^{-1}1_n)^2/1_n'D^{-1}1_n] + O(\gamma) + O(\|\theta\|^2)].$$

Thus

$$a'\Sigma_\psi a/a'K_\psi a = 1 - [(a'D^{-1}1_n)^2/(1_n'D^{-1}1_n)(a'D^{-1}a)] + O(\gamma) + O(\|\theta\|^2)$$

and therefore, $\lim_{\gamma,\theta\to 0}(a'\Sigma_\psi a/a'K_\psi a) < 1$. This lack of efficiency is surprising since $\tilde{\psi}$ is a linear function of the sample autocorrelation $Y(Z)$ and one could expect that it would be efficient in estimating the waveform autocorrelation ψ. However, if $w < n$, we do have this "naive" estimator efficient in the estimation of the first w partial correlations $\psi_1, \cdots, \psi_w$. We have

THEOREM 4. *If $w < n$, then $\tilde{\psi}(w) = \gamma^{-1}D_w^{-1}\,{}^*[Y(Z \mid w)]$ is an asymptotically efficient estimator near zero, of $\psi(w)' = (\psi_1, \cdots, \psi_w)$ where $^*[Y(Z \mid w)] = \{{}^*[Y_1(Z)], \cdots, {}^*[Y_w(Z)]\}$.*

PROOF. By applying the central limit theorem in a manner similar to that used in the proof of Theorem 3, we have from (21a) and Lemmas 5 and 6, $\mathfrak{L}\{m^{\frac{1}{2}}[\hat{\psi}(w) - \psi(w)]\} \to \mathfrak{N}(0, K_{\psi(w)})$ as $m \to \infty$, where

$$K_{\psi(w)} = \gamma^{-2}[D_w^{-1} + \gamma O(\|\theta\|^2)]$$

is the asymptotic covariance matrix of $\hat{\psi}(w)$. Comparing $K_{\psi(w)}$ with the lower bound for $\Sigma_{\psi(w)}$ as given in Corollary 2.2, we have the desired result.

Let us digress and derive an efficient estimator of ψ_0. This estimator will be part of the efficient estimator of ψ when $w \geqq n$.

THEOREM 5. *$\hat{\psi}_0 = \gamma^{-1}(1_w'\Sigma^{-1}1_w)^{-1}\,{}^*[Z]'1_w$ is an asymptotically efficient estimator of ψ_0 near zero, where $^*[Z] = m^{-1}\sum_{k=1}^m Z^{(k)}$.*

PROOF. From Lemmas 5 and 6 and the central limit theorem, $\mathfrak{L}\{m^{\frac{1}{2}}(\hat{\psi}_0 - \psi_0)\} \to \mathfrak{N}(0, k_0)$ as $m \to \infty$, where $k_0 = \gamma^{-2}[(1_w'\Sigma^{-1}1_w)^{-1} + \gamma O(\|\theta\|^2)]$ is the asymptotic variance of $\hat{\psi}_0$. Comparing k_0 with $\sigma^2_{\psi_0}$ as given by Corollaries 1.1 and 2.1, we have the desired result.

By combining ${}^*[Y(Z)]$ and $\hat{\psi}_0$ we can obtain an efficient estimator of ψ when $w \geqq n$.

THEOREM 6. *If $w \geqq n$, then*

$$\hat{\psi} = \gamma^{-1}[D^{-1} - (D^{-1}1_n1_n'D^{-1}/1_n'D^{-1}1_n)]\ {}^*[Y(Z)] + (D^{-1}1_n/1_n'D^{-1}1_n)\tfrac{1}{2}(\hat{\psi}_0)^2$$

is an asymptotically efficient estimator of ψ near zero.

PROOF. Let $T_m' = ({}^*[Y(Z)], {}^*[Z]'1_w)$. From Lemmas 5 and 6 and the central limit theorem,

$$\mathfrak{L}\left\{m^{\frac{1}{2}}\left[T_m - \gamma\begin{pmatrix} D\psi \\ 1_w'\Sigma^{-1}1_w\,\psi_0\end{pmatrix}\right]\right\} \to \mathfrak{N}(0, \mathbf{C}) \quad \text{as} \quad m \to \infty.$$

Now let $\varphi' = (\varphi_1, \cdots, \varphi_{n+1})$ and define a function g such that

$$g(\varphi) = H\begin{pmatrix}\varphi_1 \\ \vdots \\ \varphi_n\end{pmatrix} + h\varphi_{n+1}^2,$$

$$H = \gamma^{-1}(D^{-1} - (D^{-1}1_n1_n'D^{-1}/1_n'D^{-1}1_n)),$$

$$h = (D^{-1}1_n/1_n'D^{-1}1_n)\tfrac{1}{2}\gamma^{-2}(1_w'\Sigma^{-1}1_w)^{-2}.$$

Thus

$$g(T_m) = H\ {}^*[Y(Z)] + {}^*[Z]'1_w = \hat{\psi}.$$

We refer to the discussion of the delta method in Section 2. We substitute

$$\begin{aligned} g(\varphi^*) &= H\gamma D\psi + h\gamma^2(1_w'\Sigma^{-1}1_w)^2\psi_0^2 \\ &= \psi - (D^{-1}1_n/1_n'D^{-1}1_n)1_n'\psi + (D^{-1}1_n/1_n'D^{-1}1_n)\tfrac{1}{2}\psi_0^2 \\ &= \psi \end{aligned}$$

since $\frac{1}{2}\psi_0^2 = 1_n'\psi$. Then $\mathfrak{L}\{m^{\frac{1}{2}}(\hat{\psi} - \psi)\} \to \mathfrak{N}(0, K_\psi)$ as $m \to \infty$ where the asymptotic covariance of $\hat{\psi}$ is given by,

$$\begin{aligned} K_\psi &= [\partial g/\partial\psi]'\mathbf{C}[\partial g/\partial\psi] \\ &= [H, 2h\gamma(1_w'\Sigma^{-1}1_w)\psi_0]\mathbf{C}\begin{bmatrix} H \\ 2h'\gamma(1_w'\Sigma^{-1}1_w)\psi_0\end{bmatrix} \\ &= \gamma^{-2}[D^{-1} - (D^{-1}1_n1_n'D^{-1}/1_n'D^{-1}1_n) + O(\|\theta\|^2)]. \end{aligned}$$

Comparing K_ψ with Σ_ψ in Corollary 1.2, we see that $\hat{\psi}$ is asymptotically efficient.

In the special case that $w = n - 1$, using Corollary 2.3 we can prove as in the proof of Theorem 6,

THEOREM 7. *If $w = n - 1$, then*

$$\hat{\psi}_n = \tfrac{1}{2}(\hat{\psi}_0)^2 - \gamma^{-1}D_w^{-1}\ {}^*[Y(Z \mid w)]$$

is an asymptotically efficient estimator of ψ_n near zero.

5. Examples. In this section we shall give examples for some of the results in

Sections 3 and 4 for three cases: (A) white or uncorrelated noise, (B) white noise with $n = w = 2$, and (C) Gaussian Markov noise with $n = w$.

EXAMPLE A. If the noise $N(t)$ is white, the covariance matrix Σ is simply the $w \times w$ identity matrix, I_w, since we have made the normalization $\sigma^2_{N(t)} = 1$. Thus $Z = X$ and

$$Y(X)' = (\textstyle\sum_{k=1}^{w} (X_k^2 - 1), \sum_{k=1}^{w-1} X_k X_{k+1}, \cdots, \sum_{k=1}^{w-n+1} X_k X_{k+n-1}).$$

If $w \geqq n$,

$$D = E_0\, Y(X)Y(X)' = \begin{pmatrix} 2w & 0 & \cdots & & 0 \\ 0 & w-1 & & & \\ \vdots & & w-2 & & \vdots \\ & & & \ddots & 0 \\ 0 & \cdots & & 0 & w-n+1 \end{pmatrix}.$$

If $w < n$,

$$D_w = E_0 Y(X \mid w) Y(X \mid w)' = \begin{pmatrix} 2w & 0 & \cdots & 0 \\ 0 & w-1 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}.$$

From Lemma 2 we have,

$$I(\theta) = \gamma^2 J[(w/\psi_0^2) 1_n 1_n' + D + O(\gamma) + O(\|\theta\|^2)]J'$$

since $1_w'\Sigma^{-1}1_w = 1_w'1_w = w$. Thus if we have $w \geqq n$, from Corollary 1.2,

$$\Sigma_\psi = \gamma^{-2}\left[\begin{pmatrix} (2w)^{-1} & & \cdots & 0 \\ 0 & (w-1)^{-1} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & (w-n+1)^{-1} \end{pmatrix}\right.$$
$$- c^{-1}\{(2w)^{-1}, (w-1)^{-1}, \cdots, (w-n+1)^{-1}\}'$$
$$\left.\cdot\{(2w)^{-1}, (w-1)^{-1}, \cdots, (w-n+1)^{-1}\} + O(\gamma) + O(\|\theta\|^2)\right],$$

where $c = (2w)^{-1} + \sum_{k=w-n+1}^{w-1} k^{-1}$.

EXAMPLE B. *Let* $w = n = 2$,

$$I(\theta) = \gamma^2 J_2\left[(2/\psi_0^2)\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix} + O(\gamma) + O(\|\theta\|^2)\right] J_2',$$

where

$$J_2 = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_1 \end{pmatrix}.$$

Thus

$$\Sigma_\psi = \gamma^{-2}\left[\begin{pmatrix} \frac14 & 0 \\ 0 & 1 \end{pmatrix} - \frac45 \begin{pmatrix} \frac1{16} & \frac14 \\ \frac14 & 1 \end{pmatrix} + O(\gamma) + O(\|\theta\|^2)\right]$$

$$= (5\gamma^2)^{-1}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} O(\gamma) + O(\| \theta \|^2),$$

$$\hat{\psi}_0 = \gamma^{-1}(1_w' \,\Sigma^{-1} 1_w)^{-1} \sum_1^w {}^*[Z_i] = (2\gamma)^{-1}\,{}^*[(X_1 + X_2)],$$

$${}^*[X_i] = m^{-1} \sum_{k=1}^m X_i^{(k)}, \qquad i = 1, 2.$$

$$\hat{\psi} = \begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix}$$

where,

$$\hat{\psi}_1 = (5\gamma)^{-1}({}^*[X_1^2 - 1] + {}^*[X_2^2 - 1] - {}^*[X_1X_2] + \tfrac{1}{10}\hat{\psi}_0^2,$$
$$\hat{\psi}_2 = (5\gamma)^{-1}({}^*[X_1X_2] - {}^*[X_1^2 - 1] - {}^*[X_2^2 - 1]) + \tfrac{2}{5}\hat{\psi}_0^2$$

and

$${}^*[X_i^2 - 1] = m^{-1}\textstyle\sum_{k=1}^m (X_i^{(k)2} - 1) \quad \text{for} \quad i = 1, 2,$$
$${}^*[X_1X_2] = m^{-1}\textstyle\sum_{k=1}^m X_1^{(k)}X_2^{(k)}.$$

EXAMPLE C. Let $N(t)$ be Gaussian Markov noise with $EN(t + \tau)N(t) = \rho^{|\tau|}$, $0 < \rho < 1$. Then

$$\Sigma = \begin{pmatrix} 1 & \lambda & \lambda^2 & \lambda^3 & \cdots & \lambda^w \\ \lambda & 1 & \lambda & \lambda^2 & \cdots & \lambda^{w-1} \\ \lambda^2 & \lambda & 1 & \lambda & \cdots & \lambda^{w-2} \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \lambda^w & & \lambda^{w-1} & & \cdots & 1 \end{pmatrix},$$

where $\lambda = \rho^T$ with T being the pulse width.

$$(39) \qquad \Sigma^{-1} = (1 - \lambda^2)^{-1}\begin{pmatrix} 1 & -\lambda & 0 & \cdots & & 0 \\ -\lambda & 1 + \lambda^2 & -\lambda & & & \vdots \\ 0 & -\lambda & 1 + \lambda^2 & -\lambda & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 + \lambda^2 & -\lambda \\ 0 & \cdots & 0 & & -\lambda & 1 \end{pmatrix}$$

Let $w = n > 3$.
We have for $D = (d_{ij})$, from (21b),

$$d_{i,i+3} = 0 \qquad \text{for } i = 1, \cdots, n-3,$$
$$d_{i,i+2} = (1 - \lambda^2)^{-2}(n - i - 1)\lambda^2 \qquad \text{for } i = 2, \cdots, n-2,$$
$$= (1 - \lambda^2)^{-2}2(n - 2)\lambda^2 \qquad \text{for } i = 1$$
$$d_{i,i+1} = -(1 - \lambda^2)^{-2}\, 2\lambda[(n - i) + (n - i - 1)\lambda^2]$$
$$\text{for } i = 2, \cdots, n-2, n-1,$$

$$= -(1-\lambda^2)^{-2}\, 4\lambda[(n-1)+(n-2)\lambda^2] \qquad \text{for } i = 1,$$

$$d_{ii} = (1-\lambda^2)^{-2}\, 2[n+2(2n-3)\lambda^2+(n-2)\lambda^4] \qquad \text{for } i = 1,$$

$$= (1-\lambda^2)^{-2}[(n-1)+(5n-9)\lambda^2+(n-3)\lambda^4] \qquad \text{for } i = 2,$$

$$= (1-\lambda^2)^{-2}[(n-i+1)+4(n-i)\lambda^2+(n-i-1)\lambda^4] \qquad \text{for } i = 3, \cdots, n-1,$$

$$= (1-\lambda^2)^{-2} \qquad \text{for } i = n.$$

6. Estimation of θ. We have shown that $\hat{\psi}$ and $\hat{\psi}_0$ are asymptotically efficient estimators of $\psi' = (\psi_1, \cdots, \psi_n)$ and ψ_0 respectively. Unfortunately the problem of estimating $\theta' = (\theta_1, \cdots, \theta_n)$ is much more difficult. In this section we will discuss informally a method for estimating θ, based upon the estimates $\hat{\psi}$ and $\hat{\psi}_0$. However, this method seems not very practical to implement.

Recall that ψ is defined by the system of second order polynomials in the θ_i's by Equations (1) as follows:

$$\psi_1 = \tfrac{1}{2}\textstyle\sum_1^n \theta_i^2, \qquad \psi_2 = \sum_1^{n-1} \theta_i\theta_{i+1}, \cdots, \psi_n = \theta_1\theta_n$$

with $\psi_0 = \sum_1^n \theta_i$. Remember that θ is in a region such that J, the Jacobian matrix of the transformation, does not vanish and thus J^{-1} exists. However, given $\psi_0, \psi_1, \cdots, \psi_n$ there is not a unique solution to (1).

For simplicity suppose that ψ_0 and ψ are known and $\theta^{(1)}, \theta^{(2)}, \cdots, \theta^{(r)}$ are the corresponding real solutions. Then, let $\beta^{(k)} = X'\theta^{(k)}$. With each observation, depending on the position of the vector θ in the "window" X, each of the r random variables $\beta^{(k)}$ are normally distributed. In fact $\beta^{(k)}$ has mean 0 with probability $1-(n+w-1)\gamma$, or $\mu_{kj} = (S_j\theta)'\theta^{(k)}$ with probability γ for each $j = -w+1, \cdots, n-1$. Moreover given any position, $\beta^{(k)}$ has variance $\theta^{(k)\prime}\Sigma\theta^{(k)}$ which is a constant $\sigma^2 = 2[\psi_1\sigma_{11} + \psi_2\sigma_{12} + \cdots + \psi_n\sigma_{1n}]$ if Σ corresponds to stationary noise, i.e., if $\sigma_{ij} = \sigma_{i+1,j+1}$. The Cauchy-Schwartz inequality informs us that μ_{kj} achieves a unique maximum $2\psi_1$ when $j = 0$ and $\theta^{(k)} = \theta$. Thus if we set a limit $c = 2\psi_1 + \lambda\sigma$, the proportion of observations $\beta^{(k)}$ exceeding c is $p_k(\lambda) = \gamma\sum_j\{1-\Phi[\lambda+2(\psi_1+\mu_{kj})]\} + [1-(n+w-1)\gamma]\{1-\Phi[\lambda+2\psi_1]\}$, where Φ is the cdf of a standard normal variable.

By making λ large enough, the largest $p_k(\lambda)$ corresponds to the k for which $\theta^{(k)} = \theta$. Let the selection procedure be to select that $\theta^{(k)}$ for which the corresponding $\beta^{(k)}$ exceeds c the most times in the m observations. This procedure produces θ with probability approaching one providing (i) λ is sufficiently large depending on m, γ, and μ_{kj}; (ii) $m \to \infty$ and is sufficiently large depending on γ and the μ_{kj}.

However, we do not know ψ and ψ_0, but we have the estimates $\hat{\psi}$ and $\hat{\psi}_0$. From Theorems 4 and 5, $\hat{\psi} \to \psi$ and $\hat{\psi}_0 \to \psi_0$ in probability as $m \to \infty$. Then in (1), for large m, replace ψ by $\hat{\psi}$ and also ψ_0 by $\hat{\psi}_0$, and obtain the real solutions $\hat{\theta}^{(1)}, \cdots, \hat{\theta}^{(r)}$. As $m \to \infty$ each $\hat{\theta}^{(k)} \to \theta^{(k)}$ in probability. The above selection procedure will then select that $\hat{\theta}^{(k)}$ such that $\hat{\theta}^{(k)} \to \theta$ in probability as $m \to \infty$.

As yet we have not given a method for determining the real solutions of (1). As a possible contribution toward finding and counting the $\theta^{(k)}$, we now present a lemma which describes the real θ^* corresponding to (ψ_0, ψ).

LEMMA 7. *Given (ψ_0, ψ), a real solution θ^* of the system of equations given by (1) with $\psi_0 = \sum_1^n \theta_i^*$, has the form*

$$\theta^* = n^{-1}\bar{A}\varphi, \tag{40}$$

where φ is a (complex) vector whose first coordinate is ψ_0, and such that

$$|\varphi_j| = \{2\textstyle\sum_{i=1}^n \psi_i \cos 2\pi[(i-1)(j-1)/n]\}^{\frac{1}{2}}$$

$$\bar{\varphi}_{n-j+2} = \varphi_j \qquad j = 2, 3, \cdots, n$$

and with $\rho = \exp(2\pi i/n)$,

$$A = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ 1 & \rho^2 & \rho^4 & \cdots & \rho^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \rho^{n-1} & \rho^{2(n-1)} & \cdots & \rho^{(n-1)(n-1)} \end{pmatrix}.$$

PROOF. Applying (1) we have

$$|\textstyle\sum_{k=1}^n \theta_k^* \exp[2\pi i(k-1)\omega]|^2 = 2\sum_{k=1}^n \psi_k \cos 2\pi(k-1)\omega. \tag{41}$$

Setting $\omega = (j-1)/n$ in (41), with $j = 1, \cdots, n$ it follows that the absolute value of the jth component of $\varphi = A\theta^*$ is $2\{\sum_{i=1}^n \psi_i \cos 2\pi[(i-1)(j-1)/n]\}^{\frac{1}{2}}$. Observe that $\varphi_1 = \psi_0$. Since the nth roots of unity add to zero, we have $A^{-1} = n^{-1}\bar{A}$. Thus we have (40). Moreover since θ^* must be real, $\theta^* = n^{-1}\bar{A}\varphi = n^{-1}A\bar{\varphi}$. This implies $\varphi = n^{-1}A^2\bar{\varphi}$. We can check that the jth row of A^2 has zeros in all columns except for a one in the $(n-j+2)$nd column. Thus we have that $\varphi_j = \bar{\varphi}_{n-j+2}$, $j \geqq 2$.

For example if $n = 3$,

$$3\theta_1^* = \psi_0 + 2(2\psi_1 - \psi_2 - \psi_3)^{\frac{1}{2}} \cos 2\pi\omega,$$

$$3\theta_2^* = \psi_0 + 2(2\psi_1 - \psi_2 - \psi_3)^{\frac{1}{2}} \cos 2\pi(\omega + \tfrac{1}{3}),$$

$$3\theta_3^* = \psi_0 + 2(2\psi_1 - \psi_2 - \psi_3)^{\frac{1}{2}} \cos 2\pi(\omega - \tfrac{1}{3}),$$

where $0 \leqq \omega < 1$. Observe that $2\psi_1 - \psi_2 - \psi_3 \geqq 0$. However, we do not obtain a solution of (1) for all ω, only for a finite number of ω. By solving (1) for $n = 3$ in a straight-forward manner, it turns out that there are four possible real solutions and thus there are four ω (functions of ψ) which give θ_1^*, θ_2^*, and θ_3^* as real solutions.

We conjecture that there are 2^{n-1} possible real solutions to (1) such that $\psi_0 = \sum_1^n \theta_k^*$.

Finally, we must admit that in the formulation of the statistical problem we have left out certain important aspects of the PPM communication system. The informed receiver B in the system has certain *a priori* information, in addition to θ and γ, at his disposal in order to detect the recurrences of the waveform, measure the intervals between recurrences, and decode the message which A has sent out.

However, we feel that there are three basic ways in which this work is a contribution to the PPM detection problem. First, it gives a statistical formulation in terms of sampling, estimation, and detection of unknown parameters. Secondly, it shows that given the formulation, the basic parameter is the autocorrelation vector of the waveform vector and the basic statistic is the sample autocorrelation. Thirdly, using the tools of large-sample theory, it develops an asymptotically efficient estimator of the discrete autocorrelation. The estimator is fairly simple and it is hoped that it has good properties even if γ is near one and the signal-to-noise ratio is not small.

I wish to express my great gratitude to Professor H. Chernoff of the Department of Statistics, Stanford University for giving of his time, effort, and experience to the development of this research. I would also like to thank Professor M. V. Johns for his help.

REFERENCES

[1] ANDERSON, T. W., (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley, New York.

[2] CHERNOFF, H., (1956). Large-sample theory: parametric case. *Ann. Math. Statist.* **27** 1–22.

[3] CHERNOFF, H., (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24** 586–602.

[4] CRAMÉR, H., (1946). *Mathematical Methods of Statistics.* Princeton Univ. Press, New Jersey.

[5] HINICH, M. J., (1962), A model for a self-adapting filter. *Information and Control* **5** 185–203.

[6] JAKOWATZ, C., SHUEY, R. and WHITE, G. (1960). Adaptive waveform recognition. *Information Theory—Fourth London Symposium* (Colin Cherry, ed.) 317–326. Butterworths, Washington, D.C.

[7] MANN, H. and WALD, A. (1943). On stochastic limit order relationships. *Ann. Math. Statist.* **14** 217–226.

[8] WAINSTEIN, L. and ZHUBAKOV, V. (1962). *Extraction of Signals From Noise.* Prentice-Hall, New Jersey.