A FEASIBILITY STUDY OF AN ADAPTIVE BINARY DETECTOR

A Dissertation

Presented To

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Science in Electrical Engineering

---

by

Miles Melvin Bruce

August 1968

APPROVAL SHEET

This dissertation is submitted in partial fulfillment of the requirements for the degree of Doctor of Science in Electrical Engineering

Author

Approved:

Faculty Advisor

Dean, School of Engineering and Applied Science

August 1968

ACKNOWLEDGMENTS

The author wishes to acknowledge the helpful comments and suggestions of Dr. Clifford M. Siegel in his critique of the work reported here and the encouragement and assistance of NASA - Langley in the performance of this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS

a	mean of function being sampled
A	factor used in estimation of mean
A_c	channel gain
A_0	actual mean of signals intended to be a binary 0
A_1	actual mean of signals intended to be a binary 1
B	factor used in estimation of variance
b	bias
C, C_2	constants
d	distance to optimum location of threshold
$E\left\{x\right\}$	expectation, average, or mean of x
$H(z)$, $H(s)$	transfer function of a system
i, j, k, n	dummy variables
K	decision boundary
m	general term for mean
M, P, Q	constants
$n(t)$, N	noise
n_s	time constant expressed as a number of samples
$p(x)$	probability density of x
$p(x\|y)$	probability density of x given y
q_0	probability of transmitting a binary 0
SNR	signal-to-noise ratio
$s_i(t)$, S_i	waveform representing the ith symbol, $i = 0, 1$

s	complex frequency
s_k	partial sum
T	sampling period
t_c	time constant
u	input to threshold element
$u_1(z)$	unit step function
v^2	general term for variance
$w(t)$, W	input to detector
x_k	kth input data sample
$\hat{x}_k$	kth estimate of the mean of signals representing binary 1's
x	input to decision element
$\hat{X}(z)$	z-transform of $\hat{x}_k$
$\hat{y}_k$	kth estimate of the mean of signals representing binary 0's
z	variable associated with z-transform, $z = e^{sT}$
α	estimation factor used by R. G Brown
β_k	kth estimate of the threshold
$\Delta\beta$	difference of adjacent β's
ϵ	change in matched filter
σ^2	variance of function being sampled
$\hat{\sigma}_k{}^2$	kth estimate of the variance
$\sigma_n{}^2$	variance of noise

ABSTRACT

A system which is capable of acting as an adaptive binary detector is proposed and analyzed. Exponential smoothing is used for estimation of the mean. A technique similar to exponential smoothing is used for estimation of the variance. The system uses the frame synchronization code as a teacher in order to adapt itself to the characteristics of the environment. Decision Directed Measurements are used when the frame synchronization code is not available. The speed and accuracy of the different techniques are derived in this study. The optimum location of the initial conditions of the system is also determined.

CHAPTER I

INTRODUCTION

This dissertation examines the feasibility of receiving binary digital communication signals with an adaptive detector which adjusts its threshold in accordance with the need of a slowly varying or previously unknown environment. The emphasis will be upon an adaptive technique selected primarily for the simplicity of its implementation. The technique will be analyzed to show how it offers improvement over making no change in the threshold location of an optimum detector. A system of this type is needed for use in spacecraft or aircraft systems where simplicity and small size are important characteristics of a system.

For the case of a binary system operating in an environment of additive white Gaussian noise, the optimum Bayes detector consists of two matched filters, a subtracter, and a threshold device. One of the two filters is matched to the binary 0 waveform and the other to the binary 1 waveform. The received signal is applied simultaneously to the two matched filters. A decision, concerning which symbol was transmitted, is made by comparing the difference of the outputs of the two matched filters to a threshold. For many communication systems, conditions are often such that the optimum location of the threshold is fixed and known. However, there are, or may arise, conditions such that the optimum location of the threshold depends on parameters which are neither constant nor known. Conceivable examples are: (1) noise whose mean or variance is subject to change, and (2) matched filters suffering performance deterioration, a plausible condition especially when active filters are used. Such situations present the possibility that a Bayes detector which is optimum for some specific conditions will have a higher probability of error than an adaptive system after a change in the environment.

The proposed system is capable of estimating the optimum location of the threshold when the unknown or variable parameter is the nonzero mean of the noise. In the case of unequal probability of transmission of a binary 0 and binary 1, the proposed system can be used to estimate the variance of the noise, which may be the unknown or variable parameter and is necessary for the calculation of the threshold. When circuit failure or component drift in one of the matched filters causes an optimum detector to locate its threshold at a nonoptimum location, the proposed adaptive system is capable of moving the threshold to reduce the average difference between the actual threshold and the optimum location of the threshold. These situations are discussed in detail in Chapter II.

The adaptive portion of the detector receives as its input the difference of the outputs of the two matched filters. It chooses the threshold location according to calculations upon past values of its input. Estimates are made of the mean of the input to the adaptive detector when a binary 0 is transmitted, of the mean when a binary 1 is transmitted, and of the variance of the input caused by transmission of either (but not both) a binary 0 or binary 1. The location of the threshold is calculated from these estimates. Recursive equations are used to estimate the mean and variance. This paper proposes use of the frame synchronization code as the teacher in a "Learning With Teacher" mode and use of a "Decision Directed Measurement" technique when the frame synchronization code has not been located. The original contributions of this work are (1) the method of estimation of the variance (Chapters IV and V) and (2) the determination of the effect of various parameters on the convergence rate for the Decision Directed Measurement technique operating in conjunction with the estimates of the means (Chapter VI).

In the past few years the literature has contained many reports of work on adaptive detectors. Very few of these, however, contain material pertinent to the system proposed here. Papers by Abramson (Ref. 2) and by Abramson and Braverman (Ref. 5) were among the first to deal with adaptive detectors. These papers, concerned with optimal learning in a random environment, offer estimation techniques which, with modification, are useful in the proposed adaptive detector. A book by Hancock and Wintz (Ref. 7) has chapters pertaining to adaptive receivers and to learning by Decision Directed Measurement. It is generally concerned with optimum estimation methods. In addition, it presents results from computer simulations of the various learning schemes similar to those of Lindenlaub and Mix (Ref. 3). Cooper and Cooper (Ref. 14) investigate a system using learning without supervision and estimated means. Groginsky, Wilson, Middleton, Hancock, Gregg, Millard, and Kurz (Refs. 15 - 17) have described work which has been performed on adaptive detectors for operation in noise of unknown distribution. The present work is concerned with analyzing in detail a specific, simply implemented, method of adapting the threshold in an environment of white, Gaussian noise. Since the resulting system is suboptimum, the available work on optimum systems must be modified and extended.

The mean is estimated by a technique which was mentioned, but not analyzed, by Abramson and Braverman (Ref. 5) for the purpose of tracking a slowly changing parameter. The technique is similar to Kalman filtering (Ref. 18) except that it is designed on the basis of other than minimum mean-square error. Abramson (Ref. 2) gives an analysis of the recursive estimator of the mean for minimum mean-square error. Lindenlaub and Mix (Ref. 3) give the appropriate coefficients for the recursive equation in order to get minimum mean-square error for three specific autocorrelation functions of the slowly changing parameter. Lin and Yau (Ref. 19) discuss the Bayesian approach to the estimators and are concerned with minimum risk. Beine (Ref. 20) discusses an RC averager to estimate the mean which corresponds to the recursive equation used here. The most complete analysis of the recursive estimator is given by Brown (Ref. 4) and <u>his analysis is used</u> as the basis for the present work.

Dale (Ref. 21) discusses an estimation of the variance by a sum of squares. Books by Deutsch and Good (Refs. 22 and 23) treat the area of estimation theory. No appropriate references were found on estimation of the variance under the requirements of the system being investigated. This dissertation shows how an estimate of the variance can be found by modifying a method used by Brown (Ref. 4) to estimate means. The method will be analyzed to derive its accuracy.

A search of the literature for an analysis of a system using Decision Directed Measurement techniques in conjunction with a recursive estimator revealed none completely suited for detailed evaluation of the system proposed here. Lindenlaub and Mix (Ref. 3) and Hancock and Mix (Ref. 10) use Monte Carlo methods to check the convergence of several learning methods. Henry Scudder (Ref. 6) derives the asymptotic probability of error for a DDM system which makes estimates only on the binary 1 signals. Patrick and Costello (Ref. 1) derive the asymptotic probability of error for estimates on both signals but use the sample mean with an infinite number of samples as the estimator. The present work develops a computer program (Chapter VI) for performing a numerical convolution of the probability densities involved in the system and makes possible the investigation of how system operation is affected by various parameters such as signal-to-noise ratio and initial estimates of the means.

Spragins (Ref. 24) presents a review of the methods of "Learning Without Teacher" and points out that an optimum method of "Learning Without Teacher" is impractical. The present system is offered (Chapter II), from among the many suboptimum solutions to the problem, as one which is practical and simple to implement.

No construction of hardware has been performed as a part of this study. Computer simulation has been used for any work requiring investigation from an experimental viewpoint. The computer has also been used as an aid in analyzing some areas which were difficult to evaluate in closed form. Since it is anticipated that this system would operate in real time and might possibly be used in a spacecraft or aircraft, it is desirable that the system be potentially fast and small. This requirement has influenced the method of operation selected for the system.

At the present time, there is no universally accepted test for determining whether a system is an adaptive system or a learning system. The system being investigated here probably fits under most definitions of an adaptive system rather than that of a learning system. This system has a specific procedure for adjusting the threshold as a function of the past history of the input signal and does not attempt to recognize situations that it has previously encountered.

CHAPTER II

DISCUSSION OF TOTAL SYSTEM

A detector that is optimum in the Bayes sense for binary signals in additive, white, Gaussian noise (Ref. 7, p. 49) is shown in Figure 1, where

$w(t)$ = input to the detector = $A_c s_1(t) + n(t)$

A_c = channel gain

$s_1(t)$ = signal representing a binary 1

$s_o(t)$ = signal representing a binary 0

$n(t)$ = noise

b = optimum bias

This optimum detector decides that a binary 1 was sent if $u \geq 0$ and that a binary 0 was sent if $u < 0$ by use of

$$u = W^T\left(S_1 - S_o\right) - b \tag{2-1}$$

where

$$b = \frac{\sigma_n^2}{A_c} \ln K + \frac{1}{2} A_c\left(S_1^T S_1 - S_o^T S_o\right) + \bar{N}^T\left(S_1 - S_o\right) \tag{2-2}$$

In these equations the capital letters are matrices (Ref. 7, pp. 231-243) representing the functions of time shown in Figure 1. The superscript T indicates the transpose of the matrix. The factor K is a decision boundary which appears in the Bayes calculations and is

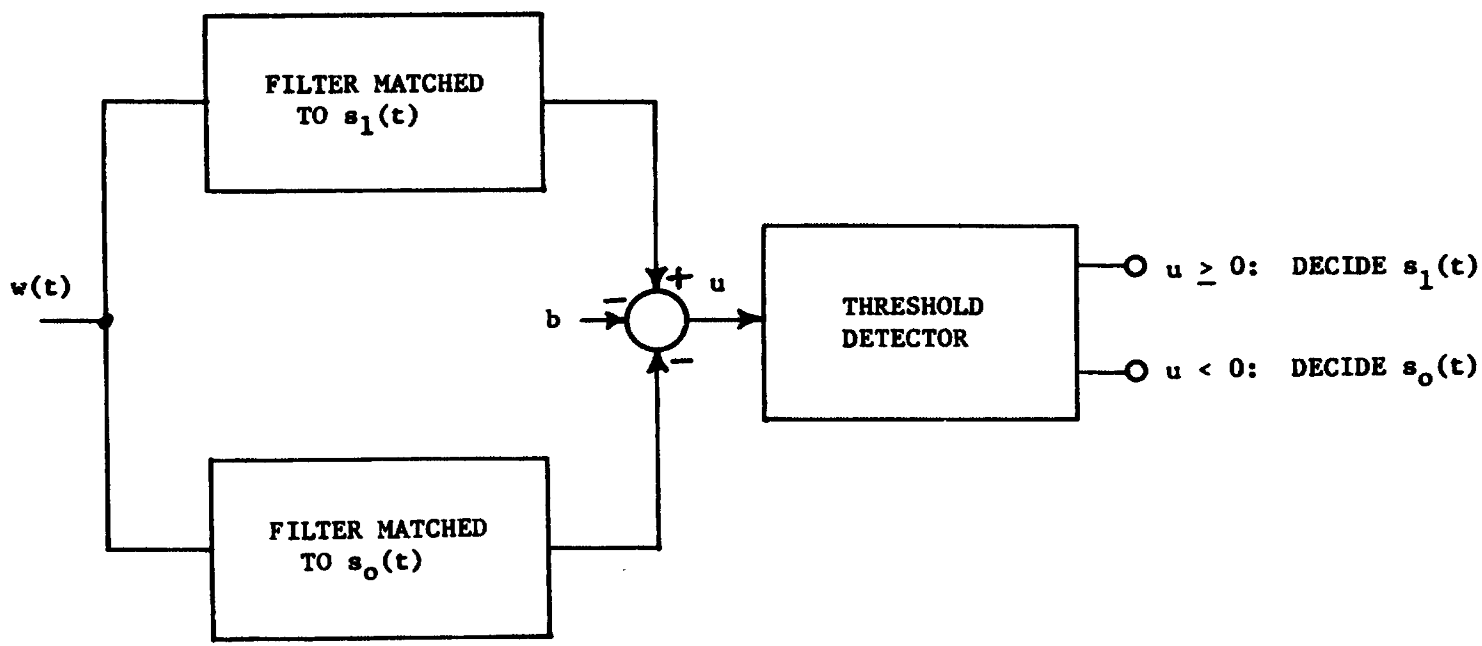

Figure 1.- Structure of binary detector.

$$K = \frac{q_o}{1 - q_o}$$

for the cost of an error in detecting a binary 1 equal to the cost of an error in detecting a binary 0 where q_o is the probability of transmitting a binary 0. The mean of the noise is represented by $\bar{N}$, and σ_n^2 is the variance of the noise. It is consistent with (2 - 1) to say that the detector decides that a binary 1 or binary 0 was transmitted by determining if $W^T(S_1 - S_o)$ is greater than or less than the threshold, b. The adaptive detector employs as its threshold an estimation, β, of the optimum bias based on past history of the difference

$$X = W^T(S_1 - S_o)$$

The estimated threshold, β, is obtained by first averaging the estimate of the mean of X when a binary 1 is transmitted and the estimate of the mean when a binary 0 is transmitted, and then adding the ratio $\frac{\sigma_n^2}{A_c}$ (the estimated variance of the noise divided by the estimated channel gain) multiplied by $ln\ K$; thus

$$\beta = \frac{1}{2}\left[E\left\{X|W = A_cS_1 + N\right\} + E\left\{X|W = A_cS_o + N\right\}\right] + \frac{\sigma_n^2}{A_c}\ ln\ K \quad (2\text{-}3)$$

The only characteristics of the system which must be known are K, the waveforms $s_1(t)$ and $s_o(t)$, and the fact that the noise is additive, white and Gaussian. The mean and variance of the noise and the channel gain can all be unknown or slowly-varying parameters.

In order to show that β corresponds to the optimum bias for perfect estimates of the means and variance, the expression for β (2-3) is rearranged to yield

$$\beta = \frac{1}{2}\left[E\left\{\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)\right\} + E\left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}\right] + \frac{\sigma_n^2}{A_c} \ln K$$

$$= \frac{1}{2}\left[A_c S_1^T S_1 - A_c S_1^T S_o + \bar{N}^T S_1 - \bar{N}^T S_o + A_c S_1^T S_1 - A_c S_o^T S_o + \bar{N}^T S_1 - \bar{N}^T S_o\right] + \frac{\sigma_n^2}{A_c} \ln K$$

$$= \frac{1}{2} A_c \left[S_1^T S_1 - S_o^T S_o\right] + \bar{N}^T\left(S_1 - S_o\right) + \frac{\sigma_n^2}{A_c} \ln K \qquad (2\text{-}4)$$

which agrees with (2-2). Since the estimations are not perfect, the adaptive detector is actually a suboptimum receiver.

The ratio $\frac{\sigma_n^2}{A_c}$ in equation (2-4) is shown in Appendix I to be estimated by

$$\frac{\sigma_n^2}{A_c} = \frac{\text{variance of } \left(A_c S_o + N\right)^T \left(S_1 - S_o\right)}{E\left\{\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)\right\} - E\left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}} \qquad (2\text{-}5)$$

The method of estimation of the means and of the variance are discussed in Chapters III and IV, respectively. The above analysis shows that the adaptive detector derives an estimate of the optimum location of the threshold when the channel gain and mean or variance of the noise are unknown. The value of the decision boundary K must be known for this situation.

For the special case of $K = 1$, the adaptive detector can also be used to increase the reliability of the receiver by making it possible to relocate the threshold in the event of a degradation in one of the matched filters in an optimum detector. The degradation would cause the optimum detector to be operating with the threshold at a nonoptimum location. The equation implemented by the degraded optimum detector is

$$u = W^T\left(S_1 + \epsilon - S_o\right) - b$$

where ϵ is the change in one of the matched filters. The threshold of the degraded optimum detector differs from the optimum location by an amount, d, given by

$$d = W^T\epsilon = \left(A_c S_i + N\right)^T\epsilon \quad ; \quad i = 0,\ 1 \tag{2-6}$$

In order to restore the threshold to the optimum location, it is necessary to determine ϵ and subtract it from the term $\left(S_1 + \epsilon - S_o\right)$. The adaptive system can be used to improve this situation when it is not practical or possible to determine ϵ and to make the necessary adjustments. This specific adaptive system is limited to situations of $K = 1$ for degradations in the matched filters because the estimation techniques used here do not give an accurate estimate of the variance when the input is degraded. Other estimation methods may permit this system to be used for $K \neq 1$, but they have not been investigated here.

The following analysis is included to show that the adaptive detector locates its threshold at

$$\beta = b + E\left\{d\right\}$$

for perfect estimations of the means. For calculation of β according to (2-3), with $K = 1$, the input is

$$X = W^T\left(S_1 + \epsilon - S_o\right)$$

and the estimates of the means are given by

$$E\left\{X \middle| W = A_cS_1 + N\right\} = E\left\{\left(A_cS_1 + N\right)^T \left(S_1 + \epsilon - S_o\right)\right\}$$

$$= A_cS_1{}^TS_1 - A_cS_1{}^TS_o + \overline{N}^T\left(S_1 - S_o\right) + \left(A_cS_1 + \overline{N}\right)^T\epsilon \qquad (2\text{-}7a)$$

$$E\left\{X \middle| W = A_cS_o + N\right\} = E\left\{\left(A_cS_o + N\right)^T \left(S_1 + \epsilon - S_o\right)\right\}$$

$$= A_cS_o{}^TS_1 - A_cS_o{}^TS_o + \overline{N}^T\left(S_1 - S_o\right) + \left(A_cS_o + 2\overline{N}\right)^T\epsilon \qquad (2\text{-}7b)$$

The threshold according to (2-3) is

$$\beta = \frac{1}{2}A_c\left(S_1{}^T S_1 - S_o{}^T S_o\right) + \overline{N}^T\left(S_1 - S_o\right) + \frac{1}{2}\left[A_cS_1 + A_cS_o + 2\overline{N}\right]^T\epsilon \qquad (2\text{-}8)$$

This differs from the optimum threshold given by (2-2) for $K = 1$ by the amount of the last term which is shown to be the expected value of d, the distance to the optimum location of the threshold. For equally probable signals, $K = 1$, the expected value of d is obtained from (2-6) to yield

$$E\left\{d\right\} = \frac{1}{2}\left[E\left\{\left(A_cS_1 + N\right)^T\epsilon\right\} + E\left\{\left(A_cS_o + N\right)^T\epsilon\right\}\right] = \frac{1}{2}\left[A_cS_1 + A_cS_o + 2\bar{N}\right]^T\epsilon$$

This corresponds to the last term of equation (2-8).

The degraded optimum detector implements

$$u = W^T\left(S_1 - S_o\right) - b + d$$

The adaptive detector, with perfect estimation, implements

$$u = W^T\left(S_1 - S_o\right) - b + d - E\left\{d\right\}$$

Since $d - E\left\{d\right\}$ is less than d on the average, the adaptive detector is closer to the optimum location of the threshold than the degraded optimum detector. This results in a suboptimum detector but would offer improvement for sufficiently large values of d over the continued use of the degraded optimum detector.

The input to the adaptive detector has been called X, which represents a matrix. In practice, the input is the sampled output of the subtracter at the end of a bit time. The input to the adaptive detector is, therefore a sequence of values, x_i; each value represents the processing of a single bit by the matched filters. Bit synchronization is assumed for this study.

For proper operation of the adaptive detector, it it necessary to perform separate estimates of the mean of x when a binary 0 has been transmitted and of the mean of x when a binary 1 has been transmitted. For this to be accomplished, it is necessary to know whether the transmitted signal was intended to be a binary 0 or binary 1 in order to know which estimator to update.

This system operates on digital communication systems in which the data are transmitted in serial fashion over a single channel. In a system of this type a known sequence of bits, called the frame synchronization code, is normally inserted into the sequence of data bits in order to synchronize the decoder located at the receiver with the encoder located at the transmitter. After synchronization has been obtained, the proposed system uses the fact that during transmission of the frame synchronization code, the correct decision is known. The system knows from the synchronization code if the received signal was intended to be a binary 0 or a binary 1 and updates the appropriate estimator. This is a form of Abramson's "Learning With Teacher" (Ref. 2) where the frame synchronization code is the teacher.

Since the "Learning With Teacher" scheme cannot be used during the transmission of actual data, the system may either cease its estimation until the appearance of the next synchronization code or may use some form of "Learning Without Teacher." The operation of the system before the synchronization code has been located also requires the use of "Learning Without Teacher" since the correct decision is not known. The proposed system uses a Decision Directed Measurement (DDM) technique similar to that discussed in Reference 3, page 13. In DDM, a decision is made with all available information and assumed to be correct. The decision determines which estimator is to be updated. Incorrect decisions are possible in DDM and cause the system to converge slower than a "Learning With Teacher" system. There are other techniques of "Learning Without Teacher," but the DDM technique appears to offer the best combination of convergence rate and implementation simplicity as shown by Lindenlaub and Mix (Ref. 3, pp. 13-37).

The DDM technique operates in such a manner that the expected value of the estimate of the mean when a binary 1 is transmitted moves to the mean of all signals above the threshold. This estimate of the mean is not unbiased when a binary 1 is transmitted because some of the signals above the threshold are due to the transmission of a binary 0 and some of the signals due to the transmission of a binary 1 fall below the threshold. Therefore, it is necessary to use the DDM technique only to move the threshold so that enough correct decisions can be made to enable the frame synchronization code to be located. The DDM scheme should not be used after the frame synchronization code is located. The "Learning With Teacher" scheme is required to give an unbiased estimate of the conditional means.

The analysis of the estimation techniques can be performed without knowing if the inputs are coming from the "Learning With Teacher" scheme or from the DDM scheme. The estimators are only required to perform computations on the data given to them. The "Learning With Teacher" or DDM technique performs the function of deciding which estimator receives each individual input sample. The convergence and accuracy of the estimations are derived as functions of the input data, and this analysis applies for both the "Learning With Teacher" and DDM cases.

Three specific problems associated with this adaptive detector are analyzed and are discussed here. These three are:

1. Estimation of the Mean
2. Estimation of the Variance
3. Operation of the system when controlled by the Decision Directed Measurement technique.

Both the estimation of the mean and the estimation of variance are performed by recursive equations. The estimate of the mean is accomplished by use of exponential smoothing (Chapter III). A technique similar to the exponential smoothing method is used for the estimate of the variance (Chapter IV). Computer simulation is used to study the operation of the Decision Directed Measurement technique and the estimation methods (Chapter VI).

Figure 2 shows a block diagram of the system discussed here. The input to this system is a sequence of random values; each value represents $W^T(S_1 - S_0)$ for a single bit. The threshold decision element examines each random value. If the random value is above the threshold, the output is the decision that the bit is a binary 1. If the value is below the threshold, the output is a binary 0. The threshold computer determines an estimated value of the threshold by calculating the terms in (2-3) from the two estimates of the means and from the estimate of the variance. The resulting value of the threshold is furnished to the threshold decision element. This is equivalent to

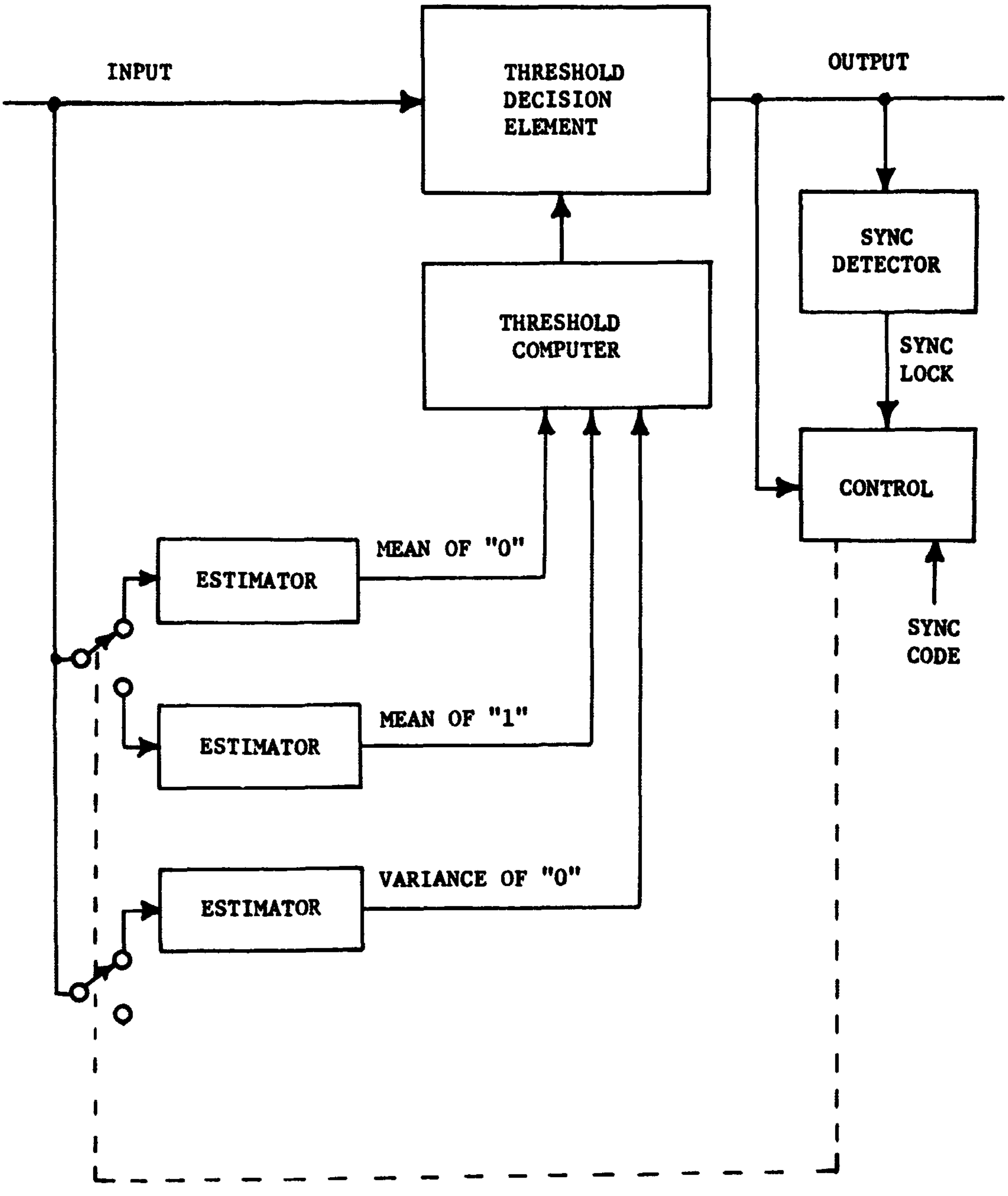

Figure 2.- Adaptive binary detector.

estimating the bias, b (2-2). The estimators use the input data to make estimates of three of its properties. The control block determines which estimators are updated. The sync detector looks for the frame synchronization code in the output of the threshold decision element. If the sync detector has properly located the frame synchronization code, the control block is directed to use the frame synchronization code to determine which set of estimators to update. If the sync detector has not located the frame synchronization code, the control block is directed to use the decisions of the threshold decision element to determine which set of estimators to update. This latter case is called Decision Directed Measurement.

The properties of interest in this investigation are accuracy, speed of response to step changes in the input, complexity of equipment required to implement the system, and required calculation time. Techniques have been selected which appear to require simple implementation and which have potentially low calculation times. In this investigation, the accuracy and speed of response are found to be variables which must be traded off against each other.

CHAPTER III

ESTIMATION OF THE MEAN

Description of the Method

One of the tasks which the adaptive detector must perform is the estimation of the mean of the received signals representing a binary 0 and representing a binary 1. These signals are $\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)$ and $\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)$, respectively. For this reason, there are two estimators of the mean in the adaptive detector; logic circuits determine which of the two is updated. The estimators are identical, hence, an analysis of one can be extended to the other. The input to the estimator is a sequence of values which represents the received data. This sequence of values consists of the differences of the outputs of the two matched filters at the ends of transmissions of successive bits. As previously mentioned, the technique for estimation should be simple, accurate, and capable of reacting quickly to abrupt changes in the characteristics of the input.

The technique selected for estimation of the mean is exponential smoothing which was introduced and analyzed by Brown (Ref. 4). He uses the following equation for the estimation of the mean with α written for $(1 - A)$:

$$\hat{x}_k = A\hat{x}_{k-1} + (1 - A)x_k; \; k = 1,2,3 \tag{3-1}$$

where

$\hat{x}_k$ = kth estimate of the mean

x_k = kth data sample

A = recursive constant; $0 < A < 1.0$

The input to the estimation equation is a series of values, x_i, which are considered to be samples of a random function. The task of the estimator is to estimate the mean, $E\{x\}$, of the random function, where the mean is defined by

$$E\{x\} = \int_{-\infty}^{\infty} xp(x)\, dx$$

where $p(x)$ is the probability density of x. Requirements for use of the above definition of the mean are given on page 64 of Reference 8.

The estimation begins with an initial guess, $\hat{x}_o$, of the mean of x. This value is used in conjunction with the first data point, x_1, to compute the next estimate, $\hat{x}_1$, of the mean. This process is continued as each succeeding input sample is applied to the estimation equation. Calculations are simple and quickly made because only two multiplications and one addition are required. Storage is needed only for $\hat{x}_{k-1}$.

Analysis of the Method

Since the input, x_i, to the estimator is a random variable, the estimated mean also is a random variable. The mean and variance of the estimated mean are used to determine the accuracy of the estimation. For proper operation of the estimator, the mean of $\hat{x}_k$ should be asymptotically unbiased (Ref. 8, p. 463), that is, the mean of $\hat{x}_k$

should approach the actual mean of the input as the number of samples handled approaches infinity. The variance of $\hat{x}_k$ is an indication of the error of the estimate and should be as small as possible. Unfortunately, small variance of the estimated mean is achieved at the expense of reaction time to abrupt changes in the mean of the input as will be shown in Figure 3.

The mean of $\hat{x}_k$ is derived as a function of k in order to show that the estimation is asymptotically unbiased. For a stationary input, Appendix II shows that the mean of $\hat{x}_k$ is

$$E\left\{\hat{x}_k\right\} = A^k \hat{x}_o + a(1 - A)\left(1 + A^2 + \cdots + A^{k-1}\right)$$

where a is the mean of the input data, x_i. Since $|A| < 1.0$,

$$\lim_{k \to \infty} A^k \hat{x}_o = 0$$

and

$$\lim_{k \to \infty} \left(1 + A + A^2 + \cdots + A^{k-1}\right) = \frac{1}{1 - A}$$

Therefore,

$$\lim_{k \to \infty} E\left\{\hat{x}_k\right\} = a \tag{3-2}$$

This demonstrates that the estimation of the mean by the exponential smoothing method is asymptotically unbiased. For nonstationary inputs, the analysis of the estimation technique still applies if the input changes very slowly with respect to the response time of the estimation.

The variance of $\hat{x}_k$ is also derived in Appendix II and is found to be

$$\text{variance of } \hat{x}_k = (1 - A)^2 \sigma^2 \sum_{i=o}^{k-1} \left(A^2\right)^i$$

where σ^2 is the variance of the input, x_i. The limiting value of the variance is found to be

$$\lim_{k \to \infty} \text{variance of } \hat{x}_k = (1 - A)^2 \sigma^2 \frac{1}{1 - A^2} = \left(\frac{1 - A}{1 + A}\right) \sigma^2 \qquad (3\text{-}3)$$

Some observations can be made at this point. Since the estimation equation is linear, the estimate of the mean is Gaussian if the input data are Gaussian (Ref. 9). The estimation technique is not limited to input data with a Gaussian distribution and should be able to operate on any distribution for which the mean exists. However, the probability density of $\hat{x}_k$ would be very difficult to calculate for distributions other than Gaussian. The Cauchy distribution is an example of a distribution which could not be used here since none of its moments exist (Ref. 8, p. 157). If the characteristics of the data are not time-varying and if the estimation began with an initial estimate $\hat{x}_o$, the actual variance of $\hat{x}_k$ is always less than the asymptotic value. This can be seen from the fact that only positive terms are added to the variance as k increases. The limiting value of the variance of $\hat{x}_k$ can be made as small as desired by making A closer to 1.0. Since the variance of the estimate is an indication of the error, the estimator can be made as accurate as desired.

In order to determine the <u>speed of response</u> to a step input, it is necessary to determine the transfer function of the estimation technique. The transfer function is determined in Appendix II and is found to be

$$H(z) = \frac{(1 - A)z}{z - A}$$

The time constant associated with this transfer function is

$$t_c = \frac{-T}{\ln A}$$

The time constant also can be expressed as the number of samples

$$n_s = \frac{-1}{\ln A}$$

This value, n_s, is positive since $0 < A < 1.0$ which means that the logarithm of A is negative. It has been shown in Equation (3-3) that A should be as near 1.0 as possible in order to reduce the variance of the estimate. However, A should be near zero in order to reduce the time required to respond to a change in the input characteristics. A potential user of this system is required to make a tradeoff study in order to determine the optimum value of A for his particular application. The variance of $\hat{x}_k$ and the time constant, n_s, are plotted in Figure 3 to aid the user in his selection of A.

An example which illustrates one procedure for selecting A follows: Due to the application of an adaptive system, it is required that the estimate of the mean be able to react in not more than 100

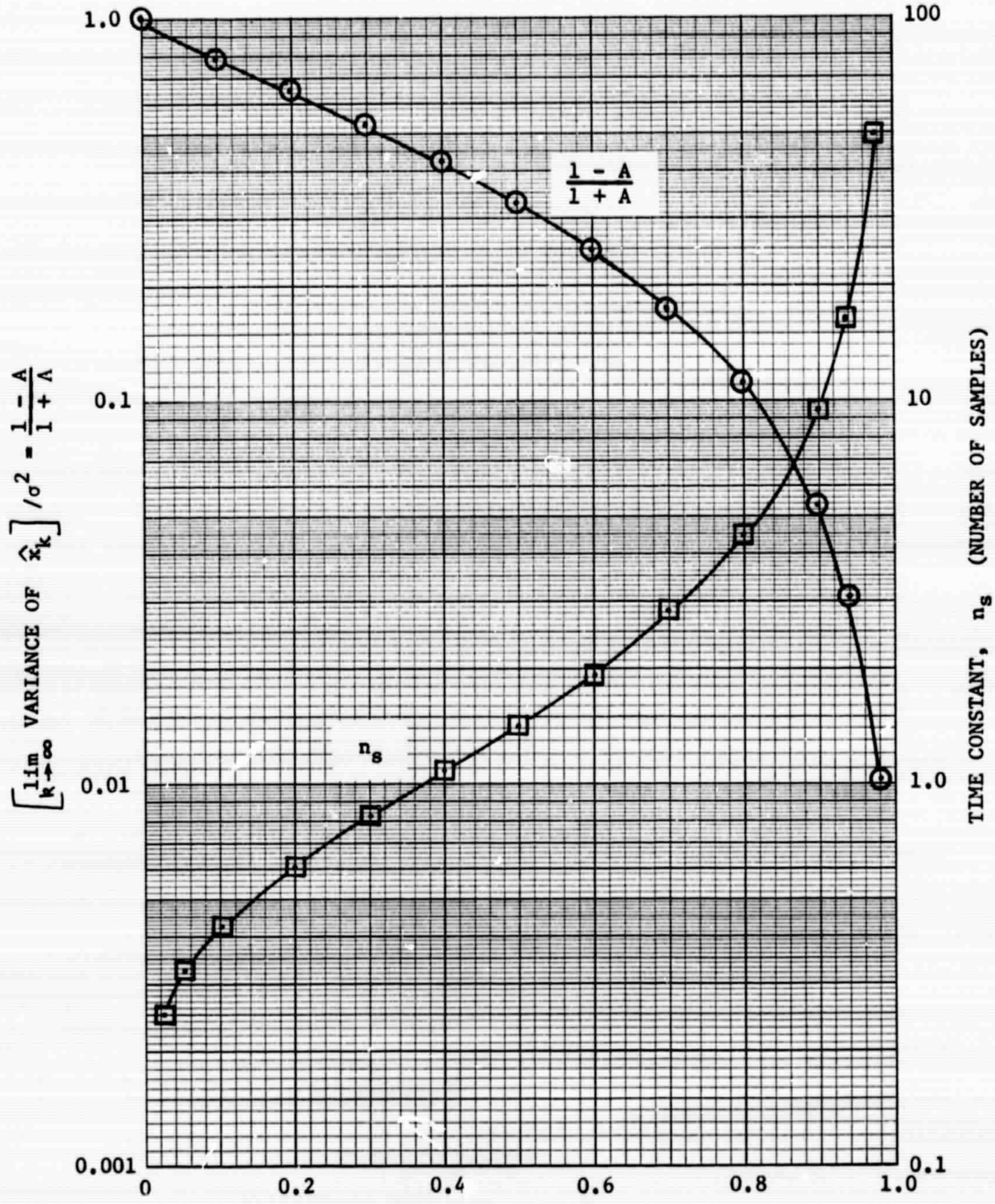

Figure 3.- Variance and time constant of the estimation of the mean.

samples to a step change in the mean of the data coming into the estimator. The accuracy of the estimator should remain as high as possible. After five time constants, the mean of the estimated mean has moved to within 99 per cent of the final value. This means that the estimator should have a time constant of 20 samples. Choosing $A = 0.95$ gives a time constant of 19.5. For this choice of A, the asymptotic variance of the estimated mean is 0.0256 times the variance of the input data.

Two other techniques which were considered for estimation of the mean are a sliding window and a running calculation of the sample mean of all previous samples. The running calculation of the sample mean is a calculation of the sample mean using all previous samples. It is not practical since it must take into account the <u>number</u> of previous samples, which could possibly exceed the capacity of the computer used for the computation during long periods of operation. It is also slow to react to changes in the data characteristics if the number of previous samples is very large. The sliding window method uses a fixed number of samples and computes their sample mean. The variance of the sample mean is $\frac{\sigma^2}{n}$ (Ref. 11, p. 246) where σ^2 is the variance of the input and n is the number of samples in the window. In order to have the same variance of estimated mean for exponential smoothing, Equation (3-3), and a sliding window,

$$\frac{\sigma^2}{n} = \left(\frac{1 - A}{1 + A}\right) \sigma^2$$

and

$$n = \frac{1 + A}{1 - A}$$

From consideration of the general range of accuracy and speed requirements, A will probably be

$$0.9 < A < 1.0$$

The following table shows the value of n required as a function of A in order to enable the sliding window system to have a variance equal to that of the recursive system:

A	n_s	n
0.9	9.491	19
.95	19.497	39
.99	99.502	199
.999	1000	1999

The table also shows the corresponding time constant, n_s, of the recursive equation. All effects of the previous characteristics of the input disappear from the sliding window technique when n samples have been processed after the step change; meanwhile the exponential smoothing technique has undergone approximately two time constants. If the initial estimate for both estimators is zero and the final estimate is 1.0, the expected value of the recursive estimate moves as $\left(1 - e^{-t/t_c}\right)$. The expected value of the sample mean of the sliding window moves linearily between zero and 1.0. For the recursive estimator, the integral of the difference between the final value and the actual value is

$$\text{ERROR} = \int_{0}^{\infty} \left[1 - \left(1 - e^{-t/t_c}\right)\right] dt$$

$$= \int_{0}^{\infty} \left(e^{-t/t_c}\right) dt$$

$$= t_c$$

The same calculation for the sample mean of a sliding window is

$$\text{ERROR} = \int_{0}^{2t_c} \left[1 - \frac{t}{2t_c}\right] dt$$

$$= t - \frac{t^2}{4t_c} \Bigg|_{0}^{2t_c}$$

$$= 2t_c - t_c$$

$$= t_c$$

This shows that the integral of the difference between the final value and the actual expected value of the mean is the same for the recursive estimator and the sample mean of a sliding window. However, the amount of equipment required for implementation of a sliding window technique due to the requirement of storing and labeling hundreds or thousands of previous samples removed the sliding window technique from further consideration in this application.

For updating the estimators during the frame synchronization code, it may be practical to use the sample mean of the synchronization code block as the estimator. However, this is not advantageous when operating in the DDM mode because if blocks of n samples are used in this mode to calculate the sample mean, the system must use the old estimate for n samples while waiting for a new estimate. This results in additional error as will be shown below. The recursive estimator with variance equal to that of the sample mean has moved two time constants closer to the new location than the sample mean during the n samples. The expected value of the sample mean of a block of n samples remains at zero for a time equal to $2t_c$ and then jumps to 1.0. The integral of the difference for the sample mean of a block of n samples is

$$\text{ERROR} = \int_0^{2t_c} (1 - 0)\, dt$$

$$= 2t_c$$

This shows that the use of blocks of n samples yields more error than use of either a recursive estimator or a sliding window. The same results are obtained for any location of the step change with respect to the location of the block of n samples. The recursive estimator has an additional advantage over the calculation of the sample mean since the accuracy of the recursive estimator can be changed by simply changing a single constant. A change in the number of samples used is required to change the accuracy of the sample mean.

CHAPTER IV

ESTIMATION OF THE VARIANCE

A technique similar to the estimation of the mean is used for the estimation of the variance. The equation used is

$$\hat{\sigma}_k^{\,2} = B\hat{\sigma}_{k-1}^{\,2} + \frac{(1 - B)}{C}\left(x_k - \hat{x}_k\right)^2 \tag{4-1}$$

where

$\hat{\sigma}_k^{\,2}$ is the kth estimate of the variance

$\hat{x}_k$ is the kth estimate of the mean

x_k is the kth data sample

B is a constant and $B < 1.0$

C is a constant

This equation uses an initial estimate of the variance, $\hat{\sigma}_o^{\,2}$, plus the received data sample and the present estimate of the mean in order to make a new estimate of the variance. The constant C will be determined later and is required to make this technique converge to the proper value, that is, to remove the bias of this estimate.

If $\hat{x}_k$ is replaced by its equivalent given by (3-1), a more usable form of (4-1) is obtained for $\hat{\sigma}_k^{\,2}$:

$$\hat{\sigma}_k^{\,2} = B\hat{\sigma}_{k-1}^{\,2} + \frac{(1 - B)}{C}\left[x_k - A\hat{x}_{k-1} - x_k + Ax_k\right]^2$$

$$= B\hat{\sigma}_{k-1}^{\,2} + \frac{A^2(1 - B)}{C}\left[x_k - \hat{x}_{k-1}\right]^2 \tag{4-2}$$

This equation for $\hat{\sigma}_k^{\ 2}$ leads to faster calculation than the preceding one since it employs $\hat{x}_{k-1}$ instead of $\hat{x}_k$. This calculation can be performed in parallel with the kth estimate of the mean instead of having to wait until $\hat{x}_k$ is calculated.

As was done in the case of the estimate of the mean, the mean and variance of the estimate of the variance are determined. The constant, C, will be selected to force the mean of the estimate of the variance to converge to the actual variance of the data being sampled. The variance of the estimate serves as an indication of the average error of the estimate.

The mean of the estimated variance is calculated for several values of k. Enough terms are used in order to recognize the series being generated. The general expression is then written, and the limiting value is determined. Thus,

$$\hat{\sigma}_1^{\ 2} = B\hat{\sigma}_o^{\ 2} + \frac{A^2(1 - B)}{C}\left[x_1 - \hat{x}_o\right]^2$$

and

$$\begin{aligned} E\left\{\hat{\sigma}_1^{\ 2}\right\} &= B\hat{\sigma}_o^{\ 2} + \frac{A^2(1 - B)}{C} E\left\{x_1^{\ 2} - 2x_1\hat{x}_o + \hat{x}_o^{\ 2}\right\} \\ &= B\hat{\sigma}_o^{\ 2} + \frac{A^2(1 - B)}{C}\left[E\left\{x_1^{\ 2}\right\} - 2\hat{x}_o E\left\{x_1\right\} + \hat{x}_o^{\ 2}\right] \\ &= B\hat{\sigma}_o^{\ 2} + \frac{A^2(1 - B)}{C}\left[a^2 + \sigma^2 - 2a\hat{x}_o + \hat{x}_o^{\ 2}\right] \\ &= B\hat{\sigma}_o^{\ 2} + \frac{A^2(1 - B)}{C}\left[\sigma^2 + \left(a - \hat{x}_o\right)^2\right] \end{aligned}$$

The same techniques are used to calculate $E\left\{\hat{\sigma}_2^{\,2}\right\}$, which gives

$$\hat{\sigma}_2^{\,2} = B\hat{\sigma}_1^{\,2} + \frac{A^2(1 - B)}{C}\left[x_2 - \hat{x}_1\right]^2 = B\left[B\hat{\sigma}_o^{\,2} + \frac{A^2(1 - B)}{C}\left(x_1 - \hat{x}_o\right)^2\right]$$

$$+ \frac{A^2(1 - B)}{C}\left[x_2 - A\hat{x}_o - (1 - A)x_1\right]^2$$

$$= B^2\hat{\sigma}_o^{\,2} + \frac{A^2B(1 - B)}{C}\left[x_1^{\,2} - 2x_1\hat{x}_o + \hat{x}_o^{\,2}\right] + \frac{A^2(1 - B)}{C}\left[x_2^{\,2} + A^2\hat{x}_o^{\,2}\right.$$

$$\left. + (1 - A)^2x_1^{\,2} - 2Ax_2\hat{x}_o - 2(1 - A)x_1x_2 + 2A(1 - A)x_1\hat{x}_o\right]$$

and

$$E\left\{\hat{\sigma}_2^{\,2}\right\} = B^2\hat{\sigma}_o^{\,2} + \frac{A^2B(1 - B)}{C}\left[E\left\{x_1^{\,2}\right\} - 2\hat{x}_oE\left\{x_1\right\} + \hat{x}_o^{\,2}\right]$$

$$+ \frac{A^2(1 - B)}{C}\left[E\left\{x_2^{\,2}\right\} + A^2\hat{x}_o^{\,2} + (1 - A)^2E\left\{x_1^{\,2}\right\} - 2A\hat{x}_oE\left\{x_2\right\}\right.$$

$$\left. - 2(1 - A)E\left\{x_1x_2\right\} + 2A(1 - A)\hat{x}_oE\left\{x_1\right\}\right]$$

The data samples are considered to be independent so that

$$E\left\{x_ix_j\right\} = E\left\{x_i\right\}E\left\{x_j\right\} \quad \text{for} \quad i \neq j$$

This yields

$$E\left\{\hat{\sigma}_2^{\,2}\right\} = B^2\hat{\sigma}_o^{\,2} + \frac{A^2B(1 - B)}{C}\left[a^2 + \sigma^2 - 2a\hat{x}_o + \hat{x}_o^{\,2}\right] + \frac{A^2(1 - B)}{C}\left[a^2\right.$$

$$+ \sigma^2 + A^2\hat{x}_o^{\,2} + (1 - A)^2\left(a^2 + \sigma^2\right) - 2Aa\hat{x}_o - 2(1 - A)a^2$$

$$\left. + 2A(1 - A)a\hat{x}_o\right]$$

$$= B^2\hat{\sigma}_o{}^2 + \frac{A^2(1 - B)}{C}\Big[(1 + B)\sigma^2 + (1 - A)^2\sigma^2$$

$$+ \left(A^2 + B\right)\left(a - \hat{x}_o\right)^2\Big]$$

If these same techniques are used, the mean value of $\hat{\sigma}_3{}^2$ and $\hat{\sigma}_4{}^2$ can be determined:

$$E\left\{\hat{\sigma}_3{}^2\right\} = B^3\hat{\sigma}_o{}^2 + \frac{A^2(1 - B)}{C}\Big[\left(1 + B + B^2\right)\sigma^2 + \left(1 + A^2 + B\right)(1 - A)^2\sigma^2$$

$$+ \left(A^4 + A^2B + B^2\right)\left(a - \hat{x}_o\right)^2\Big]$$

$$E\left\{\hat{\sigma}_4{}^2\right\} = B^4\hat{\sigma}_o{}^2 + \frac{A^2(1 - B)}{C}\Big[\left(1 + B + B^2 + B^3\right)\sigma^2 + \Big[\left(1 + A^2 + A^4\right)$$

$$+ B\left(1 + A^2\right) + B^2\Big](1 - A)^2\sigma^2 + \left(A^6 + A^4B\right.$$

$$\left. + A^2B^2 + B^3\right)\left(a - \hat{x}_o\right)^2\Big]$$

From these four mean values it is possible to recognize the general term of this series as

$$E\left\{\hat{\sigma}_k{}^2\right\} = B^k\hat{\sigma}_o{}^2 + \frac{A^2(1 - B)}{C}\left[\sigma^2 \sum_{i=o}^{k-1} B^i + \left(a - \hat{x}_o\right)^2 \sum_{i=o}^{k-1} A^{2i}B^{k-1-i}\right.$$

$$\left. + (1 - A)^2\sigma^2 \sum_{j=o}^{k-2}\left(B^j \sum_{j=o}^{k-2-j} A^{2i}\right)\right] \tag{4-3}$$

The next problem is to find the value of $E\left\{\hat{\sigma}_k{}^2\right\}$ as k approaches infinity. Since $|B|$ is less than 1.0,

$$\lim_{k \to \infty} B^k \hat{\sigma}_o^2 = 0$$

and

$$\lim_{k \to \infty} \sigma^2 \sum_{i=o}^{k-1} B^i = \sigma^2 \left[\frac{1}{1 - B}\right]$$

The limiting value of the next term in (4-3) is found from

$$\left(a - \hat{x}_o\right)^2 \sum_{i=o}^{k-1} A^{2i} B^{k-1-i} = \left(a - \hat{x}_o\right)^2 \left[B^{k-1} + A^2 B^{k-2} + A^4 B^{k-3} + \cdots + A^{2k-2}\right]$$

It is known that

$$0 < A < 1$$

$$0 < B < 1$$

Let

$$A < C_2 < 1$$

$$B < C_2 < 1$$

If C_2 is substituted for A and B in the above series, the resulting series is greater term by term than the original series involving A and B. If the limiting value of the series of C_2 is shown to approach zero, the limiting value of the series of A and B must also approach zero. Thus,

$$\sum_{i=o}^{k-1} C_2^{2i} C_2^{k-1-i} = C_2^{k-1} \sum_{i=o}^{k-1} C_2^{i}$$

This is a truncated geometric series whose partial sum, s_k (Ref. 13), is

$$s_k = C_2^{k-1} \frac{1 - C_2^k}{1 - C_2} = \frac{C_2^{k-1} - C_2^{2k-1}}{1 - C_2}$$

Since $C_2 < 1$,

$$\lim_{k \to \infty} s_k = 0$$

Therefore,

$$\lim_{k \to \infty} \left(a - \hat{x}_o\right)^2 \sum_{i=o}^{k-1} A^{2i} B^{k-1-i} = 0$$

The last term in (4-3) is

$$\left(1 - A\right)^2 \sigma^2 \sum_{j=o}^{k-2} \left(B^j \sum_{i=o}^{k-2-j} A^{2i}\right) = (1 - A)^2 \sigma^2 \Big[\Big(1 + A^2 + A^4 + \cdots$$

$$+ A^{2k-4}\Big) + B\Big(1 + A^2 + A^4 + \cdots$$

$$+ A^{2k-6}\Big) + B^2\Big(1 + A^2 + A^4 + \cdots$$

$$+ A^{2k-8}\Big) + \cdots + B^{k-3}\left(1 + A^2\right) + B^{k-2}\Big]$$

The limiting value is

$$\lim_{k \to \infty} (1 - A)^2 \sigma^2 \sum_{j=o}^{k-2} \left(B^j \sum_{i=o}^{k-2-j} A^{2i}\right) = (1 - A)^2 \sigma^2 \Big(1 + A^2 + A^4$$

$$+ \cdots\Big)\left(1 + B + B^2 + \cdots\right)$$

$$= \frac{(1 - A)^2\sigma^2}{\left(1 - A^2\right)(1 - B)}$$

$$= \frac{(1 - A)\sigma^2}{(1 + A)(1 - B)}$$

These expressions are inserted into the equation for $E\left\{\hat{\sigma}_k{}^2\right\}$ in order to find the limit as k approaches infinity. Thus,

$$\lim_{k \to \infty} E\left\{\hat{\sigma}_k{}^2\right\} = \frac{A^2(1 - B)}{C}\left[\frac{\sigma^2}{1 - B} + \frac{(1 - A)\sigma^2}{(1 + A)(1 - B)}\right]$$

$$= \frac{A^2\sigma^2}{C}\left[1 + \frac{1 - A}{1 + A}\right] = \frac{\sigma^2}{C}\left[\frac{2A^2}{1 + A}\right]$$

In order for this limit to converge to the actual variance, σ^2, of the function being sampled, we must have

$$C = \frac{2A^2}{1 + A}$$

The value of C obtained above is inserted into the estimation equation (4-2) to give

$$\hat{\sigma}_k{}^2 = B\hat{\sigma}_{k-1}{}^2 + \frac{(1 - B)(1 + A)}{2}\left[x_k - \hat{x}_{k-1}\right]^2 \qquad (4\text{-}4)$$

CHAPTER V

DERIVATION OF THE VARIANCE OF THE ESTIMATED VARIANCE

The variance of the estimated variance is also of interest since it gives an indication of the error of the estimate. Due to the complexity of the procedure of calculating the variance of the estimated variance for the general case, the derivation is performed here only for input data consisting of samples taken from a Gaussian distribution with mean of "a" and variance of "σ^2". However, the technique of estimation described in Chapter IV is not limited to this case; it applies to any probability distribution whose mean and variance exist. If the moments of a variable are expressed in terms of the mean and variance of the variable, it is found that moments of order greater than two are dependent on the probability distribution of the variable. The variance of the estimated variance is a function of the probability distribution since it involves moments of order greater than two. The moments of a Gaussian variable are shown in Appendix VII.

Since the equation used for estimation of the mean is a linear equation, the estimated mean has a Gaussian distribution if the data have a Gaussian distribution. The term, $\left(x_k - \hat{x}_{k-1}\right)$, is the difference of two terms, each of which has a Gaussian probability distribution. The probability of the difference is also Gaussian. The square of this difference has a chi-square distribution with one degree of freedom (Ref. 11, pp. 250-253).

In order to investigate the probability distribution of the estimated variance it is necessary to examine several estimation steps using

$$\hat{\sigma}_k^{\,2} = B\hat{\sigma}_{k-1}^{\,2} + \frac{(1 - B)(1 + A)}{2}\left[x_k - \hat{x}_{k-1}\right]^2$$

The initial guess, $\hat{\sigma}_o^{\,2}$, has a delta function for a probability distribution since it can have only one value. The distribution of $\hat{\sigma}_1^{\,2}$ is the weighted convolution of a delta function and a chi-square distribution with its origin shifted. The equation for $\hat{\sigma}_2^{\,2}$ is a weighted sum of $\hat{\sigma}_1^{\,2}$ and $\left(x_2 - \hat{x}_1\right)^2$. Because of the estimation technique $\hat{x}_1$ is not independent of $\hat{\sigma}_1^{\,2}$ and is fixed exactly when $\hat{\sigma}_1^{\,2}$ is determined. However, x_2 is independent of either $\hat{\sigma}_1^{\,2}$ or $\hat{x}_1$. The distribution of $\hat{\sigma}_2^{\,2}$ is a weighted convolution of the chi-square distribution representing $\hat{\sigma}_1^{\,2}$ and the distribution $p\left[\left(x_2 - \hat{x}_1\right)^2 \middle| \hat{\sigma}_1^{\,2}\right]$, which is a chi-square distribution with its mean a function of $\hat{\sigma}_1^{\,2}$. The probability distribution of any estimate of the variance by this recursive equation is a weighted convolution of the distribution of the previous estimate and a chi-square distribution whose mean is determined by the previous estimate of the variance. The probability distribution of the estimate of the variance is not determined since it is not practical to make a detailed calculation. Although the distribution of the estimate of the variance is not derived, its variance serves as a indication of the error of the estimate. The error decreases as the variance decreases.

Using the definition of variance, the variance of $\hat{\sigma}_k^{\,2}$ is

$$\text{variance of } \hat{\sigma}_k^{\,2} = E\left\{\hat{\sigma}_k^{\,4}\right\} - E^2\left\{\hat{\sigma}_k^{\,2}\right\}$$

$$= E\left\{\left[B\hat{\sigma}_{k-1}^{\,2} + \frac{(1-B)(1+A)}{2}\left(x_k - \hat{x}_{k-1}\right)^2\right]^2\right\}$$

$$- E^2\left\{B\hat{\sigma}_{k-1}^{\,2} + \frac{(1-B)(1+A)}{2}\left(x_k - \hat{x}_{k-1}\right)^2\right\}$$

$$= B^2 E\left\{\hat{\sigma}_{k-1}^{\,4}\right\} + B(1-B)(1+A)\, E\left\{\left(x_k - \hat{x}_{k-1}\right)^2 \hat{\sigma}_{k-1}^{\,2}\right\}$$

$$+ \frac{(1-B)^2(1+A)^2}{4} E\left\{\left(x_k - \hat{x}_{k-1}\right)^4\right\} - B^2E^2\left\{\hat{\sigma}_{k-1}^{\,2}\right\}$$

$$- B(1-B)(1+A)\, E\left\{\hat{\sigma}_{k-1}^{\,2}\right\} E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\}$$

$$- \frac{(1-B)^2(1+A)^2}{4} E^2\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\}$$

$$= B^2\left[E\left\{\hat{\sigma}_{k-1}^{\,4}\right\} - E^2\left\{\hat{\sigma}_{k-1}^{\,2}\right\}\right]$$

$$+ B(1-B)(1+A)\left[E\left\{\left(x_k - \hat{x}_{k-1}\right)^2 \hat{\sigma}_{k-1}^{\,2}\right\}\right.$$

$$\left. - E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} E\left\{\hat{\sigma}_{k-1}^{\,2}\right\}\right]$$

$$+ \frac{(1-B)^2(1+A)^2}{4}\left[E\left\{\left(x_k - \hat{x}_{k-1}\right)^4\right\} - E^2\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\}\right]$$

$$= B^2 \left(\text{variance of } \hat{\sigma}_{k-1}{}^2\right)$$

$$+ B(1 - B)(1 + A) \left[E\left\{x_k{}^2\hat{\sigma}_{k-1}{}^2 - 2x_k\hat{x}_{k-1}\hat{\sigma}_{k-1}{}^2\right.\right.$$

$$\left.\left. + \hat{x}_{k-1}{}^2\hat{\sigma}_{k-1}{}^2\right\} - E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} E\left\{\hat{\sigma}_{k-1}{}^2\right\}\right]$$

$$+ \frac{(1 - B)^2(1 + A)^2}{4}\left[\text{variance of } \left(x_k - \hat{x}_{k-1}\right)^2\right]$$

$$= B^2 \left(\text{variance of } \hat{\sigma}_{k-1}{}^2\right)$$

$$+ \frac{(1 - B)^2(1 + A)^2}{4}\left[\text{variance of } \left(x_k - \hat{x}_{k-1}\right)^2\right]$$

$$+ B(1 - B)(1 + A)\left[\left(a^2 + \sigma^2\right) E\left\{\hat{\sigma}_{k-1}{}^2\right\}\right.$$

$$- 2a\, E\left\{\hat{x}_{k-1}\hat{\sigma}_{k-1}{}^2\right\} + E\left\{\hat{x}_{k-1}{}^2\hat{\sigma}_{k-1}{}^2\right\}$$

$$\left. - E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} E\left\{\hat{\sigma}_{k-1}{}^2\right\}\right] \tag{5-1}$$

This last step can be made since $\hat{\sigma}_{k-1}{}^2$ and $\hat{x}_{k-1}$ are independent of x_k.

Let $k = n$ where n is large enough so that all terms in the equation for $\left(\text{variance of } \hat{\sigma}_k{}^2\right)$ except $\left(\text{variance of } \hat{\sigma}_{k-1}{}^2\right)$ have become infinitesimally close to their limiting values. The convergence of each of these terms is shown in the appendices by the derivation of their limiting values. In Appendix III these are

$$\lim_{k \to \infty} \text{variance } \left(x_k - \hat{x}_{k-1}\right)^2 = \frac{8\sigma^4}{(1 + A)^2} \tag{5-2}$$

and

$$\lim_{k \to \infty} E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} = \frac{2\sigma^2}{1 + A} \tag{5-3}$$

Appendix IV shows that

$$\lim_{k \to \infty} E\left\{\hat{x}_{k-1}\hat{\sigma}_{k-1}{}^2\right\} = a\sigma^2 \tag{5-4}$$

Appendix V shows that

$$\lim_{k \to \infty} E\left\{\hat{x}_{k-1}{}^2\hat{\sigma}_{k-1}{}^2\right\} = a^2\sigma^2 + \left[\frac{1 - A}{1 + A} + \frac{(1 - A)^2(1 - B)}{(1 + A)\left(1 - A^2B\right)}\right]\sigma^4 \tag{5-5}$$

By the choice of C in Equation (4-2) we have insured that

$$\lim_{k \to \infty} E\left\{\hat{\sigma}_{k-1}{}^2\right\} = \sigma^2 \tag{5-6}$$

Substitution of Equations (5-2) - (5-6) into Equation (5-1) yields

$$\begin{aligned}
\text{variance of } \hat{\sigma}_n{}^2 &= B^2\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right) + \frac{(1 - B)^2(1 + A)^2}{4}\left[\frac{8\sigma^4}{(1 + A)^2}\right] \\
&\quad + B(1 - B)(1 + A)\left[\left(a^2 + \sigma^2\right)\sigma^2 - 2a\left(a\sigma^2\right) + a^2\sigma^2\right. \\
&\quad \left. + \left(\frac{1 - A}{1 + A} + \frac{(1 - A)^2(1 - B)}{(1 + A)\left(1 - A^2B\right)}\right)\sigma^4 - \left(\frac{2\sigma^2}{1 + A}\right)\sigma^2\right] \qquad (5\text{-}7) \\
&= B^2\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right) + 2(1 - B)^2\sigma^4
\end{aligned}$$

$$+ B(1 - B)(1 + A)\left[a^2\sigma^2 + \sigma^4 - 2a^2\sigma^2 + a^2\sigma^2\right.$$

$$\left. + \left(\frac{1 - A}{1 + A}\right)\sigma^4 + \frac{(1 - A)^2(1 - B)}{(1 + A)\left(1 - A^2B\right)}\sigma^4 - \frac{2\sigma^4}{1 + A}\right]$$

$$= B^2\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right) + 2(1 - B)^2\sigma^4$$

$$+ B(1 - B)(1 + A)\left[\sigma^4 + \left(\frac{1 - A}{1 + A}\right)\sigma^4\right.$$

$$\left. + \frac{(1 - A)^2(1 - B)}{(1 + A)\left(1 - A^2B\right)}\sigma^4 - \frac{2\sigma^4}{1 + A}\right]$$

$$= B^2\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right) + 2(1 - B)^2\sigma^4$$

$$+ B(1 - B)(1 + A)\left[\frac{2\sigma^4}{1 + A} + \frac{(1 - A)^2(1 - B)}{(1 + A)\left(1 - A^2B\right)}\sigma^4\right.$$

$$\left. - \frac{2\sigma^4}{1 + A}\right]$$

$$= B^2\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right) + 2(1 - B)^2\sigma^4$$

$$+ \frac{B(1 - B)^2(1 - A)^2}{1 - A^2B}\sigma^4 \tag{5-8}$$

The method of determining the limiting value of variance of $\hat{\sigma}_n{}^2$ is to insert some constant, M, for $\left(\text{variance of } \hat{\sigma}_{n-1}{}^2\right)$. Several terms are determined in order to recognize the series being generated. Thus,

$$\text{variance of } \hat{\sigma}_{n-1}^{\;2} = M$$

$$\text{variance of } \hat{\sigma}_{n}^{\;2} = B^2 M + (1 - B)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \sigma^4$$

$$\text{variance of } \hat{\sigma}_{n+1}^{\;2} = B^4 M + (1 + B^2)(1 - B)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \sigma^4$$

$$\text{variance of } \hat{\sigma}_{n+2}^{\;2} = B^6 M + \left(1 + B^2 + B^4\right)\left(1 - B\right)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \sigma^4$$

The general term is

$$\text{variance of } \hat{\sigma}_{n+j}^{\;2} = B^{2(j+1)} M + (1 - B)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \sigma^4 \sum_{i=0}^{j} B^{2i}$$

The limiting value is

$$\lim_{j \to \infty} \left(\text{variance of } \hat{\sigma}_{n+j}^{\;2}\right) = M \lim_{j \to \infty} B^{2(j+1)} + (1 - B)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \sigma^4 \lim_{j \to \infty} \sum_{i=0}^{j} B^{2i}$$

Since

$$\lim_{j \to \infty} \sum_{i=0}^{j} B^{2i} = \frac{1}{1 - B^2} \quad \text{for } |B| < 1.0$$

$$\lim_{j \to \infty} B^{2(j+1)} = 0 \quad \text{for } |B| < 1.0$$

$$\lim_{j \to \infty} \left(\text{variance of } \hat{\sigma}_{n+j}^{\;2}\right) = (1 - B)^2 \left[\frac{B(1 - A)^2}{1 - A^2B} + 2\right] \left(\frac{1}{1 - B^2}\right) \sigma^4$$

$$\lim_{j \to \infty} \left(\text{variance of } \hat{\sigma}_k^{\,2}\right) = \lim_{j \to \infty} \left(\text{variance of } \sigma_{n+j}^{\,2}\right)$$

$$= \left(\frac{1 - B}{1 + B}\right)\left(\frac{B(1 - A)^2}{1 - A^2B} + 2\right) \sigma^4 \qquad (5\text{-}9)$$

No attempt is made to apply the standard mathematical tests for convergence due to the complexity of the series. The method of derivation used above shows the convergence of the series since the starting point has no effect on the limiting value of the sequence and the limiting value is determined.

A calculation which adds to the credibility of this derivation is that of the estimation of the variance when the mean is known exactly. For this case A is equal to 1.0 and the equation for the variance of the estimated variance reduces to

$$\lim_{j \to \infty} \left(\text{variance of } \hat{\sigma}_j^{\,2}\right) = \left(\frac{1 - B}{1 + B}\right) 2\sigma^4 \quad \text{for} \quad A = 1.0 \qquad (5\text{-}10)$$

This can be checked by actually calculating the mean and variance of the estimated variance with the mean known exactly. Thus,

$$\hat{\sigma}_k^{\,2} = B\hat{\sigma}_{k-1}^{\,2} + (1 - B)\left(x_k - a\right)^2 \qquad (5\text{-}11)$$

and

$$E\left\{\hat{\sigma}_o^{\,2}\right\} = \hat{\sigma}_o^{\,2}$$

$$E\left\{\hat{\sigma}_1^{\,2}\right\} = B\hat{\sigma}_o^{\,2} + (1 - B)\sigma^2$$

$$E\left\{\hat{\sigma}_2^{\,2}\right\} = B^2\hat{\sigma}_o^{\,2} + (1 + B)\ (1 - B)\sigma^2$$

$$E\left\{\hat{\sigma}_3^{\,2}\right\} = B^3\hat{\sigma}_o^{\,2} + (1 + B + B^2)(1 - B)\sigma^2$$

The general term is

$$E\left\{\hat{\sigma}_k^{\,2}\right\} = B^k\hat{\sigma}_o^{\,2} + (1 - B)\sigma^2 \sum_{j=o}^{k-1} B^j$$

Since $B < 1.0$

$$\lim_{k \to \infty} B^k\hat{\sigma}_o^{\,2} = 0 \quad \text{for} \quad |B| < 1.0$$

and

$$\lim_{k \to \infty} \sum_{j=o}^{k-1} B^j = \frac{1}{1 - B} \quad \text{for} \quad |B| < 1.0$$

Then,

$$\lim_{k \to \infty} E\left\{\hat{\sigma}_k^{\,2}\right\} = (1 - B)\sigma^2\left(\frac{1}{1 - B}\right) = \sigma^2$$

The variance of the estimation is determined by

$$\text{variance of} \quad \hat{\sigma}_o^{\,2} = 0$$

$$\text{variance of} \quad \hat{\sigma}_1^{\,2} = E\left\{\hat{\sigma}_1^{\,4}\right\} - E^2\left\{\hat{\sigma}_1^{\,2}\right\}$$

The variances for several values of k are

$$\text{variance of} \quad \hat{\sigma}_1^{\,2} = (1 - B)^2\ 2\sigma^4$$

$$\text{variance of } \hat{\sigma}_2{}^2 = 2(1 + B^2)(1 - B)^2\sigma^4$$

$$\text{variance of } \hat{\sigma}_3{}^2 = 2\left(1 + B^2 + B^4\right)(1 - B)^2\sigma^4$$

The general term is

$$\text{variance of } \hat{\sigma}_k{}^2 = 2(1 - B)^2\sigma^4 \sum_{j=o}^{k-1} B^{2j}$$

The limiting value is

$$\lim_{k \to \infty} \text{ variance of } \hat{\sigma}_k{}^2 = 2(1 - B)^2\sigma^4\left(\frac{1}{1 - B^2}\right) = \frac{2(1 - B)}{1 + B}\,\sigma^4$$

This equation checks with that obtained by letting $A = 1.0$ in the general equation (5-9).

As was found in the estimation of the mean, the limiting value of the variance of the estimate can be made as small as desired by making the estimation constant, B, closer to 1.0. Since the estimation of the mean is used in the est tion of the variance, the constant A also has an e*fect on the limiting value of the variance of the estimated variance. Again, A should be near 1.0 in order to make the variance of the estimate small.

CHAPTER VI

DECISION-DIRECTED-MEASUREMENT ESTIMATION TECHNIQUE

Description of the Method

The three previous sections of this dissertation presented an analysis of the estimation of the mean and of the estimation of the variance. The proper operation of these estimators requires that the correct answer of the decision be known so that the proper estimator can be updated. When the system is first operated in a given situation, the location of the frame synchronization code is not known. The system is required to move the threshold until enough correct decisions can be made in order to locate the frame synchronization code. A form of "Decision-Directed-Measurement" similar to that described by Lindenlaub and Mix (Ref. 3, p. 13) is used to control the system during the search for the frame synchronization code. This scheme is examined, in the pages which follow, only for the cases requiring estimates of the mean, that is, for equally probable signals. It can be employed for cases requiring estimation of the variance, but an analysis of this situation is not included here.

The Decision-Directed-Measurement (DDM) is used as outlined by the following sequence: (1) An initial guess is made of the mean, $\hat{y}_o$, of the received signal when a binary 0 is transmitted and of the mean, $\hat{x}_o$, of the received signal when a binary 1 is transmitted. (2) A first selection, β_o, for the threshold is made according to

$$\beta_o = \frac{\hat{x}_o + \hat{y}_o}{2} \tag{6-1}$$

This is used as the threshold in order to decide whether the first sample, x_1, is a binary 0 or 1. If x_1 is judged to be a binary 0, $\hat{y}_o$ is updated using the exponential smoothing equation,

$$\hat{y}_1 = A\hat{y}_o + (1 - A)x_1 \tag{6-2}$$

and the revised estimate of the threshold, β_1, is

$$\beta_1 = \frac{\hat{y}_1 + \hat{x}_o}{2} \tag{6-3}$$

If, however, x_1 is judged to be a binary 1, $\hat{x}_o$ is updated by use of

$$\hat{x}_1 = A\hat{x}_o + (1 - A)x_1 \tag{6-4}$$

and the revised estimate of the threshold is

$$\beta_1 = \frac{\hat{y}_o + \hat{x}_1}{2} \tag{6-5}$$

As more and more samples are processed, the estimated threshold moves toward the optimum location, which is the intersection of $p(x|o)$ and $p(x|1)$, and eventually gets close enough to allow so many correct decisions that the frame synchronization code may be located. When the code is located, the "Learning With Teacher" scheme is employed.

Analysis, Mathematical

Some questions which should be answered are:

1. Does the estimated threshold, in fact, move toward the optimum location?

2. What factors affect the convergence rate of the threshold?

3. What is optimum point for initial guess of the threshold location?

To answer these questions, it is helpful to examine some probability distributions. Figure 4 shows the probability density of the received signal, the initial guesses of $\hat{x}_o$ and $\hat{y}_o$, the corresponding β_o and the actual means, A_o and A_1, given by

$$A_o = \int_{-\infty}^{\infty} x_1 p(x_1 | o) \, dx_1 \tag{6-6}$$

and

$$A_1 = \int_{-\infty}^{\infty} x_1 p(x_1 | 1) \, dx_1 \tag{6-7}$$

Since any $x_1 \geq \beta_o$ is judged to be a binary 1, and any $x_1 < \beta_o$ is judged to be a binary , the conditional probabilities are

$$p(x_1 | x_1 \geq \beta_o) = \left[\frac{p(o)\ p(x_1 | o)}{\int_{\beta_o}^{\infty} p(x_1 | o)\, dx_1} + \frac{p(1)\ p(x_1 | 1)}{\int_{\beta_o}^{\infty} p(x_1 | 1)\, dx_1} \right] u_1(x_1 - \beta_o) \tag{6-8}$$

and

$$p(x_1 | x_1 < \beta_o) = \left[\frac{p(o)\ p(x_1 | o)}{\int_{-\infty}^{\beta_o} p(x_1 | o)\, dx_1} + \frac{p(1)\ p(x_1 | 1)}{\int_{-\infty}^{\beta_o} p(x_1 | 1)\, dx_1} \right] u_1(\beta_o - x_1) \tag{6-9}$$

where $u_1(z)$ is a unit step $\left(u_1 = 1 \text{ for } z > o \text{ and } u_1 = o \text{ for } z < o\right)$.

Figure 5 is a plot of $p(x_1 | x_1 \geq \beta_o)$ and $p(x_1 | x_1 < \beta_o)$.

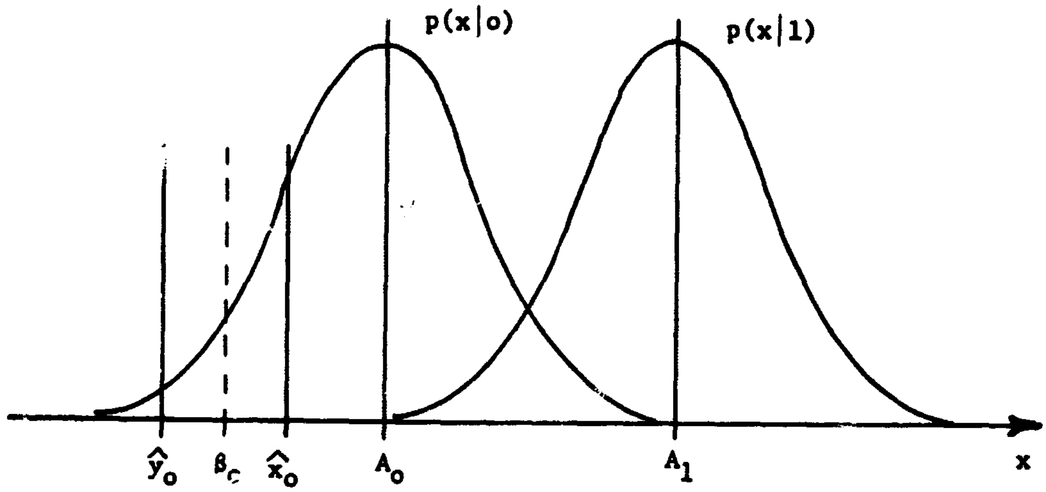

Figure 4.- Operation with initial estimates.

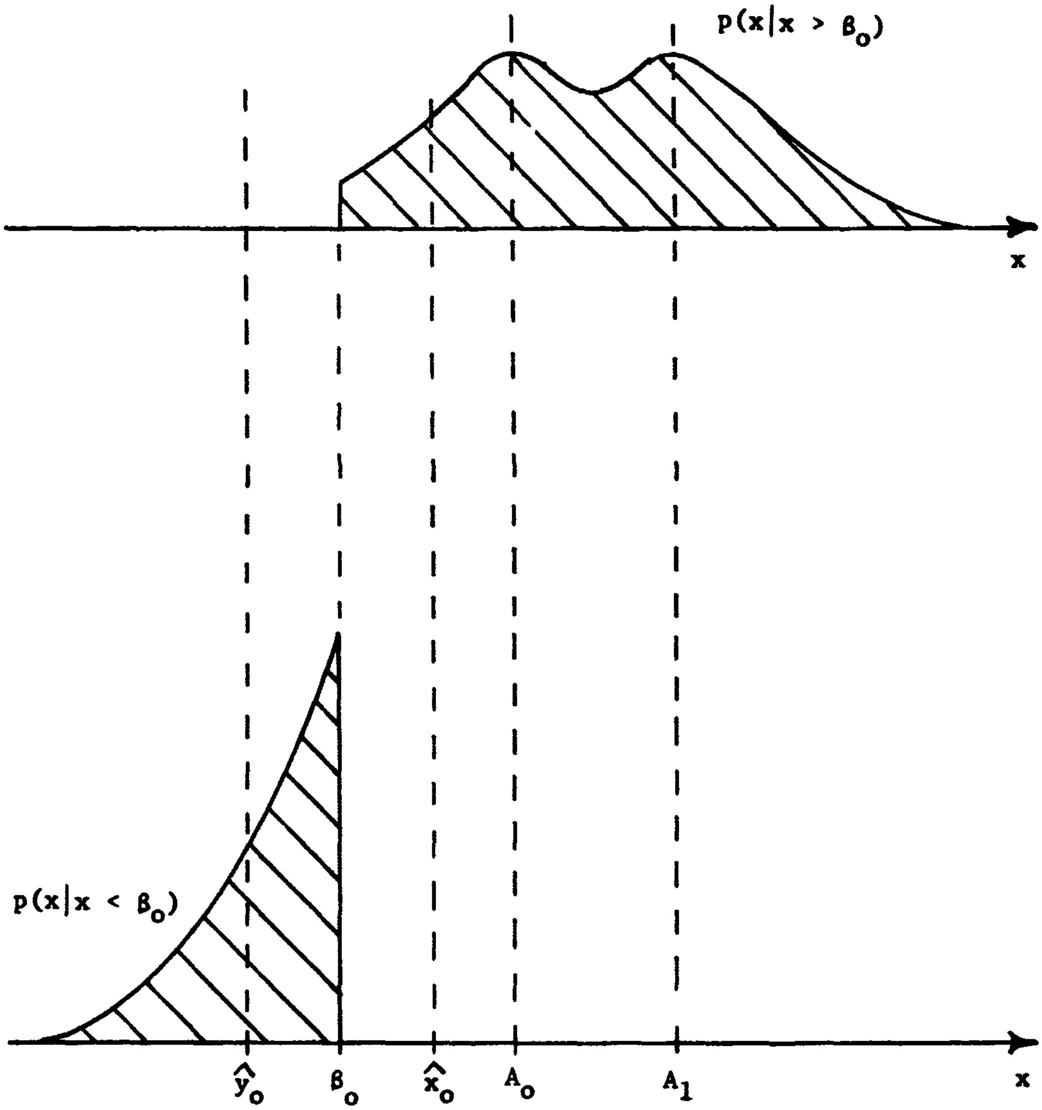

Figure 5.- Probability density of first sample after decision.

For $\hat{x}_1$ given by (6-4), the calculation of its probability density requires the conditional probability for x_1 given by (6-8). Since $\hat{x}_o$ and x_1 are independent, the probability density of $\hat{x}_1$ given that $x_1 \geq \beta_o$ is a weighted convolution of $p(\hat{x}_o)$ and $p(x_1) = p(x_1 | x_1 \geq \beta_o)$, (Ref. 11, p. 189). Similar techniques are used to determine $p(\hat{y}_1)$ with the condition of $x_1 < \beta_o$. The shapes of $p(\hat{x}_1 | x_1 \geq \beta_o)$ and $p(\hat{y}_1 | x_1 < \beta_o)$ are shown in Figure 6. The new threshold is

$$\beta_1 \begin{cases} = \dfrac{A\hat{x}_o + (1 - A)x_1 + \hat{y}_o}{2} & \text{for } x_1 \geq \beta_o \qquad (6\text{-}10) \\[2ex] = \dfrac{\hat{x}_o + A\hat{y}_o + (1 - A)x_1}{2} & \text{for } x_1 < \beta_o \qquad (6\text{-}11) \end{cases}$$

This includes the possibility that $\hat{x}_o$ is not updated if the signal is decided to be a binary 0 and that $\hat{y}_o$ is not updated if the signal is decided to be a binary 1. The probability density of β_1 is obtained by a weighted convolution of $p(\hat{x}_o)$, $p(x_1)$, and $p(\hat{y}_o)$ with appropriate use of the probability that $x_1 \geq \beta_o$ and the probability that $x_1 < \beta_o$. The process is repeated when the second sample, x_2, is received with the added complication that β_1 has a probability density. If both x_1 and x_2 are decided to be binary 0's, the calculation of β_2 will still use $\hat{x}_o$ as the estimate of A_1.

It can be seen that a mathematical analysis of this problem in closed form is virtually impossible since it involves repeated convolutions of truncated probability distributions. No general analysis of a DDM system has been located in the literature. With the

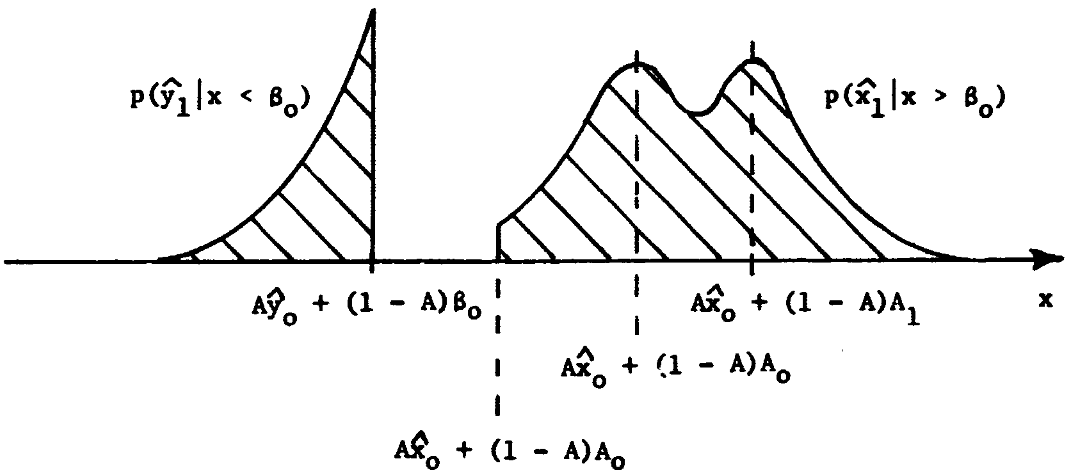

Figure 6.- Probability density of estimated means after processing of first sample.

exception of a "hang-up" for the case of no noise, there have been no cases postulated or discovered experimentally in this investigation or in the literature for binary signals for which the DDM system does not converge. The possibility of the "hang-up" is discussed and eliminated later in this report. The DDM system is designed so that the estimate of the mean of the binary 1 moves to the mean of all signals above the threshold and the estimate of the mean of the binary 0 moves to the mean of all signals below the threshold. The threshold divides the range of received signals into two portions so that the average of the estimates of the two means is equal to the boundary between the portions.

Computer Analysis

A general purpose digital computer is found to be useful for the investigation of some of the characteristics of the system. Since it is intended that digital techniques be used in the final hardware, the range of inputs to the decision element is converted from analog to digital format. Therefore, the range of inputs is separated, in effect, into 2^n discrete partitions where n is the number of bits used in the digital word. By making use of this partitioning of the input range, it is possible to perform a numerical convolution of the probability densities on a digital computer.

For the calculations here, the input range is divided into 64 levels ($n = 6$). This value is selected as providing about the minimum resolution required and as being small enough to reduce the required computer time. The probability densities $p(o)$, $p(1)$, $p(x|o)$, and $p(x|1)$ are selected for the particular test case, calculated for each of the 64 partitions, and inserted as inputs to the computer. The initial estimates, $\hat{x}_o$ and $\hat{y}_o$, are also selected and inserted as inputs. The computer uses $\hat{x}_o$ and $\hat{y}_o$ to set up a 64 x 64 array representing $p\left(\hat{x}_o, \hat{y}_o\right)$. The indices of the array represent the amplitude of the variables and the value stored in a given location represents the probability. The computer then uses each possibility of input with each possibility of threshold to generate $p\left(\hat{x}_1, \hat{y}_1\right)$. The process is repeated in order to determine $p\left(\hat{x}_i, \hat{y}_i\right)$. The average value of the threshold (as a function of number of samples) is calculated and is used as an indicator of the convergence of the threshold. Appendix VI shows the computer program used with some typical numbers as inputs.

Effect of Choice of $\hat{x}_o$ and $\hat{y}_o$ on Convergence Rate

The first characteristic of the system to be investigated is the effect of the initial guesses, $\hat{x}_o$ and $\hat{y}_o$, on the convergence rate. Since $\beta_o = \left(\hat{x}_o + \hat{y}_o\right)/2$, there are many choices of $\hat{x}_o$ and $\hat{y}_o$ which yield the same β_o. For this test all of the inputs to the computer program except $\hat{x}_o$ and $\hat{y}_o$ remain unchanged; $\hat{x}_o$ and $\hat{y}_o$ are varied with the proper relationship so that β_o remains constant. Figure 7 shows a plot of the average value of the estimated threshold location versus the number of samples processed as a function of $\hat{x}_o$ and $\hat{y}_o$. Several other cases have been run on the computer but have not been shown here since the results of Figure 7 are typical of those in the other cases. It can be observed in Figure 7 that the fastest convergence

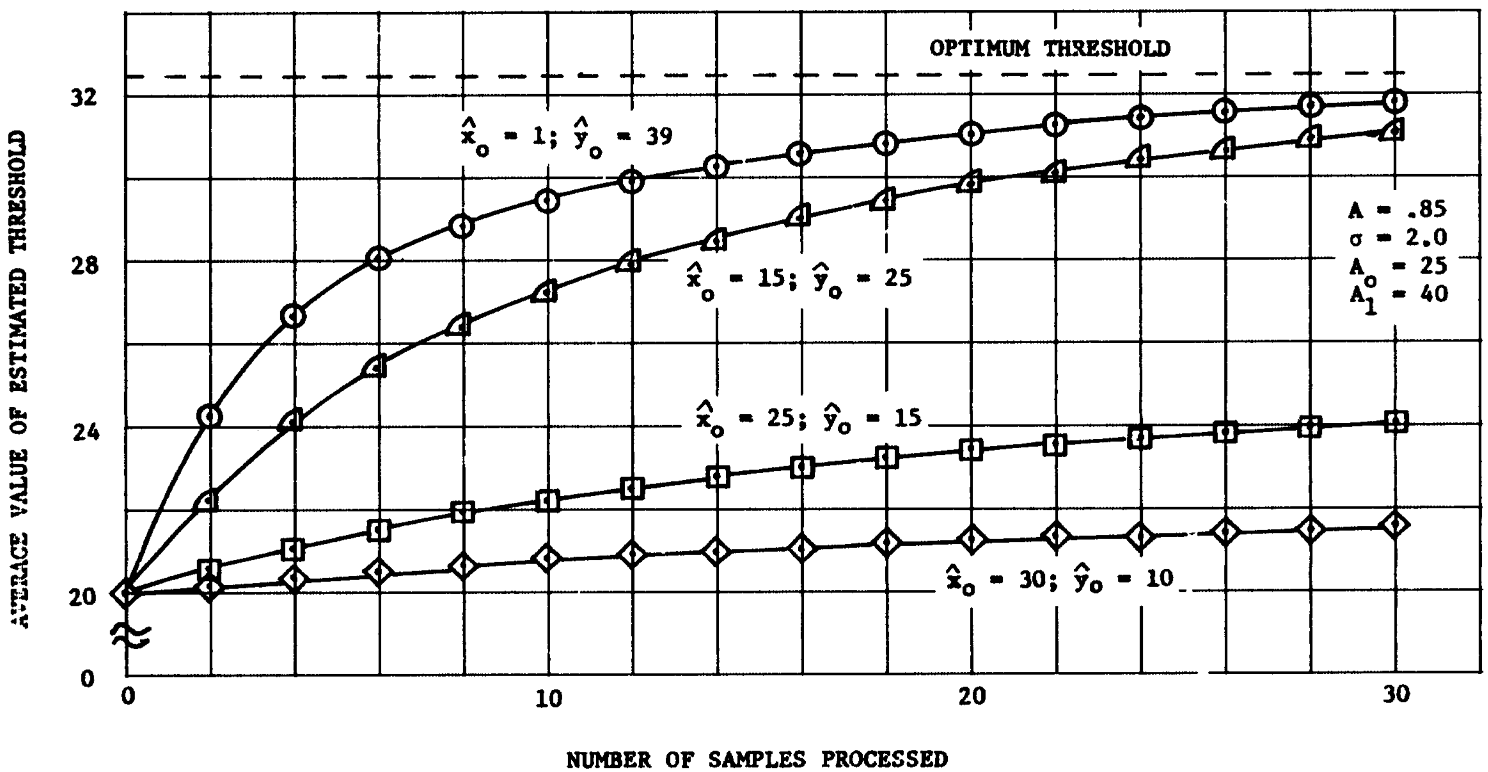

Figure 7.- Movement of threshold.

is obtained when $\hat{x}_o$ is as small as possible and $\hat{y}_o$ is as large as possible. This seems strange since $\hat{x}_o$ is the estimate of the mean of the received signal when a binary 1 is transmitted. However, this has been found to be true in all of the cases which have been investigated.

The following explanation is offered for this property. For $x_1 \geq \beta_o$, only $\hat{x}_o$ is updated. From (6-1),

$$\hat{y}_o = 2\beta_o - \hat{x}_o \tag{6-12}$$

Using this with (6-4) yields

$$\begin{aligned} \beta_1 &= \frac{\hat{x}_1 + \hat{y}_o}{2} \\ &= \frac{A\hat{x}_o + (1 - A)x_1 + 2\beta_o - \hat{x}_o}{2} \\ &= \frac{(1 - A)\left(x_1 - \hat{x}_o\right)}{2} + \beta_o \end{aligned} \tag{6-13}$$

The change, $\Delta\beta$, in the threshold is

$$\begin{aligned} \Delta\beta &= \beta_1 - \beta_o \\ &= \frac{(1 - A)\left(x_1 - \hat{x}_o\right)}{2} \end{aligned} \tag{6-14}$$

For $x_1 < \beta_o$, only $\hat{y}_o$ is updated. From (6-1),

$$\hat{x}_o = 2\beta_o - \hat{y}_o \tag{6-15}$$

Using this with (6-2) yields

$$\beta_1 = \frac{\hat{x}_o + \hat{y}_1}{2}$$

$$= \frac{2\beta_o - \hat{y}_o + A\hat{y}_o + (1 - A)x_1}{2}$$

$$= \frac{(1 - A)\left(x_1 - \hat{y}_o\right)}{2} + \beta_o \tag{6-16}$$

The change, $\Delta\beta$, in the threshold is

$$\Delta\beta = \beta_1 - \beta_o$$

$$= \frac{(1 - A)\left(x_1 - \hat{y}_o\right)}{2} \tag{6-17}$$

If β_o is less than the optimum location of the threshold, $\Delta\beta$ should be as large a positive value as possible. According to (6-14) and (6-17), $\hat{x}_o$ and $\hat{y}_o$ should both be as small as possible. Since $\beta_o = \left(\hat{x}_o + \hat{y}_o\right)/2$, one must be large if the other is small. Since the received signal is decided to be a binary 1 with greater probability than to be a binary 0, (6-14) is applicable more often than (6-17) and $\hat{x}_o$ is selected as small as possible without regard to $\hat{y}_o$.

Conversely, if β_o is greater than the optimum location of the threshold, $\Delta\beta$ should be as large a negative value as possible. Therefore, $\hat{x}_o$ and $\hat{y}_o$ should be as large as possible. Since the received signal is judged to be a binary 0 with higher probability, (6-17) applies more often. Therefore, $\hat{y}_o$ has more effect and is selected as large as possible. This reasoning gives the same results as had been observed from the computer runs; that is, $\hat{x}_o$ should be small and $\hat{y}_o$ should be large. For the case of no prior knowledge of the signal characteristics the optimum location of β_o appears to be the center of the input range. This means that $\hat{x}_o$ is the smallest possible input value and $\hat{y}_o$ is the largest possible input value.

Effect of Other Parameters on Convergence Rate

This section contains observations of the relationship of the convergence rate to other variables, such as:

1. Separation of the mean, A_1, of binary 1, from the mean, A_o, of binary 0.

2. The variance of the noise; that is, the variance of $p(x|o)$ and $p(x|1)$.

3. Separation of initial estimate of mean and actual mean.

A method is devised for measuring the convergence rate. Figure 7 shows that, for optimum location of $\hat{x}_o$ and $\hat{y}_o$, the movement of the average value of the estimated threshold has the appearance of the exponential charging of a capacitor. Although the curve for this system is not exactly an exponential, the number of samples required to move 63.2 per cent of the distance between β_o and the actual threshold is called a time constant and is used to compare different systems.

Figure 8 shows a plot of the time constant versus the deviation of the noise for three different values of A_o and A_1. The estimation factor, A, the initial estimate, β_o, of the threshold and the optimum

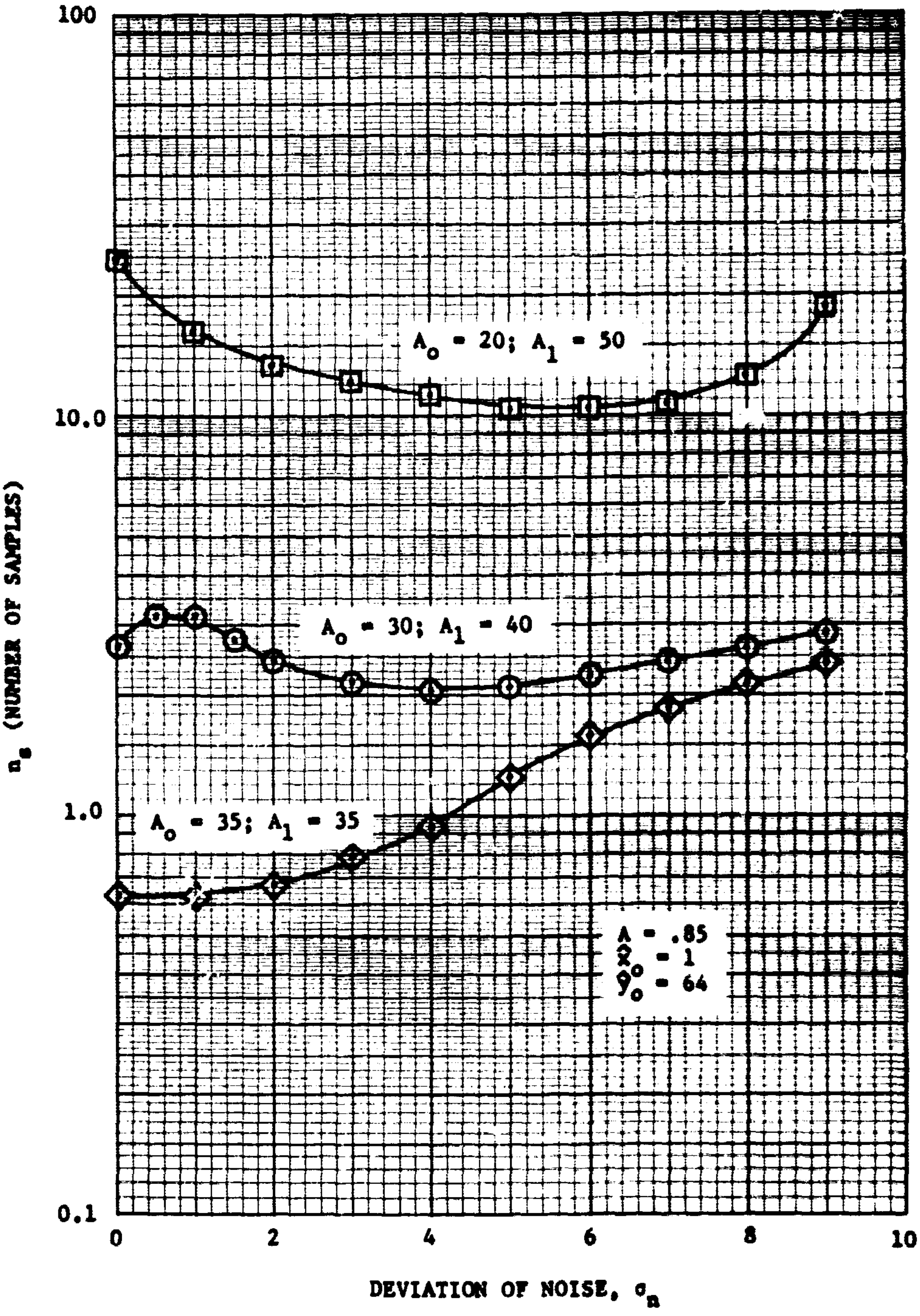

Figure 8.- Time constant of mean of estimated threshold.

location of the threshold are held constant for this test. The initial estimates, $\hat{x}_o$ and $\hat{y}_o$, are located at the points found to be optimum $\left(\hat{x}_o = 1;\ \hat{y}_o = 64\right)$ according to the discussion following equation (6-17).

It had been hoped that the convergence rate could be expressed as a function of the estimation constant, A, and of a signal-to-noise ratio defined as

$$\mathrm{SNR} = \frac{A_1 - A_o}{\sigma_n} \tag{6-18}$$

where σ_n is the deviation of $p(x|o)$ and $p(x|1)$. Figure 8 shows that the convergence rate does not depend on A, A_1, A_o, and σ_n in such a simple way. For example, for $A_1 = 50$, $A_o = 20$, $\sigma_n = 6$, and $A = 0.85$,

$$\mathrm{SNR} = \frac{50 - 20}{6} = 5$$

with Figure 8 showing that

$$\text{Time constant} = 10.32 \text{ samples} \tag{6-19}$$

Also for, $A_1 = 40$, $A_o = 30$, $\sigma_n = 2$,

$$\mathrm{SNR} = \frac{40 - 30}{2} = 5$$

Figure 8 shows that

$$\text{Time constant} = 2.47 \text{ samples} \tag{6-20}$$

Thus, it can be seen that the convergence rate is different for two cases which have the same signal-to-noise ratio. This investigation is not carried any further due to the amount of computer time required to generate each point on the curve and due to the large number of curves which are required to determine the effect of A_o, A_1, β_o, and other variables on the convergence rate.

Comparison With Learning With Teacher

The convergence rate for the Learning With Teacher scheme can be compared to the convergence rate of the DDM technique for those cases shown in Figure 8. For the case of Learning With Teacher for equally probable signals, the threshold is computed by

$$\beta_i = \frac{\hat{x}_j + \hat{y}_k}{2} \tag{6-21}$$

Since $\hat{x}_j$ and $\hat{y}_k$ are estimated by recursive equations with the same time constant, β_i will have undergone one time constant when both $\hat{x}_j$ and $\hat{y}_k$ have undergone one time constant. It takes twice as many samples for <u>both</u> $\hat{x}_j$ and $\hat{y}_k$ to undergo one time constant so that the time constant (measured in "number of samples") of β_i is twice that of either $\hat{x}_j$ or $\hat{y}_k$. For the situation shown in Figure 8, A is equal to 0.85. The time constant for either $\hat{x}_j$ or $\hat{y}_k$ is

$$\begin{aligned} n_s &= \frac{-1}{\ln A} \\ &= \frac{-1}{\ln (0.85)} \\ &= 6.15 \text{ samples} \end{aligned} \tag{6-22}$$

The time constant of β_i, the threshold, is

$$2n_s = 12.3 \text{ samples} \tag{6-23}$$

By comparing this time constant to those in Figure 8, it can be seen that the DDM technique converges faster than the Learning With Teacher scheme for small separation of A_o and A_1 and that the Learning With Teacher scheme is faster for large separation of A_o and A_1. These observations deal only with the situation shown in Figure 8. Curves similar to those in Figure 8 for other situations would be necessary in order to make a general comparison between the Learning With Teacher method and the DDM technique.

Performance of Noiseless System

For certain selections of $\hat{x}_o$ and $\hat{y}_o$ and a noiseless system (Ref. 3, p. 37), it is possible that the estimated threshold does not converge to the proper value. As an example, consider the following case:

$$p\left(x_i \mid o\right) = 1.0\ \delta\left(x_i - 20\right)$$

$$p\left(x_i \mid 1\right) = 1.0\ \delta\left(x_i - 30\right)$$

with $\hat{x}_o = 15$ and $\hat{y}_o = 10$. This case is illustrated in Figure 9. Because of the convergence of the estimation technique, $\hat{x}_k$ moves toward the mean of the signals above the threshold. For the case in Figure 9 and for $p(o) = p(1) = 0.5$,

$$\lim_{k \to \infty} E\left\{\hat{x}_k\right\} = \frac{20 + 30}{2} = 25 \tag{6-24}$$

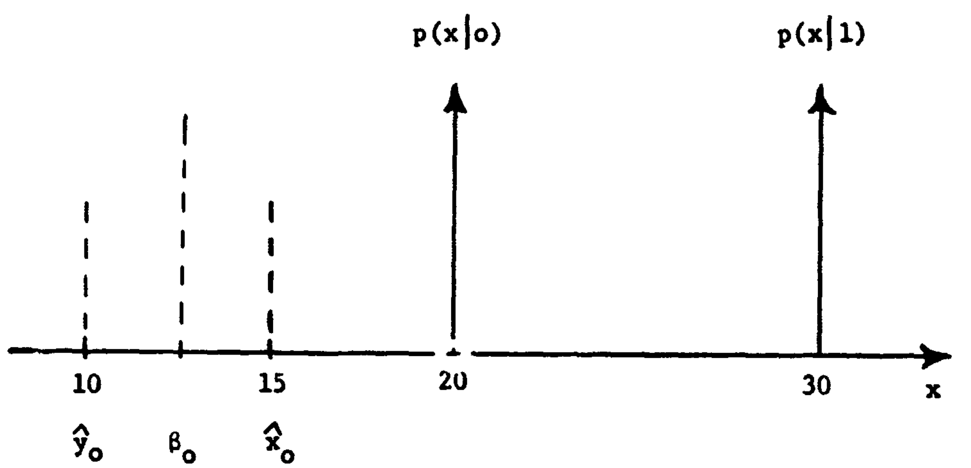

Figure 9.- Illustration of noiseless case.

There can never be any received signals falling below the threshold so that $\hat{y}_o$ is never updated. Therefore, the limiting value of the threshold is

$$\lim_{k \to \infty} E\left\{\beta_k\right\} = \frac{1}{2}\left[\lim_{k \to \infty} E\left\{\hat{y}_k\right\} + \lim_{k \to \infty} E\left\{x_k\right\}\right]$$

$$= \frac{1}{2}\left[\hat{y}_o + \lim_{k \to \infty} E\left\{\hat{x}_k\right\}\right]$$

$$= \frac{1}{2}\left[10 + 25\right] = 17.5 \qquad (6\text{-}25)$$

Therefore, the threshold has not converged to the proper value.

This possibility is eliminated if $\hat{x}_o$ and $\hat{y}_o$ are chosen as previously discussed in this section ($\hat{x}_o$ as small as possible and $\hat{y}_o$ are large as possible). This insures that $\hat{y}_o$ is always above the final location of the threshold and that $\hat{x}_o$ is always below the threshold location. In order to prevent the possibility of a "hang-up" situation occurring due to a change in signal characteristics, it is necessary to reset $\hat{x}_o$ to the smallest possible value and $\hat{y}_o$ to the largest possible value each time the location of the synchronization code is lost.

CHAPTER VII

CALCULATION OF THE PROBABILITY OF ERROR

The probability of error can be determined for a system using an estimated threshold in place of the optimum threshold. For the case of equally probable signals (K = 1) the optimum location of the threshold in the Bayes sense is

$$\beta = \frac{A_1 + A_o}{2}$$

If it is assumed that the adaptive detector has been moved sufficiently close to the optimum threshold by the Decision-Directed-Measurement technique that the Learning With Teacher scheme can be used and if it is assumed that the detector has processed a very large number of samples from a stationary environment, the asymptotic probability of error can be investigated. Since the estimates of the means are random functions, the threshold and the probability of error are random functions. The average probability of error can be obtained by using the probability of error as a function of the threshold location and the probability density of the location of the threshold.

The threshold is located by the adaptive system at

$$\beta_\infty = \frac{\hat{x}_\infty + \hat{y}_\infty}{2}$$

As discussed in Chapter III (Estimation of the Mean), the variables $\hat{x}_\infty$ and $\hat{y}_\infty$ both have a Gaussian probability density with

$$\text{variance of } \hat{x}_\infty = \text{variance of } \hat{y}_\infty = \left(\frac{1 - A}{1 + A}\right) \sigma_n^2$$

$$E\left\{\hat{x}_\infty\right\} = A_1$$

and

$$E\left\{\hat{y}_\infty\right\} = A_o$$

The two variables $\hat{x}_\infty$ and $\hat{y}_\infty$ are independent because the noise has been assumed to be white and Gaussian. The threshold, β_∞, has a Gaussian distribution with a mean of

$$E\left\{\beta_\infty\right\} = \frac{A_1 + A_o}{2}$$

and

$$\text{variance of } \beta_\infty = \frac{1}{2}\left(\frac{1 - A}{1 + A}\right) \sigma_n^2$$

The probability of error as a function of the threshold location is

$$\text{prob. of error } (\beta_\infty) = \frac{1}{2\sqrt{2\pi}\,\sigma_n} \left[\int_{\beta_\infty}^{\infty} \exp\left[-\frac{\left(x - A_o\right)^2}{2\sigma_n^2}\right] dx + \int_{-\infty}^{\beta_\infty} \exp\left[-\frac{\left(x - A_1\right)^2}{2\sigma_n^2}\right] dx\right]$$

The probability density of the threshold is

$$p(\beta_\infty) = \frac{1}{\sqrt{\pi}\sqrt{\frac{1 - A}{1 + A}}\,\sigma_n} \exp\left[-\frac{\left(\beta_\infty - \frac{A_1 + A_o}{2}\right)^2}{\left(\frac{1 - A}{1 + A}\right)\sigma_n^2}\right]$$

The average value of the probability of error is

$$E\left\{\text{prob. of error}\right\} = \int_{-\infty}^{\infty} p(\beta_\infty) \left[\text{prob. of error } (\beta_\infty)\right] d\beta_\infty$$

The average probability of error has been calculated on a digital computer for several values of A and for several values of SNR where

$$\text{SNR} = \frac{A_1 - A_o}{\sigma_n}$$

The following table shows the results of these calculations:

SNR	Average probability of error A = 0.85	A = 0.90	A = 0.95	A = 1.0
3.33	9.25×10^{-2}	9.07×10^{-2}	8.7×10^{-2}	4.74×10^{-2}
4.0	4.53×10^{-2}	4.37×10^{-2}	4.1×10^{-2}	2.27×10^{-2}
5.0	1.28×10^{-2}	1.21×10^{-2}	1.1×10^{-2}	6.21×10^{-3}
6.67	9.56×10^{-4}	8.6×10^{-4}	7.5×10^{-4}	4.30×10^{-4}
10.0	7.93×10^{-7}	6.42×10^{-7}	5.0×10^{-7}	2.87×10^{-7}
20.0	7.8×10^{-23}	3.7×10^{-23}	1.69×10^{-23}	7.62×10^{-24}

Figure 10 shows a plot of the results for A = 0.85 and A = 1.0. The results for A = 1.0 correspond to the optimum detector for this situation. All other values of A cause the average probability of error to be higher than for the case of A = 1.0. It should be noted that Figure 10 gives the average probability of error. The actual probability of error is a random function which has a minimum given by the curve for A = 1.0. The value of A can be chosen as close to 1.0 as desired in order to reduce the probability of error, but

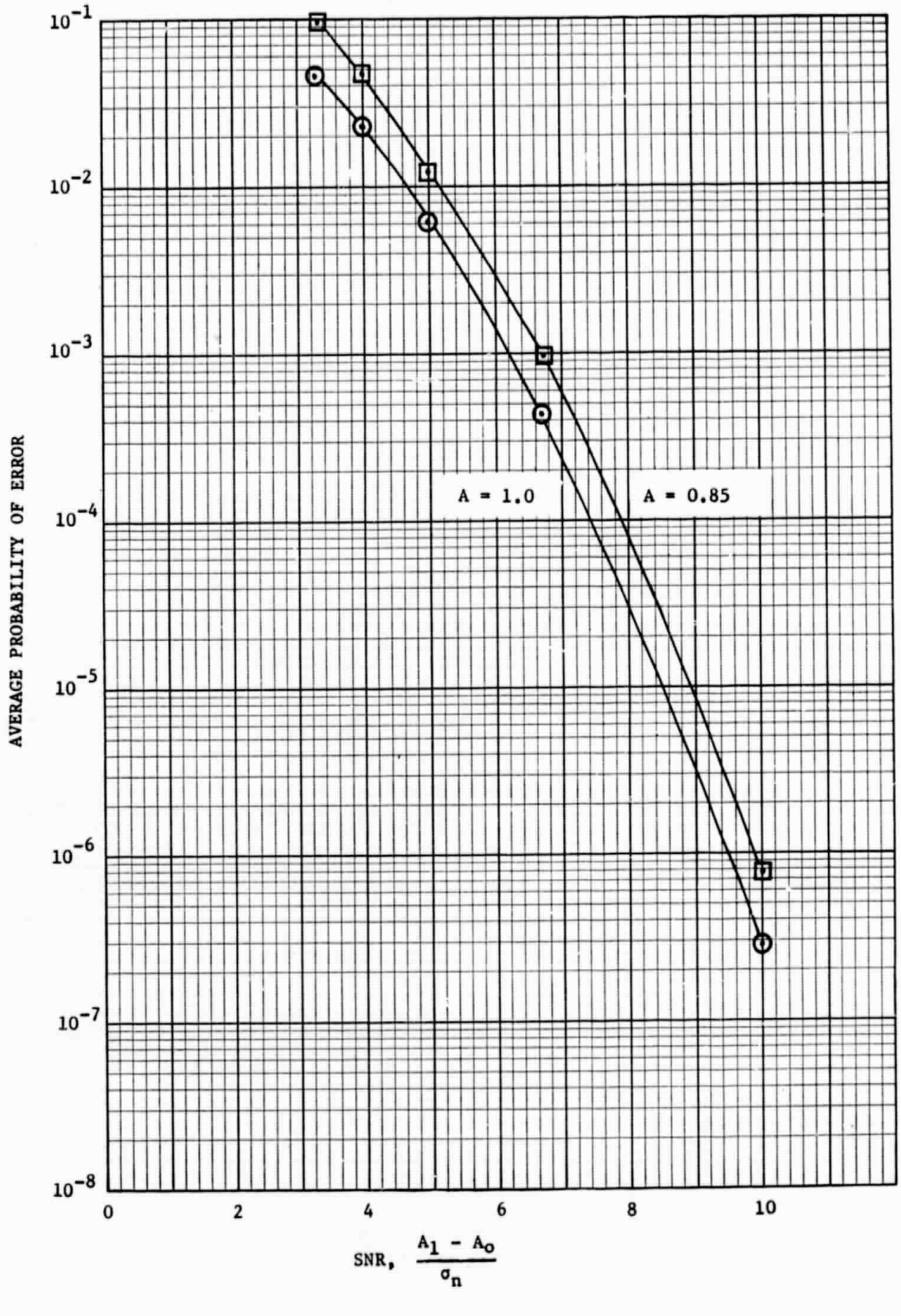

Figure 10.- Probability of error.

increasing A causes an increase in the reaction time of the estimation equation as discussed in Chapter III.

Since Patrick and Costello (Ref. 1) only make their calculations for an infinite number of samples used to compute the sample mean, the only comparison between these results and those of Reference 1 that can be made is for $A = 1$ with equally probable signals. At this point the estimates have zero variance and the adaptive detector has the same probability of error as the optimum detector. The additional error found in Reference 1 is due to an unsymmetrical bias caused by non-equally probable signals. No attempt was made by Patrick and Costello (Ref. 1) to compensate for the effect of the nonequally probable signals.

A calculation of the probability of error of the DDM system with nonequally probable signals and with A not equal to 1.0 would show the ability of the DDM system to properly locate the threshold. This analysis would be more complicated than the analysis shown in Reference 1 because values of A other than 1.0 have the effect of using less than an infinite number of samples in the estimation process and because an estimation of the variance is included in the system proposed here to reduce the additional error due to nonequally probable signals. This estimate of the variance is also a biased estimate when operating in conjunction with the DDM technique for the same reason (p. 15) that the estimates of the means were biased. Since the input to the estimator of the variance is not Gaussian in the DDM mode, the calculation of the variance of the estimated variance shown in Chapter V does not apply. This calculation of the probability of error has not been included as a part of this investigation.

An example can illustrate a situation in which the adaptive detector would give a lower probability of error than a degraded optimum detector. Assume than an optimum detector has been constructed and is operating in an environment of zero-mean, white, Gaussian noise. Let an enemy in the neighborhood of the transmitter begin to transmit a continuous sequence of signals which correspond exactly to the binary 0 signal and which are exactly in synchronization with the data bits. Before the enemy began to transmit, the equation, (2-1), implemented by the optimum detector was

$$u = W^T\left(S_1 - S_o\right) - b$$

The equation implemented by the optimum detector after the beginning of transmission of the enemy is

$$u = W^T\left(S_1 - S_o\right) - b + A_c{}'S_o^T\left(S_1 - S_o\right)$$

where $A_c{}'$ is the channel gain of the channel from the enemy transmitter to the receiver. Since $A_c{}'S_o^T\left(S_1 - S_o\right)$ is a constant and is present for both binary 0 and binary 1 signals, the additional signal $A_c{}'S_o^T\left(S_1 - S_o\right)$ has the appearance of a nonzero mean of the noise. This would cause the optimum detector to be operating with its threshold located at a nonoptimum location.

Figure 11 shows a plot of the probability of error of an optimum detector as a function of the location of the threshold. The abscissa

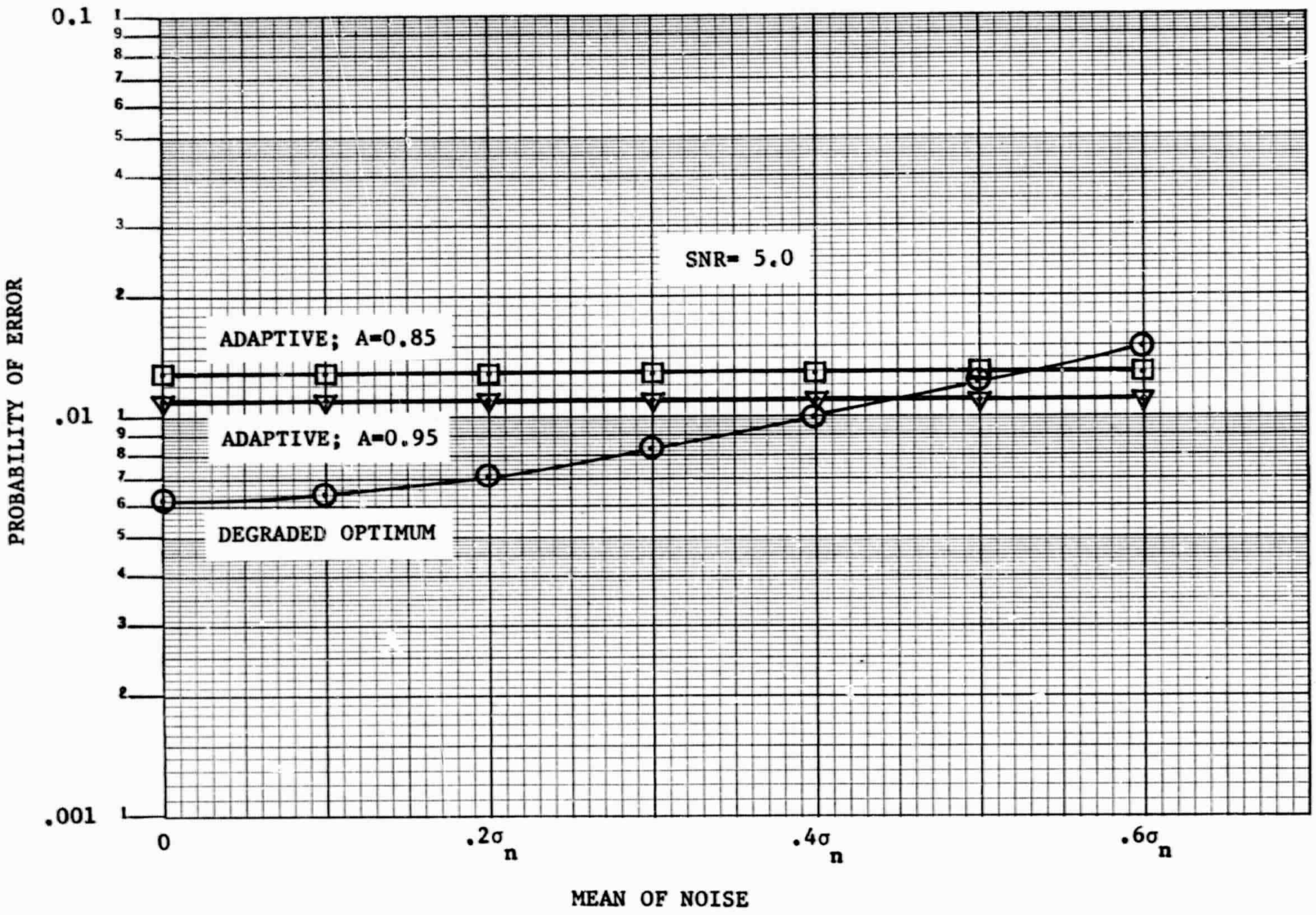

Figure 11.- Comparison of optimum and adaptive detectors.

is expressed in terms of the deviation of the noise. For instance, if the signal-to-noise ratio were 5.0 and if the mean of the noise were $0.1\ \sigma_n$, the degraded optimum detector would have a probability of error equal to 0.00643. If the mean of the noise is zero, the optimum detector has a probability of error of 0.00621.

An adaptive detector operating in this same situation has an average probability of error which does not change as the mean of the noise changes if sufficient time is allowed for any transients to disappear. The probability of error of the adaptive detector is higher than that of the optimum detector when the optimum detector is using the optimum threshold. As the mean of the noise increases, the probability of error of the degraded optimum detector increases while the probability of error of the adaptive system remains constant. Figure 11 shows the points where the probabilities of error of the two systems are equal for a given signal-to-noise ratio and for a given recursive constant. For a signal-to-noise ratio of 5.0, the adaptive system using $A = 0.85$ has a lower probability of error than a degraded optimum system for values of $A_c{}'S_o^T(S_1 - S_o)$ greater than approximately $0.53\sigma_n$. For values of A greater than 0.85, the point at which the two systems have equal probability of error is decreased. As the signal-to-noise ratio increases the point at which the two systems have equal probability of error also decreases. This fact can be shown from curves similar to Figure 11 for other signal-to-noise ratios, but have not been included here.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

The results of this investigation show that the system proposed here is capable of acting as an adaptive detector. The system requires an estimate of the mean and an estimate of the variance of a sequence of random numbers. The mean is estimated by

$$\hat{x}_k = A\hat{x}_{k-1} + (1 - A)x_k$$

and the variance is estimated by

$$\hat{\sigma}_k^{\,2} = B\hat{\sigma}_{k-1}^{\,2} + \frac{(1 + A)(1 - B)}{2}\left(x_k - \hat{x}_{k-1}\right)^2$$

If the input data have a Gaussian distribution with mean of "a" and variance of "σ^2", the estimated variance has a variance of

$$\lim_{k \to \infty} \text{variance of } \hat{\sigma}_k^{\,2} = \left(\frac{1 - B}{1 + B}\right)\left[2 + \frac{B(1 - A)^2}{1 - A^2B}\right]\sigma^4$$

The accuracy and time response of each estimator can be varied by the choice of constants A and B. The frame synchronization code is used as the teacher in a "Learning With Teacher" technique. A Decision Directed Measurement technique is used when the frame synchronization code is not available.

It has been found that in the recursive estimate of the mean, the more accurate the estimate, the slower the convergence. The optimum location of $\hat{x}_o$ and $\hat{y}_o$ when using the Decision Directed

Measurement technique is found by a computer study to be $\hat{x}_o$ as small as possible and $\hat{y}_o$ as large as possible. This selection of $\hat{x}_o$ and $\hat{y}_o$ is found to prevent the possibility of a "hang-up" when there is no noise. The asymptotic, average probability of error for the adaptive system using only the estimates of the means in the Learning With Teacher mode is found to be equal to that of the optimum detector for $A = 1.0$ and greater than the optimum detector for $A < 1.0$. However, a change in the environment can cause the adaptive detector to have a lower probability of error than an optimum detector, which cannot track the changes.

This investigation has by no means completely analyzed the proposed system. Some of the areas which offer possibilities for future work are discussed here. One of the most important areas for future work is the construction of hardware to perform the functions discussed in this dissertation. R. G. Brown (Ref. 4) mentions other techniques of estimation of the mean which are essentially higher orders of the technique used here. It would be interesting to attempt to use some of the other techniques and to compare their results to those of the exponential smoothing method. The time response of the estimation of the variance is unknown and should be determined, but it will be difficult to determine due to the nonlinearity of the estimation technique. The probability of error of the adaptive system in the Learning With Teacher mode for $K \neq 1$ also should be determined. The operation of the "Decision-Directed-Measurement" technique also has some areas which require more investigation. The effect of all the system variables on the convergence rate of the system needs to be determined. All of the techniques discussed here need to be investigated when operating in an environment of correlated noise and correlation between adjacent data samples. The probability of error for this adaptive system in the DDM mode should be calculated. The operation of the DDM mode when the variance is estimated also needs to be investigated.

PRECEDING PAGE BLANK NOT FILMED.

REFERENCES

1. Patrick, Edward A.; and J. P. Costello: Asymptotic Probability of Error Using Two Decision - Directed Estimators for Two Unknown Mean Vectors. IEEE Transactions on Information Theory, Vol. IT-14, No. 1, January 1968, pp. 160-162.

2. Abramson, N.: Learning in Pattern Recognition. First Technical Note, AF 30(602) - 2278, AD 267-798, ITT Federal Labs, Palo Alto, California, October 1961, p. 32.

3. Lindenlaub, J. C.; and D. F. Mix: Learning Theory Applied to Communications. Technical Report AFAL-TR-65-273, AD 474-097, Purdue University, Lafayette, Indiana, October 1965.

4. Brown, Robert Goodell: Smoothing Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Inc., 1962, pp. 97-205.

5. Abramson, N.; and D. Braverman: Learning to Recognize Patterns in a Random Environment. IRE Transactions on Information Theory, Vol. IT-8, No. 5, 1962, pp. 58-63.

6. Scudder, Henry J.: Adaptive Communication Receivers. Report No. 64-3, AF-AFOSR 230-63, National Science Foundation, April 29, 1964.

7. Hancock, John C.; and Paul A. Wintz: Signal Detection Theory. McGraw-Hill, 1966.

8. Fisz, Marek: Probability Theory and Mathematical Statistics. John Wiley & Sons, 1963.

9. Davenport, Wilbur B., Jr.; and William L. Root: An Introduction to the Theory of Random Signals and Noise. McGraw-Hill, 1958, pp. 154-157.

10. Hancock, J. C.; and D. F. Mix: Learning in Single and Two Threshold Decision Making Systems. Technical Report AFAL - TR-65-12, AD612909, Purdue University, Lafayette, Indiana, February 1965.

11. Papoulis, Athanasios: Probability, Random Variables, and Stochastic Processes McGraw-Hill, 1965.

12. Tou, Julius T.: Digital and Sampled-Data Control Systems. McGraw-Hill, 1959.

13. Sokolnikoff, I. S.; and R. M. Redheffer: Mathematics of Physics and Modern Engineering. McGraw-Hill, 1958, p. 115.

14. Cooper, David B.; and Paul W. Cooper: Adaptive Pattern Recognition and Signal Detection Without Supervision. IEEE International Convention Record, Part 1, March 1964, pp. 246-256.

15. Groginsky, H. L.; I. R Wilson; and David Middleton: Adaptive Detection of Statistical Signals in Noise. IEEE Transactions on Information Theory, Vol. IT-12, No. 3, July 1966, pp. 337-348.

16. Hancock, J. C.; and W. D. Gregg: On the Optimization of Mixture Resolving Signal Processing Structures. Purdue University, Lafayette, Indiana, TR-EE-66-11, October 1966.

17. Millard, J. B.; and L. Kurz: Adaptive Threshold Detection of M-ary Signals in Statistically Undefined Noise. IEEE Transactions on Communication Technology, Vol. Com-14, No. 5, October 1966, pp. 601-610.

18. Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the American Society of Mechanical Engineers, March 1960, pp. 35-45.

19. Lin, Ta-Tung; and Stephen S. Yau: Bayesian Approach to the Optimization of Adaptive Systems. IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-3, No. 2, November 1967, pp. 77-85.

20. Beine, G.: Adaptive Recognition of Random Binary Signals. Technical Report No. 80, AD-601982, Cornell University, Ithaca, New York, June 1964.

21. Dale, J. C.: Estimation of the Variance of a Stationary Random Process by Periodic Sampling. Bell System Technical Journal, July - August 1967, pp. 1283-1287.

22. Deutsch, Ralph: Estimation Theory. Prentice-Hall, Englewood Cliffs, New Jersey, 1965.

23. Good, I. J.: The Estimation of Probabilities. Research Monograph No. 30, The MIT Press, Cambridge, Massachusetts, 1965.

24. Spragins, John: Learning Without a Teacher. IEEE Transactions on Information Theory, Vol. IT-12, No. 2, April 1966, pp. 223-230.

PRECEDING PAGE BLANK NOT FILMED.

APPENDIX I

METHOD OF ESTIMATION OF $\frac{\sigma_n^2}{A_c}$

The purpose of this appendix is to show that

$$\frac{\sigma_n^2}{A_c} = \frac{\text{variance of } \left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}}{E\left\{\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)\right\} - E\left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}} \qquad \text{(A-1-1)}$$

The adaptive detector computes estimates of three properties of its input. These are

$$E\left\{X \mid W = A_c S_1 + N\right\} = E\left\{\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)\right\}$$

$$E\left\{X \mid W = A_c S_o + N\right\} = E\left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}$$

and

$$\text{variance of } \left\{X \mid W = A_c S_o + N\right\} = \text{variance of } \left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\}$$

The remaining portion of this appendix shows that the numerator of equation (A-1-1) is

$$\text{variance of } \left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\} = \sigma_n^2 \left(S_1 - S_o\right)^T \left(S_1 - S_o\right)$$

and that the denominator is

$$E\left\{\left(A_c S_1 + N\right)^T \left(S_1 - S_o\right)\right\} - E\left\{\left(A_c S_o + N\right)^T \left(S_1 - S_o\right)\right\} = A_c \left(S_1 - S_o\right)^T \left(S_1 - S_o\right)$$

The numerator is rearranged to yield

$$\text{variance of } \left(A_c S_o + N\right)^T \left(S_1 - S_o\right)$$

$$= E\left\{\left[\left(A_c S_o + N\right)^T\left(S_1 - S_o\right) - E\left\{\left(A_c S_o + N\right)^T\left(S_1 - S_o\right)\right\}\right]^T \left[\left(A_c S_o + N\right)^T\left(S_1 - S_o\right) - E\left\{\left(A_c S_o + N\right)^T\left(S_1 - S_o\right)\right\}\right]\right\}$$

Equation (2-7b) with $\epsilon = 0$ is used to reduce the complexity of the above equation, so that

$$\text{variance of } \left\{\left(A_c S_o + N\right)^T\left(S_1 - S_o\right)\right\}$$

$$= E\left\{\left[(N - \overline{N})^T\left(S_1 - S_o\right)\right]^T \left[(N - \overline{N})^T\left(S_1 - S_o\right)\right]\right\}$$

$$= E\left\{(N - \overline{N})^T (N - \overline{N})\right\} \left(S_1 - S_o\right)^T\left(S_1 - S_o\right)$$

$$= \sigma_n^{\,2}\left(S_1 - S_o\right)^T\left(S_1 - S_o\right) \qquad \text{(A-1-2)}$$

Identical results are obtained if variance of $\left\{\left(A_c S_1 + N\right)^T\left(S_1 - S_o\right)\right\}$ is computed.

The denominator of equation (A-1-1) is computed using (2-7a) and (2-7b) with $\epsilon = 0$, which yields

$$E\left\{\left(A_c S_1 + N\right)^T\left(S_1 - S_o\right)\right\} - E\left\{\left(A_c S_o + N\right)^T\left(S_1 - S_o\right)\right\}$$

$$= A_c S_1^{\,T} S_1 - A_c S_1^{\,T} S_o + \overline{N}^T\left(S_1 - S_o\right) - A_c S_o^{\,T} S_1 + A_c S_o^{\,T} S_o - \overline{N}^T\left(S_1 - S_o\right)$$

$$= A_c\left[S_1^TS_1 - S_1^TS_o - S_o^TS_1 - S_o^TS_o\right]$$

$$= A_c(S_1 - S_o)^T(S_1 - S_o) \tag{A-1-3}$$

Equation (A-1-1) results from the division of Equation (A-1-2) by Equation (A-1-3) as shown by

$$\frac{\text{variance of }\left\{(A_cS_o + N)^T(S_1 - S_o)\right\}}{E\left\{(A_cS_1 + N)^T(S_1 - S_o)\right\} - E\left\{(A_cS_o + N)^T(S_1 - S_o)\right\}}$$

$$= \frac{\sigma_n^2(S_1 - S_o)^T(S_1 - S_o)}{A_c(S_1 - S_o)^T(S_1 - S_o)}$$

$$= \frac{\sigma_n^2}{A_c}$$

APPENDIX II

ANALYSIS OF THE EXPONENTIAL SMOOTHING TECHNIQUE

This appendix is a derivation of some of the properties of the estimation of the mean by the exponential smoothing technique. The results shown here were published in Reference 4 by R. G. Brown; however, some of the results here are derived in a different manner.

$$\hat{x}_k = A\hat{x}_{k-1} + (1 - A)x_k; \; k = 1,2,3 \qquad \text{(A-2-1)}$$

where

$\hat{x}_k$ = kth estimate of the mean

x_k = kth data sample

A = recursive constant; $A < 1.0$

$\hat{x}_o$ = initial guess of mean

This equation can be compared with that in Reference 4 if $(1 - A)$ is set equal to α.

Derivation of the Mean of the Estimation

The general term of the estimation equation is rearranged by inserting an expression for $\hat{x}_{k-1}$ into the expression for $\hat{x}_k$, inserting an expression for $\hat{x}_{k-2}$, and continuing until $\hat{x}_o$ is reached. The resulting expression is

$$\hat{x}_k = A^k\hat{x}_o + (1 - A)\left[x_k + Ax_{k-1} + A^2x_{k-2} + \cdots + A^{k-1}x_1\right] \quad \text{(A-2-2)}$$

The mean value of $\hat{x}_k$ is

$$E\left\{\hat{x}_k\right\} = E\left\{A^k\hat{x}_o\right\} + (1 - A)\left[E\left\{x_k\right\} + AE\left\{x_{k-1}\right\} + \cdots + A^{k-1}\, E\left\{x_1\right\}\right]$$

The random function from which the samples, x_i, are taken is assumed to be stationary with a mean of "a" and a variance of "σ^2" so that

$$E\left\{x_i\right\} = a$$

The mean of $\hat{x}_k$ can be rewritten to yield

$$E\left\{\hat{x}_k\right\} = E\left\{A^k\hat{x}_o\right\} + (1 - A)\; E\left\{x_i\right\}\left[1 + A + \cdots + A^{k-1}\right]$$

$$= A^k\hat{x}_o + (1 - A)a\left[1 + A + A^2 + \cdots + A^{k-1}\right] \quad \text{(A-2-3)}$$

Since $|A| < 1.0$,

$$\lim_{k \to \infty} A^k\hat{x}_o = 0$$

and

$$\lim_{k \to \infty}\left(1 + A + A^2 + \cdots + A^{k-1}\right) = \frac{1}{1 - A}$$

Therefore,

$$\lim_{k \to \infty} E\left\{\hat{x}_k\right\} = a(1 - A)\left(\frac{1}{1 - A}\right)$$

$$= a \quad \text{(A-2-4)}$$

Derivation of the Variance of the Estimation

The variance of $\hat{x}_k$ is calculated by using

$$\text{variance of } \hat{x}_k = E\left\{\hat{x}_k^{\,2}\right\} - E^2\left\{\hat{x}_k\right\} \qquad \text{(A-2-5)}$$

The first two terms are

$$\text{variance of } \hat{x}_o = E\left\{\hat{x}_o^{\,2}\right\} - E^2\left\{\hat{x}_o\right\}$$

$$= \hat{x}_o^{\,2} - \hat{x}_o^{\,2} = 0 \qquad \text{(A-2-6)}$$

and

$$\text{variance of } \hat{x}_1 = E\left\{A^2\hat{x}_o^{\,2} + 2A(1 - A)x_1\hat{x}_o + (1 - A)^2 x_1^{\,2}\right\}$$

$$- \left[E\left\{A\hat{x}_o\right\} + E\left\{(1 - A)x_1\right\}\right]^2$$

$$= A^2\hat{x}_o^{\,2} + 2A(1 - A)a\hat{x}_o + (1 - A)^2\left(a^2 + \sigma^2\right)$$

$$- A^2\hat{x}_o^{\,2} - 2A(1 - A)a\hat{x}_o - (1 - A)^2 a^2$$

$$= (1 - A)^2\sigma^2 \qquad \text{(A-2-7)}$$

The variance for several values of k has been calculated by the same technique and is presented in the following table:

k	variance of $\hat{x}_k$
0	0
1	$(1 - A)^2\sigma^2$
2	$(1 + A^2)(1 - A)^2\sigma^2$
3	$(1 + A^2 + A^4)(1 - A)^2\sigma^2$
4	$(1 + A^2 + A^4 + A^6)(1 - A)^2\sigma^2$

The general expression of the variance of $\hat{x}_k$ is written from the table by inspection to yield

$$\text{variance of } \hat{x}_k = (1 - A)^2\sigma^2 \sum_{i=0}^{k-1} \left[A^2\right]^i \qquad \text{(A-2-8)}$$

As k approaches infinity,

$$\lim_{k \to \infty} \text{variance of } \hat{x}_k = (1 - A)^2\sigma^2\left(\frac{1}{1 - A^2}\right)$$

$$\lim_{k \to \infty} \text{variance of } \hat{x}_k = \left(\frac{1 - A}{1 + A}\right)\sigma^2 \qquad \text{(A-2-9)}$$

Derivation of the Time Constant of the Estimation

Since the estimation equation must also react to step changes in the mean of the incoming data, it is desirable to determine the time required to respond to a step change. The estimation equation is analyzed as if it were a filter by the use of the z-transform method (Ref. 12). The impulse response of the following equation is found:

$$\hat{x}_k = A\hat{x}_{k-1} + (1 - A)x_k$$

Let

$$\hat{x}_{-1} = 0$$

$$x_o = 1$$

$$x_i = 0 \quad \text{for} \quad i \neq 0$$

This set of conditions determines the response of the estimation technique to an input of a unit impulse at $t = 0$. From the definition of the z-transform (Ref. 12, p. 145)

$$\hat{X}(z) = \sum_{k=o}^{\infty} \hat{x}_k z^{-k}$$

$$\hat{X}(z) = (1 - A)z^o + A(1 - A)z^{-1} + A^2(1 - A)z^{-2} + \cdots$$

$$\hat{X}(z) = (1 - A)\left(1 + Az^{-1} + A^2 z^{-2} + \ldots\right)$$

$$\hat{X}(z) = (1 - A)\frac{1}{1 - Az^{-1}}$$

$$\hat{X}(z) = \frac{(1 - A)z}{z - A} \qquad \text{(A-2-10)}$$

The Laplace transform which corresponds to the z-transform given above is

$$H(s) = \frac{(1 - A)}{s - \frac{1}{T}\ln A} \qquad \text{(A-2-11)}$$

APPENDIX III

DERIVATION OF THE $\lim\limits_{k \to \infty}$ VARIANCE OF $\left(x_k - \hat{x}_{k-1}\right)^2$

The first moment of $\left(x_k - \hat{x}_{k-1}\right)^2$ is

$$E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} = E\left\{x_k^{\,2} - 2x_k\,\hat{x}_{k-1} + \hat{x}_{k-1}^{\,2}\right\}$$

$$= E\left\{x_k^{\,2}\right\} - 2E\left\{x_k\right\} E\left\{\hat{x}_{k-1}\right\} + E\left\{\hat{x}_{k-1}^{\,2}\right\}$$

This step can be made since $\hat{x}_{k-1}$ and x_k are independent.

$$\lim_{k \to \infty} E\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\} = a^2 + \sigma^2 - 2a \lim_{k \to \infty} E\left\{\hat{x}_{k-1}\right\} + \lim_{k \to \infty} E\left\{\hat{x}_{k-1}^{\,2}\right\}$$

$$= a^2 + \sigma^2 - 2a\ (a) + a^2 + \left(\frac{1 - A}{1 + A}\right)\sigma^2$$

$$= \left[1 + \frac{1 - A}{1 + A}\right]\sigma^2$$

$$= \frac{2\ \sigma^2}{1 + A}$$

The second moment is

$$E\left\{\left(x_k - \hat{x}_{k-1}\right)^4\right\} = E\left\{x_k^{\,4} - 4\ x_k^{\,3}\ \hat{x}_{k-1} + 6\ x_k^{\,2}\ \hat{x}_{k-1}^{\,2} - 4\ x_k\ \hat{x}_{k-1}^{\,3} + \hat{x}_{k-1}^{\,4}\right\}$$

As discussed in Chapter III, x_k and $\hat{x}_{k-1}$ both have a Gaussian distribution with known mean and variance and are independent. Substitutions from Appendix VII yield

$$\lim_{k \to \infty} E\left\{\left(x_k - \hat{x}_{k-1}\right)^4\right\} = a^4 + 6\,a^2\,\sigma^2 + 3\sigma^4 - 4\,a^4 - 12\,a^2\,\sigma^2 + 6\,a^4$$

$$+ 6\,a^2\,\sigma^2 + 6\left(\frac{1 - A}{1 + A}\right) a^2\,\sigma^2 + 6\left(\frac{1 - A}{1 + A}\right)\sigma^4$$

$$- 4a^4 - 12\left(\frac{1 - A}{1 + A}\right) a^2\,\sigma^2 + a^4 + 6\left(\frac{1 - A}{1 + A}\right) a^2\,\sigma^2$$

$$+ 3\left(\frac{1 - A}{1 + A}\right)^2 \sigma^4$$

$$= 3\sigma^4 + 6\left(\frac{1 - A}{1 + A}\right)\sigma^4 + 3\left(\frac{1 - A}{1 + A}\right)^2 \sigma^4$$

$$= 3\sigma^4\left[1 + \frac{1 - A}{1 + A}\right]^2$$

$$= \frac{12\,\sigma^4}{(1 + A)^2}$$

The limiting value is

$$\lim_{k \to \infty}\left[\text{variance of } \left(x_k - \hat{x}_{k-1}\right)^2\right] = \lim_{k \to \infty} E\left\{\left(x_k - \hat{x}_{k-1}\right)^4\right\}$$

$$- \lim_{k \to \infty} E^2\left\{\left(x_k - \hat{x}_{k-1}\right)^2\right\}$$

$$= \frac{12\,\sigma^4}{(1 + A)^2} - \frac{4\,\sigma^4}{(1 + A)^2}$$

$$= \frac{8\,\sigma^4}{(1 + A)^2}$$

APPENDIX IV

DERIVATION OF $\lim_{k \to \infty} E\left\{\hat{x}_k \hat{\sigma}_k^2\right\}$

$$E\left\{\hat{x}_k \hat{\sigma}_k^2\right\} = E\left\{\left[A\hat{x}_{k-1} + (1 - A)x_k\right]\left[B \hat{\sigma}_{k-1}^2 + \frac{(1 + A)(1 - B)}{2}(x_k - \hat{x}_{k-1})^2\right]\right\}$$

$$= A\ B\ E\left\{\hat{x}_{k-1} \hat{\sigma}_{k-1}^2\right\} + B(1 - A)\ E\left\{x_k \hat{\sigma}_{k-1}^2\right\}$$

$$+ \frac{A(1 + A)(1 - B)}{2}\left[E\left\{x_k^2 \hat{x}_{k-1}\right\} - 2E\left\{x_k \hat{x}_{k-1}^2\right\} + E\left\{\hat{x}_{k-1}^3\right\}\right]$$

$$+ \frac{(1 - A)(1 + A)(1 - B)}{2}\left[E\left\{x_k^3\right\} - 2E\left\{x_k^2 \hat{x}_{k-1}\right\}\right.$$

$$\left. + E\left\{x_k \hat{x}_{k-1}^2\right\}\right] \qquad \text{(A-4-1)}$$

Since x_k is independent of $\hat{x}_{k-1}$ and $\hat{\sigma}_{k-1}^2$, the expected values of the product of these variables can be separated into the product of the expected values. Using Appendix VII, Equation (A-4-1) becomes

$$E\left\{\hat{x}_k \hat{\sigma}_k^2\right\} = A\ B\ E\left\{\hat{x}_{k-1} \hat{\sigma}_{k-1}^2\right\} + B(1 - A)a\ E\left\{\hat{\sigma}_{k-1}^2\right\}$$

$$+ \frac{(1 + A)(1 - B)}{2}\left[(1 - A)\left(a^3 + 3a\ \sigma^2\right)\right.$$

$$+ (3A - 2)\left(a^2 + \sigma^2\right) E\left\{\hat{x}_{k-1}\right\}$$

$$\left. + (1 - 3A)(a)\ E\left\{\hat{x}_{k-1}^2\right\} + A\ E\left\{\hat{x}_{k-1}^3\right\}\right] \qquad \text{(A-4-2)}$$

The technique for finding the limit as k approaches infinity of this recursive equation is the same as that used in the main body of this report for the variance of the estimated variance. The index k

is set equal to n where n has a value large enough so that all terms on the right hand side of Equation (A-4-2) except $E\left\{\hat{x}_{k-1}\ \hat{\sigma}_{k-1}{}^2\right\}$ have become infinitesimally close to their limiting values. Using Appendix VII and Equations (A-2-4) and (A-2-9), Equation (A-4-2) becomes

$$\begin{aligned}
E\left\{\hat{x}_n\ \hat{\sigma}_n{}^2\right\} &= A\ B\ E\left\{\hat{x}_{n-1}\ \hat{\sigma}_{n-1}{}^2\right\} + B(1 - A)a\ \sigma^2 \\
&\quad + \frac{(1 + A)(1 - B)}{2}\Bigg[(1 - A)\left(a^3 + 3a\ \sigma^2\right) \\
&\quad + (3A - 2)\left(a^2 + \sigma^2\right)a + (1 - 3A)(a)\left(a^2 + \left[\frac{1 - A}{1 + A}\right]\sigma^2\right) \\
&\quad + A\left[a^3 + 3\left(\frac{1 - A}{1 + A}\right)a\ \sigma^2\right]\Bigg] \\
&= A\ B\ E\left\{\hat{x}_{n-1}\ \hat{\sigma}_{n-1}{}^2\right\} + (1 - A\ B)a\ \sigma^2 \qquad \text{(A-4-3)}
\end{aligned}$$

Several terms are calculated in order to recognize the series being generated. Let

$$E\left\{\hat{x}_{n-1}\ \hat{\sigma}_{n-1}{}^2\right\} = P$$

Then

$$E\left\{\hat{x}_n\ \hat{\sigma}_n{}^2\right\} = A\ B(P) + (1 - A\ B)a\ \sigma^2$$

$$E\left\{\hat{x}_{n+1}\ \hat{\sigma}_{n+1}{}^2\right\} = A^2\ B^2(P) + (1 + A\ B)(1 - A\ B)a\ \sigma^2$$

$$E\left\{\hat{x}_{n+2}\ \hat{\sigma}_{n+2}{}^2\right\} = A^3\ B^3(P) + \left(1 + A\ B + A^2\ B^2\right)(1 - A\ B)a\ \sigma^2$$

and

$$E\left\{\hat{x}_{n+i}\ \hat{\sigma}_{n+i}\right\} = A^{i+1}\ B^{i+1}(P) + (1 - A\ B)a\ \sigma^2 \sum_{j=0}^{i-1} (AB)^j \qquad (A\text{-}4\text{-}4)$$

Since $A < 1$ and $B < 1$, (A-4-4) becomes

$$\lim_{i \to \infty} E\left\{\hat{x}_{n+i}\ \hat{\sigma}_{n+i}^{\ 2}\right\} = \lim_{k \to \infty} E\left\{\hat{x}_k\ \hat{\sigma}_k^{\ 2}\right\} = \frac{(1 - A\ B)a\ \sigma^2}{1 - A\ B}$$

$$= a\ \sigma^2 \qquad (A\text{-}4\text{-}5)$$

The convergence of this series has been verified by calculating the exact expression for the series for $k = 0$, 1, and 2 but has not been included because of its length. From these expressions it is possible to recognize the general expression for the coefficients of all terms in the expression. It is found that the limit of coefficients of all terms approached zero as k approached infinity except for the coefficient of $a\sigma^2$. This coefficient is found to approach 1.0.

APPENDIX V

DERIVATION OF $\lim_{k \to \infty} E\left\{\hat{x}_k^2 \, \hat{\sigma}_k^2\right\}$

$$E\left\{\hat{x}_k^2 \, \hat{\sigma}_k^2\right\} = E\left\{\left[A\,\hat{x}_{k-1} + (1 - A)x_k\right]^2 \left[B\,\hat{\sigma}_{k-1}^2 + \frac{(1 + A)(1 - B)}{2}\left(x_k - \hat{x}_{k-1}\right)^2\right]\right\}$$

$$= A^2 B\, E\left\{\hat{x}_{k-1}^2\, \hat{\sigma}_{k-1}^2\right\} + 2A\, B(1 - A)a\, E\left\{\hat{x}_{k-1}\, \hat{\sigma}_{k-1}^2\right\}$$

$$+ B(1 - A)^2\left(a^2 + \sigma^2\right) E\left\{\hat{\sigma}_{k-1}^2\right\}$$

$$+ \frac{(1 + A)(1 - B)}{2}\left[\left(a^4 + 6a^2\,\sigma^2 + 3\sigma^4\right)(1 - A)^2\right.$$

$$+ E\left\{\hat{x}_{k-1}\right\}\left(a^3 + 3a\,\sigma^2\right)\left(-2 + 6A - 4A^2\right)$$

$$+ E\left\{\hat{x}_{k-1}^2\right\}\left(a^2 + \sigma^2\right)\left(1 - 6A + 6A^2\right)$$

$$\left. + E\left\{\hat{x}_{k-1}^3\right\}(a)\left(2A - 4A^2\right) + E\left\{\hat{x}_{k-1}^4\right\}\left(A^2\right)\right] \qquad \text{(A-5-1)}$$

Since x_k is independent of $\hat{x}_{k-1}$ and $\hat{\sigma}_{k-1}^2$ the expected value of the product of these variables is separated into the product of the expected values in the above expression.

The technique for finding the limit as k approaches infinity of this recursive equation is the same as that used in Appendix IV. The index k is set equal to n where n has a value large enough so that all terms on the right hand side of equation (A-5-1) except

$E\left\{\hat{x}_{k-1}{}^2\ \hat{\sigma}_{k-1}{}^2\right\}$ have become infinitesimally close to their limiting values. Using Appendix VII, Equation (A-5-1) becomes

$$E\left\{\hat{x}_n{}^2\ \hat{\sigma}_n{}^2\right\} = A^2\ B\ E\left\{\hat{x}_{n-1}{}^2\ \hat{\sigma}_{n-1}{}^2\right\} + 2A\ B(1 - A)a^2\ \sigma^2$$

$$+ B(1 - A)^2\ \left(a^2 + \sigma^2\right)\sigma^2 + \frac{(1 + A)(1 - B)}{2}\left[\left(a^4 + 6a^2\ \sigma^2\right.\right.$$

$$\left. + 3\sigma^4\right)(1 - A)^2 + \left(a^4 + 3a^2\ \sigma^2\right)\left(- 2 + 6A - 4A^2\right)$$

$$+ \left(a^2 + \left[\frac{1 - A}{1 + A}\right]\sigma^2\right)\left(a^2 + \sigma^2\right)\left(1 - 6A + 6A^2\right)$$

$$+ \left(a^3 + 3\left[\frac{1 - A}{1 + A}\right] a\ \sigma^2\right)(a)\ \left(2A - 4A^2\right)$$

$$\left. + \left(a^4 + 6\left[\frac{1 - A}{1 + A}\right] a^2\ \sigma^2 + 3\left[\frac{1 - A}{1 + A}\right]^2 \sigma^4\right)\left(A^2\right)\right]$$

$$= A^2\ B\ E\left\{\hat{x}_{n-1}{}^2\ \hat{\sigma}_{n-1}{}^2\right\} + \left(1 - A^2\ B\right)a^2\ \sigma^2$$

$$+ \frac{\sigma^4}{1 + A}\left[(1 - A)\left(1 - A^2\ B\right) + (1 - A)^2(1 - B)\right] \qquad \text{(A-5-2)}$$

$E\left\{\hat{x}_{n-1}{}^2\ \hat{\sigma}_{n-1}{}^2\right\}$ is set equal to an arbitrary constant Q, and several terms are calculated in order to recognize the series being generated:

$$E\left\{\hat{x}_{n-1}{}^2\ \hat{\sigma}_{n-1}{}^2\right\} = Q$$

$$E\left\{\hat{x}_n{}^2\ \sigma_n{}^2\right\} = A^2\ B(Q) + \left(1 - A^2\ B\right)a^2\ \sigma^2$$

$$+ \frac{\sigma^4}{1 + A}\left[(1 - A)\left(1 - A^2\ B\right)\right.$$

$$\left. + (1 - A)^2(1 - B)\right]$$

$$E\left\{\hat{x}_{n+1}^{\;2}\ \hat{\sigma}_{n+1}^{\;2}\right\} = A^4\ B^2(Q) + \left(1 + A^2\ B\right)\left[\left(1 - A^2\ B\right)a^2\ \sigma^2 + \frac{\sigma^4}{1 + A}\left[(1 - A)\left(1 - A^2\ B\right) + (1 - A)^2(1 - B)\right]\right]$$

The general term can be recognized to be

$$E\left\{\hat{x}_{n+j}^{\;2}\ \hat{\sigma}_{n+j}^{\;2}\right\} = A^{2(j+1)}\ B^{j+1}\ (Q) + \left[\left(1 - A^2\ B\right)a^2\ \sigma^2 + \frac{\sigma^4}{1 + A}\left[(1 - A)\left(1 - A^2\ B\right) + (1 - A)^2(1 - B)\right]\right]\sum_{i=0}^{j}\left(A^2\ B\right)^i \qquad \text{(A-5-3)}$$

Since $A^2\ B < 1.0$,

$$\lim_{k \to \infty} E\left\{\hat{x}_k^{\;2}\ \hat{\sigma}_k^{\;2}\right\} = \lim_{j \to \infty} E\left\{\hat{x}_{n+j}^{\;2}\ \hat{\sigma}_{n+j}^{\;2}\right\}$$

$$= \frac{1}{1 - A^2\ B}\left[\left(1 - A^2\ B\right)a^2\ \sigma^2 + \frac{\sigma^4}{1 + A}\left[(1 - A)\left(1 - A^2\ B\right) + (1 - A)^2(1 - B)\right]\right]$$

$$= a^2\ \sigma^2 + \sigma^4\left[\frac{1 - A}{1 + A} + \frac{(1 - A)^2\ (1 - B)}{(1 + A)\left(1 - A^2\ B\right)}\right] \qquad \text{(A-5-4)}$$

The limit of this sequence has been verified by the same method of verification discussed in Appendix IV.

APPENDIX VI

PROGRAM TO COMPUTE AVERAGE VALUE OF ESTIMATED THRESHOLD

```
      DIMENSION PT(64),PXY(64,64),STR(64,64)
      READ(5,5)SIG,AMN
    5 FORMAT(2F16.8)
      READ(5,2)(PT(I),I=1,64)
    2 FORMAT(4E16.8)
      BMN=70.0-AMN
      SNR=(AMN-BMN)/SIG
      PRINT 6,SIG,AMN,BMN,SNR
    6 FORMAT(8H1 SIGMA=F16.8,5X6HMEAN1=F16.8,5X6HMEAN0=F16.8,5X4HSNR=F16
     1.8)
      REC=0.85
      IX=30
      IY=10
      PRINT 7,IX,IY,REC
    7 FORMAT(2X14HINITIAL EST.1=I5,5X14HINITIAL EST.0=I5,5X12HREC. FACTO
     1R=F16.8)
      PRINT 8
    8 FORMAT(///2X9HSAMP. NO.8X12HAVG. THRESH./)
      DO 74 I=1,64
      DO 20 J=1,64
      IF(I-IX)15,14,15
   14 IF(J-IY)15,16,15
   16 PXY(I,J)=1.0
      GO TO 17
   15 PXY(I,J)=0.0
   17 STR(I,J)=0.0
   20 CONTINUE
   74 CONTINUE
      DO 19 IJK=1,50
      DO 75 I=1,64
```

```
      DO 23 J=1,64
      FAC=PXY(I,J)
      IF(FAC-0.0)23,23,97
   97 ITHRES=(FLOAT(I)+FLOAT(J))/2.0
      DO 76 NI=1,64
      IF(NI-ITHRES)96,98,98
   96 CAT=REC*FLOAT(J)+(1.0-REC)*FLOAT(NI)
      KAT=CAT+0.5
      IF(KAT-0)53,53,54
   53 KAT=KAT+1
   54 STR(I,KAT)=FAC*PT(NI)+STR(I,KAT)
      GO TO 76
   98 DOG=REC*FLOAT(I)+(1.0-REC)*FLOAT(NI)
      KLM=DOG+0.5
      IF(KLM-0)51,51,56
   51 KLM=KLM+1
   56 STR(KLM,J)=FAC*PT(NI)+STR(KLM,J)
   76 CONTINUE
   23 CONTINUE
   75 CONTINUE
      SUM=0.0
      DO 29 I=1,64
      DO 85 J=1,64
      SUM=SUM+STR(I,J)
   85 CONTINUE
   29 CONTINUE
      AVGT=0.0
      DO 34 I=1,64
      DO 31 J=1,64
      PXY(I,J)=STR(I,J)/SUM
```

```
      STR(I,J)=0.0
      AVGT=AVGT+(FLOAT(I)+FLOAT(J))*0.5*PXY(I,J)
   31 CONTINUE
   34 CONTINUE
      PRINT 4,IJK,AVGT
    4 FORMAT(I10,F16.8)
   19 CONTINUE
      STOP
      END
```

2.00000000	40.00000000	30.00000000		
1.11993396E-46	1.27211746E-43	1.12673493E-40	7.78235137E-38	1
4.19210440E-35	1.76127130E-32	5.77214688E-30	1.47575064E-27	2
2.94377305E-25	4.58212087E-23	5.56619973E-21	5.27769552E-19	3
3.90652497E-17	2.25771509E-15	1.01895594E-13	3.59193598E-12	4
9.89170423E-11	2.12847306E-09	3.57937164E-08	4.70516819E-07	5
4.83572127E-06	3.88643797E-05	2.44303879E-04	1.20136910E-03	6
4.62235471E-03	1.39173421E-02	3.27953085E-02	6.04887915E-02	7
8.73331968E-02	9.87067962E-02	8.73379967E-02	6.05276538E-02	8
3.30396122E-02	1.51187112E-02	9.24470942E-03	1.51187112E-02	9
3.30396122E-02	6.05276538E-02	8.73379967E-02	9.87067962E-02	10
8.73331968E-02	6.04887915E-02	3.27953085E-02	1.39173421E-02	11
4.62235471E-03	1.20136910E-03	2.44303879E-04	3.88643797E-05	12
4.83572127E-06	4.70516819E-07	3.57937164E-08	2.12847306E-09	13
9.89170423E-11	3.59193598E-12	1.01895594E-13	2.25771509E-15	14
3.90652497E-17	5.27769552E-19	5.56619973E-21	4.58212087E-23	15
2.94377305E-25	1.47575064E-27	5.77214688E-30	1.76127130E-32	16

APPENDIX VII

MOMENTS OF A GAUSSIAN VARIABLE

If x is a probabilistic variable having a Gaussian probability distribution with mean of "m" and variance of "v^2", the first four moments of x are (Ref. 11, p. 162):

$$E\{x\} = m$$

$$E\{x^2\} = m^2 + v^2$$

$$E\{x^3\} = m^3 + 3\,mv^2$$

$$E\{x^4\} = m^4 + 6m^2\,v^2 + 3v^4$$