1968

Nonparametric sequential detection

James Charles Fowler

NONPARAMETRIC SEQUENTIAL DETECTION

by

JAMES CHARLES FOWLER, 1943

A DISSERTATION

Presented to the Faculty of the Graduate School of the

UNIVERSITY OF MISSOURI AT ROLLA

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

1968

J. R. Betten
Advisor

R. E. D. Jones

F. J. Kern

G. G. Skitek

C. E. Antle

Samuel J. Dwyer III by JRB

ABSTRACT

This dissertation extends the theory of the Wilcoxon-Mann-Whitney U statistic so that this statistic can be used to perform sequential tests of hypotheses. This sequential test procedure makes use of a sequential ranking procedure similar to the one first introduced by Parent. The operating-characteristic function and average number of samples function for this new test are calculated as a function of the signal to noise ratio. The test is then shown to be efficient for several forms of alternatives with an efficiency of 95% against the Wald Sequential Probability Ratio Test for a constant signal in normal noise.

Finally, the test procedure is modified so that it is capable of making measurements on the channel in order to adapt itself to changes in the channel characteristics. Simulation results are presented to show that this adaptive detector can operate with low probability of error.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. John R. Betten for his suggestions and guidance throughout his course of work.

Expressions of thanks are also due to Dr. T. L. Noack and Dr. C. E. Antle for their help on several specific problems which arose during the course of this work.

The author also thanks his wife for her help and support during the entire course of his work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

SYMBOL	DEFINITION
α	Probability of Type I error
β	Probability of Type II error
μ_0	Mean of distribution under null hypothesis
μ_1	Mean of distribution under alternate hypothesis
σ_0^2	Variance of distribution under null hypothesis
σ_1^2	Variance of distribution under alternate hypothesis
H_0	Null hypothesis
H_1	Alternate hypothesis
λ_n	Wald's Sequential Probability Ratio statistic
$\Phi(x)$	Cumulative normal
$N(0,1)$	Normal distribution with zero mean and unit variance
$f(x)$	Probability density function
$F(x)$	Cumulative distribution function
T	Wilcoxon statistic
U	Mann-Whitney statistic
X^r	Rank vector
S_n	Sequential rank of the n^{th} sample
X	Noise channel
Y	Unknown channel
	Probability that $X > Y$

θ_1	Amount of shift associated with the alternate hypothesis
τ_n	Sequential W.M.W. U test statistic
$L(\gamma)$	Operating-characteristic function
$\overline{n(\gamma)}$	Average number of samples function

CHAPTER I

INTRODUCTION

The detection of signals buried in noise has been the principle problem facing the communications engineer since information was first transmitted to a distant point. Detecting a signal in additive noise consists of examining some unknown waveform and determining whether the waveform is random noise or signal plus random noise. Tracking distant targets by use of radar and the recovery of coded signals from distant transmitters are two examples of signal detection.

There are two basic approaches to the solution of the signal detection problem. The first is the parametric approach. The parametric detector requires knowledge about the statistical form of the input noise and the form of the signal to be detected. If the assumptions made about the form of the input are correct, this type of detector is very efficient in that it can detect the presence or absence of the given signal in a shorter time than any other type of detector. If, however, the assumptions as to the nature of the input noise are not correct or if the noise changes with time, it may be necessary to use a type of detector which does not require specific knowledge of the input waveform. This type of detector is called a nonparametric detector. It is nonparametric in the sense that the parameters of the distribution of the noise do not affect the design of the detector. This detector is not as efficient in specific cases as the parametric detector, but it is not subject to the restrictive assumptions associated with the parametric detector.

It has been found in the case of the parametric detector, that detection can be accomplished at a faster rate if a sequential decision process is used. That is, while the regular detectors examine the input for a fixed length of time and then make a decision, the sequential detector periodically tries to decide if a signal is present during the examination interval. Many times, the sequential detector can make a decision before the examination time of the fixed length detector is over.

This paper will extend the idea of sequential detection to include the nonparametric detector. The particular sequential nonparametric detector discussed here will then be compared to both the regular nonparametric detector and the sequential parametric detector in several different ways. After the theory is developed, the sequential nonparametric detector will be used to detect a pulse type signal in background noise.

The sequential nonparametric detector will then be modified so that it can adapt itself to changes in the noise or channel characteristics. After this adaptive process has been discussed, a simulation of the scheme will be made to demonstrate the ability of the adaptive detector in detecting a pulse type signal through an unknown channel.

CHAPTER II

FORMULATION OF THE DETECTION PROBLEM

2.1 Detection Using Fixed Sample Sizes

Given an observed waveform x(t), it is the function of the detector to determine whether x(t) consists of signal plus noise or noise alone. The detector will base its decision on some function of a set of samples from x(t). Thus, it is also assumed that the detector is capable of sampling x(t) at $t=t_1, t_2, \ldots, t_N$ giving the set of samples $x_1, x_2, \ldots, x_N$ where $(x = x(t_i)$, $i=1, 2, \ldots N)$. The detection problem can now be thought of as a statistical hypothesis testing problem (1). In hypothesis testing there are two alternatives, represented by the null hypothesis and the alternate hypothesis. The detector can be thought of as testing the null hypothesis (x(t) is noise alone) against the alternate hypothesis (x(t) is signal plus noise).

Fixed sample size detectors are characterized by their dichotomous decisions, i.e., either the null hypothesis, noise alone, or the alternate hypothesis, signal plus noise, is accepted. The decision is based upon the fact that some function of the sample is greater than or less than some threshold. This threshold is predetermined by the error rate which will be allowed. There are two types of errors which can be made. They are called Type I error and Type II error.

A Type I error occurs if the detector decides that a signal is present when there is no signal present. The probability of such an error is denoted by α, and it is sometimes called the probability of false alarm. A Type II error occurs if the detector decides that no signal is present when in fact a signal is present. The probability of this type of error will be denoted by β, and it is sometimes called the probability of false dismissal.

There are many special cases of the detection problem which can be obtained by making various assumptions about the signal and noise statistics. The case most often considered in texts and in the literature is that in which both the noise and signal distributions are known exactly. In this case the optimum detector is the Neyman-Pearson or likelihood detector (1,2,3). The Neyman-Pearson detector is optimum in the sense that for a given probability of Type I error, α, and for a given probability of Type II error, β, the sample size, N, is a minimum. As an example of a Neyman-Pearson detector, if the noise is assumed to have a normal distribution and the signal is a constant, the Neyman-Pearson detector is the well known t-detector (3). When the variance of the noise is known the statistic of the t-detector is given by

$$t = \frac{\left[\sum_{i=1}^{N} \frac{x_i}{N} - \mu_o\right] N^{\frac{1}{2}}}{\sigma_o} .$$

When the variance of the noise is unknown, the statistic of the t-detector is given by

$$t = \frac{\left[\sum_{i=1}^{N} \frac{x_i}{N} - \mu_o\right] \left[N(N-1)\right]^{\frac{1}{2}}}{\left\{\sum_{i=1}^{N} \left[x_i - \sum_{i=1}^{N} \frac{x_i}{N}\right]^2\right\}^{1/2}} .$$

If the variance is known, the t-detector tests the null hypothesis (the waveform x(t) has a normal distribution with mean μ_o and variance σ_o^2), i.e., x(t) is noise alone, against the alternate hypothesis (the waveform x(t) has a normal distribution with mean not equal to μ_o and variance σ_o^2), i.e., x(t) is signal plus noise. The null hypothesis is accepted if t has a value below some preset threshold level and is rejected if t is above this level. The threshold level is determined by considering t to have a normal distribution with zero mean and unit variance. For the case when the variance is unknown, the detector tests the null hypothesis (the waveform x(t) has a normal distribution with mean μ_o and unknown variance), i.e., x(t) is noise alone, against the alternate hypothesis (the waveform x(t) has a normal distribution with mean not equal to μ_o and variance unknown), i.e., x(t) is signal plus noise.

Again the null hypothesis is accepted if t has a value below the threshold level and is rejected if t is above this threshold level. Here, however, the threshold is determined by considering t to have a t-distribution with N-1 degrees of freedom.

The Neyman-Pearson detectors described above have proven to be very useful in the past, but as discussed before, their fixed sample size is sometimes a disadvantage. In some cases, for example in the phased array radar, it is advantageous to be able to make a decision as soon as possible as to the presence or absence of a signal in the input waveform. In these cases a sequential procedure may be used.

2.2 Sequential Detectors

Like the fixed sample size detector, the sequential detector must sample the input waveform x(t) and decide whether the null hypothesis (x(t) is noise alone) is true or the alternate hypothesis (x(t) is signal plus noise) is true. The sequential detector differs in two major respects from the fixed sample size detector. These are:

1. The sample size, N, is a random variable.
2. The detector must make one of three possible decisions after each sample is taken.

The sequential detector must take a sample and then calculate some function of this sample and compare it to two threshold levels. On the basis of this comparison, the detector decides to either accept the null hypothesis, accept the alternate hypothesis, or take another sample. The sequential detector can still make Type I and Type II errors as defined previously; but also, there may be a chance that the detector can not accept one of the hypotheses, and for a particular sample the test may not terminate. The design procedure for the sequential detector must take this latter possibility into consideration.

For the case when both the noise and signal distributions are known exactly, the optimum sequential detector is the Wald detector (4). The Wald detector is optimum in the sense that for a given probability of Type I error and a given probability of Type II error the average number of samples required for detection is a minimum. After each sample, the Wald detector forms the ratio

$$\lambda_m = \frac{P(x\ ,\ldots,x_m|H_1)}{P(x\ ,\ldots,x_m|H_0)}$$

where $P(x\ ,\ldots,x_m|H_1)$ is the probability that the observed sample occurred, given that the alternate hypothesis is true, and $P(x\ ,\ldots,x_m|H_0)$ represents the probability that the observed sample occurred given that the null hypothesis is true. The detector then compares

λ_m to two thresholds A and B with B<A. If $\lambda_m \geq A$, H_1 is accepted, if $\lambda_m \leq B$, H_0 is accepted, and if $B<\lambda_m<A$, another sample is taken. Wald has shown that the above test will terminate with probability one if B is a nondecreasing function of m and A is a nonincreasing function of m, where m is the number of samples. A simple example of a sequential detection problem is shown in Fig. 2.1.

2.3 Nonparametric Detectors

Both the Neyman-Pearson and Wald detectors described above are optimum in their respective ways, but there are several major drawbacks to their implementation. First, there must be a good description of both the signal and noise distributions, and if either one changes a new detector must be designed. Second, if the noise is nonstationary in nature it is impracticable to try to design a Neyman-Pearson or Wald detector because the density function of the noise is not fixed. Finally, if the noise does not have a normal distribution it is sometimes difficult to implement these optimum detectors because the mathematics becomes difficult to handle.

The need for a more general type of detector which does not have some of the above drawbacks, leads to a consideration of nonparametric detectors. These nonparametric detectors can be considered more general than

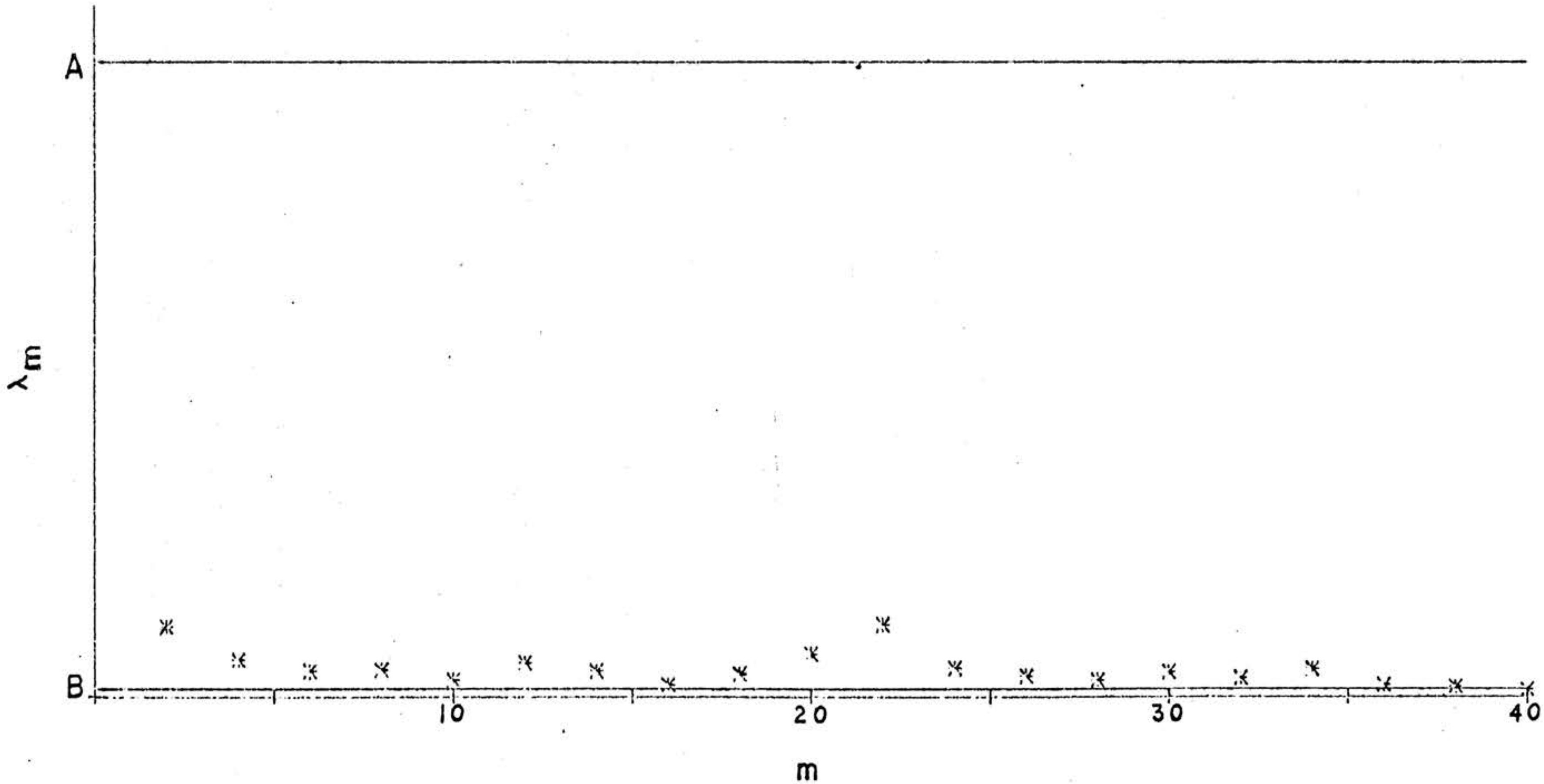

Figure 2.1 Sample Sequential Detection Problem

the parametric detectors in that a complete description of the signal and noise is not necessary. That is, a given nonparametric detector can be used for a large family of input distributions. In the following chapters, such detectors will be discussed and compared to the optimum parametric detectors.

For fixed sample size detectors the goodness criteria used in the comparison of the two detectors will be the ratio of their sample sizes, while the goodness criteria used in the comparison of two sequential detectors will be the ratio of the average number of samples necessary for each detector to operate. Thus in the case of the fixed sample size detector the best detector is the one which, for a specified α, β and signal to noise ratio, requires the smallest number of samples. For the sequential detectors the best detector is the one which, for a set α, β and signal to noise ratio, has the smallest average number of samples.

The signal to noise ratio will be defined as the ratio of the r.m.s. value of the signal to the r.m.s. value of the noise. In the following chapters f(x) is defined as the probability density function on x, i.e., the probability of x falling between x and $x+\Delta x$ is $f(x)\Delta x$, and F(x) is defined as the cumulative distribution or the probability that x takes on a value less than or equal to x, i.e.,

$$F(x) = \int_{-\infty}^{x} f(y)dy.$$

If f(x) has a normal distribution with mean zero and variance one, then it will be abbreviated as $f(x) \sim N(0,1)$. If $f(x) \sim N(0,1)$, then the cumulative of f(x) will be denoted by $\Phi(x)$. $\Phi^{-1}(b)$ will signify the number whose cumulative normal distribution, N(0,1), is b.

CHAPTER III

DESIGN OF THE NONPARAMETRIC SEQUENTIAL TEST

3.1 The Wilcoxon-Mann-Whitney U Test

The Wilcoxon-Mann-Whitney U (W.M.W. U) Test is a fixed sample size nonparametric test of hypothesis. Since this nonparametric test does not assume a form for the input noise, it is necessary for this detector to have a second input to act as a comparison channel. The second or controlled input is a noise alone channel and gives the basis for making a decision about the presence or absence of a signal in the unknown input. Thus the W.M.W. U detector is a two sample detector and the detector must be able to sample from two different input channels.

Wilcoxon proposed an equivalence test (5) for testing if the two inputs of the detector are the same. The Wilcoxon detector takes N samples from the control channel, denoted by $X=(x_1,\ldots,x_N)$. It then takes M samples from the unknown channel, denoted by $Y=(y_1,\ldots,y_M)$. The detector forms a composite sample vector $Z=(z_1,\ldots,z_{N+M})$ composed of the N samples from X and the M samples from Y. The detector now must order these samples so that $Z_1 \leq Z_2 \leq Z_3 \leq \ldots \leq Z_{N+M}$. The Wilcoxon statistic is now defined as follows

$$T = \sum_{i=1}^{N+M} w_i \tag{3.1}$$

where

$$w_i = \begin{cases} i & \text{if } Z_i \text{ is from X} \\ 0 & \text{if } Z_i \text{ is from Y.} \end{cases}$$

The Wilcoxon test statistic is now seen to be the sum of the ranks of the x_i when they are ranked together with the y_i. The null hypothesis (Y is from noise alone) is accepted if T is larger than some critical value T_c and the alternate hypothesis (Y is from signal plus noise) is accepted if $T<T_c$.

About the same time that Wilcoxon proposed his test, Mann and Whitney proposed another form of the same test (6). The Mann-Whitney detector samples from the two inputs and forms the sample vectors $X=(x_1,\ldots, x_N)$ and $Y=(y_1,\ldots,y_M)$. The test statistic is now defined as

$$U = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij} \tag{3.2}$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } x_i > y_j \\ 0 & \text{if } x_i < y_j. \end{cases}$$

The detector accepts the null hypothesis (Y is from noise alone) if U is larger than some threshold U_c and accepts the alternate hypothesis (Y is from signal plus noise) if $U<U_c$. It can be shown (see Appendix A) that

$$U = T - \tfrac{1}{2}N(N+1) \qquad (3.3)$$

Since the two statistics are linearily related, they can be considered the same. The test will be called the Wilcoxon-Mann-Whitney U Test and the statistic will be defined in the same way as Mann and Whitney defined their statistic, i.e., Eq. 3.2.

From the definition of the W.M.W. U test it should be noticed that this test can be used to detect any sort of difference between the two input channels, X and Y, but the test is especially sensitive to differences of the form

$$H_0 : G(x) = F(x)$$
$$H_1 : G(x) = F(x-\theta).$$

That is under the null hypothesis (H_0) the distribution of the Y channel, G(x), is the same as that of the X channel, $F(x)$, while under the alternate hypothesis (H_1) the Y channel is a shift in the mean from the X or control channel. This is exactly the type of alternatives found in the problem of detecting a constant signal in additive noise.

Under the null hypothesis

$$H_0 : G(x) = F(x)$$

the following is true

$$Pr(x_i<y_j) = Pr(x_i>y_j) = \tfrac{1}{2}.$$

Thus the expected value of U is seen to be

$$\begin{aligned} E(U) &= E(\sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij}) \\ &= \sum_i \sum_j E(x_{ij}) \\ &= \tfrac{1}{2}NM. \end{aligned}$$

Similiarily the variance of U under H_0 is given by

$$\begin{aligned} var(U) &= E(\sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij})^2 - \frac{N^2M^2}{4} \\ &= \sum_i \sum_j \sum_k \sum_l E(x_{ij}x_{kl}) - \frac{N^2M^2}{4}. \end{aligned}$$

It is also seen that under H_0

$$\underset{\substack{i \neq k \\ j \neq l}}{E}(x_{ij}x_{kl}) = 1/4$$

$$\underset{j \neq l}{E}(x_{ij}x_{il}) = 1/3$$

$$\underset{i \neq k}{E}(x_{ij}x_{kj}) = 1/3$$

$$E(x_{ij}x_{ij}) = \tfrac{1}{2}.$$

Then

$$\mathrm{var}(U) = NM(N + M + 1)/12 .$$

Under the alternate hypothesis

$$H_1 : G(x) = F(x-\theta)$$

the following can be shown to be true

$$P(X > Y) = \int_{-\infty}^{\infty} f(x)F(x-\theta)dx.$$

Now define

$$\gamma = \int_{-\infty}^{\infty} f(x)F(x-\theta)dx$$

and also define

$$\lambda = \tfrac{1}{2} - \gamma$$
$$\epsilon_1 = 1/3 - \int_{-\infty}^{\infty} F^2(x-\theta)f(x)dx \qquad (3.4)$$
$$\epsilon_2 = 1/3 - \int_{-\infty}^{\infty} (1 - F(x))^2 f(x-\theta)dx.$$

Then it can be seen that under H_1

$$E(U) = \gamma NM.$$

It can also be shown (6) that

$$\text{var}(U) = NM(N+M+1)/12 + NM\Big[-\lambda^2(N+M-1) + (\lambda-\epsilon_1)(M-1) + (\lambda-\epsilon_2)(N-1)\Big].$$

The following notation will now be introduced

$$\mu_0 = E(U) \quad \text{under } H_0$$
$$\sigma_0^2 = \text{var}(U) \quad \text{under } H_0$$
$$\mu_1 = E(U) \quad \text{under } H_1$$
$$\sigma_1^2 = \text{var}(U) \quad \text{under } H_1.$$

If the x_i's and y_j's are independent random variables, U, which is defined by Eq. 3.2, i.e.,

$$U = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij},$$

is the sum of NM random independent variables, since

$P(x_i > y_j)$ is independent of $P(x_k > y_l)$ if $i \neq k$ and $j \neq l$. So the central limit theorem can be applied to this sum of independent variables and the following holds for large N and M

$$\left(\frac{U - \mu_0}{\sigma_0}\right) \simeq N(0,1) \quad \text{under } H_0 \qquad (3.5)$$

and

$$\left(\frac{U - \mu_1}{\sigma_1}\right) \simeq N(0,1) \quad \text{under } H_1. \qquad (3.5)$$

The fixed sample size W.M.W. U test can now be set up in the following manner. Take N samples from X and M samples from Y, and calculate U according to Eq. 3.2. Now form the following statistic

$$W = \left(\frac{U - \mu_0}{\sigma_0}\right).$$

It can be seen that if H_0 is true $W \simeq N(0,1)$ and if H_1 is true $W \simeq N(\mu', \sigma'^2)$ where from Eq. 3.4 it is known that

$$\mu' = \frac{\mu_1 - \mu_0}{\sigma_0} < 0 \qquad (3.6)$$

and

$$\sigma' = \frac{\sigma_1}{\sigma_0} < 1. \quad (3.6)$$

With the aid of Figure 3.1 the values for the threshold (W_c), α, and β can be calculated. From examination of Figure 3.1 it is seen that

$$\alpha = P(W<W_c | H_0)$$

or

$$\alpha = \Phi(W_c). \quad (3.7)$$

The value of β can be seen to be

$$\beta = P(W>W_c | H_1)$$

or

$$\beta = 1 - \Phi\left(\frac{W_c - \mu'}{\sigma'}\right) \quad (3.7)$$

where μ' and σ' are defined by Eq. 3.6. It is now possible, using the above relationships, to design a test so that for a given α and β an N and M can be found that will let the test obtain these error probabilities.

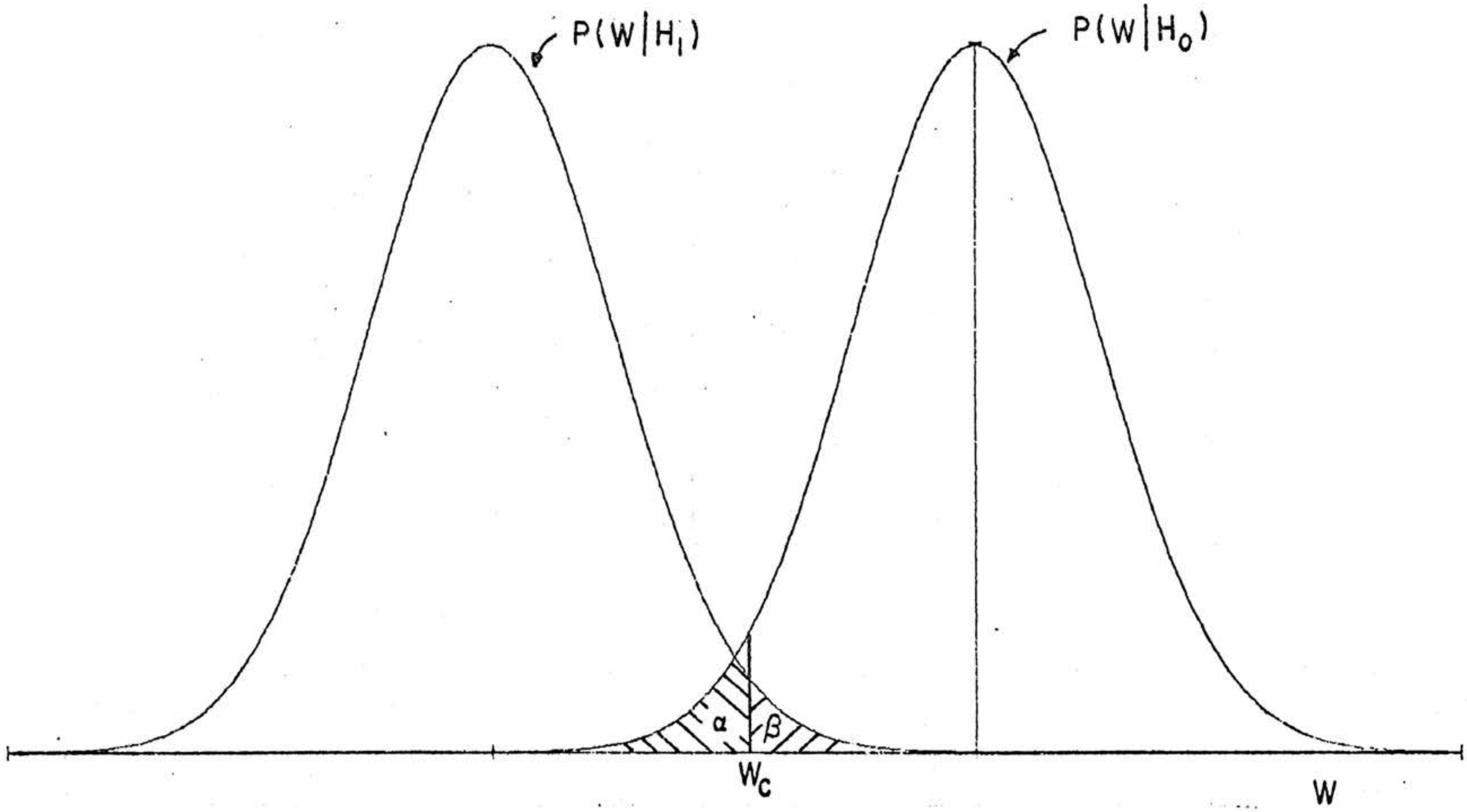

Figure 3.1 W.M.W. U Statistic Under Each Alternative

The test described above is the classical Wilcoxon-Mann-Whitney U test. It has very good power against certain types of alternatives, and can be shown (see Appendix B) to have an Assymptotic Relative Efficiency (A.R.E.) of $3/\pi$ when compared to the t-test for detecting shifts in means for normally distributed noise. In the case of nonnormal noise the W.M.W. U test can have an A.R.E. greater than one.

3.2 Sequential Wilcoxon-Mann-Whitney U Test

The high efficiency of the W.M.W. U test as compared to the t-test for normally distributed noise seems to make the idea of performing it sequentially very attractive. The need now is to find a way to implement the test in a sequential manner. This paper will calculate the T statistic as defined by Wilcoxon because of the relative ease of calculation when compared to that of the equivalent U statistic. To further speed the process, a sequential ranking procedure first devised by Parent (7) will also be used in the calculations. The traditional reranking procedure will be examined first in order to facilitate the introduction of Parent's method.

Consider the observation vector $Z=(z_1, z_2, \ldots z_n)$ where z_i, $i=1,3,5,\ldots$ are from X and z_i, $i=2,4,\ldots$ are from Y. Now in a sequential test the first observation is taken, and the observation vector is ranked.

Then the second observation is taken, and the observation vector must be again ranked. Thus, after each sample is taken, the entire observation vector must be reranked before the statistic can be calculated. This reranking after each sample is extremely time consuming and thus slows the detection process. The sequential ranking procedure provides an efficient and fast way to preserve the information found in the ranks, without reranking after each sample. Let the rank of z_i with respect to the sample $Z=(z_1,\ldots,z_i)$ be denoted by S_i. From this definition it can be seen that $S = 1$, $S = 1$ or 2, $S = 1, 2$, or 3, etc.. If a sample of size N is taken, and as each sample is taken S_i is found, then the vector $S=(S_1,S_2,\ldots,S_N)$ has a one to one correspondence to the ranks of the samples when they are ranked after the N observations (7). The following example will demonstrate this correspondence.

Consider the following observation vector

$$X=(4.0, 2.0, 5.4, 3.2, 4.7).$$

The sequential rank vector for the above observations is given by

$$S=(1, 1, 3, 2, 4).$$

The rank vector for X can now be found in the following way

$$S_2=1 \qquad x_2<x_1$$
$$S_3=3 \qquad x_2<x_1<x_3$$
$$S_4=2 \qquad x_2<x_4<x_1<x_3$$
$$S_5=4 \qquad x_2<x_4<x_1<x_5<x_3 \ .$$

Thus the rank vector is given by

$$X^r = (3,1,5,2,4)$$

where

$$X^r_{(i)} = \text{rank of } x_i \text{ in } X.$$

From the above example it can be seen that the idea of sequential ranking lends itself in a natural way to sequential testing.

The W.M.W. U statistic will now be calculated using the idea of sequential ranking. Define T_n as the value of the Wilcoxon T statistic after n samples. The observation vector $Z=(z_1,\ldots,z_n)$ is again defined as above. That is, if i is even, z_i is from Y; and if i is odd, z_i is from X. Now after n samples, the following holds

$$T_n = T_{n-1} + S_n + S_{xn} \tag{3.8}$$

if z_n was taken from X, and

$$T_n = T_{n-1} + S_{xn} \tag{3.8}$$

if z_n was taken from Y, where S_n and S_{xn} are defined as follows:

$$S_n = \text{sequential rank of } z_n$$

and

$$S_{xn} = \text{number of samples from X which are larger than } z_n.$$

S_{xn} can be found in two ways:

1. The observation z_n is ranked sequentially with respect to the z_i from X. Then this is subtracted from the total number of z_i from X.
2. The observation z_n is inversely ranked sequentially with respect to the z_i from X. That is, the largest value has rank 1, the next value 2, etc..

The second method of the two mentioned above is the one used in the detection described in the following pages.

Eq. 3.3 now becomes

$$U_n = \begin{cases} T_n - n(n+2)/8 & n \text{ even} \\ \\ T_n - (n+1)(n+3)/8 & n \text{ odd.} \end{cases} \quad (3.9)$$

The U_n statistic will now be used in the following manner to implement the sequential test. After U_n is calculated, form the following test ratio

$$\tau_n = \frac{P(U_n|H_1)}{P(U_n|H_0)} \, .$$

This is a logical extension of Wald's Sequential Probability Ratio Test to the nonparametric framework, since by examining its efficiency, it has been shown that U_n contains almost as much information about the shift in mean as does the complete sample. The fact that $P(U_n|H_i)$ is always approximately normal irrespective of the input distribution gives much weight to the use of this type of nonparametric sequential test. The decisions are made after each sample is taken in the following manner. If A and B are two thresholds with B<A then after each sample,

$$\text{if } \tau_n \leq B \quad \text{accept } H_0$$
$$\text{if } \tau_n \geq A \quad \text{accept } H_1$$

or

$$\text{if } B < \tau_n < A \quad \text{take another sample.}$$

The operation of the detector has been fully defined; it is now necessary to develop the necessary relationships to evaluate the performance of the detection scheme. The first step is to find a way to express A and B in terms of known constants.

If the observable data are continuous functions of time and if continuous sampling is used, then the statistic U could be thought of as a continuous function of time. Thus it would be possible to think of τ as a continuous function of time. Therefore, the sequential nonparametric test would terminate when $\tau(t) = A$ or $\tau(t) = B$. It is only in the case of discrete sampling that the inequalities $\tau_n > A$ or $\tau_n < B$ are possibly obtained. That is, after n-1 samples it is possible for $B < \tau_{n-1} < A$ and that the inclusion of one more sample produces either $\tau_n < B$ or $\tau_n > A$. For n large it is reasonable to assume that this excess over the boundary will be small, and for the case of the sequential W.M.W. U test it will be assumed that the test terminates with $\tau_n = A$ or $\tau_n = B$. Now suppose that the sequential test is carried out, and that after n samples $\tau_n = A$. This leads to the acceptance of H_1. Thus

$$\tau_n = \frac{P(U_n|H_1)}{P(U_n|H_0)} = A$$

implies the acceptance of H_1, or putting it another way, the acceptance of H_1 implies that

$$P(U_n|H_1) = A\ P(U_n|H_0). \tag{3.10}$$

If the set of all values of U_n which lead to the acceptance of H_1 is denoted by Γ_1, then Eq. 3.10 is equivalent to

$$\int_{\Gamma_1} P(U_n|H_1)dU_n = A \int_{\Gamma_1} P(U_n|H_0)dU_n \tag{3.11}$$

or

$$1 - \beta = A\alpha \tag{3.12}$$

Proceeding in an analogous manner it is found that the acceptance of the null hypothesis, H_0, implies

$$\int_{\Gamma_0} P(U_n|H_1)dU_n = B \int_{\Gamma_0} P(U_n|H_0)dU_n \tag{3.13}$$

where Γ_0 is the region of acceptance of H_0. Eq. 3.12 now reduces to

$$\beta = B(1 - \alpha). \qquad (3.14)$$

Now looking at Eqs. 3.12 and 3.14, it is possible to find A and B as functions of the probabilities of error, i.e.,

$$A = \frac{1-\beta}{\alpha}$$
$$B = \frac{\beta}{1-\alpha}. \qquad (3.15)$$

The next problem of interest is the actual calculation of τ_n. This involves the calculation of $P(U_n|H_0)$ and $P(U_n|H_1)$. In the preceding section of the chapter, it was found that for large n $(U_n - \mu)/\sigma$ is approximately normal. The actual density for U_n is difficult to calculate under either the null hypothesis or the alternate hypothesis. However, for large n it approaches a normal distribution, therefore, for convenience, it will be assumed that $(U_n - \mu)/\sigma$ is normal for all n. Using the above argument and Eq. 3.4 it can be seen that

$$P(U_n|H_0) = \frac{1}{\sqrt{2\pi}\,\sigma_0} e^{-\frac{1}{2}\left(\frac{U_n - \mu_0}{\sigma_0}\right)^2}$$

and

$$P(U_n|H_1) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{U_n - \mu_1}{\sigma_1}\right)^2}$$

where μ_0, μ_1, σ_0^2, and σ_1^2 are functions of n, i.e.,

$$\mu_0 = \tfrac{1}{2}NM$$

$$\mu_1 = \gamma NM$$

$$\sigma_0^2 = (NM(N+M+1))/12$$

and

$$\sigma_1^2 = NM(N+M+1)/12 + NM\left[-\lambda^2(N+M-1) + (\lambda - \epsilon_1)(M-1) + (\lambda - \epsilon_2)(N-1)\right]$$

where

N = number of samples from X

M = number of samples from Y

and λ, γ, ϵ_1, and ϵ_2 are defined by Eq. 3.4. Since $\sigma_1^2 < \sigma_0^2$ it can be assumed that $(U_n - \mu_1)/\sigma_0$ is approximately $N(0, 1)$ under the alternate hypothesis. This is not quite true, but it will produce a conservative test, since the variance of $(U_n - \mu_1)/\sigma_0$ will actually be less that one. With this assumption and upon noting that

$$N = \begin{cases} n/2 & n \text{ even} \\ (n+1)/2 & n \text{ odd} \end{cases}$$

and

$$M = \begin{cases} n/2 & n \text{ even} \\ (n-1)/2 & n \text{ odd.} \end{cases}$$

It is possible to express μ_0, μ_1, and σ_0^2 as functions of n, i.e.,

$$\mu_0 = \begin{cases} n^2/8 & n \text{ even} \\ (n^2-1)/8 & n \text{ odd} \end{cases}$$

$$\mu_1 = \begin{cases} \gamma n^2/4 & n \text{ even} \\ \gamma(n^2-1)/4 & n \text{ odd} \end{cases} \qquad (3.16)$$

$$\sigma_0^2 = \begin{cases} (n^3+n^2)/48 & n \text{ even} \\ (n^3+n^2-n-1)/48 & n \text{ odd.} \end{cases}$$

The test statistic can now be written as

$$\tau_n = e^{-\frac{1}{2}\left(\frac{U_n-\mu_1}{\sigma_0}\right)^2 + \frac{1}{2}\left(\frac{U_n-\mu_0}{\sigma_0}\right)^2}$$

or

$$\tau_n = e^{\left(\frac{\mu_1 - \mu_0}{2}\right)\left(U_n - \frac{\mu_1 + \mu_0}{2}\right)} \tag{3.17}$$

where μ_0, μ_1, and σ_0^2 are defined as above.

The next step in the analysis of the sequential W.M.W. U detector is to obtain an expression for the average number of samples necessary for detection. In order to facilitate this calculation, the following definition is made

$$\Omega_n = \ln \tau_n.$$

Now the test can be restated as

accept H_0 when $\Omega_n \leq \ln B$
accept H_1 when $\Omega_n \geq \ln A$
and when $\ln B < \Omega_n < \ln A$ take another sample.

From Eq. 3.17 it can be seen that

$$\Omega_n = \left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(U_n - \frac{\mu_1 + \mu_0}{2}\right).$$

Assume now that the test is conducted as has been described previously and that H_0 is true. Then at the time a decision is made Ω_n can either be equal to $\ln A$ or $\ln B$. If H_0 is true then Ω_n will be equal to $\ln B$

$(1-\alpha)\%$ of the time and equal to $\ln A$ $\alpha\%$ of the time. Thus the expected value of Ω_n (denoted by $\overline{\Omega}_n$) when H_0 is true is given by

$$\overline{\Omega}_n = \alpha \ln A + (1-\alpha)\ln B \, . \tag{3.18}$$

Similarly, when H_1 is true it can be argued that

$$\overline{\Omega}_n = (1-\beta)\ln A + \beta \ln B \, . \tag{3.19}$$

At the same time the expected value for U_n at the time of decision under H_0 is μ_0 and the expected value for U_n at the time of decision under H_1 is μ_1. Combining the above and Eqs. 3.18 and 3.19 yields

$$\left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(\frac{\mu_0 - \mu_1}{2}\right) = \alpha \ln A + (1-\alpha)\ln B$$

and (3.20)

$$\left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(\frac{\mu_1 - \mu_0}{2}\right) = (1-\beta)\ln A + \beta \ln B \, .$$

Substituting the values from Eq. 3.16 into Eq. 3.20 yields under H_0

$$\alpha \ln A + (1-\alpha)\ln B = \begin{cases} \dfrac{-3n^4}{2(n^3+n^2)}(\gamma-\tfrac{1}{2})^2 & n \text{ even} \\ \\ \dfrac{-3(n^2-1)^2}{2(n^3+n^2-n-1)}(\gamma-\tfrac{1}{2})^2 & n \text{ odd} \end{cases}$$

and under H_1

$$(1-\beta)\ln A + \beta \ln B = \begin{cases} \dfrac{3n^4}{2(n^3+n^2)}(\gamma-\tfrac{1}{2})^2 & n \text{ even} \\ \\ \dfrac{3(n^2-1)^2}{2(n^3+n^2-n-1)}(\gamma-\tfrac{1}{2})^2 & n \text{ odd.} \end{cases}$$

For n large $n^2-1 \approx n^2$ and $n^4/(n^3+n^2) \approx n$, solving for n in the above equations gives an approximation for $\overline{n}$, the average number of samples required for detection. The solution is

$$\overline{n} \approx \begin{cases} \dfrac{-2\ \alpha \ln A + (1-\alpha)\ln B}{3(\gamma-\frac{1}{2})^2} & \text{under } H_0 \\ \\ \dfrac{2\ (1-\beta)\ln A + \beta \ln B}{3(\gamma-\frac{1}{2})^2} & \text{under } H_1. \end{cases} \qquad (3.21)$$

Eq. 3.21 gives an approximation for the average number of samples for detection as a function of α, β, and γ. The approximations made in the calculations will become better as n gets large. Thus for small α, β, and large γ the above approximations should be very close.

Now that the test has been fully designed it is known that if H_0 is true the test will terminate with decision "H_0 true" with probability $1-\alpha$ and with decision "H_1 true" with probability α. If H_1 is true a similar knowledge is available about the acceptance of H_0 and the acceptance of H_1. The hypotheses are of the form

$$H_0 : G(x) = F(x)$$
$$H_1 : G(x) = F(x-\theta_1).$$

Therefore, the test can be thought of as testing

$$H_0 : \theta = \theta_0 = 0$$
$$H_1 : \theta = \theta_1 .$$

With this thought in mind, it is now appropriate to ask the following questions: 1) What happens if the true value of θ is not θ_0 or θ_1? That is, what happens if a mistake was made in setting up the alternate hypothesis? 2) How does the test designed to test $\theta = 0$

against $\theta = \theta_1$, perform when θ takes on values other than the designed values? The operating-characteristic function gives a measure of how well the test performs under these conditions.

The operating-characteristic function $L(\theta)$ is defined as the probability that the test terminates with decision H_0 when the true value of the unknown parameter is θ. Therefore, in the case of the sequential W.M.W. U detector it is obvious that L(0) is equal to $1-\alpha$ and that $L(\theta_1)$ is equal to β. To derive the operating-characteristic function for the sequential W.M.W. U detector, it is again necessary to assume that the excess over the boundary at the time of decision is negligible. Now consider the following

$$(\tau_m)^h = \left(\frac{P(U_m | \theta_1)}{P(U_m | \theta_0)} \right)^h$$

where $h = h(\theta)$ is some function of θ with the following restrictions (4)

$$h(\theta) \neq 0$$

and

$$\int_{\Gamma(U_m)} \left(\frac{P(U_m|\theta_1)}{P(U_m|\theta_0)} \right)^h P(U_m|\theta)dU_m = 1. \qquad (3.22)$$

The existence of a unique solution for $h(\theta)$ has been proven by Wald (4). Define

$$g(U_m|\theta) = \left(\frac{P(U_m|\theta_1)}{P(U_m|\theta_0)} \right)^{h(\theta)} P(U_m|\theta).$$

Note that $g(U_m|\theta)$ meets all of the requirements for a probability density function, i.e.,

$$g(U_m|\theta) \geq 0$$

$$\int_{\Gamma(U_m)} g(U_m|\theta)dU_m = 1.$$

Therefore, H_g can be the hypothesis that the true distribution of U_m is $g(U_m|\theta)$, and H_p can be the hypothesis that the true distribution of U_m is $p(U_m|\theta)$. Thus it is now possible to formulate a sequential test of hypothesis H_p against H_g, where the upper threshold is A^h and the lower threshold is B^h. After the m^{th} sample the test is

$$\text{if} \quad \frac{g(U_m\ \theta)}{p(U_m\ \theta)} \leq B^h \qquad \text{accept } H_p$$

$$\text{if} \quad \frac{g(U_m\ \theta)}{p(U_m\ \theta)} \geq A^h \qquad \text{accept } H_g$$

$$\text{and if } B^h < \frac{g(U_m\ \theta)}{p(U_m\ \theta)} < A^h \quad \text{take another sample.}$$

Following a procedure similar to that used in the development of Eq. 3.15, it can be shown that

$$B^h = \frac{\beta}{1-\alpha}$$

and

$$A^h = \frac{1-\beta}{\alpha}.$$

Solving the above for $1-\alpha$ yields

$$1 - \alpha = \frac{A^h - 1}{A^h - B^h}.$$

The above is by definition $L(\theta)$, thus

$$L(\theta) = \frac{A^h - 1}{A^h - B^h}. \qquad (3.23)$$

Now it becomes necessary to solve for $h(\theta)$ in order to get an expression for $L(\theta)$.

The exact solution of Eq. 3.22 for h is very difficult since the distribution for U_n is not known exactly. However, an approximation can be obtained if the normal assumption is used. That is

$$U_n \sim N(\mu, \sigma^2).$$

Instead of carrying along the variable θ, in this case it will be equivalent to carry the variable μ since μ is directly related to γ and thus also to θ. Now Eq. 3.22 becomes

$$\int_{-\infty}^{\infty} \left[\frac{\frac{1}{\sqrt{2\pi}\,\sigma_0} e^{-\frac{1}{2}\left(\frac{U-\mu_1}{\sigma_0}\right)^2}}{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{U-\mu_0}{\sigma_0}\right)^2}} \right]^h \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{U-\mu}{\sigma_0}\right)^2} dU = 1.$$

Reducing the above yields

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_0} e^{\frac{-1}{2\sigma_0^2}\left[U^2 - 2(\mu + (\mu_1 - \mu_0)h) + \mu^2 + (\mu_1^2 - \mu_0^2)h\right]} dU = 1.$$

It is known that the above integral is equal to 1 if the exponent can be written as a perfect square. This implies that

$$\left[\mu + (\mu_1 - \mu_0)h\right]^2 = \mu^2 + (\mu_1^2 - \mu_0^2)h$$

or that if $h \neq 0$

$$h = \frac{\mu_1 + \mu_0 - 2\mu}{\mu_1 - \mu_0}.$$

If γ_1, γ_0 and γ are defined by

$$\mu_1 = \gamma_1 n^2/4$$
$$\mu_0 = \gamma_0 n^2/4.$$

Then $\gamma_0 = \frac{1}{2}$ and

$$\mu = \gamma n^2/4.$$

From the above it can be seen that

$$h = \frac{\gamma_1 + \frac{1}{2} - 2\gamma}{(\gamma - \frac{1}{2})} \qquad (3.24)$$

for $\gamma \neq \dfrac{\gamma_1 + \frac{1}{2}}{2}$.

It is now necessary to find $L(\gamma)$ for the case when $\gamma = \dfrac{\gamma_1 + \frac{1}{2}}{2}$. By examining Eq. 3.23 it can be seen that $\lim_{h \to 0} L(\gamma)$ is indeterminate. However, the application of L'Hospital's rule yields

$$\lim_{h \to 0} L(\gamma) = \lim_{h \to 0} \frac{A^h \ln A}{A^h \ln A - B^h \ln B}$$

or

$$\lim_{h \to 0} L(\gamma) = \frac{\ln A}{\ln(A/B)} \quad . \qquad (3.25)$$

Now Eq. 3.25, 3.24, and 3.23 are combined to obtain the operating-characteristic function for the sequential W.M.W. U detector

$$L(\gamma) \approx \begin{cases} \dfrac{\left[\frac{1-\beta}{\alpha}\right]^{\frac{\gamma_1+\frac{1}{2}-2\gamma}{\gamma_1-\frac{1}{2}}} - 1}{\left[\frac{1-\beta}{\alpha}\right]^{\frac{\gamma_1+\frac{1}{2}-2}{\gamma_1-\frac{1}{2}}} - \left[\frac{\beta}{1-\alpha}\right]^{\frac{\gamma_1+\frac{1}{2}-2}{\gamma_1-\frac{1}{2}}}} & \gamma \neq \dfrac{\gamma_1+\frac{1}{2}}{2} \\ \\ \dfrac{\ln\,(1-\beta)/\alpha}{\ln\,(1-\alpha)(1-\beta)/\alpha\beta} \quad . & \gamma = \dfrac{\gamma_1+\frac{1}{2}}{2} \end{cases}$$

A typical plot of $L(\gamma)$ is shown in Fig. 3.2.

Knowledge of the operating-characteristic function not only gives an important index of performance for the sequential detector, but also allows the calculation of a more general form for the determination of the average number of necessary samples. If $\overline{n(\theta)}$ is defined

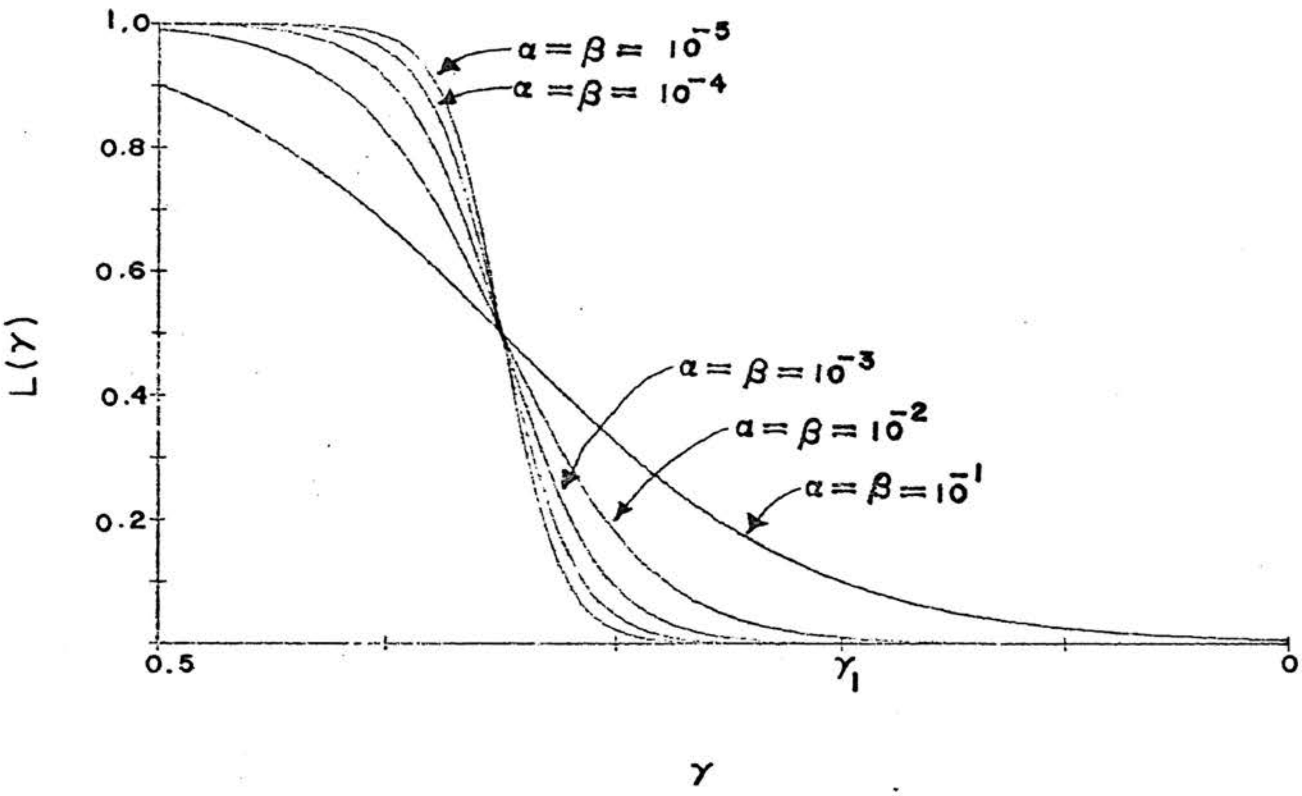

Figure 3.2 Operating-Characteristic Function for the Sequential W.M.W. U Detector

as the average number of samples necessary for termination of the test when the true value of $\theta = \theta$, then Eq. 3.21 gives an expression for $\overline{n(\theta_0)}$ and $\overline{n(\theta_1)}$. It is now possible with the use of the operating-characteristic function to calculate $\overline{n(\theta)}$ for any θ.

Again consider

$$\overline{\Omega}_n = \left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(U_n - \frac{\mu_1 + \mu_0}{2}\right).$$

The probability that the test terminates with acceptance of H_0 is now given by $L(\theta)$ and the probability of acceptance of H_1 by $1 - L(\theta)$. Thus using the argument which precipitated Eq. 3.18 it can be seen that $\overline{\Omega}_n$ can now be expressed as

$$\overline{\Omega}_n = \left[1 - L(\theta)\right] \ln A + L(\theta) \ln B \qquad (3.25)$$

At the time of decision the expected value for U_n can be expressed as

$$E(U_n) = \mu = \gamma n^2/4 .$$

This leads to the relationship

$$\left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(\mu - \frac{\mu_1 + \mu_0}{2}\right) = \left[1 - L(\theta)\right] \ln A + L(\theta) \ln B \qquad (3.26)$$

Again using the definitions of the μ_i's in Eq. 3.26 and making the same assumptions that were made in the derivation of Eq. 3.21, the value of $\overline{n(\theta)}$ becomes

$$\overline{n(\theta)} = \frac{2\left[1-L(\theta)\right]\ln A + L(\theta)\ln B}{3(\gamma_1 - \frac{1}{2})\left[2\gamma-(\gamma_1 + \frac{1}{2})\right]} \qquad \gamma \neq \frac{\gamma_1+\frac{1}{2}}{2}$$

Since γ and θ are directly related, the above will be written as

$$\overline{n(\gamma)} = \frac{2\left[1-L(\gamma)\right]\ln A + L(\gamma)\ln B}{3(\gamma_1 - \frac{1}{2})\left[2\gamma-(\gamma_1 + \frac{1}{2})\right]} \qquad (3.27)$$

for

$$\gamma \neq \frac{\gamma_1+\frac{1}{2}}{2}.$$

As in the case of the expression of $L(\gamma)$, there is an indeterminate form of the above when $\gamma = \frac{\gamma_1 + \frac{1}{2}}{2}$. Through a very lengthy development, Wald (4) was able to show that the value for $n(\gamma)$ when $\gamma = \frac{\gamma_1 + \frac{1}{2}}{2}$ can be approximated in the following manner

$$E\left\{\left[\left(\frac{\mu_1 - \mu_0}{\sigma_0^2}\right)\left(U_n - \frac{\mu_1 + \mu_0}{2}\right)\right]^2\right\} = -\ln A \ln B$$

where A and B are defined in Eq. 3.15, and where the expectation is conditional on γ. The above now becomes

$$E\left\{\left[\left(\frac{\mu_1-\mu_0}{\sigma_0^2}\right)\left(U_n-\frac{\mu_1+\mu_0}{2}\right)\right]^2\right\}=\left(\frac{\mu_1-\mu_0}{\sigma_0}\right)^2$$

or

$$\frac{(\mu_1-\mu_0)^2}{\sigma_0^2} = -\ln A \ln B .$$

Substituting in the above for the value of μ_1 and μ_0, and making the customary assumptions, yields

$$\overline{n(\gamma)} = \frac{-\ln A \ln B}{3(\gamma_1-\frac{1}{2})^2} . \qquad \gamma = \frac{\gamma_1+\frac{1}{2}}{2}$$

The final expression for $\overline{n(\gamma)}$ can now be written as

$$\overline{n(\gamma)} \approx \begin{cases} \dfrac{2\left\{\left[1-L(\gamma)\right]\ln\left(\frac{1-\beta}{\alpha}\right)+L(\gamma)\ln\left(\frac{\beta}{1-\alpha}\right)\right\}}{3(\gamma_1-\frac{1}{2})\left[2\gamma-(\gamma_1+\frac{1}{2})\right]} & \gamma\neq\dfrac{\gamma_1+\frac{1}{2}}{2} \\[2ex] \dfrac{-\ln\left(\frac{1-\beta}{\alpha}\right)\ln\left(\frac{\beta}{1-\alpha}\right)}{3(\gamma_1-\frac{1}{2})^2} . & \gamma=\dfrac{\gamma_1+\frac{1}{2}}{2} \end{cases}$$

Typical values for $\overline{n(\gamma)}$ appear in Fig. 3.3.

The sequential W.M.W. U detector has been fairly well analyzed now, since the average number of samples function ($\overline{n(\gamma)}$) gives the average number of samples to make a decision under all possible conditions.

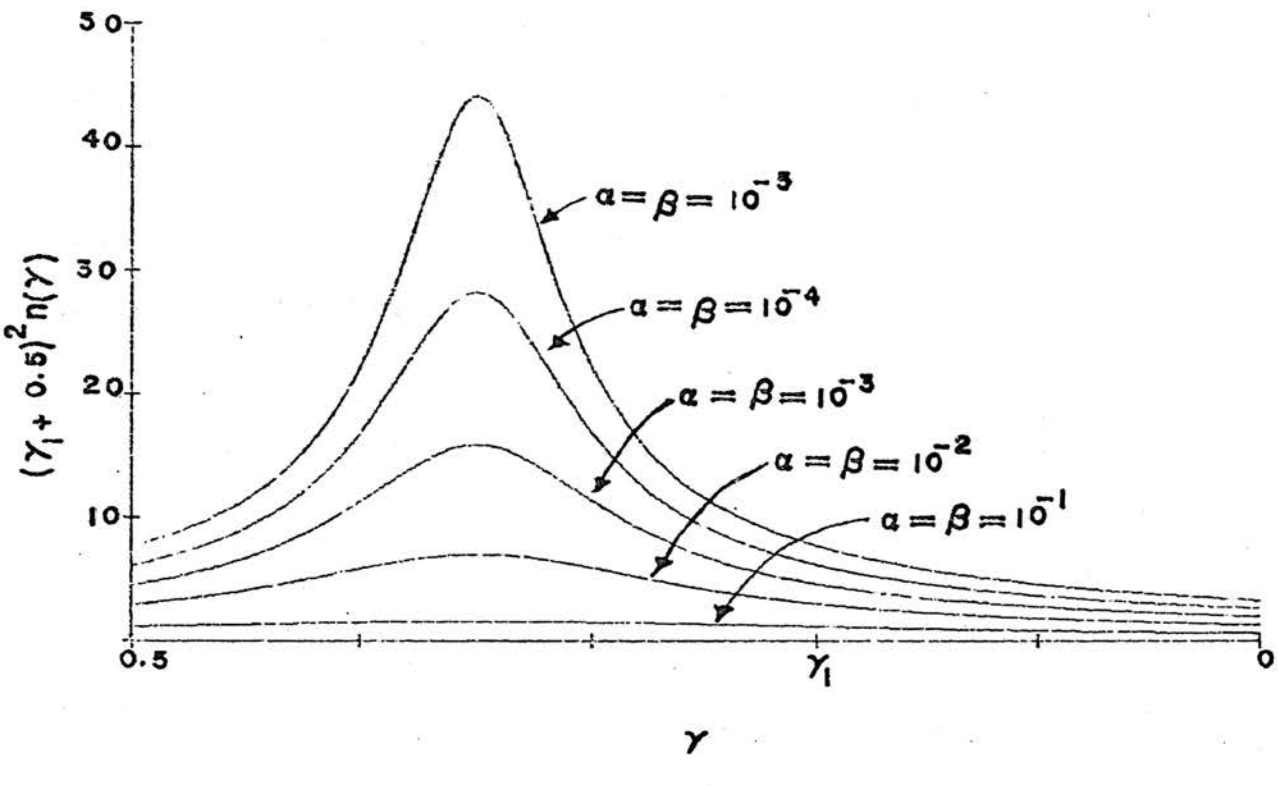

Figure 3.3 Average Number of Samples Function for the W.M.W. U Detector

Furthermore, the operating-characteristic function describes, under all possible conditions, the average number of decisions in favor of the null hypothesis and also the alternate hypothesis.

3.3 Efficiency Of The Sequential W.M.W. U Test

In the preceding section a thorough development of the test procedure along with the development of several operating characteristics for the sequential W.M.W. U test was presented. In this section the operating characteristics will be compared to those of the fixed sample size W.M.W. U detector in hopes of getting an idea of how the sequential detector compares to this other detector.

The comparison to be discussed here is one of comparing the expected number of samples for the sequential detector to the number of samples necessary for the fixed sample size W.M.W. U detector, each detecting the same error probabilities. If the sequential detector is to be considered good, it must have a smaller expected number of samples than is needed by the fixed sample size detector. The figure of merit used in the comparison is called the percentage of savings of the sequential detector. The percentage of savings is defined by

$$PS_i = 100\left[1 - \frac{\overline{n(\theta_i)}}{n_s}\right] \% \qquad (3.29)$$

where PS_i is the percentage of savings given H_i is true, $\overline{n(\theta_i)}$ is the average number of samples needed by the sequential detector given H_i, and n_s is the number of samples necessary for the fixed sample size detector to operate with the perscribed error rate.

Taking n_s to be the number of samples necessary for the fixed sample size W.M.W. U detector, and $\overline{n(\theta_i)}$ as the average number of samples needed by the sequential W.M.W. U detector, then PS_i will give an estimate of how much savings in time is incurred by changing to a sequential test. From section 1 of this chapter, it was found that

$$\alpha = \Phi(W_c)$$

and

$$\beta = 1 - \Phi\left(\frac{W_c - \mu'}{\sigma'}\right)$$

(see Eq. 3.7). Using the inverse function defined in Chapter II, the above becomes

$$W_c = \Phi^{-1}(\alpha)$$
$$\frac{W_c - \mu'}{\sigma'} = \Phi^{-1}(1-\beta). \qquad (3.30)$$

Using the approximation that $\sigma_0 \approx \sigma_1$, and using the fact that $\mu' = \frac{\mu_1 - \mu_0}{\sigma_0}$, Eq. 3.30 can be rewritten as

$$W_c = \Phi^{-1}(\alpha)$$
$$W_c + \frac{\mu_0 - \mu_1}{\sigma_0} = \Phi^{-1}(1-\beta).$$

Simplifying the above yields

$$\frac{\mu_0 - \mu_1}{\sigma_0} = \Phi^{-1}(1-\beta) - \Phi^{-1}(\alpha).$$

Using the values of μ_0, σ_0, and μ_1 given in Eq. 3.16, the above becomes:

$$\frac{\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(\alpha)\right]^2}{3(\frac{1}{2}-\gamma)^2} = \begin{cases} \dfrac{n^4}{n^3 + n^2} & n \text{ even} \\ \\ \dfrac{(n^2-1)^2}{n^3 + n^2 - n - 1} & n \text{ odd.} \end{cases}$$

Making the approximations that $n^2 - 1 \approx n^2$ and that $n^3 + n^2 \approx n^3$ the expression for n becomes

$$n_s = \frac{\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(\alpha)\right]^2}{3(\frac{1}{2}-\gamma)^2} . \tag{3.31}$$

Now using Eq. 3.31 and Eq. 3.21 it is possible to calculate PS_0 and PS_1. Performing the indicated calculations yields

$$PS_0 = 100\left[1 + \frac{2\left[\alpha \ln A + (1-\beta)\ln B\right]}{\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(\alpha)\right]^2}\right]\%$$

and (3.32)

$$PS_1 = 100\left[1 + \frac{2\left[(1-\beta)\ln A + \alpha \ln B\right]}{\left[\Phi^{-1}(1-\beta) - \Phi^{-1}(\alpha)\right]^2}\right]\% .$$

Fig. 3.4 and Fig. 3.5 give plots of PS_0 and PS_1 as a function of α and β. It is interesting to note that PS_i (as given in Eq. 3.32) is not a function of the signal to noise ratio. This is easy to understand, in that both detectors operate by using the same statistic. Thus the signal to noise ratio or equivalently γ should affect both in the same manner, indicating that the difference in sample sizes is due to the different decision rules.

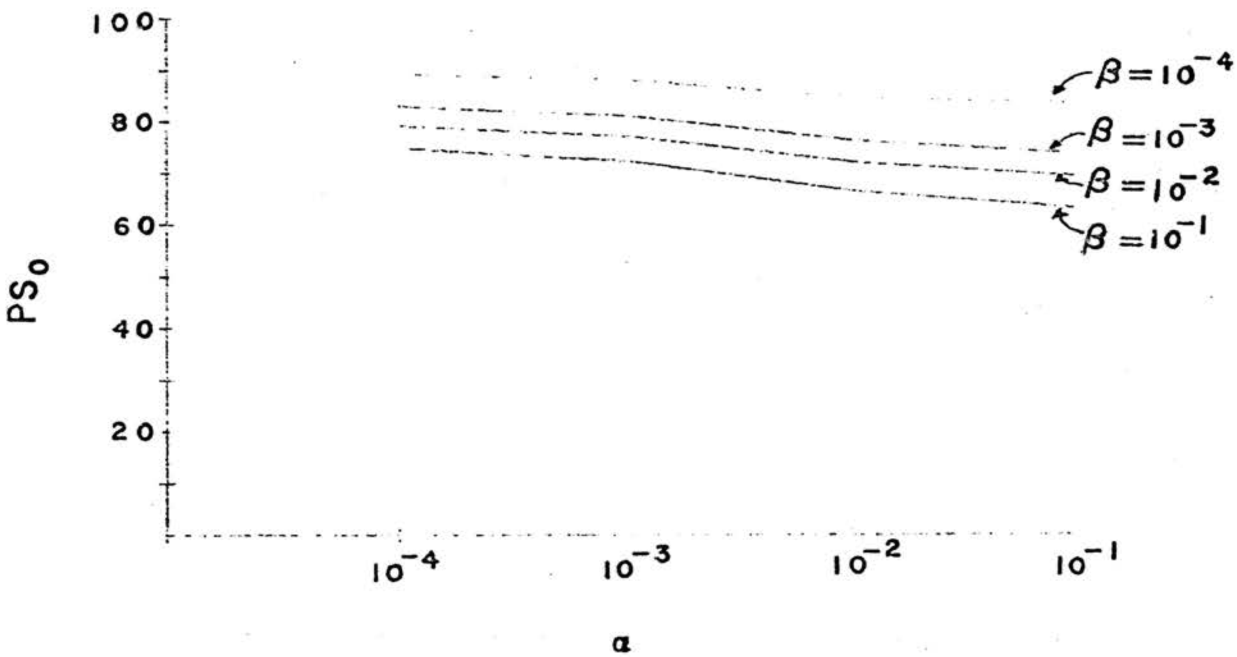

Figure 3.4 Percent Savings for the Sequential W.M.W. U Detector Under the Null Hypothesis

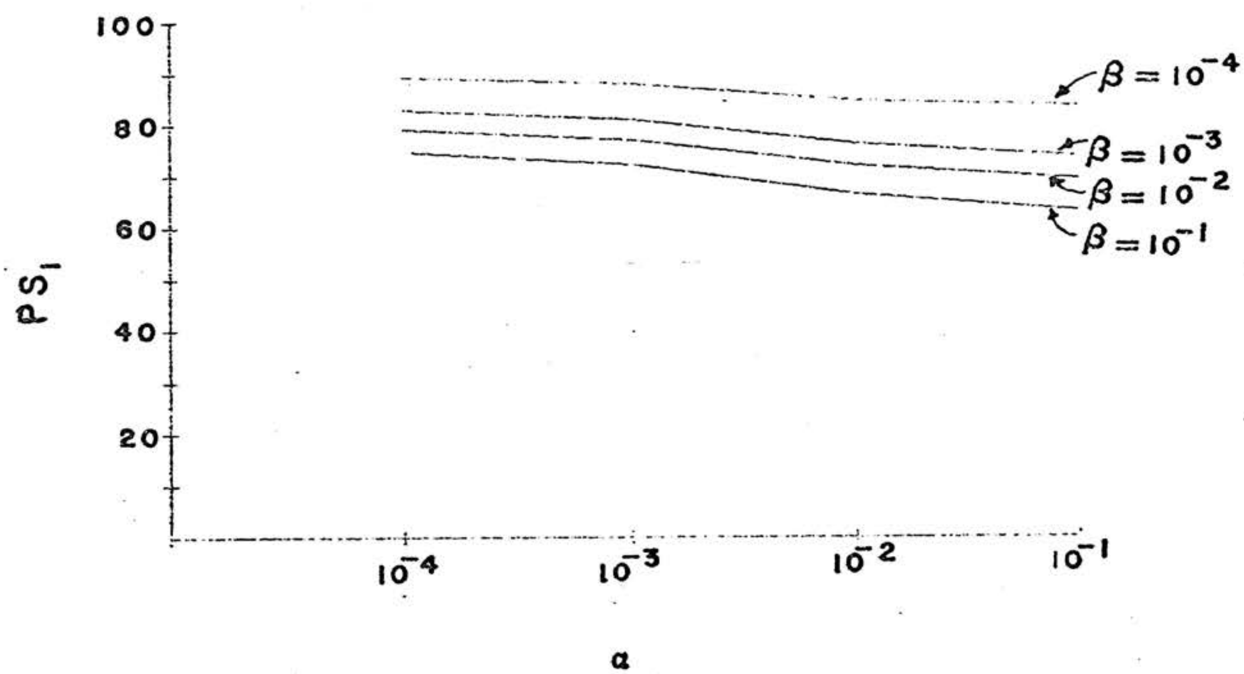

Figure 3.5 Percent Savings for the Sequential W.M.W. U Detector Under the Alternate Hypothesis

CHAPTER IV

SEQUENTIAL DETECTION OF CONSTANT SIGNALS

4.1 Detection Of A Constant Signal In Normal Noise

The optimum detectors discussed in Chapter II are usually designed for detection in normally distributed noise. So in order to be able to compare the sequential nonparametric detector to an optimum detector it is necessary to analyze the performance of the sequential detector in normally distributed noise. The alternatives are of the form

$$H_0 : \quad G(x) = \Phi(x)$$
$$H_1 : \quad G(x) = \Phi(x-\theta)$$

where $\Phi(x)$ is the cumulative normal distribution. For this case the Wald sequential probability ratio statistic is given by

$$\lambda_m = \overline{y} - \overline{x}$$

where $\overline{y}$ and $\overline{x}$ are the means of the samples drawn from the Y population and the X population respectively. If $x \sim N(0, \sigma^2)$, then the decision rule can be shown to be (8)

$$\text{if} \quad \lambda_n \geq \frac{4\sigma^2}{\theta} \ln A + \frac{n}{2}\theta \quad \text{accept } H_1$$

$$\text{if} \quad \lambda_n \leq \frac{4\sigma^2}{\theta} \ln B + \frac{n}{2}\theta \quad \text{accept } H_0$$

$$\text{and if} \quad \frac{4\sigma^2}{\theta} \ln B + \frac{n}{2}\theta < \lambda_n < \frac{4\sigma^2}{\theta} \ln A + \frac{n}{2}\theta \quad \text{take another sample.}$$

Using the above rule, it is possible to show (8) that the average number of samples necessary for detection is given by

$$n_W = \begin{cases} \frac{-8\sigma^2}{\theta^2}\left[\alpha \ln\left(\frac{1-\beta}{\alpha}\right) + (1-\alpha)\ln\left(\frac{\beta}{1-\alpha}\right)\right] & H_0 \text{ true} \\ \frac{8\sigma^2}{\theta^2}\left[(1-\beta)\ln\left(\frac{1-\beta}{\alpha}\right) + \beta \ln\left(\frac{\beta}{1-\alpha}\right)\right] & H_1 \text{ true.} \end{cases} \qquad (4.1)$$

Eq. 4.1 now gives an expression for the number of samples necessary for the Wald detector. If a sililar expression can be found for the sequential W.M.W. U detector, the two detectors can be compared in efficiency.

Eq. 3.21 gives the necessary form for n for the sequential W.M.W. U detector, but the parameter γ must be found as a function of θ before the actual comparison can take place, it is also needed before the detector can be used to detect the signal described above.

The definition for γ is given by Eq. 3.4, i.e.,

$$\gamma = P(x>y) = \int_{-\infty}^{\infty} f(x)F(x-\theta)dx.$$

Rewriting the above into a more basic form yields

$$\gamma(\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{x-\theta} f(t)f(x)dtdx.$$

Taking the derivative with respect to θ, the above becomes

$$\gamma'(\theta) = -\int_{-\infty}^{\infty} f(x-\theta)f(x)dx. \qquad (4.2)$$

For this case

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Using the above in Eq. 4.2 produces

$$\gamma'(\theta) = \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} e^{-\frac{1}{2}(x-\theta)^2} dx$$

$$= \frac{-1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(x - \frac{\theta}{2}\right)^2} e^{-\frac{\theta^2}{4}} dx$$

$$= \frac{-1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{4}}.$$

Thus

$$\gamma(\theta) = 1 - \Phi\left(\frac{\theta}{\sqrt{2}}\right) \tag{4.3}$$

where $\Phi(x)$ is the cumulative normal. Fig. 4.1 is a plot of $\gamma(\theta)$ as a function of θ.

Using the above information, it is now possible to compare the average number of samples for the sequential W.M.W. U detector. Define the percentage loss as

$$PL = 100\left[1 - \frac{\overline{n}_W}{\overline{n}}\right]\% . \tag{4.4}$$

That is, the percentage loss is a measure of the loss due to the use of the nonparametric sequential detector instead of the optimum parametric detector. Using Eq. 4.1, 4.3, 3.21, and 4.4 it is possible to evaluate PL, i.e.,

$$PL = 100\left[1 - \frac{12\left[\frac{1}{2} - \Phi\left(\frac{\theta}{\sqrt{2}}\right)\right]^2}{\theta^2}\right] \%. \tag{4.5}$$

The above equation can now be evaluated as a function of θ (See Fig. 4.2). Here it is interesting to note that PL is not a function of α, β, or the true hypoth-

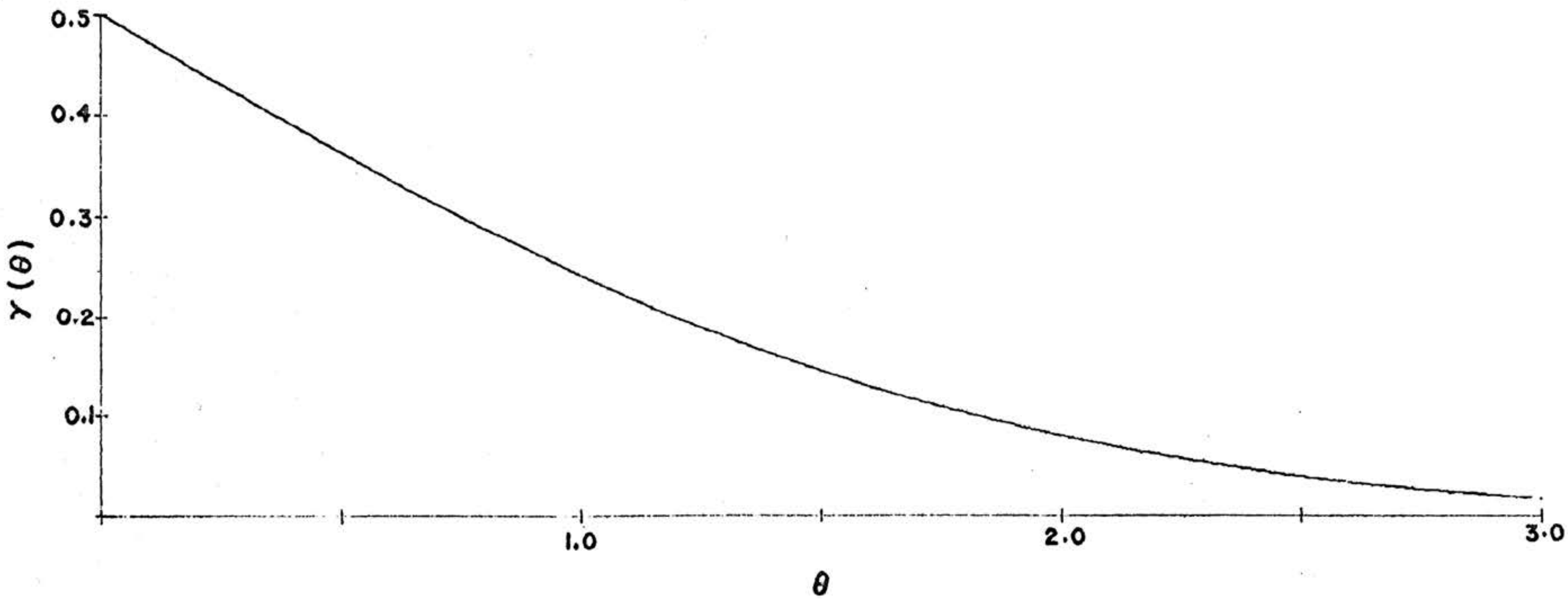

Figure 4.1 $\gamma(\theta)$ vs. θ for Normal Noise

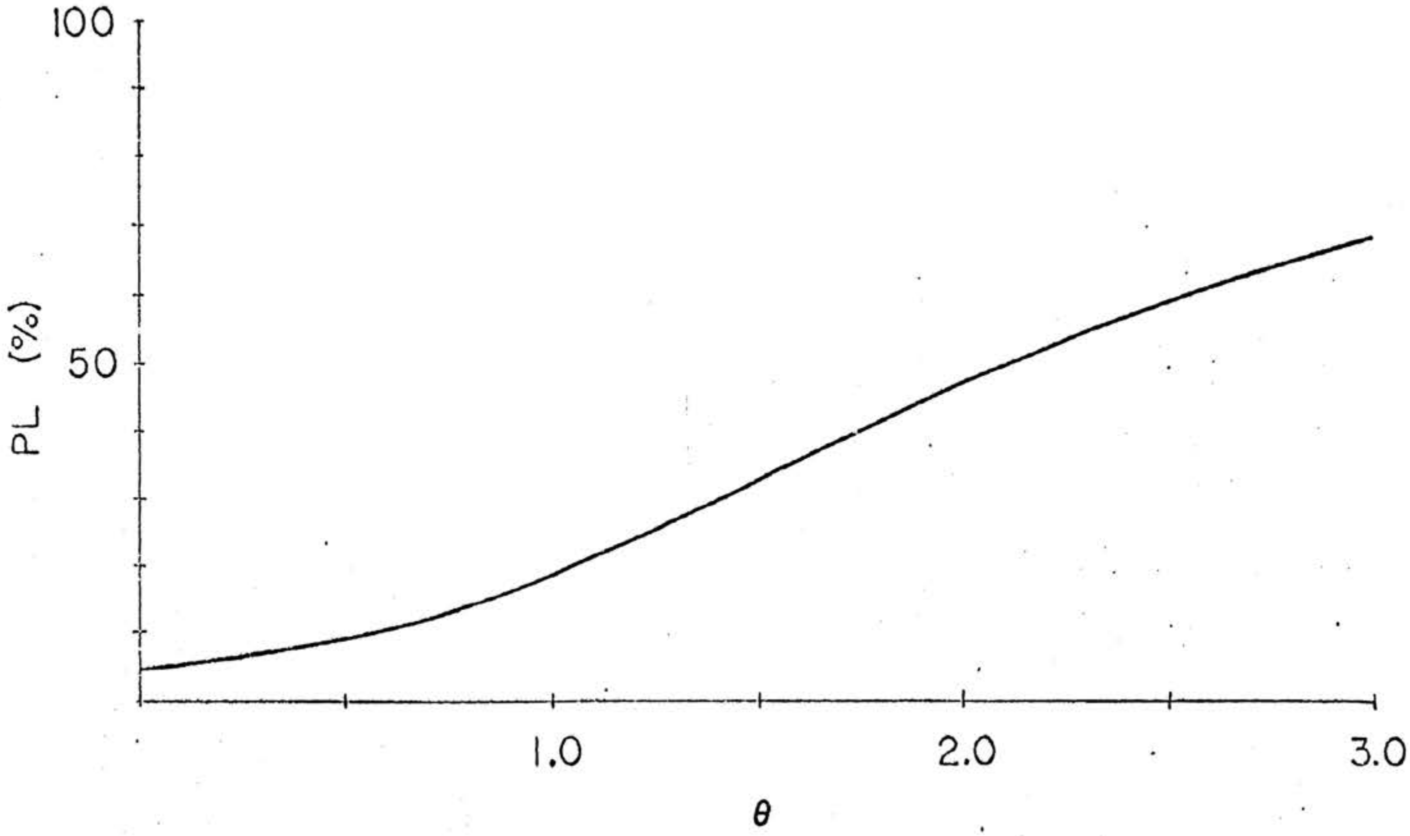

Figure 4.2 Percent Loss for the Sequential W.M.W. U Detector in Normal Noise

esis, H_i. In the two detectors compared, the decision rules were the same, the only difference comes in the fact that U_n does not contain all of the information available in the sample, so then the loss comes about by using an inefficient statistic, not an inefficient decision rule. Also note the fact that for high signal to noise ratios the Wald detector is much better than the sequential W.M.W. U detector. This occurs because the ranks of the samples do not contain information as to the amount of spread in the observations. The W.M.W. U statistic has the same value whether $x_i < y_j$ by 10 units or 1 unit. Thus while the W.M.W. U statistic is not as efficient for extremely high signal to noise ratios, in the areas of interest, however, it is very good.

Now that the detector has been analyzed and compared in several respects to other detectors, it will be used to actually detect a constant signal in normally distributed noise. The noise will be assumed to have a normal distribution with zero mean and unit variance. The value of the constant signal will be θ. Thus the signal to noise ratio as defined in Chapter II is also equal to θ. The detector will now operate in the following manner. First an estimate will be obtained for the signal to noise ratio. Once this has been obtained γ is found from Fig. 4.1. The next step is to calculate the thresholds from the desired α and β. The detector will then take a sample from the X distribution, calculate τ_n and compare it to the thresholds. If another sample is needed the detector will sample the Y distribution and again calculate τ_n and compare it to the thresholds. The detector will continue this process until one of the hypotheses can be accepted.

This problem was simulated on an I.B.M. 360 computer. The X distribution was generated as N(0,1), while the Y distribution was generated as $N(\theta,1)$ if H_1 was true and as N(0,1) if H_0 was true. The flow chart for the simulation program is shown in Fig. 4.3. This chart shows how the detection process was simulated by the computer. Fig. 4.4 and 4.5 plot two typical results for τ_n as a function of n and show how the final termination occurs as the statistic crosses the threshold. The simulation was carried out 2000 times with H_0 true and 2000 times with H_1 true. The results of this simulation are summarized in Table I.

In addition to actually simulating the experiment, the computer kept a record of how many samples were needed for detection each time a decision was made. With this record it is possible to get an idea of how the number of samples function is distributed. The results are plotted in Fig. 4.6 and 4.7. It is useful to note that while the average number of samples for

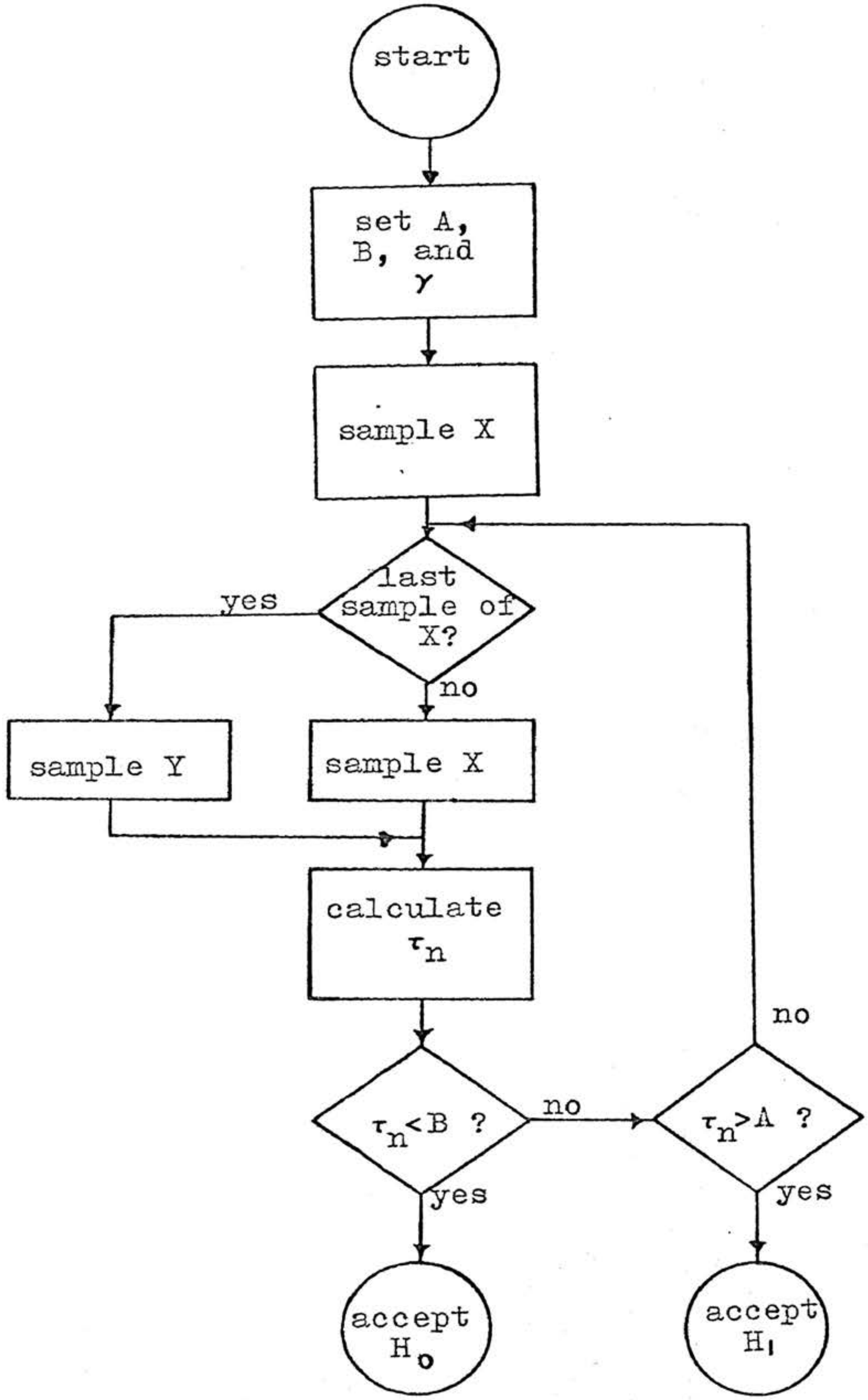

Figure 4.3 Flow Chart for Simulation of Sequential W.M.W. U Detector

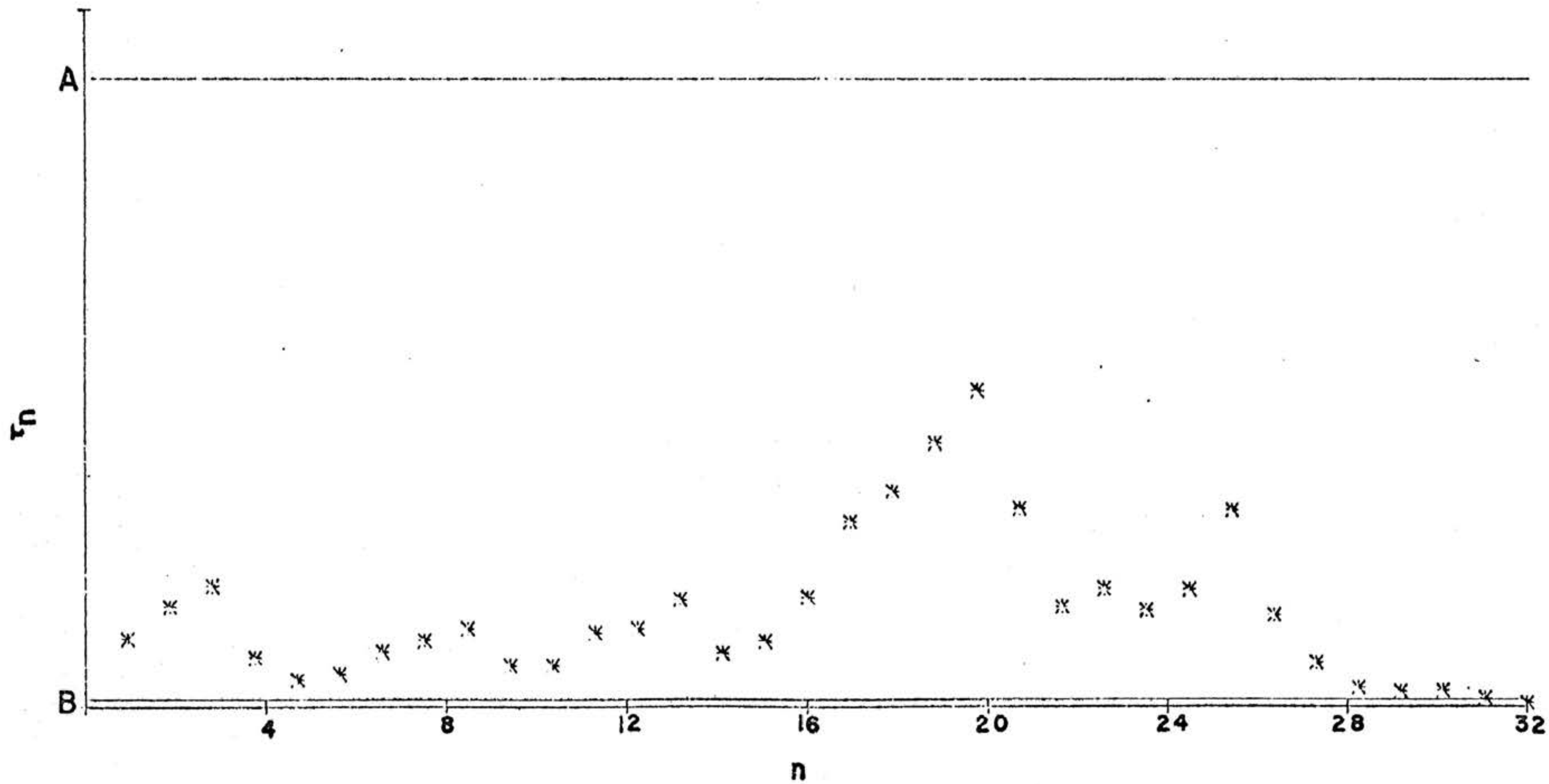

Figure 4.4 Sample Problem for W.M.W. U Detector

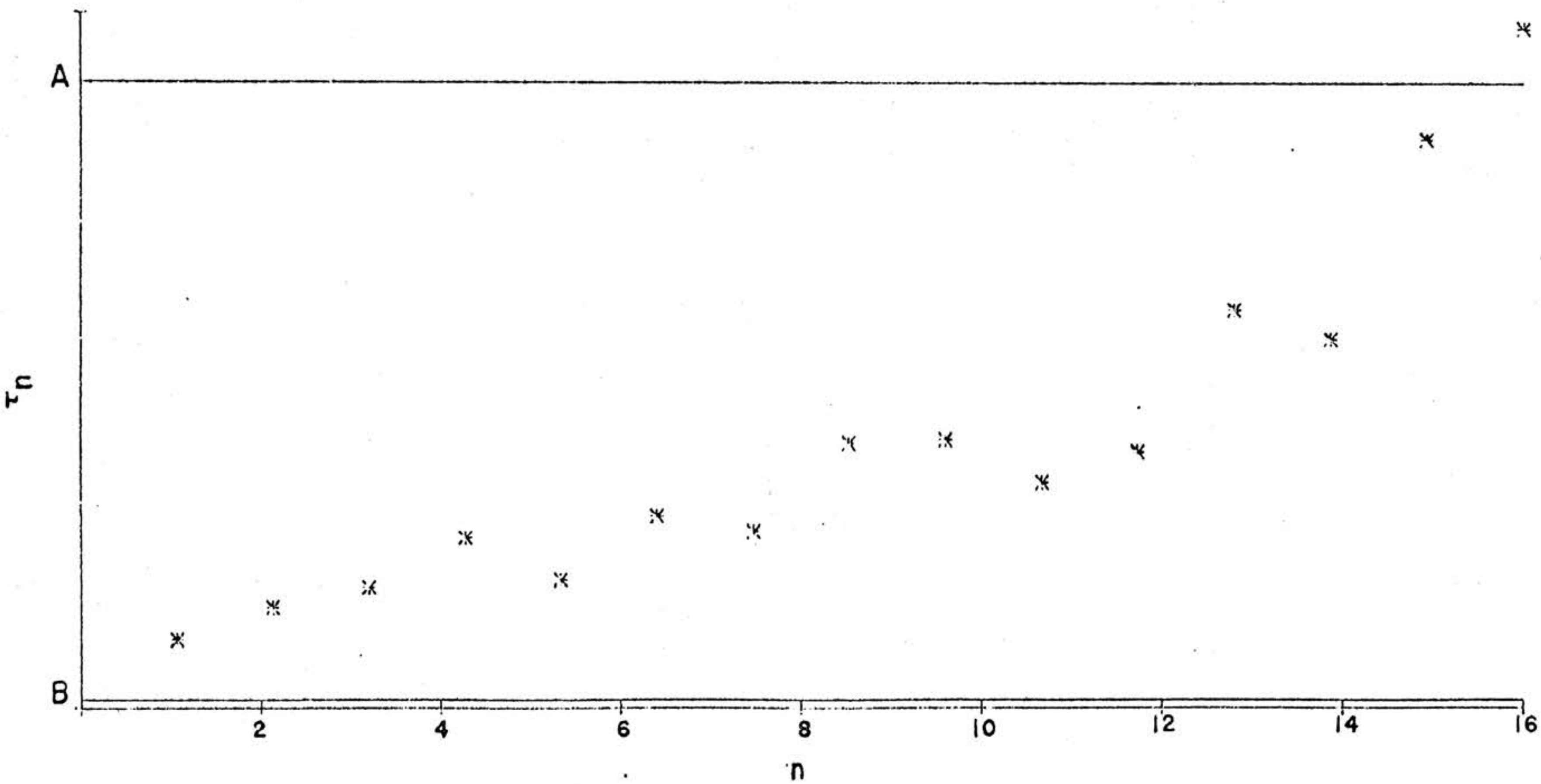

Figure 4.5 Sample Problem for W.M.W. U Detector

TABLE I

RESULTS OF W.M.W. U DETECTOR SIMULATION IN NORMAL NOISE

	H_0 TRUE	H_1 TRUE
alpha	0.10	0.10
beta	0.10	0.10
theta	1.00	1.00
number of times H_0 accepted	1863	114
number of times H_1 accepted	137	1886
simulated alpha	0.069	———
simulated beta	———	0.057
average number of samples	21.85	22.00

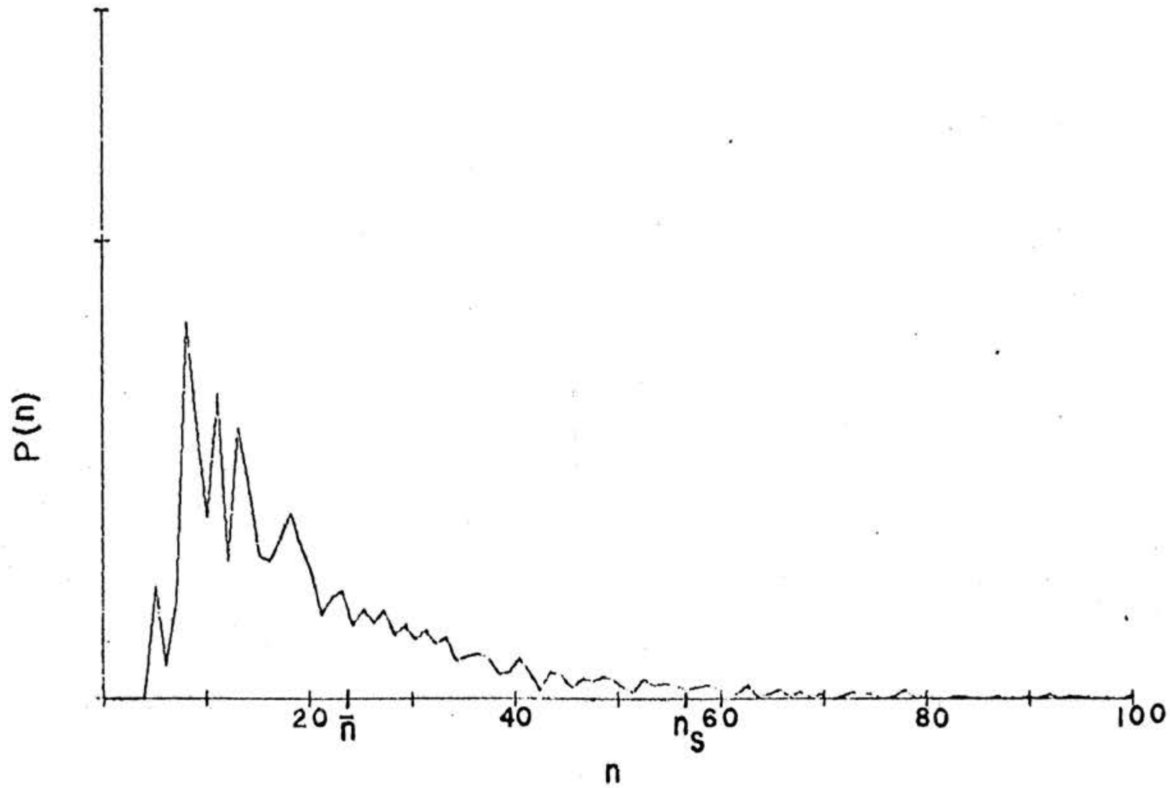

Figure 4.6 Simulated Density for Number of Samples Necessary for Detection: H_0 True

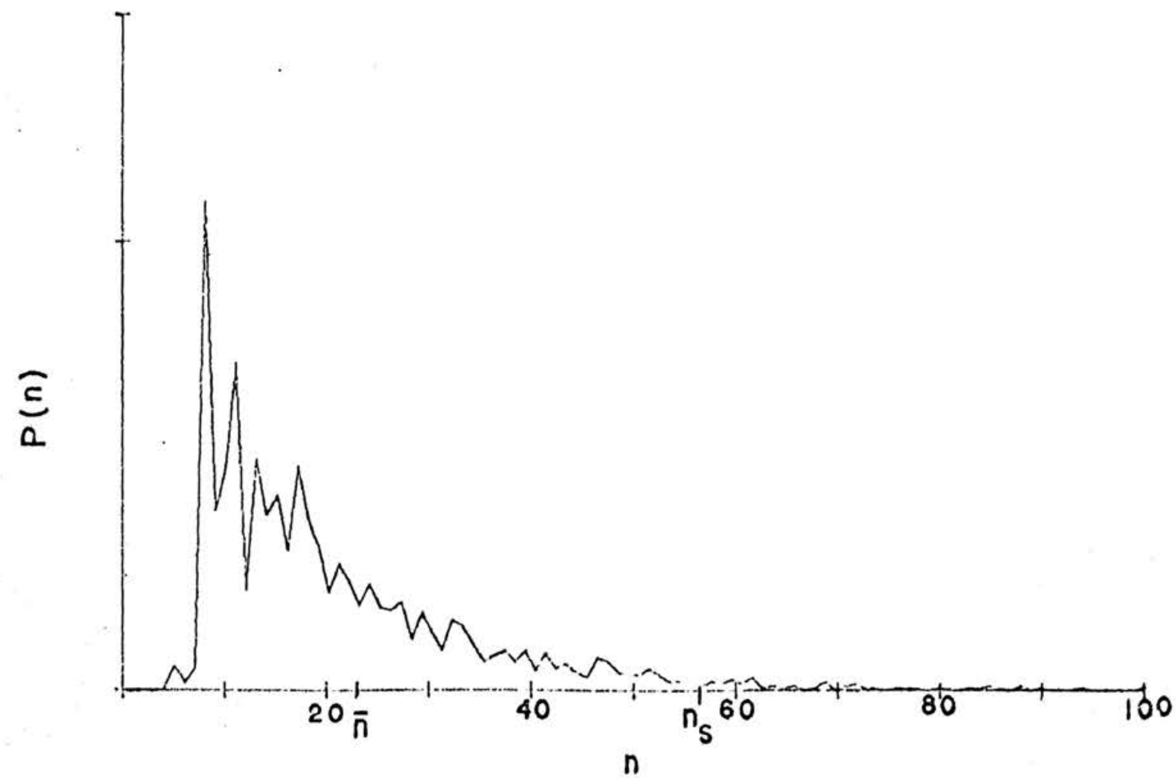

Figure 4.7 Simulated Density for Number of Samples Necessary for Detection: H_1 True

detection, $\bar{n}$, is greater than the mode, indicating a long tail on the distribution, the tail has extremely small values by the time it gets to the number of samples needed by the fixed sample size test, n_s. This is a good property in that the sequential detector will make a decision before the fixed sample size detector a large percentage of the time. It has been seen from Fig. 3.4 and 3.5 that for $\alpha = \beta = 0.1$ the savings of the sequential detector is about 60%, thus the sequential detector will, on the average, make its decision twice as fast as the fixed sample size detector.

4.2 Detection Of A Constant Signal In Nonnormal Noise

In some applications it is not possible to make the assumption that the noise is normally distributed. If this is the case it is sometimes useful to assume that the noise has a Cauchy distribution. If a variable x has a Cauchy distribution then

$$f(x) = 1/(1 + x^2). \qquad -\infty < x < \infty$$

Fig. 4.8 presents a comparison of the Cauchy distribution and the normal distribution with zero mean and unit variance. From Fig. 4.8, it is possible to see that the Cauchy noise has much more power associated with it because it does not go to zero as fast as the normal noise. Theoretically the variance for the Cauchy

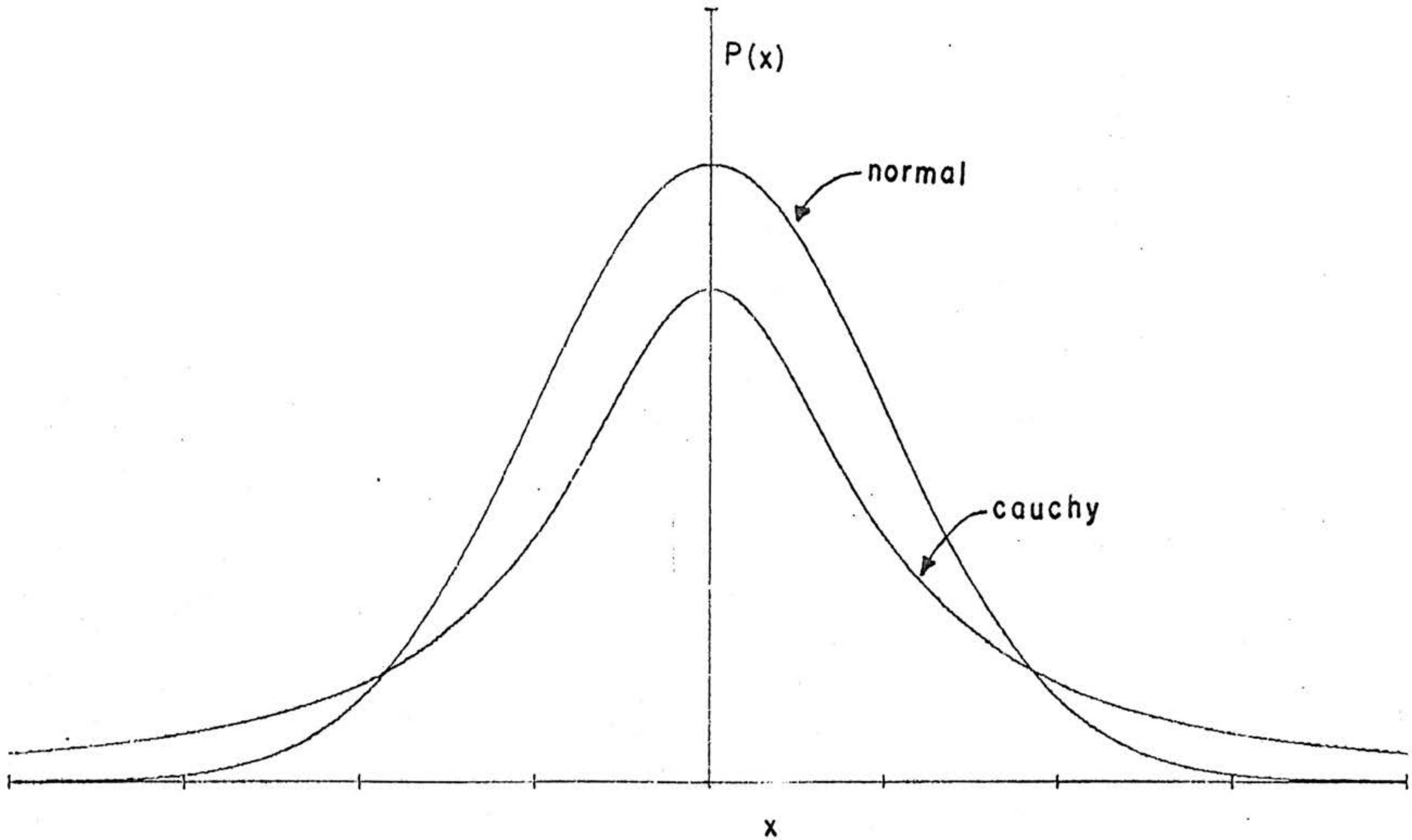

Figure 4.8 Cauchy and Normal Distributions

distribution is infinite, but this distribution can be used to approximate noise with large finite power to a better degree than can the normal distribution.

Again before the detector is used, it is necessary to obtain an estimate of $\gamma(\theta)$. For the case where the noise is distributed according to a Cauchy distribution Eq. 3.4 becomes

$$\gamma(\theta) = \int_{-\infty}^{\infty} f(x)F(x-\theta)\,dx$$

where

$$f(x) = 1/(1 + x^2)$$

and

$$F(x) = \tfrac{1}{2} - \tfrac{1}{\pi}\tan^{-1}(x).$$

An explicit expression for γ was not found, however, by using numerical integration $\gamma(\theta)$ was found. The results are plotted in Fig. 4.9.

Using the above results to set γ, the computer was again used to simulate the detection problem; however, this time Cauchy noise was used. The results of 2000 simulations of H_0 and 2000 simulations of H_1 are shown in Table II.

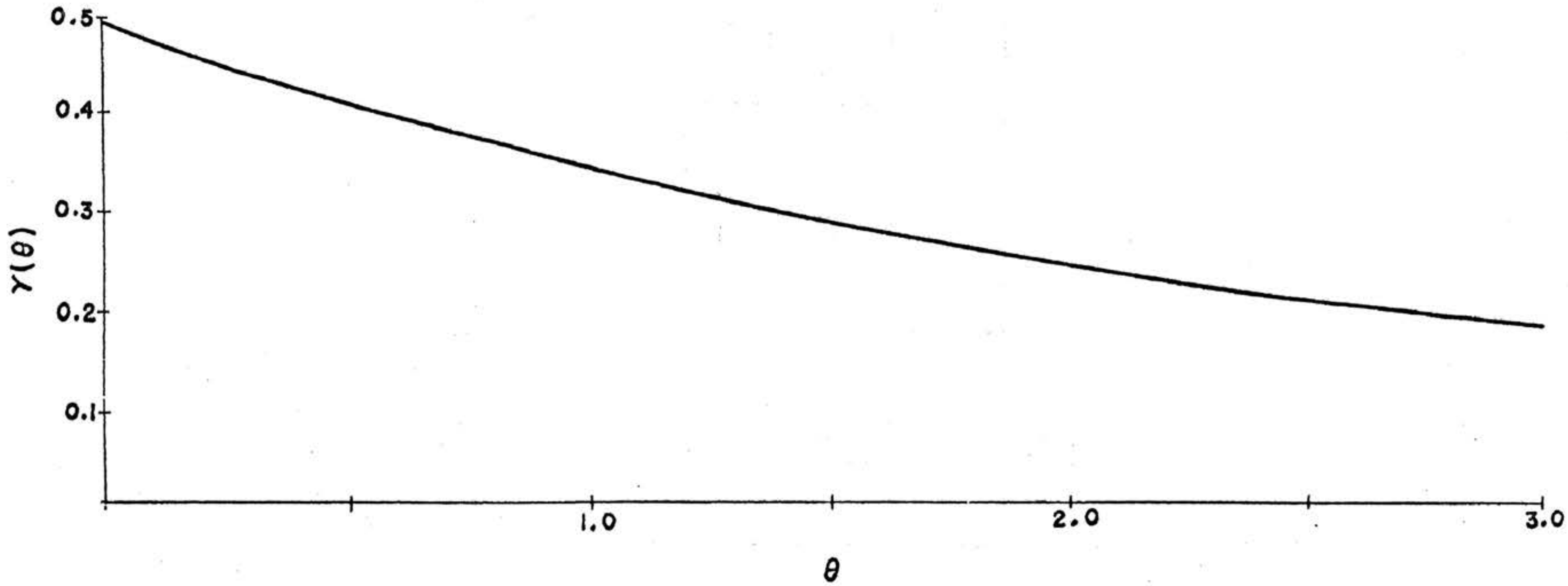

Figure 4.9 $\gamma(\theta)$ for Cauchy Noise

TABLE II

RESULTS OF W.M.W. U DETECTOR

SIMULATION IN CAUCHY NOISE

	H_0 TRUE	H_1 TRUE
alpha	0.10	0.10
beta	0.10	0.10
theta	1.75	1.75
number of times H_0 accepted	1860	188
number of times H_1 accepted	140	1812
simulated alpha	0.070	———
simulated beta	———	0.094
average number of samples	23.10	24.91

The detection process described in the above sections makes use of the fact that the signal to noise ratio was known exactly and thus γ could be calculated. If γ is not known or cannot be calculated, it can be seen from the operating-characteristic curves (Fig. 3.3) that if a conservative estimate for γ is assumed, i.e., assume γ larger than what the true value is expected to be, the detector will operate with a lower error rate than if the true value were used, but the number of samples needed for detection will be greater than would be necessary if the true value of γ were used.

CHAPTER V

ADAPTIVE SEQUENTIAL DETECTION USING THE W.M.W. U DETECTOR

The W.M.W. U detector described in the previous chapters operated on the assumption that γ was known. This is tantamount to assuming a form for the density under the alternative and assuming knowledge of the channel over which the signal is transmitted. The argument for the use of the sequential W.M.W. U detector is based on the fact that no apriori knowledge of the signal is necessary. Thus some way must be found to get rid of this restriction placed on the detector by the need for a value for the parameter γ. It was pointed out in the last chapter that if a conservative value for γ is assumed the detector will be able to detect the signal with values of α and β at least as good as the ones used in the design equations. However, since the number of samples necessary for detection goes up with the square of the difference between γ and $\frac{1}{2}$, the above method is inefficient in this respect. Under these circumstances it seems reasonable to design some sort of system which is capable of extracting information about the parameter γ from the input signal itself.

Systems which extract information about the channel from the information bearing signal as it passes through the system, and use this knowledge to help make decisions are called adaptive systems. Such systems are necessarily suboptimum in that if no apriori knowledge is assumed the detector cannot operate with maximum efficiency. These systems can, however, with good measurements, approach the optimum system.

Consider the following system. In the signalling interval the transmitter sends either s(t) or 0 (no signal) over the channel. The sequential W.M.W. U detector then samples the signal and when τ_n crosses one of the thresholds it makes a decision. The probability of error for such a scheme is given by

$$P_E = \tfrac{1}{2}\alpha + \tfrac{1}{2}L(\gamma_t)$$

where $L(\gamma_t)$ is the operating-characteristic evaluated at the true value of γ denoted by γ_t.

If the designed value of γ (γ_d) is greater than γ_t, P_E will be less than $\frac{1}{2}(\alpha + \beta)$ (the P_E when $\gamma_d = \gamma_t$). However, the number of samples for detection will be greater than the number which occurs if $\gamma_d = \gamma_t$. If $\gamma_d > \gamma_t$, P_E will be greater than $\frac{1}{2}(\alpha + \beta)$.

A reasonable solution for this problem is to have the detector measure γ in some way and use this knowledge to adapt the detection process to account for this measured γ. One way to do this is to change the value of γ each time the alternate hypothesis is accepted, the new value of γ being taken as the value observed during the detection interval. This method is reasonable, but since there is error in any measurement, it is sometimes better to use several measurements rather than one. Proceeding along this line, it is reasonable to average the value of γ observed in the K previous intervals where H_1 was accepted. While the use of K observation intervals to estimate γ implies that γ_d will not follow changes in the true value of γ as fast as when K = 1, it is also not as subject to measurement error as when K = 1. So if the channel varies slowly and remains essentially constant during several observation intervals, it is better to use this averaged γ. If the probability of signal (s) is $\frac{1}{2}$ and if the channel is essentially constant over 2K observation intervals, then the detector can use K estimates in calculating γ.

The sequential W.M.W. U detector calculates a statistic (U) which is directly related to γ. Since

$$U = \sum_i \sum_j x_{ij}$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } x_i > y_j \\ 0 & \text{if } x_i < y_j. \end{cases}$$

An estimate of γ, say $\hat{\gamma}$, can be calculated by

$$\hat{\gamma} = \begin{cases} \dfrac{4U}{n^2} & n \text{ even} \\ \\ \dfrac{4U}{n^2 - 1} & n \text{ odd} \end{cases}$$

The analytical analysis of a system like the one described above would be extremely difficult. For this reason it will not be attempted, instead a Monte Carlo simulation will be used to analyze the operation. A random number generator will determine if H_0 or H_1 is true. If H_0 is true both sample populations will be the same. If H_1 is true then $P(x>y)$ will be γ_t. The detector will then operate on the input populations and make its decision. Whenever H_1 is accepted, the detector will modify γ_d using the average of the previous K $\hat{\gamma}$'s. The simulation will count the number of errors made after the system has settled down (approximately 2K signal intervals) and determine P_E vs. γ for several values of K.

One problem in any simulation like this is a starting value for γ_d. To demonstrate the stability of this system the first K values of $\hat{\gamma}$ will be sampled from a uniform density over the interval $(0,\frac{1}{2})$. The flow chart for the simulation is given in Fig. 5.1,

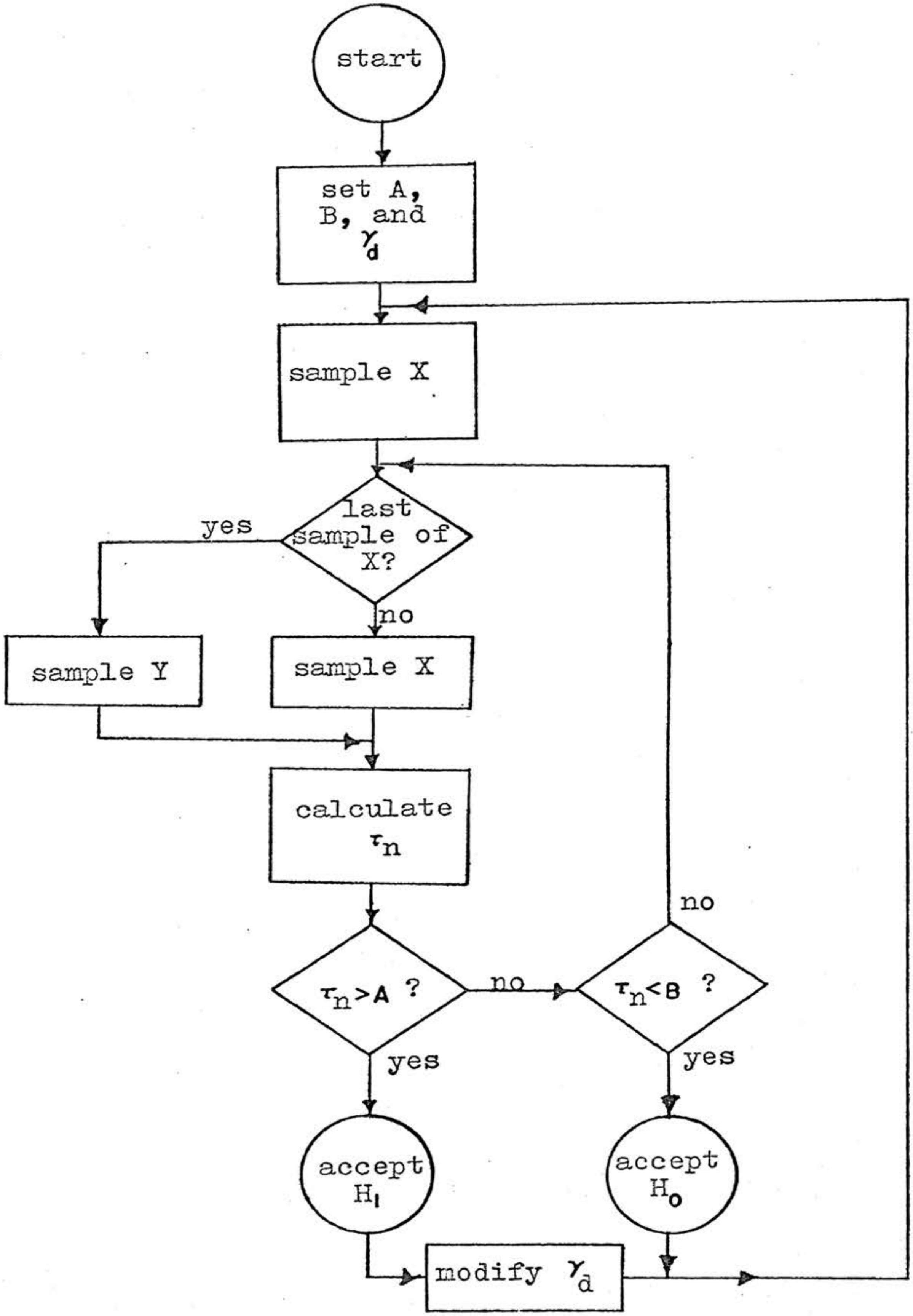

Figure 5.1 Flow Chart for Adaptive Sequential Detection

and a copy of the simulation program is shown in Appendix C. The results of the simulation are shown in Table III.

TABLE III

RESULTS OF ADAPTIVE

W.M.W. U SIMULATION

θ \ K	1	2	5	10
0.6	0.1968	0.2258	0.1939	0.1417
0.8	0.1406	0.1028	0.0714	0.0854
1.0	0.0703	0.0827	0.0449	0.0750

CHAPTER VI

CONCLUSIONS

In this paper the design of an adaptive sequential nonparametric detector was fully developed. First, it was shown that with the use of a sequential ranking procedure it is possible to adapt the Wilcoxon-Mann-Whitney U statistic to a sequential testing scheme. It was then argued that since the statistic used for the test is distribution free the new sequential test is also distribution free, i.e., the test statistic does not depend upon the distribution of the noise. Once this had been found it then became possible to develop the form of the operating-characteristic function and the average number of samples function. With these two functions it is possible to completely describe the operation of the detector under any noise conditions.

The detector was then compared to the fixed sample size detector using the same statistic, and it was found that on the average the sequential detector needs half as many samples for detection as the fixed sample size detector. The sequential nonparametric detector was then compared to the optimum parametric detector for the case of normal noise. In this comparison it was found that the nonparametric detector had very little loss in efficiency for small or moderate signal to noise ratios.

The sequential nonparametric detection process was then simulated on the computer, and it was found that the detector could operate at low error rates for given signal to noise ratios. Finally, the detector was modified so that it could make its own measurements of the pertinent channel characteristics and adapt itself to these changing characteristics. Again the problem was simulated on the computer, and it was found that this adaptive detector can operate with low error rates.

Thus it has been shown that even though the detector described here is nonparametric in nature it could operate with little loss in efficiency when compared to the optimum detector. Also, since the detector is nonparametric, it can operate over a full range of noise distributions with out any modification, while the optimum detector must be changed each time the noise changes. This versatility coupled with the relatively small loss in efficiency makes the idea of using a sequential nonparametric detector extremely practical when there may be a question as to the exact form of the noise or signal plus noise.

APPENDIX A

PROOF OF EQUIVALENCE OF THE WILCOXON AND MANN-WHITNEY STATISTICS

The definition of T is given in Eq. 3.1, i.e.,

$$T = \sum_{i=1}^{N+M} w_i$$

where

$$w_i = \begin{cases} i & \text{if } z_i \text{ is from X} \\ 0 & \text{if } z_i \text{ is from Y.} \end{cases}$$

If all of the x_i's are less than all of the y_j's then

$$T = N(N+1)/2$$

where N is the number of x_i's.

When one y_j is less than one of the x_i's the rank of that x_i is increased by one, while the rest are left the same. Thus for this case

$$T = 1 + N(N+1)/2.$$

Each time a y_j preceeds an x_i the rank of that x_i is increased by one. So if N_y is the number of times a y_j preceeds an x_i then

$$T = N_y + N(N+1)/2. \qquad (A.1)$$

But N_y is the same as U. This can be seen by examining Eq. 3.2, i.e.,

$$U = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij}$$

where

$$x_{ij} = \begin{cases} 1 & \text{if } x_i > y_j \\ 0 & \text{if } x_i < y_j. \end{cases}$$

Thus U can be thought of as the number of times a y_j preceeds an x_i. Then Eq. A.1 becomes

$$T = U + N(N+1)/2$$

or

$$U = T - N(N+1)/2. \qquad (A.2)$$

APPENDIX B

A.R.E. OF W.M.W. U TEST AGAINST t-TEST

The asymptotic relative efficiency (A.R.E.) of one test as compared to another is defined as follows (9,10): Given two detectors, each with the same α and β, the first with sample size N and the second with sample size N*, then the A.R.E. of the second with respect to the first is

$$\text{A.R.E.} = \lim_{n \to \infty} (N/N^*)$$

where θ is the signal to noise ratio.

Denote the test statistics by t_1 and t_2. Now make the following definitions

$$\left.\begin{aligned} E_{ij} &= E(t_i \mid H_j) \\ D_{ij} &= \mathrm{var}(t_i \mid H_j) \end{aligned}\right\} \qquad \begin{aligned} i &= 1,2 \\ j &= 1,2 \end{aligned}$$

and

$$E_i^{(r)}(0) = \left. \frac{\partial^r E(t_i\ \theta)}{\partial \theta^r} \right|_{\theta = 0.}$$

Now for the t-test with two populations

$$t = \bar{y} - \bar{x}.$$

If the alternate hypothesis is of the form

$$G(x) = F(x - \theta)$$

then

$$E_{10} = 0$$
$$E_{11} = \theta$$

and

$$D_{11}^2 = \frac{\sigma_x^2}{N} + \frac{\sigma_y^2}{M}$$

where σ_x^2 and σ_y^2 are the variances of the X and Y populations respectively. For the case considered here $\sigma_x^2 = \sigma_y^2 = \sigma^2$. If m_i is the first value of r such that

$$E_i^{(r)}(0) \neq 0$$

then it can be seen that in the case of the t-test

$$E_1^{(1)}(0) = 1$$

thus

$$m_1 = 1.$$

Now assume that

$$\frac{E_i^{(m_i)}(0)}{D_{i0}} \sim c_i n^{\delta_i}. \tag{B.1}$$

Then for t

$$\frac{E_1^{(1)}(0)}{D_{10}} = \frac{1}{\sigma\sqrt{\frac{1}{N} + \frac{1}{M}}} = \frac{1}{\sigma}\sqrt{\frac{NM}{N+M}}$$

since $N+M = n$ then

$$c_1 = \frac{1}{\sigma}$$
$$\delta_1 = \tfrac{1}{2}.$$

For the W.M.W. U test

$$t = \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij}$$

where x_{ij} is defined in Eq. 3.2. From Eq. 3.4 and 3.5

$$P(x_i < y_j) = \int_{-\infty}^{\infty} f(x)F(x-\theta)\,dx.$$

Now it can be seen that

$$\frac{\partial E(t_2 \mid \theta)}{\partial \theta} = -NM\int_{-\infty}^{\infty} f(x)f(x-\theta)\,dx$$

or

$$E_2^{(1)}(0) = NM\int_{-\infty}^{\infty}\left[f(x)\right]^2 dx.$$

Thus $m_2 = 1$, and from Chapter III it is known that

$$\text{var}(t_2\ 0) = NM(N+M+1)/12.$$

Then

$$\frac{E^{(1)}(0)}{D_{20}} = \sqrt{\frac{12NM}{n+1}}\int_{-\infty}^{\infty}\left[f(x)\right]^2 dx$$

$$c_2 = \sqrt{12}\int_{-\infty}^{\infty}\left[f(x)\right]^2 dx$$

$$\delta_2 = \tfrac{1}{2}.$$

If

$$\lim_{n \to \infty} \frac{E_i^m i(\theta_i)}{E_i^m i(0)} = 1 \qquad (B.2)$$

and

$$\lim_{n \to \infty} \frac{D_{i1}}{D_{i0}} = 1 \qquad (B.3)$$

then

$$A.R.E. = \lim_{n \to \infty} \left(\frac{c_2}{c_1}\right)^{\frac{1}{\delta_1}}$$

if $m_1 = m_2$ and $\delta_1 = \delta_2$. From examination of the means and variances, it is seen that the above regularity conditions (Eq. B.2 and B.3) hold so the A.R.E. is given by

$$A.R.E. = 12\sigma^2 \left[\int_{-\infty}^{\infty} \left[f(x)\right]^2 dx \right]^2 .$$

Since f(x) is a normal distribution the above becomes

$$A.R.E. = 12\sigma^2 \left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{\sigma^2}} dx \right]^2$$

$$= 12\sigma^2 \left[\frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{\sigma^2}} \, dx \right]^2$$

$$\text{A.R.E.} = 3/\pi \,. \qquad \text{(B.4)}$$

APPENDIX C

PROGRAM LISTING

A program listing for the adaptive simulation problem follows.

```
C
C      THESIS PROGRAM  JAMES FOWLER
C      ADAPTIVE SIMULATION
C      SIMULATION OF WMW-U SEQUENTIAL TEST
C      ALPHA= TYPE I ERROR
C      BETA= TYPE II ERROR
C      NT= NUMBER OF TIMES SIMULATION OCCURS
C      C1=1.0 IMPLIES H(1) TRUE
C      C1=0.0 IMPLIES H(0) TRUE
C      TA= TEST STATISTIC
C      GA= P(X > Y)
C      U= MANN-WHITNEY U STATISTIC
C      T= WILCOXON T STATISTIC
C      X(I)= SAMPLE OF X IF I ODD
C      X(I)= SAMPLE OF Y IF I EVEN
C      SGMA2= VARIANCE OF U UNDER H(0)
C      SGM2A= VARIANCE OF U UNDER H(1)
C      Z0I= U - MEAN OF U UNDER H(0)
C      Z1I= U - MEAN OF U UNDER H(1)
C      XG(I)= STORAGE OF PREVIOUS ESTIMATES OF GAMMA
C      IH= RECORD OF TRUE HYPOTHESES AND DECISIONS
C      K= NUMBER OF INTERVALS USED IN FINDING GAMMA
C      PED= DESIGNED PROBABILITY OF ERROR
C      PESS= SIMULATED STEADY STATE ERROR
C      AVNS= AVERAGE NUMBER OF SAMPLES
C      AVNSS= STEADY STATE AVERAGE NUMBER OF SAMPLES
C      ERR= NUMBER OF ERRORS
C      ERRSS= NUMBER OF ERRORS AFTER 2K TRIALS
C      MOD(I)=0   I EVEN
C      MOD(I)=1   I ODD
C      IF TAU > A ACCEPT H(1)
C      IF TAU < B ACCEPT H(0)
C
C      RANDZ:  GENERATES A GAUSSIAN VARIABLE
C      G(I)= GAUSSIAN CUMULATIVE
C      EACH INCREMENT IN I CORRESPONDS TO 0.01 UNITS
C      G(360):  CORRESPONDS TO G(0.0)
C      GE= ESTIMATE OF GAMMA
C
       DIMENSION X(900),S(900)
       DIMENSION XG(10),IH(1000)
       DIMENSION G(720)
       MOD(I)=-((I/2)+(I/2)-I)
       READ(1,101)ALPHA,BETA
       READ(1,103)(G(I),I=360,719)
       READ(1,102)NT,TH
       READ(1,102)K
       DO 2 I=1,359
       J=720-I
     2 G(I)=1.0-G(J)
C      INITIALIZE RANDOM NUMBER GENERATOR
       IX=1371897
       TH1=TH/1.414214
       NTH=(100.0*TH1)
```

```
      GA=1.0-G(NTH+360)
      A=(1.0-BETA)/ALPHA
      B=BETA/(1.0-ALPHA)
C     WRITE HEADINGS
      WRITE(3,104)B,A
      WRITE(3,113)ALPHA,BETA,NT
      WRITE(3,115)TH,GA
      XNT=NT
      T2K=2*K
      XK=K
C     INITIALIZE COUNTERS
      AVNS=0.0
      ERR=0.0
      ERRSS=0.0
      AVNSS=0.0
      GE=0.0
C     OBTAIN STARTING VALUE FOR GAMMA
      NNN=1
      DO 32 I=1,K
      CALL RANDU(IX,IY,Y)
      IX=IY
      XG(I)=0.5*Y
      NNN=NNN+1
   32 GE=GE+XG(I)/XK
      XN2K=0.0
      DO 3 LL=1,NT
C     DECIDE WHICH HYPOTHESIS IS TRUE
      CALL RANDU(IX,IY,Y)
      IX=IY
      NC1=(0.50+Y)
      IF(LL-T2K)33,34,34
   34 XN2K=1.0
   33 CONTINUE
C     INITIALIZE VARIABLES
      C=0.0
      C1=NC1
      Z0I=0.0
      Z1I=0.0
      SGMA2=0.0
      SGM2A=0.0
      U=0.0
      T=1.0
      S(1)=1.0
      TA=1.0
C     GENERATE FIRST SAMPLE
      X(1)=RANDZ(0)
      C=C1-C
C     GENERATE REST OF SAMPLES AS NEEDED
      DO 4 J=2,900
      X(J)=C*TH+RANDZ(0)
      CN=0.0
      CX=0.0
C     CALCULATE S(J)
      DO 5 I=1,J
```

```
      IF(X(I)-X(J))7,7,9
    7 CN=CN+1.0
      GO TO 5
    9 XX=MOD(I)
      CX=CX+XX
    5 CONTINUE
      S(J)=CN
C     CALCULATE T AND U
      XX=MOD(J)
      T=T+(XX*S(J))+CX
      C=C1-C
      XNX=(J+1)/2
      U=T-((XNX*(XNX+1.0))/2.0)
      J2=MOD(J)
      IF(J2)11,11,10
   10 CONTINUE
C     J IS ODD
      NM=(J-1)/2
      XNM1=NM
      XNM2=NM+1
      GO TO 13
   11 CONTINUE
C     J IS EVEN
      NM=J/2
      XNM1=NM
      XNM2=NM
   13 CONTINUE
C     CALCULATION OF TA
      XNM=NM
      XJ=J
      Z0I=U-0.50*XNM1*XNM2
      Z1I=U-GE*XNM1*XNM2
      SGMA2=(XNM1*XNM2*(XNM1+XNM2+1.0))/12.0
      SGM2A=SGMA2
      Z0=Z0I/(SQRT(SGMA2))
      Z1=Z1I/(SQRT(SGM2A))
      Z02=Z0**2
      Z12=Z1**2
      TA=EXP(0.50*(Z02-Z12))
C     COMPARE TA TO THRESHOLDS
      IF(B-TA)15,16,16
   16 CONTINUE
C     ACCEPT H(0)
      XXJ=J
      ND=0
      AVNS=AVNS+XXJ
      AVNSS=AVNSS+XN2K*XXJ
      GO TO 31
   15 IF(A-TA)17,17,4
   17 CONTINUE
C     ACCEPT H(1)
      XXJ=J
      ND=1
      AVNS=AVNS+XXJ
```

```
      AVNSS=AVNSS+XN2K*XXJ
      GO TO 31
    4 CONTINUE
C     CAN NOT MAKE A DECISION
      ND=2
      AVNS=AVNS+100.0
      AVNSS=AVNSS+100.0*XN2K
   31 CONTINUE
C     RECORD DECISION
      IH(LL)=10*NC1+ND
C     RECORD MISTAKES
      IF(NC1-ND)36,37,36
   36 ERR=ERR+1.0
      ERRSS=ERRSS+XN2K
   37 CONTINUE
C     MODIFY ESTIMATE OF GAMMA
      IF(ND-1)38,39,38
   39 GE=GE-(XG(K)/XK)
      KM1=K-1
      DO 40 II=1,KM1
      II1=K-II
   40 XG(II1+1)=XG(II1)
      XG(1)=(4.0*U)/(XXJ**2)
      GE=GE+(XG(1)/XK)
      NNN=NNN+1
   38 CONTINUE
    3 CONTINUE
C     CALCULATE ERROR RATES
      XNT1=XNT-T2K
      PE=ERR/XNT
      PESS=ERRSS/XNT1
      AVNS=AVNS/XNT
      AVNSS=AVNSS/XNT1
C     WRITE RESULTS
      WRITE(3,118)K
      WRITE(3,116)AVNS,AVNSS
      WRITE(3,117)PE,PESS
      WRITE(3,120)
      WRITE(3,119)(IH(LL),LL=1,NT)
  101 FORMAT(4E16.8)
  102 FORMAT(5X,I6,E15.8,I6)
  103 FORMAT(10F7.5)
  104 FORMAT(1H1/5X,'ACCEPT H(0) IF TA < B'/5X,'ACCEPT',
     1' H(1) IF TA > A',/5X,'TAKE ANOTHER SAMPLE IF ',
     2'B < TA < A'/5X,'B =',F10.4/5X,'A =',F10.4/)
  111 FORMAT(5X,10I6)
  113 FORMAT(/5X,'THEORETICAL ALPHA=',F6.4,/5X,
     1'THEORETICAL BETA =',F6.4,/5X,'NUMBER OF TIMES',
     2' EXPERIMENT IS REPEATED =',I6,/5X,'PROBABILITY',
     3' OF H(0) = 1/2',/)
  115 FORMAT(/5X,'SIGNAL TO NOISE RATIO =',F8.4,/5X,
     1'P(X > Y) =',F6.4/)
  116 FORMAT(//5X,'AV. NO. OF SAMPLES =',F8.3,/5X,
     2'S.S. AV. NO. OF SAMPLES =',F8.3,/5X,/)
```

```
117 FORMAT(//5X,'P. OF ERROR =',F7.4,/5X,
   1'S.S. P. OF ERROR =',F7.4,/5X,/)
118 FORMAT(//5X,'MEMORY LENGTH OF ADAPTER =',I3,/)
119 FORMAT(25I3)
120 FORMAT(1H1,/5X,'DETECTION DECISIONS AND CORRECT',
   1' VALUES',/5X,'1=H(1)    0=H(0)    2=NO DECISION',
   2/5X,'FIRST NUMBER = XMITTED SIGNAL',/5X,
   3'SECOND NUMBER = RECVD SIGNAL',/)
    STOP
    END
```

BIBLIOGRAPHY

1. C.W. HELSTROM, Statistical Theory of Signal Detection, New York; Pergamon, 1960.

2. R.V. HOGG and A.T. CRAIG, Introduction to Mathematical Statistics, New York; The Macmillan Company, 1965.

3. A.M. MOOD and F.A. GRAYBILL, Introduction to the Theory of Statistics, New York; McGraw-Hill, 1963.

4. A. WALD, Sequential Analysis, New York; John Wiley and Sons, 1963.

5. F. WILCOXON, "Individual Comparisons by Ranking Methods", Biometrics Bull., vol. 1, pp. 80-83, 1945.

6. H.B. MANN and D.R. WHITNEY, "On a Test of Whether One of Two Random Variables is Stochastically Larger Than the Other", Ann. Math Stat., vol. 18, pp. 50-60, 1947.

7. E.A. PARENT, Jr., "Sequential Ranking Procedures", Technical Report No. 80, Dept. of Statistics Stanford University, 1965.

8. J.C. HANCOCK and P.A. WINTZ, Signal Detection Theory, New York; McGraw Hill, 1966.

9. J.A. CAPON, "On the Asymptotic Efficiency of Locally Optimun Detectors", IRE Trans. on Information Theory, vol. IT-7, pp. 67-71, April 1961.

10. M.G. KENDALL and A. STUART, The Advanced Theory of Statistics, New York; Hafner, vol. 2, 1961.

11. M. KANEFSKY, "Detection of Weak Signals With Polarity Coincidence Arrays", IEEE Trans. on Information Theory, vol. IT-12, pp. 260-268, 1966.

12. M. KANEFSKY and J.B. THOMAS, "On Adaptive Nonparametric Detection Systems Using Dependent Samples", IEEE Trans. on Information Theory, vol. IT-11, pp. 521-526, 1965.

13. Y.T. CHIEN and K.S. FU, "A Modified Sequential Recognition Machine Using Time-Varying Stopping Boundaries", IEEE Trans. on Information Theory, vol. IT-12, pp. 206-214, 1966.

14. J.J. BUSSGANG and D. MIDDELTON, "Optimum Sequential Detection of Signals in Noise", IRE Trans. on Information Theory, vol. IT-1, pp. 5-18, December 1955.

15. K.S. FU and Y.T. CHIEN, "Sequential Recognition Using a Nonparametric Ranking Procedure", IEEE Trans. on Information Theory, vol. IT-13, pp. 484-492, 1967.

16. S. SIEGEL, Nonparametric Statistics, New York; McGraw-Hill, 1956.

17. L.A. WAINSTEIN and V.D. ZUBAKOV, Extraction of Signals From Noise, Englewood Cliffs, New Jersey; Prentice-Hall, 1962.

18. R.A. GROENEVELD, "A Nonparametric Rank Correlation Method for Detecting Signals in Additive Noise", IEEE Trans. on Information Theory, vol. IT-13, pp. 315-316, 1967.

19. E.L. LEHMANN, "The Power of Rank Tests", Ann. Math. Stat., vol. 24, pp. 23-43, 1953.

20. E.L. LEHMANN, "Consistency and Unbiasedness of Certain Nonparametric Tests", Ann. Math. Stat., vol. 22, pp. 165-179, 1951.

21. J.A. CAPON, "A Nonparametric Technique for the Detection of a Constant Signal in Additive Noise", 1959 Wescon Convention Record pt. 4, pp. 92-103.

22. F.J. MASSEY, "A Note on the Power of a Nonparametric Test", Ann. Math. Stat., vol. 21, pp. 440-443, 1950.

23. M.E. TERRY, "Some-Rank-Order Tests Which Are Most Powerful Against Specific Alternatives", Ann. Math. Stat., vol. 23, pp. 346-366, 1952.

24. Z.W. BIRNBAUM, "Distribution-free Tests of Fit for Continuous Distribution Functions", Ann. Math. Stat., vol. 24, pp. 1-8, 1953.

VITA

James Charles Fowler was born in Dallas, Texas, on April 26, 1943. He was graduated from Roswell High School, Roswell, New Mexico, in May, 1961. The author was graduated from New Mexico State University in June, 1965, with a degree of Bachelor of Science in Electrical Engineering. He entered the Graduate School of the University of Missouri at Rolla in September, 1965, and he received a Master of Science in Electrical Engineering in June, 1967. He is a member of Phi Kappa Phi, Eta Kappa Nu, and Sigma Pi Sigma.

The author was employed as a scientific aide by the Physical Science Laboratory, University Park, New Mexico, from March, 1963, to June, 1965, and as an assistant engineer from June, 1965, to September, 1965. He is currently a NASA Trainee at the University of Missouri at Rolla.