ENGINEERING RESEARCH INSTITUTE
UNIVERSITY OF MICHIGAN
ANN ARBOR

SIGNAL DETECTABILITY: A UNIFIED DESCRIPTION OF STATISTICAL METHODS EMPLOYING FIXED AND SEQUENTIAL OBSERVATION PROCESSES

Technical Report No. 19
Electronic Defense Group
Department of Electrical Engineering

By: W. C. Fox

Approved by: A. B. Macnee

W. W. Peterson

Project M970

TASK ORDER NO. EDG-3
CONTRACT NO. DA-36-039 sc-15358
SIGNAL CORPS, DEPARTMENT OF THE ARMY
DEPARTMENT OF ARMY PROJECT NO. 3-99-04-042
SIGNAL CORPS PROJECT NO. 29-194B-0

December, 1953

TABLE OF CONTENTS

LIST OF SPECIAL SYMBOLS AND TERMS

In Order of Appearance

Population N	Defined on page 2
Population SN	Defined on page 2
Z_n	A point in n-dimensional space
$f_N(Z_n)$	Probability density function (of dimension n) for population N
$f_{SN}(Z_n)$	Probability density function (of dimension n) for population SN
P_{SN}	Population SN's probability function, defined on page 5
P_N	Population N's probability function, defined on page 5
F	Conditional probability of a false alarm, defined on pages 6 and 17
M	Conditional probability of a miss, defined on pages 6 and 18
$\ell(Z_n)$	Likelihood ratio, defined on page 6
$A(\beta)$	Ratio criterion, defined on page 7
ROC curve	Receiver Operating Characteristic curve, defined on page 8
A_n, B_n, C_n	Sequential criteria, defined on pages 15 and 16
S_n	n-th stage sample space, defined on pages 15 and 16
E_N	N-conditional average sample number, defined on page 19
E_{SN}	SN-conditional average sample number, defined on page 19

The Following are from the Appendices

E^n	The n-dimensional Euclidean Space, E^1 is then the system of real numbers
$f_n: E^n \rightarrow E^1$	A real valued function defined on E^n
$A \times B$	The "Cartesian Product" or the set of all possible pairs of points (a, b) with a in A and b in B. For example, $E^2 = E^1 \times E^1$
$A \cup B$	The set of points which belong either to A, or to B, or to both; read "A union B"

$A \cap B$	The set of points which belong to both A and B; read "the intersection of A with B"
H_o, H_1	Hypotheses, defined on page 32
α, β	The error probabilities, defined on page 32
$E_1(r)$, $E_o(r)$	The expected values of r, defined on page 33
$A \supset B$	A "includes" or "contains" B

ACKNOWLEDGEMENTS

It is impossible to single out credit for one individual in the joint work of a team. Because of this fact, the merits of this report should reflect equally upon Messrs. W. W. Peterson and T. G. Birdsall as well as upon the author, who, however, is alone responsible for all opinions and statements of fact contained herein. In addition to acknowledging the assistance of Mr. P. C. Hayes in the calculations, the author would like to thank Geraldine L. Preston and Jenny-Lea E. Mesler for their patience and skill in preparing the text.

ABSTRACT

Signal detectability has been studied statistically from various points of view. Those involving an observation interval of fixed length are essentially equivalent, as opposed to those which involve a sequential process. Both approaches are discussed with a minimum of mathematics to provide a reasonably non-technical account of the "state of the art." Definitive comparison of the two observation processes is not possible until more general knowledge is available concerning the existence and nature of optimum sequential tests. In addition, a general mathematical formulation of sequential analysis is given in which the current theoretical obstacles in applying it to signal detectability are emphasized.

SIGNAL DETECTABILITY: A UNIFIED DESCRIPTION OF STATISTICAL METHODS EMPLOYING FIXED AND SEQUENTIAL OBSERVATION PROCESSES

1. INTRODUCTION

1.1 Prefacing Remarks

Signal detection in this report means the detection of certain functions of time (for example, voltages) called "signals" when perturbed by the addition of some other functions called "noise." No attempt will be made to consider methods of estimation of signal parameters or in general to obtain other information about the "signals."

A mathematically detailed report,[1] (hereafter referred to as Technical Report No. 13) has been made on certain statistical approaches to signal detection; that report constitutes a unified description of the subject heretofore unavailable. In addition, a number of specific applications of the resulting theory have been developed (Technical Report No. 13, Part II). However, it is felt that much of that material is inaccessible to all but a few specialists because of its highly technical nature.

Therefore, it seemed appropriate to supplement a report on the applications of sequential analysis to signal detection with a non-technical

---

[1] Peterson, W. W., and Birdsall, T. G., Theory of Signal Detectability, Part I, "The General Theory," Part II, "Applications with Gaussian Noise," Technical Report No. 13, Electronic Defense Group, Department of Electrical Engineering, University of Michigan.

description of the results of Technical Report No. 13. In this way a complete survey of the applications of statistical methods could be given in which the text would be accessible to those with a minimum of mathematical training.

1.2 The Problem of Signal Detection Formulated for Statistical Analysis

Because a receiver is essentially a linear device, noise generated by the receiver can usually be referred to the input. Thus the situation can be represented schematically as a (noiseless) receiver whose inputs are derived by adding the voltages from two sources: a "signal" generator and a "noise" generator. The totality of possible receiver inputs when the "noise" generator alone is in operation will be called "Population N." "Population SN" is the name given to the totality of possible receiver inputs when the "signal" generator and the "noise" generator are in operation simultaneously. The individual observing the receiver outputs is then being presented with a "sample" of one of the two populations, but he is in ignorance as to which population was in fact sampled, and of the probability that any particular one of them was sampled. All he knows with certainty is that one of the two was sampled. He must then judge which population was sampled.

In this discussion it should be kept in mind that the event of population SN being sampled corresponds to signal and noise being present at the receiver input. Also the event of population N being sampled means that noise alone was present at the receiver input.

2. TESTS OF FINITE SAMPLES

2.1 Finite Sampling Plans

This part of the report is concerned with a method of statistical analysis which requires[1] for raw data a finite sample; that is, a finite sequence of numbers $Z_n = (x_1, \ldots, x_n)$. In the present context, such a sample is thought of as the result of n measurements made at the receiver input. The act of making these measurements is supposed to occupy a certain interval I in time, starting at t, of length T. I is called the sample interval. Any particular scheme of making n measurements at the receiver input during the sample interval I is called a sample plan based on I.

If n were very large, a receiver which had to make the measurements called for by a sampling plan would certainly be impractical. However, the theory to be developed here is intended to specify an optimum receiver and is couched in the language of finite samples. This practical difficulty can be avoided if it be required that the sampling plan should "throw away" no information. This would mean that from each sample Z_n it would be possible to reconstruct completely the function of time present at the receiver input during the sample interval. Then the specification of the optimum receiver could be translated back to the language of receiver inputs, from that of samples.

The theory to be described below was developed on the assumption that the populations N and SN are "finite dimensional." This means that they can be constructed from some finite number of functions of time

$$\omega_1(t), \ \omega_2(t), \ \ldots, \ \omega_n(t)$$

[1] The statistical theory itself has been carried out for infinite samples (footnote 2, p. 23), but the application of it to specifying an optimum receiver for a particular case has been carried out so far for finite samples only.

by forming all possible combinations like

$$a_1\ \omega_1(t) + a_2\ \omega_2(t) + \ldots + a_n\ \omega_n(t),$$

where the coefficients $a_1, a_2, \ldots, a_n$, are any chosen numbers. The significance of this restriction will be discussed in Section 4.1, Applicability of Finite Ratio Tests.

For the purposes of the subsequent development, any sampling plan which throws away no information will be considered,[1] provided enough properties are known of the associated sample variable $Z_n = (x_1, x_2, \ldots, x_n)$ so that certain probabilities may be calculated. Specifically, the probability density functions[2] $f_N(Z_n)$ and $f_{SN}(Z_n)$ of the sample variable Z_n for the cases when Z_n is drawn from population N and from SN respectively must be known. The two basic properties of density functions are

$$f_N(Z_n) \geq 0 \qquad \int f_N(Z_n)\, dZ_n = 1$$

and

$$f_{SN}(Z_n) \geq 0 \qquad \int f_{SN}(Z_n)\, dZ_n = 1$$

where the integration symbol represents the multiple integral taken over the entire range of the sample variable Z_n.

A large part of Technical Report No. 13 is devoted to determining some circumstances where the derivation of the density functions can actually be carried out and the optimum receiver specified. These are listed in Section 4.1.

---

[1] In Appendix B it is shown that many such sampling plans are available when the populations are finite dimensional. The idea of a finite sampling plan is a device useful in performing computations for the finite dimensional case, and in approximating the infinite dimensional one. It is not essential to the theory itself.

[2] See Appendix C for a brief discussion of probability density functions, if this term is not familiar.

2.2 The Concept of a Criterion

Consider now an observer who has as available data the sample point $Z_n = (x_1, \ldots, x_n)$ given him by the receiver. The observer's job is to judge for each sample point whether or not it was taken from population SN. Although it is not possible to determine the (probably subconscious) criterion used by the observer, it is quite possible to find an external manifestation of it. Ideally all that is necessary is to submit each possible sample point Z_n to the observer and to record his judgment. This will yield a tabulation of those sample points which the observer decided were drawn from population SN. If any other observer is given this tabulation and instructed to base his decisions on it, he will behave exactly as did the first observer. Thus the tabulation of these responses can be used to replace the mental criterion employed by the observer. Such a tabulation will also be called a criterion and will be denoted by the letter A, which refers to the phraseology common in statistics of "Accepting the hypothesis that a signal is present." The tabulation of the remaining sample points, those which the observer concluded were drawn from population N, will be denoted by B.

2.3 Probabilities Associated with Criteria

There are of course as many different criteria as there are observers. Among all possible criteria it is necessary to select those that are best for various purposes. To do so, certain numerical quantities must be associated with each criterion. It will be necessary to know the probability that a sample from one of the populations will be listed in a particular criterion A. According to the standard definitions, these probabilities are given by

$$P_{SN}(A) = \int f_{SN}(Z_n)\, dZ_n \qquad \text{and}$$

$$P_N(A) = \int f_N(Z_n)\, dZ_n$$

where the multiple integral is taken over all sample points listed in the criterion A.

For example, a particular sample plan might have a density function of the form $f_N(x_1, x_2, \ldots, x_n) = K \exp-(x_1^2 + x_2^2 + \ldots + x_n^2)$. A possible criterion would consist of those sample points $Z_n = (x_1, x_2, \ldots, x_n)$ which lie outside a sphere of radius one centered at the origin. Then the integral would be taken over the exterior of this sphere.

These probabilities have a special significance. $P_N(A)$ is the conditional probability that a sample from population N will be listed in criterion A, that is, will be judged as a sample from population SN. Thus $P_N(A) = F$ is the conditional false alarm probability. Also, $P_{SN}(A)$ is the conditional probability of a certain kind of correct response called a hit (that of judging correctly that a sample is from population SN). The conditional probability of judging falsely that a sample is from population SN is therefore given by $1 - P_{SN}(A) = M$, the conditional probability of a miss. The only errors which can occur are false alarms and misses; their conditional probabilities, F and M, are called briefly the error probabilities.

A reader familiar with the formal content of probability theory should note that these quantities are true conditional probabilities: the first is conditional on the sample being drawn from population SN; the second is conditional on it being drawn from N. This is to distinguish them from a priori probabilities (the probabilities that a certain population will be sampled, for example) which are not as yet assumed known.

2.4 Likelihood Ratio and the Ratio Criteria

It is convenient to introduce a new function called the likelihood ratio, $\mathcal{L}(Z_n)$, defined as the ratio $\frac{f_{SN}(Z_n)}{f_N(Z_n)}$ for sample points $Z_n = (x_1, \ldots, x_n)$.

$\ell(Z_n)$ represents the likelihood that the sample point Z_n was drawn from SN relative to the likelihood it was drawn from N. Hence, if $\ell(Z_n)$ is sufficiently large, it would be reasonable to conclude that Z_n was in fact drawn from population SN, i.e., that Z_n should be listed in the desired "best" criterion. Thus for each number $\beta \geq 0$, a certain criterion $A(\beta)$ will be selected; $A(\beta)$ is chosen by listing each sample point Z_n for which $\ell(Z_n) \geq \beta$. The problem then reduces to that of making a wise choice of β; that is, to determine how large "sufficiently large" is. Criteria of the form $A(\beta)$ will be called ratio criteria.

A number of writers have presented varying definitions of a criterion being "optimum." It turns out that each of these optimum criteria can be expressed as a ratio criterion, so that a receiver designed to yield likelihood ratio as output could be used with any of them.

2.5 Weighted Combination Criteria

Suppose it is possible to assign a certain number β as a weighting factor representing the importance of a false alarm relative to a hit. Since $P_{SN}(A)$ is the probability of a hit and $P_N(A)$ the probability of a false alarm, it would then be reasonable to find a criterion A which maximizes the quantity

$$P_{SN}(A) - \beta P_N(A) \quad .$$

But this quantity can be written as

$$\int \left[f_{SN}(Z_n) - \beta f_N(Z_n) \right] \, dZ_n \quad ,$$

where the integration is taken over the sample points Z_n listed in A. To maximize this integral, one would list in A every sample point Z_n for which the integrand was not negative. Solving that inequality for β, one sees that A should contain those sample points Z_n for which

$$\ell(Z_n) = \frac{f_{SN}(Z_n)}{f_N(Z_n)} \geq \beta \quad .$$

Thus the desired criterion A is simply $A(\beta)$, and so it is a ratio criterion.

2.6 Neyman-Pearson Criteria

If it is critically important to keep the probability of a false alarm $P_N(A)$ below a certain level k, then it would be reasonable to choose, from among such criteria, that one which maximizes the probability of a hit. Thus Neyman and Pearson proposed as a type of optimum criterion any criterion A_k for which

(1) $P_N(A_k) \leq k$, and

(2) $P_{SN}(A_k)$ is a maximum for all the criteria A with the property $P_N(A) \leq k$.

The A_k type criterion can also be expressed as a ratio criterion. This can be made plausible as follows. To begin with, it is necessary to consider only those criteria A for which $P_N(A) = k$, because A will be taken as large as possible in order to meet condition (2). Now consider the curve given parametrically by the equations

$$x = x(\beta) = P_N\left(A(\beta)\right) \quad \text{and}$$

$$y = y(\beta) = P_{SN}\left(A(\beta)\right) \quad .$$

This curve will be called the Receiver Operating Characteristic (briefly, ROC) curve, for a receiver whose output is likelihood ratio and with which ratio criteria are being used.

The ROC curve passes through the points (0, 0) and (1, 1), the first at $\beta = \infty$, the second at $\beta = 0$. At $\beta = 0$, $\ell(Z_n) \geq \beta = 0$ for all Z_n, so A(0)

consists of all possible sample points. Thus the observer will report that every sample is drawn from SN, so he will be certain to make a false alarm and to make a hit. (This assumes that the sample points will not be drawn exclusively from one of the populations.) This can be verified, using the basic property of the density functions expressed by the following equations:

$$P_{SN}\left(A(0)\right) = \int f_{SN}(Z_n)\, dZ_n = 1 \quad \text{and}$$

$$P_{N}\left(A(0)\right) = \int f_{N}(Z_n)\, dZ_n = 1 \quad ,$$

when the integration is taken over all possible sample points Z_n. These equations mean that $x(0) = y(0) = 1$. Moreover, $x(\infty) = y(\infty) = 0$, because for $\beta = \infty$ there are no sample points Z_n with $L(Z_n) \geq \infty$; i.e., $A(\infty)$ contains no sample points at all and the operator will never report a signal is present. Therefore the operator cannot possibly make a false alarm nor can he make a hit. Thus $P_{SN}\left(A(\infty)\right) = 0$ and $P_N\left(A(\infty)\right) = 0$.

These considerations, together with those of the next section, show that the ROC curve can be sketched somewhat as in Fig. 1.

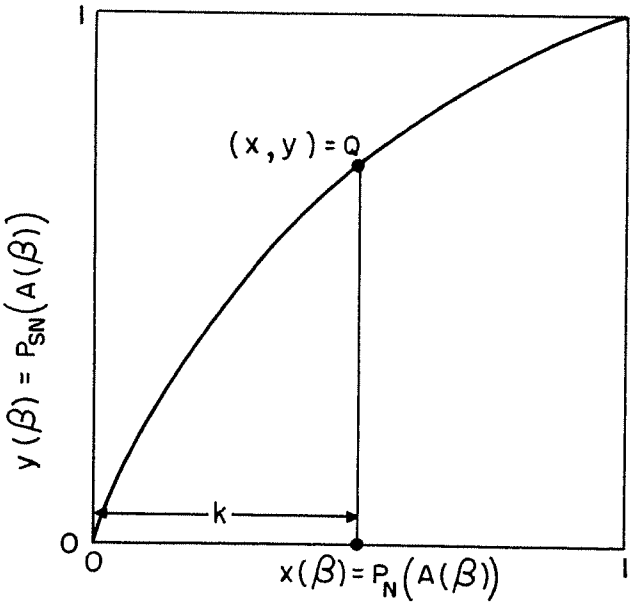

FIG. 1. TYPICAL ROC CURVE.

To determine the desired A_k, recall that all probabilities lie between zero and one, so that $P_N(A_k) = k$ is between zero and one. Then there is a point Q of the ROC curve which lies vertically above the point (k, 0). The coordinates (x, y) of Q are $x = P_N\left(A(\beta)\right) = k$ and $y = P_{SN}\left(A(\beta)\right)$, for some β, which will be written β_k. Now $A(\beta_k)$ is a possible candidate for A_k since $P_N\left(A(\beta_k)\right) = k$. Let A be any criterion with $P_N(A) = k$; it will be shown that $P_{SN}(A) \leq P_{SN}\left(A(\beta_k)\right)$, so that $A(\beta_k)$ meets the requirements of the Neyman-Pearson criterion.

From the discussion of the weighted combination criterion, it is clear that $T = P_{SN}\left(A(\beta_k)\right) - \beta_k \cdot P_N\left(A(\beta_k)\right) \geq P_{SN}(A) - \beta_k \cdot P_N(A) = T^*$. Thus

$$P_{SN}\left(A(\beta_k)\right) = T + k \quad \text{and}$$

$$P_{SN}(A) = T^* + k \ .$$

The known inequality between the T's yields the desired inequality by subtracting the last equation from the one above it.

Therefore, A_k should be chosen to be this particular $A(\beta_k)$; whence the optimum criterion proposed by Neyman and Pearson reduces to a ratio criterion.

2.7 ROC Curve

It will be desirable to digress for a moment to study the ROC curve more closely. Its value lies in the fact that if the type of criterion chosen for a particular application is a ratio criterion, $A(\beta)$, then a complete description of the detection system's performance can be read off the ROC curve. By the very definition of the ROC curve, the x coordinate is the conditional probability, F, of false alarm, and the y coordinate is the conditional probability of a hit. Similarly (1-x) is the conditional probability of being correct when noise alone is present, and (1-y) = M is the conditional probability of a

miss. Since most proposed kinds of criteria can be reduced to ratio criteria, the ROC curve assumes considerable importance.

In order to determine some of its geometric properties, it will be assumed that the parametric functions

$$x = x(\beta) = P_N\big(A(\beta)\big) \quad \text{and}$$

$$y = y(\beta) = P_{SN}\big(A(\beta)\big)$$

are differentiable functions of β. The slope m of the tangent to the ROC curve is given by the quotient $\dfrac{\left(\frac{dy}{d\beta}\right)}{\left(\frac{dx}{d\beta}\right)}$. To calculate the slope at the point $\big(x(\beta_o), y(\beta_o)\big)$, notice that among all criteria A, the quantity $P_{SN}(A) - \beta_o \cdot P_N(A)$ is maximized by $A = A(\beta_o)$. Therefore, in particular, the function

$$y(\beta) - \beta_o \cdot x(\beta) = P_{SN}\big(A(\beta)\big) - \beta_o \cdot P_N\big(A(\beta)\big)$$

has a maximum at $\beta = \beta_o$, so that its derivative must vanish there. Thus differentiating,

$$\frac{dy}{d\beta} - \beta_o \cdot \frac{dx}{d\beta} = 0 \quad \text{at} \quad \beta = \beta_o \ .$$

Solving for β_o, one obtains

$$\beta_o = \frac{\left(\frac{dy}{d\beta}\right)_{\beta = \beta_o}}{\left(\frac{dx}{d\beta}\right)_{\beta = \beta_o}} = m \ .$$

This shows that the slope of the ROC curve is given by its parameter β, and so is always positive. Hence the curve rises steadily. In addition, this means that $y(\beta)$ can be written as a single valued function of $x(\beta)$, $y = y(x)$, which is monotone increasing, and where $y(0) = 0$ and $y(1) = 1$. These remarks make fully warranted the sketch of the ROC curve given in Fig. 1.

2.8 Siegert's "Ideal Observer's" Criteria

Here it is necessary to know beforehand the a priori probabilities that population SN and that population N will be sampled. This is an additional assumption. These probabilities are denoted respectively by P(SN) and P(N). Moreover, P(SN) + P(N) = 1 because at least one of the populations must be sampled. The criterion associated with Siegert's Ideal Observer is usually defined as a criterion for which the a priori probability of error is minimized, (or, equivalently, the a priori probability of a correct response is maximized). Frequently the only case considered is that where P(SN) and P(N) are equal, but this restriction is not necessary.

Since the conditional probability F of a false alarm is known as well as the (a priori) probability of the event (that population N was sampled) upon which F is conditional, then the probability of a false alarm is given by the product

$$P(N)F \ .$$

In the same way the probability of a miss is given by

$$P(SN)M \ .$$

Because an error E can occur in exactly these two ways, the probability of error is the sum of these quantities

$$P(E) = P(N)F + P(SN)M \ .$$

It has already been pointed out that $F = P_N(A)$ and $M = 1 - P_{SN}(A)$. If these are substituted into the expression for P(E) a simple algebraic manipulation gives

$$P(E) = P(SN) - P(SN)\left[P_{SN}(A) - \frac{P(N)}{P(SN)} \cdot P_N(A)\right] \ .$$

It is desired to minimize P(E). But from the last equation this is equivalent to maximizing the quantity

$$P_{SN}(A) - \frac{P(N)}{P(SN)} \cdot P_N(A) \quad .$$

And, of course, this will yield a weighted combination criterion with $\beta = \frac{P(N)}{P(SN)}$, which is known to be simply $A(\beta)$, a ratio criterion.

2.9 The Finite Ratio Test

Once populations N and SN have been chosen, "a finite test" of these populations means a particular choice of finite sampling method and of criterion, where the requirements made by Section 2.1 are met. If the criteria are restricted to ratio criteria, then a finite test is determined by the choice of the following parameters:

- t Starting time of sample interval
- T Length of sample interval
- β Parameter of the ROC curve, from which the two conditional error probabilities and the two correct response conditional probabilities can be read off.

Such a test will be called a finite ratio test. Note especially that the ROC curve is independent of the particular sampling plan chosen.

3. SEQUENTIAL TESTS

3.1 Infinite Samples

Among the various methods of statistical analysis which have been developed, some are designed to make use of infinite samples. This does not mean that infinitely many measurements must be made in an actual application; it

involves only the theoretical possibility of doing so. If such a theoretical eventuality is allowed, one finds that in actual applications only a finite number of samples are ever needed. In fact this number may even be smaller than the number needed by a comparable standard test. These remarks will be amplified later; at the moment they should suffice to justify consideration of infinite samples.

A plan for taking an infinite sample does not necessarily entail an infinitely long interval of time. The "time base" of such a plan can be infinite--for example, by having one measurement made every second. On the other hand, a plan could call for making one measurement at each of the instants

$$t_n = 1 - \frac{1}{n} \quad , \quad n = 1, 2, \ldots \quad .$$

These instants all lie in the time interval from zero to one, and thus such a plan would involve only one unit of time at most.

Only those sampling plans for which certain statistical information is known can be used in a test. If the sampling plan has been carried out to the point where n measurements, $(x_1, x_2, \ldots, x_n)$ have been made, the variable $Z_n = (x_1, x_2, \ldots, x_n)$ is called an "n-th stage sample variable." For each stage n, the two density functions $f_N(Z_n)$ and $f_{SN}(Z_n)$ of the n-th stage sample variable Z_n must be known, where the first is the density function applicable when population N is being sampled and the second applies when population SN is sampled. The density functions at different stages may very well differ, so that actually they should be written $f_N^n(Z_n)$, and $f_{SN}^n(Z_n)$. However, the n appearing in the argument Z_n should always make the situation clear, so that the superscript n on the functions themselves will be dropped.

3.2 Sequential Tests

A sequential test will consist of two things:

1) An (infinite) sampling plan with density functions $f_N(Z_n)$ and $f_{SN}(Z_n)$
2) An assignment of certain criteria to each stage of the sampling plan.

The idea of a sequential test is as follows. First, make one measurement, x_1; if the evidence x_1 is sufficiently persuading, draw a conclusion as to whether or not a signal is present. If the evidence is not so strong, make a second measurement x_2. Then, considering the evidence (x_1, x_2), repeat the above process, and continue in a similar manner.

A particular scheme for making these decisions consists of the assignment of three criteria to each stage of the sampling process. The three criteria represent the three possible conclusions:

1) A signal is present, or
2) A signal is not present, or
3) Another measurement should be made.

At the first stage, any (real) number at all could theoretically result from the first measurement. This means that the first stage sample variable $Z_1 = (x_1)$ ranges through the entire number system, which will be written S_1 to stand for the first stage sample space. Suppose the three first-stage criteria A_1, B_1, and C_1, have been chosen. If the sample Z_1 is listed in A_1, the conclusion that a signal is present is drawn and the test terminated. If it is listed in B_1, the conclusion is that noise alone is present, and again the test is terminated. If Z_1 should be listed in C_1, another measurement will be made, and the test moves on to the second stage instead of terminating.

When the first stage criteria have been chosen, a limitation is placed on S_2, the space through which the second stage sample variable $Z_2 = (x_1, x_2)$ ranges. The only way the test can proceed to the second stage is for $Z_1 = (x_1)$ to be listed in C_1. Therefore, S_2 does not contain all possible second stage samples $Z_2 = (x_1, x_2)$ but only those for which (x_1) is listed in C_1. Three second-stage criteria, A_2, B_2, and C_2, must now be chosen from those samples Z_2 listed in S_2. They must be chosen in such a way that there are no duplications in the listings and no sample in S_2 is omitted. These criteria carry exactly the same significance as those chosen in the first stage. That is, the three conclusions that a signal is or is not present, or that the test should be continued, are drawn when the sample Z_2 is listed in A_2, B_2, or C_2 respectively.

The selection of criteria proceeds in the same way. If n-th stage criteria A_n, B_n, and C_n have been chosen, then the next stage's sample space S_{n+1} consists of those samples $Z_{n+1} = (x_1, x_2, \ldots, x_n, x_{n+1})$ for which $Z_n = (x_1, x_2, \ldots, x_n)$ was listed in C_n. Then from S_{n+1} are drawn the three (n+1) stage criteria A_{n+1}, B_{n+1}, and C_{n+1}.

When an entire sequence

$$(A_1, B_1, C_1) \ ,$$
$$(A_2, B_2, C_2) \ ,$$
$$\vdots$$
$$(A_n, B_n, C_n) \ ,$$
$$\vdots$$

of criteria is selected, a "sequential test" has been determined. This does not mean of course that the test will necessarily be particularly useful. However,

among all the possible ways of selecting a sequence of criteria and hence a sequential test, there may be particular ones which are very useful.

3.3 Probabilities Associated with a Sequential Test

If Q_n is any n-th stage criterion, then the quantities[1]

$$P_N(Q_n) = \int_{Q_n} f_N(Z_n)\, dZ_n \qquad \text{and}$$

$$P_{SN}(Q_n) = \int_{Q_n} f_{SN}(Z_n)\, dZ_n$$

represent the (N or SN) conditional probabilities that an n-th stage sample Z_n will be listed in the criterion Q_n. Some examples of the use of this notation are:

1) The n-th stage conditional error probabilities:

If population N is sampled, then the probability that the sample variable Z_n will be listed in A_n is $P_N(A_n)$. This is the N-conditional probability of a false alarm.

If population SN is sampled, then the probability that the sample variable Z_n will be listed in B_n is $P_{SN}(B_n)$. This is the SN-conditional probability of a miss.

2) The conditional error probabilities of the entire test:

$$F = \sum_{n=1}^{\infty} P_N(A_n)$$ the N-conditional probability of a false alarm, and

---

[1]The notation $\int_Q$ indicates that the integration is to be carried out over all sample points listed in Q.

$$M = \sum_{n=1}^{\infty} P_{SN}(B_n)$$ the conditional probability of a miss,

are merely the sums of the same error probabilities over all stages.

3) The conditional probabilities of terminating at stage n are

$$T_N^n = P_N(A_n) + P_N(B_n)$$

$$T_{SN}^n = P_{SN}(A_n) + P_{SN}(B_n) \quad .$$

These formulas can be justified by a simple argument. The only ways the test can terminate at stage n is for the sample variable Z_n to be listed in either A_n or B_n. The probability of this event is the sum of the probabilities of the component events which are mutually exclusive since Z_n can be listed in at most one of A_n and B_n.

4) The conditional probabilities that the entire test will terminate are

$$T_N = \sum_{n=1}^{\infty} T_N^n$$

$$T_{SN} = \sum_{n=1}^{\infty} T_{SN}^n \quad .$$

3.4 Average Sample Numbers

There are two other quantities which must be introduced. One feature of the sequential test is that it affords an opportunity of arriving at a decision early in the sampling process when the data happens to be unusually convincing. Thus one might expect that, on the average, the stage of termination of a well-constructed sequential test would be lower than could be achieved by an otherwise equal, good standard test. It is therefore important to obtain expressions

for the average or expected value of the stage of termination. As with other probabilities, there will be two of these quantities: one conditional on population N being sampled; the other conditional on population SN being sampled. They are given by

$$E_N = \sum_{n=1}^{\infty} n\, T_N^n \quad \text{and}$$

$$E_{SN} = \sum_{n=1}^{\infty} n\, T_{SN}^n \quad .$$

The letter E is used to refer to the term "expected value." The quantities E_N and E_{SN} are called the average sample numbers. The form these formulas take can be justified (somewhat freely) on the grounds that each value, n, which the variable "stage of termination" may take on must be weighted by the (conditional) probability that the variable will in fact take on that value.

It should be heavily emphasized that the average sample numbers are strictly average figures. In actual runs of a sequential test, the stages of termination will sometimes be less than the average sample numbers but will also be upon occasion much larger. Any sequential test whose average sample numbers are not finite would be useless for applications. Therefore the only ones to be considered are those with finite average sample numbers. Under this assumption, it can be shown[1] that $T_N = T_{SN} = 1$, so that the test is certain to terminate (in the sense of probability). On the other hand, if it is known that $T_N = T_{SN} = 1$, it does not always follow that the average sample numbers are finite. Such a situation would mean only that if a sequence of runs of the test

[1] See Appendix A. This particular result should be intuitively evident.

were made, each run would probably terminate, but the average stage of termination would become arbitrarily large as more runs were made.

3.5. Sequential Ratio Tests

In studying tests using finite samples it was found that the best criterion could always be expressed in terms of likelihood ratio. Therefore, it may be useful to introduce likelihood ratios at each stage of an infinite sample plan. The n-th stage likelihood ratio function $\ell(Z_n)$ is defined as the ratio $\frac{f_{SN}(Z_n)}{f_N(Z_n)}$. Optimum criteria in the finite sample tests turned out to be criteria listing all samples Z for which $\ell(Z)$ is greater than or equal to a certain number. It should be possible to choose sequential criteria (A_n, B_n, C_n) in the same way. For each stage two numbers a_n and b_n with $b_n \leq a_n$ could be chosen. Then the criteria (A_n, B_n, C_n) determined by the numbers a_n and b_n would be

A_n lists all samples Z_n of the sample space S_n for which $\ell(Z_n) \geq a_n$

B_n lists all samples Z_n of the sample space S_n for which $\ell(Z_n) \leq b_n$

C_n lists all samples Z_n of the sample space S_n for which $b_n < \ell(Z_n) < a_n$.

If criteria selected in this way meet the requirements that the average sample numbers be finite, then the resulting sequential test is called a "sequential ratio test."

3.6 Optimum Sequential Tests

Because the task of computing the various parameters (error probabilities and average sample numbers) of a sequential test is considerably more difficult than the corresponding task for the standard test, certain simplifications have been introduced. For example, each systematic study of sequential

tests has been restricted to those of the ratio type introduced in the last paragraph.

For these tests there are two ways of defining an optimum which would probably occur to one immediately. The first would say that among all ratio tests with conditional false alarm probability F, that one for which M, E_N, and E_{SN} are minimum will be called optimum. The complexities of such an extremum problem are enormous and there are no answers known as yet. The second natural possibility is to try to find, among all ratio tests with fixed error probabilities F and M, that one for which the average sample numbers E_N and E_{SN} are minimum. This is the usual sense in which the word optimum is used concerning sequential tests.

Wald has proposed a particular test as an optimum ratio test, which will will be known as the Wald test in this report. A ratio test is a Wald test if each of the sequences $\{b_n\}$ and $\{a_n\}$ are constant, that is, if $b_1 = b_n$ and $a_1 = a_n$ for all n. Moreover, Wald[1] proved under very restrictive conditions that his test is optimum. Unfortunately, his conditions are never satisfied in the case of applications to signal detectability, as is shown in Section A.4 of Appendix A. However, the absence of theoretical knowledge concerning the optimum nature of the Wald test should not be construed to ban the use of the test, but merely to temper its use with caution.

No examples of ROC curves are given for the Wald test in various cases because of the heavy computational difficulties involved. Numerical comparison of the Wald test with a finite ratio test is given in the next section.

---

[1] A. Wald, Sequential Analysis, John Wiley and Sons, 1947.

4. CONCLUSIONS

4.1 Applicability of Finite Ratio Tests

From a theory of signal detection one would hope to obtain two basic results:

1) The ROC curve, i.e., performance of an optimum receiver, and

2) Specification of an optimum receiver.

When population N is taken as finite dimensional with a white Gaussian density function, actual specification of an optimum receiver has been carried out[1] for certain particular SN populations. These cases are tabulated as follows. In the table S denotes the signal population before being perturbed by the noise.

TABLE I

S	Application
Signal Known Exactly	Coherent radar with a target of known range and character
Signal Known Except for Phase	Ordinary pulse radar with no integration and with a target of known range and character.
Signal a Sample of White Gaussian Noise	Detection of noise-like signals; detection of speech sounds in Gaussian noise.
Output of the Detector of a Broad Band Receiver	Detecting a pulse of known starting time (such as a pulse from a radar beacon) with a crystal-video or other type broad band receiver.
A Radar Case (A train of pulses with incoherent phase)	Ordinary pulse radar with integration and with a target of known range and character.

[1] Technical Report No. 13, Part II.

TABLE I (cont.)

S	Application
Signal One of M Orthogonal Signals	Coherent radar where the target is at one of a finite number of non-overlapping positions.
Signal One of M Orthogonal Signals Known Except for Phase	Ordinary pulse radar with no integration and with a target which may appear at one of a finite number of non-overlapping positions.

In all these cases, either population SN is finite dimensional, or a special method is used to reduce the problem to an equivalent finite dimensional one. Once such a reduction is achieved, a sampling plan which throws away no information can be found,[1] and the solution of the problem then consists of deriving an expression for likelihood ratio and specifying a receiver whose output will be that likelihood ratio.

However, this restriction to finite dimensionality is not at all essential. The theory concerning the existence of an optimum criterion depends only on the presence of a function to play the part of the likelihood ratio function.[2] For the purposes of initial investigation and of exposition the restriction of finite dimensionality is very convenient, for then the calculations necessary can be formulated in terms of carefully chosen sampling plans, and the expression for likelihood ratio takes a closed form. With likelihood ratio in a closed form it is not difficult to specify the optimum receiver (i.e., the receiver which has

[1] See Appendix B.

[2] Grenander, U., "Stochastic Processes and Statistical Inference," Arkiv För Matematik, Vol. 2, 195 (1950).

likelihood ratio as its output) in certain cases such as those tabulated above. Moreover, when the general form of the theory is used, actual calculations would be carried out by using finite dimensional approximations. It appears that the results already obtained concerning the optimum receiver will not be changed materially when the more general theory is used. Although to date actual ROC curves and optimum receivers have not been determined for cases susceptible to only the general theory, there is no essential obstacle to doing so.

In the absence of experimental verification of the accuracy of the ROC curve in predicting the performance of the optimum receiver, there is one remaining fact which could be interpreted as casting doubt on the reliability of the theory so far developed. Under the restrictions 1) that populations N and SN are finite dimensional and 2) that the functions of time in these populations be (real) analytic, it is possible to prove[1] that sampling plans utilizing arbitrarily small sample intervals can be found, all of which yield the same error probabilities or ROC curves.

One way to explain this anomaly is as follows:

There can be little doubt that observations restricted successively to arbitrarily small intervals cannot be equally effective in detecting a signal. At the same time there can be little doubt that extremely precise measurements cannot be made of arbitrarily small intervals. It is not at all uncommon that the assumption that errorless measurements are possible should lead to physically ridiculous conclusions. The apparent anomaly cited above can certainly be thought of as a case in point.

---

[1]See Appendix B.

4.2 Applicability of Sequential Ratio Tests

The current status of the theory of sequential ratio tests is marred by two essential defects so far as its possible applications to signal detection are concerned.

1) The Wald test is known to be optimum only relative to a very restricted class of sequential ratio tests.[1]

2) Even if the Wald test were known to be optimum in general, conditions under which its average sample numbers are finite remain unknown.

However, there are some strong reasons to believe that sequential ratio tests would be very useful if the points cited above were cleared up. The first is the point made in Section 4.1 concerning the desirability of having a practical theory which is not restricted to finite dimensional populations. Sequential analysis might be the needed key for such a theory formulated in terms of infinite sampling plans. Moreover, whether or not the Wald test is optimum, there are many instances where it compares very favorably with the finite ratio test.

For example, suppose that both populations N and SN are finite dimensional with white Gaussian density functions. In this event, successive measurements of the amplitude of the receiver input will be independent and Wald's approximation formulas for the average sample numbers of the Wald test can be used. First a particular sample interval I was chosen. Then a large number W was selected, and the functions of time present in the two populations were determined by taking all such functions which have a Fourier expansion on the interval I and deleting all terms in the expansion of frequency greater than W. This meant that the two populations were of dimension $n = 2WT$, where T is the

[1] See Appendix A for a technical discussion of this matter.

length of the sample interval. In this case the sampling plan, which consists of making n measurements equally spaced in the interval I,[1] will have the property that it throws away no information.

In order to secure as fair a comparison as possible, an infinite sampling plan was chosen for the Wald test which involved making measurements at evenly spaced intervals of length $\frac{T}{n+1}$. Thus for the first n measurements this sampling plan coincides with the finite sampling plan chosen above. A particular point (F, M) was chosen on the ROC curve of the finite ratio tests, and the average sample numbers of the Wald test whose error probabilities are (F, M) were calculated. These calculations were performed for various ratios of signal energy to noise energy and in all cases the average sample numbers came out appreciably less than the dimension n = 2 W t of the finite ratio test. Thus in this case the Wald test would terminate on the average before the entire sample interval for the finite ratio test had elapsed. The quantitative results are tabulated below.

TABLE II

Power Ratio S/N	Average Sample Numbers E_{SN}	E_N	Dimension of the Finite Ratio Test
.368	80	15	100
.9804	828	195	1,000
.02911	8754	2015	10,000
.00902	83221	20154	100,000

[1] The spacing would be $\frac{T}{n+1}$.

Although little use can be made at present of sequential analysis in signal detection, it appears that if all possibilities of obtaining the practical theory desired without the finite dimensional restriction are to be explored, then the gaps mentioned in the theory of sequential analysis should certainly receive more attention in the future.

APPENDIX A

THE MATHEMATICAL THEORY OF SEQUENTIAL TESTS

A.1 Introduction

The discussion of sequential tests given in the body of this report is somewhat novel compared to the current literature of the subject. The novelty stems primarily from a special orientation and notation. Previous work on sequential tests has been done with the chief emphasis on its application to the case of finite populations where the distribution of the n-th measurement is the same as that of the first measurement, i.e., where successive measurements are independent. Finite populations have been of special interest because of the use of sequential tests in quality control, where the assumption of independence is rarely a significant restriction. Moreover, this assumption made it possible to establish a number of formulas which are of great value in computing the varicus parameters of a sequential test. However, in the field of signal detection it is easy to find quite simple cases where successive measurements are not independent. Therefore, the use of sequential tests in this direction will depend on the extension of the general theory in the absence of the hypothesis of independence. The material of this appendix has been written with the purpose of outlining the kind of theory needed and to point out certain theoretical questions which will have to be answered before sequential analysis can be applied to signal detection. As a result, the orientation of this discussion differs from that of Wald,[1] for example;

[1] Throughout the appendix, the source for references to Wald's work is A. Wald, Sequential Analysis, John Wiley and Sons, 1947.

this change in orientation also requires a new notation, which has already been introduced above.

A.2 Sequential Tests

A sequential test is a particular combination of two basic kinds of mathematical objects, which will be called hypotheses and criteria. Let E^n denote the n-dimensional Euclidean space and μ Lebesgue measure on E^n.

A hypothesis is a family $\left\{ f_n : E^n \to E^1 \right\}$, $n = 1, 2, \ldots$, of non-negative functions subject to the conditions that, for each n,

I. $\int_{E^n} f_n \, d\mu = 1$ and

II. If A is any set in E^n for which $\int_A f_n \, d\mu$ exists, then

$$\int_A f_n \, d\mu = \int_{A \times E^1} f_{n+1} \, d\mu \ .$$

II is called the "cylindrical" property, because $A \times E^1$ can be thought of as a cylinder erected on the base A. Note that if g is a real function of one variable such that $\int_{E^1} g \, d\mu = 1$, then a hypothesis may be constructed from g by defining $f_n(x_1, x_2, \ldots, x_n)$ to be the product $\prod_{i=1}^{n} g(x_i)$. Such a hypothesis is called independent.

A criterion is a collection $\left\{ A_n, B_n, C_n \right\}$, $n = 1, 2, \ldots$, of sets subject to the conditions

III. A_n, B_n, and C_n are pairwise disjoint.

IV. $A_1 \cup B_1 \cup C_1 = E^1$ and $A_n \cup B_n \cup C_n = C_{n-1} \times E^1$ if $n > 1$.

Finally, a sequential test (of two hypotheses) consists of two hypotheses $\left\{ f_n : E^n \to E^1 \right\}$ and $\left\{ g_n : E^n \to E^1 \right\}$ together with a criterion $\left\{ A_n, B_n, C_n \right\}$

for which

V. The integrals of f_n and g_n over the sets A_n, B_n, and C_n exist.

VI. $\sum \int_{C_n} fn \, d\mu$ and $\sum \int_{C_n} g_n \, d\mu$ converge.

The chief notational difference here from that used by Wald is in the criterion. Wald supposes that, for each n, E^n has been partitioned into three mutually disjoint sets R_1^n, R_2^n, and R_3^n. Then he distinguishes between "effective" and "ineffective" samples (i.e., points of E^n) but does not assign a symbol to the "effective" samples, that is, the set $C_{n-1} \times E^1$. Because it will be necessary to compute probabilities that "effective" samples belong to, say, R_1^n, it is desirable to have a symbol for such a set. In the notation of this report, for example, $A_n = R_1^n \cap (C_{n-1} \times E^1)$.

It will be shown in a moment that the quantities appearing in VI are merely the average sample numbers diminished by one. Wald employs as an axiom the condition that the two conditional probabilities of terminating be unity, and shows that if the hypotheses are independent, then this axiom holds. However, it is doubtful that sequential tests for which the average sample numbers are infinite will ever be of real interest. Moreover, VI implies that Wald's axiom holds, and that the conditional error probabilities converge. Because VI is actually stronger than both these conditions, VI appears to be a natural and useful axiom.

Associated with the hypotheses are families $\{F_n\}$ and $\{G_n\}$ of measures defined by

$$F_n(Q) = \int_Q f_n \, d\mu \qquad \text{and}$$

$$G_n(Q) = \int_Q g_n \, d\mu \; .$$

$F_n(Q)$ is interpreted as the conditional probability that a point x of E^n is a point of Q, where the "condition" is that hypothesis $\left\{f_n: E^n \rightarrow E^1\right\}$ actually does describe the statistical properties of the sample x. A similar interpretation is made for $G_n(Q)$.

Lemma A1

$$E^1 = A_1 \cup B_1 \cup C_1 \qquad \text{and}$$

$$E^{n+1} = \left[A_{n+1} \cup B_{n+1} \cup C_{n+1}\right] \cup \left[\bigcup_{i=1}^{n} (A_i \cup B_i) \times E^{n+1-i}\right]$$

Proof: The first statement is merely IV. The second can be proved by induction. Let Q_n represent the right hand side of the second equality.

Then

$$Q_1 = \left(C_1 \times E^1\right) \cup \left(A_1 \cup B_1 \times E^1\right) \qquad \text{by IV}$$

Factoring E^1,

$$Q_1 = (C_1 \cup A_1 \cup B_1) \times E^1 = E^1 \times E^1 = E^2.$$

Thus the lemma is true when n = 1. Suppose the lemma is true for some particular number; i.e., suppose that $E^n = Q_{n-1}$. This will now be shown to imply that $E^{n+1} = Q_n$, which will complete the inductive argument. Using the inductive hypothesis, one obtains

$$\bigcup_{i=1}^{n} (A_i \cup B_i) \times E^{n+1-i} = E^1 \times \left[A_n \cup B_n \cup \left(\bigcup_{i=1}^{n} (A_i \cup B_i) \times E^{n-i}\right)\right]$$

$$= E^1 \times (Q_{n-1} - C_n)$$

$$= E^1 \times (E^n - C_n)$$

Therefore,

$$Q_n = \left[C_n \times E^1 \right] \cup \left[E^1 \times (E^n - C^n) \right]$$

$$= (C_n \times E^1) \cup (E^{n+1}) - E^1 \times C^n = E^{n+1}$$

which was to be proved.

An immediate corollary of this lemma and II is

Theorem A1

$$\sum_{i=1}^{n} \left(F_i(A_i) + F_i(B_i) \right) + F_n(C_n) = 1$$

$$\sum_{i=1}^{n} \left(G_i(A_i) + G_i(B_i) \right) + G_n(C_n) = 1$$

At this point it should be noted that $F_i(A_i) + F_i(B_i)$ and $G_i(A_i) + G_i(B_i)$ are the (conditional) probabilities that a sample $(x_1, x_2, \ldots, x_i)$ be a point of $A_i \cup B_i$, which is equivalent to the assertion that the test will terminate at exactly the i-th stage. Thus

Theorem A2

$$\sum_{i=1}^{\infty} \left(F_i(A_i) + F_i(B_i) \right) = \sum_{i=1}^{\infty} \left(G_i(A_i) + G_i(B_i) \right) = 1$$

means that the (conditional) probabilities of termination are unity. This theorem is proved by applying Theorem A1 and the fact that $\lim_{i \to \infty} F_i(C_i) = \lim_{i \to \infty} G_i(C_i) = 0$, which is a necessary condition that VI hold.

If the hypotheses $\{f_n\}$ and $\{g_n\}$ are denoted by H_o and H_1 respectively, then the quantities

$$\alpha = \sum_{i=1}^{\infty} F_i(A_i) \quad \text{and}$$

$$\beta = \sum_{i=1}^{\infty} G_i(B_i)$$

are the conditional probabilities of a type I or type II error[1] respectively.

Let r be the (random) variable denoting the stage of termination of the test. Then the conditional expected values of r are

$$E_1(r) = \sum_{n=1}^{\infty} n\, F_n\,(A_n \cup B_n) \qquad \text{and}$$

$$E_o(r) = \sum_{n=1}^{\infty} n\, G_n\,(A_n \cup B_n)$$

The question of convergence of these series is settled by the following evaluation.

Theorem A3

$$E_1(r) = 1 + \sum_{i=1}^{\infty} F_i(C_i)$$

$$E_o(r) = 1 + \sum_{i=1}^{\infty} G_i(C_i)$$

Proof: Since $C_n \times E^1 = A_{n+1} \cup B_{n+1} \cup C_{n+1}$, it is equally true that $(C_n \times E^1) - C_{n+1} = A_{n+1} \cup B_{n+1}$. Therefore $E_1(r) = F_1(A_1 \cup B_1) + \sum_{n=2}^{\infty} n\Big(F_n\,(C_{n-1} \times E^1) - F_n(C_n)\Big)$. Using the facts that $1 - F_1(C_1) = F_1(E^1) - F_1(C_1) = F_1(A_1 \cup B_1)$ and that $F_n(C_{n-1} \times E^1) = F_{n-1}\,(C_{n-1})$, (i.e., II), one obtains $E_1(r) = 1 - F_1(C_1) + \sum_{n=2}^{\infty} n\Big(F_{n-1}\,(C_{n-1}) - F_n(C_n)\Big)$. This series collapses and yields the desired result. A similar argument works for $E_o(r)$.

[1] This is the notation used by Wald; type I is a false alarm, type II is a miss.

Thus the condition VI is seen to mean that the average sample numbers must be finite. Moreover Theorem A3 represents the average sample numbers in a way that greatly facilitates comparison of average sample numbers yielded by two different criteria by comparing the sets $\{C_n\}$. It is possible that this particular representation will eventually lead to a general proof of the optimum character of the Wald test.[1]

A.3 Sequential Ratio Tests

Now that the basic properties of a sequential test have been explored, it is time to consider the problem of selecting a useful criterion for given hypotheses H_1 and H_o. Bolstered by the success of the likelihood ratio as a criterion-selecting device in the finite ratio test, one hopes it would be equally efficacious here. The likelihood ratio is usually defined as the ratio $f_n(x)/g_n(x)$. That is, at each point x^o in E^n at which the limit $\lim_{x \to x^o} f_n(x)/g_n(x) = T(x^o)$ exists, one writes $\ell(x^o) = T(x^o)$. Let S_n denote the set of all such points in E^n; S_n is called the ratio sample space. It is, of course, the domain of definition of ℓ_n; usually S_n will be all of E^n.

One would expect to construct a ratio criterion as follows from two sequences $L = \{b_n\}$ and $R = \{a_n\}$ with $0 \leq b_n \leq a_n$. Let $I_n = \{x \mid b_n < x < a_n\}$ and further let R_n and L_n be the rest of E^1 to the right and left respectively of I_n.[2]

---

[1] See below, Section A.4.

[2] The notation $\ell_n^{-1}(Q)$ denotes all points x in S_n for which $\ell_n(x)$ is in Q.

1) Let $A_1 = \ell_1^{-1}(R_1)$, $B_1 = \ell_1^{-1}(L_1)$, and $C_1 = \ell_1^{-1}(I_1)$.

2) If A_n, B_n, and C_n have been defined, let $A_{n+1} = \left(\ell_{n+1}^{-1}(R_{n+1})\right) \cap (C_n \times E^1)$, $B_{n+1} = \left(\ell_{n+1}^{-1}(L_{n+1})\right) \cap (C_n \times E^1)$, and $C_{n+1} = (C_n \times E^1) - A_{n+1} - B_{n+1}$.

The resulting sequence of sets $\{A_n, B_n, C_n\}$ is a criterion provided that $S_n \supset C_{n-1} \times E^1$ (except possibly for a set of n-measure zero). That is, the (n+1)-th stage likelihood ratio function must be defined at least for all points in $C_{n-1} \times E^1$. If the pair L, R of sequences yields a criterion which satisfies V and VI, then it is said to be admissible, and the resulting criterion is written $[L/R]$ to denote its dependence on the given sequences. Moreover, the resulting test is called a sequential ratio test. It is perhaps moot whether there are other systematic means (of generating criteria) which cannot be rejected immediately on the grounds that the resultant computational difficulties would be excessive. At any rate, the only such systematic method known as yet is that employing likelihood ratio. For that reason consideration is usually restricted to sequential ratio tests.

When $L = \{b_n\}$ and $R = \{a_n\}$ are both constant sequences, i.e., $a_1 = a_n$ and $b_1 = b_n$ for all n, and if $[L/R]$ is admissible, the resulting ratio test is called a Wald test.

A.4 Optimum Tests

To each sequential ratio test there are assigned four numbers or parameters: α, β, $E_1(r)$ and $E_0(r)$. It is desirable to choose criteria $[L/R]$ which make these numbers as small as possible. Suppose hypotheses H_1 and H_0 are given and error probabilities α and β prescribed. Then Wald defines an optimum test (at the level of α, β) as a test

1) Whose error probabilities are α and β,

2) Whose average sample numbers are minimum among all other tests whose error probabilities are also α and β.

Further, Wald conjectures that when the class of tests at the (α, β) level contains a Wald test, then the Wald Test is optimum. To support this conjecture he proves that if the hypotheses H_o and H_1 are independent and if, in the Wald test, $b_1 = b_n \leq \ell_n(x) \leq a_n = a_1$ for all x in $C_{n-1} \times E^1 = A_n \cup B_n \cup C_n$, then the Wald test is indeed optimum.

The second hypotheses says that, in the Wald test's criteria, $A_n \subset \ell_n^{-1}(a_1)$ and $B_n \subset \ell_n^{-1}(b_1)$. Moreover $\sum_{i=1}^{\infty} F(A_i) + F(B_i) = 1$. Hence for some n at least, $\ell_n^{n-1}(a_1)$ has F_n measure positive, and therefore positive Lebesgue measure; i.e., necessarily some of the point inverses of the likelihood ratio function have positive measure. In applications to signal detectability however, the likelihood ratio will be (real) analytic, so that all its point inverses have measure zero.[1] Thus Wald's theorem is of little value for such applications. In fact, whenever the functions f_n and g_n are continuous (and hence) induce measures F_n and G_n which have the property of assigning the same measure to a set as is assigned to its closure, it will follow that all point inverses of the likelihood ratio function have probability[2] zero. Therefore even under these much less restrictive conditions Wald's hypothesis does not hold.

This is one major gap in the theory of sequential analysis so far as applications to signal detection are concerned. The other question which also

---

[1] See Technical Report No. 13, Part I, Lemma 2, page 34.

[2] Having probability zero means the F_n and G_n measures both are equal to zero. Any such set will make no contributions to the parameters of the test.

remains to be answered is concerned with the consequences of knowing that Wald's test is always optimum. In this event, it would be desirable to know when the class of all sequential tests at a given (α, β) level includes a Wald test. Moreover, those pairs of hypotheses H_o and H_1 for which there is some sequential test whose average sample numbers are finite should be characterized, for these are the only hypotheses with which one would consider using a sequential test in the first place.

In connection with the question regarding the (α, β) levels at which there is a Wald test, some information is available.

Lemma A4. For every Wald test, $\alpha + \beta < 1$.

Proof: Because the given test is a Wald test, $a_n = a_1$ and $b_n = b_1$ for all n. From the inequalities

$$F_n(A_n) = \int_{A_n} f_n(x)\, d\mu(x) = \int_{A_n} \ell_n(x)\, dG_n \geq \left(\min_{x \in A_n} \ell_n(x)\right)\left(\int_{A_n} dG_n\right) = a_1 \cdot G_n(A_n),$$

$$F_n(B_n) = \int_{B_n} f_n(x)\, d\mu(x) = \int_{B_n} \ell_n(x)\, dG_n \leq \left(\max_{x \in B_n} \ell_n(x)\right)\left(\int_{B_n} dG_n\right) = b_1 \cdot G_n(B_n)$$

one obtains

$$F_n(A_n) \geq a_1\, G_n(A_n) \qquad \text{and}$$

$$F_n(B_n) \leq b_1\, G_n(B_n)$$

These expressions summed over all n yield

$$\frac{\beta}{1-\alpha} = \frac{\sum F_n(B_n)}{\sum G_n(B_n)} \leq b_1 < a_1 \leq \frac{\sum F_n(A_n)}{\sum G_n(A_n)} = \frac{1-\beta}{\alpha}$$

But $\frac{1-\beta}{\alpha} - \frac{\beta}{1-\alpha} = \frac{1-\alpha-\beta}{\alpha(1-\alpha)} > 0$ means that $1-\beta-\alpha > 0$ since the denominator is known to be positive.

APPENDIX B

SAMPLE PLANS

B.1 Introduction

The theory of the finite ratio test developed in Technical Report No. 13 depends on finding a sampling plan which throws no information away. If the measurements are to be of the instantaneous amplitude of the receiver input, then such a sampling plan on the sample interval I consists first of a basis for population N containing n linearly independent functions $\{x_i(t)\}$, i = 1, 2, ..., n, and sample points $\{t_i\}$, i = 1, 2, ..., n, in I with the property that every function w(t) in populations N and SN can be expressed as

$$w(t) = \sum_{i=1}^{n} w(t_i)\, x_i(t)$$

By measuring values of the receiver input w(t) at the sample points $\{t_i\}$ one obtains the coefficients needed to represent w(t) in terms of the known basis

functions $\{x_i(t)\}$. Such a basis together with the sample points determines an admissable sample plan on I.

B.2 If Populations N and SN are Finite Dimensional, Then There Is an Admissible Sample Plan

Since the populations are finite dimensional, there is a basis $\{y_i(t)\}$, $i = 1, 2, \ldots, n$, for them. It will be sufficient to construct a new basis $\{x_i(t)\}$ and sample points $\{t_i\}$ in I which have the property that

$$x_i(t_j) = \delta_{ij} .$$

First it is necessary to show that there are sample points $\{t_i\}$ for which $\det\left(y_i(t_j)\right) \neq 0$. This is certainly true if $n = 1$. Suppose it is true when the dimension equals n; this will imply that it is also true when the dimension equals $n + 1$. The proof goes as follows:

By the inductive hypothesis there are n sample points $\{t_i\}$, $i = 1, 2, \ldots, n$, with the property that

$$\det \begin{pmatrix} y_1(t_1) & \cdots & y_1(t_n) \\ \vdots & & \vdots \\ y_n(t_i) & \cdots & y_n(t_n) \end{pmatrix} \neq 0$$

Let

$$D(t) = \det \begin{pmatrix} y_1(t_1) & \cdots & y_1(t_n) & y_1(t) \\ \vdots & & \vdots & \vdots \\ y_n(t_1) & \cdots & y_n(t_n) & y_n(t) \\ y_{n+1}(t_1) & \cdots & y_{n+1}(t_n) & y_{n+1}(t) \end{pmatrix}$$

Then $D(t)$ can be expanded by minors along the last column, yielding $D(t) = \sum_{i=1}^{n+1} a_i\, y_i(t)$, where a_{n+1} is not zero, for it is the n-by-n determinant above. If $D(t) = 0$ for all t in the interval I, then all the a_i's must vanish because the y_i's are linearly independent on I. Hence $D(t)$ is not identically zero on I and therefore some t_{n+1} can be found for which $D(t_{n+1}) \neq 0$.

Now, in order to construct the desired basis $\{x_i(t)\}$, it is necessary only to solve the n^2 linear equations

$$\sum_{j=1}^{n} a_{ij}\, y_j(t_k) = \delta_{ik}$$

in n^2 unknowns $\{a_{ij}\}$; for if they can be solved, the desired x_i's can be chosen as $x_i(t) = \sum_{j=1}^{n} a_{ij}\, y_j(t)$. The solubility of these equations can be determined by examining the n^2 by n^2 determinant formed by their coefficients. If Q is used to denote the n x n matrix whose elements are $\{y_j(t_k)\}$, the determinant of the coefficients can be written as

$$\det \begin{pmatrix} Q & & & 0 \\ & Q & & \\ & & Q & \\ & & & \ddots \\ 0 & & & Q \end{pmatrix} = (\det Q)^n$$

which has just been shown to be non zero. Hence the equations can be solved.

B.3 Sampling in Arbitrarily Short Intervals

There are many instances where the functions of populations N and SN can be taken to be (real) analytic, as for instance when the signal is a tone modulated CW transmission. Such functions have the property that they never

vanish on any interval. This means that if $\{y_i(t)\}$, i = 1, 2, ..., n, is a basis for the populations on the interval I, then it is also linearly independent and therefore a basis on any sub-interval of I. Thus the proposition proved in B.2 by induction could be applied to determining sample points in any sub-interval of I. The rest of the demonstration in that paragraph applies to any collection of sample points as long as they are chosen in the interval I. In this way an admissible sampling plan for the interval I can be chosen with the sample points restricted to any arbitrarily small sub-interval. But for any admissible sampling plan in the given sample interval there is an optimum criterion, and so the ROC curves of any two admissible sampling plans for the given interval will be identical, since each is "optimum." If this theoretical result is interpreted literally it means that observations of the receiver input can be restricted to any small interval without impairing the effectiveness of the detection system (see page 24 for discussion of this matter).

APPENDIX C

PROBABILITY DENSITY FUNCTIONS

The only technical concept used in the body of this report which is needed in understanding the material is that of density function. The purpose of this appendix is to give a simple account of the meaning of this term.

Suppose that in a particular study only a finite number of different events are possible, for example, the events that can result from rolling dice. Then the classical definition of the probability of an event E is

$$\frac{\text{number of ways E can occur}}{\text{total number of events}} = P(E)$$

Unfortunately when the possible number of events is not finite, then the denominator of the above expression is infinite, and the quotient is zero (unless the denominator is also infinite, which only accentuates the difficulty). An example of such a situation can be constructed as follows.

Suppose a dart is to be thrown while aimed at the center of a target, where the dart's point is idealized into a mathematical point. It is (again ideally) possible to find a probability that the thrown dart will land in a certain circle by determining the frequency with which this occurs in a large number of tries. This probability may very well depend on where the circle is located on the board. In order to be able to compare the affinity of the thrown dart for circles of unequal size, one would divide the probability of each circle by the area of the circle,

$$\frac{P \text{ (dart landing in circle C)}}{\text{area of circle C}} = P^*(C)$$

Let $P^*(C)$ be called the normalized probability of hitting the circle.

In order to assign a number to each point x of the target in such a way as to represent the affinity of the dart for landing near that point, one would take an entire sequence of circles, each centered at the point, whose radii are decreasing to zero, designated by C_n. Then the limit of $P^*(C_n)$ as $n \to \infty$ could be used as the number $f(x)$ to be associated with the point x. This number may or may not be zero, which avoids the difficulty pointed out above when the classical definition of probability is applied to infinitely many events.

In addition, it can be proved that when the resulting function $f(x)$ is integrated over a circle C, the value of the integral is merely P(C) all over again. This function $f(x)$ is called the probability density function, or more simply, the density function. Its basic property is that by integrating it over a (geometric) figure, one obtains the originally assigned probability that an event (events are represented by points of the target) will be a point of the given figure. Thus the integral over the entire target, i.e., over all points, will be unity.

BIBLIOGRAPHY

On Statistical Approaches to the Signal Detectability Problem:

1. Peterson, W. W., and Birdsall, T. G., "The Theory of Signal Detectability," Technical Report No. 13, Electronic Defense Group, Department of Electrical Engineering, University of Michigan.

2. Lawson, J. L., and Uhlenbeck, G. E., Threshold Signals, McGraw-Hill, New York, 1950.

   This book is certainly the outstanding reference on threshold signals. It presents a great variety of both theoretical and experimental work. Chapter 7 presents a statistical approach of the criterion type for the signal detection problem, and the idea of a criterion which minimizes the probability of an error is introduced.

3. Davies, I. L., "On Determining the Presence of Signals in Noise," Proc. I.E.E. (London), Vol. 99, Part III, pp.45-51, March, 1952.

4. Woodward, P. M., and Davies, I. L., "Information Theory and Inverse Probability in Telecommunication," Proc. I. E. E. (London), Vol. 99, Part III, p. 37, March, 1952.

5. Woodward, P. M., and Davies, I. L., "A Theory of Radar Information," Phil. Mag., Vol. 41, p. 1001, 1950.

6. Woodward, P. M., "Information Theory and the Design of Radar Receivers," Proc. I.R.E., Vol. 39, p. 1521.

   Woodward and Davies have introduced the idea of a receiver having a posteriori probability as its output, and they point out that such a receiver gives a maximum amount of information. They have handled the case of an arbitrary signal function known exactly or known except for phase with no more difficulty than other authors have had with a sine wave signal. Their methods serve as a basis for the second part of this report.

7. Reich, E., and Swerling, P., "The Detection of a Sine Wave in Gaussian Noise," Jour. App. Phys., Vol. 24, p. 289, March, 1953.

   This paper considers the problem of finding an optimum criterion (of the second type presented in this report) for the case of a sine wave of limited duration, known amplitude and frequency, but unknown phase in the presence of Gaussian noise of arbitrary autocorrelation.

8. Middleton, D., "Statistical Criteria for the Detection of Pulsed Carriers in Noise, Jour. Appl. Phys., Vol. 24, p. 371, April, 1953.

A thorough discussion is given of the problem of detecting pulses (of unknown phase) in Gaussian noise. Both types of optimum criteria are discussed, but not in their full generality. The sequential type of test is discussed also. Middleton's equation (6.1) does not hold for the sequential test, and as a result, his calculations for the minimum detectable signal with a sequential test are incorrect.

9. Slattery, T. G., "The Detection of a Sine Wave in Noise by the Use of a Non-Linear Filter," Proc.,I. R. E., Vol. 40, p. 1232, October, 1952.

This article considers the problem of detecting a sine wave of known duration, amplitude, and frequency, but unknown phase in uniform Gaussian noise. The article contains several errors, and the results are not clearly presented.

10. Hanse, H., "The Optimization and Analysis of Systems for the Detection of Pulsed Signals in Random Noise," Doctoral Dissertation (MIT), January, 1951.

11. Schwartz, M., "A Statistical Approach to the Automatic Search Problem," Doctoral Dissertation (Harvard), June, 1951.

These dissertations both consider the problem of finding the optimum receiver of the criterion type for radar type signals.

12. North, D. O., "An Analysis of the Factors which Determine Signal-Noise Discrimination in Pulsed Carrier Systems," RCA Laboratory Report PTR-6C, 1943.

The ideas of false alarm probability and probability of detection are introduced. North argues that these probabilities will be most favorable when peak signal to average noise ratio is largest. The ideal filter, which maximizes this ratio, is derived. (This commentary is based on second-hand knowledge of the report.)

13. Kaplan, S. M., and Fall, R. W., "The Statistical Properties of Noise Applied to Radar Range Performance," Proc. I.R.E., Vol. 39, p. 56, January, 1951.

The ideas of false alarm probability and probability of detection are introduced and an example of their application to a radar receiver is given.

14. Marcum K. I., "A Statistical Theory of Target Detection by Pulsed Radar: Mathematical Appendix," Rand Corporation Report R-113, July 1, 1948.

This report contains a careful, thorough study of the mathematical problem which it considers.

On Statistics:

15. Neyman, J., and Pearson, E. S., "On the Problem of the Most Efficient Tests of Statistical Hypotheses, "Phil. Trans, Roy. Soc., Vol. 231, Series A p. 289, 1933.

16. Crámer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, 1951.

17. A. Wald, Sequential Analysis, John Wiley and Sons, 1947.

18. Grenander, U., "Stochastic Processes and Statistical Inference," Arkiv För Mathematik, Vol. 2, p. 195 (1950).

This paper presents among many things a likelihood ratio for infinite dimensional probability measure spaces which thereby relieves the Neyman-Pearson test of its restriction to finite dimensionality.

DISTRIBUTION LIST

1 copy	Director, Electronic Research Laboratory Stanford University Stanford, California Attn: Dean Fred Terman
1 copy	Commanding Officer Signal Corps Electronic Warfare Center Fort Monmouth, New Jersey
1 copy	Chief, Engineering and Technical Division Office of the Chief Signal Officer Department of the Army Washington 25, D. C. Attn: SIGGE-C
1 copy	Chief, Plans and Operations Division Office of the Chief Signal Officer Washington 25, D. C. Attn: SIGOP-5
1 copy	Countermeasures Laboratory Gilfillan Brothers, Inc. 1815 Venice Blvd. Los Angeles 6, California
1 copy	Commanding Officer White Sands Signal Corps Agency White Sands Proving Ground Las Cruces, New Mexico Attn: SIGWS-CM
1 copy	Signal Corps Resident Engineer Electronic Defense Laboratory P. O. Box 205 Mountain View, California Attn: F. W. Morris, Jr.
75 copies	Transportation Officer, SCEL Evans Signal Laboratory Building No. 42, Belmar, New Jersey For - Signal Property Officer Inspect at Destination File No. 25052-PH-51-91(1443)

1 copy	W. G. Dow, Professor Dept. of Electrical Engineering University of Michigan Ann Arbor, Michigan
1 copy	H. W. Welch, Jr. Engineering Research Institute University of Michigan Ann Arbor, Michigan
1 copy	Document Room Willow Run Research Center University of Michigan Willow Run, Michigan
10 copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan