AD NUMBER: AD0850012
LIMITATION CHANGES
TO: Approved for public release; distribution is unlimited.
FROM: Distribution authorized to U.S. Gov't. agencies and their contractors; Export Control; 1 Mar 1969. Other requests shall be referred to the Air Force Rome Air Development Center, Attn: EMCRS, Griffiss AFB, NY 13440.
AUTHORITY
RADC ltr, 17 SEP 1971

AD850012

RADC-TR-69-24
Final Technical Report
March 1969

SHORT-TIME SAMPLE ROBUST
DETECTION STUDY

L. D. Davisson
J. B. Thomas
Princeton University

Rome Air Development Center
Air Force Systems Command
Griffiss Air Force Base, New York

116

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related government procurement operation, the government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded, by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Publication of this report does not constitute Air Force approval of the report's findings or conclusions. It is published only for the exchange and stimulation of ideas.

Do not return this copy. Retain or destroy.

SHORT-TIME SAMPLE ROBUST
DETECTION STUDY

L. D. Davisson
J. B. Thomas
Princeton University

AFLC, GAFB, N.Y., 28 Mar 69-102

FOREWORD

This final technical report was prepared by Messrs. L. D. Davisson and J. B. Thomas, Princeton University, Princeton, N. J. under Contract F30602-68-C-0078, Project 4519, Task 451902. Mr. Charles N. Meyer (EMCRS) was the Rome Air Development Center project engineer.

The distribution of this report is limited by the Department of State International Traffic in Arms Regulations and Mutual Defense Assistance Control List.

This report has been reviewed and is approved:

Approved: Charles N. Meyer
CHARLES N. MEYER
Project Engineer

Approved: Richard M. Cosel
RICHARD M. COSEL
Colonel, USAF
Chief, Communications Division

FOR THE COMMANDER: Fred I Diamond
for IRVING J. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

The optimum detector is well defined when the signal and noise characteristics are completely specified through deterministic and/or probabilistic statements. Unfortunately this is not always the case and attempts to make it so through assumptions of one sort or another may lead to poor detector performance. This report studies the detection problem when certain of the signal or noise properties are unknown except perhaps within some wide class. Particular attention is directed to communications problems characterized by small time-bandwidth signals. Several applications emerge from the study. The first results from an exhaustive investigation of the nonparametric Wilcoxon rank sum detector which is found to be practical in implementation and nearly as good in performance as the optimum detector under a wide range of sample size, dependence, and noise distribution conditions. A second application from the study results from an investigation of optimum small sample linear coincidence detectors which are found to be superior to a Gaussian parametric detector when the normality assumption is violated. A third application from the study results from the analysis of a simple adaptive threshold detector which is practical to implement and attains an improvement over fixed threshold receivers. In addition there are results of an incomplete or theoretical nature whose application is either not so important or as yet well defined. These include the study of an optimum "robust" detector for nearly Gaussian noise.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF SYMBOLS

Λ = likelihood ratio

Λ_o, C = threshold

σ = noise standard deviation

λ = signal amplitude factor

K, R = covariance matrix

$v(t)$, $v_1(t)$, $v_2(t)$ = continuous time received waveforms

V = vector of observed samples, $V_i = v(i\Delta t)$

Δt = sampling interval

T = total observation time

Pr(.) = probability of event in brackets

p = probability

$\hat{p}$ = probability estimate

p(.), f(.) = probability density

$\varphi(.)$ = normal density

$\Phi(.)$ = normal distribution function

SNR = signal to noise ratio

u(.) = unit step function, $u(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$

sgn(.) = sign function, $\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$

SGN V = vector of signs of V

r(.) = unit ramp, $r(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases}$

S = signal vector or a test statistic

A.R.E. = asymptotic relative efficiency

z_i = sign of the ith ranked sample

Z = rank vector

U_N = vector of ones

W_N = noise vector

SECTION I

INTRODUCTION

In recent years there has been increasing interest in the application of nonparametric, adaptive or robust methods to the solution of practical engineering problems involving signals imbedded in additive signal-independent noise. These methods are designed to guarantee certain levels of performance over some broad class of signal and/or noise distributions (e.g. for sampled data the class of all sample-to-sample independent noise sequences with symmetric densities). In some cases an additional advantage of these techniques lies in their relative simplicity of implementation compared with the optimum procedure even if it were known.

Nonparametric statistics originated in the statistical literature. Hence certain assumptions were made in their derivation which are not always valid or desirable in engineering applications. The most important of these is the assumption of independent sampling. In an engineering situation one would like to decide on the presence or absence of a signal or to decide between several signals on the basis of time-continuous observations which are generally not incrementally independent for sufficiently small time increments. Although one can, in principle, sample the observations at such a slow rate that the sample values are statistically independent, such slow sampling is not generally desirable since it involves ignoring a sizeable portion of the available data. This report considers the effect of dependent observations on a number of nonparametric tests for both sampled and continuous time versions.

The usual measure of nonparametric test effectiveness has been the asymptotic relative efficiency (A.R.E.) of the test. The A.R.E. is defined as the relative amount of data (in terms of number of samples or observation time) required for a given (perhaps nonparametric) detector performance compared with the amount required for a second (usually the optimum) test in the limit as the signal amplitude goes to zero and the amount of data becomes infinite. This measure has limited usefulness in the study presented herein where attention is concentrated on small sample, large signal results although in some cases it is shown that A.R.E. and small sample conclusions are consistent. Hence the relative efficiency, defined as the relative signal amplitude required for a given test at a given probability of error with a fixed amount of data is used or, in some cases, probabilities of error are compared for a fixed signal amplitude.

Particular attention is devoted to nonparametric coincidence and rank tests. Suppose one channel of observation is available where $\{v(t); 0 \leq t \leq T\}$ is the observed data and the N sampled values $v(\Delta t)$, $v(2\Delta t), \ldots, v(N\Delta t = T)$ are taken. A <u>coincidence test</u> depends only on the <u>signs</u> of these observations and not their magnitudes. The <u>rank</u>, $r(v(k\Delta t))$ of $v(k\Delta t)$ is defined as the number of sample values which are smaller than or equal to $v(k\Delta t)$ in <u>magnitude</u> (disregarding the sign of the values in the ranking procedure). Let $z_1, z_2, \ldots, z_N$ be the set of <u>signs</u> of the ranked data where

$$z_1 = \begin{cases} +1 & \text{if the samllest magnitude sample if positive} \\ -1 & \text{if it is negative} \end{cases}$$

$$z_2 = \begin{cases} +1 & \text{if the second smallest magnitude sample is positive} \\ -1 & \text{if it is negative} \end{cases}$$

and so on. Rank tests are those which base signal decisions on the N values $\{z_i\}$ only. In Section II the optimum coincidence detector is considered. In Section III the optimum rank test is found for independent Guassian-distributed sample values. This is of theoretical interest primarily due to the implementational complexity involved. In Section IV a well known rank test, the Wilcoxon signed rank test, is investigated. This test is based on the sum of the ranks of the observations multiplied by their signs:

$$S_w = \sum_{i=1}^{N} z_i \, i.$$

Thus ranks count in direct proportion to their magnitude, the smaller ranked observations having a smaller effect than the larger ones on the test statistic. The Wilcoxon statistic is implementable with relative ease when it is expressed in the equivalent form:

$$S_w = \sum_{i=1}^{N} \sum_{j=1}^{i} \operatorname{sgn}(v(i\Delta t) + v(j\Delta t))$$

where "sgn" is the sign function

$$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases} .$$

This can be formed easily for small sample sizes with simple logic circuits. In Section IV it is found that the test performs nearly as well as the optimum rank test and compares favorably with the Gaussian parametric test for a range of dependence and nonnormality conditions. It is likely that tests of this form can be used in practice.

In Section V asymptotic properties of several rank tests for one and two channels of observation are investigated for large sample sizes with dependence as a guide to possible behav ›r under small sample conditions. Large sample sizes are used due to the relative ease of obtaining analytical results. It is found that the nonparametric tests investigated perform surprisingly well compared with the optimum parametric test. As this result holds for both large and small sample sizes for the Wilcoxon test, it seems likely that this will also be true for the other tests investigated in Section V primarily on large sample sizes.

An additional result of Section V is the asymptotic nomality of many rank type tests under dependence. This is primarily of theoretical interest, although it does simplify setting the test decision threshold for large sample sizes.

If the noise distributions were completely known, the optimum parametric test would be used. When very little is known of the noise distribution nonparametric tests can be used. In the intermediate case when some imprecise knowledge of the noise distribution is known, robust tests can be used. In Section VI the development of the optimum robust test is presented when the noise is "nearly" Gaussian and evaluated compared with the optimum parametric test. It is found that this robust test is superior to the optimum parametric test for Gaussian noise when the noise is not Gaussian but only "nearly" so.

In many communications applications simple schemes can be developed for adaptively improving detector performance if it is possible to distinguish that part of the observed data which is due to noise from that due to the signal. One way to do this is to start with an initial detector based on some initial guess, and then to modify it after each signal decision based upon the assumption that the correct decision wa made regarding which of M signals was sent in the most recent time interval. Such devices are called <u>decision directed receivers</u>. Although simple to implement, decision directed receivers are difficult to analyze because some of the decisions that are assumed to be correct are actually wrong, causing detector degradation instead of improvement. Under some circumstances performance can degenerate completely. In Section VII of this report one such scheme is analyzed. It is expected that techniques coming from these results will be of significant practical use.

Section VIII summarizes the important results with recommendations for future work. Section IX presents conclusions.

Many of the detailed computations are relegated to the appendices for the sake of clarity of the main presentation. Particular attention is directed to Appendix I which presents a review of pertinent statistical terminology and results.

SECTION II

LINEAR COINCIDENCE PROCEDURES FOR THE DETECTION OF KNOWN SIGNALS IN ILL-DEFINED BACKGROUND NOISE

The concept of a linear coincidence detector is an outgrowth of conventional optimum linear filtering theory suggested by the distribution-free performance of coincidence procedures in general. Coincidence procedures have previously been considered [1] for the detection of weak signals in noise of uncertain origin when the noise samples are assumed independent and identically distributed (i.i.d.) and the sample size or observation interval is large. The basis for comparison has been the asymptotic relative efficiency (A.R.E.) which is an asymptotic comparison of the relative sample sizes required by two detectors to achieve identical false alarm probability α and detection probability β in the limit of large sample sizes and vanishingly small input signals. Such comparisons are particularly appropriate to passive search and/or surveilance systems operating at threshold input signal levels where generally the assumption of large sample sizes or observation intervals is valid. The problem to be considered here, however, is the detection of known signals of arbitrary amplitude in ill-defined background noise, a situation more akin to active systems and/or the data communications problem. A performance criteria in terms of A.R.E. bears little relevance to this problem and consideration must be given to the small-sample, large-signal performance.

Some background, including a precise statement of the problem under consideration, is given together with a formulation of the locally optimum linear coincidence detector in the following material. It is found, as with most optimum nonparametric procedures on dependent data, that the implementation of the locally optimum linear coincidence detector requires parametric knowledge of the underlying distributions. To avoid this requirement for parametric knowledge, a particular suboptimum linear coincidence procedure is proposed and investigated in detail. A critique of the A.R.E. as a performance criteria for small sample conditions is given and some experimental results obtained by computer simulation are included.

A single channel detection situation is considered where the observable $\underline{V}_N = (V_1, V_2, \ldots, V_N)$ represents an N-vector of observations obtained by discrete homogeneous sampling of a stationary continuous parameter process such that $V_i = v(i\Delta t)$ as described in the Introduction and

$$\underline{V}_N = \lambda \underline{S}_N + \underline{W}_N$$

where the N-vector $\underline{S}_N$ represents a known signal or regression sequence, $\underline{W}_N$ is an N-vector of additive signal independent noise with components of variance σ^2 and normalized covariance martrix $\underline{\underline{K}}_N$ (when it exists) and where λ is a scalar signal amplitude. On the basis of the observable $\underline{V}_N$ we desire to decide between the two hypotheses

$$H_0 : \lambda = -\lambda_1$$

$$\text{vs.} \qquad H_1 : \lambda = +\lambda_1$$

for some $\lambda_1 > 0$. If the detection problem is approached from the decision theoretic viewpoint, the optimum (in the sense of Bayes, Neyman-Pearson, etc.) detector is based upon athreshold test of the likelihood ratio. If the additive noise $\underline{W}_N$ is zero-mean Gaussian with normalized covariance matrix $\underline{\underline{K}}_N$ we obtain the well-known result

$$\langle \underline{V}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle \gtrless \Lambda_0$$

where Λ_0 is a suitably chosen decision threshold and $\langle \cdot, \cdot \rangle$ denotes inner product. Letting $\underline{a}_N = \underline{\underline{K}}_N^{-1} \underline{S}_N$, the test statistic becomes

$$\langle \underline{V}_N, \underline{a}_N \rangle = \sum_{i=1}^{N} a_i V_i$$

which is merely a linear filtering operation on the observed data. If the problem were approached from the point of view of optimum linear filtering theory, the object would be to apply a linear weighting to the received data in an effort to maximize the SNR at the termination of the observation interval. The quantity so obtained is then compared to a decision threshold resulting in a decision in favor of H_0 or H_1. Again the filtering or smoothing sequence can be represented by the vector $\underline{a}_N$ with the SNR at the termination of the observation interval given by

$$\text{SNR} = \frac{\lambda^2}{\sigma^2} \quad \frac{\langle \underline{a}_N, \underline{S}_N \rangle^2}{\langle \underline{a}_N, \underline{\underline{K}}_N \underline{a}_N \rangle}$$

This quantity is easily shown to be maximized by choosing

$$\underline{a}_N = \underline{a}_N^* \stackrel{\Delta}{=} \mu \, \underline{\underline{K}}_N^{-1} \underline{S}_N$$

where μ is an arbitrary positive scale factor. This is the matched filter which is identical to the decision theoretic results for Gaussian data. In the early engineering literature (cf. [2]) maximization of SNR appeared a reasonable rule-of-thumb in any detection situation. Fortunately this ad hoc procedure coincided with the more rigorous decision theoretic results for Gaussian data. More recent results [3,4] have related this procedure to some asymptotic results as both $N \to \infty$ and $\lambda \to \infty$. Unfortunately linear filtering is ineffective against noise processes possessing broad-tailed univariate probability densities typical of say, impulse noise or jamming environments. In fact, due to the often paralyzing effect of large noise peaks on typical receiver circuitry, linear filtering is often susceptible to catastrophic degradations as the result of such peaks which may occur still only a fraction of the time. When the additive noise $\underline{W}_N$ is independent and $\underline{S}_N = \underline{U}_N$, a unit vector of ones (i.e., the d.c. signaling problem) coincidence procedures such as the sign detector have been advocated [1] as a method of avoiding the sensitivity of linear filtering to underlying statistics. The threshold test based upon the sign detector is given by

$$T_N(\underline{V}_N) = \sum_{i=1}^{N} \mathrm{Sgn}\, V_i \gtrless \Lambda_0$$

or in terms of vector-matrix notation

$$T_N(\underline{V}_N) = \langle \underline{U}_N, \mathrm{SGN}\, \underline{V}_N \rangle \gtrless \Lambda_0$$

where

$$\mathrm{SGN}\ \underline{V}_N = (\mathrm{Sgn}\, V_1, \mathrm{Sgn}\, V_2, \ldots, \mathrm{Sgn}\, V_N)$$

These detectors can be implemented using binary devices and provide a fair degree of protection against abnormal behavior while providing reasonable (at least asymptotically) efficiency under normal conditions. Unfortunately little consideration has been given to their performance in correlated noise situations and in fact no analysis has been presented when something is known of signal structure. In such situations it would appear that a linear theory could be developed to advantage for coincidence procedures of the form

$$T_N = \sum_{i=1}^{N} a_i \operatorname{Sgn} V_i$$

$$= \langle \underline{a}_N, \operatorname{SGN} \underline{V}_N \rangle$$

We shall call detectors employing statistics of this form linear coincidence detectors. An optimum coincidence detector will be one that minimizes the sample size N required to achieve a fixed false alarm probability α and detection probability β for a given input SNR. It is overly ambitious to expect that such a detector exists which minimizes the sample size uniformly in α, β and SNR. Practical considerations then dictate passing to the asymptotic case as both $N \to \infty$ and SNR $\to 0$. The result is, of course, the A.R.E. By maximizing the A.R.E. within the class of linear coincidence detectors we obtain a structure which is locally optimum. Prescription is then given to the theory that by optimizing performance for weak input signals the performance in the less critical strong signal regime will be adequate.

Under appropriate regularity conditions (cf. [5]), which are assumed to be satisfied, the A.R.E. of a detector T based upon the sequence of test statistics $\{T_N\}_N$ with respect to another detector T' employing the sequence of test statistics $\{T'_N\}_N$ can be expressed as the ratio of their efficacies; i.e.,

$$\text{A.R.E.}_{T,T'} = \frac{\mathcal{E}_T}{\mathcal{E}_{T'}}$$

where

$$\mathcal{E}_T \triangleq \lim_{N \to \infty} \frac{\left[\frac{\partial}{\partial\lambda} E_\lambda\{T_N\}\big|_{\lambda=0}\right]^2}{N \operatorname{var}_{H_0}\{T_N\}}$$

and similarly for $\mathcal{E}_{T'}$. If T is a linear coincidence detector with test statistic T_N defined for each N as above, then

$$\mathcal{E}_T = \lim_{N\to\infty} \frac{\left[\frac{\partial}{\partial\lambda} \langle \underline{a}_N, E_\lambda\{SGN\ \underline{V}_N\}\rangle\Big|_{\lambda=0}\right]^2}{N\langle \underline{a}_N, \underline{\underline{R}}_N, \underline{a}_N\rangle}$$

where $\underline{\underline{R}}_N$ is the $N \times N$ covariance matrix of the vector SGN $\underline{V}_N$ with (i, j) element

$$r_{ij} = \mathrm{Cov}(\mathrm{Sgn}\ V_i, \mathrm{Sgn}\ V_j)$$

and $E_\lambda\{SGN\ \underline{V}_N\}$ is the N-vector with i'th component $E_\lambda\{\mathrm{Sgn}\ V_i\}$, $i = 1, 2, \ldots, N$. It is assumed that the additive noise vector $\underline{W}_N$ has univariate c.d.f. $F(\cdot)$ possessing density $f(\cdot)$ symmetrical about the origin. Then

$$E_\lambda\{\mathrm{Sgn}\ V_i\} = 1 - 2F(-\lambda S_i)$$

Now for λS_i in some interval $[-h, h]$ about the origin, we can expand $F(-\lambda S_i)$ in a Taylor series in powers of λS_i with the result

$$F(-\lambda S_i) = F(0) - \lambda S_i f(0) + 0(\lambda^2)$$

and by the symmetry assumption on F we have $F(0) = 1/2$ so that

$$E_\lambda\{\mathrm{Sgn}\ V_i\} = 2f(0)\lambda S_i + 0(\lambda^2)$$

and hence

$$\frac{\partial}{\partial\lambda} \langle \underline{a}_N, E_\lambda\{SGN\ \underline{V}_N\}\rangle\Big|_{\lambda=0} = 2f(0)\langle \underline{a}_N, \underline{S}_N\rangle$$

The expression for efficacy given above then becomes

$$e_T = \lim_{N \to \infty} 4f^2(0) \frac{\langle \underline{a}_N, \underline{S}_N \rangle^2}{N \langle \underline{a}_N, \underline{\underline{R}}_N \underline{a}_N \rangle}$$

For a fixed noise distribution, the quantity

$$\epsilon_N = \frac{\langle \underline{a}_N, \underline{S}_N \rangle^2}{\langle \underline{a}_N, \underline{\underline{R}}_N \underline{a}_N \rangle}$$

serves as a local measure of SNR which if maximized for each N results in a detector structure which optimizes the A.R.E. within the class of linear coincidence detectors. Then the following reasonable definition is made:

Definition: The linear-coincidence detector based upon the sequence of statistics $\{T^*_N\}_N$ with

$$T^*_N(\underline{V}_N) = \langle \underline{a}^*_N, \text{SGN } \underline{V}_N \rangle$$

is called a locally optimum linear coincidence detector if for each N

$$\frac{\langle \underline{a}^*_N, \underline{S}_N \rangle^2}{\langle \underline{a}^*_N, \underline{\underline{R}}_N \underline{a}^*_N \rangle} = \max_{\underline{a}_N} \frac{\langle \underline{a}_N, \underline{S}_N \rangle^2}{\langle \underline{a}_N, \underline{\underline{R}}_N \underline{a}_N \rangle}$$

where the maximum is taken with respect to all $\underline{a}_N$ such that $||\underline{a}_N||^2 \triangleq \langle \underline{a}_N, \underline{a}_N \rangle < \infty$.

Now assuming that the covariance matrix $\underline{\underline{R}}_N$ is nonsingular for each N one obtains for the maximizing $\underline{a}^*_N$

$$\underline{a}^*_N = \mu \underline{\underline{R}}^{-1}_N \underline{S}_N$$

If the noise is independent $\underline{\underline{R}}_N = \underline{\underline{I}}_N$, the identity matrix, and if $\underline{S}_N = \underline{U}_N$ the optimum coincidence procedure reduces to

$$T_N^*(\underline{V}_N) = \mu \langle \underline{U}_N, \mathrm{SGN}\, \underline{V}_N \rangle$$

$$= \mu \sum_{i=1}^{N} \mathrm{Sgn}\, V_i$$

which is up to an arbitrary multiplicative factor equal to the ordinary sign test. The above result is completely analogous to the result

$$a_N^* = \mu\, \underline{\underline{K}}_N^{-1} \underline{S}_N$$

obtained for the optimum linear filter. If the output of the linear filter can be shown to satisfy the appropriate regularity conditions (in particular to be asymptotically normal), the A.R.E. of the locally optimum coincidence detector with respect to the optimum linear detector is easily shown to be given by

$$\mathrm{A.R.E.}_{C^*, L^*} = \lim_{N \to \infty} 4f^2(0)\sigma^2 \frac{\langle \underline{S}_N, \underline{\underline{R}}_N^{-1} \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle}$$

As an example, if the additive noise is white Gaussian with univariate density

$$f(\omega) = \frac{1}{\sqrt{2\pi\sigma^2}}\; e^{-1/2\,(\omega^2/\sigma^2)}$$

one obtains

$$\underline{\underline{K}}_N^{-1} = \underline{\underline{I}}_N = \underline{\underline{R}}_N^{-1}$$

so that

$$\text{A.R.E.}_{C^*, L^*} = 2/\pi$$

a known result for the sign test compared to the sample mean. Evaluation of $\text{A.R.E.}_{C^*, L^*}$ for some additional data models as well as some experimental results follow.

In the expression for $\text{A.R.E.}_{C^*, L^*}$ given above, the ratio of quadratic forms can be bounded above and below in terms of the eigenvalues of $\underline{\underline{R}}_N$ and $\underline{\underline{K}}_N$. Passing to the limit on N it is possible to bound the A.R.E. This procedure requires direct evaluation of the eigenvalues, however, and does not appear fruitful. An alternate approach is to employ some asymptotic results from linear estimation theory. In particular, let

$$\underline{W}_N = \{w_k\}_{k=1}^{N}$$

be a discrete zero-mean noise process possessing spectral density $p(\lambda)$ $\lambda \epsilon [-\pi, \pi]$ and normalized covariance function

$$r(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\lambda} p(\lambda) d\lambda$$

and corresponding covariance matrix $\underline{\underline{R}}_N$.

Now let $\underline{S}_N = \{s_k\}_{k=1}^{N}$ be a discrete stationary deterministic signal sequence, with spectral distribution function $P_s(\lambda)$ $\lambda \epsilon [-\pi, \pi]$ normalized so that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} dP_s(\lambda) = 1$$

It can be shown [6, 7] that if $p(\lambda)$ does not vanish on $[-\pi, \pi]$ and coincides a.e. with an analytic function, then

$$\lim_{N \to \infty} \frac{\langle \underline{S}_N, \underline{\underline{R}}_N^{-1} \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{S}_N \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{dP_s(\lambda)}{p(\lambda)}$$

where $\underline{\underline{R}}_N$ is the $N \times N$ covariance matrix with (i, j) element r(i-j). Likewise

$$\lim_{N \to \infty} \frac{\langle \underline{S}_N, \underline{\underline{R}}_N \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{S}_N \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} p(\lambda) dP_s(\lambda)$$

In particular, if $\underline{S}_N = \underline{U}_N$ a unit vector, then $dP_s(\lambda) = 2\pi\, \delta(\lambda) d\lambda$ so that

$$\lim_{N \to \infty} \frac{\langle \underline{S}_N, \underline{\underline{R}}_N^{-1} \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{S}_N \rangle} = \frac{1}{p(0)}$$

Applying this result to the expression for $A.R.E._{C^*, L^*}$ one obtains

$$A.R.E._{C^*, L^*} = 4f^2(0)\sigma^2 \frac{p(0)}{\tilde{p}(0)}$$

where $p(\lambda)$ is the spectral density of the raw data and $\tilde{p}(\lambda)$ is that of the infinitely-clipped data. Since by definition

$$p(\lambda) = \sum_{k=-\infty}^{\infty} \rho(k) e^{-ik\lambda}$$

it follows that

$$p(0) = \sum_{k=-\infty}^{\infty} \rho(k)$$

and similarly

$$\tilde{\rho}(0) = \sum_{k=-\infty}^{\infty} \tilde{\rho}(k)$$

with $\tilde{\rho}(k)$ the covariance function of the infinitely-clipped data samples. Thus

$$\text{A.R.E.}_{C^*, L^*} = 4f^2(0)\sigma^2 \frac{\sum_{k=-\infty}^{\infty} \rho(k)}{\sum_{k=-\infty}^{\infty} \tilde{\rho}(k)}$$

For instance, for Gauss-Markov data one has $\rho(k) = e^{-\pi|k|/\gamma}$ with $\gamma = fs/b_o$ the relative sampling rate normalized to the double-sided noise bandwidth b_o cps. Likewise

$$\tilde{\rho}(k) = \frac{2}{\pi} \; \text{Sin}^{-1} \rho(k)$$

so that

$$\text{A.R.E.}_{C^*, L^*} = \frac{2}{\pi} \; \frac{1+e^{-\pi/\gamma}}{1-e^{-\pi/\gamma}} \; \frac{1}{\frac{2}{\pi}\sum_{k=-\infty}^{\infty} \text{Sin}^{-1}\{e^{-\pi|k|/\gamma}\}}$$

A plot of the A.R.E. as a function of relative sampling rate γ is illustrated in Figure 1. Note that the A.R.E. rises monotonically from a value $2/\pi$ for $\gamma < 1$ to its limiting value

$$\lim_{\gamma \to \infty} \text{A.R.E.}_{C^*, L^*} = \frac{2}{\pi}\left(\frac{1}{\ln 2}\right) \approx .916$$

Thus for increasing relative sampling rates, the locally optimum linear coincidence detector operating on Gauss-Markov data appears an

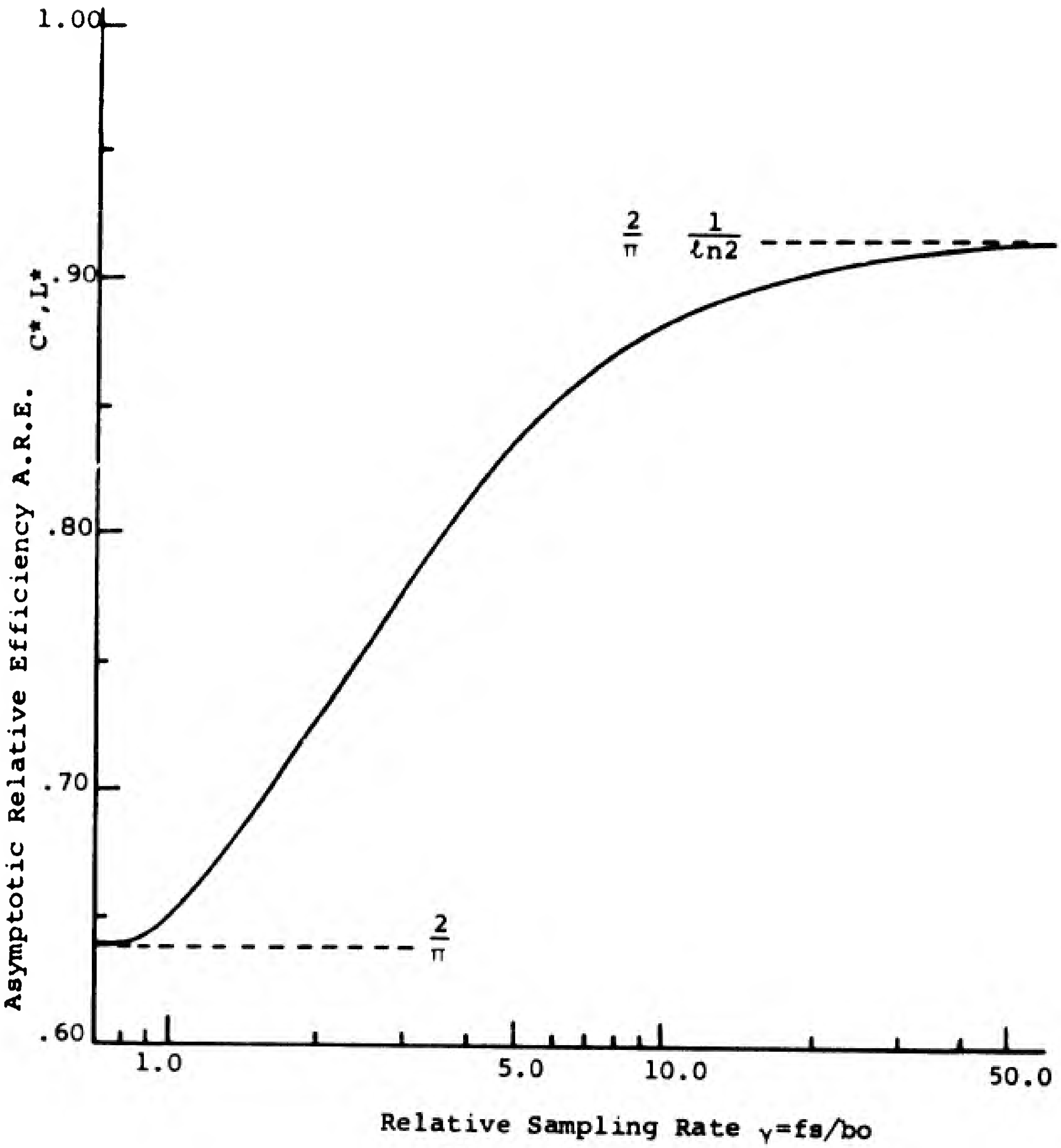


Figure 1

Asymptotic Relative Efficiency as a Function of Relative Sampling Rate for Locally Optimum Linear Coincidence Detector Operating on Gauss-Markov Data even more attractive alternative to the optimum linear detector. Another process of interest is the first-order Laplace autoregressive process described by the autoregression

$$w_n = \rho w_{n-1} + \eta_n$$

where $\rho \overset{\Delta}{=} e^{-\pi/\gamma}$ and $\{\eta_n\}_n$ is an independent identically distributed sequence with common univariate distribution H described by the mixture

$$dH(\eta) = \rho^2 \delta(\eta) d\eta + (1-\rho^2) \frac{1}{\sqrt{2}\sigma} e^{-\sqrt{2}|\eta/\sigma|} d\eta$$

where $\delta(\cdot)$ is the Dirac Delta and is used to denote a distribution degenerate at the origin. It can be shown [8] that for this process $\rho(k) = \tilde{\rho}(k)$ so that now

$$A.R.E._{C^*, L^*} = 2$$

independent of sampling rate. This compares with the well-known result A.R.E. = 2.0 for the sign detector with respect to the sample mean operating on independent Laplace data. In fact, in this case it can easily be shown that the sign detector is locally most powerful.

For more broad-tailed noise distributions it is anticipated that the improvement in A.R.E. would be even more pronounced. It is important to note that these conclusions are based upon asymptotic comparisons as $N \to \infty$ and SNR $\to 0$. Little, if anything, is known of the small-sample, large-signal behavior of coincidence detectors and since it is this regime in which interest is concentrated, the utility of asymptotic comparisons is of dubious value as a performance indicator. Nevertheless, Carlyle [9] states "....numerical studies have frequently shown that for small N there is no gross violation of the conclusions reached on an A.R.E. basis and often the latter are often reinforced for finite N." Unfortunately no published evidence is given in support of this statement and, in fact, some results obtained in the following material cast some suspicion on its validity.

The locally optimum smoothing sequence $\underline{a}_N^*$ for linear coincidence procedures of the form $T_N = \langle \underline{a}_N, \text{SGN } \underline{V}_N \rangle$ has been shown to be given by $\underline{a}_N^* = \mu \underline{\underline{R}}_N^{-1} \underline{S}_N$, where μ is an arbitrary scalar and $\underline{\underline{R}}_N$ is the covariance matrix of the infinitely clipped data. Unfortunately, even if the normalized covariance matrix $\underline{\underline{K}}_N$ of the raw data is known, evaluation of $\underline{\underline{R}}_N$ requires parametric knowledge of the underlying noise distribution. In situations where $\underline{\underline{K}}_N$ is known a priori, a reasonable suboptimum procedure might be to employ the suboptimum smoothing sequence $\underline{a}_N = \mu \underline{\underline{K}}_N^{-1} \underline{S}_N$. Actually for the Laplace autoregressive process, this coincides with the locally optimum sequence as noted previously. The resulting expression for the efficacy of the suboptimum linear coincidence detector is given by

$$\mathcal{E}_C = \lim_{N \to \infty} \frac{4f^2(0)}{N} \frac{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle^2}{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{\underline{R}}_N \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle}$$

and the A.R.E. with respect to the optimum linear detector is given by

$$\text{A.R.E.}_{C,L^*} = \lim_{N \to \infty} 4f^2(0)\sigma^2 \frac{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{\underline{R}}_N \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle}$$

Again it is possible to evaluate this asymptotic expression in terms of spectral properties as in the preceding section. In particular, let

$$\underline{\underline{K}}_N^{-1} = \underline{\underline{Q}}_N' \underline{\underline{Q}}_N$$

for some nonsingular matrix $\underline{\underline{Q}}_N$ which can always be taken to be triangular. Then the ratio of quadratic forms above becomes

$$\frac{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle}{\langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{\underline{R}}_N \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle} = \frac{\langle \underline{\Psi}_N, \underline{\Psi}_N \rangle}{\langle \underline{\Psi}_N, \underline{\underline{Q}}_N \underline{\underline{R}}_N \underline{\underline{Q}}_N' \underline{\Psi}_N \rangle}$$

where

$$\underline{\Psi}_N = \underline{\underline{Q}}_N \underline{S}_N$$

The vector $\underline{\Psi}_N$ is thus obtained by passing the signal vector $\underline{S}_N$ through a prewhitening filter represented by the matrix $\underline{\underline{Q}}_N$. If $\underline{S}_N = \underline{U}_N$ as in the preceding section, one has for the spectral distribution $P_\Psi(\lambda)$ of the "prewhitened" component $\underline{\Psi}_N$

$$dP_\Psi(\lambda) = K \frac{\delta(\lambda)}{p(0)} d\lambda$$

where $p(\lambda)$ is the spectral density of the raw data and K is a constant chosen to satisfy the normalization constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} dP_\Psi(\lambda) = 1$$

Obviously we must have $K = 2\pi p(0)$. Also observe that if $\underline{\xi}_N = \underline{\underline{Q}}_N \text{ SGN } \underline{V}_N$, then

$$\text{Cov}(\underline{\xi}_N, \underline{\xi}_N) = \underline{\underline{Q}}_N \text{ Cov}(\text{SGN } \underline{V}_N, \text{ SGN } \underline{V}_N) \underline{\underline{Q}}_N'$$

$$= \underline{\underline{Q}}_N \underline{\underline{R}}_N \underline{\underline{Q}}_N'$$

The vector $\underline{\xi}_N$ can be considered an additive noise sequence obtained by passing the infinitely-clipped data vector SGN $\underline{V}_N$ through the prewhitening filter represented by $\underline{\underline{Q}}_N$. Thus the spectral density of the ξ process is given by

$$p_\xi(\lambda) = \frac{\tilde{p}(\lambda)}{p(\lambda)}$$

and from previous comments

$$\lim_{N\to\infty} \frac{\langle \underline{\Psi}_N, \underline{\underline{Q}}_N \underline{\underline{R}}_N \underline{\underline{Q}}_N' \underline{\Psi}_N \rangle}{\langle \underline{\Psi}_N, \underline{\Psi}_N \rangle} = \frac{1}{2\pi} \int_{-\pi}^{\pi} p_\xi(\lambda) dP_\Psi(\lambda)$$

$$= \frac{\tilde{p}(0)}{p(0)}$$

Using this result the following A.R.E. is found:

$$\text{A.R.E.}_{C,L^*} = 4f^2(0)\,\sigma^2 \frac{p(0)}{\tilde{p}(0)}$$

$$= 4f^2(0)\sigma^2 \frac{\sum_{k=-\infty}^{\infty} \rho(k)}{\sum_{k=-\infty}^{\infty} \tilde{\rho}(k)}$$

This is precisely the result obtained for the locally optimum linear coincidence detector. In fact, it can easily be shown that the same result would have been obtained had we chosen

$$\underline{a}_N = \mu \underline{\underline{I}}_N^{-1} \underline{S}_N$$

with $\underline{\underline{I}}_N$ the $N \times N$ identity matrix. A similar situation in linear estimation theory was first considered by Grenander [10] and applied to a parametric detection problem by Davisson [11]. In particular, Grenander shows that in the linear estimation problem, if the spectrum of the regression sequence $\underline{S}_N$ consists of a single point, then knowledge of the true covariance matrix of the additive noise sequence provides no information with respect to the inference problem on $\underline{S}_N$ when performance is measured in terms of asymptotic efficiency. With respect to the present problem, we find that if $\underline{S}_N = \underline{U}_N$, the A.R.E. of an arbitrary linear coincidence detector with respect to an arbitrary linear detector is given by the right-hand side above. This result casts considerable doubt on the utility of A.R.E. as a valid performance criteria in the present situation. As pointed out by Davisson [11], the small sample behavior is the more appropriate performance measure. An extensive series of digital computer simulations have been initiated to enable some inferences to be drawn concerning the small-sample, large-signal behavior of linear coincidence detectors vis-à-vis linear detector structures. The results of these simulations for equiprobably and dipodal signalling are described next.

The additive noise sequence $\{w_n\}_n$ is now assumed to satisfy the first-order autoregression

$$w_n = \rho w_{n-1} + \eta_n$$

with the residual sequence $\{\eta_n\}_n$ assumed to be independent and $\rho \triangleq e^{-\pi/\gamma}$ with $\gamma = fs/b_\rho$ the relative sampling rate described previously. In particular, we consider three such processes with corresponding univariate densities $f(\omega)$ given by

1. Gauss-Markov

$$f(\omega) = \frac{1}{2\pi\beta^2} \quad e^{-1/2(\omega^2/\beta^2)}$$

2. Laplace

$$f(\omega) = \frac{1}{\sqrt{2}\,\beta} \quad e^{-\sqrt{2}\,|\omega/\beta|}$$

3. Cauchy

$$f(\omega) = \frac{1}{\pi\beta} \quad \frac{1}{1 + (\omega/\beta)^2}$$

The signal or regression sequence is given by $\underline{S}_N = \underline{U}_N$ so that the univariate distribution of the observable $\underline{V}_N = \lambda \underline{S}_N + \underline{W}_N$ is completely describable in terms of a location parameter λ and scale parameter β identical to that of the additive noise sequence $\underline{W}_N$ For the Gauss-

Markov and Laplace processes $\beta^2 = \sigma^2$ the variance of the additive noise component while no such interpretation is available, of course, for the Cauchy process where the variance is infinite. Attention is directed to the bit error probability performance of linear coincidence detectors vis-à-vis linear analog detectors as a function of the ratio $|\lambda|/\beta$ for various sample sizes N and relative sampling rates γ. In particular the suboptimum linear coincidence detectors are considered with smoothing sequence

$$\underline{a}_N = \mu \underline{\underline{K}}_N^{-1} \underline{S}_N$$

and the resulting bit error probability performance is compared with that of the optimum linear detector employing identical smoothing sequence. The normalized covariance matrix $\underline{\underline{K}}_N$ does not exist, of course, for the Cauchy process. Nevertheless, by analogy with the Gauss-Markov and Laplace process it will be convenient in this case also to take

$$\underline{a}_N = \mu \underline{\underline{K}}_N^{-1} \underline{S}_N$$

with $\underline{\underline{K}}_N$ exponential possessing (i, j) element $\rho^{|i-j|}$.

For either the coincidence detector or the analog detector, the appropriate statistic $T_N(\underline{V}_N)$ is formed and the threshold test $T_N(V_N) \gtrless 0$ is performed with the result

i) Decide H_1 if $T_N(\underline{V}_N) > 0$

ii) Decide H_0 if $T_N(\underline{V}_N) < 0$

iii) Decide H_1 with probability 1/2 if $T_N(\underline{V}_N) = 0$

For the Gauss-Markov process it can easily be shown that the bit error probability P_e of the optimum linear detector is given by

$$P_e = 1 - \Phi \left\{ \frac{|\lambda|}{\beta} \langle \underline{S}_N, \underline{\underline{K}}_N^{-1} \underline{S}_N \rangle \right\}$$

where $\Phi(\cdot)$ is the unit Gaussian c.d.f. Since $\underline{S}_N = \underline{U}_N$ and, of course, for this case $\beta = \sigma$:

$$P_e = 1 - \Phi\left\{\frac{|\lambda|}{\sigma} \sum_{i,j=1}^{N} k_{ij}^{-1}\right\}$$

where k_{ij}^{-1} is the (i, j) element of $\underline{\underline{K}}_N^{-1}$. In fact, it can be shown that

$$\sum_{i,j=1}^{N} k_{ij}^{-1} = \frac{1}{1-\rho^2}\left[2 + (N-2)(1+\rho^2) - 2(N-1)\rho\right]$$

so that P_e can easily be evaluated as a function of $|\lambda|/\sigma$, N and γ for the optimum linear detector operating on Gauss-Markov data. In the other case computational difficulties dictate resort to Monte-Carlo simulation. In particular, the sample sizes N = 2, 4, 6, 10 are considered and relative sampling rates γ = 1.0 and 10.0 are used. The results are illustrated in Figures 2-9. Although limited, these results are useful in determining to what extent, if at all, asymptotic comparisons are useful in predicting small-sample behavior and enable some inferences to be drawn on small-sample, large-signal behavior in general. Several aspects of these results deserve comment at this time.

A notable feature concerns the relative performance with increasing relative sampling rates. For each of the cases considered, the differences in performance of the linear and suboptimum coincidence detectors tends to diminish with increasing relative sampling rate. It is anticipated that the performance difference between the optimum linear and coincidence detectors would be even less. This behavior is consistent with asymptotic results obtained previously for the Gauss-Markov process (cf. Fig. 1). A more interesting behavior is exhibited by the relative performance of the linear coincidence detector (which is now locally optimum) vis-à-vis the optimum linear detector operating on Laplace autoregressive data. The linear coincidence detector was shown to have A.R.E. = 2.0 compared to the optimum linear detector and, in fact, to be locally most powerful for independent data. It is expected then that superior bit error probability performance would be obtained by employing the coincidence detector, at least for small SNR and relative sampling rate γ. The results obtained support this conclusion where it seems to be apparently true uniformly in γ while it definitely does not hold uniformly in SNR. In fact, one observes in each case a cross-over point where for SNR below this point the linear coincidence detector exhibits superior performance and decidely inferior

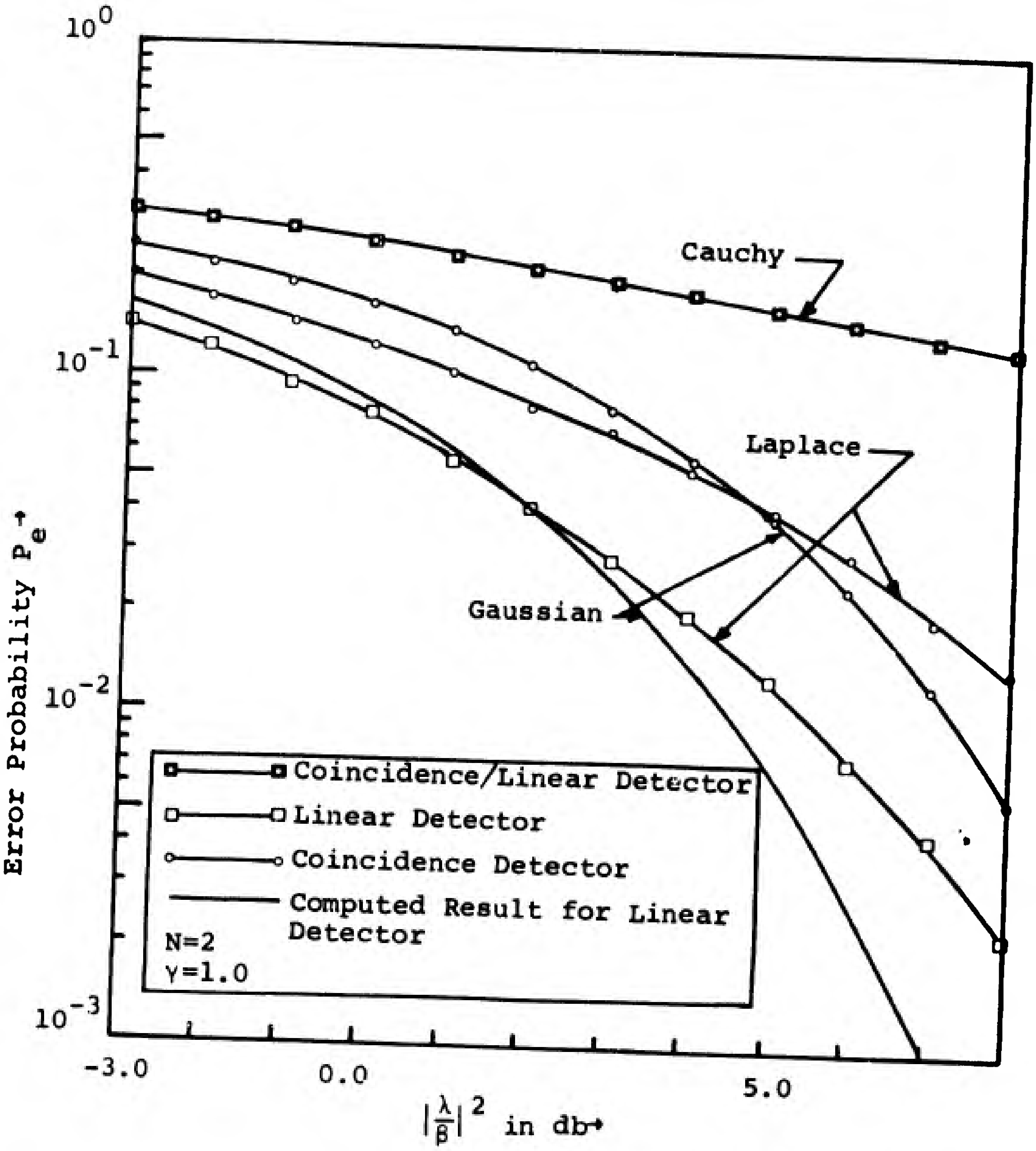


Figure 2

Error Probability vs. $|\lambda/\beta|$ for

N=2 , γ=1.0

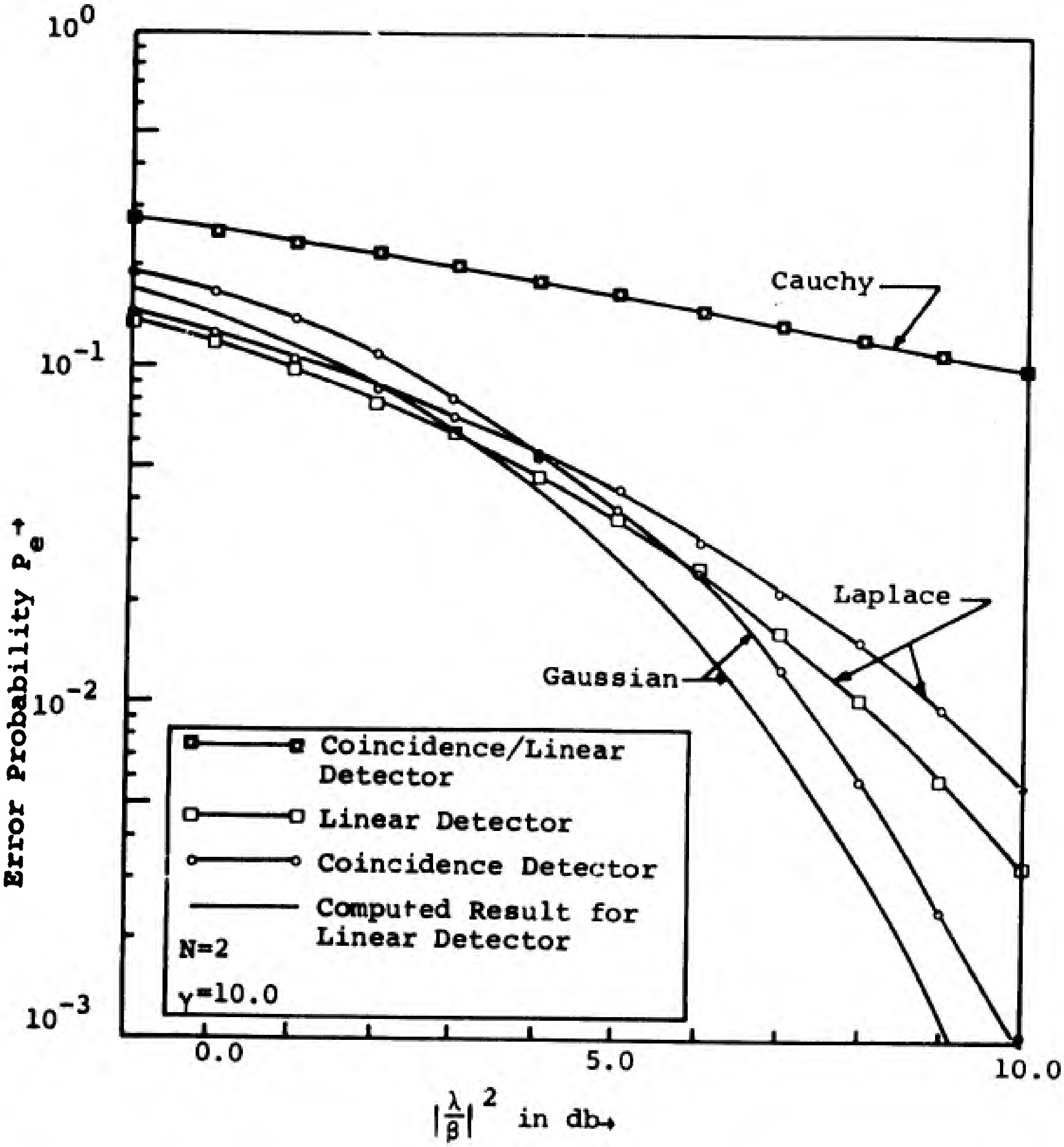


Figure 3

Error Probability vs. $|\lambda/\beta|$ for
$N=2$, $\gamma=10.0$

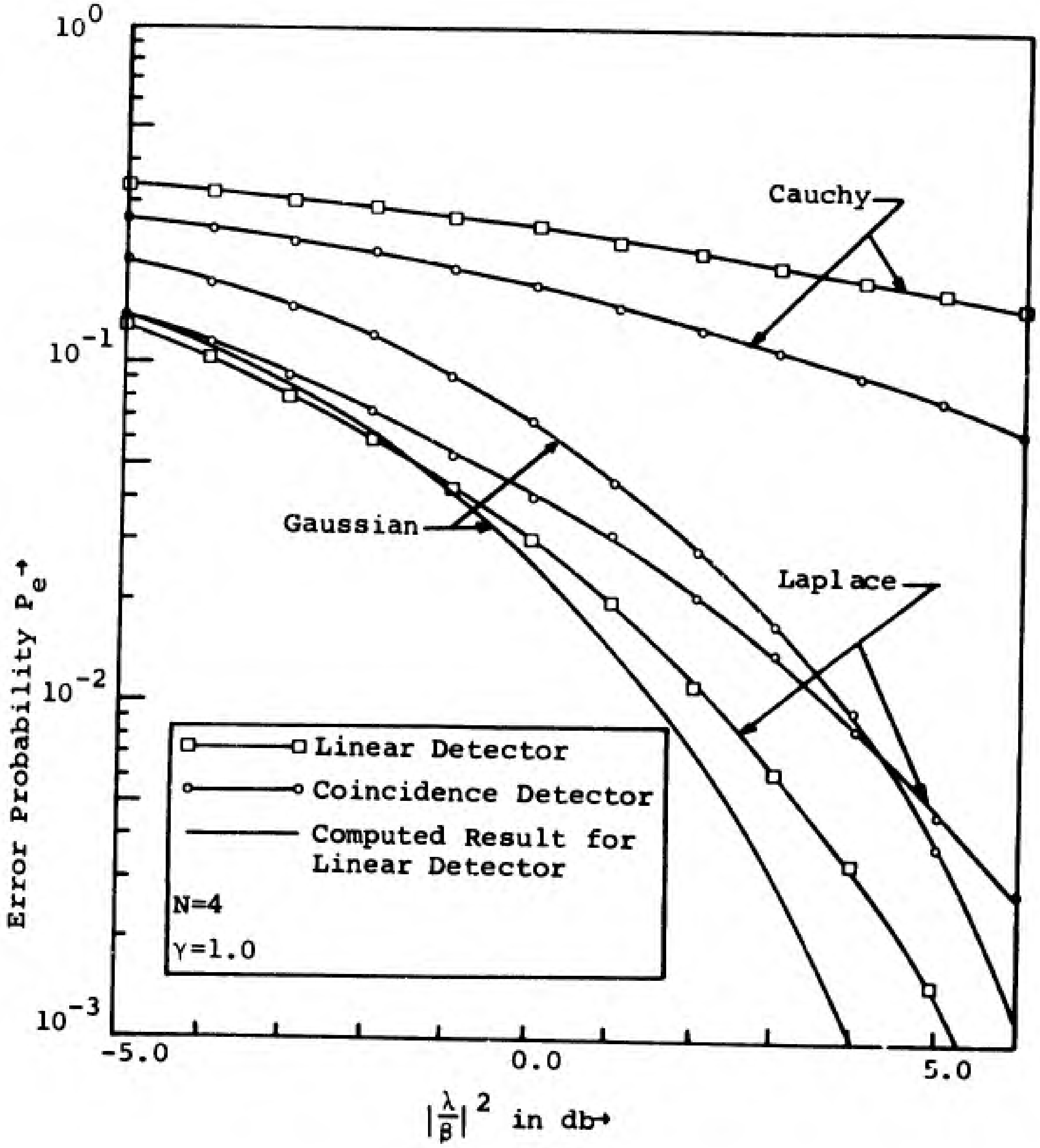


Figure 4

Error Probability vs. $|\lambda/\beta|$ for
N=4 , γ=1.0

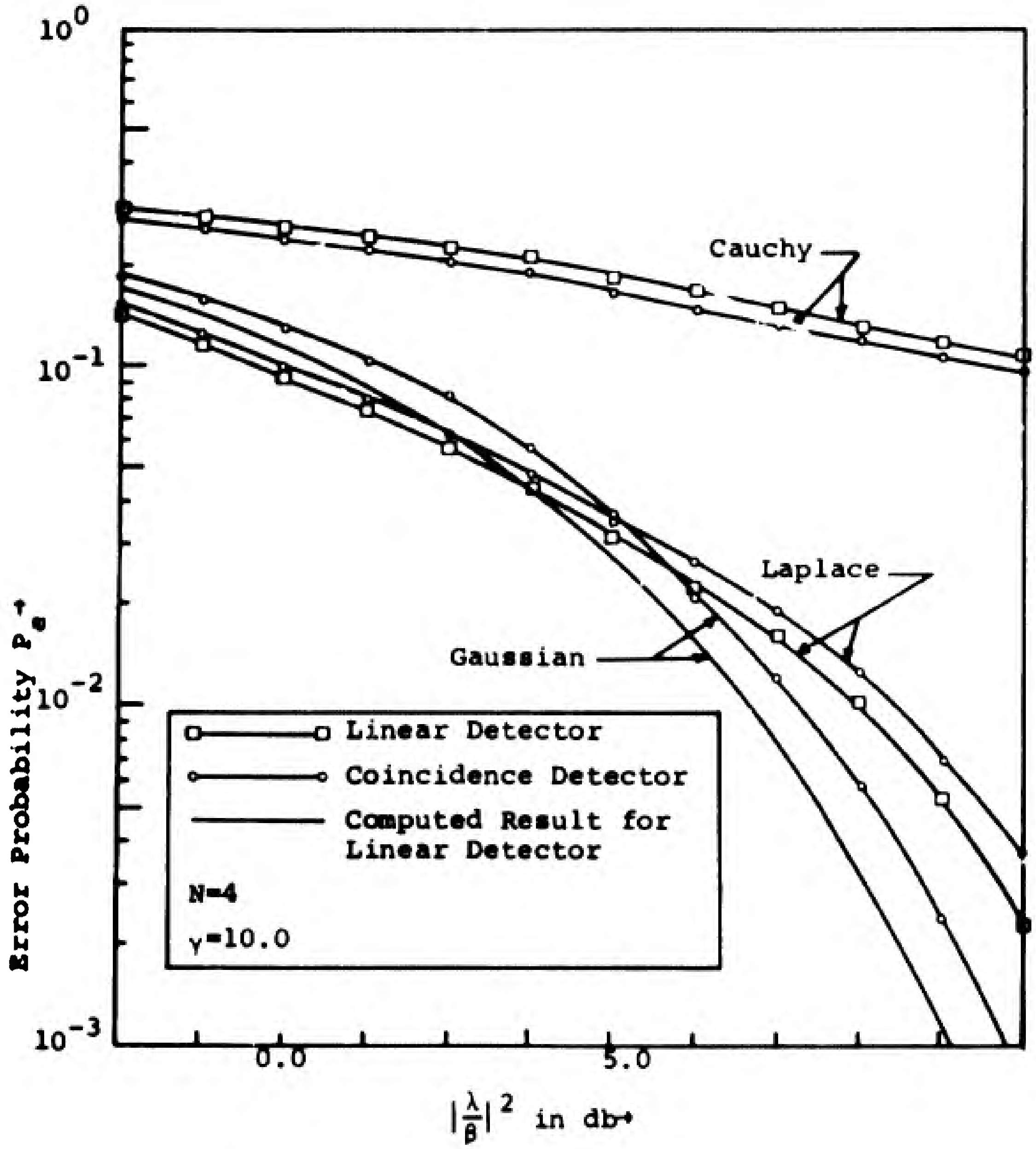


Figure 5

Error Probability vs. $|\lambda/\beta|$ for
N=4 , γ=10.0

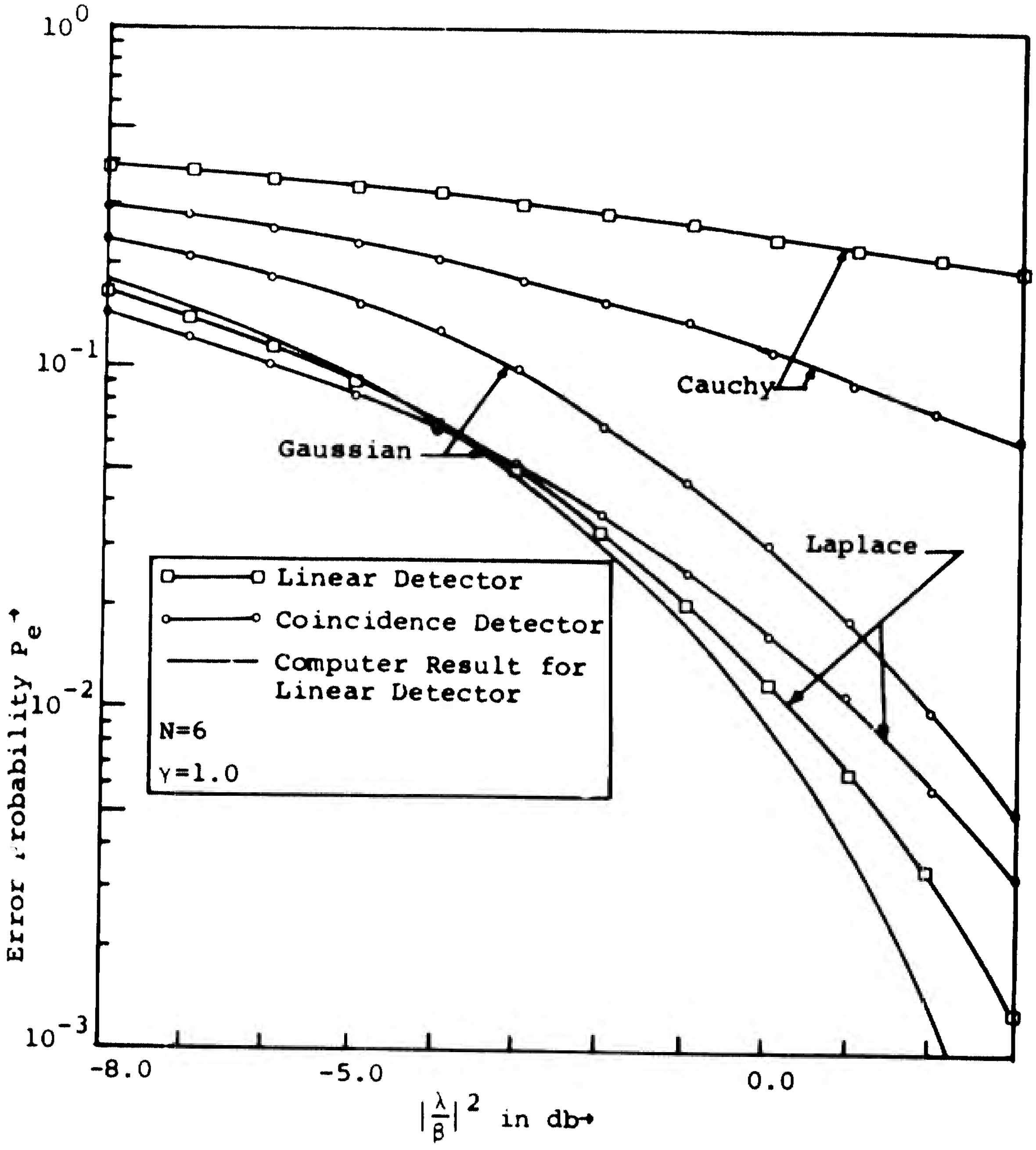


Figure 6

Error Probability vs. $|\lambda/\beta|$ for

$N=6$, $\gamma=1.0$

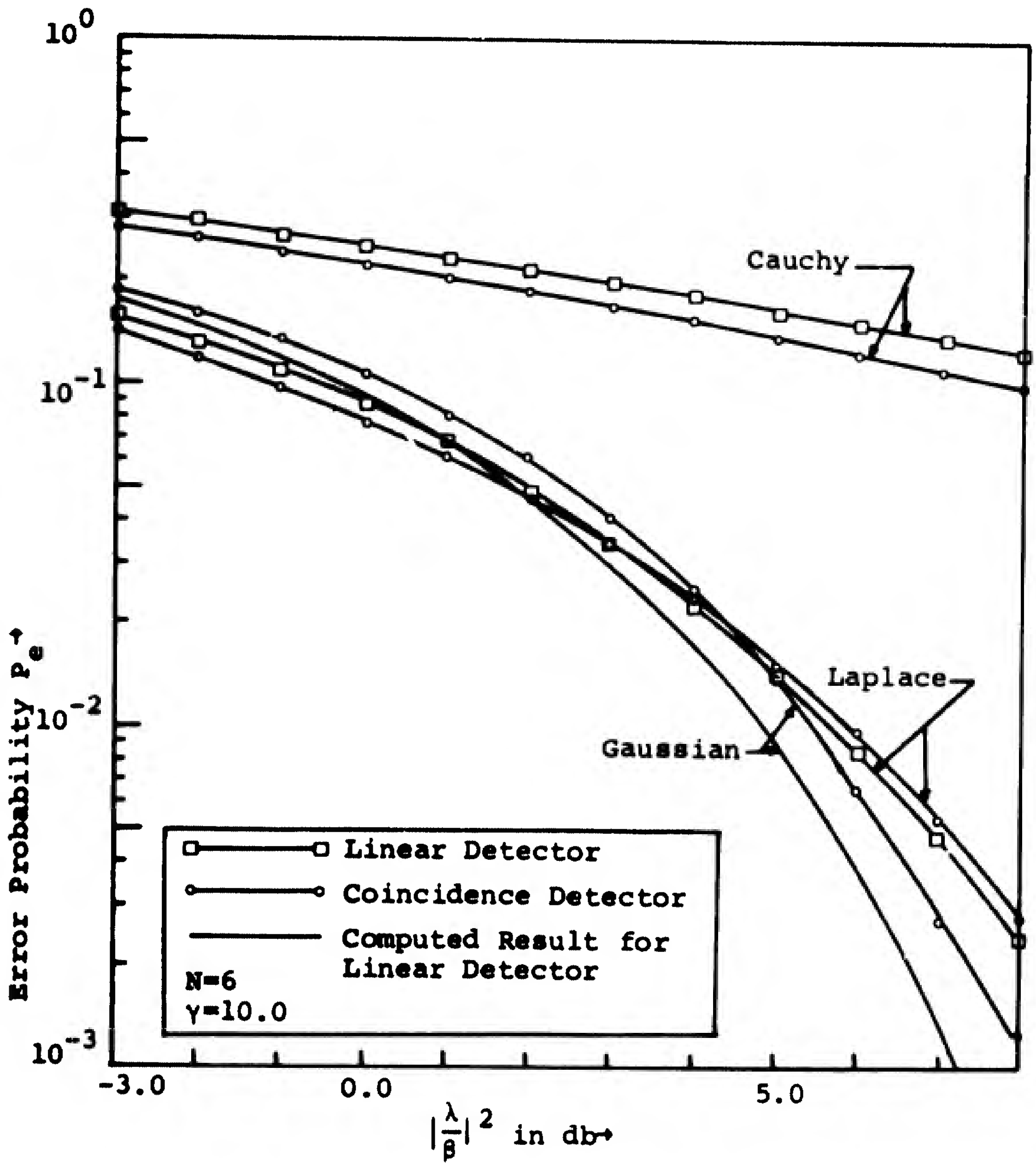


Figure 7

Error Probability vs. $|\lambda/\beta|$ for

N=6 , γ=10.0

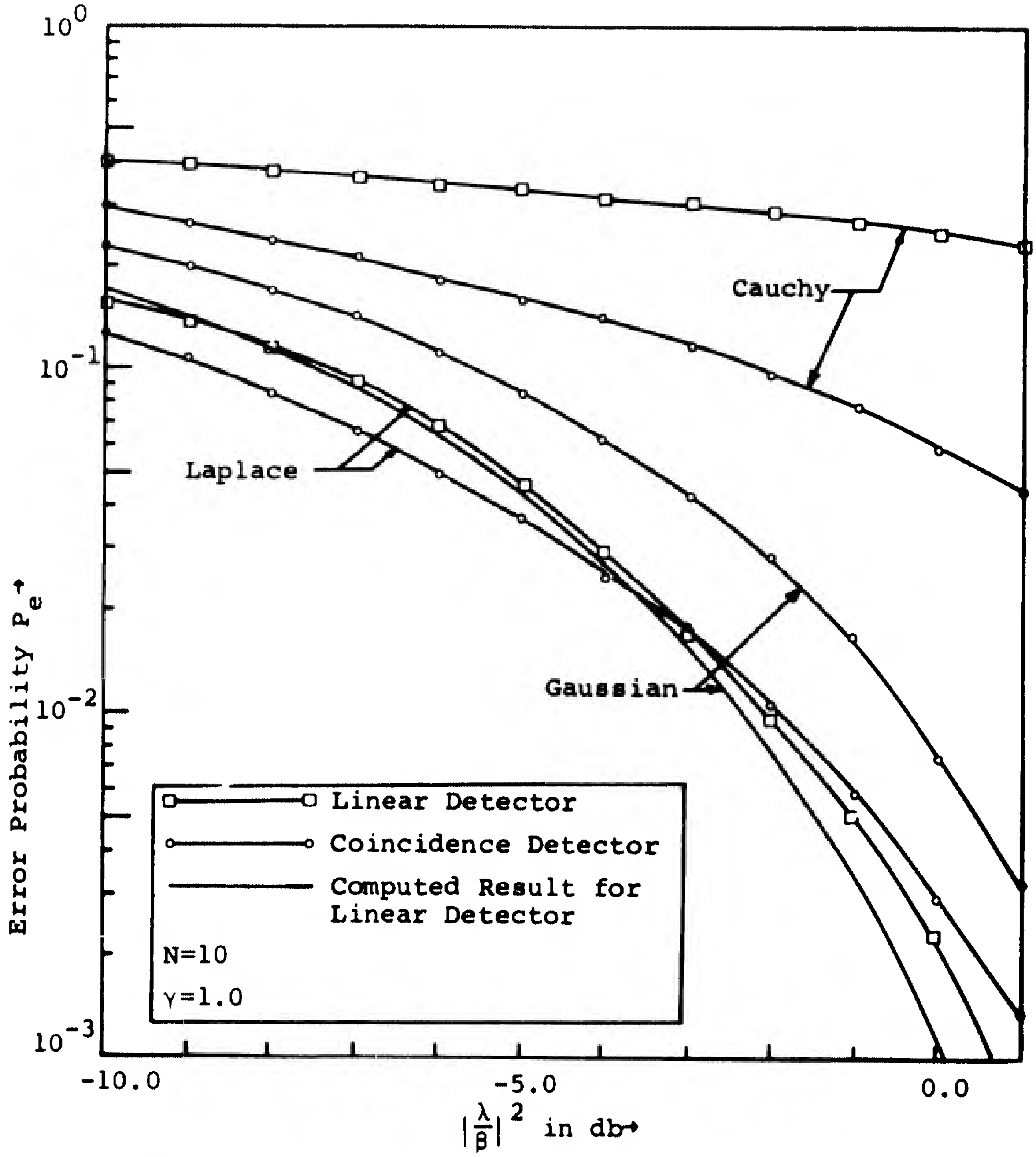


Figure 8

Error Probability vs. $|\lambda/\beta|$ for

N=10 , γ=1.0

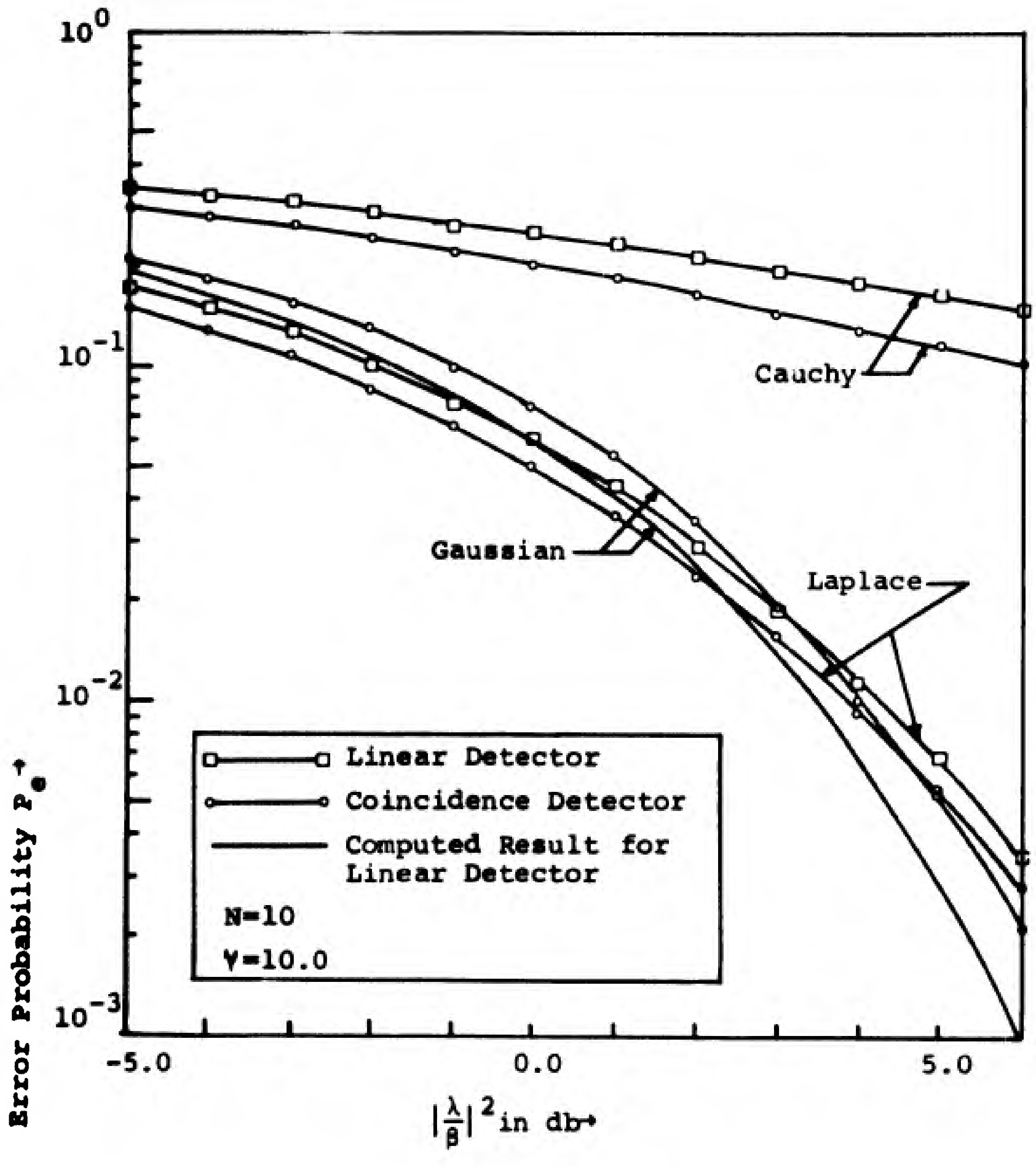


Figure 9

Error Probability vs. $|\lambda/\beta|$ for $N=10$, $\gamma=10.0$

performance at SNR above this point. The cross-over point is clearly in evidence for N = 6, 10 while for smaller sample sizes it occurs at much lower SNR outside the range for which the curves have been plotted. This is certainly the type of behavior which could not be predicted on the basis of A.R.E. It appears, however, that the SNR at which this cross-over point occurs increases monotonically in sample size so that for N in excess of 10 say, the linear coincidence detector provides decidely superior performance at SNR levels producing bit error probabilities useful for data communication purposes (i.e., P_e less than 10^{-2}).

Also to be observed is the increasingly superior performance of the coincidence detector compared to the linear detector operating upon the broad-tailed Cauchy process as the sample size N increases. This is more pronounced for small relative sampling rates. Thus one can observe (cf. Fig. 8) the reasonable performance afforded by the coincidence detector operating on Cauchy data for N larger than 10 say, while the linear detector remains completely paralyzed. For N=2 there is no noticeable differerence in performance of the two detectors. It appears then that for sample sizes ($N > 10$ say) the linear coincidence detector offers considerable advantages over the linear detector in broad-tailed noise environments typical of impulse or jamming noise.

The concept of a locally optimum coincidence detector has been introduced and consideration given to its asymptotic performance. The locally optimum coincidence detector was shown to require parametric knowledge for its implementation. As a result a particular suboptimum coincidence was proposed for the detection of known signals in ill-defined noise backgrounds which does not require parametric knowledge for its implementation. Asymptotic comparisons vis-à-vis the optimum linear detector have been shown to give identical results as the locally optimum coincidence detector. This analogy suggests, among other things, that small-sample, large-signal criteria are the more appropriate performance indicators for the detection situation under consideration. Extensive computer simulation results are given indicating the performance of coincidence detectors in small-sample, moderate-to-large SNR environments. These results suggest the appropriateness of the suboptimum linear coincidence detector in the regime of moderate sample sizes and SNR's suitable for reliable data communication purposes. Considerable protection is afforded against deviations from the almost universal assumption of Gaussian noise distributions. In fact, as we note for the case of Laplace noise, improvements vis-à-vis the optimum linear detector operating on Guassian noise are possible. Nevertheless, one must be willing to trade this feature for the accompanying degradation in bit error probability if Gaussian noise is in fact present. As an example, for N = 10, $\gamma = 1.0$, one observes from Fig. 8 that the suboptimum linear coincidence detector requires approximately 2 db more SNR to achieve $P_e = 10^{-2}$ as the optimum linear detector.

Subject to this limitation, then, the present results indicate that for moderate sample sizes the suboptimum linear coincidence detector offers an attractive alternative to the linear detector for the detection of known signals in uncertain noise environments. Obviously more extensive simulation results of this nature will be useful in presenting a more complete picture of the behavior of linear coincidence detectors.

SECTION III

OPTIMUM RANK TEST

Suppose one channel of observation is available and the sample values $v(\Delta t)$, $v(2\Delta t), \ldots, v(N\Delta t)$ are taken. Let the signal appear as one of two d.c. levels, $+\lambda$ or $-\lambda$, imbedded in additive sample-to-sample independent Gaussian noise with mean zero and variance one. It is well known that the optimum test compares the sample mean

$$S = \frac{1}{N} \sum v(i\Delta t)$$

with a threshold. As noted in the Introduction, a rank test is based on the vector of $\pm$ ones, $Z = (z_1, z_2, \ldots, z_N)$, where $z_i = +1$ if the i'th smallest sample value in magnitude is positive and $z_i = -1$ if the i'th smallest sample value is negative. An example would be the following for $N = 4$:

i	$v(i\Delta t)$
1	-2
2	-1
3	4
4	1.5

i	z_i	
1	-1	$(\lvert v(2\Delta t)\rvert = 1)$
2	+1	$(\lvert v(4\Delta t)\rvert = 1.5)$
3	-1	$(\lvert v(\Delta t)\rvert = 2)$
4	+1	$(\lvert v(3\Delta t)\rvert = 4)$

$$Z = (-1, +1, -1, +1).$$

Based on the rank vector Z, one of two decisions (hypotheses) are to be made:

H_0: signal $-\lambda$ was sent

H_1: signal $+\lambda$ was sent

Let $p_-(Z)$ and $p_+(Z)$ be the probabilities that the vector Z is observed due to the combination of signal and noise when $-\lambda$ and $+\lambda$ are sent respectively. The optimum rank test is based on the likelihood ratio

$$\Lambda = \frac{p_+(Z)}{p_-(Z)}$$

The likelihood ratio is compared with a threshold, Λ_0. If $\Lambda > \Lambda_0$, the hypothesis H_1 that $+\lambda$ was sent is accepted (which means that the probability that Z occurs when $+\lambda$ is sent is more than Λ_0 times greater than when H_0 is sent).

If $\Lambda < \Lambda_0$, H_0 is decided and if $\Lambda = \Lambda_0$ a random decision is made between H_0 and H_1. If $\Lambda_0 = 1$ and $\pm\lambda$ are a priori equally likely, the probability o. error is minimized.

In Appendix II the detailed computations are presented for the optimum rank test when C = 1 for large amplitude signals, $\lambda >> 1$ and small sample sizes. For N = 2 to 6 samples the optimum test follows:

N	Decide H_1, $+\lambda$ sent, if	Decide H_0, $-\lambda$ sent
2,3	$z_N = +1$	otherwise
4	1. $z_4 = +1$, $\sum_{i=1}^{4} z_i \geq 0$	otherwise
	or 2. $z_4 = -1$, $\sum_{i=1}^{4} z_i = 2$	
5	1. $z_5 = +1$, $\sum_{i=2}^{5} z_i \geq 0$	otherwise
	or 2. $z_5 = -1$, $\sum_{i=1}^{5} z_i = 3$	
6	1. $z_6 = +1$, $\sum_{i=3}^{6} z_i \geq 0$	otherwise
	or 2. $z_6 = +1$, $\sum_{i=1}^{6} z_i \geq 0$	
	or 3. $z_i = -1$, $\sum_{i=3}^{6} z_i = 2$,	
	$z_1 + z_2 \geq 0$	

Needless to say, the optimum rank test cannot be simply implemented although it is not prohibitive. In the next section a more practical test is investigated and found to be almost as good as the best rank test.

An important question is that of how good the optimum rank test is compared with the optimum test on the original data. This is answered in Appendix II where it is stated that the signal energy never has to be more than 4/3 greater for the optimum rank test than for the optimum test to attain the same probability of error.

SECTION IV

THE WILCOXON RANK SUM TEST

The Wilcoxon test is based on the sum of the signed ranks of the N observations:

$$S_w = \sum_{i=1}^{N} z_i \, i$$

where $z_1, z_2, \ldots, z_N$ are defined in Section II. S_w is compared with a threshold C. If $S_w > C$, H_1 is decided; otherwise H_0 is decided. This is a suboptimum rank test when the noise is Gaussian. However, it is stated in Appendix III that the Wilcoxon test is almost as good as the optimum rank test of the preceding section for sample-to-sample independent Gaussian noise. To be more precise, the Wilcoxon test requires a signal energy that is only $3/\sqrt{8} \approx 1.06$ greater than the optimum rank test for the same probability of error and a signal energy that is $\sqrt{2}$ greater than that required by the optimum test.

The above result is for independent observations. It is well-known that for a large number of independent small amplitude observations, the Wilcoxon test requires a signal energy that is only $\pi/3 \approx 1.04$ greater than the optimum test. In Section V this behavior is verified for dependent normal observations. In the remainder of this section it is shown that for some small sample nonGaussian, dependence conditions, the Wilcoxon test is almost as good or better than the optimum test for dependent normal observations.

The investigations are most easily carried out by simulations. The N sample values are generated using the following procedure. The first noise sample, $w(\Delta t)$, is generated with a random number generator in the computer followed by a transformation to obtain the desired noise probability density. Each succeeding sample value is generated by the dependence model:

$$w(i\Delta t) = \rho w((i-1)\Delta t) + \eta(i\Delta t), \quad i = 2, \ldots, N$$

where $\eta(2\Delta t), \eta(3\Delta t), \ldots, \eta(N\Delta t)$ is a sequence of independent random variables generated by the computer with a desired probability density. The noise values are added to the signal to form the observed sequence $v(\Delta t) = \pm\lambda + w(\Delta t)$, $v(2\Delta t) = \pm\lambda + w(2\Delta t), \ldots, v(N\Delta t) = \pm\lambda + w(N\Delta t)$. The Wilcoxon signed rank sum statistic is formed as described above. This is compared with the thresholds and a decision of $\pm\lambda$ made. This procedure is then repeated many times with a count of the errors kept to determine the error probability.

Figures 10-12 present the probabilities of error under various conditions. In Figs. 10 and 11, the noise is Gaussian and $\rho = 0.9$. It is seen that the probability of error is nearly the same for the Wilcoxon as for the optimum test for various numbers of samples and signal amplitudes. In Fig. 12 the noise has a Cauchy density with $\rho = 0.5$ for which the optimum normal test performs very poorly and for which the Wilcoxon test is significantly better.

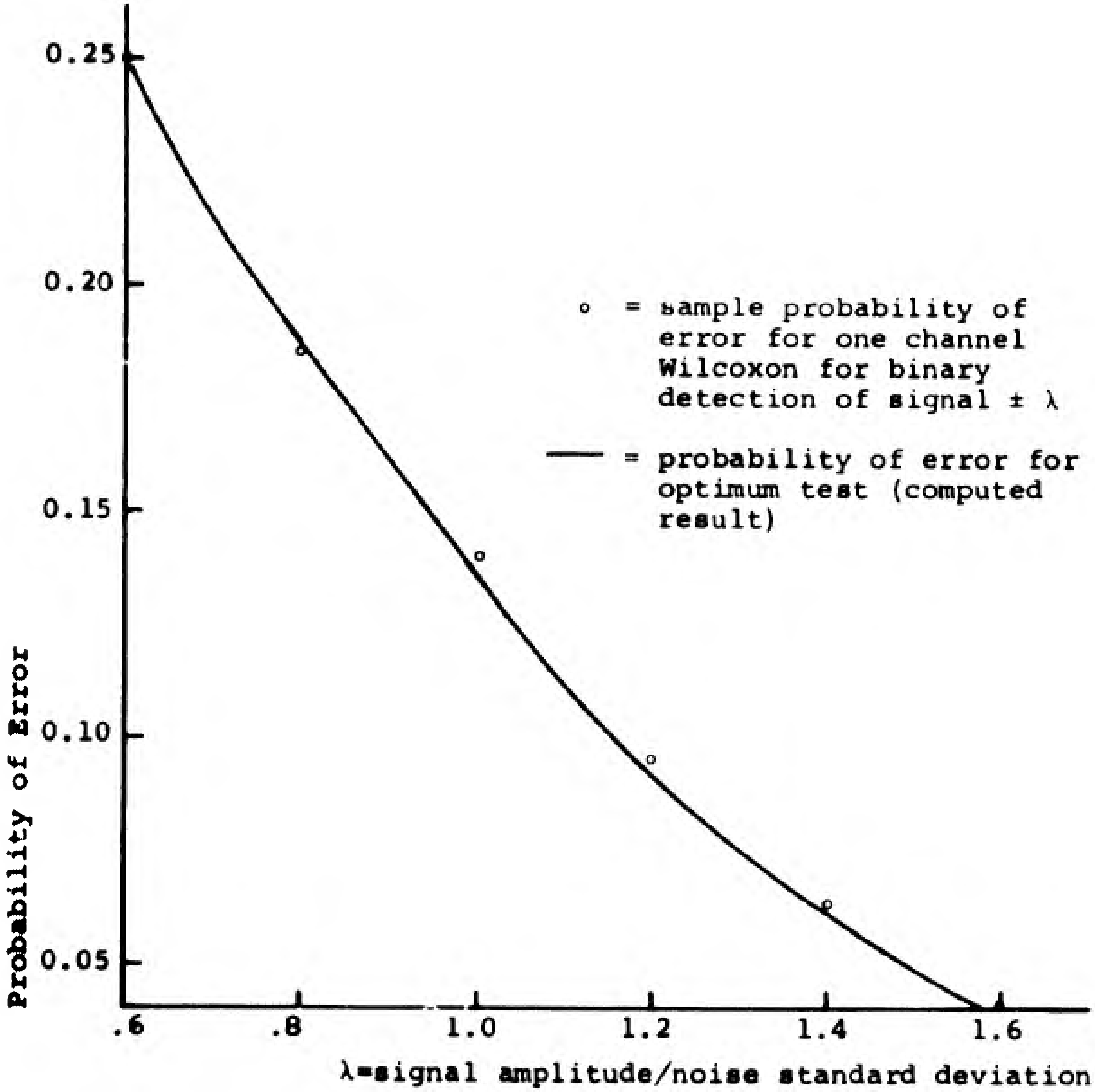


Figure 10. Probability of Error with 5 Dependent Samples vs. Signal-to-Noise Ratio Gaussian Noise $\rho = .9$

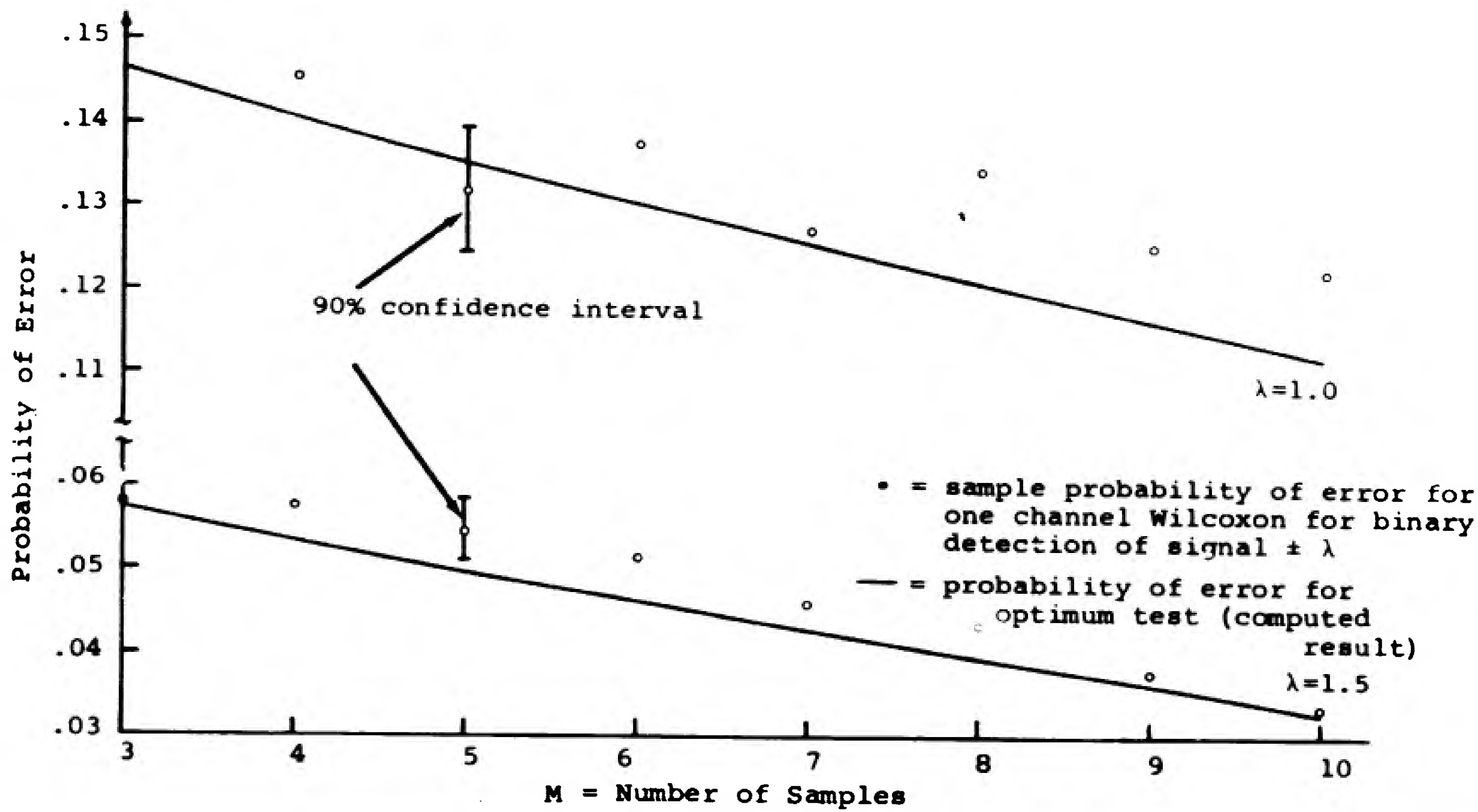


Figure 11. Error Probability for Binary Detection With Dependent Samples vs. Number of Samples Used. Gaussian Noise, ρ = .9

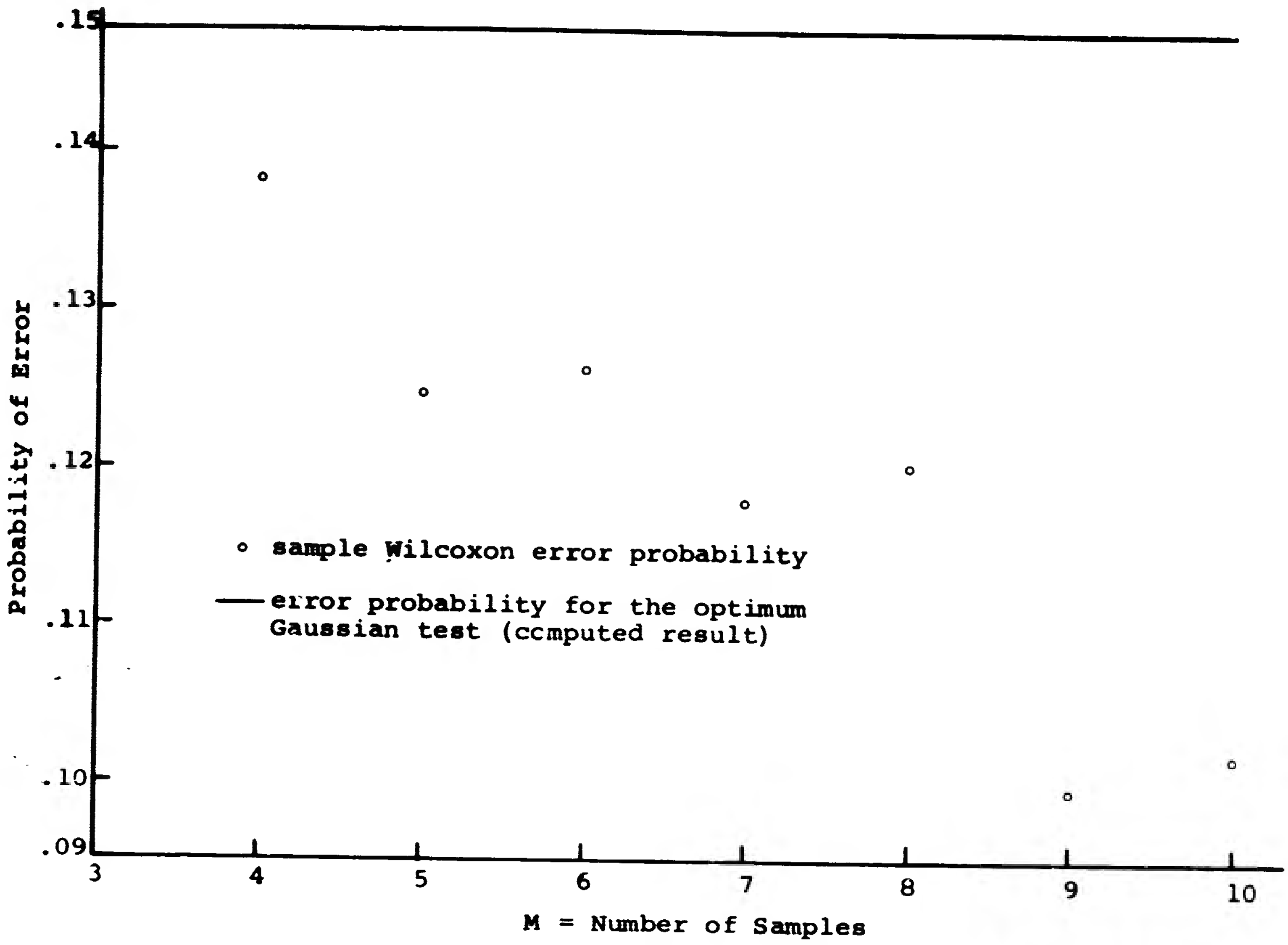


Figure 12: Error Probability for Cauchy Noise, $\lambda = 2.0$, $\rho = .9$

SECTION V

THE EFFECTS OF DEPENDENCE OF NONPARAMETRIC RANK TESTS

Although the research in this study concentrated on small sample sizes, it is nonetheless of interest to investigate the performance of nonparametric tests for large sample sizes under dependence. The reason for using large sample sizes is the relative ease of obtaining analytical results. If a nonparametric test performs well under dependence for large sample sizes, it should then perform well for small sample sizes. Comparison between the results of this section and the preceding indicate that this is true at least for the Wilcoxon test. Similar conclusions were made in Section II for coincidence tests.

In this section basically three nonparametric rank tests are considered. The first is the Wilcoxon test described in the preceding section. The second is the Mann-Whitney two channel rank test. Suppose that two channels of observation are available rather than only one with sample values $v_1(\Delta t)$, $v_1(2\Delta t),\ldots,v_1(N\Delta t)$ and $v_2(\Delta t),\ldots,v_2(N\Delta t)$ respectively. The first channel contains signal plus noise and the second channel contains noise only. The Mann-Whitney test is performed by comparing with a threshold the sum of the ranks of the signal channel observations when both channels are ranked together. This is most conveniently represented as:

$$S_{mw} = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{sgn}(v_1(\Delta t) - v_2(\Delta t))$$

where sgn is the sign function,

$$\mathrm{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$$

The third test considered in this section is the Kendall τ test. This test applies when two channels of observation are available as in the preceding paragraph, each containing the common signal plus channel-to-channel independent noise. The test is performed by comparing with a threshold the correlation between signs of sample value differences:

$$S_\tau = \sum_{i \neq j} \mathrm{sgn}[v_1(i\Delta t) - v_1(j\Delta t)]\, \mathrm{sgn}[v_2(i\Delta t) - v_2(j\Delta t)]$$

A particular variant of the above three statistics called a mixed statistic is also investigated in this section. Suppose the N observations (N paired observations if a two channel detector is used) are divided into subsets, each containing m observations (m paired observations if two channels). On each of these subsets a statistic of one of the above forms is generated as may be applicable to the particular problem at hand. If the number of subsets is denoted by p, the result of this procedure is p numbers, $S_1, S_2, \ldots, S_p$ corresponding to each of the subset statistics. The mixed statistical test is performed by comparing the sum of the p statistics,

$$S = \sum_{i=1}^{p} S_i,$$

with a threshold. The reason for doing this rather than forming the nonparametric statistic on all the data is that a double sum of N values requiring a total of N^2 additions is replaced by p double sums of m values requiring a total of $pm^2 = mN$ additions. Thus a reduction in complexity is attained. The loss in test effectiveness can be quite small for small values of m as shown in [13] for independent sampling and in this section under dependence.

Several ways of selecting the subset groupings are possible, two of which are considered here. The first, called method one, forms the subsets sequentially in real time. The disadvantage of this procedure is that for rapid sampling (small Δt), ranking effectiveness is lost due to sample dependence. This will be demonstrated later in numerical results. The second grouping, called method two, takes the observations p samples apart so that samples within each set become independent as the total number, N, of observations (and hence p) becomes large.

The detailed computations appear in Appendix III. For the purposes of presenting numerical results, a dependence model as in the preceding section was chosen. If $w_1(\Delta t), w_1(2\Delta t), \ldots, w_1(N\Delta t)$ and $w_2(\Delta t), \ldots, w_2(N\Delta t)$ are the noise samples in the two channels of observation (in the case of the Wilcoxon test only the first set is used), $w_1(i\Delta t)$ and $w_2(i\Delta t)$ satisfy

$$w_k(i\Delta t) = \rho(\Delta t) w_k((i-1)\Delta t) + \eta(i\Delta t),$$

$$k = 1,2 \qquad i = 2,3,\ldots,N.$$

Here the parameter $\rho(\Delta t)$ depends on the sampling interval:

$$\rho(\Delta t) = e^{-|\Delta t|}$$

Thus closer sampling means more highly dependent sample values.

In the case of the Kendall τ statistic the signal is random as well as the noise. The signal is taken to have correlation function

$$\rho_s(\Delta t) = (1 + \Delta t)e^{-|\Delta t|}$$

Figures 13-15 present the asymptotic relative efficiencies for the Wilcoxon, Mann-Whitney and Kendall τ statistics respectively as compared with the optimum test on the continuous data ($\Delta t = 0$) as a function of Δt. Also shown is the asymptotic relative efficiency of the optimum sampled data detector with respect to the optimum detector on continuous data. It is seen that the Wilcoxon and Mann-Whitney detectors are almost as good as the optimum sampled data detector for any sampling interval without too much further loss in efficiency when a mixed statistic with m = 10 values in each subset and the second grouping method is used. If the first method of grouping is used, however, there is a loss in A.R.E. as $\Delta t \to 0$ as predicted above. The Kendall τ detector is not as close to the optimum detector. However, this is also true under independence where it is well known that the asymptotic relative efficiency is $9/2\pi^2$ [13]. However, again the mixed statistic works well.

It is known that many nonparametric statistics including the Wilcoxon, Mann-Whitney and Kendall τ are asymptotically normally distributed when independent sampling is used. In Appendix III the theoretical details are given to establish asymptotic normality under dependence for certain of these statistics. While this is primarily of theoretical interest, in practice for large sample sizes, threshold and error probability calculations are considerably simplified through the normality assumption as the exact distribution of the statistics are difficult, if not impossible, to obtain under dependence except possibly through simulations.

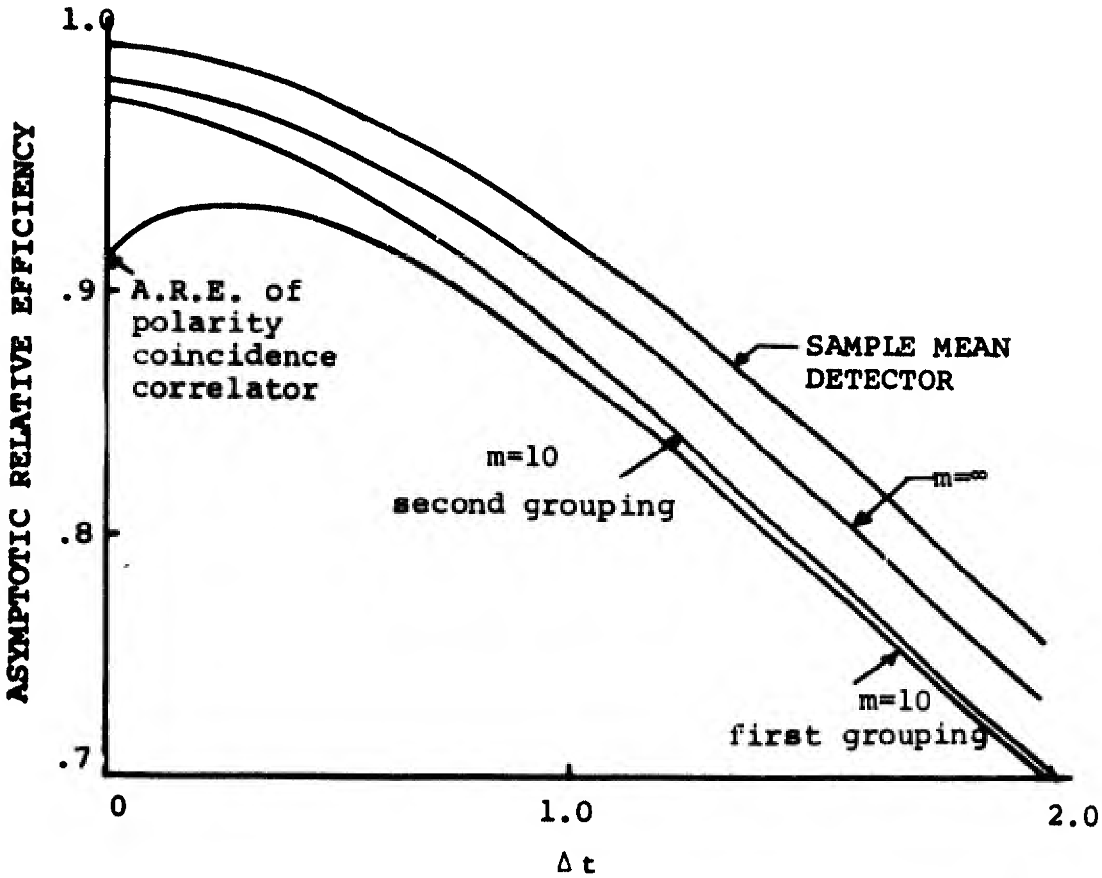


Figure 13. The Asymptotic Relative Efficiency with respect to a parametric test on continuous data for various Mann-Whitney Detectors, $\rho_n(t)=e^{-|t|}$, Δt=sampling interval.

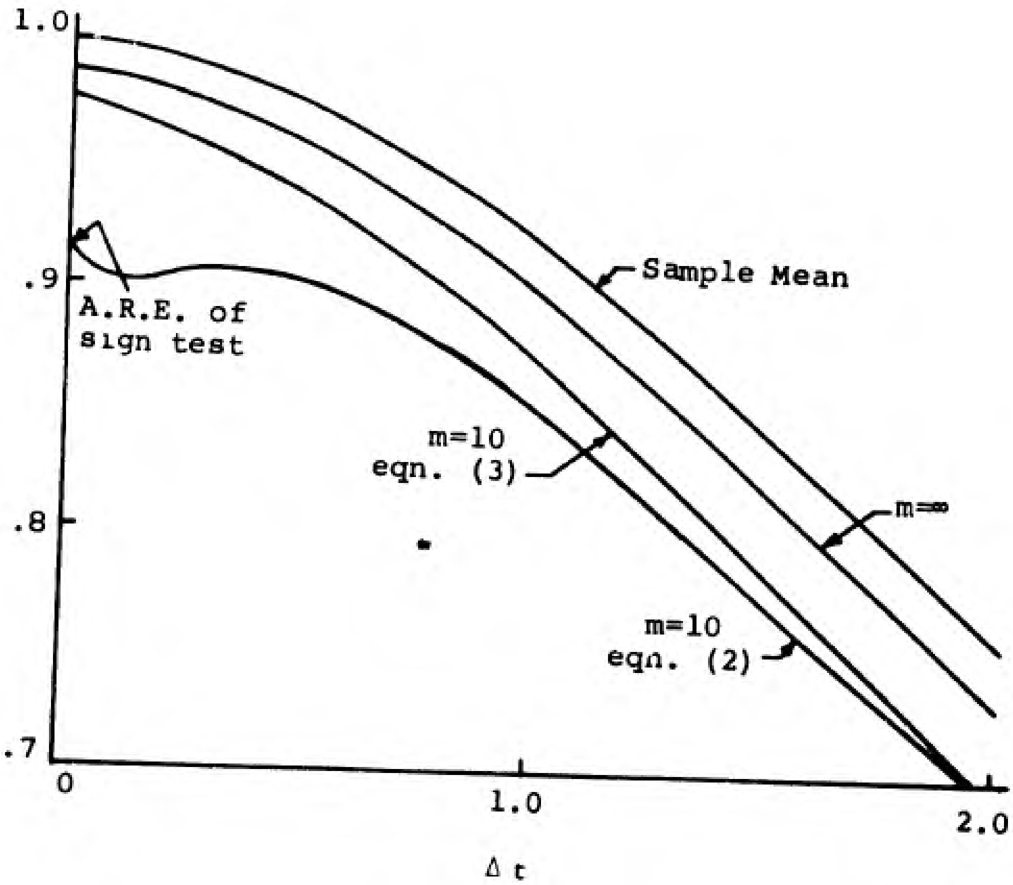


Figure 14. The Asymptotic Relative Efficiency with respect to a parametric test on continuous data for various Wilcoxon Detectors, $\rho_n(t)=e^{-|t|}$, Δt=sampling interval.

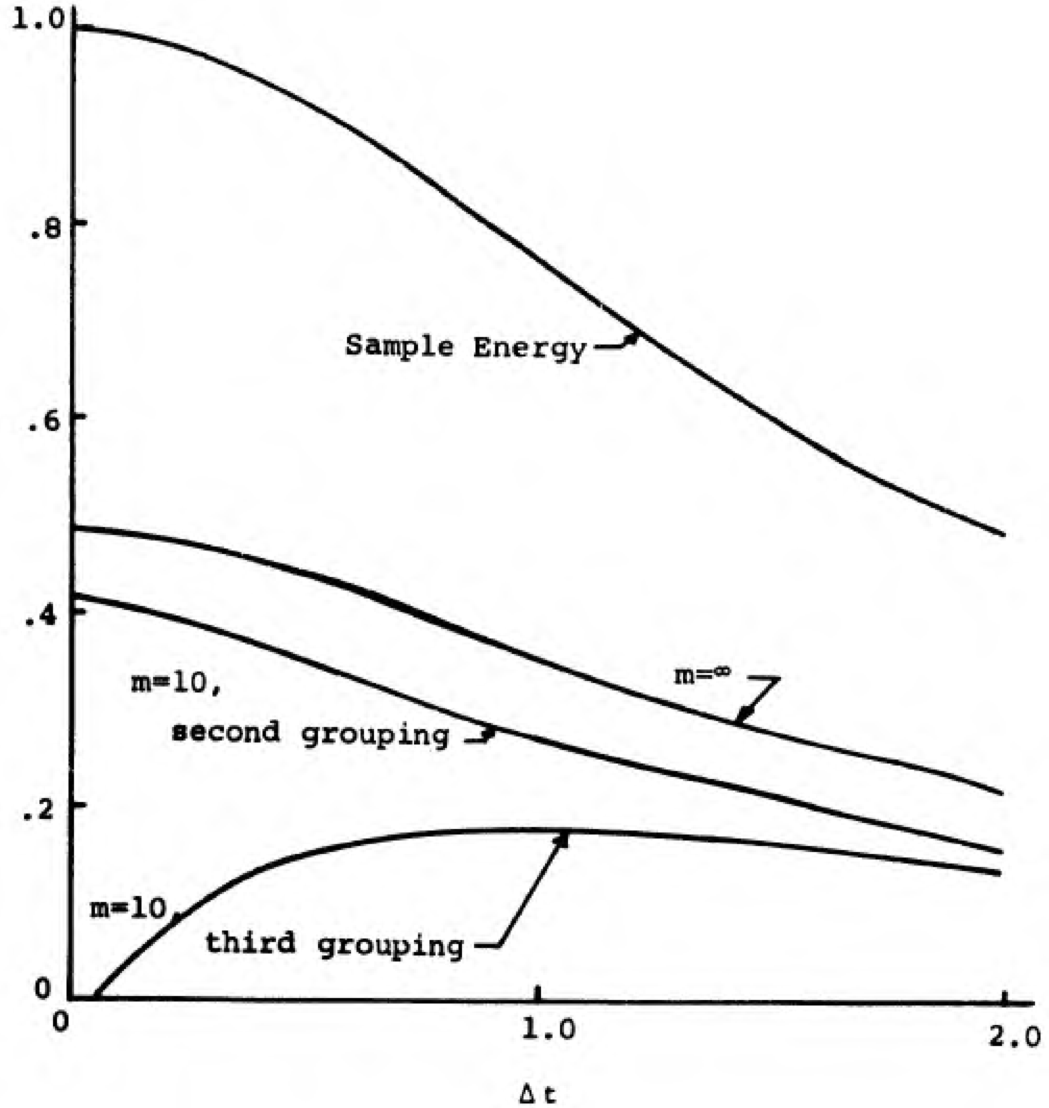


Figure 15. The Asymptotic Relative Efficiency with respect to a parametric test on continuous data for various Kendall τ Detectors, $\rho_n(t)=e^{-|t|}$, $\rho_s(t)=(1+|t|)e^{-|t|}$, Δt=sampling interval.

SECTION VI

ROBUST DETECTION PROCEDURES FOR SIGNALS OF KNOWN FORM IN NEARLY GAUSSIAN NOISE

In recent years considerable work has been done in the area of nonparametric detection. An obvious shortcoming to the use of nonparametric tests in many situations is that the assumed family of possible underlying distributions is too broad. For example, if one suspects that the underlying noise distribution is "nearly" Gaussian in some sense, with zero median, then one would expect to pay a stiff penalty for assuming that the family consists of all zero median distribution functions. The penalty paid is to be expected from the rather severe data reduction involved in nonparametric procedures, and is reflected in a relatively complex detector and/or a low asymptotic relative efficiency when the distribution is actually Gaussian.

On the other hand, the parametric approach, with the usual assumption of Gaussian noise, also has obvious shortcomings. It is known that an optimum statistical procedure based on the Gaussian assumption may perform poorly for seemingly mild deviations from the Gaussian model [14].

It therefore seems desirable to establish tests which lie somewhere between parametric tests and nonparametric tests. Such tests have been described by statisticians as robust tests. Generally speaking, a robust test is one whose performance does not deteriorate too seriously for small deviations from a nominal model. One does not expect a robust test to perform as well as the optimal test based on the nominal model when the distribution fits the model exactly; however, it should perform nearly as well. Also, the robust test should perform better than the parametric test for a large subset of the distributions which are near (according to some distance measure) to the nominal model. The test should generally perform better than a nonparametric test of equivalent complexity.

In this section is investigated the structure and performance of robust detectors for signals of known form in additive nearly Gaussian noise. By nearly Gaussian noise it is meant that the distribution which describes the actual situation may be anywhere in a well-defined neighborhood of the assumed Gaussian distribution. In particular, we use the neighborhood defined by the mixture, or ε-contaminated, model. This model is given by:

$$F(x) = (1-\varepsilon)N(\mu,\sigma) + \varepsilon H(x), \quad 0 \leq \varepsilon < 1$$

where $N(\mu,\sigma)$ is the Normal c.d.f. with mean μ and standard deviation σ, and $H(x)$ is an arbitrary distribution, reflecting possible differences between the theoretical and actual situations.* This is one of the models of

---

* Such a model, with $H(x) = N(\mu,\sigma\varepsilon)$, was considered by Tukey in connection with the robust estimation of location as early as 1949 [14].

uncertainty used by Huber [15], who has provided the theoretical framework upon which the robust procedure is based.

It is easy to see that some typical nearly Gaussian distributions (e.g., [16] the density function $f(x) = K \exp(-|x|^c/a)$, $c < 2$, but c close to 2) may be written in the form above by a suitable choice of parameters. In addition a very broad family of distributions is defined which are, in some sense, nearly normal for small ε. Thus, if a detector is robust for a family defined above, one may feel confident in using the detector in situations for which it is known that the noise is nearly normal but for which no detailed knowledge of the distributions is available.

In this section N samples from one channel of observation are taken:

$$v(i\Delta t) = \lambda + w(i\Delta t), \qquad i = 1,2,\ldots,N$$

where λ is a constant signal and the $w(i\Delta t)$ are independent, identically distributed, nearly normal noise samples. From these sample values it is desired to detect the presence or absence of a signal when the signal amplitude is known (H_0: $\lambda = 0$ against H_1: $\lambda = \lambda_0 > 0$) and the detection when the signal amplitude is unknown (H_0: $\lambda = 0$ against H_1: $\lambda > 0$). In either case it will be shown that the robust detector involves the use of a soft limiter.

Consider first the case where $\lambda_0 > 0$ is known and where the $\{w(i\Delta t)\}$ are independent, identically distributed normal random variables, N(0,1), i.e., the noise distribution is nominally the unit normal distribution. The choice of $\sigma = 1$ is only a matter of convenience. Huber's results [15] may be applied directly to this problem.

Huber's Results. Suppose that p_0 and p_1 are the probability densities of the observations $v(\Delta t), v(2\Delta t), \ldots, v(N\Delta t)$ when no signal is present and a signal is present, respectively. Then the robust test is performed by comparing a test statistic $T_N(V)$ with a threshold where

$$T_N(V) = \sum_{i=1}^{N} T(v(i\Delta t))$$

and

$$T(v) = \begin{cases} c' & p_1/p_0 \le c' \\ p_1/p_0 & c' < p_1/p_0 < c'' \\ c'' & p_1/p_0 \ge c'' \end{cases}$$

If $T_N(V)$ is greater than the threshold the decision that a signal is present is made. If it is smaller than the threshold, no signal is decided. The test is then robust in a min-max sense, with guaranteed minimum power (probability of detection) and a guaranteed maximum false alarm rate (probability of erroneously deciding a signal is present). The question as to whether or not the test has the desired robustness qualities mentioned in the Introduction remains to be determined by actual evaluation of the detector performance.

Huber's results can now be applied to the problem of testing $H_0:\lambda = 0$ against $H_1:\lambda = \lambda_0$. The densities p_0 and p_1, corresponding to H_0 and H_1, respectively, are:

$$p_0(v) = K\, e^{-x^2/2}$$

$$p_1(v) = K\, e^{-(v-\lambda_0)^2/2}$$

An individual term of the probability ratio is given by

$$\frac{p_1}{p_0} = \frac{p_1(v)}{p_0(v)} = e^{[v\theta-(\lambda_0^2/2)]}$$

Each term is to be limited at c' and c" which is equivalent to limiting the logarithm of the individual terms at log c' and log c". With log c' = a and log c" = a" the test becomes:

If $T_N(V) > C$, decide signal present.

If $T_N(v) < C$, decide no signal present

where

$$T_N(V) = \sum_{i=1}^{N} T(v(i\Delta t))$$

$$T(v) = \begin{cases} a' & v\theta - (\lambda_0^2/2) \le a' \\ v\lambda_o - (\lambda_0^2/2) & a' < v\lambda_o - (\lambda_0^2) < a'' \\ a'' & v\lambda_o - (\lambda_0^2) \ge a'' \end{cases}$$

Thus the test statistic may be formed by passing the sequence

$$v(i\Delta t)\lambda_0 - (\lambda_0^2/2)$$

through a limiter with values a' and a", and then summing. Alternatively, the data samples, $v(i\Delta t)$, may be passed through a limiter having values $AL = a'/\lambda_0 + \lambda_0/2$ and $AU = a''/\lambda_0 + \lambda_0/2$, and then summing. The two forms of the detector are shown in Figs. 16a and 16b. Different thresholds are used in the two forms of the detector. It is easy to show that the thresholds are related by

$$\frac{C}{\lambda_0} + N\frac{\lambda_0}{2} = C' \quad .$$

The limiter values are obtained by solving the following equations for a' and a" in terms of the unit normal c.d.f., $\Phi(x)$, as:

$$1 - \Phi[(a'/\lambda_0) - (\lambda_0/2)] + e^{a'}\Phi[(a'/\lambda_0) + (\lambda_0/2)] = \frac{1}{1-\epsilon}$$

$$\Phi[(a''/\lambda_0) + (\lambda_0/2)] + e^{-a''}[1 - \Phi(a''/\lambda_0 - \lambda_0/2)] = \frac{1}{1-\epsilon}$$

One is interested in solutions of these equations such that a' < a", these being solutions corresponding to c' < c". It is easily seen, noting the symmetry of the unit normal c.d.f., that if a" is a solution of the second equation, then -a" is a solution of the first. Also, the first equation is increasing in a' and the second is decreasing in a". These properties imply that the condition a' < a" is equivalent to a' < 0 < a", or simply -<u>a</u> < 0 < <u>a</u>, where <u>a</u> = a" is the solution to the second equation. Thus the quantity

$$v(i\Delta t)\lambda_0 - \lambda_0^2/2$$

is passed through a <u>symmetrical</u> **limiter** as shown in the figure. The data samples, $v(i\Delta t)$, however, are passed through a limiter with values AL and UL (Fig. 16b) and this limiter is not symmetrical; in fact, the limiter with values $AL = a'/\lambda_0 + \lambda_0/2$ and $AU = a''/\lambda_0 + \lambda_0/2$ is a limiter whose center of symmetry is located at $\lambda_0/2$, thereby limiting extreme values of both signal and no signal distributions to the same extent. The scale factor $1/\lambda_0$ results in limiting $v(i\Delta t)$ by the same relative amount that $\lambda_0 v(i\Delta t)$ is limited by the limiter with values a' and a".

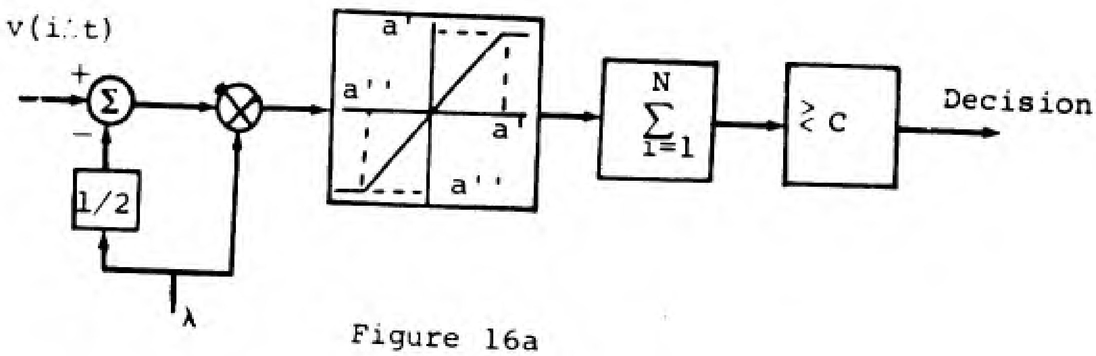


Figure 16a

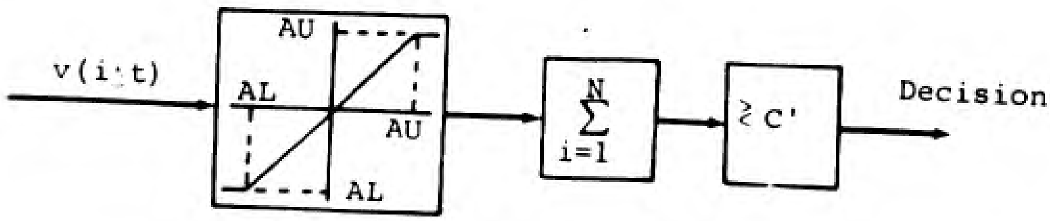


Figure 16b

Figure 16. Forms of the Robust Detector

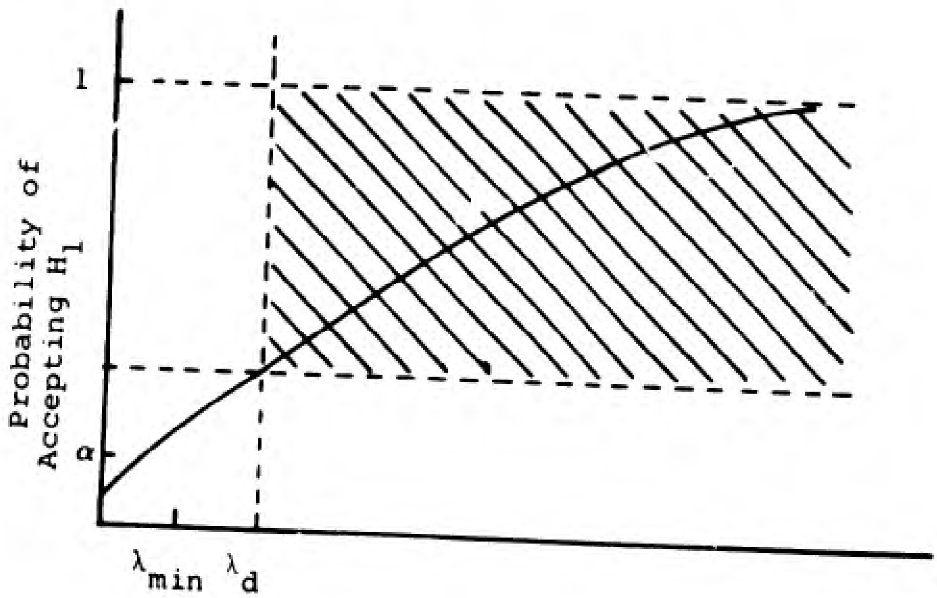


Figure 17. Probability of Accepting H_1 vs. Signal Amplitude

It is interesting to consider what happens in the limit of large and small values of λ_0 and ε. For the case where λ_0 is small, consider the behavior of the normalized statistic

$$T_N(V) = \sum_{i=1}^{N} [T(v(i\Delta t)) + \underline{a}]/2\underline{a}$$

In this case $\underline{a} \to 0$ in the limit and the statistic is equal to the number of times the observed data points are positive. In other words the test becomes the **sign test**. For the case where λ_0 is fixed and $\varepsilon \to 0$ we have $\underline{a} \to \infty$; in other words, there is no limiting at all. The test becomes the sample mean. This is certainly to be expected since in this case the noise is actually Gaussian and the sample sum is the optimum Neyman-Pearson solution. The remaining case is that for which ε is fixed and $\lambda_0 \to \infty$. In this case the limiter values approach fixed numbers, $-\underline{a}_\infty$ and $+\underline{a}_\infty$. The observations, $v(i\Delta t)$, are limited at $a_\infty/\lambda_0 + \lambda_0/2$ and $-a_\infty/\lambda_0 + \lambda_0/2$ which are both approximately $\lambda_0/2$ for large λ_0. In this case the test does not become the sample sum as might be hoped for. This simply reflects the fact that the problem is not a parametric one for $\varepsilon \to 0$ no matter how large λ_0. The reason that it is not parametric is that the ε part of the mixture can have its mass arbitrarily far out.

The choice of limiter values in the preceding depends explicitly upon the actual value of λ_0, where λ_0 was assumed to be known. However, one is frequently interested in detecting a constant signal of unknown amplitude and appropriate limiter values need to be determined for this case.

Huber has a result which may be applied to this problem [15]. Suppose $p_\lambda(v)$ is a family of densities having a monotone likelihood ratio in v. For a given fixed ε we assume λ to be greater than some minimum value, namely $\lambda > \lambda_{min}$ to prevent the classes of probability densities under signal and no signal conditions from overlapping. Then Huber's result is that the test can be designed for any $\lambda_d > \lambda_{min}$ with the assurance of a guaranteed minimum power for all $\lambda \geq \lambda_d$ along with a guaranteed maximum false alarm rate.

In other words, the power and false alarm are guaranteed to lie in the shaded regions of Fig. 17. This lower bound on the power might be adequate if λ_d and the sample size are such that the region $\lambda \geq \lambda_d$ is one of high power. However, one is generally interested in a lower bound on the power throughout the range of power. In this case the lower bound given by Huber is not adequate in the sense that it is not tight and does not hold for $0 < \lambda < \lambda_d$. It is possible, however, to find the sharpest lower bound on the power, which is denoted by $B_{\lambda_d} = B_{\lambda_d}(\lambda)$, for all $\lambda > 0$. The general form of this sharpest lower bound is given by the solid curve of Fig. 17. This bound always coincides with the power at $\lambda = \lambda_d$.

The question now arises as to what value of λ_d one should actually use. What λ_d gives the best overall performance in some sense? It appears that there is no analytical solution to this question, i.e., there exists no nice ordering relation for the tests based on various λ_d. Therefore, the selection of λ_d, which yields the limiter values, must be based on numerical evaluation of detector performance for various λ_d. The test statistic

$$T_{N_{\lambda_d}}(V) = \sum_{i=1}^{N} T_{\lambda_d}(v(i\Delta t))$$

generally has a complicated nonnormal distribution for finite N. However, for moderately light limiting and moderate signal levels the test statistic will be essentially normal for reasonably small sample sizes. Hence, calculations are made for moderately large N, assuming that $T_{N_{\lambda_d}}$ is normal.

These calculations are straightforward, but tedious. With $\varepsilon = 0.01$, false alarm rate $\alpha = 0.005$ and $N = 200$, Fig. 18 displays the minimum power over all possible noise distributions for various λ_d, the power of the robust test with $\lambda_d = 0.06$ is actually Gaussian, and the power and false alarm rate of the sample mean test for the Gaussian and for the two ε-contaminated densities,

$$q_\lambda(v) = \frac{(1-\varepsilon)}{\sqrt{2\pi}} e^{-(v-\lambda)^2/2} + \frac{\varepsilon}{\sqrt{2\pi}\;\sigma_\varepsilon} e^{-(v-\lambda)^2/2\sigma_\varepsilon^2}$$

$$\sigma_\varepsilon = 10, 30.$$

The design value $\theta_d = 0.06$ yields a nearly best minimum power curve. The corresponding limiter values are AL = −0.56, AU = +0.62. It is to be noted that although the $\lambda_d = 0.06$ curve is not uniformly best, it falls only slightly below the $\lambda_d = 0.03$ curve in the small signal region. These curves reflect the intuitive feeling that moderate limiting will result in a best robust test since harder limiting results in too much data reduction and lighter limiting results in too much sensitivity to tail behavior of the distribution. Actually, the $\lambda_d = 0.10$ curve results in nearly as good a performance; the corresponding limiter values are AL = −0.82 and AU = +0.92. This suggests

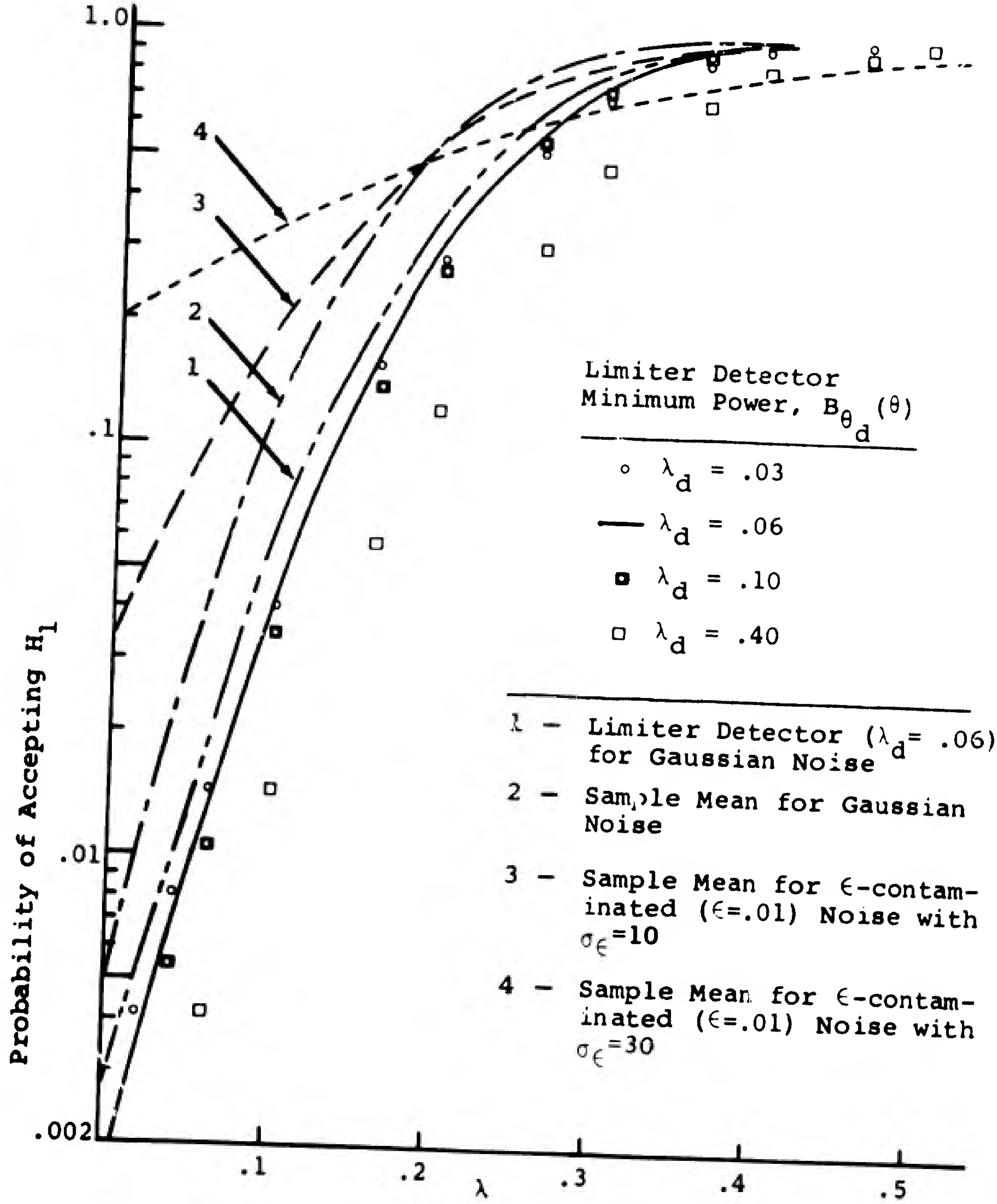


Figure 18. False Alarm and Power for Limiter Detector and Sample Mean Detector

that, in general, the minimum power will be about the same for $1.2 \lesssim$ AU-AL $\lesssim 1.8$.

Note that the behavior of the sample mean detector for the ε-contaminated models is very poor in terms of false alarm rate, e.g., for $\sigma_\varepsilon = 10$ the false alarm has increased by a factor of 7 over that where $\varepsilon = 0$, whereas the robust detector has $\alpha \leq 0.005$ for any ε contamination. This extremely good protection against false alarm variability is obtained at a moderate cost in power.

An evaluation of the performance of the detector of Fig. 18 which is referred to as a limiter can be made in terms of its A.R.E., $e_{\ell,m}$, with respect to the sample mean test. An additional comparison of interest can be made with the A.R.E., $e_{s,m}$, of the sign detector, the equivalent nonparametric test, with respect to the sample mean test.

Since Huber's model results in an overlap of hypothesis and alternate for small signals, the minimum A.R.E. may be zero. Thus one cannot, in this general situation, hope for a good lower bound on the A.R.E. This suggests restricting our attention to alternate densities which are translated versions of hypothesis densities, i.e., densities of the form $q_\lambda(v) = q_0(v-\lambda)$. This restriction is made and in addition it is assumed that the mean is zero, the variance is fixed at σ_n^2, and $\partial q_\lambda(x)/\partial\lambda$ exists, and is continuous for almost all v in some λ neighborhood of zero, there exists $G(v) \geq 0$ integrable, such that $[|\partial q_\lambda(v)/\partial\lambda|] \leq G(v)$ in some λ neighborhood of zero. These restrictions may result in a rather thin subset of the model, but a number of typical deviations from normality are still included in this subset.

Let $T = \ell$ represent the limiter-detector, $T = m$ the sample mean detector, and $T = s$ the sign detector. Also, let $\ell_{AL}^{AU}(\cdot)$ represent the functional form of the limiter with values AL and AU. Then it can be shown that [17]

$$e_{\ell,m} = \left[\int_{AL}^{AU} q_0(x)\,dx\right]^2 \frac{\sigma_n^2}{\sigma_\ell^2}$$

where σ_ℓ^2 is the variance of ℓ_{AL}^{AU} under no signal conditions.

For symmetrical distributions having a density function which is continuous at the origin, the A.R.E. of the sign detector with respect to the sample mean is given by [17]:

$$e_{s,m} = 4\,\sigma_n^2\; q_0 2(0)$$

$e_{\ell,m}$ is evaluated for two limiters, one with AL = -0.56 and AU = +0.62 (corresponding to ε = 0.01 and λ_d = 0.06 in the previous section), and one with AL = -1.3 and AU = +1.3. The models of nearly normal noise for which $e_{\ell,m}$ and $e_{s,m}$ are evaluated are:

$$\text{a)}\quad q_0(v) = \frac{(1-\varepsilon)}{\sqrt{2\pi}}\, e^{-v^2/2} + \frac{\varepsilon}{\sqrt{2\pi}\,\sigma_\varepsilon}\, e^{(-v^2)/(2\sigma_\varepsilon^2)}$$

$$\sigma_\varepsilon = 5,\; 0 \le \varepsilon \le 0.1$$

and

$$\text{b)}\quad q_0(v) = \frac{c}{2^{(1/c)+1}\,\Gamma(1/c)}\, e^{-|v|^c/2}\,, \quad 1 \le c \le 2$$

The results, shown in Fig. 19 clearly indicate the robustness of the limiter detector. For example, with $q_0(v)$ given by the above, and limiter values of AL = -0.56 and AU = +0.56, $e_{\ell,m}$ is 0.82 for ε = 0, strictly increasing with ε and equal to 2.3 for ε = 0.10. In addition $e_{\ell,m}$ is approximately 25% to 30% higher than $e_{s,m}$ for $0 \le \varepsilon \le 0.10$.

The results presented in this section indicate that the soft limiter detector is robust for detecting a constant signal in noise which is known to be nearly normal but for which little detailed knowledge is available. It appears that the power of the robust detector is relatively insensitive to limiter values within a broad region, though this question needs further study. Another open question is whether or not a robust solution exists for the problem in which the hypothesis class is given by an ε-contaminated family while the alternate class is a translated version of the hypothesis class. A solution to this problem would not be encumbered by the overlap of hypothesis and alternate in the small signal region.

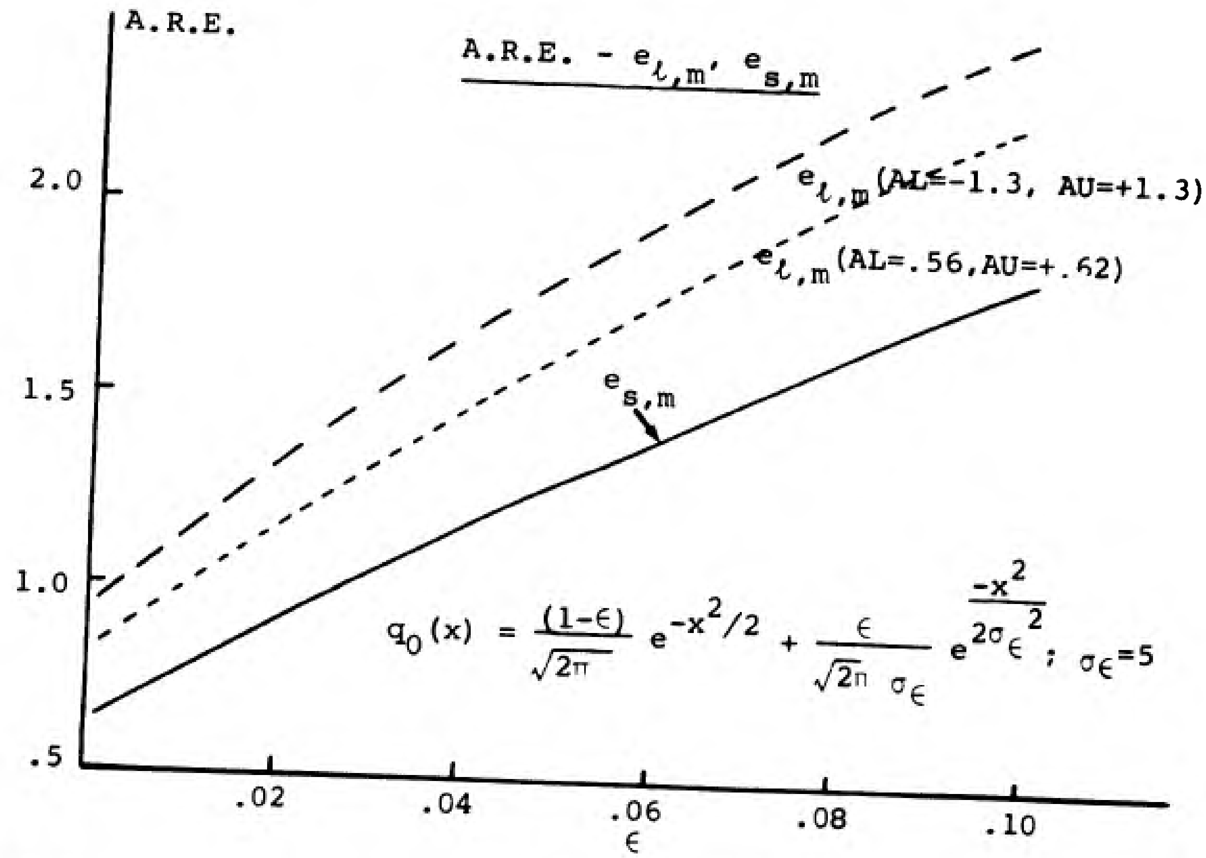


Figure 19a

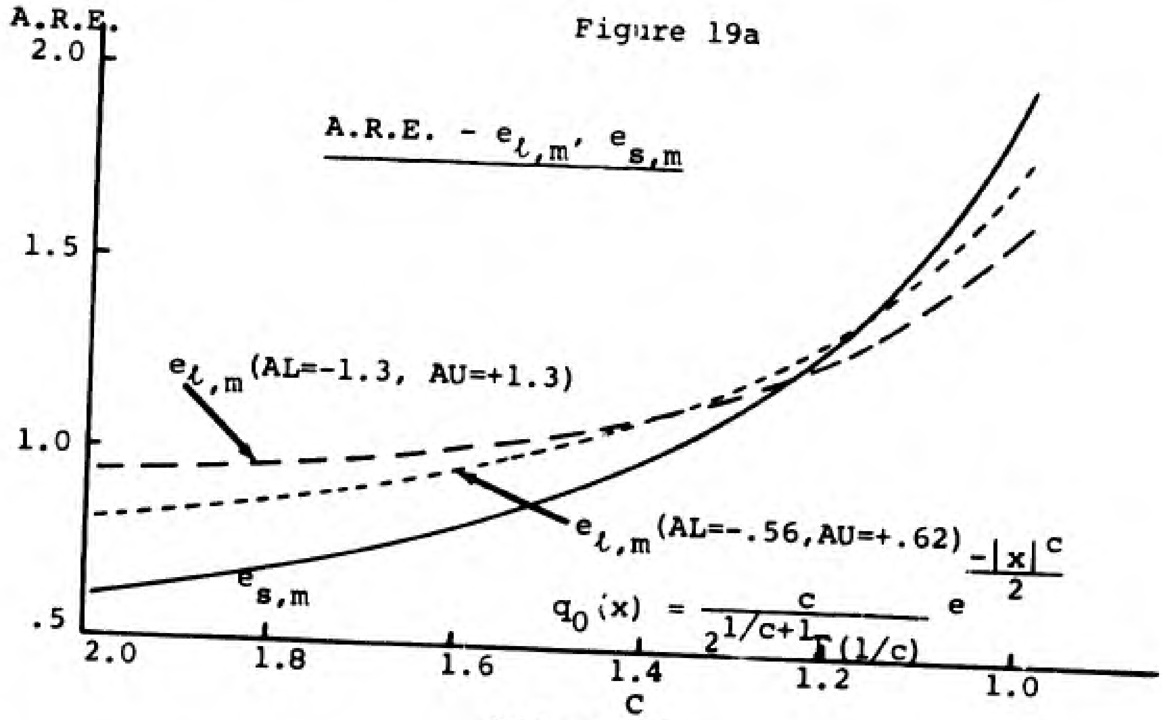


Figure 19b

Figure 19. Asymptotic Relative Efficiency of the Robust Detector

SECTION VII

A DECISION DIRECTED SCHEME FOR THRESHOLD ADJUSTMENT

In many applications the signal levels appear with an a priori probability distribution which is unknown. Thus the levels are usually assumed equally likely which is the minimax solution (the level of performance is the same as in the worst case no matter what the distribution is). To do better than this, one would like to design a procedure for adjusting the signal decision threshold so that the probability of error is minimized. Hopefully this procedure could be applied to a nonparametric detector where the noise distributions are unknown as well. In addition it should be possible to develop procedures which take advantage of dependencies in successive signal transmissions. To study the potential of such methods, threshold adjustment in independent Gaussian noise is considered where each signal transmission is independent of the past.

Suppose the signal appears as one of two levels, $+\lambda$ or $-\lambda$ with (unknown) a priori probabilities p and 1-p respectively and the Gaussian noise has zero mean and unit variance. A signal is generated every T seconds. The optimum detector decides that at time k, $+\lambda$ was sent if the k'th observation, $v(n+k)$, corresponding to the k'th signal transmission, exceeds the threshold given by:

$$\frac{1}{2\lambda} \ln \frac{1-p}{p}$$

Otherwise it is decided that $-\lambda$ was sent. If $p = 1/2$, the threshold is zero, the usual minimax solution. If $p > 1/2$, the threshold is negative, biasing the test in favor of the more likely positive observations.

A simple, easily implemented, intuitively pleasing solution to the problem of estimating the prior probability p to do better than the minimax solution is found by using the relative frequency of occurrence of the positive signal level based on the decisions made up to the present. That is, the estimate of p at time k+1 is taken to be

$$\hat{p}_{k+1} = \frac{1}{k} \sum_{i=1}^{k} u\left(v(it) - \frac{1}{2\lambda} \ln \frac{1-p_i}{\hat{p}_i} \right)$$

where u(x) is the unit step function,

$$u(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

This can be put in the form of a difference equation:

$$\hat{p}_{k+1} = \hat{p}_k - \frac{1}{k}\left[\hat{p}_k - u(v(kt) - \frac{1}{2\lambda} \quad \ell n \frac{1-\hat{p}_k}{\hat{p}_k})\right]$$

Because the current estimate of the probability p depends on the past decisions, this is called a decision directed scheme. Two questions arise immediately in the analysis of this method:

1. Can the procedure "run away" so that the estimate converges to zero or one? Conceivably it might since a series of decisions in the same direction could so bias the detector that it would continue in the same direction with a high probability so that eventually $\hat{p} = 1$ or 0.

2. If the estimate does not "run away," does it converge to p as $n \to \infty$? Conceivably it might converge to a biased estimate due to the "bootstrap" nature of the technique or not converge at all.

It turns out that the answer to the first question is that the procedure can run away but the probability is small for moderate signal-to-noise ratios. The answer to the second question is no; even if the estimate does not run away it converges to a biased estimate except when $p = 1/2$. The amount of bias depends on the signal-to-noise ratio and on p.

The stationary (i.e., $k \to \infty$) points of the distribution of $\hat{p}$ are those where in the first equation the expected value of the step change in $\hat{p}_{k+1}$ equals zero. With probability one $\hat{p}_k$ will approach one of these points:

$$E\left[\hat{p}_k - u(v(kT) - \frac{1}{2\lambda} \quad \ell n \frac{1-\hat{p}_k}{\hat{p}_k})\right] = 0 ,$$

where the expectation is taken with respect to the noise distribution as a function of $\hat{p}_k$.

An evaluation of the above equation results in a curve of one of the two forms shown in Fig. 20 depending on λ^2, the signal-to-noise ratio. λ_c is a critical value explained subsequently. Those values where the expression crosses the axis with positive slope are stable; those with negative slope are unstable (convergence occurs with probability zero). To see this, note that the average change in $\hat{p}_{k+1}$ (i.e., $-1/k$ times the expression plotted) is negative when the estimate is above a crossing with positive slope and vice versa. Hence estimates tend to return to stable points.

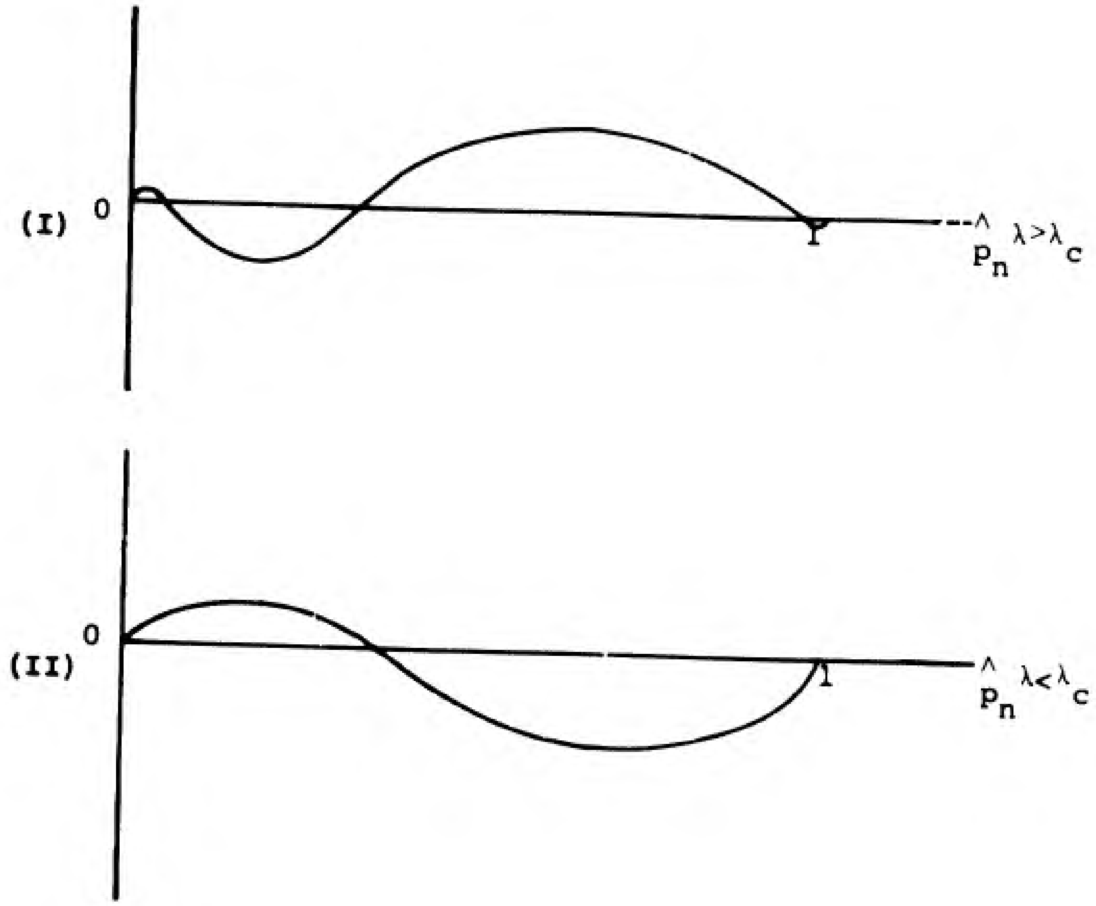


Figure 20. Form of the Expected Step Change for Signal-to-Noise Ratios Above and Below the Critical Value.

Note that in curve (I) of the figure there is a stable (although in fact generally biased) point between zero and one as well as at the ends; whereas in curve (II) the only stable points are at the ends. λ_c represents a critical signal-to-noise ratio (dependent on the true value of p) above which convergence to an intermediate value may occur and below which $\hat{p}_k$ almost surely goes to zero or one. The amount of bias when $\lambda > \lambda_c$ depends on λ and p. As $\lambda \to \infty$ or $p \to 1/2$, the bias disappears. As $\lambda \to \infty$ the probability of a runaway to the ends goes to zero. A bound has been formed on this probability which appears in Fig. 21 for $p = 1/2$. It is seen that the probability is very small for moderate signal-to-noise ratios. In Fig.22 the amount of bias in the convergent estimate is shown for several values of λ. It is seen that the bias is not excessive. The details of the computation involved appear in Appendix IV.

Of course, it would be possible in practice to prevent runaway by placing "barriers" below and above which the estimates could not proceed.

It is probably usually the case in applications that the true value p changes, perhaps slowly, with time. Hence a variation on the first equation which allows for updating is to take as the estimate

$$\hat{p}_{k+1} = \hat{p}_k - \alpha \quad \hat{p}_k - u(v(kt) - \frac{1}{2\lambda} \ell n \frac{1-\hat{p}_k}{\hat{p}_k}$$

where α is a fixed constant. Hence the technique will tend to "track" (with some lag) the true value without converging to anything.

To present specific results, Figs. 23 represent the statistical steady state distribution of $\hat{p}_\infty$, solved with a computer, one with a stable intermediate value, one without such a stable value. In the first case the probability of error using the technique is 0.161 compared with 0.158 when the true value of p is used, a negligible difference.

In the second case the probability of error is 0.448 compared with 0.308 in the case when p is known (here $\lambda < \lambda_c$, the critical value). This is, however, a very poor case in that the signal-to-noise ratio is much lower than that normally encountered.

It is felt that this technique is potentially of great practical interest.

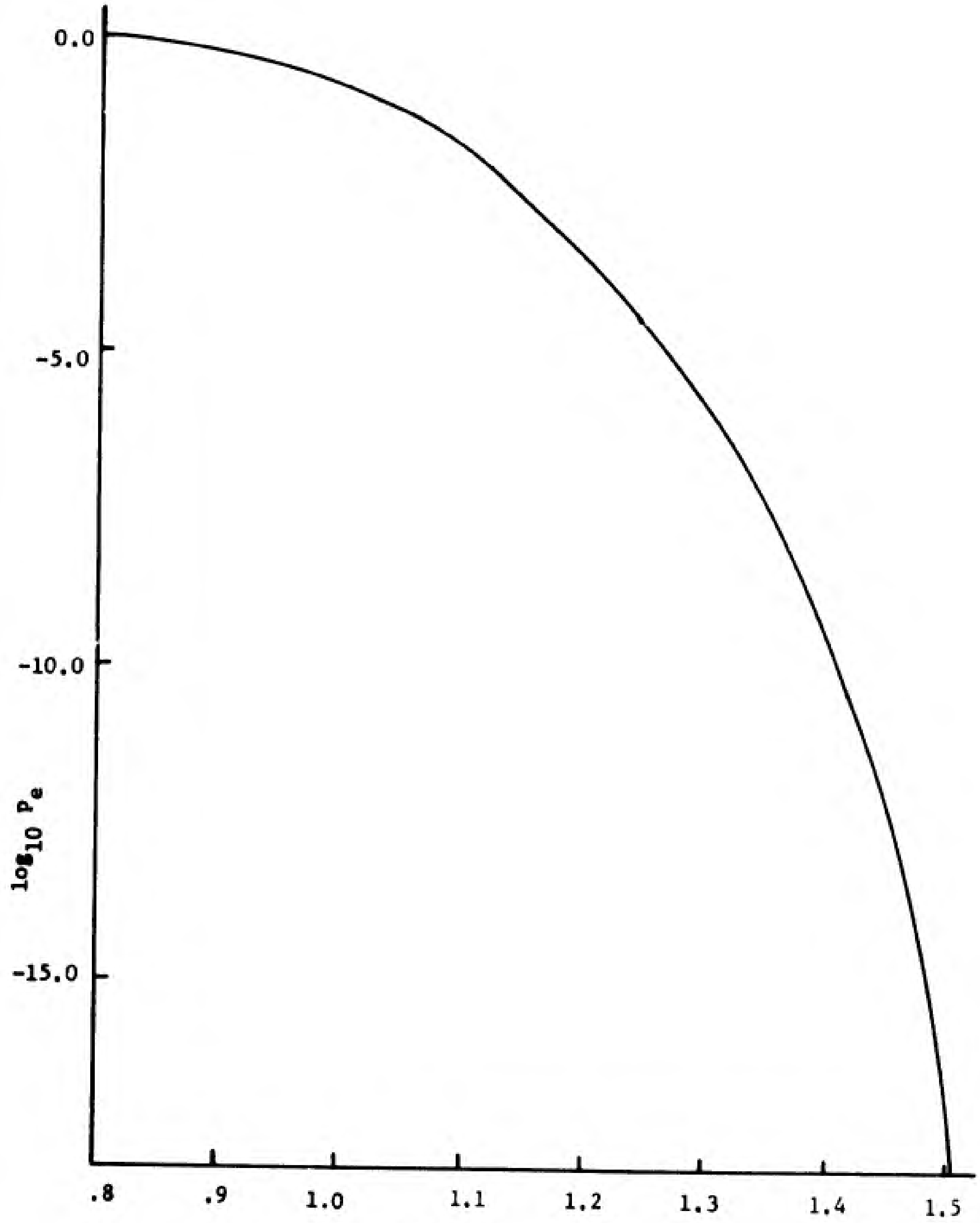


Figure 21. Bound on the Probability of Error as a Function of λ=SNR, p=0.5

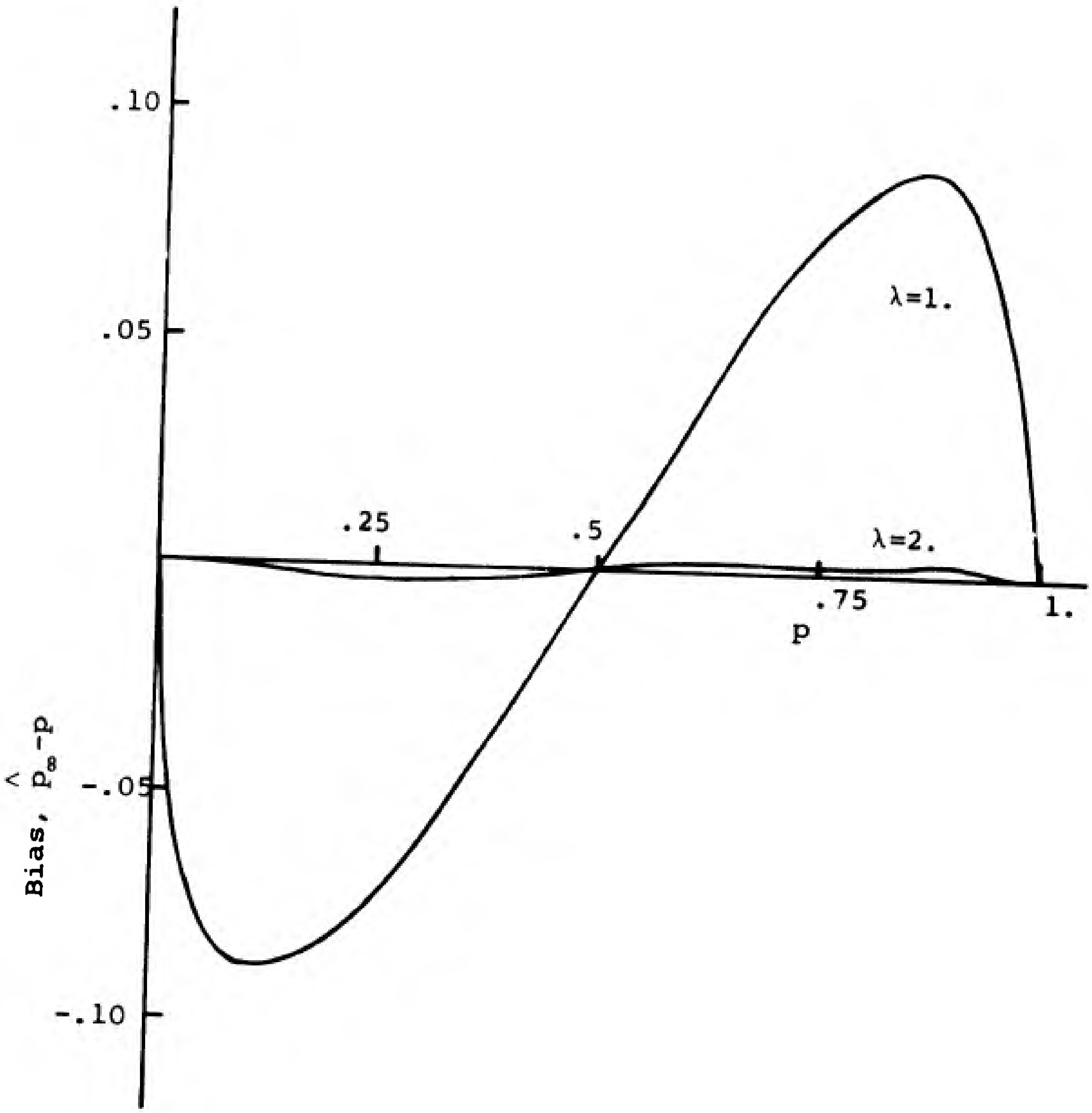
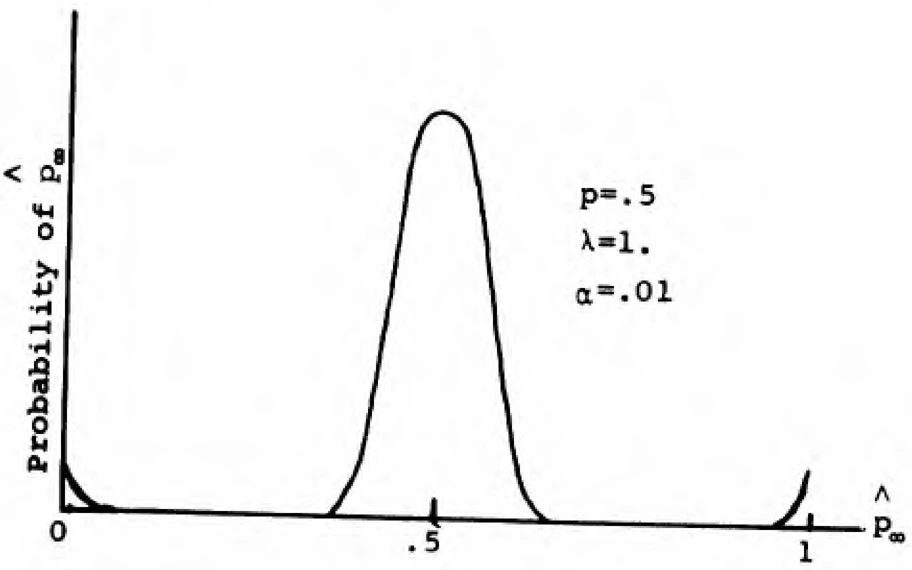


Figure 22. Bias in Estimation, $\hat{p}_\infty$-p, vs. True Prior, p



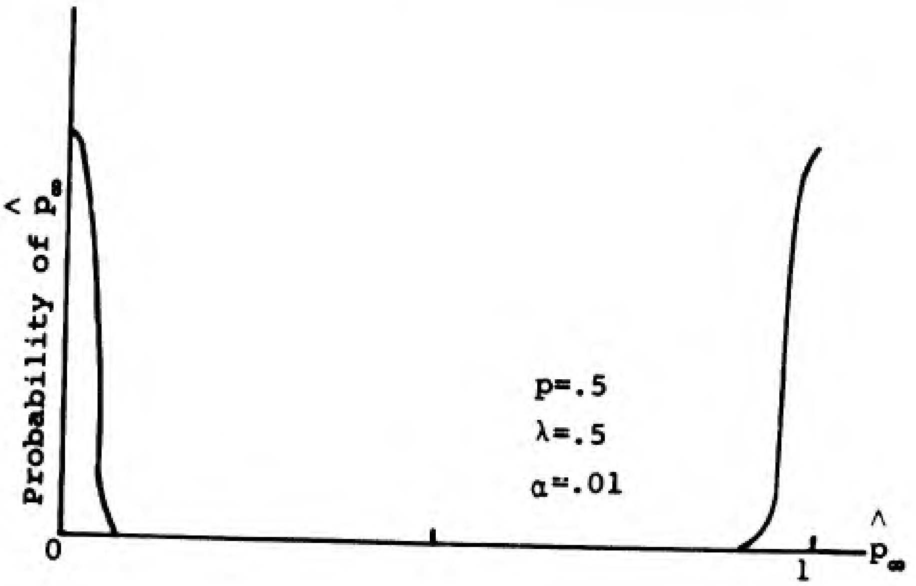


Figure 23. The Form of the Probability Curve for the Steady State Estimate for Signal-to-Noise Ratios Above and Below the Critical Value.

SECTION VIII

IMPORTANT RESULTS AND RECOMMENDATIONS FOR FUTURE WORK

Three areas considered during the contract period seem most worthy of further investigation as outlined below. These studies might be of a theoretical or an empirical nature or both.

1. The use of coincidence and rank tests for short time sample nonparametric detection. The studies during the contract period concentrated most heavily (although not exclusively) on the coherent, one channel, binary decision problem. The optimum rank test was derived for small sample, large amplitude, nominally Gaussian signals. Because of the complexity in generation, attention turned to the Wilcoxon rank sum statistic (the samples are ranked ir magnitude with the sum of the ranks of the positive observations being the threshold test statistic). This can be implemented in a relatively simple manner. It was found that the Wilcoxon statistic was nearly as good as the best statistic based on sample ranks and not too much less effective than the optimum statistic on the raw, unranked data (the statistic that would be used if the noise distribution were completely known). This conclusion was found to be valid for a range of dependence, sample size, and nonGaussian conditions. These results are highly encouraging and suggest the need for further studies. These could be of an empirical nature involving either recorded operational data if available, or generated data together possibly with a "breadboarded" receiver model. At the same time theoretical studies should continue to broaden and/or complete the results on M-ary, analog, two channel, etc. detection.

2. Robust tests which are intermediate to parametric and nonparametric tests in the degree of a priori assumptions made. Tests for more general situations than that considered in this report should be developed.

3. Towards the latter part of the contract effort, signal adaptive methods were considered to take advantage of the signal probability structure. These ideas are highly promising and seem likely to result in significant improvements in detector performance. The analysis to date has concentrated on the binary decision case with signal-to-signal independence. Greater gains can be expected when the ideas are applied where signal-to-signal dependencies are used. The theoretical problems are much more difficult in these situations, however.

It is strongly recommended that further attention be given to this area. An additional motivation is provided by the fact that much of the theory will have wider applicability than in the detection problem considered. In particular, applications to adaptive equalization and other decision directed schemes can be anticipated.

SECTION IX

CONCLUSIONS

This report has presented the most important results determined during the contract period. These results consist of several practical approaches to the problem of detecting and/or communicating in an uncertain noise and/or signal environment. Because of the broad based extent of the subject, the results in many cases, while promising, need further investigation as outlined in the report.

APPENDIX I

REVIEW OF DETECTION THEORY AND TERMINOLOGY

For theoretical studies a detection problem is normally cast as one of testing statistical hypotheses, typically the hypotheses denoted H_0 and H_1 which for the binary communication problem have the general meaning:

H_0: signal "zero" sent

H_1: signal "one" sent

For M-ary communication, M such hypotheses are defined. In the case of target detection H_0 and H_1 have the general meaning

H_0: no target present

H_1: target present

Hypothesis H_0 must be tested against hypothesis H_1, the objective being to accept H_0 and reject H_1 or vice versa. H_0 is frequently referred to as the null hypothesis or simply the hypothesis, whereas H_1 is frequently referred to as the alternate hypothesis or simply the alternative.

Statistical tests are performed on the observations containing the signal plus contaminating noise. These tests are optimized with respect to one of two criteria: the Bayes or the Neyman-Pearson criteria. Although the development starts from an entirely different point of view in each case, the form of the resulting detector is frequently (and always in this report) the same. A test statistic is formed on the observations and compared with a threshold (or thresholds) and a decision as to which hypothesis is true is made depending on the value of the test statistic compared with the threshold.

Three detection situations are distinguished in this report based on the nature of the observations. In the first situation only one channel of observation is available containing a signal of known form plus noise. In the other two situations two channels of observation are available, in one case a noisy signal of known form appears in only one channel with the second channel being a reference or "noise only" channel. In the other case a signal of unknown form is common to both channels, each containing noise which is channel-to-channel statistically independent.

At times it will be useful to distinguish between parametric and nonparametric hypotheses and tests. A parametric test is one that is derived from an exact knowledge of the data probability distribution except for a few unknown parameters (for example, independent normal observations with unknown mean and variance) some of which distinguish H_0 from H_1. Such conditions are accepted assumptions, and their verity or falsity establishes the meaningfulness of the probability statement arrived at by the parametric test. A nonparametric test is one for which a performance level is invariant within the class of distributions under H_0. A nonparametric hypothesis is one where the underlying data distribution under the hypothesis cannot be specified by a finite number of parameters (for example, all distributions with mean zero). Assumptions associated with nonparametric testing are fewer and weaker; therefore, conclusions arrived at by these tests are less sensitive to data distributional uncertainty.

Statistical tests are required; hence, statistical results must be expected (that is, errors will be made with some probability). In the event that H_1 is accepted when H_0 is really true, a type I error is said to have occurred. When the reverse happens, and H_0 is accepted when it should be rejected in favor of H_1, a type II error has been committed. For a fixed amount of data (the sample size), the probability of a type II error becomes larger as the probability of a type I error is decreased and vice versa. Ideally, some tradeoff is desired which optimizes the balance between the probabilities of making these two errors. In the case of target detection a type I error is called a false alarm.

Suppose a set of N sampled values $\{v(i\Delta t);\ i = 1,2,\ldots,N\}$ denoted for convenience by the vector V, is to be used in the statistical test. Then the hypotheses can be restated

H_0: V is from a population generated by signal zero or no target

versus

H_1: H_0 is not true.

Acceptance of H_0 implies a probabilistic judgement that the vector V is more likely to occur under H_0 than under H_1. The degree of likelihood required for acceptance of H_0 depends upon the acceptable error probabilities (or more generally the cost function) to be discussed in greater detail.

To further quantify the problem, suppose that the probability density (if the values $\{v(i\Delta t)\}$ are continuous) or the probability (if the values $\{v(i\Delta t)\}$ are discrete) of the observation vector V is given by $p_\lambda(V)$, where λ is an unknown parameter vector depending on the signal or target conditions. Suppose further that the set of all possible value of λ is denoted by Ω, called the parameter space, such that $-\infty \leq \lambda \leq \infty$. Let the parameter space be divided into two parts, Ω_{H_0} and Ω_{H_1}. If λ belongs to Ω_{H_0} (denoted by $\lambda \in \Omega_{H_0}$), H_0 is said to be true; conversely, $\lambda \in \Omega_{H_1}$ implies that H_1 is true.

Now the hypotheses to be tested become

$$H_0: \quad \lambda \in \Omega_{H_0}$$

versus

$$H_1: \quad \lambda \in \Omega_{H_1}$$

In some cases Ω_{H_0} or Ω_{H_1} may consist of a single point. In this case the hypothesis is called simple. Otherwise, the hypothesis is said to be composite. In the former case, the probability density under the simple hypothesis is uniquely determined. Suppose that H_0 and H_1 are both simple with parameters λ_{H_0} and λ_{H_1}. Then the relative probability of V occurring under H_0 is given by the likelihood ratio:

$$\Lambda = \frac{p_{\lambda_{H_0}}(V)}{p_{\lambda_{H_1}}(V)}$$

Thus, acceptance of H_0 is reasonable (and is formalized as the optimum Bayes test or the fundamental lemma of Neyman and Pearson) if

$$\Lambda > C$$

where C is a constant threshold determined a priori. Frequently the test can be expressed as a convenient function of data, for example

$$\frac{1}{N} \sum_{i=1}^{N} v(i\Delta t),$$

the sample mean. Let this function be called f(V). Then H_0 is accepted if $f(V) > C_1$.

The threshold C (or C_1) is determined to minimize the average cost (in this report the probability of error) under the Bayes criteria or to fix the false alarm probability at some level α under the Neyman-Pearson criteria.

APPENDIX II

RANK TESTS FOR NORMAL LARGE SHIFT ALTERNATIVES

The first term of an asymptotic expansion for the order vector probability is found in this section when the signal is binary and the noise is sample-to-sample independent and normal. This is used to find the optimum rank test. For equal error probabilities in detecting each of the binary signal values, the efficiency relative to the optimum parametric sample mean test is found to be greater than or equal to 3/4. Specifically, if for sample size, N, N_0 is the largest integer $\leq$ 3N/4 and N_1 is the smallest integer $\geq$ 3N/4, the efficiency is $\min(N_1, 4N_0/N)/N \geq 3/4$. The efficiency of the Wilcoxon test is found to be bounded above by $1/\sqrt{2}$ for the same error probabilities. Specifically, if N_0 is the smallest integer such that $2N_0(N_0+1) > N(N+1)$, the efficiency is $N_0/N \leq 1/\sqrt{2}$.

1. Introduction

Let $v(\Delta t), \ldots, v(N\Delta t)$ denote a sample with the normal cumulative distribution function Φ with mean λ and let $Z = (z_1, z_2, \ldots, z_N)$ be the sign-order where $z_i = 1$ if the i'th smallest observation in magnitude is positive and $z_i = -1$ otherwise. Then for testing the shift alternatives:

$$H_0: \lambda < 0$$

versus

$$H_1: \lambda > 0,$$

the most powerful (i.e. optimum) test is based on the likelihood ratio

$$\frac{p(Z)}{p(-Z)}$$

where p(Z) = probability that Z is observed when H_1 is true and by symmetry p(-Z) = probability that Z is observed when H_0 is true. For equal error probabilities under H_0 and H_1, C = 1. Using ideas similar to that of Klotz [19] and Hodges and Lehmann [20], who found the order of the error probability for large λ, the first term in an asymptotic expansion of the order vector probability can be found to get the most powerful rank test and the efficiency of rank tests.

2. Order Vector Probability

In terms of the standard normal density function φ the order vector probability is

$$P(Z) = N! \int I(t) \prod_{i=1}^{N} \varphi(t_i - z_i\lambda)\, d\underline{t},$$

where $t = (t_1, \ldots, t_N)$ and I is the indicator function

$$I(t) = \begin{cases} 1 & \text{if } 0 \leq t_1 \leq t_2 \leq \cdots \leq t_N \\ 0 & \text{otherwise} \end{cases}$$

The integrand takes on its largest value at a value of t determined by z_N and decreases exponentially for deviations from that value. This makes it possible, up to lower order terms as $\lambda \to \infty$, to break up the integral into a product of integrals which can be more readily evaluated. To illustrate this, let $N = 6$,

$$Z = (-1, -1, 1, 1, -1, 1)$$

Then

$$\max \prod_{i=1}^{6} \varphi(t_i - z_i\lambda) I(t) = \varphi^2(\lambda) \cdot \varphi^3\left(\frac{\sqrt{8}}{3}\lambda\right) \cdot \varphi(0)$$

attained for $t_1, t_2 = 0$; $t_3, t_4, t_5 = \lambda/3$; $t_6 = \lambda$. Thus, in this case, $p(Z)$ can be broken up into the product of 3 integrals since the integral over the values $0 \leq t_1 \leq t_2 \leq \cdots \leq t_6$ is the same except for lower order terms as $0 \leq t_1 \leq t_2 \leq \infty$, $0 \leq t_3 \leq t_4 \leq t_5 \leq \infty$, $0 \leq t_6 \leq \infty$. The integrand can be divided more generally by defining the sequence of "cluster" values

$0 \leq \mu_1 < \mu_2 < \ldots < \mu_n$ in terms of the integers $0 = M_0 < M_1 < M_2 < \ldots < M_n = N$ and the difference sequence $m_1 = M_1$, $m_2 = M_2 - M_1, \ldots, m_n = M_2 - M_{N-1}$ where

(1) If $\sum_{i=1}^{j} z_i \leq 0$ for every $1 \leq j \leq N$, $\mu_1 = 0$ and

M_1 is the largest value such that

$$\sum_{i=1}^{M_1} z_i \leq \sum_{i=1}^{j} z_i \text{ for every } 1 \leq j \leq N,$$

(2) If $\sum_{i=1}^{j} > 0$ for some $1 \leq j \leq N$, $\mu_1 = \frac{1}{M_1} \sum_{i=1}^{M_1} z_i > 0$ and

where M_1 is the largest value such that

$$\sum_{i=1}^{M_1} \frac{z_i}{M_1} \leq \sum_{i=1}^{j} \frac{z_i}{j}$$

(3) For $k \geq 2$,

$$\mu_k = \frac{1}{m_k} \sum_{i=M_{k-1}+1}^{M_k} z_i$$

where M_k is the largest value such that

$$\frac{1}{m_k} \sum_{i=M_{k-1}+1}^{M_k} z_i \leq \sum_{i=M_{k-1}+1}^{j} \frac{z_i}{j - M_{k-1}}, \; j \geq M_{k-1}.$$

then

$$\max_{0 \le t_1 \le \cdots \le t_N} \prod_{i=1}^{N} \varphi(t_i - z_i\lambda) = \prod_{i=1}^{n} \varphi^{m_i}\left(\sqrt{1-\mu_i^2}\;\lambda\right).$$

That this is the largest value of the integrand and that it occurs for the above sequence follows from the obvious fact that the maximum occurs when some of the t_i are equal, that by definition equality has to occur in runs with the nonnegative values between runs forming a strictly monotone increasing sequence, and finally that the sum of squares of the form $(t - x_i)^2$ is minimized for t at the centroid. It may be verified by obvious though tedious manipulation that any other sequence of runs of equality would have to result in a nonstrictly monotone sequence.

It is clear that as $\lambda \to \infty$ the probability integral is the same as the product of the integrals about the cluster values except for exponentially decreasing tail values. Hence for large λ

$$p(Z) = \prod_{i=1}^{n} Q_i$$

where for $i > 1$ and for $i = 1$ if $\mu_1 > 0$:

$$Q_i = \int_{0 \le t_1 \le \cdots \le t_{m_i}} \prod_{j=1}^{m_i} \varphi(t_i - z_{M_{i-1}+j}\lambda)\, d\underline{t}$$

$$\to \exp\left(-\frac{m_i(1-\mu_i^2)}{2}\lambda\right) \int_{-\infty \le t_1 \le \cdots \le t_m \le \infty} \prod_{j=1}^{m_i} \varphi(t_j - \mu_i\lambda)\, d\underline{t}$$

Noting that the integral is the same as the probability of an order vector without considering magnitude

$$Q_i = \frac{1}{m_i!} \exp\left(-\frac{m_i}{2}(1-\mu_i^2)\lambda\right).$$

If $\mu_1 = 0$, let $\nu_1 = \frac{1}{M_1} \cdot \sum_{i=1}^{M_1} z_i$. Then:

$$Q_1 = \int\limits_{0 \le t_1 \le \cdots \le t_{m_1}} \prod_{j=1}^{m_1} \varphi(t_j - z_j\lambda)\, d\underline{t}$$

$$= \exp\left(-\frac{m_1}{2}(1-\nu_1^2)\lambda\right) \int\limits_{0 \le t_1 \le \cdots \le t_{m_1}} \varphi(t_j - \nu_1\lambda)\,dt_j$$

Since $\nu_1 \le 0$, the limits of integration cannot be extended to $-\infty$. However, the integral is the same as the probability that a sequence of normal variates with mean $\nu_1\lambda$ is nonnegative and has a given order. Since there are $m_1!$ equally likely orderings:

$$Q_1 = \frac{1}{m_1!} \exp\left(-\frac{m_1}{2}(1-\nu_1^2)\lambda\right) \Phi^{m_1}(\nu_1\lambda)$$

$$\rightarrow \frac{1}{m_1!} \frac{\varphi^{m_1}(\lambda)}{|\nu_1\lambda|^{m_1}} \qquad \nu_1 < 0$$

$$\rightarrow \frac{1}{m_1!\, 2^{m_1}} \exp\left(-\frac{m_1}{2}\lambda\right) \qquad \nu_1 = 0$$

where the asymptotic formula $\Phi(x) \sim \frac{\varphi(x)}{|x|}$ as $x \rightarrow -\infty$ is used.

Now the likelihood ratio can be evaluated for large λ. Let $m_1, m_2, \ldots, m_n$ and $\mu_1, \ldots, \mu_n$ be defined as above and let $k_1, k_2, \ldots, k_\ell$ and $\omega_1, \ldots, \omega_\ell$ be the corresponding values for -Z. Let

$$n_0 = \begin{cases} 0 \text{ if } \omega_1 > 0 \\ k_1 \text{ if } \omega_1 = 0 \end{cases} \quad \text{and} \quad n_1 = \begin{cases} 0 \text{ if } \mu_1 > 0 \\ m_1 \text{ if } \mu_1 = 0. \end{cases}$$

Note that in general $\ell \neq n$ and $k_i \neq m_i$, although, of course,

$$\sum_{i=1}^{n} m_i = \sum_{i=1}^{\ell} k_i.$$

Then for large λ the exponential parts of the probability will dominate, i.e.

$$\frac{p(z)}{p(-z)} = K(Z)\lambda^{n_0 - n_1} \exp\left[-\frac{\lambda}{2} \sum_{i=1}^{n} m_i (1 - \mu_i^2) + \frac{\lambda}{2} \sum_{i=1}^{\ell} k_i (1 - \omega_i^2)\right]$$

where K(Z) is a function which depends on Z and not on λ. Hence the argument of the exponent generally determines the decision in the far out λ case. The decision rule is to decide H_1(that is, $+\lambda$) if

$$F(Z) = \sum_{i=1}^{n} m_i \mu_i^2 - \sum_{i=1}^{\ell} k_i \omega_i^2 > 0.$$

If $F(Z) < 0$, H_0 is decided. If $F(Z) = 0$, the decision is based on $\delta_1 - \nu_1$. If $F(Z) = 0$ and $n_0 - n_1 > 0$, H_1 is decided. If $n_0 - n_1 < 0$, H_0 is decided.

The probability of error for the optimum rank code is given by the sum of the probabilities of all rank vectors under H_1 which result in deciding H_0:

$$P_{re} = \sum_{Z:\frac{p(Z)}{p(-Z)} \leq 1} p(Z)$$

For large λ the terms in the sum which contain the minimum exponent in the asymptotic probability will deominate. Thus for any N, if

$$p(Z) = G(Z) \prod_{i=1}^{n} e^{-\frac{1}{2} m_i (1-\mu_i^2)\lambda^2}$$

$$= G(Z) \exp\left[-\frac{\lambda^2}{2} \left(\sum_{i=1}^{n} m_i (1 - \mu_i^2)\right]\right.$$

where G(Z) is a nonexponential factor and $\{m_i\}$, $\{\mu_i\}$ depend on Z, then the vector Z with the smallest value

$$M(Z) = \sum_{i=1}^{n} m_i (1 - \mu_i^2)$$

over all vectors such that $p(Z)/p(-Z) \leq 1$ will determine P_{re} as $\lambda \to \infty$. Let M_0 be that minimum value. Then asymptotically,

$$P_{re} = C_r(\lambda) e^{-(M_o/2)\lambda^2}$$

where $C(\lambda)$ is a nonexponential factor.

On the other hand, for the optimum parametric test the probability or error is asymptotically

$$P_{pe} = C_p(\lambda) e^{-(N/2)\lambda^2}$$

Hence the relative efficiency of the rank test to the parametric test is given by

$$R.E. = \frac{M_0}{N}$$

The value M_0 can be found for any N by noting from the decision procedure that there must be a vector Z such that $M_0 = M(Z)$ and in addition for which $p(Z)/p(-Z) \leq 1$, implying that $M(-Z) \leq M_0$. In other words, M_0 is the smallest number for which there is a vector Z such that

$$M_0 = M(Z) \geq M(-Z)$$

Suppose that any integer, $N_1, N/2 \leq N_1 < N$ is selected. For what vector Z is M(-Z) a minimum given that $M(Z) = N_1$? The answer is that $z_1 = z_2 = \ldots = z_{m_1} = -1$, $z_{M_1+1} = \ldots = z_n = +1$. Thus

$$M(-Z) \geq N - \frac{(2N_1 - N)^2}{N} .$$

Therefore N_1 is the smallest integer such that

$$N_1 \geq N - \frac{(2N_1 - N)^2}{N} = 4N_1(1 - \frac{N_1}{N})$$

or

$$N_1 \geq \frac{3}{4} N$$

Thus M_0 is determined by N_1 and if $N_1 > \frac{3}{4} N$, the efficiency is determined by $N_0 = N_1 - 1$ in which case M_0 is the smaller of N_1 and $4N_1(1 - N_1/N)$. Or the relative efficiency is:

$$R.E. = \min(N_1, 4N_1(1 - N_1/N))/N$$

which implies that for large N, R.E. = $\frac{3}{4}$.

A similar procedure can be used to bound the Wilcoxon efficiency

APPENDIX III

DISTRIBUTION OF TEST STATISTICS UNDER DEPENDENCE

If the Gaussian normalized correlation function $\rho(t)$ is integrable,

$$\int_{-\infty}^{\infty} |\rho(t)| \; dt < \infty$$

then the A.R.E. of one test with respect to another can be expressed as the ratio of efficacies of the test where the efficacy of statistic S^N for the statistics in this report is:

$$\varepsilon = \lim_{N\to\infty} \frac{\left\{\frac{\partial^j}{\partial\lambda^j} E[S^N]\right\}^2\Big|_{\lambda = 0}}{N \text{ var } (S^N)\Big|_{\lambda = 0}}$$

where N is the number of samples in each channel, λ is the signal amplitude, and j is the smallest value such that the derivative is nonzero.

For the sample mean

$$E[S_\mu^N] = \lambda$$

$$\lim_{N\to\infty} N \text{ var } [S_\mu^N] = \lim_{N\to\infty} \frac{\sigma^2}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} 2\rho([i-j]\Delta t)$$

$$= \lim_{N\to\infty} 2\sigma^2 \sum_{k=N}^{N} (1 - \frac{|k|}{N})\rho(k\Delta t)$$

$$= 2\sigma^2 \sum_{k=\infty}^{N} \rho(k\Delta t).$$

The last statement uses the integrability of $\rho(t)$. Thus the efficacy for the sample mean is:

$$\varepsilon_{2\mu} = \frac{1}{2\sigma^2 \sum_{k=-\infty}^{\infty} \rho(k\Delta t)}$$

For the Mann-Whitney statistic, as $\lambda \to 0$,

$$\lim_{\lambda \to 0} \frac{\partial}{\partial \lambda} E[S_M^N] = \lim_{\lambda \to 0} \frac{\partial}{\partial \lambda} \{\Pr[x(i\Delta t) - y(j\Delta t) \geq 0 | \lambda]$$

$$- \Pr[x(i\Delta t) - y(j\Delta t) < 0 | \lambda]\}$$

$$= \frac{1}{\sigma\sqrt{\pi}}$$

$$N \text{ var } [S_{MW}^N]\Big|_{\lambda=0} = \frac{1}{N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} E\{\text{sgn}[x(i\Delta t) - y(j\Delta t)]$$

$$\text{sgn}[x(k\Delta t) - y(\ell\Delta t)]\}$$

Now by the inverse sine rule, if x and y are Gaussian with mean zero and correlation coefficient ρ:

$$E[\text{sgn}(x)\text{sgn}(y)] = \frac{2}{\pi} \sin^{-1}\rho$$

Thus:

$$N \text{ var } [S_{MW}^N]\Big|_{\lambda=0} = \frac{2}{\pi N} \sum_{n=-N}^{N} \sum_{k=-N}^{N} \left(1 - \frac{|n|}{N}\right)\left(1 - \frac{|k|}{N}\right) \sin^{-1} \frac{\rho(n\Delta t) + \rho(k\Delta t)}{2}$$

This sum can be evaluated asymptotically by noting that for any $\varepsilon > 0$, there is an integer M independent of N such that

$$\left| \frac{2}{\pi N} \sum_{n=-N}^{N} \sum_{|k| \geq M} \left(1 - \frac{|n|}{N}\right)\left(1 - \frac{|k|}{N}\right) \sin^{-1} \frac{\rho(n\Delta t)+\rho(k\Delta t)}{2} - \sin^{-1} \frac{\rho(n\Delta t)}{2} \right| < \frac{\varepsilon}{2}$$

And for sufficiently large N:

$$\left| \frac{2}{\pi N} \sum_{|n|,|k|<M} \sum \left(1 - \frac{|n|}{N}\right)\left(1 - \frac{|k|}{N}\right) \sin^{-1} \frac{\rho(n\Delta t)+\rho(k\Delta t)}{2} \right| < \frac{\varepsilon}{2}$$

Thus in the limit:

$$\lim_{N\to\infty} N \operatorname{var} [S_{MW}^{N}] \Big|_{\lambda=0} = \frac{2}{\pi} \sum_{n=-\infty}^{\infty} 2 \sin^{-1} \frac{\rho(n\Delta t)}{2}$$

Thus the efficacy is:

$$\varepsilon_{MW} = \frac{1}{2\sigma^2 \sum_{n=-\infty}^{\infty} 2\sin^{-1} \frac{\rho(n\Delta t)}{2}}$$

The asymptotic relative efficiency of the Mann-Whitney with respect to the sample mean test is

$$\frac{\varepsilon_{MW}}{\varepsilon_{2\mu}} = \frac{\sum_{n=-\infty}^{\infty} \rho(n\Delta t)}{\sum_{n=-\infty}^{\infty} 2\sin^{-1} \frac{\rho(n\Delta t)}{2}}$$

The A.R.E. of the mixed statistic with the first grouping follows from the equation with N replaced by m and an additional outer summation to account for group-to-group correlation. A similar analysis results in the A.R.E. for the second grouping.

The Wilcoxon is the analog of the Mann-Whitney for the one channel case. Given the observation x(t) with mean λ it is desired to test

$$H_0: \lambda = 0$$

versus

$$H_1: \lambda > 0 \quad .$$

It can be verified that the efficacy of the one-sample mean test agrees with that of the Mann-Whitney for the two-sample mean test as defined in the preceding except for the factor of 2 in the denominator:

$$\varepsilon_{1\mu} = \frac{1}{\sigma^2 \sum_{n=-\infty}^{\infty} \rho(n\Delta t)}$$

The Wilcoxon statistic is based on the sum of the ranks in magnitude of the positive sample values. It is most conveniently represented in the form:

$$S_W^N = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (1+\delta_{ij}) \operatorname{sgn}[x(i\Delta t) + x(j\Delta t)] \gtrless C$$

The efficacy is evaluated by proceeding as before for the Mann-Whitney:

$$\lim_{\lambda \to 0} \frac{\partial}{\partial \lambda} E[S_W^N] = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1+\delta_{ij}}{\sqrt{1+\rho\,[(i-j)\Delta t]}} \quad \frac{2}{\sigma\sqrt{\pi}}$$

$$= \frac{1}{N} \sum_{n=-N}^{N} \left(1 - \frac{|n|}{N}\right) \frac{1+\delta_{on}}{1+\rho(n\Delta t)} \quad \frac{2}{\sigma\sqrt{\pi}}$$

Here the correlation appears in the **mean** as it did not for the Mann-Whitney due to the independence of the two channels there. However, using an argument similar to that in passing to the limit before:

$$\lim_{N \to \infty} \lim_{\lambda \to 0} \frac{\partial}{\partial \lambda} E[S_W^N] = \frac{2}{\sigma\sqrt{\pi}}$$

The variance also poses problems arising from dependence which were not encountered in the Mann-Whitney:

$$\lim_{N\to\infty} N \operatorname{var}(S_W^N)\Big|_{\lambda=0} = \lim_{N\to\infty} \frac{2}{\pi N^3} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} \sum_{\ell=1}^{N} \sin^{-1}$$

$$\frac{\rho[(i-k)\Delta t]+\rho[(i-\ell)\Delta t]+\rho[(j-k)\Delta t]+\rho(j-\ell)\Delta t]}{2\sqrt{1+\rho[(i-j)\Delta t]}\ \sqrt{1+\rho[(k-\ell)\Delta t]}}$$

This can be reduced by one summation to:

$$\lim_{N\to\infty} N \operatorname{var}(S_{W1}^N) = \lim_{N\to\infty} \frac{2}{\pi N^2} \sum_{n=-N}^{N} \sum_{k=-N}^{N} \sum_{\ell=-N}^{N} r[1-\max(0,\tfrac{n}{N},\tfrac{k}{N},\tfrac{\ell}{N}) + \min(0,\tfrac{n}{N},\tfrac{k}{N},\tfrac{\ell}{N})]$$

$$\sin^{-1} \frac{\rho(n\Delta t)+\rho(\ell\Delta t)+\rho[(n-k)\Delta t]+\rho[(k-\ell)\Delta t]}{2\sqrt{1+\rho(k\Delta t)}\ \sqrt{1+\rho[(n-\ell)\Delta t]}}$$

where

$$r(x) = \begin{cases} x & x > 0 \\ 0 & x \le 0 \end{cases} .$$

By an argument more involved algebraically but similar to that in form employed on the Mann-Whitney:

$$\lim_{N\to\infty} N \operatorname{var}(S_W^N) = \frac{4}{\pi} \sum_{n=-\infty}^{\infty} 2\sin^{-1} \frac{\rho(n\Delta t)}{2} .$$

Thus the efficacy is.

$$\epsilon_W = \frac{1}{\sigma^2 \sum_{n=-\infty}^{\infty} 2\sin^{-1} \frac{\rho(n\Delta t)}{2}}$$

The A.R.E. with respect to the sample mean is:

$$\frac{\varepsilon_W}{\varepsilon_{1\mu}} = \frac{\sum_{n=-\infty}^{\infty} \rho(n\Delta t)}{\sum_{n=-\infty}^{\infty} 2\sin^{-1} \frac{\rho(n\Delta t)}{2}}$$

Thus the A.R.E. is the same for the Wilcoxon as for the Mann-Whitney when the data is fully ranked. From this it can be inferred that channel-to-channel correlation does not affect the A.R.E. This is, in fact, true. It is not true, however, that the results for the Wilcoxon are the same as for the Mann-Whitney for mixed statistics with finite group size or that interchannel correlation leaves the mixed Mann-Whitney A.R.E. unaffected. The efficacy of the mixed Wilcoxon with the first sequential type of grouping is:

$$\varepsilon_{W1} = \frac{2\sum_{n=-m}^{m} \left(1 - \frac{|n|}{m}\right) \frac{1+\delta_{on}}{\sqrt{1+\rho(n\Delta t)}}}{\sigma^2 \sum_{j=-\infty}^{\infty} \sum_{n=-m}^{m} \sum_{k=-m}^{m} \sum_{\ell=-m}^{m} r\left[1-\max\left(0,\frac{n}{m},\frac{k}{m},\frac{\ell}{m}\right) + \min\left(0,\frac{n}{m},\frac{k}{m},\frac{\ell}{m}\right)\right]}$$

$$\cdot\ (1+\delta_{ok})(1+\delta_{n\ell})\sin^{-1} \frac{\rho[(jm+n)\Delta t]+\rho[(jm+\ell)\Delta t]+\rho[(jm+n-k)\Delta t]+\rho[(jm+\ell-1)\Delta t]}{2\sqrt{1+\rho(k\Delta t)}\ \sqrt{1+\rho[(n-\ell)\Delta t]}}$$

where r(x) is as defined above.

By similar calculations, the efficacy for the second method of grouping is:

$$\varepsilon_{W2} = \frac{\left[\frac{1}{\sqrt{2}} + m-1\right]^2}{\sigma^2 \sum_{n=-\infty}^{\infty} \left[m \sin^{-1}\rho(\Delta t)+2(m-1)\sin^{-1} \frac{\rho(\Delta t)}{\sqrt{2}}\ 2(m-1)(m-2)\sin^{-1} \frac{\rho(\Delta t)}{2}\right]}$$

The Kendall τ statistic is one which has application to the two channel detection of an unknown signal in noise. Let $x(t)$ and $y(t)$ be the two channel outputs:

$$x(t) = \lambda s(t) + n_x(t)$$

$$y(t) = \lambda s(t) + n_y(t)$$

where $n_x(t)$ and $n_y(t)$ are independent noise processes and $s(t)$ is a common noise independent signal process. It is desired to test

$$H_0 : \lambda = 0$$

versus

$$H_1 : \lambda \neq 0 .$$

If the signal and noise processes are mean zero Gaussian, then the optimum parametric test for independent sample value is an energy detector:

$$S_e^N = \frac{1}{N} \sum_{i=1}^{N} [x(i\Delta t) + y(i\Delta t)]^2 \gtrless C.$$

Let the normalized correlation function for the signal be $\rho_s(t)$ and the noise processes be $\rho_n(t)$, and let the noise variance be σ^2 and the signal variance be unity. Then the efficacy of the energy detector for dependent sample values is:

$$\varepsilon_e = \frac{8}{\sigma^4 \sum_{k=-\infty}^{\infty} \rho_n^2(k\Delta t)}$$

The Kendall τ nonparametric test for correlation is based on the correlation between ranks in the two channels. The statistic is most conveniently expressed as:

$$S_\tau^N = \frac{1}{N^2} \sum_{i \neq j} \operatorname{sgn}[x(i\Delta t) - x(j\Delta t)] \operatorname{sgn}[y(i\Delta t) - y(j\Delta t)] .$$

The efficacy is found by calculations similar to those in the preceding to be:

$$\frac{\varepsilon_\tau}{\varepsilon_e} = \frac{\sum\limits_{k=-\infty}^{\infty} \rho_n^2(k\Delta t)}{2 \sum\limits_{k=-\infty}^{\infty} \left[2 \sin^{-1} \frac{\rho_n(k\Delta t)}{2} \right]^2} \; .$$

It is interesting to note the similarity in form between this result and the preceding results. Note that the signal correlation is not a parameter.

The results for the mixed first and second methods of grouping are obtained by computations similar to those in the preceding material:

$$\varepsilon_{\tau 1} = \frac{4 \sum\limits_{k=1}^{m} \left(1 - \frac{k}{m}\right) \left[\frac{1-\rho_s(k\Delta t)}{1-\rho_n(k\Delta t)} \right]^2}{\sigma^4 \sum\limits_{j=-\infty}^{\infty} \sum\limits_{i=-m}^{m} \sum\limits_{\substack{k=-m \\ k \neq 0}}^{m} \sum\limits_{\substack{\ell=-m \\ \ell \neq i}}^{m} r\left[1-\max\left(0,\frac{i}{m},\frac{k}{m},\frac{\ell}{m}\right) + \min\left(0,\frac{i}{m},\frac{k}{m},\frac{\ell}{m}\right)\right]}$$

$$\left\{ \sin^{-1} \frac{\rho_n[(i+jm)\Delta t]-\rho_n[(\ell+jm)\Delta t]-\rho_n[(i-k+jm)\Delta t]+\rho_n[(\ell-k+jm)\Delta t]}{2\sqrt{[1-\rho_n(k\Delta t)][1-\rho_n(i-\ell)\Delta t]}} \right\}^2$$

$$\varepsilon_{\tau 2} = \frac{1}{\sigma^4 \frac{1}{m-1} \sum\limits_{k=-\infty}^{\infty} \left[\frac{1}{2} \left\{ \sin^{-1} \rho_n(k\Delta t) \right\}^2 + (m-2) \left\{ \sin^{-1} \frac{\rho_n(k\Delta t)}{2} \right\}^2 \right]}$$

Note that the signal correlation appears in the result for the sequential grouping but is removed in the third method of grouping.

Implicit in the foregoing computations and subsequent conclusions is the assumption that the distribution of the test statistic is asymptotically normal. In this section the assumption is justified. For definiteness the Mann-Whitney statistic is considered with the understanding that similar results hold for the Wilcoxon and Kendall τ statistics.

The theorem of Rosenblatt [21] is used to establish asymptotic normality of the mixed statistics for $m < \infty$. The definition of a strong mixing condition is required as a preliminary to a statement of the theorem. Let the process $\{x_t, t \in T\}$ be defined on a probability space $(\Omega, \mathcal{A}, P)$ where $\mathcal{A}$ is the minimum σ-field of ω sets with respect to which the x_t's are measurable. Furthermore, denote by $\mathcal{A}_r^s$ the minimum σ-field generated by the x_t's for $r < t < s$. For $\tau > 0$ let

$$\sup_{A,B} \left| P\{A \cap B\} - P\{A\}P\{B\} \right| < a_x(\tau)$$

uniformly in τ where the supremum is with respect to $A \in \mathcal{A}_{-\infty}^{t}$, $B \in \mathcal{A}_{t+\tau}^{\infty}$.

<u>Definition</u>: The process $\{x_t\}$ satisfies the strong mixing condition if $a_x(\tau) \to 0$ as $\tau \to \infty$. $a_x(\tau)$ is referred to as the mixing coefficient of the process.

Clearly if $\{x_t\}$ is strong mixing, then so is any process $\{z_t\}$, derived from it if events defined in terms of z_ν, $\nu \le t'$ depend on the x_μ's for $\mu \le t$. $\{z_t\}$ is also strong mixing if events defined in terms of z_ν, $\nu \le t'$, depend on the x_μ's for μ in a finite union of intervals and if events defined in terms of z_ν, $\nu \ge t' + \tau$, depend on the x_μ's for μ in another finite union of intervals such that the minimum separation between the two sets of intervals becomes infinite with τ.

Depending on whether sampling or continuous processing is used, define

$$H(r,s) = \sum_{t'=r}^{s} z_{t'},$$

or

$$H(r,s) = \int_r^s z_{t'}\, dt'$$

$$\sigma^2(r,s) = \mathrm{var}(H(r,s))$$

The central limit theorem can now be stated:

Theorem (Rosenblatt [21]): Let $\{z_t, t \in T\}$ be a zero mean stochastic process. The statistic $\eta(r,s) = H(r,s)/\sigma(r,s)$ is asymptotically normal if $\{z_t\}$ is strong mixing and

$$\lim_{r-s\to\infty} \frac{E[H(r,s)]^4}{\sigma^4(r,s)} < \infty .$$

Consider for the moment the Markov process defined on the integers $t = n = \ldots,-1,0,1,\ldots$ which satisfies $x_{n+1} = \rho x_n + \omega_n$ where $\{\omega_n\}$ is an independent ergodic sequence. The strong mixing properties of this process follow trivially from the easily established fact [22] that ergodic Markov processes are strong mixing.

Fortunately, for Gaussian processes, it is possible to obtain more general criteria for strong mixing. In [23] it is shown that if $\{x_t, t \in T\}$ is a stationary Gaussian process possessing spectral density $p(\lambda)$ rational in λ, then $\{x_t, t \in T\}$ is strong mixing and in fact $a_x(\tau) \to 0$ exponentially in τ.

Consider the mixed Mann-Whitney test statistic with the first method of grouping:

$$S^N_{MW1} = \frac{1}{m^2 p} \sum_{n=0}^{p-1} \sum_{i=1}^{m} \sum_{j=1}^{m} z^n_{ij}$$

with

$$z^n_{ij} \triangleq \operatorname{sgn}\,[v_1(nm\Delta t + i\Delta t) - v_2(nm\Delta t + j\Delta t)]$$

It is necessary to write the above triple sum as a single sum possessing the separation property. Let

$$z_n = \sum_{i=1}^{m} \sum_{j=1}^{m} z^n_{ij}$$

It is clear that $\{z_n\}$ is strong mixing.

Consider the second method of grouping with test statistic

$$S_{MW2}^N = \frac{1}{m^2 p} \sum_{n=0}^{p-1} \sum_{i=1}^{m} \sum_{j=1}^{m} Z_{ij}^n ,$$

$$Z_{ij}^n = \mathrm{sgn}\,[v_1(n\Delta t) + ip\Delta t) - v_2(n\Delta t + j.p\Delta t)]$$

Let

$$z_n = \sum_{j=1}^{m} \sum_{k=1}^{m} Z_{ij}^n$$

The sequence $\{z_n\}$ is not strong mixing because z_0 and z_{p-1} use adjacent values of the $\{v_1(i\Delta t)\}$ and $\{v_2(i\Delta t)\}$ processes. However, if the statistic

$$T_{MW3}^N = \frac{1}{m^2 p} \sum_{n=0}^{p-\sqrt{p}} z_n$$

is considered, rather than S_{MW2}^N, the strong mixing property of the statistic is preserved and since

$$\lim_{N\to\infty} E[(S_{MW2}^N - T_{MW3}^N)^2]/\mathrm{var}^{1/2}(S_{MW2}^N) \to 0,$$

S_{MW2}^N has the same asymptotic distribution as T_{MW2}^N .

Having established the strong mixing properties of the resultant test statistics operating on Gaussian data with rational spectral densities, it remains only to demonstrate the above moment conditions.

These will be demonstrated under the null hypothesis H_0 for the two-channel test employing the Mann-Whitney intermediate statistic and the second method of grouping. Obvious analogous results hold for the other grouping and the single-channel Wilcoxon and Kendall τ tests. Consider the statistic S_{MW1}^N given above. It has already been shown in the efficacy calculations that $\mathrm{var}(S_{MW1}^N) = O(N^2)$. Hence it is necessary to establish that

$$E[(S_{MW1}^N)^4] = O(N^4).$$

The following lemma is required as a preliminary:

Lemma: If Y_i, $i = 1,2,3,4$ are zero-mean Gaussian random variables with normalized covariance

$$\rho_{ij} \triangleq \frac{E\{Y_i Y_j\}}{E\{Y_i^2\}\ E\{Y_j^2\}}$$

so that $0 \leq |\rho_{ij}| < 1$, then

$$\left| E \prod_{i=1}^{4} \mathrm{Sgn}(Y_i) \right| \leq \frac{4}{\pi^2} \sum_{r<s} \sin^{-1}\{|\rho_{pq}|\} \sin^{-1}\{|\rho_{rs}|\};$$

$$p \neq q \neq r \neq s$$

Proof: From Pawula [24]

$$E\left\{\prod_{i=1}^{4} \mathrm{Sgn}\ Y_i\right\} = \frac{4}{\pi^2} \sum_{r<s} \rho_{rs} \int_0^1 \frac{\sin^{-1}\{\rho_{pq.rs(\alpha)}\}}{\sqrt{1-\alpha^2\rho_{rs}^2}}\, d\alpha;\ p \neq q \neq r \neq s$$

Where the quantity $\rho_{pq.rs}(\ )$ is a partial correlation coefficient [25] defined as

$$\rho_{pq.rs} = \frac{E\{\eta_{p.rs}\eta_{q.rs}\}}{E\{\eta_{p.rs}^2\}\ E\{\eta_{q.rs}^2\}}$$

with $\eta_{p.rs}$ the residual of the random variable ξ_p after the best linear estimate of ξ_p in terms of ξ_r, ξ_s is removed. The random variables ξ_i, $i = 1,2,3,4$ themselves are zero-mean Gaussian with covariance matrix

$$K_\alpha = \left(\alpha^{(1-\delta_{ij})}\rho_{ij}\right) \quad 0 < \alpha \leq 1$$

But surely the correlation coefficient between the residuals must be less than that between the variables themselves. Thus, for $p \neq q$

$$\left|\rho_{pq.rs(\alpha)}\right| \leq \alpha\left|\rho_{pq}\right|$$

so that

$$\left|E\left\{\prod_{i=1}^{4} \mathrm{Sgn}(Y_i)\right\}\right| \leq \frac{4}{\pi^2} \sum_{r<s} |\rho_{rs}| \int_0^1 \frac{\sin^{-1}\{\alpha|\rho_{pq}|\}}{\sqrt{1-\alpha^2\rho_{rs}^2}}\, d\alpha;\ p \neq q \neq r \neq s$$

$$\leq \frac{4}{\pi^2} \sum_{r<s} \sin^{-1}\{|\rho_{pq}|\} \int_0^1 \frac{|\rho_{rs}|\, d\alpha}{\sqrt{1-\alpha^2\rho_{rs}^2}}$$

where use is made of the fact $\sin^{-1}\{\alpha|\rho_{pq}|\} \leq \sin^{-1}\{|\rho_{pq}|\}$ since $0 < \alpha \leq 1$. Finally, performing the α integration in this last expression the desired result is obtained.

Now define the variables

$$Y_{i_k j_k} \triangleq v_1(n_k m\Delta t + i_k\Delta t) - v_2(n_k m + j_k\Delta t)$$

for $k = 1,2,3,4$; $i_k, j_k = 1,2,\ldots,m$; and $n_k = 0,1,2,\ldots$. With

$$\rho_{k\ell} \triangleq \frac{E\{Y_{i_k j_k} Y_{i_\ell j_\ell}\}}{E\{Y^2_{i_k j_k}\}\, E\{Y^2_{i_\ell j_\ell}\}}$$

$$= \frac{\rho[(n_k-n_\ell)\Delta t+[(i_k-i_\ell)\Delta t]+ [(n_k-n_\ell)\Delta t+(j_k-j_\ell)\Delta t]}{2}$$

Then

$$E_{H_0}\left|S_{MW2}^{N}\right|^4 = \sum_{n_1,n_2,n_3,n_4=0}\left[\sum_{i_1,i_2,i_3,i_4=1}\ \sum_{j_1,j_2,j_3,j_4=1}\right.$$

$$E_{H_0}\left\{\prod_{k=1}^{4}\mathrm{sgn}Y_{i_kj_k}\right\}$$

and from the preceding lemma

$$E_{H_0}\left\{\prod_{i=1}^{4}\mathrm{Sgn}\,Y_{i_kj_k}\right\} \leq \frac{4}{\pi^2}\sum_{r<s}\sin^{-1}\{|\rho_{rs}|\}\sin^{-1}\{|\rho_{pq}|\};$$

$$p \neq q \neq r \neq s$$

The summation on the righthand side is over the distinct combinations of four objects taken two at a time or $\binom{4}{2} = 6$ combinations. By substituting this inequality and by suitable relabeling of indices for each of the six cases it is found that the moment conditions for the grouping represented by the statistic S_{MW2}^{N} follow by similar arguments and need not be repeated. Thus the asymptotic normality of the mixed test employing the Mann-Whitney intermediate statistic for fixed finite $m < N$ is established.

For the fully ranked case $m = N$, the test statistic cannot be put in the form of a single sum of random variables appropriate for the application of the standard central limit theorems. This is true even when the data is not dependent. To circumvent this difficulty several ingenious limit theorems have been developed. Most notable in this regard are Hoeffding's Limit Theorems for U-statistics [26], and the pioneering work of Chernoff and Savage [27]. These limit theorems succeed by approximating a given statistic T_N in mean-square by another statistic T_N' which has the form of a single sum of random variables. If $\lim_{N\to\infty} E\{T_N - T_N'\}^2 = 0$ then it follows from Cramer's Theorem [28] that the two statistics will have identical asymptotic distributions. The asymptotic normality of T_N' which can be established via the standard central limit theorems, implies that of T_N.

Obviously a similar procedure can be employed in the dependent sample case where now the appropriate limit theorem is that of Rosenblatt. This is the approach employed here. The Mann-Whitney statistic is a linear rank-order statistic which can be written in the form

$$mT_N = \sum_{i=1}^{N} E_{Ni} z_{Ni}$$

where the E_{Ni} are given numbers in the interval (0,1] and $z_{Ni} = 1$ if the i'th smallest sample was from the signal channel and zero otherwise. The Mann-Whitney statistic can be written in this form with $E_{Ni} = (i/N)$.

The asymptotic distribution of statistics of this form operating on Gaussian data with rational spectral density is established by an extension of the Chernoff-Savage Theorem. The following quantities are called Chernoff and Savage: Let λ_N and m/N and assume that for all N the inequalities $0 < \lambda_0 \leq \lambda_N \leq 1-\lambda_0 < 1$ hold for some fixed $\lambda_0 \leq 1/2$. The quantity $F_m(x)$ is to be the empirical distribution function of the signal sample $\{X_i\}_{i=1}^{m}$ while $G_n(y)$ is the empirical distribution function of the reference sample $\{Y_i\}_{i=1}^{n}$. Define:

$$H_N(x) \triangleq \lambda_N F_m(x) + (1 - \lambda_N) G_n(x)$$

which will be called the combined empirical c.d.f., while the quantity

$$H(x) \triangleq \lambda_N F(x) + (1 - \lambda_N) G(x)$$

where F(x) and G(x) are the actual signal and reference sample c.d.f.'s, will be called the combined population c.d.f. Instead of the linear rank-order statistics, Chernoff and Savage have found it convenient to consider statistics T_N of the form

$$T_N = \int_{-\infty}^{\infty} J_N[H_N(x)]\,dF_m(x)$$

which is easily seen to reduce to the linear rank order statistic with $E_{Ni} = J_N(i/N)$. Obviously $J_N(\cdot)$ need only be defined on the lattice 1/N, 2/N,..., N/N, but we shall find it convenient, as Chernoff and Savage, to extend its doman of definition to (0,1] in such a way that $J_N(\cdot)$ is, in fact, continuous on this interval. Following the details of the Chernoff-Savage paper the statistic T_N can be written as

$$T_N = A + B_{1N} + B_{2N} + \sum_{i=1}^{N} C_{Ni}$$

where the A,B,C terms represent constant, first order random and higher order terms respectively. Furthermore, it is shown in the referenced paper that

$$B_{1N} = \frac{(1 - \lambda_N)}{m} \sum_{i=1}^{m} [B(X_i) - E\{B(X_i)\}]$$

$$B_{2N} = - \frac{(1 - \lambda_N)}{n} \sum_{i=1}^{n} [B^*(y_i) - E\{B(y_i)\}]$$

where

$$B(x) \triangleq \int_{x_o}^{x} J'[H(y)]dG(y)$$

$$B^*(x) \triangleq \int_{x_o}^{x} J'[H(y)]dF(y)$$

and

$$J(\zeta) \triangleq \lim_{N\to\infty} J_N(\zeta) \qquad 0 < \zeta < 1$$

is assumed to exist and is not constant. The quantity x_o is determined arbitrarily, say by $H(x_o) = 1/2$. With some relatively mild regularity conditions added to the Chernoff-Savage Theorem it can be shown by an application of the Rosenblatt Central Limit Theorem that both B_{1N} and B_{2N} are asymptotically normal, and since they are independent so also is their sum. Furthermore, the higher order C_{Ni} terms can be shown to converge to zero, at least in probability, as $N \to \infty$ so that the asymptotic normality of T_N is established by an application of Cramer's Theorem [28]. In particular,

<u>Theorem</u>: If $\{v_1(i\Delta t)\}_{i=1}^{m}$ and $\{v_2(i\Delta t)\}_{j=1}^{n}$ are mutually independent sequences

each of which is obtained by homogeneous sampling of a stationary Gaussian process with absolutely integrable covariance function, and if

1. $J(H) = \lim_{N\to\infty} J_N(H)$ exists for $0 < H < 1$ and is not constant

2. $\int_{I_N} [J_N(H_N) - J(H_N)]dF_m = o_p(1/\sqrt{N})$*

   where: $I_N \triangleq \{x: 0 < H_N(x) < 1\}$

3. $J_N(1) = o(\sqrt{N})$

4. $|J^{(i)}(H)| = |\frac{d^iJ}{dH^i}| \leq K[H(1-H)]^{-i-1/2} + \delta$

   for $i = 0,1,2$ and for some $\delta > 0$

5. $J_N(\cdot)$ is monotone increasing in its argument.

6. The limit function $J(\cdot)$ satisfies a uniform Lipshitz condition on (0,1), i.e., for each $x_o \in (0,1)$ there exists an $\delta(x_o) > 0$ and an M independent of x_o such that

$$|J(x) - J(x_o)| < M\,|x - x_o| \quad \text{for } |x-x_o| < \delta(x_o)$$

then for fixed F, G and λ_N, the statistic T_N is asymptotically nontrivially normally distributed as m,n and hence $N \to \infty$.

The proof of this theorem is fairly lengthy and will not be given here. Let us merely remark that the conditions 1-4 follow from the original Chernoff-Savage theorem while conditions 5-6 have been added to allow application of the Rosenblatt Central Limit Theorem and to establish the higher order nature of the C_{Ni} terms when the data is no longer i.i.d. The conditions 1-6 can all be shown to hold for the Mann-Whitney statistic.

Unfortunately, the preceding result does not apply directly to the single channel Wilcoxon detector or the Kendall τ detector. Some comments

---

*The notation $y = o_p(N)$ means $p - \lim_{N\to\infty} y/N = 0$

concerning the situation with respect to these latter detectors are appropriate.

If fictitious V_1 and V_2 channels are identified with the positive and negative samples and the observations ranked with respect to magnitude, it is well known [29] that the Wilcoxon statistic can be represented in the form of a linear rank statistic. It would appear then that the extended Chernoff-Savage Theorem is applicable to the single channel Wilcoxon detector. The falacy with this conclusion is the fact that while each of two sums called B_{1N} and B_{2N} above are shown to be asymptotically normal, the two sums are no longer independent and unless one can prove they are jointly asymptotically normal, one cannot conclude [30] that their sum $B_{1N} + B_{2N}$ is asymptotically normal. A proof of the joint asymptotic normality of these sums appears difficult, and consequently it is not yet possible to establish the asymptotic normality of the fully-ranked Wilcoxon detector.

The situation with respect to the Kendall τ test presents another problem as it is not a linear rank-order statistic. Nevertheless, in the independent case the Kendall τ statistic can be shown to converge in mean-square to the Spearman Rank Correlation test [31,13]which is a linear rank-order statistic. In the case of correlated input data, if this convergence continues to hold then an application of the extended Chernoff-Savage result will have established the asymptotic normality of the Kendall τ statistic. Due to the extremely complicated nature of the computations involved it has not been possible to show that this is the case even for Gaussian data. Thus the asymptotic normality of the fully-ranked Kendall τ statistic remains an open question.

APPENDIX IV

ANALYSIS OF THE DECISION DIRECTED RECEIVER

The estimate of the prior probability of a one, $\hat{p}_n$, satisfies

$$\hat{p}_{n+1} = \frac{1}{n} \sum_{i=k+1}^{n} u\left(\frac{v_i}{\sigma} - \frac{1}{2\lambda} \ln \frac{1-\hat{p}_i}{\hat{p}_i}\right) + \frac{k}{n} z_o$$

$$= \hat{p}_n - \frac{1}{n}\left[\hat{p}_n - u\left(\frac{v_n}{\sigma} - \frac{1}{2\lambda} \ln \frac{1-\hat{p}_n}{\hat{p}_n}\right)\right]$$

$u(\cdot)$ is the unit step function and z_o is an initial estimate used until time $k + 1$.

The regression function is the expected value of the bracketed change as a function of $\hat{p}$ (the subscript is now dropped for convenience:

$$m(\hat{p}, p) = E_{\hat{p}}\left[\hat{p} - u\left(\frac{v}{\sigma} - \frac{1}{2\lambda} \ln \frac{1-\hat{p}}{\hat{p}}\right)\right]$$

$$= \hat{p} - \text{Prob (deciding "one" given } \hat{p})$$

$$= \hat{p} - p\Phi\left(\lambda - \frac{1}{2\lambda} \ln \frac{1-\hat{p}}{\hat{p}}\right) - (1-p)\Phi\left(-\lambda - \frac{1}{2\lambda} \ln \frac{1-\hat{p}}{\hat{p}}\right)$$

where p is the true prior and $\Phi(\cdot)$ is the normal distribution function. The zeros of $m(\hat{p}, p)$ determine the convergent values of the estimate as $n \to \infty$. Plots of the form of $m(\hat{p}, p)$ appear in Fig. 20 in the text.

There is a critical signal-to-noise ratio λ_c such that if $\lambda > \lambda_c$, there exists a stable convergence value, $0 < \hat{p}_\infty < 1$ if $0 < p < 1$. $m(\hat{p}_\infty) - p$ is not zero, however, except as $\lambda \to \infty$ or if $p = 1/2$. This bias is calculated and appears in Fig. 21.

It is important to evaluate or at least bound the probability that the estimate converges to one or zero rather than $\hat{p}_\infty$. It is possible to get an upper bound by the following reasoning. Suppose that another estimation scheme is proposed where if p_n ever exceeds some fixed value, $x < 1$ for any n, it is absorbed (fixed) at x. Furthermore, suppose that the threshold is fixed at $(1/2\lambda)\ell n(1-x/x)$. Since $x \geq \hat{p}$, the probability of absorption at x is increased over the variable threshold case. Thus the probability of absorption at x is a bound on the probability of converging to one in the original scheme. Similarly, another scheme can be proposed where if $\hat{p}_n$ ever goes below some value y, then the estimate is absorbed (fixed) at y. The sum of the two probabilities of absorption at x or y provides an upper bound on the probability of runaway.

The probability of going above x will be considered with the understanding that an abvious similar analysis exists for going below y. This probability can be expressed as

$$\text{Prob}\left[\frac{1}{n} \sum_{i=k+1} \varphi_i + \frac{k}{n} z_o \geq x\right]$$

where for convenience φ_i is used in place of the random variable

$$u\left(\frac{v_i}{\sigma} - \frac{1}{2\lambda} \ell n \frac{x}{1-x}\right).$$

The probability can be rewritten as:

$$\text{Prob}\left[k(x - z_o) + \sum_{i=k+1}^{n} (x - \varphi_i) \leq 0\right].$$

This probability satisfies the following difference equation (cf. Feller [32]):

$$\mu_z = \pi(x)\mu_{z+x-1} + (1 - \pi(x))\mu_{z+x}$$

where

μ_z = probability of absorption starting at the value $z = k(x - z_o)$

$\pi(x)$ = probability that $\varphi_i = 1$ at any step

$= x - m(x, p)$

The solution is subject to the boundary conditions

$$\mu_z = 0 \qquad z \leq 0$$

$$\lim_{z \to \infty} \mu_z = 0$$

Proceeding in the usual fashion for linear difference equations solutions of the form $\mu_z = \gamma^z$ are sought so that γ satisfies

$$f(\gamma) = \pi(x)\gamma^{x-1} + (1 - \pi(x))\gamma^x - 1 = 0 \ .$$

Unfortunately this polynomial may have an infinitely countable set of roots. However, a bound on the solution μ_z can be found by first noting that $\gamma =$ .. is one root, and since $f(\infty) = f(0) = \infty$, there must be at least one more real root in the range $0 < \gamma < \infty$ (including possibly a double root at $\gamma = 1$). If there is a real root, γ_o, such that $0 < \gamma_o < 1$, then

$$\mu_z \leq v_z = \gamma_o^z \ .$$

This is true because $\omega_z = \gamma_o^z - \mu_z$ satisfies the difference equation, $\omega_z \to 0$ as $z \to \infty$ and $\omega_z \geq 0$ for $z \leq 0$. This implies that ω_z is a probability and hence

$$\omega_z = \gamma_o^z - \mu_z \geq 0$$

or

$$\mu_z \leq \gamma_o^z \ .$$

Fortunately there is such a root, γ_o, between zero and one if (and only if)

$$\gamma_M = \frac{(1-x)\pi(x)}{x(1-\pi(x))} < 1 \ .$$

This condition is always satisfied if $\lambda < \lambda_c$ for some x. That the above is necessary is found by differentiating $f(\gamma)$ and noting that $f'(\gamma) = 0$ has only one solution for $0 < \gamma < \infty$, the solution being for γ equal to the above expression. Hence the minimum of $f(\gamma)$ exists at the point γ_M which implies that there is a root to the left and right of there. Thus, since one root is one, there is one more positive root which is less than one if $\gamma_M < 1$.

A bound on the probability can be evaluated then by finding the real root of $f(\gamma)$ and then minimizing the bound as a function of x. This is most easily done by iteration. Starting with an initial value

$$\gamma_{oo} = \pi(x)^{1/(1-x)} ,$$

successive iterations of the following rearrangement of the equation $f(\gamma) = 0$ converge to γ_o:

$$\gamma_{o(n+1)} = (\pi(x) + (1 - \pi(x))\gamma_{on})^{1/(1-x)}$$

In fact, as $\lambda \to \infty$, $\gamma_o \to 0$. Hence,

$$\gamma_o \to [\pi(x)]^{1/(1-x)}$$

So for large λ:

$$\mu_z \leq v_z \to [\pi(x)]^{z/(1-x)}$$

$$= [\pi(x)]^{k[(x-z_o)/(1-x)]}$$

Furthermore, suppose that x is determined as a function of λ in such a way that

$$-\lambda - \frac{1}{2\lambda} \ln \frac{1-x}{x} \to -\infty .$$

To be specific, for some $0 < a < 1$, let $[x/(1-x)] = e^{-2a\lambda^2}$. Then

$$\pi(x) \to p ,$$

and

$$v_z \to p^{e^{2a\lambda^2}(1-z_o)-z_o}$$

This double exponential behavior in λ causes the probability bound to go to zero with extreme rapidity (for $p = 0.5$, $\lambda = 2.0$, $v_z = 10^{-4642}$). Hence, for even moderate signal-to-noise ratios runaway can be neglected.

The case where the estimate is updated rather than allowed to converge is handled by replacing the $1/n$ convergence factor by a fixed constant, $\alpha << 1$. If α is too large the estimate will not "settle" under stationary conditions. If α is too small the estimate will not "track" well under nonstationary conditions. Hence some compromise is called for. An examination of the stochastic difference equation reveals that the density, $f(\hat{p})$, of the estimate in steady state satisfies

$$f(\hat{p}) = \frac{1}{1-\alpha}\left[\pi\left(\frac{\hat{p}-\alpha}{1-\alpha}\right) f\left(\frac{\hat{p}-\alpha}{1-\alpha}\right) + \left(1 - \pi\left(\frac{\hat{p}}{1-\alpha}\right)\right) f\left(\frac{\hat{p}}{1-\alpha}\right)\right]$$

where $\pi(\cdot)$ as before is the probability of deciding "one" as a function of the estimate. Unfortunately it has not been possible to solve this equation exactly. However, if the continuous estimates are quantized in N increments, Δ (say $\Delta = .001$, $N = 1000$), the probability of any estimate in steady state can be determined by a computer solution.

Since the estimates continuously vary for all time, and since the probability of any estimate is nonzero (albeit small), at some future time the estimate will (with probability one) runaway to zero or one. Of interest is the mean (average) time to this event. If this is sufficiently large the probability can be neglected.

The following difference equation is satisfied by $m_{\hat{p}}$, the mean time to reach fixed values x and 1-x starting at a value $\hat{p}$:

$$m_{\hat{p}} = \pi(\hat{p}) m_{\hat{p}(1-\alpha)+\alpha} + (1 - \pi(\hat{p})) m_{\hat{p}(1-\alpha)} + 1$$

with the boundary conditions $m_{\hat{p}} = 0$, $\hat{p} \leq x$, $\hat{p} \geq 1-x$. This equation can also be solved computationally.

As a (relatively poor) approximation to the above for large signal to noise ratios, supposing that $\pi(\hat{p}) = p$ and that the estimates are quantized with an α such that at each time the estimate is changed by exacting one increment $\pm \Delta$ with $x = M\Delta$. Then the difference equation becomes

$$m_{\hat{p}} = p m_{\hat{p}+\Delta} + q m_{\hat{p}-\Delta} + 1 \ .$$

where $q = 1 - p$. Looking for solutions of the form $m_{\hat{p}} = \gamma^{\hat{p}/\Delta}$ to the homogeneous equation:

$$1 = p\gamma + q\gamma^{-1}$$

results in

$$\gamma = 1,\ q/p$$

The solution to the nonhomogeneous equation is

$$\frac{\hat{p}}{\Delta(q-p)}, \quad p \neq 1/2$$

$$-\frac{\hat{p}^2}{\Delta^2}, \quad p = 1/2$$

Combining and adjusting for boundary conditions results in:

$$m_{\hat{p}} = \frac{1}{q-p} \quad \frac{\hat{p}}{\Delta} - M + \frac{N - 2M}{(q/p)^M - (q/p)^{N-M}} \quad (q/p)^{\hat{p}/\Delta} - (q/p)^M$$

$$p \neq 1/2$$

$$= M(M - N) + N\frac{\hat{p}}{\Delta} - \left(\frac{\hat{p}}{\Delta}\right)^2$$

Suppose that $N = 1000$, $\Delta = .001$, $M = 10$ and the true prior is $p = 1/2$. Then $m_{\tilde{p}} \approx 24 \times 10^4$ by the above approximation. This is a relatively poor approximation, however, since the restoring force is greater on the ends than the one step assumption made.

REFERENCES

1. Capon, J., "Optimum Coincidence Procedures for the Detection of Weak Signals in Noise," IRE Conv. Record, 1960, pt. 4.

2. North, D.O., "An Analysis of the Factors Which Determine Signal/Noise Ratio Discrimination in Pulsed-Carrier Systems," RCA Laboratories Report PTR-6C, June 1943, Reprinted in Proc. IEEE, Vol. 51, no. 7, July 1963, pp. 1015-1027.

3. Capon, J., "On the Asymptotic Relative Efficiency of Locally Optimum Detectors," IEEE Trans. on Information Theory, Vol. IT-7, April 1961, pp. 67-72.

4. Middleton, D., "On Canonically Optimum Threshold Detection," IEEE Trans. on Information Theory, Vol. IT-12, April 166, pp. 230-243.

5. Fraser, D.A.S., Nonparametric Methods in Statistics, John Wiley & Sons, Inc., New York, N.Y., 1957, Chap. 6.

6. Grenander, U. and M. Rosenblatt, Statistical Analysis of Stationary Time Series, John Wiley & Sons, Inc., New York, N.Y., 1957, Chap. 7.

7. Grenander, U. and G. Szego, Toeplitz Forms and Their Applications, Univ. of California Press, Berkeley, California, 1955, Chap. 11.

8. Modestino, J.W., Portion of Ph.D. Thesis, (to be published), Princeton University (1969).

9. Carlyle, J.W., "Nonparametric Methods in Detection Theory," Chapter 8 in Communication Theory, A. V. Balakrishnan, ed., McGraw-Hill Book Co., Inc., New York, N.Y., 1968.

10. Grenander, U., "On the Estimation of Regression Coefficients in the Case of an Autocorrelated Disturbance," Ann. Math. Stat., Vol. 25 (1954), pp. 252-272.

11. Davisson, L.D., "Comments on 'Performance of the Neyman-Pearson Detector for Correlated Input Samples'," Proc. of IEEE, Vol. 55, No. 2, (1967), p. 239.

12. Feustel, E.A. and L.D. Davisson, "Asymptotic Efficiency of Mixed Statistical Tests," IEEE Trans. on Information Theory, Vol. IT-13, no. 2, April 1967, pp. 247-254.

13. Kendall, M.G. and A. Stuart, The Advanced Theory of Statistics, Hafner Publishing, New York, 1961.

14. Tukey, J.W., "A Survey of Sampling From Contaminated Distributions," Contributions to Probability and Statistics, (Harold Hotelling Volume) p. 448, Stanford University Press, 1960.

15. Huber, P.J., "A Robust Version of the Probability Ratio Test," Annals Math. Stat., Vol. 36, p. 1753, 1965.

16. Algazi, V.R. and R. M. Lerner, "Binary Detection in White Non-Gaussian Noise," Lincoln Laboratories, M.I.T., Technical Report DS-2138.

17. Martin, R.D., Portion of Ph.D. Thesis (to be published), Department of Electrical Engineering, Princeton University.

18. Kanefsky, M. and J. B. Thomas, "On Polarity Detection Schemes with Non-Gaussian Inputs," Journal Franklin Institute, Vol. 280, No. 2, August 1965.

19. Klotz, J.H., "Small Sample Power and Efficiency for the One-Sample Wilcoxon and Normal Scores Tests," Ann. Math. Stat., vol. 34 (1963), pp. 335-337.

20. Hodges, J.L. and E. L. Lehmann, "Probabilities of Rankings for Two Widely Separated Normal Distributions," Studies in Math. Analysis and Related Topics, Stanford University Press, 1962, pp. 146-151.

21. Rosenblatt, M., "A Central Limit Theorem and a Strong Mixing Condition," Proc. Nat. Acad. Sci., U.S.A., Vol. 42, (1956) pp. 43-47.

22. Rozanov, Yu. A., "An Application of the Central Limit Theorem," Proc. 4th Berkeley Symp. on Math. Stat. and Prob., University California Press, Berkeley, California, 1961, pp. 445-454.

23. Kolmogorov, A. and I. Rosanov, "On a Strong Mixing Condition for Stationary Gaussian Processes," Teor. Veroyatnost. i Primenen, Vol. 5 (1960).

24. Pawula, R.F., "A Modified Version of Prices Theorem," IEEE Trans. on Information Theory, Vol. IT-13, April 1967, pp. 285-288.

25. Cramer, H., Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J., 1946, pp. 305-307.

26. Hoeffding, W., "A Class of Statistics with Asymptotically Normal Distributions," Ann. Math. Stat., Vol. 19, (1948), pp. 293-325.

27. Chernoff, H. and I.R. Savage, "Asymptotic Normality and Efficiency of Certain Nonparametric Test Statistics," Ann. Math. Stat., Vol. 29, (1958), pp. 972-994.

28. Cramer, H., op. cit., p. 254.

29. Lehmann, E. L., Testing Statistical Hypotheses, John Wiley & Sons, Inc., New York, N.Y., 1959, p. 257.

30. Papoulis, A., Probability, Random Variables and Stochastic Processes, McGraw-Hill Book Co., Inc., New York, N.Y., 1965, pp. 222-224.

31. Hajek, J. and Z. Sidak, Theory of Rank Tests, Academic Press, Inc., New York, N.Y., 1962, Chap. 1.

32. Feller, W., An Introduction to Probability Theory and Its Applications, Vol. I, J. Wiley, 1957.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Princeton University Princeton NJ 08450	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP N/A

3. REPORT TITLE

Short-Time Sample Robust Detection Study

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5. AUTHOR(S) (First name, middle initial, last name)

Davisson, L. D.
Thomas, J. B.

6. REPORT DATE March 1969	7a. TOTAL NO. OF PAGES 105	7b. NO. OF REFS 32
8a. CONTRACT OR GRANT NO. F30602-68-C-0078 b. PROJECT NO. 4519 c. Task 451902 d.	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-69-24	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of RADC (EMCRS), Griffiss AFB NY 13440.

11. SUPPLEMENTARY NOTES N/A	12. SPONSORING MILITARY ACTIVITY Rome Air Development Center (EMCRS) Griffiss AFB, NY 13440

13. ABSTRACT

The optimum detector is well defined when the signal and noise characteristics are completely specified through deterministic and/or probabilistic statements. Unfortunately this is not always the case and attempts to make it so through assumptions of one sort or another may lead to poor detector performance. This report studies the detection problem when certain of the signal or noise properties are unknown except perhaps within some wide class. Particular attention is directed to communications problems characterized by small time-bandwidth signals. Several applications emerge from the study. The first results from an exhaustive investigation of the nonparametric Wilcoxon rank sum detector which is found to be practical in implementation and nearly as good in performance as the optimum detector under a wide range of sample size, dependence, and noise distribution conditions. A second application from the study results from an investigation of optimum small sample linear coincidence detectors which are found to be superior to a Gaussian parametric detector when the normality assumption is violated. A third application from the study results from the analysis of a simple adaptive threshold detector which is practical to implement and attains an improvement over fixed threshold receivers. In addition there are results of an incomplete or theoretical nature whose application is either not so important or as yet well defined. These include the study of an optimum "robust" detector for nearly Gaussian noise.

DD FORM 1 NOV 65 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Signals Communication Theory Statistical Tests Statistical Distributions Statistical Analyses						