THE THEORY OF SIGNAL DETECTABILITY:
BAYESIAN PHILOSOPHY, CLASSICAL STATISTICS,
AND THE COMPOSITE HYPOTHESIS

by

David Jaarsma

Approved by: *[signature]*

for

COOLEY ELECTRONICS LABORATORY

Department of Electrical Engineering
The University of Michigan
Ann Arbor, Michigan

February 1970

ABSTRACT

Receiver design and performance from a Bayesian viewpoint depend upon a priori specification whenever unknown parameters are encountered in the detection situation; any available information is expressed in the form of an a priori density. A sensitivity index is developed which measures the performance loss that occurs when the receiver is designed to be optimal with respect to the given a priori density $g(\cdot)$ but operates in an environment in which the a priori density $h(\cdot)$ is considered to hold. The sensitivity index is shown to possess several desirable features.

A comparison of receiver performance is made for the composite hypothesis situation. The Bayesian approach is contrasted to the classical approach. Initial indications were that classical statistics (including parameter estimation) could be closely linked to Bayesian philosophy in view of the fact that analysis according to either mode often led to the same receiver. It appeared possible that many of the classical tests could be generated from a Bayesian viewpoint by an appropriate assignment of the a priori density. Investigation revealed that this was not true in general; and the conclusion is drawn that the Bayesian approach is uniquely distinct from the classical approach. Emphasis upon unbiasedness is an admirable quality of classical statistics; yet incorporation of all prior information into the detection model by Bayesian methods is considered a superior attribute.

The externally sensed parameter (ESP) receiver is reviewed and its receiver operating characteristic (ROC) is evaluated for several examples not considered before. The ESP receiver is a hypothetical receiver whose performance with respect to the assigned a priori density serves as an upper bound to the performance of the optimum receiver.

Receiver design via numerical integration techniques is demonstrated to be feasible for composite hypothesis situations previously considered too complex to solve. The receiver is conceived of basically as a digital computer; and its performance is nearly optimal with little increase in complexity of operation. The numerical procedure is especially adept at handling diffuse a priori densities.

Receiver design via estimation techniques is considered justifiable in case optimal (Bayesian) procedures are too complex. The optimum receiver is formulated for the single composite hypothesis in such a way that when the receiver is operating in a sequential mode it appears to "estimate" the unknown parameter at each point in time. It is not an estimator in the usual sense of the word but nevertheless exhibits features and characteristics that are considered desirable of an estimator. Subject to some mild conditions the "estimator" converges asymptotically to the minimum mean square error estimator.

FOREWORD

In practice detection equipment invariably operates in a
variety of different situations. Many of the parameters considered
known in the original design theory of the equipment will be unknown
in operation. This lack of knowledge affects the available process-
ing gain from a sophisticated receiver. The lack of knowledge of
such parameters as source level, propagation loss, target strength
(that is, those parameters affecting the received signal energy) or
any parameters affecting the received noise level or spectrum,
further complicate meaningful evaluation of system performance.

Classical engineering and classical statistics propose a type
of receiver design and evaluation which is often at odds with that
developed from the more recent Statistical Theory of Signal Detect-
ability. Basically, the classical designs are made insensitive to
uncertainties or--if the uncertain parameters are too important to
be ignored--the designs are based on worst case values; performance
is necessarily evaluated on a conditional basis. Unsatisfied or con-
fused, some operational analysts add means and variances and treat
an "average situation."

The theory of signal detectability allows the designer and the
user to incorporate what they do know or can learn about the many
uncertain parameters, and, in fact, makes them responsible for

accurate specification of what they do know. Performance is evaluated by treating all possible conditions, and "optimum" means best average performance (although performance can be evaluated on a conditional basis).

This present theoretical thesis introduces mathematical investigations into (1) the performance degradation due to inaccurately specified uncertainties, (2) a re-evaluation of classical statistical principles, in an attempt to salvage relevant work and to discard misleading principles, (3) the feasibility of realizing the receiver as a digital computer, taking into account inherent discrepancies and quantization error that this type of processing will introduce.

This type of mathematical investigation, motivated by practical and physical considerations, is necessary to provide proper guidance to applied mathematicians and theoretical engineers in information theory and signal detection so that they may develop practical design principles and evaluation principles. This is necessary before such theoretical principles can be trusted to guide the development of advanced detection systems.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Cont.)

LIST OF ILLUSTRATIONS (Cont.)

LIST OF SYMBOLS

ROC	receiver operating characteristic
N	noise hypothesis
SN	signal and noise hypothesis
A	decision "signal present"
B	decision "signal absent"
$P(A|N)$	probability of false alarm
$P(A|SN)$	probability of detection
x	observation $x(t)$
$\underline{x}$	observation vector
x_i	observation sample
s	signal $s(t)$
$\underline{s}$	signal vector
s_i	signal sample
W	receiver bandwidth
T	duration of observation interval
$\Gamma(\ )$	Gamma function, $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$
σ^2	noise level, $\sigma^2 = N_0/2$
N_0	noise power per hertz
θ	unknown noise process parameter; also $\theta = \dfrac{1}{\sigma^2}$ in SKE-UNL detection problem

Θ θ parameter space

ψ unknown signal parameter

$\Phi(\ )$ normal distribution function, $\Phi(\alpha) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{\alpha} e^{-\frac{t^2}{2}} dt$

$\omega(\ )$ $\omega(x) = \Phi(x)/\Phi'(x)$

$\ell(\ )$ likelihood ratio

$\ell^*(\ )$ approximate likelihood ratio

$\ell(\ |\ )$ conditional likelihood ratio

$z(\ )$ log-likelihood ratio, $z(\ ) = \ell n\ \ell(\ )$

$z^*(\ )$ approximate log-likelihood ratio

$z(\ |\ )$ conditional log-likelihood ratio

$F(\ )$ probability distribution function

$f(\ )$ probability density function

$f^*(\ )$ approximate probability density function

$f(\ |\ )$ conditional probability density function

$g(\ )$ a priori probability density function, $g(\ )$ is "given"

$g(\ |\ )$ a posteriori probability density function

$h(\ )$ a priori probability density function, $h(\ )$ "holds"

b parameter of the a priori gamma density

b_n parameter of the a posteriori gamma density with respect to SKE-UNL detection problem

c parameter of the a priori gamma density

c_n — parameter of the a posteriori gamma density with respect to SKE-UNL detection problem

d' — detectability parameter on normal ROC

d — detectability parameter on normal ROC, $d = d'^2 = \dfrac{2E}{N_0}$

E_s — total signal energy

SKEA — signal known except for amplitude

SKE-UNL — signal known exactly, unknown noise level

α — threshold

β — threshold on likelihood ratio

Δ — threshold on log-likelihood ratio for SKE-UNL

a — amplitude; also parameter in SKE-UNL detection problem

r — parameter in SKE-UNL detection problem

(u, v) — sufficient statistics of the observation for SKE-UNL

$\xi(\ |H)$ — sufficient statistic given the hypothesis H

$\underline{\mu}$ — mean vector

Σ — autocovariance matrix

Ψ — autocovariance matrix

$\epsilon(\ )$ — $\epsilon(\underline{x}) = z(\underline{x}|h) - z[z(\underline{x}|g)|h]$; also $\epsilon(x) = z^*(x) - z(x)$

$\delta(\ )$ — transformed version of $\epsilon(\ )$

$I(g|h)$ — sensitivity index, $I(g|h) = E[\epsilon(\underline{x})|h, SN] - E[\epsilon(\underline{x})|h, N]$

J — quality index, $J = E[\epsilon(x)|SN]$

$J(1:0)$ — divergence between hypotheses

E   expected value operator

P( )   probability of ( )

w   observation statistic, $w = \underline{x}^t \, \Sigma^{-1} \, \underline{x}$

y   observation statistic, $y = \underline{s}^t \, \Sigma^{-1} \, \underline{x}$

d_0 nominal signal-to-noise ratio, $d_0 = \underline{s}^t \, \Sigma^{-1} \, \underline{s}$

m   mean of a priori amplitude density

v   variance of a priori amplitude density

ρ measure of diffuseness, $\rho = \dfrac{v}{\sigma^2}$

p   upper limit on summation index

$\phi( )$ decision rule

n   number of observations, usually $n = 2WT$

ν number of degrees of freedom for the 't' distribution

$\hat{a}$ minimum variance, unbiased estimator of the unknown amplitude a

a_i pseudo-estimator of amplitude at $t = t_i$

$\overline{a}_i$ MMSE estimator of amplitude at $t = t_i$,

$$\overline{a}_i = E(a \,|\, x_1, x_2, \, \cdots \, x_i, \, SN)$$

CO   conditionally optimum

CS   conditionally suboptimum

ESP   externally sensed parameter

OPT   optimum

$\sim$ is distributed according to

$\uparrow$	is monotone with respect to
$\triangleq$	is defined as
AGC	automatic gain control
LHS	left hand side
RHS	right hand side

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

The problem of detecting a signal in a background of noise is statistical in nature. The state of the environment can only be expressed in a probabilistic sense. Sometimes, even the statistics are not completely known. When the statistics are known except for one or more of the parameters the situation is termed a composite hypothesis problem.

Several authors applied statistical decision theory to develop a theoretical foundation for making the best possible decisions (Refs. 1-3). The theory shows that the optimum decision rule is based on likelihood ratio whenever "good" decisions are preferred to "bad" decisions (Ref. 4).

1.1 Problem Formulation

This work is conceived of as being somewhat expository in nature with one of its main goals being to explain the basic nature of detection devices and thereby reconciling Bayesian philosophy with classical statistics.

The decision process is compounded whenever uncertainties exist either in the noise and/or signal plus noise process. The theory as formulated in the early 1950's (Refs. 1-3) inherently allows for the inclusion of any available information (or lack of information) by

assigning an a priori density to the unknown parameter(s); statistical inference is accomplished by adhering to the Bayesian philosophy. Prior information is modified by actual data (observation) according to Bayes rule, i.e., the a posteriori densities are computed from the a priori densities and the observation statistics according to Bayes rule.

Bayesian philosophy is generally considered to reflect a subjective or personalistic attitude of ideally consistent people (Ref. 5). However, Bayesian philosophy (assignment of a priori densities) can equally well be interpreted to reflect the true state of the unknown parameter(s) by absorbing any environmental data. For instance, the mean of the a priori density might be chosen to be the most probable value (mode) or the mean value (ensemble average) of some function of the available data; the variance of the a priori density would be chosen to indicate the designer's degree of confidence in the data. Hence, a large variance would indicate either that no prior information is available (total ignorance) or that the confidence level in the data is virtually nil; information of this nature is described by what is referred to as a diffuse a priori density. By the same token, a small variance indicates a high level of confidence in the data; and the corresponding a priori density is highly dependent on the data.

The goals of this dissertation are:

a) To consider the effect of a priori specification on receiver design and performance. In particular, a study was conducted to

examine the effect on receiver performance of a priori densities whose means and variances were identical.

b) To examine the relationship between the Bayesian approach to detection theory and that of classical statistics. The Bayesian approach and the classical approach often result in identical decision rules (receiver action); yet they are contrasting viewpoints. Emphasis must be placed on the evaluation of receiver performance conditional to the actual value of the parameter; classical statistics never assigns a priori densities, and in fact the receiver will be operating in an environment in which the parameter is constant though unknown.

c) To develop a method of coping with the inherent integration difficulties that often arise in a composite hypothesis detection problem. Since a priori densities are never really known precisely, it is suggested that numerical integration techniques be employed. The receiver is conceived of as a digital computer; and it is shown that the degradation in performance can be kept negligible.

d) To study the use of estimation techniques in connection with the optimal procedure from the Bayesian point of view. It appears that "estimators" can be defined which yield optimal performance from the Bayesian standpoint; yet the "estimators" are closely allied to the estimators of classical statistics and exhibit characteristics which are considered desirable.

1.2 Previous Literature

There are many textbooks available today which serve as basic background information to decision theory. These textbooks include topics on mathematical statistics, hypothesis testing, statistical decision theory, random processes, information theory and communications engineering, among others (Refs. 6-18).

Much of the work that has been carried on in recent years has been concerned with detection problems in which only signal uncertainties exist (Refs. 19-21). The theory of signal detectability was extended to the double composite hypothesis problem (uncertainties exist in both the noise alone and the signal plus noise process) by Spooner (Ref. 22). Some of the present work is based on his work.

Since this work is intimately concerned with the evaluation of receiver performance in terms of receiver operating characteristics (ROC), the work by Birdsall on ROC curves and their character (Ref. 23) is of prime importance. In addition, the article by Birdsall and Tanner on psychological measures (Ref. 24) presents a useful one-number measure of receiver performance (in lieu of the two-number measure of the ROC) which is appropriate under certain circumstances. The article by Dempster and Schatzoff (Ref. 36) proposed to use the area under the ROC curve as a measure of performance level. This concept, although interesting, is of little value in comparing ROC curves. If neither ROC curve dominates the other (one ROC is superior or equal to the other at all threshold levels), the

measure is useless; even when one ROC does dominate another the proposed measure can easily lend too much significance to a particular characteristic of the ROC's (e.g., it might weight high probabilities of false alarm more than deemed desirable).

Both Bayesian analysis and classical analysis of the composite hypothesis detection problem base decisions on sufficient statistics. Ferguson (Ref. 7) and Raiffa and Schlaifer (Ref. 9) present a good discussion about sufficient statistics. An excellent discussion of reproducing densities (or natural conjugate densities) is presented by both Raiffa and Schlaifer (Ref. 9) and by Spragins (Ref. 25). Spragins aptly demonstrates the relationship of Bayes rule to the existence and functional form of a reproducing density.

The classical approach to testing statistical hypotheses is presented from a fairly theoretical point of view by Lehmann (Ref. 8). Estimation techniques are presented by Deutsch (Ref. 26) and by Van Trees (Ref. 16). A presentation on a measure of information provided by an observation was reported by Lindley (Ref. 27); further information theoretic concepts are discussed by Kullback (Ref. 13) and by Feinstein (Ref. 28).

A thorough discussion of numerical integration techniques can be found in the book by Kopal (Ref. 29); tables of the numerical constants needed are conveniently listed in The Handbook of Mathematical Functions (Ref. 30).

1.3 Procedure

Sensitivity of receiver performance to a priori specification is approached by specifying a nominal or assumed a priori density $g(\cdot)$ and then perturbing $g(\cdot)$ in such a manner that the desired a priori density $h(\cdot)$ has its first few moments equal to those of $g(\cdot)$. A sensitivity index is developed which exhibits several desirable features.

Adherence to Bayesian philosophy inherently leads to receivers based on a monotone function of likelihood ratio. Prior information (opinions, specifications or data) is specified in the form of an a priori density and is modified according to Bayes rule by taking the actual observation into account. Classical statistics, on the other hand, never assigns an a priori density, and instead attempts to design a receiver which operates "well" regardless of the actual operating conditions. Unbiasedness (performance is at least as good as chance) is a criterion often imposed by classical statistics; estimation methods are also applied by the classical approach to detection theory.

For the composite hypothesis situation numerical integration techniques can be used to design a receiver whenever integration difficulties prevent exact analysis. Once the observation statistics are specified and the a priori density assigned, the absolute observation statistics are obtained by using appropriate numerical integration techniques. The particular choice of a numerical integration method depends on the nature (region of definition, prior information available,

etc.) of the unknown parameter. Given the exact form of the a priori density, a quality index is defined, based on information theoretic measures, to determine the effect of the approximate numerical methods employed. Furthermore, the performance of the receiver based on numerical methods is compared to the receiver based on exact (not always possible to do) analysis. A nominal degradation in performance is not considered critical since a priori densities are usually quite subjective anyway.

An attempt to justify estimation as an optimal procedure (the Bayesian approach is considered optimal) is conducted by simulating receiver operation for a particular example and comparing a pseudo-estimator obtained by Bayesian methods to the minimum mean square error estimator at each increment in time.

1.4 Organization

The basic background material is presented in Chapter II. The composite hypothesis detection problem is presented; and the inherent inclusion of prior information in the form of an a priori density is emphasized. Chapter III conducts a study of the sensitivity of receiver performance to a priori specification. Separation of the processing objective from the overall objective is discussed. An index is defined which measures sensitivity of receiver performance to the particular choice of a priori density assigned. In Chapter IV receiver performance via Bayesian methods is compared to receiver

performance via classical methods. The ESP receiver is defined and its important features are discussed. Receiver design via numerical integration techniques is proposed in Chapter V for the composite hypothesis situation whenever exact analysis is difficult or impossible. A simulation of receiver operation based on a pseudo-estimator is presented in Chapter VI to justify the use of estimation methods as a "good" approach to solving signal detection problems.

1.5 Notation

The receiver, as referred to in this work, is a device which computes a test statistic (based upon the actual observation) and makes a decision as to presence or absence of a signal depending on whether or not the test statistic exceeds a pre-assigned threshold. Whenever Bayesian methods are employed, the test statistic is the likelihood ratio or a monotone function thereof. The terms "receiver" and "test statistic" will often be used interchangeably throughout the succeeding chapters.

Since this work is intended to apply primarily to fixed-time detection theory, an observation will be denoted by an n-dimensional column vector unless otherwise specified, i.e.,

$$\underline{x} = (x_1, x_2, \cdots x_n)^t$$

(A column vector will be denoted as the transpose of a row vector to save space; in this connection t denotes transpose.)

Since much of this work deals with or relates to the normal detection situation, it is convenient to define the normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

Probability density functions are denoted in symbolic form. For instance, $f(\underline{x}|H)$ is the probability density function of the observation $\underline{x}$ conditional to the hypothesis H. Likewise, $g(\theta|\underline{x}, H)$ is the a posteriori probability density function of the random variable θ given the observation $\underline{x}$ and the hypothesis H. A priori densities will generally be denoted by $g(\theta)$.

The symbol $\sim$ denotes "is distributed according to." For instance,

$$\underline{x} \sim MVN_n(\underline{\mu}, \Sigma)$$

indicates that the observation $\underline{x}$ is distributed according to a multivariate normal probability distribution of dimension n with mean vector $\underline{\mu}$ and autocovariance matrix Σ. The symbol $\dagger$ denotes "is monotone with respect to." For example, $\ell(x) \dagger x$ means that $\ell(x)$ is a monotone function of x.

The likelihood ratio of an observation $\underline{x}$ given that the assigned a priori density is $g(\theta)$ will be denoted by $\ell(\underline{x}|g)$. The likelihood ratio will be simply denoted by $\ell(\underline{x})$ when no uncertainties exist or

when no ambiguities exist as to the assigned a priori density.

The expectation operator will be denoted by E ; and the variance-covariance operator will be denoted by Var. Hence, if $\underline{\mu}$ represents the mean vector and Σ the autocovariance matrix of the observation $\underline{x}$, the following relationships exist:

$$\underline{\mu} = E(\underline{x})$$

$$\Sigma = Var(\underline{x}) = E(\underline{x}\,\underline{x}^t) - \underline{\mu}\,\underline{\mu}^t$$

The determinant of the autocovariance matrix Σ will be denoted by $|\Sigma|$.

A double notation is often used throughout this work for convenience. For example,

$$P(A \mid \frac{N}{SN}) = \int_{\ell(x) > \beta} f(x \mid \frac{N}{SN})\ dx$$

means

$$P(A \mid N) = \int_{\ell(x) > \beta} f(x \mid N)\ dx$$

$$P(A \mid SN) = \int_{\ell(x) > \beta} f(x \mid SN)\ dx$$

Other definitions and explanations will be made as they occur in the text.

CHAPTER II

REVIEW OF FIXED-TIME DETECTION THEORY

The theory of signal detectability was aptly formulated by Peterson, Birdsall and Fox (Ref. 1) in 1954. This theory is now called classical fixed-time decision theory. The basic theory was extended by Nolte, Roberts and Spooner. Nolte (Ref. 21) considered the problem of detecting a recurrence phenomenon in noise for different degrees of certainty about recurrence time while Roberts (Ref. 19) expounded on composite deferred decision theory. Roberts formulated the optimum stopping rule for the sequential detection of a composite signal for which the response time is constrained to occur within the observation interval. The cost of making an observation is taken into account. Spooner (Ref. 22) developed a general optimum processor for the double composite hypothesis situation. All uncertain noise and/or signal plus noise process parameters were expressed in terms of a priori densities. Since the present work is based on the theory of signal detectability, it is appropriate that this theory be reviewed.

2.1 Introduction

The basic signal detection problem is presented schematically in Fig. 2.1. The noise process is denoted by $n(t)$ while $s(t)$ describes the signal. The receiver is presented with an observation $x(t)$ during

11
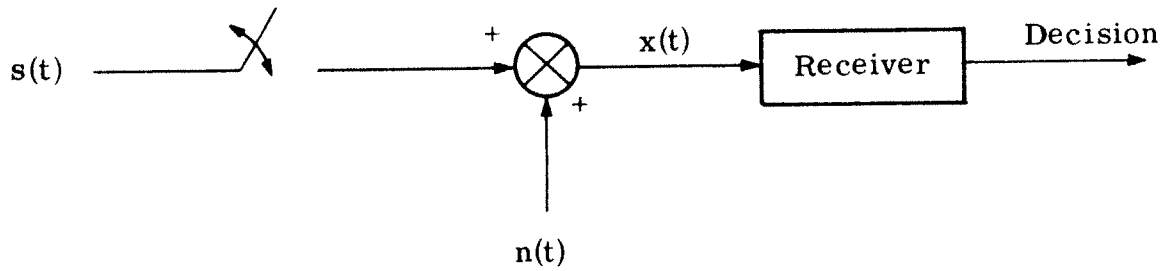
Figure 2.1. Illustration of the basic signal detection problem.

a time interval $t_0 \leq t \leq t_0 + T$. The observation consists of either

noise alone or signal plus noise. On the basis of this observation the

receiver must decide whether or not a signal had been present during

the observation interval. This observation-decision task of the

receiver can be formulated from a decision theory viewpoint as a

hypothesis test.

$$N: \quad x(t) = n(t)$$

$$SN: \quad x(t) = n(t) + s(t) \qquad t_0 \leq t \leq t_0 + T \qquad (2.1)$$

The hypotheses are mutually exclusive. That is, either the signal

was present during the entire observation interval or it was absent

during the entire observation interval.

In order to utilize statistical decision theory, it is customary

to describe the random process x(t) by a series representation. If

the observation x(t) is timelimited to an interval of length T and

Fourier Series bandlimited to a band of frequencies of width W , then

the Shannon sampling theorem can be employed to represent the observation $x(t)$ by

$$\underline{x} = (x_1, x_2, \ldots x_n)^t \tag{2.2}$$

where

$$n = 2WT \tag{2.3}$$

$$x_i = x\left(t_0 + \frac{i}{2W}\right) \qquad i = 1, 2, \ldots n \tag{2.4}$$

A more general series expansion referred to as the Karhunen-Loeve expansion is discussed in Ref. 16 and represents a random process $x(t)$ in terms of a complete orthonormal set of functions $\phi_i(t)$ so that

$$x(t) = \lim_{N \to \infty} \sum_{i=1}^{N} x_i \phi_i(t) \qquad t_0 \leq t \leq t_0 + T \tag{2.5}$$

where

$$x_i \overset{\Delta}{=} \int_{t_0}^{t_0+T} x(t) \phi_i(t) \, dt \tag{2.6}$$

Subject to certain regularity conditions, it is shown that there exists a set of $\phi_i(t)$ that leads to uncorrelated coefficients and consequently assures convergence in the mean-square sense. The second-moment characterization of the Karhunen-Loeve expansion makes it possible to represent a Gaussian process in terms of an at most countably

infinite set of statistically independent Gaussian random variables.

2.2 Receiver Design

The receiver will be presented with an observation x(t) during a time interval $t_0 \leq t \leq t_0 + T$ and make a decision as to presence or absence of a signal at the end of the observation interval. In order to optimize receiver performance, we need to specify a particular criterion.

2.2.1 Criteria.

The observation-decision task of the receiver results in either one of two possible correct decisions or one of two possible incorrect decisions. The receiver is capable of making two distinctly different responses,

> A   is the response "signal is present"
>
> B   is the response "signal is absent"

to two possible hypotheses,

> SN   the hypothesis "signal plus noise"
>
> N   the hypothesis "noise alone"

Associated with each response is the possibility of error since noise is present throughout the processing. The probabilities of error and correct decision are denoted by

> P(A|N)   probability of false alarm
>
> P(A|SN)   probability of a correct detection

P(B|N)   probability of a correct rejection

P(B|SN)   probability of a miss

These probabilities are not independent since

$$P(A|N) + P(B|N) = 1 \tag{2.7}$$

$$P(A|SN) + P(B|SN) = 1 \tag{2.8}$$

A cost may be associated with each possible outcome. The cost structure of the decision process is illustrated in Fig. 2.2.

	SN	N
A	C_D	C_F
B	C_M	C_Q

Fig. 2.2. Cost structure of the decision process

The cost of a correct detection is C_D, the cost of a false alarm is C_F, the cost of a miss is C_M and the cost of a correct rejection is C_Q.

Several different criteria may be appropriate depending upon the user's goals or objectives. The Bayes criterion minimizes the average cost C where

$$C = P(SN) \ [C_D \ P(A|SN) + C_M \ P(B|SN)]$$

$$+ \ P(N) \ [C_F \ P(A|N) + C_Q \ P(B|N)] \qquad (2.9)$$

The Neyman-Pearson criterion maximizes $P(A|SN)$ subject to the constraint

$$P(A|N) \leq \alpha$$

where α is a pre-assigned probability. The weighted combination criterion maximizes

$$P(A|SN) - w \ P(A|N)$$

where w is a positive constant relating the costs involved. Other criteria exist with analogous objectives.

2.2.2 Likelihood Ratio Receiver. It has been shown (Ref. 23) that the receiver which bases its decision on likelihood ratio yields optimum performance for a wide class of criteria including those described above. In particular the decision rule is

$$\phi(x) = \begin{cases} 1 & \ell(x) > \beta \\ r & \ell(x) = \beta \\ 0 & \ell(x) < \beta \end{cases} \qquad (2.10)$$

where

$$\ell(x) = \frac{f(x|SN)}{f(x|N)} \qquad (2.11)$$

is the likelihood ratio, β is the pre-assigned threshold and $\phi(x)$ is

the probability of deciding that a "signal is present" with $0 \leq r \leq 1$.

(Whenever $\ell(x) > \beta$, the decision is made that a "signal is present.")

The situation that occurs when $\ell(x) = \beta$ describes a randomized de-

cision rule. Whenever the observation densities $f(x|SN)$ and $f(x|N)$

are analytic, the occurrence of $\ell(x) = \beta$ has probability zero and

hence the value of r is of no consequence whatsoever.

Only the threshold depends upon the particular criterion invoked.

Birdsall (Ref. 4) has shown that the likelihood ratio receiver yields

optimum performance for any choice of criterion which considers

incorrect decisions "bad" and correct decisions "good". This power-

ful result makes it possible to evaluate the merits of any other decis-

ion device by comparing its performance to the performance of the

likelihood ratio receiver.

2.2.2.1 Simple Hypothesis. Whenever no uncertainties

exist either in the N and/or SN process, the likelihood ratio is

easily determined. It is

$$\ell(x) = \frac{f(x|SN)}{f(x|N)} \tag{2.12}$$

2.2.2.2 Composite Hypothesis. For the simple hypothesis

problem the N and SN process statistics were known precisely. For

the composite hypothesis problem either N and/or SN process

parameters are unknown. Any available information concerning these

parameters is expressed in the form of an a priori density. The available information might consist of either the designer's subjective opinion or be specified by the range of parameter definition. That is, if the current state of the unknown parameter is known reasonably well (i.e., the designer possesses a high level of confidence that the value of the unknown parameter is within a given range of values), then the unknown parameter will be modeled by a proper a priori density (the normalization constant exists so that normally the mean and variance exist). On the other hand, if the state of the unknown parameter is entirely unknown (i.e., the designer possesses total ignorance), the unknown parameter might be modeled by a diffuse a priori density (the normalization constant does not exist). A diffuse a priori density in essence expresses the opinion that all admissible states of the unknown parameter are equally likely.

2.2.2.2.1 Single Composite Hypothesis. When uncertainties exist in the signal parameter(s), the situation is referred to as a single composite hypothesis problem. If we denote the unknown signal parameter(s) by ψ, then the absolute observation statistics can be expressed in terms of the conditional observation statistics and the assumed a priori density $g(\psi)$ as

$$f(x \mid SN) = \int_{\Psi} f(x \mid \psi, SN) \, g(\psi) \, d\psi \qquad (2.13)$$

The observation statistics conditional to N remain the same since only signal uncertainties are involved. Then

$$\ell(x) = \frac{f(x \mid SN)}{f(x \mid N)}$$

$$= \int_{\Psi} \ell(x \mid \psi) \, g(\psi) \, d\psi \qquad (2.14)$$

where

$$\ell(x \mid \psi) = \frac{f(x \mid \psi, SN)}{f(x \mid N)} \qquad (2.15)$$

The total likelihood ratio is therefore obtained by averaging the conditional likelihood ratio over all possible states.

2.2.2.2.2 Double Composite Hypothesis.

In the double composite hypothesis problem uncertainties exist in both the N and SN process. Uncertainty of a noise process parameter inherently leads to a double composite hypothesis since noise is present conditional to either hypothesis. Let ψ denote the uncertain signal parameter(s) and θ the unknown noise process parameter(s), and assume θ and ψ are independent. Again, the absolute observation statistics can be expressed in terms of the conditional observation statistics and the given a priori densities $g(\psi)$ and $h(\theta)$ as

$$f(x \mid SN) = \int_{\Psi} \int_{\Theta} f(x \mid \psi, \theta, SN) \, g(\psi) \, h(\theta) \, d\theta \, d\psi \qquad (2.16)$$

$$f(x \mid N) = \int_{\Theta} f(x \mid \theta, N) \, h(\theta) \, d\theta \qquad (2.17)$$

The total likelihood ratio then becomes

$$\ell(x) = \int_{\Psi} \ell(x \mid \psi, h) \, g(\psi) \, d\psi \qquad (2.18)$$

where

$$\ell(x \mid \psi, h) = \frac{\int_{\Theta} f(x \mid \psi, \theta, SN) \, h(\theta) \, d\theta}{\int_{\Theta} f(x \mid \theta, N) \, h(\theta) \, d\theta} \qquad (2.19)$$

It is apparent that the receiver is now constrained to operate in a dual channel mode, each channel operating on a different hypothesis. This feature is inherent for the double composite hypothesis problem and will become apparent in the work which follows.

2.3 Receiver Realization

Receiver realization consists of the implementation of equipment in any fashion that realizes the likelihood ratio or a monotone function of the likelihood ratio. The receiver may be realized either sequentially in time or the entire observation may be processed. Information processing may occur in either analog or digital form. Hence, the receiver might be a matched filter or it might be a digital computer. The availability of high speed digital computers and sophisticated processing techniques has made it feasible to employ computers to do the processing for the optimum receiver in real time.

Optimum processing of the observation $x(t)$ according to Bayesian methods also permits the extraction of information useful for classification and/or estimation purposes. Prior information or

opinions are modified by data in terms of a posteriori densities.

This learning or classification output as expressed in terms of

a posteriori densities must be used only in conjunction with the detec-

tion output since learning only occurs with respect to knowledge of the

true hypothesis. Both the detection output and the classification out-

put depend upon the observation x(t) only through its sufficient statis-

tics. Receiver realization of the optimum receiver in terms of the

sufficient statistics usually reduces the dimensionality of the problem,

thereby resulting in a fixed memory receiver regardless of the length

of the observation interval T .

2.4 Receiver Evaluation

Receiver evaluation is succinctly summarized by the receiver

operating characteristic (ROC). The ROC is a convenient way of por-

traying the quality of detection (Ref. 23). It is merely a plot of the

probability of false alarm, P(A|N), versus the probability of detection,

P(A|SN), for all possible threshold levels of the receiver output. For

any receiver these probabilities are

$$P(A|N) = E(\phi(x)|N) \tag{2.20}$$

$$P(A|SN) = E(\phi(x)|SN) \tag{2.21}$$

where $\phi(x)$ is the decision rule. When the receiver is the optimum

likelihood ratio receiver these relations are

$$P(A|N) \quad = \int_{\ell(x) > \beta} f(x|N) \, dx \qquad (2.22)$$

$$P(A|SN) \quad = \int_{\ell(x) > \beta} f(x|SN) \, dx \qquad (2.23)$$

By the obvious transformation

$$\ell \;=\; \ell(x) \qquad (2.24)$$

these relations become

$$P(A|N) \quad = \int_{\beta}^{\infty} f(\ell|N) \, d\ell \qquad (2.25)$$

$$P(A|SN) \quad = \int_{\beta}^{\infty} f(\ell|SN) \, d\ell \qquad (2.26)$$

Further simplification can be obtained by utilizing a fundamental

theorem of decision theory (Ref. 23) which states "the likelihood

ratio of the likelihood ratio is the likelihood ratio." That is,

$$\ell(\ell) \;=\; \frac{f(\ell|SN)}{f(\ell|N)} \;=\; \ell \qquad (2.27)$$

so that the parametric equations describing the ROC curve for the

likelihood ratio receiver become

$$P(A|N) \quad = \int_{\beta}^{\infty} f(\ell|N) \, d\ell \qquad (2.28)$$

$$P(A \mid SN) = \int_{\beta}^{\infty} \ell \, f(\ell \mid N) \, d\ell \qquad (2.29)$$

Although only the density function of the likelihood ratio conditional to N needs to be determined to evaluate receiver performance, this task is often formidable. Existence of sufficient statistics and some facility with statistical transformations will usually simplify the problem, although it may still be necessary to resort to numerical integration techniques.

A standard for comparing ROC curves is the normal ROC. An ROC is called normal if it can be parameterized by the normal distribution as follows:

$$P(A \mid N) = \Phi(\lambda) \qquad (2.30)$$

$$P(A \mid SN) = \Phi(\lambda + d') \qquad (2.31)$$

where

$$\Phi(\lambda) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\lambda} e^{-\frac{t^2}{2}} \, dt \qquad (2.32)$$

It is evident that a normal ROC curve can be characterized by the one parameter d'. The parameter d' is usually referred to as the quality of detection. For convenience normal ROC curves are usually plotted on normal-normal paper so that normal ROC curves become linear. A family of normal ROC curves with parameter d' is plotted

on normal-normal paper and displayed in Fig. 2.3.
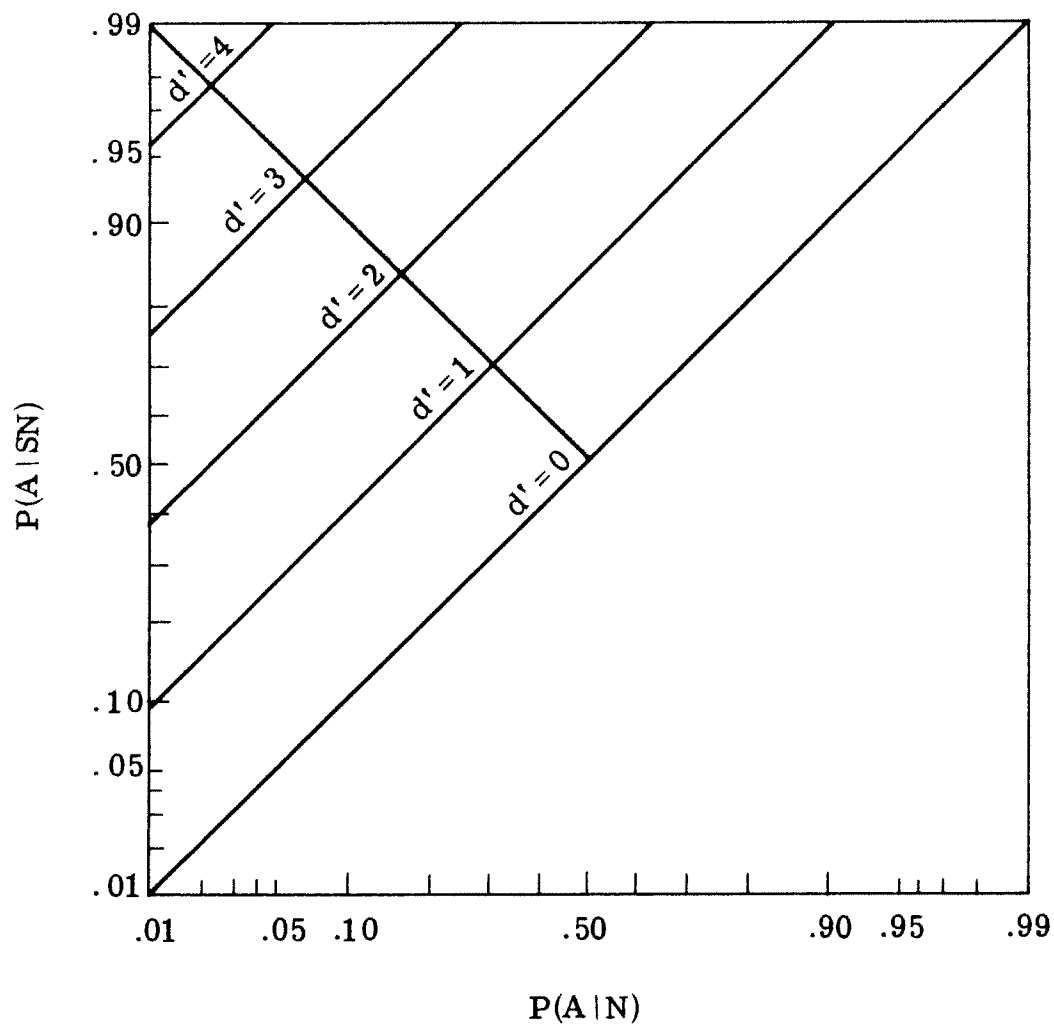
Figure 2.3. Normal ROC's with detectability parameter d'.

The utility of normal ROC curves lies in the fact that physical significance can be attributed to d'. For the detection problem signal known exactly in known Gaussian noise (SKE + KGN), the

performance of the optimum receiver is described by a normal ROC with parameter

$$d' = \sqrt{\frac{2E}{N_0}} \qquad (2.33)$$

where E is the signal energy and N_0 is the noise power per hertz.

The normal ROC provides a basis for the comparison of ROC curves. When ROC curves are almost normal the equivalent detection index d'_e as measured on the negative diagonal $[P(A|N) + P(A|SN) = 1]$ is indicative of the quality of detection and serves as a convenient quantitative measure of performance.

CHAPTER III

SENSITIVITY OF RECEIVER PERFORMANCE

TO A PRIORI SPECIFICATION

3.1 Introduction

One of the first contributions of classical fixed-time detection theory was the separation of the processing objective from the overall objective. The optimum receiver has the task of formulating the likelihood ratio $\ell(x)$ of the entire observation interval. A decision as to presence or absence of a signal is made by thresholding the likelihood ratio. The threshold value β is determined independently of the likelihood ratio. It reflects the goal of the user and may incorporate such things as costs, a priori probabilities, a desired false alarm rate, etc. The separation of roles is clearly illustrated by the block diagram of Fig. 3.1.
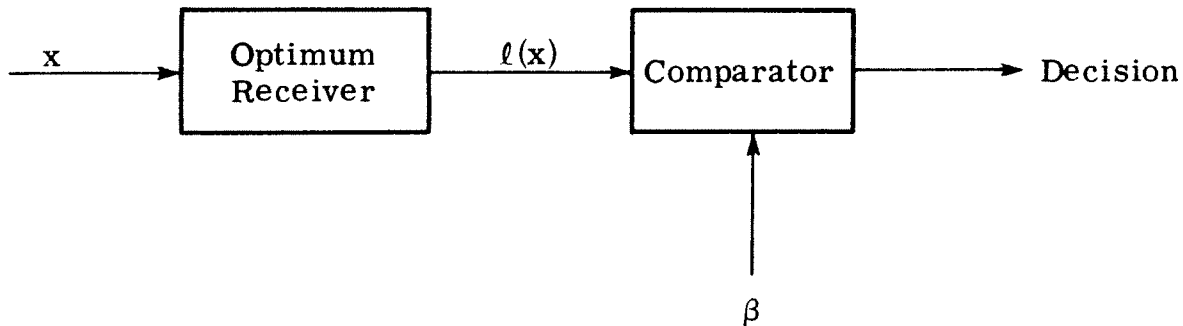
Figure 3.1. Separation of the processing objective from the overall objective.

Another important contribution to the theory of signal detectability was recently made by Birdsall (Ref. 31). He showed that the optimum processor for the composite hypothesis situation can be formulated as two cascaded processors as illustrated in Fig. 3.2.
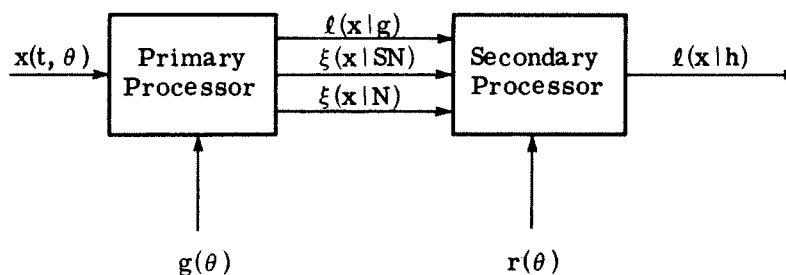
Figure 3.2 Optimum processor for the composite hypothesis problem as two cascaded processors.

For simplicity assume that the unknown parameter is θ . Furthermore, let the primary processor be designed on the basis that the unknown parameter θ has a given a priori density $g(\theta)$, but at the time of use the a priori density $h(\theta)$ holds; $h(\theta)$ is related to $g(\theta)$ by the Radon-Nikodym derivative $r(\theta)$ such that

$$h(\theta) = r(\theta) g(\theta) \tag{3.1}$$

Then the equation describing the secondary processor is essentially Eq. 39 of Ref. 31 and is

$$\ell(x \mid h) = \ell(x \mid g) \cdot \frac{\int_{\Theta} r(\theta) \, g(\theta \mid \xi(x \mid SN), SN) \, d\theta}{\int_{\Theta} r(\theta) \, g(\theta \mid \xi(x \mid N), N) \, d\theta} \qquad (3.2)$$

This relationship can be described equivalently as

$$\ell(x \mid h) = \ell(x \mid g) \cdot \frac{E[r(\theta) \mid x, g, SN]}{E[r(\theta) \mid x, g, N]} \qquad (3.3)$$

The a priori density $h(\theta)$ is the density with respect to which the user wishes to maximize performance. It may reflect his personal opinion as to the information available, or it may reflect the true state of nature with respect to the possible values θ can assume. That is, the a priori density $h(\theta)$ could either reflect a subjective point of view or conviction, or it might reflect the actual physical situation, or even a set of specifications.

The primary processor forms the likelihood ratio of the observation based on the conditional observation statistics and a natural conjugate a priori density $g(\theta)$ (provided that it exists). In addition it also determines the sufficient statistics $\xi(x \mid N)$ and $\xi(x \mid SN)$ of the observation. The secondary processor utilizes the output of the primary processor along with knowledge of the a priori density $h(\theta)$, in the form of a Radon-Nikodym derivative $r(\theta)$, to form the likelihood ratio of the observation based on the desired a priori density $h(\theta)$. This partitioning of information has made it possible to design and build a major portion of the processing equipment based on

mathematically tractable functions and to allow the exact goals of the receiver to be specified at a later time without any degradation of performance. This should not be construed to mean that the design and construction of the secondary processor is elementary. Quite to the contrary, it is often a very complex and difficult task.

3.2 Sensitivity Study

Let θ be an unknown noise process parameter; and consider a receiver designed to be optimum with respect to the given a priori density $g(\theta)$, but whose performance is evaluated with respect to the a priori density $h(\theta)$. [Mnemonically, g is "given" but h "holds".] The receiver is necessarily suboptimum with respect to the actual operating conditions. A block diagram illustrating the decision process is shown in Fig. 3.3.
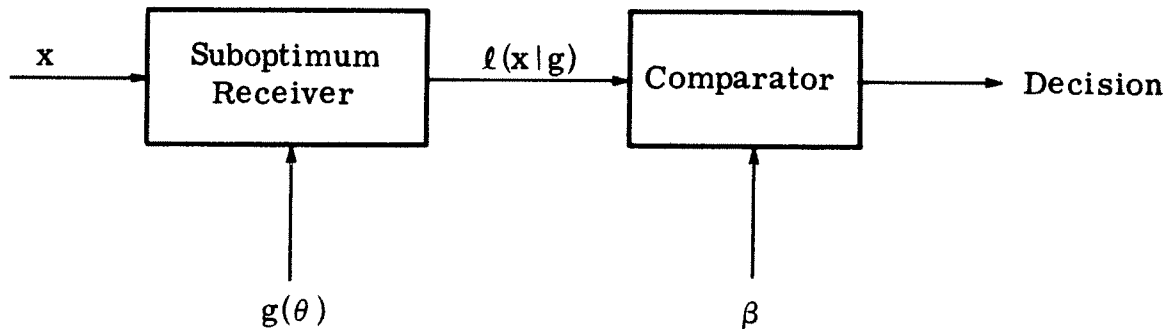
Figure 3.3. Decision process based on the given a priori density $g(\theta)$.

Performance via ROC

The performance equations of the suboptimum receiver are

$$P(A \mid g, h, \underset{SN}{N}) = \int_{\ell(x \mid g) > \beta} f(x \mid h, \underset{SN}{N}) \, dx \qquad (3.4)$$

with

$$f(x \mid h, \underset{SN}{N}) = \int_{\Theta} f(x \mid \theta, \underset{SN}{N}) \, h(\theta) \, d\theta \qquad (3.5)$$

$$f(x \mid g, \underset{SN}{N}) = \int_{\Theta} f(x \mid \theta, \underset{SN}{N}) \, g(\theta) \, d\theta \qquad (3.6)$$

and

$$\ell(x \mid g) = \frac{f(x \mid g, SN)}{f(x \mid g, N)} \qquad (3.7)$$

The optimum receiver would have based its decision on $\ell(x \mid h)$ where

$$\ell(x \mid h) = \frac{f(x \mid h, SN)}{f(x \mid h, N)} \qquad (3.8)$$

Its ROC will serve as an upper bound for the ROC of any receiver operating under these conditions. The performance equations of the optimum receiver are

$$P(A \mid h, h, \underset{SN}{N}) = \int_{\ell(x \mid h) > \beta} f(x \mid h, \underset{SN}{N}) \, dx \qquad (3.9)$$

Figure 3.4 sketches the idea of comparing the ROC of Eq. 3.4 (ROC G) to the ROC of Eq. 3.9 (ROC H).

P(A|SN)
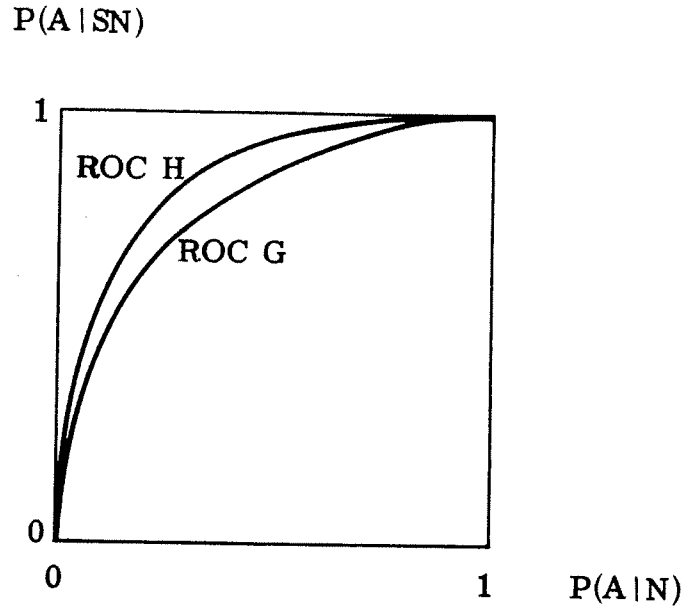
Figure 3.4. Comparison of optimum and suboptimum receiver performance.

We are not attempting to compare $g(\theta)$ to $h(\theta)$; rather, we wish to compare performance of the receivers based on $g(\theta)$ and $h(\theta)$, under the condition that $h(\theta)$ holds. Ideally, given that $h(\theta)$ holds, it is desired to compare the performance of the suboptimum receiver based on $g(\theta)$ to the performance of the optimum receiver based on $h(\theta)$ via the ROC's of Eqs. 3.4 and 3.9; but it is often very difficult to evaluate performance for a specific example, even with the use of a high speed digital computer and sophisticated numerical integration techniques.

Performance via index

An alternative way of studying sensitivity of receiver perform-
ance to a priori specification would be to construct a procedure for
ordering ROC curves, i.e., find a meaningful and consistent way of
representing the two-number description of the ROC by one number.
Several authors have studied this task; and they suggest the use of
various indices. Several of these are reviewed by Birdsall (Ref. 23).
Kullback and Leibler (Ref. 34) used the "divergence between hypotheses"

$$J(1:0) \overset{\Delta}{=} E[z(x)|SN] - E[z(x)|N] \qquad (3.10)$$

where

$$z(x) = \ln \ell(x) \qquad (3.11)$$

The divergence $J(1:0)$ was first introduced by Jeffreys (Ref. 35) in
another connection. $J(1:0)$ is a measure of the "distance" or "diver-
gence" between the hypotheses SN and N and is a measure of the
ease of discriminating between them. The divergence $J(1:0)$ has
all the properties of a metric as defined in topology except the triangle
inequality property (Ref. 13). For normal ROC curves, $J(1:0) = d$
where

$$d = \frac{2E}{N_0} \qquad \text{(signal-to-noise ratio)} \qquad (3.12)$$

Development

Following this lead and given that $h(\theta)$ holds, a performance index of the ROC based on $z(x|h)$ can be defined as

$$J_h(1:0) \overset{\Delta}{=} E[z(x|h)|h, SN] - E[z(x|h)|h, N] \qquad (3.13)$$

It is tempting to define an analogous performance index of the ROC based on $z(x|g)$ [given that $h(\theta)$ holds] as

$$J_g'(1:0) = E[z(x|g)|h, SN] - E[z(x|g)|h, N] \qquad (3.14)$$

However, the relation of $z(x|g)$ to the a priori density $h(\theta)$ is most important. Picture the situation at the moment of use of the suboptimum receiver. The user knows the value of $z(x|g)$ for the observation, say z_0, and he knows $h(\theta)$. To get the best possible performance he should base his decision on the likelihood ratio of what he has, with respect to what he knows. That is, he should calculate $\ell(z_0|h)$, or equivalently $z(z_0|h)$. We therefore give the user credit for using the suboptimum test statistic $z(x|g)$ in the best possible manner by constructing a performance index based on $z[z(x|g)|h]$, namely

$$J_g(1:0) \overset{\Delta}{=} E\{z[z(x|g)|h]|h, SN\} - E\{z[z(x|g)|h]|h, N\} \qquad (3.15)$$

The transformation $z[z(x|g)|h]$ can be considered as a means of reducing the test statistic $z(x|g)$ to a "common denominator" with respect to the a priori density $h(\theta)$.

density $g(\theta)$. Obviously, A^{-1} does not exist if the moments do not exist. Even if the moments exist, the existence of A^{-1} cannot be assured in general and must be verified for the particular application in mind. If we define

$$\underline{a} = (a_1, a_2, \ldots a_p)^t = A^{-1} \underline{\alpha} \qquad (3.30)$$

then

$$\underline{r} = (1 - r_0) \underline{a} \qquad (3.31)$$

and explicit dependence of the polynomial $r(\theta)$ on r_0 is

$$r(\theta) = r_0 + (1 - r_0) \sum_{i=1}^{p} a_i \theta^i \qquad r_0 \neq 1 \qquad (3.32)$$

Although it appears that r_0 is arbitrary, the restriction $r(\theta) \geq 0$ for all $\theta \epsilon \Theta$ limits the permissible values of r_0.

3.2.2 Signal Known Exactly, Unknown Noise Level. To determine the effect of the equality of moments on performance, a study was conducted using as an example the detection problem, signal known exactly, unknown noise level (SKE-UNL). This detection problem was extensively investigated by Spooner (Ref. 22). His study used a gamma density as the a priori density of the reciprocal of the unknown noise power. The same a priori density will be used in this study. In particular he set

$$\theta = \frac{1}{\sigma^2(x \mid N)} \tag{3.33}$$

and chose

$$g(\theta) = \frac{c^{b+1}}{\Gamma(b+1)} \, \theta^b \, e^{-c\theta} \qquad b > -1, \ c > 0, \ 0 \le \theta < \infty \tag{3.34}$$

Some typical densities of this form are shown in Fig. 3.5. The moments of $g(\theta)$ are easily computed, namely

$$\alpha_k = E(\theta^k \mid g) = \int_0^\infty \theta^k \, g(\theta) \, d\theta$$

$$= \frac{\Gamma(b+1+k)}{c^k \, \Gamma(b+1)}$$

$$= \frac{(b+1)(b+2) \ldots (b+k)}{c^k} \tag{3.36}$$

3.2.2.1 Preliminaries. It is shown in Appendix J that

$$a_i = d_i \, c^i \qquad i = 1, 2, \ldots p \tag{3.37}$$

with d_i independent of c. Substituting Eq. 3.37 into Eq. 3.32 yields $r(\theta)$ in a form which explicitly shows its dependence on r_0 and c, namely

$$r(\theta) = r_0 + (1 - r_0) \sum_{i=1}^p d_i (c\theta)^i \qquad r_0 \ne 1 \tag{3.38}$$
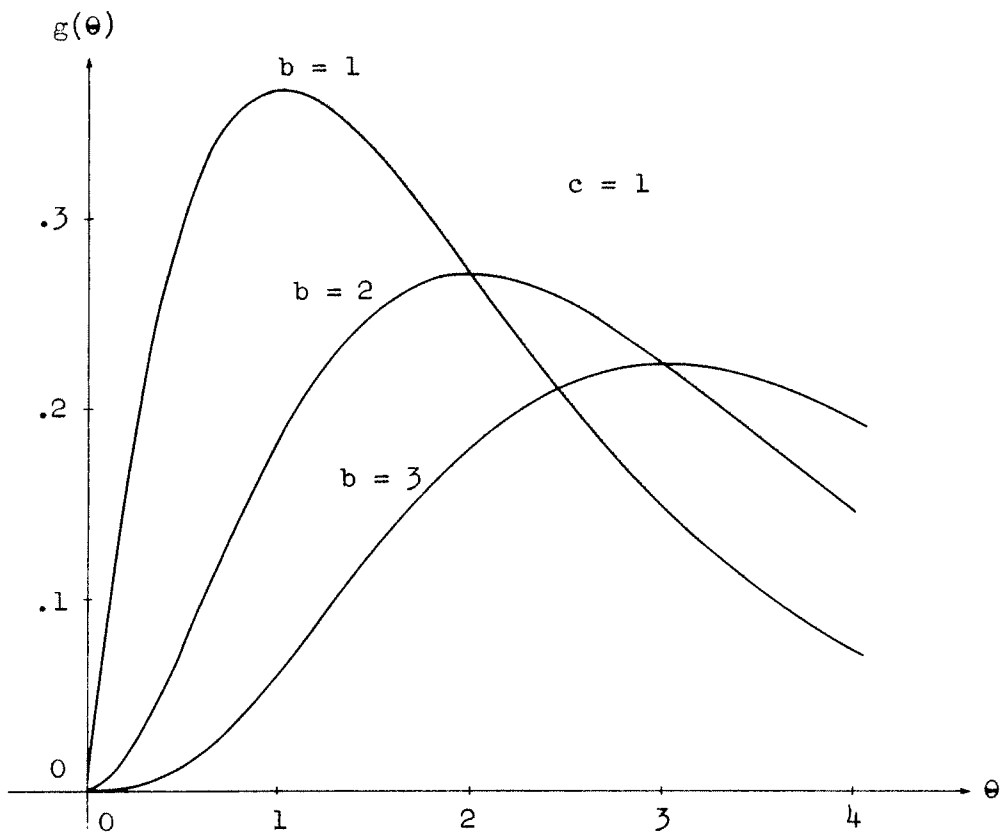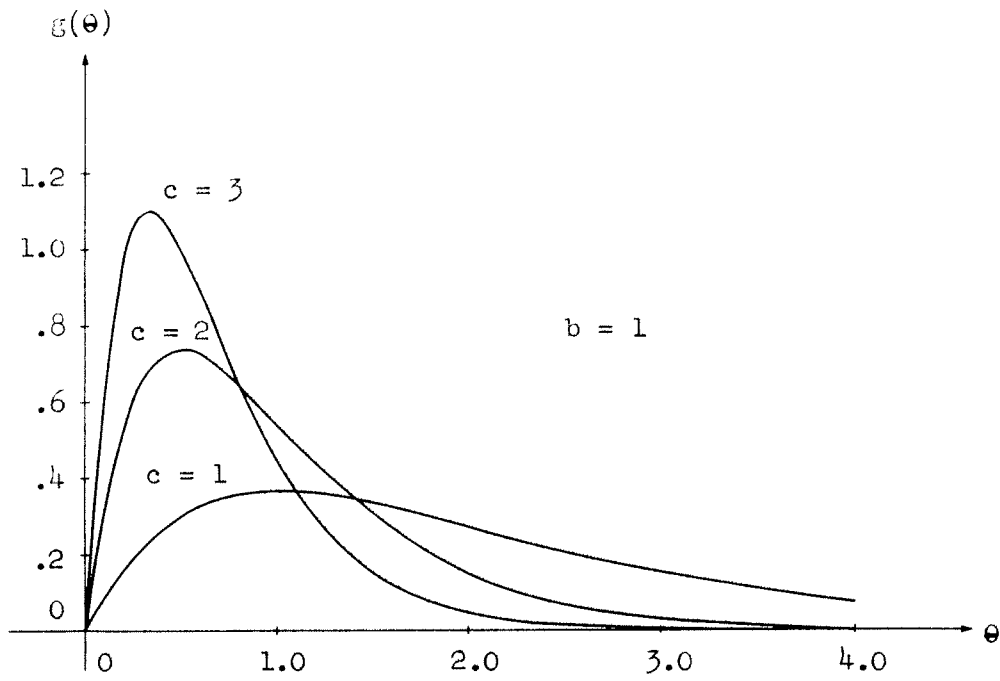
Figure 3.5. Typical probability density functions from the gamma family.

If we let

$$f_0 = r_0$$

$$f_i = (1 - r_0) d_i \qquad i = 1, 2, \dots p \qquad (3.39)$$

then the polynomial $r(\theta)$ can be expressed in the form

$$r(\theta) = \sum_{i=0}^{p} f_i (c\theta)^i \qquad (3.40)$$

with f_i independent of c.

The restriction $r(\theta) \geq 0$ for all $\theta \in [0, \infty)$ limits the permissible values of r_0. It is shown in Appendix A that

$$r_0 \in [r_0^*, 1) \quad \text{if} \quad d_p > 0$$

or $\qquad r_0 \in (1, r_0^*] \quad \text{if} \quad d_p < 0 \qquad (3.41)$

with r_0^* independent of c.

In particular, it is shown in Appendix B that for $p = 2$

$$r_0^* = b + 2 \qquad b > -1 \qquad (3.42)$$

while for $p = 3$

$$r_0^* = \begin{cases} \dfrac{2\sqrt{b+3} - b(b+3)}{2(\sqrt{b+3} + 1)} & -1 < b < 1 \\ \\ 0 & b \geq 1 \end{cases} \qquad (3.43)$$

A plot of the extreme value r_0^* is shown in Fig. 3.6 for several values of p.

A comparison of $g(\theta)$ $(r_0=1)$ and $h(\theta)$ is shown in Fig. 3.7 for p = 3. Recall that for p = 3 the means and variances match. Observe that the given a priori density $g(\theta)$ and the desired a priori density $h(\theta)$ differ noticeably for $r_0 = r_0^*$. As r_0 varies from unity to the extreme permissible value r_0^*, $h(\theta)$ changes from a unimodal to a bimodal density. (This is not obvious in the figure since the second mode occurs at $\theta > 4$.)

The effect of matching moments will be shown in Section 3.2.2.4; changes in relative performance level will be studied via the sensitivity index $I(g|h)$ as developed in Section 3.2.

3.2.2.2 Receiver Design. The receiver is a device which bases its decisions on $z(x|g)$. It is designed on the basis of the given a priori density $g(\theta)$, although in fact $h(\theta)$ holds. Receiver design for the detection problem SKE-UNL was originally developed by Spooner (Ref. 22). His work is the basis for this section.

Assume that the input observation x(t) is timelimited to the observation interval (0, T) and (Fourier Series) bandlimited to a band of frequencies of width W. Then the input waveform x(t) can be
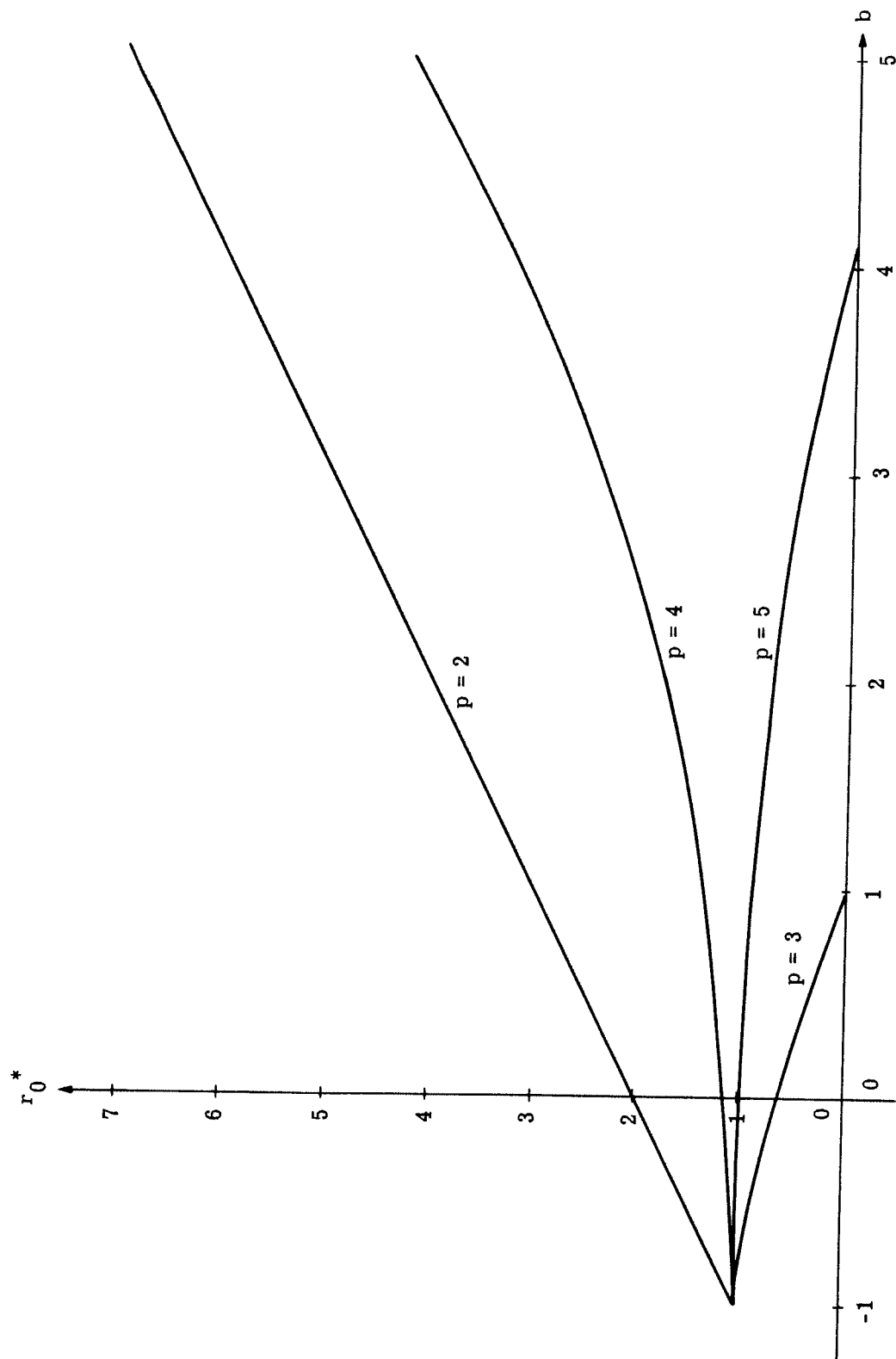
Figure 3.6. Extreme value r_0^*.
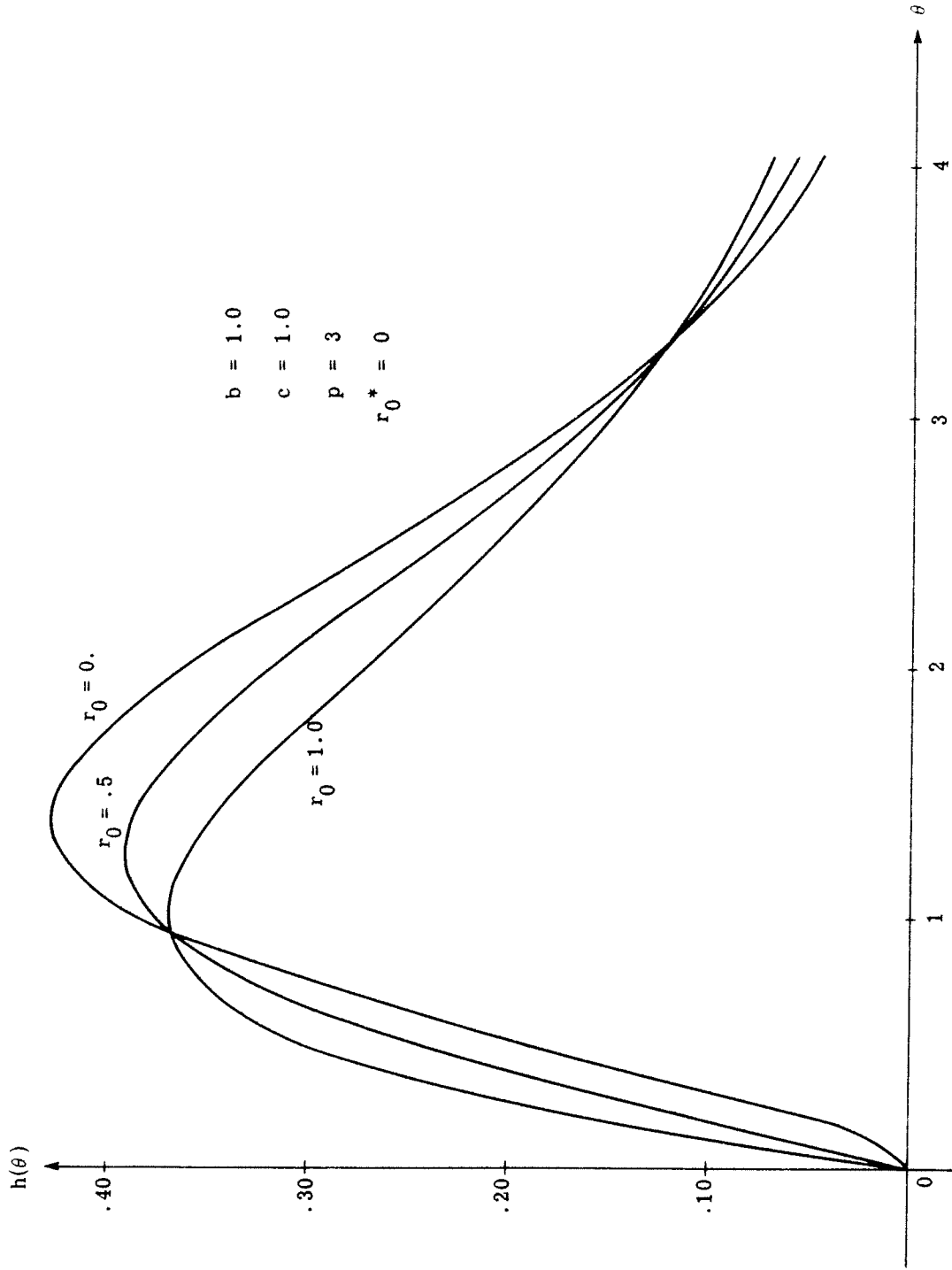
Figure 3.7. Comparison of probability density functions.

represented as a vector in 2WT dimensional space by employing the sampling theorem. Thus we can characterize x(t) by

$$\underline{x} = (x_1, x_2, \ldots, x_n)^t \tag{3.44}$$

where

n = 2WT, the dimensionality of the space

W = bandwidth of input $x(t)$

T = total duration of the observation

and

$$x_i = x\left(\frac{i}{2W}\right) \qquad i = 1, 2, \ldots, n \tag{3.45}$$

The conditional observation statistics are

$$f(\underline{x} \mid \sigma^2, N) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} x_i^2} \tag{3.46}$$

and

$$f(\underline{x} \mid \sigma^2, SN) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (x_i - s_i)^2} \tag{3.47}$$

The noise power level σ^2 is assumed to be an unknown but time-invariant parameter over the observation interval $(0, T)$. Substituting $\theta = \frac{1}{\sigma^2}$ in Eqs. 3.46 and 3.47 yields

$$f(\underline{x} \mid \theta, N) = \left(\frac{\theta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \sum_{i=1}^{n} x_i^2} \tag{3.48}$$

and

$$f(\underline{x} \mid \theta, SN) = \left(\frac{\theta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \sum_{i=1}^{n} (x_i - s_i)^2} \tag{3.49}$$

The total likelihood ratio depends upon the absolute observation statistics with respect to the given a priori density $g(\theta)$. We obtain

$$f(\underline{x} \mid g, N) = \int_{\Theta} f(\underline{x} \mid \theta, N) g(\theta) \, d\theta$$

$$= \int_{0}^{\infty} \left(\frac{\theta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \sum_{i=1}^{n} x_i^2} \frac{c^{b+1}}{\Gamma(b+1)} \theta^b e^{-c\theta} \, d\theta$$

$$= C_n \left(\sum_{i=1}^{n} x_i^2 + 2c\right)^{-(b+1+\frac{n}{2})} \tag{3.50}$$

and

$$f(\underline{x} \mid g, SN) = \int_{\Theta} f(\underline{x} \mid \theta, SN) g(\theta) \, d\theta$$

$$= C_n \left[\sum_{i=1}^{n} (x_i - s_i)^2 + 2c\right]^{-(b+1+\frac{n}{2})} \tag{3.51}$$

where

$$C_n = \frac{(2c)^{b+1} \, \Gamma(b+1+\frac{n}{2})}{\pi^{n/2} \, \Gamma(b+1)} \tag{3.52}$$

Dividing,

$$\ell(\underline{x}|g) = \frac{f(\underline{x}|g, SN)}{f(\underline{x}|g, N)} = \left[\frac{\sum_{i=1}^{n} x_i^2 + 2c}{\sum_{i=1}^{n} (x_i - s_i)^2 + 2c} \right]^{\left(b+1+\frac{n}{2}\right)} \tag{3.53}$$

The likelihood ratio $\ell(\underline{x}|g)$ depends on the observation $\underline{x}$ only through the sufficient statistics

$$\xi(\underline{x}|N) = \sum_{i=1}^{n} x_i^2 \tag{3.54}$$

$$\xi(\underline{x}|SN) = \sum_{i=1}^{n} (x_i - s_i)^2 \tag{3.55}$$

3.2.2.3 Receiver Realization. An optimum processor is any device which realizes the likelihood ratio or a monotone function of it. A particular realization is presented in Fig. 3.8. The processor is essentially a dual channel device, i.e., one channel computes a sufficient statistic conditional to N while the other channel computes a sufficient statistic conditional to SN; these outputs are

combined to form the detection output.
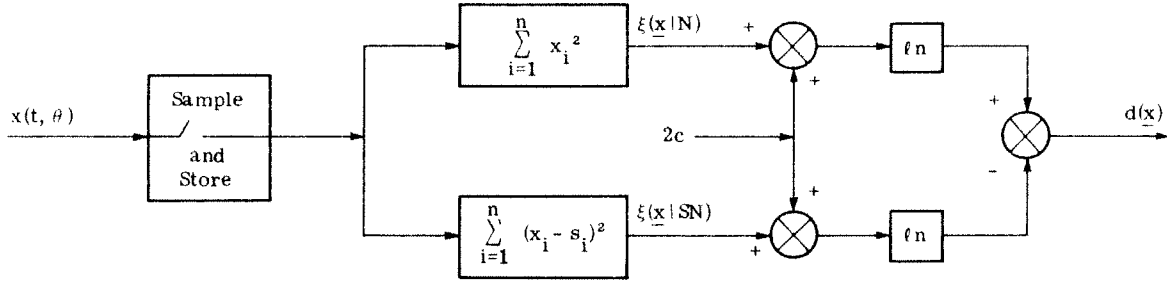
Figure 3.8. Receiver realization based on the given a priori density $g(\theta)$.

3.2.2.4 Sensitivity Index.

Performance of the receiver based on the given a priori density $g(\theta)$ will be compared to the performance of the receiver based on the a priori density $h(\theta)$ via the sensitivity index developed in Section 3.2. From Eqs. 3.17 and 3.18

$$I(g \mid h) = \int_{X} \epsilon(\underline{x}) \left[f(\underline{x} \mid h, SN) - f(\underline{x} \mid h, N) \right] dx \qquad (3.56)$$

where

$$\epsilon(\underline{x}) = z(\underline{x} \mid h) - z[z(\underline{x} \mid g) \mid h] \qquad (3.57)$$

An nth order integration is required by Eq. 3.56. The dimensionality of the problem can be reduced by the choice of an appropriate transformation.

Reduction of Dimensionality. If we let

$$E_s = \sum_{i=1}^{n} s_i^2 \tag{3.58}$$

$$u = \frac{1}{E_s} \left(\sum_{i=1}^{n} x_i s_i \right) - \frac{1}{2} \tag{3.59}$$

$$v^2 = \frac{1}{E_s} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{E_s} \sum_{i=1}^{n} x_i s_i \right)^2 \tag{3.60}$$

then the transformation developed in Appendix C can be employed to show that

$$f(u, v \mid h, N) = K_n v^{n-2} \left[\left(u + \frac{1}{2} \right)^2 + v^2 + \frac{2c}{E_s} \right]^{-(b+1+\frac{n}{2})} n(u, v) \tag{3.61}$$

and

$$f(u, v \mid h, SN) = K_n v^{n-2} \left[\left(u - \frac{1}{2} \right)^2 + v^2 + \frac{2c}{E_s} \right]^{-(b+1+\frac{n}{2})} s(u, v)$$

$$-\infty < u < \infty, \quad 0 \le v < \infty \tag{3.62}$$

with

$$K_n = \frac{2(2c/E_s)^{b+1} \, \Gamma \left(b+1+\frac{n}{2} \right)}{\sqrt{\pi} \, \Gamma \left(\frac{n-1}{2} \right) \Gamma (b+1)} \tag{3.63}$$

$$n(u, v) = \frac{1}{\Gamma(b+1+\frac{n}{2})} \sum_{i=0}^{p} g_i \left(\frac{2c}{E_s}\right)^i \left[(u+\frac{1}{2})^2 + v^2 + \frac{2c}{E_s}\right]^{-i} \qquad (3.64)$$

and

$$s(u, v) = \frac{1}{\Gamma(b+1+\frac{n}{2})} \sum_{i=0}^{p} g_i \left(\frac{2c}{E_s}\right)^i \left[(u-\frac{1}{2})^2 + v^2 + \frac{2c}{E_s}\right]^{-i} \qquad (3.65)$$

where

$$g_i = f_i \Gamma(b+1+\frac{n}{2}+i) \qquad i = 0, 1, \ldots p \qquad (3.66)$$

The coefficients g_i are independent of c; and hence the absolute observation statistics depend on E_s and c only in the form E_s/c.

Calculation of $z(\underline{x}|g)$. If we let

$$\alpha(u, v) \overset{\Delta}{=} z(\underline{x}|g) \qquad (3.67)$$

then

$$\alpha(u, v) = \ln\left[\frac{f(u, v|g, SN)}{f(u, v|g, N)}\right] \qquad (3.68)$$

Since $g(\theta)$ is a degenerate form of $h(\theta)$ obtained by letting $p = 0$ and $r_0 = 1$, we obtain

$$\alpha(u, v) = (b + 1 + \frac{n}{2}) \ln \left[\frac{(u + \frac{1}{2})^2 + v^2 + \frac{2c}{E_s}}{(u - \frac{1}{2})^2 + v^2 + \frac{2c}{E_s}} \right] \tag{3.69}$$

Calculation of $z(\underline{x} | h)$. If we let

$$\xi(u, v) \overset{\Delta}{=} z(\underline{x} | h) \tag{3.70}$$

then

$$\xi(u, v) = \ln \left[\frac{f(u, v | h, SN)}{f(u, v | h, N)} \right]$$

$$= \ln \left[\frac{s(u, v)}{n(u, v)} \right] + \alpha(u, v) \tag{3.71}$$

Calculation of $z[z(\underline{x} | g) | h]$. It is shown in Appendix K that the transformation $z[z(\underline{x} | g) | h]$ can be expressed as a ratio of integrals; evaluation can be achieved via a Gauss-Mehler integration routine. The transformation $z[z(\underline{x} | g) | h]$ appears to be monotone with respect to its argument $z(\underline{x} | g)$. As a matter of fact, the transformation is nearly the identity transformation. Figure 3.9 shows how close the transformation $z[z(\underline{x} | g) | h]$ is to the identity transformation by plotting $z[z(\underline{x} | g) | h] - z(\underline{x} | g)$ vs. $z(\underline{x} | g)$ for the "worst" ($r_0 = .01$) case. (The transformation is precisely the identity transformation if $h(\theta) = g(\theta)$, i.e., when $r_0 = 1$.)
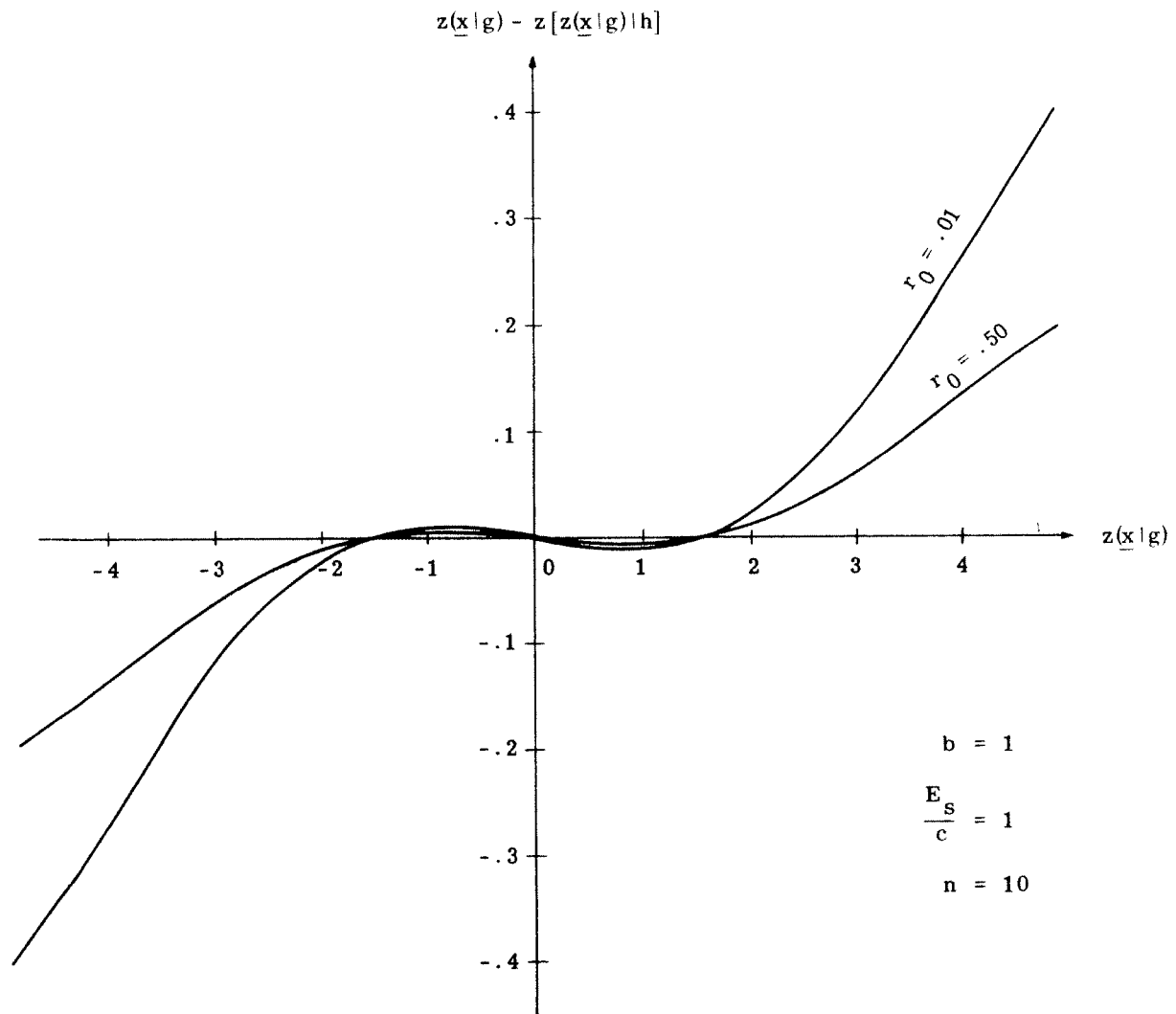
Figure 3.9. Transformation $z[z(\underline{x}|g)|h]$.

Calculation of $\epsilon(\underline{x})$. If we let

$$\delta(u, v) \stackrel{\Delta}{=} \epsilon(\underline{x}) \qquad (3.72)$$

then

$$\delta(u, v) = z(\underline{x} \mid h) - z[z(\underline{x} \mid g) \mid h]$$

$$= \xi(u, v) - z[\alpha(u, v) \mid h]$$

$$= \ell n \left[\frac{s(u, v)}{n(u, v)} \right] + \alpha(u, v) - z[\alpha(u, v) \mid h] \qquad (3.73)$$

A contour map is used to inspect $\delta(u, v)$ in more detail. Two contour maps of the upper right half plane are shown in Figs. 3.10 and 3.11. An appreciation of the nature of $\delta(u, v)$ will assist in keeping computational errors to a minimum. It can be shown that $\delta(u, v)$ is an odd function with respect to u, i.e.,

$$\delta(u, v) = -\delta(-u, v) \qquad (3.74)$$

Furthermore, there appear to be $p - 1$ or fewer distinct contours of $\delta(u, v) = 0$.

Computation of $I(g \mid h)$. The sensitivity index $I(g \mid h)$ can be expressed as

$$I(g \mid h) = E[\delta(u, v) \mid h, SN] - E[\delta(u, v) \mid h, N]$$

$$= \int_0^\infty \int_{-\infty}^\infty \delta(u, v) \, [f(u, v \mid h, SN) - f(u, v \mid h, N] \, du \, dv$$

$$(3.75)$$

Numerical Results. A Gauss-Legendre integration routine was used on an IBM 360/67 digital computer to obtain numerical results. Careful supervision of the inherent computational errors led to reasonably accurate results.
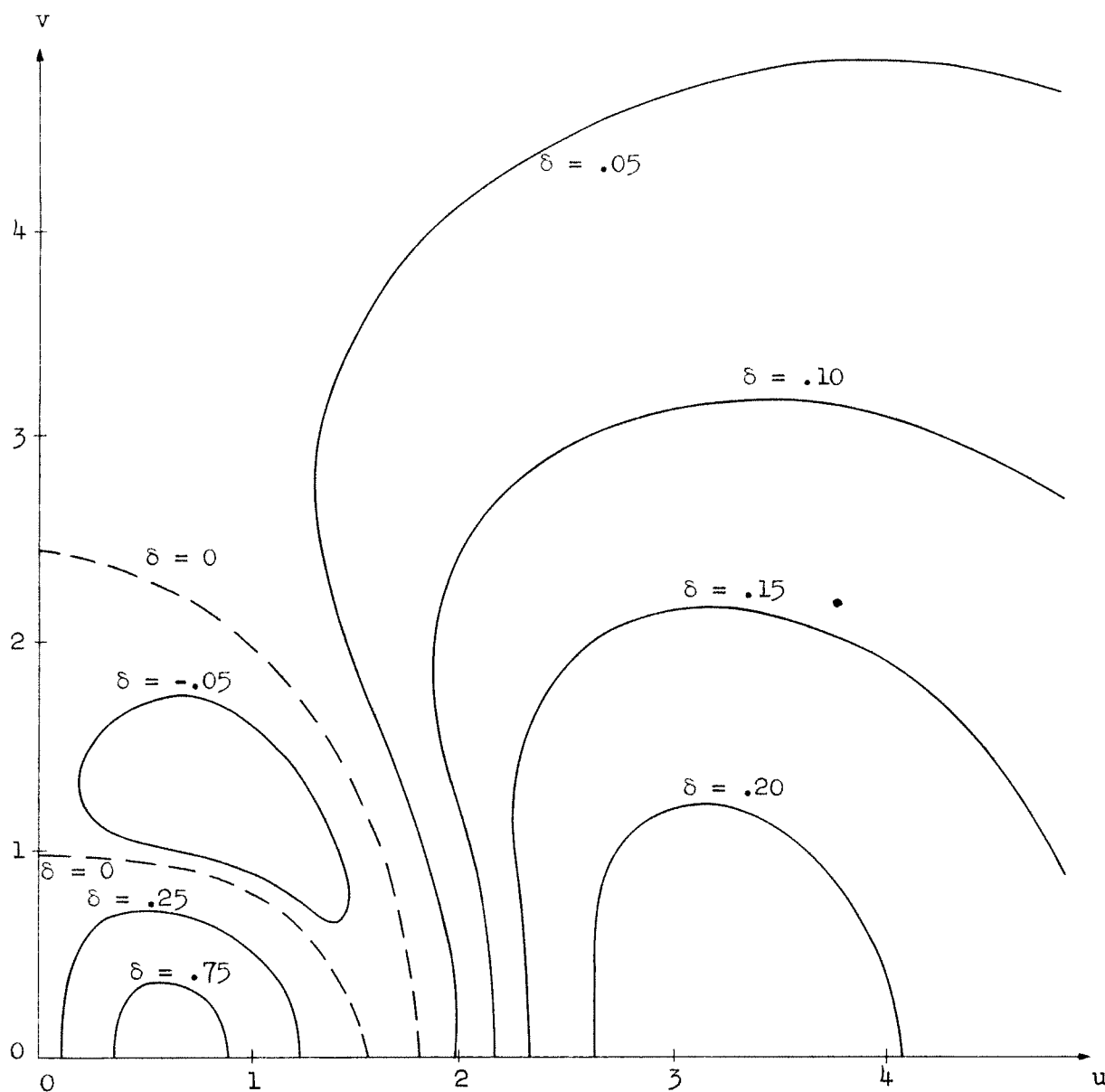
Figure 3.10. Contour map of $\delta(u,v)$, $b = 1$, $E_s/c = 1$, $n = 10$, $r_o = .5$.

The effect of r_0 on the sensitivity index $I(g|h)$ is shown in Fig. 3.12. The sensitivity index $I(g|h) = 0$ for $r_0 = 1$ [$r_0 = 1$ implies that $r(\theta) = 1$ and hence that $h(\theta) = g(\theta)$]. The sensitivity index increases monotonically from zero to its maximum value as
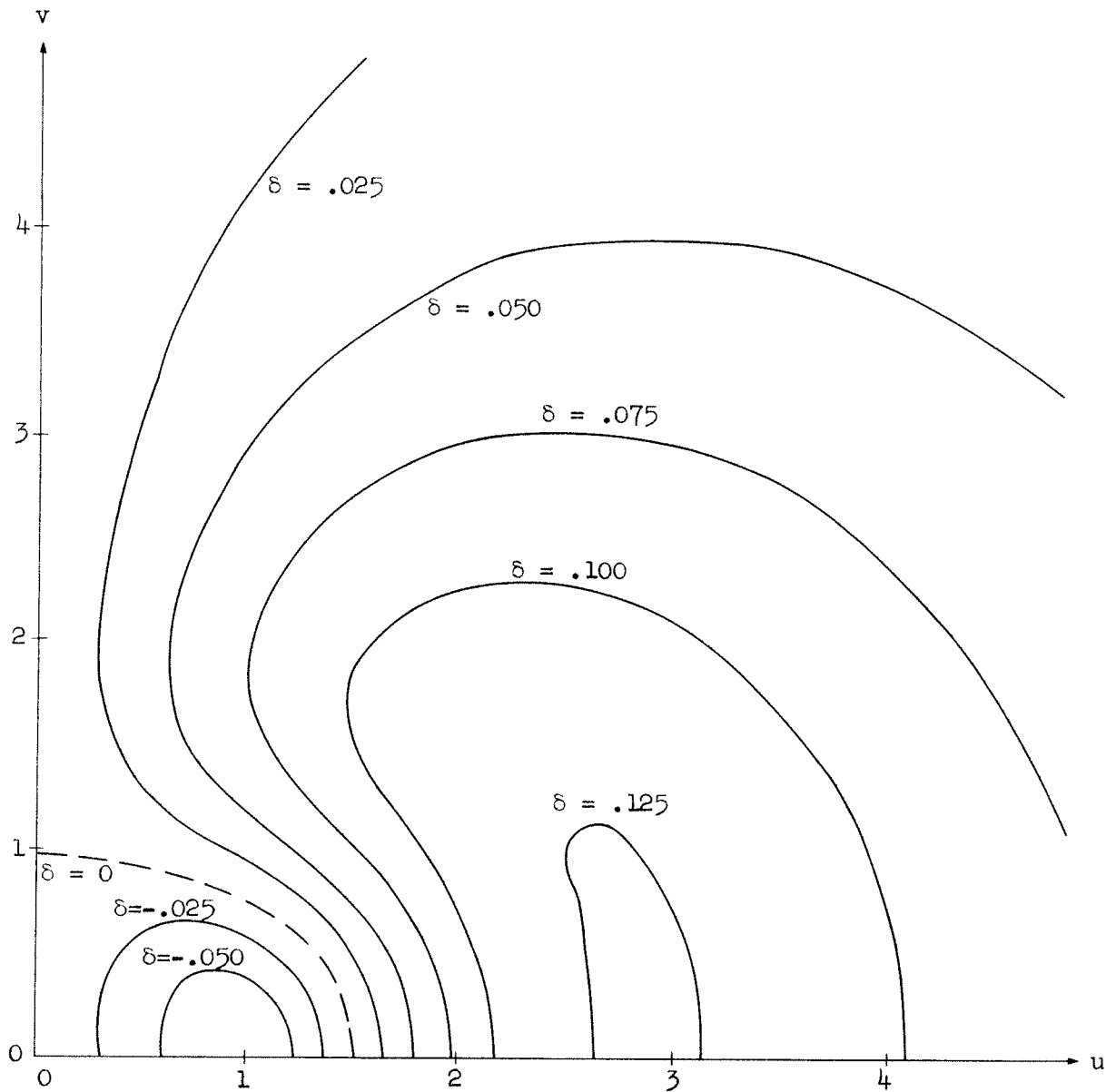
Figure 3.11. Contour map of $\delta(u,v)$, $b = 1$, $E_s/c = 1$, $n = 2$, $r_o = .5$.

$\underline{r_0}$ varies from unity to its extreme permissible value $r_0{}^* = 0$; the sensitivity index changes more rapidly as $r_0 \to r_0{}^*$. Sensitivity appears to be almost independent of the number of observations $\underline{n}$.

The numerical values of the sensitivity index $I(g \mid h)$ are
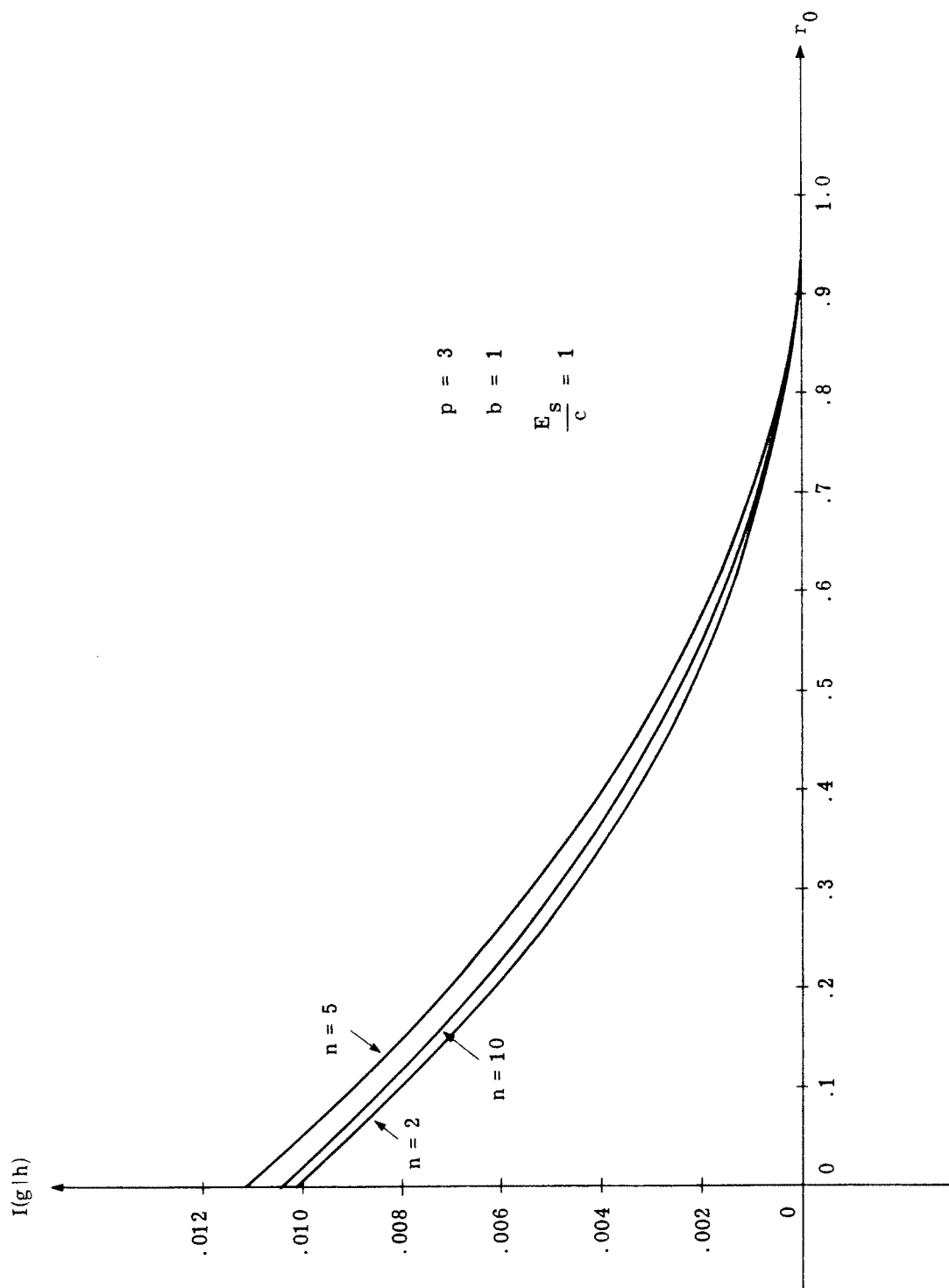
Figure 3.12. Sensitivity index, $p = 3$, $b = 1$, $E_s/c = 1$.

relatively small numbers, even for $r_0 = r_0^*$; it can be shown that

the performance index $J_h(1:0)$ is approximately equal to the expec-

ted value of the signal-to-noise ratio, i.e., $J_h(1:0) \cong E(\theta E_s \mid h) =$

$E(\theta E_s \mid g) = (b+1) \dfrac{E_s}{c}$. Thus, sensitivity of receiver performance to

a priori specification is minimal even when only the mean and variance

of $h(\theta)$ are matched to the mean and variance of $g(\theta)$.

The validity of the sensitivity index can be ascertained by com-

puting the standard deviation of $\delta(u, v)$. Figure 3.13 shows how the

standard deviation of $\delta(u, v)$ is related to r_0. The standard devia-

tion is almost linear with respect to r_0 and virtually independent of

the number of observations n. The interpretation follows: the level

of confidence one places on the sensitivity index should increase as

$r_0 \to 1$ and should be independent of the number of observations n;

in other words, one should have little doubt about the sensitivity index

as an indicator of relative performance level whenever r_0 is close to

unity regardless of n.

It is not possible to show how the parameter b affects sensi-

tivity over its entire range if r_0 is kept constant, since r_0 is lim-

ited in its range of permissible values as indicated in Fig. 3.6. To

circumvent this problem it was decided to keep the degree of perturba-

tion between minimum and maximum permissible values of r_0 constant;

and hence a new parameter d_p was defined as

58
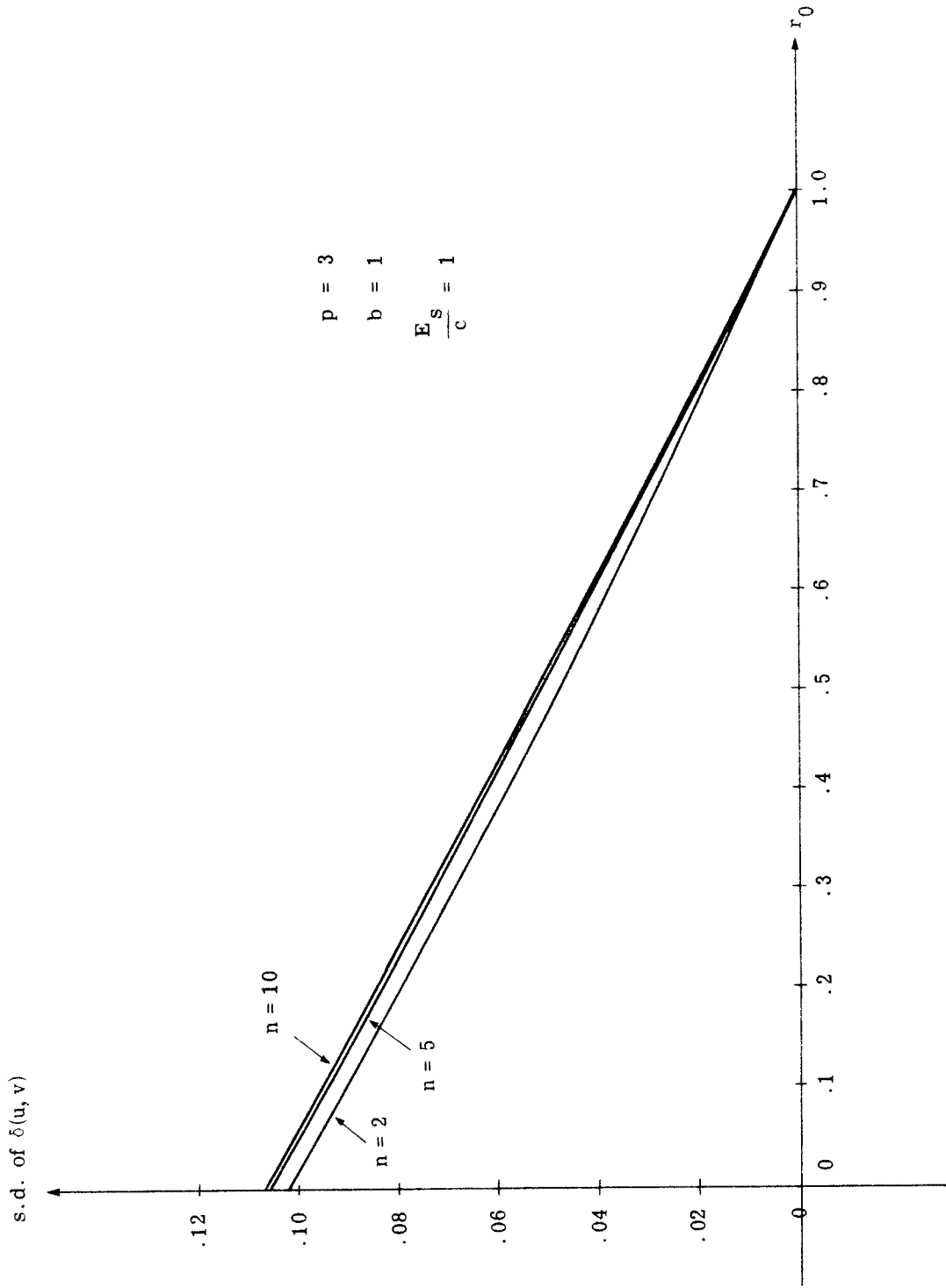
Figure 3.13. Standard deviation of $\delta(u,v)$, $p = 3$, $b = 1$, $E_S/c = 1$.

$$d_p = \frac{r_0 - r_0^*}{1 - r_0^*}$$

(3.76)

Figure 3.14 shows how the sensitivity index is related to b for $d_p = .5$ (r_0 was kept at a level halfway between its minimum and maximum permissible values.) The mode at $b = 1$ can be accounted for by referring to Fig. 3.6 and observing the transition that occurs at $b = 1$ for $p = 3$. At that point $r(\theta) = 0$ for some $\theta \in (0, \infty)$; for $p = 3$ and $b > 1$, $r(\theta) = 0$ only at $\theta = 0$ so that $r_0^* = 0$.

The effect of the parameter E_s/c on the sensitivity index is shown in Fig. 3.15. Sensitivity is nearly linear with respect to E_s/c for $0 \leq E_s/c < 2$. Intuitively this is very reasonable. An increase in signal energy E_s for c fixed results in a higher performance level for any detection situation. The relative performance level as measured by the sensitivity index is likely to follow the same pattern. The sensitivity index reaches a maximum and decreases as E_s/c gets larger and larger. Detection becomes almost a certainty as the signal energy gets large in spite of the fact that the receiver is not optimum; and hence it is difficult but irrelevant to distinguish between the performance levels of the optimum and suboptimum receivers.
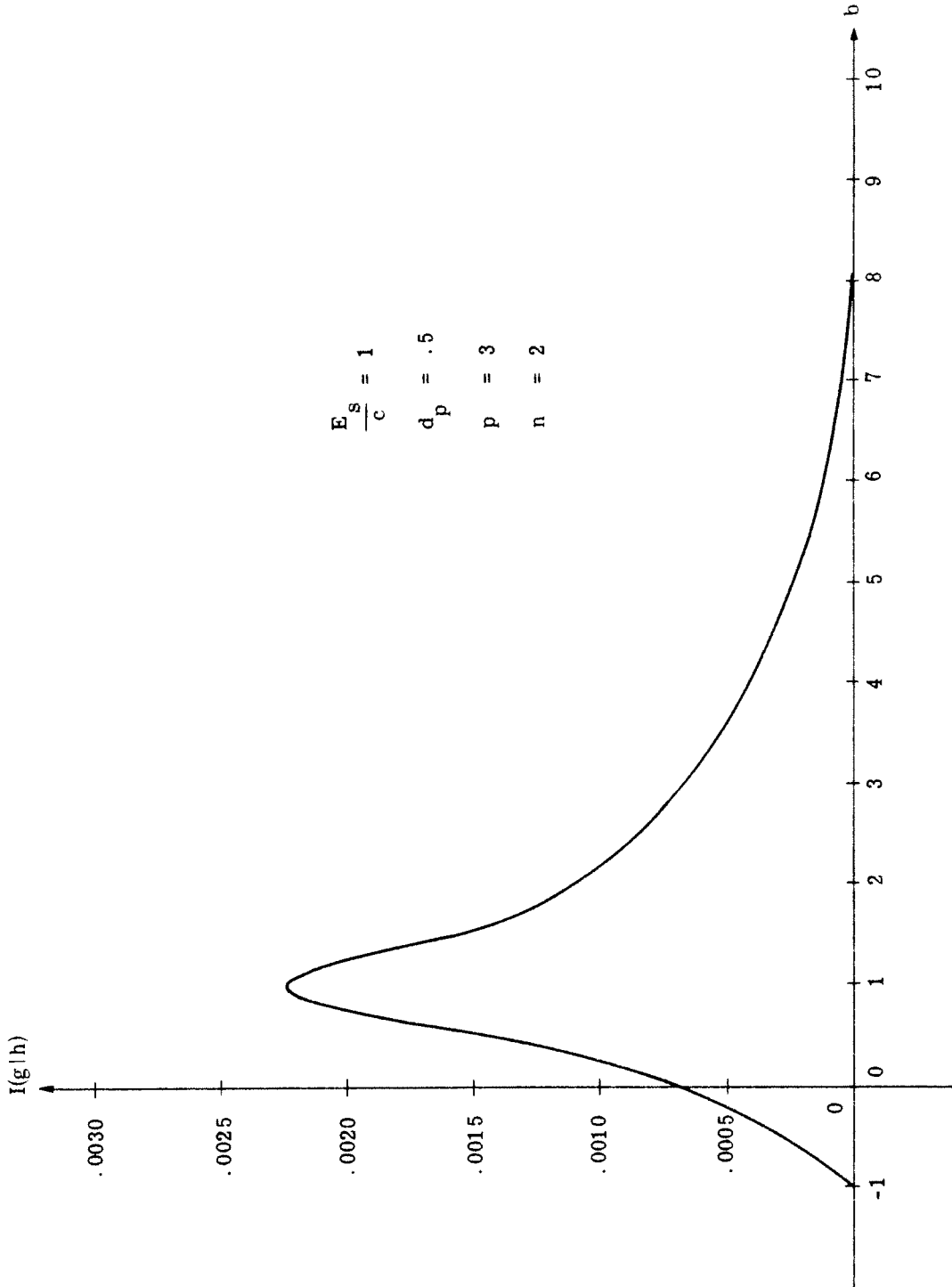
Figure 3.14. Sensitivity index, $E_s/c = 1$, $d_p = .5$, $p = 3$, $n = 2$.
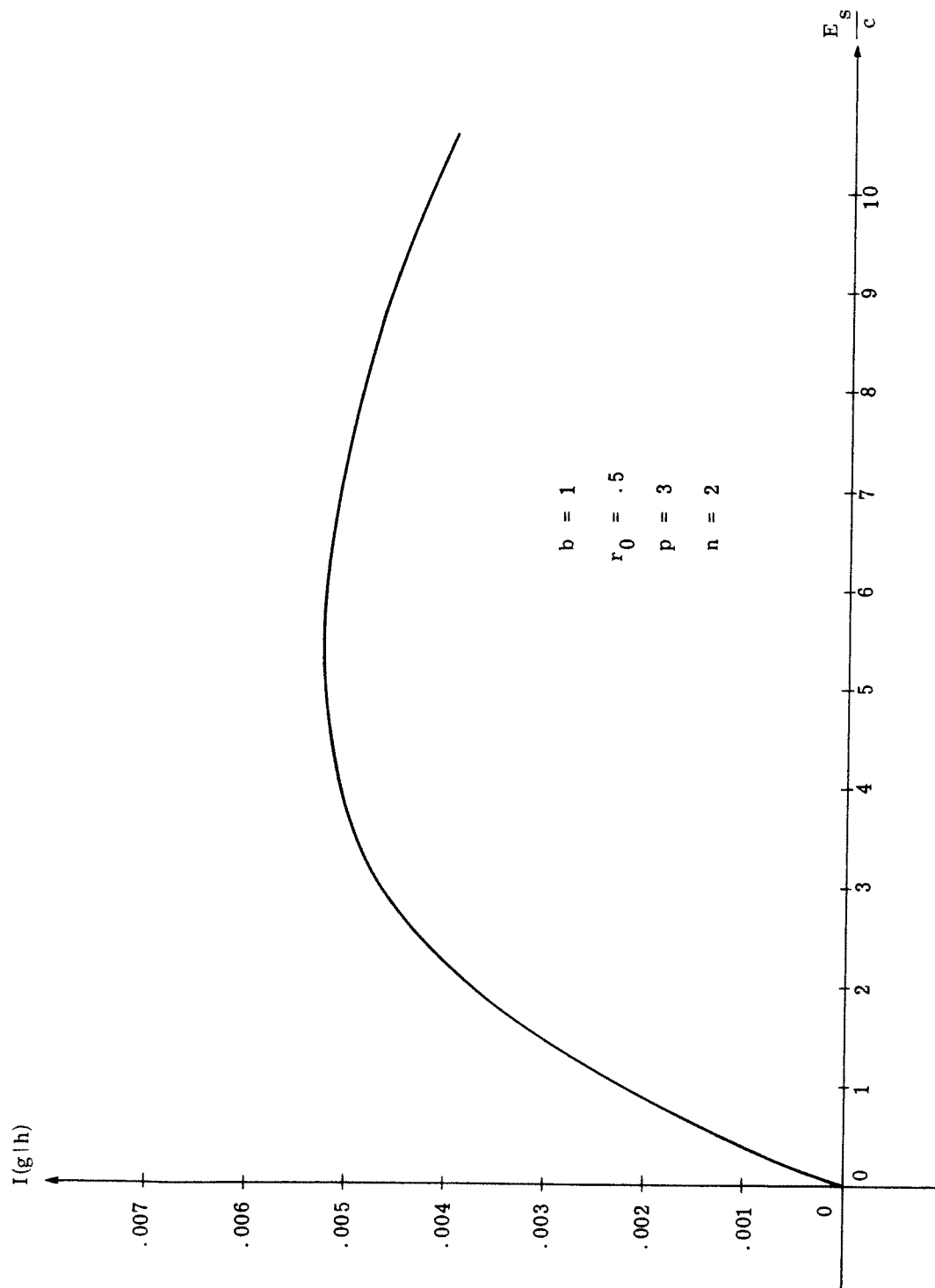
Figure 3.15. Sensitivity index, b = 1, r_0 = .5, p = 3, n = 2.

CHAPTER IV

COMPARISON OF RECEIVER PERFORMANCE FOR THE

COMPOSITE HYPOTHESIS SITUATION

Two contrasting points of view to the theory of signal detectability are presented by the Bayesian approach and the classical approach. The Bayesian approach requires the assignment of a priori densities to all unknown parameters and averages the conditional observation statistics with respect to the assigned a priori density. This allows the user to introduce physical or empirical knowledge about these parameters. Unfortunately, some view these a priori densities as simply a convenient technique to eliminate the nuisance of unknown parameters, mistaking "freedom to incorporate" for "license to eliminate."

Classical statistics, on the other hand, seeks a test which is either uniformly most powerful (UMP) or uniformly most powerful unbiased (UMPU). Basically, it seeks a test which is either the best test (maximizes the probability of detection for a given probability of false alarm) regardless of the actual value of the unknown parameter, provided only that it is admissible; or it seeks the best unbiased test (never yields performance poorer than chance) with respect to the admissible set of values of the unknown parameter. While UMPU tests often exist, UMP tests exist only in special circumstances.

Included in the classical approach are the "estimate and plug" techniques, e.g., the "generalized likelihood ratio" method.

Both the Bayesian and classical approach often lead to tests which yield identical performance in terms of ROC. This seems to be particularly true of the single composite hypothesis for situations where the parameters are "non-energy-bearing," and only when the a priori knowledge is modeled by a maximum entropy density. Even for the double composite hypothesis, application of classical estimation techniques may result in the same test obtained by Bayesian analysis. Thus, although the basic criteria employed for hypothesis testing may be dissimilar, both Bayesian and classical tests often yield equivalent performance. As a result it is conjectured that Bayesian philosophy and classical statistics can usually be reconciled with each other for the composite hypothesis.

4.1 Introduction

This chapter will consider both the Bayesian and classical approach to signal detection for the composite hypothesis and contrast performance in terms of ROC for several examples.

In reference to Bayesian philosophy the ESP receiver is postulated as a means of determining the information content and usefulness of a specific a priori density for use in receiver design. In conjunction with the ESP receiver, conditional performance curves are explained along with average performance curves. Several

pertinent examples are worked out.

From a classical point of view UMP and UMPU tests are described, and several examples are given. In addition some classical estimation techniques are employed in the task of receiver design. Indications are that "estimate-and-plug" procedures lead to the same type of receiver obtained by utilizing Bayesian techniques along with a diffuse a priori density.

4.2 Bayesian Approach

Bayesian philosophy assumes there is a known a priori density for all unknown parameters of the conditional observation densities. Let θ denote an unknown N and/or SN process parameter(s) in the subsequent development, and let $g(\theta)$ denote the a priori density assigned to θ.

4.2.1 Evaluation Methods. According to Bayesian analysis the absolute observation statistics are

$$f(x \mid {}^N_{SN}) = \int_\Theta f(x \mid \theta, {}^N_{SN}) \, g(\theta) \, d\theta \tag{4.1}$$

The total likelihood ratio is

$$\ell(x) = \frac{f(x \mid SN)}{f(x \mid N)} \tag{4.2}$$

and the conditional likelihood ratio is

$$\ell(x \mid \theta) = \frac{f(x \mid \theta, SN)}{f(x \mid \theta, N)} \tag{4.3}$$

Optimum ROC

The optimum receiver bases its decisions on $\ell(x)$ while the conditionally optimum receiver bases its decisions on $\ell(x|\theta)$. Optimum receiver performance is determined via

$$
\begin{aligned}
P^{OPT}(A \mid \underset{SN}{N}) &= \int_{\ell(x) > \beta} f(x \mid \underset{SN}{N})\, dx \\
&= \int_{\ell(x) > \beta} \left[\int_{\Theta} f(x \mid \theta, \underset{SN}{N})\, g(\theta)\, d\theta \right] dx \\
&= \int_{\Theta} \left[\int_{\ell(x) > \beta} f(x \mid \theta, \underset{SN}{N})\, dx \right] g(\theta)\, d\theta \\
&= \int_{\Theta} P^{CS}(A \mid \theta, \underset{SN}{N})\, g(\theta)\, d\theta
\end{aligned}
\tag{4.4}
$$

where

$$
P^{CS}(A \mid \theta, \underset{SN}{N}) = \int_{\ell(x) > \beta} f(x \mid \theta, \underset{SN}{N})\, dx
\tag{4.5}
$$

The parametric Eq. 4.5 describes the conditionally suboptimum ROC obtained by the receiver which operates in a conditional environment but bases decisions on $\ell(x)$ instead of $\ell(x|\theta)$. Hence, the optimum ROC is obtained by averaging the conditionally suboptimum ROC's with respect to $g(\theta)$.

ESP ROC

The externally sensed parameter (ESP) ROC has been proposed

as a convenient upper bound to the optimum ROC by Spooner (Ref. 32).
The name is derived from the hypothetic noiseless observation of the
parameter θ as well as the usual observation, x . If θ contains
"signal" parameters, the ESP receiver observes these whether the
signal is actually present or absent. The ESP receiver is a fictitious
receiver whose ROC is obtained by averaging the conditionally optimum
ROC's with respect to $g(\theta)$. Hence, it is independent of the unknown
parameter(s) θ , but dependent on the a priori density $g(\theta)$. The
parametric equations describing the ESP ROC are

$$P^{ESP}(A \mid _{SN}^{N}) = \int_{\Theta} P^{CO}(A \mid \theta, _{SN}^{N})\, g(\theta)\, d\theta \qquad (4.6)$$

where

$$P^{CO}(A \mid \theta, _{SN}^{N}) = \int_{\ell(x \mid \theta) > \beta} f(x \mid \theta, _{SN}^{N})\, dx \qquad (4.7)$$

Since the ESP ROC averages conditionally optimum ROC's while the
optimum ROC averages conditionally suboptimum ROC's, it is read-
ily apparent that the ESP ROC serves as an upper bound to the opti-
mum ROC. However, equivalence of the conditional ROC's does not
necessarily imply equivalence of the average ROC's.

Interpretation

The ESP ROC and the optimum ROC are both dependent on the
a priori density $g(\theta)$. The difference between these ROC's indicates

the degree to which the unknown parameter θ has affected perform-ance, subject to the a priori density $g(\theta)$. One might conjecture that a small difference means either that the parameter θ is irrelevant, or if relevant, that the optimum receiver acts "as if it had learned the value of θ " from the observation, x. A large difference indicates that knowledge of θ is relevant, and the optimum receiver is unable to effectively learn the value of θ from the observation, x.

4.2.2 <u>Examples</u>. Several examples have been worked to demonstrate the concepts presented above. The examples to be considered are SKEA (signal known except for amplitude) with a normal a priori density, SKEA with an a priori density that assigns zero probability to the negative axis (known sign), and SKE-UNL (signal known exactly, unknown noise level) with a gamma a priori density.

Definitions

The following definitions will be used throughout this chapter in each of the examples:

$$w = \underline{x}^t \, \Sigma^{-1} \, \underline{x} \tag{4.8}$$

$$y = \underline{s}^t \, \Sigma^{-1} \, \underline{x} \tag{4.9}$$

$$d_0 = \underline{s}^t \, \Sigma^{-1} \, \underline{s} \qquad \text{(nominal signal-to-noise ratio)} \tag{4.10}$$

4.2.2.1 Signal Known Except for Amplitude, Normal

A Priori Density. Assume that the unknown amplitude a can be modeled accurately by the normal a priori density

$$g(a) = \frac{1}{\sqrt{2\pi v}} e^{-\frac{(a-m)^2}{2v}} \tag{4.11}$$

and that the conditional observation statistics can be modeled by the nth order multivariate Gaussian densities

$$f(\underline{x} \mid a, N) = (2\pi \mid \Sigma \mid)^{-\frac{n}{2}} e^{-\frac{1}{2}\underline{x}^t \Sigma^{-1} \underline{x}} \tag{4.12}$$

$$f(\underline{x} \mid a, SN) = (2\pi \mid \Sigma \mid)^{-\frac{n}{2}} e^{-\frac{1}{2}(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s})} \tag{4.13}$$

where $\underline{s}$ is the signal "waveshape" and Σ is the noise autocovariance matrix.

In order to avoid obscuring the principles involved, the derivation of receiver operation and performance is performed in the appendices for both the conditionally optimum and suboptimum cases.

Conditionally optimum receiver

The decision rule of the conditionally optimum receiver (derived in Appendix D) is

$$
\phi(\underline{x}) = \begin{cases} 1 & y > \dfrac{\ln\beta}{a} + \dfrac{ad_0}{2} \\ \\ 0 & \text{otherwise} \end{cases} \quad \text{if} \quad a > 0
$$

or

$$
\phi(\underline{x}) = \begin{cases} 1 & y < \dfrac{\ln\beta}{a} + \dfrac{ad_0}{2} \\ \\ 0 & \text{otherwise} \end{cases} \quad \text{if} \quad a < 0 \qquad (4.14)
$$

(The nature of the decision rule $\phi(\underline{x})$ was discussed in Section 2.2.2.)

The corresponding performance equations (derived in Appendix D) are

$$
P^{CO}(A \mid a, N) = \Phi\left(-\frac{\ln\beta}{|a| \sqrt{d_0}} - \frac{|a| \sqrt{d_0}}{2}\right) \qquad (4.15)
$$

$$
P^{CO}(A \mid a, SN) = \Phi\left(-\frac{\ln\beta}{|a| \sqrt{d_0}} + \frac{|a| \sqrt{d_0}}{2}\right) \qquad (4.16)
$$

These parametric equations in β describe a normal ROC with $d' = \sqrt{d_0} \, |a|$.

Conditionally suboptimum receiver

The decision rule of the conditionally suboptimum receiver (derived in Appendix E) is

$$\phi(\underline{x}) = \begin{cases} 1 & |y + \frac{m}{v}| > \beta \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.17}$$

The corresponding performance equations (derived in Appendix E) are

$$P^{CS}(A \mid a, N) = \Phi(\lambda_1) + \Phi(\lambda_2) \tag{4.18}$$

$$P^{CS}(A \mid a, SN) = \Phi(\lambda_1 + a\sqrt{d_0}) + \Phi(\lambda_2 - a\sqrt{d_0}) \tag{4.19}$$

where

$$\lambda_1 = \frac{-\Delta + mv^{-1}}{\sqrt{d_0}} \tag{4.20}$$

$$\lambda_2 = \frac{-\Delta - mv^{-1}}{\sqrt{d_0}} \tag{4.21}$$

and

$$\Delta = \sqrt{2(v^{-1} + d_0)\left[\ell n(\beta\sqrt{1 + vd_0}) + \frac{m^2}{2v}\right]} \tag{4.22}$$

ESP ROC

The ESP ROC is determined by substituting Eqs. 4.11, 4.15 and 4.16 into Eq. 4.6. We obtain

$$P^{ESP}(A|N) = \int_{-\infty}^{\infty} \Phi\left(-\frac{\ln\beta}{|a|\sqrt{d_0}} - \frac{|a|\sqrt{d_0}}{2}\right) \frac{1}{\sqrt{2\pi v}} e^{-\frac{(a-m)^2}{2v}} da$$

$$(4.23)$$

$$P^{ESP}(A|SN) = \int_{-\infty}^{\infty} \Phi\left(-\frac{\ln\beta}{|a|\sqrt{d_0}} + \frac{|a|\sqrt{d_0}}{2}\right) \frac{1}{\sqrt{2\pi v}} e^{-\frac{(a-m)^2}{2v}} da$$

$$(4.24)$$

Numerical values for these parametric equations in β were obtained by use of numerical integration techniques.

Optimum ROC

The optimum ROC can be determined by substituting Eqs. 4.11, 4.18 and 4.19 into Eq. 4.4 and making use of Appendix H. We obtain

$$P^{OPT}(A|N) = \int_{-\infty}^{\infty} [\Phi(\lambda_1) + \Phi(\lambda_2)] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(a-m)^2}{2v}} da$$

$$= \Phi(\lambda_1) + \Phi(\lambda_2)$$

$$(4.25)$$

$$P^{OPT}(A|SN) = \int_{-\infty}^{\infty} [\Phi(\lambda_1 + a\sqrt{d_0}) + \Phi(\lambda_2 - a\sqrt{d_0})] \frac{1}{\sqrt{2\pi v}} e^{-\frac{(a-m)^2}{2v}} da$$

$$= \Phi\left(\frac{\lambda_1 + m\sqrt{d_0}}{\sqrt{1 + vd_0}}\right) + \Phi\left(\frac{\lambda_2 - m\sqrt{d_0}}{\sqrt{1 + vd_0}}\right)$$

$$(4.26)$$

Diffuseness

An inspection of Eqs. 4.18 - 4.21 shows that the conditional performance of the suboptimum receiver depends on m and v only in the ratio $\frac{m}{v}$. The ratio $\frac{m}{v}$ can be related to a measure of diffuseness or non-centrality. An a priori density is said to become diffuse as the variance v becomes large. Hence, $(\frac{m}{v})^{-1}$ can be considered a measure of diffuseness.[1]

Evaluation of receiver performance

Results are presented in Figs. 4.1 - 4.4 for specific parameter values of d_0, m and v. The parameter d_0 represents the nominal signal-to-noise ratio and was set equal to unity. Conditional ROC's (optimum and suboptimum) are shown in Figs. 4.1 and 4.3. Average ROC's (ESP and optimum) are shown in Figs. 4.2 and 4.4 for the same parameter values as the conditional ROC's.

Conditional ROC's

An inspection of the conditional ROC's of Fig. 4.1 immediately leads to the conclusion that the conditionally suboptimum ROC may become biased (yield performance poorer than chance) for some particular choice of a. This situation will occur whenever the environmental conditions (actual operating conditions) and the given a priori

---

[1]From a dimensional standpoint, $\frac{\sqrt{v}}{m}$ is a more appropriate choice, especially since $\left(\frac{\sqrt{v}}{m}\right)^{-1}$ is termed a measure of centrality.
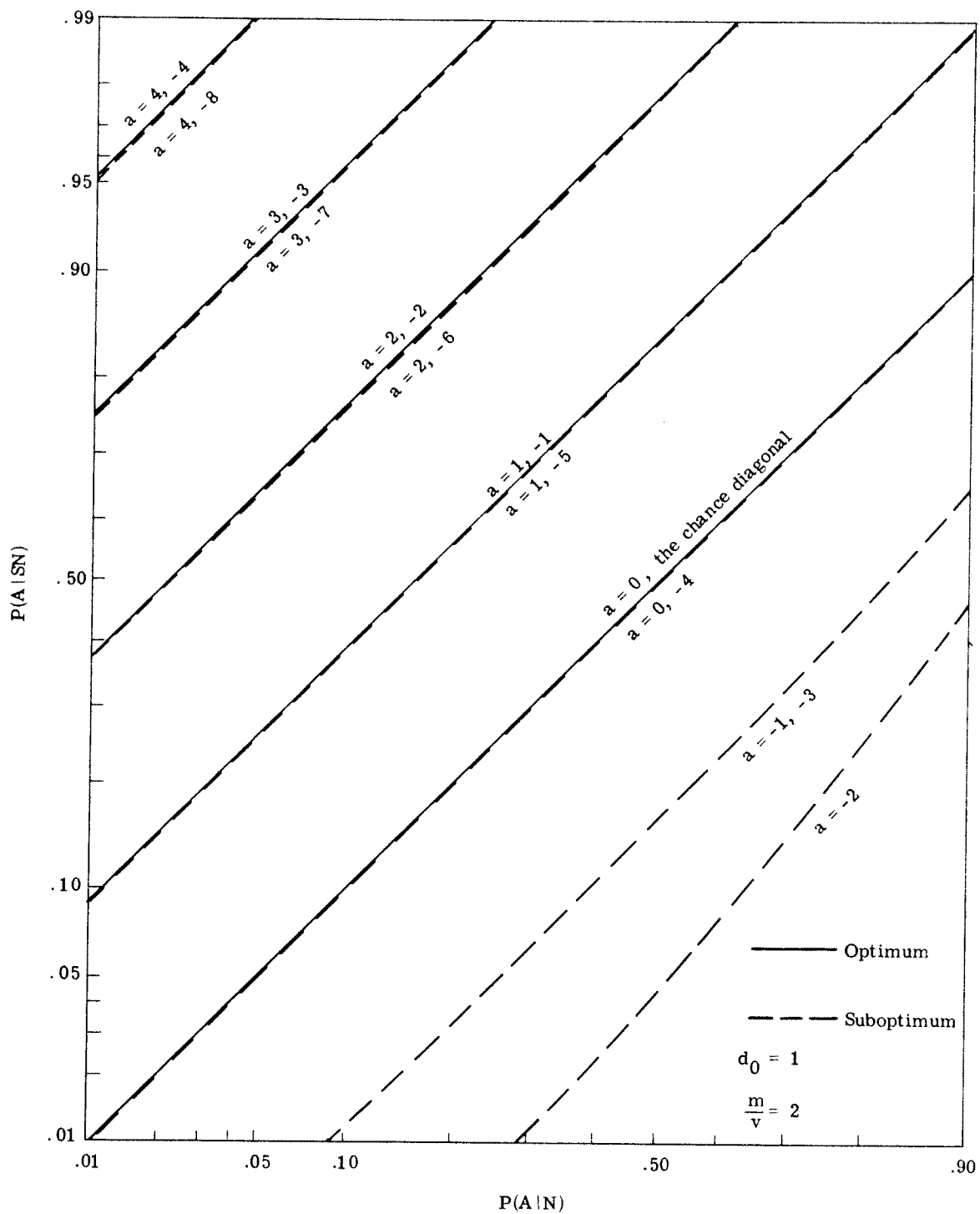
Figure 4.1. Conditional ROC's, comparison of optimum and suboptimum receiver performance, SKEA, d_o = 1, m/v = 2.
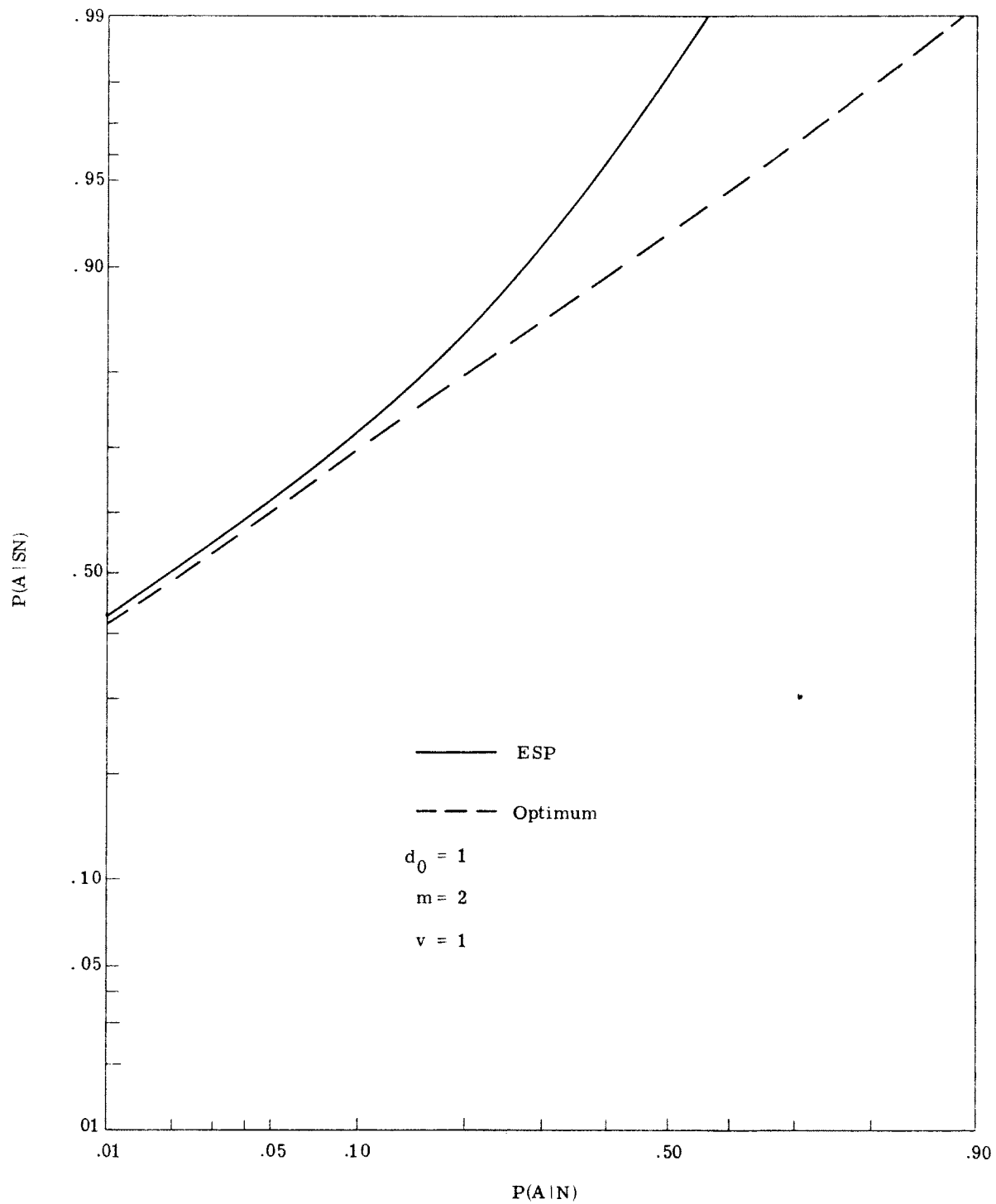
Figure 4.2. Average ROC's, comparison of ESP and optimum receiver performance, SKEA, $d_0 = 1$, $m = 2$, $v = 1$.
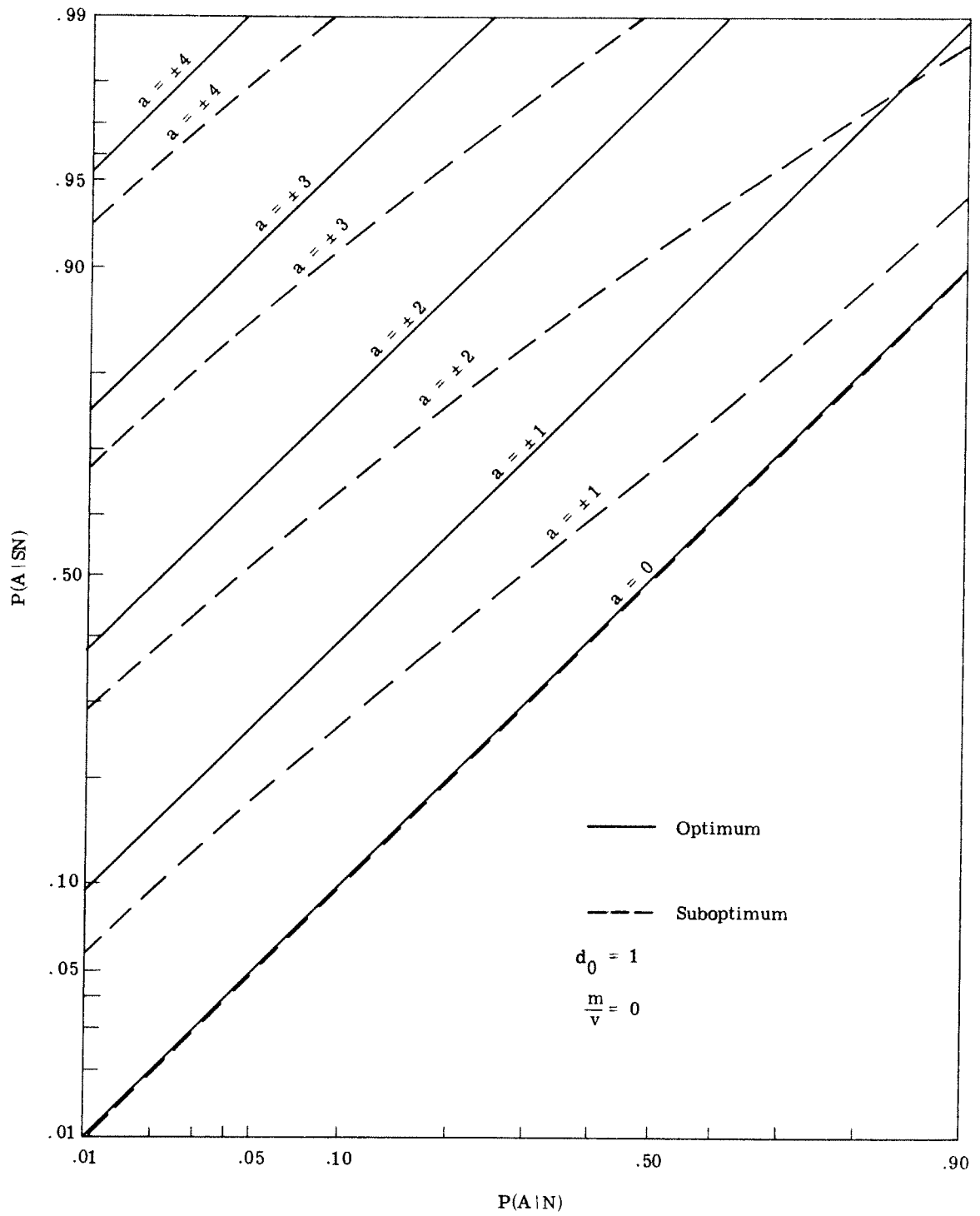
Figure 4.3. Conditional ROC's, comparison of optimum and suboptimum receiver performance, SKEA, $d_o = 1$, $m/v = 0$.
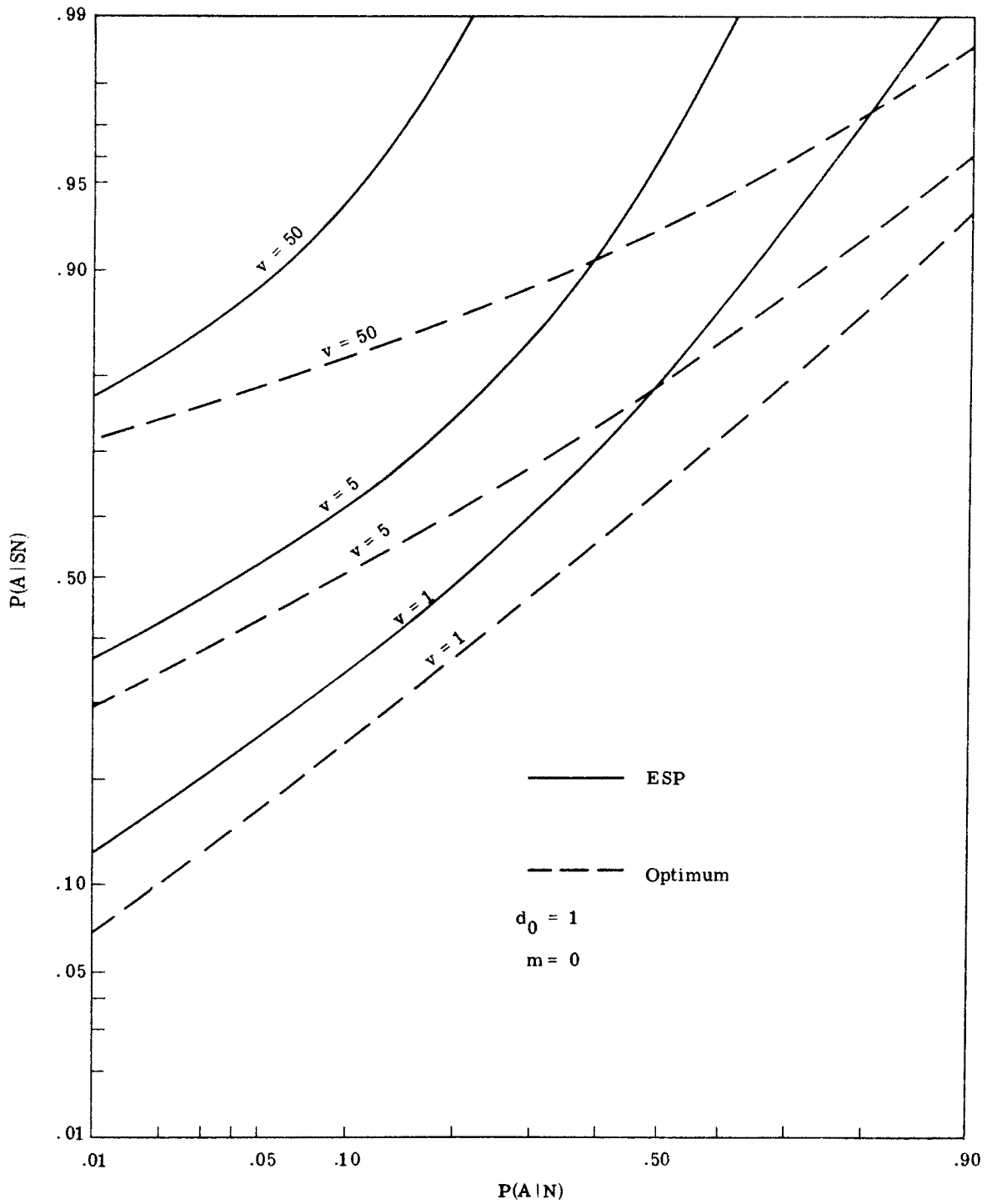
Figure 4.4. Average ROC's, comparison of ESP and optimum receiver performance, SKEA, $d_o = 1$, m = 0.

density differ widely. Hence the receiver designer must be cautioned to be very careful about the a priori density he chooses to model the physical situation. The assumption that negative amplitudes occur rarely can lead to very poor performance (worse than chance) when the environmental conditions are such that a negative amplitude does occur. The performance index d' may be reduced by as much as $2m/(v \sqrt{d_0})$. For instance, the conditionally suboptimum ROC of Fig. 4.1 obtained for a = -6 has an equivalent d' = 2 whereas precise knowledge of the amplitude would have resulted in d' = 6. By the same token, the conditionally suboptimum ROC obtained for a = 4 yields essentially the best performance possible under the circumstances. Hence it is very important to incorporate into the a priori density, in a very precise manner, any information concerning actual operating conditions. It is best to assume a diffuse a priori density $(\frac{m}{v} << 1)$ if little or no knowledge is available; use a sharp a priori density $(\frac{m}{v} >> 1)$ whenever precise knowledge is available.

The conditional ROC's of Fig. 4.3 are based on an even symmetric (m = 0) or a diffuse (m ≠ 0; as v → ∞, $\frac{m}{v}$ → 0) a priori density. Note that the conditionally suboptimum ROC's are unbiased (performance is at least as good as chance) for all admissible amplitude values. However, the conditionally suboptimum ROC's of Fig. 4.1 are superior to the conditionally suboptimum ROC's of Fig. 4.3 for positive amplitudes.

It should be clearly understood that the "optimum" receiver

produces the best average ROC, best with respect to the given a priori density $g(\cdot)$. That is why it is important to carefully choose $g(\cdot)$ to reflect the performance balance the user wishes (and not the designer's convenience). If the user does not desire best average performance, but, for example, something like maximum worst-case conditional performance,[1] then it would be advisable to drop likelihood ratio and use other design principles.

Average ROC's

The average ROC's for this detection problem are presented in Figs. 4.2 and 4.4. Since the ESP ROC is the average of the conditionally optimum ROC's and the optimum ROC is the average of the conditionally suboptimum ROC's, Fig. 4.2 is a "capsule" summary of Fig. 4.1 and Fig. 4.4 is a "capsule" summary of Fig. 4.3. Closeness of the optimum ROC to the ESP ROC tends to indicate that on the average the conditionally suboptimum ROC's are close to the conditionally optimum ROC's. In essence, comparison of the average ROC's gives some indication of the relative behavior of the conditional ROC's.

4.2.2.2 Signal Known Except for Amplitude, Known Sign.
This section will determine the operation and performance of the receiver which is optimized with respect to an a priori density that

[1] Maximum worst-case conditional performance means "go after the little signals and don't try to take advantage of the big ones."

assigns zero probability to the negative axis. [The sign of a is known, i.e., $P(a > 0) = 1$.]

Restriction of the a priori density to the class of densities which assigns zero probability to the negative axis leads to some interesting results. The unusual aspects of this detection problem will be discussed. Both conditional and average ROC's will be derived.

Conditionally optimum receiver

Conditionally optimum receiver operation and performance do not depend on the given a priori density. The decision rule and ROC were derived earlier. The decision rule is given by Eq. 4.14 and the ROC is described by Eqs. 4.15 and 4.16. Receiver operation consists of simply computing the sufficient statistic $y = \underline{s}^t \Sigma^{-1} \underline{x}$ and comparing it to a threshold which depends upon the actual amplitude a.

Conditionally suboptimum receiver

It is shown in Appendix F, subject only to the restriction that g(a) assigns zero probability to the negative axis, that suboptimum receiver operation is identical to optimum receiver operation. Again, receiver operation consists of merely comparing the sufficient statistic y to a threshold γ; the threshold γ depends upon the a priori density g(a) but is independent of the actual amplitude a.

Conditionally suboptimum receiver performance depends upon the conditional observation statistics of y. It was shown in Appendix D that

$$f(y \mid a, N) = \frac{1}{\sqrt{2\pi d_0}} \, e^{-\frac{y^2}{2d_0}} \tag{4.27}$$

and

$$f(y \mid a, SN) = \frac{1}{\sqrt{2\pi d_0}} \, e^{-\frac{(y - ad_0)^2}{2d_0}} \tag{4.28}$$

Hence, the conditionally suboptimum ROC is described by

$$P^{CS}(A \mid a, N) = \int\limits_{y > \gamma} f(y \mid a, N) \, dy$$

and

$$= \Phi(\delta) \tag{4.29}$$

$$P^{CS}(A \mid a, SN) = \int\limits_{y > \gamma} f(y \mid a, SN) \, dy$$

$$= \Phi(\delta + a \sqrt{d_0}) \tag{4.30}$$

where

$$\delta = \frac{-\gamma}{\sqrt{d_0}} \tag{4.31}$$

These performance equations yield a normal ROC with $d' = a \sqrt{d_0}$. Hence, the conditionally optimum ROC's are equivalent to the conditionally suboptimum ROC's. However, the parametric equations describing the ROC's are different. Knowledge about the exact

operating point (threshold level) becomes significant. As a result the ESP ROC and the optimum ROC are not equivalent.

ESP ROC

The performance equations of the ESP ROC are

$$P^{ESP}(A|N) = \int_0^\infty \Phi\left(\frac{-\ell n\beta}{a \sqrt{d_0}} - \frac{a \sqrt{d_0}}{2}\right) g(a) \, da \qquad (4.32)$$

and

$$P^{ESP}(A|SN) = \int_0^\infty \Phi\left(\frac{-\ell n\beta}{a \sqrt{d_0}} + \frac{a \sqrt{d_0}}{2}\right) g(a) \, da \qquad (4.33)$$

Optimum ROC

The performance equations of the optimum ROC are

$$P^{OPT}(A|N) = \int_0^\infty \Phi(\delta) \, g(a) \, da = \Phi(\delta) \qquad (4.34)$$

and

$$P^{OPT}(A|SN) = \int_0^\infty \Phi(\delta + a \sqrt{d_0}) \, g(a) \, da \qquad (4.35)$$

Evaluation of receiver performance

An example of this type was worked by Spooner (Ref. 32). The a priori density used in his study was the truncated normal density

$$g(a) = \frac{1}{\sqrt{2\pi\lambda^2} \; \Phi\left(\frac{\alpha}{\lambda}\right)} \; e^{-\frac{(a-\alpha)^2}{2\lambda^2}} \qquad 0 \leq a < \infty \qquad (4.36)$$

He presents performance curves for the ESP receiver and the optimum receiver for various values of α and λ. Basically, the average ROC's displayed the same features and characteristics as those of Figs. 4.2 and 4.4.

4.2.2.3 Signal Known Exactly, Unknown Noise Level, Gamma A Priori Density.

For the detection problem, signal known exactly, unknown noise level (SKE-UNL), the autocovariance matrix of the observation is known only to within a scale factor. If Ψ denotes the autocovariance matrix, then

$$\Psi = \sigma^2 \Sigma \qquad (4.37)$$

where Σ is known and σ^2 represents the unknown noise level. If we let

$$\theta = \frac{1}{\sigma^2} \qquad (4.38)$$

then the conditional observation statistics become

$$f(\underline{x} \mid \theta, N) = \left(\frac{\theta}{2\pi|\Sigma|}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x}^t \Sigma^{-1} \underline{x}} \qquad (4.39)$$

$$f(\underline{x} \mid \theta, SN) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2}(\underline{x} - \underline{s})^t \Sigma^{-1}(\underline{x} - \underline{s})} \qquad (4.40)$$

Reduction of dimensionality

The following transformation turns out to be very useful. (Details are discussed in Appendix G.)

If we let

$$u = \frac{y}{d_0} - \frac{1}{2} \qquad (4.41)$$

$$v^2 = \frac{w}{d_0} - \left(\frac{y}{d_0}\right)^2 \qquad (4.42)$$

where

$$w = \underline{x}^t \Sigma^{-1} \underline{x} \qquad (4.43)$$

$$y = \underline{s}^t \Sigma^{-1} \underline{x} \qquad (4.44)$$

$$d_0 = \underline{s}^t \Sigma^{-1} \underline{s} \qquad \text{(nominal signal-to-noise ratio)} \qquad (4.45)$$

then

$$f(u \mid \theta, N) = \left(\frac{\theta d_0}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta d_0 (u + \frac{1}{2})^2}{2}} \qquad -\infty < u < \infty \qquad (4.46)$$

$$f(v \mid \theta, N) = \frac{2\left(\frac{\theta \, d_0}{2}\right)^{\frac{n-1}{2}} v^{n-2} \, e^{-\frac{\theta \, d_0 \, v^2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \qquad 0 \leq v < \infty \quad (4.47)$$

$$f(u \mid \theta, SN) = \left(\frac{\theta \, d_0}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta \, d_0 (u - \frac{1}{2})^2}{2}} \qquad -\infty < u < \infty \quad (4.48)$$

$$f(v \mid \theta, SN) = f(v \mid \theta, N) \qquad\qquad\qquad 0 \leq v < \infty \quad (4.49)$$

The quantity (u, v) is a sufficient statistic of the observation and has reduced an n-dimensional problem to a 2-dimensional problem.

Conditionally optimum receiver

In terms of (u, v) the conditional likelihood ratio is

$$\ell(u, v \mid \theta) = \frac{f(u, v \mid \theta, SN)}{f(u, v \mid \theta, N)} = \frac{f(u \mid \theta, SN) \, f(v \mid \theta, SN)}{f(u \mid \theta, N) \; f(v \mid \theta, N)}$$

$$= e^{\theta \, d_0 u} \qquad\qquad (4.50)$$

Since $\ell(u, v \mid \theta) > \beta$ implies that $u > \frac{\ell n \beta}{\theta \, d_0}$, the decision rule of the conditionally optimum receiver is

$$\phi(\underline{x}) = \begin{cases} 1 & u > \frac{\ell n \beta}{\theta \, d_0} \\ \\ 0 & \text{otherwise} \end{cases} \qquad (4.51)$$

The decision region is pictured in Fig. 4.5.
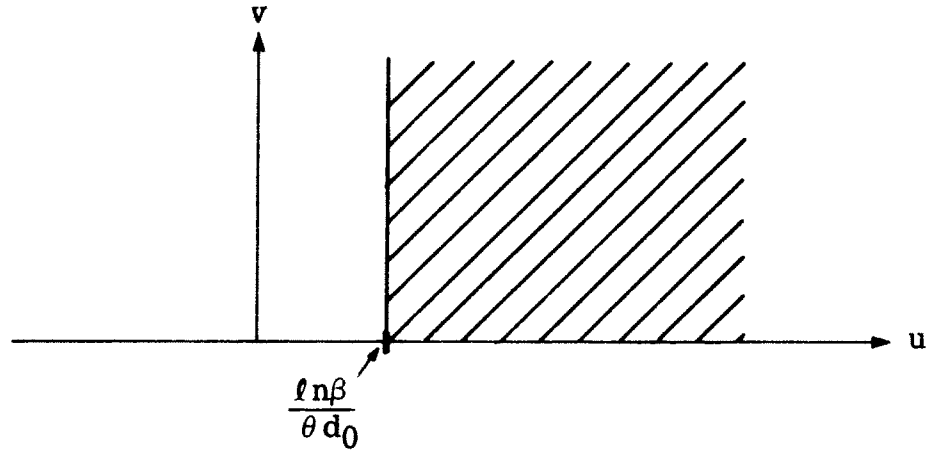
Figure 4.5. Decision region of the conditionally optimum receiver, SKE-UNL.

Dependence of the decision region only on u simplifies calculation of the conditionally optimum ROC. The performance equations are

$$P^{CO}(A \mid \theta, N) = \int_{u > \frac{\ell n \beta}{\theta \, d_0}} f(u \mid \theta, N) \, du = \Phi\left(\frac{-\ell n \beta}{\sqrt{\theta \, d_0}} - \frac{\sqrt{\theta \, d_0}}{2}\right) \tag{4.52}$$

$$P^{CO}(A \mid \theta, SN) = \int_{u > \frac{\ell n \beta}{\theta \, d_0}} f(u \mid \theta, SN) \, du = \Phi\left(\frac{-\ell n \beta}{\sqrt{\theta \, d_0}} + \frac{\sqrt{\theta \, d_0}}{2}\right) \tag{4.53}$$

Conditionally suboptimum receiver

The conditionally suboptimum receiver is based on the total

likelihood ratio. The total likelihood ratio can be determined from the absolute observation statistics

$$f(u, v \mid N) = \int_{\Theta} f(u, v \mid \theta, N) \; g(\theta) \; d\theta \qquad (4.54)$$

$$f(u, v \mid SN) = \int_{\Theta} f(u, v \mid \theta, SN) \; g(\theta) \; d\theta \qquad (4.55)$$

If we model the unknown parameter θ with a gamma a priori density

$$g(\theta) = \frac{c^{b+1}}{\Gamma(b+1)} \; \theta^b \; e^{-c\theta} \qquad c > 0, \quad b > -1, \quad 0 \le \theta < \infty \quad (4.56)$$

then performing the required integration is straightforward and yields

$$f(u, v \mid N) = K_n \; v^{n-2} \left[(u + \tfrac{1}{2})^2 + v^2 + \frac{2c}{d_0} \right]^{-(b+1+\frac{n}{2})} \qquad (4.57)$$

$$f(u, v \mid SN) = K_n \; v^{n-2} \left[(u - \tfrac{1}{2})^2 + v^2 + \frac{2c}{d_0} \right]^{-(b+1+\frac{n}{2})} \qquad (4.58)$$

where

$$K_n = \frac{2 \left(\dfrac{2c}{d_0} \right)^{b+1} \Gamma(b+1+\frac{n}{2})}{\sqrt{\pi} \; \Gamma\left(\dfrac{n-1}{2}\right) \Gamma(b+1)} \qquad (4.59)$$

Hence, the total likelihood ratio is

$$\ell(u, v) = \frac{f(u, v \mid SN)}{f(u, v \mid N)} = \left[\frac{(u + \frac{1}{2})^2 + v^2 + \frac{2c}{d_0}}{(u - \frac{1}{2})^2 + v^2 + \frac{2c}{d_0}} \right]^{(b + 1 + \frac{n}{2})} \tag{4.60}$$

and the decision rule of the conditionally suboptimum receiver is

$$\phi(\underline{x}) = \begin{cases} 1 & \ell(u, v) > \beta \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.61}$$

Some algebraic manipulation reveals that $\ell(u, v) > \beta$ implies

$$(u - a)^2 + v^2 > r^2 \quad \text{if} \quad \Delta < 1$$

$$(u - a)^2 + v^2 < r^2 \quad \text{if} \quad \Delta > 1$$

$$u > 0 \quad \text{if} \quad \Delta = 1 \tag{4.62}$$

where

$$a = \frac{1}{2} \frac{\Delta + 1}{\Delta - 1} \tag{4.63}$$

$$r^2 = \frac{\Delta}{(\Delta - 1)^2} - \frac{2c}{d_0} \tag{4.64}$$

$$\Delta = \beta^{\frac{1}{b + 1 + \frac{n}{2}}} \tag{4.65}$$

The decision region is pictured in Fig. 4.6.
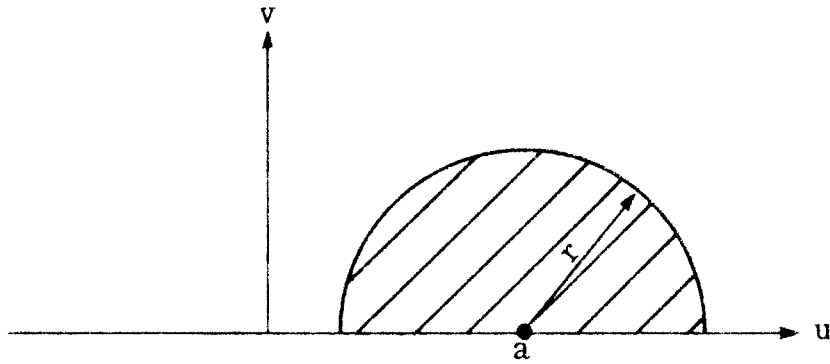
Figure 4.6. Decision region of the conditionally suboptimum receiver, SKE-UNL.

Conditionally suboptimum receiver operation consists of determining whether the observation (u, v) lies inside or outside of the semi-circular region described by Eq. 4.62. This is in contrast to the decision region of the conditionally optimum receiver (Fig. 4.5) which consisted of a half plane.

The conditionally suboptimum ROC's are described by the parametric equations

$$P^{CS}(A \mid \theta, N) = \int_{\ell(u, v) > \beta} f(u, v \mid \theta, N) \, du \, dv$$

$$= \int_0^r f(v \mid \theta, N) \int_{a - \sqrt{r^2 - v^2}}^{a + \sqrt{r^2 - v^2}} f(u \mid \theta, N) \, du \, dv$$

$$= \int_0^r f(v \mid \theta, N) \left[\Phi(R \sqrt{\theta}) - \Phi(S \sqrt{\theta}) \right] dv \qquad (4.66)$$

$$p^{CS}(A \mid \theta, SN) = \int\limits_{\ell(u,v) > \beta} f(u, v \mid \theta, SN) \, du \, dv$$

$$(4.67)$$

$$= \int_0^r f(v \mid \theta, SN) \left[\Phi(R' \sqrt{\theta}) - \Phi(S' \sqrt{\theta}) \right] dv$$

where

$$R = \left(a + \tfrac{1}{2} + \sqrt{r^2 - v^2} \right) \sqrt{d_0} \qquad (4.68)$$

$$S = \left(a + \tfrac{1}{2} - \sqrt{r^2 - v^2} \right) \sqrt{d_0} \qquad (4.69)$$

and

$$R' = \left(a - \tfrac{1}{2} + \sqrt{r^2 - v^2} \right) \sqrt{d_0} \qquad (4.70)$$

$$S' = \left(a - \tfrac{1}{2} - \sqrt{r^2 - v^2} \right) \sqrt{d_0} \qquad (4.71)$$

ESP ROC

The average ROC's can be determined in terms of the conditional ROC's. The ESP ROC is described by

$$p^{ESP}(A \mid N) = \int_0^\infty p^{CO}(A \mid \theta, N) \, g(\theta) \, d\theta$$

$$= \int_0^\infty \Phi \left(\frac{-\ell n \beta}{\sqrt{\theta \, d_0}} - \frac{\sqrt{\theta \, d_0}}{2} \right) g(\theta) \, d\theta \qquad (4.73)$$

$$P^{ESP}(A \mid SN) = \int_0^\infty P^{CO}(A \mid \theta, SN) \, g(\theta) \, d\theta$$

$$= \int_0^\infty \Phi\left(\frac{-\ln\beta}{\sqrt{\theta \, d_0}} + \frac{\sqrt{\theta \, d_0}}{2}\right) g(\theta) \, d\theta \qquad (4.74)$$

Optimum ROC

The optimum ROC can be determined from the conditionally suboptimum ROC's.

$$P^{OPT}(A \mid N) = \int_0^\infty P^{CS}(A \mid N) \, g(\theta) \, d\theta$$

$$(4.75)$$

$$= \int_0^\infty g(\theta) \int_0^r f(v \mid \theta, N) \left[\Phi(A \sqrt{\theta}) - \Phi(B \sqrt{\theta})\right] dv \, d\theta$$

Reverse the order of integration, make the substitution $\xi = \sqrt{\theta (2c + d_0 v^2)}$ and utilize Appendix H to yield

$$P^{OPT}(A \mid N) = K \int_0^r v^{n-2} \left(v^2 + \frac{2c}{d_0}\right)^{-\frac{\nu}{2}} \left[T_\nu(t_2) - T_\nu(t_1)\right] dv$$

$$(4.76)$$

where

$$\nu = 2b + n + 1 \qquad (4.77)$$

$$t_1 = \sqrt{\frac{\nu}{v^2 + \frac{2c}{d_0}}} \left(a + \tfrac{1}{2} - \sqrt{r^2 - v^2}\right) \qquad (4.78)$$

$$t_2 = \sqrt{\frac{\nu}{v^2 + \frac{2c}{d_0}}} \left(a + \tfrac{1}{2} + \sqrt{r^2 - v^2}\right) \qquad (4.79)$$

$$K = \frac{2\left(\frac{2c}{d_0}\right)^{b+1} \Gamma\left(b + \frac{n+1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right) \Gamma(b+1)} \qquad (4.80)$$

and

$$T_\nu(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\ \Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{t} \left(1 + \frac{x^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} dx \qquad (4.81)$$

Similarly,

$$P^{OPT}(A\,|\,SN) = K \int_0^r v^{n-2} \left(v^2 + \frac{2c}{d_0}\right)^{-\frac{\nu}{2}} [T_\nu(t_2') - T_\nu(t_1')]\ dv$$

$$(4.82)$$

where

$$t_1' = \sqrt{\frac{\nu}{v^2 + \frac{2c}{d_0}}} \left(a - \tfrac{1}{2} - \sqrt{r^2 - v^2}\right) \qquad (4.83)$$

$$t_2' = \sqrt{\frac{\nu}{v^2 + \frac{2c}{d_0}}} \left(a - \tfrac{1}{2} + \sqrt{r^2 - v^2}\right) \qquad (4.84)$$

Evaluation of receiver performance

Although the equations describing the various ROC's are not in a very convenient form, they are amenable to numerical integration techniques. The ROC's presented in Figs. 4.7 - 4.10 were obtained in this manner. The conditional ROC's depend only on $\frac{d_0}{c}$, $\theta \, d_0$ and n. The parameter b is absorbed into the threshold and does not affect the conditional ROC's. The average ROC's depend only on $\frac{d_0}{c}$, n and b and are independent of θ by virtue of the fact that we are looking at average performance.

Conditional ROC's

The conditional ROC's of Fig. 4.7 clearly show how uncertainty about the noise level affects performance. Whenever the threshold is unity $(\beta = 1)$, conditionally suboptimum receiver performance is equivalent to conditionally optimum receiver performance. This feature is easily explained by a closer inspection of the decision regions. The decision region of the conditionally suboptimum receiver is the semicircular region depicted in Fig. 4.6. When the threshold is set to unity, the semicircular region degenerates into the half plane $u > 0$. The decision region of the conditionally optimum receiver depicted in Fig. 4.5 is also the half plane $u > 0$ whenever the threshold is set to unity. Hence performance is equal at that particular threshold setting.

The importance of this feature lies in the fact that many binary communication systems operate under 1) the assumption of equally
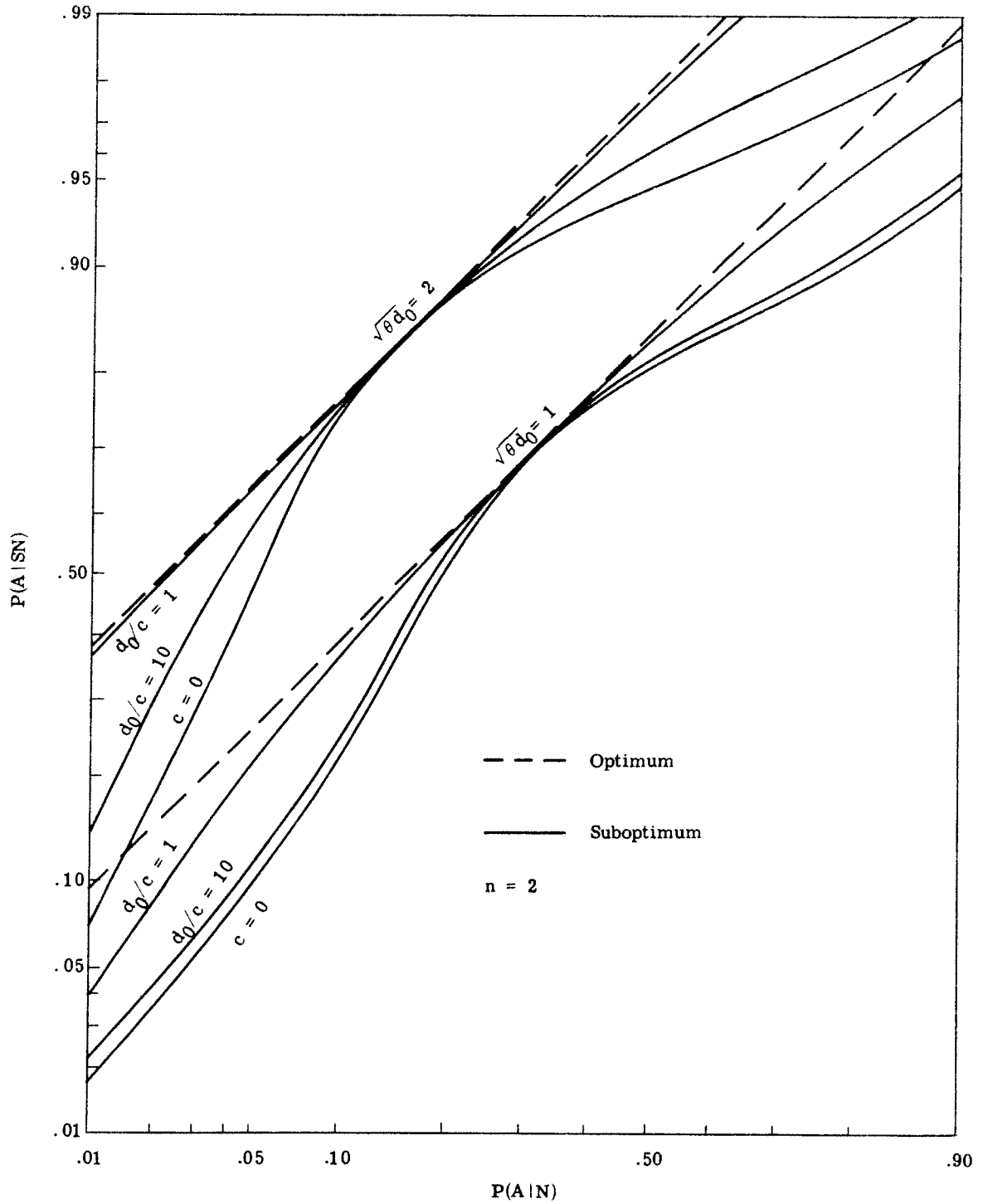
Figure 4.7. Conditional ROC's, comparison of optimum and suboptimum receiver performance, SKE-UNL, n = 2.
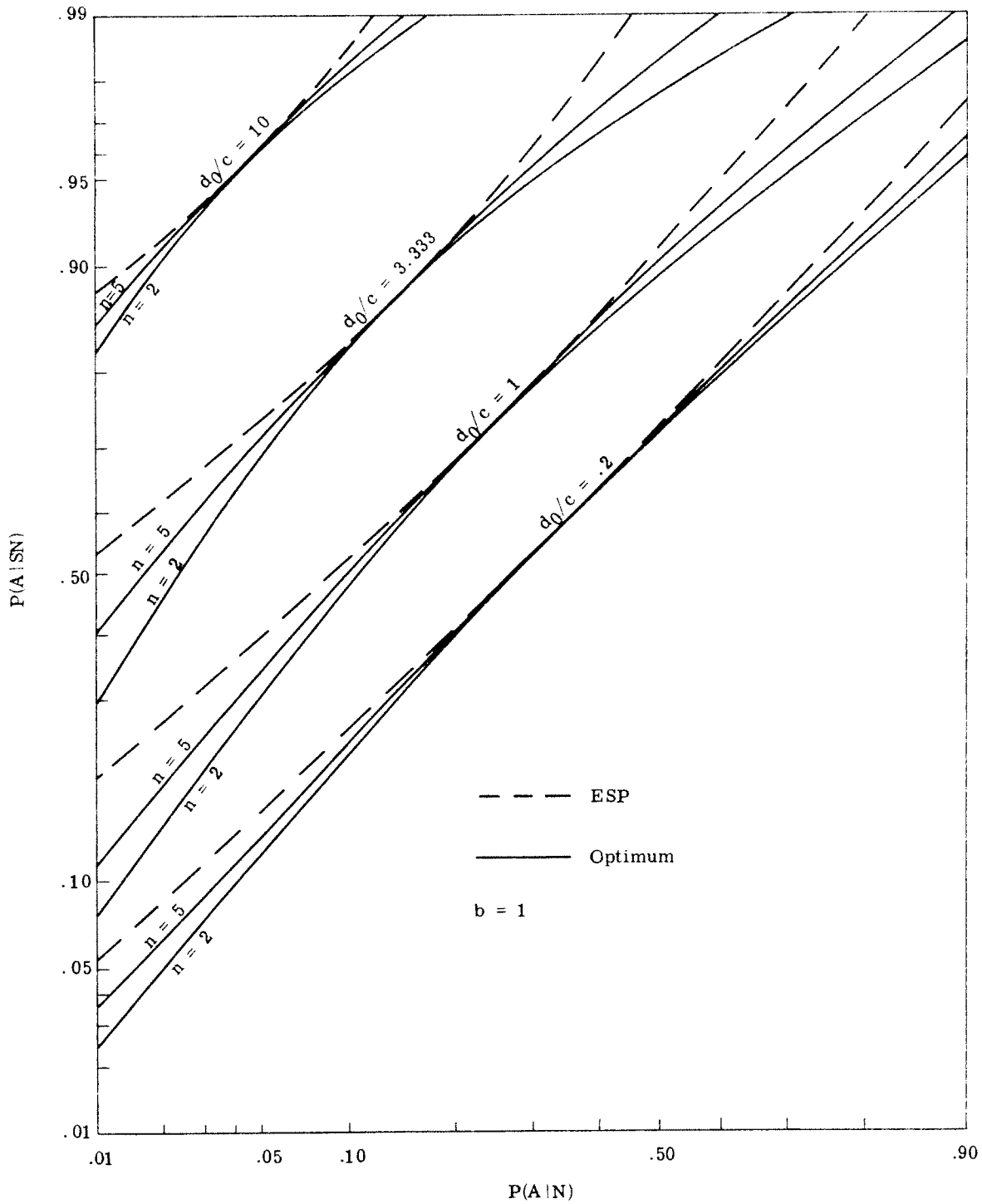
Figure 4.8. Average ROC's, comparison of ESP and optimum receiver performance, SKE-UNL, b = 1.
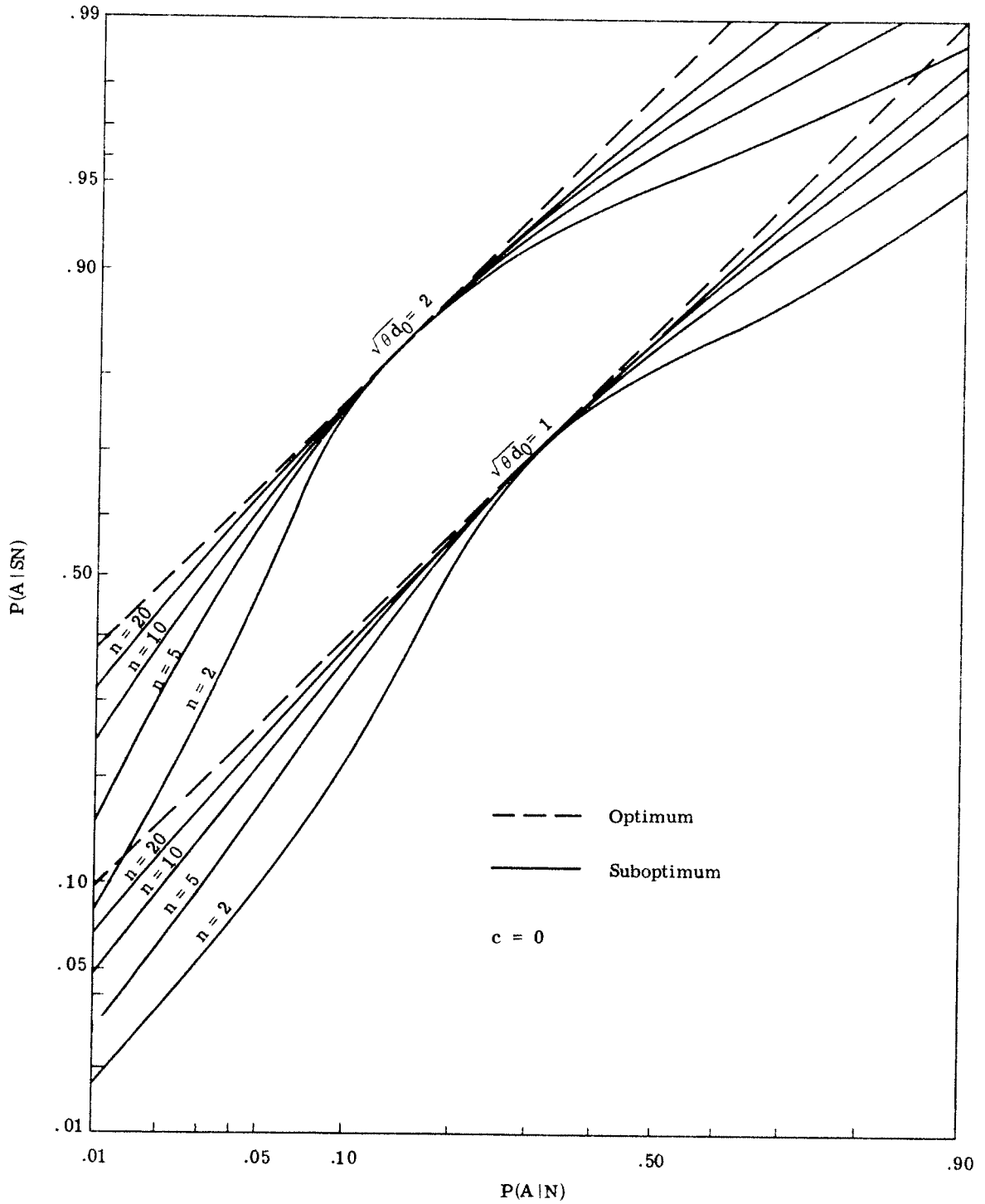
Figure 4.9. Conditional ROC's, comparison of optimum and suboptimum receiver performance, SKE-UNL, c = 0.
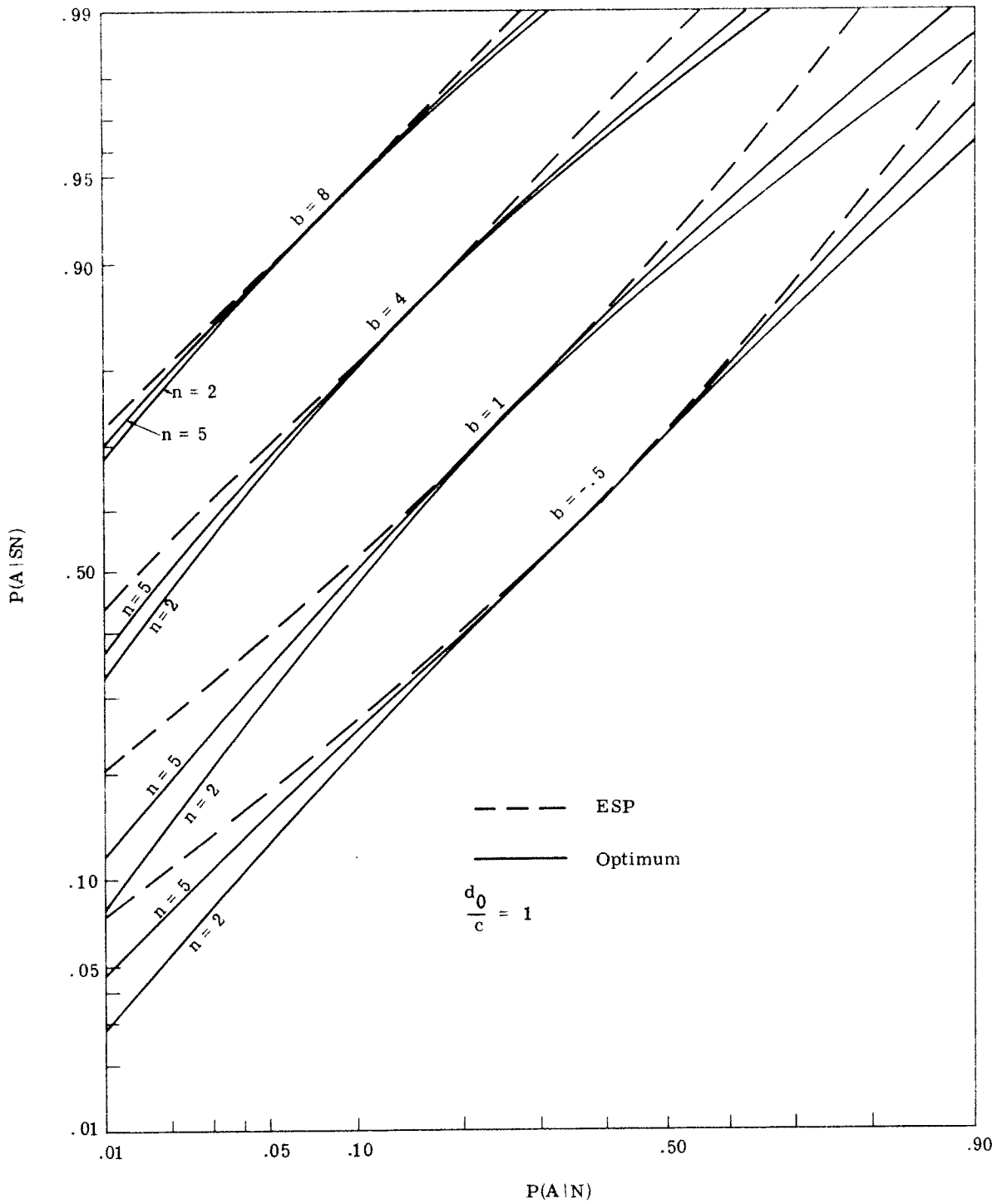
Figure 4.10. Average ROC's, comparison of ESP and optimum receiver performance, SKE-UNL, $d_o/c = 1$.

likely signals and 2) the criterion, maximize a posteriori probability. Hence, decisions are based on likelihood ratio with a threshold of unity; and performance is optimized and independent of fluctuations in noise level.

Conditionally optimum receiver performance depends solely upon the actual signal-to-noise ratio, $\theta\, d_0$, whereas conditionally suboptimum receiver performance depends primarily upon the expected value of the signal-to-noise ratio,

$$E(\theta\, d_0) \;=\; (b+1)\, \frac{d_0}{c} \tag{4.85}$$

The conditional ROC's of Fig. 4.7 indicate that conditionally suboptimum receiver performance converges to conditionally optimum receiver performance as $\frac{d_0}{c} \to 0$. The following interpretation is offered: whenever the actual signal-to-noise ratio, $\theta\, d_0$, is in close agreement with the ratio $\frac{d_0}{c}$ (the parameter b is absorbed into the threshold and affects only the particular operating point, not the ROC), a close correlation exists between conditionally optimum and conditionally suboptimum receiver performance. For instance, the conditionally suboptimum receiver designed for $\frac{d_0}{c}$ large $\left(\frac{d_0}{c} > 10\right)$ performs rather poorly whenever $\theta\, d_0$ is small ($\theta\, d_0 < 10$). Contrariwise, the conditionally suboptimum receiver designed for $\frac{d_0}{c}$ small $\left(\frac{d_0}{c} < 1\right)$ performs very well for $\theta\, d_0$ large ($\theta\, d_0 > 1$). Hence, it is very important to assign an a priori density which reflects

as closely as possible the actual environmental conditions. When

much uncertainty concerning the actual state exists, be realistic in

assigning an a priori density.

Average ROC's

The average ROC's of Fig. 4.8 are parameterized by $\dfrac{d_0}{c}$

and n and reflect the conditional ROC's of Fig. 4.7. The ESP ROC

and the optimum ROC average the conditionally optimum ROC's and

the conditionally suboptimum ROC's respectively with respect to the

given a priori density. Equality of the ESP ROC and the optimum

ROC on the negative diagonal is apparent, independent of $\dfrac{d_0}{c}$, b or

n. Performance increases as a function of $\dfrac{d_0}{c}$. In addition, the

optimum ROC converges to the ESP ROC as $n \to \infty$, although con-

vergence is not uniform along the ROC.

Diffuse a priori density

The conditional ROC's of Fig. 4.9 and the average ROC's of

Fig. 4.10 behave in an analogous manner. The conditional ROC's of

Fig. 4.9 were presented for c = 0 (or $\dfrac{d_0}{c} = \infty$) in order to make the

effect of n obvious. A value of c = 0 corresponds to a diffuse a priori

density. The conditionally suboptimum ROC's nevertheless converge

quite rapidly to the conditionally optimum ROC's independent of $\theta \, d_0$.

Hence, conditionally suboptimum receiver performance can be consid-

erably enhanced if the total observation is at least 20-dimensional.

Figure 4.10 presents the ESP and optimum ROC's for $\dfrac{d_0}{c} = 1$

(the ROC's are singular for $c = 0$). Convergence of the ESP ROC to the optimum ROC is apparent as $n \to \infty$ independent of b and $\frac{d_0}{c}$. In addition, <u>average performance improves as b increases</u>.

Summary

It has been shown that the ESP ROC serves as a useful upper bound to the optimum ROC. The ESP ROC averages the conditionally optimum ROC's while the optimum ROC averages the conditionally suboptimum ROC's. The importance of matching a priori specification to the environmental conditions has been demonstrated. In addition, performance of the conditionally suboptimum receiver has been shown to converge asymptotically to the performance of the conditionally optimum receiver as the dimensionality of the total observation is allowed to increase.

4.3 Classical Approach

The classical approach to composite hypothesis decision theory consists essentially of constructing a test (decision region) which is either uniformly most powerful (UMP) or uniformly most powerful unbiased (UMPU) or of estimating the unknown parameter(s) by some appropriate method and then proceeding as though the estimator(s) were exact, e.g., the "generalized likelihood ratio" method. An unknown parameter is not assigned an a priori density. Nevertheless, Bayesian and classical approaches to detection theory, though at odds,

often lead to tests which yield equivalent performance, especially for certain single composite hypotheses. All the tests (or receivers) in this section will be evaluated conditionally since this is the only available means of comparison.

4.3.1 Uniformly Most Powerful Tests.

A uniformly most powerful (UMP) test is basically a test which yields the same performance that would have been obtained if the unknown N or SN process parameter(s) θ were known. In addition a UMP test requires that the probability of false alarm be independent of the unknown parameter(s). In essence, the ROC corresponding to the UMP test (UMP ROC) is equivalent to the conditionally optimum ROC for all $\theta \in \Theta$. It is not particularly difficult to construct a test whose ROC is equivalent to the conditionally optimum ROC for a particular choice of θ; but to construct a test which is equivalent for all $\theta \in \Theta$ is possible only in special circumstances. A more detailed discussion of the UMP test can be found in Refs. 6-8.

Example

A UMP test is obtained for the detection problem SKEA if it is known that the amplitude is nonnegative (a $\geq$ 0, i.e., the sign of the amplitude is known). The UMP test is

$$\phi(\underline{x}) = \begin{cases} 1 & y > \alpha \\ \\ 0 & \text{otherwise} \end{cases} \tag{4.86}$$

where y is the sufficient statistic

$$y = \underline{x}^t \Sigma^{-1} \underline{s} \qquad (4.87)$$

Hence, the receiver consists of a device which computes the sufficient statistic y and compares it to a threshold α to make a decision as to presence or absence of a signal. The UMP ROC will be equivalent to the conditionally optimum ROC for all $a \in [0, \infty)$ since the test (receiver operation) is of the same form.

Comparison to Bayes test

The Bayesian analysis of this problem (Section 4.2.2.2) incorporated the knowledge $a \geq 0$ by assigning an a priori density which gave zero probability to negative values of the unknown amplitude; the corresponding receiver operation and performance are identical to that of the UMP test.

It appears as though Bayesian analysis might be sufficiently general that it could conceivably encompass classical statistics as a subset. Further investigation of this aspect of the problem was deemed worthwhile and will be pursued in later sections.

4.3.2 Uniformly Most Powerful Unbiased Tests. A uniformly most powerful unbiased (UMPU) test is basically a test which never yields performance poorer than chance and is the best test among all tests satisfying this criteria. An additional restriction placed on a UMPU test is that the probability of false alarm be independent of the

unknown parameter(s). Naturally, such a test is highly desirable,
but the existence of a UMPU test is not always assured. However,
when the conditional observation statistics belong to the exponential
family, a UMPU test exists (Refs. 7 and 8). Thus, a UMPU test
exists if the conditional observation densities are multivariate normal
densities.

Example

The detection problem **SKEA** with no restriction on the range
of the unknown amplitude (i.e., $-\infty < a < \infty$) will yield a UMPU test.
It is the familiar two-sided test (Ref. 7, p. 225).

$$
\phi(\underline{x}) = \begin{cases} 1 & |y| > \alpha \\ \\ 0 & |y| \leq \alpha \end{cases} \tag{4.88}
$$

Hence, receiver operation consists of determining the sufficient sta-
tistic y and comparing its absolute value to a threshold α.

Comparison to Bayes test

The receiver designed by Bayesian methods (Eq. 4.17) with a
symmetric a priori density (m = 0) is based on a test which is iden-
tical to the UMPU test. Assignment of an a priori density symmetric
about zero reflects the designer's knowledge (or opinion) that the sign
of the amplitude is equally likely to be + or -, the maximum entropy
condition for sign knowledge. Thus, Bayesian analysis was sufficiently

general to include the UMPU test in its repertoire and flexible enough to incorporate information ignored by classical statistics. Although Bayesian methods can lead to biased tests, this should not be considered a defect since biased ROC's will occur seldom (i.e., with small probability).

Example

A UMPU test can be obtained for the detection problem SKE-UNL if we let the observations x_i, $i = 1, 2, \ldots n$ be independent and specify the signal as constant, i.e., let

$$\Sigma = I \tag{4.89}$$

and

$$\underline{s} = s_0 \underline{1} \tag{4.90}$$

where

$$\underline{1} = (1, 1, \ldots, 1)^t \tag{4.91}$$

and s_0 is a known constant.

Subject to these conditions, the sufficient statistics w and y of the Bayes test (Eqs. 4.43 and 4.44) become

$$w = \underline{x}^t \Sigma^{-1} \underline{x} = \underline{x}^t \underline{x} = \sum_{i=1}^{n} x_i^2 \tag{4.92}$$

$$y = \underline{s}^t \Sigma^{-1} \underline{x} = \underline{s}^t \underline{x} = \sum_{i=1}^{n} x_i s_i = s_0 \sum_{i=1}^{n} x_i \qquad (4.93)$$

and the nominal signal-to-noise ratio (Eq. 4.45) becomes

$$d_0 = \underline{s}^t \Sigma^{-1} \underline{s} = \underline{s}^t \underline{s} = \sum_{i=1}^{n} s_i^2 = n s_0^2 \qquad (4.94)$$

Furthermore, the transformed random variables u and v (Eqs. 4.41 and 4.42) become

$$u = \frac{y}{d_0} - \frac{1}{2} = \frac{\bar{x}}{s_0} - \frac{1}{2} \qquad (4.95)$$

$$v^2 = \frac{w}{d_0} - \left(\frac{y}{d_0}\right)^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{d_0} \qquad (4.96)$$

where

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = s_0 (u + \tfrac{1}{2}) \qquad (4.97)$$

The UMPU test is the one-sided test (Ref. 7, p. 229)

$$\phi(\underline{x}) = \begin{cases} 1 & \frac{y}{\sqrt{w}} > \alpha \\ \\ 0 & \text{otherwise} \end{cases} \qquad (4.98)$$

This test can be expressed in terms of the standard t-test as

$$\phi(\underline{x}) = \begin{cases} 1 & \dfrac{\sqrt{n-1}\ \bar{x}}{v} > \alpha' \\[2em] 0 & \text{otherwise} \end{cases} \tag{4.99}$$

or equivalently,

$$\phi(\underline{x}) = \begin{cases} 1 & \dfrac{\sqrt{n-1}\ (u + \frac{1}{2})}{v} > \alpha'' \\[2em] 0 & \text{otherwise} \end{cases} \tag{4.100}$$

The decision region turns out to be a half-cone with its center at

$u = -\frac{1}{2}$, $v = 0$. A typical decision region is pictured in Fig. 4.11.
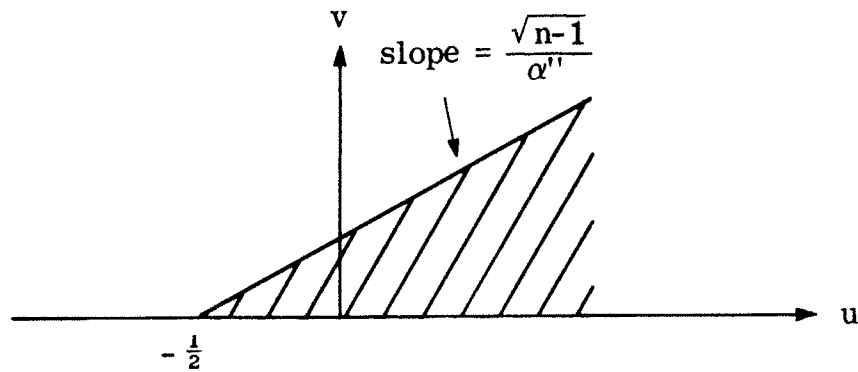
Figure 4.11. Decision region of the UMPU test, SKE-UNL.

Interpretation

The random variable $u + \frac{1}{2}$ is the signal portion of the obser-
vation, the part of the observation in the signal direction, or the part
that "looks just like" the signal. The random variable v is the noise
portion of the observation (the signal direction is excluded), the part
of the observation that can be viewed as a r.m.s. noise level estima-
tor.

The Bayesian approach decides "signal is present" only if the
observation (u, v) looks like the known signal with some modest
amount of noise added to it.

The classical solution is illustrated in Fig. 4.12. The
classical approach bases decisions on the estimated signal-to-
estimated noise ratio, $(u + \frac{1}{2})/v$; it may decide "signal is present"
even though the signal portion of the observation, $u + \frac{1}{2}$, is "far away"
from the known signal, so long as the apparent signal-to-noise ratio
is high. The classical solution corresponds to the common engineer-
ing solution of processing the signal portion of the observation through
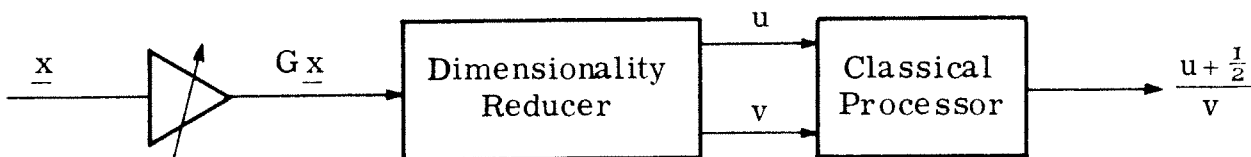
Fig. 4.12. Classical solution to the detection problem SKE-UNL.

an automatic gain control (AGC) device. The advantage of the AGC action is that the decision process is not affected by the observation level, i.e., the detection output, $(u+\frac{1}{2})/v$, is independent of the variable gain G.

Evaluation of receiver performance

The performance equations can be obtained in terms of the central and non-central t-distribution on $n-1$ degrees of freedom with noncentrality parameter δ. Specifically,

$$P(A \mid N) = 1 - T_{n-1, 0}^{(\alpha'')} \tag{4.101}$$

$$P(A \mid SN) = 1 - T_{n-1, \delta}^{(\alpha'')} \tag{4.102}$$

where

$$\delta = \sqrt{\theta \, d_0} \tag{4.103}$$

and

$$T_{n, \gamma}(t) = \frac{\sqrt{\pi}}{2^{\frac{n-3}{2}} \, \Gamma(\frac{n}{2})} \int_0^\infty x^{n-1} \, \Phi'(x) \, \Phi\left(\frac{tx}{\sqrt{n}} - \gamma\right) dx \tag{4.104}$$

Note that the probability of false alarm is entirely independent of both the signal and the noise level as was to be expected. Tests of this type are often denoted as CFAR (constant false alarm rate) tests in the literature. In addition, the probability of detection only depends

upon the actual signal-to-noise ratio, $\theta\, d_0$.

The UMPU ROC's are shown in Fig. 4.13 for various values of $\sqrt{\theta\, d_0}$ and n. At $P(A|N) = .5$ (50% false alarms) receiver performance is independent of n and has a maximum normal index of detectability, although this operating point is unlikely to be acceptable. Whenever $P(A|N) = .5$, the threshold $\alpha'' = 0$ so that the cone-shaped decision region of the UMPU test (Fig. 4.11) degenerates into the same decision region obtained for the conditionally optimum test when $\beta = e^{\frac{\theta\, d_0}{2}}$ (Fig. 4.5), namely the half-plane $u > \frac{1}{2}$. The UMPU ROC's asymptotically approach a normal ROC with performance index $d' = \sqrt{\theta\, d_0}$ as $n \to \infty$. This performance bound corresponds to the ROC of the conditionally optimum receiver for θ known. Convergence of the UMPU ROC's to the conditionally optimum performance bound appears more rapid for small values of $\sqrt{\theta\, d_0}$ $(\theta\, d_0 < 4)$.

Comparison to Bayes test

Comparing Fig. 4.6 and Fig. 4.11 gives one an appreciation of the difference in decision regions between Bayesian and classical analysis. A comparison of the performance of the UMPU test and the Bayes test (Eq. 4.61) can not be made on an absolute scale. Neither the Bayes test nor the classical test yields the best performance that could be attained if θ were known. Both tests have their merits. Obviously, the UMPU test is superior whenever $P(A|N) = .5$.
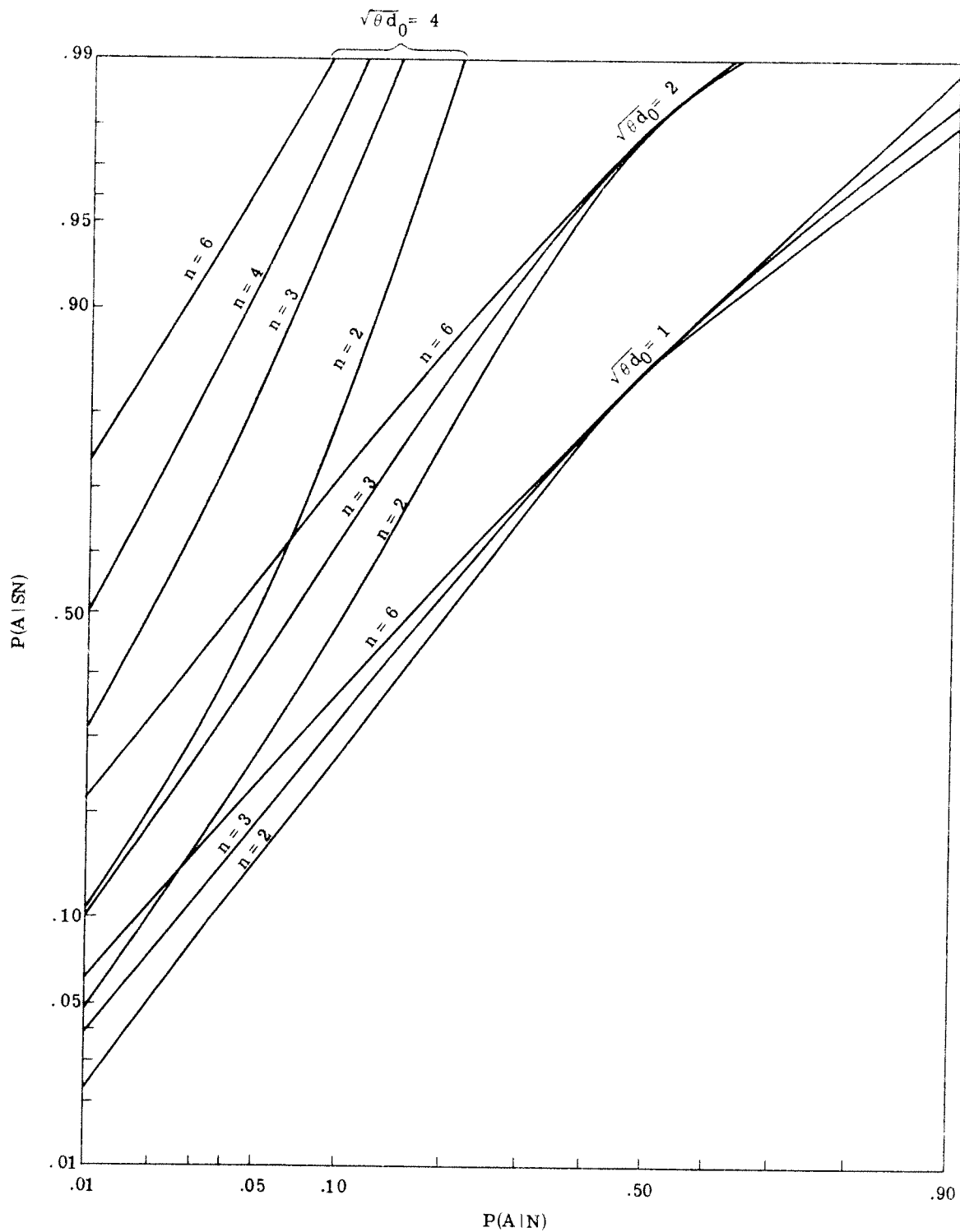
Figure 4.13. ROC's for the UMPU test, SKE-UNL.

By the same token, the Bayes test is superior on the negative dia-gonal [i.e., whenever $P(A \mid N) + P(A \mid SN) = 1$].

Existence or non-existence of an a priori density

Since the Bayes test is not equivalent to the UMPU test for any choice of the parameters of the assigned gamma a priori density, the question of whether or not equivalence can be acquired for any other choice of a priori density is raised. In particular, does there exist an a priori density $g(\cdot)$ such that

$$\ell(\underline{x} \mid g) \uparrow \frac{\underline{x}^t \Sigma^{-1} \underline{s}}{\sqrt{\underline{x}^t \Sigma^{-1} \underline{x}}} \tag{4.105}$$

That is, can we find a function $g(\cdot)$ such that

$$\frac{\int_0^\infty \theta^{\frac{n}{2}} e^{-\frac{\theta}{2}(w - 2y + d_0)} g(\theta) \, d\theta}{\int_0^\infty \theta^{\frac{n}{2}} e^{-\frac{\theta w}{2}} g(\theta) \, d\theta} \uparrow \frac{y}{\sqrt{w}} \tag{4.106}$$

It is shown in Appendix I that no such a priori density exists. Hence, the Bayesian approach and the UMPU test of classical statistics cannot be reconciled for the double composite hypothesis problem, SKE-UNL. The author believes that in general Bayesian and classical procedures (UMPU tests) yield diverse solutions to double composite hypothesis detection problems.

4.3.3 A Modified Uniformly Most Powerful Unbiased Test.

A slight modification of the UMPU test to an altered form of the t-test is of interest, namely

$$
\phi(\underline{x}) = \begin{cases} 1 & \dfrac{\sqrt{n-1}\ u}{v} > \alpha \\[3ex] 0 & \text{otherwise} \end{cases}
\tag{4.107}
$$

The decision region of this test is shown in Fig. 4.14 and is a symmetric version of the decision region of the UMPU test (Fig. 4.11).
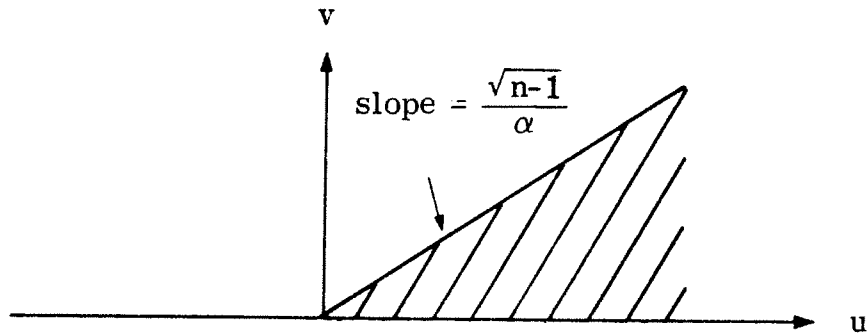
Figure 4.14. Decision region of a modified UMPU test, SKE-UNL.

The performance equations for this test are

$$
P(A \mid N) = 1 - T_{n-1, -\frac{\delta}{2}}(\alpha)
\tag{4.108}
$$

$$
P(A \mid SN) = 1 - T_{n-1, \frac{\delta}{2}}(\alpha)
\tag{4.109}
$$

Figure 4.15 displays the ROC's for various values of n. The ROC's are symmetric about the negative diagonal and equal to the performance bound on the negative diagonal. In addition the ROC's converge to the performance bound everywhere as $n \to \infty$.

Comparison to Bayes test

The modified form of the UMPU test has the same gross characteristics as the Bayes test (Eq. 4.61). A comparison of the conditional ROC's in Fig. 4.14 and Fig. 4.7 or Fig. 4.9 will reveal that the modified UMPU test is not uniformly superior to the Bayes test, or vice versa. Whereas the Bayes test may be superior in a certain range of parameter and threshold values, the modified UMPU test is superior elsewhere. If one is interested in conditional and not average performance, then the particular test one chooses to use in any given situation will depend upon the desired operating point and the surmised state of the unknown noise level.

4.3.4 Estimation Techniques. An easy and appealing way of attacking a composite hypothesis problem is to estimate the unknown parameter(s) and substitute into the observation densities as though the estimator(s) were exact. Minimization of mean square error is an appealing criteria while other desirable qualities in an estimator include unbiasedness, consistency and efficiency. Maximum likelihood techniques often result in estimators meeting one or more of
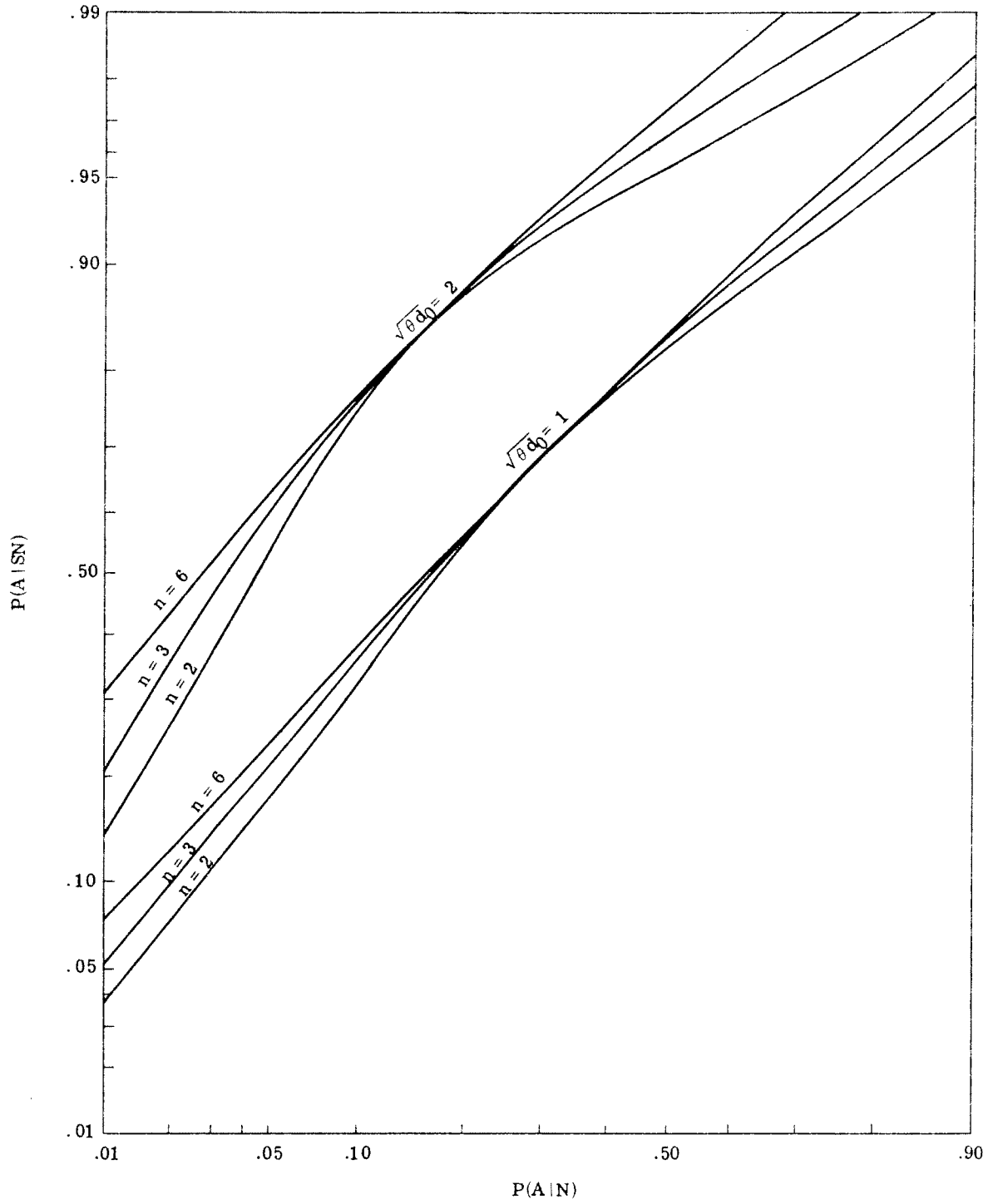
Figure 4.15.   ROC's for a modified UMPU test, SKE-UNL.

these qualities. An excellent discussion of estimation techniques can be found in Ref. 26.

Generalized likelihood ratio

A classical method of estimating unknown parameters is the method of maximum likelihood. An estimator(s) of the unknown parameter(s) is obtained by choosing the estimator $\hat{\theta}$ to be the statistic which maximizes the likelihood function $L(\theta)$ where

$$L(\theta) = f(x \mid \theta) \qquad (4.110)$$

The "generalized likelihood ratio" is obtained by treating these estimators as though they were exact, i.e.,

$$\ell(x \mid \hat{\theta}_N, \hat{\theta}_{SN}) = \frac{f(x \mid \theta = \hat{\theta}_{SN}, SN)}{f(x \mid \theta = \hat{\theta}_N, N)} \qquad (4.111)$$

where $\hat{\theta}_{SN}$ is the maximum likelihood estimator with respect to $f(x \mid \theta, SN)$ and $\hat{\theta}_N$ is the maximum likelihood estimator with respect to $f(x \mid \theta, N)$.

Example

Consider the detection of a signal with an unknown amplitude (SKEA + KGN); and let the range of admissible values be $-\infty < a < \infty$. The likelihood function is

$$L(a) = f(\underline{x} \mid a, SN) = (2\pi \mid \Sigma \mid)^{-\frac{n}{2}} e^{-\frac{1}{2}(\underline{x} - a\,\underline{s})^t \, \Sigma^{-1}(\underline{x} - a\,\underline{s})} \qquad (4.112)$$

$$= (2\pi \mid \Sigma \mid)^{-\frac{n}{2}} e^{-\frac{1}{2}(w - 2ay + a^2 d_0)} \qquad (4.113)$$

where

$$w = \underline{x}^t \, \Sigma^{-1} \, \underline{x} \qquad (4.114)$$

$$y = \underline{x}^t \, \Sigma^{-1} \, \underline{s} \qquad (4.115)$$

and

$$d_0 = \underline{s}^t \, \Sigma^{-1} \, \underline{s} \qquad (4.116)$$

Since the likelihood function $L(a)$ is maximized whenever $a = \dfrac{y}{d_0}$, choose

$$\hat{a} = \frac{y}{d_0} \qquad (4.117)$$

The "generalized likelihood ratio" is

$$\ell(\underline{x} \mid \hat{a}) = \frac{f(\underline{x} \mid a = \hat{a}, SN)}{f(\underline{x} \mid N)}$$

$$= e^{\hat{a}y - \frac{\hat{a}^2}{2} d_0}$$

$$= e^{\frac{y^2}{2d_0}} \qquad (4.118)$$

Since $\ell(\underline{x}\mid\hat{a}) \uparrow |y|$, the decision rule becomes

$$
\phi(\underline{x}) = \begin{cases} 1 & |y| > \alpha \\[2ex] 0 & \text{otherwise} \end{cases} \qquad (4.119)
$$

Comparison to Bayes test

This is precisely the decision rule obtained for the Bayes test (Eq. 4.17) when a symmetric a priori density $(m = 0)$ was assigned. Hence, application of classical estimation techniques for this single composite hypothesis problem resulted in a receiver which is identical in both operation and performance to the Bayes test which assigns a symmetric a priori density.

Example

Consider the double composite hypothesis problem, SKE-UNL. The likelihood function conditional to SN is

$$
\begin{aligned}
L(\theta\mid SN) &= \left(\frac{\theta}{2\pi\mid\Sigma\mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2}(\underline{x}-\underline{s})^{t}\Sigma^{-1}(\underline{x}-\underline{s})} \\[3ex]
&= \left(\frac{\theta}{2\pi\mid\Sigma\mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2}(w - 2y + d_0)} \qquad (4.120)
\end{aligned}
$$

and the likelihood function conditional to N is

$$L(\theta \mid N) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x} \Sigma^{-1} \underline{x}}$$

$$= \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta w}{2}} \tag{4.121}$$

The maximum likelihood estimators are

$$\hat{\theta}_{SN} = \frac{n}{w - 2y + d_0} \tag{4.122}$$

$$\hat{\theta}_N = \frac{n}{w} \tag{4.123}$$

The "generalized likelihood ratio" is

$$\ell(\underline{x} \mid \hat{\theta}_{SN}, \hat{\theta}_N) = \frac{f(\underline{x} \mid \theta = \hat{\theta}_{SN}, SN)}{f(\underline{x} \mid \theta = \hat{\theta}_N, N)}$$

$$= \left(\frac{\hat{\theta}_{SN}}{\hat{\theta}_N}\right)^{\frac{n}{2}}$$

$$= \left(\frac{w}{w - 2y + d_0}\right)^{\frac{n}{2}}$$

$$= \left[\frac{(u + \tfrac{1}{2})^2 + v^2}{(u - \tfrac{1}{2})^2 + v^2}\right]^{\frac{n}{2}} \tag{4.124}$$

where

$$u = \frac{y}{d_0} - \frac{1}{2} \qquad (4.125)$$

$$v^2 = \frac{w}{d_0} - \left(\frac{y}{d_0}\right)^2 \qquad (4.126)$$

(The random variables u and v are sufficient statistics of the observation and are defined as before.)

Comparison to Bayes test

The "generalized likelihood ratio" is the limit of the likelihood ratio of the Bayesian approach (Eq. 4.61) when a diffuse a priori density is approached (b → -1, c → 0). Again, classical analysis led to a solution which can be identified with a particular Bayes solution.

4.4 Comparison of the Bayesian Approach and the Classical Approach

A general discussion of the relation among UMP tests, "generalized likelihood ratio," and total likelihood ratio with respect to a diffuse a priori density is deemed worthwhile; it may shed some light on the specific examples considered earlier from both the Bayesian viewpoint and the classical viewpoint. Of necessity, this discussion has to be from the Bayesian viewpoint, i.e., we must assume a known a priori density for discussion's sake.

The basic equation is Bayes rule, namely

$$f(x, \theta \mid H) = f(x \mid \theta, H) \, g(\theta \mid H) = g(\theta \mid x, H) \, f(x \mid g, H) \qquad (4.127)$$

where H is either N or SN.

4.4.1 Single Composite Hypothesis. No uncertainties exist in the N process for the single composite hypothesis. Hence, the conditional likelihood ratio is

$$\ell(x \mid \theta) = \frac{f(x \mid \theta, SN)}{f(x \mid N)} \qquad (4.128)$$

Since $f(x \mid N)$ is independent of θ, choosing $\theta = \hat{\theta}(x)$ to maximize $\ell(x \mid \theta)$ is equivalent to maximizing $f(x \mid \theta, SN)$. Furthermore, if we rewrite Eq. 4.127 for $H = SN$ and $g(\theta \mid SN) = g(\theta)$, then

$$f(x \mid \theta, SN) \, g(\theta) = g(\theta \mid x, SN) \, f(x \mid g, SN) \qquad (4.129)$$

and if $g(\theta)$ is diffuse or constant [i.e., $g(\theta) = K$ for all $\theta \in \Theta$], then choosing $\theta = \hat{\theta}(x)$ to maximize $f(x \mid \theta, SN)$ is equivalent to maximizing $g(\theta \mid x, SN)$. Hence, maximum likelihood, maximum conditional likelihood ratio, and maximum a posteriori estimators are equal with respect to a diffuse a priori density.

UMP test

Theorem. If there is a uniformly most powerful statistic $t(x)$, then $t(x)$ is sufficient for total (or average) likelihood

ratio independent of the a priori density.

Proof. A uniformly most powerful statistic $t(x)$ has the property that

$$t(x) \uparrow \ell(x|\theta) \quad \text{for each} \quad \theta \in \Theta \qquad (4.130)$$

If $t(x_1) > t(x_2)$, then $\ell(x_1|\theta) > \ell(x_2|\theta)$ for each $\theta \in \Theta$. Since

$$\ell(x|g) = \int_\Theta \ell(x|\theta) \, g(\theta) \, d\theta \qquad (4.131)$$

then $\ell(x_1|g) > \ell(x_2|g)$ independent of the a priori density $g(\theta)$.

Q.E.D.

Thus, the relation between a UMP test (when it exists) and average likelihood ratio is quite strong, and does not depend upon a diffuse a priori density.

Generalized likelihood ratio

Dividing Eq. 4.129 by $f(x|N)$ yields

$$\ell(x|\theta) \, g(\theta) = g(\theta|x, SN) \, \ell(x|g) \qquad (4.132)$$

Let $\hat{\theta}(x)$ be the maximum likelihood estimator of θ, and denote the "generalized likelihood ratio" by

$$M(x) = \ell(x|\theta = \hat{\theta}(x)) \qquad (4.133)$$

If $g(\theta)$ is restricted to be a uniform or diffuse a priori density, then

$$\frac{K\,M(x)}{g(\hat{\theta}(x)|x,\,SN)} = \ell(x|g) \qquad (4.134)$$

Necessary condition

Denote the inverse image of $M(x)$ by R_α, i.e., let

$$R_\alpha = \{x\,|\,M(x) = \alpha\} \qquad (4.135)$$

A necessary condition for the relation

$$M(x) \uparrow \ell(x|g) \qquad (4.136)$$

to hold is that the inverse image of $\ell(x|g)$ must be the inverse image of $M(x)$, i.e., we must have

$$\ell(x|g) = \alpha' \qquad \text{for all } x \in R_\alpha \qquad (4.137)$$

This condition requires that

$$g(\hat{\theta}(x)|x,\,SN) = \frac{K\alpha}{\alpha'} \qquad \text{for all } x \in R_\alpha \qquad (4.138)$$

and necessarily implies that $g(\hat{\theta}(x)|x,\,SN)$ must be a single-valued function of $M(x)$. That is, we must have

$$g(\hat{\theta}(x)|x,\,SN) = p[M(x)] \qquad \text{for all } x \qquad (4.139)$$

so that

$$g(\hat{\theta}(x)|x, SN) = p(\alpha)$$

$$= \text{constant} \qquad \text{for all } x \in R_{\alpha} \qquad (4.140)$$

Example

The most well known example of this is the case of signal known except for carrier phase (Ref. 4) where the uniform a priori density yields

$$\ell(x|g) = e^{-\frac{d}{2}} I_0[r(x)] \qquad (4.141)$$

$$M(x) = e^{r(x) - \frac{d}{2}} \qquad (4.142)$$

$$g(\hat{\theta}(x)|x, SN) = \frac{e^{r(x)}}{2\pi I_0[r(x)]} = \frac{e^{\frac{d}{2}} M(x)}{2\pi I_0[\ell n M(x) + \frac{d}{2}]} \qquad (4.143)$$

where

d   = normal detection index if θ is known

r(x)   is the suitably normalized envelope of a matched filter output

$I_0(\cdot)$ is the modified Bessel function of the first kind.

Further necessary condition

If the first necessary condition is met, i.e., if

$$g(\hat{\theta}(x)|x, SN) = p[M(x)] \qquad \text{for all } x \qquad (4.144)$$

then $\ell(x\,|\,g) \uparrow M(x)$ is equivalent to

$$\frac{M(x)}{p[M(x)]} \quad \uparrow \quad M(x) \tag{4.145}$$

If the function $p(\cdot)$ is differentiable, then monotonicity of the above relation corresponds to

$$\frac{d}{dM}\left[\frac{M}{p(M)}\right] > 0 \qquad \text{for all } M \tag{4.146}$$

or equivalently,

$$\frac{p(M)}{M} > p'(M) \qquad \text{for all } M \tag{4.147}$$

Example

For the example considered above

$$p(M) = \frac{e^{\frac{d}{2}} M}{2\pi I_0[\ell n\, M + \frac{d}{2}]} \tag{4.148}$$

and $\dfrac{d}{dM}\left[\dfrac{M}{p(M)}\right] > 0$ corresponds to

$$I_0'[r(x)] > 0 \tag{4.149}$$

Since $r(x) \geq 0$ for all x and since $I_0'(\cdot) > 0$ for all positive values of its argument, $\ell(x\,|\,g) \uparrow M(x)$ for this example.

Irrelevance

This discussion is somewhat beside the point, because the "generalized likelihood ratio" is usually employed when computation of the total likelihood ratio (for any a priori density) is insurmountably difficult. If $g(\theta \mid x, SN)$ were known, then $\ell(x \mid g)$ could be found simply.

$$\ell(x \mid g) = M(x) \frac{g(\hat{\theta}(x))}{g(\hat{\theta}(x) \mid x, SN)} \tag{4.150}$$

Although we may talk of total likelihood ratio and a posteriori densities arising from diffuse a priori densities, we must recognize that they will be difficult to obtain analytically in most cases where the "generalized likelihood ratio" is employed.

Therefore, let us discuss these topics in general terms; and pictorially, not rigorously, we turn to the double composite hypothesis case for this discussion.

4.4.2 Double Composite Hypothesis. Let $\hat{\theta}_H$ maximize $f(x \mid \theta, H)$ and consider

$$M(x) = \ell(x \mid \hat{\theta}_{SN}, \hat{\theta}_N) = \frac{f(x \mid \hat{\theta}_{SN}, SN)}{f(x \mid \hat{\theta}_N, N)} \tag{4.151}$$

compared to

$$\ell(x \mid g) = \frac{\int_{\Theta} f(x \mid \theta, SN) \, g(\theta \mid SN) \, d\theta}{\int_{\Theta} f(x \mid \theta, N) \, g(\theta \mid N) \, d\theta} \qquad (4.152)$$

If $f(x \mid \theta, H)$ is "concentrated" in θ near $\hat{\theta}_H$ for both hypotheses, which we intuitively feel is necessary for $\hat{\theta}_H$ to be a good estimator, we would be willing to use the estimate-and-plug procedure of "generalized likelihood ratio." Following the logic found so useful elsewhere in "ellipse of concentration" and "equivalent square bandwidth," let us try a "region of minimum size" such that

$$\int_{\Theta} f(x \mid \theta, H) \, g(\theta \mid H) \, d\theta = f(x \mid \hat{\theta}_H, H) \int_{R(x, H)} g(\theta \mid H) \, d\theta \qquad (4.153)$$

Then

$$\ell(x \mid g) = \frac{f(x \mid \hat{\theta}_{SN}, SN) \int_{R(x, SN)} g(\theta \mid SN) \, d\theta}{f(x \mid \hat{\theta}_N, N) \int_{R(x, N)} g(\theta \mid N) \, d\theta}$$

$$= M(x) \cdot \frac{\int_{R(x, SN)} g(\theta \mid SN) \, d\theta}{\int_{R(x, N)} g(\theta \mid N) \, d\theta} \qquad (4.154)$$

This immediately suggests that if the regions $R(x, SN)$ and $R(x, N)$ are of comparable size under either hypothesis, and the a priori densities are constant and equal, the ratio of integrals will be unity. Under such conditions we would expect the "generalized likelihood

ratio" $M(x)$ and the total likelihood ratio $\ell(x|g)$ to be nearly equal.

In other words, when the a priori density is uniform or diffuse, and when the conditional observation densities are very sharp with respect to θ , and when the regions of concentration $R(x, N)$ and $R(x, SN)$ are of about equal size under both hypotheses, then the "generalized likelihood ratio" should be approximately equal to the total likelihood ratio (with respect to this diffuse a priori density).

4.5 Conclusions

Both classical tests and Bayes tests base decisions on sufficient statistics. However, the decision regions are generally of a dissimilar nature and hence receiver operation and performance are not equivalent; but whenever Bayesian analysis is based on diffuse a priori densities, the Bayes test can be identified precisely with a classical test, at least for the examples considered.

Since many of the classical tests can be obtained as a degenerate form of the Bayes test, the Bayes test is judged more flexible and representative of the actual operating state. Whereas it is a straightforward procedure to incorporate available knowledge (or ignorance) into the model by use of Bayes techniques, it is often difficult and sometimes impossible to construct a classical test that does. To a large extent the classical tests ignore available information concerning the true state of the unknown parameter(s), although

one redeeming feature of the classical approach is the emphasis on

unbiased tests, an aspect not considered from the Bayesian standpoint.

CHAPTER V

RECEIVER DESIGN VIA NUMERICAL

INTEGRATION TECHNIQUES

5.1 Motivation

It is often difficult to determine the likelihood ratio for a composite hypothesis situation. When the analytical form of the likelihood ratio eludes the designer, the electronic circuitry that comprises the receiver itself surely cannot be built. In this section an alternative approach is suggested which circumvents these difficulties and conceives of the receiver as a digital computer.

5.2 Formulation

When the conditional observation statistics are of exponential form, it is often possible to average with respect to the given a priori density $g(\theta)$ by the use of an appropriate transformation and a finite order Gauss-Hermite or Gauss-Laguerre integration routine (Ref. 29). Performance must not be affected adversely by the error introduced in this manner.

The normal detection situation falls into this class. The nth order multivariate normal density is

$$f(\underline{x}|\underline{\mu},\Psi) = (2\pi|\Psi|)^{-\frac{n}{2}} e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^t \Psi^{-1}(\underline{x}-\underline{\mu})} \tag{5.1}$$

where $\underline{\mu}$ is the mean vector and Ψ is the autocovariance matrix, i.e.,

$$\underline{\mu} = E(\underline{x}) \tag{5.2}$$

$$\Psi = E[(\underline{x} - \underline{\mu})(\underline{x} - \underline{\mu})^t] \tag{5.3}$$

In particular consider the situation where both $\underline{\mu}$ and Ψ are known except for a multiplicative constant, i.e.,

$$\underline{\mu} = a\,\underline{s} \tag{5.4}$$

$$\Psi = \sigma^2\,\Sigma \tag{5.5}$$

where a and σ^2 are unknown parameters.

If we define

$$\theta = \frac{1}{\sigma^2} \tag{5.6}$$

then

$$f(\underline{x}\,|\,a, \theta) = \left(\frac{\theta}{2\pi\,|\,\Sigma\,|}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2}(\underline{x} - a\,\underline{s})^t\,\Sigma^{-1}(\underline{x} - a\,\underline{s})} \tag{5.7}$$

Furthermore, if we let

$$d_0 = \underline{s}^t\,\Sigma^{-1}\,\underline{s} \qquad \text{(nominal signal-to-noise ratio)} \tag{5.8}$$

$$\hat{a} = \frac{\underline{s}^t\,\Sigma^{-1}\,\underline{x}}{d_0} \qquad \begin{array}{l}\text{(minimum variance, unbiased} \\ \text{estimator of } a\,)\end{array} \tag{5.9}$$

$$W = \Sigma^{-1} - \frac{\Sigma^{-1} \underline{s} \underline{s}^t \Sigma^{-1}}{d_0} \tag{5.10}$$

then the conditional observation density can be written in the form

$$f(\underline{x} \mid a, \theta) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta d_0}{2} (a - \hat{a})^2} e^{-\frac{\theta}{2} \underline{x}^t W \underline{x}} \tag{5.11}$$

Gauss-Hermite integration

Consider first the detection situation SKEA (signal known except for amplitude) with known noise parameter θ and unknown signal parameter a. Let the admissible values of the unknown amplitude a range over R where $R \equiv (-\infty, \infty)$; and let g(a) denote the given a priori density of a. The observation statistics conditional to SN are

$$f(\underline{x} \mid SN) = \int_{-\infty}^{\infty} f(\underline{x} \mid a, SN) \, g(a) \, da$$

$$\tag{5.12}$$

$$= \int_{-\infty}^{\infty} \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x}^t W \underline{x}} e^{-\frac{\theta d_0}{2} (a - \hat{a})^2} g(a) \, da$$

If we make the change of variable

$$u = \sqrt{\theta d_0 / 2} \, (a - \hat{a}) \tag{5.13}$$

then

$$f(\underline{x} \mid SN) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x}^t W \underline{x}} (\theta \, d_0/2)^{-\frac{1}{2}} \, .$$

$$\int_{-\infty}^{\infty} e^{-u^2} g\left(\hat{a} + \frac{u}{\sqrt{\theta \, d_0/2}}\right) du \qquad (5.14)$$

Let $f^*(\underline{x} \mid SN)$ be an approximation to $f(\underline{x} \mid SN)$ obtained by using a pth order Gauss-Hermite integration routine to evaluate the integral in Eq. 5.14. This gives

$$f^*(\underline{x} \mid SN) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x}^t W \underline{x}} (\theta \, d_0/2)^{-\frac{1}{2}} \sum_{i=1}^{p} w_i \, g(a_i) \qquad (5.15)$$

where

$$a_i = \hat{a} + \frac{u_i}{\sqrt{\theta \, d_0/2}} \qquad (5.16)$$

and

w_i, u_i are pth order Gauss-Hermite integration constants.

Gauss-Hermite integration is a natural choice since $f(\underline{x} \mid a, SN)$ is of exponential form and $a \in R$. The error depends upon the magnitude of the (2p)th derivative of $g(a)$ over $a \in R$. If $g(a)$ were a polynomial of order less than $2p$, the error would be zero; one might conjecture that the error would be small provided $g(a)$ behaves like a polynomial in the vicinity of $\hat{a}$ ($\hat{a}$ is a minimum variance unbiased

estimator of the unknown amplitude a .) The degree of error will be demonstrated to be minimal for a particular problem in succeeding sections.

Gauss-Laguerre integration

As a second example consider the detection situation SKE-UNL (signal known exactly, unknown noise level) with known signal parameter a and unknown noise parameter θ . Let the unknown parameter θ have a given a priori density $g(\theta)$ for $\theta \in [0, \infty)$. The observation statistics conditional to SN are

$$f(\underline{x} \mid SN) = \int_0^\infty f(\underline{x} \mid \theta, SN)\, g(\theta)\, d\theta$$

$$(5.17)$$

$$= \int_0^\infty \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2}(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s})}\, g(\theta)\, d\theta$$

If we make the change of variable

$$v = \frac{\theta}{2}(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s}) \qquad (5.18)$$

then

$$f(\underline{x} \mid SN) = \frac{2}{(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s})} \int_0^\infty \left[\frac{2v}{2\pi(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s})}\right]^{\frac{n}{2}}$$

$$e^{-v}\, g\left[\frac{2v}{(\underline{x} - a\underline{s})^t \Sigma^{-1}(\underline{x} - a\underline{s})}\right]\, dv \qquad (5.19)$$

Presume that $g(\theta)$ is of such a nature that analytical integration is difficult or impossible. An approximation to $f(\underline{x} \mid SN)$ can be obtained by employing a pth order Gauss-Laguerre integration routine, namely

$$f^*(\underline{x} \mid SN) = \frac{2(2\pi)^{-\frac{n}{2}}}{(\underline{x} - a\,\underline{s})^t \, \Sigma^{-1} \, (\underline{x} - a\,\underline{s})} \sum_{i=1}^{p} r_i \, \theta_i^{\frac{n}{2}} \, g(\theta_i) \qquad (5.20)$$

where

$$\theta_i = \frac{2\,v_i}{(\underline{x} - a\,\underline{s})^t \, \Sigma^{-1} (\underline{x} - a\,\underline{s})} \qquad (5.21)$$

and

r_i, v_i are pth order Gauss-Laguerre integration constants.

For this detection situation a Gauss-Laguerre integration routine is the natural choice since $f(\underline{x} \mid \theta, SN)$ is of exponential form and $\theta \geq 0$. The error is proportional to the magnitude of the (2p)th derivative of $g(\theta)$ over $\theta \in [0, \infty)$. If $g(\theta)$ varies slowly over its region of definition, i.e., if $g(\theta)$ is a diffuse a priori density, the error will be held within tolerable limits.

5.3 Signal Known Except for Amplitude

Apply the procedure outlined in the previous section to the detection situation SKEA. Without loss of generality this study can be restricted to the one-dimensional case. (The integral of Eq. 5.14 requires knowledge of the n-dimensional observation vector $\underline{x}$ only

in the form of the scalar random variable $\hat{a}$.) Hence, consider the one-dimensional case with conditional observation statistics

$$f(x \mid a, SN) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{1}{2\sigma^2}(x-a)^2} \qquad -\infty < x < \infty \qquad (5.22)$$

$$f(x \mid a, N) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{x^2}{2\sigma^2}} \qquad -\infty < x < \infty \qquad (5.23)$$

and a priori density

$$g(a) = \frac{1}{\sqrt{2\pi v}}\ e^{-\frac{1}{2v}(a-m)^2} \qquad -\infty < a < \infty \qquad (5.24)$$

5.3.1 <u>Receiver Design</u>. It is possible to determine the observation statistics and the total likelihood ratio of the observation for the above example since $g(a)$ is amenable to analytical integration. The availability of the exact form of the observation statistics and the total likelihood ratio makes it possible to evaluate the usefulness of numerical integration techniques for receiver design in composite hypothesis situations.

<u>Exact approach</u>

The exact observation statistics are

$$f(x \mid SN) = \int_{-\infty}^{\infty} f(x \mid a, SN) \, g(a) \, da$$

$$= \frac{1}{\sqrt{2\pi(v+\sigma^2)}} \, e^{-\frac{1}{2(v+\sigma^2)}(x-m)^2} \tag{5.25}$$

$$f(x \mid N) = f(x \mid a, N) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{x^2}{2\sigma^2}} \tag{5.26}$$

and the likelihood ratio is

$$\ell(x) = \sqrt{\frac{1}{\rho+1}} \, e^{\frac{1}{2v}\left[\frac{\rho^2}{\rho+1}(x+\frac{m}{\rho})^2 - m^2\right]} \tag{5.27}$$

where

$$\rho = \frac{v}{\sigma^2} \tag{5.28}$$

Approximate approach

If analytical integration had been impossible, a numerical integration routine could have been employed to obtain an approximate form of the observation density $f(x \mid SN)$. In particular,

$$f(x \mid SN) = \int_{-\infty}^{\infty} f(x \mid a, SN)\, g(a)\, da$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2\sigma^2}(x-a)^2}\, g(a)\, da \qquad (5.29)$$

If we make the change of variable

$$u = \frac{1}{\sqrt{2}\,\sigma}\,(a-x) \qquad (5.30)$$

then

$$f(x \mid SN) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}}\, e^{-u^2}\, g(x + \sqrt{2}\,\sigma\, u)\, du \qquad (5.31)$$

An approximation to $f(x \mid SN)$ via Gauss-Hermite integration techniques is

$$f^*(x \mid SN) = \frac{1}{\sqrt{\pi}} \sum_{i=1}^{p} w_i\, g(x + \sqrt{2}\,\sigma\, u_i)$$

$$= \frac{1}{\pi\sqrt{2v}} \sum_{i=1}^{p} w_i\, e^{-\frac{1}{2v}(x - m + \sqrt{2}\,\sigma u_i)^2} \qquad (5.32)$$

where

w_i, u_i are pth order Gauss-Hermite integration constants.

For completeness define

$$f^*(x \mid N) = f(x \mid N) = \frac{1}{\sqrt{2\pi}\ \sigma}\ e^{-\frac{x^2}{2\sigma^2}} \qquad (5.33)$$

$$\ell^*(x) = \frac{f^*(x \mid SN)}{f^*(x \mid N)} \qquad (5.34)$$

5.3.2 Receiver Realization. If $\ell(x)$ were unavailable because of integration difficulties, an alternative approach would be to base decisions on $\ell^*(x)$. The receiver would then consist essentially of a digital computer with two channels as depicted in Fig. 5.1. The one channel computes $f^*(x \mid SN)$ while the other computes $f^*(x \mid N)$. The outputs of the two channels are combined to form the natural logarithm of $\ell^*(x)$, namely $z^*(x)$; and a decision as to presence or absence of signal is made by thresholding $z^*(x)$.

5.3.3 Quality Index. One would like some assurance that the quality of decisions would not suffer drastically when decisions are based on $\ell^*(x)$ instead of $\ell(x)$, or equivalently when decisions are based on $z^*(x)$ instead of $z(x)$ where

$$z^*(x) = \ell n\ \ell^*(x) \qquad (5.35)$$

$$z(x) = \ell n\ \ell(x) \qquad (5.36)$$

Development

An indication of the "closeness" of $z^*(x)$ to $z(x)$ can be obtained via the index
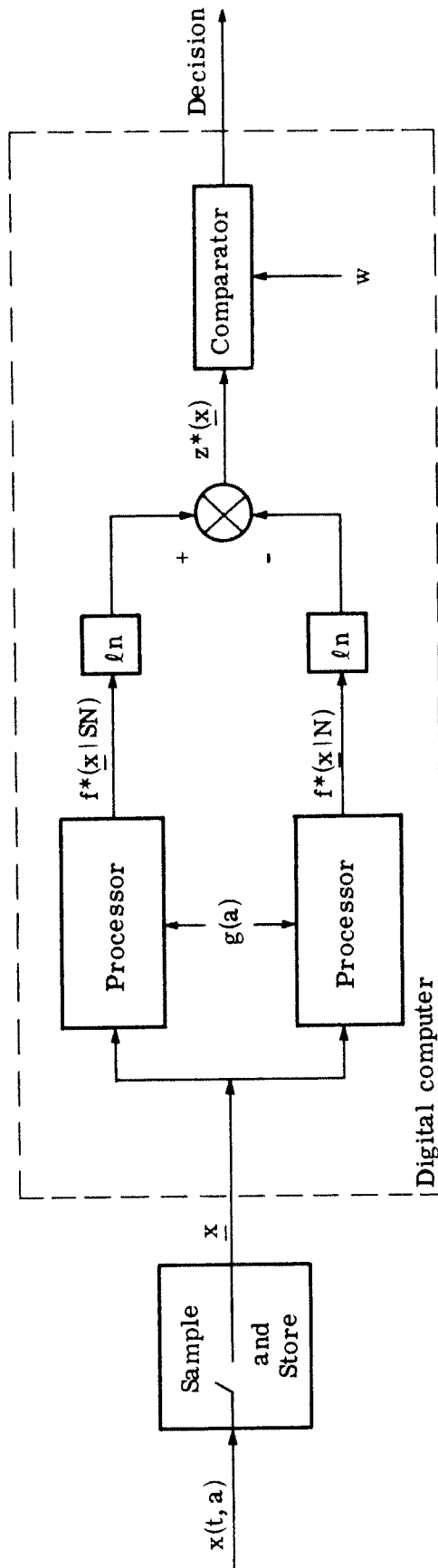
Figure 5.1. Receiver realization as a digital computer incorporating numerical integration techniques.

$$J \overset{\Delta}{=} E[\epsilon(x) \mid SN]$$

$$= \int_X \epsilon(x) \, f(x \mid SN) \, dx \qquad (5.37)$$

where

$$\epsilon(x) = z^*(x) - z(x) \qquad (5.38)$$

This index will be referred to as the quality index and is based on information theoretic concepts analogous to those employed to obtain the sensitivity index developed in Chapter III. It is "obvious" that there is little or no degradation in performance whenever the quality index J is close to zero.

Application

For the detection problem SKEA as developed in the preceding sections

$$\epsilon(x) = z^*(x) - z(x)$$

$$= \ell n \, \frac{\ell^*(x)}{\ell(x)}$$

$$= \ell n \, \frac{f^*(x \mid SN)}{f(x \mid SN)} \qquad (5.39)$$

Evaluation of the quality index yields

$$J = \int_{-\infty}^{\infty} \epsilon(x) \, f(x \mid SN) \, dx$$

$$= \int_{-\infty}^{\infty} \epsilon(x) \, \frac{1}{\sqrt{2\pi(v+\sigma^2)}} \, e^{-\frac{1}{2(v+\sigma^2)}(x-m)^2} \, dx \qquad (5.40)$$

If we make the transformation

$$y = \frac{x-m}{\sqrt{v+\sigma^2}} \qquad (5.41)$$

then

$$f(y \mid SN) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{y^2}{2}} \qquad (5.42)$$

and

$$J = E[\delta(y) \mid SN] = \int_{-\infty}^{\infty} \delta(y) \, f(y \mid SN) \, dy \qquad (5.43)$$

where

$$\delta(y) = \epsilon\left(\sqrt{v+\sigma^2} \, y + m\right) \qquad (5.44)$$

Evaluating Eq. 5.44 via Eqs. 5.25, 5.28, 5.32 and 5.39 and simplifying yields

$$\delta(y) = \ln\left\{ \sqrt{\frac{1+\rho}{\pi\rho}} \, \sum_{i=1}^{p} w_i \, e^{\frac{1}{\rho}[y^2 + 2\sqrt{2(1+\rho)} \, u_i y + 2u_i^2]} \right\} \qquad (5.45)$$

Since $\delta(y)$ depends on the parameters v, σ^2 and m only through ρ, the quality index J depends only on ρ.

An inspection of $\delta(y)$ will be helpful in understanding the behavior of J; knowledge about the nature of $\delta(y)$ will also assist in keeping computational errors to a minimum. Figures 5.2 and 5.3 present plots of $\delta(y)$ for typical parameter values of ρ and p. It can readily be shown by considering the nature of the Gauss-Hermite integration constants u_i that $\delta(y)$ is an even function of y, i.e.,

$$\delta(y) = \delta(-y) \tag{5.46}$$

Also, $\delta(y) \rightarrow \infty$ as $|y| \rightarrow \infty$ although this doesn't affect the existence of J since $f(y|SN)$ is of exponential form. In addition, $\delta(y)$ oscillates about the axis; the number of zero crossings is $2p$. Furthermore, $\delta(y)$ stays closer and closer to the axis for a longer and longer period as p increases for fixed values of ρ or for ρ increasing for fixed values of p.

Results

Numerical integration techniques were employed to evaluate the quality index J for specific values of p and ρ. The results are presented in Fig. 5.4. For fixed values of ρ the quality index J approaches zero as p increases; this indicates that $z^*(x)$ is a better and better approximation to $z(x)$ as p increases. For fixed values of p the quality index J approaches zero as ρ increases;
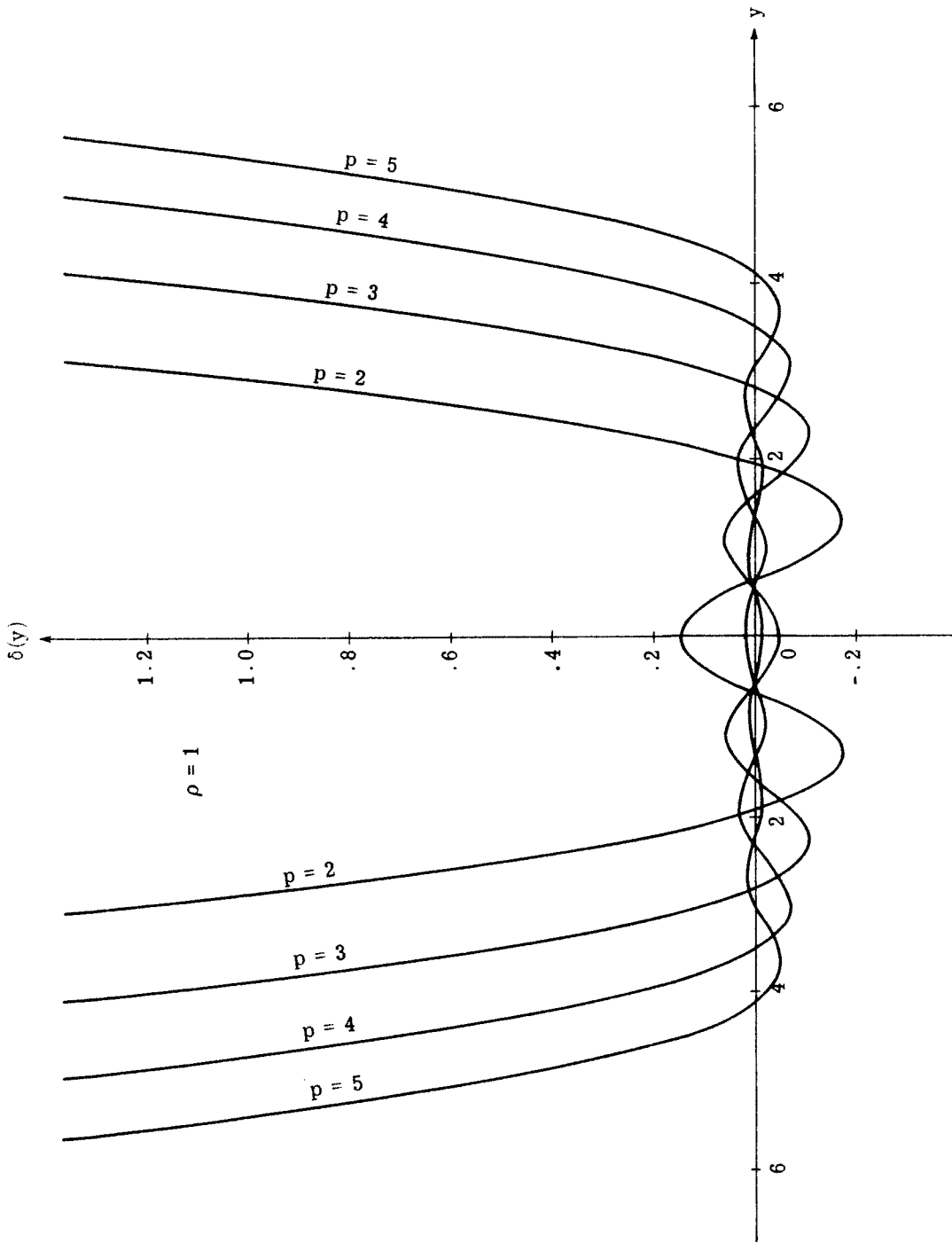
Figure 5.2. Plot of $\delta(y)$, $\rho = 1$.
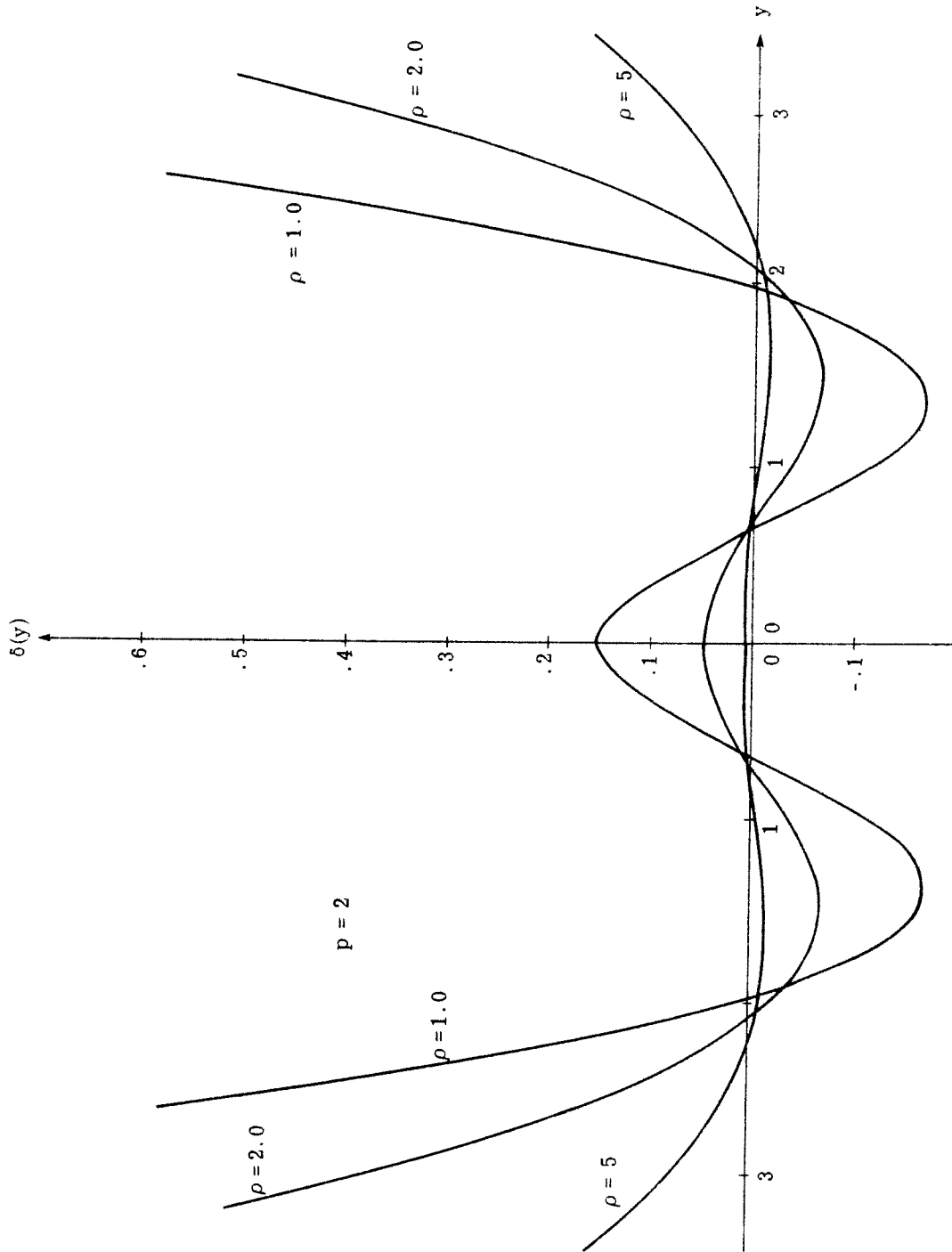
Figure 5.3. Plot of $\delta(y)$, $p = 2$.
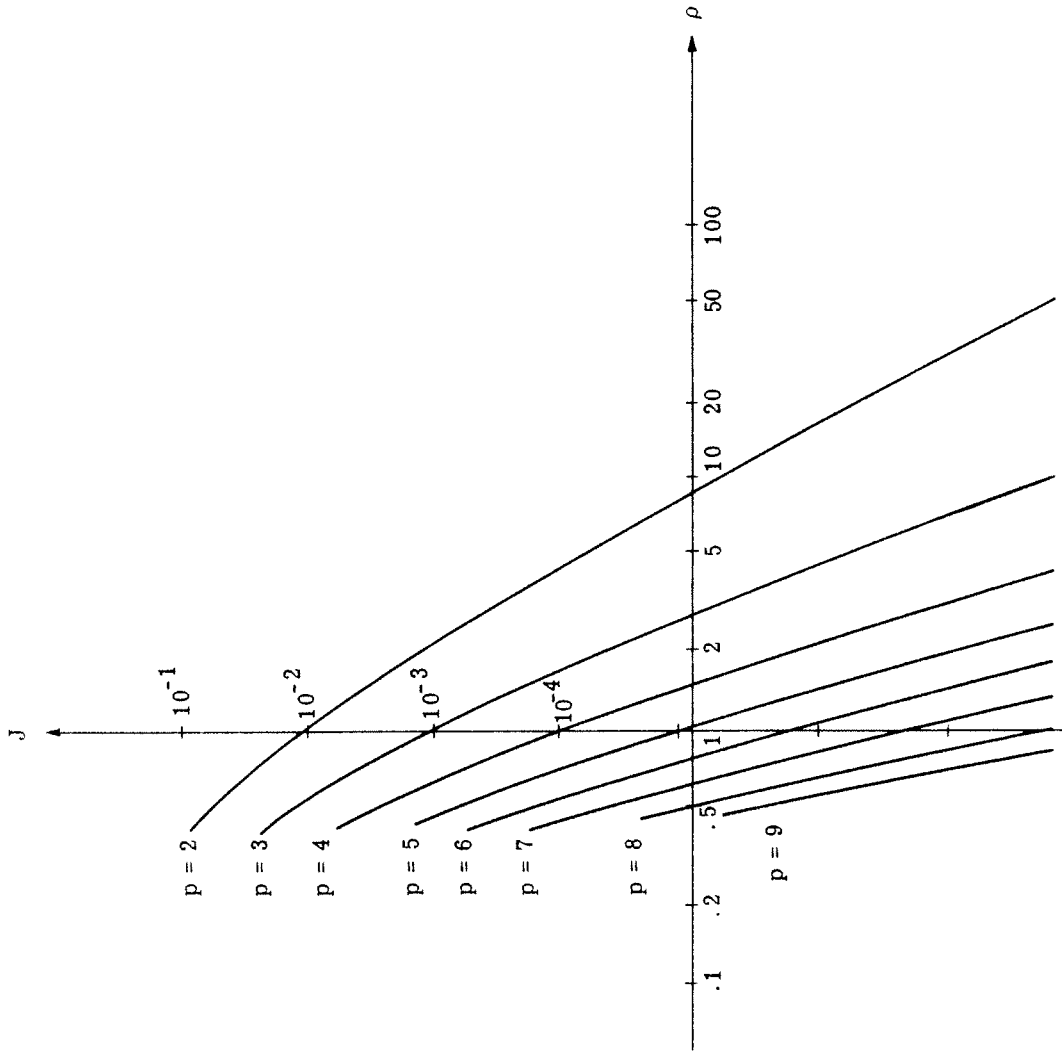
Figure 5.4. Quality index.

this shows that <u>as the a priori density g(a) becomes more and more</u>

<u>diffuse $z^*(x)$ becomes a better and better approximation to $z(x)$</u>.

The fact that the best results are obtained for large values of ρ should not be construed to mean that the entire procedure is of little significance. On the contrary, diffuse a priori densities are used to develop receiver design whenever the designer is totally ignorant of the true state of the unknown parameter, or the physical situation dictates that the unknown parameter may range over a wide latitude of equally permissible or probable values.

The validity of the results described above can be ascertained by examining the variance of $\delta(y)$, namely

$$V_J = \mathrm{Var}\left[\delta(y) \mid SN\right]$$

$$= \int_{-\infty}^{\infty} \left[\delta(y) - J\right]^2 f(y \mid SN)\, dx \qquad (5.47)$$

This quantity is plotted in Fig. 5.5 for the same parameter values as in Fig. 5.4. It shows the same gross characteristics as Fig. 5.4 and should be interpreted as follows: <u>the degree of confidence one places</u> <u>in the quality index J should increase as both p and ρ increase</u>. In other words the validity of the above-mentioned results should not be questioned for ρ and/or p large.

5.3.4 Receiver Performance.
The recommended procedure for comparing receiver performance considers first the performance of the optimum likelihood ratio receiver [decisions are based on $\ell(x)$]

CHAPTER VI

PSEUDO-ESTIMATION

6.1 Introduction

In this section an attempt is made to justify the use of estimation techniques for single composite hypothesis receiver design situations by constructing an "estimator" which yields optimum performance in the Bayesian sense. It is not an estimator in the true sense of the word but nevertheless exhibits features and characteristics that are considered desirable of an estimator. The motivation for this work came from a recent paper published by Kailath (Ref. 33) which shows that the likelihood ratio for the detection of a random signal in additive white Gaussian noise has the same form as that for a known signal in white Gaussian noise. However, the correlation integral has to be interpreted in the special sense of an Itô stochastic integral. Nevertheless, it suggests the use of an estimator-correlator as an engineering approximation to the optimum receiver.

6.2 Formulation

A word of explanation is in order. Because this chapter will be intimately concerned with a sequential processing of the observation, the notation used in this chapter will be different but self-explanatory. Rather than use a succinct notation which often leads to confusion, dependence of the observation statistics on the preceding observation

154

will be explicitly indicated.

Development of the basic theory

Assume the observation statistics are conditionally independent, and let θ denote a signal parameter with a continuous a priori density $g(\theta)$. The conditional observation statistics of the total observation are

$$f(x_1, x_2, \cdots x_n \mid \theta, SN) = \prod_{i=1}^{n} f(x_i \mid \theta, SN) \qquad (6.1)$$

$$f(x_1, x_2, \cdots x_n \mid \theta, N) = \prod_{i=1}^{n} f(x_i \mid N) \qquad (6.2)$$

and the conditional likelihood ratio of the total observation is

$$\ell(x_1, x_2, \cdots x_n \mid \theta) = \frac{f(x_1, x_2, \cdots x_n \mid \theta, SN)}{f(x_1, x_2, \cdots x_n \mid N)} = \prod_{i=1}^{n} \ell(x_i \mid \theta) \qquad (6.3)$$

where

$$\ell(x_i \mid \theta) = \frac{f(x_i \mid \theta, SN)}{f(x_i \mid N)} \qquad i = 1, 2, \cdots n \qquad (6.4)$$

is the conditional likelihood ratio of a single observation. By employing Bayes law we can write the absolute observation statistics of the total observation conditional to SN as

$$f(x_1, x_2, \cdots x_n | SN) = \frac{f(x_1, x_2, \cdots x_n | \theta, SN)\, g(\theta)}{g(\theta | x_1, x_2, \cdots x_n)} \qquad (6.5)$$

and the likelihood ratio of the total observation is

$$\ell(x_1, x_2, \cdots x_n) = \frac{f(x_1, x_2, \cdots x_n | SN)}{f(x_1, x_2, \cdots x_n | N)}$$

$$(6.6)$$

$$= \frac{g(\theta)}{g(\theta | x_1, x_2, \cdots x_n)} \prod_{i=1}^{n} \ell(x_i | \theta)$$

(The a posteriori densities depend implicitly on the hypothesis SN since θ is a signal parameter; this dependence is not expressed since it only serves to complicate the notation.) Although it appears that the RHS of Eq. 6.6 depends on θ, it is in fact independent of θ since the LHS of Eq. 6.6 is independent of θ. Let's inspect the RHS of Eq. 6.6 in more detail. Since

$$\frac{g(\theta)}{g(\theta | x_1, x_2, \cdots x_n)} = \frac{g(\theta)}{g(\theta | x_1)} \cdot \frac{g(\theta | x_1)}{g(\theta | x_1, x_2)} \cdots \frac{g(\theta | x_1, x_2, \cdots x_{n-1})}{g(\theta | x_1, x_2, \cdots x_n)}$$

$$= \prod_{i=1}^{n} \frac{g(\theta | x_1, x_2, \cdots x_{i-1})}{g(\theta | x_1, x_2, \cdots x_i)} \qquad (6.7)$$

then

$$\ell(x_1, x_2, \cdots x_n) = \prod_{i=1}^{n} \frac{g(\theta | x_1, x_2, \cdots x_{i-1})}{g(\theta | x_1, x_2, \cdots x_i)} \cdot \ell(x_i | \theta) \quad (6.8)$$

If we define

$$z(x_1, x_2, \ldots x_n) \overset{\Delta}{=} \ln \ell(x_1, x_2, \ldots x_n) \tag{6.9}$$

then

$$z(x_1, x_2, \ldots x_n) = \sum_{i=1}^{n} z(x_i | \theta) + \sum_{i=1}^{n} \ln \frac{g(\theta | x_1, x_2, \ldots x_{i-1})}{g(\theta | x_1, x_2, \ldots x_i)} \tag{6.10}$$

where

$$z(x_i | \theta) = \ln \ell(x_i | \theta) \tag{6.11}$$

If it is possible to determine θ_i such that

$$\frac{g(\theta_i | x_1, x_2, \ldots x_{i-1})}{g(\theta_i | x_1, x_2, \ldots x_i)} = 1 \qquad i = 1, 2, \ldots n \tag{6.12}$$

then

$$z(x_1, x_2, \ldots x_n) = \sum_{i=1}^{n} z(x_i | \theta_i) \tag{6.13}$$

Therefore the optimum (Bayesian sense) test statistic $z(x_1, x_2, \ldots x_n)$ can be obtained from the conditionally optimum test statistics $z(x_i | \theta_i)$, $i = 1, 2, \ldots n$ with θ_i determined in accordance with Eq. 6.12.

The simplicity of this approach is enticing. Merely determine a set of values θ_i satisfying Eq. 6.12 and use these values in the conditionally optimum receiver as though they were exact. For the

detection problem SKEA in added white Gaussian noise, the optimum

receiver becomes an "estimator-correlator," a correlator in the

sense that the optimum receiver would be a correlator if the signal

amplitude were known exactly, an estimator in the sense that "esti-

mators" of signal amplitude at each increment in time are employed

in the receiver as though they were exact.

Observations

Consider the solution(s) of Eq. 6.12 in reference to Fig. 6.1.

There may be multiple solutions to Eq. 6.12 for each i; there is at

least one solution since the a posteriori densities are continuous by

virtue of the assumption that the a priori density was continuous.

Each value of θ_i depends upon the first i observations $(x_1, x_2,$

$\cdots x_i)$ and might conceivably be interpreted as a posterior estimator

of θ at time t_i.

two possible values of θ_i
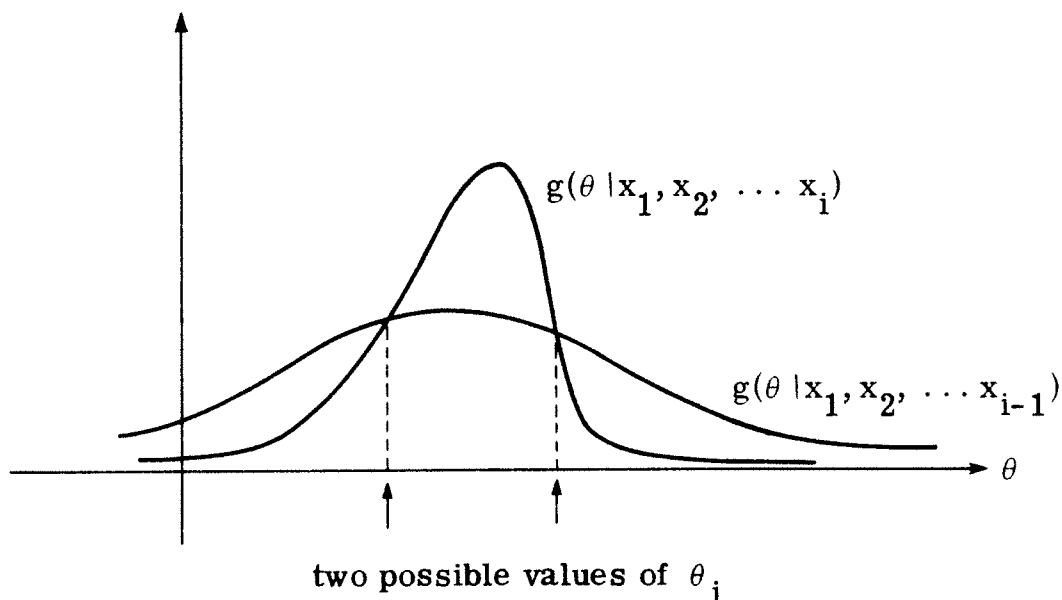
Figure 6.1. Sketch of a posteriori probability density functions.

Multiple solutions to Eq. 6.12 could cause some confusion in regard to the interpretation of θ_i as an estimator of θ, but possibly a particular solution θ_i^* exists which behaves more like a true estimator than the other solutions. To explore this possibility define the posterior mean

$$\mu_i = E(\theta \mid x_1, x_2, \ldots x_i) \tag{6.14}$$

and the posterior variance

$$\nu_i = \text{Var}(\theta \mid x_1, x_2, \ldots x_i) \tag{6.15}$$

It is common knowledge that the posterior mean μ_i is the MMSE (minimum mean square error) estimator of θ. The posterior mean μ_i is an unbiased estimator of θ if $E(\mu_i \mid \theta_0) = \theta_0$ where θ_0 is the true value of θ.

Conjecture

Conditional to the hypothesis signal and noise if the a posteriori densities are unimodal and if $\nu_i < \nu_{i-1}$, then it is conjectured that a particular solution θ_i^* exists (the pseudo-estimator or "estimator") which converges to the MMSE estimator μ_i, i.e., $\theta_i^* \to \mu_i$, or more precisely, given $\delta > 0$ there exists i_0 such that $|\theta_i^* - \mu_i| < \delta$ for $i > i_0$. Furthermore, $\mu_i \to \theta_0$, provided $P(|\theta - \theta_0| < \epsilon \mid \underline{x}_i) > 0$ for arbitrary $\epsilon > 0$, regardless of whether or not μ_i is biased, and hence $\theta_i \to \theta_0$; if μ_i is biased, then $\theta_i^* \to \mu_i$ more rapidly than $\mu_i \to \theta_0$.

Explanation

    A heuristic explanation is offered in lieu of a rigorous proof.
If $\nu_i < \nu_{i-1}$, the a posteriori densities become "sharper" as more
observations are taken into account; the mode and the mean become
almost indistinguishable for "sharp" densities and become identical
in the limit since $\nu_i \to 0$. In addition, for each i the solutions θ_i
either tend to be clustered about μ_i (usually around the mode) or
else one solution is close to μ_i (usually between posterior modes)
while the others tend to be on the "tails". Hence, as soon as the
posterior variance reaches a fairly low level, i.e., after a sufficient
number of observations, there is always one reasonable "estimator"
of θ near the posterior mean. If we denote the solution closest to
the posterior mean by θ_i^*, then it seems plausible that $\theta_i^* \to \mu_i$.
Furthermore, provided $P(|\theta - \theta_0| < \epsilon \mid \underline{x}_i) > 0$ for arbitrary $\epsilon > 0$,
$\mu_i \to \theta_0$ regardless of whether or not μ_i is biased since $\nu_i \to 0$ and
the a posteriori densities are unimodal by assumption. If μ_i were
biased, it appears reasonable that $\theta_i^* \to \mu_i$ more rapidly than $\mu_i \to$
θ_0; it may require many observations for the effect of the bias to
"wear off," especially if the a priori density was of a diffuse nature.
In either case both μ_i and θ_i^* converge to θ_0.

    The convergence of the pseudo-estimator θ_i^* to the MMSE
estimator μ_i will be demonstrated in the following example.

6.3 Signal Known Except for Amplitude in Added White Gaussian Noise

Consider the detection situation SKEA in added white Gaussian noise and attempt a comparison between the optimal Bayesian approach and estimation. The sequential nature of the problem makes it convenient to use the notation

$$\underline{x}_i = (x_1, x_2, \ldots x_i)^t \tag{6.16}$$

$$\underline{s}_i = (s_1, s_2, \ldots s_i)^t \qquad i = 1, 2, \ldots n \tag{6.17}$$

Let I_n denote the nth order identity matrix.

The hypothesis test for SKEA in added white Gaussian noise consists of

$$N: \qquad \underline{x}_n \sim MVN_n(a \, \underline{s}_n, \, \sigma^2 \, I_n)$$

$$SN: \qquad \underline{x}_n \sim MVN_n(0, \, \sigma^2 \, I_n)$$

where the unknown amplitude is denoted by a.

A priori density. Let the a priori density on a be given by the truncated normal density

$$g(a \,|\, \underline{x}_0, SN) = \frac{1}{\Phi\left(\dfrac{m}{\sqrt{v}}\right) \sqrt{2\pi v}} \, e^{-\frac{1}{2v}(a-m)^2} \tag{6.18}$$

$$0 \leq a < \infty$$

To facilitate subsequent algebraic manipulation, it is convenient to express the a priori density in the equivalent form

$$g(a \,|\, \underline{x}_0, SN) = K \, e^{Aa^2 + Ba} \qquad 0 \le a < \infty \qquad (6.19)$$

where

$$A = -\frac{1}{2v} \qquad (6.20)$$

$$B = \frac{m}{v} \qquad (6.21)$$

$$K = \frac{1}{\sqrt{v} \; \omega \left(\dfrac{m}{\sqrt{v}}\right)} \qquad (6.22)$$

and

$$\omega(x) = \frac{\Phi(x)}{\Phi'(x)} \qquad (6.23)$$

The parameters m and v can be expressed in terms of A and B as

$$m = -\frac{B}{2A} \qquad (6.24)$$

$$v = -\frac{1}{2A} \qquad (6.25)$$

Conditional log-likelihood ratio.

It has been shown previously that the conditional log-likelihood ratio of the total observation for this situation is

$$z(\underline{x}_n | a) = \left(\frac{1}{\sigma^2} \underline{x}_n^t \underline{s}_n \right) a - \frac{a^2}{2} \left(\frac{1}{\sigma^2} \underline{s}_n^t \underline{s}_n \right)$$

$$= a\, y_n - \frac{a^2}{2} d_n \tag{6.26}$$

where

$$y_n = \frac{1}{\sigma^2} \underline{x}_n^t \underline{s}_n = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i s_i \tag{6.27}$$

$$d_n = \frac{1}{\sigma^2} \underline{s}_n^t \underline{s}_n = \frac{1}{\sigma^2} \sum_{i=1}^{n} s_i^2 \tag{6.28}$$

The sequential realization of the conditional log-likelihood ratio is

$$z(\underline{x}_n | a) = \sum_{i=1}^{n} z(x_i | a) \tag{6.29}$$

where

$$z(x_i | a) = a \left(\frac{x_i s_i}{\sigma^2} \right) - \frac{a^2}{2} \left(\frac{s_i^2}{\sigma^2} \right) \tag{6.30}$$

A posteriori density

The a posteriori density is proportional to the product of the conditional likelihood ratio and the a priori density, i.e.,

$$g(a | \underline{x}_i, SN) = C \, \ell(\underline{x}_i | a)\, g(a) = C' \, e^{a y_i - \frac{a^2 d_i}{2}} \, e^{A a^2 + B a} \tag{6.31}$$

$$0 \leq a < \infty$$

Since $g(a \mid \underline{x}_i, SN)$ is a proper density, we normalize and obtain

$$g(a \mid \underline{x}_i, SN) = K_i \, e^{A_i a^2 + B_i a} \qquad\qquad 0 \leq a < \infty \qquad (6.32)$$

where

$$A_i = A - \frac{d_i}{2} = -\frac{1}{2} (v^{-1} + d_i) \qquad\qquad (6.33)$$

$$B_i = B + y_i = m \, v^{-1} + y_i \qquad\qquad (6.34)$$

$$K_i = \frac{1}{\sqrt{v_i} \; \omega\left(\dfrac{m_i}{\sqrt{v_i}}\right)} \qquad\qquad (6.35)$$

and

$$m_i = -\frac{B_i}{2A_i} = \frac{m \, v^{-1} + y_i}{v^{-1} + d_i} \qquad\qquad (6.36)$$

$$v_i = -\frac{1}{2A_i} = \frac{1}{v^{-1} + d_i} \qquad\qquad (6.37)$$

The a posteriori density can perhaps be more easily interpreted in the form

$$g(a \mid \underline{x}_i, SN) = \frac{1}{\Phi\left(\dfrac{m_i}{\sqrt{v_i}}\right) \sqrt{2\pi v_i}} \; e^{-\frac{1}{2v_i} (a - m_i)^2} \qquad 0 \leq a < \infty \qquad (6.38)$$

In this form it is readily apparent that the a posteriori density (Eq. 6.38) is of the same functional form as the a priori density (Eq. 6.18). Hence the a priori density is a reproducing density with respect to the conditional observation statistics.

MMSE estimator

The MMSE estimator of a after i observations is the posterior mean

$$\bar{a}_i = E(a \mid \underline{x}_i, SN)$$

$$= \int a \, g(a \mid \underline{x}_i, SN) \, da$$

$$= m_i \left[1 + \cfrac{1}{\cfrac{m_i}{\sqrt{v_i}} \, \omega\left(\cfrac{m_i}{\sqrt{v_i}}\right)} \right] \qquad (6.39)$$

Pseudo-estimator

The posterior variance ν_i is difficult to compute; but it can be shown that the posterior variance ν_i converges asymptotically to v_i, i.e., $\nu_i \to v_i$. In addition, it can readily be shown that $v_i < v_{i-1}$; and hence it is plausible that $\nu_i < \nu_{i-1}$. Furthermore, the a posteriori densities (Eq. 6.38) are unimodal for each i [mode = max $(0, m_i)$]. Since $P(a < 0 \mid \underline{x}_i) = 0$, asymptotic convergence of either the MMSE estimator or the pseudo-estimator to the true amplitude a will occur only if $0 \leq a < \infty$.

To obtain the set of pseudo-estimators for this detection

situation we determine that

$$\frac{g(a\,|\,\underline{x}_i,\,SN)}{g(a\,|\,\underline{x}_{i-1},\,SN)} = 1 \qquad 0 \leq a < \infty \qquad (6.40)$$

implies that

$$K_i\, e^{A_i a^2 + B_i a} = K_{i-1}\, e^{A_{i-1} a^2 + B_{i-1} a} \qquad 0 \leq a < \infty \qquad (6.41)$$

or equivalently,

$$a^2(A_i - A_{i-1}) + a(B_i - B_{i-1}) + \ell n\, \frac{K_i}{K_{i-1}} = 0 \qquad 0 \leq a < \infty \qquad (6.42)$$

Since

$$A_i - A_{i-1} = -\frac{s_i^{\,2}}{\sigma^2} \qquad (6.43)$$

and

$$B_i - B_{i-1} = \frac{x_i\, s_i}{\sigma^2} \qquad (6.44)$$

Equation 6.42 becomes

$$a^2 - 2a\left(\frac{x_i}{s_i}\right) + c_i = 0 \qquad 0 \leq a < \infty \qquad (6.45)$$

where

$$c_i = \frac{2\sigma^2}{s_i^2} \ell n \left(\frac{K_{i-1}}{K_i} \right) \qquad (6.46)$$

The pseudo-estimator(s) at time t_i is the solution(s) of Eq. 6.45. If we let

$$\alpha_i = \frac{1}{s_i} \left(x_i \pm \sqrt{x_i^2 + 2\sigma^2 \ell n \frac{K_i}{K_{i-1}}} \right) \qquad i = 1, 2, \ldots n \quad (6.47)$$

and denote the pseudo-estimator(s) by a_i, we get

$$a_i = \alpha_i \qquad 0 \leq \alpha_i < \infty, \quad i = 1, 2, \ldots n \qquad (6.48)$$

Since the a posteriori densities are continuous, there is at least one valid pseudo-estimator.

6.4 Receiver Operation

A simulation of receiver operation was conducted conditional to some actual amplitude value a. The signal waveform was chosen to be a dc signal with an energy content of unity over a total observation length of 20 time intervals. In order to observe the convergence of the pseudo-estimator described by the solution of Eq. 6.40 to some asymptotic value, the simulated run included calculation of the posterior mean (MMSE estimator) at each increment in time. These values along with the admissible pseudo-estimators are presented in the figures at each time increment as bars for improved legibility.

In addition the detection output $z(\underline{x}_{20})$ is shown on each figure to indicate the receiver's opinion as to presence or absence of a signal at the end of the observation interval. Naturally a high value is indicative of signal presence whereas a low value tends to indicate absence of a signal.

The simulated runs are summarized in Figs. 6.2 - 6.7 for three radically different information levels. Figures 6.2 and 6.3 summarize the receiver operation for a diffuse a priori density while Figs. 6.4 and 6.5 summarize receiver operation for precise knowledge. Figures 6.6 and 6.7 present a simulation of receiver operation for an intermediate information level.

Diffuse information level

Receiver operation of Figs. 6.2 and 6.3 is parameterized by $m = 2$ and $v = 100$. This represents an a priori density that is rather diffuse. In other words the receiver designer did not have a great deal of information available concerning the true value of the unknown amplitude. His ignorance necessarily forced the receiver to weigh the actual observations more heavily than the prior opinion of the designer and consequently the simulated run and detection output are heavily influenced by the data. This becomes readily apparent by inspection of the figures. Figures 6.2 and 6.3 both show the disparity between the pseudo-estimator and the posterior mean at the beginning of the observation interval although convergence of

Figure 6.2. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, diffuse information level, $a = 0$.

Figure 6.3. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, diffuse information level, $a = 4$.

the pseudo-estimator to the posterior mean has pretty well taken place at the end of the observation interval independent of the true amplitude. Note however that the posterior mean has not yet converged to the true amplitude. In both Figs. 6.2 and 6.3 a general inclination of the posterior mean to converge to the true amplitude already appears, but in both cases convergence is not yet apparent. This tends to indicate that the pseudo-estimator is in fact closely related to the posterior mean long before the posterior mean has converged to the true amplitude. This observation lends some credence to the use of estimation techniques as nearly optimal procedures for composite hypothesis detection and/or estimation problems.

Precise information level

When the receiver designer has some very precise opinions and/or knowledge concerning the true amplitude, the actual physical situation had better bear him out or else his job may be in jeopardy. Precise opinions are reflected in the a priori density in terms of small variances, and heavily influence the operation of the receiver. Figures 6.4 and 6.5 show the dominant effect of precise prior opinions and the insignificant effect of the actual data on the simulated run. The receiver is firmly convinced that it knows the actual amplitude and acts accordingly regardless of the data. Hence, precise opinions must be corroborated by the actual situation in order to yield optimal behavior.
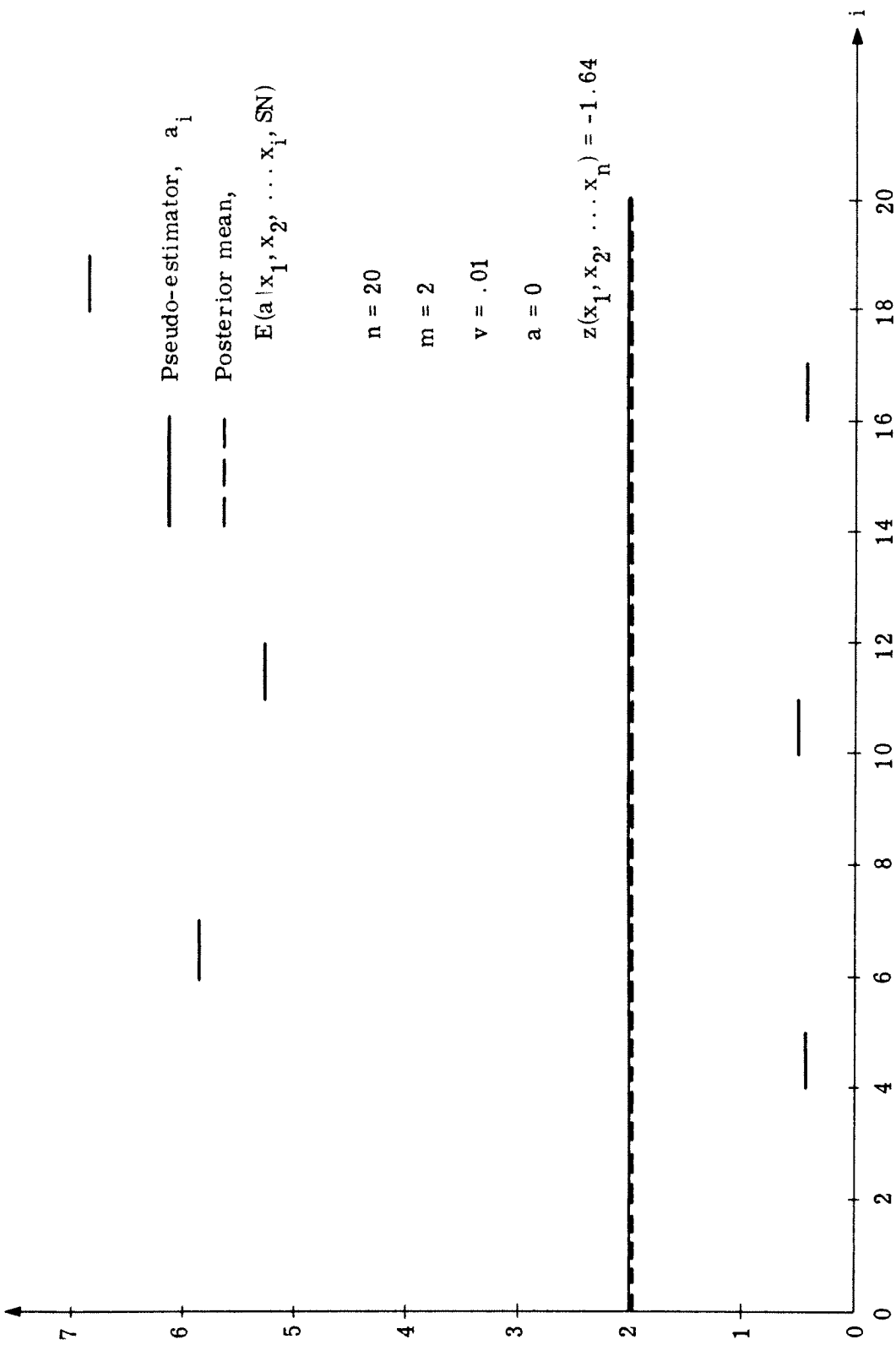
Figure 6.4. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, precise information level, $a = 0$.
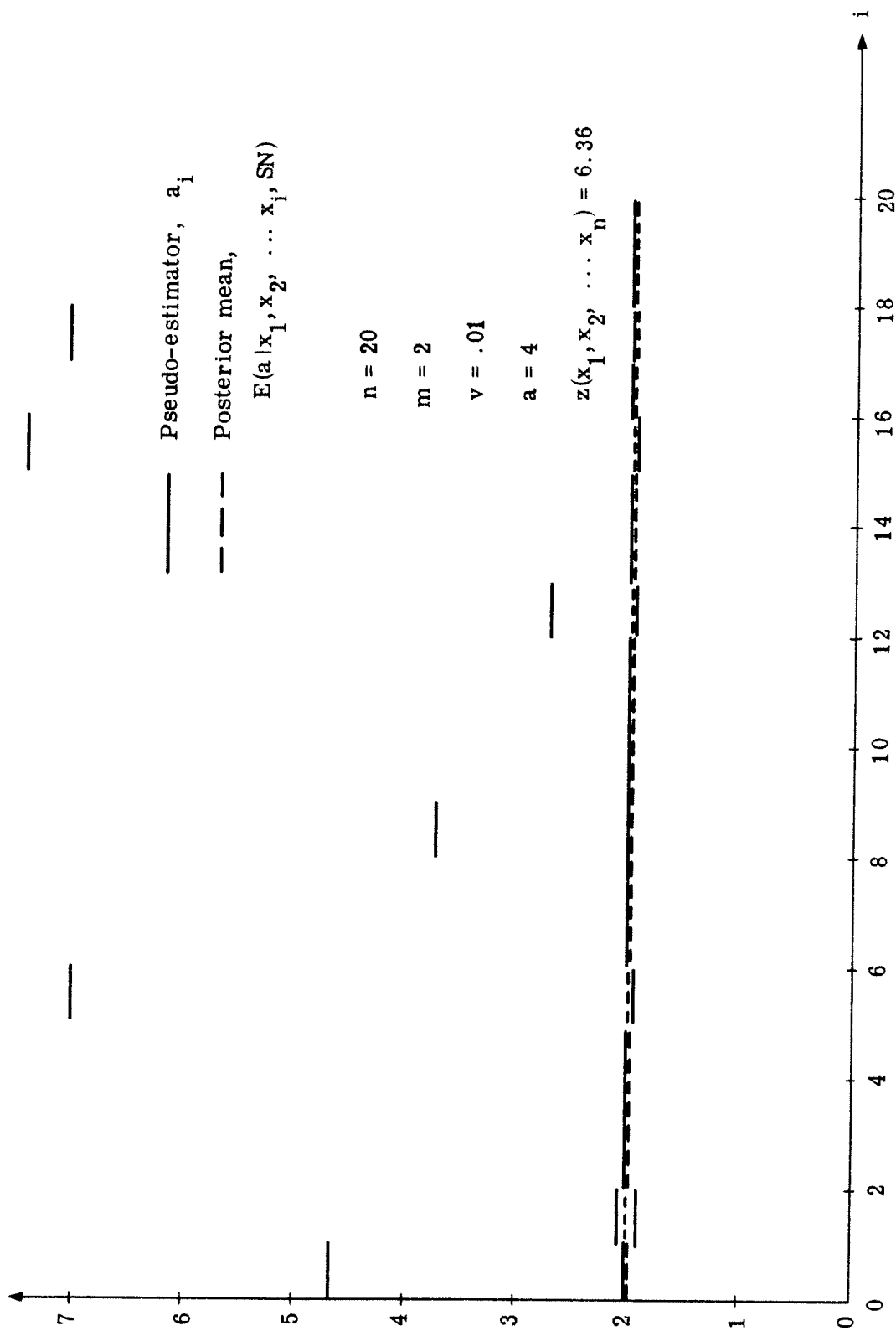
Figure 6.5. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, precise information level, a = 4.

Intermediate information level

For the intermediate level of knowledge the results of Figs. 6.6 and 6.7 show a "middle of the road" policy. Receiver operation is influenced by both prior knowledge and data. The effect of prior knowledge is most noticeable at the beginning of the observation interval while the effect of the actual data outweighs a priori knowledge at the end of the observation interval. Hence the receiver has obtained a delicate balance between the two extremes of knowledge, and yet its action can be closely allied to the action of a receiver which is designed on the basis of estimation techniques.
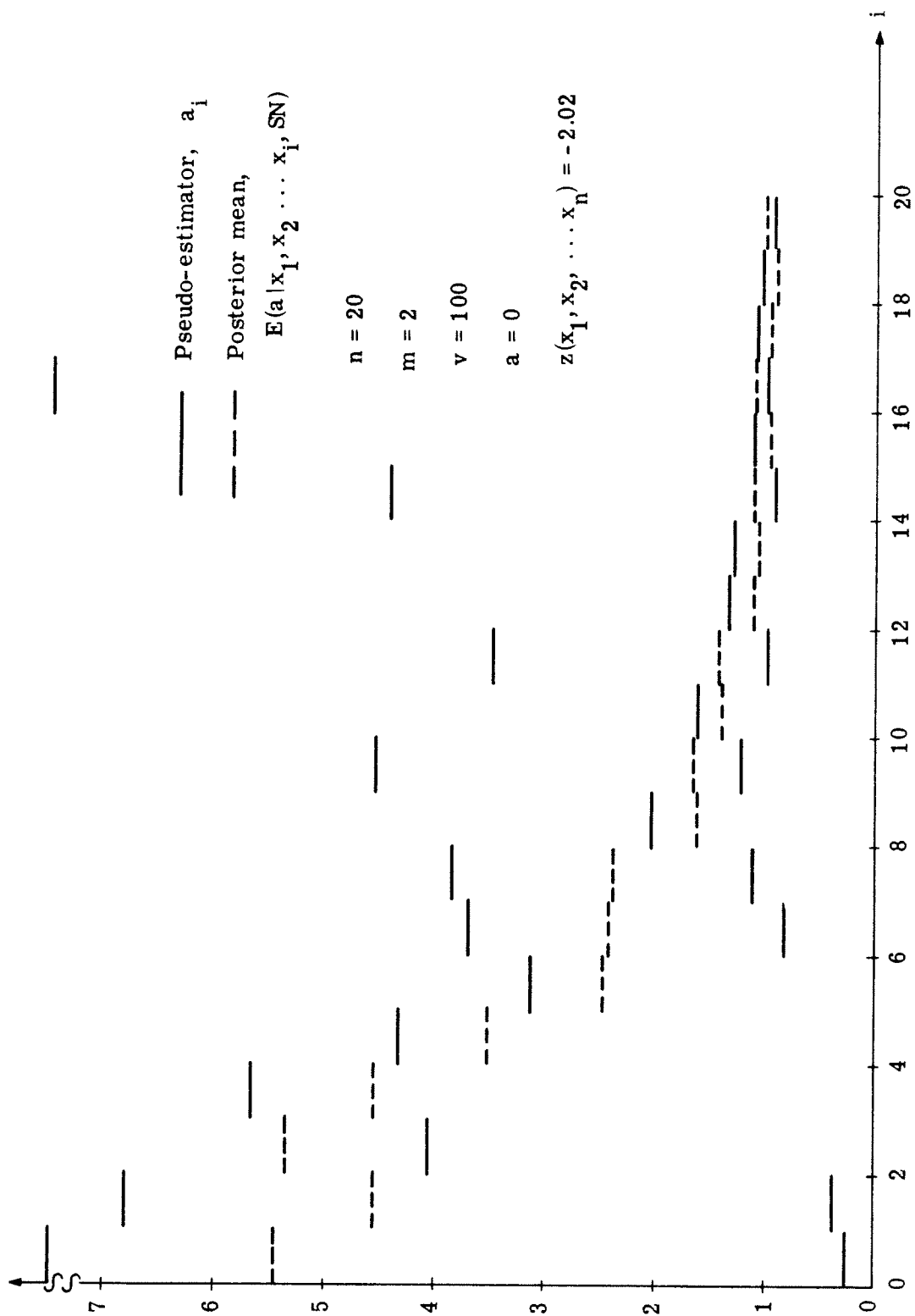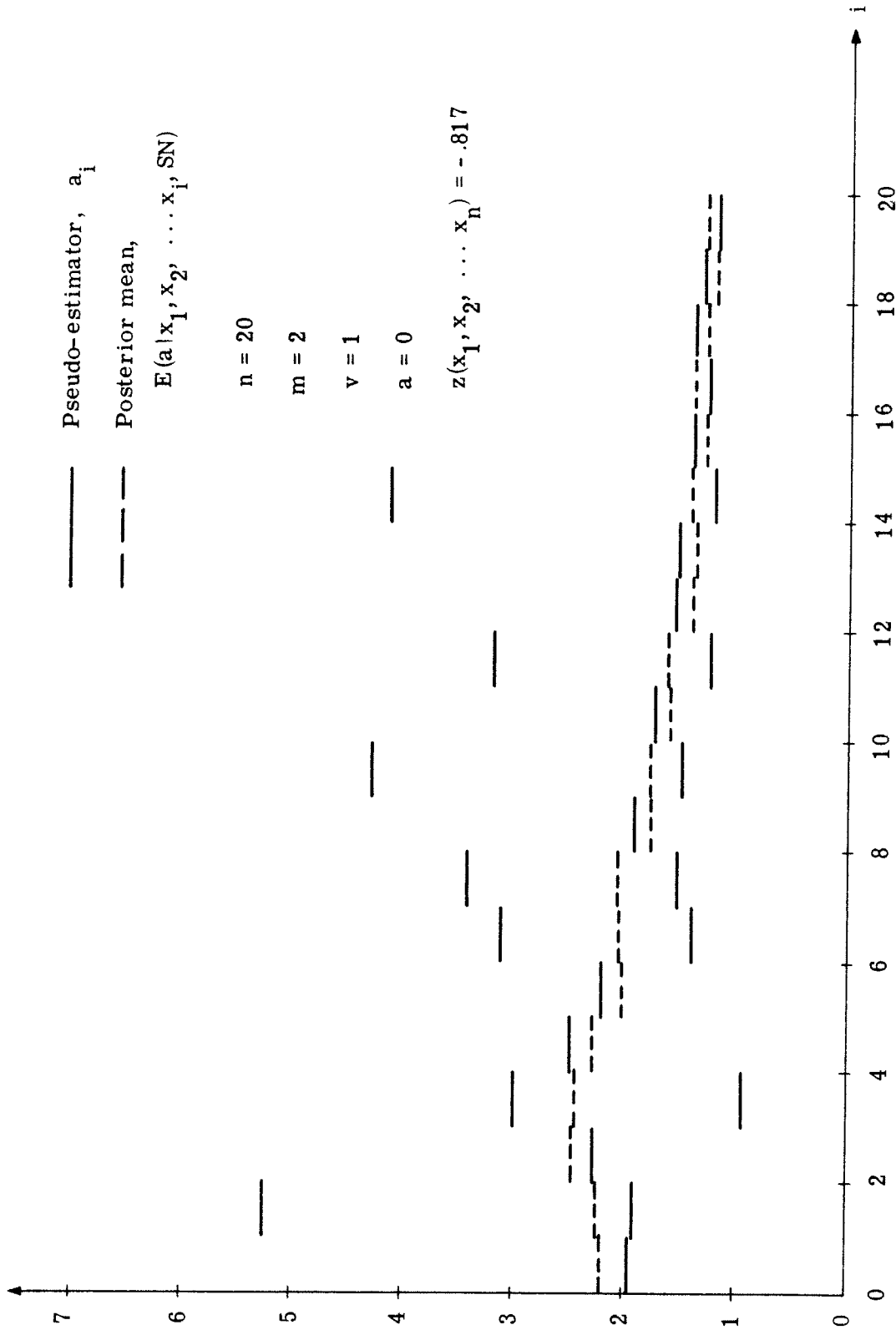
Figure 6.6. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, intermediate information level, a = 0.
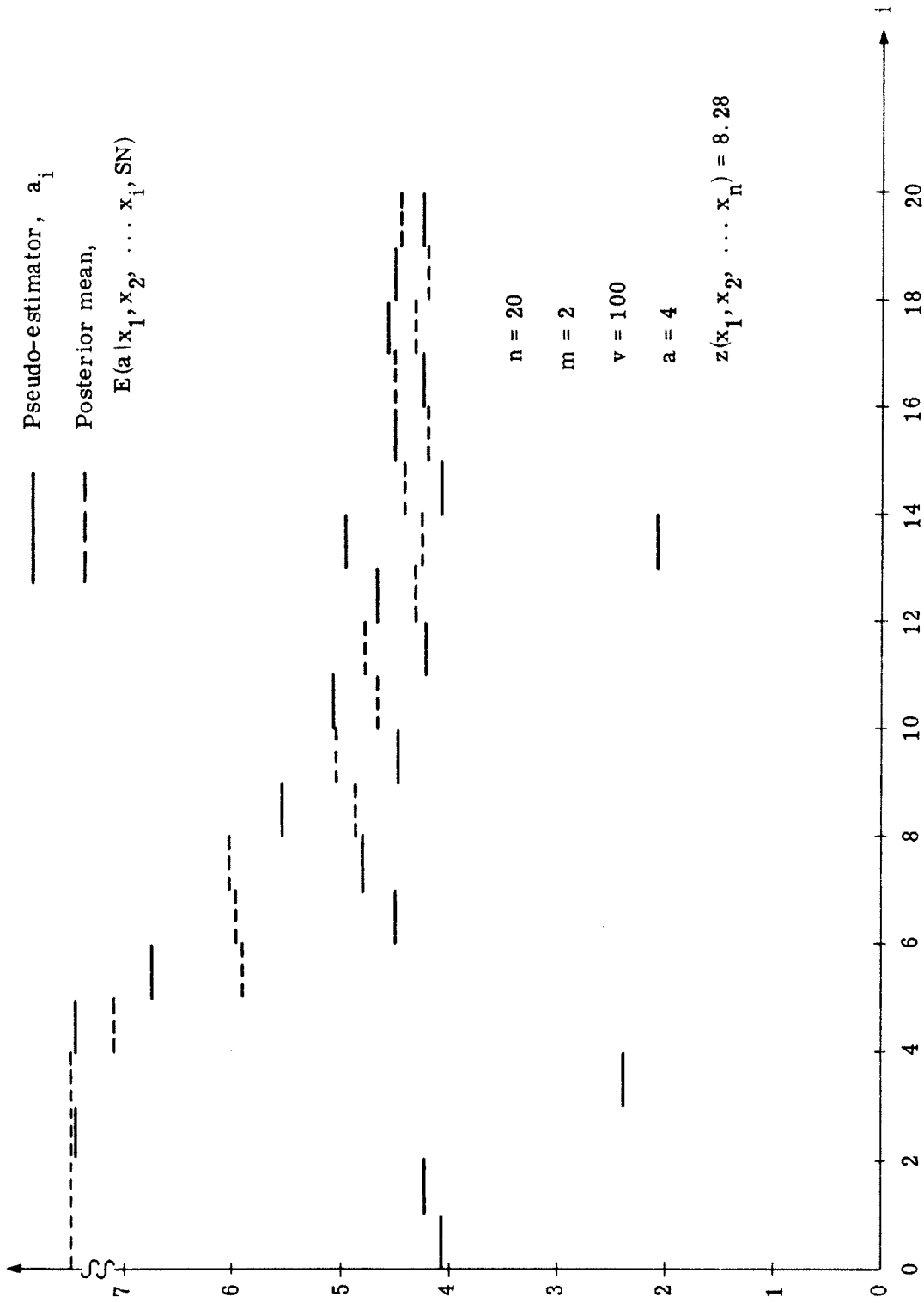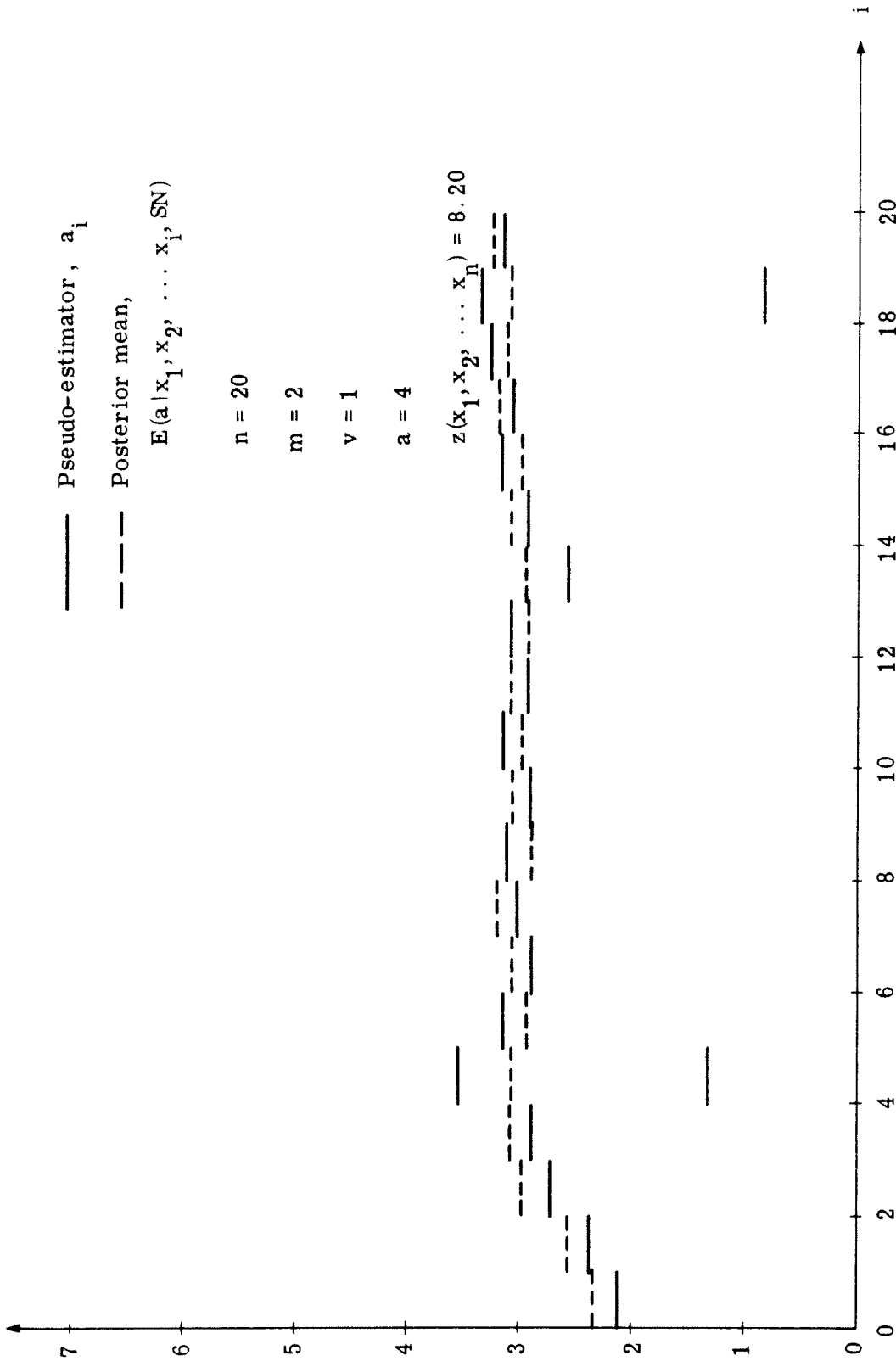
Figure 6.7. Simulation of receiver operation, comparison of the pseudo-estimator and the posterior mean, SKEA, intermediate information level, $a = 4$.

CHAPTER VII

SUMMARY

7.1 Conclusions

One of the major goals of this work was to examine the effect of the particular choice of a priori density on receiver performance for the composite hypothesis detection problem. This study was conducted by constructing a sensitivity index which measures the performance loss that occurs when the receiver is based on the a priori density $g(\cdot)$ but in fact $h(\cdot)$ holds. The sensitivity index was shown to possess several desirable features.

The exact form of the a priori density is never known precisely since the a priori density reflects the subjective viewpoint of the designer. There is a lot of leeway in choosing both the functional form and the parameters of the a priori density. The sensitivity index was employed to show that performance depends primarily on the first few moments of the a priori density. Functional form was important only to the extent that the entire range of parameter definition was modeled. The design of a receiver which accurately reflects and incorporates the information at hand can be accomplished by choosing an a priori density whose mean corresponds to the designer's confidence level in his choice of the mean.

Comparison of receiver performance for the composite hypothesis situation was done in Chapter IV. The Bayesian philosophy was contrasted with the classical approach for several examples. The Bayesian approach to detection problems is more versatile than the classical approach. The Bayesian approach inherently incorporates all prior information into the model by the assignment of an a priori density. By using a degenerate form of the a priori density (a diffuse a priori density), the Bayesian approach leads to the identical receiver obtained by at least one of the methods of classical statistics for all the examples considered. In particular it appears as though the UMPU test of classical statistics (if it exists) is equivalent to the Bayes test corresponding to a diffuse a priori density for the single composite hypothesis. No such relationship exists for the double composite hypothesis.

Emphasis upon unbiasedness by classical statistics is certainly desirable; Bayesian analysis can and often does result in a test which is not unbiased. The nature of the Bayes model should, however, preclude the notion that Bayes tests are "bad" since biased tests seldom (with small probability) occur.

The concept of an ESP receiver was useful in comparing Bayes tests. The ESP ROC served as a convenient upper bound ROC to the optimum ROC with respect to the assigned a priori density and thereby served to indicate how much useful information was conveyed by the a priori density. The ESP ROC was shown to be the average of

conditionally optimum ROC's whereas the optimum ROC was shown to be the average of conditionally suboptimum ROC's. Hence the ESP and the optimum ROC can be viewed as "capsule" summaries of the performance of the conditionally optimum and suboptimum receivers respectively.

Receiver design via numerical integration techniques was proposed in Chapter V for composite hypothesis situations whenever exact analysis was difficult or impossible. It was shown that by using a numerical method which was "matched" to the detection situation, a receiver design could be achieved that was almost optimal (in the Bayesian sense) and yet was readily implemented. The receiver is conceived of as an analog-to-digital converter in conjunction with a digital computer. The availability of high speed electronic components makes it possible to operate such a device in real time.

In Chapter VI a study was conducted which attempted to justify estimation as a technique which is almost optimal in the Bayesian sense. The impetus for this study was provided by a recent paper published by Kailath (Ref. 33). For the single composite hypothesis a pseudo-estimator is constructed which is optimal in the Bayesian sense and compared to the minimum mean square error estimator. Simulation of receiver operation for a particular example shows a close correlation between estimation techniques and exact Bayesian analysis and lends credence to the use of estimate-and-plug techniques as "good" techniques.

7.2 Contributions

A sensitivity index was developed to measure the performance loss that occurs when a receiver is designed to be optimum with respect to the given a priori density $g(\cdot)$, but in fact the a priori density $h(\cdot)$ is considered to hold; the important features of the sensitivity index were discussed. The sensitivity index was used to show that the functional form of the a priori density chosen to model the state of the unknown parameter of a composite hypothesis detection problem has little effect on receiver performance. Of prime importance are the first few moments of the a priori density, especially the mean and variance. The designer should choose the mean to be that value which, in his opinion, most accurately represents the state of the parameter; and the variance should be chosen to reflect the designer's confidence level (how sure is he?) in his choice of the mean. Hence, the functional form of the a priori density can be chosen to be mathematically tractable.

Bayesian analysis and classical statistics were compared and reconciled. It was shown that Bayesian analysis was more versatile and that one or more of the classical tests could often be obtained by a Bayes test based on a diffuse a priori density. The Bayesian approach is judged to be superior to the classical approach primarily because the Bayesian approach incorporates all available information into the model via a priori densities; classical procedures ignore such information.

The externally sensed parameter (ESP) receiver was reviewed and its ROC was evaluated for several examples never considered previously. Its function as an upper bound ROC and its relationship to the conditionally optimum ROC's was pointed out.

The feasibility of receiver design via numerical integration techniques was shown. Conceived of as a digital computer, the receiver is nearly optimal in performance with little increase in complexity. The numerical procedure is especially adept (little or no error) at handling a priori densities which are of a diffuse nature. Hence, receiver design for composite hypothesis situations has been demonstrated to be feasible for many problems previously considered too complex to solve.

Estimation techniques for single composite hypothesis decision problems have been shown to correspond closely to optimal Bayesian analysis by the construction of a pseudo-estimator. Conditional to the hypothesis signal and noise this pseudo-estimator was shown to exhibit many features considered desirable of an estimator and to asymptotically approach the MMSE (minimum mean square error) estimator. The close correlation between the pseudo-estimator and the MMSE estimator serves to point out that receiver design via estimation techniques can surely be considered justifiable in case the optimal (Bayesian) procedure is too complex.

APPENDIX A

NATURE OF THE SET R_0

This appendix will determine the nature of the set R_0 where

$$R_0 = \{r_0 \mid r(\theta) \geq 0, \quad 0 \leq \theta < \infty, \quad r_0 \neq 1\} \qquad (A.1)$$

and

$$r(\theta) = r_0 + (1 - r_0) \sum_{i=1}^{p} d_i (c\theta)^i \qquad (A.2)$$

with d_i independent of c.

Independence of c. If we let

$$\phi = c\theta \qquad (A.3)$$

$$s(\phi) = r\left(\frac{\phi}{c}\right) \qquad (A.4)$$

then

$$R_0 = \{r_0 \mid s(\phi) \geq 0, \quad 0 \leq \phi < \infty, \quad r_0 \neq 1\} \qquad (A.5)$$

Since $s(\phi)$ is independent of c, R_0 is also independent of c.

Range of permissible values. Without loss of generality, let $c = 1$ so that

182

$$r(\theta) = r_0 + (1 - r_0) \sum_{i=1}^{p} d_i \theta^i \qquad \text{(A.6)}$$

Manipulating Eq. A.6 we get

$$r(\theta) = (1 - r_0) [d_0 + r_0(\theta)] \qquad \text{(A.7)}$$

where

$$d_0 = \frac{r_0}{1 - r_0} \qquad \text{(A.8)}$$

$$r_0(\theta) = \sum_{i=1}^{p} d_i \theta^i \qquad \text{(A.9)}$$

Then $r(\theta) \geq 0$ for $\theta \epsilon [0, \infty)$ implies

$$d_0 + r_0(\theta) \geq 0 \quad \text{and} \quad 0 \leq r_0 \leq 1 \quad \text{if} \quad d_p > 0$$

or

$$d_0 + r_0(\theta) \geq 0 \quad \text{and} \quad r_0 \geq 1 \quad \text{if} \quad d_p < 0 \qquad \text{(A.10)}$$

The extreme permissible value of r_0 can be determined by choosing d_0 to be an extremum. This is easily accomplished by choosing

$$d_0^* = \begin{cases} - \min\limits_{\theta \epsilon [0, \infty)} r_0(\theta) & d_p > 0 \\ \\ - \max\limits_{\theta \epsilon [0, \infty)} r_0(\theta) & d_p < 0 \end{cases} \qquad \text{(A.11)}$$

Then the extreme permissible value r_0^* is

$$r_0^* = \frac{d_0^*}{d_0^* + 1} \qquad (A.12)$$

and

$$R_0 = \{r_0 | r_0^* \leq r_0 < 1\} \qquad d_p > 0$$

$$(A.13)$$

$$R_0 = \{r_0 | 1 < r_0 \leq r_0^*\} \qquad d_p < 0$$

APPENDIX B

EXTREME VALUES OF r_0

In this appendix we will calculate the extreme value $r_0{}^*$ for $p = 2$ and $p = 3$.

It was shown in Appendix A that

$$r_0{}^* = \frac{d_0{}^*}{d_0{}^* + 1} \qquad \text{(B.1)}$$

where

$$d_0{}^* = \begin{cases} - \min_{\theta \in [0, \infty)} r_0(\theta) & d_p > 0 \\ \\ - \max_{\theta \in [0, \infty)} r_0(\theta) & d_p < 0 \end{cases} \qquad \text{B.2)}$$

$$r_0(\theta) = \sum_{i=1}^{p} d_i \, \theta^i \qquad \text{(B.3)}$$

The coefficients d_i were derived in Appendix J and are

$$\underline{d} = (d_1, d_2, \ldots d_p)^t = \Gamma^{-1} \underline{\gamma} \qquad \text{(B.4)}$$

where

$$\Gamma = (\gamma_{i+j-1}) = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_p \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{p+1} \\ \vdots & \vdots & & \\ \gamma_p & \gamma_{p+1} & \cdots & \gamma_{2p-1} \end{bmatrix} \qquad (B.5)$$

$$\underline{\gamma} = (\gamma_1, \gamma_2, \cdots \gamma_p)^t \qquad (B.6)$$

and

$$\gamma_i = (b+1)(b+2) \cdots (b+i) \qquad (B.7)$$

<u>p = 2.</u> From Eq. B.4

$$d_1 = \frac{2}{b+1} \qquad (B.8)$$

$$d_2 = -\frac{1}{(b+1)(b+2)} \qquad (B.9)$$

Since $d_2 < 0$, choose

$$d_0^* = -\max_{\theta \, \epsilon \, [0, \infty)} r_0(\theta) \qquad (B.10)$$

where

$$r_0(\theta) = d_1 \theta + d_2 \theta^2 \qquad (B.11)$$

The critical values occur at $\theta = 0$ and at those values of θ for which $r_0{}'(\theta) = 0$. But $r_0{}'(\theta) = 0$ implies that $\theta = -\dfrac{d_1}{2d_2}$ with

$$r_0 \left(- \frac{d_1}{2d_2} \right) = - \frac{d_1{}^2}{4d_2} \quad . \quad \text{Hence}$$

$$d_0{}^* = - \max \left(0, - \frac{d_1{}^2}{4d_2} \right)$$

$$= \frac{d_1{}^2}{4d_2} \tag{B.12}$$

and
$$r_0{}^* = \frac{d_0{}^*}{d_0{}^* + 1} = b + 2 \tag{B.13}$$

$\underline{p = 3.}$ From Eq. B.4

$$d_1 = \frac{3}{b + 1} \tag{B.14}$$

$$d_2 = - \frac{3}{(b+1)(b+2)} \tag{B.15}$$

$$d_3 = \frac{1}{(b+1)(b+2)(b+3)} \tag{B.16}$$

Since $d_3 > 0$, choose

$$d_0{}^* = - \min_{\theta \epsilon [0, \infty)} r_0(\theta) \tag{B.17}$$

where

$$r_0(\theta) = d_1 \theta + d_2 \theta^2 + d_3 \theta^3 \tag{B.18}$$

$$f(u' \mid \theta) = \left(\frac{\theta E_s}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta E_s u'^2}{2}} \qquad -\infty < u' < \infty \tag{C.4}$$

$$f(v \mid \theta) = \frac{2\left(\frac{\theta E_s}{2}\right)^{\frac{n-1}{2}} v^{n-2} e^{-\frac{\theta E_s v^2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \qquad 0 \leq v < \infty \tag{C.5}$$

where

$$E_s = \sum_{i=1}^{n} s_i^2 \tag{C.6}$$

Proof:

If we let $\underline{y} = A\underline{x}$ define an orthonormal transformation $(A^t A = I)$ with

$$y_1 = \sum_{i=1}^{n} \left(\frac{s_i}{\sqrt{E_s}}\right) x_i \tag{C.7}$$

then

$$f(\underline{y} \mid \theta) = \left(\frac{\theta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \sum_{i=1}^{n} y_i^2} \tag{C.8}$$

If we let

$$z_i = \frac{y_i}{\sqrt{E_s}} \qquad i = 1, 2, \ldots, n \tag{C.9}$$

then

$$f(\underline{z} \mid \theta) = \left(\frac{\theta E_s}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta E_s}{2} \sum_{i=1}^{n} z_i^2} \tag{C.10}$$

Next, let

$$u' = z_1 = \frac{1}{E_s} \sum_{i=1}^{n} x_i s_i \tag{C.11}$$

$$v^2 = \sum_{i=2}^{n} z_i^2 = \frac{1}{E_s} \sum_{i=1}^{n} x_i^2 - \left(\frac{1}{E_s} \sum_{i=1}^{n} x_i s_i\right)^2 \tag{C.12}$$

Since z_1 is independent of $(z_2, z_3, \ldots, z_n)$, this implies that u' is independent of v and furthermore

$$u' \sim N\left(0, \frac{1}{\theta E_s}\right) \tag{C.13}$$

$$\theta E_s v^2 \sim \chi^2_{n-1} \tag{C.14}$$

That is,

$$f(u' \mid \theta) = \left(\frac{\theta E_s}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta E_s u'^2}{2}} \qquad -\infty < u' < \infty \tag{C.15}$$

$$f(v \mid \theta) = \frac{2\left(\frac{\theta E_s}{2}\right)^{\frac{n-1}{2}} v^{n-2} e^{-\frac{\theta E_s v^2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \qquad 0 \leq v < \infty \tag{C.16}$$

Q.E.D.

2) Transformation with respect to absolute densities

This section will determine the effect of the transformation of Eqs. C.1 and C.2 on the absolute densities with respect to the a priori density

$$h(\theta) = r(\theta) g(\theta) \tag{C.17}$$

where

$$g(\theta) = \frac{c^{b+1}}{\Gamma(b+1)} \theta^b e^{-c\theta} \qquad b > -1, \quad c > 0, \quad 0 \le \theta < \infty \tag{C.18}$$

$$r(\theta) = \sum_{i=0}^{p} r_i \theta^i \tag{C.19}$$

From Eq. 3.40 an equivalent expression for the polynomial $r(\theta)$ is

$$r(\theta) = \sum_{i=0}^{p} f_i (c\theta)^i \tag{C.20}$$

with f_i independent of c.

Average the conditional densities derived in part 1) to obtain

$$f(u', v \mid h) = \int_{\Theta} f(u', v \mid \theta) h(\theta) d\theta$$

$$= \int_{\Theta} f(u' \mid \theta) f(v \mid \theta) h(\theta) d\theta$$

$$= K \int_0^\infty \theta^{\frac{1}{2}} e^{-\frac{\theta E_s u'^2}{2}} \theta^{\frac{n-1}{2}} v^{n-2} \, .$$

$$e^{-\frac{\theta E_s v^2}{2}} \sum_{i=0}^p f_i (c\theta)^i \theta^b e^{-c\theta} \, d\theta$$

$$= K_n v^{n-2} \left(u'^2 + v^2 + \frac{2c}{E_s} \right)^{-(b+1+\frac{n}{2})} m(u', v)$$

$$-\infty < u' < \infty, \quad 0 \le v < \infty \qquad (C.21)$$

with

$$K_n = \frac{2(2c/E_s) \, \Gamma\left(b + 1 + \frac{n}{2}\right)}{\sqrt{\pi} \, \Gamma\left(\frac{n-1}{2}\right) \, \Gamma(b+1)} \qquad (C.22)$$

$$m(u', v) = \frac{1}{\Gamma\left(b + 1 + \frac{n}{2}\right)} \sum_{i=0}^p g_i \left(\frac{2c}{E_s}\right)^i \left(u'^2 + v^2 + \frac{2c}{E_s} \right)^{-i} \qquad (C.23)$$

and

$$g_i = f_i \, \Gamma\left(b + 1 + \frac{n}{2} + i\right) \qquad i = 0, 1, \ldots p \qquad (C.24)$$

APPENDIX D

OPTIMUM RECEIVER OPERATION AND PERFORMANCE

FOR SKE + KGN

The operation and performance of the optimum receiver for SKE + KGN (signal known exactly in known Gaussian noise) will be derived in this appendix. From Eqs. 4.12 and 4.13 the conditional observation statistics are

$$f(\underline{x}\,|\,a, SN) = (2\pi|\Sigma|)^{-\frac{n}{2}}\, e^{-\frac{1}{2}\underline{x}^t \Sigma^{-1} \underline{x}} \tag{D.1}$$

$$f(\underline{x}\,|\,a, N) = (2\pi|\Sigma|)^{-\frac{n}{2}}\, e^{-\frac{1}{2}(\underline{x}-a\underline{s})^t \Sigma^{-1}(\underline{x}-a\underline{s})} \tag{D.2}$$

Receiver operation

The conditional likelihood ratio is

$$\ell(\underline{x}\,|\,a) = \frac{f(\underline{x}\,|\,a, SN)}{f(\underline{x}\,|\,a, N)} = e^{a y - a^2 \frac{d_0}{2}} \tag{D.3}$$

where

$$y = \underline{s}^t \Sigma^{-1} \underline{x} \qquad \text{(observation vector)} \tag{D.4}$$

$$d_0 = \underline{s}^t \Sigma^{-1} \underline{s} \qquad \text{(nominal signal-to-noise} \tag{D.5}$$
$$\text{ratio)}$$

The random variable y is a sufficient statistic of the observation $\underline{x}$

194

since $\ell(\underline{x}\,|\,a)$ depends upon $\underline{x}$ only through y. Subsequent discussion of receiver operation and performance will center about the sufficient statistic y. Since $\ell(\underline{x}\,|\,a) > \beta$ implies that

$$y > \frac{\ell n\ \beta}{a} + \frac{ad_0}{2} \qquad a > 0$$

$$\tag{D.6}$$

$$y < \frac{\ell n\ \beta}{a} + \frac{ad_0}{2} \qquad a < 0$$

the optimum receiver operation consists of simply thresholding the sufficient statistic y.

Receiver performance

In order to determine the ROC, it is necessary to determine the statistics of y. Since y is a linear transformation of multivariate normal random variables, y is also a normal random variable. Hence, we need to determine only its mean and variance conditional to both N and SN for a complete description. We obtain

$$E(y\,|\,SN) = E(\underline{s}^t\ \Sigma^{-1}\ \underline{x}) = \underline{s}^t\ \Sigma^{-1}\ E(\underline{x}\,|\,SN) = \underline{s}^t\ \Sigma^{-1}\ a\,\underline{s}$$

$$= ad_0 \tag{D.7}$$

$$Var(y\,|\,SN) = Var(\underline{s}^t\ \Sigma^{-1}\ \underline{x}) = \underline{s}^t\ \Sigma^{-1}\ Var(\underline{x}\,|\,SN)\,(\underline{s}^t\ \Sigma^{-1})^t$$

$$= \underline{s}^t\ \Sigma^{-1}\ \Sigma\ \Sigma^{-1}\ \underline{s}$$

$$= d_0 \tag{D.8}$$

Likewise,

$$E(y \mid N) \quad = 0 \qquad (\text{since } a = 0) \qquad \qquad (D.9)$$

$$Var(y \mid N) = d_0 \qquad \qquad (D.10)$$

Hence

$$f(y \mid a, N) \quad = \frac{1}{\sqrt{2\pi d_0}} \; e^{-\frac{y^2}{2d_0}} \qquad \qquad (D.11)$$

$$f(y \mid a, SN) = \frac{1}{\sqrt{2\pi d_0}} \; e^{-\frac{(y - ad_0)^2}{2d_0}} \qquad \qquad (D.12)$$

The performance equations are

$$P^{CO}(A \mid {}^N_{SN}) \quad = \int\limits_{\ell(\underline{x} \mid a) > \beta} f(y \mid a, {}^N_{SN}) \; dy \qquad \qquad (D.13)$$

Utilizing Eq. D.6 and simplifying yields

$$P^{CO}(A \mid N) \quad = \Phi \left(- \frac{\ln \beta}{\sqrt{d_0} \; |a|} - \frac{\sqrt{d_0} \; |a|}{2} \right) \qquad \qquad (D.14)$$

$$P^{CO}(A \mid SN) \quad = \Phi \left(- \frac{\ln \beta}{\sqrt{d_0} \; |a|} + \frac{\sqrt{d_0} \; |a|}{2} \right) \qquad \qquad (D.15)$$

APPENDIX E

SUBOPTIMUM RECEIVER OPERATION AND

PERFORMANCE FOR SKE + KGN

The operation and performance of a suboptimum receiver for SKE + KGN (signal known exactly in known Gaussian noise) will be derived in this appendix. Receiver operation will be based on the total likelihood ratio for SKEA (signal known except for amplitude) with a normal a priori density on the unknown amplitude. Receiver performance will be evaluated conditional to the actual amplitude a.

Receiver operation

The conditional likelihood ratio was derived in Appendix D and is

$$\ell(\underline{x}\,|\,a) = e^{\,ay - a^2\,\frac{d_0}{2}} \tag{E.1}$$

where

$$y = \underline{s}^t \, \Sigma^{-1} \, \underline{x} \qquad \text{(sufficient statistic of the observation)} \tag{E.2}$$

$$d_0 = \underline{s}^t \, \Sigma^{-1} \, \underline{s} \qquad \text{(nominal signal-to-noise ratio)} \tag{E.3}$$

From Eq. 4.11 the a priori density on the unknown amplitude is

$$g(a) = \frac{1}{\sqrt{2\pi v}} \; e^{-\frac{(a-m)^2}{2v}} \qquad (E.4)$$

Subsequent algebraic manipulation will be simplified if $g(a)$ is expressed in the form

$$g(a) = K \, e^{A a^2 + Ba} \qquad (E.5)$$

where

$$A = -\frac{1}{2v} \qquad (E.6)$$

$$B = \frac{m}{v} \qquad (E.7)$$

$$K = \frac{e^{-\frac{m^2}{2v}}}{\sqrt{2\pi v}} \qquad (E.8)$$

The parameters v and m can be expressed in terms of A and B as

$$v = -\frac{1}{2A} \qquad (E.9)$$

$$m = -\frac{B}{2A} \qquad (E.10)$$

The total likelihood ratio $\ell(\underline{x})$ is the average of the conditional likelihood ratio $\ell(\underline{x}|a)$ with respect to the a priori density $g(a)$, i.e.,

$$\ell(\underline{x}) = \int \ell(\underline{x}|a) \, g(a) \, da$$

$$= \int_{-\infty}^{\infty} e^{\,ay - a^2 \frac{d_0}{2}} \, K \, e^{Aa^2 + Ba} \, da$$

$$= K \int_{-\infty}^{\infty} e^{A'a^2 + B'a} \, da$$

$$= \frac{K}{K'} \qquad (E.11)$$

where

$$A' = A - \frac{d_0}{2} = -\frac{1}{2} \, (v^{-1} + d_0) \qquad (E.12)$$

$$B' = B + y = m \, v^{-1} + y \qquad (E.13)$$

$$K' = \frac{e^{-\frac{m'^2}{2v'}}}{\sqrt{2\pi v'}} \qquad (E.14)$$

and

$$v' = -\frac{1}{2A'} = (v^{-1} + d_0)^{-1} \qquad (E.15)$$

$$m' = -\frac{B'}{2A'} = (v^{-1} + d_0)^{-1} \, (mv^{-1} + y) \qquad (E.16)$$

Substituting,

$$\ell(\underline{x}) = \sqrt{\frac{v'}{v}} \; e^{\frac{m'^2}{2v'} - \frac{m^2}{2v}} \tag{E.17}$$

Receiver operation is determined by the nature of the decision region $\{\underline{x} \mid \ell(\underline{x}) > \beta\}$. It is easy to verify that $\ell(\underline{x}) \uparrow \mid y + mv^{-1} \mid$. Thus the suboptimum receiver operation consists of thresholding $\mid y + mv^{-1} \mid$. In particular, the decision rule is

$$\phi(\underline{x}) = \begin{cases} 1 & \mid y + mv^{-1} \mid > \Delta \\ \\ 0 & \text{otherwise} \end{cases} \tag{E.18}$$

where

$$\Delta = \sqrt{2(v^{-1} + d_0) \left[\ell n \; (\beta \sqrt{1 + vd_0}) + \frac{m^2}{2v} \right]} \tag{E.19}$$

Receiver performance

The statistics of y were already determined in Appendix D and are

$$f(y \mid a, N) = \frac{1}{\sqrt{2\pi d_0}} \; e^{-\frac{y^2}{2d_0}} \tag{E.20}$$

$$f(y \mid a, SN) = \frac{1}{\sqrt{2\pi d_0}} \; e^{-\frac{(y - ad_0)^2}{2d_0}} \tag{E.21}$$

The performance equations are

$$P^{CS}(A \mid {}_{SN}^{N}) = \int_{|y + mv^{-1}| > \Delta} f(y \mid a, {}_{SN}^{N}) \, dy \qquad (E.22)$$

Performing the integration and simplifying yields

$$P^{CS}(A \mid N) = \Phi(\lambda_1) + \Phi(\lambda_2) \qquad (E.23)$$

$$P^{CS}(A \mid SN) = \Phi(\lambda_1 + a\sqrt{d_0}) + \Phi(\lambda_2 - a\sqrt{d_0}) \qquad (E.24)$$

where

$$\lambda_1 = \frac{-\Delta + mv^{-1}}{\sqrt{d_0}} \qquad (E.25)$$

$$\lambda_2 = \frac{-\Delta - mv^{-1}}{\sqrt{d_0}} \qquad (E.26)$$

APPENDIX F

MONOTONICITY OF THE LIKELIHOOD RATIO

FOR A SPECIAL CASE

This appendix will show that the total likelihood ratio for SKEA is strictly monotone increasing in the sufficient statistic

$$y = \underline{s}^t \, \Sigma^{-1} \, \underline{x}^t \qquad (F.1)$$

whenever the assumed a priori density of the unknown amplitude has zero probability on the negative axis.

From Appendix D the conditional observation statistics of y are

$$f(y \,|\, a, N) = \frac{1}{\sqrt{2\pi d_0}} \, e^{-\frac{y^2}{2d_0}} \qquad (F.2)$$

$$f(y \,|\, a, SN) = \frac{1}{\sqrt{2\pi d_0}} \, e^{-\frac{(y - a d_0)^2}{2d_0}} \qquad (F.3)$$

and the conditional likelihood ratio of y is

$$\ell(y \,|\, a) = \frac{f(y \,|\, a, SN)}{f(y \,|\, a, N)} = e^{ay - a^2 \frac{d_0}{2}} \qquad (F.4)$$

202

The total likelihood ratio $\ell(y)$ for a single composite hypothesis is the average of the conditional likelihood ratio $\ell(y|a)$ with respect to the assigned a priori density $g(a)$, i.e.,

$$\ell(y) = \int_{-\infty}^{\infty} e^{ay - a^2 \frac{d_0}{2}} g(a)\, da \qquad (F.5)$$

To determine whether or not $\ell(y)$ is a monotone function of its argument, check its derivative

$$\ell'(y) = \int_{-\infty}^{\infty} a\, e^{ay - a^2 \frac{d_0}{2}} g(a)\, da$$

$$= \int_{-\infty}^{\infty} a\, \frac{f(y|a, SN)}{f(y|N)}\, g(a)\, da$$

$$= \frac{1}{f(y|N)} \int_{-\infty}^{\infty} a\, f(y|SN)\, g(a|y)\, da$$

$$= \ell(y)\, E(a|y) \qquad (F.6)$$

Since $\ell(y)$ is a non-negative function of its argument, the monotonicity of $\ell(y)$ depends only upon the posterior mean $E(a|y)$. Furthermore, the posterior mean will be positive provided only that $g(a)$ has zero probability on the negative axis. Therefore, the monotonicity of $\ell(y)$ is assured whenever $g(a)$ has zero probability on the negative axis.

APPENDIX G

A USEFUL TRANSFORMATION OF

NORMAL RANDOM VARIABLES

This appendix will show that the transformation

$$u' = \frac{y}{d_0} \tag{G.1}$$

$$v^2 = \frac{w}{d_0} - \left(\frac{y}{d_0}\right)^2 \tag{G.2}$$

where

$$y = \underline{s}^t \Sigma^{-1} \underline{x} \tag{G.3}$$

$$w = \underline{x}^t \Sigma^{-1} \underline{x} \tag{G.4}$$

$$d_0 = \underline{s}^t \Sigma^{-1} \underline{s} \tag{G.5}$$

will convert the conditional density

$$f(\underline{x} \mid \theta) = \left(\frac{\theta}{2\pi \mid \Sigma \mid}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{x}^t \Sigma^{-1} \underline{x}} \tag{G.6}$$

to the conditional densities

$$f(u' \mid \theta) = \left(\frac{\theta \, d_0}{2\pi}\right)^{\frac{1}{2}} e^{\frac{-\theta \, d_0 \, u'^2}{2}} \qquad -\infty < u' < \infty$$

$$\text{(G.7)}$$

$$f(v \mid \theta) = \frac{2 \left(\dfrac{\theta \, d_0}{2}\right)^{\frac{n-1}{2}} v^{n-2} e^{-\frac{\theta \, d_0 \, y^2}{2}}}{\Gamma\left(\dfrac{n-1}{2}\right)} \qquad 0 \le v < \infty \quad \text{(G.8)}$$

Proof:

If we make the transformation

$$\underline{r} = A \, \underline{x} \qquad \text{(G.9)}$$

with A determined such that

$$A \, \Sigma \, A^t = I \qquad \text{(G.10)}$$

or equivalently

$$A^t A = \Sigma^{-1} \qquad \text{(G.11)}$$

then

$$f(\underline{r} \mid \theta) = \left(\frac{\theta}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\theta}{2} \underline{r}^t \underline{r}} \qquad \text{(G.12)}$$

Apply the transformation of Appendix C, namely

$$u' = \frac{\underline{\eta}^t \underline{r}}{E_s} \qquad \text{(G.13)}$$

$$v^2 = \frac{\underline{r}^t \underline{r}}{E_s} - \left(\frac{\underline{\eta}^t \underline{r}}{E_s}\right)^2 \qquad (G.14)$$

with

$$\underline{\eta} = A \underline{s} \qquad (G.15)$$

$$E_s = \underline{\eta}^t \underline{\eta} \qquad (G.16)$$

to obtain

$$f(u' \mid \theta) = \left(\frac{\theta E_s}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta E_s u'^2}{2}} \qquad -\infty < u' < \infty \qquad (G.17)$$

$$f(v \mid \theta) = \frac{2 \left(\frac{\theta E_s}{2}\right)^{\frac{n-1}{2}} v^{n-2} e^{-\frac{\theta E_s v^2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \qquad 0 \leq v < \infty \qquad (G.18)$$

Since

$$\underline{\eta}^t \underline{r} = (A \underline{s})^t (A \underline{x}) = \underline{s}^t A^t A \underline{x} = \underline{s}^t \Sigma^{-1} \underline{x} = y \qquad (G.19)$$

$$\underline{r}^t \underline{r} = (A \underline{x})^t (A \underline{x}) = \underline{x}^t A^t A \underline{x} = \underline{x}^t \Sigma^{-1} \underline{x} = w \qquad (G.20)$$

$$\underline{\eta}^t \underline{\eta} = (A \underline{s})^t (A \underline{s}) = \underline{s}^t A^t A \underline{s} = \underline{s}^t \Sigma^{-1} \underline{s} = d_0 \qquad (G.21)$$

the transformation

$$u' = \frac{y}{d_0} \tag{G.22}$$

$$v^2 = \frac{w}{d_0} - \left(\frac{y}{d_0}\right)^2 \tag{G.23}$$

results in

$$f(u'|\theta) = \left(\frac{\theta\, d_0}{2\pi}\right)^{\frac{1}{2}} e^{-\frac{\theta\, d_0\, u'^2}{2}} \qquad -\infty < u' < \infty \tag{G.24}$$

$$f(v|\theta) = \frac{2\left(\frac{\theta\, d_0}{2}\right)^{\frac{n-1}{2}} v^{n-2}\, e^{-\frac{\theta\, d_0\, y^2}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} \qquad 0 \le v < \infty \tag{G.25}$$

Q.E.D.

APPENDIX H

USEFUL INTEGRAL RELATIONS

The following integral relations turn out to be useful in analytically evaluating receiver performance.

1) $$\int_0^\infty x^{\nu-1} \, \Phi'(x) \, \Phi\left(\frac{tx}{\sqrt{\nu}}\right) \, dx = \frac{2^{\frac{\nu-3}{2}} \, \Gamma\left(\frac{\nu}{2}\right)}{\sqrt{\pi}} \, T_\nu(t) \qquad \text{(H.1)}$$

where

$$T_\nu(t) = \frac{1}{\sqrt{\nu\pi}} \, \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \, \int_{-\infty}^t \left(1 + \frac{u^2}{\nu}\right)^{-\left(\frac{\nu+1}{2}\right)} \, du \qquad \text{(H.2)}$$

2) $$\int_{-\infty}^\infty \Phi(ax + b) \, \Phi'(x) \, dx = \Phi\left(\frac{b}{\sqrt{1 + a^2}}\right) \qquad \text{(H.3)}$$

APPENDIX I

NON-EXISTENCE OF AN A PRIORI DENSITY*

This appendix will show that there exists no a priori density $g(\cdot)$ such that the solution to the double composite hypothesis detection problem SKE-UNL (signal known exactly, unknown noise level) results in a likelihood ratio which is monotone in the test statistic of the UMPU test. Specifically, it will be shown that there exists no function $g(\cdot)$ such that

$$\frac{\int_0^\infty \theta^{\frac{n}{2}} e^{-\frac{\theta}{2}(w-2y+d_0)} g(\theta)\, d\theta}{\int_0^\infty \theta^{\frac{n}{2}} e^{-\frac{\theta w}{2}} g(\theta)\, d\theta} = m\left(\frac{y}{\sqrt{w}}\right) \qquad (I.1)$$

where $m(\cdot)$ is any monotone function of its argument.

<u>Proof</u>:

Restate the hypothesis in terms of the function $h(\cdot)$ where

$$h(w) = \int_0^\infty \theta^{\frac{n}{2}} e^{-\frac{\theta w}{2}} g(\theta)\, d\theta \qquad (I.2)$$

---

*Credit is due to R. Heitmeyer for the technique of this proof.

APPENDIX J

EXISTENCE OF A^{-1} AND CALCULATION OF COEFFICIENTS

This appendix will show that it is plausible to assume the existence of the matrix A^{-1} where

$$A = (\alpha_{i+j-1}) = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_p \\ \alpha_2 & \alpha_3 & \cdots & \alpha_{p+1} \\ \vdots & \vdots & & \\ \alpha_p & \alpha_{p+1} & \cdots & \alpha_{2p-1} \end{bmatrix} \tag{J.1}$$

and

$$\alpha_k = \frac{(b+1)(b+2)\cdots(b+k)}{c^k} \tag{J.2}$$

Furthermore, if

$$\underline{a} = (a_1, a_2, \ldots a_p)^t = A^{-1} \underline{\alpha} \tag{J.3}$$

$$\underline{\alpha} = (\alpha_0, \alpha_1, \ldots \alpha_{p-1}) \tag{J.4}$$

then the coefficients a_i can be written in the form

$$a_i = d_i c^i \qquad i = 1, 2, \ldots p \tag{J.5}$$

with d_i independent of c.

212

<u>Existence of A^{-1}.</u>

Write the moments α_i in the form

$$\alpha_i = \gamma_i \, c^{-i} \qquad \text{(J.6)}$$

with

$$\gamma_i = (b+1)(b+2) \ldots (b+i) \qquad \text{(J.7)}$$

Define

$$C = \begin{bmatrix} c^{-1} & & & 0 \\ & c^{-2} & & \\ & & \ddots & \\ 0 & & & c^{-p} \end{bmatrix} \qquad \text{(J.8)}$$

$$\Gamma = (\gamma_{i+j-1}) = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_p \\ \gamma_2 & \gamma_3 & \cdots & \gamma_{p+1} \\ \vdots & & & \\ \gamma_p & \gamma_{p+1} & \cdots & \gamma_{2p-1} \end{bmatrix} \qquad \text{(J.9)}$$

so that the matrix A can be expressed in the form

$$A = c \, C\Gamma C \qquad \text{(J.10)}$$

and $\underline{\alpha}$ can be written as

$$\underline{\alpha} = c\,C\,\underline{\gamma} \tag{J.11}$$

where

$$\underline{\gamma} = (\gamma_1, \gamma_2, \cdots \gamma_p)^t \tag{J.12}$$

Since $c > 0$, the diagonal matrix C is obviously invertible. Hence the matrix A^{-1} exists provided only that the matrix Γ^{-1} exists. The existence of Γ^{-1} will be assured provided the determinant of the matrix Γ is unequal to zero. For $p = 2$

$$|\Gamma| = (b+1)^2 (b+2) \neq 0 \tag{J.13}$$

while for $p = 3$

$$|\Gamma| = 2(b+1)^3 (b+2)^2 (b+3) \neq 0 \tag{J.14}$$

since $b > -1$. Computer evaluation of $|\Gamma|$ for $p > 3$ and for specific values of b appeared to indicate the existence of Γ^{-1} in general and hence the existence of A^{-1} in general.

Calculation of coefficients.

If we assume that Γ is in fact invertible, then

$$\underline{a} = A^{-1}\underline{\alpha} = (cC\Gamma C)^{-1} c\,C\,\underline{\gamma}$$

$$= (c^{-1} C^{-1} \Gamma^{-1} C^{-1})\,c\,C\,\underline{\gamma}$$

$$= C^{-1} \Gamma^{-1}\,\underline{\gamma}$$

$$= C^{-1}\underline{d}$$

where

$$\underline{d} = (d_1, d_2, \ldots d_p)^t = \Gamma^{-1} \underline{\gamma} \qquad (J.16)$$

The vector $\underline{d}$ is independent of c since both Γ and $\underline{\gamma}$ are independent of c. In addition

$$C^{-1} = \begin{bmatrix} c & & & 0 \\ & c^2 & & \\ & & \ddots & \\ 0 & & & c^p \end{bmatrix} \qquad (J.17)$$

so that $\underline{a} = C^{-1} \underline{d}$ implies

$$a_i = d_i c^i \qquad i = 1, 2, \ldots p \qquad (J.18)$$

with d_i independent of c.

APPENDIX K

THE TRANSFORMATION $z[z(\underline{x} \mid g) \mid h]$

This appendix will construct a procedure for calculating $z[z(\underline{x} \mid g) \mid h]$. The transformation of Eqs. 3.59 and 3.60 has simplified the task considerably; we need only determine $z[\alpha(u, v) \mid h]$ where

$$\alpha(u, v) \stackrel{\Delta}{=} z(\underline{x} \mid g)$$

$$= \left(b + 1 + \frac{n}{2}\right) \ln \left[\frac{\left(u + \frac{1}{2}\right)^2 + v^2 + \frac{2c}{E_s}}{\left(u - \frac{1}{2}\right)^2 + v^2 + \frac{2c}{E_s}}\right] \tag{K.1}$$

with $f(u, v \mid h, SN)$ and $f(u, v \mid h, N)$ given by Eqs. 3.61 and 3.62.

Distribution Function

Let H denote either the N or SN hypothesis. The distribution function of $\alpha(u, v)$ can be determined via

$$F(t \mid h, H) \stackrel{\Delta}{=} P[\alpha(u, v) < t \mid h, H] \tag{K.2}$$

It can readily be shown that

216

$$F(t \mid h, H) = \begin{cases} 0 & t \leq -t_0 \\ P[(u, v) \in R_t \mid h, H] & -t_0 < t < t_0 \\ 1 & t \geq t_0 \end{cases} \qquad (K.3)$$

where

$$t_0 = \left(b + 1 + \frac{n}{2}\right)\ell n \left[1 + \frac{E_s}{4c} + \sqrt{\frac{E_s}{4c}\left(\frac{E_s}{4c} + 2\right)}\right] \qquad (K.4)$$

$$R_t = \begin{cases} (u, v) \mid v \geq 0 \ \text{and} \ (u - a)^2 + v^2 > r^2 \ \text{if} \ 0 < t < t_0 \ ; \ v \geq 0 \ \text{and} \\ (u - a)^2 + v^2 < r^2 \ \text{if} \ -t_0 < t < 0 \ ; \ \text{or} \ v \geq 0 \ \text{and} \qquad (K.5) \\ u < 0 \ \text{if} \ t = 0 \end{cases}$$

and

$$a = \frac{1}{2} \cdot \frac{\Delta + 1}{\Delta - 1} \qquad (K.6)$$

$$r^2 = \frac{\Delta}{(\Delta - 1)^2} - \frac{2c}{E_s} \qquad (K.7)$$

$$\Delta = e^{\frac{t}{b + 1 + \frac{n}{2}}} \qquad (K.8)$$

For $|t| < t_0$, the distribution function is

$$F(t \mid h, H) = \int_0^r \int_{u_1(\Delta, v)}^{u_2(\Delta, v)} f(u, v \mid h, H) \, du \, dv \qquad -t_0 < t < 0$$

$$F(t \mid h, H) = \int_0^\infty \int_{-\infty}^0 f(u, v \mid h, H) \, du \, dv \qquad t = 0 \qquad \text{(K.9)}$$

$$F(t \mid h, H) = 1 - \int_0^r \int_{u_1(\Delta, v)}^{u_2(\Delta, v)} f(u, v \mid h, H) \, du \, dv \qquad 0 < t < t_0$$

where

$$u_1(\Delta, v) = a - \sqrt{r^2 - v^2} \qquad \text{(K.10)}$$

$$u_2(\Delta, v) = a + \sqrt{r^2 - v^2} \qquad \text{(K.11)}$$

Density Function

The density function can be obtained by differentiating the distribution function.

$$f(t \mid h, H) = \frac{d}{dt} F(t \mid h, H) \qquad \text{(K.12)}$$

We can direct our attention to the interval $0 < t < t_0$; the interval $-t_0 < t < 0$ can be dealt with in a similar manner. Using the chain rule for differentiation

$$f(t \mid h, H) = \frac{d\Delta}{dt} \cdot \frac{d}{d\Delta} F(t \mid h, H) \qquad (K.13)$$

For $0 < t < t_0$, employ Leibnitz's rule to obtain

$$f(t \mid h, H) = \frac{d\Delta}{dt} \cdot \left\{ \int_0^r \left[f(u_2(\Delta, v), v \mid h, H) \cdot \frac{\partial}{\partial \Delta} u_2(\Delta, v) \right. \right.$$

$$\left. \left. - f(u_1(\Delta, v), v \mid h, H) \cdot \frac{\partial}{\partial \Delta} u_1(\Delta, v) \right] dv \right\} \qquad (K.14)$$

Simplifying,

$$f(t \mid h, H) = \frac{\Delta}{(\Delta - 1)^2 \left(b + 1 + \frac{n}{2}\right)} \left\{ \int_0^r \left[f(u_2(\Delta, v), v \mid h, H) \cdot u_2(\Delta, v) \right. \right.$$

$$\left. \left. + f(u_1(\Delta, v), v \mid h, H) u_1(\Delta, v) \right] \frac{dv}{\sqrt{r^2 - v^2}} \right\} \qquad 0 < t < t_0$$

$$(K.15)$$

If we make the transformation

$$w = 2\left(\frac{v}{r}\right)^2 - 1 \qquad (K.16)$$

then

$$f(t \mid h, H) = \int_{-1}^1 \frac{f_0(w \mid H)}{\sqrt{1 - w^2}} \, dw \qquad 0 < t < t_0 \quad (K.17)$$

where

$$f_0(w \mid H) = \frac{\Delta}{2(\Delta - 1)^2 \left(b + 1 + \frac{n}{2}\right)} \left[f(u_1(\Delta, v), v \mid h, H) \cdot u_1(\Delta, v) \right.$$

$$\left. + f(u_2(\Delta, v), v \mid h, H) \cdot u_2(\Delta, v) \right] \Bigg|_{v = r\sqrt{\frac{1+w}{2}}} \qquad \text{(K.18)}$$

A Gauss-Mehler quadrature routine is ideal for evaluating Eq. K.17 for specific values of t. An mth order Gauss-Mehler quadrature routine evaluates Eq. K.17 via

$$f(t \mid h, H) \cong \frac{\pi}{m} \sum_{j=1}^{m} f_0\left[\cos\left(\frac{2j-1}{2m} \right) \pi \mid H \right] \qquad 0 < t < t_0 \qquad \text{(K.19)}$$

This method is exact if $f_0(w \mid H)$ is a polynomial in w of degree $\leq$ $2m - 1$ (Ref. 29, p. 381).

An entirely analogous procedure for $-t_0 < t < 0$ leads to

$$f(t \mid h, H) \cong -\frac{\pi}{m} \sum_{j=1}^{m} f_0\left[\cos\left(\frac{2j-1}{2m} \right) \pi \mid H \right] \qquad -t_0 < t < 0 \qquad \text{(K.20)}$$

The Transformation $z[\alpha(u, v) \mid h]$

If we let

$$t = \alpha(u, v) \qquad \text{(K.21)}$$

then

$$z(t \mid h) = \ln\left[\frac{f(t \mid h, SN)}{f(t \mid h, N)} \right] \qquad 0 < |t| < t_0 \qquad \text{(K.22)}$$

It can readily be shown from symmetry considerations that the transformation $z(t \mid h)$ is an odd function, i.e.,

$$z(t \mid h) = -z(-t \mid h) \qquad 0 < |t| < t_0 \qquad \text{(K. 23)}$$

and hence from continuity that

$$z(t \mid h) = 0 \qquad \qquad t = 0 \qquad \text{(K. 24)}$$

REFERENCES

1. W. W. Peterson, T. G. Birdsall and W. C. Fox, "The theory of signal detectability," IRE Trans. on Information Theory, IT-4, 1954.

2. D. Middleton and D. Van Meter, "Modern statistical approaches to reception in communication theory," IRE Trans. on Information Theory, IT-4, 1954.

3. U. Grenander, "Stochastic processes and statistical inference," Arkiv det Mat., 1, 1950.

4. T. G. Birdsall, "Application of detection theory in bandlimited Gaussian noise," Random Processes, Linear Systems and Radar, The University of Michigan Summer Conferences, 1964.

5. L. J. Savage, The Foundations of Statistics, John Wiley and Sons, Inc., New York, 1950.

6. H. Cramér, Mathematical Statistics, Princeton University Press, Princeton, 1951.

7. T. S. Ferguson, Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York, 1967.

8. E. L. Lehmann, Testing Statistical Hypotheses, Wiley, New York, 1959.

9. H. Raiffa and R. O. Schlaifer, Applied Statistical Decision Theory, Division of Research, Graduate School of Business Administration, Harvard University, Boston, 1961.

10. D. Blackwell and M. A. Gershick, Theory of Games and Statistical Decisions, Wiley, New York, 1954.

11. C. W. Helstrom, Statistical Theory of Signal Detection, Pergamon Press, New York, 1960.

12. J. C. Hancock, Signal Detection Theory, McGraw-Hill, New York, 1966.

REFERENCES (Cont.)

13. S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.

14. A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.

15. W. G. Davenport and W. L. Root, An Introduction to the Theory of Random Signals and Noise, McGraw-Hill, New York, 1958.

16. H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, John Wiley and Sons, Inc., New York, 1968.

17. N. M. Blachman, Noise and Its Effect on Communication, McGraw-Hill, New York, 1966.

18. J. M. Wozencraft and I. M. Jacobs, Principles of Communication Engineering, Wiley, New York, 1967.

19. R. A. Roberts, Theory of Signal Detectability: Composite Deferred Decision Theory, Cooley Electronics Laboratory Technical Report No. 161, The University of Michigan, Ann Arbor, March 1965.

20. O. Ye Antonov, "Optimum detection of signals in non-Gaussian noise--detection of a signal of unknown amplitude and phase," Radio Engineering and Electronic Physics, vol. 12, May 1967.

21. L. W. Nolte, Adaptive Realizations of Optimum Detectors for Synchronous and Sporadic Recurrent Signals in Noise, Cooley Electronics Laboratory Technical Report No. 163, The University of Michigan, Ann Arbor, Michigan, March 1965.

22. R. L. Spooner, Theory of Signal Detectability: Extension to the Double Composite Hypothesis Situation, Cooley Electronics Laboratory Technical Report No. 192, The University of Michigan, Ann Arbor, Michigan, April 1968.

23. T. G. Birdsall, The Theory of Signal Detectability: ROC Curves and Their Character, Ph.D. dissertation at The University of Michigan, Ann Arbor, Michigan, 1966.

REFERENCES (Cont.)

24. W. P. Tanner and T. G. Birdsall, "Definitions of d' and η as psychological measures," The Journal of the Acoustical Society of America, vol. 30, October 1958.

25. J. Spragins, "A note on the iterative application of Bayes' rule," IEEE Trans. on Information Theory, IT-11, October 1965.

26. R. Deutsch, Theory of Estimation, Prentice Hall, Englewood Cliffs, New Jersey, 1965.

27. D. V. Lindley, "On a measure of information provided by an expectation," Annals of Mathematical Statistics, December 1956.

28. A. Feinstein, Foundations of Information Theory, McGraw-Hill, New York, 1966.

29. Z. Kopal, Numerical Analysis, John Wiley and Sons, Inc., New York, 1961.

30. M. Abramowitz and L. A. Stegun (Eds.), Handbook of Mathematical Functions, National Bureau of Standards, AMS 55, Washington, D. C., 1964.

31. T. G. Birdsall, Adaptive Detection Receivers and Reproducing Densities, Cooley Electronics Laboratory Technical Report No. 194, The University of Michigan, Ann Arbor, Michigan, July 1968.

32. R. L. Spooner, "Comparison of receiver performance: the ESP receiver," The Journal of the Acoustical Society of America, vol. 45, no. 1, January 1969.

33. T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," IEEE Trans. on Information Theory, vol. IT-15, no. 3, May 1969.

34. S. Kullback and R. A. Leibler, "On information and sufficiency," Annals of Mathematical Statistics, vol. 22, no. 1, 1951, pp. 79-86.

35. H. Jeffreys, Theory of Probability (2nd ed.), Oxford Univ. Press, 1948.

36. A. P. Dempster and M. Schatzoff, "Expected significance level as a sensitivity index for test statistics," Journal of the American Statistical Association, Vol. 60, June 1965, pp. 420-437.

DISTRIBUTION LIST

	No. of Copies
Office of Naval Research (Code 468) Navy Department Washington, D. C. 20360	2
Director, Naval Research Laboratory Technical Information Division Washington, D. C. 20360	6
Director Office of Naval Research Branch Office 1030 East Green Street Pasadena, California 91101	1
Office of Naval Research San Francisco Annex 1076 Mission Street San Francisco, California 94103	1
Office of Naval Research New York Annex 207 West 24th Street New York, New York 10011	1
Director Office of Naval Research Branch Office 219 South Dearborn Street Chicago, Illinois 60604	1
Commanding Officer Office of Naval Research Branch Office Box 39 FPO New York 09510	8
Commander, Naval Ordnance Laboratory Acoustics Division White Oak, Silver Spring, Maryland 20910	1

225

DISTRIBUTION LIST (Cont.)

	No. of Copies
Commanding Officer and Director Naval Electronics Laboratory San Diego, California 92152	1
Commanding Officer and Director Navy Underwater Sound Laboratory Fort Trumball New London, Connecticut 06321	1
Commanding Officer Naval Air Development Center Johnsville, Warminister, Pennsylvania	1
Commanding Officer and Director David Taylor Model Basin Washington, D. C. 20370	1
Superintendent Naval Postgraduate School Monterey, California 93940 Attn: Professor L. E. Kinsler	1
Commanding Officer Navy Mine Defense Laboratory Panama City, Florida 32402	1
Superintendent Naval Academy Annapolis, Maryland 21402	1
Commander Naval Ordnance Systems Command Code ORD-0302 Navy Department Washington, D. C. 20360	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Commander Naval Ship Systems Command Code SHIPS-03043 Navy Department Washington, D. C. 20360	1
Commander Naval Ship Systems Command Code SHIPS-1630 Navy Department Washington, D. C. 20360	1
Chief Scientist Navy Underwater Sound Reference Division Post Office Box 8337 Orlando, Florida 38200	1
Defense Documentation Center Cameron Station Alexandria, Virginia	20
Dr. Melvin J. Jacobson Rensselaer Polytechnic Institute Troy, New York 12181	1
Dr. Charles Stutt General Electric Company P. O. Box 1088 Schenectady, New York 12301	1
Dr. J. V. Bouyoucos General Dynamics/Electronics 1400 N. Goodman Street P. O. Box 226 Rochester, New York 14609	1
Mr. J. Bernstein EDO Corporation College Point, New York 11356	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. T. G. Birdsall Cooley Electronics Laboratory The University of Michigan Ann Arbor, Michigan 48105	1
Dr. John Steinberg Institute of Marine Science The University of Miami Miami, Florida 33149	1
Dr. R. A. Roberts Department of Electrical Engineering University of Colorado Boulder, Colorado	1
Commander Naval Ordnance Test Station Pasadena Annex 3203 E. Foothill Boulevard Pasadena, California 91107	1
Dr. Stephen Wolff John Hopkins University Baltimore, Maryland 21218	1
Dr. M. A. Basin Litton Industries 8000 Woodley Avenue Van Nuys, California 91409	1
Dr. Albert Nuttall Litton Systems, Inc. 335 Bear Hill Road Waltham, Massachusetts 02154	1
Dr. Philip Stocklin Box 360 Raytheon Company Newport, Rhode Island 02841	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. H. W. Marsh Raytheon Company P.O. Box 128 New London, Connecticut 06321	1
Mr. Ken Preston Perkin-Elmer Corporation Electro-Optical Division Norwalk, Connecticut 06852	1
Mr. Tom Barnard Texas Instruments Incorporated 100 Exchange Park North Dallas, Texas 75222	1
Dr. John Swets Bolt, Beranek and Newman 50 Moulton Street Cambridge 38, Massachusetts	1
Dr. H. S. Hayre The University of Houston Cullen Boulevard Houston, Texas 77004	1
Dr. Robert R. Brockhurst Woods Hole Oceanographic Institute Woods Hole, Massachusetts	1
Cooley Electronics Laboratory The University of Michigan Ann Arbor, Michigan 48105	50
Director Office of Naval Research Branch Office 495 Summer Street Boston, Massachusetts 02210	1

DISTRIBUTION LIST (Cont.)

	No. of Copies
Dr. L. W. Nolte Department of Electrical Engineering Duke University Durham, North Carolina	2
Mr. F. Briggson Office of Naval Research Representative 121 Cooley Building The University of Michigan Ann Arbor, Michigan 48105	1
Dr. Ronald Spooner Bolt, Beranek and Newman, Inc. 1501 Wilson Boulevard Arlington, Virginia 22209	1
Computer Science Library University of Pittsburgh Pittsburgh, Pennsylvania 15213	1