Technical Report

T-4

Nonparametric Detection Using Extreme-Value Theory

Prepared for

National Aeronautics and Space Administration

Electronics Research Center

under

NASA GRANT NGR 33-006-020

Prepared by

Laurence B. Milstein

Donald L. Schilling, Associate Professor

Jack K. Wolf, Associate Professor

Department of Electrical Engineering

Polytechnic Institute of Brooklyn

FACILITY FORM 602

N68-18070 (ACCESSION NUMBER) (THRU)

87 (PAGES) 1 (CODE)

CR-93261 (NASA CR OR TMX OR AD NUMBER) 07 (CATEGORY)

Technical Report

T-4

Nonparametric Detection Using Extreme-Value Theory

Prepared for

National Aeronautics and Space Administration

Electronics Research Center

under

NASA GRANT NGR 33-006-020

Prepared by

Laurence B. Milstein

Donald L. Schilling, Associate Professor

Jack K. Wolf, Associate Professor

Department of Electrical Engineering

Polytechnic Institute of Brooklyn

ACKNOWLEDGEMENT

This research was partially supported by the Air Force Office of Scientific Research under Contract AF 49(638)-1600, and under NASA Grant NGR 33-006-020.

Special thanks are due to the National Aeronautics and Space Administration for support given under the NASA Predoctoral Traineeship Program.

AN ABSTRACT

NONPARAMETRIC DETECTION USING EXTREME-VALUE THEORY

by

Laurence B. Milstein

Advisors: Jack K. Wolf
Donald L. Schilling

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Electrical Engineering).

This paper concerns itself with the detection of a binary signal in additive, but statistically unknown, noise. The signal will be either a constant signal, or a slowly fading signal. The noise will be arbitrary except for the one restriction that its probability density function exhibit some type of exponential behavior on its "tails."

The detector will be based upon Gumbel's extreme-value theory (EVT). Extreme-value theory is a branch of mathematical statistics which considers the asymptotic distributions of the maximum and minimum samples from sets of independent and identically distributed random variables. This theory will be used to obtain estimates of the optimum threshold and the probability of error of a binary detector. Confidence intervals are obtained for all estimates.

A comparison is made between the EVT detector and another nonparametric detector, one which is based upon the rank test. It is shown that in certain situations, the EVT detector becomes identical to the Neyman-Pearson detector, and therefore will outperform the rank or any other nonparametric detector.

When the signal fades, it is shown that the EVT detector becomes adaptive and can track the fade. Computer simulations are run for a fading signal, and the results verify the theory.

Finally, while the above results are obtained with the help of an initial learning period, a study is made, for the case of detecting a constant signal in additive noise, of the performance of the detector when the learning period is removed. It is shown that for low error rates, the estimates will converge to values close to those obtained when the learning period is present. A computer simulation is run for this case, and again the results verify the theory.

TABLE OF CONTENTS

			Page
Introduction			
Chapter 1		Basic Theory	3
	1.1	Extreme-Value Theory (EVT)	3
	1.2	Relationship between EVT and Detection Theory	7
	1.3	Estimation of Parameters and Confidence Intervals	10
	1.4	Numerical Examples	16
	1.5	Review	19
Chapter 2		Nonparametric Detection using EVT	21
	2.1	The EVT Dectector	21
	2.2	Examples	24
	2.2.1	Narrowband FM for high SNR and Gaussian Noise	24
	2.2.2	On-Off Systems (Noncoherent Detection)	25
	2.2.3	On-On System	
	2.2.4	Impulse Noise	29
	2.3	Comparison with Rank Test	32
Chapter 3		Fading	39
	3.1	Fixed Threshold	39
	3.2	Adaptive Detection	45
	3.2.1	Adaptive EVT Detector	45
	3.2.2	Confidence Intervals	47
	3.2.3	Computer Simulated Results	50
Chapter 4		Decision-Directed Measurements	57
	4.1	Elimination of Learning Period	57
	4.2	Confidence Intervals and Computer Simulated Results	61

Table of Contents (contd)

Summary		71
Appendix A.	Density Function for Minimum of n Independent Samples	73
Appendix B.	Asymptotic Normality for Functions of Nonidentically Distributed Random Variables	76
References		78

LIST OF FIGURES

Figure		Page
2.1	Binary FSK Detector	27
2.2	On-Off Detector	30
3.1	Comparison of Optimum and Estimated Thresholds (A = 5)	53
3.2	Comparison of Number of Errors made by Optimum and EVT Detectors (A = 5)	54
3.3	Comparison of Optimum and Estimated Thresholds (A = 6)	55
3.4	Comparison of Number of Errors made by Optimum and EVT Detectors (A = 6)	56
4.1	Sample Mean as Threshold Estimate	58
4.2	Convergence of Threshold Estimate Using Decision-Directed Measurements	70

INTRODUCTION

Nonparametric detection is that branch of communication theory which concerns itself with detecting signals in statistically unknown noise. In this report, a low-pass binary signal is assumed, and in most cases, the system under consideration is an on-off system. In this type of problem, the optimum threshold (that is, the threshold which results in minimum probability of error) could be found by setting up the likelihood-ratio if the probability density of noise and the probability density of signal-plus-noise were known. For equal apriori probabilities of transmission, the threshold would occur at the point of intersection of the two probability density functions. When the density functions are unknown, suboptimum detection schemes must be used.

It will be the object of this report to obtain a nonparametric detector which yields an error rate comparable to that obtained using an optimum parametric detector. The nonparametric detector employed is based upon extreme-value theory.

Extreme-value theory is a branch of mathematical statistics which deals with the asymptotic probability distributions of extreme samples taken from sets of independent and identically distributed random variables. This theory is used in a nonparametric detection scheme by using the knowledge of the behavior of the extremes to obtain knowledge of the behavior of the initial variate on its "tails". In other words, if we know the properties of the maximum of a set of n independendent and identically distributed random variables, we will use this information to obtain the properties of the original random variable in the vicinity of the maximum value.

In Chapter 1, a brief review of extreme-value theory is given,

and the general method of using this theory to detect signals in unknown noise is presented. The only restriction that is placed on the noise and signal-plus-noise densities is that they exhibit some type of exponential behavior on their "tails".

Chapter 2 considers specific examples of detecting a constant signal in unknown noise. In addition, a comparison is made with another nonparametric detector, namely the rank detector. It is shown that the extreme-value theory (EVT) detector can, in certain situations, perform as well as the optimum Neyman-Pearson detector.

In Chapter 3, a fading signal is considered. Two different schemes are presented. The first is appropriate when the detector uses a constant threshold, and the second forms the basis for an adaptive receiver. In the latter case, computer simulations are run, and the results are seen to verify the theory.

The procedure presented in the first three chapters requires that the receiver employ an initial learning period, during which the detector samples the noise and the signal-plus-noise separately. Chapter 4 considers a scheme using decision-directed measurements, in which the learning period is eliminated. Again, computer simulations are run and can be seen to substantiate the results predicted by the theory.

CHAPTER 1 - BASIC THEORY

1.1 Extreme-Value Theory

Extreme-value theory (EVT) is the theory of the probability distributions of the extremes of sets of independent and identically distributed random variables. That is, given n independent random variables $x_1 \ldots x_n$, all of which have the same probability density, what is the distribution of the largest (or smallest) sample?

It is well known how to obtain the exact distributions for these extremes, but, in general, these functions are quite complicated. Gumbel has derived simple asymptotic expressions (for large n) for the distributions of the extremes, and these results are the basis of this report.[1.1, 1.2]

Gumbel's results are subdivided into three categories, according to the type of distribution for the random variable x. The first, and most important, is the exponential-type, which deals with those distributions which approach either unity on their right-hand tail or zero on their left-hand tail at least as fast as an exponential.

The second category deals with unlimited distributions which only possess a finite number of moments (e.g., Cauchy density). A density is limited to the right if it is identically zero for x greater than some number C_1, and limited to the left if it is identically zero for x less than some number C_2. Otherwise, it is unlimited. Limited distributions are considered in Gumbel's third category. Since the last two types can be obtained from the exponential-type by simple transformations, and since the most common noise densities encountered in practice fall into the exponential class, only that class will be considered in this report.

Let $f(x)$ and $F(x)$ be the probability density function and the cumulative distribution function respectively of the random variable x. Then x will be said to be of the exponential-type if it satisfies one of the

two following equations:[1.3]

For maximum values:

$$\lim_{x \to \infty} \frac{f(x)}{1-F(x)} = \lim_{x \to \infty} -\frac{f'(x)}{f(x)} , \tag{1.1}$$

where

$$f'(x) = \frac{df(x)}{dx} .$$

For minimum values:

$$\lim_{x \to -\infty} \frac{f(x)}{F(x)} = \lim_{x \to -\infty} \frac{f'(x)}{f(x)} . \tag{1.2}$$

The most common examples of densities of this type are the exponential distribution itself, the Rayleigh distribution, and the chi-square distribution for maximum values, and the normal distribution for both maximum and minimum values.

Equation (1.1) ((1.2)) insures that the right (left) hand tail of the distribution behaves as an exponential. Gumbel shows this in his derivation of the extreme-value densities[1.4] (see Appendix A). Starting with an arbitrary density that satisfies the first condition, for example, and using a Taylor series expansion about a parameter he labels u_n, Gumbel shows that, asymptotically,

$$F(x) = 1 - \frac{1}{n} e^{-\alpha_n (x-u_n)} ,$$

where α_n and u_n are defined as

$$F(u_n) = 1 - \frac{1}{n} , \tag{1.3}$$

and

$$\alpha_n = nf(u_n), \tag{1.4}$$

n being the number of samples from which the maximum is chosen.

u_n is called the expected largest value. Its relation to n can be thought of as follows: n is the number of samples one would have to take, on the average, to find a sample which is greater than u_n.

α_n is called the intensity function. In general, intensity functions $\alpha(x)$ are defined[1.5] such that

$$\alpha(x)\,dx = \frac{f(x)\,dx}{1-F(x)},$$

and represent the probability that a value, known to be equal to or greater than x, is between x and $x+dx$.

The analogous parameters for minimum values, u_1 and α_1, are defined as follows:

$$F(u_1) = \frac{1}{n}, \tag{1.5}$$

and

$$\alpha_1 = n\,f(u_1). \tag{1.6}$$

Since the Taylor series expansion is taken about u_n (or u_1), the important point to note is that for any distribution in the exponential class, except the exponential distribution itself, the Taylor series coefficients are all approximations. This means, of course, that when using Gumbel's results, they must only be used in an appropriate range about u_n (or u_1).

Gumbel proceeds from this point to derive the final form of the distribution function for the extreme-value. That form is a double exponential. Specifically,

$$F_{max}(x) = e^{-e^{-\alpha_n(x-u_n)}}, \tag{1.7}$$

and

$$F_{min}(x) = 1-e^{-e^{\alpha_1(x-u_1)}}. \tag{1.8}$$

Since we are not interested in the distribution of the extreme-

values, but rather in the distribution of the initial variate on its tails, these last two formulas are not the important results. The limiting forms for the initial variate (see Appendix A) are

$$\underset{x \to \infty}{F(x)} = 1 - \frac{1}{n} e^{-a_n(x-u_n)} \quad , \tag{1.9}$$

and

$$\underset{x \to -\infty}{F(x)} = \frac{1}{n} e^{a_1(x-u_1)} \quad . \tag{1.10}$$

It should be noted that the probability of the initial variate being greater than a fixed number can be obtained from the probability that the maximum value is greater than that number directly. This can be shown easily as follows:

Let F(x) be the initial distribution, and $\Phi(x)$ be the distribution of the maximum value. Since $\Phi(x) = [F(x)]^{n^{1.6}}$, we can always invert the above equation, and, knowing $\Phi(x)$, obtain F(x). However, this is only convenient if we are dealing with a single point, and even then it is more trouble than is necessary. If we should want to use F(x) in an equation, say a likelihood-ratio equation, the equation would become extremely complex.

Therefore, we will now leave Gumbel's final result, and go back to his intermediary result, namely, equations (1.9) and (1.10).

1.2 Relationship between EVT and Detection Theory

In the previous work in nonparametric detection, the detection schemes were based upon various nonparametric statistics, most notably the rank test or some modification of the rank test.[1.7, 1.8, 1.9, 1.10] The main trouble with these tests is that they do not use the amplitude information available from the data samples, but rather use information such as the algebraic sign of the sample or the relative ordering of the sample by size (rank).

It is the purpose of this report to present a detection scheme which does use amplitude information. This scheme is based on the observation that the only parts of the unknown distributions that are of interest are the tails of those distributions. This is precisely what we can obtain from Gumbel.

More specifically, if we knew the probability densities for noise and signal-plus-noise, what we would do would be to form the likelihood-ratio and solve for the threshold. Since we assume we do not know the distributions, we first take an initial series of measurements, estimate u_1, u_n, α_1, and α_n, and then form the likelihood-ratio. That is,

$$f_n(x) \sim \frac{\alpha_n}{n_n} e^{-\alpha_n(x-u_n)}, \tag{1.11}$$

and

$$f_{s+n}(x) \sim \frac{\alpha_1}{n_1} e^{\alpha_1(x-u_1)} . \tag{1.12}$$

Thus,

$$\frac{f_n(x_t)}{f_{s+n}(x_t)} = \frac{\frac{\alpha_n}{n_n} e^{-\alpha_n(x_t-u_n)}}{\frac{\alpha_1}{n_1} e^{\alpha_1(x_t-u_1)}} = \beta , \tag{1.13}$$

where β is the ratio of the apriori probabilities of sending signal-plus-noise and noise only.

Solving eq. (1.13) for x_t yields

$$x_t = \frac{u_1 \alpha_1 + u_n \alpha_n + \ln \left(\frac{\alpha_n}{\alpha_1} \frac{n_1}{n_n} \frac{1}{\beta}\right)}{\alpha_1 + \alpha_n} \quad . \tag{1.14}$$

For $\beta = 1$, the false alarm probability, P_{fa}, and the false dismissal probability, P_{fd}, are :

$$P_{fa} = \frac{1}{n_n} e^{-\alpha_n (x_t - u_n)} , \tag{1.15}$$

and

$$P_{fd} = \frac{1}{n_1} e^{\alpha_1 (x_t - u_1)} \quad . \tag{1.16}$$

Of course, the u's and α's employed in eq. (1.15) and eq. (1.16) are only estimates of the true values, since the actual distributions are unknown. However, we will initially assume the u's and α's are known exactly. This then gives an upper bound to the performance of the system, and thus gives an approximation to the actual system behavior. (That is, if, knowing the actual values of u and α, we cannot predict with reasonable accuracy the threshold and probability of error, then we cannot hope to do so using estimated values.)

Let us illustrate the above concepts by considering a numerical example. Table (1) shows results obtained when x has a normal density with zero mean and unit variance. It was computed by choosing a threshold, calculating the actual P_{fa} (that is, finding the area under an N(0, 1) curve for all x greater than the threshold), and then using eq. (1.15) at

the same threshold for different values of n. It can be seen that there is a range of n giving acceptable estimates for the probability of error. That is, to measure a probability of error equal to P_e, there are certain acceptable sample sizes n which can be used. This result can be seen from the Taylor series. Since the series was expanded about u_n, and since the coefficients in the series are only approximations (except for the exponential distribution), the only exact value we can find is $F(u_n)$.

However, at $x = u_n$, we have

$$P_e = 1 - F(u_n) = \frac{1}{n}, \tag{1.17}$$

If we are in a reasonably restricted range about $x = u_n$, we will not have an exact estimate of $F(x)$, but we will have a "good" estimate. "Good" here is defined as whatever is acceptable to the particular situation.

Hence, if we have a communication system in which the probability of false alarm is P_{fa}, the optimum value of n would be $n = \frac{1}{P_{fa}}$. Also, in estimating the threshold of the system, we would want u_n to be as close as possible to the actual threshold, since the further away we are, the more inaccurate is the Taylor series.

TABLE (1)

Approximations to P_{fa} for Various Values of n.

n \ Actual P_{fa}	1.35×10^{-3}	3.16×10^{-5}	2.87×10^{-7}
10^2	1.66×10^{-3}	11.5×10^{-5}	80.9×10^{-7}
10^3	1.36×10^{-3}	4.74×10^{-5}	16.8×10^{-7}
10^4	1.71×10^{-3}	3.3×10^{-5}	6.41×10^{-7}
10^5	2.9×10^{-3}	3.29×10^{-5}	3.73×10^{-7}
10^6	5.83×10^{-3}	3.73×10^{-5}	2.98×10^{-7}

1.3 Estimation of Parameters and Confidence Intervals

Returning to the nonparametric communcation problem, we must first find a method to estimate the unknown parameters u_1, a_1, u_n and a_n. There are a variety of ways of accomplishing this, but only the simplest of these is presented here.[1.11]

To this end, we must return to Gumbel's double exponential distribution. Using Gumbel's terminology, we call $y = a_n(x-u_n)$ the reduced largest variate (the word "reduced" is used because y is dimensionless). If we let $\varphi(y)$ be the probability density function of y, we have

$$\varphi(y) = e^{-y-e^{-y}} . \tag{1.18}$$

The generating function of y is then

$$G_n(t) = \int_{-\infty}^{\infty} e^{yt} e^{-y-e^{-y}} dy . \tag{1.19}$$

Let $z = e^{-y}$.

This gives

$$G_n(t) = -\int_{\infty}^{0} e^{yt-z} dz = \int_{0}^{\infty} e^{-t\ln z} e^{-z} dz \tag{1.20}$$

$$= \int_{0}^{\infty} z^{-t} e^{-z} dz = \Gamma(1-t) ,$$

where

$$\Gamma(t) = \int_{0}^{\infty} z^{t-1} e^{-z} dz$$

is the gamma function.

Similarly, we can show that the generating function of the reduced smallest variate is

$$G_1(t) = \Gamma(1+t) .$$

Since $\Gamma'(1) = -\gamma$, γ being Euler's constant, and $\Gamma''(1) = \frac{\pi^2}{6} + \gamma^2$, we can obtain the first and second moments of y, and from these, the

mean and variance of the extreme-values themselves.

This results in the following expressions:

$$E(x_n) = E(x_{max}) = u_n + \frac{\gamma}{\alpha_n} \tag{1.21}$$

$$E(x_1) = E(x_{min}) = u_1 - \frac{\gamma}{\alpha_1} \tag{1.22}$$

$$\sigma_n^2 = var(x_{max}) = \frac{\pi^2}{6} \frac{1}{\alpha_n^2} \tag{1.23}$$

$$\sigma_1^2 = var(x_{min}) = \frac{\pi^2}{6} \frac{1}{\alpha_1^2} \tag{1.24}$$

If we now replace the theoretical means and variances with the sample means and sample variances of the extreme values, we will have two sets of two equations and two sets of two unknowns, and we can therefore solve for the u's and α's .

As mentioned above, the procedure is used because of its simplicity, not because it is optimum. If more accurate estimates are needed, there are various maximum-likelihood estimates that can be used.[1.12, 1.13, 1.14] The trouble with these estimates is the difficulty in solving the resulting equations.

Returning to the straightforward estimates, the obvious questions are how good are they, and, more important, how good are the resultant estimates of the probability of error and the threshold?

The fact that the sample mean and sample standard deviation are asymptotically normal leads to results which are tractable.

The following asymptotic results are given by Gumbel:[1.15] The limiting means of the sample mean and sample standard deviation are the corresponding population values (that is, the actual mean and standard deviation of the distribution). Also,

$$\text{var}\,(\hat{\sigma}) = \frac{\sigma^2}{N}\left(\frac{\beta_2 - 1}{4}\right), \tag{1.25}$$

and

$$\text{var}\,(a\bar{x}_o + b\hat{\sigma}) = \frac{\sigma^2}{N}\left[a^2 + ab\sqrt{\beta_1} + b^2(\beta_2 - 1)/4\right], \tag{1.26}$$

where $\hat{\sigma}$ is the sample standard deviation, $\bar{x}_o$ is the sample mean, $\beta_1 = \mu_3^2\,\mu_2^{-3}$, and $\beta_2 = \mu_4\,\mu_2^{-2}$, μ_r being the rth central moment. Finally, N is the number of extremes that are used, each extreme coming from n independent samples (i.e., there will be a total of nN samples).

Again using the generating function for the reduced variate, we can compute β_1 and β_2. The final results are: $\sqrt{\beta_1} = 1.3$ and $\beta_2 = 5.4$ for maximum values, and $\sqrt{\beta_1} = -1.3$ and $\beta_2 = 5.4$ for minimum values.

We therefore conclude that the estimates of u and $\frac{1}{a}$ are asymptotically normal, unbiased estimates with variances given as follows:

$$\text{var}\left(\frac{1}{\hat{a}}\right) = \frac{1.1}{Na^2}, \tag{1.27}$$

and

$$\text{var}(\hat{u}) = \frac{1.04}{Na^2}. \tag{1.28}$$

In order to obtain confidence intervals on the estimates for the threshold and probability of error, we need the following results from Cramer:[1.16]

1. If an arbitrary function of sample moments is continuous in some neighborhood of the corresponding population moments, if the function has continuous first and second derivatives with respect to the sample moments in that neighborhood, and if the sample moments are sums of independent identically distributed random variables which obey the standard (Lindeberg-Lévy) central-limit theorem, then the function is asymptotically normal with parameters $\eta_{\hat{f}}$ and $\sigma_{\hat{f}}^2$.

2. If f, f_{m_1}, and f_{m_2} denote the value of the function f and its two first order partial derivatives (assuming it is a function of just two sample moments) evaluated at the point $m_1 = E(m_1)$ and $m_2 = E(m_2)$ (that is, the population values), then

$$\eta_{\hat{f}} = f \ , \tag{1.29}$$

and

$$\sigma^2_{\hat{f}} = \mathrm{var}(m_1)\, f^2_{m_1} + 2\,\mathrm{cov}(m_1, m_2)\, f_{m_1} f_{m_2} + \mathrm{var}(m_2)\, f^2_{m_2} \ . \tag{1.30}$$

For the estimate of probability of error, two asymptotic distributions will be given. One will be for the case when x_t is a constant, such as an on-on system having a threshold which is always zero.[1.17] The other will be appropriate when x_t is estimated from the u's and α's by eq. (1.14).

When x_t is constant, the estimate for the false-alarm probability, written in terms of the sample mean and sample variance of maximum values, is

$$\hat{P}_{fa} = \frac{1}{n}\, e^{-\hat{\alpha}_n(x_t - \hat{u}_n)} = \frac{1}{n}\, e^{-\frac{\pi}{\sqrt{6}\sqrt{\hat{\sigma}_n^2}}\left[x_t - \bar{x}_n + \frac{\gamma\sqrt{6}\sqrt{\hat{\sigma}_n^2}}{\pi}\right]} ,$$

or,

$$\hat{P}_{fa} = \frac{1}{n}\, e^{\gamma}\, e^{-\frac{\pi}{\sqrt{6}}\left[\frac{x_t - \bar{x}_n}{\sqrt{\hat{\sigma}_n^2}}\right]} \ . \tag{1.31}$$

If we let P_{fa} equal the value of $\hat{P}_{fa}$ evaluated at $\bar{x}_n = E(\bar{x}_n)$, and $\hat{\sigma}_n^2 = \sigma_n^2$, and $P_{fa\,\bar{x}_n}$ and $P_{fa\,\hat{\sigma}_n^2}$ the corresponding values of the partial derivatives of $\hat{P}_{fa}$ with respect to $\bar{x}_n$ and $\hat{\sigma}_n^2$ respectively, then,

$$\eta_{\hat{P}_{fa}} = P_{fa} \ , \tag{1.31}$$

and

$$\sigma^2_{\hat{P}_{fa}} = \mathrm{var}(\bar{x})\,[P_{fa\bar{x}_n}]^2 + 2\,\mathrm{cov}(\bar{x}_n, \hat{\sigma}^2_n)\,P_{fa\bar{x}_n}\,P_{fa\,\hat{\sigma}^2_n} + \\ + \mathrm{var}(\hat{\sigma}^2_n)[P_{fa\,\hat{\sigma}^2_n}]^2, \tag{1.33}$$

are respectively the mean and variance of the asymptotic normal distribution. A specific example will be given later.

The $\mathrm{cov}(\bar{x}_n, \hat{\sigma}^2_n)$ is shown in Cramér to be

$$\mathrm{cov}(\bar{x}_n, \hat{\sigma}^2_n) = \frac{\mu_3}{N}\,.$$ [1.18]

For the means and variances of the limiting distributions of $\hat{x}_t$ and the estimate of error probability when the threshold is a random variable, we must apply the two dimensional version of Cramér's theorem, since we now have functions involving sample moments of two distributions, namely moments of minimum values, and moments of maximum values. If, for either case, f represents the function to be estimated, $\hat{f}$ represents the estimate, and $f_{\bar{x}_n}$, $f_{\hat{\sigma}^2_n}$, $f_{\bar{x}_1}$, and $f_{\hat{\sigma}^2_1}$ represent the appropriate partial derivatives (all evaluated at $\bar{x}_n = E(\bar{x}_n)$, $\hat{\sigma}^2_n = \sigma^2_n$, $\bar{x}_1 = E(\bar{x}_1)$, and $\hat{\sigma}^2_1 = \sigma^2_1$), we have

$$\eta_{\hat{f}} = f\,, \tag{1.34}$$

and

$$\sigma^2_{\hat{f}} = \mathrm{var}(\bar{x}_n)\,f^2_{\bar{x}_n} + 2\,\mathrm{cov}(\bar{x}_n, \hat{\sigma}^2_n)\,f_{\bar{x}_n} f_{\hat{\sigma}^2_n} + \mathrm{var}(\hat{\sigma}^2_n)\,f^2_{\hat{\sigma}^2_n} \\ + \mathrm{var}(\bar{x}_1)\,f^2_{\bar{x}_1} + 2\,\mathrm{cov}(\bar{x}_1, \hat{\sigma}^2_1) f_{\bar{x}_1} f_{\hat{\sigma}^2_1} + \mathrm{var}(\hat{\sigma}^2_1)\,f^2_{\hat{\sigma}^2_1}\,. \tag{1.35}$$

These results show that, asymptotically, the estimates of interest have a limiting normal distribution. These distributions have as their means the functions evaluated at the point where the sample moments are equal to their expected values, that is, the population moments.

Also, the limiting distributions have variances which go to zero as $\frac{1}{N}$.

Finally, it should be emphasized that the functions we are estimating are the results based on the theoretical values of Gumbel. For example, when we say that the mean of the limiting normal distribution of the false alarm estimate is

$$\eta_{\hat{P}_{fa}} = P_{fa},$$

P_{fa} is equal to $\frac{1}{n} e^{-a_n(x_t-u_n)}$, not the exact false alarm error we would have if we knew the actual density function of the noise. In other words, it should always be remembered that we are estimating statistically not what we really want, but rather an approximation to what we want.

1.4 Numerical Examples

Several examples will now be given so that it can be seen what type of confidence intervals these densities produce. In all cases except one, the confidence intervals will be functions of the actual parameters, which are, of course, unknown. Therefore, in order to evaluate the intervals, it is necessary to assume the actual form of the probability density functions.

The one exception to the above is the confidence interval on α. We could find a confidence interval on α directly by using equations (1.29) and (1.30). However, if we obtain intervals on $\frac{1}{\alpha}$, using the fact that it is asymptotically unbiased along with eq. (1.27), our results will be simpler.

That the confidence interval on $(\frac{1}{\hat{\alpha}})$ does not depend on any parameters can be shown as follows:

$$\Pr[a\alpha \leq \hat{\alpha} \leq b\alpha] = \Pr[\frac{1}{b\alpha} \leq \frac{1}{\hat{\alpha}} \leq \frac{1}{a\alpha}] = \frac{1}{\sqrt{2\pi(\frac{1.1}{N\alpha^2})}} \int_{1/b\alpha}^{1/a\alpha} e^{-[1/\hat{\alpha}-1/\alpha]^2 / 2\frac{1.1}{N\alpha^2}} d(1/\hat{\alpha})$$

$$= \frac{1}{\sqrt{2\pi}} \int_{(\frac{1}{b}-1)/\sqrt{1.1/N}}^{(\frac{1}{a}-1)/\sqrt{1.1/N}} e^{-\frac{x^2}{2}} dx = \varphi\left(\frac{(\frac{1}{a})-1}{\sqrt{1.1/N}}\right) - \varphi\left(\frac{(\frac{1}{b})-1}{\sqrt{1.1/N}}\right),$$

where $\varphi(x)$ is the cumulative N(0, 1) distribution (i.e.,

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy).$$

Table (2) below summarizes some numerical results for N = 20.

TABLE (2)

Confidence Intervals on $\hat{\alpha}$

a	b	$\Pr[a\alpha \leq \hat{\alpha} \leq b\alpha]$
.5	2	$\varphi(4.27)-\varphi(-2.135) = .985$
.67	1.5	$\varphi(2.135)-\varphi(-1.41) = .899$
.8	1.25	$\varphi(1.07) - \varphi(-.854) = .661$

To obtain confidence intervals on $\hat{u}_n$, it is necessary to assume a specific distribution. Therefore, using an N(0, 1) with n = 100, we obtain the parameters $\alpha_n = 2.66$ and $u_n = 2.33$. Proceedings as above, we can show that the

$$\Pr[au_n \le \hat{u}_n \le bu_n] = \phi[(b-1)(.439\ \alpha_n u_n)] - \varphi[(a-1)(.439\alpha_n u_n)].$$

Specifically, with the above numbers and N = 20, we have

$$\Pr[.9u_n \le \hat{u}_n \le 1.1u_n] = \varphi(2.73) - \varphi(-2.73) = .99367,$$

and

$$\Pr[.8u_n \le \hat{u}_n \le 1.2u_n] = \varphi(1.37) - \varphi(-1.37) = .829 .$$

It can be seen that u_n can be estimated much more accurately than α_n.

As a final example, we will use equations (1.32) and (1.33) to obtain a confidence interval on the false-alarm estimate. From eq. (1.31), we obtain

$$P_{fa} = \frac{1}{n}\ e^{-\alpha_n(x_t - u_n)} ,$$

$$P_{fa\,\bar{x}} = \left.\frac{\partial \hat{P}_{fa}}{\partial \bar{x}}\right|_{\substack{\bar{x} = u_n + \frac{\gamma}{\alpha_n} \\ \hat{\sigma}_n = \frac{\pi}{\sqrt{6}\,\alpha_n}}} = \frac{\alpha_n}{n}\ e^{-\alpha_n(x_t - u_n)} ,$$

and

$$P_{fa\,\hat{\sigma}^2} = \left.\frac{\partial \hat{P}_{fa}}{\partial \hat{\sigma}^2}\right|_{\substack{\bar{x} = u_n + \frac{\gamma}{\alpha_n} \\ \hat{\sigma}_n = \frac{\pi}{\sqrt{6}\alpha_n}}} = \frac{1}{n}\frac{3\alpha_n^3}{\pi^2}\left(x_t - u_n - \frac{\gamma}{\alpha_n}\right) e^{-\alpha_n(x_t - u_n)} .$$

Also, we have[1.19]

$$\mathrm{var}(\bar{x}_n) = \frac{\sigma_n^2}{N} ,$$

$$\mathrm{var}(\hat{\sigma}_n^2) = \sigma_n^2\ \frac{\beta_2 - 1}{N} + O\left(\frac{1}{N^2}\right),$$

and

$$\mathrm{cov}(\bar{x}_n, \sigma_n^2) = \frac{\mu_3}{N} .$$

Inserting these expressions in equations (1. 32) and (1. 33), along with the appropriate moments calculated from eq. (1. 20), and forming the ratio of the mean of $\hat{P}_{fa}$ to its standard deviation, we obtain

$$\frac{P_{fa}}{\sigma_{\hat{P}_{fa}}} = \left[\frac{6N}{\pi^2\left[1+\frac{2.404(6)^{3/2}}{\pi^3}\frac{(x_t-u_n)}{\sigma_n} + \frac{1.1(x_t-u_n)^2}{\sigma_n^2}\right]} \right]^{1/2} .$$

We can now compute the confidence interval as follows:

$$P_r[a\,P_{fa} \leq \hat{P}_{fa} \leq b\,P_{fa}] = \varphi\left(\frac{(b-1)\,P_{fa}}{\sigma_{\hat{P}_{fa}}}\right) - \varphi\left(\frac{(a-1)P_{fa}}{\sigma_{\hat{P}_{fa}}}\right).$$

As a specific example, consider a Rayleigh random variable, that is, one whose probability density function is

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & x > 0 \\ 0 & x < 0 \end{cases} .$$

Using values of $\sigma = 1$, $x_t = 4.03$, $n = 100$, $u_n = 3.04$, $\alpha_n = 3.04$, and $N = 20$, we obtain $\frac{P_{fa}}{\sigma_{\hat{P}_{fa}}} = 1.31$, and

$$P_r[\frac{P_{fa}}{2} \leq \hat{P}_{fa} \leq 2\,P_{fa}] = \varphi\,(1.31) - \varphi\,(-.66) = .65,$$

and

$$P_r[\frac{P_{fa}}{3} \leq \hat{P}_{fa} \leq 3\,P_{fa}] = \varphi\,(2.62) - \varphi(-.87) = .804 .$$

1.5 Review

Before proceeding to the study of communication systems for which this method can be applied, it might be well to review what has been done in this chapter.

The fundamental theory for a nonparametric detector has been given. This theory is based on Gumbel's theory of extreme-values. The one restriction to the noise considered in this report is that it be of the exponential-type. This however, is not a necessary restriction, since other forms of Gumbel's asymptotes can be used.

Using this theory, estimates of those parts of the noise and signal-plus-noise distributions appropriate to probability of error measurement and threshold determination (i.e., the tails of the distributions) are formed. What remains to be seen is how it compares to optimum parametric detection in various communication systems, and also how it compares to other nonparametric schemes. It should again be stressed that the reason why it should perform nearly as well as parametric systems, and also possibly better than other nonparametric systems, is that it uses the amplitude information of the received samples as opposed to rank or other more qualitative information.

Finally, it should be pointed out that there are other ways to estimate probability functions besides using Gumbel's theory. However, these alternate methods have drawbacks to them which are eliminated by using EVT.

One such method is to construct bar graph-type estimates of the probability density function,[1.20] commonly called histograms.[1.21] However, histograms are only accurate around the center of the distribution, not around the tails, and in any communication system with low

error rates, as stated above, the tails are the most important part of the distribution. Also, histograms are not unique, since they depend upon the size of the amplitude windows that are used.

Another approach would be to use either empirical distribution functions, that is, functions G(x) which are defined such that

$$G(x) = \frac{1}{n} \sum_{i}^{n} U(x-x_i),$$

where U(x) is the unit step function and x_i, $i=1, \ldots n$, are the sample values, or functions of empirical distribution functions.[1.22, 1.23] However, empirical distribution functions are much more difficult to deal with analytically than is extreme-value theory, and they also have the same drawback as histograms in that they are most useful in the center of the distribution, not on the tails.

CHAPTER 2 - NONPARAMETRIC DETECTION USING EVT

2.1 The EVT Dectector

In this chapter, the object is to detect a low-pass binary signal in stationary additive noise (the one exception will be when impulse noise is introduced). In most examples considered here, we will attempt to decide between noise only or signal-plus-noise (i. e., an on-off system). In all cases, the apriori probabilities of transmission of the binary signals will be assumed to be equal.

It was pointed out previously that to estimate a probability of error, $P(\varepsilon)$, we would want n to equal $\frac{1}{P(\varepsilon)}$. However, since we do not know beforehand what probability of error exists, we do not know what value of n to use. Therefore, a trial-and-error procedure is employed.

Consider first the noise distribution. An initial value of n will be picked, say n_{1n}. Nn_{1n} samples will be taken and $\hat{u}_{1n}$ will be calculated. In this scheme, it is not necessary to use the estimate for α_n. This is desirable, since u_n can be estimated much more accurately than α_n.

At this point, it should be noted that the Nn_{1n} samples will be obtained during a learning period from a controlled noise distribution. That is, samples will be taken from a time waveform which is known not to contain signal.

The sample size is then increased from Nn_{1n} to Nn_{2n}, and $\hat{u}_{2n}$ is calculated. For $n_{2n} > n_{1n}$, u_{2n} will be greater than u_{1n}, as can be seen from the definition of u_n, eq. (1. 3). This procedure will be repeated, say, m times. m will be determined is as follows: Besides the control set of noise samples, we will need another control set, this time of signal-plus-noise. The same procedure is used on this control set, except that minimum instead of maximum values are used. That is, instead of obtaining an increasing sequence of $\hat{u}_n$'s,

we will now obtain a decreasing sequence of $\hat{u}_l$'s. The termination point occurs when the two sequences "intersect" (that is, where $\hat{u}_n \approx \hat{u}_l$). The point of intersection is the threshold of the system, and if the final sample sizes were Nn_{mn} and $Nn_{\ell l}$, then the estimates of the two types of error are

$$\hat{P}_{fa} = \frac{1}{n_{mn}}, \tag{2.1}$$

and

$$\hat{P}_{fd} = \frac{1}{n_{\ell l}} . \tag{2.2}$$

It might be wondered at this point why the expressions for the threshold and probability of error are different from those obtained in Chapter 1. Actually, these expressions are a special case of the results derived there. If, in the results of Chapter 1, u_l is set equal to u_n and both u_l and u_n are set equal to the threshold x_t, and if furthermore, it is noted that for equal apriori probabilities, $f_n(x_t) = f_{s+n}(x_t)$, the results of Chapter 1 reduce to the above results. The advantage of using this special case is that the results here, are, at least theoretically, exact (i.e., if we could estimate u_n and u_l exactly, the threshold and probabilities of error would be exact). This is not true in the more general case, since, as was pointed out before, the two Taylor series are only exact at the points u_n and u_l respectively.

At the start of the procedure, the change from one value of n to another value can be large. As the two sequences approach each other, the change in n can be made much smaller. However, since the values of $\hat{u}_n$ and $\hat{u}_l$ are only estimates of u_n and u_l respectively, and therefore not exact, there is no point in changing n by too small an amount.

Finally, if n becomes too large to fit Nn samples in the allotted learning period, we can always stop at some tolerable value and go back to the procedure given in Chapter 1, that is, estimate both the u's and the α's.

2.2 Examples

2.2.1 Narrowband FM for high SNR and Gaussian Noise

This is probably the simplest situation, since for high SNR, the output of a narrowband FM discriminator can be represented as just a signal in additive gaussian noise. The high SNR and small modulation index enables one to assume the number of errors that occur because of spikes can be neglected.[2.1] For this situation, as well as any other case in which the noise has a symmetric density and the signal-plus-noise distribution is just a shift in mean of the noise distribution, either the method of Chapter 1 or the method of this chapter gives theoretically exact results, as is demonstrated below.

For symmetric distributions, $\alpha_1 = \alpha_n$ for the same number of samples. Since the signal distribution is just a shift of the noise distribution, and since the α's are not affected by a shift, α_n of the noise density will equal α_1 of the signal-plus-noise density.

Also, for a symmetric density with mean equal to μ, $u_1 = 2\mu - u_n$. If the density is now shifted by an amount A, both u_1 and u_n will shift by that amount. Therefore, the relationship between u_n for the noise distribution, and u_1 for the signal distribution, is $u_1 = A - u_n$. Hence, from eq. (1.14),

$$x_t = \frac{u_1\alpha_1 + u_n\alpha_n + \ln\frac{\alpha_n}{\alpha_1}}{\alpha_1 + \alpha_n} = \frac{u_1 + u_n}{2} = \frac{A - u_n + u_n}{2} = \frac{A}{2} ,$$

which, of course, is the optimum threshold.

To obtain some feeling as to how well the parameters can be estimated for a normal density, calculations were made with the aid of a table of normal random variables with $n = 100$ and $N = 10, 20,$ and 40. The results are given below in Table (3).

TABLE (3)

Estimates of u_n and α_n from N(0, 1) Density

	N=10	N 20	N=40	Actual Value
$\hat{u}_n$	2.133	2.201	2.312	2.33
$\hat{\alpha}_n$	2.38	2.36	2.5	2.664

2.2.2 On-Off Systems (Noncoherent Detection)

If a noncoherent detector is used to detect a constant signal in additive gaussian noise, the densities at the output of the envelope detectors are Rayleigh for noise alone, and Rician for signal-plus-noise.[2.2] That is, the probability density function (pdf) for noise alone is

$$f_n(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} U(x) , \tag{2.3}$$

and the pdf for signal-plus-noise is

$$f_{s+n}(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} e^{-\frac{A^2}{2\sigma^2}} I_o \left(\frac{xA}{\sigma^2}\right) U(x) , \tag{2.4}$$

where U(x) is the unit step function.

The Rician density might at first seem to present a difficulty, since the portion of it that interest us is its left-hand tail, which goes to zero as x , not $e^{-|x|}$ as is required of exponential-type distributions. If, however, we are dealing with high SNR , a large portion of the left-hand tail of the Rician is dominated by an exponential behavior.[2.3] Specifically, for large SNR , the behavior of the Rician at the threshold will be exponential.

As an example, let us assume we have a Rician pdf. with parameters A=8 and σ =1, and a Rayleigh pdf with parameter σ =1. The optimum threshold, for equal apriori probabilities, is then 4.33 . Using a "good" value of $n=10^4$, we have $u_n = \alpha_n$ = 4.291 for the Rayleigh,

and $u_1 = 4.38$ and $\alpha_1 = 4.21$ for the Rician. If we insert these values into the formula for the threshold, eq. (1.14), we calculate $x_t = 4.35$, which is in excellent agreement with the optimum value.

To go to a more realistic situation, using random Rayleigh and Rician numbers, a system with $A = 7.35$, $\sigma = 1$, $n = 100$, and $N = 20$ was simulated. Table (4) summarizes the results. Note that $n = 100$ is not near $\frac{1}{P_{fa}}$ or $\frac{1}{P_{fd}}$. Yet the system performs well, showing that the initial estimates can be off and reasonable results can still be expected. Also note that we are working with a relatively high probability of error (i.e., we are not very far out on the tails). This is significant because the exponential approximations, eqs. (1.9) and (1.10), which result from eqs. (1.1) and (1.2), become more accurate the further out we are on the tails. In other words, for lower probabilities of error, we can expect even better results.

TABLE (4)

Comparison of Experimental Results with Optimum Values

Variable	Optimum and/or Actual Value	Experimental Result
u_n	3.04	3.11
α_n	3.04	2.84
u_1	5.1	4.95
α_1	2.65	3.01
x_t	4.03	4.04
P_{fd}	$.404 \times 10^{-3}$	$.64 \times 10^{-3}$
P_{fa}	$.33 \times 10^{-3}$	$.71 \times 10^{-3}$

If a square law detector had been used instead of an envelope detector, the two probability distributions would have been as follows:

$$f_n(z) = \frac{1}{2} e^{-z/2} U(z) ,$$

and

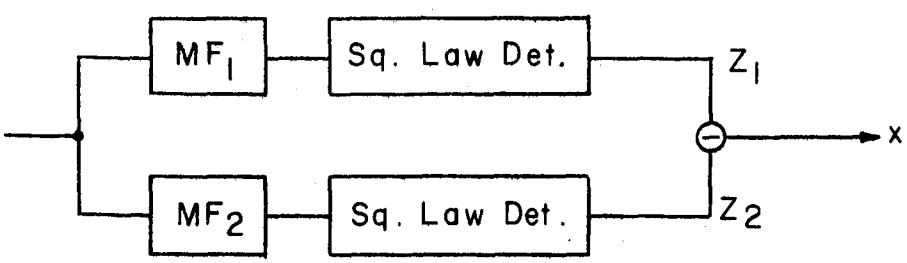
$$f_{s+n}(z) = \frac{1}{2}\, e^{-A^2/2}\, e^{-z/2}\, I_o(A\sqrt{z})\, U(z) \ .$$

As before, for high SNR, the signal-plus-noise distribution is basically an exponential at the threshold.

As an example, using exact values for u_n, a_n, u_1, and a_1, if we again take $A = 8$ and $n = 10^4$, the optimum threshold is 18.8, while the threshold predicted by eq. (1.14) is 18.5.

2.2.3 On-On-System

Consider an FSK system which transmits either $\omega_o + A_o$ or $\omega_o - A_o$. The signals are one of two orthogonal waveforms. The received signal will be the difference between the outputs of the two square-law detectors in Fig. (2.1). The input noise is white and gaussian.



Binary FSK Dectector

Figure (2.1)

In this type of system, the threshold is always zero, so that the only estimate required is that of the probability of error. To see whether Gumbel's theory applies, we need the density at the output. This density is basically the convolution of the densities of z_1 and z_2, the outputs of the two square law detectors. Assume signal one was sent. The density of z_1 is given by eq. (2.4), and the density of z_2 is given by eq. (2.3).

Therefore,

$$f(x) = \int_{-\infty}^{\infty} f_1(z-x)\, f_2(z)\, dz$$

$$= \frac{1}{4} \int_{-\infty}^{\infty} e^{-\frac{(z-x)}{2}}\, U(z-x)\; e^{-\frac{A^2}{2}}\; e^{-z/2}\; I_o(A\sqrt{z})\, U(z)\, dz$$

$$= \frac{1}{4}\, e^{x/2} \int_{a}^{\infty} e^{-z}\, e^{-A^2/2}\, I_o(A\sqrt{z})\, dz\,, \qquad (2.5)$$

$$\text{where } a = \begin{cases} x & x > 0 \\ 0 & x < 0 \end{cases} .$$

Making the change of variable $z = \frac{y^2}{2}$, we obtain

$$f(x) = \frac{1}{4}\, e^{x/2} \int_{\sqrt{2a}}^{\infty} y\, e^{-y^2/2}\, I_o(y\sqrt{\tfrac{A}{2}})\; e^{-\frac{A^2}{2}}\, dy\,, \qquad (2.6)$$

which can be integrated to yield

$$f(x) = \begin{cases} \frac{1}{4}\, e^{-A^2/4}\; e^{x}\; Q(A/\sqrt{2}\,,\, \sqrt{2x}) & x > 0 \\ \frac{1}{4}\, e^{-A^2/4}\; e^{x} & x < 0 \end{cases}\,, \qquad (2.7)$$

where $Q(a, b) = \int_{b}^{\infty} x\; e^{-(\frac{a^2+x^2}{2})}\; I_o(ax)\, dx$ is Marcum's Q-function.[2.4, 2.5]

By noting that $Q(a, 0) = 1$, and by again dealing with a large SNR (so that $A^2/4 >> x$ for x near zero), it can be seen that in the vicinity of the threshold $x_t = 0$, the density of x behaves as an exponential.

If both signals are of equal strength, the two types of errors will be the same. Also, since we already know the threshold, only one learning sequence of $\hat{u}$'s is necessary, and this sequence can stop whenever the value of $\hat{u}$ becomes zero or close enough to zero to give the required accuracy.

2.2.4 Impulse noise

When impulse noise is present, as in telephone lines, detection becomes much more difficult. This is because impulse noise is nonstationary and therefore difficult to model, and also because it is not always present (that is, it is not continuous in time).[2.6, 2.7, 2.8]

The problem of finding a good representative model will simply be ignored. The model used for a single impulse will be a mathematical delta function with random energy and a random time delay.

The second problem will be resolved by using two different thresholds. The first threshold will correspond to an impulse (or impulses) being absent during the bit interval, and the second threshold will be used when impulses are present.

The presence or absence of an impulse will be determined at the input of the system, not the output. This is because impulses at the input can usually be recognized on sight, whereas once they have been passed through a detection system, they basically act the same way a signal does, as an increase in the mean of the distribution (or, equivalently, as an increase in the DC component of the time waveform).

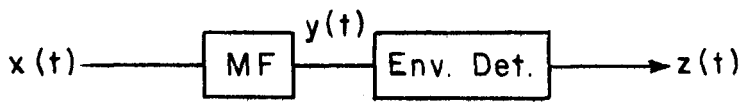
Picking the first threshold (without impulses) will be done as before. It must be remembered, however, that during the learning periods for noise and signal-plus-noise, only those samples should be used where an impulse did not occur.

Picking the second threshold is much more difficult. We now need two additional learning periods, one with " regular" (say gaussian) noise plus impulse noise, and one with signal-plus-" regular" noise plus impulse noise. However, this is not the difficulty. The trouble is that we do not know the height of any impulse that occurs, plus we do not know how many will occur in a given bit interval. We could probably find

some distribution for both the energy and the number of impulses in a given interval of time; however, due to the nonstationarity of impulse noise, these might not be meaningful. Therefore, what will be done will be to choose some average height and some average number of pulses per interval, based on information received during the learning period. In effect, what we are doing is simply shifting the first threshold upward by some "reasonable" amount.

More than two thresholds could be used, but for the sake of simplicity, we will use only two.

The specific system to be analyzed is given in Fig. (2.2). It is the same on-off system that was previously discussed, except that now, impulse noise has been added to the gaussian noise.



On-Off Detector

Figure (2.2)

$$x(t) = A \sin(\omega t + \theta) + n_g(t) + B\,\delta[t-(n-1+k)T] \qquad (2.8)$$

for $(n-1)T \leq t \leq nT$ and $0 < k < 1$, and where θ, k, and B are random variables.

The fact that only one impulse was used instead of a train of impulses does not lessen the generality of the example, because once the impulse goes through the matched filter, it becomes, as mentioned previously, just an increase in DC value. Therefore, putting one impulse in of unknown height is effectively the same as putting more than one impulse in, the sum of whose heights add up to the first one.

The output of the matched filter (matched to $\sin \omega t$), is

$$y(t) = u(t) \sin(\omega t + \theta) - v(t) \cos(\omega t + \theta), \text{ where}$$

$$u(T) = \int_0^T n_g(\tau) \cos(\omega\tau + \theta) d\tau + B \cos(\omega (n-1+k)T + \theta) = N_g + N_{impulse},$$

and

$$v(T) = \frac{AT}{2}\delta + \int_0^T n_g(\tau) \sin(\omega\tau + \theta) d\tau + B \sin(\omega(n-1+k)T + \theta)$$

$$= \frac{AT}{2}\delta + N_g' + N'_{impulse} . \qquad \delta = \begin{cases} 1 & \text{signal present} \\ 0 & \text{signal absent} \end{cases} .$$

Therefore, the output of the envelope detector will, in either case, be Rician distributed. If signal is present, it will have a spectral component $\frac{AT}{2} + N'_{impulse}$, and if signal is absent, it will have a spectral component of $N_{impulse}$.

We have shown that for high SNR, the left-hand tail of a Rician behaves as an exponential near the threshold, and since the right-hand tail behaves as an exponential, we are justified in using Gumbel's theory.

Consequently, the first threshold we would use, i.e., the threshold when impulses are not present, would be found exactly as in the example on on-off systems in Section. 2.2.2 .

The threshold when impulses are present would be that determined by two Rician densities, with both spectral components being based upon some average value of B.

2.3 Comparison with Rank Test

The rank test is a nonparametric test[2.9] which has been frequently used as the basis for a nonparametric detector. The test works as follows: A learning period is available during which noise samples are taken. These samples form a control set. A set of unknown samples are then taken, all of which come from the same density, and are interordered with the control set such that the smallest sample is first and the largest sample is last. Each sample is now given a rank equal to its ordering position (i.e., the smallest sample has rank = 1, the second smallest has rank = 2, etc.).

After this has been done, the ranks of the unknown samples are added together, and if this sum is greater than a predetermined threshold, it is decided that signal is present.

The intuitive justification for this method is that if signal is absent, both the control set of samples and the received set of samples are from the same distribution, namely the noise distribution, and therefore they should order amongst each other fairly uniformly.

On the other hand, if signal is present, the unknown set should be shifted to the right due to its higher mean, and therefore the ranks in this case should be higher than the ranks in the first case.

The rank statistic has been proven to be asymptotically normal under both signal and no signal conditions.[2.10, 2.11] Since we need both densities in order to evaluate the system, their means and variances will now be computed.

Let the control set of noise samples consist of n x's, and let the unknown set consist of m y's. The x's will be assumed to come from an N(0, 1) distribution, and the y's from an N(A, 1) distribution.

Let T be the statistic we seek. That is, $T = \sum_{i=1}^{m} r_i$, where r_i is the rank of the ith y.

Define a new statistic W equal to the total number of x's that are less than y's . For example, if we have the sequence $x_1\ x_2\ y_1\ x_3\ y_2\ y_3\ x_4$, W would equal 8 , since there are two x's less than y_1 , and three x's less than y_2 and y_3 .

We will first compute the mean and variance of W , and then use them to obtain the mean and variance of T .

Let $s_{ij} = x_i - y_j$. Then s_{ij} has an N(-A, 2) density. Therefore, $Pr(x_i < y_j) = Pr(x_i - y_j = s_{ij} < 0) =$

$$\int_{-\infty}^{0} \frac{1}{\sqrt{4\pi}} e^{-(s_{ij}+A)^2/4} d s_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{A/\sqrt{2}} e^{-r^2/2} dr = \varphi\left(\frac{A}{\sqrt{2}}\right) .$$

Let $z_{ij} = \begin{cases} 1 & x_i < y_j \\ 0 & x_i > y_j \end{cases}$. Then $W = \sum_{i,j} z_{ij}$.

This yields

$$E(z_{ij}) = 1\,\varphi\left(\frac{A}{\sqrt{2}}\right) + 0\left[1 - \varphi\left(\frac{A}{\sqrt{2}}\right)\right] = \varphi\left(\frac{A}{\sqrt{2}}\right) ,$$

$$E(z_{ij}^2) = \varphi\left(\frac{A}{2}\right) ,$$

and therefore

$$var(z_{ij}) = \varphi\left(\frac{A}{\sqrt{2}}\right) \left[1 - \varphi\left(\frac{A}{\sqrt{2}}\right)\right] .$$

Similarly,

$$E(W) = E\left[\sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij}\right] = mn\,\varphi\left(\frac{A}{\sqrt{2}}\right),$$

and

$$var(W) = cov\left(\sum_{i}^{n} \sum_{j}^{m} z_{ij}, \sum_{h}^{n} \sum_{k}^{m} z_{hk}\right) = \sum_{i}^{n} \sum_{j}^{m} \sum_{h}^{n} \sum_{k}^{m} cov\,(z_{ij}, z_{hk}).$$

To obtain the $cov(z_{ij}, z_{hk})$, note that

$$E(z_{ij}\, z_{hk}) = 1 \times Pr[(x_i < y_j),\ (x_h < y_k)] + 0 \times Pr\,[\text{any other combination}]$$
$$= Pr[(x_i < y_j),\ (x_h < y_k)] .$$

To calculate this quantity, which changes with different combinations

of subscripts, there are four possible cases which must be considered.

(1) For $i = h$, and $j = k$, $P_1 = \varphi(\frac{A}{\sqrt{2}})$.

(2) For $i \neq h$, and $j \neq k$, $P_2 = \varphi^2(\frac{A}{\sqrt{2}})$.

(3) For $i = h$ and $j \neq k$, $P_3 = Pr[x_i < \min(y_j, y_k)]$.

(4) For $i \neq h$ and $j = k$, $P_4 = P_r[\max(x_i, x_h) < y_j]$.

Let us look at case (3). Let $s = \min(y_j, y_k)$. Then

$F(s) = 1 - [Pr(y > s)]^2 = 1 - [1-\varphi(s-A)]^2$, or

$$f(s) = 2[1-\varphi(s-A)] \frac{1}{\sqrt{2\pi}} e^{-(s-A)^2/2} .$$

This yields

$$Pr(x_i < s) = \int_{-\infty}^{\infty} f_{x_i}(x_i) dx_i \int_{x_i}^{\infty} f_s(s)\, ds =$$

$$\int_{-\infty}^{\infty} f_{x_i}(x_i)\, dx_i \int_{x_i}^{\infty} 2[1-\varphi(s-A)] d\,\varphi(s-A) =$$

$$\left[-\int_{-\infty}^{\infty} f(x_i)\, d\, x_i [1-\varphi(s-A)]^2\right]_{x_i}^{\infty} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [1-\varphi(x_i-A)]^2 e^{-x_i^2/2} dx_i = P_3.$$

To calculate P_4, we must compute the probability $Pr[\max(x_i, x_h) < y_j]$. The derivation is very similar to the above one and will not be given. The end result is that $P_4 = P_3$. Therefore, we have

$$cov(z_{ij}, z_{hk}) = \begin{cases} \varphi(\frac{A}{\sqrt{2}})[1-\varphi(\frac{A}{\sqrt{2}})] & i=h, \quad j=k \\ 0 & i \neq h, \quad j \neq k \\ P_3 - \varphi^2(\frac{A}{\sqrt{2}} & i=h, \quad j \neq k \text{ or} \\ & i \neq h, \quad j = k \end{cases}$$

Finally,

$$var(W) = \sum_i^n \sum_j^m \sum_h^n \sum_k^m cov(z_{ij}, z_{hk}) =$$

$$\varphi(\frac{A}{\sqrt{2}})[1-\varphi(\frac{A}{\sqrt{2}})]\, mn + [P_3 - \varphi^2(\frac{A}{\sqrt{2}})][nm(m-1) + mn(n-1)] .$$

P_3 cannot be evaluated in closed form and is done numerically for each individual case.

It now remains to express T in terms of W and to obtain its mean and variance.

Let r_i be the rank of the ith y . This means there are $r_i - i$ x's less than y_i . This can be seen by noting that, if r_i is the rank of y_i , there are $r_i - 1$ total samples less than y_i , and i-1 y's less than y_i , thus leaving $r_i - 1 - (i-1) = r_i - i$ x's less than y_i .

Now we sum $r_i - i$ from 1 to m .

$$\sum_{i=1}^{m} (r_i - i) = T - \frac{m(m+1)}{2} ,$$

since

$$\sum_{i=1}^{m} r_i \triangleq T.$$

But

$$\sum_{i=1}^{m} (r_i - i) = W ,$$

so we have

$$T = W + \frac{m(m+1)}{2} ; \tag{2.9}$$

It now follows that

$$E(T) = \frac{m(m+1)}{2} + E(W) = \frac{m(m+1)}{2} + mn\varphi\left(\frac{A}{\sqrt{2}}\right), \tag{2.10}$$

and

$$\mathrm{var}(T) = \mathrm{var}(W) . \tag{2.11}$$

For noise only, A = 0 , and the above results reduce to

$$E(T) = \frac{m(m+1)}{2} + \frac{mn}{2} = \frac{m(n+m+1)}{2} , \tag{2.12}$$

and

$$\mathrm{var}(T) = \frac{mn}{4} + \frac{1}{12}[nm(m+n-2)] = \frac{1}{12}[nm(m+n+1)]. \tag{2.13}$$

When using these results, it must be remembered that they are only true asymptotically, that is, for n and m very large.

Before we can compare this method of detection with the EVT detector proposed in Chapter 1, the differences between the two techniques must be recognized and reconciled in some way in order to come up with a meaningful comparison.

The rank test starts out with n control samples of noise, and then makes a decision of noise or signal on the basis of m test samples. On the other hand, the method based upon EVT uses Nn_n control noise samples and Nn_1 control signal-plus-noise samples, but then makes its decision on just one test sample. Furtheremore, note that the rank test has to be used as a radar-type detector, that is, a detector which picks a threshold which satisfies one type of error requirement, say the false alarm probability, and accepts the resulting false dismissal error.[2.12] This is because the optimum threshold comes from a likelihood-ratio, whereas the rank test was obtained independently of the likelihood-ratio . In contrast, one of the main objects of the method proposed here is to estimate the optimum threshold.

Therefore, the following method of comparison will be used: The learning period of the rank test will consist of $n = N(n_n+n_1)$, the number of samples in both learning periods of the EVT test. However, to compensate for the fact that the rank test uses m test samples, as opposed to only one for the EVT test, the noise power of the rank detector will be increased by a factor of m (i. e., the voltage SNR of the rank test will be decreased by a factor of $\frac{1}{\sqrt{m}}$ from that of the EVT test). This is reasonable, because in order to take m times as many independent samples in the same one-bit interval, the bandwidth of the system using the rank test must be increased by a factor of m . But this is just another way of saying that the noise power is increased by a factor of m .

Finally, to avoid the problem of the two systems arriving at thresholds in different manners, a range of thresholds will be determined for the EVT detector (the range being chosen such that x_t lies within $\pm 10\%$ of u_n with a confidence coefficient of greater than 99%), and the rank detector will be given the same false alarm probability at each threshold. The corresponding false dimissal probabilities of the two will be compared.

Specifically, for the EVT test, we will take $n_n = n_1 = 10^3$, $N = 100$, and a voltage SNR = 6 . This means that for the rank test, $n = 2 \times 10^5$. Also for the rank test, we will take $m = 100$, which means the SNR for it equals $\frac{6}{\sqrt{m}} = .6$.

The values picked for the rank test results in $\varphi(\frac{A}{\sqrt{2}}) = \varphi(\frac{.6}{\sqrt{2}}) = \varphi(.425) = .664$, and $P_3 = .51$. Table (5) summarizes the results.

TABLE (5)

Comparison of Rank Test with EVT Test

EVT Threshold	Rank Threshold	P_{fa}	P_{fd} (rank)	P_{fd} (EVT)
2.78 (= .9u_n)	11.61×10^6	2.72×10^{-3}	8.2×10^{-4}	6.41×10^{-4}
3.09 (= u_n)	11.79×10^6	1.01×10^{-3}	2.42×10^{-3}	1.81×10^{-3}
3.4 = (1.1u_n)	11.96×10^6	3.37×10^{-4}	6.39×10^{-3}	4.66×10^{-3}

It can be seen that while the EVT detector gives a smaller false dismissal probability, the difference is not great. The significant result is a considerable savings in receiver complexity. This results because the rank detector must, for every decision, rank the 100 test samples amongst the 2×10^5 learning samples and sum their ranks. The EVT detector merely compares the amplitude of a single sample with the threshold.

Finally, it might be argued that this result was just a coincidence (i. e., given different values of n, m, and the rank SNR, the rank test might outperform the EVT test). This however, can be shown to be impossible. The reason is as follows: Consider what the optimum Neyman-Pearson detector would do if it knew the noise was gaussian. It would, after forming the likelihood-ratio and choosing the false alarm probability it wanted, decide a signal was present each time the threshold corresponding to that specific false alarm probability was exceeded. That is, the optimum Neyman-Pearson detector, in this situation, would perform precisely as the EVT detector. Therefore, by definition of a Neyman-Pearson detector, [2.13], for that false alarm rate, neither the rank detector nor any other detector working under the same conditions could do better than the EVT detector. At best, any other detector would do as well.

Summarizing, we can now see two advantages of the EVT test. One, the ease of using it after the learning period is over, and two, the possibility of actually doing as well as the optimum parametric detector.

CHAPTER 3 - FADING

3.1 Fixed Threshold

In this chapter, the detection of a fading signal will be considered in two ways. The first considers the use of a constant threshold, and the second considers an adaptive detector. The fading is assumed to be slow with respect to bit duration, and is assumed to be governed by a Rician probability distribution.[3.1, 3.2] A Rician density was chosen rather than the more common Rayleigh density in order to prevent very large fades. The reason for this, is that since we are using EVT, we want to apply it on the tails of the densities, and this will not be the case for deep fades.

The method used in this first section is only valid when we are dealing with a signal whose amplitude does not fluctuate greatly. This restriction is relaxed in Section 3.2. Because the variance is small, the probability density is approximated by a gaussian density. Specifically, we assume the fading has an amplitude distribution given by

$$f_A(A) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(A - A_o)^2}{2\sigma^2}} , \qquad (3\text{-}1)$$

where A_o is the spectral component of the fade. If we assume the density of signal-plus-noise remains functionally the same (undistorted) for any value of A, the only change being a shift in mean, the parameter α_1 will remain constant and the parameter u_1 will vary in the same way as the mean. This can be seen from the defining equations (1.5) and (1.6) for u_1 and α_1 by replacing F(x) with F(x-b), b being an arbitrary shift.

Therefore, if eq. (3.1) is the density of A, the density of u_1 is

$$f_{u_1}(u_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u_1 - u_o)^2}{2\sigma^2}} , \tag{3.2}$$

where u_o is the value of u_1 that corresponds to $A = A_o$. Averaging our estimate of the false dismissal probability over the fade gives

$$< P_{fd} > = \int_{-\infty}^{\infty} \frac{e^{a_1(x_t - u_1)}}{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u_1 - u_o)^2}{2\sigma^2}} du . \tag{3.3}$$

Note that while the expression $\frac{e^{a_1(x_t - u_1)}}{n}$ is an approximation which is only valid for certain values of u_1, we are averaging it over all possible values. This is the reason why it was important to have the possible fading amplitudes confined to a limited region. In other words, this is why we chose a Rician fading pdf with a large A_o and small σ instead of Rayleigh fading. Also note that the false alarm estimate does not have to be averaged, since the noise does not fade.

Equation (3.3) can be integrated by combining the exponentials and completing the square in the exponent. The result is

$$< P_{fd} > = \frac{1}{n} e^{a_1(x_t - u_o)} e^{\frac{a_1^2 \sigma^2}{2}} . \tag{3.4}$$

Next, we must establish the region of validity of eq. (3.4). We originally found where EVT was applicable by considering the Taylor series derivation. However, now that we have integrated over one of the parameters, it is no longer obvious where Gumbel's theory should apply.

To solve this problem, we first write that $G(x_t)$, the true average false dismissal probability, is

$$G(x_t) = \int_{-\infty}^{\infty} P_{fd}(A) f_A(A) dA . \tag{3.5}$$

Next, we set eq. (3.4) equal to eq. (3.5):

$$\frac{1}{n} e^{a_1(x_t-u_o)} e^{\frac{a_1^2 \sigma^2}{2}} = G(x_t) \quad . \tag{3.6}$$

Solving, we obtain

$$u_o - x_t = \frac{a_1 \sigma^2}{2} - \frac{1}{a_1} \ln \left[n\, G(x_t) \right] \quad . \tag{3.7}$$

However, $G(x_t)$ is unknown, so we must use some approximation to obtain the desired region. As a first approximation, let us assume that $\frac{1}{a_1} \ln \left[nG(x_t) \right]$ is negligable, thus leaving

$$u_o - x_t \approx \frac{a_1 \sigma^2}{2} \quad . \tag{3.8}$$

As a numerical example, consider a system operating in the presence of additive gaussian noise. The average false dismissal probability is then

$$\int_{-\infty}^{\infty} \varphi(x_t - A)\, f_A(A)\, dA = \int_{-\infty}^{\infty} \int_{-\infty}^{x_t} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-A)^2}{2}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(A-A_o)^2}{2\sigma^2}} dy dA. \tag{3.9}$$

This can be evaluated by making the following change of variables:

$$y = x + A + x_t \; .$$

Equation (3.9) then simplifies to

$$\int_{-\infty}^{\infty} \int_{-\infty}^{-A} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+x_t)^2}{2}} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(A - A_o)^2}{2\sigma^2}} dx\, dA =$$

$$\Pr\,(x < -A) = \Pr\,(x + A = z < 0) \quad .$$

Since $x = N(-x_t, 1)$, and $A = N(A_o, \sigma^2)$, we have

$$z = N(A_o - x_t, 1 + \sigma^2) \quad .$$

Therefore,

$$\Pr(z < 0) = \varphi\left(\frac{x_t - A_o}{\sqrt{1+\sigma^2}}\right) = 1 - \varphi\left(\frac{A_o - x_t}{\sqrt{1+\sigma^2}}\right). \qquad (3.10)$$

The results of the comparison are given in Table (6), computed with $a_1 = 3.944$ and $n = 10^4$.

Table (6)

Range of Validity of Fading Approximation

σ	$\frac{a_1 \sigma^2}{2} = \Delta$	Δ_{exp}	$\Delta(nG(x_t))$
1	1.68	1-2	.97 - 5.6
$\sqrt{2}$	3.36	3-5	1.43×10^{-2} - .58
2	6.72	8-11	1.96×10^{-6} - $.133 \times 10^{-2}$

In this table, Δ_{exp} is the experimental range of differences between u_o and x_t over which eq (3.4) was a "reasonable" estimate of eq. (3.10). "Reasonable" here means within a factor of 4 or 5.

$\Delta\left[nG(x_t)\right]$ is the corresponding range in the product $nG(x_t)$, which was implicitely assumed to be approximately unity in the derivation leading to eq. (3.8). It can be seen that this approximation can be somewhat violated without significantly affecting the results. We will improve upon this result later.

At this point, the question that must be resolved is how do we estimate the parameters u_o, a_1, and σ ? Barricelli proposed the following method when he was using EVT to study climatic variations.[3.3]

Consider again the double exponential distribution of Gumbel. If we average that density over the fade, we obtain

$$H(x) = \int_{-\infty}^{\infty} a_1 \, e^{a_1(x-u_1) - e^{a_1(x-u_1)}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(u_1-u_o)^2}{2\sigma^2}} . \qquad (3.11)$$

But this can be written as

$$H(x) = a_1 e^{+a_1 u_1 - e^{+a_1 u_1}} * f_{u_1}(u_1) \quad , \tag{3.12}$$

that is, a convolution between a double exponential and the density of u_1.

From Section 1.3, we know the mean of the density $f(z) = e^{z - e^z}$ is γ, while its variance is $\frac{\pi^2}{6}$. Since the mean and variance of a density which is obtained by convolving two other densities is just the sum of the individual means and sum of the individual variances of the two convolved densities, we have

$$E(x) = \frac{\gamma}{a_1} + u_o \quad , \tag{3.13}$$

and

$$\text{var}(x) = \frac{\pi^2}{6 a_1^2} + \sigma^2 \quad . \tag{3.14}$$

Therefore, if we again replace the population mean and population variance by the corresponding sample values, again computed from the N minima of n samples, we obtain

$$\bar{x}_{min} = \frac{\gamma}{\hat{a}_1} + \hat{u}_o \quad , \tag{3.15}$$

and

$$\text{var}(x_{min}) = \frac{\pi^2}{6 \hat{a}_1^2} + \hat{\sigma}^2 \quad . \tag{3.16}$$

Since we now have three unknowns and two equations, we need one more equation. This equation can be obtained by measuring the fade separately. That is, since the variance of u_1 is the same as the variance of A, we do not have to compute $\hat{\sigma}^2$ from extereme-values. Rather, we can compute the mean of each set of n samples, and then compute the variance of those means. Since we are assuming slow fading. (fading which is approximately constant over n bits), this will be a reasonable estimate of σ^2.

Finally, we have to determine how we are going to choose n and x_t. We had some idea of the region where eq. (3.4) is valid by using the approximation leading to eq. (3.8). However, in order to estimate the initial density parameters (that is, the parameters of the density before it is averaged over the fade), n must be in the vicinity of $1/F(x_t)$. Once we average over the fade, our final probability of error $G(x_t)$ will be considerably larger than $F(x_t)$, and therefore we know that the assumption $nG(x_t) \approx 1$ will always be violated. Keeping this in mind, the actual detection procedure will be as follows: pick an average false dismissal probability $G(x_t)$, pick a value of n reasonably larger than $1/G(x_t)$ (say by a factor of 100), and compute x_t from eq. (3.7):

$$x_t = u_o - \frac{a_1 \sigma^2}{2} - \frac{1}{a_1} \ln \left[nG(x_t) \right] .$$

At this value of x_t, compute $\hat{P}_{fa}$, and then calculate the total probability of error:

$$\hat{P}(\varepsilon) = \frac{1}{2} \left[\hat{P}_{fa} + G(x_t) \right] ,$$

which is valid for equal apriori probabilities.

Having done this, we will now pick another $G(x_t)$ and repeat the entire procedure. We will continue this over a wide range of x_t's and finally select that x_t which corresponds to the smallest value of the estimate of the minimum probability of error, $\hat{P}(\varepsilon)$.

It must be pointed out, however, that the accuracy of the value of x_t decreases as the variance of the fade increases. That is why this method is valid for signals which only fade over a small range of values.

3.2 Adaptive Detection

3.2.1 Adaptive EVT Detector

In order to design an adaptive detector using EVT, we must find a way to track the fade by means of changing u_1, since that is the only parameter that is changing with time. (Recall that the density of the signal-plus-noise is assumed to be undistorted by the fade, thus keeping α_1 constant, and u_n and α_n are constant since the noise does not fade).

The first change that must be made is in the learning period. We will assume, as in the previous section, that the fade is constant over n bits, and that we want Nn total bits in the learning period. However, whereas previously we were able to use N minimum values to estimate u_1 and α_1, we can not do that now because the N minima all come from different distributions.

We can, however, transform all the minima from different densities into new "minima" from the same distribution.

To see this, consider the following: the only differences in the densities from which the minima originally come are their mean values, due of course, to the assumption of the fade not causing any distortion in the densities. Now let us examine two densities which differ by only a shift. Consider $f_x(x)$ and $f_y(y) = f_x(x - A)$. If $\varphi_{x_1}(x)$ and $\varphi_{y_1}(x)$ are the densities of the respective minimum values, we have

$$\varphi_{x_1}(x) = n\left[1 - F_x(x)\right]^{(n-1)} f_x(x) \quad , \tag{3.17}$$

and

$$\varphi_{y_1}(y) = n\left[1 - F_y(y)\right]^{(n-1)} f_y(y) =$$

$$n\left[1 - F_x(x - A)\right]^{(n-1)} f_x(x - A) = \varphi_{x_1}(x - A) \; . \tag{3.18}$$

In other words, if two densities differ only by a shift, the densities of

their minima differ by that same shift.

Therefore, if we have N minima from N densities having means $A_1, A_2, \ldots A_N$, and we want to compute Gumbel's parameters which correspond to, say, the Nth density, we can do so by taking the N-1 remaining minima and adding to them the difference between A_N and the mean corresponding to the minimum we are changing. For example, if the ith minimum has a mean equal to A_i, the quantity $x'_{i_1} = x_{i_1} + (A_N - A_i)$ will be distributed in the same way as the minimum value corresponding to the Nth density, namely x_{N_1}.

We are now in a position to discuss the detection procedure. The parameters corresponding to the last (most recent) value of the fade will be calculated using the adjusted minimum values as discussed above. We will then use a slightly modified version of the usual method for estimating u_1 and α_1. This modification, which only involves the value of n that we use, will also apply when u_n and α_n are estimated. Previously, we were very careful about the value of n we used. Now, because the error probabilities will constantly change, we will arbitrarily pick a value of n such that $\frac{1}{n}$ is somewhere in the vicinity of the error probabilities we expect. Also, with respect to modifying the minimum values, since we do not know the actual values of the N means $A_1 \ldots A_N$, we will use the sample means of each set of n bits to estimate the true means.

Having the four parameters $\hat{u}_1$, $\hat{\alpha}_1$, $\hat{u}_n$, and $\hat{\alpha}_n$, we will use eq. (1.14) to estimate the threshold, and then use this threshold until we have detected n bits as signal. Of these n bits, some will be wrong, but for low probability of error, not many. We will then compute the mean of the new samples, use this as the new reference mean, and transform the original N minima of the learning period in precisely the same way as

was done in the learning period. Having a new value for $\hat{u}_1$ (a new value for $\hat{a}_1$ is not necessary), we will obtain a new threshold with which to detect until n more bits are decided as containing signal. This process will then be continually repeated.

There is one point that should be noted, however, and that is that we have changed one of our original assumptions. Whereas we originally assumed the fade was constant over n consecutive bits, now we are assuming the fade is constant over n bits which are detected as signal. This period will be longer than the original period, because for equal apriori probabilities of transmission, signal bits will only be transmitted half the time. Therefore, we are now assuming the fade remains constant over approximately 2n consecutive bits.

3.2.2 Confidence Intervals

Finding confidence intervals in the fading case is more difficult than in the case of a constant signal, but nevertheless, theoretical results can be obtained. It was previously shown, in the non-fading case, that all the estimates were asymptotically normal. It will now be shown that this result is valid in the fading case, except that these asymptotic results have to be averaged over another asymptotically normal density. To show this, consider again the way the N minimum samples are being altered. The jth minimum sample, $x_{j\,min}$, will become z_j, where

$$z_j = x_{j\,min} - \frac{1}{n}\sum_{i}^{n} x_{ij} + \frac{1}{n}\sum_{i}^{n} x_{iR} \qquad 1 \leq j \leq N,$$

and where the two summations are the estimates of the means of the jth fading amplitude of the learning period, and of the reference fading amplitude respectively.

There are two situations to consider. The first is the case when $N = R$, which is the learning period; the second is when $N > R$, where the detector is in the process of adapting its threshold.

In the first case, it will be noticed that the two summations cancel for $j = N = R$. This is because the last minimum is a perfect reference in the learning period without any adjustment of its mean. On the other hand, for all $j \leq N - 1$, the term $\frac{1}{n} \sum_{i}^{n} x_{iR}$ acts as a constant. It is also correlated to z_N, since $z_N = x_{N\,min}$ is the minimum of the samples composing the sum $\sum x_{iR}$ (recall $R = N$). Therefore, if we assume that $\sum x_{iR} = \text{const}$, and neglect the last minimum z_N, the remaining N-1 adjusted minima will all be independent. They are also identically distributed, as can be seen by recalling that the signal-plus-noise density was assumed not to become distorted as the signal faded. Therefore, the only difference between any of the x's is their mean . But the z's were formed specifically to adjust for this difference in means, so that the z's end up all having the same distribution. This, of course, assumes that $E\left[\sum x_{iR}\right] = A_R$, which will not be the case if there are errors present. However, as mentioned above, for low error rates, very few bits will be in error, and they will be ignored. If greater accuracy is desired, the methods of Chapter 4 can be used.

Because of this simplification, when we compute the confidence intervals for the learning period, we will make our proofs significantly easier if we only use the first N - 1 minima, and disregard the last one.

This is not necessary when the detector is out of the learning period, because, with $R > N$, the sum $\frac{1}{n} \sum_{i}^{n} x_{iR}$ acts as a constant for all N adjusted minima, and all N z_j's have the same density.

Therefore, in the derivation that follows, N minima will be used,

but it should be changed to N - 1 if the results are used for the learning period.

Because the variables z_i are independent and identically distributed, given that $\frac{1}{n}\sum_{i}^{n} x_{iR}$ = const, we are able to use the previous results given in Chapter 1, and then average over the distribution of $\frac{1}{n}\sum_{i=1}^{n} x_{iR}$, which will be asymptotically normal. However, to use our previous results, we need the first four central moments of the z_j's.

To this end, consider the non-constant part of z_j, $x_{j\,min} - \frac{1}{n}\sum_{i}^{n} x_{ij}$. This is a sum of dependent random variables, since the last n terms are all correlated with $x_{j\,min}$. If we assume the x_{ij}'s have means and variances equal to n_j and σ_j^2 respectively, then

$$E\left[z_j \middle| \frac{1}{n}\sum_{i}^{n} x_{iR} = \text{const}\right] = u_{1j} - \frac{\gamma}{a_{1j}} - n_j + \frac{1}{n}\sum_{i}^{n} x_{iR} ,$$

and

$$\text{var}\left[z_j \middle| \frac{1}{n}\sum_{i}^{n} x_{iR} = \text{const}\right] = \frac{\pi^2}{6a_{1j}^2} + \frac{\sigma_j^2}{n} - \frac{2}{n}\sum_{i=1}^{n} \text{cov}\,(x_{j\,min}, x_{ij}).$$

As is usually the case in calculating confidence intervals, we need the actual values of the unknown parameters, namely the n_j's, σ_j's, etc. In fact, we will see below that we need the actual densities.

In order to calculate the covariances, we need

$$\text{cov}\,(x_{j\,min}, x_{ij}) = E(x_{j\,min}, x_{ij}) - E(x_{j\,min}) E(x_{ij}).$$

Since x_{ij} is any sample from the n samples from which $x_{j\,min}$ is the minimum, we can calculate the cov $(x_{j\,min}, x_{ij})$ as follows:

Let $y_{1j} \ldots y_{nj}$ be the ordered set of x_{ij}, with y_{1j} the smallest and y_{nj} the largest. Then

$$\text{cov}\,(x_{j\,min}, x_{ij}) = \sum_{k=1}^{n} \text{cov}\left[x_{j\,min}, y_{kj} \middle| x_{ij} = y_{kj}\right] \Pr\left[x_{ij} = y_{kj}\right] .$$

However, the probability that x_{ij} is the kth order statistic is $\frac{1}{n}$.

Therefore,

$$\operatorname{cov}(x_{j\,min}, x_{ij}) = \frac{1}{n} \sum_{k=1}^{n} \operatorname{cov}\left[x_{j\,min}, y_{kj} \middle| x_{ij} = y_{kj}\right] .$$

The covariances in the above sum can be calculated, at least in theory, by considering the joint density of the kth order statistic with the first order statistic. In general, the density of the ith and j^{th} order statistics is given by[3.4]

$$f(x_i, x_j) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} F(x_i)^{i-1} \left[F(x_j) - F(x_i)\right]^{j-i-1} F(x_j)^{n-j} f(x_i) f(x_j),$$

$$\text{for} \qquad -\infty < x_i \leq x_j < \infty .$$

We can similarly calculate the third and forth central moments, and therefore use all the results arrived at previously. That is, since the confidence intervals for functions of the sample mean and sample variance of independent identically distributed random variables only depend upon the first four central moments of the distribution, which we can (in principle) calculate, we can (in principle), calculate the desired confidence intervals.

Finally, we must remember that all these results are conditioned on $\frac{1}{n} \sum x_{iR}$ = const, and therefore our answer has to be averaged over its distribution, which, as pointed out above, is asymptotically normal.

3.2.3 Computer Simulated Results

In order to obtain some numerical verification for this method, two computer simulations were run. The system simulated was an on-off system with a matched filter and an envelope detector. The signal was a carrier whose amplitude was fading, and the noise was gaussian. Therefore, at the output of the system, the density of the noise was Rayleigh (σ = 1), while the density of signal-plus-noise was Rician (σ = 1)

with time varying parameter A(t).

The mean value of the fade in the first simulation was 5, and the mean value was 6 in the second simulation. In both cases, n = 50 and N = 20.

Note that this is not the case we analyzed, since we assumed the signal-plus-noise density remains undistorted as the signal fades, and this is not true of a Rician density. We will see however, that the system will nevertheless perform well, indicating, as has been the case in previous examples, that the assumptions are not critical.

The slow fade was simulated by a waveform which was constant over n signal decisions, and then the next fade was correlated to the ten previous values of the fade using a correlation coefficient of approximately 90% between any two adjacent fading levels.

The results of the first simulation are shown in Figs. (3.1) and (3.2). Figure (3.1) illustrates how this system tracks the fade, and compares this result with an illustration of how the optimum system would track the fade. It can be seen that even though the fade becomes quite deep, the system still performs well, and still recovers when the amplitude increases.

Figure (3.2) compares the number of errors this system makes with the number of errors the optimum system would make. In both figures, it should be noted that the optimum (or parametric) system employed here is one that knows both the exact values of the reference fade plus the actual densities of noise and signal-plus-noise. In reading Fig. (3.2), each horizontal line between two adjacent abscissa values is the number of errors made between threshold changes.

Also, in both figures, the first abscissa point corresponds to the

learning period, so that point (2) on the horizontal scale corresponds to the first threshold adjustment.

Figures (3.3) and (3.4) are the corresponding results of the second simulation. There is, however, one point to note regarding Fig. (3.3). Whereas in Fig. (3.1) we obtained the worst results for deep fades, in Fig. (3.3) we have the worst results when the amplitude was large (that is, when the signal did not fade deeply). The explanation is that when the mean of the fading amplitude was changed from 5 to 6, no corresponding change was made in the value of n. In other words, when the signal had a large amplitude in the second simulation, the value of $n = 50$ was too small to give an accurate approximation to the density functions at the threshold.

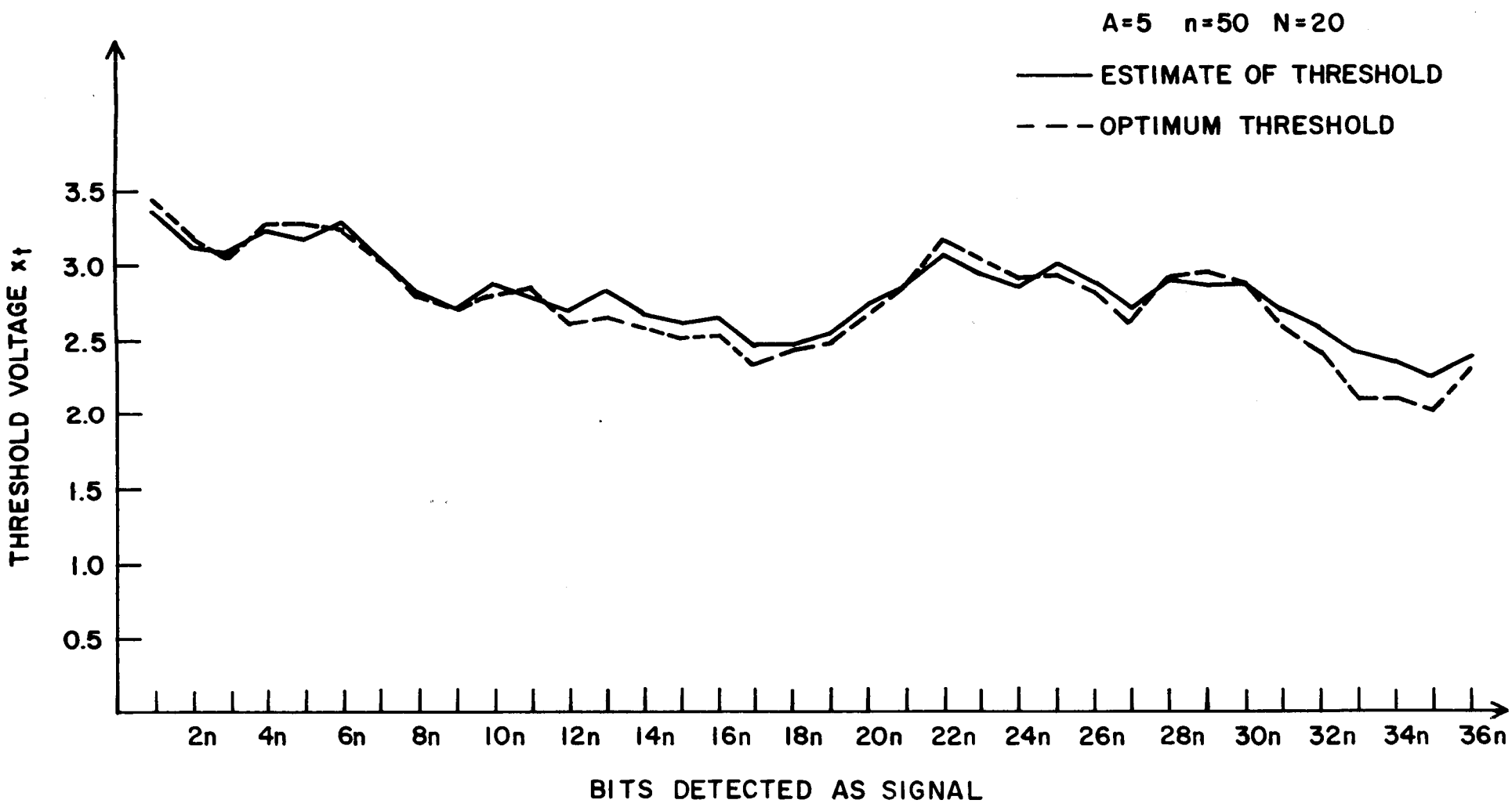


Figure (3.1) Comparison of Optimum and Estimated Thresholds

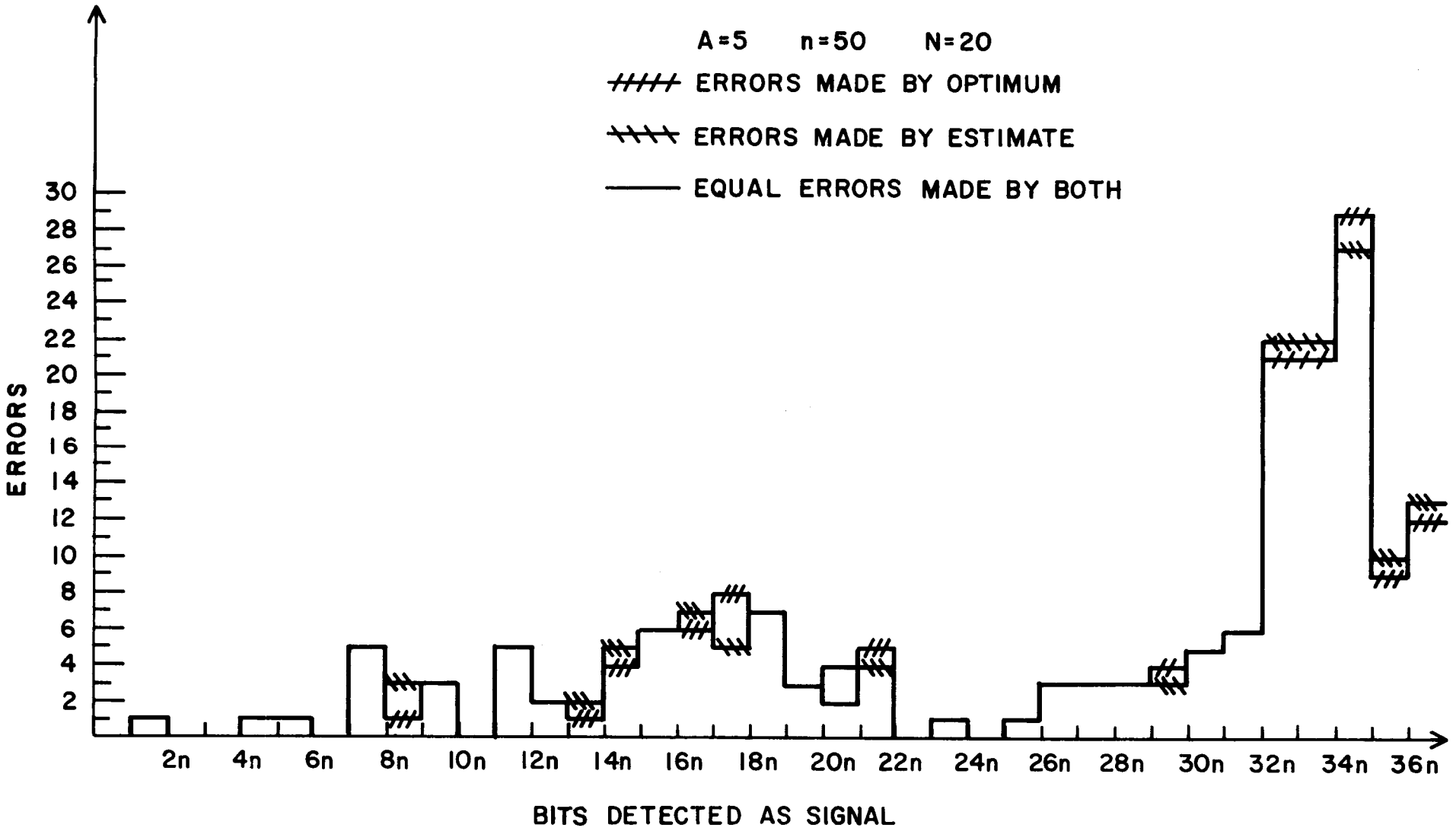


Figure (3.2) Comparison of Number of Errors made by Optimum and EVT Detectors

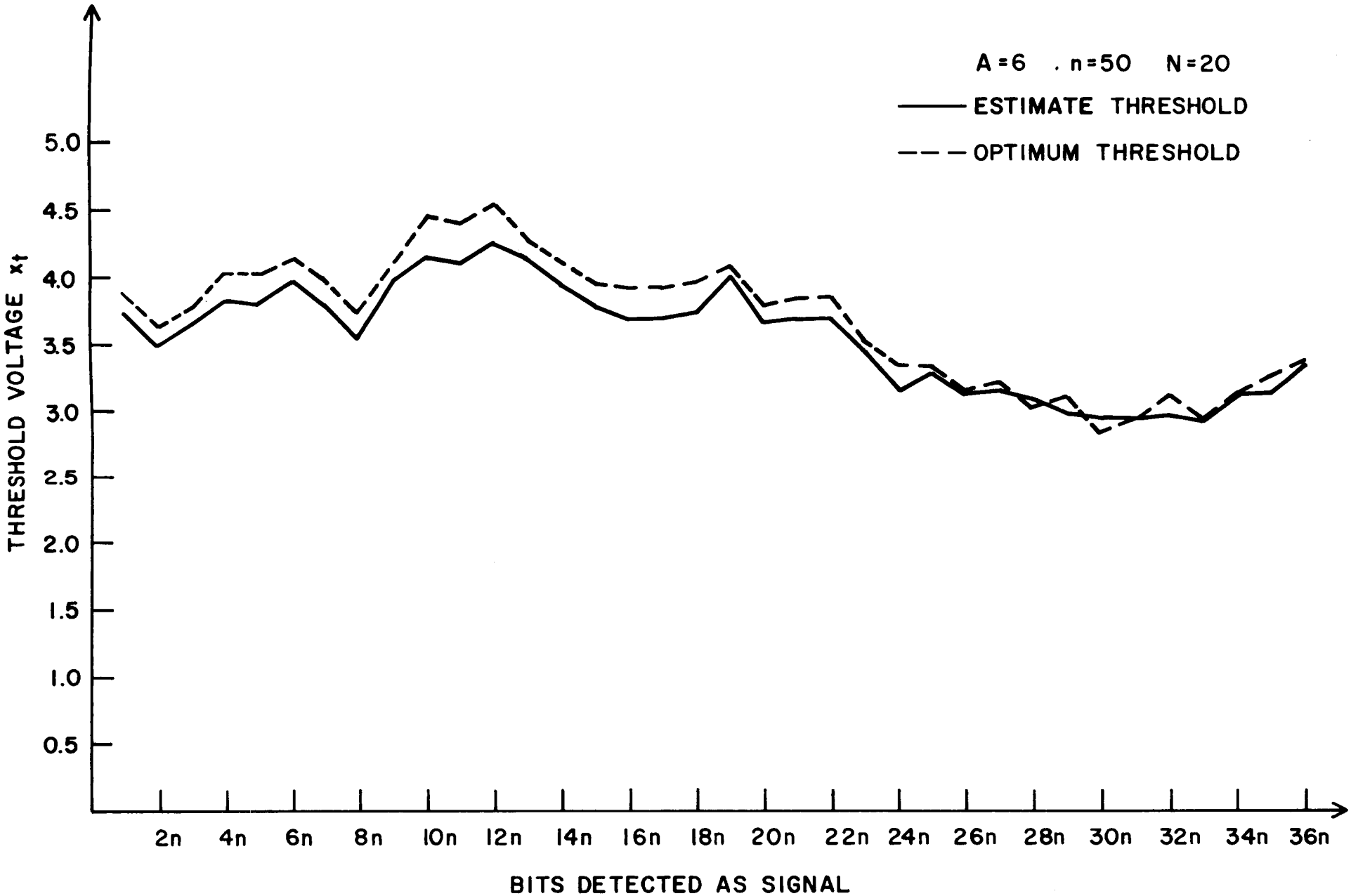


Figure (3.3) Comparison of Optimum and Estimated Thresholds

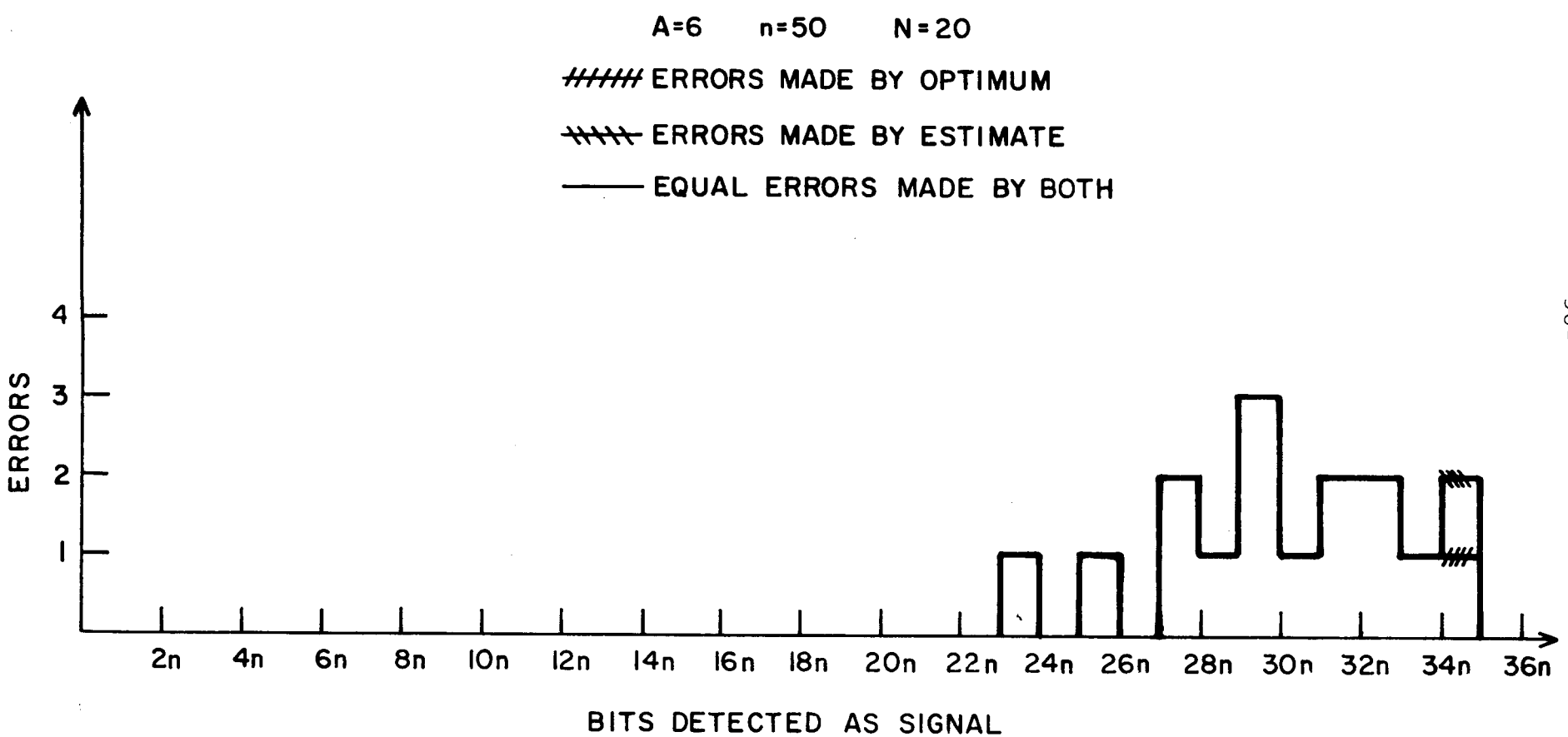


Figure (3.4) Comparison of Number of Errors made by Optimum and EVT Detectors

CHAPTER 4. - DECISION-DIRECTED MEASUREMENTS[4.1]

4.1 Elimination of Learning Period

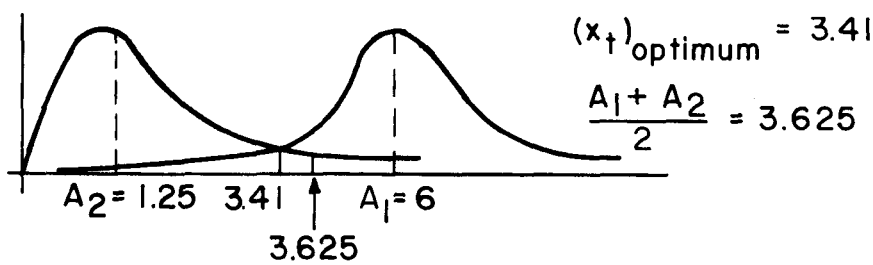
In all of the previous examples, the key factor which enabled us to detect signals, in noise having an unknown probability density, was the learning period, during which time the appropriate density functions were estimated. Because of the time consumed by the learning period, it is worthwhile to consider a scheme which does not require such a learning period.

In this new scheme, it will be necessary for the detector to estimate Gumbel's parameters (after an appropriate time delay) on the basis of its own decisions, some of which will be wrong. That is, we will consider a detector which makes its estimates of the parameters without the benefit of knowing that those bits which were detected as noise are, in fact, noise, and that those bits detected as signal-plus-noise do, in fact, contain signal.

The first thing we must do is to specify an initial threshold with which to begin detecting. We have to be careful here, because the estimates we make of Gumbel's parameters will be obtained from the samples we detect with this initial threshold. That is, no matter what threshold we pick (including the optimum threshold) we will make a number of incorrect decisions, and these incorrect decisions will affect our estimates. Obviously then, if we initially pick a very bad threshold, we will get worse estimates then if the threshold was better.

Since we are starting with no more information than that the densities considered have exponential-type tails, we certainly cannot expect to initially arrive at any optimum threshold. However, we can come up with a reasonable threshold by taking, say s samples, where s will be more

than 2nN, and computing the mean of those s samples. That this is a reasonable value to pick as a threshold can be seen as follows: Since we are dealing with large SNR, the optimum threshold will be somewhere out on the tails of both distributions. Likewise, if we take the average of the means of the two densities, this will also lie on the tails of both distributions. A specific example is shown in Figure (4.1). It can be seen that the two thresholds are reasonably close, and that starting with almost no knowledge whatsoever, the average of the population means makes a good initial threshold.



Sample Mean as Threshold Estimate

Figure (4.1)

The densities in Fig. (4.1) are Rayleigh with parameter $\sigma = 1$, and Rician with parameters $\sigma = 1$ and $A = 6$.

There is still one more problem. If we agree to use the average of the two means, the best way to estimate it would be by the average of the sample means. However, since in our initial s samples, we do not know how many of them came from each density, we cannot estimate the individual sample means.

What we can do, though, is compute the overall mean of the s samples. This will be a reasonable estimate of the average of the individual means if the apriori probabilities of transmission are equal.

Having decided then on an initial threshold, the next step will be to go back and detect the s samples we have accumulated. This is accomplished by calling all samples below the threshold, noise, and all those above the threshold, signal (plus noise).

It can now be seen how large s must be. Since we need nN samples to estimate u_1 and α_1, and the same number to estimate u_n and α_n, we must, once we start detecting the s samples, make at least nN noise decisions, and nN signal decisions. Because of the equal apriori probabilities, s will be approximately 2nN, but, except in an unusual case, will have to be larger.

Note that we have not yet decided how we are going to pick n. That the value of n we choose can be critical can be seen as follows: Consider our method for estimating the parameters. It is based on the sample mean and sample standard deviation of the maximum or minimum values. But once we determine, say, the maximum of n samples from samples which were below some threshold, we are automatically bounding the value that the maximum sample can take. Suppose now, that the value of n chosen corresponds to a value of u_n which is greater than the threshold. This means that the estimate we make of u_n will have to be wrong. That is, since the N maxima we are using are all less than the threshold, their average is certainly less than the threshold. Therefore, since the estimate of u_n is this average minus the sample standard deviation times a positive constant (i.e., minus a positive number), $\hat{u}_n$ will necessarily be less than the threshold.

Now assume we choose an n which corresponds to a u_n less than the threshold, but "near" it. Since u_n is near the threshold, at least some of the maximum values we would obtain if there were no threshold

will probably be greater than the threshold. In other words, we will probably be using values we believe to be maxima from sets of n noise samples which are either not from the noise distribution (they would then be samples of the signal-plus-noise distribution which fell below the threshold and consequently were detected as noise), or from the noise distribution, but not the largest of n consecutive noise samples. The latter situation would arise if either some signal-plus-noise samples were among the n samples believed to be noise, or if a noise sample larger than the threshold was actually transmitted during the period when the n samples were collected.

The net affect of this, with respect to the resulting parameter estimates, is a value of $\hat{u}_n$ which is approximately equal to the threshold, and a value of $\hat{a}_n$ much larger than the true value a_n. This can easily be seen by the following reasoning: Since most or all the N maxima will be near the threshold, the sample mean of those maxima will be near the threshold. Also, for the same reason, the sample standard deviation of the maxima will be much smaller than it should be. Consequently, $\hat{u}_n$ will be approximately equal to the sample mean, and $\hat{a}_n$, being inversely proportional to the sample standard deviation, will be much too large.

Therefore, to ensure that most of the maximum values we use are correct, we pick n so that the corresponding u_n is quite a bit lower than the threshold. This, of course, will probably take a trial-and-error procedure, since we have no idea how to choose n initially.

Since all of the above reasoning applies to the minimum values of the signal-plus-noise density, the value of $\hat{u}_1$ should be reasonably larger than the threshold.

At this point, we have just about caught up to ourselves. Having

detected 2nN of the original s samples, and having estimated Gumbel's parameters, we can compute a new threshold, use that threshold to detect the remaining s - 2nN samples, and then start detecting new samples. As we are detecting incoming bits, we will stop after making either n signal decisions or n noise decisions, recalculate the appropriate parameters, and again change our threshold accordingly.

4.2 Confidence Intervals and Computer Simulated Results

In order to obtain confidence intervals, we need the distribution of the maximum and minimum of sets of n variables which are no longer all signal-plus-noise samples, or all pure noise samples. The n bits now contains samples from both distributions, and to find the density functions of the extreme-values from the combined set, we proceed as follows:

Let z be an arbitrary sample. If we have n z's which we have detected to be noise, then, by definition, all the z's will be below the threshold x_t.

Therefore, to find the distribution of z, we write

$$\Pr\left[z \leq x \mid z \leq x_t\right] = \frac{\Pr\left[z \leq x,\ z \leq x_t\right]}{\Pr\left[z \leq x_t\right]} = \frac{\Pr\left[z \leq \min(x, x_t)\right]}{\Pr(z \leq x_t)} .$$

But,

$$\Pr\left[z \leq \min(x, x_t)\right] = \Pr\left[z \leq \min(x, x_t) \mid z = \text{signal}\right] \Pr\left[z = \text{signal}\right]$$
$$+ \Pr\left[z \leq \min(x, x_t) \mid z = \text{noise}\right] \Pr\left[z = \text{noise}\right]$$

$$= \begin{cases} F_{s+n}(x)\left(\frac{1}{2}\right) + F_n(x)\left(\frac{1}{2}\right) & x \leq x_t \\ F_{s+n}(x_t)\left(\frac{1}{2}\right) + F_n(x_t)\left(\frac{1}{2}\right) & x > x_t \end{cases} .$$

Also,

$$\Pr(z \leq x_t) = \Pr(z \leq x_t \mid z = \text{signal}) \Pr(z = \text{signal})$$

$$+ \Pr(z \leq x_t \mid z = \text{noise}) \Pr(z = \text{noise})$$

$$= F_{s+n}(x_t)\left(\tfrac{1}{2}\right) + F_n(x_t)\left(\tfrac{1}{2}\right).$$

Therefore,

$$\Pr\left[z \leq x \mid z \leq x_t\right] = \begin{cases} \dfrac{F_n(x) + F_{s+n}(x)}{F_n(x_t) + F_{s+n}(x_t)} & x \leq x_t \\ 1 & x > x_t \end{cases}.$$

The distribution of z_n, the maximum of n z's, is given by

$$F_{z_n}(x) = \left[\Pr[z \leq x \mid z \leq x_t]\right]^n = \begin{cases} \left[\dfrac{F_n(x) + F_{s+n}(x)}{F_n(x_t) + F_{s+n}(x_t)}\right]^n & x \leq x_t \\ 1 & x > x_t \end{cases}. \quad (4.1)$$

In order to obtain a result in a more useful form, we substitute the exponential approximations we have been using for $F_n(x)$ and $F_{s+n}(x)$ into the above equation. That is,

$$F_{z_n}(x) = \begin{cases} \dfrac{\left[1 - \dfrac{e^{-a_n(x-u_n)}}{n} + \dfrac{e^{a_1(x-u_1)}}{n}\right]^n}{\left[1 - \dfrac{e^{-a_n(x_t-u_n)}}{n} + \dfrac{e^{a_1(x_t-u_1)}}{n}\right]^n} & x \leq x_t \\ 1 & x > x_t \end{cases}. \quad (4.2)$$

However, we know that

$$\lim_{n \to \infty}\left[1 + \frac{y}{n}\right]^n = e^y.$$ [4.2]

Therefore, for large n, we have

$$F_{z_n}(x) = \begin{cases} \dfrac{e^{-\left[e^{-a_n(x-u_n)} - e^{a_1(x-u_1)}\right]}}{e^{-\left[e^{-a_n(x_t-u_n)} - e^{+a_1(x_t-u_1)}\right]}} & x \leq x_t \\ 1 & x > x_t \end{cases}. \quad (4.3)$$

Also,

$$f_{z_n}(x) = \left[a_n e^{-a_n(x-u_n)} + a_1 e^{a_1(x-u_1)}\right] F_{z_n}(x)\left[1 - U(x-x_t)\right], \quad (4.4)$$

where $U(y)$ is the unit step function.

If we now consider those samples which have been detected as signal, we must find the density of the minimum of n z's which are all greater than the threshold. Its cumulative distribution is then given by

$$\Pr\left[z \leq x \mid z \geq x_t\right] = \frac{\Pr\left[z \leq x,\ z \geq x_t\right]}{\Pr\left[z \geq x_t\right]} = \frac{\Pr\left[x_t \leq z \leq x\right]}{\Pr\left[z \geq x_t\right]} .$$

Proceeding as before, we have

$$\Pr\left[x_t \leq z \leq x\right] = \Pr\left[x_t \leq z \leq x \mid z = \text{signal}\right]\Pr\left[z = \text{signal}\right]$$
$$+ \Pr\left[x_t \leq z \leq x \mid z = \text{noise}\right]\Pr\left[z = \text{noise}\right]$$
$$= \begin{cases} \frac{1}{2}\left[F_{s+n}(x) - F_{s+n}(x_t) + F_n(x) - F_n(x_t)\right] & x \geq x_t \\ 0 & x < x_t \end{cases} .$$

Also,

$$\Pr\left[z \geq x_t\right] = \Pr\left[z \geq x_t \mid z = \text{signal}\right]\Pr\left[z = \text{signal}\right] + \Pr\left[z \geq x_t \mid z = \text{noise}\right]\Pr\left[z = \text{noise}\right]$$
$$= \left[1 - F_{s+n}(x_t)\right]\frac{1}{2} + \left[1 - F_n(x_t)\right]\frac{1}{2} .$$

Therefore, we have

$$\Pr\left[z \leq x \mid z \geq x_t\right] = \begin{cases} \dfrac{F_n(x) - F_n(x_t) + F_{s+n}(x) - F_{s+n}(x_t)}{2 - F_n(x_t) - F_{s+n}(x_t)} & x \geq x_t \\ 0 & x < x_t \end{cases} , \quad (4.5)$$

and

$$1 - \Pr\left[z \leq x \mid z \geq x_t\right] = \begin{cases} \dfrac{2 - F_n(x) - F_{s+n}(x)}{2 - F_n(x_t) - F_{s+n}(x_t)} & x \geq x_t \\ 1 & x < x_t \end{cases} . \quad (4.6)$$

If we now substitute for $F_n(x)$ and $F_{s+n}(x)$ the exponential approximations, we obtain

$$1 - \Pr\left[z \le x \mid z \ge x_t\right] = \begin{cases} \dfrac{1 + \frac{e^{-a_n(x-u_n)}}{n} - \frac{e^{a_1(x-u_1)}}{n}}{1 + \frac{e^{-a_n(x_t-u_n)}}{n} - \frac{e^{a_1(x_t-u_1)}}{n}} & x \ge x_t \\ 1 & x < x_t \end{cases} . \quad (4.7)$$

Finally, the distribution of z_1, the minimum of n z's, is given by

$$F_{z_1}(x) = 1 - \left[1 - \Pr\left[z \le x \mid x \ge x_t\right]\right]^n = \begin{cases} 1 - \dfrac{e\left[e^{-a_n(x-u_n)} - e^{a_1(x-u_1)}\right]}{e\left[e^{-a_n(x_t-u_n)} - e^{a_1(x_t-u_1)}\right]} & x \ge x_t \\ 0 & x < x_t \end{cases} , \quad (4.8)$$

and the density of z_1 is

$$f_{z_1}(x) = \left[a_n e^{-a_n(x-u_n)} + a_1 e^{a_1(x-u_1)}\right]\left[1 - F_{z_1}(x)\right] U(x-x_t) . \quad (4.9)$$

If we have N sets of n z's detected at a given threshold as noise, and another N sets of n z's detected at that threshold as signal (plus noise), we can use all the previous results of Chapter 1 for calculating confidence intervals, and in determining the values towards which the estimates converge, since we again have functions of identically distributed random variables. The only information needed are the values of the first four central moments. However, because of the difficulty in evaluating the appropriate integrals, the moments must either be calculated numerically on a computer, or some approximations made for the integrals so that closed form results can be obtained.

Tables (7) and (8) summarize the results of two computer simulations. Table (7) was run with a Rician density ($A = 6.5$, $\sigma = 1$) for signal and a Rayleigh density ($\sigma = 1$) for noise, and Table (8) was run with the

same noise density, but the Rician density had its parameter "A" changed from A = 6.5 to A = 7.0. Both simulations were made with N = 20. The threshold for Table (7) was x_t = 3.86, and the threshold for Table (8) was x_t = 4.1. The column labeled "optimum value" contains the theoretical values corresponding to the extremes of n samples taken with a learning period (i.e., no incorrect decisions). The column labeled "asymptotic value" gives the values which resulted from the above analysis (that is, results obtained from eqs. (4.4) and (4.9)), and finally, the last column gives the numerical results from the actual computer simulation.

Table (7)

Parameter	Optimum Value	Asymptotic Value	Simulation Result
u_1	4.54	4.58	4.69
a_1	2.44	5.72	3.77
u_n	2.8	2.86	3.0
a_n	2.8	3.06	3.23
x_t	3.64	3.91	3.89

Table (8)

Parameter	Optimum Value	Asymptotic Value	Simulation Result
u_1	5.04	5.07	5.18
a_1	2.48	3.64	3.09
u_n	2.8	2.84	3.02
a_n	2.8	2.86	2.77
x_t	3.88	4.05	4.14

It should be noted that the results of Chapter 1 only apply to functions of identically distributed random variables. Thus, the above results are only valid when all of the samples are detected using a single threshold.

However, if we recall the detection procedure, after the initial time delay (when all samples are detected at the same threshold), the new samples are detected at a different threshold each time either n signal decisions or n noise decisions are made. Also, in any given set of n samples, detected as either signal or noise, there will probabily be more than one threshold used, because while we are in the process of making, say n signal samples, we are also collecting $\ell < n$ noise samples. After the nth signal decision is made, we change the threshold, and then continue detecting. It is obvious then, that by the time we make n noise decisions, we will have used at least two thresholds, since the first ℓ noise decisions were made with one threshold, and then that threshold was adjusted at least once before the remaining $n - \ell$ noise decisions were made.

The simplest solution (although most costly in terms of time and effort) is to redetect all the previous samples each time the threshold is changed. In other words, instead of only using a given threshold to make n new signal decisions or n new noise decisions, we can use that threshold to redetect all the previous samples for which decisions were already made. If this procedure is used, then eqs. (4.4) and (4.9) would always be applicable, and therefore all the results of Chapter 1 would be appropriate.

If the cost of storing and redetecting the old samples is prohibitive, the above analysis can be thought of as an upper-bound to the accuracy of the estimates. The term "upper bound" is used here only in an intuitive sense. The justification for it is that as the threshold estimate improves, fewer errors should be made, and therefore, the parameters should become closer to what they would be if we had a learning period.

If neither of the above two approaches is satisfactory, confidence

intervals can be obtained using the nonidentical random variables. We will, however, make the following approximation: If we are taking the maximum (or minimum) sample from n independent samples, we will assume they were all detected with the same threshold. As mentioned above, this will almost always not be the case, but for equal apriori probabilities of transmission, it will be very nearly true for large n, since for each n signal bits that are transmitted, approximately the same number of noise bits will be transmitted. Also note that this assumption, even though violated, is not very critical, because the change in adjacent threshold values should not be very great (see Fig. (4.2)). In other words, the only difference in n adjacent signal decisions or noise decisions is that not all of them have been detected with the same threshold. However, since it is highly unlikely (for low error rates and equal apriori transmission probabilities) that more than two thresholds would have been used, and since the difference in the two thresholds is small, the random variables resulting from decisions using these thresholds will be almost identically distributed.

Finally, it should be realized that this approximation is not theoretically necessary, but is made so that we can use the analysis leading to eqs. (4.4) and (4.9) (that is, so that we can approximate the density functions by their asymptotes and thereby obtain a simple expression for the densities of the extremes). If the random variables are not identically distributed, we cannot obtain a simple exponential asymptote for their densities, but we can still write down the exact distritution of the extremes (assuming specific forms for the initial density as has usually been the case when confidence intervals were required). For example, if we have made $i < n$ noise decisions at one threshold, and $n - i$ noise decisions at another threshold, the distribution of the maximum of the set is

$$F_{max}(x) = F_1^i(x)\, F_2^{n-i}(x) \quad ,$$

where $F_1(x)$ is the distribution corresponding to the first threshold, and $F_2(x)$ is the distribution corresponding to the second threshold.

Proceeding now with the above assumption, the extreme from any set of n random variables all detected at the same threshold has a density which is given by either eq. (4.4) or eq. (4.9).

If we take N maxima and N minima, where each one came from a group of n samples detected with a different threshold, and compute the four Gumbel parameters, estimates of the probability of error, and the threshold, we have functions of sample moments of nonidentically distributed random variables.

To prove that all the estimates are normal random variables, we start with a theorem given by R. de Misès.[4.3] A sum of independent nonidentically distributed random variables is asymptotically normally distributed if the following conditions are satisfied:

Let x_i, $i = 1 \ldots n$, be any random variable with mean μ_i and variance σ_i^2. Then the quantity

$$C_i = \int |x_i - \mu_i|^{2+\varepsilon} \; f(x_i)\, dx_i$$

must be bounded for some $\varepsilon > 0$, and $n^{\frac{-2}{2+\varepsilon}} \sum_{i=1}^{n} \sigma_i^2$ must go to infinity as n goes to infinity.

Using this theorem, it can easily be shown that $\sum_{i=1}^{n} x_i$ and $\sum_{i=1}^{n} x_i^2$ are asymptotically normal, where the x_i are distributed according to either eq. (4.4) or eq. (4.9). This can be seen by noting that these densities are bounded on one side and go to zero as a double exponential on the other side, so that all moments will exist of both x_i and x_i^2.

In Appendix B, we use the above result to prove that the sample variance is also asymptotically normal. This will enable us to extend

Cramér's theorem on functions of identically distributed random variables to our present case, and therefore we will have the result that all our estimates are asymptotically normal.

Finally, a computer simulation was run to supply experimental verification for the theory. The simulation showed that the threshold changed by small amounts with each adaptation, so the approximation that that n consecutive decisions were made with only one threshold was reasonable.

Figure (4.2) shows the result of the computer simulation of the system. The signal density was a Rician with parameters $A = 6.5$ and $\sigma = 1$, and the noise came from a Rayleigh density with parameter $\sigma = 1$. During the time delay, 2500 samples were collected, and the overall sample mean was used as the first threshold. The samples were then detected at that threshold until there were 1000 signal decisions and 1000 noise decisions. In each group of 1000 samples, N was set equal to 20, and therefore $n = 50$. The four Gimbel parameters were then calculated and this led to the first threshold adjustment. After that, the threshold was continually changed each time either 50 signal decisions or 50 noise decisions were made.

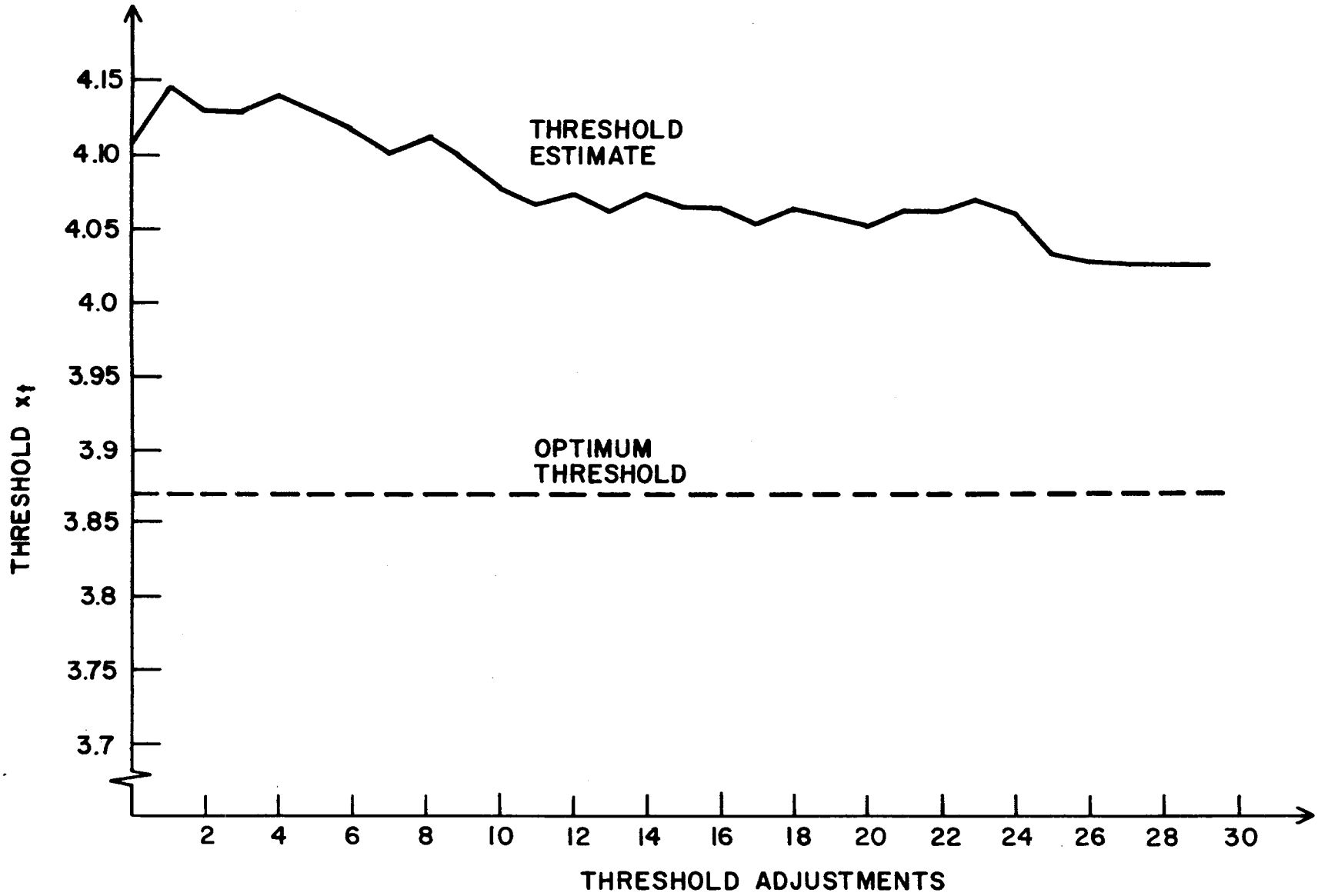


Figure (4.2) Convergence of Threshold Estimate Using Decision Directed Measurements

Summary

A nonparametric detector which can operate in any environment which has noise having exponential-type characteristics has been studied. The detector was based upon extreme-value theory.

In Chapter 1, a review of extreme-value theory was given, and it was shown how this theory was to be used to detect digital signals in additive noise. EVT has previously been used in signal detection by engineers at JPL.[S.1] This work differs from their's in several respects. Only a single exponential is used to estimate the probability of error, instead of a double exponential. Also, one of the main objectives of this report was to arrive at an estimate for the optimum threshold of the system. In this last respect, this work differs from other nonparametric detectors, since, to the author's knowledge, most nonparametric detectors are radar-type detectors in that they choose the desired false alarm rate and accept whatever false dismissal rate that results.

In Chapter 2, specific examples were given for detecting a constant signal in additive noise, and a comparison was made between the EVT detector and the rank detector. The two significant results of that comparison are that the EVT detector, in certain situations, will perform as well as the optimum Neyman-Pearson parametric detector, and therefore better than the rank or any other nonparametric detector, and that it will do so with significantly less effort than that required by other nonparametric detectors.

Chapter 3 considers two detectors which can be used when the signal is fading. The first detector uses a fixed threshold, and is not of much practical use because it can only perform satisfactorily when the variance of the fade is quite small. The second detector is an adaptive

detector, and, to the author's knowledge, this method of using extreme-value theory with a time-varying parameter has not been considered before. Computer simulated results were presented for the adaptive detector, and the computer results were in agreement with the theoretical results.

Finally, in Chapter 4, we eliminated the learning period that was essential to the results of the first three chapters, and resorted to decision-directed measurements. The asymptotic distributions of the estimates were derived, and it was shown that satisfactory results could be obtained if the error probabilities were low. As in Chapter 3, numerical results obtained by a computer simulation were presented.

Among the problems which have not been solved and which seem to be worth investigating is that of choosing an optimum value of n in two situations.

The first case is when the learning period is present, but, regardless of what error probability is to be estimated, only $nN = K$ samples can be taken. That is, given K samples in a learning period, what is the best way to divide them into N groups of n samples?

The second situation arises when the learning period is eliminated. Since picking n too small results in inaccurate theoretical estimates (because of the Taylor series only being accurate in a limited region about the u's), and choosing n too large results in inaccurate statistical estimates of the parameters, a compromise has to be made. However, for any given error probability, it is not known how to make that compromise.

Appendix A - Density Function for Minimum of n Independent Samples

The following is the derivation leading up to equation (1.8), and is essentially the same as Gumbel's derivation[A.1] which results in equation (1.7).

Expanding F(x), the distribution of the initial variate, about $x = u_1$, we have

$$F(x) = F(u_1) + \frac{x - u_1}{1!} f(u_1) + \frac{(x - u_1)^2}{2!} f'(u_1) + \dots , \qquad (A.1)$$

where

$$F(u_1) = \frac{1}{n} .$$

Factoring $\frac{1}{n}$ out of eq. (A.1), we obtain

$$F(x) = \frac{1}{n}\left[1 + \sum_{1}^{\infty} \frac{(x - u_1)^k}{k!} n f^{(k-1)}(u_1)\right] . \qquad (A.2)$$

If eq. (1.2) is now used at the point $x = u_1$, we obtain

$$\frac{f(u_1)}{F(u_1)} = \frac{f'(u_1)}{f(u_1)} .$$

But

$$a_1 = n f(u_1) ,$$

so we can write $nf'(u_1)$ as follows:

$$n f'(u) = n f(u_1) \frac{f(u_1)}{F(u_1)} = (a_1)\left(\frac{\frac{a_1}{n}}{\frac{1}{n}}\right) = a_1^2 .$$

Differentiating both sides of eq. (1.2), and evaluating at $x = u_1$, we obtain

$$\frac{f''(u_1)}{f(u_1)} - \frac{[f'(u_1)]^2}{f^2(u_1)} = \frac{f'(u_1)}{F(u_1)} - \frac{f^2(u_1)}{F^2(u_1)} ,$$

or, again using eq. (1.2),

$$n f''(u_1) = n \frac{f(u_1) f'(u_1)}{F(u_1)} = n \frac{\left(\frac{a_1}{n}\right)\left(\frac{a_1^2}{n}\right)}{\frac{1}{n}} = a_1^3 \quad .$$

We will now assume that

$$n f^{(k)}(u_1) = a_1^{k+1} \quad , \tag{A.3}$$

and will show that this implies

$$n f^{(k+1)}(u_1) = a_1^{k+2} \quad . \tag{A.4}$$

By mathematical induction, it will then follow that the equation holds for all k.

To show this, consider

$$f^{(k)}(u_1) = \frac{a_1^{k+1}}{n} = f(u_1)\, a_1^k = \frac{\left[f(u_1)\right]^{k+1}}{\left[F(u_1)\right]^k} \quad ,$$

or

$$\frac{f^{(k)}(u_1)}{f(u_1)} = \left[\frac{f(u_1)}{F(u_1)}\right]^k \quad . \tag{A.5}$$

Again differentiating both sides of the equation, we obtain

$$\frac{f^{(k+1)}(u_1)}{f(u_1)} - \frac{f^{(k)}(u_1)\, f'(u_1)}{\left[f(u_1)\right]^2} = \frac{k\left[f(u_1)\right]^{k-1} f'(u_1)}{\left[F(u_1)\right]^k} - \frac{k\left[f(u_1)\right]^{k+1}}{\left[F(u_1)\right]^{k+1}} \quad .$$

Using equation (A.5), we have

$$\frac{f^{(k+1)}(u_1)}{f(u_1)} = (k+1)\frac{\left[f(u_1)\right]^{k-1} f'(u_1)}{\left[F(u_1)\right]^k} - k\frac{\left[f(u_1)\right]^{k+1}}{\left[F(u_1)\right]^k} =$$

$$(k+1)\frac{\left(\frac{a_1}{n}\right)^{k-1}\frac{a_1^2}{n}}{\left(\frac{1}{n}\right)^k} - \frac{k\left(\frac{a_1}{n}\right)^{k+1}}{\left(\frac{1}{n}\right)^{k+1}} = (k+1)\, a_1^{k+1} - k a_1^{k+1} = a_1^{k+1} \quad ,$$

or,

$$n f^{(k+1)}(u_1) = a_1^{k+1} \quad .$$

Therefore, we can rewrite eq. (A.2) as

$$F(x) = \frac{1}{n}\left[1 + \sum_{1}^{\infty} a_1^k \frac{(x - u_1)^k}{k!}\right] , \tag{A.6}$$

or,

$$F(x) = \frac{1}{n} e^{a_1 (x - u_1)} , \tag{A.7}$$

which is the asymptotic form we have used throughout this report.

To complete this derivation, it simply must be recalled that if $\Phi_1(x)$ is the cumulative distribution of the minimum of n samples, then

$$\Phi_1(x) = 1 - \left[1 - F(x)\right]^n = 1 - \left[1 - \frac{e^{a_1 (x - u_1)}}{n}\right]^n ,$$

and

$$\lim_{n \to \infty} \Phi_1(x) = 1 - e^{-e^{a_1(x - u_1)}} ,$$

which is eq. (1.8).

Appendix B - Asymptotic Normality for Functions of Nonidentically Distributed Random Variables

We will prove that if we have nonidentical random variables such that $\sum x_i$ and $\sum x_i^2$ are both asymptotically normal, then $\frac{1}{n}\sum(x_i - \bar{x})^2$ is also asymptotically normal. The proof is simply an extension of a proof given by Cramér[B. 1] for identically distributed random variables.

Let

$$\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 ,$$

$$\sigma^2 = E(\hat{\sigma}^2) ,$$

$$m_k = \frac{1}{n}\sum x_i^k \qquad k = 1, 2 ,$$

$$n_k = E(m_k) ,$$

$$\sigma^2_{m_k} = \text{var}(m_k) ,$$

$$z_1 = \frac{m_1 - n_1}{\sigma_{m_1}} , \qquad \text{and}$$

$$z_2 = \frac{m_2 - n_2}{\sigma_{m_2}} .$$

Then we have

$$(\hat{\sigma}^2 - \sigma^2) = \frac{1}{n}\sum x_i^2 - E\left[\frac{1}{n}\sum x_i^2\right] - \left[\bar{x}^2 - E(\bar{x}^2)\right] ,$$

or

$$(\hat{\sigma}^2 - \sigma^2) = \sigma_{m_2} z_2 - \left[(z_1 \sigma_{m_1} + n_1)^2 - E(\bar{x}^2)\right] =$$

$$\sigma_{m_2} z_2 - 2n_1\sigma_{m_1} z_1 - \sigma^2_{m_1} z_1^2 + \left[E(\bar{x}^2) - n_1^2\right] =$$

$$\sigma_{m_2} z_2 - 2n_1 \sigma_{m_1} z_1 - \sigma^2_{m_1} z_1^2 + \sigma^2_{m_1} .$$

Consider the term $\sigma^2_{m_1} z_1^2$. We have $E\left[\left|\sigma^2_{m_1} z_1^2\right|\right] = \sigma^2_{m_1} E(z_1^2) = \sigma^2_{m_1}$ times a number independent of n, since z_1, by definition, is asymptotically normal with zero mean and variance equal to unity. Also,

$\sigma^2_{m_1} = \text{var}(\bar{x}) = 0\left(\frac{1}{n}\right)$. Therefore, by Tekebycheff's inequality[B.2], $\sigma^2_{m_1} z_1^2$ converges in probability to zero.

Therefore, $\hat{\sigma}^2 - \sigma^2$ is a linear function of two asymptotically normal random variables, and is therefore itself asymptotically normal.[B.3]

Having established the asymptotic normality of $\hat{\sigma}^2 - \sigma^2$, we can now, step by step, use the proof of Cramérs theorem referred to in Chapter 1 (see eqs. (1.29) and (1.30)) and therefore conclude that the functions of the nonidentically distributed random variables given in Chapter 4 are also asymptotically normal.

References

Chapter 1

(1.1) E. J. Gumbel, Statistical Theory of Extreme Values and Some Practical Applications (Washington, 1954).

(1.2) E. J. Gumbel, Statistics of Extremes (New York, 1958).

(1.3) ibid., pp. 118-122.

(1.4) ibid., pp. 166-169.

(1.5) ibid., p. 20.

(1.6) John C. Ashlock and Sandra M. Lurie, On the Application of Extreme-Value Statistics to Command Oriented Problems, Technical Report No. 32-1025, Jet Propulsion Laboratory, Pasadena, California, October 15, 1966.

(1.7) J. Capon, "A Nonparametric Technique for the Detection of a Constant Signal in Additive Noise," 1959 IRE WESCON Convention Record, pt. 4, pp. 92-103.

(1.8) H. L. Groginsky, L. R. Wilson, and D. Middleton, "Adaptive Detection of Statistical Signals in Noise," IEEE Transactions on Information Theory, Vol. IT-12, pp. 337-348, July, 1966.

(1.9) R. F. Daly and C. K. Rushforth, "Nonparametric Detection of a Signal of Known Form in Additive Noise," IEEE Transactions on Information Theory, Vol. IT-11, pp. 70-76, January 1965.

(1.10) E. A. Feustel and L. D. Davisson, "The Asymptotic Relative Efficiency of Mixed Statistical Tests," IEEE Transactions on Information Theory, Vol. IT-13, pp. 247-255, April, 1967.

(1.11) Gumbel, Statistics of Extremes, pp. 174-175.

(1.12) M. Muntner, "Error Control on Real Channels," Ph. D. Dissertation, Polytechnic Institute of Brooklyn, Brooklyn, New York, 1968.

(1.13) Ashlock.

(1.14) Gumbel, Statistics of Extremes, pp. 231-235.

(1.15) ibid., pp. 9-11.

(1.16) H. Cramér, Mathematical Methods of Statistics (Princeton, 1946), pp. 366-367.

(1.17) M. Schwartz, W. Bennett, and S. Stein, Communication Systems and Techniques (New York, 1966) p. 406.

(1.18) Cramér, P. 348.

(1.19) ibid., pp. 345-349.

(1.20) M.H. Bickelhaupt, Jr., "Errors in the Measurement of the Probability Density Function," Tech. Rept. No. RADC-TR-65-432, Rome Air Development Center, November, 1965.

(1.21) C.H. Goulden, Methods of Statistical Analysis, (New York, 1956) 2nd Ed., pp. 23-24.

(1.22) G.R. Cooper and J.A. Tabaczynski, "Estimation of Probability Density and Distribution Functions," Report No. TR-EE65-15, Purdue University, August 1965.

(1.23) J. Millard and L. Kurz, "Nonparametric Signal Detection - An Application of the Kolmogorov-Smirnov Cramér-Von Mises Tests," Dept. of Elec. Eng. Tech. Report 400-127, New York University, January, 1966.

Chapter 2

(2.1) Schwartz, Bennett, and Stein, pp. 120-154.

(2.2) ibid., pp. 20-26.

(2.3) M. Schwartz, Information Transmission, Modulation, and Noise (New York, 1959), pp. 407-409.

(2.4) Schwartz, Bennett, and Stein, p. 28.

(2.5) J.I. Marcum, "A Statistical Theory of Target Detection by Pulsed Radar," IRE Trans. on Information Theory, Vol. IT-6, pp. 59-267, April, 1960.

(2.6) A. Bodonyi, "Effects of Impulse Noise on Digital Data Transmission," IRE Trans. on Communication Systems, Vol. CS-9, pp. 335-361, December, 1961.

(2.7) P. Mertz, "Model of Impulsive Noise for Data Transmission," IRE International Convention Record, Part 5, pp. 247-259, 1960.

(2.8) L. Halsted, "On Binary Data Transmission Error Rates Due to Combinations of Gaussian and Impulse Noise," IEEE Transactions on Communications Systems, Vol. CS-11, pp. 428-435, December, 1963.

(2.9) A.M. Mood and F.A. Graybill, Introduction to the Theory of Statistics (New York, 1963) 2nd Ed., pp. 417-419.

(2.10) D.A.S. Fraser, Nonparametric Methods in Statistics (New York, 1957), p. 235.

(2.11) J. Capon, "A Note on the Asymptotic Normality of the Mann-Whitney-Wilcoxen Statistic," Journal of the American Statistical Association, Vol. 56, pp. 687-691, September, 1961.

(2.12) Schwartz, Bennett, and Stein, pp. 57-63.

(2.13) idem.

Chapter 3

(3.1) E.I. Nelson, D.L. Schilling, K.K. Clarke, "The Response of an FM Discriminator to an FM Signal in Randomly Fading Channels," Tech. Report T-1, Dept. of Elec. Eng., Polytechnic Institute of Brooklyn, Chapter 2, 1967.

(3.2) Schwartz, Bennett, and Stein, pp. 372-374.

(3.3) N.A. Barricelli, "Les Plus Grands et Plus Petits Maxima ou Minima Annuels d'une Variable Climatique," Archiv for Mathematik og Naturvidenskab, Oslo, 46 (No. 6), pp. 155-194, 1943.

(3.4) Gumbel, Statistics of Extremes, p. 53.

Chapter 4

(4.1) J.G. Proakis, P.R. Drouilhet, Jr., R. Price, "Performance of Coherent Detection Systems Using Decision Directed Channel Measurement," IEEE Trans. on Comm. Systems, Vol. CS-12, pp. 54-63, March, 1964.

(4.2) M. Abramowitz, I.A. Stegun, Handbook of Mathematical Functions, (Washington, 1964), p. 70.

(4.3) R. de Misès, "Les Lois de Probabilité pour les Fonctions Statistiques," Annales de L'Institut Henri Poincaré, 6 (1936) pp. 185-212.

Summary

(S.1) Ashlock.

Appendix A

(A.1) Gumbel, Statistics of Extremes, pp. 166-169.

Appendix B

(B.1) Cramér, pp. 364-366.

(B.2) ibid., p. 182.

(B.3) Fraser, p. 210.