THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

STATISTICAL THEORY FOR THE DETECTION OF SIGNALS UNDER LINEAR SCALE TRANSFORMATIONS

by

Gibb Blanks Matlock

Technical Report No. 68
Department of Statistics THEMIS Contract

April 20, 1970

Research sponsored by the Office of Naval Research
Contract N00014-68-A-0515
Project NR 042-260

DEPARTMENT OF STATISTICS
Southern Methodist University

STATISTICAL THEORY FOR THE DETECTION OF SIGNALS UNDER LINEAR SCALE TRANSFORMATIONS

A Dissertation Presented to the Faculty of the Graduate School

of

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

by

Gibb Blanks Matlock

(M.A., University of Texas, 1959)

April 20, 1970

Matlock, Gibb Blanks

B.A., The University of Texas, 1957

M.A., The University of Texas, 1959

Statistical Theory for the Detection of Signals Under Linear Scale Transformations

Adviser: Professor and Director of Graduate Studies Donald B. Owen

Doctor of Philosophy degree conferred May 24, 1970

Dissertation completed April 20, 1970

Most of the literature of signal detection assumes a parametric signal model of the form $f(t) = \beta S(t - t_0)$ where the amplitude β and the time of arrival t_0 are unknown. Many of the questions remain unanswered about signals of the form $\beta S(at - t_0)$ where a is an unknown scale parameter. Several basic results are presented about the reception of signals of this more general form.

The Likelihood Ratio Test for detection is developed and curves of probability of detection as a function of signal-to-noise ratio are given for various false alarm rates. Detection in the case of multiple observations is also considered.

Estimation of the unknown signal parameters β, a, t_0 and the unknown noise variance σ^2 is treated. The maximum likelihood or least squares estimators for these parameters are given, along with an iterative computational technique. The large sample distribution of the estimators is also given.

Two types of signal classification problems are discussed and the Bayes decision rules for their solutions are presented.

ACKNOWLEDGEMENTS

The author wishes to express his gratitude to his dissertation director, Dr. Donald B. Owen for his guidance and encouragement during the research period. Thanks is also given to the other members of the dissertation committee, Dr. Richard P. Bland, Dr. A.M. Kshirsagar, and Dr. Kenneth W. Heizer for their excellent instruction and patient consultation.

Thanks are also due to my employer, Texas Instruments Incorporated, who has generously supported my education through its Ph.D. Honors Co-op Program.

The author wishes to thank his wife and family for their love and support throughout this educational endeavor.

Finally my thanks to the Publications Department of Texas Instruments for the typing and final production of this dissertation.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER I

STATEMENT OF THE PROBLEM

1.1 Introduction

The general topic of signal detection has been for many years the subject of some very fruitful statistical research, resulting in a voluminous literature. There are a few fundamental papers (i.e., Slepian[10], Middleton and Van Meter[7], and Chapter 2 of Van Trees[11]) which indicate how to formulate detection and estimation problems for a general signal whose exact form depends on several unknown parameters. But by far the largest part of the literature assumes a signal model of the form

$$f(t) = \beta S(t - t_0)$$

where only amplitude β and the time of arrival t_0 are unknown. This assumption permits the theory to be developed in much sharper detail than is possible for the more general signal model. Many good textbooks are available which consider detailed questions about the detection and estimation of signals of this form, such as Van Trees[11], Helstrom[2], Middleton[6], Davenport and Root[1], Y.W. Lee[5] and Lawson and Uhlenbeck[4]. These textbooks with very specific results are usually more valuable to the working system designer than are the more fundamental papers upon which they are partially based.

This dissertation is an attempt to extend the general topic of signal detection in another direction, and that is to begin a collection of detailed results pertaining to a slightly more general signal model of the form

$$f(t) = \beta S(at - t_0)$$

where the new parameter a represents an unknown linear scale distortion. Results will be obtained in the following chapters concerning some of the questions ordinarily asked about detection, parameter estimation, and classification.

1.2 The Classical Problems

The reception of signals in the presence of noise presents many statistical decision problems. These can be roughly divided into three areas, signal detection, parameter estimation, and classification. These types of decision problems can be illustrated with an example. Figure 1.1 shows a signal reception process involving all three types of decision problems.

Suppose that a voltage or waveform v(t) is observed at the terminals of a receiver. This received waveform is observed over an interval of time that could contain a signal $S(\underline{\theta},t)$, and the observer must choose between two possible alternative hypotheses: (H_0) there is no signal, and the input consists of noise alone, $V(t) = N(t)$, or (H_1) the signal is present, and $V(t) = N(t) + S(\underline{\theta}, t)$. This decision represents the detection part of the process, and because of the stochastic nature of the noise, it can be formulated as a statistical test of hypotheses.

If the decision is yes, there is a signal $S(\underline{\theta}, t)$ present; the next problem in the sequence is to estimate the unknown parameters $\underline{\theta}$ of the signal. All of the standard techniques of statistical estimation theory are applicable subject only to limitations on the amount of prior knowledge or assumptions available concerning the statistics of the signal and the noise or system costs. A common example of a signal parameter to be estimated occurs in radar, where the signal is displaced in time an amount t_0 which is proportional to the range to the reflecting target. By estimating the signal parameter t_0 based on the observed data $V(t) = N(t) + \beta S(t - t_0)$ the radar system can then indicate the range to target.

The last of the three decision problems is that of classifying the received signal. Classification means the assignment of the signal to one of a set of preassigned classes. Decoding the signal and deciding which letter of the alphabet it represents is an example of classification. Deciding, on the basis of its observed parameters which type of emitter the signal came from is another example. In some cases the classification problem is more easily formulated as a test of multiple hypotheses. Suppose there are M sources active and the *kth* source emits the signal $S_k(t)$, $k = 1, 2, \ldots, M$. After observing the received waveform V(t), the observer must choose between the M + 1 hypotheses:

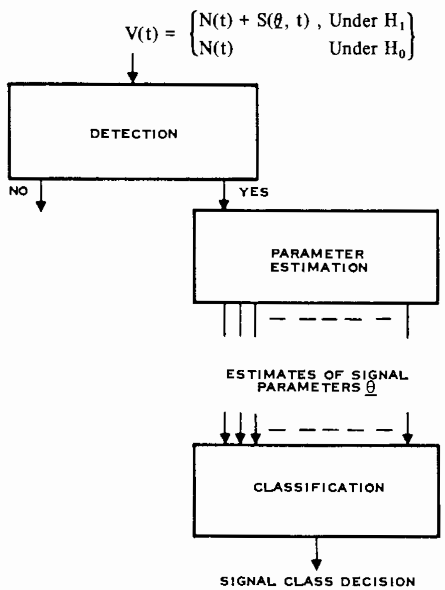


Figure 1.1 Signal Reception Process

$$
\begin{aligned}
H_0 &: V(t) = N(t) \\
H_1 &: V(t) = N(t) + S_1(t) \\
H_2 &: V(t) = N(t) + S_2(t) \\
&\vdots \\
H_M &: V(t) = N(t) + S_M(t)
\end{aligned}
$$

This formulation is slightly different in that it includes the detection problem. This latter formulation is sometimes called the M-ary detection problem.

These three general categories encompass a large part of the statistical decision theory used in the study of communication systems, but are by no means exhaustive. Also there are many variations of the problems within each of the three categories that are worthy of separate studies in themselves. The material in this dissertation will include only the most basic of the questions that could be asked about the reception of signals from the generalized class described below.

1.3 Linear Scale Distortion

Throughout this dissertation we shall say that $f(t)$ and $S(t)$ are the same under a linear scale transformation if there are real numbers a and t_0 such that $f(t) = S(at - t_0)$. There are many ways in which signals may undergo linear scale distortions. The most obvious is the Doppler effect. If the signal $S(t)$ is transmitted and there is a relative velocity v/c between the transmitting and receiving platforms (v/c merely expresses the velocity v in units of c, the propagation velocity of the signal), then the signal $S(at - t_0)$ will be observed at the receiver, where $a = 1 + v/c$. In radar applications and in electromagnetic communication systems, the ratio of velocities v/c and the ratio of signal bandwidth to center frequency are so small that scale distortion in these cases can be handled rather simply. Almost all signals in these problems are modeled as in (Helstrom[2], page 13).

$$S(t) = A(t) \cos [\omega t + \phi(t)]$$

where $A(t)$ represents an amplitude modulation, and $\phi(t)$ represents a phase modulation. Since $1 + v/c$ is very little different from one in the radar application, and since the carrier frequencies ω are so great compared to the signal bandwidths, the Doppler effect is approximated as a shift in the carrier frequency ω. The envelope function A(t) and the phase function $\phi(t)$ are assumed to be invariant. This approximation works well in radar but is completely worthless in many other problems. In both passive and active sonar, for example, the signal bandwidths and velocity ratios are such that more exact treatment is required to yield usable answers.

Other problems which involve linear scale transformations include the following.

1. In automatic processing of speech data, there is a scaling problem because some speakers utter the same words or signals faster than others. Indeed the same speaker does not always talk at the same rate.
2. Medical data and other biological data have a scale ambiguity because growth rates vary, or pulse rates vary or the rate of any physical process may vary, although the signal, or phenomenon signature for which one is searching may be invariant under a linear scale transformation.
3. For a two dimensional scale transformation, consider that in image recognition there is a magnification parameter which is not always available to the observer.

It may be argued that many scale distortions like the above are not actually linear but are linear to a first order approximation, so it is felt that a theory based on linear distortions would handle these problems better than that which assumes no scale distortion.

The problems discussed in this dissertation, then, will be concerned with the observation of samples V_i from the process $V(t) = N(t) + \beta S(at - t_0)$, or

$$V_i = V(t_i) = N_i + \beta S(at_i - t_0)$$

where

S(t) is the known signal form

β is the unknown amplitude of the signal

a is the unknown linear scale distortion parameter

$\frac{t_0}{a}$ is the unknown time of arrival of the signal

N_i are noise samples of known statistics, usually assumed to be white, or uncorrelated Gaussian random variables, unless otherwise specified

$t_i = i \cdot \Delta t$ are sample times regularly spaced at intervals of Δt

The samples V_i will be observed for a length of time, T, long enough so that either the signal is entirely contained within the interval (0, T) or it is not present at all. In the literature of signal detection theory when this assumption is made the signal is said to be well imbedded in the interval. On the basis of these observed samples the observer will be asked to judge whether or not the signal, S, is present in any form, and if it is, to estimate its parameters. In the M-ary detection problem, S may be one of several known signal forms $S_1, S_2, \ldots, S_M$, and the observer will be asked to judge which one it actually is, based on the observations V_i.

If the scale distortion parameter a is equal to one, these questions become the classical ones that are treated in detail in the textbooks already referenced. It will be the contribution of this paper to extend these results to signals having undergone linear scale distortions and to find the new decision rules implied.

CHAPTER II

DETECTION IN WHITE GAUSSIAN NOISE

2.1 Signal Detection and Statistics

Suppose the known signal form S(t) is zero outside the interval $(0, \tau)$ as depicted in Figure 2.1A, where $\tau \leqslant T$. The epoch τ will be called the signal duration and T will be called the observation time. Given the observed samples $V_i = V(t_i)$ over the interval (0, T) the observer must decide whether or not there is any signal of the form $S(at - t_0)$ completely contained within the observation interval. Figure 2.1B shows an example of how $S(at - t_0)$ might be related to the observation interval (0, T). If the signal is present, it will be added to a noisy background, so no matter what procedure he uses to make this decision there is always a chance that he may be wrong. Figure 2.1C shows how the combined observation of signal plus noise might appear. Although the detection decision can be treated as a Bayes decision problem if loss functions and prior distributions can be specified, it will be treated here as a Test of Hypothesis. The error probability that will be specified is the probability of declaring that there is a signal present when in reality there is none. This is called the single scan false alarm probability. To state the Test of Hypothesis formally use the model

$$V(t) = N(t) + \beta S(at - t_0).$$

On the basis of this model and the samples V_i, the observe must test the hypothesis

$H_0 : \beta = 0$ against

$H_a : \beta \neq 0.$

The Likelihood Ratio Test will be used for making this decision. This detection criteria is called the theory of the "Neyman-Pearson Observer" in radar texts, presumably because Neyman and Pearson first proposed the Likelihood Ratio Test in 1928. Let $f(\underline{V}|\beta)$ denote the conditional joint likelihood function of the observations $\underline{V}$ given the value of β. This likelihood function also depends on the values of the unknown

n_0 denote the number of sample points in the interval $(0, t_0/a)$,

n_1 denote the number of sample points in the interval $(0, t_0/a + \tau/a)$,

n denote the total number of sample points in the interval (0, T)

Assuming the alternative hypothesis H_a to be true, that is $\beta \neq 0$, the samples V_i are given by

$$V_i = N_i + \beta S(at_i - t_0). \quad i = 1, 2, \ldots, n$$

Assuming the noise samples N_i to be white (or uncorrelated) and Gaussian with mean zero and variance σ^2, the joint likelihood of the observed V_i is the product

$$f(\underline{V}|\beta \neq 0) = L_1 \cdot L_2 \cdot L_3$$

where

$$L_1 = \prod_{i=1}^{n_0} \frac{1}{\sqrt{2\pi}\sigma} e^{-V_i^2/2\sigma^2}$$

$$L_2 = \prod_{i=n_0+1}^{n_1} \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2\sigma^2)[V_i - \beta S(at_i - t_0)]^2}$$

$$L_3 = \prod_{i=n_1+1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-V_i^2/2\sigma^2} .$$

$$f(\underline{V}|\beta \neq 0) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \cdot \prod_{i=1}^{n_0} e^{-V_i^2/2\sigma^2} \cdot \prod_{i=n_1+1}^{n} e^{-V_i^2/2\sigma^2} \cdot \prod_{i=n_0+1}^{n_1} e^{-(1/2\sigma^2)[V_i - \beta S(at_i - t_0)]^2}$$

The natural log of this likelihood function becomes

$$\ln f(\underline{V}|\beta \neq 0) = \frac{-n}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n_0} V_i^2 - \frac{1}{2\sigma^2} \sum_{i=n_1+1}^{n} V_i^2 - \frac{1}{2\sigma^2} \sum_{i=n_0+1}^{n_1} [V_i - \beta S(at_i - t_0)]^2 \quad . \tag{2.2.3}$$

To find the value of β which maximizes this expression, we set

$$0 = \frac{d}{d\beta} \ln f(\underline{V}|\beta \neq 0) \quad .$$

$$0 = \frac{1}{\sigma^2} \sum_{i=n_0+1}^{n_1} [V_i - \beta S(at_i - t_0)]\ S(at_i - t_0) \tag{2.2.4}$$

or

$$\beta = \frac{\sum_{i=n_0+1}^{n_1} V_i\ S(at_i - t_0)}{\sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2}$$

Substituting this value of β back into Equation (2.2.3), we find

$$\underset{\beta}{\text{Max}} \ln f(\underline{V}|\beta \neq 0) = \frac{-n}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[\sum_{i=1}^{n_0} V_i^2 + \sum_{i=n_1+1}^{n} V_i^2\right] - \frac{1}{2\sigma^2} \sum_{i=n_0+1}^{n_1} \left[V_i - \frac{S(at_i - t_0) \sum_{j=n_0+1}^{n_1} V_j\, S(at_j - t_0)}{\sum_{j=n_0+1}^{n_1} [S(at_j - t_0)]^2} \right]^2 \quad .$$

$$\underset{\beta}{\text{Max}} \ln f(\underline{V}|\beta \neq 0) = \frac{-n}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} V_i^2 + \frac{1}{2\sigma^2} \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\ S(at_i - t_0)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2} \tag{2.2.5}$$

Let $\hat{a}$ and $\hat{t}_0$ denote the values of a and t_0 which maximize

$$\frac{\left[\sum_{i=n_0+1}^{n_1} V_i\ S(at_i - t_0)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2} \quad . \tag{2.2.6}$$

where n_0 and n_1 are also functions of a and t_0. More explicitly

n_0 = largest integer less than $(t_0/a\Delta t)$

n_1 = largest integer less than $(t_0/a + \tau/a)/\Delta t$.

Note that for any value of σ^2, the numbers $\hat{a}$ and $\hat{t_0}$ also maximize $f(\underline{V}|\beta \neq 0)$. So

$$\underset{\beta,\, a,\, t_0}{\text{Max}} \ln f(\underline{V}|\beta \neq 0) = \frac{-n}{2} \ln (2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n} V_i^2 + \frac{1}{2\sigma^2} \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\, S(\hat{a}t_i - \hat{t_0})\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t_0})]^2} . \tag{2.2.7}$$

where $\hat{a}$, $\hat{t_0}$ are the quantities which maximize the expression (2.2.6). To find the value of σ^2 which maximizes $\ln f(\underline{V}|\beta \neq 0)$, we differentiate Equation (2.2.7).

$$0 = \frac{d}{d\sigma^2} \underset{a,t_0,\beta}{\text{Max}} \ln f(\underline{V}|\beta \neq 0)$$

$$0 = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} V_i^2 - \frac{1}{2\sigma^4} \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\, S(\hat{a}t_i - \hat{t_0})\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t_0})]^2} \tag{2.2.8}$$

or

$$\hat{\sigma^2} \equiv \frac{1}{n}\left\{\sum_{i=1}^{n} V_i^2 - \frac{\left[\sum_{i=n_0+1}^{n_1} V_1\, S(\hat{a}t_i - \hat{t_0})\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t_0})]^2}\right\} .$$

So the second term of Equation (2.2.2) for $\ln \lambda$ is obtained by substituting this value for $\hat{\sigma^2}$ into Equation (2.2.7), or

$$\underset{\beta,a,t_0,\sigma^2}{\text{Max}} \ln f(\underline{V}|\beta \neq 0) = -\frac{n}{2} \ln (2\pi\hat{\sigma^2}) - \frac{1}{2\hat{\sigma^2}}\left\{\sum_{i=1}^{n} V_i^2 - \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\, S(\hat{a}t_i - \hat{t_0})\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t_0})]^2}\right\} .$$

$$\underset{\beta,a,t_0,\sigma^2}{\text{Max}} \ln f(\underline{V}|\beta \neq 0) = -\frac{n}{2} \ln (2\pi\hat{\sigma^2}) - \frac{n}{2} \tag{2.2.9}$$

Now the first term of $\ln \lambda$ is

$$\underset{a,t_0,\sigma^2}{\text{Max}} \ln f(\underline{V}|\beta = 0) = \underset{\sigma^2}{\text{Max}} \left[-\frac{n}{2} \ln (2\pi\sigma^2) - \frac{1}{2\pi\sigma^2}\sum_{i=1}^{n} V_i^2\right] .$$

Taking the derivative with respect to σ^2, we get

$$0 = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n} V_i^2 ,$$

or

$$\tilde{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} V_i^2 , \tag{2.2.10}$$

and

$$\underset{a,t_0,\sigma^2}{\text{Max}} \ln f(\underline{V}|\beta = 0) = -\frac{n}{2} \ln (2\pi\tilde{\sigma^2}) - \frac{n}{2} . \tag{2.2.11}$$

Finally we find ln λ by subtracting Equation (2.2.9) from Equation (2.2.11).

$$\ln\lambda = \frac{n}{2}\ln(2\pi\hat{\sigma}^2) - \frac{n}{2}\ln(2\pi\tilde{\sigma}^2)$$

$$= \frac{n}{2}\ln(\hat{\sigma}^2/\tilde{\sigma}^2)$$

$$-2\ln\lambda = -n\ln\left\{1 - \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\, S(\hat{a}t_i - \hat{t}_0)\right]^2}{\sum_{i=1}^{n} V_i^2 \sum_{j=n_0+1}^{n_1} [S(\hat{a}t_j - \hat{t}_0)]^2}\right\} \qquad (2.2.12)$$

Now since any statistical uncertainty involved in this problem arises from the noise samples V_i which we have assumed to be normally distributed, it will be assumed that regularity conditions hold, hence the joint maximum likelihood estimators $\hat{\beta}$, $\hat{\sigma}^2$, $\hat{a}$, $\hat{t}_0$ tend to multivariate normal random variates and are asymptotically efficient (Kendall[3], page 54).

With these conditions, the asymptotic distribution of the test statistic ($-2\ln\lambda$) under the null hypothesis is central χ^2 with one degree of freedom, a result originally due to Wilks[12]. The single degree of freedom is implied because only the value of the scalar parameter β was tested.

Using the test statistic in the form of Equation (2.2.12) permits us to choose the detection threshold or critical value for the test independent of the noise power or length of record n.

From a table of the χ^2 distribution (Owen[8], page 50) the critical value can be read as a function of γ, which is the significance level of the detection test, or one minus the single scan false alarm probability. For example, a false alarm probability of 0.01 implies $\gamma = 0.99$ and the detection threshold is $\ell = 6.635$. Anytime the test statistic (2.2.12) exceeds 6.635, a signal detection will be declared. Values of the detection threshold for three other typical false alarm rates are shown at the top of Figure 2.2.

FALSE ALARM PROBABILITY	0.1	0.05	0.01	0.001
DETECTION THRESHOLD ℓ	2.690	3.840	6.635	10.839

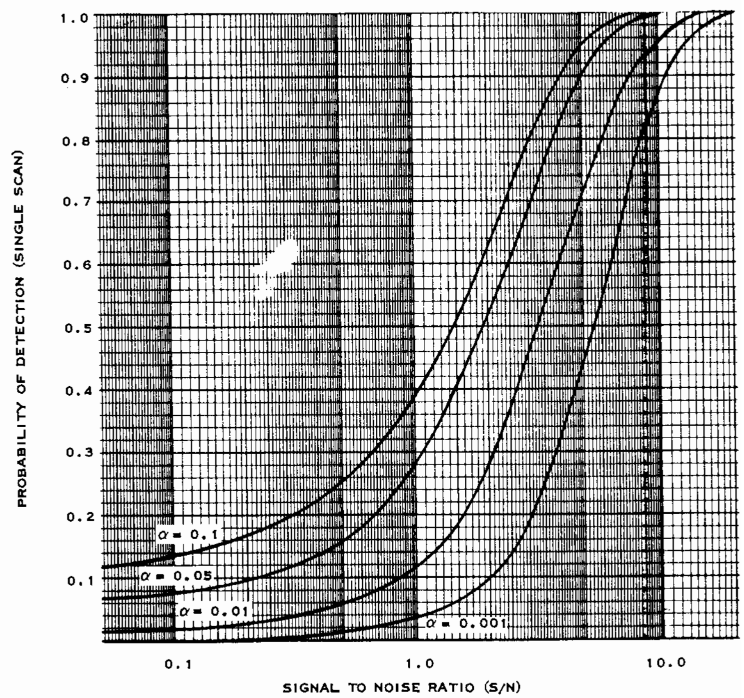


Figure 2.2 Probability of Detection Versus (S/N)

Alternately, the test can be expressed in terms of the sample correlation coefficient

$$r^2 = \frac{\left[\sum_{i=n_0+1}^{n_1} V_i\, S(\hat{a}t_i - \hat{t}_0)\right]^2}{\sum_{i=1}^{n} V_i^2 \sum_{j=n_0+1}^{n_1} [S(\hat{a}t_j - \hat{t}_0)]^2} \qquad (2.2.13)$$

From Equation (2.2.12), we see that the critical region $-2 \ln \lambda \geqslant \ell$ is equivalent to

$$r^2 \geqslant 1 - e^{-\ell/n}. \qquad (2.2.14)$$

Consequently, a signal detection could be declared anytime the sample correlation coefficient r^2 exceeds this value.

The probability of detection of a given signal with a given amplitude β_a, or the power of the test at $\beta = \beta_a$ can also be evaluated. Kendall[3] gives the procedure on page 231. Since there is only one degree of freedom involved in this test, the power can be obtained from the χ^2 distribution function. Under the alternate hypothesis $\beta \neq 0$, the test statistic $(-2 \ln \lambda)$ is asymptotically distributed as a noncentral χ^2 with one degree of freedom and noncentrality parameter

$$V_{\beta\beta}^{-1} (\beta - \beta_0)^2 = V_{\beta\beta}^{-1} \cdot \beta^2$$

where $V_{\beta\beta}$ is the variance of the ML estimator β. The inverse of $V_{\beta\beta}$ is shown in Chapter III of this dissertation to be given by

$$V_{\beta\beta}^{-1} = -E\left(\frac{\partial^2 \ln L}{\partial \beta^2}\right) = \frac{1}{\sigma^2} \sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2 .$$

So the noncentrality parameter above becomes

$$V_{\beta\beta}^{-1}(\beta - \beta_0)^2 = \frac{\beta^2}{\sigma^2} \sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2$$

Evaluated at $\beta = \beta_a$, the given signal coefficient, this noncentrality parameter becomes

$$\frac{\sum_{n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\sigma^2}.$$

We will now relate this noncentrality parameter to the signal-to-noise ratio which is defined to be the total signal energy divided by the average noise power per unit bandwidth, and for which we will write S/N.

$$\frac{\sum_{n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\sigma^2} = \frac{\sum_{n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\frac{1}{n} E\left(\sum_{1}^{n} N_i^2\right)}$$

Multiplying numerator and denominator by the time increment Δt, the sums become approximations for integrals, and the noncentrality parameter can be written as

$$\frac{\sum_{n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\sigma^2} = \frac{\int_{t_0/a}^{(t_0+\tau)/a} [\beta_a S(at - t_0)]^2 \, dt}{\frac{1}{n} E\left[\int_0^T N^2(t) \, dt\right]}$$

Multiplying the identity $(n/T) \cdot \Delta t = 1$ into the denominator, we have

$$\frac{\sum_{n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\sigma^2} = \frac{\int_{t_0/a}^{(t_0+\tau)/a} [\beta_a S(at_i - t_0)]^2 dt}{\Delta t \, E\left[\frac{1}{T}\int_0^T N^2(t) \, dt\right]}$$

The numerator of the right side is seen to be the total signal energy, or the S part of S/N. The expectation in the denominator is seen to be the average noise power, or energy per unit time. Let N* denote this total average noise power.

$$N^* = E\left[\frac{1}{T}\int_0^T N^2(t) \, dt\right]$$

Then the noncentrality parameter can be written

$$\frac{\sum_{i=n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2}{\sigma^2} = \frac{S}{\Delta t \cdot N^*}$$

Since the data is sampled at a rate of $1/\Delta t$, the total average noise power N* appears spread over the Nyquist bandwidth of $1/2\Delta t$. Consequently, the average noise power per unit bandwidth, or the N part of S/N is equal to the average noise power N* divided by the bandwidth $1/2\Delta t$.

$$N = 2 \cdot \Delta t \cdot N^*$$

The noncentrality parameter becomes

$$\frac{1}{\sigma^2} \sum_{i=n_0+1}^{n_1} [\beta_a S(at_i - t_0)]^2 = 2 \cdot (S/N) \cdot$$

So the probability of detection (on a single scan) of any signal in white Gaussian noise is seen to depend only on the signal-to-noise ratio, and not on the form of the signal. The values of this probability of detection or power of the test can be obtained from the χ^2 distribution function. From Kendall[3], page 231, we write the expression for the power in terms of the noncentrality parameter 2 (S/N) for a LR detector with a single scan false alarm probability of α. Substituting 2 (S/N) for the noncentrality parameter in Kendall's expression, we have

$$P = \int_{\left[\frac{1+2(S/N)}{1+4(S/N)}\right]\chi^2_\alpha(1)}^{\infty} d\chi^2 \left[1 + \frac{4(S/N)^2}{1 + 4(S/N)}\right] \qquad (2.2.15)$$

The curves in Figure 2.2 show this single scan probability of detection as a function of signal-to-noise ratio (S/N) for several values of the single scan false alarm probability α. The values were obtained by a two-way linear interpolation between the values given in the Biometrika Tables[9], page 122.

2.3 Detection by Multiple Observations

Suppose the same signal of known form is contained in several segments of independent noise, as is the case in active sonar and in some passive applications. That is,

$$V_1(t) = N_1(t) + \beta S(at - t_0)$$
$$V_2(t) = N_2(t) + \beta S(at - t_0)$$
$$\vdots$$
$$V_m(t) = N_m(t) + \beta S(at - t_0)$$

where again β, a, t_0 are unknown parameters to be estimated and $N_1(t), N_2(t), \ldots, N_m(t)$ are independent Gaussian processes.

Again suppose that these functions are sampled at times $t_i = i \cdot \Delta t$ yielding the samples $V_{k,i} = V_k(t_i)$, and suppose that the noise samples $N_{k,i} = N_k(t_i)$ are uncorrelated.

We first show that the numbers S_v^2 and V_i, $i = 1, 2, \ldots, n$ are sufficient statistics, when

$$V_i = \sum_{k=1}^{m} V_{k,i}, \qquad i = 1, 2, \ldots, n$$

$$S_v^2 = \sum_{k=1}^{m} \sum_{i=1}^{n} (V_{k,i})^2 \cdot$$

The joint likelihood of the $V_{k,i}$ is

$$L = L_1 \cdot L_2 \cdot L_3$$

where

$$L_1 = \prod_{i=1}^{n_0} \prod_{k=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-V_{k,i}^2/2\sigma^2}$$

$$L_2 = \prod_{i=n_0+1}^{n_1} \prod_{k=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(1/2\sigma^2)[V_{k,i} - \beta S(at_i - t_0)]^2}$$

$$L_3 = \prod_{i=n_1+1}^{n} \prod_{k=1}^{m} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-V_{k,i}^2/2\sigma^2}$$

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{nm} \cdot \left\{ \begin{array}{l} \displaystyle\prod_{i=1}^{n_0} e^{-1/2\sigma^2 \sum_{k=1}^{m} V_{k,i}^2} \\ \displaystyle\cdot \prod_{i=n_0+1}^{n_1} e^{-1/2\sigma^2 \sum_{k=1}^{m} [V_{k,i} - \beta S(at_i - t_0)]^2} \\ \displaystyle\cdot \prod_{i=n_1+1}^{n} e^{-1/2\sigma^2 \sum_{k=1}^{m} V_{k,i}^2} \end{array} \right\}$$

Squaring the quadratic factor and combining terms, we get

$$L = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{nm} \cdot \left\{ \begin{array}{l} \prod_{i=1}^{n} e^{-1/2\sigma^2 \sum_{k=1}^{m} V_{k,i}^2} \\ \cdot \prod_{i=n_0+1}^{n_1} e^{(1/\sigma^2)\beta S(at_i - t_0) \sum_{k=1}^{m} V_{k,i}} \\ \cdot \prod_{i=n_0+1}^{n_1} e^{-(m/2\sigma^2)\beta^2 [S(at_i - t_0)]^2} \end{array} \right\} .$$

$$L = (2\pi\sigma^2)^{-nm/2} \cdot \left\{ \begin{array}{l} e^{-(1/2\sigma^2) S_v^2} \\ \cdot \prod_{i=n_0+1}^{n_1} e^{(1/\sigma^2)\beta S(at_i - t_0) V_i} \\ \cdot \prod_{i=n_n+1}^{n_1} e^{-(m/2\sigma^2)\beta^2 [S(at_i - t_0)]^2} \end{array} \right\}$$

Since the joint likelihood function depends on the data only through the numbers S_v^2 and V_i, $i = 1, 2, \ldots, n$, these are shown to be jointly sufficient statistics. Operationally, this means that one can summarize the data in this way, namely, compute

$$S_v^2 = \sum_{k=1}^{m} \prod_{i=1}^{n} V_{k,i}^2$$

and

$$V_i = \sum_{k=1}^{m} V_{k,i}, \qquad i = 1, 2, \ldots, n \quad \text{, since}$$

the likelihood ratio statistic must be a function of these sufficient statistics. The derivation of the test in this form will proceed almost exactly as in the previous case.

CHAPTER III

ESTIMATION OF SIGNAL PARAMETERS

3.1 The Estimators

The next problem to be discussed is that of estimating the signal parameters, once it is decided that a signal is present. Two of the most powerful techniques for estimating unknown parameters in the presence of statistical uncertainty are the method of maximum likelihood and the method of least squares. Since we have assumed the noise variables N_i in the model $V_i = N_i + \beta S(at_i - t_0)$ to be normally distributed, these two methods produce identical results in this problem. This can be seen in Equation (2.2.3). The quantities $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$ which minimize the quantity

$$\sum_{i=n_0+1}^{n_1} [V_i - \beta S(at_i - t_0)]^2$$

will also maximize the log likelihood function $\ln f(\underline{V}|\beta \neq 0)$. The approach chosen here will be the method of maximum likelihood. The maximum likelihood estimator of the unknown parameters β, a, t_0, σ^2 is defined to be the set of numbers $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$, $\widehat{\sigma^2}$ which maximizes the likelihood function, evaluated at the observed values V_i, $i = 1, 2, \ldots, n$. This is exactly the same set of parameters that was chosen to maximize the denominator of the likelihood ratio in Equation (2.2.1), so most of the work of finding the maximum likelihood estimators was done in Chapter II. The results will be restated below.

From Equation (2.2.6) we get the ML estimators $\hat{a}$ and $\hat{t}_0$. The values of a and t_0 which maximize $f(\underline{V}|\beta = 0)$ are the $\hat{a}$ and $\hat{t}_0$ which maximize the expression (2.2.6). So

$$\frac{\left[\sum_{i=n_0+1}^{n_1} V_i S(\hat{a}t_i - \hat{t}_0)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t}_0)]^2} = \underset{a,\,t_0}{\text{Max}} \frac{\left[\sum_{i=n_0+1}^{n_1} V_i S(at_i - t_0)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2} \qquad (3.1.1)$$

From Equation (2.2.8) we restate the ML estimator of σ^2.

$$\widehat{\sigma^2} = \frac{1}{n}\left\{\sum_{i=1}^{n} V_i^2 - \frac{\left[\sum_{i=n_0+1}^{n_1} V_i S(\hat{a}t_i - \hat{t}_0)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t}_0)]^2}\right\} \qquad (3.1.2)$$

And finally, from Equation (2.2.4) it can be inferred that the ML estimator for β is given by

$$\hat{\beta} = \frac{\sum_{i=n_0+1}^{n_1} V_i S(\hat{a}t_i - \hat{t}_0)}{\sum_{i=n_0+1}^{n_1} [S(\hat{a}t_i - \hat{t}_0)]^2} \,. \qquad (3.1.3)$$

Under the general regularity conditions, maximum likelihood estimators are consistent and efficient for large n. Furthermore the joint ML estimators tend to a multivariate normal distribution as n gets large. These results are given in Kendall[3], page 54. The variance-covariance matrix of this limiting distribution is calculated in part 3 of this chapter.

3.2 Successive Approximation of Estimators

The estimators $\hat{\beta}$ and $\hat{\sigma^2}$ have been obtained in explicit form above, so their computation presents no particular difficulty once the values of $\hat{a}$ and $\hat{t}_0$ are known. No method for calculating these two estimators is given, however, since they are defined above to be those values of a and t_0 which maximize the right side of Equation (3.1.1). Conceivably one could perform an exhaustive grid search, that is, to calculate the quantity

$$Q(a, \tau) = \frac{\left[\sum_{i=n_0+1}^{n_1} V_i S(at_i - \tau)\right]^2}{\sum_{i=n_0+1}^{n_1} [S(at_i - \tau)]^2} \tag{3.2.1}$$

for each of a large number of discrete number pairs a, τ and select the largest. Since a two-parameter search can require excessive computations, an iterative technique is suggested to reduce the number of computations. A sparse grid search would still be required to provide a starting point for the iterative procedure. Newton's method for finding the maximum value of Equation (3.2.1) will be developed below.

Suppose a_i, τ_i are the values at the *ith* stage of the iterative process which are considered to be fairly close to the ML estimators $\hat{a}$ and $\hat{t}_0$, and it is desired to find new values a_{i+1} and τ_{i+1} which are even closer. The truncated Taylor Series for Q(a, τ) about the point (a_i, τ_i) is

$$Q(a, \tau) = Q(a_i, \tau_i) + (a - a_i)\left.\frac{\partial Q}{\partial a}\right|_{a_i,\tau_i} + (\tau - \tau_i)\left.\frac{\partial Q}{\partial \tau}\right|_{a_i,\tau_i}$$
$$+ \frac{1}{2}\left\{(a - a_i)^2\left.\frac{\partial^2 Q}{\partial a^2}\right|_{a_i,\tau_i} + (\tau - \tau_i)^2\left.\frac{\partial^2 Q}{\partial \tau^2}\right|_{a_i,\tau_i}\right\} + (a - a_i)(\tau - \tau_i)\left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i}$$

Differentiating with respect to a and τ, we get the two equations

$$\frac{\partial Q}{\partial a} = \left.\frac{\partial Q}{\partial a}\right|_{a_i,\tau_i} + (a - a_i)\left.\frac{\partial^2 Q}{\partial a^2}\right|_{a_i,\tau_i} + (\tau - \tau_i)\left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i}$$

$$\frac{\partial Q}{\partial \tau} = \left.\frac{\partial Q}{\partial \tau}\right|_{a_i,\tau_i} + (a - a_i)\left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i} + (\tau - \tau_i)\left.\frac{\partial^2 Q}{\partial \tau^2}\right|_{a_i,\tau_i}$$

We wish to find the point (a, τ) such that the two derivatives on the left side of these equations are zero, since this is a necessary condition for Q to have its maximum value at (a, τ). Accordingly, we shall choose the next values in the approximation process to be the values a_{i+1}, τ_{i+1} which make the two derivatives equal to zero at that point.

$$0 = \left.\frac{\partial Q}{\partial a}\right|_{a_i,\tau_i} + (a_{i+1} - a_i)\left.\frac{\partial^2 Q}{\partial a^2}\right|_{a_i,\tau_i} + (\tau_{i+1} - \tau_i)\left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i}$$
$$0 = \left.\frac{\partial Q}{\partial \tau}\right|_{a_i,\tau_i} + (a_{i+1} - a_i)\left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i} + (\tau_{i+1} - \tau_i)\left.\frac{\partial^2 Q}{\partial \tau^2}\right|_{a_i,\tau_i}$$

We can now solve the above 2 X 2 linear system to find the two increments ($a_{i+1} - a_i$) and ($\tau_{i+1} - \tau_i$) necessary to find the new approximations for the point (a, τ) at which the two derivatives are zero. This condition alone does not guarantee the maximum, but, if the sparse grid search has located the starting point close enough to the maximum, then the "stationary point," (a, τ), approximated in this way will actually be the ML estimator ($\hat{a}$, $\hat{t}_0$).

Let

$$Q_a = \left.\frac{\partial Q}{\partial a}\right|_{a_i,\tau_i} \qquad Q_\tau = \left.\frac{\partial Q}{\partial \tau}\right|_{a_i,\tau_i} \qquad Q_{a\tau} = \left.\frac{\partial^2 Q}{\partial a \partial \tau}\right|_{a_i,\tau_i}$$
$$Q_{aa} = \left.\frac{\partial^2 Q}{\partial a^2}\right|_{a_i,\tau_i} \qquad Q_{\tau\tau} = \left.\frac{\partial^2 Q}{\partial \tau^2}\right|_{a_i,\tau_i}$$

Then the solutions of the linear system are:

$$\tau_{i+1} - \tau_i = (Q_\tau \cdot Q_{aa} - Q_a \cdot Q_{a\tau})/D$$

$$a_{i+1} - a_i = (Q_a \cdot Q_{\tau\tau} - Q_\tau \cdot Q_{a\tau})/D$$

where

$$D = (Q_{a\tau} \cdot Q_{a\tau} - Q_{\tau\tau} \cdot Q_{aa}) \ .$$

So if the *ith* approximation (a_i, τ_i) is close to the desired values $(\hat{a}, \hat{t}_0)$ which maximize Q (and also ln L), then the values (a_{i+1}, τ_{i+1}) computed by adding the increments above should be even closer. The derivatives above can be evaluated numerically. Since these are functions of the observations, they are themselves random variables, and the probability is zero that the denominator D above will be zero. If this ever happens, however, the values of the iterates a_i, τ_i can be changed a small amount, and the process can proceed. This successive approximation technique for ML estimators has been used by the author in detecting signals of the type shown in Figure 2.1 with no particular difficulty. As signal-to-noise ratios get lower, however, there are more local maxima, so the starting points must be chosen with more care. When S/N gets lower than 0.1, the process occasionally will not even converge.

3.3 Distribution of Estimators

The estimators $\widehat{\sigma^2}$, $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$ are the coordinates of the point at which the likelihood function reaches its maximum. Since the likelihood function also depends on the noise values N_i observed, the peak will be displaced in a random manner. The estimators, which are measures of the location of the peak, are therefore random variables. In this section we will derive the variances and covariances of $\widehat{\sigma^2}$, $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$.

From Kendall[3] page 54, we see that the joint ML estimators tend, under regularity conditions, to a multivariate normal distribution, with dispersion matrix whose inverse is given by

$$(V_{rs}^{-r}) = -E\left(\frac{\partial^2 \ln L}{\partial\theta_r \partial\theta_s}\right) = E\left(\frac{\partial \ln L}{\partial\theta_r} \cdot \frac{\partial \ln L}{\partial\theta_s}\right) \quad , \qquad (3.3.1)$$

Remembering the expression

$$\ln L = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n_0} V_i^2 - \frac{1}{2\sigma^2}\sum_{i=n_1+1}^{n} V_i^2 - \frac{1}{2\sigma^2}\sum_{i=n_0+1}^{n} [V_i - \beta S(at_i - t_0)]^2 \qquad (3.3.2)$$

we begin the derivation by writing the four derivatives

$$\frac{\partial \ln L}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=n_0+1}^{n_1} S(at_i - t_0)\,[V_i - \beta S(at_i - t_0)] \qquad (3.3.3)$$

$$\frac{\partial \ln L}{\partial a} = \frac{\beta}{\sigma^2}\sum_{i=n_0+1}^{n_1} t_i\, S'(at_i - t_0)\,[V_i - \beta S(at_i - t_0)] \qquad (3.3.4)$$

$$\frac{\partial \ln L}{\partial t_0} = \frac{-\beta}{\sigma^2}\sum_{i=n_0+1}^{n_1} S'(at_i - t_0)\,[V_i - \beta S(at_i - t_0)] \qquad (3.3.5)$$

$$\frac{\partial \ln L}{\partial\sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\left\{\sum_{i=1}^{n_0} V_i^2 + \sum_{i=n_1+1}^{n} V_i^2 + \sum_{i=n_0+1}^{n_1} [V_i - \beta S(at_i - t_0)]\right\}^2 \qquad (3.3.6)$$

We will first show that the estimator $\widehat{\sigma^2}$ is asymptotically independent of the other three estimators $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$. That is, all of the covariances (V_{β,σ^2}), (V_{a,σ^2}), (V_{t_0,σ^2}) are zero.

$$V_{\beta,\sigma^2}^{-1} = -E\left(\frac{\partial^2 \ln L}{\partial\beta\partial\sigma^2}\right)$$

$$= -E\left\{\frac{\partial}{\partial\beta}\,\frac{1}{2\sigma^4}\sum_{i=n_0+1}^{n_1}[V_i - \beta S(at_i - t_0)]^2\right\}$$

$$= E\left\{\frac{1}{\sigma^4}\sum_{i=n_0+1}^{n_1} S(at_i - t_0)\,[V_i - \beta S(at_i - t_0)]\right\}$$

$$= E\left\{\frac{1}{\sigma^4}\sum_{i=n_0+1}^{n_1} S(at_i - t_0)\,[N_i]\right\}$$

$$= \frac{1}{\sigma^4}\sum_{i=n_0+1}^{n_1} S(at_i - t_0)\cdot E(N_i)$$

$$V_{\beta,\sigma^2}^{-1} = 0$$

Similar steps will show the other two inverse elements (V_{a,σ^2}^{-1}) and (V_{t_0,σ^2}^{-1}) to be zero.

So the inverse of the dispersion matrix has the form

$$V^{-1} = \left[\begin{array}{c|ccc} V_{\sigma^2,\sigma^2}^{-1} & 0 & 0 & 0 \\ \hline 0 & V_{\beta,\beta}^{-1} & V_{\beta,a}^{-1} & V_{\beta,t_0}^{-1} \\ 0 & V_{\beta,a}^{-1} & V_{a,a}^{-1} & V_{a,t_0}^{-1} \\ 0 & V_{\beta,t_0}^{-1} & V_{a,t_0}^{-1} & V_{t_0,t_0}^{-1} \end{array}\right]$$

$$= \left[\begin{array}{c|c} V_{\sigma^2,\sigma^2}^{-1} & \underline{0}' \\ \hline \underline{0} & A^{-1} \end{array}\right] \qquad (3.3.7)$$

It follows that the dispersion matrix itself is of the form

$$V = \begin{bmatrix} V_{\sigma^2,\sigma^2} & \underline{0}' \\ \underline{0} & A \end{bmatrix} \qquad (3.3.8)$$

and consequently the three covariances involving σ^2 are zero. The variance of $\widehat{\sigma^2}$ can now be evaluated simply.

$$V_{\sigma^2,\sigma^2}^{-1} = -E\left(\frac{\partial^2 \ln L}{\partial\sigma^2}\right)$$

Substituting in the derivative of Equation (3.3.6) with respect to σ^2, we have

$$V_{\sigma^2,\sigma^2}^{-1} = -E\left(\frac{n}{2\sigma^4}\right) + \frac{1}{\sigma^6}E\left\{\sum_{i=1}^{n_0} V_i^2 + \sum_{i=n_1+1}^{n} V_i + \sum_{i=n_0+1}^{n_1}[V_i - \beta S(at_i - t_0)]^2\right\}$$

$$= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}E\left\{\sum_{i=1}^{n}(N_i)^2\right\}$$

$$= -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\left\{n\sigma^2\right\}$$

$$V_{\sigma^2,\sigma^2}^{-1} = \frac{n}{2\sigma^4}$$

Because Equation (3.3.7) contains the null vectors, it follows that the large sample variance of $\widehat{\sigma^2}$ is given by

$$\mathrm{Var}\,(\widehat{\sigma^2}) = V_{\sigma^2,\sigma^2} = \frac{2\sigma^4}{n} \quad . \tag{3.3.9}$$

The other elements of the inverse matrix A^{-1} can be found by similar steps of differentiation and expectation. The results are given below. To facilitate the expression of results, we shall adopt the shorthand vector notation indicated. Also $S'(t)$ denotes d/dt $S(t)$.

$$V_{\beta\beta}^{-1} = \frac{1}{\sigma^2} \sum_{i=n_0+1}^{n_1} [S(at_i - t_0)]^2 = \frac{1}{\sigma^2} (\underline{S})^T (\underline{S}) \tag{3.3.10}$$

$$\begin{aligned} V_{\beta a}^{-1} &= \frac{\beta}{\sigma^2} \sum_{i=n_0+1}^{n_1} t_i\, S'(at_i - t_0)\, S(at_i - t_0) \\ &= \frac{\beta}{\sigma^2} (\underline{tS'})^T (\underline{S}) \end{aligned} \tag{3.3.11}$$

$$\begin{aligned} V_{\beta,t_0}^{-1} &= -\frac{\beta}{\sigma^2} \sum_{i=n_0+1}^{n_1} S'(at_i - t_0)\, S(at_i - t_0) \\ &= -\frac{\beta}{\sigma^2} (\underline{S}')^T (\underline{S}) \end{aligned} \tag{3.3.12}$$

$$\begin{aligned} V_{a,a}^{-1} &= \frac{\beta^2}{\sigma^2} \sum_{i=n_0+1}^{n_1} t_i^2\, [S'(at_i - t_0)]^2 \\ &= \frac{\beta^2}{\sigma^2} (\underline{tS'})^T (\underline{tS'}) \end{aligned} \tag{3.3.13}$$

$$\begin{aligned} V_{a,t_0}^{-1} &= -\frac{\beta^2}{\sigma^2} \sum_{i=n_0+1}^{n_1} t_i\, [S'(at_i - t_0)]^2 \\ &= -\frac{\beta^2}{\sigma^2} (\underline{tS'})^T (\underline{S}') \end{aligned} \tag{3.3.14}$$

$$\begin{aligned} V_{t_0,t_0}^{-1} &= \frac{\beta^2}{\sigma^2} \sum_{i=n_0+1}^{n_1} [S'(at_i - t_0)]^2 \\ &= \frac{\beta^2}{\sigma^2} (\underline{S}')^T (\underline{S}') \end{aligned} \tag{3.3.15}$$

Since each of the elements above of the matrix A^{-1} has the factor $1/\sigma^2$, we shall write

$$A^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} (\underline{S})^T(\underline{S}) & \beta(\underline{S})^T(\underline{tS'}) & -\beta(\underline{S})^T(\underline{S}') \\ \beta(\underline{tS'})^T(\underline{S}) & \beta^2(\underline{tS'})^T(\underline{tS'}) & -\beta^2(\underline{tS'})^T(\underline{S}') \\ -\beta(\underline{S}')^T(\underline{S}) & -\beta^2(\underline{S}')^T(\underline{tS'}) & \beta^2(\underline{S}')^T(\underline{S}') \end{bmatrix} \tag{3.3.16}$$

$$A^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} (\underline{S})^T \\ \beta(\underline{tS'})^T \\ -\beta(\underline{S}')^T \end{bmatrix} \cdot [(S),\ \beta(\underline{tS'}),\ -\beta(S')] \tag{3.3.17}$$

When the signal form $S(t)$ is specified, the matrix A of variances and covariances of the estimators can be calculated for various values of the true parameters β, a, t_0. The fact that the joint ML estimators are consistent implies that as the number of elements in the vector $(\underline{S})$ gets large, the bias of each of the estimators approaches zero. So if the sampling rate is quite large (or the vector $\underline{S}$ has many elements) the joint ML estimators $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$ are approximately distributed according to the multivariate normal distribution with mean (β, a, t_0) and variance-covariance matrix A. The joint density is given by

$$\frac{|A^{-1}|^{1/2}}{(2\pi)^{3/2}} e^{-1/2} (\hat{\beta} - \beta, \hat{a} - a, \hat{t}_0 - t_0) A^{-1} \begin{pmatrix} \hat{\beta} - \beta \\ \hat{a} - a \\ \hat{t}_0 - t_0 \end{pmatrix}$$

Substituting Equation (3.3.17) for A^{-1}, the quadratic form in the exponential becomes

$$\frac{1}{\sigma^2} (\hat{\beta} - \beta, \hat{a} - a, \hat{t}_0 - t_0) \begin{bmatrix} (\underline{S})^T \\ \beta(t\underline{S}')^T \\ -\beta(\underline{S}')^T \end{bmatrix} [(\underline{S}), \beta(\underline{tS}'), -\beta(\underline{S}')] \begin{bmatrix} \hat{\beta} - \beta \\ \hat{a} - a \\ \hat{t}_0 - t_0 \end{bmatrix}$$

Or, letting Q denote the quadratic form, we have

$$Q = \frac{1}{\sigma^2} \sum_{i=n_0+1}^{n_1} \left\{ (\hat{\beta} - \beta) S(at_i - t_0) + \beta[(\hat{a} - a)t_i - (\hat{t}_0 - t_0)] S'(at_i - t_0) \right\}^2 \qquad (3.3.18)$$

where

$$S'(t) = \frac{d}{dt} S(t)$$

So the limiting joint distribution of the estimators $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$ can be written as

$$\frac{|A^{-1}|^{1/2}}{(2\pi)^{3/2}} e^{-(1/2) \cdot Q}$$

where Q is specified in Equation (3.3.18). Evaluation of the determinant $|A^{-1}|$ from Equation (3.3.17) would then allow one to use this large sample approximation for the

probability that the error in any one of the parameter estimates would exceed a given value.

Since $\hat{\sigma^2}$ was shown to be independent of $\hat{\beta}$, $\hat{a}$, $\hat{t}_0$, its limiting distribution is normal with mean σ^2 and variance $2\sigma^4/n$.

CHAPTER IV

CLASSIFICATION OF SIGNALS

4.1 Classification Based on Parameter Estimates

The diagram in Figure 1.1 was referred to as an example of the problems encountered in the reception process. It will become apparent in this chapter that this example is not general enough to represent all of the problems that might arise when more than one class of signals is considered.

The situation visualized in Figure 1.1 is perhaps the simpliest of the classification problems which involve linear scale transformations. That is, there is only one signal form S(t), but the signal could have been generated from one of several sources. As the diagram indicates, the signal will first be detected, its parameters estimated, and then judgement must be made, on the basis of the parameter estimates, as to which source generated the received signal. To make this judgement with a reasonable degree of success, there must be some difference between the parameters of the signals generated from different sources. This difference can be either deterministic or statistical. In this section it will be assumed that the true parameters β, a, t_0 of the detected signal are themselves chance variables, samples from a distribution characteristic of the source class from which the signal was generated. The distribution of the true signal parameters β, a, t_0 of signals generated from the *kth* source will be assumed to be multivariate normal with mean (β_k, a_k, t_{0k}) and variance-covariance matrix Σ_k. If the signal does indeed belong to the *kth* source class, then the joint conditional distribution $g(\beta, a, t_0|k)$ is given by

$$\frac{1}{(2\pi)^{3/2}\ |\Sigma_k^{1/2}|}\, e^{-1/2\ (\beta - \beta_k,\ a - a_k,\ t_0 - t_{0k})\ \Sigma_k^{-1}\ \begin{pmatrix} \beta - \beta_k \\ a - a_k \\ t_0 - t_{0k} \end{pmatrix}}$$

But since the joint conditional distribution $f(\hat\beta, \hat a, \hat t_0|\beta, a, t_0)$ is given in Equation (3.3.19), we can write the probability distribution of the estimators $(\hat\beta, \hat a, \hat t_0)$ conditional upon k, the index of the source class.

$$L(\hat\beta, \hat a, \hat t_0|k) = \int_{\beta}\int_{a}\int_{t_0} f(\hat\beta, \hat a, \hat t_0|\beta, a, t_0) \cdot g(\beta, a, t_0|k)\, d\beta \cdot da \cdot dt_0 \qquad (4.1.2)$$

Once the likelihood functions above have been obtained, for all source classes k = 1, 2, . . . M, a variety of classification procedures are applicable. We shall present here the Bayes classification rule, because it is general enough to include several of the other methods. This procedure assumes a great deal of prior knowledge about various elements of the problem.

Let $p_1, p_2, \ldots, p_M$ denote the prior probability of occurrence or relative frequency of occurrence of the M source classes. Let C(j|i) denote the cost of classifying an observation into class j when the signal really came from source class i. We assume that the p_i and the C(j|i) are known for all i and j. Suppose that a signal is detected, and its parameters are estimated to be $(\hat\beta, \hat a, \hat t_0)$. The conditional probability of this signal having come from source k is

$$P(k|\hat\beta\, \hat a, \hat t_0)$$

Using Bayes rule, this can be expressed as

$$\frac{p_k\ L(\hat\beta, \hat a, \hat t_0|k)}{\sum_{i=1}^{M} p_i\ L(\hat\beta, \hat a, \hat t_0|i)}$$

If we classify this signal as having come from source j, the conditional expected loss for making this decision given the estimated parameters $(\hat{\beta}, \hat{a}, \hat{t}_0)$ is

$$\sum_{k=1}^{M} C(j|k)\, P(k|\hat{\beta}, \hat{a}, \hat{t}_0) = \sum_{k=1}^{M} C(j|k) \frac{p_k\, L(\hat{\beta}, \hat{a}, \hat{t}_0|k)}{\sum_{i=1}^{M} p_i\, L(\hat{\beta}, \hat{a}, \hat{t}_0|i)} . \qquad (4.1.3)$$

We minimize this expected loss if we choose j so as to minimize Equation (4.1.3). All that is necessary is to calculate the term

$$\sum_{k=1}^{M} C(j|k)\, p_k\, L(\hat{\beta}, \hat{a}, \hat{t}_0|k)$$

for all j and select the j which gives the minimum. This is the Bayes classification procedure for minimizing the expected loss. It is customary to think of this rule as dividing the space of observations into m mutually exclusive and exhaustive regions R_1, $R_2, \ldots, R_M$ such that if an observation falls within region R_k the Bayes rule will assign it to source class k. The hope is that this mapping will result in a convenient expression of the classification rule, as for example a linear discriminant. In the modern age of high speed computers, however, the advantages offered by such a mapping are not as important as they once were. In some cases the expression (4.1.4) itself can be used as a computational procedure without difficulty.

An important subcase of the Bayes rule occurs for the simple loss function, defined to be

$$C(j|i) = \begin{Bmatrix} 0, \text{ if } i = j \\ 1, \text{ if } i \neq j \end{Bmatrix} .$$

In this case, the computation can be simplified somewhat. The summation (4.1.4) can be written as

$$\sum_{k=1}^{M} p_k\, L(\hat{\beta}, \hat{a}, \hat{t}_0|k) - p_j\, L(\hat{\beta}, \hat{a}, \hat{t}_0|j) \qquad (4.1.4)$$

and is seen to be minimum for the same j for which

$$p_j\, L(\hat{\beta}, \hat{a}, \hat{t}_0|j) \qquad (4.1.5)$$

is maximum. Since all errors were weighted equally in the loss function, this is the rule which minimizes the probability of misclassification, or the average error rate. Since the expression (4.1.5) is the joint *a posteriori* likelihood function for the four variables $(\hat{\beta}, \hat{a}, \hat{t}_0, k)$, the Bayes rule for the simple loss function can be considered to be the maximum likelihood decision for the source class k.

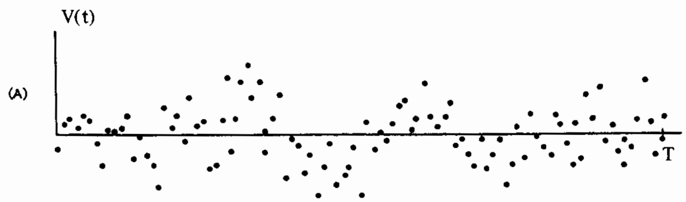
4.2 M–ary Detection

Another general problem in signal classification can be described as follows. Suppose the observer is given the data $V_i = V(t_i)$, $i = 1, 2, \ldots, n$ and is required to choose one of the M + 1 hypotheses

$$H_0: V(t) = N(t)$$
$$H_1: V(t) = N(t) + \beta S_1(at - t_0)$$
$$H_2: V(t) = N(t) + \beta S_2(at - t_0)$$
$$\vdots$$
$$H_M: V(t) = N(t) + \beta S_M(at - t_0)$$

where $S_1(t)$, $S_2(t), \ldots, S_M(t)$ are distinct, known signal forms. This problem is illustrated in Figure 4.1. Again, it will be assumed that the signal parameters are unknown to the observer. The Bayes rule for this classification decision will now be discussed.

For each hypothesis H_k, the maximum conditional likelihood function is given by Equation (2.2.9) to be



HYPOTHESES $H_0: V(t) = N(t)$

$H_1: V(t) = N(t) + \beta S_1(at - t_0)$

$H_2: V(t) = N(t) + \beta S_2(at - t_0)$

$\vdots$

$H_M: V(t) = N(t) + \beta S_M(at - t_0)$

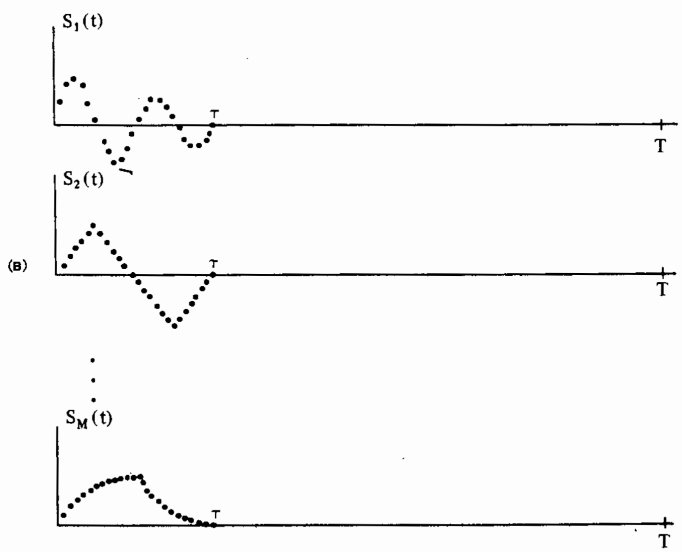


Figure 4-1. Signal Ensemble for M–ary Detection

$$L(\underline{V}|k) = -\frac{n}{2}\ln(2\pi\widehat{\sigma_k^2}) - \frac{n}{2} \tag{4.2.1}$$

where $\widehat{\sigma_k{}^2}$ is given by Equation (2.2.8) to be

$$\widehat{\sigma_k^2} = \frac{1}{n}\left\{\sum_{i=1}^{n} V_i^2 - \frac{\left[\sum_{i=n_o+1}^{n_1} V_i S_k(\widehat{a_k} t_i - \widehat{t_{o\,k}})\right]^2}{\sum_{i=n_o+1}^{n_1} [S_k(\widehat{a_k} t_i - \widehat{t_{o\,k}})]^2}\right\} \tag{4.2.2}$$

for $k \neq 0$, and from Equation (2.2.10) we have

$$\widehat{\sigma_o^2} = \frac{1}{n}\sum_{i=1}^{n} V_i^2 \tag{4.2.3}$$

The estimators $\widehat{\sigma_k{}^2}$, $\widehat{\beta_k}$, $\widehat{a_k}$, $\widehat{t_{o\,k}}$ are now different for each of the M + 1 hypotheses.

Letting $p_0, p_1, \ldots, p_M$ denote the prior probabilities of each of the M + 1 hypotheses, the conditional probability of the hypothesis H_k given the data $\underline{V}$ can be inverted using Bayes Rule to give

$$P(k|\underline{V}) = \frac{p_k\, L(\underline{V}|k)}{\sum_{i=0}^{M} p_i\, L(\underline{V}|i)}$$

The conditional expected loss for making decision j becomes

$$\sum_{k=0}^{M} C(j|k)\, P(k|\underline{V}) = \sum_{k=0}^{M} C(j|k) \frac{p_k\, L(\underline{V}|k)}{\sum_{i=0}^{M} p_i\, L(\underline{V}|i)} \tag{4.2.4}$$

where, as before, C(j|i) denotes the cost of choosing hypothesis H_j when H_i is really true.

The expected loss in Equation (4.2.4) is minimized by choosing the j for which the numerator,

$$\sum_{k=0}^{M} C(j|k)\, p_k\, L(\underline{V}|k) \quad ,$$

is the least. Substituting Equation (4.2.1) for L(V|k) yields the expression

$$\sum_{k=0}^{M} C(j|k)\, p_k \left[-\frac{n}{2} \ln (2\pi\widehat{\sigma_k^2}) - \frac{n}{2} \right] \tag{4.2.5}$$

The Bayes rule for the M–ary Detection problem can now be stated: Choose the hypothesis H_j for which the expression (4.2.5) is minimized where the $\widehat{\sigma_k^2}$ are given by Equation (4.2.2) and Equation (4.2.3). If the loss function is simple and equal prior probabilities of occurrence (p_k) k = 1, 2, . . . , M are assumed, the rule becomes simply: Choose the hypothesis H_j for which the estimate $\widehat{\sigma_j^2}$ of the noise variance is the least.

LIST OF REFERENCES

1. Davenport, W.B. and Root, W.L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
2. Helstrom, C.W., *Statistical Theory of Signal Detection*, Pergamon, New York, 1960.
3. Kendall, M.G., and Stuart, A., *The Advanced Theory of Statistics, Vol. 2*, Hafner, New York, 1961.
4. Lawson, J.L., and Uhlenbeck, G.E., *Threshold Signals*, McGraw-Hill, New York, 1950.
5. Lee, Y.W., *Statistical Theory of Communication*, Wiley, New York, 1960.
6. Middleton, D., *An Introduction to Statistical Communication Theory*, McGraw-Hill, New York, 1960.
7. Middleton, D. and Van Meter, D., "Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory," *J. Soc. Ind. Appl. Math*, 3, 192 (1955); 4, 86 (1956).
8. Owen, D.B., *Handbook of Statistical Tables*, Addison-Wesley, Reading, Massachusetts, 1962.
9. Pearson, E.S., and Hartley, H.O., *Biometrika Tables for Statisticians, Volume I*, Cambridge University Press, Cambridge, England, 1954.
10. Slepian, D., "Estimation of Signal Parameters in the Presence of Noise," *IRE Trans. on Information Theory*, PGIT-3, 68 (March 1954).
11. Van Trees, H.L., *Detection, Estimation and Modulation Theory, Part I*, Wiley, New York, 1968.

12. Wilks, S.S., "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *Ann. Math. Statist.*, 9, 60 (1938).