THE REPRESENTATION OF BIVARIATE DENSITIES WITH APPLICATIONS IN DETECTION AND ESTIMATION

George V. Moustakides

A DISSERTATION
PRESENTED TO THE
FACULTY OF PRINCETON UNIVERSITY
IN THE CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE BY THE
DEPARTMENT OF
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

October 1983

Congratulations on an outstanding dissertation!

John B Thomas

THE REPRESENTATION OF BIVARIATE DENSITIES WITH APPLICATIONS IN DETECTION AND ESTIMATION

ABSTRACT

This dissertation is divided into two parts. The first part is a study of the representation of bivariate densities. The second part is the application of such representations to signal detection and estimation under the assumption of dependent observations.

In the first part we present three models for bivariate densities. The first is the classical model of expanding the bivariate density into a series using the orthonormal polynomials defined by the marginal densities. The second model is the Fréchet class of bivariate densities, which is simple and has interesting properties. The third model is the expansion of a bivariate density into a Fourier series. This model turns out to be simple as well as possessing interesting closed form expressions. The dissertation concentrates principally on the last two models because they are less classical than the first. Their richness is examined and useful properties are obtained.

In the second part of the dissertation, the problem of the detection of signals in additive noise and the problem of the estimation of a location parameter are considered. The main thrust is to solve these two problems assuming some form of dependency between the observations. For the detection problem the efficacy is used as the performance measure and for the estimation problem the asymptotic variance. The detection and estimation scheme that is considered here is sums of memoryless

transformations of the observations. This scheme is tractable because, in order to calculate its performance measure, knowledge only of the bivariate densities between observations is required. This is exactly the point where the two parts of the dissertation are connected. Using the models of the first part, optimum detection and estimation schemes are derived. The general approach that is used is the min-max approach. This approach allows the knowledge of the dependency between the observations to be limited into quantities that can be estimated in practice, such as marginal densities and correlation coefficients, for example. This is definitely an advantage over an approach that optimizes the performance measure and that requires exact knowledge of the statistics between the observations.

ACKNOWLEDGEMENT

The author would like to express his appreciation to the faculty of the EECS department of Princeton University. In particular, special thanks go to his advisor Professor John B. Thomas and also to Professors Stuart Schwartz and Bede Liu for their careful reading of this dissertation.

The author would like to thank the people of the United States who were the real funders of this work.

This work was funded by the U.S. Army Research Office under Grant DAAG29-82-K-0095 and by the National Science Foundation through Grant ECS-79-18915.

ΣΤΗΝ ΜΗΤΕΡΑ

γιωργος

july 1983

Table of Contents.

THE REPRESENTATION OF BIVARIATE DENSITIES WITH APPLICTIONS IN DETECTION AND ESTIMATION.

CHAPTER I.

INTRODUCTION

In detection theory, one of the most commonly encountered problems is the detection of known signals in additive noise. As is well known, designing optimum tests in the Bayes sense, or the Neyman-Pearson sense, results in likelihood ratio tests. For the case of independent observations these tests have a very simple form which is shown in Figure 1.1.1. The observed sequence is passed through zero memory nonlinearities and the sum is compared to a threshold for the decision. Even though the independence assumption makes things easy from an analytical point of view, it is far from being a realistic assumption.

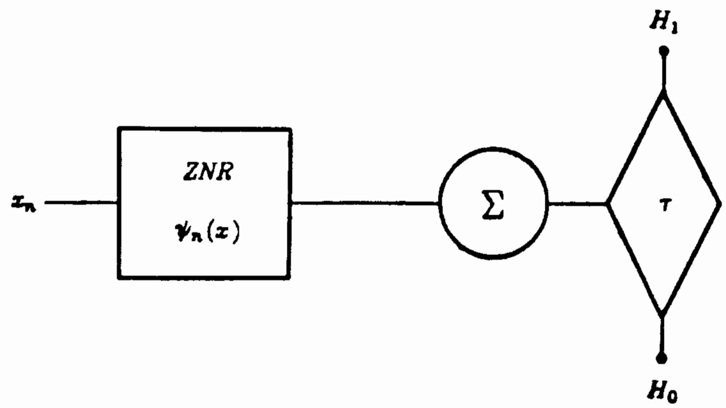

Figure 1.1.1. Optimum detection scheme for independent observations.

If the restriction that the observation sequence is independent is dropped, the optimum tests, in the two senses described before, no longer have the form of Figure 1.1.1. Under dependency, in order to define the optimum schemes we need a knowledge of the multivariate statistics of the observation sequence. Unfortunately, not many models for multivariate statistics are known. The major exception, of course, is the Gaussian case. This is exactly what makes this approach very difficult.

One way of overcoming the problem of representation of the multivariate density is to try to find suboptimal schemes. In other words, we define a class of tests that are of interest to us, the allowable class. We then try to find the optimum test in this class, by optimizing some performance measure. Here, the allowable class will be all the tests that can be represented by Figure 1.1.1. Later, in Chapter III, we will specify more accurately the class.

Optimizing inside the allowable class creates other problems. First, we must select a performance measure that is tractable for our class of tests. It turns out that the Bayes risk or the Neyman-Pearson criterion are not tractable. They require, again, a knowledge of the multivariate statistics. Even with this knowledge, it is not certain that the optimization is an easy task. Clearly there is a need of another measure. As we will see the *efficacy* possesses all of the properties that we will need. The efficacy is an asymptotic measure of performance. The ratio of the efficacies of two tests, under very general conditions, is equal to the Asymptotic Relative Efficiency (ARE) of the two tests. In detection the efficacy has clear meaning for a large number of observations and for the weak signal case. It is a quantity that depends on the test and on the statistics of the observation sequence. Its most important property is that, in order to be calculated, there is needed knowledge only of the second-order statistics of the sequence. In other words we need to know all of the bivariate densities involved. Of course this is still a lot to know, but compared to the multivariate statistics that is required for the other approaches, it is, without any doubt, preferable. In addition, use of bivariate densities is the first natural step into a study of dependency. As we will see in Chapters IV and V, for our approaches, we do not really need to know exactly the bivariate densities. Quantities such as the correlation coefficients will be enough to define optimum tests from the allowable class.

Before discussing the contents of each chapter, we must say a few things about the estimation problem, since this term is included in the title of this work. Here we are interested in estimating a location parameter. As we will explain in Chapter III, with the asymptotic approach that we take here, this problem turns out to be a simple version of the detection problem. Thus, the two problems will be put under the same mathematical framework, which is maximization of the efficacy.

As we said before, bivariate densities are important in our study. Thus Chapter II is devoted to the presentation of three models of bivariate densities. The First Model is the expansion of the density in a series, using the orthonormal polynomials defined by the marginals. This part is a brief description of the main existing results that will be also used in our study. The Second Model is a rather simple but very interesting model. We concentrate on this model because it is encountered in later chapters. Its richness is checked by finding the bounds for the correlation coefficient. Finally the Third Model is the expansion in Fourier series. This model has very nice closed form expressions for the bivariate densities and is developed carefully. Unfortunately, it is not used in any of the later chapters.

In Chapter III the problems of detection and estimation are defined and formulated under the same mathematical framework. It is shown that, from a mathematical point of view, the estimation problem is a special case of the detection problem. The important parts of this chapter are the two appendices. They contain proofs of the Central Limit Theorem for quantities of interest to us.

In Chapter IV we solve the following problem: Given the marginal density and the correlation coefficients of a stationary sequence, what is a tractable way to define an optimum detection or estimation scheme. We use a min-max approach and, for the bivariate densities involved, we assume only that they can be expanded into a diagonal series using orthonormal polynomials. Clearly this approach is of practical interest since there are ways of estimating the marginal density and the correlation coefficient.

In Chapter V we go one step beyond Chapter IV. We assume that we do not know exactly the marginal density. Instead we assume that it belongs to an ε-contaminated class. For the bivariate densities, we assume that they belong to a class that is more general than the Second Model. Optimality is again defined in a min-max way. As it will turn out, the least favorable densities will belong to the Second Model. The dependency class we are using here is limited. It does not contain densities that are drastically different from the independence case. The same holds for the optimum schemes; they do not differ greatly from the independence assumption schemes. But, as we will see from the examples that are presented, the performances of the two schemes are very different. This means that even small dependency can change the performance drastically.

Finally in Chapter VI we give our conclusions and we discuss some ideas for further study in this area.

Before leaving the introduction it is appropriate to say a few things about the numbering of the chapters, sections etc. The chapters are numbered 1 through 6. For the sections we have two numbers. The first denotes the chapter in which they occur and the second denotes the section itself. Thus Section 1.3 means the third section of the Chapter I. Everything else except the sections is characterized by three numbers. Thus, subsections, equations, figures and tables have three numbers. The first number denotes the chapter, the second the section in which they occur and finally the third number characterizes their order. Thus, Figure 3.1.1 means the first figure of the first section of Chapter III. In the text, when three numbers occur in parentheses, they denote an equation. All the other types will be stated explicitly and their numbers will not be in brackets. Finally, the appendices are treated as sections. Only, instead of a second number, they have a letter. Thus Appendix 4.A denotes the first appendix of Chapter IV. The same is true for the equations that occur in the appendices. Their numbering is composed of three elements and the second, which denotes the appendix, is a letter. Thus Equation (3.B.2) means the second equation of the second appendix of Chapter III.

CHAPTER II.

MODELS FOR BIVARIATE DENSITIES.

2.1 Preliminaries.

In this chapter we present three models for representing bivariate densities when the marginal densities are given. The first model is the classical model of expansion of a bivariate density into a double series using the orthonormal polynomials defined by the marginal densities. The second model is the Fréchet class of bivariate densities which is very simple compared to the first one. Finally the third model is the Fourier series expansion model. This model turns out to have very nice closed form expressions and to be simple as well.

Before examining each model separately we present a few general properties for bivariate densities that are important in our study. Let $f_x(x)$ and $f_y(y)$ be two univariate densities and $F_x(x)$ and $F_y(y)$ their univariate cumulative distributions. Let also $f(x,y)$ be a bivariate density and $F(x,y)$ its cumulative distribution. We define as $\Lambda_{f_x f_y}$ the class of bivariate cumulative distributions $F(x,y)$ that satisfy

$$F(x,+\infty) = F_x(x) \quad or \quad \int_{-\infty}^{\infty} f(x,y)dy = f_x(x)$$

$$F(+\infty,y) = F_y(y) \quad or \quad \int_{-\infty}^{\infty} f(x,y)dx = f_y(y) \qquad (2.1.1)$$

For simplicity we will also say that $f(x,y)$ belongs to $\Lambda_{f_x f_y}$ if its $F(x,y)$ belongs to $\Lambda_{f_x f_y}$.

Let us now define

$$F^*(x,y) = \min\left[\ F_x(x), F_y(y)\ \right] \tag{2.1.2}$$

$$F_*(x,y) = \max\left[\ F_x(x) + F_y(y) - 1,\ 0\ \right] \tag{2.1.3}$$

It can be shown that both functions defined in (2.1.2) and (2.1.3) are legitimate bivariate cumulative distributions with marginals $F_x(x)$ and $F_y(y)$,i.e. they belong to the class $\Lambda_{f_x\, f_y}$ and moreover

$$F_*(x,y) \le F(x,y) \le F^*(x,y) \tag{2.1.4}$$

The proof is easy and we will not present it here. As we can see these two distributions define boundaries for the class $\Lambda_{f_x\, f_y}$. Expression (2.1.4) is by no means sufficient for a function to be a cumulative distribution from the class $\Lambda_{f_x\, f_y}$. Notice that the above property holds for distributions and not for densities. The densities of these two limiting distributions are degenerate. It can be also shown that they maximize and minimize the correlation coefficient, among all distributions from the class $\Lambda_{f_x\, f_y}$. Actually they have this property over a whole class of functions, the class of L-superadditive functions (for definition see [1]). A model for bivariate densities is considered "good" if it can represent these two limiting cases. We will see that the first and third model have this property but not the second.

2.2 First Model: Expansion in Orthonormal Polynomials.

This is the most well known representation. The general form is

$$f(x,y) = f_x(x) f_y(y) \left\{ 1 + \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \alpha_{n,m} \varphi_n(x) \vartheta_m(y) \right\} \tag{2.2.1}$$

Where $\{ \varphi_n(x) \}_{n=0}^{\infty}$ is a complete orthonormal system of polynomials in the $L_2(f_x)$ Hilbert space and $\{ \vartheta_m(y) \}_{m=0}^{\infty}$ in $L_2(f_y)$. Equality in (2.2.1) is in the sense that

$$\lim_{M,N \to \infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \{ \frac{f(x,y)}{f_x(x)f_y(y)} - 1 - \sum_{n=1}^{N}\sum_{m=1}^{M} \alpha_{n,m}\varphi_n(x)\vartheta_m(y) \}^2 f_x(x)f_y(y)dxdy = 0 \tag{2.2.2}$$

The representation in (2.2.1) is very general; thus it is difficult to find conditions for the coefficients $\alpha_{n,m}$ to yield a valid density (i.e. a nonnegative function). Special cases have been considered. The most important special case is the diagonal expansion, which is of the form

$$f(x,y) = f_x(x)f_y(y)\{ 1 + \sum_{n=1}^{\infty} \alpha_n \varphi_n(x)\vartheta_n(y) \} \tag{2.2.3}$$

This diagonal expansion has several nice properties; for example, the Gaussian bivariate density can be expanded in this form [2]. We can see that if we put $\alpha_n = 1$ for all n, we get the density of $F^*(x,y)$ and for symmetric marginals if we put $\alpha_n = (-1)^n$ for all n, we get the density of $F_*(x,y)$.

Comments. There are several problems with this model. In order to be able to apply it, the marginal densities must have all their moments finite, otherwise the two orthonormal families cannot be defined. For example, densities like the Cauchy are not good candidates. The second serious problem is the lack of tractable sufficient conditions for the coefficients α_n in (2.2.3) in order for $f(x,y)$ to be nonnegative. There are several necessary conditions for this representation though. The most common is the one for the case where the two marginals are the same and have unbounded support in both directions, then

$$a_n = \int_{-1}^{1} z^n h(z) dz \tag{2.2.4}$$

where $h(z)$ is a univariate density supported in (-1,1). This result was first found in [2] for the case where the marginal density is the Gaussian. For this case it is shown that (2.2.4) is also sufficient. For the expansion in (2.2.3) when the marginals are different, the coefficients satisfy a similar expression as in (2.2.4). More information can be found in [3-8]. The representation can be generalized to include off-diagonal terms [9]. Also in [10] it is shown that any bivariate density can be expanded into a diagonal series when the canonical eigenfunctions of $f(x,y)$ are used.

Another problem for this model is the slow convergence of the series. For the case of unbounded support, keeping only a finite number of terms of the series results always in a function that takes on negative values, i.e. it is not a density. This is a very serious weakness from a practical point of view.

This model of bivariate densities will be used in Chapter IV where optimum structures for detection and estimation will be found.

2.3 Second Model: Fréchet Class.

This model was introduced by Fréchet in [11]. It is very simple and has interesting properties. Actually it does not have many of the problems that the First Model has, but, on the other hand, it is not very rich. For example it does not contain the two limiting densities given by (2.1.2) and (2.1.3). Let us first introduce the model

$$f(x,y) = f_x(x) f_y(y) \{ 1 + \Phi(x)\Theta(y) \} \tag{2.3.1}$$

In order for (2.3.1) to be a density with marginals $f_x(x)$ and $f_y(y)$ the following conditions must hold

$$\int_{-\infty}^{\infty} \Phi(x) f_x(x) dx = \int_{-\infty}^{\infty} \Theta(y) f_y(y) dy = 0 \tag{2.3.2}$$

For $f(x,y)$ to be nonnegative it is necessary that both $\Phi(x)$ and $\Theta(y)$ be bounded. Thus let us assume that

$$-m_x \leq \Phi(x) \leq k_x m_x$$

$$-m_y \leq \Theta(y) \leq k_y m_y \tag{2.3.3}$$

where m_i , k_i $i = x,y$ are nonnegative constants. More precisely $-m_x$ can be defined as the essential inf of $\Phi(x)$ with respect to the marginal $f_x(x)$ and $k_x m_x$ the essential sup. Similarly for m_y and $k_y m_y$.To make the function $f(x,y)$ defined in (2.3.1) nonnegative, we need

$$1 \geq m_x m_y \max \{ k_x, k_y \} \tag{2.3.4}$$

Conditions (2.3.2) and (2.3.4) are necessary and sufficient to make $f(x,y)$ a valid bivariate density. Clearly the marginals can be any univariate density and not necessarily one that has all its moments. In Figure 2.3.1 we can see three dimensional graphs of the following bivariate density that has Cauchy marginals

$$f(x,y) = \frac{\frac{1}{\pi^2}}{(1+x^2)(1+y^2)} \left\{ 1+4d\left[\frac{1}{2} - \frac{1}{1+x^2}\right]\left[\frac{1}{2} - \frac{1}{1+y^2}\right] \right\} \tag{2.3.5}$$

for different values of the parameter d .

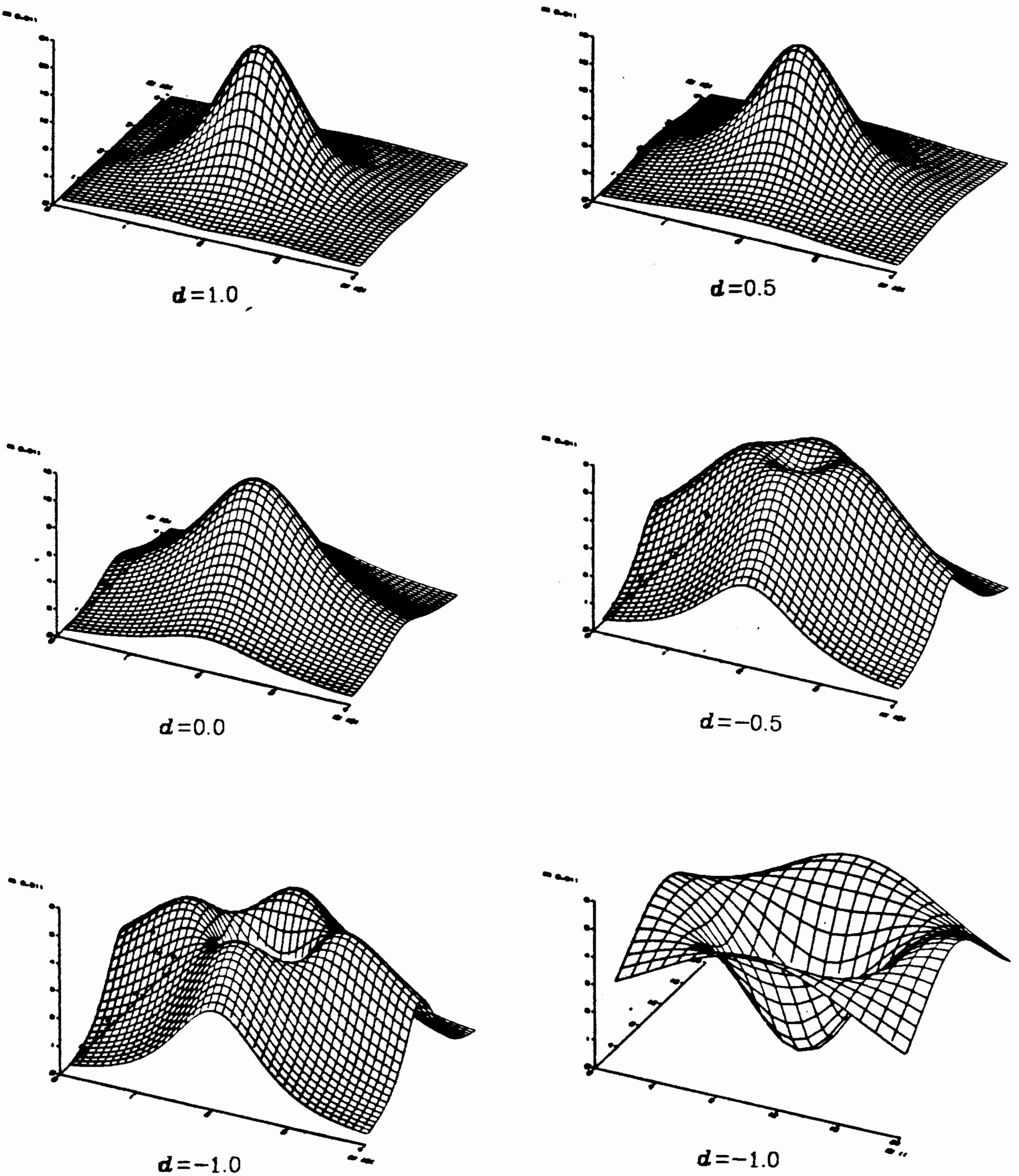

Figure 2.3.1 Bivariate densities with Cauchy marginals using the Second Model.

The Second Model is used in [12] for the study and comparisons of certain estimators. Also in [13] the richness of this model is considered by finding the bounds for the correlation coefficient, for the case where the two marginals are symmetric. In the next section we present a method for finding these bounds for the general case. We do not know if it is the same method as in [13]. We present it though, because it is interesting to see the derivation of these bounds and also to get a better feeling for this model, since we are going to use it in the following chapters.

2.3.1 Bounds for the Correlation Coefficient of the Second Model.

For simplicity we assume that the two marginal densities have mean zero and variance equal to unity. The correlation coefficient then becomes

$$\rho = \left[\int_{-\infty}^{\infty} x \Phi(x) f_x(x) dx \right] \left[\int_{-\infty}^{\infty} y \Theta(y) f_y(y) dy \right] \tag{2.3.6}$$

As a first step we assume that m_i , k_i $i = x, y$ defined in (2.3.3) are given and we will try to find the forms of $\Phi(x)$ that maximize and minimize the expression

$$\sigma_x(k_x) = \int_{-\infty}^{\infty} x \Phi(x) f_x(x) dx \tag{2.3.7}$$

Proposition 2.3.1 The functions $\Phi_U(x)$ and $\Phi_L(x)$ that maximize and minimize (2.3.7) are given by

$$\Phi_U(x) = \begin{cases} -m_x & \text{for } x < x_U \\ k_x m_x & \text{for } x \geq x_U \\ c_U m_x & \text{for } x = x_U \end{cases} \tag{2.3.8}$$

$$\Phi_L(x) = \begin{cases} k_x m_x & \text{for } x \le x_L \\ -m_x & \text{for } x > x_L \\ c_L m_x & \text{for } x = x_L \end{cases} \tag{2.3.9}$$

where x_U and x_L are two points that satisfy

$$F_x(x_U-) \le \frac{k_x}{1+k_x} \le F_x(x_U)$$

$$F_x(x_L-) \le \frac{1}{1+k_x} \le F_x(x_L) \tag{2.3.10}$$

where we assumed that $F_x(x)$ is right continues. Also c_U and c_L are defined to make $\Phi_U(x)$ and $\Phi_L(x)$ zero mean as follows:

$$c_U = \begin{cases} \dfrac{F_x(x_U-) - k_x\big(1 - F_x(x_U)\big)}{F(x_U) - F(x_U-)} & \text{when } F_x(x_U) - F_x(x_U-) > 0 \\ \text{anything} & \text{otherwise} \end{cases}$$

$$c_L = \begin{cases} \dfrac{1 - F_x(x_L) - k_x F_x(x_L-)}{F(x_L) - F(x_L-)} & \text{when } F_x(x_L) - F_x(x_L-) > 0 \\ \text{anything} & \text{otherwise} \end{cases}$$

(2.3.11)

Proof. The proof is given in Appendix 2.A. Using Proposition 2.3.1 we have

$$\sigma_x^L(k_x) = \int_{-\infty}^{\infty} x\Phi_L(x)f_x(x)dx \le \int_{-\infty}^{\infty} x\Phi(x)f_x(x)dx$$

$$\le \int_{-\infty}^{\infty} x\Phi_U(x)f_x(x)dx = \sigma_x^U(k_x) \tag{2.3.12}$$

By putting $\Phi(x) = 0$ we see that

$$\sigma_x^L(k_x) \le 0 \le \sigma_x^U(k_x) \tag{2.3.13}$$

From now on, for simplicity, we will assume that the marginals $f_x(x)$ and $f_y(y)$ have no point masses. Using the property that $f_x(x)$ is zero mean we have

$$\sigma_x^U(k_x) = m_x\ A_x(k_x) \quad \text{and} \quad \sigma_x^L(k_x) = -\,m_x\ B_x(k_x) \tag{2.3.14}$$

where we defined

$$A_x(k_x) = (1 + k_x)\int_{x_U}^{\infty} x f_x(x)dx \tag{2.3.15}$$

$$B_x(k_x) = (1 + k_x)\int_{x_L}^{\infty} x f_x(x)dx \tag{2.3.16}$$

Notice that

$$B_x(k_x) = A_x(\frac{1}{k_x})k_x \tag{2.3.17}$$

Proposition 2.3.2 The two functions $A_x(k_x)$ and $B_x(k_x)$ are increasing in k_x .

Proof . The proof is straightforward. It is based on the fact that if $k_1 \ge k_2$ then the $\Phi_U(x)$ and $\Phi_L(x)$ that correspond to m_x and k_2 are just some other $\Phi(x)$ functions that satisfy the bounds defined by m_x and k_1 .

Now we can finally find the bounds for the correlation coefficient. Define $A_y(k_y)$ and $B_y(k_y)$ for the density $f_y(y)$ in the same way that $A_x(k_x)$ and $B_x(k_x)$ were defined for the density $f_x(x)$.

Proposition 2.3.3 The correlation coefficient ρ for the Second Model can take any value in the interval $[\rho_L, \rho_U]$ where

$$-\rho_L = \max_k A_x(k) A_y(\frac{1}{k}) \tag{2.3.18}$$

$$\rho_U = \max_k \frac{A_x(k) A_y(\frac{1}{k})}{k} \tag{2.3.19}$$

Proof. The proof is in the Appendix 2.A. To obtain a better idea for the bounds let us substitute (2.3.15) and (2.3.16) in (2.3.18) and (2.3.19); we get

$$-\rho_L = \max_k \frac{(1+k)^2}{k} \int_{x_U}^{\infty} x f_x(x) dx \int_{y_L}^{\infty} y f_y(y) dy \tag{2.3.20}$$

$$\rho_U = \max_k \frac{(1+k)^2}{k} \int_{x_U}^{\infty} x f_x(x) dx \int_{y_U}^{\infty} y f_y(y) dy \tag{2.3.21}$$

where we have defined

$$x_U = F_x^{-1}(\frac{k}{1+k}) \qquad y_U = F_y^{-1}(\frac{k}{1+k})$$

$$y_L = F_y^{-1}(\frac{1}{1+k}) \tag{2.3.22}$$

For the case where one of the marginals is symmetric, say $f_y(y)$, we have that $y_L = -y_U$ and thus we get

$$|\rho| \le \max_k \frac{(1+k)^2}{k} [\int_{x_U}^{\infty} x f_x(x) dx][\int_{y_U}^{\infty} y f_y(y) dy] \tag{2.3.23}$$

If in addition $f_x(x) = f_y(x) = f(x)$ then, using (2.3.22) and the symmetry of the marginal, we get

$$|\rho| \le \max_{x \ge 0} \frac{[\int_x^\infty z f(z) dz]^2}{F(x)[1 - F(x)]} \tag{2.3.24}$$

In Table 2.3.1 we can see values of ρ_U and $-\rho_L$ for the case where the marginals are the same and for various types of the common marginal.

Table 2.3.1 Values for $-\rho_L$ and ρ_U

marginal	$-\rho_L$	ρ_U
$0.5\,[\delta(x+1)+\delta(x-1)]$	1.0	1.0
Uniform	0.75	0.75
Gaussian	0.64	0.64
Double Exponential	0.5	0.5
Cauchy	0.0	0.0
Single Exponential	0.48	0.65

2.3.2 Random Processes with Bivariate Densities from the Second Model.

In this subsection we will find sequences of strictly stationary random variables that have bivariate densities from the Second Model. The importance of this problem will be apparent in Chapter V where the existence of such sequences will be important for the validity of our results. We will show a way of defining two types of sequences, a) Markov and b) M-dependent sequences. Thus let $X = \{ X_n \}_{n=1}^{\infty}$ be a strictly stationary random sequence. To define any of the two types, we need to define the multivariate density of the sequence.

Markov Sequences. To define a multivariate density of a strictly stationary Markov sequence we need only the bivariate density of two consecutive members of the sequence. Clearly this bivariate density will be from the Second Model since we put this as a requirement. Thus let

The multivariate density then takes the form

$$f(x_1,x_2,\dots,x_n) = \left[\prod_{k=1}^{n} f(x_k) \right] \prod_{k=1}^{n-1} \left\{ 1 + \Phi(x_k)\Theta(x_{k+1}) \right\} \qquad (2.3.26)$$

By direct integration on (2.3.26) we can see that all the bivariate densities satisfy the Second Model, because

$$f_k(x_j,x_{k+j}) = f(x_j)f(x_{k+j})\left\{ 1 + \alpha^{k-1}\Phi(x_j)\Theta(x_{k+j}) \right\} \qquad (2.3.27)$$

where

$$\alpha = \int_{-\infty}^{\infty} \Phi(x)\Theta(x)f(x)dx \qquad (2.3.28)$$

As we can see from (2.3.27), if $|\alpha| > 1$ then, for high enough k, we can make the bivariate density negative. Thus we conclude that $|\alpha| \le 1$.

M-Dependent Sequences. This is the second most important class of sequences. We will present a method here for defining M-dependent sequences when the bivariate densities are of the form

$$f(x_j,x_{k+j}) = f(x_j)f(x_{k+j})\left\{ 1 + \gamma_k\Phi(x_j)\Phi(x_{k+j}) \right\} \qquad (2.3.29)$$

where $\gamma_k = 0$ for $k > \mathrm{M}$ and $|\Phi(x)| \le 1$.

Let $\{\alpha_i\}_{i=1}^{\infty}$ be any strictly stationary M-dependent sequence and such that all the α_i are supported on [-1,1]. Define

$$f(x_1,x_2,\dots,x_n) = \left[\prod_{i=1}^{n} f(x_i) \right] E_\alpha\left\{ \prod_{i=1}^{n}\left[1 + \alpha_i\Phi(x_i)\right]\right\} \qquad (2.3.30)$$

where by E_α we mean expectation with respect to the sequence $\{ \alpha_i \}$.

Proposition 2.3.4 If

$$E\{\alpha_j \alpha_{k+j}\} = \gamma_k \tag{2.3.31}$$

then (2.3.30) is a multivariate density that has bivariate densities given by (2.3.29).

Proof. The proof is easy. First we can see that the expression in (2.3.30) is nonnegative. Also by interchanging integration and expectation in (2.3.30) we can show that

$$f(x_1,\dots,x_j,x_{j+M+1},\dots,x_n) = f(x_1,\dots,x_j) f(x_{j+M+1},\dots,x_n) \tag{2.3.32}$$

which is the condition for M-dependence. The problem is how to generate a sequence $\{\alpha_j\}$ that satisfies (2.3.31). This can be done by using a simple moving average model, as follows:

$$\alpha_j = \sum_{k=0}^{M} c_k y_{j-k} \tag{2.3.33}$$

where the $\{y_j\}$ is an i.i.d sequence supported on [-1,1] and c_k are $M+1$ real constants with $\sum_{k=0}^{M} |c_k| = 1$. Also the c_k are selected in order that (2.3.31) is satisfied. Unfortunately not every γ_j can be realized with this method. For example when M = 1 we can realize only $|\gamma_1| \le \frac{1}{4}$.

2.4 Third Model: Fourier Series Expansion.

From the previous two models we have seen that the most difficult problem is to ensure nonnegativity of the representation. With the approach we take here, we require from the beginning our expression to be nonnegative. First we will assume that the two marginals $f_x(x)$ and $f_y(y)$ are uniform on $[0,1]$. As we will see, this assumption does not limit us because, with a simple transformation, we can get any marginal we like. The model we propose here is the following:

$$f(x,y) = \left| \sum_{k=1}^{\infty} \alpha_k \Phi_k(x) \Theta_k(y) \right|^2 \qquad (2.4.1)$$

Notice that nonnegativity is obtained automatically. We will not say anything yet about the functions $\Phi_k(x)$ and $\Theta_k(y)$ but we will restrict them considerably as we go on. The only thing we assume at this point is that they are square integrable functions on $[0,1]$ and that they can be complex. Without loss of generality we assume that

$$\int_0^1 |\Phi_k(x)|^2 dx = \int_0^1 |\Theta_k(y)|^2 dy = 1 \qquad (2.4.2)$$

We can write (2.4.1) as follows:

$$f(x,y) = \sum_{k=1}^{\infty} |\alpha_k|^2 |\Phi_k(x)\Theta_k(y)|^2 + \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \alpha_k \overline{\alpha}_m \Phi_k(x)\Theta_k(y)\overline{\Phi}_m(x)\overline{\Theta}_{(y)}$$

$$(2.4.3)$$

where the over-bar means complex conjugate. In order for this expression to be a bivariate density with uniform marginals it must satisfy

$$\int_0^1 f(x,y)dx = \int_0^1 f(x,y)dy = 1 \qquad (2.4.4)$$

Notice now that if $\{\Phi_k(x)\}_{k=1}^{\infty}$ and $\{\Theta_k(y)\}_{k=1}^{\infty}$ are two sets of orthonormal function satisfying

$$|\Phi_k(x)| = |\Theta_k(y)| = 1 \qquad \text{for every } k \tag{2.4.5}$$

and also

$$\sum_{k=1}^{\infty} |\alpha_k|^2 = 1 \tag{2.4.6}$$

Then (2.4.4) is satisfied. Clearly with (2.4.5) we restrict considerably our class of functions. The functions that can satisfy (2.4.5) are of the form

$$\Phi_k(x) = e^{j\varphi_k(x)}$$

$$\Theta_k(y) = e^{j\vartheta_k(y)} \tag{2.4.7}$$

for some real functions $\varphi_k(x)$ and $\vartheta_k(y)$. We must now select these functions in order that the two sets $\{e^{j\varphi_k(x)}\}_{k=1}^{\infty}$ and $\{e^{j\vartheta_k(y)}\}_{k=1}^{\infty}$ be two orthonormal families. Following the idea of Fourier series, we consider the case

$$\varphi_k(x) = 2\pi k\,\varphi(x)$$

$$\vartheta_k(y) = 2\pi k\,\vartheta(y) \tag{2.4.8}$$

for some real measurable functions $\varphi(x)$ and $\vartheta(y)$. What we need now is

$$\int_0^1 e^{j2\pi(k-m)\varphi(x)}dx = \int_0^1 e^{j2\pi(k-m)\vartheta(y)}dy = \delta_{k,m} \tag{2.4.9}$$

Let us define a class L of functions. A function $\varphi(x)$ belongs to this class if it satisfies

$$\int_0^1 e^{j2\pi n\varphi(x)}dx = \delta_{n,0} \quad \text{for } n = 0,\pm1,\pm2,\ldots \tag{2.4.10}$$

Let us try to define a little more specifically the class L. Define

$$r(\tau) = m\ \{\ (\ x : \varphi(x) < \tau\ ) \cap [0,1]\ \} \tag{2.4.11}$$

where by $m\{A\}$ we denote the Lebesque measure of the set A. Let also $s(r)$ denote the support of $d(r(\tau))$. Now we can write the integral in (2.4.10) as

$$\int_{s(r)} e^{j2\pi n\tau} dr(\tau) = \delta_{n,0} \tag{2.4.12}$$

If $r(\tau)$ has a derivative then we can also write

$$\int_{r'(\tau)>0} e^{j2\pi n\tau}\, r'(\tau) d\tau = \delta_{n,0} \tag{2.4.13}$$

If, in addition, $\varphi(x)$ is strictly monotone we can easily find $r(\tau)$. For example, if it is strictly increasing, then

$$r(\tau) = \varphi^{-1}(\tau) \tag{2.4.14}$$

Equations (2.4.12) and (2.4.13) are Fourier integrals. They require that the Fourier transforms of the functions have zeros at the points $\pm 2\pi, \pm 4\pi, \ldots$ and to be equal to unity at zero. This is as far as we go in specifying the class L .

Let as assume now that $\varphi(x)$ and $\vartheta(y)$ belong to the class L . If we substitute (2.4.6),(2.4.7) and (2.4.8) in (2.4.3) we have

$$f(x,y) = 1 + \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \alpha_k \overline{\alpha}_m e^{j2\pi(k-m)\varphi(x)}\, e^{j2\pi(k-m)\vartheta(y)} \tag{2.4.15}$$

By defining $n = k - m$ we get

$$f(x,y) = 1 + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \beta_n e^{j2\pi n(\varphi(x) + \vartheta(y))} \tag{2.4.16}$$

where

$$\beta_n = \sum_{k=1}^{\infty} \alpha_k \alpha_{k+n} \tag{2.4.17}$$

2.4.1 Closed Form Expressions for the Third Model.

Using the results from Section 2.4 we will try to find useful closed form expressions for Equation (2.4.16).

Let us define a set A_z of all points $z \in [-½, ½)$ for which there exist an integer k such that

$$z + k = \varphi(x) + \vartheta(y) \tag{2.4.18}$$

for some pair $(x,y) \in [0,1]\times[0,1]$. Define also the function

$$h(z) = 1 + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \beta_n e^{j2\pi n z} \tag{2.4.19}$$

Clearly if, in (2.4.16) we have $f(x,y) \geq 0$, then $h(z) \geq 0$ for $z \in A_z$ and vice versa. Thus, it is enough to investigate the nonnegativity of $h(z)$ in order to ensure nonnegativity of $f(x,y)$. If A_z is essentially equal to $[-½, ½)$, the function $h(z)$ is a univariate density on the same interval. If A_z is a strict subset of $[-½, ½)$, the function $h(z)$ will be nonnegative on A_z and can be extended into anything on the set $[-½, ½) \sim A_z$ (even negative), under the constraint that the integral of $h(z)$ over the interval $[-½, ½]$ is equal to unity.

Using the above result, we have a nice way of generating bivariate densities with uniform marginals. Let $\varphi(x)$ and $\vartheta(y)$ be two functions from the class L. Let $h(z)$ be any univariate density on $[-½, ½)$ that is square integrable (i.e has a Fourier series expansion and is equal to it in the mean square sense). Then

$$f(x,y) = h([\varphi(x) + \vartheta(y)]mod_{½}) \tag{2.4.20}$$

is a bivariate density with uniform marginals, where by $[z]mod_{½}$ we mean the operation of subtracting integers from z until the result is in the interval $[-½, ½)$. Another way of explaining this modulo operation is that, we periodically extend $h(z)$ in the two directions. If we want a result for some other than the uniform marginals, we apply a transformation using the cumulative distributions, and we obtain

$$f(x,y) = f_x(x)f_y(y)h\left(\left[\varphi(F_x(x)) + \vartheta(F_y(y))\right]mod_{½}\right) \tag{2.4.21}$$

Comments. We can see that this model can represent the two limiting densities. For example we can take $h(z) = \delta(z)$, $\varphi(x) = x$ and $\vartheta(y) = -y$. This selection results in $F^*(x,y)$. The second important property of this class is that all the moments of the marginals need not exist. As we will see in the next subsection, the Cauchy density yields very nice examples of bivariate densities. Clearly, there remains the question of how big is the class we have defined here. Another question is what physical mechanism could generate these densities. For example we know that some Fokker-Planck equations have solutions that can be represented by the First Model [14, pp. 178-179]

2.4.2 Examples.

As we have stated, this class can represent bivariate densities with marginals that do not have all their moments. A good example for such a case is the Cauchy density given by

$$f(x) = \frac{\frac{1}{\pi}}{1 + x^2} \tag{2.4.22}$$

with cumulative distribution

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x) \tag{2.4.23}$$

We will consider the case $\varphi(x) = x$ and $\vartheta(y) = -y$. In other words the orthonormal families in (2.4.7) will be the regular Fourier family. For the univariate density $h(z)$ we select

$$h(z) = \frac{\sin(\frac{\tau\pi}{2})}{1 - \cos(\frac{\tau\pi}{2})\cos(2\pi z)} \tag{2.4.24}$$

This univariate density has a very nice property. For $\tau = 1$ it is uniform, as $\tau \to 0$ it tends to a δ-function at zero and as $\tau \to 2$ it tends to a δ-function at ½. Applying now (2.4.21) we have that

$$f(x,y) = \frac{\frac{1}{\pi^2}}{(1+x^2)(1+y^2)} \frac{\sin(\frac{\tau\pi}{2})}{1 - \cos(\frac{\tau\pi}{2})\cos\left(2\tan^{-1}x - 2\tan^{-1}y\right)}$$

(2.4.25)

Using elementary trigonometry we have that

$$f(x,y) = \frac{\sin(\frac{\tau\pi}{2})}{\pi^2\{(1 + x^2)(1 + y^2) - \cos(\frac{\tau\pi}{2})\left[(1 - x^2)(1 - y^2) + 4xy\right]\}}$$

(2.4.26)

In Figure 2.4.1 we can see three-dimensional graphs of this bivariate density for different values of the parameter τ. It is interesting to notice that when $\tau \to 0$ the density clearly concentrates on the line $y = x$ and when $\tau \to 2$, it concentrates on the hyperbola $xy = -1$.

The above can be applied easily to any marginal density of the form

$$f(x) = K \frac{\prod_{i=1}^{N} (x^2 + b_i^2)}{\prod_{i=1}^{M} (x^2 + a_i^2)} \tag{2.4.27}$$

where K is a normalizing constant. Obviously the bivariate density will be more complicated than in (2.4.26).

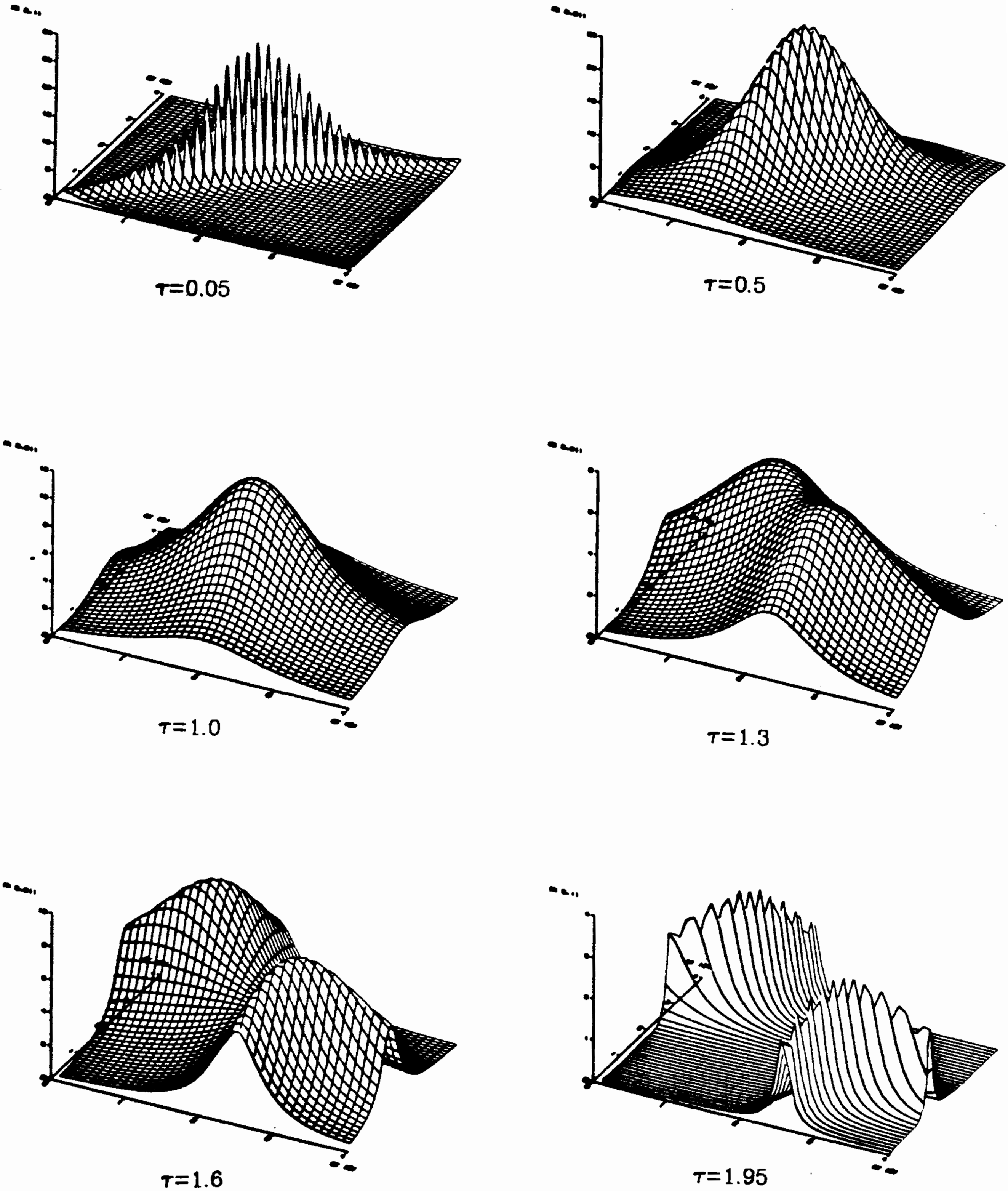

Figure 2.4.1 Bivariate densities with Cauchy marginals using the Third Model.

REFERENCES.

[1] S. Cambanis, G. Simons and W.Stout, " Inequalities for E K(X,Y) when the Marginals are Fixed," *Z. Wahrsch. Verw. Geb.*, 36 (1976),pp. 287-294.

[2] O.V. Sarmanov and Z.N. Bratoeva, " Probabilistic Properties of Bilinear Expansions of Hermite Polynomials," *Theor. Prob. Appl.*, Vol. 12, pp. 470-481, 1967.

[3] S. Tyan and J. B. Thomas , " Characterization of a Class of Bivariate Distribution Functions," *Journ. Multiv. Anal.*, Vol. 5, No. 2, pp. 227-235, June 1975.

[4] S.G. Tyan, H Derin and J.B. Thomas, "Two Necessary Conditions on the Representation of Bivariate Distributions by Polynomials," *Ann. of Stat.*, Vol.4, No. 1, pp. 216-222, 1976.

[5] S.G. Tyan, "The Structure of Bivariate Distribution Functions and their Relation to Markov Processes," *Ph.D Dissertation*, Dep. Electr. Engin., Princeton Univ., Princeton N.J., 1975.

[6] G.L. Wise, "The Bivariate Structure and Representation of Random Processes," *Ph.D Dissertation*, Dep. Electr. Engin., Princeton Univ., Princeton N.J., 1974.

[7] J.A.McFadden, "A Diagonal Expansion in Gegenbauer Polynomials For a Class of Second Order Probability Densities," *SIAM Jour. of App. Math.*, Vol. 14, No. 6, pp. 1433-1436, Nov. 1966.

[8] J. Abrahams and J.B. Thomas, "A Note on the Characterization of Bivariate Densities with Gaussian Marginals Densities," *SIAM Jour. of App. Math.*, Vol. 37, No. 2, pp. 433-435, Oct. 1979.

[9] H. Derin, "On the Properties of Series Expansions of Bivariate Distributions," *Ph.D Dissertation*, Dep. of Electr. Engin., Princeton Univ., Princeton N.J., 1972.

[10] H.O. Lancaster, "The Structure of Bivariate Distributions," *Ann. Math. Stat.*, Vol. 29, pp. 729-736, 1958.

[11] M. Fréchet, "Sur les Tableau de Correlation dont les Marges sont Données," *Ann. Univ. Lyon*, III ser., A14 (1951), pp. 53-77.

[12] D.J.G. Farlie, "The Performance of Some Correlation Coefficients for a General Bivariate Distribution," *Biometrica*, Vol. 47, pp. 307-323, 1960.

[13] O.V. Sarmanov, " The Maximal Correlation Coefficient (The Symmetric Case)," *Doklady AN SSSR 120*, No. 4 (1958), pp. 715-718. (In Russian.)

[14] E. Wong, "Stochastic Processes in Information and Dynamical Systems," *Krieger Publ. Comp.*, Huntington New York, pp. 178-179, 1979.

[15] G. Szegö, "Orthonormal Polynomials," *Amer. Math. Soc. Colloq. Pub. 23*, Amer. Math. Soc., Providence, 1967.

APPENDIX 2.A

Proof of Proposition 2.3.1. We present only the proof for $\Phi_U(x)$. The proof for $\Phi_L(x)$ is similar. The reason we have defined $\Phi_U(x)$ this way is to be able to take into account point masses that might exist at x_U. First we must prove that $\Phi_U(x)$ satisfies the bounds defined by (2.3.3). This is trivial for the case $x > x_U$ and $x < x_U$. When $x = x_U$ we must show that

$$-1 \leq c_U \leq k_x \tag{2.A.1}$$

This is easy because $-1 \leq c_U$ is equivalent, after some operations, to $F_U(x_U) \geq \dfrac{k_x}{1 + k_x}$ and $c_U \leq k_x$ to $F_x(x_U-) \leq \dfrac{k_x}{1 + k_x}$ which, because of (2.3.10) are valid inequalities. Now we define the difference

$$D = \int_{-\infty}^{\infty} x[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx =$$

$$\int_{-\infty}^{x_U} x[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx + \int_{x_U}^{\infty} x[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx +$$

$$\int_{\{x_U\}} x[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx \tag{2.A.2}$$

In the first integral in (2.A.2) the difference $\Phi_U(x) - \Phi(x)$ is nonpositive and in the second nonnegative. Thus we can write for D

$$D \geq \int_{-\infty}^{x_U} x_U[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx + \int_{x_U}^{\infty} x_U[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx +$$

$$\int_{\{x_U\}} x_U[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx = x_U\int_{-\infty}^{\infty}[\ \Phi_U(x) - \Phi(x)\ ]f_x(x)dx = 0 \tag{2.A.3}$$

And this concludes the proof.

Proof of Proposition 2.3.3. We have to show that the correlation coefficient can take any value between the values ρ_L and ρ_U given by (2.3.18) and (2.3.19).

From Equation 2.3.12 we can have bounds for the term $\sigma_x(k_x)$, defined in Equation (2.3.7). Thus, combining Equations (2.3.6), (2.3.12) and (2.3.14) yields

$$\begin{gathered} -m_x m_y \max\{ A_x(k_x)B_y(k_y) , B_x(k_x)A_y(k_y) \} \le \rho \le \\ m_x m_y \max\{ A_x(k_x)A_y(k_y) , B_x(k_x)B_y(k_y) \} \end{gathered} \tag{2.A.4}$$

Because of symmetry in (2.A.4) without loss of generality we can assume that $k_x \ge k_y$. Then, Condition (2.3.4) for nonnegativity becomes

$$\frac{1}{k_x} \ge m_x m_y \tag{2.A.5}$$

We can see that, by increasing k_y and keeping it less or equal than k_x, we do not change (2.A.5). On the other hand the bounds in (2.A.4) increase in the right directions because of the monotonicity of $A(k)$ and $B(k)$. Thus by selecting $k_y = k_x$ and also $m_x m_y = \frac{1}{k_x}$ we can have the largest bounds for a given k_x as follows:

$$\begin{gathered} -\frac{1}{k_x}\max\{ A_x(k_x)B_y(k_x) , B_x(k_x)A_y(k_x) \} \le \rho \le \\ \frac{1}{k_x}\max\{ A_x(k_x)A_y(k_x) , B_x(k_x)B_y(k_x) \} \end{gathered} \tag{2.A.6}$$

Now, using the definitions of ρ_L and ρ_U from (2.3.18) and (2.3.19), we can see that

$$\frac{A_x(k_x)A_y(k_x)}{k_x} \le \rho_U \tag{2.A.7}$$

Using (2.3.17) we have that

$$\frac{B_x(k_x)B_y(k_x)}{k_x} = \frac{A_x(\frac{1}{k_x})A_y(\frac{1}{k_x})}{\frac{1}{k_x}} \le \rho_U \tag{2.A.8}$$

Thus we proved $\rho \le \rho_U$. In a similar way we can prove $\rho_L \le \rho$. To show now that we can actually achieve any value in the interval $[\rho_L, \rho_U]$, let $\Phi_*(x)$ and $\Theta_*(y)$ be a pair that achieves one of the two bounds, let us say ρ_U. Then the following expression is a bivariate density in the class $\Lambda_{f_x f_y}$.

$$f(x,y) = f_x(x)f_y(y)\{\, 1 + s\Phi_*(x)\Theta_*(y)\, \} \tag{2.A.9}$$

with s a parameter in $[0,1]$. By changing s we can achieve any value of $\rho \in [0,\rho_U]$. And this concludes the proof.

CHAPTER III.

OPTIMUM DETECTION AND ESTIMATION

3.1 Preliminaries.

This chapter deals with two problems: a) The problem of detection of weak signals and b) The problem of estimating a location parameter. In this section we define explicitly these two problems and state our assumptions. We also put these two problems into the same mathematical formulation. Thus, by solving one problem, we have immediately the solution to the other. In Section 3.2 we present the solution to the common mathematical problem using as our dependency model the Second Model of Section 2.3. Finally, because we consider asymptotic measures and quantities, in Appendices 3.A and 3.B we prove the Central Limit Theorem for cases of interest to us.

At this point let us present our assumptions. Since we will be dealing with series of random variables, for convenience we denote them with capital bold-face letters.

Assumptions. Let $\mathbf{N} = \{N_i\}_{i=1}^{\infty}$ be a strictly stationary sequence of random variables. Define as M_a^b the σ-algebra generated by the random variables $\{ N_a , \ldots , N_b \}$. Let $f(x)$ be the common marginal density. We assume that it has a derivative a.e. and finite Fisher's information.

We call the sequence $\mathbf{N}$ a *φ-mixing* sequence, if there exists a sequence of real numbers $\{\varphi_k\}$ such that, if A an event from M_1^k and B an event from M_{k+j}^{∞}, then

$$1 \geq \varphi_1 \geq \varphi_2 \geq \cdots \geq 0 \qquad (3.1.1)$$

$$| P(A \cap B) - P(A)P(B) | \leq \varphi_j P(A) \qquad (3.1.2)$$

where P is the probability measure that characterizes $\boldsymbol{N}$. We also call a sequence $\boldsymbol{N}$ a *symmetrically φ-mixing* sequence if (3.1.1) is satisfied and instead of (3.1.2) we have the following

$$| P(A \cap B) - P(A)P(B) | \leq \varphi_j \min \{ P(A) , P(B) \} \qquad (3.1.3)$$

Either of the two sequences will be called *acceptable*, if

$$\sum_{k=1}^{\infty} \varphi_k^{1/2} < \infty \qquad (3.1.4)$$

For simplicity we will call the acceptable (symmetrically) φ-mixing sequences, (symmetrically) $A\varphi$-mixing . Clearly a symmetrically $A\varphi$-mixing sequence is always $A\varphi$-mixing . Conditions (3.1.3) and (3.1.4) say that the dependency between future and past tends to zero as their "distance" becames larger, in a uniform way. The classes of sequences we have defined here are similar to the ones defined in [1]. They will help us to prove asymptotic normality under dependency. We can see that all the M-dependent sequences belong to these classes. We now state two lemmas that will give us some properties of our φ-mixing classes.

Lemma 3.1.1. Let N be a φ-mixing sequence and Ξ and Θ two random variables defined on M_1^k and M_{k+j}^{∞} respectively. If $E\{ |\Xi|^r \} < \infty$ and $E\{ |\Theta|^q \} < \infty$ with $\frac{1}{r} + \frac{1}{q} = 1$, then

$$| E\{\Xi\Theta\} - E\{\Xi\}E\{\Theta\} | \leq 2\varphi_j^{1/r} E^{1/r}\{|\Xi|^r\} \, E^{1/q}\{|\Theta|^q\} \qquad (3.1.5)$$

where by $E\{\ \}$ we mean expectation.

***Proof*.** For a proof see [1, p. 170].

***Lemma* 3.1.2.** Let $\boldsymbol{N}$ be a symmetrically φ-mixing sequence and let $\psi(x)$ be a bounded function with $E\{\psi(N_1)\} = 0$. Then

$$E\{\ \psi(N_1)\ /\ N_{j+1}\ \} \leq 2\varphi_j C \tag{3.1.6}$$

$$E\{\ \psi(N_{j+1})\ /\ N_1\ \} \leq 2\varphi_j C \tag{3.1.7}$$

where C is a bound for $\psi(x)$.

***Proof*.** We will prove the first inequality. In a similar way we can prove the second since everything is symmetric. Let $A \in M_{j+1}^{j+1}$ and I_A be the random variable that takes the value one when $N_{j+1} \in A$ and zero otherwise. Define

$$\lambda(N_{j+1}) = E\{\ \psi(N_1)\ /\ N_{j+1}\} \tag{3.1.8}$$

then, as we know,

$$E\{\psi(N_1)I_A\} = E\{\lambda(N_{j+1})I_A\} \tag{3.1.9}$$

From Lemma 3.1.1, by setting $\Xi = \psi(N_1)$ and $\Theta = I_A$, also $r = \infty$ and $q = 1$, we have

$$|\ E\{\psi(N_1)I_A\}\ | \leq 2\varphi_j C\ E\{I_A\} \tag{3.1.10}$$

Thus for $E\{I_A\} > 0$, we have

$$\frac{|\ E\{\lambda(N_{j+1})I_A\}\ |}{E\{I_A\}} \leq 2\varphi_j C \tag{3.1.11}$$

Since the above is true for any measurable set A, we have that

$$|\lambda(N_{j+1})| \le 2\varphi_j C \tag{3.1.12}$$

with probability one.

Now we will state and formulate the two problems. We begin with the detection problem.

3.1.1. Optimum Detection of Weak Signals.

The problem we would like to solve here is the following : We are given a set of observations $\{X_i\}_{i=1}^{n}$ and we would like to decide between the two hypotheses

$$\begin{aligned} H_0&: \quad X_i = N_i \\ H_1&: \quad X_i = N_i + \delta s_i \quad i = 1,\ldots,n \end{aligned} \tag{3.1.13}$$

where $\boldsymbol{N} = \{N_i\}_{i=1}^{\infty}$ is a symmetrically $A\varphi$-mixing sequence, $\{s_i\}$ is a signal sequence and δ a parameter that tends to zero.

We would like the decision scheme for solving this problem to have the form

$$T_n(X) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} s_k \psi(X_k) \tag{3.1.14}$$

$$u(T_n(X)) = \begin{cases} 1 & \text{if} \quad T_n(X) > \tau \\ p & \text{if} \quad T_n(X) = \tau \\ 0 & \text{if} \quad T_n(X) < \tau \end{cases} \tag{3.1.15}$$

where $\psi(x)$ is some zero memory nonlinearity and $u(T_n(\boldsymbol{X}))$ is the probability of deciding H_1. The constants τ and p are selected to keep the false alarm probability at a certain level. We chose this form of decision scheme because of its simplicity and also because it is Neyman-Pearson optimal for i.i.d. sequences. This detector is known as the

Nonlinear Correlator (NC).

The performance of the above scheme will be defined in terms of the *Efficacy*. The efficacy is an asymptotic measure of performance. Under reasonable conditions, the ratio of efficacies is equal to the ARE (Asymptotic Relative Efficiency). We will give sufficient conditions for this statement to be valid. Let us now define the efficacy as

$$eff(\psi,N) = \lim_{n\to\infty} \frac{\left[\frac{\partial}{\partial\delta} E_\delta\{T_n(X)\}\,|_{\delta=0}\right]^2}{nE_0\{\,[T_n(X)]^2\,\}} \tag{3.1.16}$$

where E_δ and E_0 denote expectation under H_1 and H_0 respectively. In order now to be able to simplify this expression we will restrict the class of sequences $\{s_i\}$ and the nonlinearity $\psi(x)$. We assume that the signal sequence is bounded and that the following limits exist:

$$\nu_j = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n-j} s_k s_{k+j} \qquad j = 0,1,2,\ldots \tag{3.1.17}$$

For simplicity we assume that $\nu_0 = 1$; then it is easy to see that

$$|\nu_j| \le 1 \tag{3.1.18}$$

For the nonlinearity $\psi(x)$ the restrictions we impose will define the class Ψ of allowable nonlinearities. Let us define $\delta_n = \frac{k}{\sqrt{n}}$ where k is any nonnegative constant. For $\psi(x)$ we assume that it is measurable with $E\{\psi(N_1)\} = 0$ and $E\{[\psi(N_1)]^2\} < \infty$ and also that

$$\frac{\partial}{\partial\delta}\int_{-\infty}^{\infty}\psi(x)f(x-\delta)dx\,\,|_{\delta=0} = \int_{-\infty}^{\infty}\frac{\partial}{\partial\delta}\psi(x)f(x-\delta)\,|_{\delta=0}dx \tag{3.1.19}$$

$$\int_{-\infty}^{\infty}\psi(x)f'(x)dx < 0 \tag{3.1.20}$$

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} \psi(x) f'(x - \delta_n) dx = \int_{-\infty}^{\infty} \psi(x) f'(x) dx \qquad (3.1.21)$$

$$\lim_{t \to 0} E \left\{ [\psi(N_1 - t) - \psi(N_1)]^2 \right\} = 0 \qquad (3.1.22)$$

$$\sigma_0^2(\psi) = E\left\{\psi(N_1)^2\right\} + 2 \sum_{j=1}^{\infty} \nu_j E\left\{\psi(N_1)\psi(N_{j+1})\right\} > 0 \qquad (3.1.23)$$

The above conditions, even though they look complicated, are usually valid for well behaved functions. They are sufficient for the validity of the Pitman-Noether theorem, i.e. for proving that the ratio of efficacies is equal to the ARE. Under the above assumptions the efficacy takes the form

$$eff(\psi, N) = \frac{\left[\int_{-\infty}^{\infty} \psi(x) f'(x) dx\right]^2}{\sigma_0^2(\psi)} \qquad (3.1.24)$$

The efficacy and $\sigma_0^2(\psi)$ play an important role in our approach. Since the efficacy is our measure of performance, maximizing it will define the nonlinearity $\psi(x)$ in an optimum way. From (3.1.23) notice that in order to calculate (3.1.24) we need knowledge of all the bivariate densities. Thus we can see now where the models presented in Chapter II can be of some use. The quantity $\sigma_0^2(\psi)$ is the variance of the random variable $T = \lim_{n \to \infty} T_n(X)$. As we show in Appendix 3.A this limit exists under H_0 and is Gaussian. Using $\sigma_0^2(\psi)$ we can set the threshold τ. If n is large then we can assume that $T_n(\boldsymbol{X})$ is "close" to T and thus

$$P\left(T_n(\boldsymbol{X}) > \gamma / H_0\right) \approx P\left(T > \tau / H_0\right) = 1 - \Phi\left(\frac{\tau}{\sigma_0(\psi)}\right) \qquad (3.1.25)$$

This approach does not specify p since the event $T = \tau$ has zero probability.

The problem of maximizing the efficacy is considered in [2] for M-dependent sequences and in [3] for symmetrically $A\varphi$–mixing sequences and the constant signal case. Now we present a theorem which gives an integral equation that the optimum nonlinearity has to satisfy.

Theorem 3.1.1. The optimum nonlinearity $\psi_o(x)$ that maximizes (3.1.24) satisfies

$$f(x)\psi_o(x) = -f'(x) - \sum_{j=1}^{\infty} \nu_j \int_{-\infty}^{\infty} [f_j(x,y) + f_j(y,x)]\psi_o(y)dy \qquad (3.1.26)$$

where $f_j(x,y)$ is the bivariate density of N_1 and N_{j+1}.

Proof. The proof follows exactly the same steps as in [3, Theorem 1].

3.1.2 Estimation of a Location Parameter.

Here we would like to solve the following problem: We are given a set of observations $\{X_i\}_{i=1}^{n}$. We assume that

$$X_i = N_i + \delta \qquad (3.1.27)$$

Where $\boldsymbol{N}$ is a zero-mean symmetrically $A\varphi$–mixing sequence and δ is an unknown shift that we wish to estimate. If T_n is the estimate of δ we would like it to be the solution of an equation that has the form

$$\sum_{i=1}^{n} \psi(X_i - T_n) = 0 \qquad (3.1.28)$$

where $\psi(x)$ is some zero memory nonlinearity. These estimators are called M-estimators. For reasons of consistency we must assume that $\psi(x)$ satisfies $E\{\psi(N_1)\} = 0$, otherwise T_n does not tend to δ as n goes to infinity. Thus, the class Ψ of allowable nonlinearities is defined for functions $\psi(x)$ that are continuous and satisfy $E\{\psi(N_1)\} = 0$ and $E\{[\psi(N_1)]^2\} < \infty$ and also

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} [\psi(x - \frac{k}{\sqrt{n}}) - \psi(x)]^2 dx = 0 \tag{3.1.29}$$

$$\frac{\partial}{\partial \delta} \int_{-\infty}^{\infty} \psi(x) f(x - \delta) \Big|_{\delta=0} = \int_{-\infty}^{\infty} \frac{\partial}{\partial \delta} \psi(x) f(x - \delta) \Big|_{\delta=0} dx \tag{3.1.30}$$

$$\int_{-\infty}^{\infty} \psi(x) f'(x) dx < 0 \tag{3.1.31}$$

$$\sigma_0^2(\psi) = E\{\psi^2(N_1)\} + \sum_{j=1}^{\infty} E\{\psi(N_1)\psi(N_{j+1})\} > 0 \tag{3.1.32}$$

where k is any real number. If, in addition we assume that $\psi(x)$ is monotone, then we can show that the normalized error $n^{1/2}(T_n - \delta)$ is asymptotically Gaussian with mean zero and variance $A(\psi,N)$, given by

$$A(\psi,N) = \frac{\sigma_0(\psi)}{\left| \int_{-\infty}^{\infty} \psi(x) f'(x) dx \right|} \tag{3.1.33}$$

where $\sigma_0(\psi)$ is defined in (3.1.32). In Appendix 3.B we prove this statement. Notice that the assumption that $\psi(x)$ is monotone is restrictive. However, it simplifies the proof of the Central Limit Theorem. Probably, as in the i.i.d. case [4], there might be other conditions that will lead to a larger class. As we can see, (3.1.33) is the inverse of the square root of the efficacy defined in (3.1.24) for the case $\nu_j = 1$ for every j. A way of measuring the performance of the estimator is the asymptotic variance of the error. The smaller the variance, the better the estimation. Clearly a variance equal to zero is the perfect situation. Our goal then is to minimize (3.1.33) or equivalently, to maximize (3.1.24). Thus the two problems of detection and estimation are now under the same mathematical formulation, which is maximization of the efficacy.

Comment. The two Ψ classes defined for the two problems are not the same. The reason we have so many conditions in the detection problem is because we want to ensure that ratios of efficacies yield AREs. Not all of the conditions are needed for proving normality. In the estimation problem the conditions are needed only to prove normality.

Before going to the next section, where we present solutions of Equation (3.1.26), we prove the following theorem

***Theorem* 3.1.2.** Let N be a symmetrically $A\varphi$–mixing sequence. Then, the solution to (3.1.26) cannot be a bounded function unless the locally optimum nonlinearity $l(x) = -\frac{f'(x)}{f(x)}$ is bounded.

Proof. The proof is straightforward. Let us assume that $\psi_o(x)$, the solution to (3.1.26), is bounded, but $l(x)$ is not. Then, using (3.1.26)

$$|f'(x)| \le f(x)C + \sum_{j=1}^{\infty} |\nu_j| \left\{ \left| \int_{-\infty}^{\infty} f_j(x,y)\psi_o(y)dy \right| + \left| \int_{-\infty}^{\infty} f_j(y,x)\psi_o(y)dy \right| \right\} \tag{3.1.34}$$

where C is a bound for $\psi_o(x)$. Using (3.1.6) and (3.1.7) we get

$$|\, l(x) \,| \le C\Big(1 + 4\sum_{j=1}^{\infty} |\nu_j|\varphi_j\Big) \le C\Big(1 + 4\sum_{j=1}^{\infty} \varphi_j^{1/2}\Big) < \infty \tag{3.1.35}$$

which is a contradiction since we assumed that $l(x)$ is unbounded.

The purpose of Theorem 3.1.2 was to show that it is unreasonable to expect that symmetrically $A\varphi$–mixing processes with marginals like the Gaussian will have bounded optimum nonlinearities.

3.2 Optimum Nonlinearities for the Second Model.

Let us assume that the bivariate densities of N_1 and N_{j+1} are given by

$$f_j(x,y) = f(x)f(y)\{1 + \gamma_j \Phi(x)\Theta(y)\} \tag{3.2.1}$$

where we assume that $\sum_{j=1}^{\infty} |\gamma_j| < \infty$. In Subsection 2.3.2 we have seen that there exist stationary sequences that satisfy (3.2.1). If we substitute (3.2.1) into (3.1.26) we get

$$f(x)\psi_o(x) = -f'(x) - \gamma\{A\Phi(x) + B\Theta(x)\}f(x) \tag{3.2.2}$$

where we have defined

$$A = \int_{-\infty}^{\infty} \Theta(x)\psi_o(x)f(x)dx \tag{3.2.3}$$

$$B = \int_{-\infty}^{\infty} \Phi(x)\psi_o(x)f(x)dx \tag{3.2.4}$$

$$\gamma = \sum_{j=1}^{\infty} \nu_j \gamma_j \tag{3.2.5}$$

Thus we can see that the solution to (3.1.26) has the form

$$\psi_o(x) = l(x) - \gamma A\Phi(x) - \gamma B\Theta(x) \tag{3.2.6}$$

To find A and B we substitute (3.2.6) in (3.2.3) and (3.2.4) and we get two equations with unknowns A and B . If we solve them we have

$$A = \frac{\sigma_{\Theta l}\left[1 + \gamma\sigma_{\Theta\Phi}\right] - \gamma\sigma_{\Phi l}\sigma_{\Theta\Theta}}{\left[1 + \gamma\sigma_{\Phi\Theta}\right]^2 - \gamma^2\sigma_{\Theta\Theta}\sigma_{\Phi\Phi}} \tag{3.2.7}$$

$$B = \frac{\sigma_{\Phi l}\left[1 + \gamma\sigma_{\Theta\Phi}\right] - \gamma\sigma_{\Theta l}\sigma_{\Phi\Phi}}{\left[1 + \gamma\sigma_{\Phi\Theta}\right]^2 - \gamma^2\sigma_{\Theta\Theta}\sigma_{\Phi\Phi}} \tag{3.2.8}$$

where, for any two functions $C(x)$ and $D(x)$, we have defined

$$\sigma_{CD} = \int_{-\infty}^{\infty} C(x)D(x)f(x)dx \tag{3.2.9}$$

For the case where $\Phi(x) = \Theta(x)$ we have that the optimum nonlinearity becomes

$$\psi_o(x) = l(x) - 2\gamma A\Phi(x) \tag{3.2.10}$$

and A is given by

$$A = \frac{\sigma_{\Phi l}}{1 + 2\gamma\sigma_{\Phi\Phi}} \tag{3.2.11}$$

3.2.1 Application to Markov Processes.

Let us assume that N is strictly stationary Markov and that $\nu_j = 1$ for every j. If the bivariate density of two consecutive members of the sequence is given by

$$f(x,y) = f(x)f(y)\left\{1 + \Phi(x)\Theta(y)\right\} \tag{3.2.12}$$

then from (2.3.26) we have that the multivariate density is given by

$$f(x_1, \ldots, x_n) = \Big[\prod_{k=1}^{n} f(x_k) \Big] \prod_{k=1}^{n-1} \Big\{ 1 + \Phi(x_k)\Theta(x_{k+1}) \Big\} \qquad (3.2.13)$$

To show that this defines a symmetrically $A\varphi$−mixing sequence, let $A \in M_1^k$ and $B \in M_{k+n}^q$ with q an integer such that $q \geq k+n$. Let A_i be the range of values of the random variable N_i given the event A. This is a Borel subset of the real line. Define $A' = A_1 \times A_2 \times \cdots \times A_k$. Similarly define B_j to be the range of N_j under B and define $B' = B_{k+n} \times \cdots \times B_q$. Then because the sequence is Markov, we have

$$P(A \cap B) = \int_{A' \times B'} f(x_1,\ldots,x_{k-1}) \frac{f_n(x_k, x_{k+n})}{f(x_k) f(x_{k+n})} f(x_{k+n},\ldots,x_q) d\mathbf{x} \qquad (3.2.14)$$

Using (2.3.27) we have

$$|P(A \cap B) - P(A)P(B)| \leq |\alpha|^{n-1} C \int_{A' \times B'} f(x_1,\ldots,x_k)\, f(x_{k+n},\ldots,x_q) d\mathbf{x} =$$

$$|\alpha|^{n-1} C P(A) P(B) \qquad (3.2.15)$$

where C is a bound for $|\Phi(x)\Theta(y)|$ and α is equal to $E\{\Phi(x)\Theta(x)\}$ (see (2.3.28)). Since (3.2.15) is true for any q we have that our sequence is $A\varphi$−mixing when $|\alpha| < 1$. It is also symmetrically $A\varphi$−mixing because

$$P(A)P(B) \leq \min \{ P(A)\, , P(B) \} \qquad (3.2.16)$$

As an example let us assume that

$$f(x,y) = f(x) f(y) \{ 1 + d\, [1 - 2F(x)]\, [1 - 2F(y)] \} \qquad (3.2.17)$$

where $F(x)$ is the cumulative distribution of $f(x)$ and d is a parameter from the interval [-1,1]. If we calculate the necessary quantities we get that

$$\psi_o(x) = -\frac{f'(x)}{f(x)} + \frac{12d}{3+d}\left[\int_{-\infty}^{\infty} f^2(x)\right]\{1 - 2F(x)\} \qquad (3.2.18)$$

Notice that for $d > 0$, which corresponds to a positive correlation, then for symmetric marginals, $\psi_o(x)$ is always more conservative than $l(x)$. In other words for $x > 0$ we have $\psi_o(x) \leq l(x)$. If we apply (3.2.18) for the Gaussian and the Cauchy densities we have

$$\psi_G(x) = x + \frac{12}{\sqrt{4\pi}}\frac{d}{3+d}\{1 - 2\Phi(x)\}$$

$$\psi_C(x) = \frac{2x}{1+x^2} - \frac{12d}{\pi^2(3+d)}\tan^{-1}(x) \qquad (3.2.19)$$

In Figure 3.2.1 (a,b) we can see the forms of $\psi_G(x)$ and $\psi_C(x)$ for different values of the parameter d. As we can see the Gaussian case is affected only around the origin and not at the tails. With the Cauchy things are quite different. The optimum nonlinearity is not a blanker any-more. Actually, it weighs the large values negatively (for $d > 0$).

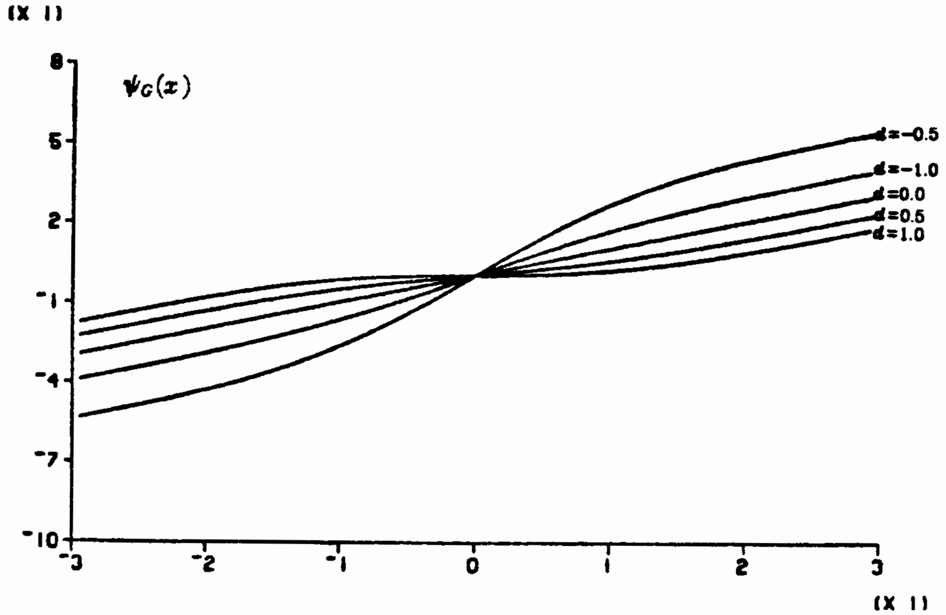

Figure 3.2.1 (a) Optimum nonlinearities for the Gaussian marginal case.

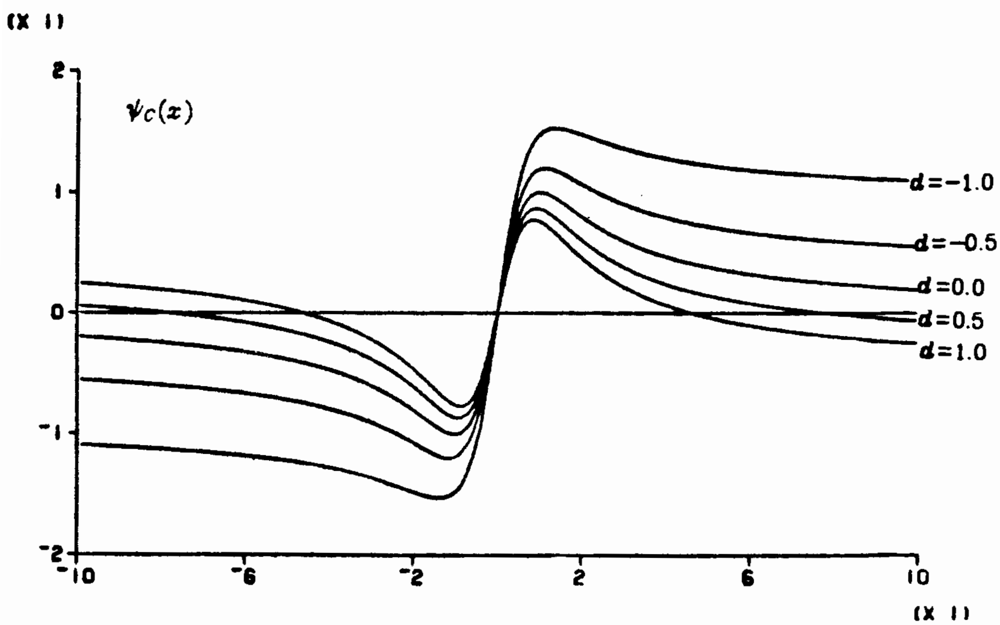

Figure 3.2.1 (b) Optimum nonlinearities for the Cauchy marginal case.

REFERENCES.

[1] J. Billingsley, Convergence of probability measures. New York, Wiley 1968.

[2] H.V. Poor and J.B. Thomas, "Memoryless discrete time detection of a constant signal in m-dependent noise," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 54-61, Jan. 1979.

[3] D.R. Halverson and G.L. Wise, "Discrete time detection in φ-mixing noise," IEEE Trans. Inf. Theory, vol. IT-26, pp. 189-198, March 1980.

[4] P.J. Huber, "The behavior of the maximum likelihood estimates under nonstandard conditions," *Proc. 5-th Berkeley Symp. Math. Stat. and Prob.*, vol. 1, Univ. California Press, Berkeley.

[5] ---, "Robust estimation of a location parameter," *Ann. Math. Stat.*, vol. 35, pp. 73-101, 1964.

[6] ---, Robust Statistics. New York, Wiley 1980.

APPENDIX 3.A

Asymptotic Normality for the Detection Problem.

In this appendix we prove that, under the assumptions we made in Subsection 3.1.1 about the sequence N, the signal $\{s_i\}$ and the non-linearity $\psi(x)$, the quantity $T_n(X)$ defined in (3.1.14) is asymptotically a Gaussian random variable. Notice that this is not an immediate consequence of [1, Theorem 20.1, p. 174], because the sequence $\{s_i\psi(N_i)\}$ is not stationary. For the proof we are going to use a more general theorem [1, Theorem 19.2, p. 157] and we will prove that our case satisfies all the assumptions of this theorem. Before doing so let us show that under the $A\varphi$−mixing assumption the quantity $\sigma_0(\psi)$ of (3.1.23) is well defined.

Lemma 3.A.1. Let N be a $A\varphi$−mixing sequence. Let $\{s_i\}$ be a bounded signal sequence that satisfies (3.1.17), let also $\psi(x)$ be a measurable zero memory nonlinearity that satisfies $E\{\psi(N_1)\} = 0$ and $E\{\psi^2(N_1)\} < \infty$. Then $\sigma_0(\psi)$ is absolutely summable.

Proof. The proof is easy. Using (3.1.18), Lemma 3.1.1 for $q = r = 2$ and stationarity, we have

$$|\sigma_0^2(\psi)| \le E\{\psi^2(N_1)\} + 4\sum_{j=1}^{\infty} |\nu_j| \varphi_j^{1/2} E\{\psi^2(N_1)\} \le$$
$$E\{\psi^2(N_1)\}\{1 + 4\sum_{j=1}^{\infty}\varphi_j^{1/2}\} < \infty \tag{3.A.1}$$

and this concludes the proof. We now state the theorem that gives sufficient conditions for normality [1, p. 157].

Theorem 3.A.1. Let **U** be a sequence of random variables; define

$$R_n = \sum_{i=1}^{n} U_i \tag{3.A.2}$$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} U_i \qquad t \in [0,1] \tag{3.A.3}$$

where by [ ] we mean the integer part. Then if

i. $X_n(t)$ has asymptotically independent increments.

ii. $E\{X_n(t)\} \to 0$ and $E\{X_n^2(t)\} \to \sigma^2 t$

iii. $X_n^2(t)$ is uniformly integrable.

iv. For every ε there exist a $\beta > 1$ and an integer n_0 such that

$$P\left\{\max_{i \le n} |R_{k+i} - R_k| \ge \beta\sigma\sqrt{n}\right\} \le \frac{\varepsilon}{\beta^2} \quad \text{for every } k \text{ and } n \ge n_0$$

then $X_n(t)$ tends in distribution to a Brownian motion.

Proof. For the proof see [13,Theorem 8.4 and Theorem 19.2]. The next theorem gives the limiting forms for the quantities of our interest.

Theorem 3.A.2. Let N be a $A\varphi$−mixing sequence, let $\{s_i\}$ be a bounded signal sequence that satisfies (3.1.17). Let $\psi(x)$ be a measurable zero memory nonlinearity that satisfies $E\{\psi(N_1)\} = 0$ and $E\{[\psi(N_1)]^2\} < \infty$ and also Conditions (3.1.22) and (3.1.23). If we define $\lambda(t) = E\{\psi(N_1+t)\}$, then

$$T_n^0(X) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \psi(N_i) \overset{D}{\to} \mathrm{N}(0,\sigma_0(\psi)) \tag{3.A.4}$$

$$T_n^1(X) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \psi(N_i + \frac{ks_i}{\sqrt{n}}) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} s_i \lambda(\frac{ks_i}{n^{1/2}}) \overset{D}{\to} \mathrm{N}(0,\sigma_0(\psi)) \tag{3.A.5}$$

where $\xrightarrow{D}$ means convergence in distribution.

Proof. We will now show that our case satisfies the conditions given in Theorem 3.A.1. We will first prove (3.A.4). The proof is based on the fact that an $A\varphi$−mixing sequence satisfies all the above conditions (see [13, Theorem 20.1]). Let us define

$$R_n = \sum_{i=1}^{n} s_i \psi(N_1) \tag{3.A.6}$$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} s_i \psi(N_i) \tag{3.A.7}$$

$$S_n = \sum_{i=1}^{n} \psi(N_i) \tag{3.A.8}$$

$$Y_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \psi(N_i) \tag{3.A.9}$$

Notice that $X_n(t)$ is the process which we want to show satisfies the conditions of Theorem 3.A.1 and that $Y_n(t)$ is the stationary case which we know satisfies the conditions.

To show Condition i., let $0 \le u_1 \le v_1 < u_2 \le v_2 < \cdots < u_r \le v_r \le 1$ and $b = \min_i(u_i - v_{i-1})$. Also let $A_i \in M_{[nu_i]}^{[nv_i]}$. Then from (3.1.2) using induction we can show

$$|P(\bigcap_{i=1}^{r} A_i) - \prod_{i=1}^{r} P(A_i)| \le r\varphi_{[nb]} \to 0 \tag{3.A.10}$$

For Condition ii., we have $E\{X_n(t)\} = 0$. Also for $t > 0$

$$E\{X_n^2(t)\} = \frac{[nt]}{n} E\Big\{\Big[\frac{1}{\sqrt{[nt]}} \sum_{i=1}^{[nt]} s_i \psi(N_i)\Big]^2\Big\} \to t\,\sigma_0^2(\psi) \tag{3.A.11}$$

To show Condition iii., we have to show that

$$\lim_{\alpha\to\infty} \sup_n \int_{|X_n|>\alpha} X_n^2(t)dP = 0 \qquad \text{for every } t \tag{3.A.12}$$

Let M be a bound for the sequence $\{|s_i|\}$. Then since $|\frac{s_i}{M}| \le 1$ we have for the two events

$$\left\{ \frac{1}{\sqrt{n}} \left|\sum_{i=1}^{[nt]} \frac{s_i}{M}\psi(N_i)\right| > \alpha \right\} \subset \left\{ \frac{1}{\sqrt{n}} \left|\sum_{i=1}^{[nt]} \psi(N_i)\right| > \alpha \right\} \tag{3.A.13}$$

Thus if we define $\lambda_1(\omega) = P\{X_n^2(t) > \omega\}$ and $\lambda_2(\omega) = P\{Y_n^2(t) > \omega\}$ we have that

$$\lambda_1(M^2\omega) \le \lambda_2(\omega) \tag{3.A.14}$$

Because both $X_n(t)$ and $Y_n(t)$ have finite variance we can write

$$\int_{X_n^2(t)>\alpha^2M^2} X_n^2(t)dP = \alpha^2M^2\lambda_1(\alpha^2M^2) + \int_{\alpha^2M^2}^{\infty} \lambda_1(\omega)d\omega = M^2\Big(\alpha^2\lambda_1(\alpha^2M^2) + \int_{\alpha^2}^{\infty}\lambda_1(\omega M^2)d\omega\Big)$$

$$\le M^2\Big(\alpha^2\lambda_2(\alpha^2) + \int_{\alpha^2}^{\infty}\lambda_2(\omega)d\omega\Big) = \quad M^2 \int_{|Y_n^2(t)|>\alpha^2} Y_n^2(t)dP \tag{3.A.15}$$

and because $Y_n(t)$ is the stationary case we have

$$\lim_{\alpha\to\infty} \sup_n \int_{|X_n(t)|>\alpha M} X_n^2(t)dP \le \lim_{\alpha\to\infty} \sup_n M^2 \int_{|Y_n(t)|>\alpha} Y_n^2(t)dP = 0 \tag{3.A.16}$$

To prove Condition iv., notice that

$$P\left\{ \max_{i\le n} |R_{k+i} - R_k| \ge \beta\sigma\sqrt{n} \right\} = P\left\{ \bigcup_{i=1}^{n} \left\{ \left|\sum_{j=1}^{i} s_{j+k}\psi(N_j)\right| \ge \beta\sigma\sqrt{n} \right\} \right\} \tag{3.A.17}$$

and thus using (3.A.13)

$$P\left\{\max_{i\le n}|R_{k+i}-R_k|>M\beta\sigma\sqrt{n}\right\}\le P\left\{\bigcup_{i=1}^{n}\{|S_i|>\beta\sigma\sqrt{n}\}\right\}$$

$$=P\left\{\bigcup_{i=1}^{n}\{|S_{i+k}-S_k|>\beta\sigma\sqrt{n}\}\right\}=P\left\{\max_{i\le n}|S_{i+k}-S_k|>\beta\sigma\sqrt{n}\right\}<\frac{\varepsilon}{\beta^2} \tag{3.A.18}$$

for sufficiently large β and n. The last inequality is true because it comes from the stationary case. And if we define $\varepsilon' = \varepsilon M^2$ and $\beta' = \beta M$ we prove iv.

To prove (3.A.5) fortunately things are much easier. We will prove that the difference of $T_n^0(X)$ and $T_n^1(X)$ tends in the mean-square sense to zero. We have that

$$E\{[T_n^1(X)-T_n^0(X)]^2\}=\frac{1}{n}E\{[\sum_{i=1}^{n}s_i U_i]^2\} \tag{3.A.19}$$

where we define

$$U_i=\psi(N_i+\frac{ks_i}{n^{1/2}})-\psi(N_i)-\lambda(\frac{ks_i}{n^{1/2}}) \tag{3.A.20}$$

Notice that

$$E\{U_i^2\}=E\{[\psi(N_i+\frac{ks_i}{n^{1/2}})-\psi(N_i)]^2\}-\lambda^2(\frac{ks_i}{n^{1/2}}) \tag{3.A.21}$$

Because $\{s_i\}$ is a bounded sequence, given $\varepsilon>0$ from (3.1.22) we have that, for large enough n, $E\{U_i^2\}$ becomes smaller than ε for every i. And using Lemma 3.1.1

$$E\{[T_n^1(X)-T_n^0(X)]^2\}\le\frac{M^2}{n}\{\sum_{i=1}^{n}U_i^2+4\sum_{k<j}^{n}\sum^{n}\varphi_{j-k}^{1/2}[E\{U_j^2\}]^{1/2}[E\{U_k^2\}]^{1/2}\}\le$$

$$M^2\{1 + 4\sum_{i=1}^{\infty}\varphi_i^{1/2}\}\varepsilon \tag{3.A.22}$$

which proves the mean square convergence. Clearly now we have that the two quantities will also converge in distribution to the same limit and thus we have that (3.A.5) is also true.

APPENDIX 3.B

Asymptotic Normality for the Estimation Problem.

As we said in Subsection 3.1.2 we will prove normality only for the case where $\psi(x)$ is a monotone function. We follow the same idea as in [4]. Let us define

$$\lambda(\xi-\delta) = \int_{-\infty}^{\infty} \psi(x-\xi+\delta) f(x) dx \tag{3.B.1}$$

Since $\psi(x)$ satisfies $E\{\psi(N_1)\} = 0$, we have that

$$\lambda(0) = 0 \tag{3.B.2}$$

Theorem 3.B.1 Let N be a $A\varphi$−mixing sequence. Let also $\psi(x)$ be a continuous nondecreasing nonlinearity with $E\{\psi(N_1)\} = 0$ and $E\{[\psi(N_1)]^2\} < \infty$, satisfying (3.1.29), (3.1.30), (3.1.31) and (3.1.32). If $\{X_i = N_i + \delta\}$ for $i = 1,2,\ldots,n$, then the solution T_n to the Equation (3.1.28) satisfies

$$n^{1/2}(T_n - \delta) \overset{D}{\rightarrow} \mathrm{N}(0, A(\psi, \boldsymbol{N})) \tag{3.B.3}$$

There is only one small problem. Because we include nonlinearities that are not strictly monotone, it is possible for an equation like (3.1.28) to have as a solution a whole interval. In such case we assume that T_n is the lower end of the interval. Because of (3.1.22) as $n \rightarrow \infty$ the measure of this interval tends to zero and we will have a unique solution.

***Proof*.** To show (3.B.3), it is enough to show that for every real g we have

$$\lim_{n\to\infty} P\{n^{1/2}T_n < g\} = \Phi\left(\frac{g\lambda'(0)}{\sigma_0} \right) \tag{3.B.4}$$

where $\Phi(x)$ is the normalized Gaussian cumulative distribution and where, for simplicity, we have substituted σ_0 for $\sigma_0(\psi)$. Because of the monotonicity of $\psi(x)$ and the way T_n is defined, we have

$$[T_n < \frac{g}{n^{1/2}}] \subseteq [\sum_{j=1}^{n} \psi(N_j - \frac{g}{n^{1/2}}) \le 0] \subseteq [T_n \le \frac{g}{n^{1/2}}] \tag{3.B.5}$$

where by [ ] we denote the set where the relation is valid. Thus it is enough to prove that

$$P\{\sum_{j=1}^{n} \psi(N_j - \frac{g}{n^{1/2}}) \le 0\} \rightarrow \Phi\left(\frac{g\lambda'(0)}{\sigma_0} \right) \tag{3.B.6}$$

We will now prove a lemma that will help us to prove (3.B.6).

Lemma 3.B.1. Let us define

$$Y_n = \frac{1}{n^{1/2}\sigma_0} \sum_{j=1}^{n} \psi(N_j - \frac{g}{n^{1/2}}) - \frac{n^{1/2}}{\sigma_0}\lambda\left(\frac{g}{n^{1/2}} \right)$$

$$Z_n = \frac{1}{n^{1/2}\sigma_0} \sum_{j=1}^{n} \psi(N_j) \tag{3.B.7}$$

Then we have

$$\lim_{n\to\infty} E\{ (Y_n - Z_n)^2\} = 0 \tag{3.B.8}$$

Proof. We have that

$$E\{ (Y_n - Z_n)^2\} = \frac{1}{n\sigma_0^2} E\{[\sum_{j=1}^{n} U_j]^2\} \tag{3.B.9}$$

where we have defined

$$U_j = \psi(N_j - \frac{g}{n^{1/2}}) - \psi(N_j) - \lambda(\frac{g}{n^{1/2}}) \tag{3.B.10}$$

Using stationarity and Lemma 3.1.1, we have that

$$E\{ (Y_n - Z_n)^2\} \le \frac{1}{\sigma_0^2}(1 + 4\sum_{j=1}^{\infty} \varphi_j^{1/2})E\{U_1^2\} \tag{3.B.11}$$

Now we can see that

$$E\{U_1^2\} = E\{ [\psi(x - \frac{g}{n^{1/2}}) - \psi(x)]^2\} - \lambda^2(\frac{g}{n^{1/2}}) \tag{3.B.12}$$

and because of (3.1.22) we have that

$$\lim_{n\to\infty} E\{U_1^2\} = 0 \tag{3.B.13}$$

and this concludes the proof.

From Lemma 3.B.1 we can deduce that, if Z_n tends in distribution to some random variable, then Y_n will tend to the same random variable. But since the sequence $\{\psi(N_j)\}$ is $A\varphi$−mixing , we know that

$$\frac{1}{n^{1/2}\sigma_0} \sum_{j=1}^{n} \psi(N_j) \overset{D}{\to} \mathrm{N}(0,1) \tag{3.B.14}$$

Also because of (3.1.30) we have

$$\lim_{n\to\infty} n^{1/2} \lambda(\frac{g}{n^{1/2}}) = g\lambda'(0) \tag{3.B.15}$$

Thus we can conclude

$$\frac{1}{n^{1/2}\sigma_0} \sum_{j=1}^{n} \psi(N_j - \frac{g}{n^{1/2}}) \overset{D}{\to} \mathrm{N}(\frac{g\lambda'(0)}{\sigma_0},1) \tag{3.B.16}$$

and thus (3.B.5) is true. And this concludes the proof of Theorem 3.B.1.

CHAPTER IV.

OPTIMUM DETECTION WITH MINIMAL KNOWLEDGE OF DEPENDENCY

4.1 Preliminaries.

In this chapter our goal is to define optimum nonlinearities with only a vague knowledge of the dependency structure. Specifically, we assume that the noise sequence is one-dependent and strictly stationary and that the bivariate density of consecutive members of the sequence belongs to the First Model introduced in Section 2.1. In addition to this general information, we also assume that we know the correlation coefficient between consecutive members of the sequence and also the common marginal density. These assumptions are very nice from a practical point of view because there are ways of estimating marginal densities and correlation coefficients [2,3]. Let us now define things more formally.

Let $f(x)$ be a symmetric density with unbounded support, such that all its moments exist and its Fisher's information is finite. Assume that the orthonormal polynomials $\varphi_n(x)$ defined by $f(x)$ form a complete orthonormal system in the $L_2(f)$ Hilbert space. We are interested in bivariate densities that can be represented as

$$f(x,y) = f(x)f(y)\left\{\sum_{n=0}^{\infty} \alpha_n \varphi_n(x)\varphi_n(y)\right\} \tag{4.1.1}$$

which is the diagonal expansion for the First Model. As we discussed in Section 2.1, a necessary condition for (4.1.1) to be a valid density is

$$\alpha_n = \int_{-1}^{1} z^n h(z) dz \qquad n = 0,1,2,... \tag{4.1.2}$$

where $h(z)$ is a univariate density supported in $(-1,1)$. We can easily see that α_1 is equal to the correlation coefficient of x and y. As we said, the noise sequence N is assumed to be one-dependent. Before continuing, we will prove a lemma that gives us more information about the coefficients α_n of (4.1.1) for one-dependent sequences.

***Lemma* 4.1.1** Let $f(x,y)$ be a bivariate density of consecutive members of a one-dependent stationary sequence. Assume that it can be represented by Equation (4.1.1). Then,

$$| \alpha_n | \leq ½ \tag{4.1.3}$$

***Proof*.** Let $\{c_j\}$ be any sequence of real numbers. Then

$$E\{[\sum_{j=1}^{m} c_j \varphi_n(N_j)]^2\} \geq 0 \tag{4.1.4}$$

Manipulating (4.1.4) we get

$$\sum_{j=1}^{m} c_j^2 + 2[\sum_{j=1}^{m-1} c_j c_{j-1}]\alpha_n \geq 0 \tag{4.1.5}$$

By putting $c_j = 1$ for every j or $c_j = (-1)^j$ for every j and also letting $m \to \infty$, we get $\alpha_n \geq -½$ and $\alpha_n \leq ½$ respectively. And this concludes the proof. Notice also that because of (4.1.2), if α_2 is no greater than ½ in absolute value, then so is any other α_n for $n > 2$.

We now define a class F_ρ of functions $f(x,y)$. A function $f(x,y)$ belongs to F_ρ if it satisfies Equations (4.1.1) and (4.1.2) for some density $h(z)$ supported on [-1,1] and if

$$\alpha_1 = \int_{-1}^{1} z h(z) dz = \rho \tag{4.1.6}$$

where ρ is the known correlation coefficient. By allowing $h(z)$ to be supported on $[-1,1]$ we allow degenerate functions in the class F_ρ. Clearly F_ρ contains the bivariate densities of all one-dependent sequences that can be represented by (4.1.1). As we can see, we do not take into account that $|\alpha_2| \le \frac{1}{2}$. But we will see that this requirement is automatically satisfied for a region of values of ρ.

The detection problem we would like to solve is the one we defined in Subsection 3.1.1. As we said, our performance measure is the efficacy, which for one-dependent sequences takes the form

$$eff\left(\psi(x), f(x,y)\right) = \frac{\left[\int_{-\infty}^{\infty} \psi(x) f'(x) dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x) f(x) dx + 2\nu_1 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y) f(x,y) dx dy} \tag{4.1.7}$$

where $f(x,y)$ is the bivariate density of N_1 and N_2.

For the nonlinearity $\psi(x)$, we assume that it belongs to the class Ψ_m of all odd symmetric polynomials that have order up to $2m-1$. Since by assumption the polynomials are dense in $L_2(f)$, by letting $m \to \infty$, we have that Ψ_∞ is the class of all odd symmetric nonlinearities that satisfy $E\{\psi^2(N_1)\} < \infty$. The restriction to odd symmetric polynomials is reasonable because we can prove that, for every function in the class F_ρ, the nonlinearity that maximizes (4.1.7) is odd symmetric.

4.2 Optimum Nonlinearity.

The nonlinearity that maximizes (4.1.7) is related to the actual bivariate density $f(x,y)$. Since we do not assume knowledge of this density, we define the optimum nonlinearity in a min-max way. In other words, we would like to find a pair $\psi_r(x) \in \Psi_m$ and $f_l(x,y) \in F_\rho$ such that the following saddle-point relation is satisfied :

$$eff\left(\psi(x),f_l(x,y)\right) \le eff\left(\psi_r(x),f_l(x,y)\right) \le eff\left(\psi_r(x),f(x,y)\right) \tag{4.2.1}$$

for every $\psi(x) \in \Psi_m$ and every $f(x,y) \in F_\rho$.

First we will find the function from F_ρ that minimizes (4.1.7) for a given $\psi(x)$. Let us, for simplicity, assume that $\nu_1 > 0$ and we will comment in Section 4.3 for the case $\nu_1 < 0$. Thus, minimizing (4.1.7) is equivalent to the following maximization

$$\sup_{f(x,y)\in F_\rho} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y)f(x,y)dxdy \tag{4.2.2}$$

Since $\psi(x)$ is odd symmetric, it can be expanded using only the odd symmetric orthonormal polynomials. Let

$$\psi(x) = \sum_{n=1}^{m} \psi_n \varphi_{2n-1}(x) \tag{4.2.3}$$

Using (4.2.3) we can write

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(x)\psi(y)f(x,y)dxdy = \sum_{n=1}^{m} \psi_n^2 \alpha_{2n-1} = \sum_{n=1}^{m} \psi_n^2 \int_{-1}^{1} z^{2n-1}h(z)dz$$

$$= \int_{-1}^{1} \left[\sum_{n=1}^{m} \psi_n^2 z^{2n-1} \right] h(z) dz = \int_{-1}^{1} A(z) h(z) dz \tag{4.2.4}$$

where we defined as $A(z)$

$$A(z) = \sum_{n=1}^{m} \psi_n^2 z^{2n-1} \tag{4.2.5}$$

The manipulations in (4.2.4) hold for the case $m = \infty$ as well, because the series is absolutely convergent for $|z| \leq 1$. We can also interchange summation and integration in (4.2.4) using bounded convergence. The maximization problem now reduces to the following:

$$\sup_{h(z)} \int_{-1}^{1} A(z) h(z) dz \tag{4.2.6}$$

given that

$$\int_{-1}^{1} z h(z) dz = \rho \tag{4.2.7}$$

Notice the following properties of $A(z)$: It is increasing and bounded in [-1,1], it is analytic in $(-1,1)$, it is odd symmetric and, for $z > 0$, it is convex. A typical form of $A(z)$ is given in Figure 4.2.1. Now let $B(z) = \lambda z + \mu$ be the line that passes through the point $(1, A(1))$ and is tangent to $A(z)$ at $-z_0$, (see Figure 4.2.1). Then, for every z, we have that

$$B(z) \geq A(z) \tag{4.2.8}$$

Notice that the point $-z_0$ can be found by solving the equation

$$\frac{A(1) - A(-z_0)}{1 + z_0} = A'(-z_0) \tag{4.2.9}$$

The point z_0 is unique for the case where there exists at least one $\psi_n \neq 0$ for $n > 1$. This is true because then $A'(z)$ is strictly increasing.

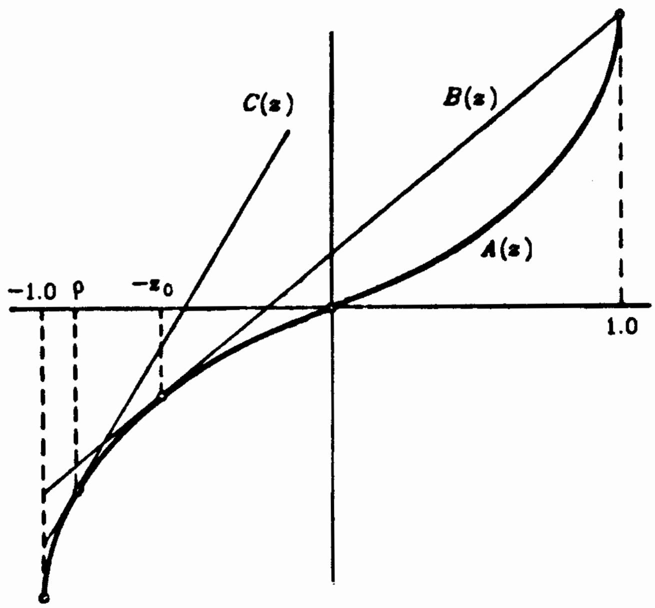

Figure 4.2.1 Typical form of $A(z)$ and of the tangent lines $B(z)$ and $C(z)$.

We now prove a proposition that gives the solution to the maximization problem defined by (4.2.6) and (4.2.7).

***Proposition* 4.2.1.** The density $h(z)$ that solves the maximization problem defined by (4.2.6) and (4.2.7) is given by one of the following two cases:

Case 1. If $z_0 \geq -\rho$, the maximum is achieved by

$$h_M(z) = \frac{z_0 + \rho}{1 + z_0}\delta(z - 1) + \frac{1 - \rho}{1 + z_0}\delta(z + z_0) \tag{4.2.10}$$

Case 2. If $z_0 < -\rho$, the maximum is achieved by

$$h_M(z) = \delta(z - \rho) \qquad (4.2.11)$$

***Proof*.** We can see that, in both cases, $h_M(z)$ is a valid univariate density satisfying (4.2.7). For Case 1 maximizing (4.2.6) is equivalent to the following:

$$\sup_{h(z)} \left\{ \int_{-1}^{1} A(z)h(z)dz - \lambda\rho - \mu \right\} \qquad (4.2.12)$$

But, using (4.2.8), we have

$$\int_{-1}^{1} A(z)h(z)dz - \lambda\rho - \mu = \int_{-1}^{1} \big(A(z) - \lambda z - \mu\big)h(z)dz \le 0 \quad (4.2.13)$$

Equality to zero is achieved when $h(z)$ is supported only on points where $A(z) = \lambda z + \mu$. For Case 2 the proof proceeds in a similar way. Instead of the line $B(z)$ that is tangent at $-z_0$, we use the line $C(z)$ that is tangent at ρ . The same arguments are valid because this line, as a result of the convexity, is always above A(z), (see Figure 4.2.1).

The function $f(x,y) \in F_\rho$ that corresponds to (4.2.10) is given by

$$f_M(x,y) = (1-p)f(x)\delta(x-y) + pf(x)f(y)\left\{ \sum_{n=0}^{\infty} (-z_0)^n \varphi_n(x)\varphi_n(y) \right\}$$

$$(4.2.14)$$

where

$$p = \frac{1-\rho}{1+z_0} \qquad (4.2.15)$$

Let us now find the optimum nonlinearity $\psi(x) \in \Psi_m$, when the function $f(x,y)$ has a form similar to the one given by (4.2.14). Since $f(x)$ has finite Fisher's information and is symmetric, we can write

$$-\frac{f'(x)}{f(x)} = \sum_{n=1}^{\infty} \beta_n \varphi_{2n-1}(x) \tag{4.2.16}$$

The efficacy for a $\psi(x) \in \Psi_m$ takes the form

$$eff\left(\psi(x), f_M(x,y)\right) = \frac{\left[\sum_{n=1}^{m} \psi_n \beta_n\right]^2}{\sum_{n=1}^{m} \psi_n^2 + 2\nu_1 \sum_{n=1}^{m} \psi_n^2\left[(1-p) - p z_0^{2n-1}\right]}$$

$$= \frac{\left[\sum_{n=1}^{m} \psi_n \beta_n\right]^2}{\sum_{n=1}^{m} \psi_n^2\left[1+2\nu_1 - 2\nu_1 p(1 + z_0^{2n-1})\right]} \tag{4.2.17}$$

Equation (4.2.17) is maximized when $\psi_n = \psi_n^*$, where

$$\psi_n^* = \frac{k\beta_n}{\sqrt{1+2\nu_1 - 2\nu_1 p(1 + z_0^{2n-1})}} \qquad n = 1,2,\ldots,m \tag{4.2.18}$$

and k is an arbitrary constant. Thus, for a given $\psi(x)$, the function $f(x,y)$ that minimizes the efficacy is given by (4.2.14). On the other hand if $f(x,y)$ has the form of (4.2.14) then the optimum $\psi(x)$ satisfies (4.2.18).

In order to find the pair we are looking for, we have to satisfy (4.2.14) and (4.2.18) simultaneously. We will assume that Case 1 of Proposition 4.2.1 will occur and that our $\psi_r(x)$ satisfies (4.2.18) for some $z_0 = z_r$. Thus for $\psi_r(x)$ we only have to specify z_r in some way. For every z_r

and the corresponding $\psi_r(x)$, we define a function $A_r(z)$, similar to $A(z)$ defined in (4.2.5) as follows:

$$A_r(z) = \sum_{n=1}^{m} \frac{\beta_n^2 z^{2n-1}}{1+2\nu_1 - 2\nu_1 p (1 + z_r^{2n-1})} \tag{4.2.19}$$

where p is given by

$$p = \frac{1-\rho}{1+z_r} \tag{4.2.20}$$

Notice that for simplicity we assumed $k = 1$. Since, from (4.2.1), we would like $f_l(x,y)$ to minimize the $eff\left(\psi_r(x), f(x,y)\right)$, the function $f_l(x,y)$ must have a form similar to (4.2.14) with $z_0 = z_r$. In order for this form to minimize the efficacy, z_r must be a solution of an equation similar to (4.2.9). In other words

$$\frac{A_r(1) - A_r(-z_r)}{1+z_r} = A_r{}'(-z_r) \tag{4.2.21}$$

Substituting (4.2.19) into (4.2.21) and multiplying by $(1 + z_r)$, after canceling common terms, we find that (4.2.21) reduces to

$$\sum_{n=2}^{m} \frac{\beta_n^2 (1 + z_r^{2n-1})}{1+2\nu_1 - 2(1-\rho)\nu_1 \dfrac{1+z_r^{2n-1}}{1+z_r}} = (1+z_r) \sum_{n=2}^{m} \frac{(2n-1)\beta_n^2 z_r^{2(n-1)}}{1+2\nu_1 - 2(1-\rho)\nu_1 \dfrac{1+z_r^{2n-1}}{1+z_r}} \tag{4.2.22}$$

Equation (4.2.22) has z_r as its only unknown. In Appendix 4.A, we show that a positive solution always exists and that it is no less than ½. This means that we always have $z_r \geq \rho$ and that we do not contradict our assumption that Case 1 will occur. We will have a contradiction though, if we assume that Case 2 will occur.

Theorem 4.2.1. Let z_r be a solution to (4.2.22). Define

$$f_l(x,y) = \frac{z_r + \rho}{1 + z_r} f(x)\delta(x-y) + \frac{1-\rho}{1 + z_r} f(x)f(y)\left\{ \sum_{n=0}^{\infty} (-z_r)^n \varphi_n(x)\varphi_n(y) \right\}$$

$$\psi_r(x) = k \sum_{n=1}^{m} \frac{\beta_n}{\sqrt{1+2\nu_1 - 2(1-\rho)\nu_1 \dfrac{1 + z_r^{2n-1}}{1 + z_r}}} \varphi_{2n-1}(x) \qquad (4.2.23)$$

where k is an arbitrary constant. Then $\psi_r(x)$ and $f_l(x,y)$ satisfy (4.2.1).

Proof. The proof is an immediate consequence of the way that z_r is defined. The left inequality of (4.2.1) is satisfied, because $\psi_r(x)$ satisfies (4.2.18). The right inequality is satisfied, because $f_l(x,y)$ minimizes the $eff\left(\psi_r(x), f(x,y)\right)$.

Comments. It is worth noticing a few things: When either $m = 2$ or $\beta_n = 0$ for $n = 3,\ldots,m$, the only nonnegative solution to (4.2.22) is $z_r = 1/2$, regardless of β_2, as long as $\beta_2 \neq 0$. Also when $\nu = 1$, then for any $m < \infty$, when $\rho \to -1/2$ we have that $\psi_r(x) \to x$ (after it is properly normalized), so long as $\beta_1 \neq 0$. Notice that $z_r \neq 0$ even when $\rho = 0$, which means that the independence assumption is not necessarily correct, when the correlation coefficient is zero.

4.2.1 Examples.

Let us now present three examples and let us, for simplicity, assume that $\nu_1 = 1$. In the first two examples, as we will see, we get the Ψ_∞ solution. But in the third example things are more complicated and more interesting.

First Example. We consider the case where $f(x)$ is the standard $N(0,1)$ Gaussian density. The orthonormal polynomials are the Hermite polynomials. Since the locally optimum nonlinearity is linear, in other words equal to $\varphi_1(x)$, we have that $\beta_n = 0$ for $n \geq 2$. From (4.2.23) we conclude that $\psi_r(x)$ will be linear too.

Second Example. We consider $f(x)$ to be the following

$$f(x) = 0.0802\, e^{-\frac{x^4}{8.754}} \tag{4.2.24}$$

The numbers are selected in order to have a variance equal to unity. The locally optimum nonlinearity is

$$-\frac{f'(x)}{f(x)} = 0.457\, x^3 \tag{4.2.25}$$

To represent this function using the orthonormal polynomials, we need only $\varphi_1(x)$ and $\varphi_3(x)$. From (4.2.23) we see that this is also the case for $\psi_r(x)$. We have that

$$\varphi_1(x) = x \qquad \varphi_3(x) = 1.333\, x^3 - 2.916\, x \tag{4.2.26}$$

and also

$$\beta_1 = 1.0 \qquad \beta_2 = 0.343 \tag{4.2.27}$$

Since $\beta_n = 0$ for $n \geq 3$ we get that $z_r = \frac{1}{2}$. The optimum nonlinearity, after it is normalized, becomes

$$\psi_r(x) = x + \sqrt{\frac{1+2\rho}{1+\rho}} \left\{ 0.373\, x^3 - 0.8165\, x \right\} \tag{4.2.28}$$

In Figure 4.2.2 we can see $\psi_r(x)$ for different values of the correlation coefficient ρ.

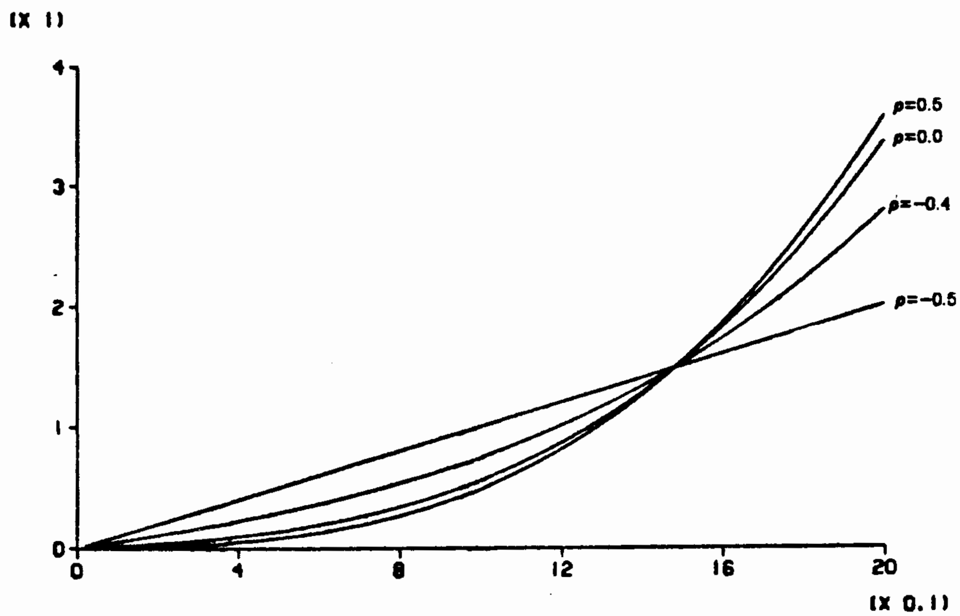

Figure 4.2.2 Optimum nonlinearities for the Second Example.

Third Example. We will pay a little more attention to this example because it gives us an idea about the usefulness of the method. Consider $f(x)$ to be a contaminated Gaussian density of the form

$$f(x) = \begin{cases} \dfrac{0.857}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} & \text{for } |x| \leq 1 \\ \dfrac{0.857}{\sqrt{2\pi}} e^{\frac{1}{2}-|x|} & \text{for } |x| \geq 1 \end{cases} \tag{4.2.29}$$

The locally optimum nonlinearity then is

$$-\frac{f'(x)}{f(x)} = \begin{cases} x & \text{for } |x| \le 1 \\ sgn(x) & \text{for } |x| \ge 1 \end{cases} \qquad (4.2.30)$$

In Table 4.2.1 we give the first five odd symmetric orthonormal polynomials and the corresponding β_n . The a_n in the table correspond to the coefficient of x^n and, because the polynomials are odd symmetric, we have only the odd-term coefficients.

***Table* 4.2.1** Orthonormal polynomials and corresponding coefficients of the expansion of the locally optimum nonlinearity.

Order	β_n	a_9	a_7	a_5	a_3	a_1
$\varphi_1(x)$	0.667					0.667
$\varphi_3(x)$	-0.242				0.045	-0.547
$\varphi_5(x)$	0.152			9.2×10^{-4}	- 0.068	0.483
$\varphi_7(x)$	-0.112		8.9×10^{-6}	-2×10^{-3}	0.081	-0.441
$\varphi_9(x)$	0.089	5.0×10^{-8}	-2.5×10^{-5}	3.1×10^{-3}	-0.091	0.411

In Table 4.2.2 the values of z_r are given for different values of the parameter m of the class Ψ_m and the correlation coefficient ρ . We can see that z_r changes very little over the range of values of the two parameters.

***Table* 4.2.2** Values of z_r for different values of m and ρ .

m \ ρ	-0.5	-0.3	-0.1	0.0	0.1	0.3	0.5
2	0.5	0.5	0.5	0.5	0.5	0.5	0.5
3	0.529	0.531	0.532	0.533	0.533	0.534	0.534
4	0.546	0.550	0.552	0.552	0.553	0.554	0.555
6	0.716	0.718	0.719	0.719	0.720	0.720	0.720
8	0.736	0.738	0.739	0.739	0.739	0.740	0.740
10	0.751	0.753	0.754	0.754	0.755	0.755	0.755

In Figure 4.2.3 we can compare the polynomials $p_m(x)$ given by

$$p_m(x) = \sum_{n=1}^{m} \beta_n \varphi_{2n-1}(x) \qquad (4.2.31)$$

with the locally optimum nonlinearity defined in (4.2.30). For large x things are very bad.

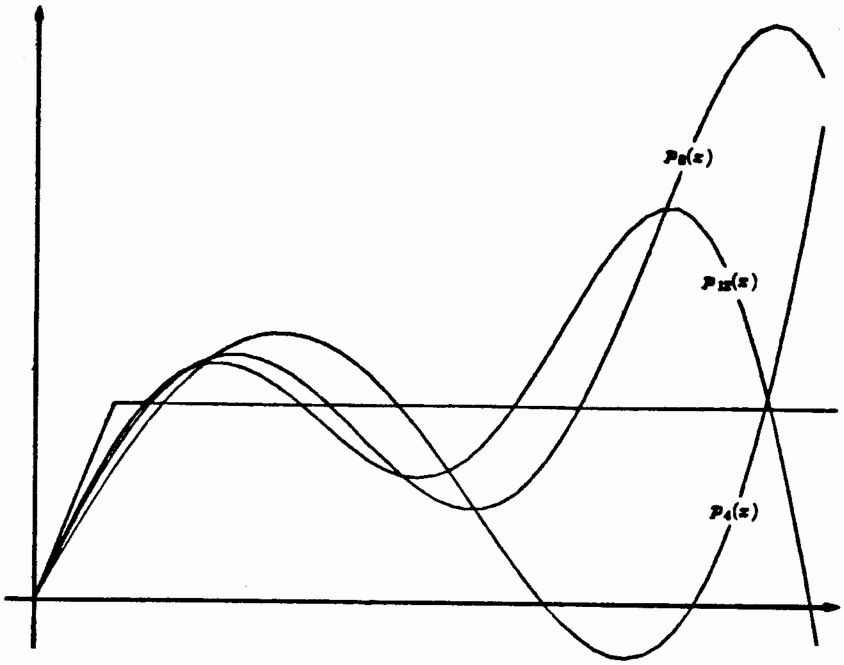

Figure 4.2.3 The polynomials $p_m(x)$ and the locally optimum nonlinearity.

In Figure 4.2.4 we can see the polynomial $p_{10}(x)$ and the optimum nonlinearities for different values of the parameter ρ.

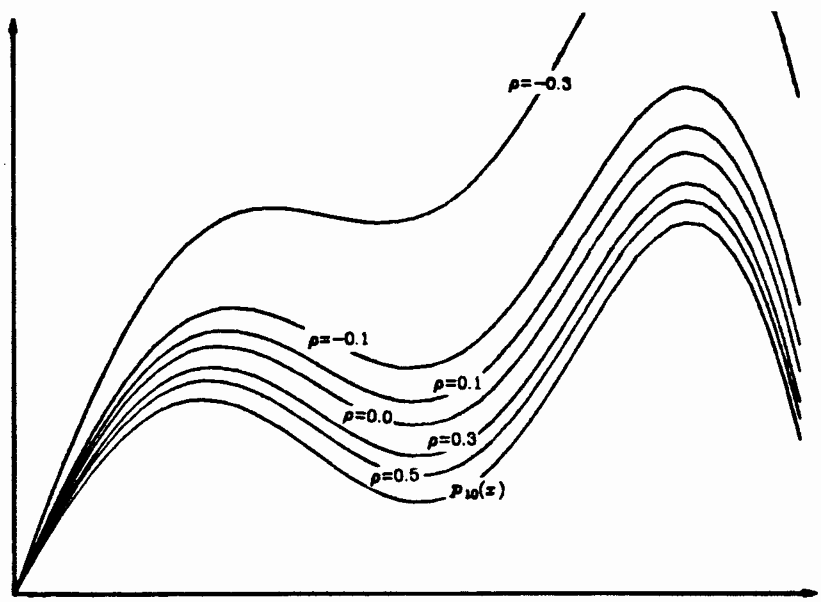

Figure 4.2.4 Optimum nonlinearities for the case $m = 10$.

Comment. To see for what values of ρ and m we can have $0 \le \alpha_2 \le 1/2$, notice that the α_2 for the density $f_l(x,y)$ defined in (4.2.23) becomes

$$\alpha_2 = (1 - p) + z_r^2 p \tag{4.2.32}$$

where p is defined by (4.2.20). In order to have $\alpha_2 \le 1/2$, after using (4.2.20), we need

$$1 - \frac{1}{2(1-\rho)} \ge z_r \tag{4.2.33}$$

Now because $z_r \ge 1/2$, it turns out that a necessary condition for (4.2.33) to be true is that $\rho \le 0$. For $\rho = -0.5, -0.3,$ and -0.1 we have that the left side of (4.2.33) becomes 0.667, 0.615 and 0.545 respectively. We can compare these bounds with the z_r given in Table 4.2.2. The first two satisfy (4.2.33) for $m \le 4$ and the third for $m \le 3$.

4.3 Other Cases and Generalizations.

The Case $\nu_1 < 0$. In Section 4.2 we solved the problem for the case $\nu_1 > 0$. Now we will consider the case $\nu_1 < 0$. If we define $x = -w$, because of the symmetry of $f(x)$, we have that $f(-w,y)$ is a bivariate density satisfying (4.1.1) and (4.1.2) with w in place of x. Because of the odd symmetry of $\psi(x)$ we have that the efficacy becomes

$$eff\,(\psi(w), f(-w,y)) = \frac{[\int_{-\infty}^{\infty} \psi(w) f'(w) dw]^2}{\int_{-\infty}^{\infty} \psi^2(w) f(w) dw + (-\nu_1) \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \psi(w)\psi(y) f(-w,y) dw dy} \tag{4.3.1}$$

This is exactly the same as in Section 4.2 since $-\nu_1 > 0$. The important point is that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} wyf(-w,y)dwdy = -\rho \tag{4.3.2}$$

In other words we solve the problem for $-\nu_1$ and $-\rho$ using Theorem 4.2.1 and in the result we substitute x with $-x$. But since the nonlinearities are odd symmetric, we can see that we will have the same nonlinearity. The least favorable function $f_l(x,y)$ will be different however.

4.3.1 Generalization to the M-Dependent Case.

The M-dependent case can be solved using the results of Section 4.2. The only problem here is that finding the worst case requires some experimentation. Let us for simplicity assume that $\nu_j = 1$ for $j = 1,\ldots,M$. The efficacy then takes the form:

$$eff(\psi(x),\{f_j(x,y)\}) = \frac{[\int_{-\infty}^{\infty}\psi(x)f'(x)dx]^2}{\int_{-\infty}^{\infty}\psi^2(x)dx + 2\sum_{j=1}^{M}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi(x)\psi(y)f_j(x,y)dxdy} \tag{4.3.3}$$

For this case we assume that we know the correlation coefficients ρ_j between N_1 and N_{j+1} for $j = 1,2,\ldots,M$. These coefficients have to satisfy the positive definiteness condition. Here we will need only two necessary conditions which we present them in the following lemma.

Lemma 4.3.1. Let N be an M-dependent stationary sequence and let the correlation coefficient between N_1 and N_{j+1} for $j = 1,2,\ldots,M$ be

ρ_j. Then, the following is true:

$$\sum_{j=1}^{M} \rho_j \geq -\tfrac{1}{2} \tag{4.3.4}$$

$$|\rho_j| \leq \tfrac{1}{2} \quad \text{for } j > \frac{M}{2} \tag{4.3.5}$$

***Proof*.** We will prove (4.3.4) only, because (4.3.5) can be proved in a similar way as (4.1.3) was proved. Assume, for simplicity, a variance equal to unity. Then

$$E\{[N_1 + \cdots + N_n]^2\} = n + 2(n-1)\rho_1 + \cdots + 2(n-M)\rho_M \geq 0 \tag{4.3.6}$$

on letting $n \to \infty$, we get what we want.

Notice now that we can maximize separately each term in the denominator sum of (4.3.3), using Proposition 4.2.1. We can see that we will have the same $A(z)$ and thus the same z_0, for all j, but different p and different $f_M(x,y)$ functions. Thus, to each j, there corresponds a p_j and a $f_M^j(x,y)$. The parameters p_j and the functions $f_M^j(x,y)$ will depend on whether Case 1 or Case 2 of Proposition 4.2.1 is true. Unfortunately we cannot know *a priory* which case will be true, as we did in the one-dependent case. Using the worst case functions $f_M^j(x,y)$ from the classes F_{ρ_j}, the efficacy takes the form

$$eff\left(\psi(x),\{f_M^j(x,y)\}\right) = \frac{\left[\sum_{n=1}^{m} \psi_n \beta_n\right]^2}{\sum_{n=1}^{m} \psi_n^2 + 2\sum_{j=1}^{M}\sum_{n=1}^{m} \psi_n^2\{1-p_j + p_j z_j^{2n-1}\}} \tag{4.3.7}$$

where

$$p_j = \frac{1-\rho_j}{1+z_0} \quad \text{and} \quad z_j = -z_0 \quad \text{when } z_0 \geq -\rho_j$$
$$p_j = 1 \qquad \text{and} \quad z_j = \rho_j \qquad \text{when } z_0 < -\rho_j \tag{4.3.8}$$

The optimum ψ_n that maximize (4.3.7) are given by

$$\psi_n^* = \frac{k\beta_n}{\sqrt{2M+1 - 2\sum_{j=1}^{M} p_j(1-z_j^{2n-1})}} \tag{4.3.9}$$

The denominator of (4.3.9), because of (4.3.4), can be shown to be nonnegative. Following steps similar to those in Section 4.2, we define a function $A_r(z)$ as follows:

$$A_r(z) = \sum_{n=1}^{m} \frac{\beta_n^2 z^{2n-1}}{2M+1-2\sum_l(1+\rho_l^{2n-1}) - 2\left[\sum_k(1+\rho_k)\right] \frac{1+z_r^{2n-1}}{1+z_r}} \tag{4.3.10}$$

where l runs through all indices that satisfy Case 2 and k through the ones that satisfy Case 1. The equation that defines z_r is

$$\frac{A_r(1) - A_r(-z_r)}{1 + z_r} = A_r'(-z_r) \tag{4.3.11}$$

Depending on how we separate the set of M indices into two sets for l and k, we get a different $A_r(z)$ function and a different equation (4.3.11). A solution of (4.3.11) will be acceptable when it does not contradict the assumptions we made for separating the indices. In order to make this statement more clear, notice that, because of (4.3.5), we know that for $j > \frac{M}{2}$ we have Case 1. For $j \leq \frac{M}{2}$ if we order the correlation coefficients as follows

$$-\rho_{(1)} \le -\rho_{(2)} \le \cdots \le -\rho_{([\frac{M}{2}])} \tag{4.3.12}$$

where (i) corresponds to the index of the coefficient that is in the i-th place, then, there are at most $[\frac{M}{2}]+1$ different intervals in which z_r can be. If, for example, we assume that $-\rho_{(j)} \le z_r < -\rho_{(j+1)}$ where $j \le [\frac{M}{2}]+1$; then k will take the values $(1),(2),\ldots,(j)$ and $([\frac{M}{2}])+1,\ldots,M$; on the other hand, l will take the values $(j+1),\ldots,([\frac{M}{2}])$. This separation of indices will define an equation (4.3.11) and a solution z_r of this equation. Now if this solution does not contradict the assumption that $-\rho_{(j)} \le z_r < -\rho_{(j+1)}$ then it is acceptable and can be used to solve our problem. It is still an open problem whether there always exists a solution to (4.3.11).

Comments. We have presented a method for finding an optimum nonlinearity for detection when dependency is present. For the dependency, we have assumed knowledge only of the correlation coefficient between M consecutive observations. Even though this method is tractable from an analytical point of view, it produces some problems in practice. The generation of the orthonormal polynomials is difficult for high orders. If we consider the polynomials as an approximation to the optimum nonlinearity for the Ψ_∞ case, their convergence is slow in cases where this optimum nonlinearity is bounded. That is because we approximate a bounded function using unbounded polynomials. Also, from Equation (4.2.23), we can see that the density $f_l(x,y)$ contains a delta-function component. This function is not a good candidate for a bivariate density of a one-dependent sequence. The reason we get this form of worst case density is because we optimize using only necessary and not sufficient conditions. By requiring the functions $f(x,y)$ to satisfy more necessary conditions, we could probably get better results. For example, in the one-dependent case, we could take into account (4.1.3) and restrict further the class F_ρ of allowable functions $f(x,y)$, but a more complicated analysis results. Besides, as we have seen, (4.1.3) is satisfied for a range of values of ρ with the analysis that we present in Section 4.2. This means that we do not gain much by using (4.1.3). A better idea might be to solve this problem for Markov sequences. For these sequences, any form of the bivariate density is allowable.

REFERENCES.

[1] J.H. Miller and J.B. Thomas, "Detection for discrete time signals in non-Gaussian noise," *IEEE Trans. Inf. Theory*, IT-18, pp. , March 1972.

[2] S.C. Schwartz, "Estimation of probability density by an orthogonal series," *Ann. Math. Stat.*, vol. 38, No 4, Aug. 1967.

[3] ---, "On the estimation of a Gaussian convolution probability density," *SIAM J. Appl. Math.*, vol. 17, No 2, March 1969.

[4] G. Szegö, Orthonormal Polynomials, *Amer. Math. Soc. Publ. 23*, Math. Soc., Providence, 1967.

[5] M. Abramowitz and I.A. Stegun, Handbook of mathematical functions. Dover Publications, New York, 1970.

APPENDIX 4.A

Existence of a solution z_r. We would like to show existence of a solution for Equation (4.2.22). First notice that, since by assumption $|\nu_1| \leq 1$,if we use (4.2.20), then

$$1 + 2\nu_1 - 2p\nu_1(1 + z_0^{2n-1}) \geq 1 + 2\nu_1\rho \geq 0 \tag{4.A.1}$$

This is important because in (4.2.18) we take the square root of this expression. Let us now define as $D(z_r)$ and $G(z_r)$ the left and right side of Equation (4.2.22), respectively. Notice that each term in these two expressions is a continuous function of z_r . Using bounded convergence, we can show that, for $\rho > -\frac{1}{2}$, the functions $D(z_r)$ and $G(z_r)$ are continuous and absolutely summable on $[0,1]$ and $[0,1)$, respectively. By direct calculation, we have

$$D(0) \geq G(0) \tag{4.A.2}$$

We also have that

$$G(z_r) \geq (1 + z_r) \sum_{n=2}^{m} \frac{\beta_n^2 z_r^{2(n-1)}}{1+2\nu_1 - 2(1-\rho)\nu_1 \dfrac{1 + z_r^{2n-1}}{1 + z_r}} \tag{4.A.3}$$

The right side of (4.A.3) is continuous and absolutely summable on $[0,1]$ and thus, in the limit as $z_r \to 1$, we get

$$\liminf_{z_r \to 1} G(z_r) \geq \frac{2}{1 + 2\nu_1\rho} \sum_{n=2}^{m} \beta_n^2 = D(1) \tag{4.A.4}$$

Continuity of the two expressions, combined with (4.A.2) and (4.A.3), proves existence of a solution in the interval $[0,1)$. To show now that this solution cannot be less than ½ it is enough to show that every term in the difference $D(z_r) - G(z_r)$ is nonnegative for $z_r \leq ½$. In other words, it is enough to show that for $0 \leq z_r \leq ½$

$$(1 + z_r^{2n-1}) - (2n - 1)(1 + z_r)\, z_r^{2(n-1)} \geq 0 \tag{4.A.5}$$

or equivalently

$$1 - (2n - 2)z_r^{2n-1} - (2n - 1)z_r^{2(n-1)} \geq 0 \tag{4.A.6}$$

The left side of (4.A.6) is decreasing with $z_r \in [0,1]$; thus it is enough to show (4.A.6) for $z_r = ½$ or, after some manipulation, to show

$$1 \geq \frac{3n - 2}{4^{n-1}} \tag{4.A.7}$$

This is true for every $n \geq 1$. Thus the solution to (4.2.22) satisfies $0 \leq z_r \leq ½$. This concludes the proof.

CHAPTER V.

MIN-MAX DETECTION AND ESTIMATION

5.1 Preliminaries.

The problem we would like to solve in this chapter is similar in nature to the one in Chapter IV. Here though we go one step further. In Chapter IV we assumed that we know the marginal density of the observation sequence. Here we assume that this marginal is not exactly known. Instead, we assume that it belongs to a known ε-contamination class of densities. Before going into further details, let us refer to to some earlier work in this area.

As we said in Chapter III our measure of performance for both detection and estimation problems is the efficacy. It is well known that, for i.i.d. observations, the efficacy is maximized when the nonlinearity $\psi(x)$ in (3.1.14) or (3.1.29) is given by the locally optimum nonlinearity defined by the common marginal density. When the density is not known exactly, optimality is usually defined in a min-max way. Following the ideas of Huber on robust estimation and hypothesis testing [1,2] the min-max nonlinearities for detection are derived in [3,4,5] for the i.i.d. case and for densities belonging to an $\varepsilon-$ contamination class. In [3,4] the densities are also assumed symmetric. In [5] symmetry is assumed inside an interval around the origin.

All of these approaches assume independent observations. For the dependent case, as we have seen in Chapter III, in order to calculate the efficacy we must know the bivariate densities of all pairs of observations. The nonlinearity that maximizes the efficacy is given by Theorem 3.1.1.

Min-max detection with dependent observations is considered in [8]. Following similar ideas from [9,10] the min-max nonlinearity is derived under the assumption that the observations are generated by a moving average process and are weakly dependent. In [11] the problem of min-max detection of a constant signal in stationary Markov noise is considered. It is shown that, for a special class of Markov noise processes, the min-max nonlinearity is very closely related to the one for the i.i.d. case. Here we consider an extension of this result. The class of sequences that we define here are more general than the Markov sequences defined in [11]. We continue now by defining our class.

Let $\boldsymbol{N} = \{N_i\}_{i=1}^{\infty}$ be a strictly stationary noise sequence. Define as M_a^b the σ-algebra generated by the random variables $\{ N_a, N_{a+1}, \ldots, N_b \}$. Let $f(x)$ be the common marginal density for the random variables N_i. We assume that this density is symmetric, that it has a continuous derivative different from zero a.e. with respect to $f(x)$ and that it has finite Fisher's information. Here we consider a subclass of the acceptable φ-mixing sequences defined in Section 3.1. We say that a sequence N belongs to the class S if it is an acceptable φ-mixing sequence and also satisfies the following conditions concerning the bivariate and univariate densities of two components N_k and N_{k+n}. If A is an event for N_k and B for N_{k+n} then, for every k and n, we have

$$| P(A \cap B) - P(A)\, P(B) | \leq \gamma_n P(A)\, P(B) \tag{5.1.1}$$

with

$$\sum_{n=1}^{\infty} \gamma_n < \infty \tag{5.1.2}$$

and also

$$f(x) = (1-\varepsilon)g(x) + \varepsilon h(x) \tag{5.1.3}$$

Notice that (5.1.1) is different from (3.1.2) since it is defined only for two random variables. Also the right side of (5.1.1) involves the product of the two marginal probabilities rather than one marginal as in (3.1.2). Even though every bivariate density satisfies (3.1.2) for some φ_n (for example $\varphi_n = 1$) such is not the case for (5.1.1). Finally (5.1.3) defines an ε−contamination model for the marginal density $f(x)$. We assume that $g(x)$ is a known, symmetric, strongly unimodal density, with continuous derivative different from zero a.e. with respect to $g(x)$ and with finite Fisher's information. For $h(x)$ we assume that it is a symmetric density; and ε a known constant in $[0,1)$.

The class of allowable nonlinearities depend on the problem we consider. For the detection problem, this class is defined by conditions (3.1.19 - 3.1.23) and for the estimation problem is defined by (3.1.30 - 3.1.33). Here we will assume one more thing about the nonlinearity $\psi(x)$; we assume that

$$\psi(x) = -\psi(-x) \tag{5.1.4}$$

This assumption is reasonable because all of the locally optimum nonlinearities are odd symmetric. As we recall, the efficacy has the form

$$eff(\psi(x), N) = \frac{\left[\int_{-\infty}^{\infty} \psi(x) f'(x) dx\right]^2}{\sigma_0^2(\psi)} \tag{5.1.5}$$

where $\sigma_0^2(\psi)$ is defined as

$$\sigma_0^2(\psi) = E\{\psi^2(N_1)\} + 2\sum_{i=1}^{\infty} \nu_j E\{\psi(N_1)\psi(N_{j+1})\} \tag{5.1.6}$$

Notice that we consider the more general case encountered in the detection problem. We now prove a lemma that gives us a property that characterizes the class S.

Lemma 5.1.1 Let $\psi(x)$ be a measurable function with $E\{[\psi(N_1)]^2\} < \infty$. Let also $\boldsymbol{N} \in S$. Then

$$|E\{\psi(N_1)\psi(N_{j+1})\} - [E\{\psi(N_1)\}]^2| \le \gamma_j [E\{|\psi(N_1)|\}]^2 \tag{5.1.7}$$

Proof. It is enough to show (5.1.7) for simple functions. Thus let

$$\psi(x) = \sum_{i=1}^{K} \psi_i I_{A_i} \tag{5.1.8}$$

Let B_i be the event $\{N_1 \in A_i\}$ and C_i the event $\{N_{j+1} \in A_i\}$. Then, using (5.1.1), we obtain

$$|E\{\psi(N_1)\psi(N_{j+1})\} - [E\{\psi(N_1)\}]^2| \le \sum_{i=1}^{K}\sum_{l=1}^{K} |\psi_i\psi_l| \; |P(B_i \cap C_l) - P(B_i)P(C_l)|$$

$$\le \gamma_j \sum_{i=1}^{K}\sum_{l=1}^{K} |\psi_i\psi_l| P(B_i)P(C_l) = \gamma_j [E\{|\psi(N_1)|\}]^2 \tag{5.1.9}$$

And this concludes the proof. We now define the optimum nonlinearity.

5.2 Min - Max Nonlinearity.

The problem we would like to solve is the following: Find a nonlinearity $\psi_r(x) \in \Psi$ and a sequence $\boldsymbol{N}_r \in S$ such that

$$\sup_{\psi(x) \in \Psi} \; \inf_{\boldsymbol{N} \in S} \; eff(\psi(x), \boldsymbol{N}) = eff(\psi_r(x), \boldsymbol{N}_r) \tag{5.2.1}$$

We now proceed as follows: for a given $\psi(x)$ we find the $\mathbf{N} \in S$ that minimizes the efficacy. Then the resulting expression is maximized over the nonlinearity $\psi(x)$. The minimization is done in two steps. First we keep the marginal density fixed and minimize over all sequences that have the same marginal and then we minimize over the marginal. If the marginal is fixed, we can see from (5.1.5) that, in order to minimize the efficacy, we need to maximize $\sigma_0^2(\psi)$. Using (5.1.7) and remembering that $\psi(x)$ is odd symmetric (zero-mean), we have

$$\sigma_0^2(\psi) \le \int_{-\infty}^{\infty} \psi^2(x) f(x) dx + 2\left(\sum_{j=1}^{\infty} |\nu_j| \gamma_j\right)\left[\int_{-\infty}^{\infty} |\psi(x)| f(x) dx\right]^2 \quad (5.2.2)$$

The series in (5.2.2) is summable because

$$\sum_{j=1}^{\infty} |\nu_j| \gamma_j \le \sum_{j=1}^{\infty} \gamma_j < \infty \quad (5.2.3)$$

We have equality in (5.2.2) when the bivariate densities $f_j(x,y)$ of N_1 and N_{j+1} are given by

$$f_j(x,y) = f(x) f(y) \{1 + \gamma_j sgn(\nu_j) sn_\psi(x) sn_\psi(y)\} \quad (5.2.4)$$

The function $sn_\psi(x)$ is defined to be odd symmetric and for $x > 0$ is equal to the sign of $\psi(x)$ when $\psi(x) \neq 0$ and may take any value in $[-1,1]$ when $\psi(x) = 0$. Also $sgn(\nu_j)$ is equal to the sign of ν_j when $\nu_j \neq 0$. When $\nu_j = 0$ the bivariate density can be anything. The odd symmetry of $sn_\psi(x)$ is important because it makes $f_j(x,y)$ a legitimate bivariate density with marginal $f(x)$. Even though these densities are of the right form it is possible that there is no sequence in S that will have them as bivariate densities. Here we will assume that such a sequence always exists and, in the examples we present, we show a way to construct its multivariate density. We must point out that, if we cannot show the existence of a sequence in S, then this approach does not necessarily lead to the min-max solution. Let us now substitute (5.2.2) into the expression for the efficacy and call the resulting expression eff^*. Thus we obtain

$$eff^*(\psi(x), f(x)) = \frac{\left[\int_{-\infty}^{\infty} \psi(x) f'(x)dx\right]^2}{\int_{-\infty}^{\infty} \psi^2(x) f(x)dx + \gamma\left[\int_{-\infty}^{\infty} |\psi(x)| f(x)dx\right]^2} \quad (5.2.5)$$

where $\gamma = 2\sum_{j=1}^{\infty} |\nu_j| \gamma_j$. The case $\gamma = 0$ is of no importance since it is no different from the i.i.d. case. Thus we assume $\gamma > 0$. Next, we have to find a pair $\psi_r(x)$ and $f_r(x)$ such that

$$\sup_{\psi(x) \in \Psi} \inf_{f(x)} eff^*(\psi(x), f(x)) = eff^*(\psi_r(x), f_r(x)) \quad (5.2.6)$$

It turns out that this new min-max problem defined by (5.2.6) has a saddle point; in other words the pair $\psi_r(x)$ and $f_r(x)$ satisfies the following double inequality

$$eff^*(\psi(x), f_r(x)) \le eff^*(\psi_r(x), f_r(x)) \le eff^*(\psi_r(x), f(x)) \quad (5.2.7)$$

for any $\psi(x) \in \Psi$ and any $f(x)$ satisfying (5.1.3). Any pair that satisfies (5.2.7) is known to satisfy (5.2.6). Thus we will solve (5.2.7) instead of (5.2.6). The left inequality in (5.2.7) indicates that $\psi_r(x)$ is the optimum nonlinearity for $f_r(x)$ when the criterion function is the eff^*. The following theorem gives the form of this optimum nonlinearity in terms of the marginal density.

Theorem 5.2.1. Let $f(x)$ be a symmetric density with finite Fisher's information and continuous derivative different from zero a.e. with respect to $f(x)$. Then the optimum nonlinearity $\psi_0(x)$ that maximizes the eff^* is given by

$$\psi_0(x) = -\frac{f'(x)}{f(x)} - \mu\pi_0(x) \tag{5.2.8}$$

where $\pi_0(x)$ is defined as follows

$$\pi_0(x) = \begin{cases} -\frac{1}{\mu}\frac{f'(x)}{f(x)} & \text{for} \quad -1 \le -\frac{1}{\mu}\frac{f'(x)}{f(x)} \le 1 \\ 1 & \text{for} \quad 1 \le -\frac{1}{\mu}\frac{f'(x)}{f(x)} \\ -1 & \text{for} \quad -1 \ge -\frac{1}{\mu}\frac{f'(x)}{f(x)} \end{cases} \tag{5.2.9}$$

and μ is a positive constant that satisfies the equation

$$S(\mu) = \mu + \frac{\int_{-\infty}^{\infty} f'(x)\pi_0(x)dx}{\frac{1}{\gamma} + \int_{-\infty}^{\infty} \pi_0^2(x)f(x)dx} = 0 \tag{5.2.10}$$

The proof of this theorem is given in Appendix 5.A. From (5.2.8) and (5.2.9) we see that $\psi_0(x)$ is odd symmetric and closely related to the locally optimum nonlinearity. The function $\pi_0(x)$ is defined in such a way that $\psi_0(x)$ becames zero whenever $-\frac{f'(x)}{f(x)}$ takes on values between $-\mu$ and μ. Now we are ready to define the pair that satisfies the saddle point relation (5.2.7). Since $\psi_r(x)$ is optimum for $f_r(x)$ we need to define only $f_r(x)$ and this is done in the following theorem.

Theorem 5.2.2 The density $f_r(x)$ that gives the solution to the saddle point problem defined by (5.2.7) is the following

$$f_r(x) = \begin{cases} (1-\varepsilon)g(x_1)e^{x_1(x+x_1)} & \text{for } x \le -x_1 \\ (1-\varepsilon)g(x) & \text{for } |x| \le x_1 \\ (1-\varepsilon)g(x_1)e^{-x_1(x-x_1)} & \text{for } x \ge x_1 \end{cases} \qquad (5.2.11)$$

where $x_1 \ge 0$ and such that $f_r(x)$ has total mass equal to unity.

Proof. This density is exactly the one defined by Huber in [1,2] for the i.i.d. case. It belongs to the ε−contamination class with a density $h_r(x)$ that places all the mass outside the interval $[-x_1,x_1]$

$$\varepsilon h_r(x) = \begin{cases} (1-\varepsilon)[g(x_1)e^{x_1(x+x_1)} - g(x)] & \text{for } x \le -x_1 \\ 0 & \text{for } |x| \le x_1 \\ (1-\varepsilon)[g(x_1)e^{-x_1(x-x_1)} - g(x)] & \text{for } x \ge x_1 \end{cases} \qquad (5.2.12)$$

The nonnegativity of $h_r(x)$ can be proved using the strong unimodality of $g(x)$ (see [1,2]). To find the nonlinearity $\psi_r(x)$, we use Theorem 5.2.1 and obtain

$$\psi_r(x) = \begin{cases} 0 & \text{for } 0 \le x \le x_2 \\ -\dfrac{g'(x)}{g(x)} + \dfrac{g'(x_2)}{g(x_2)} & \text{for } x_2 \le x \le x_1 \\ -\dfrac{g'(x_1)}{g(x_1)} + \dfrac{g'(x_2)}{g(x_2)} & \text{for } x_1 \le x \end{cases} \qquad (5.2.13)$$

For $x \le 0$ we recall that $\psi_r(x)$ is odd symmetric. We define x_2 as

$$-\frac{g'(x_2)}{g(x_2)} = \mu \qquad (5.2.14)$$

where μ is a solution to the equation defined by (5.2.10). In order for (5.2.13) to be valid the x_2 defined by (5.2.14) must satisfy $0 \le x_2 \le x_1$. In Appendix 5.A we show that such an x_2 always exists. Up to this point, because of Theorem 5.2.1, we have that $\psi_r(x)$ and $f_r(x)$ satisfy the left inequality in (5.2.7). To prove that they also satisfy the right inequality, notice that, since $g(x)$ is strongly unimodal, we have that $\psi_r(x)$ is a nondecreasing function. If we define

$$M = -\frac{g'(x_1)}{g(x_1)} + \frac{g'(x_2)}{g(x_2)} \tag{5.2.15}$$

then, since $\psi_r(x)$ is odd and nondecreasing, we have $|\psi_r(x)| \le M$. Call $n(f)$ and $d(f)$ the numerator and the denominator of the $eff^*(\psi_r(x), f(x))$; then

$$n(f) = \Big[\int_{-\infty}^{\infty} \psi_r(x) f'(x) d(x)\Big]^2 = \Big[(1-\varepsilon)\int_{-\infty}^{\infty} \psi_r(x) g'(x) dx + \varepsilon \int_{-\infty}^{\infty} \psi_r(x) h'(x) dx\Big]^2 \tag{5.2.16}$$

Because $\psi_r(x)$ is nondecreasing, the two terms in the last expression are nonpositive; thus

$$n(f) \ge \Big[(1-\varepsilon)\int_{-\infty}^{\infty} \psi_r(x) g'(x) dx\Big]^2 = n(f_r) \tag{5.2.17}$$

also

$$d(f) = \int_{-\infty}^{\infty} \psi_r^2(x) f(x) dx + \gamma\Big[\int_{-\infty}^{\infty} |\psi_r(x)| f(x) dx\Big]^2$$

$$\le (1-\varepsilon)\int_{-\infty}^{\infty} \psi_r^2(x) g(x) dx + \varepsilon M^2 + \gamma\Big[(1-\varepsilon)\int_{-\infty}^{\infty} |\psi_r(x)| g(x) dx + \varepsilon M\Big]^2$$

$$= d(f_\tau) \tag{5.2.18}$$

Thus $f_\tau(x)$ simultaneously minimizes the numerator and maximizes the denominator of the $eff^*(\psi_\tau(x), f(x))$ which means that the right inequality in (5.2.7) is also satisfied. This concludes the proof.

Returning now to our original min-max problem defined in (5.2.1), we have that $\psi_\tau(x)$ is the nonlinearity defined in (5.2.13) and $\mathbf{N}_\tau$ is any sequence from S that has bivariate densities given by

$$f_j^\tau(x,y) = f_\tau(x)f_\tau(y)\{1 + \gamma_j sgn(\nu_j) sn_{\psi_\tau}(x) sn_{\psi_\tau}(y)\} \tag{5.2.19}$$

where $f_\tau(x)$ is defined in (5.2.11).

Up to this point we have solved the min-max problem concerning the efficacy. This is good enough for the estimation problem. For the detection problem though, more has to be done. We must somehow define the threshold τ of (3.1.15). What we would like to do is to define τ in such a way that we will have

$$\sup_{\mathbf{N} \in S} P_{FA}(\psi_\tau, \mathbf{N}) \leq \alpha \tag{5.2.20}$$

where P_{FA} denotes the asymptotic false alarm probability. In order now to satisfy (5.2.20), we must set the threshold for the detection structure

$$T_n^\tau(\mathbf{X}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(X_i) \tag{5.2.21}$$

Notice that $T_n^\tau(\mathbf{X})$ under H_0 is Gaussian in the limit as we proved in Appendix 3.A. Hence if (5.2.20) is satisfied for the sequence that has the maximum asymptotic variance, it will be satisfied for any sequence. But the asymptotic variance of (5.2.21) is the square root of the denominator of the $eff\,(\psi_r(x),N)$. This denominator is maximized when $N = N_r$ and the maximum value is equal to $d(f_r)$ defined in (5.2.18). Thus the threshold τ is given by

$$1 - \Phi\left(\frac{\tau}{[d(f_r)]^{1/2}} \right) = \alpha \qquad (5.2.22)$$

where $\Phi(x)$ is the $N(0,1)$ Gaussian cumulative distribution. Note that the sequence N_r achieves simultaneously the worst performance for the efficacy and for the false alarm probability.

Comment. For the case where we cannot prove existence of a sequence in S that has bivariate densities defined by (5.2.4), we still satisfy a min-max relation. Only instead of the class S of sequences, we will consider the class of bivariate densities that satisfy (5.1.1 - 5.1.3). In other words, the min-max problem will be defined over a larger class of bivariate densities than the one that S defines. Thus the lower performance bound of the min-max problem will not be the best for the class S .

5.2.1 Examples.

As we can see from Theorem 5.2.2 the min-max nonlinearity and its worst performance depend on the density $g(x)$ and the constants ε and γ and not on the actual sequences $\{\gamma_i\}$ and $\{s_i\}$. In the following, we give tables for the point x_2 and the performance of the min-max nonlinearity, for $g(x)$ the Gaussian $N(0,1)$. In Table 5.2.1 are given the values of x_2 for different γ and ε . The parameter x_1 depends only on ε and it turns out that as, $\gamma \to \infty$, then $x_2 \to x_1$ .Thus the last column

($\gamma = \infty$ in the table) gives also the values for x_1 .

Table 5.2.1 Values for x_2 (last column values for x_1).

ε \ γ	1.0	2.0	3.0	4.0	5.0	10.0	20.0	∞
0.001	0.436	0.635	0.764	0.860	0.935	1.173	1.412	2.630
0.01	0.431	0.626	0.753	0.845	0.917	1.140	1.353	1.945
0.05	0.410	0.519	0.704	0.784	0.845	1.019	1.161	1.399
0.1	0.386	0.549	0.648	0.716	0.766	0.902	1.000	1.140
0.15	0.361	0.510	0.597	0.655	0.697	0.807	0.883	0.980
0.2	0.337	0.472	0.549	0.600	0.636	0.730	0.788	0.862
0.3	0.291	0.402	0.436	0.502	0.530	0.595	0.637	0.685
0.4	0.247	0.337	0.385	0.416	0.436	0.485	0.515	0.550
0.5	0.204	0.276	0.314	0.337	0.353	0.390	0.412	0.436
0.8	0.080	0.107	0.121	0.130	0.135	0.147	0.154	0.162

Table 5.2.2 gives the values of the ARE of $\psi_r(x)$ versus the locally optimum nonlinearity $-\frac{f'_r(x)}{f_r(x)}$ when the underlying sequence is the N_r . Notice that this locally optimum nonlinearity would have been the one to use if we had falsely assumed that the observations were i.i.d.

Table 5.2.2 ARE of $\psi_r(x)$ versus the locally optimum nonlinearity.

ε \ γ	1.0	2.0	3.0	4.0	5.0	10.0	20.0	∞
0.001	1.08	1.19	1.29	1.37	1.45	1.74	2.11	4.48
0.01	1.08	1.18	1.26	1.34	1.40	1.63	1.89	2.68
0.05	1.07	1.14	1.21	1.26	1.30	1.46	1.56	1.77
0.1	1.06	1.12	1.17	1.20	1.23	1.32	1.39	1.49
0.15	1.05	1.10	1.13	1.16	1.18	1.24	1.29	1.35
0.2	1.04	1.08	1.11	1.13	1.15	1.19	1.22	1.27
0.3	1.03	1.06	1.08	1.09	1.10	1.12	1.14	1.16
0.4	1.02	1.04	1.05	1.06	1.07	1.08	1.09	1.10

Now we present two cases where the theory in Section 5.2 can be applied.

M−dependent case. Assume $\gamma_j = 0$ for $j > M$. Here $\gamma = 2\gamma_1|\nu_1|$

and the bivariate densities defined in (5.2.4) take the form

$$f_1(x,y) = f(x)f(y)\{1 + \gamma_1 sgn(\nu_1) sn_\psi(x) sn_\psi(y)\} \qquad j = 1,\ldots,M$$

$$f_j(x,y) = f(x)f(y) \qquad j > M \qquad (5.2.23)$$

In Subsection 2.3.2 we have seen a way of generating strictly stationary M-dependent sequences that have bivariate densities of this form.

Markov case. Assume $\gamma_j = m^j$ with $0 \le m < 1$ and $\nu_j \ge 0$. This is the case treated in [11]. Here $\gamma = 2\sum_{j=1}^{\infty} m^j \nu_j$ and

$$f_j(x,y) = f(x)f(y)\{1 + m^j sn_\psi(x) sn_\psi(y)\} \qquad (5.2.24)$$

If the function $sn_\psi(x)$ takes values of only +1 or -1 (always possible), then the bivariate density of consecutive points becomes

$$f_1(x,y) = f(x)f(y)\{1 + m sn_\psi(x) sn_\psi(y)\} \qquad (5.2.25)$$

As we have seen in Subsection 2.3.2, a density of this form defines a Markov sequence that has bivariate densities given by (5.2.24). And this is done by defining $\Phi(x) = m\ sn_\psi(x)$ and $\Theta(y) = sn_\psi(y)$. Since $sn_\psi(x)$ can take values only ± 1, it is easy to see that the α defined in (2.3.28) is equal to m. Also in Subsection 3.2.1 we have seen that these sequences are symmetrically $A\varphi$-mixing sequences.

REFERENCES .

[1] P.J. Huber, " Robust estimation of a location parameter " *Ann. Math. Stat.*, vol 35, pp 73-101, 1964.

[2] ---, " A robust version of the probability ratio test," *Ann. Math. Stat.*, vol 36, pp 1753-1758, 1965.

[3] R.D. Martin and S.C. Schwartz, " Robust detection of a known signal in nearly Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-17, pp 50-56, Jan. 1971.

[4] S.A. Kassam and J.B. Thomas, " Asymptotically robust detection of a known signal in contaminated non-Gaussian noise," *IEEE Trans. Inf. Theory*, vol. IT-22, pp 22-26, Jan. 1976.

[5] S.A. Kassam, G.V. Moustakides and J.G. Shin, " Robust detection of known signals in asymmetric noise," *IEEE Trans. Inf. Theory*, vol. IT-28, pp 84-91, Jan. 1982.

[6] H.V. Poor and J.B. Thomas, " Memoryless discrete time detection of a constant signal in m−dependent noise," *IEEE Trans. Inf. Theory*, vol. IT-25, pp 54-61, Jan. 1979.

[7] D.R. Halverson and G.L. Wise, " Discrete time detection in φ−mixing noise," *IEEE Trans. Inf. Theory*, vol. IT-26, pp 189-198, March 1980.

[8] H.V. Poor, " Signal detection in the presence of weakly dependent noise - Part II: Robust detection," *IEEE Trans. Inf. Theory*, vol. IT-28, pp 744-752, Sept. 1982.

[9] S.L. Portnoy, " Robust estimation in dependent situations," *Ann. Statist.*, vol. 5, pp 22-43, 1977.

[10] ---," Further remarks on robust estimation in dependent situations," *Ann. Statist.*, vol.7, pp 224-231, 1979.

[11] G.V. Moustakides and J.B. Thomas, " Min-max detection of a weak signal in stationary Markov noise," *Proc. Twent. Ann. Allerton Conf.*, pp 713-722, Oct. 1982.

[12] N.N. Vorob'ev, "Consistent families of measures and their extensions," *Th. Prob. Appl.* Vol. VII no 2, pp 147-163, 1962.

[13] J. Billingsley, Convergence of probability measures. New York : Wiley, 1968.

APPENDIX 5.A

Proof of Theorem 5.2.1. Notice that in (5.2.5) the value of the eff^* does not change if we multiply the nonlinearity by a constant. Thus we maximize the numerator assuming that the denominator has some fixed value. Using (3.1.20), this is equivalent to maximizing the following expression

$$H(\psi) = -\int_{-\infty}^{\infty}\psi(x)f'(x)dx - \rho\left\{\int_{-\infty}^{\infty}\psi^2(x)f(x)dx + \gamma\Big[\int_{-\infty}^{\infty}|\psi(x)|f(x)dx\Big]^2\right\} \tag{5.A.1}$$

where ρ is a Lagrange multiplier. We will show that (5.A.1) is maximized by

$$\psi_0(x) = \frac{1}{2\rho}\Big[-\frac{f'(x)}{f(x)} - \mu\pi_0(x)\Big] \tag{5.A.2}$$

where μ and $\pi_0(x)$ were defined in (5.2.9) and (5.2.10). Let $\psi_1(x)$ be some other nonlinearity from the class Ψ. Define the following variation

$$J(\xi) = -\int_{-\infty}^{\infty}\Big[(1-\xi)\psi_0(x) + \xi\psi_1(x)\Big]f'(x)dx - \rho\left\{\int_{-\infty}^{\infty}\Big[(1-\xi)\psi_0(x) + \xi\psi_1(x)\Big]^2 f(x)dx\right.$$

$$\left. + \gamma\Big[\int_{-\infty}^{\infty}\Big\{(1-\xi)|\psi_0(x)| + \xi|\psi_1(x)|\Big\}f(x)dx\Big]^2\right\} \tag{5.A.3}$$

where $\xi \in [0,1]$. Notice that $J(0) = H(\psi_0)$ and $J(1) = H(\psi_1)$. By manipulating (5.A.3) we can write it as

$$J(\xi) - J(0) = I_1 + I_2 + I_3 \tag{5.A.4}$$

where

$$I_1 = \xi \int_{-\infty}^{\infty} \Big\{ -f'(x) - 2\rho\psi_0(x)f(x)$$

$$- 2\gamma\rho\Big[\int_{-\infty}^{\infty} |\psi_0(z)| f(z)dz\Big]\pi_0(x)f(x)\Big\}\Big[\psi_1(x)-\psi_0(x)\Big]dx \qquad (5.A.5)$$

$$I_2 = -2\xi\gamma\rho\Big[\int_{-\infty}^{\infty} |\psi_0(z)| f(z)dz\Big]\Bigg(\int_{-\infty}^{\infty}\Big\{ |\psi_1(x)| - |\psi_0(x)| \Big\} f(x)dx$$

$$- \int_{-\infty}^{\infty} \pi_0(x)\Big[\psi_1(x) - \psi_0(x)\Big]f(x)dx\Bigg) \qquad (5.A.6)$$

$$I_3 = -\xi^2\rho\Bigg(\int_{-\infty}^{\infty}\Big[\psi_1(x) - \psi_0(x)\Big]^2 f(x)dx$$

$$+ \gamma\Big[\int_{-\infty}^{\infty}\Big\{ |\psi_1(x)| - |\psi_0(x)| \Big\} f(x)dx\Big]^2\Bigg) \qquad (5.A.7)$$

To prove that $J(0)$ is the maximum it is enough to show that $J(\xi) - J(0) \le 0$ or that $I_i \le 0$ for i=1,2,3. From the definition of $\pi_0(x)$ in (5.2.9) notice the following:

$$|\psi_0(x)| = \psi_0(x)\pi_0(x) \qquad (5.A.8)$$

$$|\pi_0(x)| \le 1 \qquad (5.A.9)$$

If we multiply (5.A.2) by $\pi_0(x)f(x)$ and integrate and also use (5.2.10), we obtain

$$\int_{-\infty}^{\infty} |\psi_0(x)| f(x)dx = \frac{1}{2\rho}\Big[-\int_{-\infty}^{\infty} f'(x)\pi_0(x)dx - \mu\int_{-\infty}^{\infty}\pi_0^2(x)dx\Big] = \frac{\mu}{2\rho\gamma}$$

(5.A.10)

Substituting (5.A.10) in the expression for I_1 and using (5.A.2), we get zero. On using (5.A.8), the term I_2 becomes

$$I_2 = -2\xi\rho\gamma\Big[\int_{-\infty}^{\infty}|\psi_0(z)|f(z)dz\Big]\left[\int_{-\infty}^{\infty}\{|\psi_1(x)| - \pi_0(x)\psi_1(x)\}f(x)dx\right] \tag{5.A.11}$$

Because of (5.A.9) we have $|\psi_1(x)| \geq \pi_0(x)\psi_1(x)$ and thus for, $\rho > 0$, the I_2 becomes nonpositive. Finally for, $\rho > 0$, the I_3 is clearly nonpositive too. If we define $\rho = \frac{1}{2}$ then (5.A.2) becomes the same as (5.2.8).

In order to complete the proof of Theorem 5.2.2 we must show that the equation defined in (5.2.10) has always a solution. Using continuity arguments it is enough to show existence of two points μ_1 and μ_2 such that $S(\mu_1)S(\mu_2) \leq 0$. Notice that, as $\mu \to 0$, then $-\frac{1}{\mu}\frac{f'(x)}{f(x)} \to \pm\infty$ except on sets of f-measure zero. Thus $\pi_0(x) \to sgn\big(-\frac{f'(x)}{f(x)}\big) = -sgn\big(f'(x)\big)$. Substituting in (5.2.10), we find

$$S(0) = -\frac{\int_{-\infty}^{\infty}|f'(x)|dx}{\frac{1}{\gamma}+1} < 0 \tag{5.A.12}$$

Now using (5.A.9) and the Schwarz inequality, we have

$$\Big|\int_{-\infty}^{\infty}f'(x)\pi_0(x)dx\Big| \leq \int_{-\infty}^{\infty}|f'(x)|dx \leq \Big[\int_{-\infty}^{\infty}\Big(\frac{f'(x)}{f(x)}\Big)^2 f(x)dx\Big]^{1/2} = \big[I(f)\big]^{1/2} < \infty \tag{5.A.13}$$

where $I(f)$ is Fisher's Information. Thus the second term in (5.2.10) is bounded by $\gamma[I(f)]^{1/2}$ and as $\mu \to +\infty$ we have that $S(\mu) \to +\infty$ or it becomes positive. And this concludes the proof of Theorem 5.2.1

Existence of x_2. In Theorem 5.2.2 we assumed that there exists an x_2, with $0 \le x_2 \le x_1$, that satisfies (5.2.14), where μ satisfies (5.2.10). Now, because $S(0) < 0$, if we show that $S\left(-\frac{g'(x_1)}{g(x_1)}\right) \ge 0$, then there exists a solution to (5.2.10) which will satisfy $0 < \mu \le -\frac{g'(x_1)}{g(x_1)}$ and, because of the monotonicity of $-\frac{g'(x)}{g(x)}$, we will have $0 \le x_2 \le x_1$. To prove this, notice that the locally optimum nonlinearity for the $f_r(x)$ defined in (5.2.11) is

$$-\frac{f_r'(x)}{f_r(x)} = \begin{cases} -\dfrac{g'(x)}{g(x)} & \text{for } |x| \le x_1 \\ -\dfrac{g'(x_1)}{g(x_1)} sgn(x) & \text{for } |x| \ge x_1 \end{cases} \tag{5.A.14}$$

Thus, for $\mu = -\frac{g'(x_1)}{g(x_1)}$, we are always in the first case of Equation (5.2.9) and we have

$$\pi_0(x) = -\frac{1}{\mu}\frac{f'_r(x)}{f_r(x)} \tag{5.A.15}$$

Substituting into (5.2.10) we obtain

$$S\left(-\frac{g'(x_1)}{g(x_1)}\right) = -\frac{g'(x_1)}{g(x_1)}\left[1 - \frac{I(f_r)}{\frac{1}{\gamma}\left[\frac{g'(x_1)}{g(x_1)}\right]^2 + I(f_r)}\right] \ge 0 \tag{5.A.16}$$

And this proves the existence of x_2.

CHAPTER VI.

CONCLUSIONS

This work has concentrated on finding optimum detection and estimation schemes under the assumption of dependency of the observation sequence. Interesting results are obtained that can be used when the knowledge about dependency is only partial; for example, when it is limited to a knowledge of the correlation coefficients.

The principal approach that was used was a min-max approach. Even though min-max approaches are usually conservative, it turned out that here they helped us to overcome the problem of lack of knowledge of all second-order statistics. This is very important from a practical point of view since there exist methods for estimating first-order statistics and correlation coefficients, but very little can be done for second-order properties.

For topics for further study we can mention the following: In Chapter IV, as we have seen, the optimization was done over a larger class of bivariate densities than the one that one-dependent sequences define. The main reason was the lack of more information about the bivariate densities of such sequences. It might be possible to prove that, for one-dependent sequences, the coefficients of the expansion in (2.2.3) satisfy

$$\alpha_n = \int_{-\frac{1}{2}}^{\frac{1}{2}} z^n h(z) dz \qquad (6.1.1)$$

with $h(z)$ a density possibly concentrated on $[-\frac{1}{2}, \frac{1}{2}]$. If this is true, then the least favorable bivariate density will have a much more reasonable form compared to the degenerate form it has now. It is easy to see though that Proposition 4.2.1 will still be valid with a little modification. Something more interesting would be to solve the same problem for Markov sequences. We can see that for the case of optimizing nonlinearities in ψ_2, for Markov sequences, we get the same solution as the one-dependent case. Things become considerably more difficult when the order of the polynomials is higher. Finally, it would be interesting to find the min-max solution to this problem, when we make absolutely no assumption about the form of the bivariate density. In other words, we have the marginal and the correlation coefficient and nothing else.

All the approaches we have taken were characterized by an *a priori* setting of the class of the second-order statistics. It seems that an approach which takes into account some dynamical model for generating the sequences will be more interesting. For example, we can assume that the noise sequence is generated by the following stochastic difference equation

$$N_n = g(N_{n-1}) + W_n \tag{6.1.2}$$

where $g(x)$ is some nonlinearity and $\{W_n\}$ is a white Gaussian noise sequence. We could use the efficacy as measure again and try to define the optimum nonlinearity that will maximize it. Probably (6.1.2) will lead to some form of difference equation for the determination of the bivariate densities, something that corresponds to the Fokker-Planck equations for the continuous time case. Using the eigenfunctions of a related Sturm-Liuville problem, we can possibly have a formal representation of the bivariate densities involved. Combining all this information, it might turn out that the optimum nonlinearity has some tractable connection with the eigenfunctions of the Sturm-Liuville problem. This approach is definitely more interesting than the one we used here because there is no need to define *a priori* a class of bivariate densities. Also, all of the bivariate densities involved will be used only in a formal way, hoping that at the end there will be no need to find them explicitly.