THRESHOLD ORDER OF A SWITCHING FUNCTION AND STOCHASTIC DEPENDENCE OF INPUTS

T. KRISHNAN

Dissertation submitted to the Indian Statistical Institute in partial fulfilment of the requirements of the degree of Doctor of Philosophy (Ph.D)

Research and Training School
INDIAN STATISTICAL INSTITUTE
Calcutta-35
INDIA

December 1966

The soul's dark cottage, battered and decayed,
Lets in new light through chinks that Time has made.
Stronger by weakness, wiser men become
As they drawn near to their eternal home:
Leaving the old, both worlds at once they view,
That stand upon the threshold of the new.

Edmund Waller
'Of the Last Verses in the Book'

C O N T E N T S

P R E F A C E

This dissertation is being submitted to the Indian Statistical Institute, in support of the author's application for the degree of Doctor of Philosophy. This dissertation embodies the work carried out by the author at the Indian Statistical Institute under the supervision of Professor C. R. Rao. He is grateful to Professor Rao for his constant encouragement and inspiration and for the facilities provided at the Institute.

The main part of the results of Chapter 5 have already been published [20]. The results of Chapters 2 and 3 and of Chapter 4 are being reviewed for publication. Referees of these papers have helped me in making a number of improvements, for which I am grateful to them.

Dr. Robert O. Winder of RCA Laboratories was kind enough to go through a previous version of this work, to make helpful comments and suggestions, and to point out mistakes in Tables 4.3 and 4.4. The author is indebted to Dr. Winder for this help. I am indebted to Dr. Laveen Kanal of Philco Corporation for his comments on the use of Bahadur's expansion (vide p. 114).

Professor C. L. Sheng of the University of Ottawa read the original version of this dissertation and gave comments. I thank him for that. Thanks are due to Dr. Michael L. Dertouzos for his comments and suggestions on this work.

Shri Arijit Roychowdhury and Miss Mary Kuriyan have prepared tables of 3-input switching functions, classified by the 14 symmetry types and have computed the Coleman and Chow parameters of each type. In Tables 4.3 and 4.4, these tables have been used. The author is indebted to them for permission to use their tables.

The author is indebted to Dr. E.V.Krishnamurthy for going through the manuscript and for a number of discussions, comments and suggestions on the content as well as on the presentation. He wishes to thank Shri T. Krishnamurthi, Shri K. Viswanath, Shri U.S.R. Murty and Shri T. J. Rao for comments and suggestions and for help during the final stages of the work. To Shri Gour Mohon Das, he owes thanks for producing the typescript.

T. KRISHNAN

CHAPTER 1

INTRODUCTION :

THRESHOLD ORDER OF A SWITCHING FUNCTION

'There is nothing more difficult to take in hand, more perilous to conduct, or more uncertain in its success, than to take the lead in the introduction of a new order of things.

Nicolò Machiavelli
Translation: W.K. Marriott

1.0 Summary

Linear-input elements or threshold elements are briefly discussed and their limitations are pointed out. The concept of Threshold order of a Switching Function is introduced, and some networks that realize a switching function of threshold order r are proposed. The need for a non-linear-input logic is pointed out, quoting several authors. A summary of the results in the subsequent Chapters is presented.

1.1 On Linear-Input Logic

Recent developments in the hardware of switching circuits have produced considerable interest in a new kind of switching element, called a threshold element, or a linear-input element. Before these developments, the AND OR gates or the NAND or NOR gates were used as the basic logical building blocks by the computer designer, since they offered cheaper and easier means of construction. Threshold logic elements seem to offer a more convenient set of building blocks, since a given amount of logical circuitry could be realised with much fewer threshold gates than AND/OR, or NAND or NOR gates and hence would be more economical, reliable and would permit more easy maintenance. Further, the threshold gate has been found to have wide applications in the studies on artificial intelligence such as the construction of communication and decision networks, pattern classifiers, learning machines, probability transformers and decoders.

Another reason for interest in threshold logic is that it offers a satisfactory method to construct a mathematical model of neuron, the decision element in the human central nervous system.

Threshold logic has been studied under very many different names such as linearly separable logic, majority decision logic, voting logic and linear input logic.

A threshold logic element consists of a set of n binary inputs $x_1, x_2, \ldots, x_n$ on which weights $a_1, a_2, \ldots, a_n$ (real numbers) are applied; these are connected to an algebraic adder whose output goes to a discriminator with a threshold T $(= - a_o)$; the output of the discriminator is true if its input is $\geq a_o$ and is false if its input is $< a_o$. This is explained in Fig. 1.1. In other words, a threshold logic

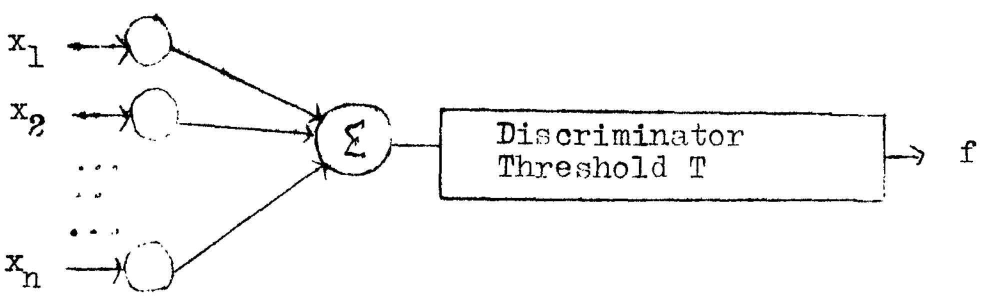

Fig. 1.1

Threshold Logic Gate

gate computes a (binary) threshold function f of n (binary) inputs $x_1, x_2, \ldots, x_n$, where a threshold function is a binary function for which real numbers $a_0, a_1, a_2, \ldots, a_n$

exist such that

$$f(x_1, x_2, \ldots, x_n) \text{ is true if } a_o + \sum_{i=1}^{n} a_i x_i \geq 0$$
$$f(x_1, x_2, \ldots, x_n) \text{ is false if } a_o + \sum_{i=1}^{n} a_i x_i < 0. \qquad (1.1)$$

The a_i $(i \neq 0)$ are called <u>weights</u>, and $-a_o$ is the <u>threshold</u>. Geometrically, a threshold function is a switching function whose true vertices are separated from its false vertices by an (n-1) dimensional hyperplane.

Extensive work has been done and is being done on the theory and applications of threshold logic. Winder [39], [40] briefly reviews the work on threshold logic.

It is interesting to note that the threshold functions form only a very small proportion of the set of switching functions, and as n increases this proportion decreases, as shown in the following table (first three columns from Winder [41]).

Table 1.1

Number of Threshold Functions

No. of inputs	No. of Switching Functions 2^{2^n}	No. of Threshold Functions	Proportion
0	2	2	1
1	4	4	1
2	16	14	0.875
3	256	104	0.40625
4	65536	1882	0.0287
5	$\sim 4.3 \times 10^{19}$	94572	0.21993×10^{-4}
6	$\sim 1.8 \times 10^{19}$	15028134	0.834900×10^{-12}
7	$\sim 3.4 \times 10^{38}$	8378070864	$0.246413800 \times 10^{-28}$

Hence, apart from the complex problems of testing and realization of switching functions by a single threshold element whenever possible, a major problem is the realization of a given arbitrary switching function as a network of threshold gates in an economical manner. Although the problems of the first kind have been solved to a great extent, those of the second kind are still largely unsolved.

Threshold functions, being realizable as a linear combination of inputs, can be justifiably called <u>linear-input functions</u>. Hence non-threshold functions can otherwise be called <u>non-linear-input functions</u>.

1.2 Probability Distributions of Binary Variables

One of the main mathematical tools for the study of switching functions, particularly the threshold functions, is the representation of such functions in various parametric forms. Several such representation theories are available in the literature. In linear-input logic and more so in the non-linear-input logic, such representations -not only of switching functions but also of probability distributions of the input vectors - play a major role. The probability distributions, in fact, play a dual role:

(1) In problems of a stochastic nature, for instance, in information and coding problems, they enter quite naturally and inevitably.

(2) Probabilistic methods lead to deterministic results, for instance, in the characterization of threshold functions and in the representation of functions of binary variables.

Hence, in Chapter 2, we gather some results on the representation of real-valued functions of binary variables, taking particular interest in switching functions and probability distributions. We also prove some results regarding such representations which will be of use in subsequent Chapters.

The literature on threshold logic abounds in results which show, from a number of viewpoints, the strong connections between threshold functions and independently distributed inputs. The main part of this report shows that the connections between switching functions and probability distributions are much deeper in the sense that a switching function of threshold order r corresponds to a probability distribution of order r (See Definition 5.2), from these several viewpoints. Minsky and Selfridge [28] have noted the severity of the independence assumption and point out the need for consideration of higher-order joint probabilities. However, they do not seem to formulate it specifically, nor do they consider the minimal choice of joint probabilities (such as we have done in terms of the order of the distribution) and hence have not noted the connection between threshold order and order of probability distribution.

1.3 Need for a Non-Linear-Input Logic

Several research workers on threshold logic have realized the limitations of threshold logic in some situations, and have felt the need for a more general fixed logic like the one we present here. We quote below relevant portions from some of these authors:

Chow [8] who presented 'an equivalence between threshold functions and statistical recognition with independent distribution' writes '... recognition with statistical independence represents merely a special case of statistical recognition, in the same sense that threshhold functions constitute a very small subset of the set of all Boolean functions.' He further suggests that 'in considering recognition schemes, we should not be confined to cases of statistical independence only.'

Hawkins in his review of Self-Organizing Systems [13]: Minsky and Selfridge discuss the assumptions of independence and state that, in its absence, the only alternative appears to be the calculation of higher-order joint probabilities.

The last remark serves to point up what is probably the central problem in statistical recognition and learning network synthesis. This is the selection and analysis of suitable nonlinear function of input variables. It is apparent that the simple linear expansion ...... will frequently be inadequate to approximate an arbitrary output function. However, general criteria do not exist for the selection of more complex functions of inputs which ideally should be as few in number as possible, readily mechanized, and capable of modification in such a way that over-all network learning will occur.'

Mays [24, p.3]: 'As a result of studying threshold logic several geometric and algebraic concepts were generated to help explain the limitations of threshold logic... the concepts do shed light on the limitations of threshold logic and they are included in this report with the hope that some other researcher will find them useful ... One of the concepts is that the input variables to a threshold-logic device span only (n+1) dimensions, whereas all possible switching functions span 2^n dimensions. It is suggested that one way to realize more functions would be to use fixed logic to generate inputs that would span more than (n+1) dimensions.'

While discussing the use of a threshold function for a recognition procedure with independence assumptions, Winder [40, pp. 5, 6] writes: 'Saul Amarel (RCA Laboratories) has suggested to me that the independence assumption is a severe one, and that only a small proportion of threshold functions might actually be realized by the above procedure .. .. Thus we have shown that for any interesting threshold functions, with equi-probable input combinations, the independence assumptions in fact don't hold. The theorem can probably be extended to more general input distributions -possibly to arbitrary distributions.'

1.4. Definition of Threshold Order

We present, in this report, another approach to the problem of realization of non-linear input functions. This approach consists in the introduction of fixed logic gates which are similar to threshold gates but are more general. The main features of such a gate are:

(1) its properties are quite similar to those of a threshold gate,

(2) it covers the complete set of switching functions, including the threshold functions,

(3) it is a natural generalization of the concept of a threshold function,

(4) a measure of the complexity of a switching function in terms of its non-linearity of inputs is available in what is called its threshold order; this helps to realize the function in an optimal network.

We define a threshold function of order r as follows:

Definition 1.1: A switching function $f(x_1, x_2, \ldots, x_n)$ of n binary variables $x_1, x_2, \ldots, x_n$ is said to be a threshold function of order r $(0 \leq r \leq n)$, if there exists a set of

$$t_r = 1 + n + \binom{n}{2} + \ldots + \binom{n}{r} \qquad (1.2)$$

real numbers

$$w_0 \, ; \quad w_{i_1 i_2 \cdots i_j} \, , \qquad i_1 < i_2 < \cdots < i_j$$
$$j = 1, 2, \ldots, r,$$

such that

$$w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i_1 < i_2 = 1}^{n} w_{i_1 i_2} \; x_{i_1} x_{i_2} + \cdots +$$

$$+ \sum_{i_1 < i_2 < \cdots < i_r = 1}^{n} w_{i_1 i_2 \cdots i_r} \; x_{i_1} \, x_{i_2} \cdots \, x_{i_r} \geq 0$$

if f is true, (1.3)

" " " " < 0

if f is false.

<u>Definition 1.2:</u> The smallest integer r satisfying (1.3) is called the <u>threshold order</u> of the switching function f.

<u>Remark 1.1:</u> A threshold function, according to our terminology, is a threshold function of order 1.

<u>Remark 1.2:</u> Since the definition requires $r \leq n$, the question arises whether every switching function has a threshold order. In the sequel, we give an affirmative

answer. In what follows, we use interchangeably the terms 'switching function of threshold order r' and 'threshold function of order r.'

Even though any two arbitrary values can be used for inputs as well as outputs, we would find it convenient to use two sets of values 0, 1; -1, +1. Throughout this report, we use the notation x_i when the values are 0, 1 and y_i when the values are ± 1. We denote the output by f when it is 0 or 1 and F when it is ± 1. Hence

$$\left. \begin{aligned} y_i &= 2x_i - 1, \\ F &= 2f - 1 . \end{aligned} \right\} \qquad (1.4)$$

Example 1.1: Consider a 3-variable switching function $f(x)$ ($= f(x_1, x_2, x_3)$), given by the normal form

$$x_1 x_2 + \bar{x}_2 x_3 \qquad (1.5)$$

($\bar{x}_i$ is complement of x_i). It is well known [26] that a function which cannot be expressed in a normal form, without involving both x_i or $\bar{x}_i$ for any i, cannot be a threshold function. Here f is not a threshold function. However, f is a threshold function of order 2, since it satisfies conditions (1.3) if we choose

$$a_0 = -2;\ a_1 = 0,\ a_2 = 0,\ a_3 = 4;\ a_{12} = 4,\ a_{23} = -4,\ a_{31} = 0. \quad (1.6)$$

This is shown in Table 1.2 below, where

$$T(x) = T(x_1, x_2, x_3) = a_0 + a_1 x_1 + a_2 x_2 + a_3 x_3 + a_{12} x_1 x_2 + a_{23} x_2 x_3 + a_{31} x_3 x_1. \quad (1.7)$$

Table 1.2

Example of a Second-order Threshold Function

x_1	x_2	x_3	x_1x_2	x_2x_3	x_3x_1	f(x)	T(x)
0	0	0	0	0	0	0	-2
0	0	1	0	0	0	1	2
0	1	0	0	0	0	0	-2
0	1	1	0	1	0	0	-2
1	0	0	0	0	0	0	-1
1	0	1	0	0	1	1	3
1	1	0	1	0	0	1	3
1	1	1	1	1	1	1	3

The idea of using product terms in a functional expansion to produce non-linear-input functions is natural and indeed has occurred to some previous workers. In his celebrated review paper, Hawkins [13] discusses this problem and presents a brief review. Huffman [16] and von Heerden [14] use such an expansion in their work on coding problems. However, none of these authors has investigated the minimum such set of product functions and hence has not reduced the complications of the representations. Further, they are concerned with essentially Boolean operations, whereas our treatment is in terms of real algebraic operations, which are particularly suited to problems of realization. The main contribution of the present work is the reduction of such complexity by introducing the threshold order, which has minimal properties in terms of such product functions. This answers the point raised by Hawkins [13] 'general criteria do not exist for the selection of more complex function of inputs ...'

Kaszerman [18] has suggested a non-linear summation threshold device, somewhat similar to the one presented here. However, he has not formulated any theory but has just given a method of generating a non-linear surface that would separate the true and false vertices of a switching function. As we shall show in Chapter 4, his method leads to very inoptimal (as he has himself admitted) implementation, from the point of view of the number of terms involved in the surface and of the weighting factors for inputs. This is owing to his failure to take into account the threshold order of the function, which, as we show, is the most natural starting point of a non-linear summation threshold device. Mattson [23] also discusses the application of product function in a non-linear expansion.

A decision element to realize a threshold function of order r is illustrated in Figs. 1.2 and 1.3, in which a threshold function of second order, in three inputs, is considered. As a network, the memory is distributed in the element of Fig. 1.2 and localized in Fig. 1.3. These are based on the networks proposed by Gose [12], for the computation of real-valued functions of binary inputs. Fig.1.2 is similar to the model of Kaszerman [18] for the non-linear summation threshold device. A threshold element of order r of n inputs can be constructed in an analogous manner.

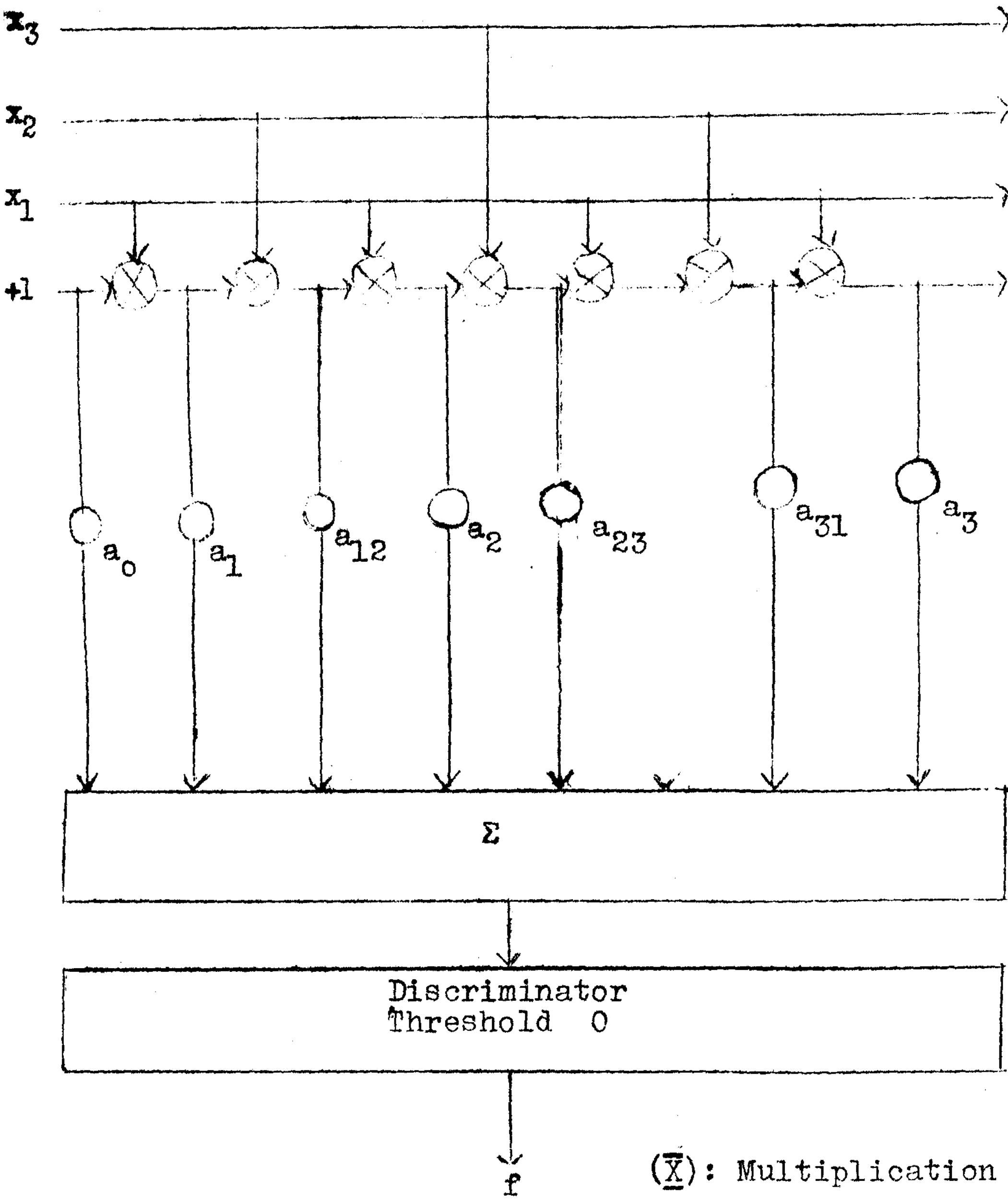

Fig. 1.2

Threshold Logic Element of Order 2 with 3 inputs

(distributed memory)

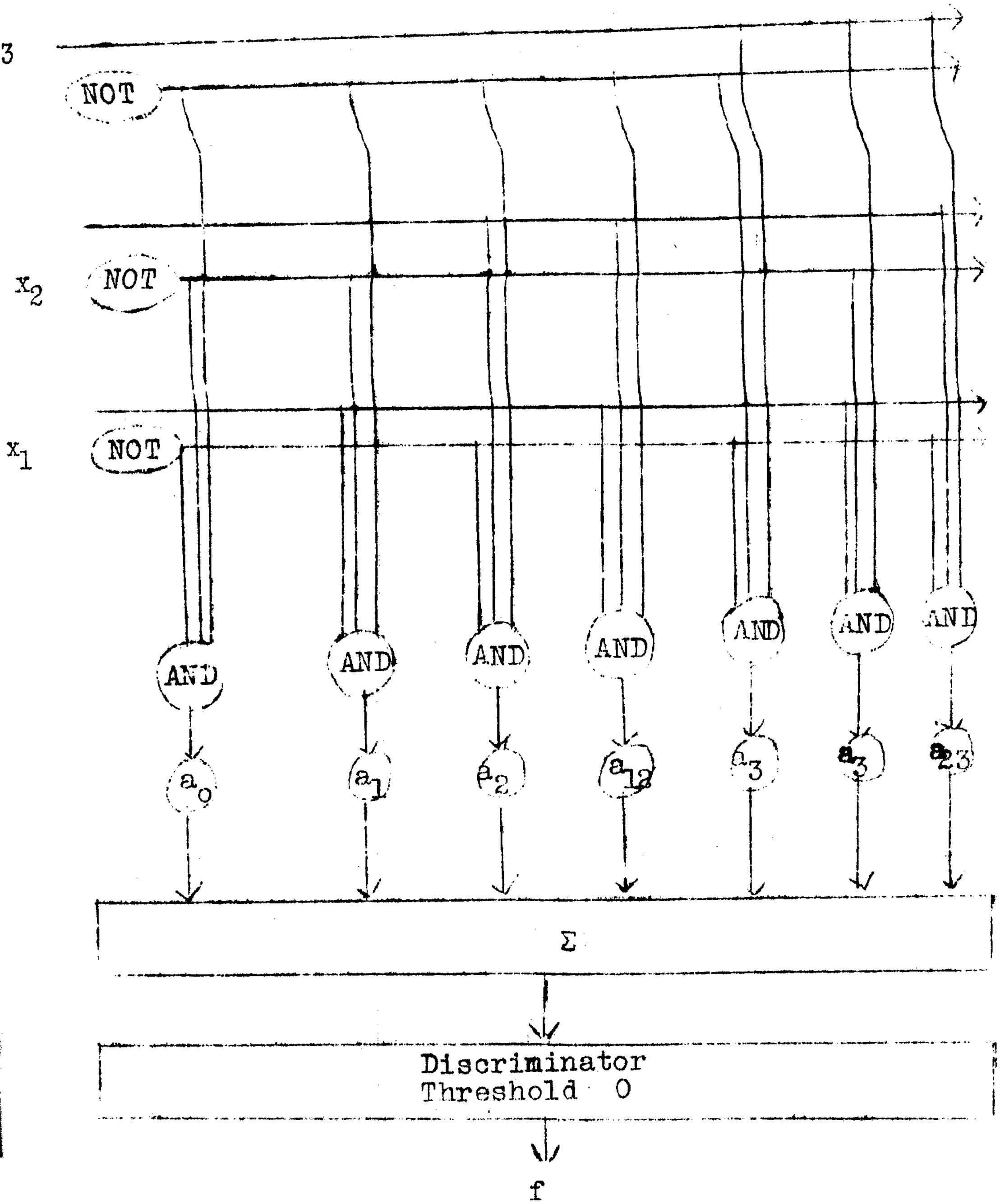

Fig. 1.3

Threshold Logic Element of Order 2
with 3 inputs
(Localized Memory)

1.5 Summary of Results

The investigation reported in these pages is an attempt towards the removal of such severe assumptions and towards the development of a theory, that has so far been limited to small class of switching functions, namely threshold functions, to handle any switching function.

In Chapter 2, we review various theories of representations of switching functions and probability distributions, and present the connections between them. Here we generalize the notion of Chow parameters and prove its connection to Golomb parameters.

In Chapter 3, we examine the various kinds of dependence as applied to a system of binary variables. In particular, we introduce the notions of dependence of order r logical, linear and stochastic - and examine the interconnections between them. We derive the structure of a set of logically r-dependent binary variables in terms of the number of points. We also show that stochastic dependence of order r implies logical dependence of order $\geq r$, which in turn implies linear dependence of order $\geq r$. Even though these results are not used for the particular kind of non-linear threshold device we are discussing, they are useful as necessary conditions for stochastic dependence, which is related to such non-linearity in terms of inputs, of switching functions. Further, these results are in the same spirit as generalization from order 1 to order r.

In Chapter 4, we first present some immediate consequences of the definition of a threshold function of order r. Then we examine the relationship to a non-linear summation device proposed by Kaszerman[18] and show how our theory results in a better realization of the function. Then we go on to the problems of characterizations of a threshold function of order r and show (1) the characterizing property of Chow parameters (2) the realization with Chebyshev approximations and (3) the validity of the Ho-Kashyap algorithm for testing and realization. We present an enumeration and a tabulation of switching functions of three inputs, by their threshold order.

In Chapter 5, we consider the application of the threshold logic gate of order r to such problems as pattern recognition and decoding. We show that the recognition with statistical dependence of order r (different from the concept of Chapter 3) requires a threshold gate of order at most r. We also show that a threshold gate of order r can be used as a decoder in the presence of dependent noise of order r.

CHAPTER 2

REPRESENTATIONS OF FUNCTIONS OF BINARY VARIABLES

' Nothing can please many, and please long, but just representations of a general nature. Particular manners can be known to few, and, therefore, few only can judge how nearly they are copied.'

Samuel Johnson
'Preface to Shakespeare.

2.0 Summary

The results of Partanen on the representation of switching functions and probability distributions using Boolean algebra are presented. Then the results of Bahadur are presented, which start with the vector space V of real-valued functions on the set of 2^n points of n binary variables and construct an orthonormal basis using probabilistic methods. It is shown that the basis obtained by Coleman to represent a switching function is a particular

case of this. Switching functions and probability distributions are particularly studied using this representation. Some results based on this framework are derived for use in subsequent chapters. The Rademacher-Walsh functions as basis of V and parameters of a switching function defined by Golomb are then presented. The concept of Chow parameters of a switching function is extended to a set of 2^n parameters. The relationships between all these representations are presented. Incidentally, a simple proof is given for a result of Gose used in his adaptive network. Several examples are worked out.

2.1 Introduction

The study of real-valued functions of n binary variables $x = (x_1, x_2, \ldots, x_n)$ is greatly facilitated by different representations of such functions. The starting point of such representations is the consideration of the space of such functions as a vector space and the determination of suitable bases for the vector space. In this chapter, we discuss three such bases; one is obtained with the help of Boolean algebra and thus considers only switching functions; the second due to Bahadur [2] is obtained on the basis of some probabilistic considerations, and, although it is particularly suited to represent switching functions and probability functions, it takes into account all real-valued functions; the third, which is the well-known work of Rademacher and Walsh [11], [38], also deals with real-valued functions and is ideal to represent switching functions. The second and third bases are related in the sense that a particular case of the second leads to the same representation as the third in the case of switching functions.

Nambiar [32] presents a discussion of the representation of probability distributions and their approximations.

In Section 2.3 (Theorems 2.6 and 2.7) we derive some results in the framework of Bahadur's, for use in subsequent chapters.

In Section 2.5, we present a generalization of Chow parameters (Definition 2.6) associated with a switching function and prove its relationship (Theorem 2.8) to the other parametric representations of switching functions.

In Section 2.6, on the basis of these results, we give a simple proof of a result of Gose [19] which he uses in an adaptive network proposed by him.

The results of Partanen [35] though not used subsequently have been summarised in Section 2.2, in view of the fact that these are available in a psychology publication and hence may not be widely known to switching theorists; it further enables us to give a fairly complete account of the results available on the representation of switching functions.

2.2 Boolean Algebraic Representation

Most of the results presented in this Section are found in Partanen [35].

Consider a Boolean algebra B generated by a set of n binary variables $x_1, x_2, \ldots, x_n$, under the operations $(\mp)$ (sum modulo 2 addition) and . (Boolean multiplication). We omit . in what follows; for example, xy for x.y. Then B is a vector space over GF (2).

DEFINITION 2.1: An element of B is said to be a basic element if it can be expressed as a product of n distinct factors, each factor being either x_i or $\bar{x}_i$.

There are 2^n such basic elements of B, denoted by

$$b^{(n)} = (b_0^{(n)}, b_1^{(n)}, b_2^{(n)}, \ldots, b_{2^n-1}^{(n)}) \qquad (2.1)$$

and it can be shown that

<u>**THEOREM 2.1:** The set of basic elements is a basis of the vector space B</u>.

Using the notation $\times$ for the Kronecker product of two matrices, $b^{(n)}$ can be written recursively in an elegant form as follows:

$$b^{(1)} = \begin{pmatrix} \bar{x}_1 \\ x_1 \end{pmatrix} \; ; \quad b^{(i)} = b^{(i-1)} \times \begin{pmatrix} \bar{x}_i \\ x_i \end{pmatrix} . \qquad (2.2)$$

For instance, $b^{(2)} = (\bar{x}_1 \bar{x}_2, x_1 \bar{x}_2, \bar{x}_1 x_2, x_1 x_2)$.

Another interesting set of elements is what is known as the set of sum variables denoted by

$$a^{(n)} = (a_0^{(n)} , a_1^{(n)}, a_2^{(n)}, \ldots, a_{2^n-1}^{(n)}), \qquad (2.3)$$

defined recursively as follows:

$$a^{(1)} = \begin{pmatrix} 0 \\ x_1 \end{pmatrix}, \quad \bar{a}^{(1)} = \begin{pmatrix} 1 \\ \bar{x}_1 \end{pmatrix}$$

$$a^{(i)} = a^{(i-1)} \times \begin{pmatrix} 1 \\ \bar{x}_i \end{pmatrix} + a^{(i-1)} \times \begin{pmatrix} 0 \\ x_i \end{pmatrix} \tag{2.4}$$

For instance, $a^{(2)} = (0, x_1, x_2, x_1 + x_2)$.

The relation between $a^{(n)}$ and $b^{(n)}$ is exhibited by

THEOREM 2.2: $a^{(n)} = A_n b^{(n)}; \quad \bar{a}^{(n)} = \bar{A}_n b^{(n)}$, (2.5)

where $A_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, $A_i = \begin{bmatrix} A_{i-1} & A_{i-1} \\ A_{i-1} & \bar{A}_{i-1} \end{bmatrix}$, $\bar{A}_i = \begin{bmatrix} \bar{A}_{i-1} & \bar{A}_{i-1} \\ \bar{A}_{i-1} & A_{i-1} \end{bmatrix}$

The set B is just the set of 2^{2^n} switching functions of n variables. Hence any switching function is expressible uniquely as a combination of components of $b^{(n)}$ with coefficients 0 or 1. This is the familiar normal form obtainable from the prime implicant table.

Let us now consider probability distributions on X, the set of 2^n points $x = (x_1, x_2, \ldots, x_n)$. It is clear that the joint probability distribution of $(x_1, x_2, \ldots, x_n)$

is completely specified by a set of 2^n-1 parameters, since the total of the function-values over the sample space is 1. There are many methods of specifying the parameters. One of these is the set of probabilities associated with the basic elements, that is,

$$p_\ell^{(n)} = \Pr \left\{ b_\ell^{(n)} = 1 \right\} , \ell = 0, 1, 2, \ldots, 2^n-1 \tag{2.6}$$

subject to $\sum_\ell p_\ell^{(n)} = 1$.

$$\text{Let } p^{(n)} = (p_0^{(n)}, p_1^{(n)}, p_2^{(n)}, \ldots, p_{2^n-1}^{(n)}). \tag{2.7}$$

It is easy to see that the 2^n events $b_\ell^{(n)} = 1$, correspond to 2^n elements of the sample space X of 2^n points.

Let q represent similar probabilities on the sum variables. That is,

$$q^{(n)} = (q_0^{(n)}, q_1^{(n)}, q_2^{(n)}, \ldots, q_{2^n-1}^{(n)}) \tag{2.8}$$

$$\text{where} \quad q_\ell^{(n)} = \Pr. \left\{ a_\ell^{(n)} = 1 \right\} . \tag{2.9}$$

Then the relationship between $q^{(n)}$ and $p^{(n)}$ is:

<u>**THEOREM 2.3:**</u> $$q^{(n)} = A_n p^{(n)}; \quad \bar{q}^{(n)} = \bar{A}_n p^{(n)}, \tag{2.10}$$

and $$p^{(n)} = 2^{-n} H_n (\bar{q}^n - q^n), \tag{2.11}$$

where $$H_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}; \quad H_i = \begin{bmatrix} H_{i-1} & H_{i-1} \\ H_{i-1} & -H_{i-1} \end{bmatrix}$$

2.3 Bahadur's Representation

The results of this Section are mainly based on the work of Bahadur [2].

The set of real-valued functions on X is a 2^n-dimensional vector space V over the reals. A basis of V is constructed in the following manner. Let

$$0 < \alpha_i < 1, \quad i = 1, 2, \ldots, n.$$

Consider the functions

$$z_i = \frac{x_i - \alpha_i}{\sqrt{\alpha_i(1 - \alpha_i)}} \tag{2.12}$$

Let

$$S = \{1;\ z_1, z_2, \ldots, z_n;\ z_1 z_2, \ldots, z_{n-1} z_n;\ z_1 z_2 z_3 \ldots; \\ \ldots;\ z_1 z_2 \ldots,\ z_n\} \tag{2.13}$$

be 2^n functions in V. A probability distribution $p^*(x)$ is defined on X as

$$p^*(x) = \prod_{i=1}^{n} \alpha_i^{x_i} (1 - \alpha_i)^{1-x_i}, \tag{2.14}$$

made up of independent binomial components. Then

$$E_{p^*}(z_i) = 0, Var_{p^*}(z_i) = 1, \quad i = 1,2,\ldots, n. \tag{2.15}$$

An inner product is defined on V as

$$(f, g) = E_{p^*}(fg), \quad f, g \in V. \tag{2.16}$$

Then

THEOREM 2.4: S is an orthonormal basis of V.

The coefficients r_s corresponding to $s \in S$, in the representation of any function f with S is given by

$$r_\ell = \sum_{x \in X} f(x)\ s(x)\ p^*(x), \tag{2.17}$$

$$\ell = 0,\ 1,\ 2, \ldots,\ 2^n - 1.$$

By choosing $\alpha_i = \frac{1}{2}$, and denoting z_i by y_i in this case, we have $y_i = 2x_i - 1$ and thus $y_i = +1$ if $x_i = 1$ and $y_i = -1$ if $x_i = 0$. So all the elements of S take -1 and $+1$ values. In this case we denote S by S^*. This is very convenient to represent switching functions. Representating r in this case by d, the correspondence between d and p is given by

$$d = p^T H_n \cdot \tag{2.18}$$

This basis is independently obtained by Coleman [9] and is written in the alternative form

$$(-1)^{k_1 x_1 + k_2 x_2 + \cdots + k_n x_n}, \tag{2.19}$$

$k_1, k_2, \ldots, k_n = 0$ or 1.

Let us consider the representation of an arbitrary probability distribution $p(x)$ on X. For this, let us choose

$$\alpha_i = \Pr.\left\{ x_i = 1 \mid p(x) \right\} , \quad i=1,2,\ldots, n. \tag{2.20}$$

Consider $p(x)|p^*(x)$. Then

$$\rho_\ell = E_p(s_\ell), \quad j = 0,1,2,\ldots, 2^n-1 . \tag{2.21}$$

Thus $\rho_i = 0$, $i = 1,2,\ldots, n$ (that is, those associated with $s_j = x_1, x_2,\ldots, x_n$). Hence

THEOREM 2.5 A probability distribution $p(x)$ has a representation

$$\frac{p(x)}{p^*(x)} = 1 + \sum_{j=2}^{n} \rho_{i_1 i_2 \ldots i_j} z_{i_1} z_{i_2} \ldots z_{i_j} . \tag{2.22}$$

Since any function has a representation in terms of elements of S, we can also represent a probability distribution by the parameters of $\log p(x)$ or $\log \frac{p(x)}{p^*(x)}$ provided $p(x) > 0$, for all $x \in X$.

DEFINITION 2.2: The parameters r_ℓ defined in (2.21) are called the correlation parameters associated with the probability distribution $p(x)$.

We now prove some new results which will be of use in the following chapters.

THEOREM 2.6: The set of functions

$$T = \{1;\ x_1,\ x_2,\dots,x_n;\ x_1x_2,\dots,x_{n-1}x_n; x_1x_2x_3,\dots;\dots;\dots;\ x_1x_2\dots,x_n\} \qquad (2.23)$$

is a basis of V.

Proof: The matrix A representing the functions $h(x) \in T$ as linear combinations of $g(x) \in S$ is obtained as follows. The function $y_{i_1}\ y_{i_2}\dots\ y_{i_j}$ is a linear combination of the 2^j functions constructed from $x_{i_1}, x_{i_2},\dots,\ x_{i_j}$ in the same manner as (2.23), that is,

$$\{x_{i_1}\ x_{i_2},\dots,x_{i_j}; x_{i_1}x_{i_2},\dots; x_{i_1}x_{i_2}x_{i_3},\dots; x_{i_1}x_{i_2}\dots x_{i_j}\} \qquad (2.24)$$

and the coefficients are

$$(-1)^j\ \{1;\ -2(1,1,\dots1);\ 2^2(1,1,\dots,1);\dots;\dots;2^j\}. \qquad (2.25)$$

Let us denote the set like (2.24) obtained from the variable in any $t \in T$, by M_t and also, let

$$M_t^* = M_t - \{x_{i_1}, x_{i_2}, \ldots \} . \tag{2.26}$$

Removing the powers of 2 from each column of the matrix A thus obtained, we have a term

$$N_n = \sum_{i=1}^{n} i\binom{n}{i} = n2^{n-1} \tag{2.27}$$

Let a matrix B be given recursively by

$$B_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad B_{i+1} = \begin{bmatrix} B_i & 0 \\ -B_i & B_i \end{bmatrix} \tag{2.28}$$

For instance,

$$B_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ -1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \end{bmatrix} \tag{2.29}$$

The matrix B corresponds to the order

$$\{1; x_1, x_2, x_1x_2, x_3, x_1x_3, x_2x_3, x_1x_2x_3, \ldots; x_1x_2\ldots x_n\}. \quad (2.30)$$

Let B_n^* be a matrix obtained from B_n by a rearrangement of the rows of B_n to correspond to the order of (2.23). Then

$$A = 2^{N_n} \; B_n^* \; . \qquad (2.31)$$

B_n (or B_n^*) can be reduced to a $2^n \times 2^n$ unit matrix by elementary row operations of subtracting from the row corresponding to $x_{i_1} x_{i_2} \ldots x_{i_j}$ all the rows corresponding to elements of (2.24). Hence the set of functions T is linearly independent.

Hence Theorem 2.6.

Let $p_{r_t}(x)$ denote the joint distribution of $(x_{i_1}, x_{i_2}, \ldots, x_{i_{r_t}})$ for $t \in T$. Also let

$$\Pi_t(x) = \frac{p_t(x)}{p_{i_1}(x)\, p_{i_2}(x) \ldots p_{i_{r_t}}(x)} \; . \qquad (2.32)$$

Considering

$$\prod_t(x) \quad \text{as} \quad \frac{p_t(x) \prod_{j \neq i_v} p_j(x)}{p_1 p_2 \cdots p_n} \tag{2.33}$$

$$(v = 1, 2, \ldots, r_t)$$

and using Theorem 2.5, we have

$$\prod_t(x) = 1 + \sum_{\ell \in M_t} \rho_\ell \, g_\ell(x), \quad t \in T^*, \tag{2.34}$$

where $T^* = T - \{x_1, x_2, \ldots, x_n\}$. (2.35)

Let $\operatorname{Sgn}(t) = (-1)^{r_t}$. (2.36)

The following theorem expresses $p(x)$ in terms of distributions of lower order.

<u>THEOREM 2.7</u>: <u>For a probability distribution $p(x)$ on X,</u>

$$\frac{p(x)}{p^*(x)} = (-1)^n \Big[(n-1) - \sum_{j=2}^{n} \sum_{i_1 < i_2 < \cdots < i_j} (-1)^{i_j} \prod_{i_1 i_2 \ldots}(x) \Big]$$

(2.37)

<u>Proof</u>: <u>Case (i)</u>: <u>No $\rho_{i_1 i_2 \ldots i_j}$ is zero, $j \geq 2$.</u>

Consider the linear transformation σ on the vector space V, defined by

$$\sigma [\, z_{i_1} z_{i_2} \cdots z_{i_j}] = \pi_{i_1 i_2 \cdots i_j}(x), \quad 2 \leq j \leq n-1$$

$$\sigma[\, z_1 z_2 \cdots z_n] = z_1 z_2 \cdots z_n . \qquad (2.38)$$

Let $B = \{ 1;\ \pi_{12}, \pi_{13}, \ldots;\ \pi_{123}, \ldots; \ldots; \ldots; z_1 z_2 \ldots z_n \}$. (2.39)

Because of (2.34), the square matrix of order $2^n - n$ of σ with respect to A as well as B is

$$M = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
1 & \rho_{12} & 0 & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
1 & 0 & \rho_{23} & 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
1 & 0 & 0 & \rho_{31} & 0 & 0 & \cdot & \cdot & \cdot & \cdot & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
1 & \rho_{12} & \rho_{23} & \rho_{31} & 0 & \cdot & \cdot & \rho_{123} & 0 & \cdot & \cdot & 0 \\
1 & \rho_{12} & 0 & 0 & \rho_{24} & \rho_{14} & \cdot & \cdot & \rho_{124} & \cdot & \cdot & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix} \qquad (2.40)$$

and with reference to B as basis $\sigma \left[\dfrac{p(x)}{p^*(x)} \right]$ is

$$\xi = (1;\ \rho_{12},\ \rho_{23};\ \ldots;\ \rho_{123};\ldots;\ldots;\ldots;\ \rho_{12\ldots n}). \tag{2.41}$$

M can be reduced by elementary row operations to a diagonal matrix with 1 and the ρ's, the row operations are subtractions from the row corresponding to $x_{i_1} x_{i_2} \ldots x_{i_j}$, all the rows corresponding to (2.24). Hence it is non-singular if and only if none of the ρ's is zero. Now

$$M^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
-\frac{1}{\rho_{12}} & \frac{1}{\rho_{12}} & 0 & 0 & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
-\frac{1}{\rho_{23}} & 0 & \frac{1}{\rho_{23}} & 0 & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
-\frac{1}{\rho_{31}} & 0 & 0 & \frac{1}{\rho_{31}} & \cdots & \cdot & \cdot & \cdot & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\frac{2}{\rho_{123}} & \frac{-1}{\rho_{123}} & \frac{-1}{\rho_{123}} & \frac{-1}{\rho_{123}} & \cdots & \frac{1}{\rho_{123}} & 0 & 0 & \cdots & 0 \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\frac{1}{\rho_{1234}} & \frac{1}{\rho_{1234}} & \cdot & \cdot & \cdots & \frac{-1}{\rho_{1234}} & \frac{1}{\rho_{1234}} & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & \cdot \\
0 & 0 & 0 & \cdot & \cdots & \cdot & \cdot & \cdot & \cdots & 1
\end{bmatrix} \tag{2.42}$$

$$\sigma^{-1}\ \sigma\ \left(\frac{p(x)}{p^*(x)}\right) = \frac{p(x)}{p^*(x)} ,$$

and with reference to B, this is

$$\xi\ M^{-1} = (-1)^n[(n-1);-1,-1,\ldots;1,1,\ldots;\ldots;(-1)^n\ \varrho_{12\ldots n}] \qquad (2.43)$$

Hence the theorem for case (i).

<u>Case (ii)</u>: <u>Some $\varrho_{i_1 i_2 \ldots i_j}$ are zero, $j \geq 2$.</u>

Consider the linear transformation σ restricted to these $g(x) \in S$, for which $\varrho_g \neq 0$. The matrix M is of the same form as before, excepting that the rows and columns corresponding to the g for which $\varrho_g = 0$ are absent. Let ξ be a vector in which the element corresponding to $g = 0$ is absent. Then M as before, is non-singular and M^{-1} is of the same form except that the rows and columns corresponding to $\varrho_g = 0$ are absent. Then

$$\xi\ M^{-1} = (-1)^n[(n-1);\ 1,1,\ldots 1;\ -1,-1,\ldots -1;\ldots(-1)^n\ \varrho_{12\ldots n}] \qquad (2.44)$$

$$+ \sum_{\substack{\text{over the set for which} \\ \varrho_g = 0}} (-1)^j \; [(j-1); \sum_{\substack{\text{over the set } M_t^* \\ \text{generated by } g \\ \text{for which } \varrho_g = 0}} (-1) \;] \cdot \qquad (2.45)$$

Since, by case (i)

$$\pi_{i_1 i_2 \ldots i_j}(x) = (-1)^j \; [(j-1); \sum_{\substack{\text{over these } M_t \text{ generated by} \\ x_{i_1} x_{i_2} \ldots x_{i_j}}} (-1)^j \;] \qquad (2.46)$$

$$+ \varrho_{i_1 i_2 \ldots i_j}(x),$$

and

$$\varrho_{i_1 i_2 \ldots i_j} = 0,$$

$$\xi M^{-1} = (-1)^n [(n-1); \sum_{\text{over all } i_1, i_2, \ldots i_j} (-1)^j \;; (-1)^n \varrho_{12\ldots n}] \qquad (2.47)$$

Hence theorem for Case (ii) as well.

Hence Theorem 2.7.

2.4 Rademacher-Walsh Representation

A set of 2^n functions are defined on the half-open interval $[0, 1)$ as follows:

$$r_0(t) = 1, \qquad 0 \leq t < 1$$

$$r_1(t) = 1, \qquad 0 \leq t < \tfrac{1}{2}$$

$$= -1, \qquad \tfrac{1}{2} \leq t < 1;$$

and

$$T(t) = 2t, \qquad 0 \leq t < \tfrac{1}{2}$$

$$= 2t-1; \qquad 0 \leq t < 1$$

$$r_i(t) = r_{i-1}(T(t)), \quad 0 \leq t < 1, \quad i = 1,2,\ldots, n.$$

Defining inner product on the space V as

$$(f, g) = \int_0^1 f(t)g(t)\,d\,t, \qquad f, g \in V,$$

the $(n+1)$ functions $r_i(t)$, $i = 0, 1, 2, \ldots, n$ and their distinct products $r_{i_1} r_{i_2} \cdots r_{i_j}$, $j = 0,1,2,\ldots,n$ ($j = 0$ gives constant 1) form an orthonormal basis of V[38]. These functions are constants on the 2^n intervals $[\frac{\ell}{2^n} \leq t < \frac{\ell+1}{2^n})$, $\ell = 0, 1, 2, \ldots, 2^n-1$, and can be obtained by assigning value 1 in one interval and -1 in the others. The coefficients of this basis are obtained with the help of parameters defined by Golomb [11] as follows:

<u>DEFINITION 2.4:</u> The measure of a switching function f, denoted by $m[f]$, is the number of points at which it takes the true value, or, in our notation,

$$m[f] = \sum_{x \in X} f(x_1, x_2, \ldots, x_n) \qquad (2.48)$$

<u>DEFINITION 2.5:</u> The 2^n parameters defined by

$$g(x_{i_1} x_{i_2} \cdots x_{i_k}) = m[f(\mp)x_{i_1}(\mp)x_{i_2}(\mp)\ldots(\mp)x_{i_k}] \qquad (2.49)$$

$$i_1 < i_2 < \ldots < i_k ,$$

($k = 0$ gives $m[f]$.),

are called the <u>Golomb parameters</u> of the switching function f.

It is known [11] that the coefficients of the Rademacher-Walsh expansion are obtained as

$$\frac{(-1)^k}{2^n} [2g(x_{i_1} x_{i_2} \cdots x_{i_k}) - 2^n], \qquad (2.50)$$

and these are the same as the d parameters of Coleman in (2.18). This is obtained by considering the range of the Rademacher-Walsh functions as $\{-1, 1\}$ instead of $\{0, 1\}$ and equating the coefficients to the Golomb coefficients.

2.5 Chow Parameters

Chow [4] defined a set of $(n+1)$ parameters for a switching function and proved that these parameters are characteristic of threshold functions. We define a set of 2^n parameters similar to these to characterize any switching function completely by its threshold order (see Chapter 1, Definition 1.2; Chapter 4, Section 1).

Let $F_{y_{i_1} y_{i_2} \cdots y_{i_r}}$ $(F = 2f-1,\ y_i = 2x_i - 1)$ denote a switching function on the 2^{n-r} points obtained by the valuation $y_{i_1} = y_{i_2} = \cdots = y_{i_r} = 1$. Let the set of 2^r valuations of $y_{i_1}, y_{i_2}, \ldots, y_{i_r}$ be denoted by $Y_{i_1 i_2 \ldots i_r}$

The following Lemmas are obtained immediately.

LEMMA 2.1

$$m\,[F_{y_{i_1} y_{i_2} \cdots y_{i_r}}] + m[F_{y_{i_1} y_{i_2} \cdots \bar{y}_{i_j} \cdots y_{i_r}}]$$

$$= m\,[F_{y_{i_1} y_{i_2} \cdots y_{i_{j-1}} y_{i_{j+1}} \cdots y_{i_r}}]. \qquad (2.51)$$

LEMMA 2.2

$$m[F_{y_{i_1} y_{i_2} \cdots y_{i_r}}] + m[F_{y_{i_1} y_{i_2} \cdots y_{i_r}}] = 2^{n-r} \qquad (2.52)$$

LEMMA 2.3

$$\sum_{\substack{(y_{i_1}, y_{i_2}, \ldots, y_{i_r}) \\ \in Y_{i_1 i_2 \ldots i_r}}} m[F_{y_{i_1} y_{i_2} \ldots y_{i_r}}] = 2^{n-r} \qquad (2.33)$$

DEFINITION 2.6 The 2^n Chow parameters are given by

$$ch(y_{i_1} y_{i_2} \ldots y_{i_r}) = \sum_{\substack{(y_{i_1}, y_{i_2}, \ldots, y_{i_r}) \\ \in Y_{i_1 i_2 \ldots i_r}}} (y_{i_1} y_{i_2} \ldots y_{i_r}) F_{y_{i_1} y_{i_2} \ldots y_{i_r}} \qquad (2.54)$$

$$i_1 < i_2 < \ldots < i_r, \quad r = 1, 2, \ldots, n.$$

and $ch(y_o) = m[F] - 2^{n-1}$ (2.55)

We now prove a result relating Chow parameters to the coefficients of the Coleman basis (2.19).

THEOREM 2.8:

$$ch(y_{i_1} y_{i_2} \ldots y_{i_r}) = 2^{n-1}\, d_{i_1 i_2 \ldots i_r}, \qquad (2.56)$$

where $d_{i_1 i_2 \ldots i_r}$ is the coefficient of $y_{i_1} y_{i_2} \ldots y_{i_r}$ in the Coleman basis.

Proof: $2^n d_{i_1 i_2 \ldots i_r} = \Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} \quad F(y)$

$$= \Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} \quad m[F_{y_{i_1} y_{i_2} \ldots y_{i_r}}] - m[\bar{F}_{y_{i_1} y_{i_2} \ldots y_{i_r}}]$$

$$= \Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} \quad m[F_{y_{i_1} y_{i_2} \ldots y_{i_r}}]$$

$$- \Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} \quad 2^{n-r} - m[F_{y_{i_1} y_{i_2} \ldots y_{i_r}}]$$

using Lemma 2.2.

$$= 2\, \Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} \quad m[F_{y_{i_1} y_{i_2} \ldots y_{i_r}}]$$

[since $\Sigma\, y_{i_1} y_{i_2} \ldots y_{i_r} = 0$]

$$= 2ch\ (y_{i_1} y_{i_2} \ldots y_{i_r})$$

Hence the Theorem.

The Theorem given below follows from the orthogonality of the various bases:

THEOREM 2.9: If $\phi_\ell(x)$, $\ell = 0, 1, 2, \ldots, 2^n-1$ are the elements of any of the bases of Bahadur, Coleman, and Rademacher-Walsh, and if

$$f(x) = \sum_\ell a_\ell\, \phi_\ell(x) \tag{2.57}$$

then the least square approximation of $f(x)$ by any subset of $\phi_\ell(x)$, say, by $\phi_{\ell_i}(x)$, $i = 1, 2, \ldots, k$ is given by

$$\sum_{i=1}^{k} a_{\chi_i} \phi_{\chi_i}(x). \qquad (2.58)$$

2.6 An Alternative Proof of Gose's Result

Gose [12] has suggested an adaptive network for producing real-valued functions of binary inputs. An important factor that helps to make the network adaptive is that the two transformations T_1 and T_2 that he defines are inverses of each other. His proof is rather long and we give a short proof of this fact using the Coleman basis.

Define linear transformation T_1 and T_2 on V as follows:

$$T_1 : \quad w_g = (f,g) = 2^{-n} \sum_{x \in X} (fg), \quad g \in S^* \qquad (2.59)$$

w is a 2^n-vector.

$$T_2: \quad f(x) = \sum_{g \in S^*} w(x)g(x), \qquad x \in X \qquad (2.60)$$

f(x) is a 2^n-vector.

THEOREM 2.8: (Gose [12]): T_1 and T_2 are inverses of each other.

Proof: Since S^* is a basis, the weights associated with f(x) in (2.59) are unique. Thus from the w_g of (2.59) used in (2.60) we get f(x). Thus $T_1 T_2$ is the identity transformation. Hence $T_2 = T_1^{-1}$ or $T_1 = T_2^{-1}$. Alternatively, the images of $g \in S^*$ by the transformation T_1 are 2^n basic elements $b_\ell^{(n)}$ of B defined in (2.1) but over the field of real numbers. These functions take values 1 at one point of the 2^r points and 0 on the rest. Thus they are linearly independent. Hence T_1 is non-singular. The images of elements of $b^{(n)}$ by T_2 can be easily seen to be the elements of S^*. Hence T_2 is the inverse of T_1.

2.7 Examples

We present examples of the various parametric representions of switching functions and probability distributions, discussed in the previous sections.

Table 2.1

3-input Switching and Probability Functions

ℓ	x_1	x_2	x_3	f	F	p_1	p_2	p_3
0	0	0	0	0	-1	.125	.2500	.1250
1	0	0	1	1	1	.125	.1250	.1250
2	0	1	0	0	-1	.125	.0625	.0625
3	0	1	1	0	-1	.125	.0625	.0625
4	1	0	0	0	-1	.125	.0625	.0625
5	1	0	1	1	1	.125	.0625	.0625
6	1	1	0	1	1	.125	.1250	.1250
7	1	1	1	1	1	.125	.2500	.3750

F and f are same functions; with different truth values. In terms of the basis (2.1) f is represented by (0, 0, 0, 1, 1, 1, 0, 1) since

$$f = \bar{x}_1\bar{x}_2x_3 + x_1\bar{x}_2x_3 + x_1x_2\bar{x}_3 + x_1x_2x_3.$$

Noting that the p's of (2.7) are the same as the columns of a probability distribution, the q's of (2.8) are obtained from p_2 in Table 2.1 as

(0; .5, .5; .375; .5; .375, .25; .5).

Now let us consider the Coleman basis.

Table 2.2

Coleman Basis

	0	1	2	3	4	5	6	7
0	1	-1	-1	-1	1	1	1	-1
1	1	-1	-1	1	-1	-1	-1	1
2	1	-1	1	-1	-1	-1	1	1
3	1	-1	1	1	-1	1	-1	-1
4	1	1	-1	-1	-1	1	1	1
5	1	1	-1	1	-1	-1	1	-1
6	1	1	1	-1	1	-1	-1	-1
7	1	1	1	1	1	1	1	1

The coefficients for the Coleman basis for various functions are given by:

#: $\frac{1}{8}$ (4; 2, 0, 2; 2, -2, 0; 0)

F : $\frac{1}{8}$ (0; 4, 0, 4; 4, -4, 0; 0)

p_1: $\frac{1}{8}$ (1, 0, 0, 0; 0, 0, 0; 0)

p_2: $\frac{1}{8 \times 16}$ (16; 0, 0, 0; 8, 4, 4; 0)

$\log_{10} p_2 = -\frac{1}{8}(7.8628;\ 0,\ 0,\ 0;\ -1.0160, -0.6020,\ -0.6020; 0)$

Notice that both p_1 and p_2 yield $P(x_i = 1) = \frac{1}{2}$, $i = 1, 2, 3$ and hence $z_i = y_i$ and so the representation by (2.22) is the same as by Coleman basis. For p_3, the probabilities $P(x_i = 1)$ are

$P(x_i = 1) = 0.625, \quad i = 1, 2, 3.$

$\alpha_i(1 - \alpha_i) = .234375, \quad \sqrt{\alpha_i(1 - \alpha_i)} = 0.484.$

The Bahadur basis is constructed in Table 2.3.

Table 2.3

Bahadur Basis

	1	z_1	z_2	z_3	z_1z_2	z_2z_3	z_3z_1	$z_1z_2z_3$
0	1	0.7748	0.7748	0.7748	0.6003	0.6003	0.6003	0.4651
1	1	0.7748	0.7748	-1.2913	0.6003	-1.0005	-1.0005	-0.7752
2	1	0.7748	-1.2913	0.7748	-1.0005	-1.0005	0.6003	-0.7752
3	1	0.7748	-1.2913	-1.2913	-1.0005	1.6675	-1.0005	1.2919
4	1	-1.2913	0.7748	0.7748	-1.0005	-0.6003	-1.0005	-0.7752
5	1	-1.2913	0.7748	-1.2913	-1.0005	-1.0005	1.6675	1.2919
6	1	-1.2913	-1.2913	0.7748	1.6675	-1.0005	-1.0005	1.2919
7	1	-1.2913	-1.2913	-1.2913	1.6675	1.6675	1.6675	-2.1532

Then

$$p^*(x) = \prod_{i=1}^{n} (0.625)^{x_i} (0.375)^{1-x_i},$$

and

$$\frac{p(x)}{p^*(x)} = (1;\ 0,\ 0,\ 0;\ 0.6336,\ 0.4669,\ 0.4669;\ -0.6021).$$

Let us now compute the Golomb parameters of f as follows:

Table 2.4

Computation of Golomb Parameters

	x_1	x_1	x_3	f	$f(\pm)x_1$	$f(\pm)x_2$	$f(\pm)x_3$	$f(\pm)x(\pm)x_2$	$f(\pm)x_2(\pm)x_3$	$f(\pm)x_3(\pm)x_1$	$f(\pm)x_1(\pm)x_2(\pm)x_3$
0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	1	1	1	0	1	0	0	0
2	0	1	0	0	0	1	0	1	1	0	1
3	0	1	1	0	0	1	1	1	0	1	0
4	1	0	0	0	1	0	0	1	0	1	1
5	1	0	1	1	0	1	0	0	0	1	1
6	1	1	0	1	0	0	1	1	0	0	1
7	1	1	1	1	0	0	0	1	1	1	0

The Golomb parameters are:

$$(4;\ 2,\ 4,\ 2;\ 6,\ 2,\ 4;\ 4).$$

From this one gets the Rademacher-Walsh coefficients using (2.50) as :

$$\frac{1}{8}\ (0;\ 4,\ 0,\ 4;\ 4,\ -4,\ 0;\ 0)$$

which are the same as the d-coefficients obtained from Coleman basis for F. The Chow parameters are obtained as:

$$(0;\ 2,\ 0,\ 2;\ 2,\ -2,\ 0;\ 0).$$

These are seen to satisfy the relations (2.50) and (2.54) - (2.56).

CHAPTER 3

DEPENDENCE OF SWITCHING FUNCTIONS

Independence? That's middle class blasphemy. We are all dependent on one another, every soul of us on earth'.

George Bernard Shaw
'Pygmalion'

3.0 Summary

The notions of linear, logical and stochastic dependence, of n dichotomies, of order $r \leq n$, are introduced. The connections between linear, logical and stochastic dependence are investigated. This includes generalizations of some well-known results on logical independence and the result that logical independence is a necessary consequence of stochastic independence, of a system of dichotomies. The concepts of logical independence and dependence have been applied to switching functions and relay-contact network by previous authors.

3.1 Introduction

In previous discussions, we have used the terms 'binary variables' generally to mean 'inputs', and, 'switching function' to generally mean output. However, the inputs can themselves be considered trivial switching functions and hence as outputs. Since the discussion in this Chapter does not differentiate inputs and outputs, we avoid these two terms in preference to a common term 'dichotomies'.

Several notions of independence of mathematical objects exist, each one finding use in different contexts. Many of these notions are of special significance when applied to a system of dichotomies. Kjellberg [19] gives an account of the results in this regard and discusses several interesting connections between the notions of logical, stochastic, linear and functional independence of a system of dichotomies. He also indicates an application of these notions to relay-contact network. He reports that an interpretation of independence has been applied to switching functions by Muller [29].

It is shown by Kjellberg [19] that 'logical independence of a system of dichotomies is a necessary condition for the stochastic independence of the corresponding events'. This means that if a system of n dichotomies is stochastically independent, then, it is necessary that all 2^n combinations of values have non-zero probabilities. We investigate here the situation in which logical independence, and, as a consequence, stochastic independence is lost, and present some methods of describing the strength of the dependence-logical as well as stochastic. The notion of logical dependence of order r ($\leq n$) of n dichotomies is straightforward, and applies to a system of partitions of a set E just as well as to a system of dichotomies. The notion of stochastic dependence of order r ($\leq n$) which applies to n dichotomous variables, is introduced with the help of a result of Bahadur on the representation of the joint probability distribution of n dichotomies. In other words, we investigate here the dependence of functions defined on the subsets of 2^n points of n dichotomous variables.

After presenting some general consequences of the definition of logical dependence of order r, we proceed to present results connecting linear and logical dependence and stochastic and logical dependence. We show that linear dependence of order r is a necessary consequence of logical dependence of order r, and that logical dependence of order at least r is a necessary consequence of stochastic dependence of order r. This gives an idea of the nature of possible probability distributions that can be defined on subsets, of various kinds, of the set of all 2^n possible combinations of n dichotomies. As in the case of independence, the converse is not true that logical dependence of order r implies stochastic dependence of order r; however, a probability distribution which gives stochastic dependence of order r, can be found on dichotomies which are logically dependent of order r. Finally, we define another kind of stochastic dependence of order r and investigate the nature of logical dependence associated with it.

3.2 Logical Dependence

We extend the notion of logical independence, restricting our attention to finite sets.

DEFINITION 3.1: Consider a finite set E partitioned into a system of n disjoint nonempty subsets as follows:

$$E = A_o^i \cup A_1^i \cup \ldots \cup A_{n_i}^i \qquad (3.1)$$

$$A_j^i \cap A_h^i = \phi, \text{ the null set, if } j \neq h.$$

Then we say that this system of partitions is logically <u>dependent of order</u> r (or <u>logically r-dependent</u>) if

(i) any arbitrary collection of components, one from each of any collection of $r(\leq n)$ partitions is nonempty and

(ii) at least one collection of components, one from each of any collection of $(r+1)$ partitions is empty.

For $r = n$, (ii) is understood to be vacuous.

<u>REMARK 3.1.</u> If the system of partitions E is logically n-dependent, then the system is logically independent.

<u>REMARK 3.2.</u> A system is logically r-dependent, if any sub-system with r **partitions** is logically independent and at least one sub-system of $(r+1)$ partitions is not logically independent.

<u>REMARK 3.3.</u> The logical dependence of a system of dichotomies $x_1, x_2, \ldots, x_n$ will be viewed with reference to a particular kind of partition. Let X be the set of 2^n vectors $x = (x_1, x_2, \ldots, x_n)$. Let $A \subseteq X$, and

$$A_0^i = \{x = (x_1, x_2, \ldots, x_n) : x_i = 0\} \tag{3.2}$$

$$A_1^i = \{x = (x_1, x_2, \ldots, x_n) : x_i = 1\} \tag{3.3}$$

such that

$$A = A_0^i \cup A_1^i \tag{3.4}$$

Obviously,

$$A_0^i \cap A_1^i = \phi \tag{3.5}$$

<u>DEFINITION 3.2.</u> A set A of vectors of n <u>dichotomous</u> variables is said to be <u>logically</u> <u>r-dependent</u>, if the system of n partitions

$$A = A_0^i \cup A_1^i, \qquad i = 1,2,\ldots, n \tag{3.6}$$

is logically r-dependent.

<u>EXAMPLES 3.1.</u>

Set	Order of Logical Dependence
1). X : 2^n points $(x_1, x_2, \ldots, x_n)$	n
2). ϕ	0
3). (0, 0,..., 0)	0
4), (1, 1,..., 1)	0
5). (0, 0,..., 0)(1, 1,..., 1)	1

One can talk of the order of logical dependence of n dichotomous variables $x_1, x_2, \ldots, x_n$ with reference to a probability distribution on X as follows:

<u>DEFINITION 3.3</u> A set of n dichotomous variables $x_1, x_2, \ldots, x_n$ with a <u>probability distribution</u> $p(x_1, x_2, \ldots, x_n)$ is said to be <u>logically r-dependent</u> if the set $A = \{(x_1, x_2, \ldots, x_n): p(x_1, x_2, \ldots, x_n) > 0\}$ is logically r-dependent.

This definition will be useful in connecting logical and stochostic dependence of dichotomies.

We shall now present some immediate consequences of the foregoing definitions, which are direct generalizations of the results on independence found in Kjellberg [19]; we omit the proofs, since they are fairly straight-forward:

<u>THEOREM 3.1.</u> <u>If a system of n partitions is logically r-dependent, then the order of logical dependence of any sub-system of ℓ partitions is ℓ if $\ell \leq r$, and $\geq r$ if $\ell > r$ and there is at least one sub-system of (r+1) partitions, which is logically r-dependent.</u>

<u>THEOREM 3.2</u> . <u>If a system of n partitions of a set E divide E into $m_1, m_2, \ldots, m_n$ components, and if the system is logically r-dependent, then E must contain at least</u>

$$\max_{\substack{i_1 < i_2 < \ldots < i_r \\ = 1,2,\ldots,n}} m_{i_1} m_{i_2} \cdots m_{i_r} \tag{3.7}$$

<u>elements.</u>

THEOREM 3.3. If a system of n dichotomies is logically r-dependent, then the set must contain at least 2^r elements, or the order of logical dependence of a system of dichotomies with p elements is at most $\log_2 p$.

THEOREM 3.4. The order of logical dependence of a set of switching functions (excluding the constants 0 and 1) on a logically r-dependent set of dichotomies is $\leq r$.

THEOREM 3.5: Every function in a set of logically r-dependent switching functions on a logically r-dependent set of dichotomies takes the values 0 and 1, each for at least 2^{r-1} combinations of values of the variables.

THEOREM 3.6: A product of $p (\leq r)$ functions (or their complements) of a set of logically r-dependent switching functions on a logically r-dependent set of dichotomies takes the values 0 and 1, each for at least 2^{r-p} combinations of values of the variables.

REMARK 3.4: Let $u_1, u_2, \ldots, u_n$ be a set of logically r-dependent switching functions on a set of logically r-dependent dichotomies $x_1, x_2, \ldots, x_n$. Then, in general, $x_1, x_2, \ldots, x_n$ and $u_1, u_2, \ldots, u_n$ are not one-to-one functions.

EXAMPLES 3.2 Consider the case $n=3$. The set of points $A = (0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 1)$ is a logically 2-dependent set. If a probability distribution is defined on (x_1, x_2, x_3) such that these four points carry positive probabilities, say, $\frac{1}{4}$ each, and the other points carry zero probability, then we would say that under this probability distribution, (x_1, x_2, x_3) is a set of logically 2-dependent dichotomies. Consider two sets of switching functions f_1, f_2, f_3 and g_1, g_2, g_3 on A as follows:

x_1	x_2	x_3	f_1	f_2	f_3	g_1	g_2	g_3
0	0	1	1	1	1	1	1	1
0	1	0	0	0	0	1	0	0
1	0	0	0	0	1	0	1	0
1	1	1	1	0	1	0	0	1

f_1, f_2, f_3 are logically 1-dependent, g_1, g_2, g_3 are logically 2-dependent, since the set of values is the same as A. In this case, if one of g_1, g_2, g_3 does not take values 0 or 1, at least at two points, then it is not possible for them to be 2-dependent. The product of any two functions of g_1, g_2, g_3 takes values 0 or 1 at least at one point.

3.3 Logical and Linear Depedence

We introduce the notion of linear dependence of order r and establish its relation to logical dependence of order r.

DEFINITION 3.4 A set A of n dichotomies $x_1, x_2, \ldots, x_n$ is said to be linearly dependent of order r if in the vector space of real-valued functions defined on $A \subseteq X$, any subset of r of the dichotomies is linearly independent.

REMARK 3.5: When $r=n$, $x_1, x_2, \ldots, x_n$ are linearly independent.

THEOREM 3.7: A set of dichotomies defined on a logically independent set is linearly independent.

Proof: If a set A is logically independent, then A = X and hence the set of functions (2.23) defined on X is independent; in particular $x_1, x_2, \ldots, x_n$ are independent. Hence Theorem 3.7.

THEOREM 3.8. A set of dichotomies defined on a logically r-dependent set is linearly r-dependent.

Proof: If a set A is linearly r-dependent, then it contains at least 2^r points. Let V_A be the vector space of real-valued functions defined on A. The dimension of V_A is $\geq 2^r$. The set of real-valued functions on any r of the dichotomies is a subset of V_A. Since any set of r dichotomies is logically independent, any set of r dichotomies is linearly independent, in fact, any set of 2^r functions of the form (2.24) is linearly independent.

Hence Theorem 3.8.

3.4 Logical and Stochastic Dependence

DEFINITION 3.9: The order of stochastic dependence of a set A of dichotomies ($A \subseteq X$) with a probability distribution p(x) on it, is the co-order of p(x) extended to X by considering $p(x) = 0$ for all $x \in X - A$.

REMARK 3.6: The usual notion of stochastic independence is obtained when the order of stochastic dependence is equal to 0 or 1.

The following Theorems show that the order and co-order are related to the joint distribution function of subsets of the n dichotomies and we give explicit expressions in terms of the joint distributions. They follow easily from the steps of the proof of Theorem 2.7.

THEOREM 3.9. *If a probability distribution p(x) on X is of order r, then*

$$\frac{p(x)}{p^*(x)} = (-1)^{n-1} + \sum_{i=r+1}^{n-1} (-1)^i (i-1)$$

$$+(-1)^n \left[1 + \sum_{i=r+1}^{n-1} \binom{n}{i}\right] \sum_{\substack{i_1<i_2<\ldots<i_j \\ j \leq r}} \left[(-1)^j \pi_{i_1 i_2,\ldots i_j}(x)\right]. \tag{3.8}$$

THEOREM 3.10. If a probability distribution $p(x)$ on X is of co-order r, then

$$\frac{p(x)}{p^*(x)} = (-1)^n \left[(n-1) \sum_{\substack{i_1<i_2<\ldots<i_j \\ j > r}} (-1)^j \pi_{i_1 i_2,\ldots,i_j}(x)\right] \tag{3.9}$$

Theorem 3.11 below connects stochastic and logical dependence of order r.

LEMMA 3.1 (Kjellberg [19]): Logical independence of a system of dichotomies is a necessary condition for their stochastic independance.

Proof: Easily follows from Theorem 2.5.

THEOREM 3.11. If a system of dichotomies is stochastically r-dependent, then the order of its logical dependence is at least r.

Proof: Since all $\rho_{i_1 i_2,\ldots,i_j} = 0$ for $j \leq r$, all sets of r or less of dichotomies are stochastically independent by Theorem 3.3 or by (2.34.). Hence any set of r (or less) dichotomies is logically independent, by Lemma 3.1. Hence by Remark 3.2, the order of logical dependence is at least r.

The following Theorem brings out a feature of the probability under a certain kind of logical dependence, which in a sense is complementary to logical dependence of order r.

THEOREM 3.12: If a system of dichotomies under a probability distribution is such that any set of r of the dichotomies is logically dependent, then at a set of points in X, $p_g(x) = 0$ for all g for which $\geq r$ (including $p(x)$).

Proof: For a choice of r dichotomies, there are 2^r possible intersections, at least one of which has a zero probability. Let D_g be the set of dichotomies of the r_g variables in g. Let

$$d_g = (x^g_{i_1}, x^g_{i_2}, \ldots, x^g_{i_{r_g}}) \in D_g. \tag{3.10}$$

Then let $A_g^{d_g}$ be the intersection of the subsets one from each of the r_g partitions of X according to the value combination d_g, that is,

$$A_g^{d_g} = A_{i_1}^{r_1^g} \cap A_{i_2}^{r_2^g} \cap \dots \cap A_{i_{r_g}}^{r^g} . \qquad (3.11)$$

Let $D^*_{d_g}$ be the set of 2^{n-r_g} combinations of the $(n-r_g)$ dichotomies not contained in g, for a given set of values d_g for the r_g dichotomies in g. Each set A^{d_g} contains exactly 2^{n-r_g} of the 2^n dichotomous vectors, given by fixing the values of the r_g dichotomies $x^g_{i_1}, x^g_{i_2}, \dots, x^g_{i_{r_g}}$ and by taking all 2^{n-r_g} combinations of the remaining dichotomies. If this system of dichotomies is to be logically dependent, then at least 2^{n-r_g} points have to carry zero probability. So, if each set of r dichotomies is to be logically dependent, then such a condition is to be satisfied for any combination of r dichotomies. This implies that for any g with $r_g = r$, p(x) assigns zero probability to a set of points $M^*_{d_g}$ for a particular d_g. Hence at any point $x \in M^*_{d_g}$, $p(x) = 0$ and $p_g(x) = \sum_{x \in D^*_{d_g}} p(x) = 0$. Since all sets of

(r+1) dichotomies are also logically dependent in this case, $p_g(x) = 0$ for $r_g > r$.

Hence Theorem 3.12.

COROLLARY 3.1 If a system of dichotomies under a probability distribution of order r is such that any set of $(r+1)$ of the dichotomies is logically dependent, then at a set of points in X,

$$\sum_{\substack{i_1<i_2<\cdots<i_j \\ j\leq r}} [(-1)^j \pi_{i_1 i_2 \cdots i_j}(x)] = \text{constant.} \tag{3.12}$$

Proof: For $x \in M^*_{d_g}$ by Theorem 2.7,

$$0 = \pi(x) = (-1)^n (n-1) + (-1)^n \sum_{\substack{i_1<i_2<\cdots<i_j \\ j\leq r}} (-1)^j \pi_{i_1 i_2, \cdots i_j}(x). \tag{3.13}$$

Using Theorem 3.9, we get

$$\sum_{i=r+1}^{n-1} (-1)^i (i-1) + (-1)^n \left[\sum_{i=r+1}^{n} \binom{n}{i}\right] \sum_{\substack{i_1<i_2\cdots<i_j \\ j\leq r}} (-1)^j \pi_{i_1 i_2 \cdots i_j}(x) = 0. \tag{3.14}$$

Hence $\displaystyle\sum_{\substack{i_1<i_2<\cdots<i_j \\ j\leq r}} (-1)^j \pi_{i_1 i_2 \cdots i_j} = \text{const.}$

$$= \frac{\sum_{i=r+1}^{n-1} (-1)^i (i-1)}{(-1)^n \sum_{i=r+1}^{n-1} \binom{n}{i}} \quad \text{for all } x \in D^*_{d_g} \tag{3.15}$$

Hence Corollary 3.1.

REMARK 3.7: We have established in this Chapter that logical dependence implies linear dependence. The converse however is not true. For example, consider the set of four points $A = \{(0, 0, 1), (0, 1, 0), (1, 0, 0), (1, 1, 1)\}$ and the set of dichotomies on A, $f_1 = (1, 0, 0, 0)$, $f_2 = (0, 1, 0, 0)$, $f_3 = (0, 0, 1, 0)$, $f_4 = (0, 0, 0, 1)$ which are linearly 4-dependent; however, they are **not logically** 4-dependent.

REMARK 3.8: Similarly, logical dependence does not imply stochastic dependence. For, one can construct any number of probability distributions (say, with probability $\frac{1}{2}$ for $(0, 0, \ldots, 0)$ and, $\frac{1}{2(2^n - 1)}$ for the rest.) which make $x_1, x_2, \ldots, x_n$ stochastically dependent but logically independent.

~~(say, with probability $\frac{1}{2}$ for $(0, 0, \ldots 0)$ and $\frac{1}{2(2^n - 1)}$ for the rest.)~~

CHAPTER 4

PROPERTIES AND CHARACTERIZATIONS OF THRESHOLD ORDER

'The business of a poet, said Imlac, is to examine, not the individual, but the species, to remark general properties and large appearances. He does not number the streaks of the tulip'.

Samuel Johnson
'Rasselas'

4.0 Summary

Some simple properties of a threshold function of order r are presented. Characterizations of a threshold function of order r through Chow parameters, Chebyshev approximation and Ho-Kashyap algorithm are given, which lead to the solution of testing and realization problems. A comparison of these techniques is made with the Kaszerman's procedure of developing a non-linear surface for realizing a switching function and it is shown that our techniques are superior. Enumeration and tabulation of threshold function of order 3 are given. Some comments and problems are presented at the end.

4.1 Simple Properties

The theory of testing and realization of a threshold function of order r follows very closely the corresponding theory of threshold functions. This is achieved by taking the orthogonal functions, which are products of input variables. The only change that is needed for the case of a threshold function of order r is that the matrices, inequalities, etc., are to be augmented by these functions of order $\leq r$, that is, functions which are products of $\leq r$ input variables. A number of proofs of results of this Chapter are similar to the proofs for case $r = 1$; hence we do not present in this Chapter, proofs of such theorems, but give appropriate references, where the case $r = 1$ is discussed.

Since the two sets of values $\{0, 1\}$ and $\{-1, 1\}$ are used for inputs and outputs, in different contexts, we present in Theorem 4.1 below the relation between the weight vectors of a threshold function of order r for these two cases.

Towards this, we describe a matrix A_n as follows:

$$B_1 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} ; \qquad B_{i+1} = \begin{bmatrix} B_i & 0 \\ -B_i & B_i \end{bmatrix} \qquad (4.1)$$

Each row of the matrix B_n corresponds to an $i_1 i_2 \dots i_j$, which can be obtained from the order set out in (2.30). Then multiply all elements of this row by 2^j to get matrix A_n. The row corresponding to $i_1 i_2 \dots i_j$ contains zero after the t_j-th term. Let $\tilde{a}_{i_1 i_2 \dots i_j}$ denote this vector of t_j elements.

THEOREM 4.1: Let f be a threshold function of order r realized by a set of t_r weights denoted by a vector $w_{i_1 i_2 \dots i_j}$. Then with the truth values ± 1, F is a threshold function of order $\leq r$ realized by the weights

$$b_{i_1 i_2 \dots i_j} = a_{i_1 i_2 \dots i_j} \; w_{i_1 i_2 \dots i_j} \tag{4.2}$$

$$i_1 < i_2 < \dots < i_j, \quad 0 \leq j \leq r.$$

Proof: By (2.31), the matrix that represents the linear transformation from (2.30) to similar set of functions in y, is A_n. Hence the function

$$\sum_{j=0}^{r} \sum_{i_1 < i_2 < \dots < i_j} w_{i_1 i_2 \dots i_j} \, x_{i_1} x_{i_2} \dots x_{i_j}$$

$$= \Sigma \, \Sigma \, \tilde{a}_{i_1 i_2 \dots i_j} \, \tilde{w}_{i_1 i_2 \dots i_j} \, y_{i_1} y_{i_2} \dots y_{i_j}$$

$$= \Sigma \, \Sigma \, b_{i_1 i_2 \dots i_j} \, y_{i_1} y_{i_2} \dots y_{i_j}. \tag{4.3}$$

Hence Theorem 4.1.

<u>EXAMPLE 4.1</u>: The function $x_1 x_2 + \bar{x}_2 x_3$, a threshold function of second order is realized by the weights

$$\begin{Bmatrix} -2 & 0, & 0, & 4; & 4, & -4, & 0 \\ w_0 ; & w_1, & w_2, & w_3; & w_{12}, & w_{23}, & w_{31} \end{Bmatrix}$$

The relevant rows of the matrix A_3 in this case are:

$$\begin{array}{lcl}
\tilde{a}_0 : & 1 & (1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) \\
\tilde{a}_1 : & \frac{1}{2} & (1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) \\
\tilde{a}_2 : & \frac{1}{2} & (1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0) \\
\tilde{a}_{12} : & \frac{1}{4} & (1\ \ 1\ \ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 0) \\
\tilde{a}_3 : & \frac{1}{2} & (1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0) \\
\tilde{a}_{23} : & \frac{1}{4} & (1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 0) \\
\tilde{a}_{31} : & \frac{1}{4} & (1\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0) .
\end{array}$$

The b coefficients are hence

$$(0;\ \ 1,\ \ 0,\ \ 1;\ \ 1,\ \ -1,\ \ 0).$$

Let $v^{(r)}$ denote a vector of t_r functions

$$x_{i_1} x_{i_2} \cdots x_{i_j} , \quad i_1 < i_2 < \cdots < i_j,\ 0 \leq j \leq r,$$

(for $j = 0$, this is 1). Let $V_1(f)$ be the set of true vertices of a switching function f and $V_0(f)$ the set of false vertices.

Then we have the following results as for case $r = 1$.

1. A threshold function of order r can be realized with integral weights. That is, a switching function $f(x_1, x_2, \ldots, x_n)$ is a threshold function of order r, if and only if there exist t_r integers,

$$N_o, N_{i_1}, N_{i_1 i_2}, \ldots, N_{i_1 i_2 \ldots, i_r} \tag{4.4}$$

$$i_1 < i_2 < \ldots < i_r,$$

(denoted by a t_r-vector $N^{(r)}$),

such that on the true vertices $v_{A1}^{(r)}, v_{A2}^{(r)}, \ldots, v_{An_a}^{(r)}$

$$N^{(r)} v_{Ai}^{(r)} \leq 1, \quad i = 1, 2, \ldots, n_a, \tag{4.5}$$

and on the false vertices $v_{B1}^{(r)}, v_{B2}^{(r)}, \ldots, v_{Bn_b}^{(r)}$

$$N^{(r)} v_{Bj}^{(r)} \geq 0, \quad j = 1, 2, \ldots, n_b, \tag{4.6}$$

$$n_n + n_b = 2^n.$$

2. The threshold order is an intrinsic property of the switching function and thus does not depend upon the choice of truth values for the inputs and outputs. Any pair of numbers can be chosen for inputs and outputs and the function retains its threshold order.

Theorem 4.1 gives a method of obtaining weights from (0, 1) system to (-1, +1) system.

3. The dual and complement of a threshold function of order r are also of threshold order r.

4. In fact, all members of a symmetry class of switching functions, obtained by permutation of variables, complementation of variables, and complementation of functions, are of the same threshold order.

5. A switching function is of threshold order r, if and only if for any 2^n non-negative numbers $c_i \geq 0$ $(1 \leq i \leq 2^n)$, the relations

$$\sum_{i=1}^{n_a} c_i = \sum_{i=n_a+1}^{2^n} c_i, \tag{4.7}$$

and

$$\sum_{i=1}^{n_a} c_i v_i^{(r)} = \sum_{i=n_a+1}^{2^n} c_i v_i^{(r)}, \tag{4.8}$$

imply $c_i = 0$ for all $i = 1,2,\ldots, 2^n$.

6. Let us define t_r parameters of a switching function f as follows:

$$a^{(r)}(f) = (a_{i_1 i_2 \dots i_j}(f)), \quad i_1 < i_2 < \dots < i_j$$

$$0 \le j \le r$$

where $a_{i_1 i_2 \dots i_j}(f) = \sum_{v \in V_1(f)} v^{(r)}_{i_1 i_2 \dots i_j}(f)$. (4.9)

6.1. Let f and g be switching functions of n variables. If

$$a^{(r)}(f) = a^{(r)}(g),$$

then either both f and g are of threshold order r or both are not.

6.2. Let f and g be switching functions of n variables, with $a^{(r)}(f) = a^{(r)}(g)$. If f is of threshold order r, then $g = f$.

6.3. <u>Corollary</u>: Let f and g be two distinct switching functions of n variables. If $a^{(r)}(f) = a^{(r)}(g)$, then f is of threshold order $\geq r$.

6.4. <u>Corollary</u>: Let f be a threshold function of order r. If $a_{i_1 i_2 \dots i_j}$ is the same for all $(i_1, i_2, \dots, i_j)$ and if this is true for all $j = 1, 2, \dots, r$, then f is a symmetric function.

7. If f is a threshold function of order r, then

$$f + x_{i_1} x_{i_2}, \dots, x_{i_j}, \quad 0 \le j \le r, \quad i_1 < i_2 < \dots < i_j,$$

is a threshold function of order r.

8. Let n_a input combinations yield a value 1 to a switching function f and n_b combinations yield a value 0 ($n_a + n_b$ not necessarily 2^n, that is, 'don't care' outputs are allowed). Correspondingly let the $n_a \times t_r(n)$ matrix of upto rth order product functions, be called α - matrix and $n_b \times t_r(n)$ matrix be called β-matrix. Adding columns of ones and zeros to each of these matrices, complementing β matrix and adding to α matrix results in a matrix, say α^* Then the following is a generalization of Akers' [1] result:

A function f $(x_1 \; x_2, \ldots, x_n)$ with $[\alpha^*_{ik}]$, i = 1,2 i = 1,2,..., t_r, $k = n_a + n_b$ is a threshold function of order r, if and only if $[\alpha^*_{ik}]$ when solved as a two-person zero-sum game has a value > 1/2.

This game theoretic solution is a variation of the linear programming solution, Minnick [27], Muroga [31].

4.2 Boundary Matrix

The theory of eliminating redundancy in inequalities and of partial specification of a truth table in case of threshold functions has been neatly formulated by Mays [24], [25], using the concept of a boundary matrix. In this Section, we note that the concept of a boundary matrix and these results of Mays, hold for threshold function of order r with suitable definitions.

DEFINITION 4.1. (Mays [24], [25]): A matrix M is called a boundary matrix with respect to a matrix A if it has the following properties:

1. The rows of M are taken from the rows of A.
2. $M \cdot g > 0$ implies $A \cdot g > 0$, for some vector g.
3. The number of rows of M is minimum consistent with the first two conditions.

DEFINITION 4.2: A matrix M is called the boundary matrix of a threshold function of order r, if it is a boundary matrix of the matrix A, defined with respect to the t_r functions,

$$x_{i_1} \; x_{i_2} \cdots x_{i_j} \tag{4.10}$$

$$i_1 < i_2 < \cdots < i_j, \quad 0 \leq j \leq r.$$

Then the following results are obtained by a proof similar to that of **Mays** for $r = 1$.

THEOREM 4.2. There exists a unique boundary matrix for any matrix A representing a threshold function of order r.

THEOREM 4.3: The rank of the boundary matrix is equal to the rank of A.

The number of rows in the boundary matrix of a threshold function of order r is at least t_r. It is possible that the specification of t_r rows of a truth table, completely specifies a threshold function of order r.

4.3 Chebyshev Approximation

An alternative to the realization of threshold function by solving inequalities using linear programming is the Chebyshev approximation, as outlined by Kaplan and Winder [17]. We give, in this Section, appropriate definitions from which results similar to Kaplan and Winder follow for the case of threshold functions of order r.

DEFINITION 4.3: The approximation of order r of a switching function F(y) is defined to be the one obtained using the Coleman coefficients with $\leq$ r subscripts, that is, using the first t_r subscripts in the Coleman expansion.

We have shown that these coefficients can be computed with the help of the measures of suitable switching functions on Y and on its subsets.

DEFINITION 4.4: The approximation of order r of a switching function defined by

$$\sum_{j=0}^{r} \sum_{i_1<i_2<\dots<i_j} d_{i_1 i_2 \dots, i_j} y_{i_1} y_{i_2} \dots, y_{i_j} \tag{4.11}$$

(denoted in vector form by d.s)

is said to be <u>best in the Chebyshev sense</u> (<u>C-best of order r</u>), if

$$\max_{y} |f(y) - d.s.| \qquad (4.12)$$

is minimum.

<u>THEOREM 4.4</u>: (1) <u>A switching function F is a threshold function of order r, if and only if, F has approximation of order r, with maximum deviation less than one.</u>

(2) <u>F is a threshold function of order r, if and only if, F is realized by its own C-best approximation of order r.</u>

This enables the following classification of switching functions, by means of their threshold orders.

<u>THEOREM 4.5.</u> <u>Any switching function of n variables is a threshold function of order</u> $r \leq n$.

<u>Proof</u>: Since any switching function has an exact representation of order at most n, by (2) of Theorem 4.4, F is a threshold function of order at most n.

Hence Theorem 4.5.

<u>EXAMPLE 4.3</u>: Consider the 3-input function of Example 4.1, namely $x_1x_2 + \bar{x}_2 x_3$. The first order Coleman coefficients are

$$\frac{1}{8}\left\{0;\ 4,\ 0,\ 4\right\},$$

and the first and second order Coleman coefficients are

$$\frac{1}{8}\left\{ 0;\ 4,\ 0,\ 4;\ \ 4,\ -4,\ 0 \right\}$$

The following Table shows the realization by first and second order weights, denoted by $T_1(x)$ and $T_2(x)$ respectively.

Table 4.1

Example of Chebyshev approximation

y	F(y)	$T_1(y)$	$T_2(y)$
(-1, -1, -1)	-1	-1	-1
(-1, -1, 1)	1	0	1
(-1, 1, -1)	-1	-1	-1
(-1, 1, 1)	-1	0	-1
(1, -1, -1)	-1	0	-1
(1, -1, 1)	1	1	1
(1, 1, -1)	1	0	1
(1, 1, 1)	1	1	1

It is easy to see that T_1 does not realize F, whereas T_2 does. Hence the threshold order of F is 2.

4.4 Ho-Kashyap Algorithm

Ho and Kashyap [15] recently presented an algorithm for linear inequalities and applied it to the problem of threshold realization of a switching function. We show in this Section that their algorithm is applicable to the problem of realization of a switching function of threshold order r.

Consider a vector $u^{(r)}$ of t_r functions

$$y_{i_1} y_{i_2} \cdots y_{i_j}, \quad i_1 < i_2 < \cdots < i_j, \quad 0 \leq j \leq r. \tag{4.13}$$

Let $F(y_1, y_2, \ldots, y_n)$ be a switching function with true vertices

$$u_{Ai}^{(r)}, \quad i = 1, 2, \ldots, n_a, \tag{4.14}$$

and false vertices

$$u_{Bj}^{(r)}, \quad j = 1, 2, \ldots, n_b. \tag{4.15}$$

The problem is to find a t_r-vector α such that

$$\alpha^T u_{Ai}^{(r)} > 0, \quad i = 1, 2, \ldots, n_a, \tag{4.16}$$

$$\alpha^T u_{Bj}^{(r)} < 0, \quad j = 1, 2, \ldots, n_b, \tag{4.17}$$

or writing,

$$A = \begin{bmatrix} u_{A1}^{(r)} \\ u_{A2}^{(r)} \\ \vdots \\ u_{An_a}^{(r)} \\ -u_{B1}^{(r)} \\ -u_{B2}^{(r)} \\ \vdots \\ -u_{Bn_b}^{(r)} \end{bmatrix} \tag{4.18}$$

we have (4.16) and (4.17) as

$$\underset{2^n \times t_r}{A} \;\; \underset{t_r \times 1}{\alpha} > 0 \tag{4.19}$$

We state here the

<u>Ho Kashyap Algorithm [15]</u>:

(1) <u>Problem:</u> Find α(m-vector) and β (N-vector) such that

$$\left.\begin{array}{l} A\,\alpha = \beta \\ \beta > 0 \end{array}\right\}, \tag{4.20}$$

or

find α and $\beta > 0$, such that

$$J = \|A\,\alpha - \beta\|^2 \tag{4.21}$$

is minimized.

(2) Algorithm: Denoting by $A^{\#}$ the generalized inverse of A ,

$$\left.\begin{aligned} \alpha(0) &= A^{\#} \beta(0), \ \beta(0) > 0, \quad \text{otherwise arbitrary} \\ y(i) &= A\,\alpha(i) - \beta(i) \\ \alpha(i+1) &= \alpha(i) + \varrho\, A^{\#} [y(i) + |y(i)|] \\ \beta(i+1) &= \beta(i) + \varrho [y(i) + |y(i)|] . \end{aligned}\right\} \tag{4.22}$$

This algorithm converges exponentially to the solution, if one exists, in a finite number of steps. It also tests the consistency of the set of inequalities.

In our case, $m = t_r$ and $N = 2^n$.

LEMMA 4.1: $A^T A = (2^n)\ I.$ (4.23)

(I : unit matrix)

Proof: By Theorem 2.4, the set of functions (2.13) is orthonormal under the inner product

$$2^{-n} \sum_{x \in X} f(x)\ g(x) \tag{4.24}$$

and hence all the sum of squares of columns of A are 2^n and the sum of products of different columns are zero. Hence Lemma 4.1.

Thus $A^{\#} = \dfrac{A^T}{2^n}$. (4.25)

Hence the following algorithm.

THEOREM 4.6. The following is an algorithm for testing and realization of a threshold function of order r:

$$\left.\begin{aligned} \alpha(0) &= \frac{1}{2^n} A^T \beta(0), \quad \beta(0) > 0, \text{ otherwise arbitrary} \\ y(i) &= A\,\alpha(i) - \beta(i), \\ \alpha(i+1) &= \alpha(i) + \rho A^T[y(i) + |y(i)|], \quad 0 < \rho \leq 1 \\ \beta(i+1) &= \beta(i) + \rho\,[y(i) + |y(i)|]. \end{aligned}\right\} \quad (4.26)$$

The non-positivity of the vector $y(i)$ at any stage of the iteration shows that the function is realizable by a threshold function of order r. The algorithm ends when $\alpha(i+1) = \alpha(i)$ and the $\alpha(i)$ at this stage gives the realizing weights.

If we choose $\beta^T(0) = [1, 1, \ldots, 1]$, then as in the case $r = 1$, the vector differences between mean vectors of u_A and u_B is the $\alpha(0)$, which is the same as the Chow parameters discussed in Section 2.5.

EXAMPLE 4.4: We shall take the 3-input function $x_1x_2 + \bar{x}_2 x_3$, considered in Example 4.1.

$$A = \{1, 5, 6, 7\}, \qquad n_a = 4$$

$$B = \{0, 2, 3, 4\}, \qquad n_b = 4.$$

Applying Ho-Kashyap algorithm for $r = 1$, we find that the function is not of threshold order 1, as follows:

$$\beta^T(0) = \{1, 1, 1, 1, 1, 1, 1, 1\},$$

$$A = \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$$

$$\alpha^T(0) = \frac{1}{8}\{0, -2, 0, 2\}.$$

$$y^T(0) = \frac{1}{8}(4, 0, -4, 0, -4, 0, -4, 0) - (1,1,1,1,1,1,1,1)$$

$$= \left(-\frac{1}{2}, -1, -\frac{3}{2}, -1, -\frac{3}{2}, -1, -\frac{3}{2}, 0\right).$$

The non-positivity of $y(0)$ shows that the function is not a threshold function.

Then, considering the case $r = 2$, we have

Then, considering the case $r = 2$, we have

$$A = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ -1 & -1 & 1 & 1 & 1 & -1 & 1 \end{bmatrix},$$

$$\alpha^T(0) = \frac{1}{8}\,(0,\ -2,\ 0,\ 2,\ 2,\ -2,\ 0).$$

$$y^T(0) = \frac{1}{8}\,(8,0,0,0,0,0,0,8) - (1,1,1,1,1,1,1,1)$$

$$= (0,\ -1,\ -1,\ -1,\ -1,\ -1,\ -1,\ 0).$$

$$|y^T(0)| = (0,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 0)$$

$$y^T(0) + |y^T(0)| = (0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0).$$

Hence $\alpha(1) = \alpha(0)$.

Thus the function is of threshold order 2 and it is realized by the inputs

$$\frac{1}{8}\,(0;\ -2,\ 0,\ 2;\ 2,\ -2,\ 0).$$

4.5 Kaszerman's Model

Kaszerman [18] has given the following procedure to generate a surface for a non-linear threshold device for a switching function. We show in this Section that our concept of threshold order and its characterizations have resulted in an optimality in non-linear realization, that is not obtained by Kaszerman's procedure.

Kaszerman's Procedure:

1. Express the switching function as a sum of products or reduce it to minimum sum form.

2. For each term write an equation of the form

$$\alpha_k = m - \sum_{i=1}^{m} x_i + \sum_{j=1}^{\ell} x_j, \quad (m + \ell \leq n)$$

where x_i's are the m uncomplemented variables and x_j's are the complemented variables.

3. Then

$$\varphi = - \prod_k \alpha_k \tag{4.27}$$

gives the surface.

EXAMPLE 4.5: Consider $f = x_1 x_2 + \bar{x}_1 \bar{x}_2$. By Kaszerman's procedure, the second-order surface is obtained as

$$\Phi = - x_1^2 - x_2^2 + 2x_1 x_2 + \frac{1}{2} \ .$$

The Coleman coefficients are

$$(0,\ 0,\ 0,\ 4)$$

that is, the surface is

$$y_1y_2 = 4x_1x_2 - x_1 - 2x_2 + 1$$

$$\text{or} \quad -x_1 - x_2 + 2x_1x_2 + \frac{1}{2}\ ,$$

which is the same as Φ since $x_i^2 = x_i$. Hence these two procedures coincide.

EXAMPLE 4.6: This example is intended to show that even though sometimes the orders of the realizing functions are the same by the two procedures, our procedure leads to a better realization.

Consider the 3-input function $x_1x_2 + \bar{x}_2x_3$.
For this

$$\alpha_1 = 2 - x_1 - x_2,$$

$$\alpha_2 = 1 + x_2 - x_3.$$

Thus

$$\Phi_1 = -(2 - x_1 - x_2)(1 + x_2 - x_3)$$

$$= -2 + x_1 + 2x_3 + x_1x_2 - x_2x_3 - x_3x_1.$$

The Coleman coefficients for this function are

$$\frac{1}{8}\left\{-2;\ 1,\ 0,\ 2;\ 1,\ -1,\ -1;\ 0\right\},$$

which lead to the coefficients of x functions as

$$\frac{1}{8}\left\{-1;\ 0,\ 0,\ 2;\ 1;\ -1,\ 0;\ 0\right\},$$

that is, the surface is,

$$\Phi_2 = -1 + 2x_3 + 2x_1x_2 - 2x_2x_3\,,$$

which is much simpler, for realizing, than ϕ_1, since it contains less number of input functions and smaller weights.

EXAMPLE 4.7. Now we give an example in which Kaszerman's procedure constructs a surface of order higher than the threshold order of a switching function.

Consider the switching function of 3-inputs,

$$\bar{x}_1\ \bar{x}_2\ \bar{x}_3 + x_1x_2 + \bar{x}_2x_3.$$

Kaszerman's surface is

$$\Phi_1 = -(2 - x_1 - x_2)(1 + x_2 - x_3)(3 - x_1 - x_2 - x_3)$$
$$= -6+4x_1+2x_2+6x_3-3x_2x_3-4x_3x_1 + x_1x_2x_3,$$

which is a third order surface.

However, the Coleman coefficients are

$$\frac{1}{4}\ (1;\ 1,\ -1,\ 1;\ 3,\ -1,\ 1;\ 1)$$

and the first and second order coefficients realize the function as shown below.

Table 4.2

A function of Threshold order 2,

of 3-inputs

x	F	Realization
(0, 0, 0)	1	3/8
(0, 0, 1)	1	5/8
(0, 1, 0)	0	-3/8
(0, 1, 1)	0	-5/8
(1, 0, 0)	0	-3/8
(1, 0, 1)	1	3/8
(1, 1, 0)	1	3/8
(1, 1, 1)	1	5/8

The function is a second order function and is realized by

$$(1;\ 1,\ -1,\ 1;\ 3,\ -1,\ 1),$$

which is considerably simpler than ϕ_1.

Since we have shown that our procedure uses the threshold order and is also equivalent to the programming approach of minimizing the total weight, it is clear that there are no cases in which the realizing function by our procedure will have greater weights or of greater order of surface.

4.6 Miscellaneous Comments and Problems

So far, we have confined our attention to only such problems of realization as are solved by the solution of a set of inequalities and variations of this solution such as linear programming, game theory and Chebyshev approximation. The generalization to an case of order r was achieved by suitably taking the orthogonal functions. There are many other aspects of threshold logic, the generalizations of which to the case of order r are equally, if not more, interesting where this technique is not of use. We have not been able to get the mathematical framework that would make this generalization easy. For instance, the most elegant necessary condition for linear separability, namely, unateness [26], should have correspondence to something like unateness of order r, one way of defining which would be as follows:

DEFINITION 4.5: A switching function is said to be unate in $(x_{i_1}, x_{i_2}, \ldots, x_{i_r})$, $(r \leq n)$, if there is a normal expansion in which whenever all these r variables occur together in a term they occurs in only one of the 2^r possible products of uncomplemented and complemented variables.

EXAMPLE 4.8: The function of four variables

$$x_1 \bar{x}_2 x_3 x_4 + x_1 \bar{x}_2 \bar{x}_3 \bar{x}_4$$

is unate in (x_1, x_2) whereas

$$x_1 x_2 x_3 x_4 + x_1 \bar{x}_2 x_3 \bar{x}_4$$

is not unate in (x_1, x_2).

<u>DEFINITION 4.6</u>: A switching function is said to be <u>unate of order r</u> if it is unate in every set of $\binom{n}{r}$ combinations of r variables out of n.

<u>EXAMPLE 4.9</u>: The function

$$x_1 \bar{x}_2 x_3 + x_4 + x_1 \bar{x}_2 \bar{x}_3 \bar{x}_4$$

is not unate of order 2 and

$$x_1 \bar{x}_2 x_3 + \bar{x}_2 x_3 + \bar{x}_1 x_4$$

is unate of order 2.

It is to be noted that a function which is unate of order r need not be unate of order < r and a function unate of order r is unate of order r+1.

<u>EXAMPLE 4.10</u>: The function

$$x_1 \bar{x}_2 x_3 + \bar{x}_2 x_3 + \bar{x}_1 x_4$$

is unate of order 2 but not of order 1.

Then we have a problem which is a generalization of the celebrated result of McNaughton [26]:

PROBLEM 4.1: Is a threshold function of order r necessarily unate of order r?

An alternative way to derive results for threshold functions of order r, using the orthogonal functions, appears to be the use of the results of a threshold function to the t_{r-1} inputs obtained by products upto order r. These two approaches are different because in the former case, the problem is a complete specification of outputs on 2^n input-combinations and in the latter case it is a partially specified one of 2^n input combinations out of $2^{t_r - 1}$ being specified. It may be possible to solve some problems of order r by viewing this as 'don't care' at the other $2^{t_r - 1} - 2^n$ points.

The theory is very similar to the theory of multiplexing of inputs introduced by von Neumann [33], where with n inputs and k multiples, in the space of 2^{nk} inputs, the switching function is defined only on 2^n points. On the other hand, multiplexing schemes for threshold elements of order r may be developed on lines similar to Pierce [36] and Liu and Liu [21].

One of the greatest advantages of threshold gates is that a number of switching functions can be expressed as a combination of simple threshold gates. In a similar fashion it may be possible to realize a switching function more economically by combination of gates of order one or more. In this connection, we feel that the ease with which a switching function is realized as a combination of threshold gates, should have something to do with its threshold order. It is clear that any switching function is realized by threshold gates at two levels since any switching function can be written in terms of AND/OR gates. But, in general, such a realization needs many inputs at the second level. The problem would be made simpler if the following problem is solved.

PROBLEM 6.2: Is it possible to realize any threshold function of order r of n inputs, at two levels, with n or less inputs in the first level and r or less inputs in the second level ?

There is an interesting generalization of threshold function by Ercoli and Mercurio[10] that is different from ours.

DEFINITION 4.7 [10]: A switching function of n inputs $x_1, x_2, \ldots, x_n$ is said to be a function with m thresholds $t_1, t_2, \ldots, t_m$, if there exist real numbers $w_1, w_2, \ldots, w_n$ such that (defining $t_0 = -\infty$, $t_{m+1} = +\infty$),

$$t_{i-1} < \sum_{i=1}^{n} w_i x_i \leq t_i, \quad i = 1,2,\ldots, m+1. \qquad (4.28)$$

It is immediately clear that any switching function has a maximum of 2^n thresholds. Also, a threshold function has a single threshold. This definition still retains the linearity of the inputs and hence from the point of view of circuitry is simpler than a threshold function of order > 1. So then a problem is to find out if it is possible to realize arbitrary switching functions with less than 2^n thresholds and to find the number of thresholds. Since any switching function is completely characterized by 2^n parameters and n parameters are already chosen as the weights it appears that it is possible to restrict the number of thresholds to $2^n - n$. For the same reason, it also appears that it is impossible to realize arbitrary switching functions with just n thresholds. A piece of interesting work in this direction has been done by Ercoli and Mercurio [10]. Then a problem to be investigated is the connection between the threshold order of a switching function and the number of thresholds of a switching function.

Owing to the similarity of the network of a threshold function of order r to the network of Gose [12] for any real-valued function of inputs, this network can be made adaptive in a manner analogous to Gose's.

4.7 Enumeration and Tabulation

Table-look-up has been found to be one simple way of testing and realization of threshold functions. Such tables of threshold functions have been computed upto 7 inputs. Muroga, Toda and Kondo [30] present tables upto six inputs and Winder [41], [42] presents tables of n inputs.

These authors have utilized a number of properties of threshold functions to simplify the computations and presentation of such tables. In this section, we present an enumeration and tabulation of switching functions of three inputs by their threshold order.

The only simplifying idea that we have used is the symmetry type of switching function introduced by Golomb [11]. Since functions belonging to the same symmetry class have the same threshold order, we have

(1) classified the 256 functions into 14 symmetry classes,

(2) computed the parameters for one function of each class;

(3) found the threshold order of this function. This helps us to enumerate the number of functions with each threshold order. The results are presented below.

Denoting a switching function by the vertices with the output 1 we have the following table.

Table 4.3

Table of Threshold Functions of order r.

Class	Typical functions	No. of functions	d-parameter	Threshold order
I	Σ	2	(-8 0, 0, 0, 0, 0, 0, 0)	0
II	Σ_0	16	(-6,-2,-2,-2,-2,-2,-2,-2)	1
III	$\Sigma_{0,7}$	8	(-4, 0, 0, 0, 4, 4, 4, 0)	2
IV	$\Sigma_{0,4}$	24	(-4, 0,-4,-4, 0, 4, 0, 0)	1
V	$\Sigma_{4,7}$	24	(-4, 4, 0, 0, 0, 4, 0, 4)	2
VI	$\Sigma_{0,4,7}$	48	(-2, 2,-7,-2,-2, 2, 2,-2)	2
VII	$\Sigma_{0,4,5}$	48	(-2, 2,-6,-2,-2, 2, 2,-2)	1
VIII	$\Sigma_{0,3,5}$	16	(-2,-2,-2, 2,-2, 2, 2,-6)	2
IX	$\Sigma_{0,3,4,7}$	6	(0, 0, 0, 0, 0, 8, 0, 0)	2
X	$\Sigma_{0,4,5,7}$	24	(0, 4,-4, 0, 0, 4, 4, 0)	2
XI	$\Sigma_{0,1,2,3}$	6	(0,-8, 0, 0, 0, 0, 0, 0)	1
XII	$\Sigma_{0,1,2,7}$	24	(0,-4, 0, 0, 4, 0, 4, 4)	2
XIII	$\Sigma_{0,1,2,4}$	8	(0,-4,-4,-4, 0, 0, 0, 4)	1
XIV	$\Sigma_{1,2,4,7}$	2	(0, 0, 0, 0, 0, 0, 0, 8)	3
	Total	256		

This yields the following table.

Table 4.4

Enumeration of Threshold Function of 3-inputs

Threshold order	No. of functions
0	2
1	102
2	150
3	2
Total	256

CHAPTER 5

APPLICATIONS OF THRESHOLD ORDER TO DEPENDENT INPUTS

'The bearings of this observation lays in the application on it.'

Bunsby

'Dombey and Son'

5.0 Summary

In this chapter, we consider the application of a threshold gate of order r to such problems as pattern recognition and decoding. We show that the recognition with statistical dependence or order r (different from the concept in Chapter 3) requires a threshold gate of order at most r. We also show that a threshold gate of order r can be used as a decoder in the case of dependent noise of order r.

5.1 Introduction

Despite the wide variety of applications of threshold gates, particularly in problems of a stochastic nature, such as pattern recognition, coding and decoding networks, and, probability transformers, a major disadvantage of a threshold gate has been that it can be generally used only in the presence of independent noise. This is owing to the fact that threshold functions form a very small proportion of switching functions. In Chapter 1, we have elaborated this point already.

Hence there is a need to find logic gates more complicated than a threshold gate, which can handle dependent noise but the complications of which should be minimal in the sense that with less and less dependence of noise, the gate should be less and less complicated. We show here that this is precisely achieved using our notion of threshold order, by defining the order of stochastic dependence of inputs and by relating the two.

The results obtained in this Chapter are generalizations of the results of Chow [7] and of Massey [22], with regard to statistical recognition and threshold decoding respectively. The relaxation of the assumption of independent distributions is achieved by an application of the results of Bahadur [2], in defining the order of stochastic dependence. Chow's 'equivalence' between threshold functions and independent noise is then generalized to that between threshold functions of order r and statistical recognition with noise of order r.

The term 'equivalence' is to be interpreted as follows: If binary n-tuples are subject to a certain independent noise, they can afterwards be best classified (given certain conditional statistics) by a threshold gate. As we shall show below this 'equivalence' is not to be interpreted to mean that threshold gates cannot be used to discriminate in the presence of non-independent noise, nor, in a more general way, that 'noise of order r' requires threshold gates of order r, for classification; it means that a noise of order r requires a threshold gate of order at most r.

It appears that the possibility of the use of the results of Bahadur in relaxing the assumptions of independent components made by Winder and Chow is suggested in the recent book by Nilsson [30 p.62]. However, he does not appear to have formulated the concept of the order of a probability distribution.

Braverman [3] reviews theories of pattern recognition and Sheng [37] applies threshold gate to probability transformers.

5.2 Definitions

By Theorem 2.4, for a probability distribution $p(x)$, which is non-zero at every point $x \in X$, log $p(x)$ can be uniquely expressed as a linear combination of the set of functions S, and let us denote the coefficients by λ with corresponding suffixes.

DEFINITION 5.1: $\lambda_{i_1 i_2 \cdots i_r}$ is called a log-correlation parameter of order r, with reference to S.

DEFINITION 5.2: A distribution on the set X of n-dimensional binary vectors $x = (x_1, x_2, \ldots, x_n)$ is said to be of order r $(r \leq n)$, if all its log-correlation parameters of order $> r$ are zero and at least one of the log-correlation parameters of order r is not zero.

REMARK 5.1. Theorem 2.4 establishes the existence (and uniqueness) of the order of any probability distribution on X and that the order is $\leq n$.

REMARK 5.2. One might think that the natural way to formulate this definition is in terms of correlation parameters. Our object in using log-correlation parameters, is, of course, to retain the correspondence between dependence of order r and threshold function of order r. However, this definition may not appear unnatural to statisticians, who would realize the dominant role played by log density throughout statistical theory. As observed by Bahadur [2], independence is dependence of order 0, in which case this definition is clearly the same in terms of density or log density, that is, in terms of correlation parameters or log-correlation parameters.

5.3 Stochastic Dependence and Threshold Order

Consider an alphabet with two characters a_1 and a_2 each represented by a binary vector $x = (x_1, x_2, \ldots, x_n)$. Being subject to noise, x is a random variable on the sample space X of 2^n vectors of the n-dimensional cube. The probability distribution of x depends on the alphabet a_k and we denote by $P(x|a_k)$ the conditional probability distribution of x given a_k, and by p_k the a priori probability that character a_k occurs, $k = 1,2$. A recognition rule of Chow [5], [7], based on statistical decision approach is to identify pattern x as character a_1, if

$$R(x) = 1$$

and as character a_2, if

$$R(x) = 0$$

where

$$R(x) = \begin{cases} 1 & \text{if} \quad p_1 P(x|a_1) \geq p_2 \, P(x|a_2) \\ 0 & \text{if} \quad p_1 P(x|a_1) < p_2 P(x|a_2). \end{cases} \tag{5.1}$$

The following is a generalization of Chow's theorem [7].

THEOREM 5.1: A switching function $f(x_1, x_2, \ldots, x_n)$ is of threshold order r, if and only if, there exist a binomial distribution $p = (p_1, p_2)$ and two conditional probability distributions $P(x|a_1)$ or order r_1 and $P(x|a_2)$ of order r_2, neither of which vanishes at any point $x \in X$, such that max. $(r_1, r_2) = r$ and that their associated recognition function is $f(x_1, x_2, \ldots, x_n)$.

Proof: (i) 'If' part:

$$\text{Let} \quad \beta_i^{(k)} = P(x_i = 0|a_k), \; \alpha_i^{(k)} = 1 - \beta_i^{(k)} , \tag{5.2}$$

$$i = 1, 2, \ldots, n, \quad k = 1,2.$$

and

$$P^*(x|a_k) = \prod_{i=1}^{n} (\alpha_i^{(k)})^{x_i} (1 - \alpha_i^{(k)})^{1-x_i}. \tag{5.3}$$

Consider

$$\log\ [P(x|a_k)/P^*(x|a_k)]_{k=1,2}. \tag{5.4}$$

This can be written uniquely as a linear combination of 2^n functions S of (2.13). These 'z' functions are again linear combinations of the corresponding 'x' functions. By Theorem 2.6, expression (5.4) can be written uniquely as a linear combination of functions T of (2.23). Let the coefficients be

$$B_k = \left\{ b_0^{(k)};\ b_1^{(k)},\ b_2^{(k)},\ldots,\ b_n^{(k)};\ b_{12}^{(k)}, b_{23}^{(k)},\ldots;\ b_{123}^{(k)},\ \ldots\ b_{12\ldots n}^{(k)} \right\} \tag{5.5}$$

$$k = 1,2.$$

Hence if (5.4) is of order r_k, then in (5.5) all b's with more than r_k suffixes vanish. The rest of the proof consists in taking log on the right side of (5.1) using this fact, and, in verifying that $R(x)$ is of threshold order r, $r = \max. (r_1, r_2)$. Thus $\log\ [p_k P(x|a_k)]$ is equal to

$$\log p_k + \sum_{i=1}^{n} \log \beta_i^{(k)} + \sum_{i=1}^{n} x_i \log \frac{1-\beta_i^{(k)}}{p_i^{(k)}} + b_0^{(k)} +$$

$$+ \sum_{i=1}^{n} b_i^{(k)} x_i + \sum_{i_1<i_2} b_{i_1 i_2}^{(k)} x_{i_1} x_{i_2} + \ldots +$$

$$+ \sum_{i_1<i_2<\ldots<i_r} b_{i_1 i_2 \ldots i_r}^{(k)} x_{i_1} x_{i_2} \ldots x_{i_r},$$

and F being a recognition function, it satisfies (1.3), the condition for a threshold function of order r, if we define

$$w_0 = \log \frac{p_1}{p_2} + \sum_{i=1}^{n} \log \frac{\beta_i^{(1)}}{\beta_i^{(2)}} + (b_o^{(1)} - b_o^{(2)}), \quad (5.6)$$

$$w_i = \log \frac{\beta_i^{(2)}(1 - \beta_i^{(1)})}{\beta_i^{(1)}(1 - \beta_i^{(2)})} + (b_i^{(1)} - b_i^{(2)}), \quad (5.7)$$

$$i = 1, 2, \ldots, n.$$

$$\left.\begin{array}{l} w_{i_1 i_2} = (b_{i_1 i_2}^{(1)} - b_{i_1 i_2}^{(2)}), \quad i_1 < i_2 \\ \ldots \qquad \ldots \qquad \ldots \\ \ldots \qquad \ldots \qquad \ldots \\ w_{i_1 i_2 \ldots i_r} = (b_{i_1 i_2 \ldots i_r}^{(1)} - b_{i_1 i_2 \ldots i_r}^{(2)}), \; i_1 < i_2 < \ldots < i_r \end{array}\right\} (5.8)$$

$$i_1, i_2, \ldots, i_r = 1, 2, \ldots, n.$$

(ii) 'Only if' part:

This part is established by showing that, given w's, (5.6)-(5.8) are consistent with a set of p's and β's between 0 and 1, and a set of b's.

Firstly, let us choose.

$$\beta_i^{(2)} = \frac{1}{2}$$

and

$$\beta_i^{(1)} = 1/[1 + \exp. \{ w_i - (b_i^{(1)} - b_i^{(2)}) \}] . \qquad (5.9)$$

All the b's except b_o can be arbitrarily chosen for one distribution and for the other

(i) those with a single suffix can be arbitrarily chosen and

(ii) the rest except b_o can be obtained from (5.8) since w's are given.

Since

$$P(x|a_k) = P^*(x|a_k)\exp.\ b_o^{(k)}\ \exp.\ S_k, \quad k = 1,2,$$

where S_k are sums generated by the aforesaid b coefficients. Clearly $P(x|a_1)$ and $P(x|a_2)$ are ≥ 0, since each component is ≥ 0. Then let us choose

$$b_o^{(1)} = -\log_e [\ \Sigma\ P^*(x|a_1)\ \exp.\ S_1], \qquad (5.10)$$

$$b_o^{(2)} = -\log_e [\ \Sigma\ P^*(x|a_2)\ \exp.\ S_2], \qquad (5.11)$$

so that

$$\sum_x P(x|a_1) = \sum_x P(x|a_2) = 1.$$

Then let us choose

$$p_1 = 1-p_2 = \frac{\exp.[w_o - \{b_o^{(1)} - b_o^{(2)}\}]}{\exp.[w_o - \{b_o^{(1)} - b_o^{(2)}\}] + \prod_{i=1}^{n} \frac{\beta_i^{(1)}}{\beta_i^{(2)}}} . \qquad (5.12)$$

Then $P(x|a_1)$ and $P(x|a_2)$ according to this choice are indeed probability distributions and p's and β's clearly lie between 0 and 1, all consistent with (5.6)-(5.8). Hence Theorem 5.1.

<u>REMARK 5.3:</u> A convenient choice of arbitrary b values is indeed zero, in which case, choosing the second distribution for such a choice we obtain

$$P(x|a_2) = 2^{-n} \qquad (5.13)$$

which means that $b_o^{(2)} = 0$ and the distribution is uniform.

<u>REMARK 5.4:</u> Chow's theorem falls out as a particular case $r = 1$.

<u>REMARK 5.5:</u> The orthonormality of the basis is not used in the proof since our concern is only the existence. However, if one desires to compute the b coefficients, the orthonormality is to be used, the coefficients of the z functions are the inner products of log $[P(x)|P^*(x)]$ with the corresponding z functions.

<u>REMARK 5.6</u> Given a threshold function of order r, this theorem ensures the existence of one set of conditional probability distributions of order $\leq r$. However, there may exist other conditional probability distributions of order $> r$, for which the given threshold function of order r may be a recognition function. Thus it is possible to use threshold gate of order r for noise of order more than r; in particular, it is possible to use (first order) threshold gate to discriminate in the presence of non-independent noise.

Using Bayes' law, Winder[40] obtains a correspondence between statistical recognition with independently distributed inputs and threshold functions at any confidence level $\Lambda > \frac{1}{2}$. Working on similar lines as above we obtain a similar correspondence between distribution of order r and threshold gate of order r. We classify a pattern into a particular class, if given an observation ξ,

$$\Pr\,(f\,(\xi) = 1) \geq \Lambda\,. \tag{5.14}$$

Bayes' law leads to

$$\Pr\,(x = \xi \mid f(x) = 1) \geq \Lambda\,\frac{\Pr(x=\xi)}{\Pr(f(x)=1)}. \tag{5.15}$$

Expressing

$$Pr(x=\xi) = Pr(x=\xi\,|\,f(x) = 1)\ Pr(f(x) = 1)$$
$$+ P(x=\xi\,|\,f(x) = 0)\ Pr(f(x) = 0), \qquad (5.16)$$

We see that by taking logarithms this results in a threshold function of order r.

5.4 An Example

Consider the case $n = 3$ and a switching function given by the weights

$$w_0=0;\ w_1 = w_3=4,\ w_2 = 0;\ w_{12}= 4,\ w_{23}= -4,\ w_{31}= 0.$$

Kaplan-Winder theorem shows that this function is not of first threshold order, since the first order weights, which are the same as the ones given above, do not realize the function.

Choose all $b^{(2)}$ except $b_0^{(2)}$ as zero. Also, choose $b_1^{(1)} = b_2^{(1)} = b_3^{(1)} = 0$. Then, we have, by (5.8), $b_{12}^{(1)} = 4$, $b_{23}^{(1)} = -4$, $b_{31}^{(1)} = 0$, and by (5.11), $b_0^{(2)} = 0$. Thus, $P(x|a_2) = \frac{1}{8}$. By (5.9), we have

$\beta_1^{(1)} = \beta_3^{(1)} = 1|(1 + \exp. 4) = 0.0180$, $\beta_2^{(1)} = 0.5$.

Thus $\alpha_1^{(1)} = \alpha_3^{(1)} = 0.9820$, $\alpha_2^{(1)} = 0.5$. Then (5.10) gives $b_0^{(1)} = -0.376396$, (5.12) gives $p_1 = 1-p_2 = 0.9991082$;

$p_2 = 0.0008915$. Thus

$$P(x|a_1) = P^*(x|a_1) . \exp.[-0.376396 + 4x_1x_2 - 4x_2x_3].$$

Then the recognition function is computed in Table 5.1.

Table 5.1

Computing the Recognition Function

x	f(x)	$P(x\|a_1)$	$P(x\|a_2)$	$p_1P(x\|a_1)$	$p_2P(x\|a_2)$	R(x)
(0,0,0)	0	0.000111	0.125	0.0001109	0.0001114	0
(0,0,1)	1	0.006065	0.125	0.006060	0.0001114	1
(0,1,0)	0	0.000111	0.125	0.0001109	0.0001114	0
(0,1,1)	0	0.000111	0.125	0.0001109	0.0001114	0
(1,0,0)	0	0.000111	0.125	0.0001109	0.0001114	0
(1,0,1)	1	0.330908	0.125	0.330643	0.0001114	1
(1,1,0)	1	0.331186	0.125	0.330921	0.0001114	1
(1,1,1)	1	0.330908	0.125	0.330643	0.0001114	1

It is easily seen that F(x) is the same as R(x). This verifies Theorem 5.1.

5.5 Threshold Decoding

Massey [22] has shown that a threshold logic gate can be used to instrument certain decoding rules for a binary memoryless channel with additive noise. His decoding rule is based on the average probability of error as a criterion of goodness and is hence a posteriori probability decoding or simply APP decoding. Considering a noise sequence $e_1, e_2, \ldots, e_n$, the decoding algorithm will assign to e_m $(m = 1,2,\ldots, n)$, that value V for which the conditional probability

$$\Pr\left\{ e_m = V \mid \{ A_i \} \right\}$$

is a maximum where A_i is a composite parity check, $i = 1, 2, \ldots, J$, orthogonal on e_m (for definition see [22], p. 6).

We drop here the assumption of independence of the noise sequence and generalize this result showing that a threshold gate of order r can be used to instrument APP decoding rules for a binary channel with noise, dependent of order r.

Assuming that the joint distribution of A_i conditionally on $e_m = 1$ and $e_m = 0$ are of order r_1 and r_2

respectively with $r = \max. (r_1, r_2)$, the following Theorem is obtained by the same techniques as in Section 5.3 , based on the proof of Massey [27].

THEOREM 5.2: For a binary channel with additive noise, the APP decoding rule is: If the conditional probability distribution of the orthogonal parity checks A_i on e_m, given $e_m = 1$ and $e_m = 0$ are of orders r_1 and r_2 respectively, with $r = \max. (r_1, r_2)$, then choose e_m using a threshold function of order r, given by the weights (5.6) - (5.8), where p_1 and p_2 are a priori probabilities of $e_m = 1$ and $e_m = 0$ respectively, and b with superscripts 1 and 2 are respectively the coefficients in the expansion of $\log [A_i \mid e_m = 1]$ and $\log [A_i \mid e_m = 0]$ and $\beta_i^{(1)} = P(x_i = 0 \mid e_m = 1)$ and $\beta_i^{(2)} = P(x_i = 0 \mid e_m = 0)$.

REMARK 5.7: It may be possible to prove that the order of the conditional distributions of A_i given $e_m = 1$ and $e_m = 0$ is less than or equal to the order of the distribution of the noise sequence $e_1, e_2, \ldots, e_n$. In such a case, only the assumption of order r of the noise sequence will do for the Theorem.

RELATED WORK

In Chapter 5 (page 101, last paragraph), we refer to Nilsson's book [34] regarding the relaxation of the assumption of independence using Bahadur's results. Dr. Laveen Kanal has kindly pointed out to the author that Bahadur's expansion was first introduced to the pattern recognition-threshold logic community by him in the papers [43] - [46]. Dr. Kanal writes: ... the relevance of Bahadur's expansion to threshold functions both for the case of statistically independent and dependent inputs was pointed out in [43] and also mentioned in [44]. The reasons which make approaches similar to that of Bahadur's unsuitable for practical applications were briefly described in [45] and have been presented in greater detail in [46]. The author regrets having overlooked Dr. Kanal's papers. Even though Dr. Kanal does not appear to explicitly introduce the notion of threshold order and prove a theorem like out Theorem 5.1, the idea of the use of Bahadur's expansion in this context is found in his papers. The author is thankful to Dr. Kanal for his comments.

After writing this dissertation, I came to know of the recent work of Haring [47] in which he proves a number of interesting results on 'multiple- threshold threshold functions' discussed in our pages 94-96 connecting it to a threshold function with a larger number of inputs. Dr. Haring informed the author that Spann of Massachusetts Institute of Technology, in his doctoral dissertation has established some connections between these multi-threshold elements and threshold order. Spann seems to have called a threshold function of order r as a multiple-weight threshold function. The author has not yet had access to the results of Spann.

REFERENCES

[1] Akers, S.B.: Threshold Logic and Two-Person, Zero-Sum Games, Switching Circuit Theory and Logical Design, pp. 27-33, 1961.

[2] Bahadur, R.R.: A Representation of the Joint Distribution of Responses to n Dichotomous Items. Studies in Item Analysis and Prediction ed: H. Solomon, Stanford University Press, Ch. 9, pp. 158, 168, 1961.

[3] Braverman, D.: Theories of Pattern Recognition, in A.V. Balakrishnan, ed. Advances in Communication Systems, Academic Press, London, pp. 33-56. 1964.

[4] Chow, C.K.: On the Characterization of Threshold Functions. Switching Circuit Theory and Logical Design. AIEE Proc. S-134, pp. 34-38, 1961.

[5] -------- An optimum character recognition system using decision function. IRE Transations on Electronic Computers, Vol. EC-6 pp.247-254, 1957.

[6] ----------- Boolean Functions Realizable with Single Threshold Devices. Proceedings of the IRE, Vol. 49, pp. 370-371, 1961.

[7] --------- A Recognition Method Using Neighbor Dependence. IRE Transactions on Electronic Computers, Vol. EC-11, pp. 683-690, 1962.

[8] ---- ---- Statistical Independence and Threshold Functions. IEEE Transactions on Electronic Computers, Vol. EC-14, pp. 66-68, 1965.

[9] Coleman, R.P.: Orthogonal Functions for the Logical Design of Switching Circuits. IRE Transactions on Electronic Computers, Vol. EC-10, pp. 379-382, 1961.

[10] Ercoli, P. and Mercurio, L: Threshold Logic with one or more than one threshold. Proc. IFIP Conference 62, Munich, pp. 741-746, 1962.

[11] Golomb, S.W.: On the classification of Boolean Functions. IRE Transactions on Circuit Theory, Vol. CT-6, (Special Suppl.), pp. 176-186, 1959.

[12] Gose, E.E.: An Adaptive Network for Producing Real Functions of Binary Inputs. Information and Control, Vol. 8, pp. 111-123, 1965.

[13] Hawkins, J.K.: Self-organizing Systems - A Review and Commentary. Proc. IRE, Vol. 49, pp. 31-48, 1961.

[14] van Heerden, P.J.: Analysis of Binary Time Series in Periodic Functions. IRE Trans. on Electronic Computers, Vol. EC-8, pp. 228-229 1959.

[15] Ho, Y-C. and Kashyap, R.L.: An Algorithm for Linear Inequalities and its Applications. IEEE Transactions on Electronic Computers, Vol. EC-14, pp. 683-688, 1965.

[16] Huffman, D.A.: The Synthesis of Linear Sequential Coding Networks. Proc. Third Symp. on Information Theory, London, pp. 77-97, 1955.

[17] Kaplan, K.R. and Winder, R.O: Chebyshev approximation and Threshold Functions. IEEE Trans. on Electronic Computers, Vol. EC-14, pp. 250-252, 1965.

[18] Kaszerman, P.: A Non-Linear Summation Threshold Device. IEEE Transactions on Electronic Computers, Vol. EC-12, pp. 914-915, 1963.

[19] Kjellberg, G.: Logical and other kinds of independence. Proceedings of an International Symposium on the Theory of Switching. Part I Harvard University Press. Cambridge Mass. pp. 117-124, 1959.

[20] Krishnan, T.: On the Threshold order of a Boolean Function. IEEE Transactions on Electronic Computers, Vol. EC-15, pp. 369-372, 1966.

[21] Liu, C.L. and Liu, J.W.S. On a Multiplexing Scheme for Threshold Logical Elements. Information and Control Vol. 8, pp. 282-294, 1965.

[22] Massey, J.L.: Threshold Decoding. M.I.T. Press Cambridge Mass. 1963.

[23] Mattson, R.L.: A self-organizing Binary System. Proceedings of the Eastern Joint Computer Conference, pp. 212-217, 1959.

[24] Mays, C.H.: Adaptive Threshold Logic. Stanford Electronics Laboratories, Stanford. Technical Report No.1557-1, 1963.

[25] --------- The Boundary Matrix of Threshold Functions. IEEE Transactions on Electronic Computers, Vol. EC - 14 , 1965, pp. 65-66,

[26] McNaughton, R.: Unate Truth Functions. IRE Transactions on Electronic Computers, Vol. EC-10, pp. 1-6, 1961.

[27] Minnick, R.C.: Linear-input Logic. IRE Transactions on Electronic Computers, Vol. EC-10, pp. 6-16, 1961.

[28] Minsky, M. and Selfridge, O.G. Learning in Random Nets. M.I.T. Lincoln Lab., Lexington.Mass., Report No.54G-0024; June 1960.

[29] Muller, D.E.: Application of Boolean algebra to switching circuit design and error detection. IRE Transactions on Electronic Computers Vol, EC-3, pp. 6-12, 1954.

[30] Muroga, S., Toda, I. and Kondo, M.: Majority Decision Functions of upto Six Variables. Mathematics of Computation, Vol.16 pp. 459-472, 1962.

[31] Muroga, S.: Majority Logic and Problems of Probabilistic Behavior. Self.Organizing Systems 1962, Eds.: Yovits,M.C. Jacobi, G.T. and Goldstein, G.D., Spartan, Washington, pp. 243-281, 1962.

[32] Nambiar, K.K.: Representation of Distributions of Dichotomous Random Variables. Doctoral Dissertation, Univ. Pennsylvania, 1965.

[33] Neumann, J. von: Probabibilistic Logics and the Synthesis of Reliable Organisms from unreliable components. Automata Studies, pp. 43-98, Princeton 1956.

[34] Nilsson, N. J.: Learning Machines. McGraw-Hill 1965.

[35] Partanen, J.: A Decomposition of Dichotomous Random Variables into Independent Components of Variation . Item Analysis, Test Design and Classification. Ed. H. Solomon, Stanford University Press, pp. 321-359, 1965.

[36] Pierce, W.H.: Interwoven Redundant Logic. Journal of the Franklin Institute, Vol. 277, pp. 55-85, 1964.

[37] Sheng, C.L.: Threshold Logic Elements Used as a Probability Transformer. Journal of the Association of Computing Machinery, Vol. 12, pp. 262-276, 1965.

[38] Walsh, J.F.: A closed Set of Normal Orthogonal Functions. American Journal of Mathematics. Vol. 45, pp. 5-24, 1923.

[39] Winder, R. O.: Some Recent Papers in Threshold Logic Proceedings of IRE, Vol. 49, p. 1100, 1961.

[40] ---------- Threshold Logic in Artificial Intelligence. Artificial Intelligence, IEEE Publication, S. 142, pp. 107-128, 1963.

[41] ---------- Enumeration of Seven-Argument Threshold Functions. IEEE Transactions on Electronic Computers, Vol. EC-14, pp. 315-325, 1965.

[42] ---------- Threshold Functions Through n=7, AFCRL Rept. 64-925, RCA Laboratories, Princeton, N.J., 1964.

[43] Kanal, L. Evaluation of a Class of Recognition Networks. Biological Prototypes and Synthetic Systems, Ed.: E.E. Bernard and M.R.Kane, New York: Plenum Press, pp. 261-269, 1962.

[44] Kanal, L., Slaymaker, F., Smith, D. and Walker, W.: Basic Principles of Some Pattern Recognition Systems. Proceedings National Electronics Conference, Chicago, pp. 279-295, 1962.

[45] Abend, K., Harley, T. J. and Kanal, L. N.: Classification of Binary Random Patterns. IEEE Trans. on information Theory, Vol. IT-11, pp. 538-544, 1965.

[46] Kanal, L. and Abend, K.: Adaptive Modeling of Likelihood Classification-1. Technical report on RADC contract, AF 30 (602)-3623, Feb. 1966.

[47] Haring, D. R.: Multiple- Threshold Threshold Functions. IEEE Transactions on Electronic Computers, Vol. EC-15, pp. 45-65, 1966.
