This is to certify that the

thesis entitled

ASSESSING UNCERTAINTY IN MEDICAL DIAGNOSIS BY FOUR PROBABILITY MODELS

presented by

Raywin Rufus Huang

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Department of Counseling, Personnel Services and Educational Psychology (Statistics & Research Design)

Major professor

Date June 14, 1977

ASSESSING UNCERTAINTY IN MEDICAL DIAGNOSIS BY FOUR PROBABILITY MODELS

By

Raywin Rufus Huang

AN ABSTRACT OF A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services
and Educational Psychology

1977

ABSTRACT

ASSESSING UNCERTAINTY IN MEDICAL DIAGNOSIS
BY FOUR PROBABILITY MODELS

By

Raywin Rufus Huang

Uncertainty plays a pernicious role in medical diagnosis. This dissertation defines uncertainty as not having knowledge of the relational structure of the disease outcome and a set of symptoms in the true state of nature. Conditional probability is used as the fundamental measure of uncertainty. Four probability models, namely (1) the Bayesian model, (2) the Binary Regression model, (3) the Logistic Discrimination model, and (4) the Entropy Minimax Pattern Discovery model, are presented as well as their mathematical algorithms for generating the conditional probability of a disease outcome given a set of symptoms. An algorithm is also developed to simulate different classes or levels of uncertainty within the structure of the diagnostic problem. Each model is applied to each class to derive its parameters and each model is cross-validated to an equivalent sample for the purposes of (1) determining the stability of each model's estimated parameters in terms of sensitivity, specificity, and predictive value, and (2)

to model the clinical situation where the physician is cross-validating his set of strategies to new cases on the basis of prior information. Each model is also evaluated in terms of a utility function, losses and gains. Some special classes of the uncertainty structure are also simulated and each model is evaluated by the same methodology and with the same evaluation indices. The models are then applied to different relational structures and are then evaluated in terms of sensitivity, specificity, predictive value and utility function. These results are then compared to prior findings. The findings of this dissertation are as follows:

1. Overall, sensitivity increases for all models as the correlation with the disease outcome increases.

2. There is a "hump" or convex effect for sensitivity for all models except the Bayesian (B), Bayesian with the Bahadur's expansion (BB), and the Entropy Minimax Pattern Discovery (EMPD) models, in situations where the symptoms have a low correlation with the disease outcome. That is, the maximum sensitivity is not when the intercorrelation between the symptoms is greatest but when the symptoms are moderately intercorrelated! This phenomenon did not appear in situations where the symptoms have a high correlation with the occurrence of the disease. In fact, sensitivity increases as the intercorrelations increase under the latter situation.

3. The values for sensitivity did not differ among models in situations where highly interrelated symptoms are also highly related to the occurrence of the disease. In other words, when the relational structure is highly correlated, it does not matter which model one uses if sensitivity is chosen as a criterion for selection models.

4. The "pit" or concave effect of specificity across binary regression models occurs when, given those situations where the symptoms are highly correlated with the disease outcome, the intercorrelations between the symptoms increase. This means that specificity is at a minimum when the symptoms are moderately related.

5. The "hump" or convex effect is also found for predictive values in the same way as the sensitivity index, that is, when the symptoms have a low correlation with the occurrence of the disease.

6. With the presence of a suppressor symptom, it does not matter what measure one uses as a criterion for selecting models as all models perform the same for all prediction efficiency indices.

7. If a model is chosen with the criterion as having the best sensitivity, it is at a cost of losing specificity and vice versa. In other words, there are no models that have the best of both indices for all classes considered in this dissertation. The statement holds when one looks

across classes and within classes of problems. This also means that there is no single model that performs consistently better for each class or across classes in terms of sensitivity and specificity.

8. A decision function analysis was performed. Penalty (negative) weights were given for the two diagnostic errors (i.e., Type I and Type II) and no credit was given to the correct diagnosis. The binary least square model (BLS) and the binary weighted least square model (BWLS) showed the smallest loss when the symptoms had a low correlation with the disease's occurrence but themselves had high intercorrelations. However, when considering gains, with credits given to the correct diagnosis, but the same penalty weights, the Bayesian model (B) had the most gain when the intercorrelations among the symptoms were low but the correlation between the symptoms and outcome was high. The logistic discrimination model (LD) had the most gain when the symptoms had a low correlation with each other but had a high correlation with the occurrence of the disease outcome. The LD model also had the most gain when the symptoms were moderately interrelated with each other and the symptoms had a low correlation with the disease. If one disregards the intercorrelation among symptoms, the LD model had the highest gain whether the symptoms had a high or low correlation with the occurrence of the disease. That is, the

best model to use to maximize gain in the absence of knowledge about the relationship among and between symptoms and disease outcomes, is the LD model.

Implications and applications of these findings to diagnostic problem solving are also presented.

TO MY BELOVED WIFE, RITA

AND SON, WINSON

ACKNOWLEDGMENTS

I would like to express my sincere and deepest appreciation to the following individuals who have played a major role in one way or the other in the making of this dissertation.

To Dr. Maryellen McSweeney who has unselfishly offered her time and effort throughout my graduate academic career at Michigan State University. She has not only enriched my academic experience but has also set me the example of a person with modesty and humility.

To Dr. Howard Teitelbaum whose kindness expressed to my account in numerous ways, has given me great encouragement and moral support throughout my graduate career. He has been more than simply an adviser and friend and I have enjoyed tremendously and with great respect working with him.

To Dr. Arthur Elstein for his many stimulating and enjoyable discussions which brought much delightful inspiration to this dissertation. He has led me to new levels of intellectual capabilities.

To Dr. James Potchen who has offered me much valuable advice in making this dissertation relevant to a real clinical setting.

To Dr. William Schmidt for his assistance in statistical methodology of this dissertation which I am most grateful.

To my beloved wife, Rita, whose love and patience act as a buoy to the process of the making of this dissertation.

To my father, Rufus Huang, who has given me great moral encouragement and support throughout my academic experience.

To Grace Rutherford who has so professionally typed and prepared this manuscript.

And to my Lord Jesus Christ, who has provided me the strength and will to overcome the hurdles towards the making of this dissertation.

It is my hope that these efforts so kindly and graciously offered to me by the above individuals will not go in vain and I will see that their qualities will be passed on to others in the manner they have been so generously passed on to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

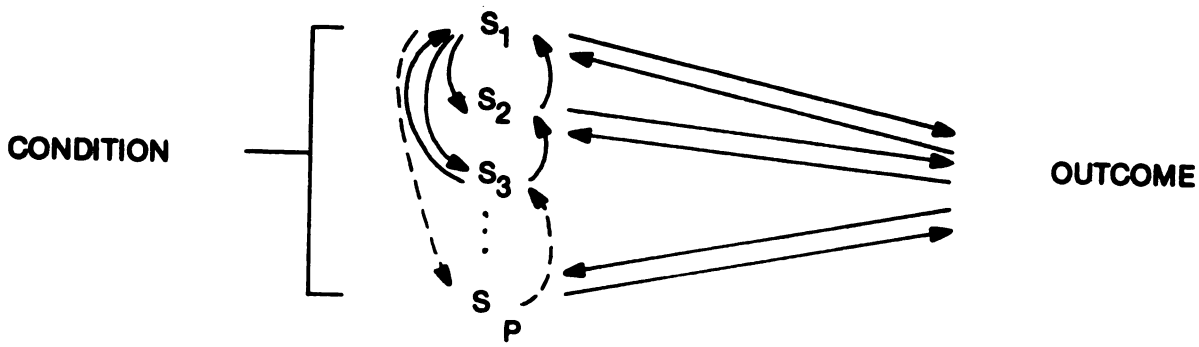
INTRODUCTION

Uncertainty is a root of indecision. It is like a disease to the process of decision making and it curtails human performances. Swet (1961) found that performance in signal detection by human subjects decreased as the amount of uncertainty increased. Uncertainty, however, is defined in many ways. Webster (1974) defined it as a quality or state of being indefinite, indeterminate, problematical, dubious and fitful. Bowman (1964) defined it as a situation characterized (either objectively or subjectively) by incomplete predictability of alternative events.

Cohen (1973) distinguished two categories of uncertainty, namely the intrinsic and the extrinsic. Intrinsic uncertainties arise from imprecicion, ambiguity, and limitation of the data on which the decision is to be made. Extrinsic uncertainties refer to the failure on the part of the data interpreter in translating the data, otherwise known as "observer error." Kaplan (1964), on the other hand, defined two different kinds of uncertainty. One kind is risk where there is a knowledge of a law that operates in nature but involves a purely random element.

The outcome, despite a given probability, remains unassured. The other kind of uncertainty is referred to as statistical ignorance where the law of operation itself is unknown. Ignorance arises not necessarily because of non-specifiable circumstances but rather because there is a lack of the occurrence of enough significant outcomes so that deterministic probabilities can be assigned to these outcomes. Kaufmann (1968) classified levels of uncertainty according to the degree of knowledge available. One level is non-structural uncertainty, that is, when the states of the system are unknown at any point in time. Structural uncertainty occurs when the state of the system, despite being known in general, is not known at any given time. The condition when the states of the system with its laws of probability are known at any time, is called chance. Certainty is a state where the system is known and it can be described at any point of time.

This dissertation defines uncertainty as not having the knowledge of the structure of the relationship between an outcome and certain sets of conditions and the interrelationships among the attributes in the condition (Figure 1.1). The term condition is referred to as a universal space in which the attributes are its elements. Predictors, indicators, attributes, independent variables, and exogeneous variables will be used synonymously with



The set $S = \{S_1, S_2, \ldots, S_p\}$ is a subset of the universal set of

$$U = \{S_1, S_2, \ldots, S_\infty\}$$

Note: The arrows leading from the attributes to the outcome imply that the attributes precede the outcome and are causative to the outcome.

The arrows leading from the outcome to the attributes imply that the attributes are subsequent to the outcome and are purely symptomatic in nature.

Figure 1.1 The Relational Structure Between Attributes and the Occurrence of an Outcome and Among Attributes in the True State of Nature.

symptoms and/or signs whereas dependent variable, the criterion and endogeneous variables will be used synonymously with the disease outcome. The terms variables or attributes will refer, in general, to both disease outcome and symptoms. The term relational structure will refer to the relations between the disease outcome and the symptoms and the interrelations among the symptoms.

One underlying assumption of this definition of uncertainty is that there exists a well defined and structural deterministic relationship between an occurrence of a disease and certain sets of conditions in the true state of nature. This has two implications. One implying that certain sets of conditions precede the disease and are causative agents to the disease outcome. The other implication is the set of conditions are subsequent to the disease outcome and are purely symptomatic in nature. This stipulated definition of uncertainty also leads to the formulation of the uncertainty principle which states that only when full knowledge of this (true state of nature) relational structure is obtained, can the outcome of any diseases be stochastically predicted without error with reference to the given known conditions. Hence, when only partial or imperfect knowledge of this relational structure is obtained, uncertainty arises and thus leads to random guesses.

The paradox of this principle is that even when full knowledge is gained, which demands the collection of exhaustive information relating to the disease, the prediction of a specific disease outcome is still subject to error. This is due to the complex relations among variables as illustrated in Figure 1.2. For instance, symptom S is related to both disease D_1 and D_2 (denoted by SD_1 and SD_2). The absence of the symptom, $\bar{S}$, is also related to the outcome of both of these diseases denoted by $\bar{S}D_1$ and $\bar{S}D_2$. Adding to this complexity of relationship, the presence of the symptom is not necessarily related to the outcome of either D_1 or D_2 denoted by $S\bar{D}_1$ and $S\bar{D}_2$. Hence, the presence or absence of the symptom, S, could not determine exactly the occurrence of either D_1 or D_2 for a single case. Error is, therefore, an inescapable consequence. Nonetheless, this principle holds over a large number of cases. That is, when the relational structures are known, the prediction of the proportion of cases having the disease will be without error. But it should be noted that this is accomplished only when full and perfect knowledge is obtained.

This imperfect and incomplete knowledge of the relational structure is caused by the complexity of the relational structure itself which leads to the difficulty of obtaining this knowledge as Hammond et al. (1975) noted:

> Knowledge of the environment is difficult to acquire because of casual ambiguity and because of the probabilistic intangled relations among environmental variables. (emphasis mine)

In spite of this difficulty, partial knowledge of the environment and its relational structure can still be gained from samples from the complex state of nature. These samples constitute imperfect information about the universal relational structure. This sample information, unfortunately, leads to inferences about the state of nature in probabilistic terms such as "likely," probable," "perhaps," or "maybe" which constitute many human beliefs. Inference of the true state of nature becomes then an art of estimation. These probabilistic beliefs prompted Tversky (1974) to describe uncertainty as an essential element of the human condition. It should be noted that prediction and inferences are used synonymously.

The ambiguity of the structural relationship of the true state of nature constitutes uncertainty, and this ambiguity is due to partial knowledge arising from insufficient information. This, in turn, is primarily due to the complexity of the true relational structure and secondarily due to methodological limitations in obtaining full and complete information about the structure.

Uncertainty in Medicine

Disease is defined as literally meaning "lack of ease" or the pathological condition of the body that presents a group of symptoms peculiar to it and which sets the condition apart as an abnormal entity differing from other normal or pathological body states (Taber, 1970) and symptom simply denotes the manifestation of the disease. Medical diagnosis is then an art of identifying the correct disease with reference to certain set of symptoms as Wakefield (1955) remarked: "Diagnosis is the art and the science of recognizing the presence of the absence of disease from signs, symptoms. . . ."

The Dorland's Illustrated Medical Dictionary defined diagnosis as the art of distinguishing one disease from another.

In a different perspective, medical diagnosis is also an art of probabilistic inferences or prediction. "Medicine is a science of uncertainty and an art of probability," was the dictum of Sir William Osler (Bean, 1950). Lusted (1968) further remarked the following: "The uncertainty about the correlation of signs, symptoms and disease makes medical diagnosis a matter of probability."

Engel and Davis (1963) distinguished five orders (levels) of medical uncertainty with variation of etiology within each order. They are presented as follows:

1. Diagnosis of the First Order: the diagnostic situation where the disease is considered to be well defined and the etiology of the disease is, in most instances, clear and the disease picture does not vary much from person to person or from environment to environment.

2. Diagnosis of the Second Order: diagnosis with well defined etiology but the disease picture has greater variability from patient to patient and from environment to environment.

3. Diagnosis of the Third Order: the diagnosis is clearly descriptive and the etiology is unknown.

4. Diagnosis of the Fourth Order: the general type of reaction is recognized but the specific cause is not known and individual and environmental variation occurs.

5. Diagnosis of the Fifth Order: the diagnosis is based on the constellation of signs and symptoms which comprise the disease picture. However, the etiology of the disease is unknown.

This dissertation will consider Engel and Davis' last order of diagnostic certainty.

Engel and Davis (1963) concluded their thesis by stating the following:

> Thus, inherent in every diagnosis is a factor of uncertainty, greater in some and less in others. These uncertainties are partially related to our imperfections of knowledge concerning health and disease. (emphasis mine)

Consider the situation where a patient is at the physician's office showing a set of symptoms or signs. The physician has a prior knowledge of the disease as represented in a disease-symptom matrix, something like Figure 1.3, with p symptoms and N patients. The ones and

	(Disease) Outcome	Condition S_1	S_2	S_3 . . .	S_p
Patient 1	1	1	0	0	1
Patient 2	0	0	1	1	0
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
Patient N	1	1	0	1	1

Figure 1.3 Representation of the Physician's Prior Knowledge of a Disease With Its Symptom.

zeroes represent the presence or absence of the disease or symptoms. It is worthy to note that only two possible disease outcomes will be considered in this dissertation, namely, the presence of a disease, denoted by D and the absence of the disease, denoted by $\bar{D}$, and that emphasis is placed on discrete symptoms. Such form of diagnosis is referred to as symptom diagnosis (Rinalde et al., 1963) or a diagnosis of the fifth order. The physician, based on this prior information can proceed with the medical diagnosis in two possible ways. One way is by probabilistic explanation and may be schematized as follows:

> From the matrix, the probability for disease D to have symptom or symptoms S is high. The patient has symptom or symptoms S, (therefore it is highly probable that) the patient has disease D.

This is known as the laws of probabilistic form (Hempel, 1966); the explicans implies the explicandum not with deductive certainty but with near certainty or with high probability.

The other way is by pattern recognition: that is to say, the selection of a number of possibilities which come nearest to explaining the signs or symptoms. The process of matching the disease with symptoms was noted by Harvey and Bordley (1970). Alternatively, the physician considers the process which enumerates in orderly fashion the various diseases which give rise to particular signs or symptoms.

These two methods represent two diagnostic paradigms but the final diagnosis, by either method, is still characterized in the form of "odds," "risk," and "chances." Medical uncertainties undoubtedly play a detrimental role in human welfare. It is a challenge to assess these uncertainties in the hope of reducing them.

The Structure of Uncertainty

Since the relational structure in the true state of nature is unknown, the main problem is how can "uncertainty" be conceptualized such that it can be systematically and formally investigated? The key to this problem is by theoretically partitioning the relational structure into

possible exhaustive states. This is done by arbitrarily dividing the degree of relationship among attributes into categories and likewise the degree of relationship with the disease. Figure 1.4 shows one way of partitioning uncertainty into these possible classes.

The number of dividing levels is totally at the discretion of the investigator. As the number of levels increases, the structure of uncertainty is increasingly defined. The mixture of the classes also constitutes states of uncertainty.

		Intercorrelation of the Symptoms		
		Low	Medium	High
Correlation with the Disease	Low	I	II	III
	Medium	IV	VI	VI
	High	VII	VIII	IX

Figure 1.4 The Structure of Uncertainty.

Hence, uncertainty is "captured" into a well defined bounded framework, making assessment possible.

Quantification of Uncertainty

Bearing in mind with the above uncertainty structure, a step is taken further to derive a quantitative measure. Since the occurrence of any event cannot be deterministically defined, the occurrence of any event can only be stochastically derived. This means that with certain sets of a known condition, the occurrence of an event appears only n% of the time or n times out of a hundred. The (100 - n)% times that the event does not occur with relation to the set of conditions is either due to imperfect or partial knowledge from insufficient information or due to error.

In deriving a quantified measure for uncertainty, consider a disease, D, has n number of cases in a population of size N. Assuming equally likely outcomes, the probability of D occurring in this population is simply:

$$P(D) = n/N \tag{1}$$

For a given sign or symptom, S, the probability that D will occur conditioned upon the occurrence of S is:

$$P(D|S) = P(D \cap S)/P(S) \tag{2}$$

where $P(D \cap S)$ is the probability that both the disease and the symptom will occur and $P(S)$ is the probability that S will occur in the population of size N. The probability, $P(D|S)$, is known as the conditional probability or posterior probability. In this dissertation, it will be referred to as diagnostic probability. Equation 2 can be elaborated by the following 2 x 2 matrix as illustrated in Table 1.1:

Table 1.1

The Possible Distribution of Cases by Both Disease and Symptom Outcome

		Symptom S: 1	Symptom S: 0	
Disease D	1	n_1	n_2	$\sum_{i=1}^{4} n_i = N$
	0	n_3	n_4	

where n_1 is the frequency or number of cases having the symptom and the disease, n_2 is the number of cases having the disease but no symptom, n_3 is the number of cases having the symptom but not the disease, and n_4 is the number of cases not having either the symptom or the disease. Hence, assuming equally likely events, the above probabilities can be written as:

$$P(D \cap S) = n_1/N \tag{3}$$

$$P(D) = (n_1 + n_2)/N \tag{4}$$

$$P(S) = (n_1 + n_3)/N \tag{5}$$

Hence, equation (2) can be written as follows:

$$P(D|S) = n_1/(n_1 + n_3) \tag{6}$$

Extending to p number of signs or symptoms, the probability that D will occur conditioned upon the occurrence of the symptoms will be:

$$P(D|S_1,S_2,\ldots,S_p) = P(D \cap S_1 \cap S_2 \cap \ldots \cap S_p)/P(S_1 \cap S_2 \cap \ldots \cap S_p) \tag{7}$$

where $P(S_1 \cap S_2 \cap \ldots \cap S_p)$ is the probability that the symptoms jointly occur. The left side of the term of equations (2) and (7) can be interpreted as the probability of occurrence of the disease given the occurrence of the symptom or p symptoms. Let $\overline{D}$ and $\overline{S}_i$ denote the absence of the disease and the ith symptom, respectively. Then $P(\overline{D}|\overline{S})$ would be the probability of the disease's not occurring given the absence of the symptom. Likewise, for $P(\overline{D}|S_1,S_2,\ldots,S_i\ldots S_p)$ would be the probability of the disease's not occurring given the occurrence of (p - 1) symptoms and the absence of the ith symptom. It is worthy to note that for p number of signs or symptoms, there will be 2^p number of possible combinations or patterns. Let $k=1\ldots2^p$ denote one of the

possible patterns and let $\underline{x}_k$ denote the vector of the pattern, then equation (7) can be rewritten as:

$$P(D|\underline{x}_k) = P(D \cap \underline{x}_k)/P(\underline{x}_k) \tag{8}$$

With more than one symptom, the situation can be presented as in Table 1.2. The probability that the disease will occur given the $\underline{x}_k$ pattern and assuming equally likely events is:

$$P(D|\underline{x}_k) = m_{1k}/(m_{1k} + m_{2k}) \tag{9}$$

The conditional probability of the symptom(s) given the disease, $P(S|D)$ or $P(\underline{x}_k|D)$ can be written for a single symptom as:

$$P(S|D) = P(S \cap D)/P(D) \tag{10}$$

or in the case of p symptoms as:

$$P(\underline{x}_k|D) = P(\underline{x}_k \cap D)/P(D) = m_{1k}/M_1 \tag{11}$$

These probabilities are used to derive the diagnostic probabilities with respect to the base rate of the disease which will be presented in the following chapter. $P(S|D)$ is also known as the <u>likelihood probability</u>.

The conditional probabilities derived from equations (2) and (8) are <u>exact probabilities</u>. They are derived directly from allocating observed cases according to the disease outcome and the symptom pattern outcomes.

Table 1.2

The Frequency Distribution of Cases by Disease Outcome and Symptomatic Patterns

Symptom Pattern	Disease	
	D	$\bar{D}$
$\underline{x}_1 = (S_1, S_2, \ldots, S_p)$	m_{11}	m_{21}
$\underline{x}_2 = (S_1, S_2, \ldots, \bar{S}_p)$	m_{12}	m_{22}
.	.	
.	.	
.	.	
$\underline{x}_k = (S_1, \bar{S}_2, \ldots, \bar{S}_p)$	m_{1k}	m_{2k}
.	.	.
.	.	.
.	.	.
$\underline{x}_h = (\bar{S}_1, \bar{S}_2, \ldots, \bar{S}_p)$	m_{1h}	m_{2h}
	M_1	M_2

m_{ij} = the number of cases having the ith disease outcome and the jth pattern;

M_i = the total number of cases for the ith disease; and

$h = 2^p$ where p is the number of symptoms.

The Diagnostic Situation and the Diagnostic Problem

There are at least two paradigmatic ways of diagnosing under uncertainty: (1) the probabilistic explanation and (2) pattern recognition.

The two paradigms can be illustrated as follows. Probabilistic explanation can be seen in a physician's checking off the diseased and non-diseased cases in a set of new patients based on his prior experience and the manner used to integrate this information or the method used for diagnosis. Pattern matching can be seen in the attempt of a disease clinic to detect the high risk group for a particular disease with respect to certain symptoms or signs. This latter situation is known as mass screening. One example of mass screening would be the detection of breast cancer.

With respect to these two paradigms, the crucial question to be raised is, which way is better? Better will be considered in terms of diagnostic accuracy and in terms of utility, losses or gains in dollars, or mortality.

The Purpose and Strategy of the Study

The purpose of this dissertation is to answer this question of diagnosing under uncertainty through quantification methodology. Quantitative methods or probability

models are chosen because they are a set of systematic and formal procedures capable of deriving an optimal solution from a complicated entanglement of variables in the true state of nature.

This dissertation will investigate the performance of four different statistical models in assessing uncertainty. The paradigms and associated probability models are:

Paradigm	Models
A. Probabilistic explanation	Bayesian
B. Pattern recognition	1. Binary Regression a. Ordinary Least Squares b. Weighted Least Squares c. Ridge Regression d. Weighted Ridge
	2. Logistic Discrimination
	3. Entropy Minimax Pattern Discovery

The strategy of this study begins by simulating the structure of uncertainty by a set of mathematical algorithms shown in Figure 1.4. Each simulated class with a fixed population is randomly divided into equal halves, one representing the prior information available and the other half representing the "unknown." Although the two halves have the same relational structure statistically, the second half is still referred to as the

"unknown." Statistical models are then applied to the first sample to derive its parameters and these parameters are used in turn to predict the outcome in the second "unknown" sample. This latter procedure is known as cross-validation. This step will assess the stability of the estimates from a statistical point of view. It is also analogous to the practice of medicine where a physician is constantly cross-validating his decision algorithms when a conclusion is reached after examining two patients presenting with the same sign and symptoms. How well each model cross-validates is measured by a set of efficiency indices. The values of each index will indicate the accuracy and error of each model. Each model is also evaluated in terms of utility or worth. The efficiency indices and utility measures are then compared with each other to determine the best model under different degrees of uncertainty. The models will also be examined under different relational structures between the two halves of each population.

The above procedure will deal with the following questions.

1. Which probability model has the best performances in terms of efficiency indices and utility across classes of uncertainty?

2. Which probability model performs the best in terms of efficiency indices and utility within each class of uncertainty?

The remainder of this dissertation is organized in the following manner. Chapter II presents the description and derivation of the probability models used for this study. Chapter III derives and discusses the algorithm used to generate the simulated data employed in deriving the estimates for each model. Chapter IV presents the data analysis and results of applying each probability model for each degree of uncertainty. The analysis is in terms of efficiency indices and utility functions. The models are then cross-validated with the same and with different relational structures between samples. Chapter V presents general findings and recommendations for further research. An implication of this study to decision making in a real medical setting is also discussed in Chapter V.

CHAPTER II

PROBABILITY MODELS

The probability models which follow the two main paradigms of medical diagnosis to derive the diagnostic probabilities are selected in this thesis are as follows:

Paradigm	Models
A. Probabilistic explanation	Bayesian
B. Pattern recognition	1. Binary Regression a. Ordinary Least Squares b. Weighted Least Squares c. Ridge Regression d. Weighted Ridge
	2. Logistic Discrimination
	3. Entropy Minimax Pattern Discovery

A description of each model within each paradigm is now presented.

Probabilistic Explanation

Bayesian Model (B)

This model was originated by Rev. Thomas Bayes (1763) and is simply formulated as:

$$P(D|S) = \frac{P(D)P(S|D)}{P(D)P(S|D) + P(\bar{D})P(S|\bar{D})} \tag{2.1}$$

where the probabilities are explained in the previous chapter. In the situation of determining the diagnostic probability when the symptom, S, is not present. The probability becomes:

$$P(D|S) = \frac{P(D)P(\bar{S}|D)}{P(D)P(\bar{S}|D) + P(\bar{D})P(\bar{S}|\bar{D})} \; . \tag{2.2}$$

However,

$$P(\bar{S}|D) = 1 - P(S|D) \tag{2.3}$$

and

$$P(\bar{S}|\bar{D}) = 1 = P(S|\bar{D}) \tag{2.3.1}$$

so that equation (2.2) can be rewritten as:

$$P(D|S) = \frac{P(D)(1 - P(S|D))}{P(D)(1 - P(S|D)) + P(\bar{D})(1 - P(S|\bar{D}))} \; . \tag{2.4}$$

For convenience in computation, a new variable, a, is defined to associate with the symptom. The new variable, a, will take a value of 1 if the symptom is present and 0 if it is absent. Hence, equation (2.1) and equation (2.4) are combined as follows:

$$P(D|a) = \frac{P(D)(aP(S|D) + (1-a)(1 - P(S|D))}{P(D)(aP(S|D)+(1-a)(1-P(S|D))+P(\bar{D})(aP(S|\bar{D})+(1-a)(1-P(S|\bar{D}))} \tag{2.5}$$

In extending to the situation of two symptoms, equation (2.1) becomes:

$$P(D|S_1 \cap S_2) = \frac{P(D)P(S_1|D)P(S_2|D \cap S_1)}{P(D)P(S_1|D)P(S_2|D \cap S_1) + P(\overline{D})P(S_1|\overline{D})P(S_2|\overline{D} \cap S_1)} \tag{2.6}$$

and for p number of symptoms and letting $D_1 = D$ and $D_2 = \overline{D}$, the formula becomes:

$$P(D_1|S_1 \cap S_2 \cap \ldots \cap S_p) = \frac{P(D_1)P(S_1|D_1)P(S_2|D_1 \cap S_1)\ldots P(S_p|D_1 \cap S_1 \cap \ldots \cap S_{p-1})}{\sum_{i=1}^{2} P(D_i)P(S_1|D_i)P(S_2|D_i \cap S_1)\ldots P(S_p|D_i \cap S_1 \cap \ldots \cap S_{p-1})} \tag{2.7}$$

When the symptoms are independent, equation (2.7) becomes:

$$P(D_1|S_1 \cap S_2 \cap \ldots \cap S_p) = \frac{P(D_1)P(S_1|D_1)P(S_2|D_1)\ldots P(S_p|D_1)}{\sum_{i=1}^{2} P(D_i)P(S_1|D_i)P(S_2|D_i)\ldots P(S_p|D_i)} \tag{2.8}$$

or simply written as:

$$P(D_1|S_1 \cap S_2 \cap \ldots \cap S_p) = \frac{P(D_1)\prod_{j=1}^{p} P(S_j|D_1)}{\sum_{i=1}^{2} P(D_i)\prod_{j=1}^{p} P(S_j|D_i)} \tag{2.9}$$

Similarly, when the symptoms are not present, the diagnostic probability becomes:

$$P(D_1|S_1 \cap S_2 \cap \ldots \cap S_p) = \frac{P(D_1) \prod_{j=1}^{p} (1 - P(S_j|D_1))}{\sum_{i=1}^{2} P(D_i) \prod_{j=1}^{p} (1 - P(S_j|D_i))} \tag{2.10}$$

In combining equation (2.9) and equation (2.10), let A denote the complete set of a_j (i.e., $A = \{a_1, a_2, \ldots, a_p\}$). The diagnostic probability becomes:

$$P(D_1|A) = \frac{P(D_1) \prod_{j=1}^{p} ((a_j P(S_j|D_1) + (1-a_j)(1-P(S_j|D_1)))}{\sum_{i=1}^{2} P(D_i) \prod_{j=1}^{p} (a_j P(S_j|D_i) + (1-a_j)(1-P(S_j|D_i)))} \quad a_i \in A \tag{2.11}$$

In deriving the diagnostic probability for various combinations of the symptoms, say $\underline{x}_k$, the formula can be written simply as:

$$P(D_1|\underline{x}_k) = \frac{P(D_1)P(\underline{x}_k|D_1)}{\sum_{i=1}^{2} P(D_i)P(\underline{x}_k|D_i)} \tag{2.12}$$

The Bayesian model has two underlying assumptions. They are:

1. Independence among the symptoms--the occurrence of one symptom is not related to other symptoms.
2. The set of diseases, D_i, in this dissertation i=1,2, is exhaustive and mutually exclusive--the diseases are distinct from each other.

In adjusting the model to correlated symptoms, Scheinok (1972) proposed a solution by using the Bahadur's distribution (1961) and this is presented as follows. Let $P(\underline{x}_k)$ denote $P(D|\underline{x}_k)$. Then for p number of independent symptoms, $P(\underline{x}_k)$ becomes:

$$P(\underline{x}_k) = \prod_{i=1}^{p} \alpha_1^{x_i} (1-\alpha)^{1-x_i} \qquad (2.13)$$

$$= P'(\underline{x}_k)$$

where α_i is the marginal probability or base rate for symptom i and x_i takes a value of either "1" or "0" depending on the presence or absence of the symptom, respectively. Let:

$$z_i = (x_i - \alpha_i)/(\alpha_i(1-\alpha_i))^{\frac{1}{2}} \qquad (2.14)$$

where i=1...p and z_i becomes the standardized variable for the ith symptom with the following property:

$z \sim N(0, 1)$.

Then the following correlation parameters of second, third, ..., pth order can be defined as:

$$r_{ij} = E_p(z_i z_j)$$

$$r_{ijh} = E_p(z_i z_j z_h)$$

.
.
.

$$r_{12\ldots p} = E_p(z_1 z_2 \ldots z_p)$$

where E_p denotes expectation with respect to the distribution of $P(\underline{x}_k)$ which is presented as:

$$P(\underline{x}_k) = P'(\underline{x}_k) f(x_k) \tag{2.15}$$

where

$$f(\underline{x}_k) = 1 + \sum_{i<j} r_{ij} z_i z_j + \sum_{i<j<k} r_{ijk} z_i z_j z_k + \ldots + \sum r_{1,2\ldots n} z_1 z_2 \cdots z_p \tag{2.16}$$

From Table 1.1, the probability that the pattern $\underline{x}_k$ given D is estimated by:

$$P(\underline{x}_k | D) = M_{1k}/M_1 = \alpha_k \tag{2.17}$$

Bailey (1965) suggested an alternative estimator as follows:

$$P^*(\underline{x}_k | D) = (m_{1k} + 1)/(M_1 + 2) = \alpha_k^* \tag{2.18}$$

Then in estimating the correlations for D, the following is used:

$$r_{ij} = 1/M_1 \sum_{\ell=1}^{M_1} z_{i\ell} z_{j\ell}$$

$$r_{ijk} = 1/M_1 \sum_{\ell=1}^{M_1} z_{i\ell} z_{j\ell} z_{k\ell}$$

$$\vdots$$

$$r_{12\ldots p} = 1/M_1 \sum_{\ell=1}^{M_1} z_{1\ell} z_{2\ell} \cdots z_{p\ell}$$

These correlation estimates are then substituted into equation (2.16) and the probability estimates into equation (2.14). $P(\underline{x}_k)$ is derived and this is substituted into equation (2.12) to obtain the desired diagnostic probability.

These computations can become tedious even with a small number of symptoms. Davies (1972), however, demonstrated that with correlated symptoms, the diagnostic probability is simply the proportion of patients having the disease, out of the total having the symptom pattern, $\underline{x}_k$, which is the exact formula of equation (9) in the previous chapter.

Further references on the Bayesian model and its application can be found in studies by Warner et al. (1961), Fraser and Franklin (1974), Overall et al. (1963), Lusted (1968), Vanderokas (1967), Barnoon and Wolfe (1972), Cornfield (1967), Hall (1967), Parker (1967), Schmidt (1971), and Gustfatason (1969).

Pattern Recognition

Binary Regression

Ordinary Least Squares (BLS). For p number of symptoms, the binary regression model is formulated as follows:

$$Y_i = a + \sum_{j=1}^{p} b_j S_{ij} + e_i \qquad (2.19)$$

where

Y_i = the disease of the ith patient (i=1,...,N) and Y_i takes on a value of one or zero depending on the presence or absence of the disease outcome, respectively;

a = constant term;

b_j = regression weight or coefficient for the jth symptom;

S_{ij} = the jth symptom for the ith patient (j=1,...,p) and it takes on a value of one or zero depending on the presence or absence of the symptom for the ith patient, respectively; and

e_i = random error of the ith patient with the following property:

$$e \sim N(0,\sigma^2).$$

The matrix formulation of the binary regression model is as follows:

$$Y = SB + E \qquad (2.19.1)$$

where

Y = vector of disease outcome of ones and zeroes of dimensions N x 1;

S = matrix of symptoms with ones or zeroes of dimensions N x (p + 1) including the constant term;

B = vector of regression parameters, including the constant term, of dimensions (p + 1) x 1; and

E = vector of random errors of dimensions N x 1 with the following property;

$$E \sim N(0, I\sigma^2)$$

The objective of the binary regression model is to minimize the sum of squared errors, E'E. This is done by reformulating equation (2.19.1) as:

$$E = Y - SB. \tag{2.19.2}$$

Premultiplying both sides of equation (2.19.2) by its irrespective transpose to get positive squared errors, the above equation becomes:

$$\begin{aligned} E'E &= (Y - SB)'(Y - SB) \\ &= Y'Y - 2Y'SB + B'S'SB \end{aligned} \tag{2.19.3}$$

Differentiating equation (2.19.3) with respect to B and setting the resultant matrix equation equal to 0 yields

$$\frac{\partial E'E}{\partial BB'} = -2S'Y + 2S'SB = 0 \tag{2.19.4}$$

Replacing B by $\underline{b}$ to denote an estimated parameter, the previous step provides the <u>normal equations</u> as follows:

$$(S'S)\underline{b} = S'Y \tag{2.19.5}$$

If the p normal equations are independent, (S'S) is non-singular, and its inverse, $(S'S)^{-1}$, exists. Then equation (2.19.5) can be rewritten as:

$$\underline{b} = (S'S)^{-1}S'Y \tag{2.19.6}$$

The solution $\underline{b}$ has the following properties:

1. It is an estimate of B which minimizes the sums of squares error E'E irrespective of any distributional properties of the errors.
2. The elements of $\underline{b}$ are linear functions of the observations $Y_1, Y_2, \ldots, Y_n$ and provide unbiased estimates of the elements of B, irrespective of distributional properties of the errors.

The procedure from equation (2.19.2) to the derived estimate $\underline{b}$ in equation (2.19.6) is known as the ordinary least squares procedure. Since both the disease outcome and symptoms take on value of ones or zeroes, Haase (1976) found that the prediction outcome ($\hat{Y}$) which is derived from (2.19.1) as:

$$\hat{Y} = S\,\underline{b}$$

or

$$\hat{Y} = b_0 + S^*\underline{b}^*$$

where:

S^* = a N x p matrix of symptoms;

b_0 = estimated constant term; and

b^* = a pxl vector of estimated regression weights without the constant term,

is in fact the diagnostic probability. Hence, for a given pattern, say $\underline{x}_k$, the diagnostic probability is given by:

$$P(D|\underline{x}_k) = b_0 + \underline{b}^*\underline{x}_k \,. \qquad (2.20)$$

It is very important to note at this point that the estimated diagnostic probability's value is highly dependent upon the values of the estimated regression weights. These weights are derived from the matrix $(S'S)^{-1}$ and vector S'Y.

The ordinary least squares solution to the binary regression model is inappropriate when the errors have unequal variances or are intercorrelated. In the former case, the variance-covariance matrix of the errors is a diagonal matrix with unequal diagonal elements. In the latter case, the off-diagonal elements of the variance-covariance matrix are non-zero values so that the matrix is still symmetric but no longer diagonal. Weighted least squares with a properly estimated variance-covariance matrix can be used to correct for either or both of these problems.

For cases in which the observations are highly intercorrelated, or multicollinear, the matrix (S'S) approaches singularity. In such situations, the variances of the estimated regression weights become highly unstable, resulting in a highly unstable binary regression equation which is sensitive to changes in the data set. Ridge regression is an appropriate technique to use in this situation.

In some instances the estimated diagnostic probability, $P(D|\underline{x}_k)$, can become greater than one, an overestimate, or

become less than zero, an underestimate for all three forms of binary regression. As noted earlier, the values of the estimated diagnostic probability depend on the values of the estimated regression weights which are also subject to underestimation and overestimation. In either case, the diagnostic probability is reset to one if it is greater than one and reset to zero if it is less than zero.

The description of the two forms of solution to the ordinary least square binary regression will be presented in the following two sections. Further reference on the ordinary least squares binary regression can be found in Draper and Smith (1966), Cohen and Cohen (1975) and Kerlinger and Pedhazur (1973). References on the application of the binary regression model in medicine can be found in Feldstein (1966) and Elwood and Mackenzie (1971).

Weighted Least Squares (BWLS). As stated earlier, when the error variance is heterogeneous, it is necessary to amend the estimation procedure by using the weighted least squares method. The key step is to transform the outcome Y_i, to a new variable, Z_i, such that it satisfies the condition of homogeneous error variances. The new transformed model becomes:

$$Z = SQ + F \qquad (2.22)$$

where

Z = vector of new transformed observation of dimensions (N x 1);

Q = vector of new weights transformed from vector B of (2.19.1). This vector has dimensions ((p+1) x 1);

S = matrix of symptoms of ones and zeroes with dimensions (N x (p+1); and

F = vector of errors of the new transformed model of dimensions (N x 1) with the following properties:

$$F \sim N(0, I\sigma^2)$$

Computationally, the weighted least squares can be performed by the two stage method (Neter and Wasserman, 1974) and is described in the following manner:

1. obtain the estimated outcome ($\hat{Y}_i$) for each patient or case by the ordinary least squares method; and
2. define a new variable for each case, w_i, by:

$$w_i = 1/(\hat{Y}_i(1 - \hat{Y}_i))$$

obtaining a diagonal matrix, W, of rank (N x N).

The new regression weights or the "weighted" regression weights, $\underline{q}$, become:

$$\underline{q} = (S'WS)^{-1}S'WY \qquad (2.23.1)$$

where $\underline{q}$ is the estimate of Q.

Hence, the estimated diagnostic probability for pattern, $\underline{x}_k$, for the weighted least squares model is given by:

$$P(D|\underline{x}_k) = q_0 + \underline{q}^*\underline{x}_k \qquad (2.24)$$

where

q_0 = the constant term estimated by the weighted regression model; and

$\underline{q}^*$ = vector of new regression weights estimated by the weighted regression model.

Ridge Regression (BR). In the situation of highly correlated symptoms, the variances of the regression estimates become highly unstable when derived by the ordinary least squares procedure. The estimated values of the regression weights will change with slight changes in the data set. Thus, there is difficulty in determining the contribution of each symptom to the outcome of the disease. Therefore, it is necessary to stabilize the variance of the regression estimates. One technique is by the ridge regression method (Hoerl, 1964, 1970a, 1970b; Marquardt & Snee, 1975). The method consists of adding a constant term, c where c lies between 0.1 and 1.0, to the estimation procedure,

$$\underline{b}^* = (S'S + c)^{-1} S'Y \qquad (2.24.1)$$

and at a certain point of c, say c*, the variance of the estimated regression weights become stabilized. That is, let $V(b^*_{ij})$ be the variance of the ith estimate regression weight at point c_j and ε be a predefined amount of change such that the following condition holds:

$$\{V(b^*_{i,j+1}) - V(b^*_{ij})\} \leq \varepsilon \ .$$

Then when the variance of estimated regression weights is plotted against the various values of c, the following properties will be found:

1. at a certain value of c, say c*, the variances of the regression weights will all stabilize and have the general characteristics of an orthogonal system;
2. the weights will not have unreasonable values with respect to the symptom for which they represent rates of change; and
3. any weights with apparently incorrect signs at c = 0 will have changed to have the proper signs.

The curve connecting the points for all values of c is known as the ridge trace, and the above properties not only hold for a single estimate for a single symptom but for all p estimates.

Weighted Ridge Regression (BRWLS). This model combines ridge regression and the weighted least squares method. The resulting weighted ridge coefficients or weights are substituted into the first stage of weighted least squares instead of the ordinary least squares regression weights. The second stage of the weighted least squares procedure remains the same. This model is intended

to correct both for heterogenous error variances and highly interrelated symptoms.

Logistic Discrimination (LD)

A second pattern recognition method maximizes the relationship between the presence or absence of the disease and a linear combination of the symptoms. This method was developed by R. A. Fisher (1936) and is commonly known as linear discriminant analysis. The description of this technique can be found in Morrison (1967), Tatsuoka (1971), Van de Geer (1971), Timm (1974), and Bock (1975).

Conventional discriminant analysis only applies to variables that are continuous, so when the variables are dichotomous in nature it becomes inappropriate. Anderson (1972a, 1972b, 1973, 1974) proposed logistic discrimination which was originally introduced by Cox (1966) as a solution to this problem. Essentially, the mathematical representation of the model is equivalent to the binary regression model which can be presented as follows:

$$Y = SB + E$$

where

Y = a vector of observations or disease outcomes with ones and zeroes of dimensions (N x 1);

S = a matrix of symptoms of ones and zeroes of dimensions (N x (p+1)), including the constant term;

B = vector of discriminant weights of dimension $(p+1 \times 1)$; and

E = vector of random error of dimensions $(N \times 1)$.

However, the algorithm in deriving the estimated discriminant weights is different from that used in the binary regression model. In developing the mathematical algorithm for this model, let a_{ij} represent the ith row and jth column of the matrix S. Then the diagnostic probability of an outcome, say D, is given by (Cox, 1970):

$$P(D|S) = \frac{e^{a_i B}}{1 + e^{a_i B}} \tag{2.25}$$

and its complement is:

$$P(\bar{D}|S) = \frac{1}{1 + e^{a_i B}} \tag{2.26}$$

The above two equations can be rewritten as the log odds ratio as:

$$\lambda_i = \log_e \frac{P(D|S)}{P(\bar{D}|S)} = a_i B \, . \tag{2.27}$$

Then the likelihood of $Y_1, Y_2, \ldots, Y_N$ independent dichotomous outcomes is:

$$\frac{\prod_{i=1}^{N} e^{a_i B}}{\prod_{i=1}^{N} (1+e^{a_i B})} = \frac{\exp(\sum_{s=1}^{p} B_s t_s)}{\prod_{i=1}^{N} (1+e^{a_i B})} \tag{2.28}$$

where

$$t_s = \sum_{i=1}^{N} a_{is} Y_i$$

Thus, the log likelihood of the above equation becomes

$$L(B) = \sum_{s=1}^{p} B_s T_s - \sum_{i=1}^{N} \log(1 + e^{a_i B}) \qquad (2.30)$$

The solutions for the estimated parameters are derived by the Newton-Raphson iterative numerical procedure as described by Bock (Bose, 1970) which is illuatrated as follows: Let e be a criterion value to stop iteration

Γ_i be the value of the parameters at the ith iteration;

Δ_i be the increment value at the ith iteration

then

$$\Delta_i = -(I)^{-1} F \qquad (2.31)$$

where I is the matrix of the second derivative and F is the vector of first derivatives of equation (2.30) with respect to B (Cox, 1970). The increment is, therefore, simply the product of the negative of the inverse of the second derivative and the vector of the first derivatives. The values for the parameters at the (i + 1)th iteration are:

$$\Gamma_{i+1} = \Gamma_i + \Delta_i ; \qquad (2.32)$$

the iteration stops when the vector of F is less than or equal to e (i.e., $F \leq e$), and the final Γ is the solution for the estimated discriminant weights B. These estimated weights are then substituted into equation (2.25) to derive the estimated diagnostic probabilities.

There are two key assumptions to this model. They are:

1. the populations, the diseased and the non-diseased populations, are multivariate normal with equal variance-covariance matrices; and
2. the populations are multivariate independent and dichotomous in nature.

Other techniques for deriving the solution besides the Newton-Raphson solution can be found in Walker and Duncan (1967) and Jones (1975).

Application of this model to medicine can be found in Truett et al. (1967), Halperin et al. (1971) and Hartz and Rosenberg (1975).

The Entropy Minimax Pattern Discovery (EMPD)

The term entropy refers to the statistical measure of uncertainty. This method was developed from information theory by Christenson (1967, 1968, 1972, and 1973). The key concept of this method is to define symptom subsets that are capable of acting as predictors of a disease outcome. If the presence or absence of a symptom

contributes significantly to a change in the probability of a given outcome, it will be classified as a determinant of the outcome of the disease. The term determinant does not imply deterministic or causative in nature. The purpose of this model is to minimize uncertainty. The model assumes that the measurement of uncertainty has the following properties:

1. uncertainty is a continuous function of the probabilities of various outcomes;
2. greater relative weights are given to occurrence of rare events than to occurrence of common events because rare events convey more information than events that agree with previous prediction; and
3. additivity--the uncertainty associated with two or more independent sources is just the algebraic sum of uncertainty associated with each taken separately.

Given the above properties and give n possible outcomes, each with probability of occurrence P_n, Shannon and Weaver (1964) postulated the measure for the average information per outcome for the discrete case is

$$H = E(-\log_2 P(x)) = -\sum_{i=1}^{n} P_i \log_2 P_i \qquad (2.33)$$

where $\sum_{i=1}^{n} P_i = 1.$

The function H is maximized when $P_i = 1/n$ for all i. To derive the maximum of the above equation, take the derivative of H with respect to P_i

$$\frac{\partial H}{\partial P_i} = -(\log_2 e + \log_2 P_i) + (\log_2 e + \log_2 P_n) \qquad (2.34)$$
$$= -\log_2 (P_i/P_n)$$

Setting

$$\frac{\partial H}{\partial P_i} = 0$$

equation (2) becomes

$$H_{max} = -\sum_{i=1}^{n} (1/n)\log_2(1/n) = \log_2 n \qquad (2.35)$$

The attributes can be partitioned into cells and can be repartitioned into sets of disjointed cells whose sum fill the space. This repartitioning of cells is referred to as screening. Hence, the probability of an outcome for a given cell and jth screening is given by

$$P(D|\text{ith cell, jth screening}) = P_{D|ij}$$

The measure of uncertainty for the ith cell and jth screening becomes

$$H_{ij} = -\sum P_{D|ij} \log_2 P_{D|ij}$$

Summing across outcomes, the measure of uncertainty for the jth cell is

$$H_j = - \sum_{i=1}^{k} P_{ij} \sum_{d=1}^{D} P_{d|ij} \log_2 P_{d|ij} \quad (k = \text{no. of cells}) \qquad (2.36)$$

where

$$P_{ij} = \frac{n_{ij} + u_{ij}}{n + u_j} ;$$

$$P_{d|ij} = \frac{n_{ijk} + w_{ijk}}{n_{ij} + w_{ij}} ;$$

n = total number of events in the sample;

n_{ijk} = number of events with outcome D in the ith cell and jth screening;

n_{ij} = total number of events across D for the ith cell and jth screening;

w_{ijk} = theoretical number of outcome event; and

w_{ij} = the total sum of theoretical event for both happening and non-happening outcome.

The ratios have the following meanings:

$\frac{w_{ijk}}{w_{ij}}$ = priori probability of the D outcome in the jth cell; and

$\frac{u_{ij}}{u_j}$ = priori probability of finding even in the ith cell.

The measure H_j determines how successful the information is in separating the outcome into individual cells in the feature space (i.e., the amount by which the screening has reduced the average uncertainty in predicting an outcome given a set of attributes).

The "best" screening that partitioned the feature space is the one that minimized uncertainty or entropy. The final results are the probabilities of outcome for various patterns of attributes.

These models may be summarized in Table 2.1.

The following chapter will present the theoretical foundation and the algorithm to simulate each individual class of the uncertainty structure when the levels of the correlation have been predefined as presented in Figure 1.4.

Table 2.1

Summary of Probability Models

	Paradigm		Name	Model	Assumptions/Properties
A.	Probabilistic explanation		Bayesian	$\theta = \frac{P(\underline{S}\|D_1)P(D_1)}{\sum_{i=1}^{2} P(\underline{S}\|D_i)P(D)_i}$	1. Symptoms are mutually independent 2. The diseases are exhaustive and mutually exclusive
			Bayesian with Bahadur's correction	Bahadur's expansion factor	1. The diseases are exhaustive and mutually exclusive
B.	Pattern recognition	1.	Binary regression		
			a. Ordinary least squares	$\theta = B_0 + \sum BS + e$	$e \sim N(0, I\sigma^2)$
			b. Weighted least squares	$\theta = Q_0 + \sum QS + f$	$f \sim N(0, I\sigma^2)$
			c. Ridge regression	$\theta = B_0^* + \sum B^*S + g$	
			d. Weighted ridge	$\theta = B_0^* + \sum B^{**}S + h$	
		2.	Logistic discrimination	$\theta = \frac{e^{-B_0 + \sum BS}}{1 + e^{-B_0 + \sum BS}}$	1. The symptoms are multivariate normal with equal dispersion matrices 2. The symptoms are multivariate independent dichotomous
		3.	Entropy minimax pattern discovery	$H = - \sum \theta \log_2 \theta$	1. Continuous function of the probabilities of various outcomes 2. Greater relative weights are given to occurrence of common events 3. Additivity

CHAPTER III

SIMULATION

Prior to simulating the classes in Figure 1.3, consider the situation for a single symptom. The probability that it will have n_1 number of occurrences in a population size of N is simply:

$$P(n_1) = \frac{N!}{n_1!\ n_2!}\ P^{n_1}\ (1-P)^{n_2} \tag{3.1}$$

where $N = n_1 + n_2$ and where P is the marginal proportion of the symptom. This is known as the binomial distribution (Hasting and Peacock, 1975). It is noted that the values of n_1 can be greater or equal to zero and less than or equal to N. Extending this to p number of mutually exclusive symptoms, the joint distribution of the symptoms, $n_1, n_2, \ldots n_p$, where n_j is the number of occurrences for the jth symptom with marginal proportion P_j, is (Johnson, 1969) as follows:

$$P(n_1, n_2, \ldots, n_p) = N! \prod_{j=1}^{p} (P_j^{n_j}/n_j!) \tag{3.2}$$

where $n_j \geq 0$ and $N = \sum_{j=1}^{p} n_j$. This distribution is known as the multinomial distribution.

Since the symptoms are not necessarily mutually exclusive, attention should be given to their intercorrelation and also the correlation of each with the occurrence of the disease. This can be represented in the following matrix (Figure 3.1), R_i, where i denotes the ith class as presented in Figure 1.3:

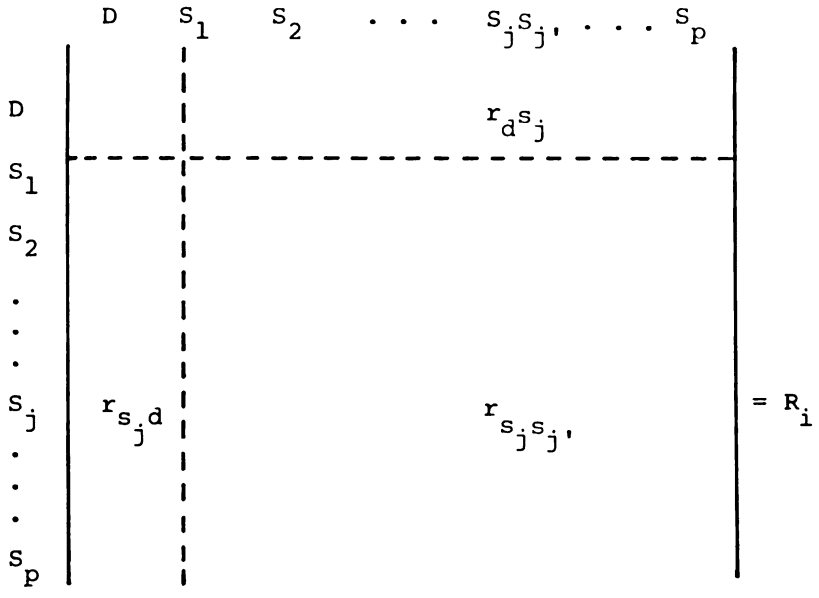


Figure 3.1 The Relational Structure of a Disease and p Symptom.

where r_{ds_j} is the correlation between the disease outcome and the jth symptom and $r_{s_j s_{j'}}$, $(j \neq j')$, is the intercorrelation between the jth and j'th symptom. Since the disease and the symptoms are dichotomous with only ones and zeroes, denoting their presence and absence,

respectively, the correlations are phi-coefficients. In terms of probability, this coefficient can be represented as follows:

$$r_{ds_j} = \phi_{phi} = (P(D \cap S_j) - P(D)P(S_j))/((P(D)(1-P(D))(P(S_j)(1-P(S_j))^{\frac{1}{2}} \quad (3.3)$$

Given the marginal proportions or base rates of the disease and the symptoms, P_d and P_{s_j}, respectively, the above correlation matrix is reformulated into a variance and covariance matrix, Σ, as in Figure 3.2:

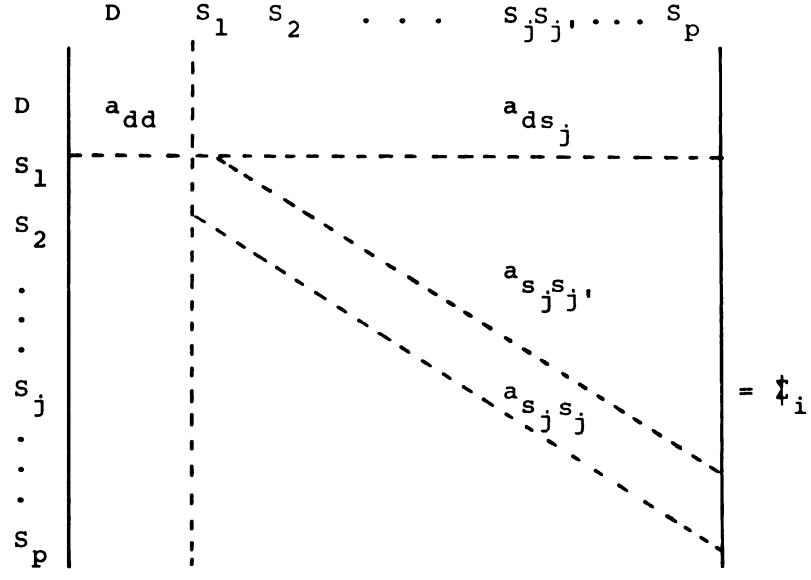


Figure 3.2 The Covariance Structure of a Disease and p Symptoms.

where

$$a_{dd} = P_d(1-P_d);$$

$$a_{s_j s_j} = P_{s_j}(1-P_{s_j}); \text{ and}$$

$$a_{ds_j} = r_{ds_j}(a_{dd}a_{s_j s_j})^{\frac{1}{2}}; \text{ and}$$

$$a_{s_j s_{j'}} = r_{s_j s_{j'}}(a_{s_j s_j}a_{s_j s_{j'}})^{\frac{1}{2}}.$$

It should be noted that a term of the form, P(1-P), is the variance of the disease or the symptom.

The Simulation Computer Routine

A computer program was written by Scheifly (1974), to generate a multivariate continuous distribution with a given mean vector and variance-covariance matrix. In modifying the program to generate the multinomial distribution, the steps comprising this generation are as follows:

1. Generation of independent random variables which are uniformly distributed between zero and one.
2. The generated uniform variates are then combined to form normal deviates with zero means and with the identity matrix for the covariance matrix.

3. The normal deviates are then generally transformed to obtain the desired structure of means and variance-covariance structure.
4. The resultant matrix is then transformed back into probability terms and each variate is assigned a one or zero according to whether the probability is greater or less than the marginal probability of that variable.

The following description elaborates the above steps.

The uniform random variate is generated by the mixed congruential method (Mihram, 1972). This technique can be represented by the following equation:

$$U_k = (aU_{k-1} + c) \pmod{m} \quad k=1,2,...$$

where a and c are constants, U_k is the kth recursion, and U_0 is known as the seed set in the initial recursion. The residual is then divided by P. The values of a, c, and P are chosen as to maximize the period of the generator that produce numbers which behave as if they are random. In terms of these three constants, the kth pseudorandom variate in the sequence is given by

$$U_k = a^k U_0 + c(a^k - 1)(a - 1) \pmod{m} \tag{3.5}$$

The generated sequence of uniform variates are then converted to normal variates by the Teichroew's technique (Knuth, 1968) which is an approximation of the inverse of the probability function for the standard normal distribution. His procedure generates 12 independent random variables, $U_1, U_2, \ldots, U_{12}$, uniformly distributed between zero and one. Then, R is defined as follows:

$$R = (U_1 + U_2 + \ldots + U_{12} - 6)/4. \tag{3.6}$$

The normal variate, z, is then approximated by

$$z = ((((a_1R^2 + a_2)R^2 + a_3)R^2 + a_4)R^2 + a_5)R \tag{3.7}$$

where

$a_1 = .029899776;$

$a_2 = .008355968;$

$a_3 = .076542912;$

$a_4 = .252408784;$ and

$a_5 = 3.949846138.$

This z is only a point in the N x p matrix. In order to obtain the total entries of the matrix, this z is generated N x p times. The result matrix is Z of dimensions N x p.

In transforming the matrix Z to the desired matrix, X, which is distributed with the given mean vector, $\underline{u}$, and variance-covariance matrix, Σ, the following linear transformation is necessary:

$$\underset{N\ P}{X} = TZ + \underline{u}$$

where

$TT' = \ddagger$ (T is the cholesky factor of $\ddagger$).

In transforming the entries of the matrix X into discrete entries, the following procedure is used:

$$z^*_{ij} = (x_{ij} - u_i)/J_i \tag{3.9}$$

where u_i is the given mean and J_i is the ith given standard deviation. The new variable, z^*_{ij}, is converted into a probability or the area under the standard normal curve by numerical approximation according to the following equation:

$$P(z^*_{ij}) = \int_{\infty}^{z^*_{ij}} \frac{1}{(2\pi)^{\frac{1}{2}}} \cdot e^{\frac{1}{2}z^{*2}_{ij}} dz \tag{3.10}$$

By the rejection method (Hasting and Peacock, 1975), the entry on the ith column and jth row is assigned a zero or one according to the following rule:

$$y^*_{ij} = \begin{cases} 1 & \text{if } P(z^*_{ij}) < P_i \\ 0 & \text{if } P(z^*_{ij}) > P_i \end{cases} \tag{3.11}$$

where P_i is the given marginal proportion of the ith symptom or the disease.

In this dissertation, the marginal proportion or the base rate for the disease is set at 0.2 and the marginal proportions for the symptoms, of which there are three, is set at 0.5. Given such marginal proportions, the maximum positive correlation between symptoms and disease is 0.5 and among symptoms is 1.0. The number of cases, N, is set at 300.

Given the above, Table 3.1 represents the partitioning of uncertainty. It must be noted carefully that this table is a reformulation of Figure 1.4 (page 10). Because the maximum correlation between symptoms and the disease is 0.5, the medium and the high categories will be absorbed under the label "High."

Table 3.1

The Simulated Structure of Uncertainty

		Intercorrelations of the Symptoms		
		Low	Medium	High
		0.00-0.30	0.31-0.50	0.51-1.00
Correlation with the disease	Low 0.00-0.30	I	II	III
	High 0.31-0.50	IV	V	VI

It should be noted that the resultant matrix of Y* of zero and one entries is generated from an underlying continuous distribution and the correlations computed from his matrix are in fact tetrachoric correlations. The relationship between the phi-coefficient, ρ_{ij}, and the tetrachoric coefficient, ρ'_{ij}, between the ith and jth symptom, is developed by Pearson (1900) and cited by Lord and Novick (1974) as:

$$\frac{\sigma_i \sigma_j \rho_{ij}}{\delta(\gamma_i)\delta(q_j)} = p'_{ij} + \frac{1}{2}\gamma_i\gamma_j\rho'^{2}_{ij} + \frac{1}{6}(\gamma_i^2 - 1)(\gamma_j^2 - 1)\rho'^{3}_{ij}$$

$$+ \frac{1}{24}\gamma_i\gamma_j(\gamma_i^2 - 3)(\gamma_j^2 - 3)\rho'^{4}_{ij}$$

$$+ \frac{1}{120}(\gamma_i^4 - 6\gamma_i^2 + 3)(\gamma_j^4 - 6\gamma_j^2 + 3)\rho'^{5}_{ij} + \ldots \qquad (3.12)$$

where γ_i and γ_j are the cutoff points for the ith and jth variable, respectively, and σ_i and σ_j are its standard deviations. In the special case where $P_i = P_j = 0.5$, the relationship is simplified as:

$$\rho_{ij} = \sin(\pi\rho_{ij}/2) \qquad (3.13)$$

The following chapter will present the analysis on the generated samples by the above simulation algorithm. The analysis will include (1) the statistical test of equivalence between the generated sample and the predefined population for each individual class in the uncertainty structure,

(2) the statistical test of equivalence between the randomly split samples for each individual class in the uncertainty structure, (3) the statistical test of severity of multicollinearity for each generated class, (4) the evaluation of each probabilistic model in terms of discrepancy indices, (5) the evaluation of each probabilistic model in terms of prediction indices, (6) the evaluation of relative performance of each probabilistic model in terms of utility functions, losses, and gains, (7) the evaluation of relative performance of each probabilistic model for three special classes of the uncertainty structure with the same evaluation indices and utility function, and finally (8) an application to a set of real data.

CHAPTER IV

DATA ANALYSIS

The population correlation matrix, R_p, and the variance-covariance matrix, Σ_p, are defined and shown in Appendix A. The sample variance-covariance matrices, Σ_s, and correlation matrices, R_s, were then obtained by the generation routine described in Chapter III, also shown in Appendix A. The sample variance-covariance matrices are then tested to determine if they are statistically equivalent to the population matrices. This procedure translates into testing the following hypothesis:

$$H_0: \Sigma_s = \Sigma_p$$

against the alternative:

$$H_1: \Sigma_s \neq \Sigma_p .$$

The test statistic used (Morrison, 1976) is as follows:

$$L = v(\log_e|\Sigma_p| - \log_e|\Sigma_s| + \text{tr}\ \Sigma_s\Sigma_p^{-1} - p) \qquad (4.1)$$

where p is the number of symptoms plus the disease outcome and v is equal to (N - 1) where N is the population size. The test statistic, L, is distributed as a chi-square

variate with $\frac{1}{2}(p(p+1))$ degrees of freedom if the null hypothesis is true. For moderate N, Bartlett (1954) has suggested the scaled statistic as:

$$L' = \left\{ 1 - \frac{1}{6\ (N-1)} (2p + 1 - 2/(p+1)) \right\} L \qquad (4.1.1)$$

as an improvement on the chi-square approximation. The results of the tests are presented in Table 4.1

Table 4.1

Test of Fit for Sample Variance-Covariance Matrices With the Specified Population Variance-Covariance Matrices (N = 300; p = 4)

Class	L	L'	df	Significance Probability P
I	11.03	10.98	10	.50
II	10.32	10.27	10	.50
III	11.05	11.00	10	.50
IV	9.97	9.93	10	.50
V	10.48	10.43	10	.50
VI	12.47	12.41	10	.25

*Significant at the 0.5 level.

From these results, the sample variance-covariance matrices are not significantly different from the specified population variance-covariance matrices.

Each class of 300 cases was then shuffled and randomly divided into two equal sub-samples of 150 each, called Sample I and Sample II. The resultant variance-covariance matrices of the "split" samples for each class are also shown in Appendix A. The hypothesis tested becomes:

$$H_0: \; \Sigma_1 = \Sigma_2$$

against the alternative:

$$H_1: \; \Sigma_1 \neq \Sigma_2$$

The test statistic used (Morrison, 1976) is as follows:

$$M = \sum_{i=1}^{2} n_i \log_e |\Sigma_p,| - \sum_{n=1}^{2} n_i \Sigma_i \qquad (4.1.2)$$

where n_i is equal to $(N_i - 1)$, N_i is the sample size of the ith sample and Σ_p, is the pooled matrix of Σ_1 and Σ_2. Box (1949) has found that if the scale factor,

$$G = 1 - \frac{2p^2 + 3p - 1}{6(p + 1)} \left(\sum_{i=1}^{2} \frac{1}{n_1} - \frac{1}{\sum_{i=1}^{2} n_i} \right), \qquad (4.1.3)$$

is introduced into equation (4.1.2) (i.e., G x M), GM is approximately distributed as a chi-square variae with degrees of freedom equal to $\frac{1}{2}(p(p+1))$. The results of the tests for equivalence between split samples are presented in Table 4.2.

Table 4.2

Test of Equivalence of Split-Samples Variance-Covariance Matrices ($N_1 = N_2 = 150$; $p = 4$)

Class	GM	df	Significance Probability P
I	1.39	10	.99
II	3.39	10	.99
III	6.44	10	.75
IV	1.55	10	.99
V	10.29	10	.50
VI	2.02	10	.99

Again the split-half samples for each class show no statistical differences at the 0.5 level of significance. From these results, it can be concluded that the similarities between the specified population and the sample and between split samples are statistically assured.

Before proceeding further into the analysis, the first sample, Sample I, of each class is tested for the severity of multicollinearity (i.e., highly interrelated symptoms). This is achieved by testing the following hypothesis:

$$H_0: \ |S'S| = 1$$

against the alternative:

$$H_1: \ |S'S| < 1$$

where S is the matrix of symptoms of dimension (p x p). If S is a standardized matrix, then (S'S) will be a correlational matrix and the testing hypothesis can be reformulated as:

$$H_0: (S'S) = I$$

against the alternative:

$$H_1: (S'S) \neq I$$

where I is an identity matrix of the same rank. Barlett (1950) formulated the following test statistic:

$$\lambda = -((N-1) - 1/6(2p+5))\log_e|S'S|$$

where λ distributed as a chi-square variate with degrees of freedom $\frac{1}{2}(p(p-1))$. The results for testing for multicollinearity for each class are presented in Table 4.3.

Table 4.3

Test of Severity of Multicollinearity
for Sample I of Each Class

Class		df	Significance Probability P
I	7.94	3	.025*
II	54.04	3	.005*
III	149.98	3	.005*
IV	13.99	3	.005*
V	76.07	3	.005*
VI	174.16	3	.005*

*Significant at the 0.5 alpha level.

Each class shows the presence of multicollinearity, even in Class I and Class IV which were supposed to have low intercorrelated symptoms. This is not at all surprising. Correlation coefficients of 0.16 are significantly different from zero at the 0.05 level for 150 cases and 3 symptoms. Since most of the correlations among the symptoms exceed that value, the situation represents a significant multicollinear condition.

Since there are only three symptoms, there are $2^3 = 8$ possible combinational patterns. The exact probability for a disease to be positive (present) with pattern $\underline{x}_k$ is calculated as follows:

$$P(D|\underline{x}_k) = \frac{\text{(number of patients with disease that has pattern } \underline{x}_k)}{\text{(number of patients with pattern } \underline{x}_k)}$$

The exact probabilities calculated from the preceding formula for each class and pattern are presented in Table 4.4.

Table 4.4

Exact Probabilities, $P(D|\underline{x}_k)$ for Each Class of the First Sample

Pattern	I	II	III	IV	V	VI
111	.36	.41	.27	.58	.57	.43
110	.22	.13	.12	.21	.36	.20
100	.08	.14	.10	.00	.00	.00
001	.00	.07	.33	.00	.00	.00
011	.12	.24	.00	.00	.00	.00
101	.37	.57	.36	.17	.07	.14
010	.18	.15	.25	.06	.07	.00
000	.08	.03	.10	.00	.00	.00

In evaluating the discrepancies between the estimated probabilities from the exact probabilities for each model, the following three discrepancy measures are used:

1. Mean Square Deviation (MSD):

$$MSD_j = \sum_{i=1}^{p} \frac{(\hat{p}_{jk} - p_{\underline{k}}^*)^2}{2^p} \tag{4.3}$$

where $\hat{p}_{ij}$ denotes the estimated probability for the ith pattern by the jth model and $p_{\underline{k}}^*$ (where $p_{\underline{k}}^* = P(D|\underline{x}_k)$ is the exact probability for the kth pattern. It is worthy to note that the numerator of equation (4.3) is the squared difference between the model's estimates and the true probability. Squaring the differences insures a positive value. Dividing by the denominator which is the number of combinations or patterns and summing across all 2^p possible combinations, equation (4.3) gives the mean square differences within a class.

2. Weighted Mean Square Deviation (WMSD):

$$WMSD_j = \sum_{k=1}^{2^p} \frac{(\hat{p}_{jk} - p_k^*)^2 \cdot \hat{p}_{jk}}{2^p} \tag{4.4}$$

Equation (4.4) is essentially the same as equation (4.3) with the exception of the inclusion of the multiplier, $\hat{p}_{jk}$, in the numerator. This means that more weight is given to the higher probability estimates by the model. In other words, higher probability estimates for the kth pattern by the jth model are assumed to have greater squared differences and vice versa.

3. Misclassification Rate (MR):

$$MR_j = \sum_{d=1}^{2} \sum_{k=1}^{2^p} \hat{p}_{jkd}\bar{p}_d \qquad (4.5)$$

where $\bar{p}_d$ is the marginal probability or the base rate of the dth disease not occurring (i.e., $\bar{p}_d = 1 - p_d$) where d is equal to 1 and 2 in this dissertation. The term, $\hat{p}_{jkd}$, is the estimated probability for the dth disease outcome. It is noted that $\hat{p}_{ij1} = \hat{p}_{ij} = P(D|\underline{x}_k)$ and $\hat{p}_{ij2}$ is simply equal to $(1 - \hat{p}_{ij1})$. Then the multiplication of the terms gives the expected misclassification rate for the kth pattern in the event that the dth disease does not occur. Summing across all possible 2^p patterns, equation (4.5) gives the <u>total</u> expected misclassification rate within a class.

The probability models with the exception of the Bayesian model are then applied to the first sample to obtain the estimated parameters for each symptom as shown in Appendix B. These parameters are then used to derive the conditional or diagnostic probabilities p_{jk} for each class, pattern and model as presented in Appendix C. The performances of these probability models in terms of the above three indices are then obtained and they are presented in Table 4.5.

Table 4.5

Comparison of Models in Terms of Discrepancy Indices

	Discrepancy Indices	Classes													
		I	R[a]	II	R	III	R	R*[b]	IV	R	V	R	VI	R	R**[c]	R***[d]
	SMD	.000	1	.000	1	.000	1	1	.000	1	.000	1	.000	1	1	1
B	MWSD	.000	1	.000	1	.000	1	1	.000	1	.000	1	.000	1	1	1
	MR	.306	4	.335	6	.306	3	4.6	.278	2	.280	1	.255	1	1.3	2.9
	SMD	.006	4	.012	4	.012	4	4	.010	3	.016	5	.013	3	3.6	3.8
BLS	MWSD	.001	4	.003	4	.002	4	4	.002	3	.004	4	.002	3	3.6	3.8
	MR	.311	6	.325	4	.320	5	5	.304	4	.328	6	.309	5	5	5
	SMD	.007	6	.012	4	.027	5	5	.014	4	.025	6	.017	6	5.3	5.15
BWLS	MWSD	.001	4	.003	4	.009	7	5	.004	4	.007	6	.003	5	5	5
	MR	.312	7	.325	4	.355	7	6	.305	5	.335	7	.312	7	6.3	6.15
	SMD	.014	7	.017	7	.030	7	7	.020	5	.031	7	.014	4	5.3	6.65
BR	MWSD	.001	4	.004	7	.002	4	5	.005	5	.009	7	.002	3	5	5
	MR	.265	2	.290	1	.236	1	1.3	.328	6	.327	5	.306	4	5	3.15
	SMD	.006	4	.012	4	.027	5	4.3	.034	6	.013	4	.015	5	5	4.65
BRWLS	MWSD	.001	4	.003	4	.004	6	4.6	.007	6	.004	4	.003	5	5	4.8
	MR	.301	3	.320	3	.293	2	2.6	.269	1	.326	4	.309	5	3.3	2.9
	SMD	.005	3	.008	3	.009	3	3	.002	2	.000	1	.000	1	1	2
LD	MWSD	.000	1	.002	3	.001	2	2	.000	1	.000	1	.000	1	1	1.5
	MR	.308	5	.315	2	.315	4	3.6	.293	3	.285	2	.256	2	2.3	2.9
	SMD	.000	1	.000	1	.001	2	1.3	.042	7	.001	3	.024	7	5.6	3.4
EMPD	MWSD	.000	1	.000	1	.001	2	1.3	.024	7	.000	1	.010	7	5	3.15
	MR	.031	1	.339	7	.331	6	4.6	.328	6	.295	3	.303	3	4	4.3

[a] R = ranking of model within classes.

[b] R* = average ranking of model across classes I, II, and III (pool situation where the symptoms are lowly correlated with the disease outcome).

[c] R** = average ranking of model across classes IV, B, and VI (pool situation where the symptoms are highly correlated with the disease outcome).

[d] R**** = average ranking all classes.

In the situation in which the symptoms have a low correlation with the disease outcome, the Bayesian model (B) has the smallest mean squared deviation (MSD), followed by the Entropy Minimax Pattern Discovery (EMPD) model. The Binary Ridge (BR) model has the highest MSD. In terms of having the smallest weighted mean squared deviation (WMSD), the B model again ranks first with the EMPD model ranking second. The Binary weighted least square regression model (BWLS) ranks first in having the highest WMSD and also in having the smallest misclassification rate (MR) when the Binary Ridge regression (BR) is used as the solution.

In the situation where the symptoms are highly correlated with the disease outcome, the logistic discrimination model (LD) and the B models have the smallest MSD and WMSD. The EMPD has the highest MSD and WMSD. The B model ranks first in terms of having the smallest MR followed by the LD model. The Binary weighted least squares (BWLS), BLS and BR models, all three solutions to the BLS model, have the highest MR, implying these solutions did not improve the BLS model with respect to MR.

When the situations of (1) symptoms having high correlation with the disease outcome and (2) the symptoms having low correlations with the disease outcome are pooled, the B model has the smallest MSD, WMSD, and MR followed by the LD model. The BR, BWLS, and the BRWLS did not improve the ranking of all three indices for the BLS model.

However, the main thrust of this dissertation is diagnostic prediction. The term prediction entails a different and perhaps future event. The indices discussed so far address the quality of the probability models that are available for building a basis for judgment. The assessment of prediction will be done by cross-validating each model on Sample II, the statistically equivalent counterpart to Sample I. This notion is similar to the process used by physicians of determining the best procedure to treat a class of problems (model selection, Sample I) and then evaluating its efficiency and correctness on new patients with the same problem (cross-validating, Sample II).

The process of cross-validation can result in four possible outcomes. They are:

1. the correct prediction or identification of a truly diseased case, also known as true positives (TP).
2. the correct prediction or identification of a truly non-diseased case, also known as true negatives (TN).
3. the incorrect prediction or identification of a truly non-diseased case as having the disease, also known as false positives (FP).
4. The incorrect prediction or identification of a truly diseased case as not having the disease, otherwise known as false negatives (FN).

These outcomes can be represented in the following figure (Figure 4.1):

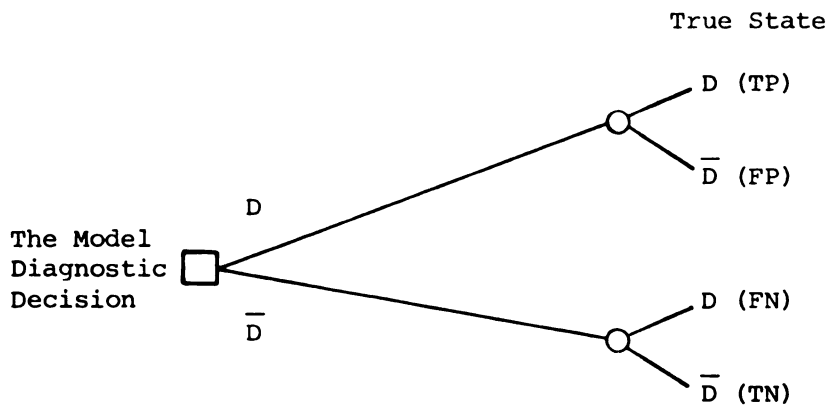


Figure 4.1 Possible Outcome of a Diagnostic Decision.

The above situation can also be reformulated into the following table (Table 4.6):

Table 4.6

Possible Distribution of Cases by Model Decision and the True Outcome

		True State: D	True State: $\overline{D}$
Model's Decision	D	TP(n_1)	FP(n_2)
	$\overline{D}$	FN(n_3)	TN(n_4)

$$\sum_{i=1}^{4} n_1 = N$$

where n_1 and n_4 are the numbers of patients with true positives and true negatives, respectively, n_2 and n_3 are the numbers of patients with false positives and false negatives, respectively. From Table 4.6, the following indices henceforth termed as prediction indices are defined:

1. Sensitivity (SEN): The ability of the model to predict the proportion of patients who truy have the disease. The formula is:

$$SEN = n_1/(n_1 + n_3).$$

The standard error of SEN is found to be:

$$SE(SEN) = \{(\hat{p}_1\hat{q}_1)/(n_1 + n_3)\}^{\frac{1}{2}} \quad \text{where } \hat{p}_1 = n_1/(n_1 + n_3)$$

$$\text{and} \quad \hat{q}_1 = (1 - \hat{p}_1).$$

Hence, the confidence interval for SEN becomes:

$$SEN \pm z_\alpha \cdot SE(SEN) \quad \text{where } 1 - \alpha = \text{confidence level.}$$

The greater the sensitivity, the greater the accuracy of the model in predicting the occurrence of the disease.

2. Specificity (SPEC): The ability of the model to predict the proportion of patients who truly do not have the disease. The formula is:

$$SPEC = n_4/(n_2 + n_4)$$

The standard error of SPEC is found to be:

$$SE(SPEC) = \{(\hat{p}_2\hat{q}_2)/(n_2 + n_4)\}^{\frac{1}{2}} \quad \text{where } \hat{p}_2 = n_4/(n_2 + n_4)$$

$$\text{and} \quad \hat{q}_2 = (1 - \hat{p}_2).$$

Hence, the confidence interval for SPEC becomes:

$$SPEC \pm z_\alpha \cdot SE(SPEC).$$

The greater the specificity, the greater the accuracy of the model in predicting the non-occurrence of the disease.

3. Predictive value (PRED): The Proportion of patients who truly have the disease among those predicted by the model to have it. The formula is:

$$PRED = n_1/(n_1 + n_2)$$

The standard error of PRED is found to be:

$$SE(PRED) = \{(\hat{p}_3\hat{q}_3)/(n_1 + n_2)\}^{\frac{1}{2}} \quad \text{where } \hat{p}_3 = n_1/(n_1 + n_2)$$

$$\text{and} \quad \hat{q}_3 = (1 - \hat{p}_3).$$

Hence, the confidence interval for PRED becomes:

$$PRED \pm z_\alpha \cdot SE(PRED).$$

The greater the predictive value, the more accurate or "precise" is the prediction of the model.

It should be noted that when the values for both SEN and SPEC are one, it implies that the PRED is also one. However, a PRED of one does not necessarily imply a SEN value of one or a SPEC value of one. A SEN value of zero would imply a PRED value of zero and vice versa.

4. Type I error (E1): The proportion of patients which the model predicted as not having the disease among those who truly have the disease. The formula is:

$$E1 = n_3/(n_1 + n_3)$$

or simply the complement of SEN, i.e.,

$$E1 = 1 - SEN.$$

It can also be written in the form of a conditional probability as:

$$E1 = P(\bar{D}_m | D_s)$$

where $\bar{D}_m$ is the model's diagnosis as not having the disease and D_s denotes the true state as having the disease. The standard error of E1 is equivalent to the standard error of SEN as E1 is the complement of SEN. Hence, the confidence interval for E1 is simply:

$$E1 \pm z_\alpha \cdot SE(SEN).$$

5. Type II error (E2): The proportion of cases which the model predicts as having the disease when the patients are in fact non-diseased. The formula is:

$$E2 = n_2/(n_2 + n_4),$$

or simply is the compliment of SPEC; i.e.,

$$E2 = 1 - SPEC.$$

Written as a conditional probability:

$$E2 = P(D_m | \bar{D}_s)$$

where D_m is the model's diagnosis as having the disease and $\bar{D}_s$ denotes the true state as not having the disease. The standard error of E2 is equivalent to the standard error of SPEC as E2 is the complement of SPEC. Hence, the confidence interval for E2 is simply:

$$E2 \pm z_{\alpha} \cdot SE(SPEC).$$

The conventional rule in allocating patients with pattern $\underline{x}_k$ as having the disease, D, or not having the disease, $\bar{D}$, is as follows:

Diagnostic Rule	Decision
$P(D \mid \underline{x}_k) > P(\bar{D} \mid \underline{x}_k)$	Disease
$P(\bar{D} \mid \underline{x}_k) > P(D \mid \underline{x}_k)$	Non-Disease
$P(\bar{D} \mid \underline{x}_k) = P(D \mid \underline{x}_k)$	Equivocal

These decision rules, however, are arbitrary and they are at the discretion of the decision maker. In this dissertation, the criterion of allocation is chosen at π where π is equal to the base rate of the disease in the first sample. Hence, these decision rules are reformulated as follows:

Diagnostic Rules	Decision
$P(D\|\underline{x}_k) \geq \pi$	Disease
$P(\bar{D}\|\underline{x}_k) > \pi$	Non-Disease

This has, in effect, eliminated the equivocal decision and has the advantage of increasing sensitivity of the model to detect diseased cases which have a low base rate. The price of using this decision rule, however, is maximizing the probability of identifying a case as diseased when, in fact, it is non-diseased.

However, this is seen as better than identifying a case as non-diseased when, in fact, it is a diseased case. The reason for this is explained by Neyman (1950) as follows:

> [If the patient is actually well, but the hypothesis that he is sick is accepted, a Type 2 error] then the patient will suffer some unjustified anxiety and, perhaps, will be put to some unnecessary expense until further studies of his health will establish that any alarm about the state of his chest is unfounded. Also, the unjustified precautions ordered by the clinic may somewhat affect its reputation. On the other hand, should the hypothesis (of sickness) be true and yet the accepted hypothesis be (that he is well, a Type 1 error), then the patient will be in danger of losing the precious opportunity of treating the incipient disease in its beginning stages when the cure is not so difficult. Furthermore, the oversight by the clinic's specialist of the dangerous condition would affect the clinic's reputation even more than the unnecessary alarm. From this point of view, it appears that the error of rejecting the hypothesis (of sickness) when it is true is far more important to avoid than the error of accepting the hypothesis (of illness) when it is false. (1950, p. 270, emphasis added)

Increasing the opportunity of committing the former error to reduce the risk of the latter error is one of the pervasive and fundamental rules in medicine which may be stated as: "When in doubt, continue to suspect illness." The logic of this decision rule rests on two assumptions (Scheff, 1963). They are:

1. Disease is usually a determinate, inevitably unfolding process, which, is undetected and untreated, will grow to a point where it endangers the life or limb of the individual, and in the case of contagious disease, the lives of others.
2. Medical diagnosis unlike legal judgment, is not an irreversible act which does untold damage to the status and reputation of the patient.

He further states that: "In light of these two assumptions, it is far better for the physician to chance a Type 2 error than a Type 1 error."

The results for the cross validation in terms of these prediction indices for each model and for each class are presented in Appendix D. These results are then re-tabulated in Table 4.7 to show those classes within each model with the highest and lowest values for SEN, SPEC, and PRED.

Table 4.7

The Class Where Each Model Has the Highest and Lowest Predictive Indices

	Highest			Lowest		
Models	SEN	SPEC	PRED	SEN	SPEC	PRED
Bayesian	VI	IV	IV	I	III	III
Bayesian w/Bahadur Binary	VI	IV	IV	I	III	III
Binary:						
Ordinary Least Squares	V	IV	IV	I	I	I
Weighted Least Squares	V	IV	IV	I	I	I
Ridge Regression	IV,V	III	IV	III	V	III
Weighted Ridge	V	IV	IV-VI	I	II	I
Logistic Discrimination	VI	IV	IV	I	III	I
Entropy Minimax Pattern Discovery	VI	IV	IV	I	III	III

Class VI is the class where the B, BB, LD, and EMPD models have the highest SEN, and Class V is the class where the BLS and its solutions have the highest SEN. In terms of SPEC, all models with the exception of BR have the highest SPEC in Class IV where all models also have the highest PRED. Hence, Class IV, V, and VI, where the symptoms are highly related to the disease outcome, are optimal situations for these models in terms of the three predictive efficiency indices.

All models with the exception of BWLS have the lowest SEN for Class I, and Class III has the lowest SPEC for B, BB, LD, and EMPD. Class III also has the lowest PRED for B, BB, BWLS, and EMPD, and Class I has the lowest PRED for BLS, BR, and LD. Class I and Class III, where the symptoms are lowly correlated with the disease outcome, are "pit" situations for these models in terms of these indices as all models have the lowest values in this class. From the above results, the B, BB, LD, and EMPD models perform similarly in having the highest and lowest predictive efficiency indices.

To determine the relative performance of these models within each class, Table 4.8 is reformulated. The binary regression models have the highest SEN across all classes except Class III. The BR model has the lowest SEN in Classes I, II, and III, and BRWLS model has the lowest SEN in Class IV, and LD and B models have the lowest SEN for Classes V and VI. In terms of having the highest SPEC, the BR model performs the best in Classes I to III and the LD model from IV to VI. In terms of PRED, BRWLS performs the best in Classes I and III while the LD model performs most optimally in Classes II, IV, V, and VI.

Table 4.8

Comparison Across Models Within Class

Predictive Indices		Class					
		I	II	III	IV	V	VI
SEN	Highest	BLS & BWLS	BRWLS & BWLS	B, BB & EMPD	BR	BLS, BWLS, BR, & BRWLS	All except B & LD
	Lowest	BR	BR	BR	BRWLS	BB, B, LD, & EMPD	B & LD
SPEC	Highest	BR	BR	BR	LD & EMPD	B, BB, LD, & EMPD	B & LD
	Lowest	BLS & BWLS	BLS, BWLS, & BRWLS	B, BB, & EMPD	BR	BLS & BWLS	BLS & BRWL
PRED	Highest	BRWLS	B, BB, LD, & EMPD	BWLS & BRWLS	LD & EMPD	B, BB, LD, & EMPD	B & LD
	Lowest	BLS & BWLS	BLS, BRWLS, & BWLS	BR	BRWLS	BLS & BWLS	BLS, BR, & BRWLS

Those classes where the symptoms have low correlation with the disease outcome are now considered. Table 4.9 shows the values and ranking for the three predictive indices across models with their condition. The BLS and BWLS models have the highest SEN but have the lowest SPEC. The BR model has the lowest SEN but ranks first in having the highest SPEC. The B, BB, LD, BRWLS, and EMPD have the best predictive value.

Table 4.9

Performances of Each Model in Terms of Prediction Indices When the Symptoms Are Lowly Correlated With the Disease and Their Ranking

Model	SEN	Rank	SPEC	Rank	PRED	Rank
Bayesian	.76	3	.61	4	.32	1
BB	.76	3	.61	4	.32	1
BLS	.78	1	.54	8	.29	7
BWLS	.78	1	.56	7	.30	6
BR	.31	8	.84	1	.21	8
BRWLS	.73	7	.62	2	.32	1
LD	.75	6	.63	3	.32	1
EMPD	.76	3	.61	4	.32	1

When the symptoms are highly correlated with the disease, Table 4.10, the BR model has the highest SEN, implying that the ridge solution improves the ordinarily BLS in SEN but at the price of losing SPEC. The B has the lowest SEN but has the highest SPEC and PRED. The LD and EMPD share in having the highest PRED.

Table 4.10

Performances of Each Model in Terms of Prediction Indices When the Symptoms Are Highly Correlated With the Disease and Their Ranking

Model	SEN	Rank	SPEC	Rank	PRED	Rank
Bayesian	.86	7	.79	1	.49	1
BB	.88	5	.79	1	.49	1
BLS	.96	2	.64	7	.37	7
BWLS	.92	3	.65	6	.38	5
BR	.98	1	.64	7	.38	5
BRWLS	.90	4	.66	5	.37	7
LD	.86	7	.79	1	.49	1
EMPD	.88	5	.78	4	.49	1

When the conditions of high and low intercorrelated symptoms are pooled, the BLS model has the highest SEN and the BR model has the lowest SEN and PRED but with the highest SPEC as shown in Table 4.11. The BWLS model has the lowest SPEC. The BB model ranks first in having the highest PRED followed by the B, LD, and EMPD models.

Summarizing the above results, Table 4.12 is formulated, as can be seen below.

Table 4.11

Performances of Each Model in Terms of Prediction Indices for Pooled Situation and Their Ranking

Model	SEN	Rank	SPEC	Rank	PRED	Rank
Bayesian	.81	5	.70	3	.40	2
BB	.82	3	.70	3	.41	1
BLS	.87	1	.69	6	.33	7
BWLS	.85	2	.61	8	.34	5
BR	.65	8	.74	1	.30	8
BRWLS	.81	5	.64	7	.34	5
LD	.80	7	.71	2	.40	2
EMPD	.82	3	.70	3	.40	2

Table 4.12

Performance of Models Relative to the Predictive Indices Across Correlational Patterns of Disease and Symptoms

	Correlation With the Disease Outcome						
	Intercorrelation Among Symptoms						
	Low			High			
	Low	High	Pooled	Low	High	Pooled	Overall
SEN	BLS, BWLS	B, BB, EMPD	BLS, BWLS	BR	BB, BLS, BWLS, BR, BRWLS, EMPD	BLS	BLS
SPEC	BR	BR	BR	LD	LD	LD, B, BB	BR
PRED	BRWLS	BRWLS	B, BB, BRWLS, LD, EMPD	LD	LD	LD, BB, B, EMPD	BB

A key question could be asked as to the cost of using these models in each class (i.e., when the symptoms have low correlation with the disease outcome or when the symptoms are highly correlated with the disease outcome). In examining these models to answer this question, the following table represents the consequences for various outcomes. This table is also known as the utility matrix.

Model's Diagnostic Action	True State: D	True State: $\bar{D}$
D	w_{11}	w_{12}
$\bar{D}$	w_{21}	w_{22}

where w_{ij} represents the arbitrary weight given to each outcome. These weights can be in the form of mortality or cost in dollars. They could either be gain--(positive in value) or a loss--(negative in value) or zero (neither gain or loss). Hence, a decision function, E(D), is defined for the mth model as follows:

$$E(D_m) = \sum_{j=1}^{2} \sum_{i=1}^{2} w_{ij} p_i^* \hat{p}_{ijm}$$

where p_i^* denotes the marginal probability or base rate of the disease. In this dissertation, $p_1^* = P(D)$ and $P_2^* = P(\bar{D})$. $\hat{p}_{ijm}$ is the probability of patients having the disease predicted by the mth model for the ith and jth outcome.

To emphasize the Type 1 and Type 2 error differences, let us assume the following weights: $w_{21} = -2$, $w_{12} = -1$, and $w_{11} = w_{22} = 0$, which means that the penalty for committing a Type 1 error is twice as costly as that for committing a Type 2 error, and there is no credit or gain given to the right diagnosis. A loss function can now be

formulated since there will be only loss and no gains. This can be shown in the following matrix.

		True State: D	True State: $\overline{D}$
Model's Diagnostic Action	D	0	-1
	$\overline{D}$	-2	0

The model's performance in terms of this loss function is presented in Table 4.13.

The LD model has the least loss followed by the B and BR models, and the BLS model has the most loss when the symptoms are lowly correlated with the disease outcome. The B and LD models share in having the least loss and the BB model has the greatest loss when the symptoms are highly correlated with the disease outcome. For both situations, the B and LD models again have the least loss with the BLS and BWLS models having the greatest loss.

If credits are given to the right diagnosis by setting $w_{11} = 2$, $w_{22} = 1$, then the amount of credit given to a correct diagnosis of a truly diseased patient is worth twice as much as a correct diagnosis of a truly non-diseased patient. The matrix below references the result of setting $w_{12} = -1$ and $w_{22} = 1$.

Table 4.13

Utility Table in Terms of Losses for Each Model and for Each Class

	Models															
Classes	B	R	BB	R	BLS	R	BWLS	R	BR	R	BRWLS	R	LD	R	EMPD	R
I	.41	2	.40	1	.50	7	.50	7	.41	2	.41	2	.41	2	.41	2
II	.37	1	.44	8	.42	5	.43	6	.39	4	.43	6	.37	1	.37	1
III	.44	6	.44	6	.42	5	.38	1	.40	3	.38	1	.40	3	.44	6
	.40	2	.42	5	.45	7	.44	6	.40	2	.41	4	.39	1	.41	4
IV	.22	1	.22	1	.23	4	.27	5	.27	5	.32	8	.22	1	.30	7
V	.21	2	.20	1	.38	7	.38	7	.32	5	.32	5	.21	2	.21	2
VI	.22	2	.40	8	.29	5	.26	4	.29	5	.29	5	.21	1	.22	2
	.21	1	.27	8	.30	5	.30	5	.29	4	.31	7	.21	1	.24	3
Pooled	.30	1	.34	4	.37	7	.37	7	.35	5	.36	6	.30	1	.32	3

Model's Diagnostic Action	True State: D	True State: $\bar{D}$
D	2	-1
$\bar{D}$	-2	1

In terms of gain, as shown in Table 4.14, the LD model has the most gain followed by the EMPD, BRWLS, and B models with the BR model having the least gain when the symptoms are lowly correlated with the disease outcome. And when the symptoms are highly correlated with the disease outcome, the B and LD models share the highest gain with BRWLS having the least gain. Again, combining the conditions where (1) the symptoms have a low correlation with the disease and (2) the symptoms have a high correlation with the disease produces the following results. The LD model had the most gain with the B model ranking second. The binary regression models have the smallest gain. These results show that solutions resulting from the binary regression model and the Bayesian model which are intended to correct for highly interrelated symptoms did not improve significantly in reducing loss nor in increasing gain.

The summary of these resultant losses and gains are presented in the following matrix:

Table 4.14

Utility Table in Terms of Gains for Each Model and for Each Class

	Models													
Classes	B	R	BB	R	BLS	R	BWLS	R	BR	R	BRWLS	R	LD	R	EMPD	R
I	.38	2	.40	1	.19	8	.20	7	.38	2	.38	2	.38	2	.38	2
II	.46	1	.32	8	.35	5	.34	6	.41	2	.34	7	.46	1	.46	1
III	.32	8	.32	6	.36	5	.44	1	.40	3	.44	1	.40	3	.32	6
	.39	2	.35	5	.30	6	.33	5	.19	8	.39	2	.41	1	.39	2
IV	.76	1	.76	1	.73	4	.65	7	.66	6	.56	8	.76	1	.68	5
V	.78	2	.80	1	.44	7	.44	7	.56	5	.56	5	.78	2	.78	2
VI	.78	2	.40	8	.62	5	.68	4	.62	5	.61	7	.78	1	.76	2
	.77	1	.65	4	.63	5	.59	7	.61	6	.57	8	.77	1	.74	3
Pooled	.58	2	.50	4	.46	7	.46	7	.50	4	.48	6	.59	1	.56	3

	Least Loss	Best Gain
Low correlation with the disease	LD	LD
High correlation with the disease	LD, B	LD, B
Pooled	LD, B	LD

It should be borne in mind that the above evaluation is contigent on the choice of weights and the fact that the second sample has the same relational structure as the first. It is now of interest to see how the models would perform when the estimated parameters are applied to a second sample that has a different relational structure than the first, keeping the weights constant. This situation would represent the case when a sample of information is gained and the relational structure is derived based on that sample and assumed to hold for all subsequent samples. That is, the derived parameters from this initial sample are "blindly" generalized to a second sample which has an unknown relational structure. The results of using one class to generalize to another class in terms of the prediction indices are shown in Appendix E.

In Table 4.15, the rows represent the knowledge of the sample equivalent to the class and the columns represent the model that has the highest SEN, SPEC, and PRED across the classes. These classes have a different structure from

Table 4.15

The Best Model in Terms of Predictive Efficiency Indices Under Different Relational Structural Situation

Situation	Highest SEN	Highest SPEC	Highest PRED
A	BB	EMPD	LD, EMPD, B
B	BB	EMPD	LD
C	BLS	LD	B, LD
D	BLS	LD	LD, B

the sample relational structure from which they are developed. For simplicity, only four classes were chosen, namely, Class I, III, IV, and VI. These classes permit comparisons of the effects of low versus high intercorrelations among symptoms and low versus high correlations with the disease. The case of intermediate intercorrelations among the symptoms was ommited. Let the following notation represent these situations:

A. Prior knowledge of the relational structure of I and predicting across relational structure III, IV, and VI.

B. Prior knowledge of the relational structure of III and predicting across relational structures I, IV, and VI.

C. Prior knowledge of the relational structure of IV and predicting across the relational structures I, III, and VI.

D. Prior knowledge of the relational structure of VI and predicting across relational structures, I, III, and IV.

In terms of the predictive efficiency indices, the results are shown in Table 4.15.

The BB model has the highest SEN with the EMPD model having the highest SPEC and LD with the highest PRED for situations A and B. The BLS model has the highest SEN and the LD model has the highest SPEC and PRED along with the B model in situations C and D.

In terms of utility, the B, LD, and EMPD models have the least loss for situation A. The LD model has the least loss for situation B. The B, BB, and LD models have the least loss in situations C and B, and the LD model has the least loss in situation D. Across all four situations, the LD model has the least loss followed by the B and BB models. This is shown in Table 4.16.

The BB model has the most gain in situation A with the BLS model having the least gain. The LD model has the most gain in situation B and also in situation C along with the B model. The B and LD models also have the most gain in situation D. This is shown in Table 4.17.

Table 4.16

The Performance of the Probabilistic Models in Terms of Losses When Cross Validating to a Different Relational Structure

Prior Knowledge	B	R	BB	R	BLS	R	LD	R	EMPD	R
I	.28	1	.29	4	.36	5	.28	1	.28	1
III	.46	4	.42	2	.43	3	.35	1	.46	4
IV	.32	1	.32	1	.34	4	.32	1	.35	5
VI	.32	1	.33	3	.38	5	.32	1	.34	4
	.34	2	.34	2	.37	5	.32	1	.36	4

Table 4.17

The Performance of the Probabilistic Models in Terms of Gains When Cross Validating to a Different Relational Structure

Prior Knowledge	B	R	BB	R	BLS	R	LD	R	EMPD	R
I	.64	2	.71	1	.48	5	.64	2	.64	2
III	.28	4	.36	2	.34	3	.50	1	.28	4
IV	.55	1	.54	3	.52	4	.55	1	.50	5
VI	.56	1	.54	3	.43	5	.56	1	.52	4
	.51	3	.54	2	.44	5	.56	1	.48	4

These results can be summarized as follows:

	Least Loss	Highest Gain
A	LD, EMPD, & B	BB
B	LD	LD
C	B, BB, & LD	B & LD
D	LD, B	LD, B

Special Classes

Besides the above six classes generated, three "special" classes were generated to have the following properties.

1. Mixed Class: The mixture of the six classes, i.e., there are highly correlated symptoms and also lowly correlated symptoms and some are highly correlated or lowly correlated with the disease outcome.

2. Suppressor Class: The presence of a symptom which is highly correlated with other symptoms but has low correlation, near zero, with the disease outcome, i.e., if ith is the symptom, then, r_{ij} = high and r_{iD} = 0. This symptom is known as the suppressor symptom (Lubin, 1957; Conger and Jackson, 1972).

3. High Correlated Class: An extreme class of high correlation among symptoms and high correlation with the disease.

The population and sample variance and covariance matrices for these special classes are presented in Appendix F. Using the same test of equivalence, the following results were obtained (Table 4.18).

Table 4.18

Test of Equivalence for Special Classes

Class	L	L'	d.f.	p
Mixed	152.04	151.32	10	.005*
Suppressor	4.83	4.81	10	.999
High correlation	36.94	36.76	10	.025*

*Significant at the 0.05 level.

Despite the statistical lack of equivalence between the population and sample variance-covariance matrices for the mixed and high correlation classes, the two classes still represent the intended situations and hence, would not be a major concern for later analysis and interpretation.

The three samples were then randomly split into two equal halves and the two sub-samples were then tested for equivalence as shown in Table 4.19.

Table 4.19

Test of Equivalence for Sub-Samples for Special Classes

Class	2	d.f.	Significance Probability P
Mixed	1.56	10	.995
Suppressor	15.77	10	.10
High correlation	2.86	10	.99

The estimated parameters were derived from the first sample as before and they were cross-validated with the second sample. The performances in terms of the prediction indices are presented in Appendix G. Table 4.20 shows the summary results for the special classes.

Table 4.20

The Models That Perform Relatively the Best in Terms of Predictive Indices

Class	Highest SEN	Highest SPEC	Highest PRED
Mixed	All	B, BB, LD, EMPD	B, BB, LD, EMPD
Suppressor	All	All	All
High correlation	All	B, BB, LD	B, BB, LD

In terms of decision function values and using the same weights as given before, Table 4.21 and Table 4.22 show the values of loss and gains, respectively.

In all three special classes, all models surprisingly perform the same in terms of sensitivity! In terms of having the highest SPEC and PRED, in the mixed class, all models except the binary regression models, BLS, BWLS, and BR, perform the same. In the suppressor class, all models have the same performances in terms of SEN, SPEC, and PRED. In the high correlation class, all models have the same SEN, and B, BB, LD models have the highest SPEC and PRED.

In terms of loss, all models except the binary regression models have the same amount of loss in the mixed class. In the suppressor class, all models have the same amount of loss. In the high correlation class, the B, BB, and LD models have the least loss while the BR model has the most loss.

In terms of gain, the B, BB, LD, and EMPD models have the most gain in the mixed class. In the suppressor class there is no difference in gain for all models. In the high correlation class, the EMPD model has the most gain while the BWLS model has the least gain.

Table 4.21

Loss Function for Various Models
for Special Classes

	Models													
Class	B	R	BB	R	BLS	R	BWLS	R	BR	R	LD	R	EMPD	R
Mixed	.19	1	.19	1	.26	5	.33	7	.31	6	.19	1	.19	1
Suppressor	.40	1	.40	1	.40	1	.42	1	.42	1	.42	1	.42	1
High correlation	.30	1	.30	1	.32	5	.36	7	.34	6	.30	1	.31	4

Table 4.22

Gain Function for Various Models
for Special Classes

	Models													
Class	B	R	BB	R	BLS	R	BWLS	R	BR	R	LE	R	EMPD	R
Mixed	.82	1	.82	1	.68	5	.53	7	.57	6	.82	1	.82	1
Suppressor	.36	1	.36	1	.36	1	.36	1	.36	1	.36	1	.36	1
High correlation	.60	2	.60	2	.55	5	.48	7	.52	6	.60	2	.88	1

Clinical Application

The data selected for application are from a study on brain scans by Potchen (1975) from July 1974 to June 1975 at Johns Hopkins Hospital, Maryland. The procedure involved in his study is shown in Figure 4.2. The instrument that was used is presented in Appendix H. The patients with the given symptoms were recorded on a questionnaire and these were given to physicians to determine the probability of having an abnormal scan for each patient and whether a brain scan was necessary. The final results were confirmed by the brain scan when the patient was referred for such action. In this application, only those patients that were referred for brain scan were used. There are altogether 86 patients in which 8 patients had abnormal brain scans which means tumor growth in the brain. Since the application is about symptomic diagnosis, only the signs and symptoms were selected. They are:

1. headaches;
2. seizure;
3. cortical deficit;
4. motor deficit;
5. sensory abnormality; and
6. visual field defect.

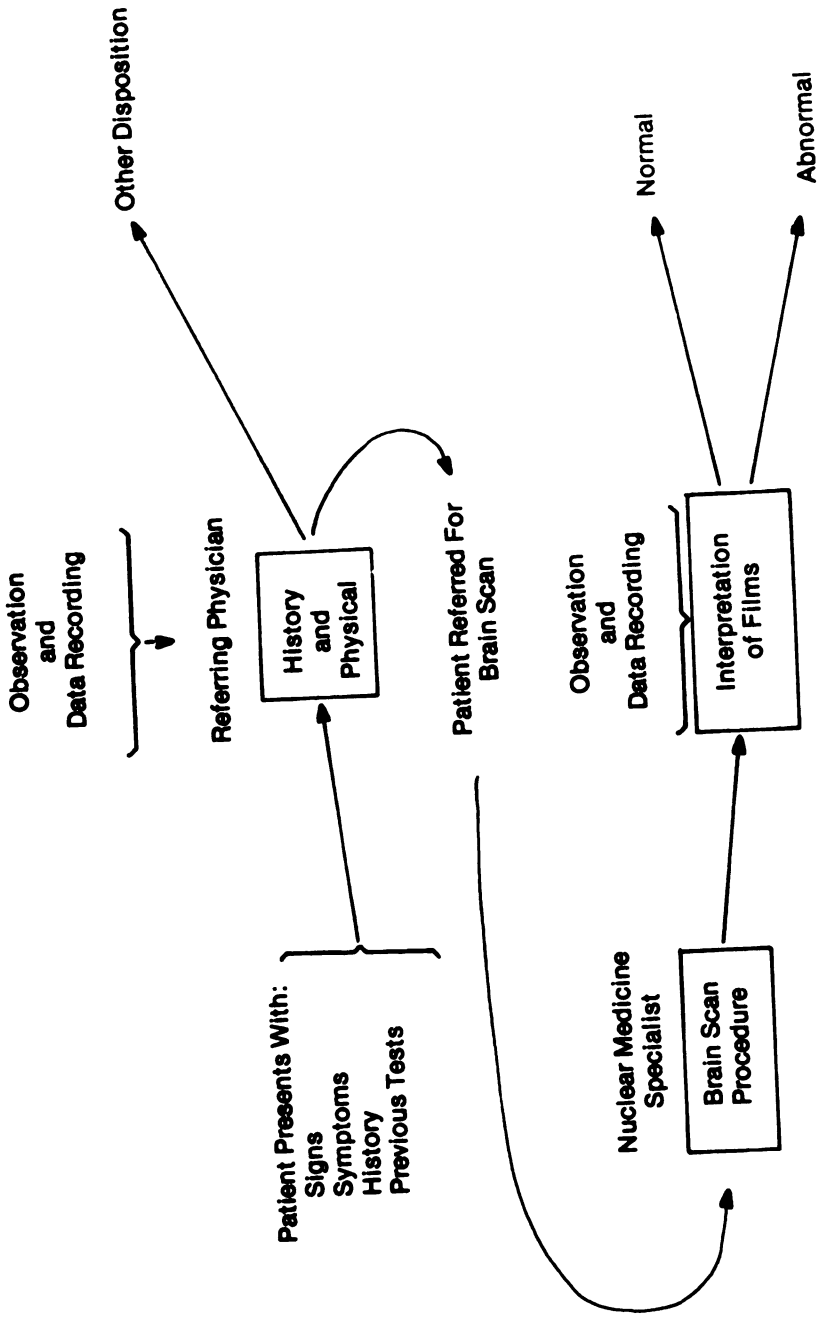


Figure 4.2 Flow Diagram Indicating the Events Involved in Completing a Single Brain Scan Questionnaire.

The normality or the abnormality of the final brain scan will be considered as the disease outcome, the 86 patients will be considered as the "population" of brain tumor suspected patients with the given six symptoms--the set of conditions. It should be borne in mind that with such few patients, the following results can only be considered a pilot study or preliminary investigation for the models.

With the same procedure, the 86 patients were split into two equal halves of 43. Each half having 4 abnormally scanned patients. The variance-covariance matrix for the "population" and the samples are shown in Appendix J. The test of equivalence for the split samples and test for multicollinearity are presented in Table 4.23.

Table 4.23

Test for Equivalence and Multicollinearity

Test	2	d.f.	Significance Probability P
Equivalence	.001	28	.99
Multicollinearity	20.65	15	.25

The estimated parameters are shown in Appendix F. The decision point π is set at .10 (π = 8/86 = .10). The results of computing the prediction indices for the models are shown in Table 4.24. From the table, the weighted ridge solution surprisingly improves the sensitivity of the ordinary least squares binary model. The entropy model has the highest specificity with the binary model having the least specificity. However, the binary model has the highest predictive value.

Table 4.24

Prediction Indices for Various Models for Brain Scan

	Models								
Indices	B	BB	BLS	BWLS	BR	BRWLS	LD	EMPD	PSP[a]
SEN	.00	.50	.50	.25	.75	.75	.25	.00	.75
SPEC	.90	.95	.54	.65	.69	.41	.74	.97	.41
PRED	.00	.00	1.00	.06	.20	.12	.09	.00	.09

[a]PSP = physician subjective probability derived from category IV, section 1a, on the questionnaire as shown in Appendix H.

With respect to the values of the decision function with the weights as given on page 82, Table 4.25 shows the values for both the loss and gain for various models for the brain scan data.

Table 4.25

Decision Function Values for Various Models on Brain Scan

	Models								
Function	B	BB	BLS	BWLS	BR	BRWLS	LD	EMPD	PSP[a]
Loss	.29	.14	.51	.46	.33	.58	.38	.03	.25
Gain	.52	.81	.08	.17	.44	.07	.34	.84	.63

[a]PSP = physician subjective probability derived from category IV, section 1a, on the questionnaire as shown in Appendix H.

Hence, from the results, the EMPD model is the best model in terms of utility; the least loss and the most gain, in screening or predicting brain scan patients.

What are the significant findings from the analyses in this chapter? What can these models have to offer for diagnostic problem-solving? How do these models relate to a real clinical setting? And how can one go about using these probabilistic models for diagnostic problem solving? These issues and other important issues will be discussed in the following chapter.

CHAPTER V

SUMMARY AND DISCUSSION

The significant findings in this thesis may be summarized as follows:

1. Overall, sensitivity increases for all models as the correlation with the disease outcome increases.

2. There is a "hump" or convex effect for sensitivity for all models except the Bayesian (B), Bayesian with the Bahadur's expansion (BB) and the Entropy Minimax Pattern Discovery (EMPD) models, in situations where the symptoms have a low correlation with the disease outcome. That is, the maximum sensitivity is not when the intercorrelation between the symptoms is greatest but when the symptoms are moderately intercorrelated as shown in Appendix L. This phenomenon did not show in situations where the symptoms have a high correlation with the occurrence of the disease. In fact, sensitivity increases as the intercorrelations increase under the latter situation as shown in Appendix M.

3. The values for sensitivity did not differ among models in situations where highly interrelated symptoms are also highly related to the occurrence of the disease. In other words, when the relational structure is highly

correlated, it does not matter which model one uses if sensitivity is chosen as a criterion for selection models.

4. The "pit" or concave effect of specificity across binary regression models occurs when, given those situations where the symptoms are highly correlated with the disease outcome, the intercorrelations between the symptoms increase. This is also shown in Appendix M. This means that specificity is at a minimum when the symptoms are moderately related.

5. The "hump" or convex effect is also found for predictive values in the same way as the sensitivity index, that is, when the symptoms have a low correlation with the occurrence of the disease as shown in Appendix L.

6. With the presence of a suppressor symptom, it does not matter what measure one uses as a criterion for selecting models as all models perform the same for all prediction efficiency indices.

7. If a model is chosen with the criterion of having the best sensitivity, it is at a cost of losing specificity and vice versa. In other words, there are no models that have the best of both indices for all classes considered in this dissertation. The statement holds when one looks across classes and within classes of problems. This also means that there is no single model that performs consistently better for each class or across classes in terms of sensitivity and specificity.

8. A decision function analysis was performed. Penalty (negative) weights were given for the two diagnostic errors (i.e., Type 1 and Type 2) and no credit is given to the correct diagnosis. The binary least square model (BLS) and the binary weighted least square model (BWLS) showed the smallest loss when the symptoms had a low correlation with the disease's occurrence but themselves had high intercorrelations. However, when considering gains, with credits given to the correct diagnosis, but the same penalty weights, the Bayesian model (B) had the most gain when the intercorrelations among the symptoms was low but the correlation between the symptoms and outcome was high. The logistic discrimination model (LD) had the most gain when the symptoms had a low correlation with each other but had a high correlation with the occurrence of the disease outcome. The LD model also had the most gain when the symptoms were moderately interrelated with each other and the symptoms had a low correlation with the disease. If one disregards the intercorrelation among symptoms, the LD model had the highest gain whether or not the symptoms had a high or low correlation with the occurrence of the disease. That is, the best model to use to maximize gain in the absence of knowledge about the relationship among and between symptoms and disease outcomes, is the LD model.

9. Summarizing the above results, the following table (Table 5.1) can be formulated. The columns denote the kind of relational structure cross-validated and the rows represent the criterion for selecting models. For example, assume one wants to cross-validate under the assumption of an unknown relational structure and also chooses specificity (SPEC) as the criterion. One would go to the intersection of column two (Unknown) and the second row (SPEC) and conclude that the B model should be used.

Table 5.1

Decision Table in Choosing Models With Respect to Prediction Index and Kind of Cross-Validated Relational Structure

Criterion for Selecting Models	Same	Unknown
SEN	BLS	BLS
SPEC	BR	B
PRED	BB, LD, EMPD	LD
LOSS	LD, B	LD
GAIN	EMPD, LD	LD

Further Recommendations

The purpose of this thesis was to demonstrate how different statistical models perform when applied to different relational structures and under differing degrees of uncertainty. The following are some recommendations for further research:

1. Vary the base rates of the disease and the symptoms and determine the changes of the prediction efficiency indices for various models.
2. Increase the number of disease categories beyond the two that were considered in this dissertation (i.e., D_1, D_2, ... D_d).
3. Change the direction of the intercorrelation among the symptoms and with the occurrence of the disease to negative and determine the changes in prediction efficiency indices for various models.
4. Vary the decision rules and determine the changes in prediction efficiency indices for various models.
5. For symptoms that have high intercorrelations, combine symptoms to form "factors" by means of factor analysis and principal components techniques and use these generated factors or components to predict the occurrence of the disease.

Clinical Implications

What sort of implications do these models and the findings have for an empirical clinical setting? First of all, these models are attempts to quantify uncertainty. They are a set of mathematical algorithms to generate indices from a complicated universe in order to enable decision-making to be less difficult and to be more effective. They are not meant to replace the human decision maker but rather to supplement the decision process. They act as an additional source of information for the decision maker. The model's relationship with the human decision maker may be illustrated as in Figure 5.1.

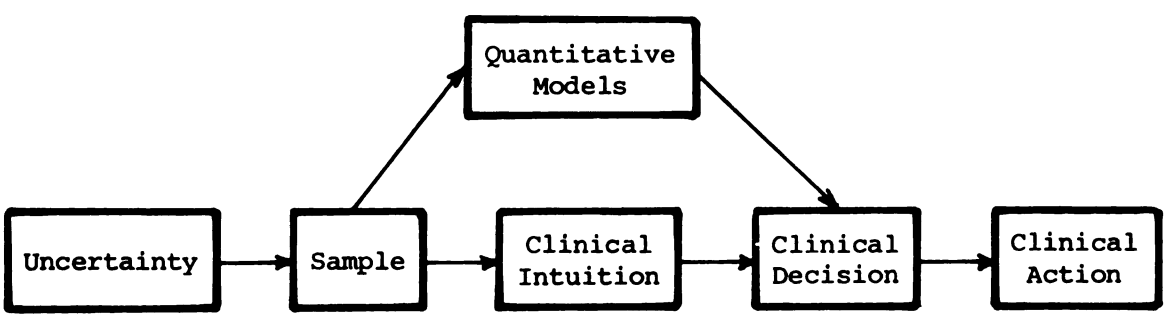


Figure 5.1 The Relationship Between Quantitative Model Decision and Clinical Intuition.

After one obtains the additional information from the quantitative models, one can choose the following three alternatives: (1) ignore the prediction made by the quantitative models and follow clinical information, (2) modify the clinical impression on the basis of the information provided by the quantitative models, or (3) abandon clinical intuition in favor of the quantitative choice. It should be borne in mind that the final and full responsibility of medical diagnosis lies on the physician and not on a set of mathematical algorithms, regardless of which of the three alternatives is chosen.

When the physician's clinical intuition is in agreement with the quantitative prediction, there is no problem and the quantitative prediction is seen as "reinforcing" clinical intuition. However, when clinical intuition is in disagreement with the quantitative prediction, the physician should weigh all the evidence by objectively examining the validity of his own intuition and the validity of the assumptions of the quantitative models to generate the prediction. If the model's assumptions are violated, then he should take alternative (1) (i.e., abandon the quantitative prediction and follow his own intuition). However, if the physician feels that for some reason his clinical intuition is somehow suspect, then it is recommended that he take alternative (3) (i.e., abandon his own

intuition and follow the model's prediction). Again the physician must bear the responsibility of abandoning his own intuition and abiding by the quantitative prediction. For all clinical decisions, if the crux is to determine whether the patient has a disease, and there is doubt, despite all possible evidence gathered by both the human decision maker and the models, it is better to diagnose the patient as having the disease. This follows the axiom, "If in doubt, diagnose illness." The above situations can be illustrated in the following table (Table 5.2).

Table 5.2

Final Decision by Clinical Intuition and Quantitative Prediction

		D	$\bar{D}$
Clinical Intuition	D	D	D
	$\bar{D}$	D	$\bar{D}$

Schema for Application of the Models to a Diagnostic Problem

To apply these probabilistic models to a diagnostic problem, the following steps should be taken:

1. Select the disease to be identified.
2. Identify the set of signs or symptoms which are thought to occur jointly with the disease. That is, in effect, similar to identifying the signs or

symptoms which are related to the occurrence of the disease without implying causality between the symptoms and the occurrence of the disease.

3. Collect all available cases of the set of signs or symptoms. It is to be noted that the frame-of-reference for the data collection is with respect to the set of signs or symptoms and not with respect to the occurrence of the selected disease. Hence, the collected data will include those cases that the selected disease and those cases that have other diseases or no diseases of interest.
4. From the collected data and for each individual case, code a one (1) if the case shows the presence of the selected disease and code a zero (0) if the case shows other diseases or no disease. Likewise, use the same scheme of coding with the signs or symptoms for each individual case. The resultant coded data will resemble the data matrix shown in Figure 1.3.
5. Define an uncertainty structure by dividing the magnitudes of the intercorrelations among the signs or symptoms into levels and likewise with the magnitudes of the correlation of the signs or symptoms with the occurrence of the disease. Then label,

numberically or alphabetically, the cells or classes in the uncertainty structure. The above two procedures will result in the following figure:

		Intercorrelation Among Symptoms		
		Level 1	Level 2	Level 3
Correlation of Symptoms With Disease	Level 1	I	II	III
	Level 2	IV	V	VI
	Level 3	VII	VIII	IX

It should be noted that the levels need not be of equal intervals.

6. From the new coded data matrix, compute all possible pairwise correlations among the symptoms and the correlations between the symptoms and the occurrence of the disease by using the phi-coefficient formula (Cohen and Cohen, 1976), obtaining the correlational matrix as shown in Figure 3.1. The computed correlational matrix constitutes the relational structure of the disease and the symptoms.
7. Identify the cell or class where the computed relational structure is the closest in value with the results in step 5. This is done by either

a. "eyeballing" the values of the computed relational structure along with the values of the correlations in each individual classes in the uncertainty structure and selecting the class that bears the most resemblance (a purely subjective judgment) to the computed relational structure, or

b. performing a statistical test of equivalence between the classes and the computed relational structure. This is, in effect, testing the following hypothesis:

$$H_0: \quad R_s = R_{ci}$$

against the alternative,

$$H_1: \quad R_s \neq R_{ci}$$

where R_{ci} = the ith class in the uncertainty structure. It is worthy to note that the values in the correlation matrix, R_{ci}, are the median values of the two intervals of the ith class (i.e., if the level of correlation among symptoms is 0.5 to 0.7 and the level of correlation of symptoms and the disease is 0.0 to 0.30 for the ith class, the median values for the correlational matrix, R_{ci}, are 0.6 and 0.15, respectively).

R_s = the computed relational structure from the collected data.

Such test of equicorrelation patterns of relational structures can be found in Morrison (1976, p. 276).

8. Select a criterion, sensitivity, specificity, or predictive value, according to the following rule:

Situation	Criterion Selected
The consequences of committing a Type 1 error is more serious than committing a Type 2 error	Sensitivity
The consequences of committing a Type 2 error is more serious than committing a Type 1 error	Specificity
The consequences of committing both Type 1 and Type 2 errors are of no difference	Sensitivity or Specificity

9. Use Table 4.8 where the rows represent the criterion to be selected and the columns represent the classes of the uncertainty structure. The intersection of the rows and the columns represents the probability model or models that perform relatively the best with respect to the selected criterion in that particular class.

10. Use that model for diagnostic prediction for the selected disease in maximizing the chosen criterion.

The summary of the above ten steps is represented in Figure 5.2.

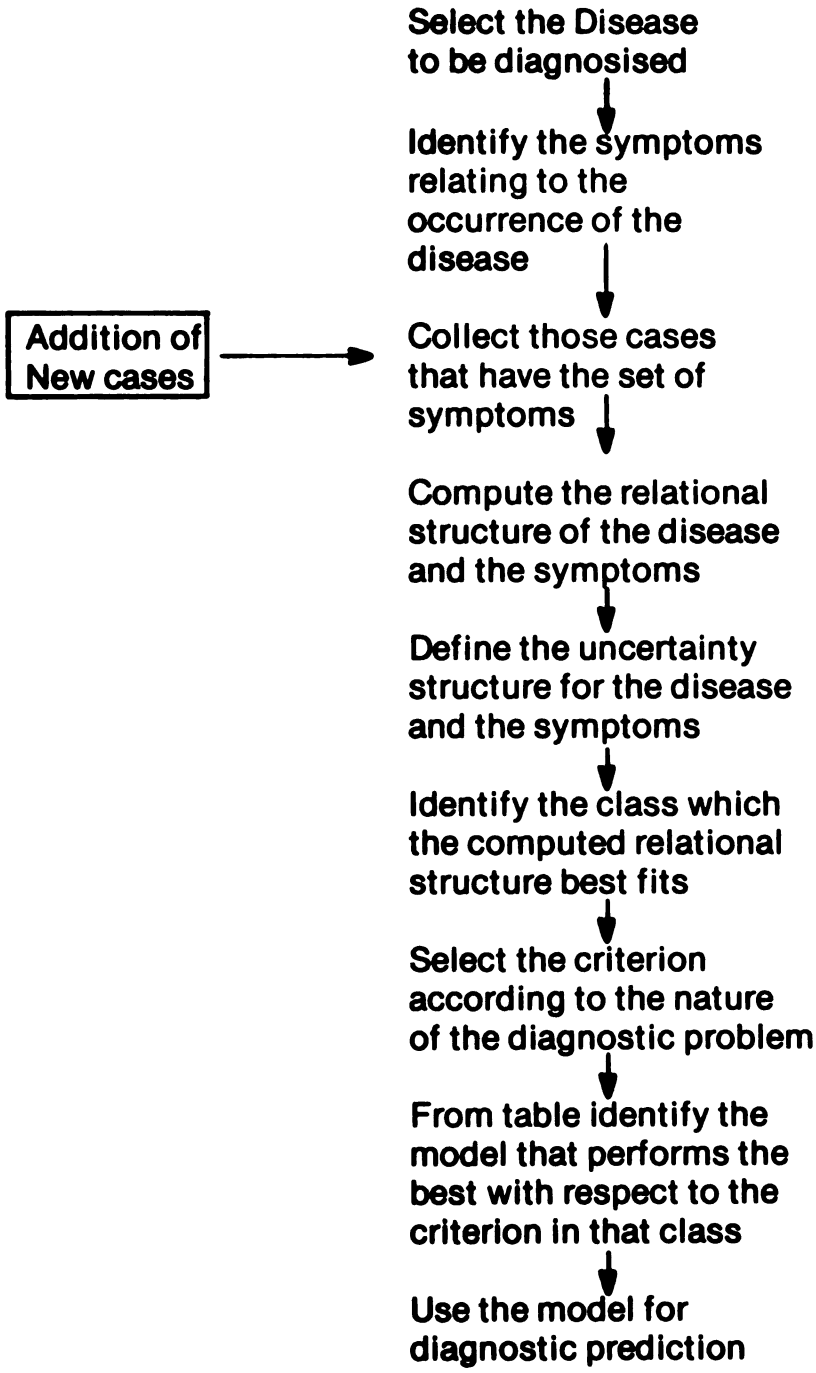


Figure 5.2 Schema for Application of the Models to a Diagnostic Problem.

An Example in Breast Cancer

Consider the problem of detecting breast cancer which is one of the major causes of death among women. It is widely recognized that the early detection of this cancer will reduce its mortality rate. However, the term "early" has equivocal meanings as it denotes the absence of any signs or symptoms at the onset stage of the cancerous growth. The only "signs" or "symptoms" for such an early detection are sociological cues: the patient's familial history of breast cancer, the patient's pregnancy and menarche history, and other cues which are not directly related to the cancer. These cues are known as risk factors and they constitute the physician's index-of-suspicion. The diagnostic problem is then to use these risk factors to identify the high risk group of patients as having breast cancer. The risk factors that are known to be highly related to the occurrence of breast cancer are (1) age, (2) socioeconomic status, (3) age at menarche, (4) age at pregnancy, (5) age at menopause, (6) familial history of breast cancer, and (7) number of pregnancies. Gather all available cases that have these risk-factors. An excellent data source would be from mass screening centers. A portion of the collected cases will be confirmed breast cancer cases (coded as ones) and the other portion of cases will be non-confirmed breast cancer

cases (coded as zeroes). Each risk factor is dichotomized by setting a subjective cut-off point and coding a value of one if the value of the risk-factor exceeds the cut-off point and coding a value of zero if the value of the risk-factor lies below the cut-off point. The next step is then to define an uncertainty structure. The following matrix is one possible definition of the uncertainty structure:

		Intercorrelations Among the Risk Factors		
		< 0.20	0.21-0.50	> 0.50
Correlations of the Risk Factors with Breast Cancer	> 0.20	I	II	III
	0.21-0.50	IV	V	VI
	< 0.50	VII	VIII	IX

Then compute all possible pairwise correlations among the risk-factors and the correlations between the risk-factors and the occurrence of breast cancer, thereby deriving the relational structure of the risk factors and breast cancer. Using the relational structure and the uncertainty structure, identify the class of the relational structure by either strategy as mentioned in Step 7. Misclassifying a breast cancer case as non-breast cancer case (Type 1) has more serious consequences than classifying a non-breast cancer case as a cancer case (Type 2 error), since the former

action means later detection and delayed therapy which might lead to death; consequently, sensitivity is preferable to specificity as a criterion for selecting models. Finally, from Table 4.8, with sensitivity as the criterion and the class that has been identified for the relational structure, say hypothetically Class IV, the binary ridge regression model is the best model relative to the other models, for identifying the high risk breast cancer group.

It should be borne in mind that Table 4.8 is generated from the assumption that the base rate for the selected disease is 0.2 and the base rates for the symptoms are 0.5. However as further studies which use the same methodology as this dissertation, investigate the effects of varying the base rates for the disease and the symptoms, this assumption can be relaxed.

Quantitative Models in Medical Decision Making

The use of quantitative models can achieve three main objectives which are merits of the models in their own right. They:

1. Combine probabilistic reasoning and uncertainty of the data in a formal explicit system rather than by intuition to achieve more efficient and consistent information processing.

2. Provide a systematic processing of uncertainty that takes account of all available information for decision making and find the optimal weighting combination of symptoms, ensuring that each contributes properly to the disease outcome.
3. Develop formulae, rules, or strategies for optimal consistent information processing in the presence of uncertainty.

There are three areas in which quantitative models can assist in better medical decision making. They are (1) teaching tools, (2) patient management, and (3) public policy. These areas might be considered as follows:

1. Teaching Tool: Elstein (1976) has noted that strategies for different degrees of uncertainty have been made explicit by quantitative models. Hence, they can become a learning device for the novice in finding strategies and rules for identification of a disease. Consider the detection of breast cancer. The problem is to find the "high risk" group without referring every case for radiological examination. Radiological examinations have turned out to be hazardous to health. Blair (1976) has found that radiation has killed as many patients as breast cancer itself. Yet, radiological examinations or techniques remain the best device for detecting breast cancer despite their potential hazards. Hence, the crux of the problem is

to (1) assume the patient has breast cancer and to refer the patient for radiological examination knowing that exposure to radiation is hazardous, a possible Type 2 error, or (2) assume the patient does not have breast cancer with the danger of committing a Type 1 error. When the explicit rules and strategies for making this crucial decision have been generated by quantitative models, the novice could learn from these rules to make his decision.

2. Patient Management: In situations of diagnostic ambiguity, the physician has difficulty in taking clinical action. But when rules and strategies for diagnosing the disease have been made explicit, this information becomes a frame-of-reference for diagnosis hence removing the ambiguity of the situation. An excellent example for patient management is the common symptom, headache. MacBryde and Blacklow (1970) have listed fifteen diseases associated with the symptom, headache, among which is brain tumor. Each disease demands unique treatment and therapy. The kinds of treatment range from administering an aspirin to brain surgery. Each treatment procedure demands cost, time, and potential hazard. The problem is to identify the disease correctly in order to give the correct form of patient management.

3. Public Policy Making: In the area of health care, there are many decisions involving the expenditure of large dollar amounts for public health programs. As one instance, a debate is current between the Department of Public Health and the third party carriers as to whether physicians be for "whole body" computed assisted tomography (CT) scans. This is only symptomatic of the impact of technology on medical diagnosis. The question becomes, how should one and when should one use these expensive and sometimes potentially hazardous diagnostic techniques. Further, who will pay for it; how much will be paid; and how often will these procedures be paid for, become a series of questions that are entering into health policy. It is anticipated that the application of the quantitative models studied in this dissertation will help provide answers to questions such as these. For example, if one can determine the efficacy and correctness of a clinical diagnosis through the use of quantitative models, then the procedures used to reach the diagnosis would be strengthened and consequently, be candidates for reimbursement. If, however, the weight assigned to particular procedures is low, which would indicate little or no contribution to the overall clinical diagnosis, then the procedures needed to obtain the information as to whether the symptom is present or absent should be scrutinized for reimbursement.

APPENDICES

APPENDIX A

THE RELATIONAL STRUCTURE OF THE POPULATION, THE GENERATED SAMPLE AND THE SPLIT SAMPLES

NOTE

The correlational matrices (R) and the variance-covariance (Σ) matrices in this appendix and the following appendices should be interpreted as follows:

$$
\begin{array}{c} \\ S_1 \\ S_2 \\ S_3 \\ D \end{array}
\begin{array}{c}
\begin{array}{cccc} S_1 & S_2 & S_3 & D \end{array} \\
\left[\begin{array}{cccc}
a_{11} & & & \\
a_{12} & a_{22} & \text{(Symmetric)} & \\
a_{13} & a_{23} & a_{33} & \\
a_{1d} & a_{3d} & a_{dd} &
\end{array}\right]
\end{array}
$$

where

$a_{11}, a_{22}, a_{33}, a_{dd}$ = 1 if it is a correlational matrix and the variances of symptom 1, 2, and 3, and the disease, respectively, if it is a variance-covariance matrix;

a_{12} = correlation between symptom 1 and symptom 2 if it is a correlational matrix, and the covariance of symptom 1 and symptom 2 if it is a variance-covariance matrix;

a_{13} = correlation between symptom 1 and symptom 3 if it is a correlational matrix, and the covariance of symptom 1 and symptom 3 if it is a variance-covariance matrix;

a_{23} = correlation between symptom 2 and symptom 3 if it is a correlational matrix, and the covariance of symptom 1 and symptom 2 if it is a variance-covariance matrix; and

a_{1d}, a_{2d}, a_{3d} = the correlation of the occurrence of disease with symptom 1, 2, and 3, respectively, for the correlational matrix and the covariance of the disease with symptom 1, 2, and 3, respectively, for the variance-covariance matrix.

Relational Matrices of Populations and Samples

	Class I	Class II	Class III
R_p	1.00 0.20 1.00 0.20 0.20 1.00 0.20 0.20 0.20 1.00	1.00 0.40 1.00 0.40 0.40 1.00 0.20 0.20 0.20 1.00	1.00 0.60 1.00 0.60 0.60 1.00 0.20 0.20 0.20 1.00
$\hat{R}_s$	1.00 0.10 1.00 0.13 0.14 1.00 0.21 0.10 0.11 1.00	1.00 0.40 1.00 0.35 0.33 1.00 0.24 0.24 0.31 1.00	1.00 0.56 1.00 0.67 0.62 1.00 0.23 0.18 0.23 1.00

	Class IV	Class V	Class VI
R_p	1.00 0.20 1.00 0.20 0.20 1.00 0.40 0.40 0.40 1.00	1.00 0.40 1.00 0.40 0.40 1.00 0.40 0.40 0.40 1.00	1.00 0.60 1.00 0.60 0.60 1.00 0.40 0.40 0.40 1.00
$\hat{R}_s$	1.00 0.15 1.00 0.23 0.21 1.00 0.41 0.34 0.32 1.00	1.00 0.31 1.00 0.44 0.42 1.00 0.45 0.41 0.37 1.00	1.00 0.56 1.00 0.67 0.61 1.00 0.43 0.44 0.42 1.00

Variance-Covariance Matrices of Population and Samples

	Class		
	I	II	III
Σ_p	.250 .050 .250 .040 .050 .250 .040 .040 .040 .160	.250 .100 .250 .100 .100 .250 .040 .040 .040 .160	.250 .150 .250 .150 .150 .250 .040 .040 .040 .160
$\hat{\Sigma}_s$	.250 .020 .250 .030 .030 .250 .040 .020 .020 .150	.250 .100 .250 .080 .080 .250 .040 .040 .060 .160	.250 .140 .250 .168 .155 .250 .045 .037 .046 .154
	IV	V	VI
Σ_p	.250 .050 .250 .050 .050 .250 .080 .080 .080 .160	.250 .100 .250 .100 .100 .250 .080 .080 .080 .160	.250 .150 .250 .150 .150 .250 .080 .080 .080 .60
$\hat{\Sigma}_s$	.250 .040 .250 .060 .050 .250 .080 .060 .060 .150	.250 .080 .250 .110 .100 .250 .090 .080 .070 .160	.250 .140 .250 .170 .150 .250 .080 .080 .080 .150

Correlational Matrices of Sub-Samples

	Class		
	I	II	III
$\hat{R}_1$	1.00 0.16 1.00 0.15 0.09 1.00 0.23 0.13 0.11 1.00	1.00 0.43 1.00 0.30 0.35 1.00 0.29 0.22 0.29 1.00	1.00 0.52 1.00 0.63 0.60 1.00 0.16 0.10 0.18 1.00
$\hat{R}_2$	1.00 0.03 1.00 0.11 0.17 1.00 0.18 0.08 0.10 1.00	1.00 0.38 1.00 0.41 0.32 1.00 0.20 0.26 0.33 1.00	1.00 0.60 1.00 0.72 0.64 1.00 0.29 0.27 0.28 1.00
	IV	V	VI
$\hat{R}_1$	1.00 0.21 1.00 0.36 0.26 1.00 0.36 0.43 0.34 1.00	1.00 0.41 1.00 0.49 0.36 1.00 0.49 0.44 0.32 1.00	1.00 0.59 1.00 0.68 0.60 1.00 0.43 0.43 0.40 1.00
$\hat{R}_2$	1.00 0.10 1.00 0.12 0.16 1.00 0.46 0.26 0.30 1.00	1.00 0.22 1.00 0.39 0.49 1.00 0.42 0.38 0.42 1.00	1.00 0.55 1.00 0.67 0.65 1.00 0.48 0.46 0.45 1.00

Variance-Covariance Matrices of Sub-Samples

	Class		
	I	II	III
$\hat{\Sigma}_1$	.250 .040 .250 .030 .040 .250 .040 .020 .020 .150	.240 .100 .240 .070 .080 .250 .050 .040 .050 .160	.240 .120 .240 .150 .150 .250 .030 .020 .030 .150
$\hat{\Sigma}_2$	.250 .010 .250 .020 .040 .250 .030 .010 .020 .150	.250 .090 .250 .090 .070 .240 .040 .050 .060 .170	.250 .150 .250 .180 .160 .250 .050 .050 .050 .150
	IV	V	VI
$\hat{\Sigma}_1$	.250 .050 .250 .090 .060 .250 .060 .080 .060 .150	.250 .100 .250 .120 .090 .250 .090 .080 .060 .150	.240 .140 .250 .160 .150 .250 .080 .080 .070 .140
$\hat{\Sigma}_2$	.250 .020 .240 .030 .040 .250 .080 .050 .050 .150	.240 .050 .250 .090 .120 .250 .080 .070 .080 .150	.250 .140 .250 .160 .160 .240 .090 .080 .080 .140

APPENDIX B

THE ESTIMATED PARAMETERS FOR EACH PROBABILITY MODEL FOR EACH CLASS

Estimated Parameters Binary Regression (LS)

	Class					
	I	II	III	IV	V	VI
Constant	.03	.03	.10	-.14	-.07	-.04
Symptom 1	.17	.14	.07	.27	.28	.17
Symptom 2	.07	.09	-.02	.19	.22	.18
Symptom 3	.06	.14	.11	.12	.03	.09

Estimated Parameters Binary Regression (Weighted LS)

	Class					
	I	II	III	IV	V	VI
Constant	.07	.02	.11	-.27	.04	.02
Symptom 1	.16	.09	.05	.23	.28	.14
Symptom 2	.08	.14	.02	.24	.14	.17
Symptom 3	.01	.02	.09	.34	-.05	.03

Estimated Parameters Binary Regression (Ridge)

	Class					
	I	II	III	IV	V	VI
Constant						
Symptom 1	.05	.12	.05	.19	.06	.12
Symptom 2	.11	.06	.01	.14	.16	.13
Symptom 3	.05	.12	.06	.10	.19	.09

Estimated Parameters Binary Regression (Weighted Ridge)

	Class					
	I	II	III	IV	V	VI
Constant						
Symptom 1	.16	.15	.07	-.07	.18	.14
Symptom 2	.11	.10	.18	.13	.19	.16
Symptom 3	.06	.15	.06	.15	.04	.07

Estimated Parameters Logistic Discrimination (LD)

	Class					
	I	II	III	IV	V	VI
Constant	-2.75	-3.07	-2.10	-6.78	-6.69	-9.25
Symptom 1	1.20	1.12	.52	3.48	3.27	5.87
Symptom 2	.53	.40	-.14	2.15	3.04	1.69
Symptom 3	.44	1.31	.75	1.46	.66	1.46

APPENDIX C

THE ESTIMATED DIAGNOSTIC PROBABILITIES FOR EACH 2^P PATTERN FOR EACH PROBABILITY MODEL FOR EACH CLASS

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3
Bayesian (B)

	Class					
Pattern	I	II	III	IV	V	VI
111	.35	.42	.27	.59	.60	.43
110	.21	.13	.10	.22	.35	.18
100	.08	.14	.09	.00	.00	.00
001	.00	.07	.26	.00	.00	.00
011	.14	.24	.00	.00	.00	.00
101	.38	.63	.35	.18	.06	.13
010	.18	.15	.25	.05	.06	.00
000	.08	.02	.10	.00	.00	.00

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3
Binary Regression (BLS)

	Class					
Pattern	I	II	III	IV	V	VI
111	.33	.39	.29	.45	.47	.39
110	.27	.25	.18	.32	.43	.31
100	.20	.16	.18	.13	.22	.12
001	.09	.17	.22	.00	.00	.04
011	.17	.26	.22	.17	.18	.23
101	.26	.30	.29	.25	.25	.21
010	.11	.11	.11	.05	.15	.14
000	.03	.02	.11	.00	.00	.00

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3 Binary Regression (BWLS)

	Class					
Pattern	I	II	III	IV	V	VI
111	.31	.39	.27	.54	.41	.34
110	.30	.25	.18	.19	.46	.33
100	.22	.16	.16	.00	.32	.16
001	.07	.17	.19	.06	.00	.05
011	.15	.26	.22	.30	.13	.22
101	.23	.30	.25	.29	.27	.19
010	.14	.11	.13	.00	.18	.19
000	.07	.02	.11	.00	.04	.02

Estaimted Diagnostic Probabilities for 2^p Possible Pattern for p = 3 Binary Regression-Ridge (BR)

	Class					
Pattern	I	II	III	IV	V	VI
111	.22	.30	.12	.43	.42	.35
110	.16	.18	.06	.32	.23	.26
100	.05	.12	.05	.18	.06	.12
001	.05	.12	.06	.10	.19	.09
011	.17	.17	.07	.24	.36	.23
101	.10	.24	.11	.29	.26	.22
010	.11	.06	.01	.14	.16	.13
000	.00	.00	.00	.00	.00	.00

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3 Binary Regression (BRWLS)

	Class					
Pattern	I	II	III	IV	V	VI
111	.34	.40	.31	.21	.42	.37
110	.27	.25	.25	.06	.37	.30
100	.16	.14	.07	.00	.18	.14
001	.06	.15	.06	.15	.04	.07
011	.18	.25	.24	.28	.24	.23
101	.22	.29	.13	.08	.23	.19
010	.11	.10	.18	.13	.19	.16
000	.00	.00	.00	.00	.00	.00

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3 Logistic Discrimination (LD)

	Class					
Pattern	I	II	III	IV	V	VI
111	.36	.44	.28	.65	.57	.44
110	.27	.17	.15	.29	.40	.15
100	.17	.12	.17	.05	.03	.03
001	.09	.14	.21	.00	.00	.00
011	.14	.20	.18	.05	.05	.00
101	.25	.34	.33	.18	.06	.13
010	.09	.06	.10	.01	.02	.00
000	.06	.04	.11	.00	.00	.00

Estimated Diagnostic Probabilities for 2^p Possible Pattern for p = 3
Entropy Minimax Pattern Discovery (EMPD)

	Class					
Pattern	I	II	III	IV	V	VI
111	.36 (.17)	.41 (.25)	.28 (.35)	.58 (.23)	.57 (.26)	.44 (.36)
110	.24 (.09)	.15 (.09)	.17 (.03)	.23 (.07)	.38 (.07)	.25 (.02)
100	.09 (.11)	.15 (.09)	.11 (.19)	.01 (.02)	.07 (.04)	.06 (.01)
001	.03 (.02)	.10 (.04)	.37 (.08)	.01 (.02)	.01 (.02)	.00 (.02)
011	.14 (.06)	.25 (.09)	.06 (.01)	.58 (.02)	.04 (.01)	.44 (.02)
101	.38 (.10)	.57 (.04)	.37 (.08)	.18 (.08)	.07 (.04)	.16 (.06)
010	.19 (.08)	.18 (.05)	.28 (.35)	.08 (.04)	.10 (.04)	.00 (.02)
000	.09 (.11)	.04 (.05)	.11 (.19)	.01 (.02)	.01 (.02)	.00 (.02)

Note: The entries in the parentheses are the entropy values (H) of the particular pattern within the particular class.

APPENDIX D

THE PREDICTIVE INDICES OF EACH MODEL FOR EACH CLASS

Predictive Indices
Bayesian

	Class					
Indices	I	II	III	IV	V	VI
SEN	.65 ± .17	.81 ± .14	.82 ± .13	.78 ± .14	.90 ± .10	.92 ± .09
SPEC	.66 ± .08	.63 ± .08	.54 ± .08	.84 ± .06	.78 ± .07	.77 ± .07
PRED	.31 ± .11	.37 ± .11	.29 ± .10	.52 ± .15	.50 ± .13	.47 ± .13
E1	.35 ± .17	.19 ± .14	.18 ± .13	.21 ± .14	.10 ± .10	.07 ± .09
E2	.34 ± .08	.37 ± .08	.46 ± .08	.16 ± .06	.22 ± .07	.22 ± .07

Note: Results are reported as estimate ± standard error.

Predictive Indices
Bayesian (BB)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.65 ± .17	.81 ± .14	.82 ± .13	.78 ± .14	.90 ± .10	.96 ± .07
SPEC	.66 ± .08	.63 ± .18	.54 ± .08	.84 ± .06	.78 ± .07	.74 ± .07
PRED	.31 ± .11	.37 ± .11	.29 ± .10	.52 ± .15	.50 ± .13	.45 ± .12
E1	.34 ± .17	.19 ± .14	.18 ± .13	.21 ± .14	.10 ± .10	.04 ± .07
E2	.34 ± .08	.37 ± .08	.46 ± .08	.16 ± .06	.21 ± .07	.26 ± .07

Predictive Indices
Binary Regression (LS)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.72 ± .16	.83 ± .13	.78 ± .14	.93 ± .09	1.00 ± .00	.96 ± .07
SPEC	.51 ± .08	.54 ± .08	.58 ± .08	.75 ± .07	.53 ± .08	.65 ± .08
PRED	.25 ± .09	.33 ± .10	.30 ± .10	.45 ± .12	.34 ± .10	.38 ± .11
E1	.27 ± .16	.16 ± .13	.22 ± .14	.07 ± .09	.00 ± .00	.04 ± .07
E2	.49 ± .08	.45 ± .08	.42 ± .08	.25 ± .07	.47 ± .08	.35 ± .08

Note: Results are reported as estimate ± standard error.

Predictive Indices
Binary Regression (WLS)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.72 ± .16	.84 ± .13	.78 ± .14	.82 ± .13	1.00 ± .00	.96 ± .07
SPEC	.51 ± .08	.54 ± .08	.64 ± .08	.75 ± .07	.53 ± .08	.69 ± .08
PRED	.25 ± .07	.33 ± .10	.33 ± .11	.42 ± .13	.34 ± .10	.40 ± .11
E1	.27 ± .16	.16 ± .13	.22 ± .14	.18 ± .13	.00 ± .00	.04 ± .07
E2	.49 ± .08	.46 ± .08	.36 ± .08	.25 ± .07	.47 ± .08	.31 ± .08

Predictive Indices
Binary Regression-Ridge (BR)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.27 ± .15	.66 ± .17	.00 ± .08	1.00 ± .00	.00 ± .00	.96 ± .07
SPEC	.84 ± .06	.68 ± .08	1.00 ± .00	.66 ± .08	.60 ± .08	.65 ± .08
PRED	.29 ± .17	.35 ± .12	.00 ± .00	.40 ± .11	.38 ± .10	.38 ± .11
E1	.72 ± .15	.34 ± .17	1.00 ± .00	.00 ± .00	.00 ± .00	.04 ± .07
E2	.15 ± .06	.32 ± .08	.00 ± .00	.33 ± .08	.40 ± .08	.35 ± .08

Note: Results are reported as estimate ± standard error.

Predictive Indices
Binary Regression (BRWLS)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.65 ± .17	.84 ± .13	.71 ± .16	.75 ± .15	1.00 ± .00	.96 ± .07
SPEC	.66 ± .08	.54 ± .08	.67 ± .08	.72 ± .08	.60 ± .08	.65 ± .08
PRED	.32 ± .11	.33 ± .10	.33 ± .11	.38 ± .12	.38 ± .10	.38 ± .11
E1	.34 ± .17	.15 ± .13	.29 ± .16	.25 ± .15	.00 ± .00	.04 ± .07
E2	.34 ± .08	.46 ± .08	.33 ± .08	.28 ± .08	.39 ± .08	.35 ± .18

Predictive Indices
Logistic Discrimination (LD)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.65 ± .17	.81 ± .14	.78 ± .14	.78 ± .14	.90 ± .10	.92 ± .09
SPEC	.66 ± .08	.63 ± .08	.61 ± .08	.84 ± .06	.78 ± .07	.77 ± .07
PRED	.31 ± .11	.37 ± .11	.32 ± .10	.52 ± .15	.50 ± .13	.47 ± .13
E1	.34 ± .17	.19 ± .14	.22 ± .14	.21 ± .14	.10 ± .10	.07 ± .09
E2	.34 ± .08	.37 ± .08	.39 ± .08	.16 ± .06	.21 ± .07	.22 ± .07

Note: Results are reported as estimate ± standard error.

Predictive Indices
Entropy Minimax Pattern Discovery (EMPD)

	Class					
Indices	I	II	III	IV	V	VI
SEN	.65 ± .17	.81 ± .14	.82 ± .13	.78 ± .14	.90 ± .10	.96 ± .07
SPEC	.66 ± .08	.63 ± .08	.54 ± .08	.84 ± .06	.78 ± .07	.74 ± .07
PRED	.31 ± .11	.37 ± .11	.29 ± .10	.52 ± .15	.50 ± .13	.45 ± .12
E1	.34 ± .17	.18 ± .14	.18 ± .13	.21 ± .14	.10 ± .10	.04 ± .07
E2	.34 ± .08	.37 ± .08	.46 ± .08	.16 ± .06	.21 ± .07	.26 ± .07

APPENDIX E

THE PREDICTIVE INDICES OF EACH MODEL WHEN PREDICTING TO A DIFFERENT RELATIONAL STRUCTURE

Prediction from Prior Knowledge of One Relational Structure to Another Different Relational Structure
Bayesian (B) Model

Prior Knowledge		Class I	Class III	Class IV	Class VI	Average
I	SEN	.65	.78	.93	.96	.89
	SPEC	.66	.66	.74	.70	.70
	PRED	.31	.34	.45	.41	.40
	E1	.35	.22	.07	.03	.11
	E2	.34	.34	.25	.29	.29
III	SEN	.55	.82	.86	.92	.78
	SPEC	.65	.54	.66	.70	.67
	PRED	.28	.29	.37	.40	.35
	E1	.45	.18	.14	.08	.22
	E2	.35	.46	.34	.30	.33
IV	SEN	.45	.71	.78	.92	.69
	SPEC	.76	.67	.84	.74	.72
	PRED	.30	.33	.52	.45	.36
	E1	.55	.28	.21	.08	.31
	E2	.24	.33	.16	.26	.28
VI	SEN	.45	.71	.78	.92	.65
	SPEC	.76	.67	.83	.77	.75
	PRED	.30	.33	.52	.47	.38
	E1	.55	.28	.22	.07	.35
	E2	.24	.33	.17	.22	.25

Prediction from Prior Knowledge of One Relational Structure to Another Different Relational Structure
Bayesian W/Baduhur (BB) Model

Prior Knowledge		Class I	Class III	Class IV	Class VI	Average
	SEN	.65	.78	.96	.96	.90
	SPEC	.66	.66	.66	.69	.67
I	PRED	.31	.34	.39	.41	.38
	E1	.34	.21	.03	.03	.09
	E2	.34	.34	.33	.30	.32
	SEN	.62	.82	.89	.96	.82
	SPEC	.54	.54	.53	.60	.56
III	PRED	.25	.29	.30	.35	.30
	E1	.38	.18	.11	.04	.18
	E2	.46	.46	.47	.40	.44
	SEN	.44	.71	.78	.96	.70
	SPEC	.76	.70	.84	.73	.73
IV	PRED	.30	.36	.52	.44	.37
	E1	.55	.29	.21	.03	.29
	E2	.23	.30	.16	.26	.26
	SEN	.45	.71	.85	.96	.67
	SPEC	.76	.70	.79	.74	.75
V	PRED	.30	.36	.48	.45	.38
	E1	.55	.29	.14	.04	.33
	E2	.23	.30	.21	.26	.25

Prediction from Prior Knowledge of One Relational Structure to Another Different Relational Structure
Binary Regression (BLS) Model

Prior Knowledge		Class I	Class III	Class IV	Class VI	Average
	SEN	.76	.78	.92	.96	.89
	SPEC	.51	.59	.60	.62	.60
I	PRED	.55	.31	.35	.35	.34
	E1	.24	.22	.07	.03	.11
	E2	.49	.41	.39	.37	.39
	SEN	.62	.78	.86	.92	.80
	SPEC	.51	.58	.51	.65	.56
III	PRED	.23	.30	.28	.37	.29
	E1	.38	.22	.14	.08	.20
	E2	.48	.42	.49	.35	.44
	SEN	.65	.78	.93	.96	.80
	SPEC	.65	.66	.75	.70	.67
IV	PRED	.31	.34	.45	.41	.35
	E1	.34	.22	.07	.03	.20
	E2	.34	.34	.25	.29	.32
	SEN	.72	.78	1.00	.96	.83
	SPEC	.52	.62	.66	.65	.60
VI	PRED	.26	.32	.40	.38	.33
	E1	.27	.21	.00	.04	.16
	E2	.47	.38	.33	.35	.39

Prediction from Prior Knowledge of One Relational Structure to Another Different Relational Structure
Entropy Minimax Pattern Discovery (EMPD) Model

Prior Knowledge		Class I	III	IV	VI	Average
I	SEN	.65	.78	.93	.96	.89
	SPEC	.66	.66	.75	.70	.70
	PRED	.31	.34	.45	.41	.40
	E1	.34	.22	.07	.04	.11
	E2	.34	.34	.25	.30	.30
III	SEN	.62	.82	.89	.96	.72
	SPEC	.54	.54	.53	.60	.56
	PRED	.25	.29	.30	.35	.30
	E1	.38	.18	.11	.04	.28
	E2	.46	.46	.47	.40	.44
IV	SEN	.45	.71	.78	.92	.69
	SPEC	.76	.67	.84	.74	.72
	PRED	.30	.33	.52	.45	.36
	E1	.55	.28	.21	.08	.31
	E2	.24	.33	.16	.26	.28
VI	SEN	.45	.71	.78	.96	.65
	SPEC	.76	.67	.83	.74	.75
	PRED	.30	.33	.52	.45	.38
	E1	.55	.28	.22	.04	.35
	E2	.24	.33	.17	.26	.25

Prediction from Prior Knowledge of One Relational Structure to Another Different Relational Structure
Logistic Discrimination (LD) Model

Prior Knowledge		Class I	III	IV	VI	Average
I	SEN	.65	.78	.93	.96	.89
	SPEC	.66	.66	.74	.70	.70
	PRED	.31	.34	.45	.41	.40
	E1	.34	.22	.07	.03	.11
	E2	.34	.34	.25	.29	.29
III	SEN	.55	.78	.86	.92	.78
	SPEC	.65	.61	.66	.70	.67
	PRED	.28	.32	.37	.40	.35
	E1	.45	.22	.14	.08	.22
	E2	.35	.39	.34	.30	.33
IV	SEN	.45	.71	.78	.96	.71
	SPEC	.76	.70	.84	.74	.73
	PRED	.31	.36	.52	.44	.37
	E1	.55	.29	.21	.03	.29
	E2	.23	.30	.16	.25	.26
VI	SEN	.27	.71	.60	.92	.53
	SPEC	.84	.72	.92	.77	.83
	PRED	.29	.37	.63	.47	.43
	E1	.72	.28	.39	.07	.46
	E2	.15	.27	.08	.22	.17

APPENDIX F

THE RELATIONAL STRUCTURE FOR EACH SPECIAL CLASS

Correlational and Variance-Covariance Matrices for Special Classes

	Class		
	Mixed	Suppressor	High Correlation
R_p	1.00 0.20 1.00 0.40 0.60 1.00 0.50 0.50 0.20 1.00	1.00 0.40 1.00 0.01 0.10 1.00 0.09 0.40 0.20 1.00	1.00 0.80 1.00 0.80 0.80 1.00 0.50 0.50 0.50 1.00
R_s	1.00 0.55 1.00 0.27 0.37 1.00 0.46 0.48 0.26 1.00	1.00 0.36 1.00 0.01 0.10 1.00 0.08 0.37 0.15 1.00	1.00 0.70 1.00 0.77 0.71 1.00 0.46 0.47 0.41 1.00
Σ_p	.250 .200 .250 .100 .125 .250 .090 .100 .040 .160	.250 .100 .250 .002 .020 .250 .020 .080 .040 .160	.250 .400 .250 .400 .400 .250 .100 .100 .100 .160
Σ_s	.250 .140 .250 .070 .100 .250 .100 .100 .050 .150	.250 .090 .250 .003 .030 .250 .018 .070 .030 .160	.250 .350 .250 .380 .350 .250 .080 .080 .070 .140

Correlational and Variance-Covariance Matrices
for Special Classes for Split-Samples

	Class		
	Mixed	Suppressor	High Correlation
R_1	1.00 0.49 1.00 0.24 0.34 1.00 0.46 0.48 0.25 1.00	1.00 0.24 1.00 0.08 0.14 1.00 0.10 0.46 0.04 1.00	1.00 0.76 1.00 0.84 0.76 1.00 0.48 0.47 0.45 1.00
R_2	1.00 0.59 1.00 0.29 0.40 1.00 0.49 0.50 0.28 1.00	1.00 0.49 1.00 0.06 0.06 1.00 0.08 0.28 0.27 1.00	1.00 0.78 1.00 0.83 0.78 1.00 0.45 0.46 0.42 1.00
Σ_1	.250 .120 .250 .060 .090 .250 .090 .090 .040 .150	.250 .060 .250 .020 .030 .240 .020 .090 .008 .160	.250 .380 .250 .420 .380 .250 .090 .090 .080 .140
Σ_2	.250 .140 .250 .070 .100 .250 .090 .090 .050 .150	.250 .120 .250 .010 .020 .250 .010 .050 .050 .150	.250 .400 .250 .400 .400 .250 .080 .080 .070 .140

APPENDIX G

THE ESTIMATED DIAGNOSTIC PROBABILITIES AND PREDICTIVE INDICES FOR EACH MODEL WITHIN EACH SPECIAL CLASS

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^p Number of Patterns of Special Classes
Bayesian (B)

	Class		
Pattern	Mixed	Suppressor	High Correlation
111	.54	.40	.43
110	.42	.42	.30
100	.00	.00	.00
001	.00	.04	.00
011	.00	.35	.00
101	.00	.00	.00
010	.00	.45	.00
000	.00	.01	.00

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^p Number of Patterns of Special Classes
Bayesian W/Baduhur (BB)

	Class		
Pattern	Mixed	Suppressor	High Correlation
111	.5250	.4000	.4262
110	.4118	.4118	.2500
100	.0000	.0000	.0000
001	.0000	.0476	.0000
011	.0000	.3333	.0000
101	.0000	.0000	.0000
010	.0000	.4444	.0000
000	.0000	.0400	.0000

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^p Number of Patterns of Special Classes
Binary Regression (BLS)

Estimated Parameters	Class		
	Mixed	Suppressor	High Correlation
Constant	-.08	.04	-.02
Symptom 1	.22	-.01	.20
Symptom 2	.25	.37	.17
Symptom 3	.06	-.02	.04
Pattern	**Mixed**	**Suppressor**	**High Correlation**
111	.45	.38	.39
110	.39	.40	.35
100	.14	.03	.18
001	.00	.02	.01
011	.23	.39	.19
101	.19	.01	.22
010	.17	.41	.15
000	.00	.04	.00

Estimated Parameters and Estimated Diagnostic Probabilities for 2^p Number of Patterns of Special Classes Binary Regression (BWLS)

Estimated Parameters	Class		
	Mixed	Suppressor	High Correlation
Constant	.11	.04	.005
Symptom 1	.12	-.04	.22
Symptom 2	.23	.38	.15
Symptom 3	-.13	.00	.002
Pattern	**Mixed**	**Suppressor**	**High Correlation**
111	.32	.38	.37
110	.45	.38	.33
100	.22	.00	.22
001	.00	.04	.00
011	.20	.42	.15
101	.09	.00	.24
010	.34	.42	.15
000	.11	.04	.00

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^p Number of Patterns of Special Classes
Binary Regression-Ridge (BR)

Estimated Parameters	Classes		
	Mixed	Suppressor	High Correlation
Constant			
Symptom 1	.06	.00	.12
Symptom 2	.17	.21	.12
Symptom 3	.16	.06	.09
Pattern	**Mixed**	**Suppressor**	**High Correlation**
111	.39	.23	.34
110	.23	.21	.25
100	.06	.00	.12
001	.16	.02	.09
011	.33	.23	.21
101	.22	.02	.21
010	.17	.21	.12
000	.00	.00	.00

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^P Number of Patterns of Special Classes
Logistic Discrimination (LD)

Estimated Parameters	Class		
	Mixed	Suppressor	High Correlation
Constant	-19.15	-3.54	-17.40
Symptom 1	9.35	-.08	8.13
Symptom 2	9.44	3.27	8.17
Symptom 3	.46	-.16	.80
Pattern	**Mixed**	**Suppressor**	**High Correlation**
111	.52	.37	.57
110	.41	.41	.25
100	.00	.03	.00
001	.00	.02	.00
011	.00	.39	.00
101	.00	.02	.00
010	.00	.43	.00
000	.00	.03	.00

Estimated Parameters and Estimated Diagnostic Probabilities
for 2^P Number of Patterns of Special Classes
Entropy Minimax Pattern Discovery (EMPD)

	Class		
Pattern	Mixed	Suppressor	High Correlation
111	.53 (.26)	.40 (.19)	.43 (.40)
110	.42 (.11)	.41 (.11)	.31 (.02)
100	.00 (.02)	.01 (.02)	.06 (.01)
001	.00 (.02)	.05 (.09)	.006 (.02)
011	.03 (.02)	.35 (.09)	.43 (.02)
101	.00 (.02)	.01 (.02)	.06 (.01)
010	.03 (.02)	.45 (.05)	.006 (.02)
000	.00 (.02)	.05 (.09)	.006 (.02)

Note: The entries in the parentheses are the entropy values (H) for that particular pattern within that particular class.

Predictive Indices for Special Classes for Various Models

Models	Indices	Class		
		Mixed	Suppression	High Correlation
B	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.76 ± .07	.60 ± .08	.63 ± .08
	PRED	.49 ± .12	.31 ± .10	.37 ± .11
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.24 ± .07	.40 ± .08	.37 ± .08
BB	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.76 ± .07	.60 ± .08	.63 ± .08
	PRED	.49 ± .12	.31 ± .10	.37 ± .11
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.24 ± .07	.40 ± .08	.37 ± .08
BLS	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.67 ± .08	.60 ± .08	.59 ± .08
	PRED	.41 ± .11	.31 ± .10	.35 ± .10
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.33 ± .08	.40 ± .08	.41 ± .08
BWLS	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.58 ± .08	.60 ± .08	.55 ± .08
	PRED	.35 ± .10	.31 ± .10	.33 ± .10
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.42 ± .08	.40 ± .08	.45 ± .08
BR	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.61 ± .08	.60 ± .08	.58 ± .08
	PRED	.37 ± .10	.31 ± .10	.34 ± .10
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.39 ± .08	.40 ± .08	.42 ± .08
LD	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.76 ± .07	.60 ± .08	.63 ± .08
	PRED	.49 ± .12	.31 ± .10	.37 ± .11
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.24 ± .07	.41 ± .08	.37 ± .08
EMPD	SEN	1.00 ± .00	.76 ± .15	1.00 ± .00
	SPEC	.76 ± .07	.60 ± .08	.61 ± .08
	PRED	.49 ± .12	.31 ± .10	.36 ± .10
	E1	.00 ± .00	.24 ± .15	.00 ± .00
	E2	.24 ± .07	.41 ± .08	.39 ± .08

Note: Results are reported as estimate ± standard error.

APPENDIX H

BRAIN SCAN EVALUATION QUESTIONNAIRE

BRAIN SCAN EVALUATION STUDY QUESTIONNAIRE Date: No:

COO-2427-5

I. Patient Identification

JHH History Number _____

Patient Name _____

Sex: - - - - - O Male
O Female

Age: _____

O Outpatient
O Inpatient

Fill in or use JHH Patient Identification Card.

II. General Information

1. Physician Filling Out Form _____

2. Date _____

3. Is the decision to do a brain scan based (in part) on the results of another diagnostic procedure? O yes O no

If yes, what was it? _____

4. Has the patient had a:

	yes	no	normal	abnormal
Lumbar Puncture	O	O	O	O
EEG	O	O	O	O
Skull X-ray	O	O	O	O
Arteriorgram	O	O	O	O
Echo	O	O	O	O

III. Definitions of Motivational Factors

1. Efficacy:

The use of a diagnostic procedure is motivated by efficacy if the outcome of the procedure could contribute to-or effect-a change in the course of the patient's disease.

2. Defense:

A procedure is being used defensively if its use is motivated by either potential peer incrimination or legal responsibility.

3. Innovation-Curiosity

Innovation-Curiosity is the principal motivating force if the objective in ordering a procedure is simply to find out what the result will be for this particular case. It may even help the patient.

IV. Historical Data

1. Headache - - - - - - - - - - - - - O yes O no

   a) Duration - - - - - - - - - - O <1 Week
      O 1 Week to 1 Mo.
      O 1 Mo. to 3 Mo.
      O >3 Mo.

   b) Continuity - - - - - - - - O Continuous
      O Intermittent

   c) Severity - - - - - - - - - O Mild
      O Moderate
      O Severe

   d) Location if diffuse - - - - O Bilateral
      O Unilateral

   e) Location if focal - - - - - O Retroorbital
      O Frontal
      O Temporal
      O Parietal
      O Occipital

2. Seizure - - - - - - - - - - - - - O yes O no

   a) Number of Episodes - - - - O Single (First)
      O Multiple >10
      O Multiple <10 (Longstanding)

   b) Location - - - - - - - - - - O Generalized
      O Focal

   c) Type - - - - - - - - - - - - O Major Motor
      O Minor Motor
      O Temporal Lobe
      O Other

   d) Is Seizure Pattern - - - - O yes
      Changing? O no

   e) Pertinent Family History O yes
      of Seizures O no
      O unknown

3. Neoplasm - - - - - - - - - - - - O yes O no O suspect

   a) Location - - - - - - - - - - O Brain
      O Lung
      O Breast
      O Other

   b) Pertinent Family History O yes
      of Neoplasm O no
      O unknown

4. History of Trauma - - - - - - - O yes O no

V. Physical Examination

1. Cortical Deficit - - - - - - - - ○ yes ○ no

   a) State of Consciousness (Indicate by an X anywhere on the scale)
   1 Normal
   2
   3
   4
   5 Abnormal

   b) Generalized Deficit - - - - ○ Dementia ○ Other Abnormality

   If "Other" what is it? _____

   c) Focal Deficit - - - - - - - - ○ yes ○ no

   If yes, what is it? _____

2. Motor Deficit - - - - - - - - - - ○ yes ○ no

   a) If yes: Location _____
   Lateralization _____
   Severity
   1 Mild
   2
   3
   4
   5 Severe

   b) Ataxia - - - - - - - - - - - ○ yes ○ no

   Type? _____

   c) Involuntary Movement - - - - ○ yes ○ no

   Type? _____

   d) Reflex Abnormality - - - - - ○ yes ○ no

   Type? _____

   e) Abnormal Gait - - - - - - - - ○ yes ○ no

3. Sensory Abnormality - - - - - - - ○ yes ○ no

   If yes, what is it? _____

4. Visual Field Defect - - - - - - - ○ yes ○ no

   If yes, what is it? _____

5. Alteration of Brain Stem Function Including Eye Movements - - - - ○ yes ○ no

   If yes, what is it? _____

COO-2427-5

VI. Prospective Outcomes

1. Subjective Probabilities (Mark anywhere on the scale, probabilities need not total 100%)

a) In your opinion what is the probability that this brain scan will be normal

0% 20% 40% 60% 80% 100%

b) With what probability do you suspect each of the following diagnosis?

Note: Probabilities need not total 100%

0% 20% 40% 60% 80% 100%

Subdural Hematoma

Vascular Malformation

Stroke (e.g., TIA, Hemorrhage or Infarction)

Cerebral Infection

Cerebral Tumor (Primary)

Cerebral Tumor (Metastasis)

Other Pathology

c) What is your Presumtive Diagnosis? _____

d) What do you feel the odds are that your diagnosis is correct?

Certain 10:1 Even 1:1 Remote 1:10

e) Will you alter your management of this patient if the result of this brain scan is:

(i) Normal O yes O no

(ii) Abnormal O yes O no

Taking total motivation to request brain scan as 100 - How do you distribute your subjective motivation over the following reasons (as defined above) for requesting this examination?

Efficacy	_____%
Defense	_____%
Innovation-Curiosity	_____%
Other	_____%
Total	100

APPENDIX I

THE RELATIONAL STRUCTURE FOR THE BRAIN SCAN STUDY

Correlation Matrix and Base Rates of the Symptoms and the Disease for Brain Scan

Headache	1.000						
Seizure	-.006	1.000					
Cort. def.	-.256	-.079	1.000				
Motor Def.	-.081	-.172	.318	1.000			
Sen. ab.	.054	-.035	-.151	.215	1.000		
Visual	-.052	-.113	.151	.031	-.190	1.000	
Outcome	-.015	-.130	.215	.225	.097	.064	1.000

$P_{headache} = .52$

$P_{seizure} = .31$

$P_{cort.\ def.} = .31$

$P_{motor\ def.} = .30$

$P_{sen.\ ab.} = .24$

$P_{visual} = .17$

$P_{outcome} = .10$

APPENDIX J

THE RELATIONAL STRUCTURE FOR THE SPLIT SAMPLES OF THE BRAIN SCAN STUDY

Correlation and Variance-Covariance Matrices for Brain-Scan Data for Split Samples

	Sample I							Sample II						
	1.00							1.00						
	-.03	1.00						.05	1.00					
	-.28	.12	1.00					-.23	-.30	1.00				
R	-.08	.01	.15	1.00				-.06	-.39	.49	1.00			
	-.01	-.23	-.10	.34	1.00			.15	.18	-.21	.06	1.00		
	.01	-.05	.09	.09	-.27	1.00		-.12	-.16	.20	-.03	-.09	1.00	
	.05	-.23	.12	.12	.33	.07	1.00	-.08	-.02	.31	.34	-.16	.05	1.00
	.250							.250						
	-.010	.250						.010	.200					
	-.060	.030	.220					-.050	-.060	.210				
‡	-.020	.002	.030	.220				-.010	-.070	.090	.200			
	-.002	-.050	-.020	.070	.200			.030	.030	-.040	.010	.170		
	.002	-.010	.020	.020	-.050	.140		-.020	-.030	.030	-.005	-.010	.150	
	.010	-.030	.020	.020	.040	.010	.080	-.010	-.001	.040	.040	-.180	.005	.080

APPENDIX K

THE ESTIMATED PARAMETERS AND PREDICTIVE INDICES OF EACH MODEL FOR THE BRAIN SCAN STUDY

Estimated Parameters for Brain Scan by Various Models

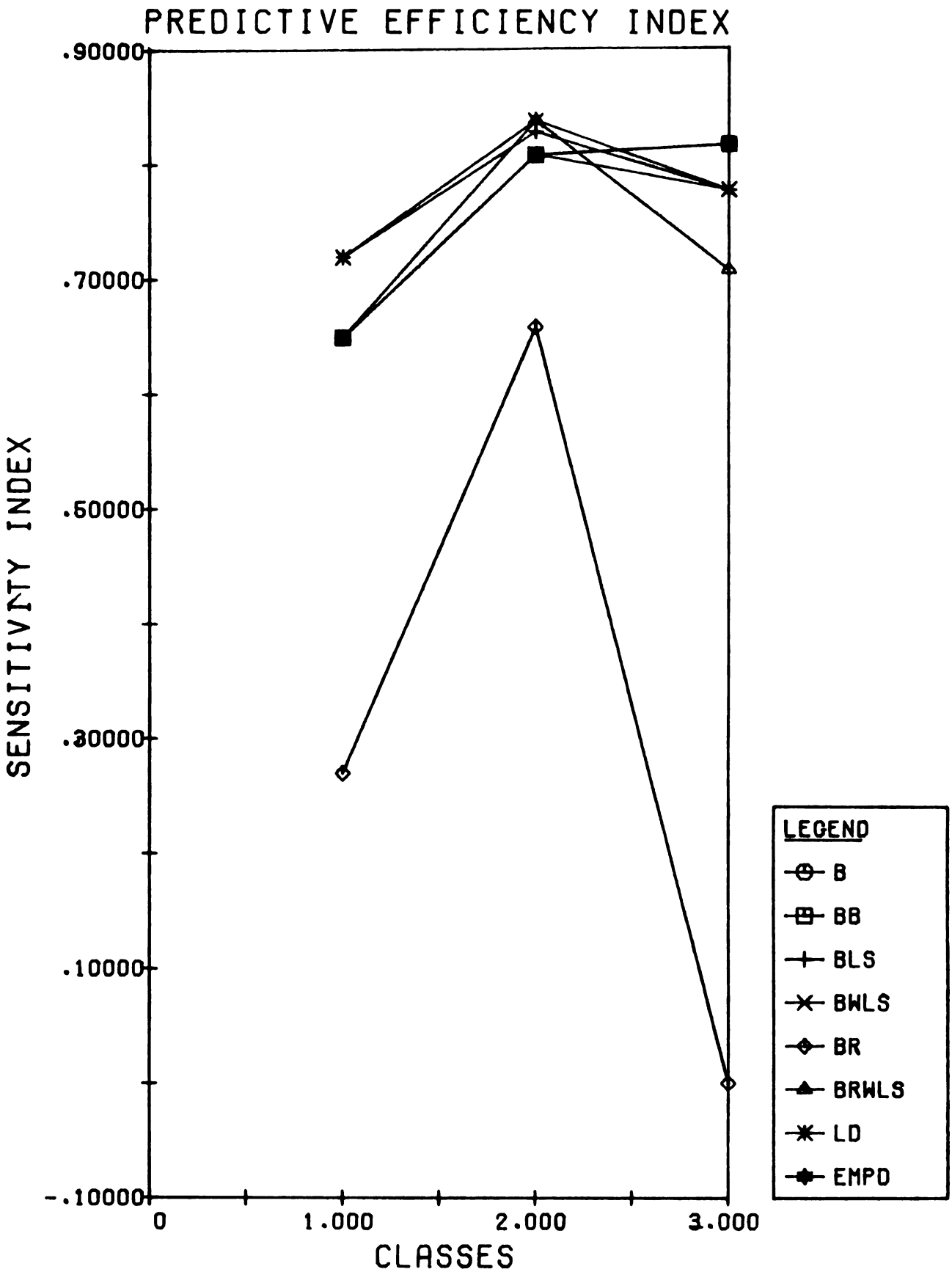
	Models				
Symptom	BLS	BWLS	BR	BRWLS	LD
Constant	-.02	-.06	.00		-11.23
Headache	.06	.13	.03	1.96	.58
Seizure	-.09	.41	-.07	-6.38	-8.57
Cortical deficit	.12	-.29	.12	.39	1.58
Motor deficit	-.03	.08	.06	-2.15	-.47
Sensory abnormal	.24	.15	.05	.25	9.89
Visual defect	.12	.09	.02	-.05	9.16

Prediction Indices for Various Models for Brain Scan

	Models						
Indices	BB	BLS	BWLS	BR	BRWLS	LD	EMPD
SEN	.00	.50	.25	.75	.75	.20	.00
SPEC	.95	.54	.61	.69	.41	.74	.97
PRED	.00	.10	.06	.20	.11	.09	.00
E1	1.00	.50	.75	.25	.25	.75	1.00
E2	.05	.46	.38	.31	.59	.26	.02

APPENDIX L

GRAPH FOR THE PREDICTIVE INDICES FOR EACH PROBABILITY MODEL WHEN THE INTERCORRELATION OF THE SYMPTOMS ARE LOW


PREDICTIVE EFFICIENCY INDEX
.90000
.70000
.50000
.30000
.10000
-.10000
SENSITIVITY INDEX
0
1.000
2.000
3.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD

.900

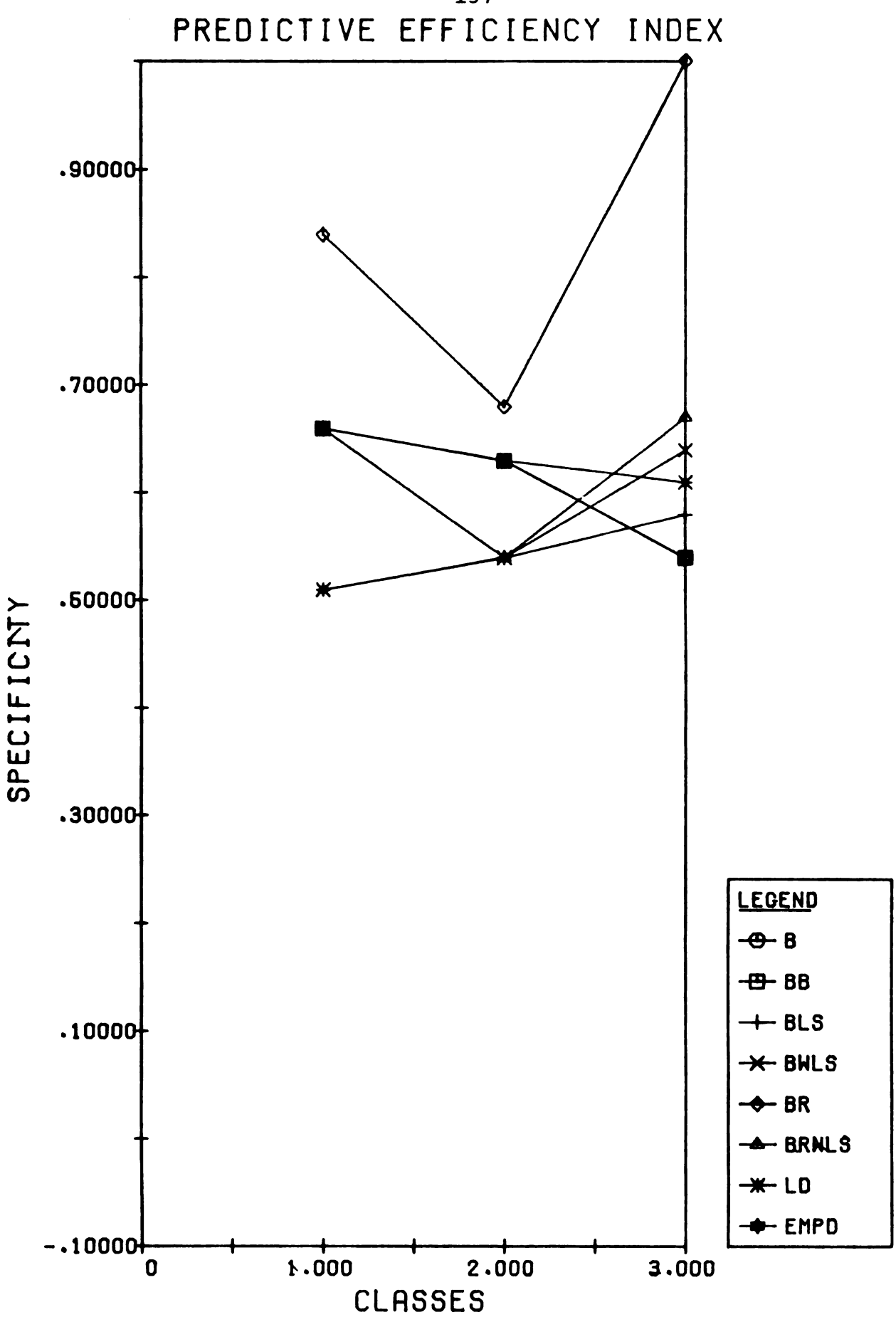
.700

SPECIFICITY .5000

.3000

.1000

-.10000


PREDICTIVE EFFICIENCY INDEX
SPECIFICITY
.90000
.70000
.60000
.30000
.10000
-.10000
0
1.000
2.000
3.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD

PREDICTIVE VALUE

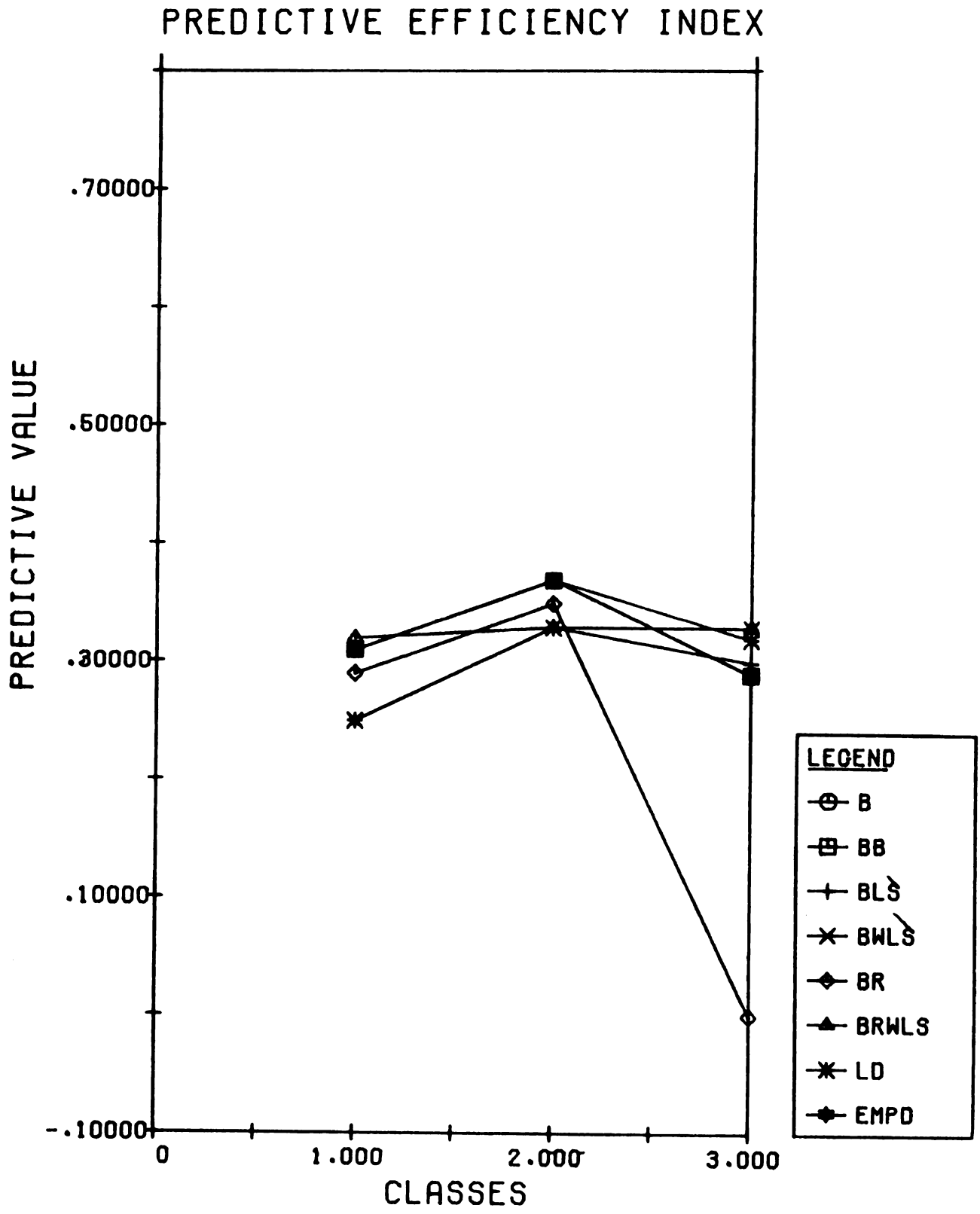
.7000

.5000

.3000

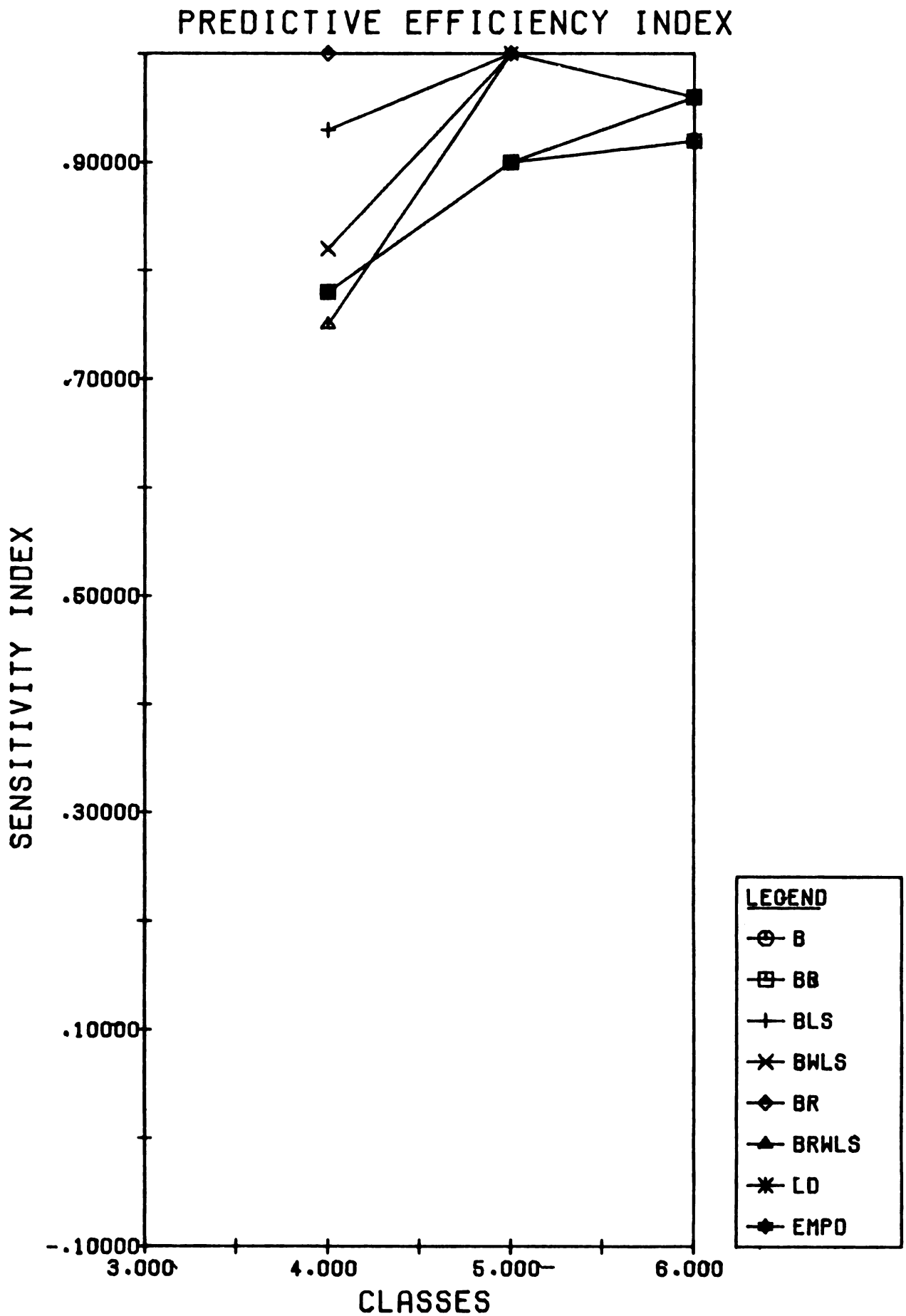
.1000

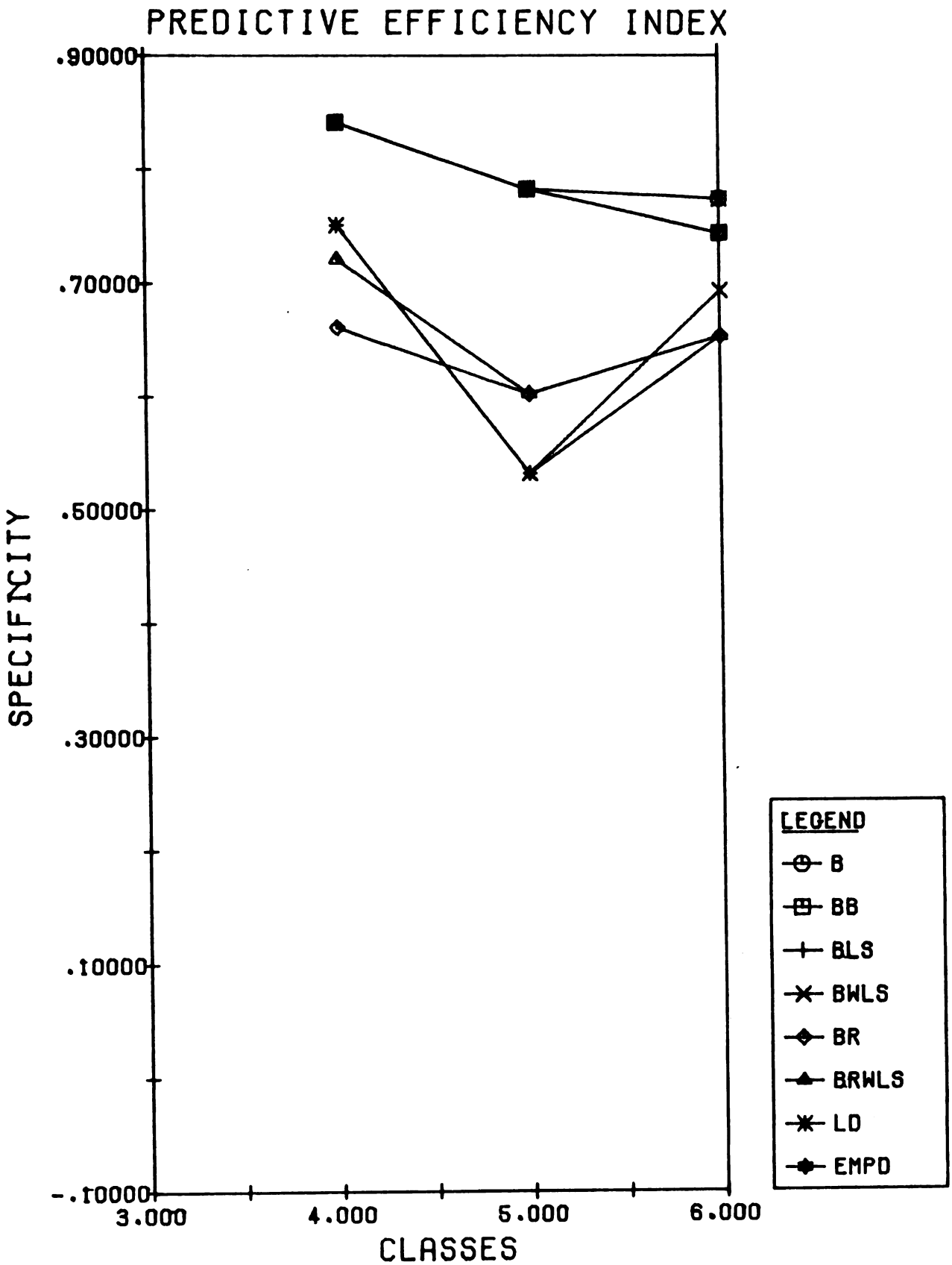
-.1000

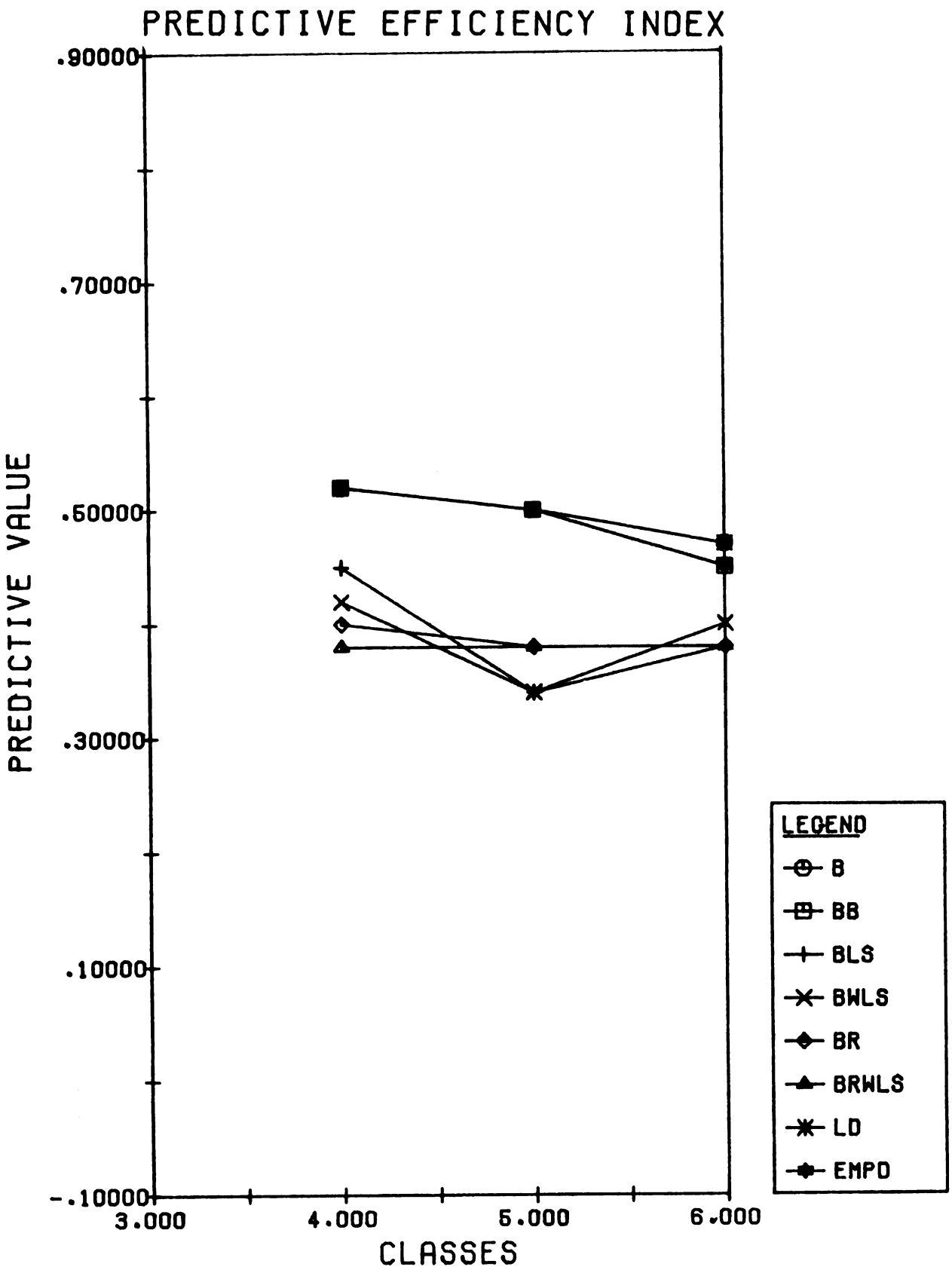

PREDICTIVE EFFICIENCY INDEX
PREDICTIVE VALUE
.70000
.50000
.30000
.10000
-.10000
0
1.000
2.000
3.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD

APPENDIX M

GRAPH FOR THE PREDICTIVE INDICES FOR EACH PROBABILITY MODEL WHEN THE INTERCORRELATION OF THE SYMPTOMS ARE HIGH


PREDICTIVE EFFICIENCY INDEX
SENSITIVITY INDEX
.90000
.70000
.50000
.30000
.10000
-.10000
3.000
4.000
5.000
6.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD


PREDICTIVE EFFICIENCY INDEX
.90000
.70000
.50000
.30000
.10000
-.10000
SPECIFICITY
3.000
4.000
5.000
6.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD


PREDICTIVE EFFICIENCY INDEX
.90000
.70000
.50000
.30000
.10000
-.10000
PREDICTIVE VALUE
3.000
4.000
5.000
6.000
CLASSES
LEGEND
B
BB
BLS
BWLS
BR
BRWLS
LD
EMPD

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, J. A. Separate sample logistic discrimination. Biometrika, 1972, 59, 19-34.

Anderson, J. A. Logistic discrimination with medical application. In T. Cacoullos (Ed.), Discriminative Analysis and Application. New York: Academic Press, 1973.

Anderson, J. A. Diagnosis by logistic discrimination function; further practical problems and results. Applied Statistics, 1974, 23(3), 397-404.

Anderson, J. A., Whaley, K., Williamson, J., & Buchanan, W. W. A statistical aid to the diagnosis of kertoconjunctivitis sioca. Quarterly Journal of Medicine, New Series, 1972, 41, 175-189.

Anderson, T. W. An introduction to multivariate statistical analysis. New York: Wiley and Sons, 1958.

Bahadur, R. R. A representation of the joint distribution of responses to n dichotomous items. In H. Solomon (Ed.), Studies in item analysis and prediction. Stanford: Stanford University Press, 1961, pp. 1568-1588.

Bailey, N. T. J. Probability methods of diagnosis based on small samples. Biology and Medicine. London: Her Majesty's Stationery Office, 1965.

Bartlett, M. S. A note on multiplying factors for various chi-squared approximations. Journal of the Royal Statistical Society, 1954, 16, 296-298.

Barnoon, S., & Wolfe, H. Measuring the effectiveness of medical decisions. Springfield: Charles C. Thomas, 1972.

Bayes, T. Essays towards solving a problem in the doctrine of chances. Biometrika, 45, 293-315.

Bean, W. B. (Ed.). Aphorisms from Osler's bedside teaching and writings. New York: Schuman, 1950.

Blair, J. C. Mammography: A contrary view. Annals of Internal Medicine, 1976, 84, 77-84.

Bock, R. D. Estimating multinomial response relation. In Boss (Ed.), Essays in probability and statistics. Chapel Hill: University of North Carolina Press, 1970.

Bock, R. D. Multivariate statistical methods in behavioral research. New York: McGraw-Hill Book Co., 1975.

Bowman, M. J. Risk and uncertainty. In J. Gould and W. Kalb (Eds.), A dictionary of the social science. New York: The Free Press, 1964.

Box, G. E. A general distribution theory for a class of likelihood criteria. Biometrika, 1950, 36, 317-346.

Christensen, R. A. A general approach to pattern discovery. Technical Report, 1967, 20, University of California, Berkeley, Computer Center.

Christensen, R. A. Pattern discovery program for analyzing qualitative and quantitative data. Behaviour Science, 1968, 13, 423.

Christensen, R. A. Entropy minimax method of pattern discovery and probability determination. New York: Arthur D. Little, Inc., 1972.

Christensen, R. A. Entropy to minimax: A non-bayesian approach to probability estimation from empirical data. Paper presented at the International Conference on Cybernetics and Soc., Boston, November 1973.

Cohen, J. Psychological probability. Cambridge: Schenkman Publishing Co., 1973.

Cohen, J., & Cohen, P. Applied regression analysis for the behavioral science. New Jersey: Lawrence Erlbaum Assn. Inc. Publishers, 1975.

Conger, A. J., & Jackson, D. N. Suppressor variables, prediction, and the interpretation of psychological relationships. Educational and Psychological Measurement, 1972, 32, 579-599.

Cornfield, J. Bayes theorem. Review of the International Statistical Institute, 1967, 35, 34.

Cox, D. R. Some procedures associated with the logistic qualitative response curve. In F. N. David (Ed.), Research paper in statistics: Festschrift for J. Neyman. London: Wiley, 1966, pp. 55-71.

Cox, D. R. The analysis of binary data. London: Methuen, 1970.

David, A. P. Properties of diagnostic data distribution. Biometrics, 1976, 32, 647-658.

Davies, P. Symptom diagnosis using bahadur's distribution. International Journal Bio-Medical Computing, 1972, 3, 307-312.

Dorland's Illustrated Medical Dictionary. Philadelphia: W. B. Saunder, 1974.

Draper, N. R., & Smith, H. Applied regression analysis. New York: John Wiley & Sons, 1966.

Elwood, J. H., & Mackenzie, G. The measurement and comparison of infant mortality risk by binary multiple regression analysis. Journal of Chronic Disease, 1971, 93-106.

Elstein, A. S. Clinical judgment: Psychological research and medical practice. Science, 1976, 194, 696-700.

Engel, R. L., & Davis, B. J. Medical diagnosis: Present, past, and future. Archives of Internal Medicine, 1963, 112, 512-519.

Feldstein, M. S. A binary variable multiple regression model of analysing factors affecting perinatal mortality and other outcomes of pregnancy. Journal of Royal Statistical Society, Ser. A, 1966, 129, 61-73.

Fisher, R. A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936, 8, 376-386.

Fraser, P. M., & Franklin, D. A. Mathematical models for the diagnosis of liver disease. Quarterly Journal of Medicine, New Series, 1974, 169, 73-88.

Galen, R., & Gambino, S. R. Beyond normality: The predictive value and efficiency of medical diagnosis. New York: Wiley, 1975.

Gustafson, D. H. Evaluation of probabilistic information processing in medical decision making. Organic Behavior and Human Performance, 1969, 4, 20-34.

Haase, R. The use of multiple regression analysis to generate conditional probabilities about the occurrence of events. Educational and Psychological Measurement, 1976, 36, 135-140.

Hall, G. H. Clinical application of bayes' theorem. Lancet, 1967, 2, 555.

Halperin, M., Blackwelder, W. C., & Verter, J. I. Estimation of the multivariate logistic risk function: A comparison of the discriminant function and maximum likelihood approaches. Journal of Chronic Disease, 1971, 24, 125-158.

Hammond, K. R., Thomas, T. R., Brehmer, B., & Steinmann, D. O. Social judgment theory. In M. F. Kaplan and S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press Inc., 1975.

Hartz, S., & Rosenberg, L. The computation of maximum likelihood estimates for the multiple logistic risk function for use with categorical data. Journal of Chronic Disease, 1975, 28, 421-429.

Harvey, A. M., & Bordley, J. B. Differential diagnosis. Philadelphia: W. B. Saunders Co., 1970.

Hasting, N. A. J., & Peacock, J. B. Statistical distribution. London: Butterworths, 1974.

Hempel, C. G. Philosophy of natural science. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Hoerl, A. E. Application of ridge analysis to regression problem. Chemical Engineering Progress, 1962, 58(3), 54-59.

Hoerl, A. E., & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970a, 12, 55-67.

Hoerl, A. E., & Kennard, R. W. Ridge regression: Applications to nonorthogonal problems. Technometrics, 1970b, 12, 69-82.

Johnson, N. Discrete distribution. Boston: Houghton Mifflin Co., 1969.

Jones, R. Probability estimation using a multinomial logistic function. Journal of Statistical Comp. Simulation, 1975, 3, 315-329.

Kaplan, A. The conduct of inquiry. San Francisco: Chandler Publishing Co., 1964.

Kaufmann, A. The science of decision making. New York: McGraw-Hill Book Co., 1968.

Kerlinger, F. N., & Pedhazur, E. J. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, Inc., 1973.

Knuth, D. The art of computer programming: Seminumerical algorithms (Vol. 2). Cambridge: Addison-Wesley Pub. Co., 1969.

Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Cambridge: Addison-Wesley Publishing Co., 1974.

Lubin, A. Some formulae for use with suppressor variables. Educational and Psychological Measurement, 1957, 17, 286-296.

Lusted, L. Introduction to medical decision making. Springfield: Charles C. Thomas, 1968.

MacBryde, C. M., & Blacklow, R. S. Signs and symptoms. Philadelphia: J. B. Lippincott Co., 1970.

Marquardt, D. W., & Snee, R. D. Ridge regression in practice. The American Statistician, 1975, 29(1), 3-20.

Mihram, G. A. Simulation: Statistical foundations and methodology. New York: Academic Press, 1972.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill Book Co., 1976.

Neter, J., & Wasserman, W. Applied linear statistical models. Homewood, Ill.: Richard D. Irwin Co., 1974.

Neyman, J. First course in statistics and probability. New York: Holt, 1950.

Overall, J. E., Manhattan, K., & Williams, C. Conditional probability program for diagnosis of thyroid function. Journal of American Medical Association, 1963, 183(5), 307-313.

Parker, R. D., & Lincoln, T. L. Medical diagnosis using bayes' theorem. Health Service Research, 1967, 2, 34.

Pearson, K. On the correlation of characters not quantitatively measurable. Royal Society Philosophical Transactions, Ser. A, 1900, 195, 1-47.

Potchen, E. J. Biologic considerations in anatomic imaging with radionuclides. Energy Research and Development Administration, Division of Biomedical and Environmental Research, Final Progress Report, July 1974-June 1975.

Rinaldo, J. A., Scheinok, P., & Rupe, C. E. Symptom diagnosis: A mathematical analysis of epigastric pain. Annals of Internal Medicine, 1963, 59(2), 145-154.

Scheff, T. J. Decision rules, types of error and their consequences in medical diagnosis. Behavioral Science, 1963, 8, 97-107.

Scheifly, V. M. Analysis of repeated measures data: A simulation study. Unpublished doctoral dissertation, 1974.

Scheinok, P. Symptom diagnosis: Bayes theorem and bahadur's distribution. International Journal of Bio-Medical Computing, 1972, 3, 17.

Schonbein, W. Analysis of decisions and information in patient management. In E. J. Potchen (Ed.), Current concepts in radiology. St. Louis: C. V. Mosby Co., 1975.

Shannon, E. C., & Weaver, W. The mathematical theory of communication. Urbana: University of Illinois Press, 1964.

Swets, J. A. Detection theory and psychophysics: A review. Psychometrika, 1961, 26, 49-63.

Taber, C. W. Taber's cyclopedic medical dictionary. Philadelphia: F. A. Davis Co., 1970.

Tatsuoka, M. M. Multivariate analysis in educational and psychological research. New York: Wiley, 1971.

Tou, J. T., & Gonzalez, R. C. Pattern recognition principles. London: Addison-Wesley Publishing Co., 1974.

Truett, J., Cornfield, J., & Kannel, W. B. A multivariate analysis of the risk of coronary heart disease in framingham. Journal of Chronic Disease, 1967, 20, 511-524.

Tversky, A. Assessing uncertainty. Journal of Royal Statistical Association, Ser. B, 1974, pp. 148-159.

Van de Geer. Introduction to multivariate analysis for the social science. San Francisco: W. H. Freeman and Co., 1971.

Vanderplas, J. M. A method of determining probabilities for correct use of bayes' theorem in medical diagnosis. Computers Biomedical Research, 1967, 1, 215-220.

Wakefield, E. G. Clinical diagnosis. New York: Appleton-Century-Crofts, Inc., 1955.

Walker, S. H., & Duncan, D. Estimation of the probability of an event as a function of several independent variables. Biometrika, 1967, 54, 167-179.

Warner, H. R., Toronto, A. F., Veasey, G., & Stephenson, R. A mathematical approach to medical diagnosis. Journal of American Medical Association, 1961, 177, 177-183.

Webster's New Collegiate Dictionary. Springfield, Mass.: G. & C. Merriam Co., Publishers, 1974.