THEMIS SIGNAL ANALYSIS STATISTICS RESEARCH PROGRAM

A DUALITY PROPERTY FOR BAYES RULES WITH APPLICATIONS

by

Charles Ray Johnson

Technical Report No. 75
Department of Statistics THEMIS Contract

Juee 26, 1970

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

DEPARTMENT OF STATISTICS
Southern Methodist University

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

A well known result in statistical decision theory is that many minimax rules also solve related Bayes or extended Bayes problems. This paper investigates the inverse of the above result which is that many Bayes or extended Bayes rules solve related minimax problems. This relationship, which does not seem to have been treated before, is an interesting one in that it gives an "objective" justification for a Bayes rule. Because of the similarity of this relationship to the duality principle in linear programming it has been given here the name "duality property for Bayes rules." A new concept called a "least favorable loss function," which shows promise of being useful because it allows the introduction of prior information into minimax problems, has been developed from the duality property. The definition, structure, interpretation, and application of the duality property are discussed in later chapters.

Since the proofs and discussions of this property of Bayes rules require the notation and an understanding of statistical decision theory, this chapter first reviews the necessary theory before introducing the concept of a duality property. In addition, brief outlines of the other chapters of this paper are presented.

Fundamental Ideas of Statistical Decision Theory

The notation and approach in this review of decision theory is taken primarily from Ferguson [5]. The roots of decision theory can be traced to game theory and utility theory as formulated by von Neuman and Morgenstern [17]. Wald recognized the usefulness of game theory in statistics; did research on this subject from 1939 to 1951, [18], [19], [20], [21], and [22]; and published a book containing his results [23] in 1950. Some of Wald's work was amplified in the text by Blackwell and Girshick [1] and in the text by Weiss [24]. Savage [14] employed the utility theory formulated by von Neuman and Morgenstern [17] and the personalistic definition of probability to justify and strengthen the decision theory approach to statistics. A later important work in decision theory and Bayesian statistics is by Raiffa and Schlaifer [12]. A sampling of the research that supported the above texts is reported in [2], [3], [4], [6], [7], [8], [10], [11], and [16].

The elements of game theory for a game in which nature takes the role of one of the players and the statistician takes the role of the other player are:

1. A non-empty set, Θ, of possible states of nature, referred to as the parameter space (space of actions for nature).
2. A non-empty set, A, of actions available to the statistician.
3. A loss function, $L(\theta, a)$, a real valued function with θ in Θ and a in A [for this paper $L(\theta, a) \geq 0$]. The statistician's loss, or payoff, is $L(\theta, a)$ when his action is a and nature's action is θ.

The game is for nature to choose a θ in Θ (θ is then said to be the "true state of nature"); and the statistician, without knowledge of nature's choice, is to select an action a in A. The game then terminates with the statistician losing an amount $L(\theta, a)$. Thus, the game is described by the **triplet** $(\Theta, \mathcal{A}, L)$. A very readable and entertaining introduction to the theory of games is contained in [25].

Statistical decision theory alters game theory by introducing an experiment that the statistician can use to gain knowledge about θ, the true state of nature. In other words, the game is defined as in game theory but with the statistician choosing an action without being totally ignorant about nature's choice of θ. Therefore, a statistical decision problem is defined to be a game $(\Theta, \mathcal{A}, L)$ coupled with an experiment involving a sample space $\mathcal{X}$, a random variable X (with outcome x in $\mathcal{X}$) whose distribution function $F(x|\theta)$ depends on the state θ in Θ chosen by nature. It should be noted that X, a, and θ might be vectors or even of a more general nature.

On the basis of the outcome of the experiment, x, the statistician chooses an action $\delta(x)$. The function δ, which maps $\mathcal{X}$ onto $\mathcal{A}$ is a non-randomized decision rule or elementary strategy for the statistician. The loss incurred by the statistician due to his use of δ is hence a random variable. The expectation of this loss as a function of θ is termed the "risk function," $R(\theta, \delta)$, where

$$R(\theta, \delta) = E[L(\theta, \delta)] = \int_{\mathcal{X}} L(\theta, \delta)\,dF(x|\theta) \quad .$$

In all the following work only non-randomized decision rules whose risk functions exist are considered and the prefix "non-randomized" is dropped in all further discussion.

The risk function is the sole criteria used by the statistician to select decision rules. This use of the risk function is justified by utility theory in [17], also see, for example, the discussions in [5] and

[12]. The ideal decision rule would be one with uniformly minimum risk for all θ in Θ. However, since such an ideal decision rule does not exist in general, other criteria for ranking decision rules must be considered.

Complete, Minimal Complete, and Admissible Classes

To aid in the explanation of procedures for ranking decision rules the following definitions are made:

A decision rule δ_1 is said to be "as good as" a rule δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all θ in Θ.

A decision rule δ_1 is said to be "better than" a rule δ_2 if $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ for all θ in Θ, and $R(\theta, \delta_1) < R(\theta, \delta_2)$ for at least one θ in Θ.

A decision rule δ_1 is said to be "risk equivalent" to δ_2 if $R(\theta, \delta_1) = R(\theta, \delta_2)$ for all θ in Θ.

One method of partially ranking decisions rules is to separate them into two sets -- one being a set of rules anyone of which would be "acceptable" and the other being a set of "unacceptable" rules. One such set of acceptable rules is given by the definition of a "complete class" as follows:

Let D be the set of all decision rules and C a subset of D, then C is said to be a "complete class" if given any rule δ in D not in C there exists a rule δ_0 in C that is better than δ.

If the statistician could find a complete class then he would select a decision rule in C in preference to a rule not in C. It should be noted that a complete class always exists since D, the set of decision rules, itself is a complete class.

Rather than choosing a decision rule from just any complete class, the statistician would prefer to identify and select a rule from the smallest set or from the "minimal complete" class which is defined as:

> A class C of decision rules is said to be "minimal complete" if C is complete and if no proper subclass of C is complete.

For some problems every complete class if infinite and a minimal complete class will not exist.

Another important concept is the one of "admissibility":

> A decision rule δ is said to be "admissible" if there exists no rule better than δ. The set of all admissible rules is called an "admissible class" of decision rules.

Admissible rules do not always exist, but if they do exist and can be identified, then the class of all admissible rules is a set of acceptable rules from which the statistician can choose. The class of all admissible rules, A, and a complete class of rules, C, are related in that $A \subset C$. In addition, if a minimal complete class exists, it consists of exactly the admissible rules (Theorem 1, pg. 56 of Ferguson [5]).

The Bayes Principle

The complete and admissible concepts are methods of identifying sets of acceptable decision rules from which the statistician can choose. Unfortunately these concepts do not produce a full ranking (i.e., a linear ordering) of rules within the complete or admissible classes to provide the statistician with definite advice about which decision rule to select. The Bayes principle, which does produce such a full ranking among all decision rules, provides a criterion by which the statistician can identify an **optimal** rule. The Bayes principle introduces the notion of a distribution called a "prior distribution," on Θ, the parameter space. The Bayes risk, $r(G, \delta)$, of a rule δ with respect to the prior distribution, $G(\theta)$, is then defined as:

$$r(G, \delta) = E_{\theta}[R(\theta, \delta)] = \int_{\Theta} R(\theta, \delta) dG(\theta) .$$

The prior distribution need not be interpreted as the distribution of the random variable θ for the Bayes principle to be useful. An alternate and in some instances more appealing interpretation of the prior distribution is that it is just a normalized risk weighting function (which, in some cases, may be derived from information known about θ prior to the experiment). Regardless of the interpretation for the prior distribution the Bayes risk is used to rank competing decision rules. A "Bayes rule" is defined as:

> A decision rule δ^* is said to be a "Bayes rule" with repect to the prior distribution $G(\theta)$ if
>
> $$r(G, \delta^*) = \inf_{\delta} r(G, \delta) .$$

If more than one rule is Bayes, then, any admissible Bayes rule is selected (if a Bayes rule is unique it will necessarily be admissible). In some instances the Bayes rule may not exist, in which case the statistician might select an "ε-Bayes rule" which is defined as:

> A decision rule δ^* is said to be an ε-Bayes rule with respect to the prior distribution $G(\theta)$ if for $\varepsilon > 0$
>
> $$r(G, \delta^*) \leq \inf_{\delta} r(G, \delta) + \varepsilon .$$

The statistician might prefer a stronger rule (if it exists) such as an "extended Bayes rule" which is:

> If δ^* is an ε-Bayes rule for every $\varepsilon > 0$ then δ^* is said to be "extended Bayes."

The Minimax Principle

Another method for ranking decision rules is the minimax principle which uses the quantity $\sup_\theta R(\theta, \delta)$ as a figure of merit. A "minimax rule" is defined as follows:

> A decision rule δ^* is said to be a "minimax rule" if
>
> $$\sup_\theta R(\theta, \delta^*) = \inf_\delta \sup_\theta R(\theta, \delta) .$$

The derivation of the name "minimax" can be seen if min and max are substituted for inf and sup in the above definition. A minimax rule may not be unique. In fact, if the supremum of the risk function is infinite for all decision rules, then every rule is by definition a minimax rule. Just as a Bayes rule may not exist a minimax rule may not exist for a particular problem, leading to the definition of ε-minimax and extended minimax rules similar in spirit to the definitions of ε-Bayes and extended Bayes rules. In cases of non-uniqueness any admissible minimax rule is selected. A unique minimax rule is necessarily admissible.

Relationship Between Bayes and Minimax Rules

In many problems minimax rules are related to Bayes rules in that a minimax rule is also a Bayes rule for some prior distribution. The prior distribution for which a minimax rule is also a Bayes rule is called a "least favorable distribution." For many other problems minimax rules are also extended Bayes rules. The condition for minimax rules to be either Bayes or extended Bayes rules is that

$$\sup_{\tau} \inf_{\delta} r(\tau, \delta) = \inf_{\delta} \sup_{\tau} r(\tau, \delta)$$

as given in pg. 57 of [5], (the minimax theorem). The underlying conditions for the minimax theorem to hold were derived by Wald in [20] and these conditions are satisfied by a large number of practical decision theory problems.

An extremely useful theorem (see [20] for a formal proof) providing a method for finding minimax rules, for many statistical decision theory problems is the following

Theorem: A Bayes rule or extended Bayes rule δ^* with constant risk, $R(\theta, \delta^*) = k$, is a minimax rule.

We have one final definition:

Any Bayes rule with constant risk is called an "equalizer rule."

A Duality Property for Bayes Rules

The above text discussed the Bayes principle without any mention of the criticism which has been voiced to this approach. The principle criticism centers on the prior distribution which the critics claim is based on subjective judgements. It is said that the Bayes principle is not as objective as classical procedures. Although more serious doubts about the objectivity of some of the classical procedures have been expressed, the taint of non-objectivity still hinders the full acceptance of the Bayes principle as a useful statistical tool. Therefore, any results which provide some "objective" justification for the Bayes principle would be important additions to this area of statistics. One application of the duality property for Bayes rules, as discussed in the following paragraphs, provides one such "objective" justification for the Bayes principle.

The relation that many minimax rules are also the solution for a Bayes or an extended Bayes problem with a least favorable prior distribution has been discussed above. This paper reports on some research on an investigation of the inverse to the above relationship. That is: is a Bayes rule (or extended Bayes rule) also the solution of a related minimax problem? This research, reported in Chapter II, shows that indeed a Bayes rule or extended Bayes rule (with modest restrictions on its risk function) for a given loss function, sample distribution and prior distribution is also the solution to a minimax problem with the same sample distribution, but with a different (but related) loss function.

The related minimax problem for which the Bayes rule is also a solution is called the "dual problem" in the following discussion. The result that Bayes rules are solutions to both an original Bayes problem and a dual minimax problem is similar to the duality principle that has proven so useful in linear programming. Therefore, this dual problem result for Bayes rules is called a "duality property." One of the most important results of the duality property is that it provides an additional justification for the Bayes principle. As can be seen in the next chapter, the dual and original problems are fundamentally different problems and not simply restatements of the same problem. Therefore, a Bayes rule is a solution of a different problem, which does not depend upon a prior distribution, and which may be as important as the original problem. Thus, the duality property is an additional and "objective" justification for application of the Bayes principle which is an important result for the reasons discussed earlier.

Other results of the duality property can be obtained through interpretation and application of the dual problem as is done in Chapter III. The natural application of the duality property is as a procedure for finding minimax rules. However, the most useful application of the duality property might well result from the fact that the Bayes rule is an equalizer rule for the dual minimax problem. One interpretation of this result is that the new loss function for the dual problem is a "least favorable loss function" that plays an analogous role to a least favorable distribution. The concept of a least favorable loss function, which is thought to be new, has the potential of being a valuable tool in engineering design problems where prior information is available as is illustrated in Chapter III. A collection of least favorable loss functions and their associated minimax rules for several well known problems are included in Chapter IV.

CHAPTER II

PROOF AND STRUCTURE OF THE DUALITY PROPERTY

The first chapter introduced the concept of a duality property for Bayes rules. This chapter defines this duality property more completely and proves a theorem containing a set of sufficient conditions for a rule to possess the duality property. An investigation of the structure of the dual minimax problem is also made in this chapter.

Definition of the Duality Property

The duality property for Bayes rules of this paper is defined in terms of decision theory problems, Bayes problems and minimax problems. Therefore, a shorthand notation for these problems will be introduced first to facilitate further discussion. A "decision problem" will be designated by the symbols D(L, F) and is defined as:

> A decision problem "D(L, F)" is to select a decision rule δ^* when the loss function is $L(\theta, \delta)$ with θ in Θ and the sample distribution is $F(x|\theta)$ with x in $\mathcal{X}$.

A "Bayes problem" will be designated by the symbols B(L, F, G) and is defined as

> A Bayes problem "B(L, F, G)" is a decision problem D(L, F) in which an optimum rule δ^* is any rule such that
>
> $$r(G, \delta^*) = \inf_{\delta} r(G, \delta)$$
>
> for the designated prior distribution $G(\theta)$.

An "extended Bayes problem" will be designated by the symbols EB(L, F, G) and is defined as:

> An extended Bayes problem "EB(L, F, G)" is a decision problem D(L, F) in which an optimum rule δ^* is an extended Bayes rule.

A minimax problem is designated by the symbol M(L, F) and is defined as:

> A minimax problem "M(L, F)" is a decision problem D(L, F) in which an optimum rule is any rule for which
>
> $$\sup_{\theta} R(\theta, \delta^*) = \inf_{\delta} \sup_{\theta} R(\theta, \delta) .$$

In any discussion about two or more of these problems we assume Θ and $\mathcal{A}$ are the same and hence have not been included in the notation. (A more general notation, for example, would have D(Θ, $\mathcal{A}$, L, F) instead of D(L, F).)

The duality property of this paper is defined to be:

> Let δ^* be a Bayes rule for the problem B(L, F, G) or an extended Bayes rule for the problem EB(L, F, G). Then δ^* is said to possess the "duality property" if δ^* is a minimax rule for the minimax problem M(L*, F) where L* is some loss function.

In the following discussion the Bayes or extended Bayes problem of this definition will be referred to as the "original problem" while the minimax problem of the definition will be called the "dual problem." Other useful duality properties might also be defined but the above definition will be used exclusively in the following work. The duality property theorem in the next paragraph gives a set of sufficient conditions for Bayes or extended Bayes rule to possess the duality property.

Duality Property Theorem

Theorem: Let δ^* be a Bayes rule [or an extended Bayes rule] for the Bayes problem B(L, F, G) [or for the extended Bayes problem EB(L, F, G)]. Then δ^* possesses the duality property, that is δ^* is a solution of the problem M(L*, F), if

1. $\mathcal{X}$ and Θ are the real line;
2. $R(\theta, \delta^*) > 0$ for all θ in Θ;
3. $R(\theta, \delta^*)$ is bounded on every interval of Θ;
4. $r(G, \delta^*)$ exists and is finite;
5. $L^*(\theta, \delta) = L(\theta, \delta)/R(\theta, \delta^*)$.

Proof: First it is proven that $H(\theta)$, where

$$H(\theta) = \int_{-\infty}^{\theta} \frac{R(t, \delta^*)}{r(G, \delta^*)}\, dG(t)$$

is a distribution function. From the definitions of $r(G, \delta^*)$, $G(\theta)$ and the fact that $R(t, \delta^*)$ is strictly positive, it is easy to see that (i) $\lim_{\theta \to -\infty} H(\theta) = 0$, (ii) $\lim_{\theta \to \infty} H(\theta) = 1$, and (iii) $H(a) - H(b) > 0$ for all $a > b$ in Θ.

To complete the proof that $H(\theta)$ is a distribution function it is necessary to show (iv) that $H(\theta)$ is continuous on the right **at** every point θ_0 in Θ, or that for all $h > 0$

$$\lim_{h \to 0} [H(\theta_0 + h) - H(\theta_0)] = 0 ,$$

or

$$\lim_{h \to 0} \left[\int_{\theta_0}^{\theta_0+h} \frac{R(t, \delta^*)}{r(G, \delta^*)} \, dG(t) \right] = 0 ,$$

and since $R(t, \delta^*)$ is bounded in the interval (θ_0, θ_0+h) where M is the upper bound

$$\lim_{h \to 0} \left[\frac{M}{r(G, \delta^*)} \int_{\theta_0}^{\theta_0+h} dG(t) \right] = 0 ,$$

or

$$\lim_{h \to 0} \left[\frac{M}{r(G, \delta^*)} \Big(G(\theta_0+h) - G(\theta_0) \Big) \right] = 0 ,$$

and since $G(t)$ is a distribution function this is a valid expression. Therefore, $H(\theta)$ is continuous on the right and is a distribution function.

The next step is to show that if δ^* is a Bayes rule for the original problem $B(L, F, G)$ then δ^* is a **Bayes** rule for the problem $B\big(L/R(\theta, \delta^*), F, H\big)$, where H is defined above. The Bayes risk for the problem $B\big(L/R(\theta, \delta^*), F, H\big)$ is $r^*(H, \delta)$ where

$$r^*(H, \delta) = \int_\Theta \int_X \frac{L(\theta, \delta)}{R(\theta, \delta^*)} \, dF(x|\theta) \, dH(\theta)$$

$$= \int_\Theta \int_X \frac{L(\theta, \delta)}{R(\theta, \delta^*)} \, dF(x|\theta) \, \frac{R(\theta, \delta^*)}{r(\theta, \delta^*)} \, dG(\theta)$$

$$= \frac{1}{r(G, \delta^*)} \int_\Theta \int_X L(\theta, \delta) \, dF(x|\theta) \, dG(\theta)$$

$$= \frac{1}{r(G, \delta^*)} \int_\Theta R(\theta, \delta) \, dG(\theta)$$

or

$$r^*(H, \delta) = \frac{r(G, \delta)}{r(G, \delta^*)}$$

and we want to choose δ to minimize $r^*(H, \delta)$. It should be recognized that $r(G, \delta^*)$ in the expression for $r^*(H, \delta)$ is a constant and that since δ^* minimizes $r(G, \delta)$ it must also minimize $r^*(H, \delta)$. Therefore,

$$r^*(H, \delta^*) = \inf_{\delta} r^*(H, \delta)$$

which is the desired result.

The final step in the proof is to show that δ^* yields constant risk for the dual problem $M\big(L/R(\theta, \delta^*), F\big)$. The risk for the rule δ^*, $R^*(\theta, \delta^*)$, in the dual problem is

$$R^*(\theta, \delta^*) = \int_X \frac{L(\theta, \delta^*)}{R(\theta, \delta^*)} \, dF(x|\theta)$$

$$= \frac{R(\theta, \delta^*)}{R(\theta, \delta^*)} = 1$$

which is the result needed.

In summary, the Bayes rule, δ^*, for the original problem $B(L, F, G)$ is an equalizer Bayes rule for the problem $B\big(L/R(\theta, \delta^*), F, H\big)$ and is therefore a minimax rule for the dual problem $M\big(L/R(\theta, \delta^*), F\big)$ as required to complete the proof of this part of the theorem.

Now suppose δ^* is an extended Bayes rule. Then for $\varepsilon_0 > 0$ there exists a prior distribution $G_0(\theta)$ such that

$$r(G_0, \delta^*) \leq \inf_{\delta} r(G_0, \delta) + \varepsilon_0 \, .$$

As before, $r(G_0, \delta^*)$ is a known constant. Therefore, for every $\varepsilon > 0$ there exists an $\varepsilon_1 > 0$ such that $\varepsilon = \varepsilon_1/r(G_0, \delta^*)$ and

$$r(G_0, \delta^*) \le \inf_\delta r(G_0, \delta) + \varepsilon_1$$

since δ^* is an extended Bayes rule. Now consider

$$H_0(\theta) = \int_{-\infty}^{\theta} \frac{R(\theta, \delta^*)}{r(G_0, \delta^*)} \, dG_0(\theta)$$

and the problem $EB\big(L/R(\theta, \delta^*), F, H_0\big)$ with Bayes risk

$$r^*(H_0, \delta) = \frac{r(G_0, \delta)}{r(G_0, \delta^*)} \quad .$$

Therefore,

$$\inf_\delta r^*(H_0, \delta) + \varepsilon = \frac{\inf_\delta r(G_0, \delta)}{r(G_0, \delta^*)} + \varepsilon$$

but from above

$$\inf_\delta r(G_0, \delta) \ge r(G_0, \delta^*) - \varepsilon_1$$

and $\varepsilon = \varepsilon_1/r(G_0, \delta^*)$ so

$$\begin{aligned} \inf_\delta r^*(H_0, \delta) + \varepsilon &\ge \frac{r(G_0, \delta^*) - \varepsilon_1}{r(G_0, \delta^*)} + \frac{\varepsilon_1}{r(G_0, \delta^*)} \\ &\ge \frac{r(G_0, \delta^*)}{r(G_0, \delta^*)} \end{aligned}$$

and this means that

$$\inf_\delta r^*(H_0, \delta) + \varepsilon \ge r^*(H_0, \delta^*) \quad .$$

Therefore, δ^* is an extended Bayes rule as well as an equalizer rule for the problem, $EB\big(L/R(\theta, \delta^*), F, H_0\big)$ and is consequently a minimax rule for the dual problem $M(L/R(\theta, \delta^*), F)$. This result completes the proof of the theorem.

Structure of the Dual Problem

The dual minimax problem was constructed from the original Bayes problem. The structure of the dual problem will be examined in more detail with respect to the structure of the original problem. In particular, it is shown in the following that if A is the class of all admissible rules for the original problem then A is also the class of all admissible rules for the dual problem. In a similar way it is shown that if C is a complete class of decision rules for the original problem then C is also a complete class for the dual problem. Since the two problems are equivalent in these important properties, it is natural to inquire if the application of the Bayes principle in the original problem and the minimax principle in the dual problem produce the same ranking of decision rules. If the ranking were the same, it could then be said that for all practical purposes the dual problem is simply a restatement of the original problem. The answer to this ranking inquiry is given in the following text by an example involving two rules δ_1 and δ_2 in which δ_1 is the preferred rule in the original Bayes problem and δ_2 is the preferred rule in the dual minimax problem. In other words, the dual problem, although similar to the original problem in several important properties, is not simply a restatement of the original Bayes problem.

Consider first a complete class structure theorem for the dual problem as follows:

<u>Theorem</u>: Let C_1 be any complete class for the original problem , then C_1 is a complete class for the dual problem. Let C_2 be any complete class for the dual problem, then C_2 is a complete class for the original problem.

Proof: The proof of both parts to the theorem is by contradiction. Suppose C_1 is any complete class of decision rules for the original problem $D(L, F)$. Suppose C_1 is not a complete class for the dual problem $D(L/R(\theta, \delta^*), F)$. This means there exists a δ_1, not in C_1, such that if $R^*(\theta, \delta)$ is the risk function for the dual problem then

$$R^*(\theta, \delta_1) \leq R^*(\theta, \delta')$$

for some δ' in C_1 and all θ in Θ, and

$$R^*(\theta', \delta_1) < R^*(\theta', \delta')$$

for some δ' in C_1 and θ' in Θ. But these inequalities can be expressed in terms of the risk for the original problem as

$$R(\theta, \delta_1)/R(\theta, \delta^*) \leq R(\theta, \delta')/R(\theta, \delta^*)$$

and

$$R(\theta', \delta_1)/R(\theta', \delta^*) < R(\theta', \delta')/R(\theta', \delta^*)$$

and multiplying each side of both inequalities by $R(\theta', \delta^*)$ yields

$$R(\theta, \delta_1) \leq R(\theta, \delta')$$
$$R(\theta', \delta_1) < R(\theta', \delta')$$

which says that δ_1 is also better than δ' for the original problem. But this contradicts the assumption that C_1 was a complete class. Therefore, if C_1 is any complete class for original problem it is also a complete class for the dual problem. In a similar way (by retracing the above steps) it can be shown that if C_2 is any complete class for the dual problem $D(L/R(\theta, \delta^*), F)$ then it is also a complete class for the original problem

$D(L, F)$. This then completes the proof of the theorem.

The next consideration is an admissibility theorem that shows that the original and dual problems have the same class of admissible rules.

Theorem: Let A_1 be the class of admissible rules for the original problem $D(L, F)$ and A_2 be the class of admissible rules for the dual problem $D\big(L/R(\theta, \delta^*), F\big)$ then $A_1 = A_2$.

Proof: First, it is shown that A_1 is a subclass of A_2, then that A_2 is a subclass of A_1 and as a consequence of these two relationships $A_1 = A_2$. The proof is by contradiction and is similar to the proof of the complete class theorem above. Suppose A_1 is not a subclass of A_2. This means that for any δ' in A_1 there exists a δ_1 and θ' such that if $R^*(\theta, \delta)$ is the risk for the dual problem then

$$R^*(\theta, \delta_1) \leq R^*(\theta, \delta')$$

for all θ in Θ, and

$$R^*(\theta', \delta_1) < R^*(\theta', \delta')$$

which (when operated upon as in the proof of the above complete class theorem) leads to

$$R(\theta, \delta_1) \leq R(\theta, \delta')$$

for all θ in Θ, and

$$R(\theta', \delta_1) < R(\theta', \delta')$$

which implies that δ_1 is better than δ' for the original problem and this contradicts the assumption that A_1 is the class of admissible rules for the original problem.

$$\delta^* = \frac{x + \frac{1}{2}}{\left(\frac{1}{2} + \frac{1}{2} + 4\right)} = \frac{1}{10}\,(2x + 1)$$

and the risk is

$$R(\theta, \delta^*) = \frac{1}{100}\,[1 + 12\,\theta(1 - \theta)]$$

now consider the rules (estimators)

$$\delta_1 = \frac{x+1}{6} \quad \text{and} \quad \delta_2 = \frac{x}{4}$$

which yield risks and Bayes risks of

$$R(\theta, \delta_1) = \frac{1}{36} \quad , \quad r(G, \delta_1) = \frac{1}{36} \quad ,$$

$$R(\theta, \delta_2) = \frac{1}{4}\,\theta(1 - \theta) \quad , \quad r(G, \delta_2) = \frac{1}{32} \quad .$$

Therefore, δ_1 is preferred to δ_2 in the original Bayes problem because δ_1 has the smaller Bayes risk.

For the dual minimax problem the risk function becomes

$$R^*(\theta, \delta_1) = \frac{1}{36}\left(\frac{100}{1 + 12\,\theta(1-\theta)}\right)$$

with

$$\max_{\theta}\left\{R^*(\theta, \delta_1)\right\} = \frac{100}{36}$$

and

$$R^*(\theta, \delta_2) = \frac{1}{4}\,\theta(1 - \theta)\left(\frac{100}{1 + 12\,\theta(1-\theta)}\right)$$

now to determine $\max_{\theta} R^*(\theta, \delta_2)$ let $x = \theta(1 - \theta)$; which means $0 \leq x \leq \frac{1}{4}$; then

$$R^*(\theta,\ \delta_2) = \frac{100}{4}\left(\frac{x}{1 + 12x}\right)$$

$$\frac{dR^*(\theta,\ \delta_2)}{dx} = \frac{100}{4}\left(\frac{1 + 12x - 12x}{(1 + 12x)^2}\right) = \frac{100}{4}\left(\frac{1}{(1 + 12x)^2}\right)$$

which is not 0 for any x such that $0 < x < \frac{1}{4}$. Therefore the maximum of $R^*(\theta,\ \delta_2)$ must occur on a boundary point and since $R^*(0,\ \delta_2) = 0$, the maximum must occur when $x = \frac{1}{4}$ or $\theta = \frac{1}{2}$ which means that

$$\max_{\theta} R^*(\theta,\ \delta_2) = \frac{100}{4(1 + 12/4)^2} = \frac{100}{4 \cdot 16} = \frac{100}{64}$$

and

$$\max_{\theta} R^*(\theta,\ \delta_1) > \max_{\theta} R^*(\theta,\ \delta_2)$$

and δ_2 is preferred to δ_1 in the dual problem of this example.

The conclusion can be drawn that although the two problems are equivalent with respect to admissibility and complete class concepts as well as having a common optimum solution, they are not totally equivalent in a practical sense because they do not necessarily provide the same ranking of rules as was shown in the above example.

CHAPTER III

INTERPRETATION AND APPLICATION OF THE DUALITY PROPERTY

The duality property for Bayes rules is similar to the duality principle of linear programming in many respects (see [15] for an explanation of linear programming). The duality principle plays an important role in linear programming problems, for example, it reduces the computational requirements to solve linear programming problems through application of the primal-dual algorithm. In addition, in some problems the dual problem has been interpreted to be as significant as the primal problem. As an example, in economics problems the primary variables might be interpreted as unit costs with the objective of the primal problem being to minimize total costs. The dual problem in this instance has unit prices as variables with the objective of maximizing total profit.

If a similar important interpretation of the duality property for Bayes rules could be made, then the significance of this property would be greatly enhanced. Unfortunately, research has revealed no simple universally recognizable interpretation of the duality property similar to the cost-price duality relationship of linear programming. However, several interpretations and applications have been made and are recorded in the following text.

First, and most importantly, the duality property provides an "objective" justification for Bayes rules. Applications of the duality property to problems with indefinite loss functions are presented. A new concept called "least favorable loss function" which can be used to introduce prior information into minimax problems is developed. The "least favorable loss function" concept appears to be different from the G-minimax idea which also considers prior information. Finally, the application of the duality property as a technique for finding minimax rules is illustrated by several examples.

An "Objective" Justification for Bayes Rules

The most straight forward and important interpretation of the duality property is to view it simply as an additional property of Bayes rules. The fact that many Bayes (or extended Bayes) rules also solve a dual problem provides an additional and "objective" justification for using the Bayes rules. This justification is important because Bayes rules have not gained universal acceptance due to the criticism that they are not "objective." The argument is that the prior distribution is based upon subjective judgements and for this reason Bayes rules are not "objective." This type of argument is very common and we need not document it extensively here. We will just quote Kempthorne who in [9] made the following statement:

> "It would be wonderful (perhaps) if one could get a Bayesian type of answer without the total arbitrariness of the Bayesian arguments."

The duality property provides an answer of sorts to this request because many Bayes rules are also solutions to dual minimax problems which do not depend upon an "arbitrary" prior distribution.

The Bayes rule of the original problem is a minimax rule for the dual problem which has a modified loss function. A minimax rule can be thought of as protecting against the worst case. A worst case design criteria is popular in engineering, therefore, this justification for Bayes rules may appeal to engineers.

In addition to being a minimax rule for the dual problem the original Bayes rule is also an equalizer rule for the dual problem. The appeal of an equalizer rule is that it protects against error equally for all values of θ. This equal protection in the dual problem with a modified loss function should also appeal to engineers because it also represents a worst case viewpoint together with the equal protection property.

Another point that should not be overlooked is that if the original Bayes rule is admissible in the original problem then it is also admissible in the dual problem (as was proven in Chapter II). There is a wide set of conditions under which Bayes rules are admissible (see pages 60, 61, 62, and 69 of [5]). Therefore in many cases the original Bayes rule is likely to be an admissible rule for the dual problem.

From this discussion it can be seen that the original Bayes rule is an equalizer, minimax rule that in many instances is admissible for the dual problem. These are strong attributes for this decision rule and they enhance the "objective" justification provided by the duality property.

Applications to Problems with Indefinite Loss Functions

In many decision theory problems a great deal of work is expended in developing definite loss functions related only to that problem. In many other problems the loss functions are selected with much less thought and are therefore of an indefinite nature. Modifications to these indefinite loss functions should be viewed without alarm and accepted if they are interpretable and useful. The duality property suggests that the statistician, after solving for the Bayes (or extended Bayes) rule in a problem with an indefinite loss function, should also examine the dual problem to determine if it can be interpreted and explained to strengthen the case for the Bayes rule he has found.

There are three types of loss functions commonly used in estimation problems:

$$L_1(\theta, \delta) = (\theta-\delta)^2$$

$$L_2(\theta, \delta) = |\theta-\delta|$$

$$L_3(\theta, \delta) = 0 \quad \text{if } |\theta-\delta| \leq c$$

$$= 1 \quad \text{if } |\theta-\delta| > c$$

The following examples, each based on one of the above indefinite loss functions in the original problem, show that the modified loss function in the dual problem can be just as meaningful as the original loss function in some instances.

The first example is to find a Bayes estimator for the parameter, θ, in the Poisson distribution with a sample of size 1, loss function $L(\theta, \delta) = (\theta-\delta)^2$, and a gamma prior distribution with parameters (5, .1). The distribution functions are: x = value of the observation

$$F(x|\theta) = \sum_{i=0}^{x} \frac{e^{-\theta}\theta^{i}}{i!} \qquad \text{if } x = 0, 1, \cdots$$

$$G(\theta) = \int_0^{\theta} \frac{5^{.1}e^{-5t}t^{-.9}}{\Gamma(.1)}\,dt \quad \text{if } \theta > 0 \quad .$$

The posterior distribution in this case is

$$K(\theta|x) = \int_0^{\theta} \frac{6^{x+.1}t^{x-.9}}{(x+.1)}\,e^{-6t}dt \quad \text{if } \theta > 0$$

The Bayes estimator in this problem where the loss is $L(\theta, \delta) = (\theta-\delta)^2$ is the mean of $K(\theta|x)$ (see page 46 of [5]). Therefore, δ^* the Bayes rule

$$L^*(\theta,\ \delta) = \frac{36}{25}\,\frac{(\theta-\delta)^2}{\theta^2} \quad .$$

The constant multiplier $\frac{36}{25}$ in this loss function does not influence the choice of δ. Therefore, the dual problem for all practical purposes can be said to have the loss function

$$L^*(\theta,\ \delta) = \frac{(\theta-\delta)^2}{\theta^2} \quad .$$

This loss function is the square of the percentage loss which is another much used loss function. Therefore, $\delta^* = \frac{x+.1}{6}$ solves a dual minimax problem which is just as meaningful as the original Bayes problem for which it is also a solution.

Another problem that illustrates this application of the duality property is as follows. The problem is to find the Bayes estimator for θ in the uniform distribution with one observation when

$$F(x|\,\theta) = \int_0^x \frac{1}{\theta}\,dt \quad , \quad 0 < x < \theta$$

$$L(\theta,\ \delta) = |\,\theta-\delta\,|$$

$$G(\theta) = \int_0^\theta 100te^{-10t}dt$$

The Bayes rule, δ^*, for this problem is

$$\delta^* = x - .1\ \ln(.5) = x + .069315 \quad .$$

And, the risk for this rule is

$$R(\theta,\ \delta^*) = \frac{1}{2\theta}\,[(\theta - .069315)^2 + .005]$$

This means that the modified loss function for the dual problem, $L^*(\theta, \delta)$, is

$$L^*(\theta, \delta) = \frac{2\theta|\theta-\delta|}{(\theta-.069315)^2 + .005}$$

and that except for small θ's, $L^*(\theta, \delta)$ is approximately

$$L^*(\theta, \delta) \doteq \frac{2\theta|\theta-\delta|}{\theta^2} ,$$

or

$$L^*(\theta, \delta) \doteq \frac{2|\theta-\delta|}{\theta} .$$

And, $L^*(\theta, \delta)$ is approximately twice the percentage loss. This means that δ^* is a rule that for all practical purposes minimizes the maximum mean percentage loss. In addition, the percentage loss is just as meaningful a loss function as the absolute error loss function of the original problem.

The modified loss function in the next example is not immediately interpretable and its meaning is questionable. The problem is to find the Bayes estimator for θ, the mean of a normal distribution with unit variance, from one observation where

$$F(x|\theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta)^2} dt ,$$

$$\begin{aligned} L(\theta, \delta) &= 0 \quad \text{if } |\theta-\delta| \leq 1 \\ &= 1 \quad \text{if } |\theta-\delta| > 1 , \end{aligned}$$

$$G(\theta) = \int_{-\infty}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt .$$

The Bayes rule, δ^*, for this problem is

$$\delta^* = \frac{x}{2} \ .$$

The risk function for δ^* is

$$R(\theta, \delta^*) = 1 - \phi(\theta+2) + \phi(\theta-2)$$

where $\phi(\cdot)$ is the standard normal distribution in this instance the modified loss function, $L^*(\theta, \delta)$, becomes

$$\begin{aligned} L^*(\theta, \delta) &= 0 && \text{if } |\theta-\delta| \leq 1 \\ &= 1/[1 - \phi(\theta+2) + \phi(\theta-2)] && \text{if } |\theta-\delta| > 1 \ . \end{aligned}$$

It is not immediately apparent that this loss function has inherent meaning and the statistician would have to make an intensive numerical investigation of $L^*(\theta, \delta)$ to see if it was at all relevant to the problem at hand. This example has been included to show the form of the modified loss function when the original loss function is

$$\begin{aligned} L(\theta, \delta) &= 0 \quad \text{if } |\theta-\delta| \leq c \\ &= 1 \quad \text{if } |\theta-\delta| > c \ . \end{aligned}$$

In addition, the example illustrates the fact that additional work might be required to interpret the modified loss function.

It can be concluded from these examples that after a Bayes (or extended Bayes) problem has been solved then it is well worth the effort to examine the dual problem to see if this problem will enhance the value of the Bayes rule.

Least Favorable Loss Function

One interpretation of the duality property is that the loss function in the dual minimax problem plays a role similar to that of the least favorable distribution used in constructing many minimax rules. Therefore, the new loss function could be called a "least favorable loss function" as is done in the following discussion. This new concept could be applied in the following way. Consider a problem in which there were some control over the selection of the loss function or some uncertainty as to the appropriate loss function and a constant risk rule is required. The duality property allows the selection of both a least favorable loss function and an associated minimax rule in parametric form. These parameters can be selected to introduce prior information into the problem.

One practical example where a least favorable loss function approach might be useful is in the design of electronic sensing equipment. Suppose the objective of the equipment is to provide an image of the radiation (infrared, radiometric, ultraviolet, etc.) in the sensors field of view. The radiation is converted by a linear transformation to an electronic signal, x, by a detector and preamplifier. Noise in the atmosphere, detector, and amplifier are superimposed on the true signal, θ. Therefore, x is a random variable. Suppose the distribution function for x is normal with mean θ and variance 1, that is

$$F(x|\theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t-\theta)^2} dt \quad .$$

The problem now is to determine an estimate δ (a function of the measurement x) of the true signal θ. In this instance, there is considerable experimental information available on targets of interest such as trucks, artillery, etc. Suppose this prior information indicates that most of the targets of interest have values of θ between 10 and 30 with the peak at 20. This prior information can be introduced into the problem by the use of a least favorable loss function $L(\theta, \delta) = \lambda(\theta)(\theta-\delta)^2$ where $\lambda(\theta)$ is a weighting function to be based upon the prior information. A least favorable loss function for this problem is:

$$L^*(\theta, \delta) = \frac{(1+\beta)^2}{\beta^2 + (\theta-\alpha)^2} (\theta-\delta)^2 \qquad \alpha > 0,\ \beta > 0\ .$$

The problem now is to determine α and β to reflect the prior information. The weighting function

$$\lambda(\theta) = \frac{(1+\beta)^2}{\beta^2 + (\theta-\alpha)^2}$$

is maximized when $\theta = \alpha$ for a fixed β. This means that in this example $\alpha = 20$. Then, since very few targets of interest will have $\theta < 10$, β can be selected such that

$$\frac{\lambda(20)}{\lambda(10)} = .2 \quad , \quad \text{or} \quad \frac{\beta^2}{\beta^2 + 100} = .2 \quad , \quad \text{or} \quad \beta = 5$$

which means that the weight given to signals outside the interval $10 \leq \theta \leq 30$ will be less than twenty percent of the weight given to the signal at $\theta = 20$. This means the system will be designed to be most sensitive in the region that the most important targets will appear, which is the desired result. The minimax estimate for the above least favorable loss function in this problem with $\alpha = 10$, $\beta = 5$, is

$$\delta^* = \frac{10 + 5x}{6}\ .$$

If the sensing equipment is designed to display δ^*, as defined above, then it can be said that the design is optimum in the sense that the maximum weighted mean square risk has been minimized with the weighting function reflecting available prior information. This example illustrates the use of the least favorable loss function concept in concrete engineering design terms.

Another application of this concept is in structuring the following reliability experiment in which the prior information is quite vague. The object of the experiment is to estimate the mean time between failure, θ, of a particular type of equipment. The manufacturer and buyer have agreed that a proper loss function would be of the form:

$$L(\theta, \delta) = \lambda(\theta)(\theta-\delta)^2 .$$

They also agree that they wish to select $\lambda(\theta)$ and the estimate δ to provide a constant risk which does not favor either side. In other words they are searching for a least favorable loss function and the related minimax rule. Further it is agreed that three units of the equipment would be tested to failure and the three times to failure x_1, x_2, x_3 would be used to form the estimator. The times to failure are assumed to be exponentially distributed with distribution function

$$F(x_i|\theta) = \int_0^{x_i} \frac{1}{\theta} e^{-\frac{t}{\theta}} dt; \; i = 1, 2, 3; \; x_i > 0$$

Here $x = \frac{\sum_{i=1}^{3} x_i}{3}$ is a sufficient statistic and therefore x is the statistic that will be used to form the estimator. Using the duality property with

$$L(\theta, \delta) = (\theta-\delta)^2 ,$$

$$F(x|\theta) = \int_0^x \left(\frac{1}{\theta}\right)^3 t^2 e^{-\frac{t}{\theta}}\, dt\ , \qquad x > 0$$

and using the inverted gamma distribution as a prior

$$G(\theta) = \int_0^\theta \frac{e^{-\frac{\alpha}{t}}\ \alpha^\beta \left(\frac{1}{t}\right)^{\beta+1}}{\Gamma(\beta)}\, dt\ , \qquad \theta > 0$$

the least favorable loss function is

$$L^*(\theta,\ \delta) = \frac{(\beta+2)^2}{\left[(\beta-1)^2 + 3\right]\theta^2 - 2\alpha(\beta-1)\theta + \alpha^2}\ (\theta-\delta)^2$$

and the associated minimax rule is

$$\delta^* = \frac{\alpha + x}{\beta + 2}\ .$$

Now $\lambda(\theta)$, where

$$\lambda(\theta) = \frac{(\beta+2)^2}{\left[(\beta-1)^2 + 3\right]\theta^2 - 2\alpha(\beta-1)\theta + \alpha^2}\ ,$$

is a modifier of the term $(\theta-\delta)^2$ in $L^*(\theta,\ \delta)$ which shapes $L^*(\theta,\ \delta)$ for different values of θ. Suppose now that both the manufacturer and buyer are agreed that the equipment probably has a θ on the order of 100 hours and that $\lambda(\theta)$ should peak at this value. Also there is little chance that θ will be 50 hours or less and it is agreed that $\lambda(\theta)$ for all $\theta < 50$ hours should be less than 1/3 of the value of $\lambda(\theta)$ for $\theta = 100$ hours. Also, if $\theta > 100$ hours both parties will be equally happy. Therefore, they are indifferent to the value of $\lambda(\theta)$ in the region in which $\theta > 100$ where $\lambda(\theta)$ can assume as small a value as needed to assure that the above two requirements are met. The problem now is to determine α and β so that the two conditions are met. A trial and error procedure was used to find that if $\alpha = 560$, $\beta = 6$ then the maximum of $\lambda(\theta)$ occurs at $\theta = 100$, and that

$$\frac{\lambda(\theta < 50)}{\lambda(\theta = 100)} \leq \frac{1}{3} .$$

These values mean that the least favorable loss function is

$$L^*(\theta, \delta) = \frac{36}{28\theta^2 - 5600\theta + 313{,}600} (\theta-\delta)^2$$

for which

$$\delta^* = \frac{560 + x}{8}$$

is the associated minimax rule.

A Procedure for Solving Minimax Problems

Some statisticians prefer to approach statistical problems from a minimax point of view. These statisticians might interpret the duality property as providing another tool with which to solve minimax problems. The following application illustrates the usefulness of the duality property in minimax problems.

The problem is to find minimax estimators for θ, the parameter of a Bernoulli distribution from a sample of size n; $(x_1, x_2, \cdots, x_n)$, $x_i = 0, 1$; for the following set of loss functions:

$$L(\theta, \delta) = \frac{(\theta-\delta)^2}{\theta^r (1-\theta)^s}$$

where $r = 0, 1$; $s = 0, 1$. Each of these loss functions are interpretable and have been used previously in minimax problems of this type. To apply the duality property, the first step is solve a Bayes problem with loss function

$$L(\theta, \delta) = (\theta-\delta)^2$$

and a Beta prior distribution with parameters (α, β). This loss function and prior were selected because it is known from previous work (page 91 of [5]) that the resultant risk function is of the form $a\theta^2 + b\theta + c$, with a, b, c depending on α and β. This means the modified loss function of the dual problem will be of the form

$$L^*(\theta, \delta) = \frac{(\theta-\delta)^2}{a\theta^2 + b\theta + c}$$

which has at least the potential (if α and β are selected correctly) of being one of the loss functions

$$L(\theta, \delta) = \frac{(\theta-\delta)^2}{\theta^r(1-\theta)^s} \qquad r = 0, 1;\ s = 0, 1 .$$

And these are the loss functions of the minimax problem outlined above. For this problem $x = \sum_{i=1}^{n} x_i$ is a sufficient statistic and will therefore be used as the sample value. This means the elements of the Bayes problem are

$$F(x|\theta) = \sum_{i=0}^{[x]} \binom{n}{i} \theta^i(1-\theta)^{n-i} \quad , \quad 0 \le x \le n \quad ,$$

$$G(\theta) = \int_0^\theta \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}\, dt \quad , \quad 0 \le \theta \le 1 \quad ,$$

and

$$L(\theta, \delta) = (\theta-\delta)^2 \quad .$$

The Bayes rule (estimator) for this problem is

$$\delta^* = \frac{\alpha+x}{\alpha+\beta+n}$$

and the risk for δ^* is

$$R(\theta,\ \delta^*) = \frac{[(\alpha+\beta)^2 - n]\theta^2 + [n - 2(\alpha+\beta)]\theta + \alpha^2}{(n+\alpha+\beta)^2} \ .$$

By the duality theorem $\delta^* = \frac{\alpha+x}{\alpha+\beta+n}$ is also a minimax estimator for θ when the loss function is

$$L^*(\theta,\ \delta) = \frac{(\theta-\delta)^2}{R(\theta,\ \delta^*)} = \frac{(n+\alpha+\beta)^2(\theta-\delta)^2}{[(\alpha+\beta)^2 - n]\theta^2 + [n - 2\alpha(\alpha+\beta)]\theta + \alpha^2} \ .$$

If α and β could take on the value of 0 then the minimax problem could be solved as follows:

(1) If $\alpha = \beta = \frac{\sqrt{n}}{2}$

then $L^*(\theta,\ \delta) = \frac{4(n + \sqrt{n})^2}{n}(\theta-\delta)^2$

which means $r = 0$, $s = 0$

and the minimax rule is $\delta^* = \frac{\frac{\sqrt{n}}{2} + x}{\sqrt{n} + n}$.

(2) If $\alpha = 0$, $\beta = \sqrt{n}$

then $L^*(\theta,\ \delta) = \frac{(n + \sqrt{n})^2}{n}\ \frac{(\theta-\delta)^2}{\theta}$

which means $r = 1$, $s = 0$

and the minimax rule would be $\delta' = \frac{x}{\sqrt{n} + n}$.

(3) If $\alpha = \sqrt{n}$, $\beta = 0$

then $L^*(\theta, \delta) = \frac{(n + \sqrt{n})^2}{n} \frac{(\theta-\delta)^2}{1-\theta}$

which means $r = 0$, $s = 1$

and the minimax rule would be $\delta' = \frac{x + \sqrt{n}}{n + \sqrt{n}}$.

(4) If $\alpha = 0$, $\beta = 0$

then $L^*(\theta, \delta) = \frac{n(\theta-\delta)^2}{\theta(1-\theta)}$

which means $r = 1$, $s = 1$

and the minimax rule would be $\delta' = \frac{x}{n}$.

But $\alpha > 0$, and $\beta > 0$ therefore the above results in (2), (3) and (4) are not Bayes rules. However, these rules can be shown to be extended Bayes rules and are for this reason minimax rules for the dual problems. Thus, (1) (2), (3) and (4) are solutions to the minimax problem posed at the first of this section.

To illustrate the proof that these rules are extended Bayes rules it is shown that $\frac{x}{n}$ [from (4)] is an extended Bayes rule in the following text.

The Bayes risk for the Bayes rule, δ^*, is

$$r(B, \delta^*) = \frac{\alpha}{(n+\alpha+\beta)^2(\alpha+\beta)} \left\{ \left[\frac{(\alpha+\beta)^2 - n}{\alpha+\beta+1} \right] (\alpha+1) + n - \alpha(\alpha+\beta) \right\} .$$

The Bayes risk for the rule $\delta' = \frac{x}{n}$ is

$$r(B, \delta') = \frac{1}{n} \left[\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)} \right] .$$

Now let $\alpha = \beta$. Then

$$r(B, \delta^*) = \frac{\alpha}{2(n+2\alpha)(2\alpha+1)}$$

and

$$r(B, \delta') = \frac{\alpha}{2n(2\alpha+1)}$$

Let $\varepsilon > 0$. An $\alpha > 0$ is sought such that

$$r(B, \delta') \leq r(B, \delta^*) + \varepsilon$$

or

$$\frac{\alpha}{2n(2\alpha+1)} \leq \frac{\alpha}{2(n+2\alpha)(2\alpha+1)} + \varepsilon$$

or

$$\frac{\alpha}{2(2\alpha+1)}\left[\frac{1}{n} - \frac{1}{n+2\alpha}\right] \leq \varepsilon$$

let $\alpha_0 = \varepsilon$. Then $\frac{1}{2(2\alpha+1)}$ will always be a positive fraction as is $\left[\frac{1}{n} - \frac{1}{n+2\alpha}\right]$ and therefore

$$\frac{\alpha_0}{2(2\alpha_0+1)}\left[\frac{1}{n} - \frac{1}{n+2\alpha_0}\right] \leq \varepsilon$$

as required and δ' is an extended Bayes rule.

This example illustrates the use of the duality property to find minimax rules. Other minimax problems can be attacked in exactly the same way. One bonus that may accompany this technique is that many Bayes rules are also admissible and as is seen in Chapter II these rules are also automatically admissible minimax rules in the dual problem.

These interpretations and applications of the duality property have been presented here for two reasons. First, to indicate that there are reasonably important problems for which the duality property yields useful results. Second, the derivations and results are intended to provide a clarification of the duality property theorem of Chapter II. The next chapter contains a collection of least favorable loss functions and their associated minimax rules in hopes they prove useful.

CHAPTER IV

SOME "LEAST FAVORABLE LOSS FUNCTIONS"

Chapter III introduced the concept of a least favorable loss function and discussed two examples illustrating their use. This chapter presents least favorable loss functions for several well known problems for reference and to illustrate their forms. First, three least favorable loss functions will be discussed for the problem D(L, F) where $F(x|\theta)$ is the normal distribution with unknown mean and known variance. Least favorable loss functions for other problems are presented in a table with little comment.

First consider the problem D(L, F) where $F(x|\theta)$ is normal and σ^2 is known, that is

$$F(x|\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}}$$

and a sample of size n is taken. Let

$$u = \sum_{i=1}^{n} x_i/n$$

then

$$\delta^* = \frac{\frac{\sigma^2}{n}\alpha + \beta u}{\frac{\sigma^2}{n} + \beta} \quad , \quad \text{where } \beta > 0$$

is the minimax rule for all three of the following least favorable loss functions:

$$L_1^*(\theta,\ \delta) = \frac{k^2(\theta-\delta)^2}{(1+a^2)}$$

$$L_2^*(\theta,\ \delta) = \frac{k|\theta-\delta|}{\left[|a|(2\phi\{a\}-1) + 2\exp\left(-\frac{a^2}{2}\right)\right]}$$

$$L_3^*(\theta,\ \delta) = \frac{1}{1-\phi\left\{a+\frac{\sqrt{n}\,c}{\sigma}k\right\}+\phi\left\{a-\frac{\sqrt{n}\,c}{\sigma}k\right\}} \quad \text{if} \quad |\theta-\delta| > c$$

$$= 0 \quad \text{if} \quad |\theta-\delta| \leq c$$

where

$$a = (\theta-\alpha)\frac{\sigma}{\beta\sqrt{n}}$$

$$k = \left(1+\frac{n}{\sigma^2}\beta\right)/\beta$$

and

$\phi\{\cdot\}$ = cumulative standard normal distribution function.

These three least favorable loss functions are all for the same problem and therefore can be compared and the one most appropriate for the problem at hand can be used. Also these three loss functions (each of which was derived from a different loss function L in an original Bayes problem B(L, F, G)) show the variation in least favorable loss functions caused by the choice of loss function in the original Bayes problem.

Of the three loss functions $L_1^*(\theta,\ \delta)$ is the simplest and is the easiest to work with. The original loss from which $L_1^*(\theta,\ \delta)$ was derived was $L(\theta,\ \delta) = (\theta-\delta)^2$. Table 1 contains least favorable loss functions and other information for several important problems when the original loss function is $L(\theta,\ \delta) = (\theta-\delta)^2$.

TABLE 1

Examples of Least Favorable Loss Functions

Sample Distribution	Least Favorable Loss Function* $[L^*(\theta, \delta)]$	Minimax Rule	Statistic (u)
Normal (variance unknown) $F(x\|\theta) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\frac{(t-\mu)^2}{\theta}} dt$	$\frac{\left(\beta + \frac{1}{2}n\right)^2 (\theta-\delta)^2}{\theta^2\left(\frac{1}{2}n + \beta^2\right) - 2\alpha\beta\theta + \alpha^2}$	$\frac{2\alpha+u}{2\beta+n}$	$u = \sum_{i=1}^{n} (x_i-\mu)^2$
Binomial $F(x\|\theta) = \sum_{i=0}^{[x]} \theta^i (1-\theta)^{1-i}$; $x = 0, 1$	$\frac{(n+\alpha+\beta)^2(\theta-\delta)^2}{\theta^2[(\alpha+\beta)^2 - n] - \theta[2\alpha(\alpha+\beta) - n] + \alpha^2}$	$\frac{\alpha+u}{\alpha+\beta+n}$	$u = \sum_{i=1}^{n} x_i$
Poisson $F(x\|\theta) = \sum_{i=0}^{[x]} \frac{e^{-\theta}\theta^i}{i!}$	$\frac{(n+\beta)^2(\theta-\delta)^2}{\theta^2\beta^2 - \theta[2\alpha\beta-n] + \alpha^2}$	$\frac{\alpha+u}{\beta+n}$	$u = \sum_{i=1}^{n} x_i$
Exponential $F(x\|\theta) = \int_{0}^{x} \frac{1}{\theta} e^{-\frac{t}{\theta}} dt$	$\frac{(n+\beta-1)^2(\theta-\delta)^2}{\theta^2[n + (\beta-1)^2] - \theta 2\alpha[\beta-1] + \alpha^2}$	$\frac{\alpha+u}{n+\beta-1}$	$u = \sum_{i=1}^{n} x_i$

$*\alpha \geq 0$, $\beta \geq 0$, n = sample size

The first column of the table lists the sample distribution for each problem considered. The least favorable loss function, as a function of two parameters, is listed in the second column. The minimax rule, which are also equalizer rules, for the problem M(F, L*) is also included in the table. The last column contains the definition of the statistic used in the minimax rule. Note the similarity of all of the least favorable losses and the related minimax rules.

These least favorable loss functions have been included here to encourage their use in the problems as outlined above. In addition, it is hoped that this entire discussion will stimulate the application of least favorable loss functions wherever appropriate.

LIST OF REFERENCES

1. Blackwell, D. and M. A. Girshick (1954). Theory of Games and Statistical Decisions. Wiley, New York, N. Y.

2. DeGroot, M. H. and M. M. Rao (1963). "Bayes estimation with convex loss," Annals of Mathematical Statistics, 34, 839-846.

3. Fabius, J. (1964). "Asymptotic behaviour of Bayes estimates," Annals of Mathematical Statistics, 35, 846-856.

4. Farrell, R. H. (1967). "Weak limits of sequences of Bayes procedures in estimation theory," Proceedings 5th Berkeley Symposium on Mathematical Statistics and Probability.

5. Ferguson (1967). Mathematical Statistics, A Decision Theoretic Approach. Academic Press, New York, N. Y.

6. Fisher, Sir Ronald (1962). "Some examples of Bayes method of experimental determination of probabilities apriori," Journal of Royal Statistical Society, B, 24, 118-124.

7. Girshick, M. A. and L. J. Savage (1951). "Bayes and minimax estimates for quadratic loss functions," Proceedings of the Second Berkely Symposium on Mathematical Statistics and Probability, University of California Press, 53-74.

8. Hodges, J. L. and E. L. Lehmann(1950). "Some problems in minimax point estimation," Annals of Mathematical Statistics, 21, 182-197.

9. Kempthorne, (1969). Discussion on "The Bayesian outlook and its application" by J. Cornfield. Handout for Iowa Meeting, April, 1969.

10. Kiefer, J. (1953). "On Wald's complete class theorem's," Annals of Mathematical Statistics, 24, 70-75.

11. Lehmann, E. L. (1947). "On families of admissible tests," Annals of Mathematical Statistics, 18, 97-104.

12. Raiffa, H. and R. Schlaifer (1961). Applied Statistical Decision Theory. Graduate School of Business, Harvard.

13. Savage, L. J. (1951). "The theory of statistical decision," Journal of American Statistical Association, 46, 55-67.

14. Savage, L. J. (1954). The Foundations of Statistics. Wiley, New York, N.Y.

15. Simonnard, M. (1968). Linear Programming. Prentice Hall.

16. Sprott, D. A. (1960). "Necessary restrictions for distributions a posteriori," Journal Royal Statistical Society, B, 22, 313-318.

17. von Neuman, J. and Oskar Morgenstern (1944). Theory of Games and Economic Behavior, 3rd edition. Princeton University Press, 1953, Princeton, N. J.

18. Wald, A. (1939). "Contributions to the theory of statistical estimation and testing hypothesis," Annals of Mathematical Statistics, 10, 299-326.

19. Wald, A. (1942). "On the principles of statistical inference," Notre Dame Math Lectures, No. 1. Notre Dame, Indiana.

20. Wald, A. (1945). "Statistical decision functions which minimizes the maximum risk," Annals of Mathematics, 46, 265-280.

21. Wald, A. (1946). "An essentially complete class of admissible decision functions," Annals of Mathematical Statistics, 18, 549-555.

22. Wald, A. and J. Wolfowitz (1951). "Characterization of the minimal complete class of decision functions when the number of distributions and decisions are finite," Proceedings of 2nd Berkely Symposium on Mathematical Statistics and Probability. University of California Press.

23. Wald, A. (1950). Statistical Decision Functions. Wiley, New York, N. Y.

24. Weiss, Lionel (1961). Statistical Decision Theory. McGraw Hill, New York, N. Y.

25. Williams, J. D. (1954). The Complete Strategyst. McGraw Hill, New York, N. Y.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) SOUTHERN METHODIST UNIVERSITY	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	
	2b. GROUP UNCLASSIFIED	
3. REPORT TITLE A Duality Property for Bayes Rules with Applications		
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report		
5. AUTHOR(S) (First name, middle initial, last name) Charles R. Johnson		
6. REPORT DATE June 26, 1970	7a. TOTAL NO. OF PAGES 47	7b. NO. OF REFS 25
8a. CONTRACT OR GRANT NO. N00014-68-A-0515	9a. ORIGINATOR'S REPORT NUMBER(S) 75	
b. PROJECT NO. NR 042-260		
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.		
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.		
11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Office of Naval Research	

13. ABSTRACT

Many minimax rules are also solutions to a related Bayes or extended Bayes problem. This paper investigates the inverse to the above result which is that many Bayes (or extended Bayes) rules are also solutions to a related minimax problem. This relationship has been given herein the name "duality property for Bayes rules".

Interpretations and applications of the duality property are discussed. For example the duality property provides an "objective" justification for Bayes rules in that the dual problem does not depend upon a prior distribution. This is important from a military viewpoint because some problems such as threat classification of enemy radars and final attack of enemy submarines can be formulated very easily within the Bayesian framework. Minimax rules for several problems are presented to illustrate the obvious application to solving minimax problems. Minimax rules are useful in making decisions on tactics in order to minimize the maximum expected loss for a limited war or counterinsurgency situation. The loss function for the dual minimax problem in parametric form is defined here as a "least favorable loss function" because it plays a similar role to the least favorable distribution. The least favorable loss function notion is applied to reliability test and reconnaissance sensor designs to illustrate how this concept can be used to introduce prior information into a minimax problem. Least favorable loss functions for several well known problems are included for reference.