

A Modification of Brown's
Technique for Proving Inadmissibility*

by

James O. Berger
Purdue University

Department of Statistics
Division of Mathematical Sciences
Mimeograph Series #79-33

December 1979

*This research was supported by the Alfred P. Sloan Foundation, by NSF Grant MCS 78-02300, and by NIH Grant R01-GM21215.

A Modification of Brown's Technique for Proving Inadmissibility*

James O. Berger

Purdue University and Stanford University

Summary

A technique is presented for proving that a decision rule is inadmissible. The technique is based on finding a better decision rule for large values of the parameter, and is hence similar to that discussed in Brown (1980). (See also Brown (1979).) The method can be useful when the parameter space is not closed, a situation in which it can be difficult to apply Brown (1980).

*This research was supported by the Alfred P. Sloan Foundation, by NSF Grant MCS 78-02300, and by NIH Grant R01-GM21215.

1. Introduction

We adopt the usual decision theoretic framework in which a random variable $X \in \mathcal{X}$, with associated probability measure F_θ , is observed. It is desired to make some decision concerning the unknown parameter $\theta \in \Theta$. Letting G denote the available set of actions, the loss incurred in taking action $a \in G$ when θ is the true value of the parameter will be denoted $L(\theta, a)$. (This and all other functions we consider will be assumed to be appropriately measurable.) For simplicity, we assume that $L(\theta, a) \geq -K > -\infty$. A nonrandomized decision rule, $\delta(x)$, is a function from \mathcal{X} into G , with $\delta(x)$ being the action taken when x is observed. Let \mathcal{D} be a set of nonrandomized decision rules under consideration. (The theory developed below applies also to randomized decision rules, but, since the only applications we know of are to situations in which the nonrandomized decision rules form a complete class, we will, for simplicity, restrict attention to nonrandomized decision rules.) The risk function of a decision rule δ is the expected loss, to be denoted $R(\theta, \delta) = E_\theta[L(\theta, \delta(X))]$. A decision rule δ is said to be inadmissible (with respect to \mathcal{D}) if there exists a decision rule δ^* such that $R(\theta, \delta^*) \leq R(\theta, \delta)$ for all θ , with strict inequality for some θ . Otherwise, δ is said to be admissible (with respect to \mathcal{D}). A decision rule δ^π is said to be Bayes with respect to a finite measure π on Θ if it minimizes the Bayes risk $r(\pi, \delta) = E^\pi[R(\theta, \delta)]$ among all decision rules in \mathcal{D} . The finite measure π will be called a prior distribution, and assume that attention is to be restricted to a class Θ^* of prior distributions.

The standard method of showing that a decision rule is inadmissible is simply to construct a better decision rule. This is often difficult, however, and so Brown (1979) developed a heuristic method with which one can verify whether or not a decision rule is heuristically inadmissible. A technique for implementing the heuristics is given in Brown (1980) (see also Brown (1979)), and requires only the construction of a better decision rule for "large θ ".

A technique similar to that in Brown (1980) can be derived from another viewpoint, namely as a consequence of Stein's necessary and sufficient condition for admissibility (Stein (1955)). This technique also allows one to conclude that a decision rule is inadmissible by finding a better decision rule for "large θ ". The advantage of this alternate technique is that it can sometimes deal with parameter spaces that are not closed, a situation which causes difficulties in the use of Brown (1980). On the other hand, the auxiliary conditions in Brown (1980) tend to be much easier to deal with than those needed here. Further discussion of this difference will be given after the presentation of the technique.

2. Proving Inadmissibility

In some decision problems it is possible to represent the Bayes risk of certain decision rules, with respect to any prior $\pi \in \Theta^*$, by

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) \pi(d\theta) = \int_{\Gamma} R^*(\gamma, \delta) \tilde{\pi}(d\gamma) + K(\pi) ,$$

where $K(\pi)$ is a constant (independent of δ), γ corresponds to

some transformation of θ , Γ contains the image of θ under this transformation, and $\tilde{\pi}$ is related to the measure induced on Γ by π . There are two reasons for introducing such a reformulation. First, the γ formulation may be much easier to work with, such as in Berger and Zaman (1980), where θ is p -dimensional while Γ is one-dimensional. Second, it can be possible to work with a closed space Γ when the original θ is open. This can greatly simplify the analysis. For many problems, of course, no such reduction is needed (or possible). It is probably easiest to interpret the basic result which follows by ignoring the reduction and thinking in terms of the original θ problem (i.e., identify γ with θ , $\tilde{\pi}$ with π , etc.).

Theorem. Let $\delta^0 \in \mathcal{D}$ be an admissible decision rule for which (2.1) holds. Assume that there exists a sequence $\{\pi_n\}$ of finite measures in Θ^* , with corresponding Bayes rules δ^n such that $r(\pi_n, \delta^n) < \infty$, and a nonnegative function $h(\gamma)$, which is strictly positive on the interior of Γ , such that

- (i) $\tilde{\pi}_n(C) \geq 1$ ($n = 1, 2, \dots$), for some compact set C in the interior of Γ ;
- (ii) $\lim_{n \rightarrow \infty} \int_{\Theta} [R(\theta, \delta^0) - R(\theta, \delta^n)] \pi_n(d\theta) = 0$;
- (iii) the measures

$$\mu_n(d\gamma) = h(\gamma) \tilde{\pi}_n(d\gamma) / \int_{\Gamma} h(\gamma) \tilde{\pi}_n(d\gamma)$$

are probability measures and converge weakly to a probability measure μ on Γ .

Let δ^* be a decision rule such that (2.1) holds and the function

$$g(\gamma) = [h(\gamma)]^{-1} [R^*(\gamma, \delta^0) - R^*(\gamma, \delta^*)]$$

is continuous on Γ and is positive outside some compact set $B \subset \Gamma$.

Then it must be true that

$$(2.2) \quad \int_{\Gamma} g(\gamma) \mu(d\gamma) \leq 0 .$$

Proof. Assume the contrary, namely that

$$(2.3) \quad \int_{\Gamma} g(\gamma) \mu(d\gamma) > 0 .$$

Since π_n is a finite measure and $R(\theta, \delta) \geq -K$, it is clear that $r(\pi, \delta) = E^\pi[R(\theta, \delta)]$ exists for all $\delta \in \mathcal{D}$ (though it is possibly infinite). From the definition of δ^n , it follows that

$$\int R(\theta, \delta^n) \pi_n(d\theta) \leq \int R(\theta, \delta^*) \pi_n(d\theta) .$$

Also, condition (ii) of the theorem and the assumption that

$r(\pi_n, \delta^n) < \infty$ imply that $r(\pi_n, \delta^0) < \infty$ for, say, $n \geq N$. Combining

this with the above result and defining

$$c_n = \int_{\Gamma} h(\gamma) \tilde{\pi}_n(d\gamma) ,$$

it follows that, for $n \geq N$,

$$\begin{aligned}
(2.4) \quad \int_{\Theta} [R(\theta, \delta^0) - R(\theta, \delta^n)] \pi_n(d\theta) &\geq \int_{\Theta} [R(\theta, \delta^0) - R(\theta, \delta^*)] \pi_n(d\theta) \\
&= \int_{\Gamma} [R^*(\gamma, \delta^0) - R^*(\gamma, \delta^*)] \tilde{\pi}_n(d\gamma) \\
&= c_n \int_{\Gamma} g(\gamma) \mu_n(d\gamma) .
\end{aligned}$$

Define (for $k > 0$)

$$g_k(\gamma) = \begin{cases} g(\gamma) & \text{if } g(\gamma) \leq k \\ k & \text{if } g(\gamma) > k \end{cases} .$$

Since g is continuous, g_k is continuous. Also,

$$\sup_{\gamma} |g_k(\gamma)| \leq \max\{\sup_{\gamma \in B} |g(\gamma)|, k\} < \infty ,$$

since B is a compact set and g is continuous. Hence g_k is a continuous bounded function on Γ , so that

$$(2.5) \quad \lim_{n \rightarrow \infty} \int_{\Gamma} g_k(\gamma) \mu_n(d\gamma) = \int_{\Gamma} g_k(\gamma) \mu(d\gamma) .$$

Defining $K = \sup_{\gamma \in B} g(\gamma)$, it is also clear that $g_k(\gamma) \leq g(\gamma)$ for $k \geq K$, which together with (2.5), gives that (for $k \geq K$)

$$(2.6) \quad \lim_{n \rightarrow \infty} \int_{\Gamma} g(\gamma) \mu_n(d\gamma) \geq \int_{\Gamma} g_k(\gamma) \mu(d\gamma) .$$

Note, on the other hand, that the g_k are increasing in k , so that

the monotone convergence theorem implies that

$$\lim_{k \rightarrow \infty} \int g_k(\gamma) \mu(d\gamma) = \int g(\gamma) \mu(d\gamma) .$$

Combining this with (2.6) and (2.3) leads to the conclusion that

$$(2.7) \quad \lim_{n \rightarrow \infty} \int_{\Gamma} g(\gamma) \mu_n(d\gamma) \geq \int g(\gamma) \mu(d\gamma) > 0 .$$

Next, let $\varepsilon = \inf_{\gamma \in C} h(\gamma)$. Observe that ε is greater than zero, since C is a compact set in the interior of Γ , and h is strictly positive and continuous on the interior of Γ . By condition (i) of the theorem, it follows that

$$c_n = \int_{\Gamma} h(\gamma) \tilde{\pi}_n(d\gamma) \geq \varepsilon \tilde{\pi}_n(C) \geq \varepsilon .$$

Together with (2.4) and (2.7), this gives that

$$\lim_{n \rightarrow \infty} \int_{\Theta} [R(\theta, \delta^0) - R(\theta, \delta^n)] \pi_n(d\theta) \geq \varepsilon \int_{\Gamma} g(\gamma) \mu(d\gamma) > 0 .$$

But this contradicts condition (ii) of the theorem. Hence (2.3) cannot hold, completing the proof. ||

The above theorem can be used to show that a decision rule δ^0 is inadmissible, by first showing that if δ^0 were admissible, then conditions (i), (ii), and (iii) would be satisfied, but that there exists an appropriate δ^* for which (2.2) is violated. This contradiction establishes the inadmissibility of δ^0 .

The verification of conditions (i) and (ii) of the theorem can generally be carried out using Stein's necessary and sufficient condition for admissibility (Stein (1955)). Indeed (i) and (ii) are frequently just more explicit statements of Stein's necessary condition. (Farrell (1968a and 1968b) establish this for a wide variety of problems.)

The verification of condition (iii) of the theorem tends to be quite difficult. It involves showing that if (ii) is satisfied, then a properly normalized subsequence of the π_n will converge to a probability measure. It should be possible to demonstrate this fairly easily for problems of estimating, under a quadratic loss, the natural parameter (vector) of a distribution from the (multivariate) exponential family. See Brown (1971) and Berger and Srinivasan (1978) for indications of why this is so. For other problems, however, no general method of verification seems available.

The rule δ^* is found by trying to determine a rule which is better than δ^0 for "large θ " (or "large γ "). This is usually much easier than trying to find a rule that is better for all θ . Indeed the problem typically reduces to that of solving a differential inequality. (See Brown (1979).) When a better rule for large θ is found, it will frequently be the case that the integral in (2.2) over these large θ is infinite. It will then follow easily that (2.2) is violated, so that δ^0 is inadmissible.

Two final points deserve mention. First, the measure μ can sometimes be explicitly represented in terms of δ^0 , or, at least, the integral in (2.2) can be replaced by an integral involving δ^0 , but not μ . This is a very desirable simplification.

Finally, the verification that $g(\gamma)$ is positive outside some compact set can be hard when Γ is not closed. For example, if $\Gamma = (0, \infty)$, then one must show that δ^* is better than δ^0 for small γ , as well as for large γ . This can be annoying, and is indeed the difficulty encountered in using Brown (1980). The difficulty can be circumvented, here, by trying to choose a closed Γ . If Θ is closed to start with, Brown (1980) will usually be easier to apply than Theorem 1, since Brown (1980) does not require the verification of conditions (i) through (iii).

For an explicit application of the above theory, the reader is referred to Berger and Zaman (1979).

References

- Berger, J. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. Ann. Statist. 6, 783-801.
- Berger, J. and Zaman, A. (1979). Inadmissibility in the control problem. Technical Report No. 52, Division of Biostatistics, Stanford University.
- Brown, L. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann. Math. Statist. 42, 855-903.
- Brown, L. (1979). A heuristic method for determining admissibility of estimators - with applications. Ann. Statist. 7, 960-994.
- Brown, L. (1980). A minimax theorem and a necessary condition for admissibility. To appear in Ann. Statist.
- Farrell, R. H. (1968a). On a necessary and sufficient condition for admissibility of estimators when strictly convex loss is used. Ann. Math. Statist. 38, 23-28.
- Farrell, R. H. (1968b). Towards a theory of generalized Bayes tests. Ann. Math. Statist. 38, 1-22.
- Stein, C. (1955). A necessary and sufficient condition for admissibility. Ann. Math. Statist. 26, 518-522.