Admissibility for Vector-valued Loss Functions

by

George P. McCabe, Jr.

Department of Statistics

Division of Mathematical Sciences

Mimeograph Series No. 272

November, 1971

ABSTRACT

Admissibility for Vector-valued Loss Functions

An admissibility criterion for vector-valued loss functions is proposed and the relationship between this concept and the classical criterion in terms of linear combinations of the components is examined. Generalizations of some standard admissibility theorems are given.

Admissibility for Vector-valued Loss Functions

by

George P. McCabe, Jr.

Purdue University

1. Introduction. One of the basic concepts in the formulation of a statistical problem in decision theoretic terms is the loss function. For a given state of nature and a particular action, one expresses the consequences of the action in terms of this real valued function. It may be validly argued that the construction of the loss function is not within the domain of statistics, i.e. the loss function must be presented to the statistician along with the other relevant aspects of the model.

In some instances, however, it is possible to derive results which are valid for all loss functions in a given class. The usefulness of such results is apparent in light of the fact that often the construction of the loss function is exceedingly difficult.

For many practical problems, the loss function is a linear combination of two or more functions which may be viewed as components of loss. For example, in an estimation problem, one may consider one component to be the squared error and another to be the cost of the observations.

For estimating the mean vector of a multivariate normal distribution with identity covariance matrix, a standard loss function is the sum of the squared errors for each component. If one considers the estimation of each component to be a separate problem then the sample means are admissible estimates. However, for the composite problem, Stein (1956) has shown that these estimates are inadmissible when the number of components is greater than two. Thus, when components of loss are added, one cannot always extrapolate results from the component problems.

In this paper, an admissibility criterion for evaluating decision procedures in terms of the individual components of a loss vector is proposed. Some relationships between this concept and the classical criterion for appropriate linear combinations are examined.

2. Basic notation and definitions.

We assume $(X, \mathcal{B})$ is a measurable space and $P = \{P_\omega: \omega \in \Omega\}$ is a set of probabilities with the property that $(X, \mathcal{B}, P_\omega)$ is a probability space for each $\omega \in \Omega$. The set Ω is called the parameter space. Let $\mathcal{A}$ denote the action space and $\mathcal{D}$ be the set of decision functions. An element $d \in \mathcal{D}$ is a function from X to $\mathcal{A}$. Equivalently, $\mathcal{D}$ could be some class of suitably defined randomized rules (see Ferguson (1967) for definitions). For any $\omega \in \Omega$ and $a \in \mathcal{A}$, we denote the loss vector by

$$\ell(a,\omega) = (\ell_1(a,\omega),\ldots,\ell_k(a,\omega))'$$

where $\ell_i(a,\omega) \geq 0$ for all i, and the risk vector by

$$r(d,\omega) = E_\omega \ell(d(x),\omega) = (r_1(d,\omega),\ldots,r_k(d,\omega))'$$

where

$$r_i(d,\omega) = E_\omega \ell_i(d(x),\omega), \qquad i=1,\ldots,k \ .$$

For any vector $b = (b_1,\ldots,b_k)'$ with $b_i > 0$ for all i and $\Sigma b_i=1$, let

$$L_b(a,\omega) = b'\ell(a,\omega) = \sum_{i=1}^{k} b_i \, \ell_i(a,\omega)$$

and

$$R_b(d,\omega) = E_\omega L_b(d(x),\omega) = \sum_{i=1}^{k} b_i \, r_i(d,\omega) \; .$$

Definitions. (a) $d \in \mathcal{D}$ is admissible (L_b) if there does not exist $d' \in \mathcal{D}$ such that

(2.1) $$R_b(d',\omega) \le R_b(d,\omega) \quad \text{for all } \omega,$$

and

(2.2) $$R_b(d',\omega) < R_b(d,\omega) \quad \text{for some } \omega.$$

(b) $d \in \mathcal{D}$ is admissible (ℓ) if there does not exist $d' \in \mathcal{D}$ such that

(2.3) $$r_i(d',\omega) \le r_i(d,\omega) \quad \text{for all } (\omega,i) \; ,$$

and

(2.4) $$r_i(d',\omega) < r_i(d,\omega) \quad \text{for some } (\omega,i) \; .$$

(c) $d \in \mathcal{D}$ is Bayes-$\Pi(\ell)$ for a prior Π on Ω if there does not exist $d' \in \mathcal{D}$ such that

(2.5) $$B_i(d') \le B_i(d) \quad \text{for all } i \; ,$$

and

(2.6) $$B_i(d') < B_i(d) \quad \text{for some } i \; ,$$

where

$$B_i(d) = \int r_i(d,\omega)\, d\Pi(\omega) \quad (2.7)$$

3. Some theorems. To avoid unnecessary complications in what follows, we assume

$$r_i(d,\omega) < \infty \quad \text{for all } (\omega,i,d) .$$

Theorem 1. If d is admissible (L_b) for some b, then d is admissible (ℓ).

Proof. Suppose d is admissible (L_b) and not admissible (ℓ). Then there exists $d' \in \mathcal{D}$ such that (2.3) and (2.4) hold. But since $b_i > 0$ for all i, this implies that (2.1) and (2.2) hold, i.e. d is inadmissible (L_b). □

Stein's (1956) work provides a counterexample to the converse of the above theorem. Note that the restriction $\Sigma b_i = 1$ is merely a convenient normalization. Conditions under which admissibility (ℓ) implies admissibility (L_b) can be derived. Theorem 2 gives these conditions for the case of two components of loss. For any $b = (b_1,b_2)$, we identify the vector by its first component, i.e. $b_1 = b$, $b_2 = 1-b$. For any two decisions d and d', let

$$s_{i\omega}(d,d') = r_i(\omega,d') - r_i(\omega,d) \text{ for } i = 1,2 ,$$

$$t_\omega(d,d') = s_{2\omega}(d,d')/(s_{2\omega}(d,d') - s_{1\omega}(d,d')) ,$$

$$\Omega_0(d,d') = \{\omega:\ s_{1\omega}(d,d') = s_{2\omega}(d,d')\} ,$$

$$\Omega_{00}(d,d') = \{\omega:\ s_{1\omega}(d,d') = s_{2\omega}(d,d') = 0\} ,$$

$$\Omega_1(d,d') = \{\omega:\ s_{1\omega}(d,d') < s_{2\omega}(d,d')\} ,$$

$$\Omega_2(d,d') = \{\omega:\ s_{1\omega}(d,d') > s_{2\omega}(d,d')\} ,$$

$$c_1(d,d') = \max\{0, \sup(t_\omega(d,d'): \omega \in \Omega_1(d,d'))\} ,$$

and

$$c_2(d,d') = \min\{1, \inf(t_\omega(d,d'): \omega \in \Omega_2(d,d'))\} .$$

Let A denote the following four conditions:

(1) $t_\omega(d,d') = c_1(d,d')$ for all $\omega \in \Omega_1(d,d')$,

(2) $t_\omega(d,d') = c_2(d,d')$ for all $\omega \in \Omega_2(d,d')$,

(3) $\Omega_0(d,d') = \Omega_{00}(d,d')$, and

(4) one or more of the following:

(a) $\Omega_1(d,d') = \phi$,

(b) $\Omega_2(d,d') = \phi$,

(c) $c_1(d,d') = c_2(d,d')$.

Let

$$I(d,d') = \begin{cases} (c_1(d,d'), c_2(d,d')) & \text{if A holds,} \\ [c_2(d,d'), c_2(d,d')] & \text{otherwise.} \end{cases}$$

Note that $I(d,d')$ may consist of a single point or be degenerate. For any $d \in \mathfrak{D}$, let

$$\mathfrak{D}_d = \mathfrak{D} - \{d': s_{i\omega}(d,d') = 0 \text{ for all } (i,\omega)\}$$
$$- \{d': s_{2\omega}(d,d') = s_{1\omega}(d,d') > 0 \text{ for some } \omega\} ,$$

and

$$\mathcal{B}_d = (0,1) - \bigcup_{d' \varepsilon \mathcal{D}_d} I(d,d') \ .$$

Theorem 2. If d is admissible (ℓ) then d is admissible (L_b) if and only if $b \varepsilon \mathcal{B}_d$.

Proof. First note that L_b-admissibility is defined only for $0 < b < 1$. Assume d is admissible (ℓ). If $b \not\varepsilon \mathcal{B}_d$ then there exists a $d' \varepsilon \mathcal{D}_d$ such that $b \varepsilon I(d,d')$ or equivalently,

$$c_1(d,d') \leq b \leq c_2(d,d')$$

with strict inequality if A holds. Now, for all $\omega \varepsilon \Omega_1(d,d')$,

$$t_\omega(d,d') \leq b$$

and thus,

$$bs_{1\omega}(d,d') + (1-b)\ s_{2\omega}(d,d') \leq 0 \ .$$

From the definition of $s_{i\omega}(d,d')$ it follows that

$$br_1(\omega,d') + (1-b)r_2(\omega,d') \leq br_1(\omega,d) + (1-b)r_2(\omega,d)$$

or equivalently,

$$R_b(\omega,d') \leq R_b(\omega,d) \ , \tag{3.1}$$

with strict inequality for at least one $\omega \varepsilon \Omega_1(d,d')$ unless $t_\omega(d,d') = b$ for all $\omega \in \Omega_1(d,d')$. Note that this exception includes the possibility that $\Omega_1(d,d') = \phi$. A similar argument shows that (3.1) holds for all $\omega \in \Omega_2(d,d')$ with strict inequality for at least one such ω unless $t_\omega(d,d') = b$ for all $\omega \in \Omega_2(d,d')$. In addition, it is easy to see that (3.1) holds for all $\omega \in \Omega_0(d,d')$ with strict inequality for some ω in this set unless $\Omega_0(d,d') = \Omega_{00}(d,d')$. Hence, (3.1) holds for all $\omega \in \Omega$ unless $t_\omega(d,d') = b$ for all $\omega \in \Omega_1(d,d')$, $t_\omega(d,d') = b$ for all $\omega \in \Omega_2(d,d')$ and $\Omega_0(d,d') = \Omega_{00}(d,d')$. Since $0 < b < 1$ and $c_1(d,d') \leq b \leq c_2(d,d')$, the above exception is equivalent to condition A. However, if A holds then $c_1(d,d') < b < c_2(d,d')$ from the definition of $I(d,d')$. If, in addition, either $\Omega_1(d,d') \neq \phi$ or $\Omega_2(d,d') \neq \phi$ then there exists at least one ω for which the inequality in (3.1) is strict. Alternatively, if $\Omega_1(d,d') = \Omega_2(d,d') = \phi$ then $\Omega_0(d,d') = \Omega$ and the equality $\Omega_0(d,d') = \Omega_{00}(d,d')$ cannot hold because such d' have been excluded from $\mathcal{D}_d$. Thus, if $b \notin \mathcal{B}_d$, then there exists a $d' \in \mathcal{D}_d$ which is better, i.e. d is inadmissible (L_b).

On the other hand, if $b \in \mathcal{B}_d$ and d is inadmissible (L_b) then there exists a rule $d' \in \mathcal{D}$ such that

(3.2) $$R_b(\omega,d') \leq R_b(\omega,d) \text{ for all } \omega \in \Omega$$

and

(3.3) $$R_b(\omega_0,d') < R_b(\omega_0,d) \text{ for some } \omega_0 \in \Omega \text{ .}$$

For $\omega \in \Omega_0(d,d')$, the inequality (3.2) is equivalent to

(3.4) $$s_{1\omega}(d,d') = s_{2\omega}(d,d') \leq 0 \text{ .}$$

Also, by virtue of (3.3), it is not possible to have

$$s_{i\omega}(d,d') = 0 \text{ for all } i \text{ and } \omega .$$

Therefore, $d' \in \mathcal{D}_d$. Now, for $\omega \in \Omega_1(d,d')$, (3.2) implies

$$br_1(\omega,d') + (1-b)r_2(\omega,d') \leq br_1(\omega,d) + (1-b)r_2(\omega,d)$$

or equivalently

$$t_\omega(d,d') \leq b .$$

Similarly, for $\omega \in \Omega_2(d,d')$,

$$b \leq t_\omega(d,d') .$$

Therefore,

$$c_1(d,d') \leq b \leq c_2(d,d') . \qquad (3.5)$$

If the inequalities in (3.5) are strict, then clearly $b \in I\ (d,d')$ which implies the contradiction that $b \notin \mathcal{B}_d$. If condition A does not hold, then the interval $I(d,d')$ is closed and the same contradiction results. Finally, suppose that A holds and either $b = c_1(d,d')$, $b = c_2(d,d')$ or both. If the ω_0 for which (3.3) holds is in $\Omega_0(d,d')$ then the inequality in (3.4) must be strict. However, this contradicts part (3) of condition A. Therefore, ω_0 must be in either $\Omega_1(d,d')$ or $\Omega_2(d,d')$. Suppose that $b = c_1(d,d')$. Then, $\omega_0 \notin \Omega_1(d,d')$ since part (1) of condition A would be violated. Therefore, $\omega_0 \in \Omega_2(d,d')$. But by part (4) of condition A, this implies that either $\Omega_1(d,d') = \phi$, $c_1(d,d') = c_2(d,d')$, or both. The first alternative is incompatible with the condition $b > 0$ whereas the second implies

$$t_{\omega_0}(d,d') = b$$

which violates (3.3). A similar argument holds for the case where $b = c_2(d,d')$. Therefore, $b \notin \mathcal{B}_d$ and the theorem is proved. $\square$

Example. Suppose that $\Omega = \{1,2\}$ and $\mathcal{D}$ is a family of rules indexed by a parameter ε, with $0 \leq \varepsilon \leq 1/2$, having the following risk functions:

$$r_1(1,\varepsilon) = 1 + \varepsilon^2$$

$$r_1(2,\varepsilon) = 1 - 2\varepsilon$$

$$r_2(1,\varepsilon) = 1 - \varepsilon$$

$$r_2(2,\varepsilon) = 1 - \varepsilon \ .$$

Clearly, all rules are admissible (ℓ). It is easy to verify that

$$I(\varepsilon,\delta) = \begin{cases} \phi & \text{if } \varepsilon > \delta \ , \\ [0,\ 1/(1+\varepsilon+\delta)] & \text{if } \varepsilon < \delta \ . \end{cases}$$

Hence,

$$\mathcal{B}_\varepsilon = [1/(1+2\varepsilon),\ 1) \ .$$

Thus, the rule $\varepsilon = 0$ is admissible (L_b) for all $b \in (0,1)$, whereas for any other rule, there is an interval of values of b for which it is admissible (L_b).

This example suggests the following corollary:

Corollary. If for any rule d, there is a subset $\mathcal{D}' \subset \mathcal{D}$ such that

(a) $\sup_{d' \in \mathcal{D}'} \sup\{t_\omega(d,d') : \omega \in \Omega_1(d,d')\} = 1$

and

$$c_2(d,d') = 0 \quad \text{for all } d' \in \mathcal{D}' ,$$

or

(b) $\inf_{d' \in \mathcal{D}'} \inf\{t_\omega(d,d') : \omega \in \Omega_2(d,d')\} = 0$

and

$$c_1(d,d') = 1 \quad \text{for all } d' \in \mathcal{D} ,$$

then d is inadmissible (L_b) for all $b \in (0,1)$.

Many of the standard admissibility theorems can be generalized for vector valued loss functions. The following are some examples:

Theorem 3. If Ω is finite and Π assigns positive probability to each point, then any Bayes-$\Pi(\ell)$ rule is admissible (ℓ).

Proof. Let $\Omega = (\omega_1,\ldots,\omega_n)$ and $\Pi = (\pi_1,\ldots,\pi_n)$ and suppose d is Bayes-$\Pi(\ell)$ but not admissible (ℓ). Then there exists $d' \in \mathcal{D}$ such that (2.3) and (2.4) hold. But this clearly implies (2.5) and (2.6), a contradiction. □

A rule d is almost (Π) admissible (ℓ) if there does not exist $d' \in \mathcal{D}$ such that

(3.6) $$r_i(d',\omega) \le r_i(d,\omega) \text{ for all } i \text{ and } \omega ,$$

and

(3.7) $$r_i(d',\omega) < r_i(d,\omega)$$

for some i and all ω in a set of positive Π-probability.

<u>Theorem 4.</u> If d is Bayes-$\Pi(\ell)$ then d is almost (Π) admissible (ℓ).

<u>Proof.</u> If d is Bayes-$\Pi(\ell)$ and not almost (Π) admissible (ℓ) then there exists $d' \in \mathcal{D}$ such that (3.6) and (3.7) hold. This, however, implies the contradiction that (2.5) and (2.6) must hold. □

Two rules d and d' are <u>equivalent</u> if

(3.8) $$r_i(d,\omega) = r_i(d',\omega) \text{ for all } i \text{ and } \omega .$$

A Bayes-$\Pi(\ell)$ rule d is <u>essentially unique</u> if any other rule having the same Bayes risk vector is equivalent to d, i.e. if d' is such that $B_i(d) = B_i(d')$ for $i = 1,\ldots,k$, then d' is equivalent to d.

<u>Theorem 5.</u> A Bayes-$\Pi(\ell)$ rule is admissible (ℓ) if it is essentially unique.

<u>Proof.</u> Suppose d is an essentially unique Bayes-$\Pi(\ell)$ rule which is inadmissible (ℓ). Then there exists $d' \in \mathcal{D}$ such that (2.3) and (2.4) hold. This implies $B_i(d') \leq B_i(d)$ for all i. But since d is Bayes-$\Pi(\ell)$, this is equivalent to $B_i(d') = B_i(d)$ which implies (3.8) since d is essentially unique. Hence, (2.4) cannot hold and d must be admissible (ℓ). □

<u>4. Conclusions.</u> Given the appropriate definitions, many additional results can be obtained. The usefulness of the concept of vector-valued loss functions, however, must be measured in terms of its applicability. Many so-called optimal properties may be reformulated in terms of admissibility (ℓ) for a certain loss vector. The optimal property of the sequential probability ratio test is an

example. In addition, many other statistical procedures are evaluated in terms of more than one loss function. In subset selection procedures one considers the probability of correct selection and the size of the subset, while for confidence intervals, one is interested in the coverage probability and the length of the interval. This paper represents an attempt at extending the structure of decision theory so that more problems may be meaningfully formulated within its framework.

REFERENCES

[1] Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

[2] Stein, C. (1956) Inadmissibility of the usual estimator for the mean of a multi-variate normal distribution. Proc. Third Berkeley Symp. Math. Statist. 1: 197-206.