THE DIFFERENTIAL INEQUALITY OF A STATISTICAL ESTIMATION PROBLEM

LAWRENCE D. BROWN
Department of Mathematics
White Hall
Cornell University
Ithaca, New York 14853, U.S.A.

I. MOTIVATION

This introductory chapter describes some well known facts about a familiar statistical estimation problem. It is then described how these facts motivate the study of a class of differential inequalities. These differential inequalities are the main topic of this paper with the main results stated in Part II. Some further comments about the statistical applicability of these results concerning differential inequalities are contained in part III. Part IV contains proofs and some additional results of a more technical nature.

1. UNBIASED ESTIMATE OF THE DIFFERENCE IN RISKS.

Let $X$ be a $k$-variate normal vector with mean $\theta \in \mathbb{R}^k$ and covariance matrix $I$. It is desired to estimate $\theta$ with the customary loss function $L(\theta, a) = ||a - \theta||^2$. Write a non-randomized estimator $\delta: \mathbb{R}^k \rightarrow \mathbb{R}^k$ in the form $\delta(x) = \delta_0(x) + \gamma(x)$ where $\delta_0(x) = x$. Integration by parts yields the following expression for the risk, $\mathcal{R}(\theta, \delta) = E_\theta(L(\theta, \delta(X)))$:

$$\mathcal{R}(\theta, \delta) = k + E_\theta(2 \nabla \cdot \gamma(X) + \|\gamma(X)\|^2) = k + E_\theta(R_0 \gamma(X)).$$

(1.1)


Let

$$\gamma(x) = -x \min \left( \frac{c(k-2)}{||x||^2}, 1 \right), \quad 0 < c < 2, \quad k \geq 3. \quad (1.2)$$

An easy calculation shows that $R_0 \gamma(x) < 0$ for all $x$. Hence $\mathcal{R}(\theta, \delta) < k = \mathcal{R}(\theta, \delta_0)$ which shows that $\delta_0$ is inadmissible for $k \geq 3$. The estimator $\delta(x) = x + \gamma(x)$ used in this calculation is of course the positive part James-Stein (1960) estimator.

Suppose, now, that $\delta_\gamma(x) = x + \gamma(x)$ is a proposed estimator and $\delta_{(\gamma + \lambda)}(x) = x + \gamma(x) + \lambda(x)$ is a proposed competing estimator. Then from (1.1)

$$\Delta(\theta) = \mathcal{R}(\theta, \delta_{(\gamma + \lambda)}) - \mathcal{R}(\theta, \delta_\gamma)$$

299
where the operator \( R_\gamma \) is defined by the above. It follows from (1.3) that the estimator \( \delta_\gamma \) is inadmissible if there exists an absolutely continuous function \( \lambda \neq 0 \) such that
\[
R_\gamma \lambda(x) \leq 0 \quad \forall x. \tag{1.4}
\]
(Note that if (1.1) holds then \( R_\gamma (\lambda/2)(x) \leq -\|\lambda(x)\|^2/4 \) and hence is negative wherever \( \lambda(x) \neq 0 \).)

The expression (1.3) suggests as a heuristic step approximating \( \Delta(\theta) \) by \( R_\gamma \lambda(\theta) \). In other words, one thinks of \( R_0 \gamma(\theta) \) as the risk function associated with \( \gamma \) (corresponding to \( \delta_\gamma \)), and thinks of \( R_\gamma \lambda = R_0 \gamma - R_0 (\gamma + \lambda) \) as the difference in risks between \( \gamma \) and \( (\gamma + \lambda) \). If there is a \( \lambda \neq 0 \) for which (1.4) holds one concludes that the corresponding \( \delta_\gamma \) is inadmissible. If not, one conjectures that \( \delta_\gamma \) is admissible.

The objective of this paper is to study differential inequalities like (1.4) in the hope of providing a useful tool for generating valid conjectures about admissibility and of increasing understanding about the mathematical structure of admissibility.

2. BAYES PROCEDURES AND BAYES RISK.

Consider the same normal distribution estimation problem as in Section 1. Let \( g \) be a prior density (with respect to Lebesgue measure). Following an integration by parts the Bayes procedures can be written as
\[
\delta_g^*(x) = x + \frac{\nabla g^*(x)}{g^*(x)} = x + \gamma_g^* \tag{2.1}
\]
where
\[
g^*(x) = (g^*(\varphi))(x) = \int (2\pi)^{-k/2} \exp(-\|x - \theta\|^2/2) g(\theta) d\theta.
\]
(See Stein (1965) and Brown (1971).) Furthermore the Bayes risk can be written as
\[
R(g, \delta_g^*) = k - \int \| \frac{\nabla g^*(x)}{g^*(x)} \|^2 g^*(x) dx. \tag{2.2}
\]
(See Stein (1959) and Brown (1971).)

A tempting heuristic is to approximate \( \delta_g^* \) by
\[
g_g^* \approx x + \frac{\nabla g(x)}{g(x)} \tag{2.3}
\]
(and hence to approximate \( \gamma_g^* \) by \( \nabla g(x)/g(x) \)) and accordingly to approximate
the Bayes risk by

\[ R(g, \delta_2) = k - \int \| \nabla g(x) / g(x) \|^2 g(x) dx. \] (2.4)

(This idea has recently been exploited in interesting ways in Bickel (1981), Berger (1982), and Marazzi (1985). See also Brown (1979a), (1980b), (1985).)

Although not immediately apparent, expressions related to (2.3) and (2.4) can be used to provide a practical test as to whether inequalities like (1.4) possess non-trivial solutions. This test is described in Theorems 1 and 3, stated in Section 5.

These results provide a link between the heuristics suggested below (1.4) and those surrounding (2.3) and (2.4) above. The existence of such a link could indeed be anticipated on the basis of Blyth's (1951) method, which is the most flexible tool for proving admissibility of an estimator. This method requires construction of a sequence of prior distributions whose Bayes risks behave in an appropriate fashion, namely \( R(g, \delta) - R(g, \delta_{(x)}) = o(\int_{\|g\| \leq 1} \theta(\theta) d\theta) \). In fact Stein (1955) shows that an appropriately formulated version of Blyth's method is a necessary and sufficient condition for admissibility of \( \delta \). See also Farrell (1966) for a different formulation and proof of this fact due to Le Cam. Thus an expression such as (2.2) for the Bayes risk relates to a test for admissibility. This same relation turns out to be valid in the heuristic domain; thus, the expression on the right of (2.4) relates to a test for insolubility of the differential inequality (1.4).

3. PLAN OF THE PAPER.

Part II describes the main results concerning differential inequalities. In that chapter differential operators like \( R_0 \) appearing in (1.1) are treated as if \( R_0 \gamma \) were actually the risk function of the estimator corresponding to \( \gamma \). In this way admissibility relative to \( R_0 \) can be defined and an expression can be given (analogous to (2.1)) which defines Bayes estimators.

Theorems 1 and 3 in Section 5 are the principal results of Part II. They provide necessary and sufficient conditions for admissibility relative to \( R_0 \). These conditions parallel the previously mentioned statistical result of Stein and Le Cam, but are often easier to implement. The proofs of these results are deferred to Part IV of this paper. Part II concludes with a section of examples illustrating the application of Theorems 1-4 to specific operators \( R_0 \) of statistical interest.

Part III recapitulates the statistical motivation for the development of Part II in a somewhat more general context than that already given in this introduction. Then two specific statistical results are given as easy applications of the theory so far derived. The first of these applications concerns the previously mentioned positive part James-Stein estimator. This estimator is known to be
statistically inadmissible. However it is admissible relative to the differential operator \( R_0 \). This means it is impossible to find a statistically dominating estimator for which the unbiased estimate of the difference in risk is always non-negative.

The second application concerns a Brownian motion with constant drift, \( \mu \), stopped by the boundaries of a truncated sequential probability ratio test. It is shown that the ordinary, maximum likelihood estimator of \( \mu \) is inadmissible. This result may at first appear surprising since for a different stopping rule (such as stopping at a fixed time) the maximum likelihood estimator is admissible.

II. MAIN RESULTS AND EXAMPLES

4. DESCRIPTION OF THE PROBLEM.

Consider the (non-linear) operator \( R_0 \) defined by

\[
R_0 \gamma = 2D\gamma + \gamma ' B \gamma
\]

when \( \gamma : \mathbb{R}^k \to \mathbb{R}^k \) is a differentiable vector valued function and where \( D \) is the (linear) first order differential operator

\[
D\gamma = \sum_i a_i \lambda_i + \sum_{ij} a_{ij} \frac{\partial \lambda_i}{\partial x_j}.
\]

Here \( B = (b_{ij}) \) is a symmetric positive definite matrix valued function, \( A = (a_{ij}) \) is a non-singular matrix valued function and \( a_i, a_{ij}, b_{ij} \) are everywhere continuously differentiable. (These non-singularity and differentiability assumptions can be somewhat relaxed, but appear to include the application of major statistical interest.)

Because of the statistical motivation discussed in Sections 1 and 8 we call \( R_0 \gamma \) the risk function of \( \gamma \). A function \( \gamma \) is then called admissible if

\[
R_0(\lambda + \gamma) \leq R_0 \gamma \Rightarrow \lambda \equiv 0.
\]  

Otherwise \( \gamma \) is inadmissible. (Later comments provide a generalized definition of \( R_0 \), and consequently of admissibility, which is consistent with the above when \( \gamma, \lambda \) are differentiable.)

The definition of admissibility can be phrased somewhat differently. Let

\[
R_\gamma \lambda = R_0(\gamma + \lambda) - R_0 \gamma = 2D\lambda + 2\gamma ' B \lambda + \lambda ' B \lambda.
\]

Thus \( \gamma \) is inadmissible if there is a \( \lambda \neq 0 \) such that \( R_\gamma \lambda \leq 0 \). If \( \lambda \) satisfies \( R_\gamma \lambda \leq 0 \) then we say that \( \gamma_1 = \gamma + \lambda \) is as good as \( \gamma \). If \( R_\gamma \lambda \leq 0 \) with strict inequality for at least one value of \( x \) then we say that \( \gamma_1 \) is better than \( \gamma \).

Note that the operators \( R_0 \) and \( R_\gamma \) are strictly convex. In particular, if \( R_\gamma \lambda \leq 0 \) then \( R_\gamma(\lambda/2) \leq 0 \) and is \( < 0 \) on \( \{ x: \lambda(x) \neq 0 \} \). Hence if \( \gamma \) is
inadmissible, and \( \gamma_1 = \gamma + \lambda \) is as good as \( \gamma \), then \( \gamma_2 = \gamma + \lambda/2 \) is better than \( \gamma \). In short, \( \gamma \) is admissible if and only if there is no function \( (\gamma_2) \) which is better than \( \gamma \).

In order to develop the general theory it is necessary to provide a generalized definition of \( R_0 \lambda \), and a corresponding interpretation of the notion of admissibility. For a non-negative measurable function \( f: \mathbb{R}^k \to [0, \infty) \) let \( L_2(f) \) denote the Hilbert space of measurable functions \( \gamma: \mathbb{R}^k \to \mathbb{R}^k \) with norm defined by

\[
\|\gamma\|_2^2 = \int \|\gamma(x)\|^2 f(x) \, dx.
\]

We use the symbol \((\cdot, \cdot)\) for the inner product in the space \( L_2(1) \). Let \( \varphi \) be a continuously differentiable non-negative function, vanishing off a compact subset of \( \mathbb{R}^k \). Let \( D^* \) denote the formal dual of \( D \), whose coordinate functions are given by

\[
(D^*\varphi)_i = a_i\varphi - \sum_j \frac{\partial}{\partial x_j} (a_{ij}\varphi), \quad i = 1, \ldots, k. \tag{4.5}
\]

Then, if \( \gamma \) is as above, \( (D\gamma, \varphi) = (\gamma, D^*\varphi) \) with the obvious notational convention that \( (\gamma, D^*\varphi) = \int (\gamma(x) \cdot D^*\varphi(x)) \, dx \). Let \( \phi^+ \) be the set of all functions, \( \varphi \), as above, which also satisfy

\[
\frac{D^*\varphi}{\varphi} \in L_2(\varphi). \tag{4.6}
\]

(Since \( A^*A \) is uniformly positive definite on any compact subset of \( \mathbb{R}^k \) and \( \varphi \) vanishes off a compact subset, (4.6) can be easily seen to be equivalent to \( \nabla \varphi \in L_2(\varphi) \), which is in turn equivalent to \( \nabla (\varphi^{1/2}) \in L_2(1) \).)

For \( \varphi \in \phi^+ \) and any measurable \( \gamma \) define \( R_0 \gamma \) through

\[
(R_0 \gamma, \varphi) = \begin{cases} 2(\gamma, D^*\varphi) + (\varphi'B\gamma, \varphi) & \text{if } \gamma \in L_2(\varphi) \\ \infty & \text{if } \gamma \notin L_2(\varphi) \end{cases}. \tag{4.7}
\]

Note that \( \gamma \in L_2(\varphi) \) implies \( (R_0 \gamma, \varphi) < \infty \). Let \( \mathcal{F} = \{ \gamma: \gamma \in L_2(\varphi) \forall \varphi \in \phi^+ \} \).

Then if \( \gamma \in \mathcal{F} \), \( \gamma \in \phi^+ \), and \( \gamma \) is measurable define \( R_\gamma \lambda \) by

\[
(R_\gamma \lambda, \varphi) = (R_0 (\gamma + \lambda), \varphi) - (R_0 \gamma, \varphi). \tag{4.4'}
\]

The function \( \gamma \in \mathcal{F} \) is now defined to be \textit{admissible} (relative to \( R_0 \)) if there does not exist another function \( \gamma_2 \) which is \textit{better than} \( \gamma \) in the sense that

\[
(R_0 \gamma_2, \varphi) \leq (R_0 \gamma, \varphi) \forall \varphi \in \phi^+ \tag{4.8}
\]

with strict inequality for some \( \varphi \in \phi^+ \), or equivalently if \( (R_\gamma \lambda, \varphi) \leq 0 \forall \varphi \in \phi^+ \) with strict inequality for some \( \varphi \in \phi^+ \). If \( \gamma \) is continuously differentiable then admissibility according to (4.8) implies admissibility according to (4.3), in view of remarks in the paragraph following (4.4).

The above also yields the basic formula expressed in the following propo-
Proposition 1. Suppose $g \in \phi^+$. Let

$$\gamma_g = -B^{-1} \frac{D^+ g}{g} \text{ a.e. } (g(x)dx).$$

(4.9)

Then

$$(R_0 \gamma_g, g) = \inf \{(R_0 \gamma_g, g) : \gamma \in L_2(g)\} > -\infty.$$  

(4.10)

PROOF. The hypotheses of the proposition guarantee the validity of the following:

$$(R_0 \gamma_g, g) = (2\gamma' \frac{D^+ g}{g} + \gamma' B \gamma, g)$$

$$\geq \left( \left( \frac{D^+ g}{g} \right)' B^{-1} \left( \frac{D^+ g}{g} \right), g \right)$$

$$= (R_0 \gamma_g, g) > -\infty$$

since $2v'w + v'Mv \geq -(w'M^{-1}w)$ for any vectors $v$, $w$ and positive definite matrix $M$; and the minimum occurs at $v = -M^{-1}w$.

Theorem 1, below, shows that every admissible procedures is of the form (4.9), however for a wider collection $\{g\} \text{ than hypothesized in Proposition 1.}$

In the statistical context, $g$ would be called a prior (with compact support), $\gamma_g$ would be called the Bayes procedure for $g$, and its form is given by (4.9).

5. MAIN RESULTS.

The latter part of this paper is devoted to proving the basic results concerning admissibility relative to $R_0$ given in Theorems 1-4 below.

Theorem 1. $\gamma \in \mathcal{F}$ is admissible only if

$$\gamma = \gamma_g = -B^{-1} \frac{D^+ g}{g}$$

(5.1)

for some continuous differentiable $g \neq 0$. (In (5.1) observe the convention $\frac{0}{0} = 0$.)

Theorem 1 is proved in Section 13. It can be refined to apply also to admissible $\gamma \in \mathcal{F}$. However such functions are of less interest and we omit the details. A similar remark is also true of some later results.

To prepare for Theorem 2 note that if $\gamma_g$ is given and $h > 0$ (a.e.) then

$$\gamma_h \gamma_g = \gamma_g + \lambda_h$$

where

$$\lambda_h = \gamma_h \gamma_g - \gamma_g = -B^{-1} \frac{D^+ h}{h}$$

(5.2)
and the operator $\overline{D}$ is defined by
\[ g\overline{D}^*(h) = D^*(hg) - hD^*(g). \]
A direct calculation yields
\[ \overline{D}^* h = -A\nabla h. \]
Note that $\overline{D}^*$ is independent of $g$; hence $\lambda_h$ is as well.

Suppose $\lambda_h$ is given by (5.2). Define $Q = A'B^{-1}A$. Then a further direct calculation yields
\[
R_{\gamma_h} \lambda_h = -\left(2/g\right)\overline{D}(gB^{-1}D)h - \left(\frac{\nabla h}{h}\right)' \cdot A'B^{-1}A \left(\frac{\nabla h}{h}\right)
= \left(\frac{2}{gh}\right)\nabla \cdot (gQ\nabla h) - \left(\frac{\nabla h}{h}\right)' \cdot Q \left(\frac{\nabla h}{h}\right). \tag{5.3}
\]
(If $h$ is not continuously differentiable, etc., then the expression in (5.3) should be interpreted in an appropriate generalized sense as in (4.7).)

The first assertion of the next theorem is immediate from (5.3) and the remarks preceding the definition of admissibility; and the brief proof of the second assertion will be given in Section 14.

Theorem 2. Let $\gamma_g \in \mathcal{F}$. Suppose $h > 0$ (a.e. $(dx)$) is continuously differentiable
\[
\nabla \cdot (gQ\nabla h) \leq 0.
\tag{5.4}
\]
Then $\gamma_g$ is inadmissible and $\gamma_h$ is better than $\gamma_g$.

Let $\mu \in \mathbb{R}_h$, $\tau > 0$. Suppose $h$ is non-constant, continuously differentiable on $\{x: \|x - \mu\| > \tau\}$, and satisfies
\[
\nabla \cdot (gQ\nabla h) \leq 0 \text{ for } \|x - \mu\| > \tau \tag{5.5(i)}
\]
\[ h(x) = 1 \text{ for } \|x - \mu\| \leq \tau \tag{5.5(ii)} \]
\[ 0 < h \leq 1. \tag{5.5(iii)} \]
Then $\gamma_h$ is better than $\gamma_g$.

Note that the left hand side of (5.5(i)), viewed as an operator on $h$, is a self adjoint elliptic operator on $\mathbb{R}^k$. Hence the problem of constructing $h$ satisfying (5.5) is a familiar type of exterior Dirichlet problem.

The following two further basic results are proved in Part IV. These serve to characterize the admissible functions. Note, in particular, that subject (perhaps) to the smoothness condition in Theorem 4 the existence of $h$ satisfying (5.5) is necessary as well as sufficient for $\gamma_g \in \mathcal{F}$ to be inadmissible.

Theorem 3. Let $\gamma = \gamma_g \in \mathcal{F}$, with $g$ bounded on $\{x: \|x\| \leq 1\}$. Then $\gamma$ is admissible if and only if there is a sequence $\{h_i\}$ of non negative functions such
that

\[ h_i(x) = 1 \quad \|x\| \leq 1 \quad (5.6(i)) \]

\[ h_i(x) = 0 \quad \|x\| \geq i \quad (5.6(ii)) \]

\[ ((\nabla h_i)'Q(\nabla h_i), g) \to 0 \text{ as } i \to \infty. \quad (5.6(iii)) \]

Any sequence satisfying (5.6) must satisfy

\[ \lim_{i \to \infty} h_i(x) = 1. \quad (5.7) \]

If there is a sequence satisfying (5.7), (5.6(iii)), and \( \lim_{\|x\| \to \infty} h_i(x) = 0 \) then \( \gamma_g \) is admissible. (In (5.6) one may replace the set \( \{x: \|x\| \leq 1\} \) by any other bounded open set.)

**Theorem 4.** Let \( \mu, r \) be as in Theorem 2. Suppose \( g > 0 \), \( g \) is bounded on \( \{x: \|x - \mu\| < r\} \), and the coefficients of \( Q \) and \( g \), and their derivatives, are Holder continuous on \( \|x - \mu\| > r \). The following conditions are then each necessary and sufficient for inadmissibility of \( \gamma = \gamma_g \in F \)

(a) The existence of a function \( h \) satisfying conditions (5.5) of Theorem 2.

(b) The existence of a function \( h \) satisfying (5.5(i)), (5.5(ii)) and

\[ \liminf_{\|x\| \to \infty} h(x) = 0. \]

(c) The non-existence of a function \( h \) satisfying (5.5(i)), (5.5(ii)), and

\[ \lim_{\|x\| \to \infty} h(x) = \infty. \]

In (a), (b), (c) above, the requirement may also be added that equality holds in (5.5(i)). Furthermore, if any of (a), (b), (c) are satisfied for some \( \mu \in \mathbb{R}^d \), \( r > 0 \) then they will all be satisfied for any \( \mu \in \mathbb{R}^d \), \( r > 0 \) for which the smoothness conditions on \( Q, g \) are satisfied.

For the definition of Holder continuity see, for example, Dynkin (1965, Section 5.22). The assumption of Holder continuity is surely stronger than needed for the validity of the assertions of Theorem 4. It may even be that some, or all, of these assertions remain valid with no supplementary conditions placed on \( Q \) and \( g \) provided the operator in (5.5(i)) is interpreted in the suitable generalized fashion.

Here is a useful corollary to these results; the first part of which is only a minor extension of Theorem 2. This corollary is also proved in Part IV.

**Corollary 1.** Suppose \( h \) satisfies (5.5). Then for \( 0 < c \leq 2, \, \gamma_g + c\lambda_h \) dominates \( \gamma_g \), where \( \lambda_h \) is defined in (5.2). If, also,

\[ \lim_{\|x\| \to \infty} h(x) = 0 \]

and \( \|\nabla h\| \) is locally bounded then \( \gamma_h c_g \) is admissible for \( c \geq 1 \).

(The boundedness condition on \( \|\nabla h\| \) in the above is convenient but much
stronger than necessary.)

6. DISCUSSION.

According to the preceding results the admissible vector-valued functions \( \gamma \) correspond to the real-valued functions \( g \) for which \( \gamma_g \) is admissible. Furthermore, admissibility of \( \gamma_g \) depends on \( A \) and \( B \) only through \( Q = A'B^{-1}A \). In particular the coefficients \( \{ a_i \} \) do not affect the admissibility of \( \gamma_g \) for given \( g \) although they do, of course, affect its form.

The above comments, and other simple consequences of the preceding theorems are contained in the following corollary, whose proof is left to the reader.

**Corollary 2.** Admissibility of \( \gamma_g \in \mathcal{F} \) depends on \( A, B \) only through \( Q \). Consider two operators of the form (4.1) with respective matrices \( Q_1 \) and \( Q_2 \) and functions \( g_1 \) and \( g_2 \). Suppose \( \max \{ \operatorname{eig}(g_2 Q_2) Q_1 \} \leq K \) for some constant, \( K < \infty \). If \( \gamma_{g_1} \) \((\gamma_{g_2}, \text{resp.})\) is admissible (inadmissible, resp.) in the first (second) problem then it is admissible (inadmissible) in the other.

For example, suppose \( Q \) is fixed and \( 0 < g_1 \leq g_2 \) (a.e.). If \( \gamma_{g_2} \) is admissible so is \( \gamma_{g_1} \); and if \( \gamma_{g_1} \) is inadmissible so is \( \gamma_{g_2} \).

Suppose \( \gamma \) is of the form (5.1) (i.e., \( \gamma = \gamma_g \) for some \( g \) and is inadmissible). Then - subject (possibly) to smoothness conditions on \( g \) - an improvement \( \gamma_{g,h^2} \) can be found by solving the appropriate partial differential equation for \( h \) (5.5(i)), with equality). On the other hand if \( \gamma \) is not of the form (5.1) then Theorem 1 states that \( \gamma \) is inadmissible but neither the Theorem nor its proof provides a similarly convenient formula enabling one to write down an improvement on \( \gamma \). This is a defect of the current presentation, as well as of the statistical theorems which parallel Theorem 1.

7. EXAMPLES.

The most important and most familiar example involves the situation where \( Q = I \).

**Example 1.** Suppose \( a_1 \equiv 0 \) and \( A = B = I \). Then

\[
\gamma_g = \frac{\nabla g}{g} = \nabla (\ln g).
\]

The basic equation (5.5(i)) for the exterior problem of Theorem 2 is

\[
\nabla \cdot (g \nabla h) = 0.
\]

This can also be rewritten as the uniformly elliptic equation

\[
\nabla^2 h + \frac{\nabla g}{g} \cdot \nabla h = 0. \tag{7.1}
\]

If \( g \equiv 1 \) then \( \gamma_g = 0 \), and inadmissibility of \( 0 \) is thus equivalent to non-trivial
bounded solubility of the exterior Dirichlet problem
\[ \nabla^2 h = 0 \text{ for } \|x\| > 1, \quad h = 1 \text{ for } \|x\| = 1. \]
Consequently 0 is admissible here if and only if \( k = 1 \) or 2. (see e.g. Brown (1971).)

More generally, if \( g \) is spherically symmetric an explicit solution of (7.1) yields that \( \gamma_0 \) is inadmissible if and only if
\[ \int_0^\infty (r^{k-1} G(r))^{-1} dr < \infty \quad (7.2) \]
where \( g(x) = G(\|x\|) \). See e.g. Meyers and Serrin (1960), Brown (1971, p.898-900), or Orey (1977). If \( \gamma_0 \) is inadmissible then
\[ h(x) = \int_0^\|x\| (r^{k-1} G(r))^{-1} dr \text{ for } \|x\| > r \]
\[ = 1 \text{ for } \|x\| \leq r \]
is a solution to the exterior Dirichlet problem. Hence by Corollary 1 for \( 1 \leq c \leq 2 \)
\[ \gamma_0 \epsilon(x) = \gamma_0(x) - \frac{cx}{\|x\|^k g(x) h(x)} \text{ if } \|x\| > r \]
\[ \gamma_0(x) \text{ if } \|x\| \leq r \]
is an admissible function dominating \( \gamma_0 \).

When \( g \equiv 1 \) then \( \gamma_0 \equiv 0 \), and letting \( r \to 0 \) in the above (which can be justified) yields
\[ \gamma(x) = -c \frac{(k-2)x}{\|x\|^2} \quad k \leq 3, \quad 1 \leq c < 2 \quad (7.3) \]
as an admissible function dominating \( \gamma_0 \equiv 0 \).

If, instead of \( A = B = I \), and \( a_i \equiv 0 \) one has only \( A'B^{-1}A = I \), or even \( 0 < \epsilon < \text{mineig} A' B^{-1} A \leq \text{maxeig} A' B^{-1} A < \epsilon^{-1} \), then the above statements concerning admissibility remain valid, except that the formulae for \( \gamma_0 \) and \( \gamma_0 \epsilon \) must be modified in accordance with (5.1), (5.2).

If \( g \) is not spherically symmetric then it may be more difficult to determine whether \( \gamma_0 \) is admissible. There is a test due to Meyers and Serrin (1960) for solubility of the exterior Dirichlet problem valid for elliptic operators more general than those of the form (5.5(i)). In the case at hand this test is weaker than a combination of Corollary 2 and the preceding test for spherically symmetric functions, \( g \). More general tests for the case \( A = B = I \) are described in Brown (1971, Theorems 6.4.2 to 6.4.4). (Substitute \( g(x) \) of the current manuscript for \( f^*(x) \) in (6.4.1), op.cit. The uniform boundedness condition in Theorem 6.4.3 (op.cit.) is unnecessary for the current application.)

In Brown (1980a) we encountered two interesting examples of (1) with \( a_i \equiv \)}
DIFFERENTIAL INEQUALITY OF AN ESTIMATION PROBLEM

0, and \( A = B \) non constant.

**Example 2.** Let \( a_i = 0 \) and let \( A = B \) be diagonal with diagonal elements \( b_{ii} = e^{\alpha x_i} \) where \( \alpha \) is an arbitrary scalar. (Unboundedness of the coefficients in \( A = B \) could be avoided by multiplying both by \( (\Sigma e^{\alpha x_i})^{-1} \).) The expression (5.5), defining the exterior problem can be simplified by letting
\[
p(x) = e^{\alpha x_i} g(x)
\]
for it then becomes simply
\[
\Sigma \xi_i(x) \frac{\partial}{\partial x_i} \left( p \frac{\partial}{\partial x_i} h \right) = 0
\]
with \( \xi_i(x) = \exp(-\alpha \sum_{j \neq i} x_j) \). Also,
\[
(\gamma_h)i = \xi_i \frac{\partial}{\partial x_i} p.
\]

A particularly interesting case is that for \( p = 1 \), for which \( \gamma_h \equiv 0 \). In dimension \( k = 1 \) the problem is, of course, identical with that in Example 1, and so \( \gamma_h \equiv 0 \) is admissible. However \( \gamma_h \equiv 0 \) is inadmissible for \( k \geq 2 \). To see this let
\[
h(x) = (\Sigma e^{-\alpha x_i})^{-1}.
\]
Then \( h \) is a solution to (7.4) on all of \( R^k \), and inadmissibility of \( \gamma_h \) follows from Theorem 2. Here \( (\gamma_{h})(i) = (\sum h)_i = \alpha (k - 1)e^{-\alpha x_i}/(\Sigma e^{-\alpha x_i}) \). An unusual feature of this problem (in addition to the fact that inadmissibility of \( \gamma_h \equiv 0 \) occurs when \( k = 2 \), rather than when \( k = 3 \) as in Example 1) is that the dominating function \( \gamma_{h} \) points toward the vector \((\text{sgn} \alpha)\cdot \infty \) rather than towards \( 0 \) (or some other fixed \( \mu \in R^k \)) as was the case throughout Example 1.

Corollary 1 does not apply to \( \gamma_{h} \) since \( \lim_{\|x\| \to \infty} h(x) \) does not exist. However, it is easy to see that \( c \gamma_{h} = c \gamma_{h} \) dominates \( \gamma_{h} \) for \( 0 < c < 2 \). In fact, \( R_0((1 + \beta) \gamma_{h}) = R_0((1 - \beta) \gamma_{h}) \) for \( \beta > 0 \). Hence \( c \gamma_{h} \) is actually inadmissible for \( c \neq 1 \). For \( c = 1 \), \( c \gamma_{h} = \gamma_{h} \) is admissible. To see this consider the system (5.6) for the procedure \( \gamma_{h} \). The left side of equation (5.6(iii)) becomes
\[
\sum_i e^{a x_j} e^{-\Sigma a x_i} (\Sigma e^{-a x_i})^{-1} k + 1 \left[ \frac{\partial h_i}{\partial x_j} \right]^2.
\]
When \( k = 2 \) the coefficient of \( \frac{\partial h_i}{\partial x_j} \) satisfies \( e^{a x_j} e^{-\Sigma a x_i} (\Sigma e^{-a x_i})^{-1} k + 1 \)
\[
= |\sum e^{\alpha (x_i - x_j)}|^{-1} \leq 1.
\]
It follows via Theorem 3 that since \( \gamma \equiv 0 \) is admissible for \( k = 2 \) in Example 1 it must also be admissible here. When \( k \geq 3 \) the relevant coefficients are not bounded. However the resulting form of (5.6(iii)) is qualitatively similar to the variational problem considered in Brown (1979a, p.988-990). The functions \( h_i \) used there may be adapted to prove admissibility. (The admissibility of \( \gamma_{h} c \) if and only if \( c = 1 \) should carry over to the parallel statistical contexts in Brown (1980b) and Berger (1980); but was not anticipated.
in those manuscripts. See also Das Gupta (1986).

**Example 3.** Let \( a_i \equiv 0 \) and let \( A = B \) be diagonal with diagonal elements \( b_{ii} = (1 + x_i^2)^{r/2} \). (Again, the equation can be multiplied throughout to avoid unboundedness of the coefficients.) Here, set \( p(x) = (\pi(1 + x_i^2)^{r/2})g(x) \). Then

\[
(\gamma_i)_i = \hat{\gamma}_i \frac{\partial}{\partial x_i} p
\]

with \( \hat{\gamma}_i(x) = \pi_j \phi_j(1 + x_j^2)^{-r/2} \). The expression (14) becomes

\[
\Sigma \hat{\gamma}_i(x) \frac{\partial}{\partial x_i} (p \frac{\partial}{\partial x_i} h) = 0.
\]

Again the case \( p \equiv 1 \), for which \( \gamma_i \equiv 0 \), is of particular interest. This function is of course admissible when \( k = 1 \), as in Example 1. Further, if \( 0 \leq r < 1 \) and \( 2 \leq k \leq (2 - r)/(1 - r) \) then \( \gamma_i \equiv 0 \) is also admissible; and if \( r \geq 1 \), \( \gamma_i \equiv 0 \) is admissible for any \( k \). Otherwise \( \gamma_i \equiv 0 \) is inadmissible. Thus, the dividing dimension between admissibility and inadmissibility can here be any value from 1 to \( \infty \) (including \( \infty \)), depending on the choice of \( r \).

The proof of the above assertions can conveniently be accomplished via Theorem 4. Suppose \( 0 \leq r < 2 \) and \( k < (2 - r)/(1 - r) \). Then \( \alpha = 2 - r > 1 \) and \( s = (k\alpha - k - \alpha)/\alpha > 0 \). Let \( h = (\Sigma(1 + x_i^2)^{\alpha/2})^{-r} \). Then,

\[
\nabla \cdot (gQ \nabla h) = -\frac{s\alpha}{\pi(1 + x_i^2)^{r/2}}
\]

\[
\times \sum \left[ 1 - \frac{r x_i^2}{1 + x_i^2} \right] \frac{\Sigma(1 + x_i^2)^{\alpha/2} - \alpha(1 + x_i^2)^{\alpha/2}}{(\Sigma(1 + x_i^2)^{\alpha/2})^{r+2}}
\]

\[
< -\frac{s\alpha}{\pi(1 + x_i^2)^{r/2}} \frac{k(1 - r) - \alpha(s + 1)}{\Sigma(1 + x_i^2)^{s+1}} < 0
\]

as in Berger (1980). Hence \( \gamma_i \) is inadmissible; and, by Corollary 1, \( \gamma_i + c\lambda h \) dominates \( \gamma_i \) for \( 0 < c \leq 2 \) and is admissible for \( 1 \leq c \leq 2 \).

Conversely, suppose \( 2 \geq \alpha = 2 - r > 1 \) and \( k \leq (w - r)/(1 - r) \). Let \( h = \ell \Sigma(1 + x_i^2)^{\alpha/2} \). Then

\[
\nabla \cdot (gQ \nabla h) = \frac{\alpha}{\pi(1 + x_i^2)^{r/2}}
\]

\[
\times \sum \left( 1 - \frac{r x_i^2}{1 + x_i^2} \right) \frac{\Sigma(1 + x_i^2)^{\alpha/2} - \alpha x_i^2(1 + x_i^2)^{\alpha/2}}{(\Sigma(1 + x_i^2)^{\alpha/2})^2}
\]

\[
\leq \frac{\alpha}{\pi(1 + x_i^2)^{r/2}} \sum \frac{x_i^2((1 - r)\Sigma(1 + x_i^2)^{\alpha/2} - \alpha(1 + x_i^2)^{\alpha/2}}{(\Sigma(1 + x_i^2)^{\alpha/2})^2}
\]
\[
\frac{\alpha \left( \frac{\Sigma_i \frac{x_i^2}{1+x_i^2}}{1+\frac{x_i^2}{1+x_i^2}} \right) (1 - r - (\alpha/k)\Sigma(1 + x_i^2)^{\alpha/2})}{\pi(1 + x_i^2)^{\alpha/2}(\Sigma(1 + x_i^2)^{\alpha/2})^2} 
\leq 0.
\]

(7.5)

Admissibility of \( \gamma \) follows from Theorem 4. (Brown (1980b) contains a different proof, utilizing essentially Theorem 3, and omits the case \( k - (2r)/(1 - r) \).)

The diagonal terms of \( gQ \) are \( g(x) \). Note that these terms decrease in \( r \) for fixed \( x \). The assertions involving \( r < 0 \) and \( r \geq 1 \) therefore follow from Corollary 2 and the above.

III. STATISTICAL APPLICATIONS

8. RELATION TO STATISTICS.

The considerations of Part II were already motivated by the particular statistical application described in Sections 1 and 2 of Part I. In this section we describe a somewhat more general statistical motivation for the differential inequality development of Part II. We will return to the normal distribution example of Part I at the end of this section and in Section 9.

In the statistical setting a random variable taking value \( x \in \mathbb{R}^k \) is to be observed. This variable has probability density \( p_\theta(x) \). The value of the parameter \( \theta \in \mathbb{R}^k \) is unknown, and is to be estimated. The estimation rule is specified by a measurable function \( \delta: \mathbb{R}^k \to \mathbb{R}^k \); where \( \delta(x) \) describes the estimate of \( \theta \) to be made if the value \( x \) is observed. The discrepancy between the estimate \( d = \delta(x) \) and the true (but unknown) value \( \theta \) is measured by the loss \( (d - \theta)'B(\theta)(d - \theta) \) with \( B(\theta) \) as in (4.1). Consequently for given estimation rule, \( \delta(\cdot) \), one examines the expected loss under \( \theta \):

\[
\mathcal{R}(\theta, \delta) = \int (\delta(x) - \theta)'B(\theta)(\delta(x) - \theta)p_\theta(x)dx.
\]

\( \mathcal{R}(\cdot, \delta) \) is called the risk function for \( \delta \). One says that \( \delta \) is inadmissible if there exists an estimation rule \( \delta' \) such that

\[
\mathcal{R}(\theta, \delta') \leq \mathcal{R}(\theta, \delta) \text{ for all } \theta \in \mathbb{R}^k,
\]

with strict inequality for some \( \theta \in \mathbb{R}^k \). Otherwise, \( \delta \) is admissible.

Define \( \gamma(x) = \delta(x) - x \). Consider a second estimator, \( \delta' \), defined by \( \delta'(x) = x + \gamma(x) + \lambda(x) \). Let \( \mathcal{R}_\gamma \lambda: \mathbb{R}^k \to \mathbb{R}^1 \) be defined by \( \mathcal{R}_\gamma \lambda(\theta) = \mathcal{R}(\theta, \delta') - \mathcal{R}(\theta, \delta) \).

Algebraic manipulation (adding and subtracting \( \mathcal{R}(\theta, x) \)) yields

\[
\mathcal{R}_\gamma \lambda = 2\lambda + \epsilon ((2\gamma + \lambda)'B\lambda)
\]
where the (integral) operators $\mathcal{D}$ and $\mathcal{E}$ are defined by
\[
\mathcal{D} \lambda(\theta) = \int \lambda'(x) B(\theta)(x - \theta) p_\theta(x) \, dx
\]
\[
\mathcal{E} u(\theta) = \int u(x) p_\theta(x) \, dx.
\]
(Hence $\mathcal{E}((2\gamma + \lambda)' B \lambda = \int (2\gamma(x) + \lambda(x))' B(\theta) \lambda(x) p_\theta(x) \, dx$.)

Suppose $\lambda$ is differentiable, and write the Taylor expansion
\[
\lambda_i(x) = \lambda_i(\theta) + \left( \sum_j \frac{\partial \lambda_i(\theta)}{\partial x_j} \cdot (x - \theta)_j \right) + \rho, \quad i = 1, \ldots, k,
\]
where here (and later) $\rho$ is a generic symbol used to denote the appropriate remainder (error) term. Then
\[
\mathcal{D} \lambda = \sum_i a_i \lambda_i + \sum_{i,j} a_{ij} \frac{\partial \lambda_i}{\partial x_j} + \rho
\]
where
\[
a_i(\theta) = \int (B(\theta)(x - \theta))_i p_\theta(x) \, dx
\]
\[
a_{ij}(\theta) = \int (B(\theta)(x - \theta))_i (x - \theta)_j p_\theta(x) \, dx.
\]
Consequently
\[
\mathcal{D} \lambda \approx D \lambda
\]
with $D$ defined in (4.2) and $\{a_i\}, \{a_{ij}\}$ as above. Keeping only the leading term of the Taylor expansion of $u$ one may write
\[
\mathcal{E} u \approx u.
\]

To this degree of approximation, then
\[
\mathcal{R}_\gamma \lambda \approx R_\gamma \lambda \tag{8.1}
\]

with $R_\gamma \lambda$ defined in (4.4). (These approximations are not particularly good, but they seem to suffice if, for example, $p_\theta(x) \leq w(||x - \theta||)$ where $w$ is a sufficiently rapidly decreasing function, and if $\lambda$ is nearly linear in the sense that $|\frac{\partial^2 \lambda(x)}{\partial x_i \partial x_j}|$ is small compared to $||\nabla \lambda(x)||$, $1 \leq i, j \leq k$. Solutions to $R_\gamma \lambda \leq 0$ seem empirically to frequently have this property, at least asymptotically as $||x|| \to \infty$. See Brown (1979a), (1980b).)

According to the above definition $\delta = x + \gamma$ is inadmissible if there exists a $\lambda$ such that $R_\gamma \lambda \leq 0$ but $R_\gamma \lambda \neq 0$. Consequently, to examine for inadmissibility of $\delta$ it makes sense to look for non-trivial solutions to the differential inequality $R_\gamma \lambda \leq 0$ - in other words to check whether $\gamma$ is inadmissible under $R_\gamma$. 
The operators dual to \( D \) and \( \varepsilon \) are, respectively,
\[
D^* g = \int B(\theta)(x - \theta) g(\theta) p_\theta(x) d\theta
\]
and
\[
\varepsilon^* g = \int g(\theta) p_\theta(x) d\theta.
\]
Suppose \( g \) is a given probability density with respect to Lebesgue measure. The expectation under \( g \) of \( \mathcal{R}(\cdot, \delta) - \mathcal{R}(\cdot, x) = \mathcal{R}_0 \) is
\[
(\mathcal{R}_0 \gamma, g) = 2(\gamma, D^* g) + (\gamma^T B \gamma, \varepsilon^* g).
\]
Consequently \((\mathcal{R}_0 \gamma, g)\) is minimized with respect to \( \gamma \) by the choice
\[
\gamma^*_g = -\frac{D^* g}{\varepsilon^* g}.
\]
The procedure \( \delta^*_g(x) = x + \gamma^*_g(x) \) is called in statistical terminology the Bayes procedure corresponding to prior density \( g \). It is plausible from the above that
\[
D^* g \approx D^* g \text{ and } \varepsilon^* g \approx g
\]
so that
\[
\gamma^*_g \approx \gamma_g,
\]
where \( \gamma_g \) is defined in (4.9).

When \( \gamma^*_g \) is given by (8.2) for \( g \) an arbitrary (nonnegative) function rather than a probability density, it is then called a generalized Bayes procedures. The definitions of \( D^* g, \varepsilon^* g, \) and \( \gamma^*_g \) can, and should, naturally be generalized to include non negative measures in place of the densities, \( g \).

Phrased in statistical language, Theorem 1 states that the generalized Bayes procedures relative to the operator \( \mathcal{R}_0 \) include all admissible procedures. J. Sacks (1963) first proved a theorem having such a conclusion in certain one dimensional statistical settings. Other “Sacks” type theorems for the statistical problem can be found in Farrell (1966), Brown (1971), Berger and Srinivasan (1977), and Brown (1986). All of these theorems require fairly specific conditions on the statistical problem - none approach the generality of Theorem 1. Thus Theorem 1 motivates the conjecture that Sacks type theorems should be valid much more generally than has so far been established.

As will be explained in Part IV, Theorem 3 is a form of the Stein-Le Cam necessary and sufficient condition for admissibility in statistics. Theorem 4 provides a convenient test for admissibility of \( \gamma_g \) under \( \mathcal{R}_0 \). To the extent that the approximations (8.1) and (8.3) are valid this same test should be applicable to test admissibility of \( \gamma^*_g \) under \( \mathcal{R}_0 \). Portions of the indicated test for admissibility under \( \mathcal{R}_0 \) were suggested heuristically in much more detail in Brown (1979a) and again in Brown (1985). (Brown (1979a) also applies in a
much more general statistical framework than that described above.)

Sections 1 and 2 have already mentioned the special problem where $p_\theta$ is normal with mean $\theta$, covariance $I$, and $B = I$. In that case

$$\mathcal{R}_0 = \mathbb{E}_0(R_0\gamma)$$

and

$$\mathcal{R}(\theta, x + \gamma) = \mathbb{E}_0(R_0\gamma) + k$$

with $A = B = I$ so that $R_0\gamma = 2\nabla \cdot \gamma + \|\gamma\|^2$. (See Stein (1973).) Consequently, inadmissibility of $\gamma$ under $R_0$ rigidly implies inadmissibility of $\gamma$ under $R_0$. The converse is also valid when $\gamma = \gamma_0^*$ so long as $\gamma_0^*$ is bounded (and somewhat more generally). See Brown (1971) and Srinivasan (1981). However, an example in Brown (1979b) shows that this converse result is not valid without some sort of condition on $\gamma_0^*$. This example, therefore, also demonstrates that there are limits to the heuristics suggested following (8.1) and (8.3) even in the vest of statistical problems.

9. JAMES-STEIN ESTIMATION.

Suppose in the setting of the previous section that $p_\theta$ is normal with mean $\theta$, covariance $I$, and that $B = I$. The maximum likelihood estimator is, of course, $\delta_0(x) = x$. According to (1.1) and (7.3) a better estimator is $\delta(x) = (1 - e^{\frac{k-2}{2}})x, 1 \leq x < 2, k \geq 3$, which is the James-Stein (1960) estimator. Since $1 - e^{\frac{k-2}{2}}$ can be negative it can be shown that an even better estimate is

$$\delta^+(x) = \left(1 - e^{\frac{k-2}{2}}\right)^+x = x + \gamma^+(x), 1 \leq x < 2, k \geq 3,$$

the positive part James-Stein estimator, as in (1.2). (See, e.g., Brown (1986 Exercise 4.11.1).)

As previously noted, for this problem the operator $R_0$ is given by $R_0\gamma = 2\nabla \cdot \gamma + \|\gamma\|^2$. Hence $\gamma_0 = -\frac{D\gamma}{\partial x} = -\frac{\partial \gamma}{\partial x}$ as in (5.1). Let $g(x) = G(|x|)$ where

$$G(r) = \begin{cases} 
  e^{-r^2/2} & \text{if } 0 \leq r \leq c(k-2) \\
  c(k-2)e^{-c(k-2)^2/2} & \text{if } r > c(k-2) 
\end{cases}$$

Then $\gamma_0 = \gamma^+ = -x \min(\frac{c(k-2)}{\|x\|^2}, 1)$, as in (9.1).

(As already noted in Section 1, $R_0\gamma^+ < 0$ when $k \geq 3$ which establishes that $\delta_0(x) = x$ is then inadmissible as a statistical estimator of $\theta$.)

It is known that $\delta^+$ is itself inadmissible as a statistical estimator since it is not an analytic function. (See Brown (1971) or (1986).) However, the function $\gamma^+$ is admissible by (7.2) for every $k \geq 3$ relative to the operator $R_0$. This means that there does not exist a function $\lambda$ such that $R_0\gamma^+ + \lambda - R_0(\gamma^+ + \lambda) > 0$. Consequently, it is impossible to find an estimator $\delta = \delta^+ + \lambda$ whose unbiased estimate of risk is always less than that of $\delta^+$. In short, the customary method for finding an estimator dominating $\delta^+$ cannot succeed.
Differential Inequality of an Estimation Problem

10. Estimation After Sequential Stopping.

Let \( \{Z(t)\} \) be a Brownian motion with constant drift \( \mu \) and variance one per unit time, so that \( Z(t) \sim N(\mu t, t) \). Let \(-\infty < a < 0 < b < \infty, 0 < m \leq \infty, \) and let \( T \) denote the random stopping time

\[
T = \inf \{ t : Z(t) \geq b \text{ or } Z(t) \leq a \text{ or } t \geq m \}. \tag{10.1}
\]

Thus \( Z(t) \) is the value of the process after stopping according to a truncated or untruncated sequential probability ratio test. Suppose it is desired to estimate \( \mu \) on the basis of the sufficient statistics \( (T, Z(T)) \) under ordinary quadratic loss \( L(\mu, d) = (d - \mu)^2 \).

Given that \( (T, Z(T)) = (t, z) \), the likelihood function is proportional to \( \varphi(t^{1/2}(\mu - z/t)) \), where \( \varphi \) denotes the standard normal density function. Let \( g = 1 \) be the uniform prior density. The generalized Bayes procedure is then \( \delta_1^*(t, z) = z/t \). In what follows \( (z/t) \) will be identified as \( \delta_0 \) so that \( \gamma_1^* = 0 \).

Define \( w = z/t \) and

\[
h(w) = \frac{1}{m} \text{ if } a/m < w < b/m
\]

\[
= \frac{w}{a} \text{ if } w \leq a/m
\]

(with the obvious conventions when \( m = \infty \)). Note that since \( (T, Z(T)) + (t, z) \)

\( h(w) = t^{-1} \) because of (10.1).

Now consider an arbitrary prior density, \( g \). The generalized Bayes procedure \( \delta_0^* = (z/t) + \gamma_0^* \) satisfies

\[
\gamma_0^*(t, z) = \frac{\int (\mu - z/t) \varphi(t^{1/2}(\mu - z/t)) g(\mu) d\mu}{\int \varphi(t^{1/2}(\mu - z/t)) g(\mu) d\mu}
\]

\[
= \frac{\tilde{g}'(z/t)}{t \tilde{g}(z/t)} \tag{10.2}
\]

where

\[
\tilde{g}(w) = \int t^{-1/2} \varphi(t^{-1/2}(\mu - w)) g(\mu) d\mu
\]

\[
= \int h^{1/2}(w) \varphi(h^{1/2}(\mu - w)) g(\mu) d\mu.
\]

(Formula (10.2) is verified by integrating by parts in the numerator.) Hence (10.2) may be rewritten in terms of the single variable \( w \) as

\[
\gamma_0^*(w) = -\frac{h(w) \tilde{g}'(w)}{\tilde{g}(w)}. \tag{10.2'}
\]

Next note that \( t^{-1/2} \varphi(t^{-1/2}(\mu - w)) \) is a probability density in \( \mu \), and that for large values of \( |w| \), corresponding to small values of \( t \), this density is highly
concentrated about its mean \( w \). Hence for sufficiently smooth functions \( g \)
\[
\tilde{g}(w) \sim g(w) \text{ and } \tilde{g}'(w) \sim g'(w) \text{ as } \vert w \vert \to \infty.
\]
(10.3)

(More information about asymptotic approximations like (10.3) is contained in Brown (1971) and Brown (1979a).)

We can thus write
\[
\gamma_0^+ \approx \frac{k(w)g'(w)}{g(w)},
\]
(10.4)

(The approximation in (10.4) is asymptotically accurate as \( \vert w \vert \to \infty \) and may be poor for small \( \vert w \vert \).) It is suggested in (8.2) and (8.3) that \( \gamma_0^+(w) \approx \frac{D^* g(w)}{g(w)} \).

Consequently one considers the parallel question of admissibility of the function \( \gamma = 0 \) relative to the operator \( R_0 \) associated with \( D^* g = -hg' \). According to the standard formula (4.5) this means \( D \gamma = (h \gamma)' \), and \( R_0 \gamma = 2D \gamma + \gamma^2 \) as in (4.1).

\( \gamma_0 = 0 \) is inadmissible relative to this operator. For example, take \( B > 3b^2/2m^2 \) and
\[
\gamma(w) = \begin{cases} 
\frac{w^2}{b(b + w^2)}w & \geq b/m, \\
\frac{mw}{(B + w^2)a/m < w < b/m,} \\
\frac{w^2}{a(B + w^2)}w & \leq a/m.
\end{cases}
\]
(10.5)

Then
\[
R_0(\gamma) = 2(h \gamma)' + \gamma^2 = \frac{-w^2}{b^2(B + w^2)}w < b/m
\]
and
\[
R_0(\gamma) = 2(h \gamma)' + \gamma^2 = \frac{-2Bm^2 + 3m^2w^2 - a/m < w < b/m,}{}
\]
\[
\frac{-w^2}{a^2(B + w^2)}w < a/m,
\]
so that \( R_0 \gamma(w) < 0 \) for all \( w \). The function \( \gamma \) in (10.5) is the generalized Bayes function \( \gamma_0 \) corresponding to \( g = (B + w^2)^{-1/2} \) as given by the right side of (10.4).

The preceding facts motivate the conjecture that the statistical estimator \( \delta_1^*(t, z) = z/t \) is inadmissible for this optional stopping problem. This is somewhat surprising since this same estimator is admissible for the fixed sample size problem - i.e., when \( T = m \). (See Elyth (1951).)

One is furthermore led to conjecture that an estimator which statistically dominates \( \delta_1^* \) should be of the form \( \gamma(t) + \gamma^*(z/t) \) with \( \gamma^*(w) \sim \gamma(w) \) as \( \vert w \vert \to \infty \). (In other words, \( \gamma^*(w) \sim -1/b \) as \( w \to \infty \) and \( \delta^*(w) \sim 1/\alpha \) as
Furthermore, such a $\gamma^*$ should be generalized Bayes with respect to a prior density $g^*$, say, which behaves like $g(\theta) = (1 + \theta^2)^{-1/2}$ when $|\theta| \to \infty$. Such a procedure would also be admissible. One cannot conclude without a finger analysis how $\gamma^*(x)$ (or $g^*(\theta)$) should behave for small values of $|x|$ (or $|\theta|$) since the approximation (10.4) is not necessarily good in that region.

It is possible to easily indicate how to prove the preceding inadmissibility conjecture. Let $W = Z/T$. A routine calculation shows that

$$E_{\mu}(W) \sim \frac{\mu + 1/b}{\mu + 1/a} \text{ as } \mu \to -\infty$$

and

$$\text{Var}_{\mu}(W) \sim \frac{\mu/b}{\mu/a} \text{ as } \mu \to -\infty$$

Choose $\gamma^*(x)$ to behave as $|x| \to \infty$ like $\gamma$ in (10.5). A very simple choice for $\gamma^*$ is $\gamma^*(x) = x^1/b$ or $x^1 + (1/a)$ if $x > 0$ or $x < 0$. Then (10.6) plus some simple error bounds yields

$$\mathcal{R}(\mu, \delta^*_1) - \mathcal{R}(\mu, \delta^*_1 + \gamma^*) \sim \frac{-1}{b^2} \text{ as } \mu \to \infty$$

$$= \frac{-1}{a^2} \text{ as } \mu \to -\infty.$$  

(10.8)

It follows from Brown (1980a) that $\delta^*_1$ can be admissible only if it is proper Bayes. By virtue of (10.2') this would imply $\tilde{g}(w) \equiv \text{const.}$ (A small technical point: (10.2) and (10.2') and the following are written only for (generalized) prior densities; however they can easily be extended to prior distributions as needed for this step of the argument.)

Suppose $\tilde{g}(w) \equiv \text{const.}$ Note that

$$H(\mu) = \int h^{1/2}(w) \varphi(h^{1/2}(w)(\mu - w)) dw \to 1 \text{ as } |w| \to \infty.$$  

Hence $\infty = \int \tilde{g}(w) dw = \int g(\mu) H(\mu) d\mu$ which implies $\int g(\mu) d\mu = \infty$. Therefore $g(\mu)$ cannot both be proper and satisfy $\tilde{g}(w) \equiv \text{const.}$ It follows that $\delta^*_1$ is not admissible.

The preceding leaves open the question of finding an admissible estimator dominating $\delta^*_1$. However, it does indicate how such an estimator would behave as $|x| \to \infty$ and it suggests searching among estimators which are generalized Bayes for prior densities satisfying $g(\theta) \sim |\theta|^{-1}$ as $|\theta| \to \infty$. Actually, there should exist prior densities satisfying $g(\theta) \sim |\theta|^{-\beta}$, $1 \leq \beta \leq 2$, yielding estimators which dominate $\delta^*_1$; of these, the ones with $g(\theta) \sim |\theta|^{-1}$ will perform best as $|\theta| \to \infty$.  

11. A Subcompactness Theorem.

Let \( \gamma: \mathbb{R}^k \to \mathbb{R}^k \) be measurable. Define
\[
\gamma_d(x) = \begin{cases} 
\gamma(x) & \text{if } \|\gamma(x)\| \leq d \\
0 & \text{otherwise.}
\end{cases}
\]

Suppose \( \{\gamma_\alpha: \alpha \in A\} \) is a net of measurable functions. Let \( S_d = \{x: \|x\| < d\} \). Then there is a subnet \( \{\alpha' \in A\} \) and a measurable function \( \gamma \) such that
\[
\gamma_{\alpha'} \to \gamma \text{ weak* in } L_2(\chi_{S_d})
\]
for every \( d \in \mathbb{R} \) where \( \chi \) denotes the indicator function. It follows also that
\[
\gamma_{\alpha'}(\varphi) \to \gamma(\varphi) \text{ weak* in } L_2(\varphi) \text{ for all } \varphi \in \Phi^+.
\]
Consequently if \( \varphi \in \Phi^+ \) and
\[
\limsup \|\gamma_\alpha\|_{\varphi} < \infty,
\]
then \( \gamma_{\alpha'} \to \gamma \) weak* in \( L_2(\varphi) \).

We also need

**Lemma 1.** Let \( \varphi \in \Phi^+ \). Suppose \( \gamma \in nL_2(\varphi) \). Then there are positive constants \( C_i, i = 1, 2 \), such that for all measurable \( \lambda \)
\[
(R_{\gamma, \lambda}, \varphi) \geq -C_1 \|\lambda\|_{\varphi} + C_2 \|\lambda\|_{\varphi}^2.
\]

**REMARK.** In view of (11.1) it makes sense to define \( R_{\gamma, \lambda}, \varphi \) = \( \infty \) if \( \gamma \in L_w(\varphi), \lambda \not\in L_2(\varphi) \). This definition is consistent with \( (4.4') \). We also define \( R_{\gamma, \lambda}, \varphi \) = 0 if \( \gamma \not\in L_w(\varphi) \) which is consistent with \( (4.4') \) under the convention \( -\infty - \infty = 0 \).

**PROOF.**
\[
(R_{\gamma, \lambda}, \varphi) = 2\lambda \cdot \left( \frac{D^* \varphi}{\varphi}, \varphi \right) + ((2\gamma + \lambda)^T B \lambda, \varphi)
\]
\[
\geq -2\left( \frac{D^* \varphi}{\varphi}, \varphi \right) \|\lambda\|_{\varphi} + b_1 \|\lambda\|_{\varphi} + b_2 \|\lambda\|_{\varphi}^2
\]
where \( b_1 = \sup \{ \max \text{eig} B(x): \varphi(x) > 0 \} < \infty, b_2 = b^2(\varphi) = \inf \{ \min \text{eig} B(x): \varphi(x) > 0 \} > 0 \).

**Lemma 2.** Let \( \varphi \in \Phi^+ \). Suppose \( \gamma_{\alpha} \to \gamma \) weak* in \( L_2(\varphi) \). Then
\[
\liminf (R_{\gamma_{\alpha}, \varphi}) \geq (R_{0\gamma}, \varphi).
\]

**PROOF.** The lemma follows from the definition \( (4.7) \) of \( (R_{0\gamma}, \varphi) \) and the fact that the map \( \gamma \to (\gamma^T B \gamma, \varphi) \) is weak* lower semicontinuous on \( L_2(\varphi) \).

Let \( T \) denote the space of functions
\[
T = \{ r: \Phi^+ \to (-\infty, \infty]: \exists \gamma \in \Phi^+: r(\varphi) \geq C(\varphi) \}
\]
with \( C(\varphi) = (R_{0\gamma}, \varphi) > -\infty \) as in \( (4.10) \). Give \( T \) the compact (Tychonoff)
topology of pointwise convergence - i.e., $T = X_{\varphi \in \Phi} \cap [C(\varphi), \infty]$. Let

$$\hat{\Gamma} = \{ r \in T : \exists \gamma \in r(\varphi) \geq (R_{0} \gamma, \varphi) \forall \varphi \in \Phi^{+} \}.$$ 

Theorem 5. $\hat{\Gamma}$ is compact.

Proof. Let $\{r_{\alpha}\} \subset \hat{\Gamma}$ be a convergent net in $T$, $r_{\alpha} \to r \in T$. By definition there exists a net $\{\gamma_{\alpha}\}$ such that $(R_{0} \gamma_{\alpha}, \varphi) \leq r_{\alpha}(\varphi)$ for all $\varphi \in \Phi^{+}$. Let $\{\gamma_{\alpha}\}$ be a weak* convergent subnet as discussed prior to Lemma 1, with limit $\gamma$.

If $r(\varphi) < \infty$ then $\limsup (R_{0} \gamma_{\alpha}, \varphi) \leq r(\varphi) < \infty$. By Lemma 1 this implies $\limsup \|\gamma_{\alpha}\varphi\| < \infty$. Hence $\gamma_{\alpha} \to \gamma$ weak* in $L_{2}(\varphi)$. Lemma 2 yields

$$\limsup (R_{0} \gamma, \varphi) \leq \liminf (R_{0} \gamma_{\alpha}, \varphi) \leq \liminf r_{\alpha}(\varphi) = r(\varphi).$$

Hence $r \in \hat{\Gamma}$; which proves that $\hat{\Gamma}$ is compact.

(Parallel arguments in the statistical setting appear in Le Cam (1955).)

It will be useful to augment the collection $\Phi^{+}$. Let $\{\varphi_{i} : i = 1, \ldots \} \subset \Phi^{+}$. Let $\beta_{i} = (1 \vee b_{2}^{-1}(\varphi_{i}))$ with $b_{2}$ as defined in the proof of Lemma 1. Let $\alpha_{i} \geq 0$, $i = 1, \ldots$, and $\varphi = \Sigma \alpha_{i} \varphi_{i}$. Then, for $p = 0, 1$,

$$(D*\varphi)^{p}B^{-p}(D*\varphi) \leq \Sigma \alpha_{i} \varphi_{i} \beta_{i} \left( \frac{D*\varphi_{i}}{\varphi_{i}} \right)^{2} \geq \Sigma \alpha_{i} \varphi_{i} \Sigma \alpha_{i} \beta_{i} \left( \frac{D*\varphi_{i}}{\varphi_{i}} \right)^{2} = \varphi \Sigma \alpha_{i} \beta_{i} \varepsilon_{i},$$

with $\varepsilon_{i} = \left\| \frac{D*\varphi_{i}}{\varphi_{i}} \right\| < \infty$. Suppose $\Sigma \alpha_{i} \beta_{i} \varepsilon_{i} < \infty$, which implies $B^{-1} \frac{D*\varphi}{\varphi} \in L_{w}(\varphi)$ for $p = 0, 1$. As in the proof of Proposition 1 this guarantees that $(R_{0} \gamma_{\varphi}, \varphi) > -\infty$.

Let

$$\Phi^{*} = \{ \varphi : \exists \{ \varphi_{i} \in \Phi^{+} : i = 1, \ldots \}, \{ \alpha_{i} \geq 0 \} \exists \Sigma \alpha_{i} \beta_{i} \varepsilon_{i} < \infty;$$

and $\varphi = \Sigma \alpha_{i} \varphi_{i}$. Let

$$T^{*} = X_{\varphi \in \Phi^{*}} \cap [C(\varphi), \infty]$$

and

$$\hat{T}^{*} = \{ r \in T^{*} : \exists \gamma \in r(\varphi) \geq (R_{0} \gamma, \varphi) \forall \varphi \in \Phi^{*} \}.$$ 

Since $\hat{T}^{*}$ is a continuous image of $\hat{\Gamma}$ it follows directly from Theorem 5 that $\hat{T}^{*}$ is compact.

12. The Admissible Functions.

Consider the set $\hat{\Gamma}$ corresponding to the operator $R_{0}$. A function $\gamma$ is admissible if and only if $(R_{0} \gamma, \varphi)$ is on the lower left boundary of $\hat{\Gamma}$. The following theorem therefore is really just a characterization of lower left boundary points.
of compact subsets of $T$.

The first proof of this result in a statistical context is due to C. Stein (1955). The proof which follows is due to Le Cam (around 1960, unpublished) and appears in Farrell (1966, 1968).

Theorem 6. $\gamma$ is admissible if and only if for every $\varepsilon > 0$ and $\varphi \in \Phi^*$, such that $(R_0\gamma, \varphi) < \infty$ there exists a $g \in \Phi^*$ such that

$$g(x) \geq \varphi(x) \quad \forall x \in \mathbb{R}^k,$$

and

$$(R_\gamma(\gamma_o - \gamma_o), g) = (R_0\gamma_o, g) - (R_0\gamma, g) \geq -\varepsilon. \quad (12.1)$$

$\Phi$ may replace $\Phi^*$ in the above statement.

If there is any $g$ satisfying (12.1) for which $R_\gamma(\gamma_o - \gamma)$ is well defined and satisfies (12.2) then $\gamma$ is admissible.

Proof. Suppose $\gamma$ is not admissible and $(R_0\gamma, \varphi) < \infty$. Then, as noted earlier, there is a $\gamma^{**} = \gamma + \lambda^{**}$ such that $R_\gamma \lambda^{**} \leq 0$ and $(R_\gamma \lambda^{**}, \varphi) = \zeta < 0$. Suppose $g$ satisfies (12.1). Then

$$(R_\gamma(\gamma - \gamma_o), g) \leq (R_\gamma \lambda^{**}, g) = (t_\gamma \lambda^{**}, \varphi) + (R_\gamma \lambda^{**}, g - \varphi) \leq (R_\gamma \lambda^{**}, \varphi) = \zeta < 0.$$ 

It follows that (12.2) fails for all $\varepsilon < \zeta$. This proves the sufficiency of (12.1), (12.2).

Conversely, suppose $\gamma$ is admissible. Fix $\varphi \in \Phi^*$ such that $(R_0\gamma, \varphi) < \infty$. Let $r_\varepsilon \in T^*$ be given by $r_\varepsilon(\cdot) = (R_0\gamma, \cdot) - \varepsilon \chi(\varphi, \cdot)$. Since $r_\varepsilon \notin \hat{T}^*$ it may be separated from $\hat{T}^*$ by a linear functional on $T$. That is, there exists $\{0 \neq \beta_i \in \mathbb{R}, \varphi_i \in \Phi^*, i = 1, \ldots, I\}$ such that

$$\sum_{i=1}^I \beta_i r_\varepsilon(\varphi_i) < \inf_{r \in \hat{T}^*} \Sigma \beta_i r(\varphi_i). \quad (12.3)$$

It follows that all $\beta_i > 0$, for otherwise the right side of (12.3) is $-\infty$. Also, one of the $\varphi_i = \varphi$, for otherwise the left side of (12.3) is $\geq$ the right side. Say $\varphi_1 = \varphi$. Let $g = \beta_1^{-1} \sum_{i=1}^I \beta_i \varphi_i$. Then $g \in \Phi^*$, satisfies (12.1), and

$$(R_0\gamma, g) = \beta_1^{-1} \sum_{i=1}^I \beta_i (R_0\gamma, \varphi_i) > \beta_1^{-1} \sum_{i=1}^I \beta_i r_\varepsilon(\varphi_i)$$

$$= \beta_1^{-1} \sum_{i=1}^I \beta_i (R_0\gamma, \varphi_i) - \varepsilon = (R_\gamma, g) - \varepsilon$$

by (12.3) and the definition of $r_\varepsilon$. This verifies (12.2) and completes the proof. \qed
13. Proof of Theorem 1.

Note that for \( g \in \Phi^* \)
\[
(R_\gamma, \lambda)^* g = (2\lambda, D^* g) + ((2\gamma + \lambda)\gamma^* \lambda, g)
\]
\[
\geq - \left( \left( \gamma + B^{-1} \frac{D^* g}{g} \right)^t B \left( \gamma + B^{-1} \frac{D^* g}{g} \right), g \right)
\]
\[
= (R_\gamma, \lambda, g, g)
\]
(13.1)
where \( \lambda, g = -(\gamma + B^{-1} \frac{D^* g}{g}) = \lambda - \gamma. \)

Suppose \( \gamma \in F \). Then there exists a \( \varphi \in \Phi^* \) such that \( \varphi > 0 \) a.e. \( dx \) and \( R_0 \gamma, \varphi < \infty \). Thus there is a sequence \( \{g_n\} \in \Phi^* \) such that \( g_n \geq \varphi \) and
\[
(\lambda_{\gamma, g_n}, \lambda_{\gamma, g_n}^*, \varphi) \to 0
\]
(13.2)
by (13.1) and Theorem 6. From the definition of \( D^* \)
\[
\lambda_{\gamma, g_n} = \gamma + B^{-1} A^{-1} \nabla g_n = \gamma + B^{-1} A^{-1} \nabla h_n
\]
(13.3)
where \( \gamma = -\gamma - B^{-1} Aa + B^{-1} (\sum_j \frac{g_j}{g_i} - a_{ij}) \) and \( h_n = \lambda^* g_n. \) (13.2) implies that \( \lambda_{\gamma, g_n} \to 0 \) in \( L_2(\Omega S_\delta) \) for every \( d < \infty \). This, in turn, implies that \( \nabla h \to -A^{-1} B\gamma \) strongly in \( L_2(\Omega S_\delta) \). Each \( h_n \) is continuous, and one may assume that \( h_n(0) = 0 \) since addition of a constant to \( h \) leaves (13.3) unaffected. It follows that there is a limiting \( h \), for which \( h_n \to h \) in \( L_2(\Omega S_\delta) \) and \( \nabla h_i \to \nabla h = -A^{-1} B\gamma \) in \( L_2(\Omega S_\delta) \). This yields that \( g = \lambda h \) and \( \gamma = -B^{-1} \frac{D^* g}{g} \), which verifies (5.1) and completes the proof of Theorem 1.


As previously noted, the assertions of the first paragraph of Theorem 2 follow immediately from (5.3). The second paragraph does also, of (5.3) is interpreted in the suitable generalized sense. To see this observe, using (5.3), and Green's theorem, that for \( h \) as in (5.5) and \( \varphi \in \Phi^* \)
\[
(R_\gamma, \lambda h, \varphi) = (R_\gamma, \lambda h, \varphi \chi_{\{h \leq 1 - \epsilon\}}) + (R_\gamma, \lambda h, \varphi \chi_{\{h < 1 - \epsilon\}})
\]
\[
= (R_\gamma, \lambda h, \varphi \chi_{\{h \geq 1 - \epsilon\}}) + \int_{\{\chi h \leq 1 - \epsilon\}} (\nabla \cdot Q \nabla h)(x) \varphi(x) dx
\]
\[
+ \int_{\{\chi h = 1 - \epsilon\}} (Q \nabla h)(x) \cdot n(x) \varphi(x) dx
\]
\[
- \int \left( \frac{\nabla h(x)}{h(x)} \right)' Q(x) \left( \frac{\nabla h(x)}{h(x)} \right) \varphi(x) dx
\]
\[
\leq (R_\gamma, \lambda h, \varphi \chi_{\{h \geq 1 - \epsilon\}}) - \int \left( \frac{\nabla h(x)}{h(x)} \right)' Q(x) \left( \frac{\nabla h(x)}{h(x)} \right) \varphi(x) dx
\]
\[
- \int \left( \frac{\nabla h(x)}{h(x)} \right)' Q(x) \left( \frac{\nabla h(x)}{h(x)} \right) \varphi(x) dx
\]
where \( n(x) \) denotes the outward normal to \( \{ x : h(x) < 1 - \varepsilon \} \), since (5.5(i)) is valid on \( \{ x : h(x) < 1 - \varepsilon \} \), since \( \nabla h(x) \) is a non-negative multiple of \( n(x) \) on \( \{ x : h(x) = 1 - \varepsilon \} \), and since \((\varphi^2\chi_{\{1 > h \geq 1 - \varepsilon\}}) \to 0 \) as \( \varepsilon \to 0 \). The inequality in (14.1) is the desired result.

\[ \square \]

15. **PROOF OF THEOREM 3.**

Suppose \( h \) satisfies (5.6). A direct calculation from (5.3) yields

\[ (R_{\nu^2}\lambda \nabla^2, h^2 g) = -((\nabla h_i)^t Q(\nabla h_i), g) \to 0. \]  

(15.1)

Hence \( \gamma_g \) is admissible by Theorem 6. The last assertion of the theorem is proved in the same way.

Conversely, suppose \( \gamma_g \in \mathcal{F} \) is admissible. According to Theorem 6 (for the collection \( \Phi \)) there exists a sequence \( \{ \varphi_j \} \subset \Phi \) with \( \varphi_j(x) \geq g \) for \( \| x \| \leq 1 \), and such that \( \lim (R_{\nu^2}\lambda \varphi_j, \varphi_j) = 0 \). (Here we use the fact that \( \sup \{ g(x) : \| x \| \leq 1 \} < \infty \).)

Let \( \nu_j = (\varphi_j)^{1/2} \) and \( \varphi_j = \inf \{ \nu_j, g \} \), \( \nu_j = (\varphi_j^2)^{1/2} \). Then

\[ (R_{\nu^2}\lambda \varphi_j, \varphi_j) = ((\nabla \nu_j)^t Q(\nabla \nu_j), \nu_j) \geq -((\nabla \nu_j)^t Q(\nabla \nu_j), \nu_j) \]

\[ = (R_{\nu^2}\lambda \varphi_j, \varphi_j) \to 0. \]  

(15.2)

Furthermore each \( \nu_j \) has compact support, since \( \varphi_j \in \Phi \). It follows that one may choose \( h_i = \nu_j(x) \) for appropriate indices \( j(i) \), and \( \{ h_i \} \) will satisfy (5.6), as desired. Finally, (5.7) follows from (15.2) and (5.6(i)).

\[ \square \]

16. **PROOF OF THEOREM 4 AND COROLLARY 1.**

The inequality (5.5(i)) can be written as

\[ g \nabla \cdot (Q \nabla h) + (\nabla g)^t Q(\nabla h) \leq 0. \]  

(16.1)

The operator on the left of (16.1) is locally uniformly elliptic on \( \| x - \mu \| > r \) with coefficients which are Hölder continuous. Under these conditions the equivalence of the conditions (a), (b), (c) of Theorem 4 is well known, with equalities or inequalities and for any, or all \( \mu \in R^h, r > 0 \). Further, the solutions \( h \) in Theorem 4 can be chosen to be (twice) continuously differentiable on \( \{ x : \| x - \mu \| > r \} \) and to have bounded derivatives. See Orey (1967, p.18 and 37-41) for a concise proof. (The function, \( h \), of (b) is called a semi-barrier; and the function, \( h \), of (c) is called an anti-barrier. See e.g. Meyers and Serrin (1969).)

By Theorem 2, condition (a) implies inadmissibility of \( \gamma_g \). Conversely, suppose (c) is false, so that there exists a function, \( h \), satisfying (5.5(i)), and (5.5(ii), and

\[ \lim_{\| x \| \to \infty} h(x) = \infty. \]  

(16.2)

Let \( h_j(x) = [(j - h(x))/(j - 1)]^+. \) Then \( h \) has bounded derivatives, \( h_j(x) = 1 \)
for \( \| z - \mu \| \leq r = 0 \) for \( \| z \| \) sufficiently large, and, using (5.5(i))
\[
\int (\nabla h_f(z))^T Q(z) \nabla h_f(z) g(z) \, dz \\
\leq \int_{\| z - \mu \| = r} g(z)(\nabla h(x))^T Q(z)(n(x)) \, dz \\
= (j - 1)^{-1} \int_{\| z - \mu \| = r} g(z)(\nabla h(x))^T Q(z)(n(x)) \, dz \\
\to 0 \text{ as } j \to \infty.
\]
It follows from Theorem 3 that \( g_\alpha \) is admissible. This completes the proof of
Theorem 4. Note that the smoothness conditions on \( g \) and \( Q \) are not used here
(except peripherally to guarantee that \( h \) have bounded derivatives).

To prove Corollary 1 note that if \( h \) satisfies (5.5(i)) then
\[
\nabla \cdot (h^\alpha g_\alpha \nabla (h^{1-\alpha})) \leq 0;
\]
and \( h^{1-\alpha} \) satisfies (16.2). Admissibility of \( g_\alpha c_\alpha \) then follows as in (16.3). This
proves the corollary.

\[ \square \]

ACKNOWLEDGMENTS

Research supported in part by National Science Foundation Research Grants
MCS 7824175 and DMS 85-06447.

BIBLIOGRAPHY

families with applications to simultaneous estimation of gamma scale

Berger, J. (1982). Estimation in continuous exponential families: Bayesian
estimation subject to risk restrictions and inadmissibility results. In *Stat.


when the parameter space is restricted. *Ann. Statist.* 9, 1301–1305.

Blyth, C. R. (1951). On minimax statistical decision procedures and their

Brown, L. (1971). Admissible estimators, recurrent diffusions, and insoluble

Brown, L. (1973). Correction to “Admissible estimators, recurrent diffusions,


