APPROXIMATION OF IMPROPER PRIOR MEASURES
BY PRIOR PROBABILITY MEASURES

BY

CHARLES STEIN

TECHNICAL REPORT NO. 12
July 31, 1964

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT GP-40

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
APPROXIMATION OF IMPROPER PRIOR MEASURES

BY PRIOR PROBABILITY MEASURES

by

Charles Stein

TECHNICAL REPORT NO. 12

July 31, 1964

PREPARED UNDER THE AUSPICIES

OF

NATIONAL SCIENCE FOUNDATION GRANT GP-40

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
APPROXIMATION OF IMPROPER PRIOR MEASURES

BY PRIOR PROBABILITY MEASURES

By

Charles Stein
Stanford University

1. Introduction

It is known that, ordinarily, any admissible decision procedure for a statistical decision problem is, in a fairly strong sense, a limit of Bayes procedures, and in many cases such a limit must be a formal Bayes procedure with respect to a prior measure which may be improper (unbounded). This paper is, for the most part, a non-rigorous attempt at finding, in a reasonably explicit form, conditions for such a formal Bayes procedure to be admissible. In addition, a small amount of effort is devoted to the question of approximation of improper prior measures by prior probability measures without regard to the question of admissibility.

In section 2 we introduce a distance between probability measures, to be applied to posterior distributions, first suggested by Kakutani [ ]. We then define the separation of one prior probability measure, \( \pi^{(1)} \) from a second, \( \pi^{(2)} \) to be the expected value, under \( \pi^{(1)} \), of the squared separation between the posterior distributions corresponding to \( \pi^{(1)} \) and \( \pi^{(2)} \). Roughly speaking, this measures how far off you will be on the average if you compute posterior distribution using \( \pi^{(2)} \) as the prior distribution when \( \pi^{(1)} \) is the true prior distribution. When this separation is small we shall say that \( \pi^{(2)} \) is well approximated by \( \pi^{(1)} \). The separation, and the notion of
good approximation, are, in general, highly asymmetric. The definition can be extended to the case where $\Pi^{(2)}$ is an improper (unbounded) prior measure subject only to (2.9). A result of Stone [13] concerning the approximate consistency of confidence intervals based on Student's $t$ with a Bayes approach is reformulated, without proof, in this terminology.

In section 3, a rough argument is given to indicate that, for sufficiently smooth decision problems, the excess risk incurred by using $\Pi^{(2)}$ as the prior distribution when $\Pi^{(1)}$ is the true prior distribution is bounded by the separation of $\Pi^{(1)}$ from $\Pi^{(2)}$. It looks as if this approximation is reasonably sharp if the decision problem is sufficiently complicated, for example, if all parameters are required to be estimated with roughly equal emphasis. In section 4, an approximation is computed for the separation between $\Pi^{(1)}$ and $\Pi^{(2)}$ in the case where $\Pi^{(2)}$ is an improper prior measure and $\Pi^{(1)}$ is absolutely continuous with respect to $\Pi^{(2)}$ with a density that is smooth (on a large scale). A remainder is given explicitly, but not in a convenient form. In section 5, it is shown that, when the approximations of sections 2 and 4 are valid, the author's necessary and sufficient condition for admissibility [12] yields a more readily usable sufficient condition that is related to a generalized exterior Dirichlet problem. In section 6, this reduced problem is studied, very incompletely.

This work suffers from two major defects. First, as I have already mentioned, the main results must be considered to be conjectures rather
than theorems, since the arguments in sections 3 and 5 are not rigorous. The lack of rigor in section 3 should not be difficult to overcome, but a rigorous treatment of the argument in section 5 will require a careful examination of the remainder in section 4. Perhaps more important is the fact that the admissibility of a statistical procedure is of little direct practical interest. Nevertheless, in many special cases, the study of admissibility yields useful information. For most problems concerning multivariate normal distributions, traditional procedures exploiting the invariance of the problem under translations or linear transformations (or both) seem to be inadmissible. If the characterization of admissible procedures suggested by this paper can be made precise and somewhat more explicit, it may help us find procedures, better than the usual ones, that can seriously be recommended in practice. This should be important in those problems, constituting a large proportion of statistical practice, in which classical large sample theory is not fully applicable because of the large number of unknown parameters.

I am indebted to many people who listened patiently when I presented some of these ideas in a way that was even more chaotic than the presentation in this paper. In particular, the basic idea is, in part, due to J. Kiefer.
2. A definition of separation between two prior distributions.

Let \( X \) be a set (the sample space), \( \mathcal{B} \) a \( \sigma \)-algebra of subsets of \( \mathcal{X} \), \( \lambda \) a \( \sigma \)-finite measure on \( \mathcal{B} \), \( \mathcal{I} \) a set (the parameter space), \( \mathcal{C} \) a \( \sigma \)-algebra of subsets of \( \mathcal{I} \), and \( p \) a \( \mathcal{B} \mathcal{C} \)-measurable non-negative valued function on \( \mathcal{X} \times \mathcal{I} \) such that

\[
\int p(x|\theta) \, d\lambda(x) = 1,
\]

for all \( \theta \in \mathcal{I} \). We suppose we are to observe a random point \( X \) in \( \mathcal{X} \) distributed, for some unknown \( \theta \in \mathcal{I} \), according to the probability density \( p(\cdot|\theta) \) with respect to \( \lambda \), so that the probability that \( X \) is in a set \( B \in \mathcal{B} \) is

\[
P_{\theta}(B) = \int_B p(x|\theta) \, d\lambda(x).
\]

We are interested in making inferences about \( \theta \) or making decisions having consequences depending on \( \theta \). For various reasons, as indicated in later sections, we are interested in the case where \( \theta \) is a random point of \( \mathcal{I} \), in which case we write \( \Theta \) rather than \( \theta \). Then (2) is to be interpreted as the conditional probability that \( X \in B \) given \( \Theta = \theta \). If \( \Theta \) is distributed according to the probability measure (prior distribution) \( \Pi \) on \( \mathcal{C} \), the conditional probability that \( \Theta \in C \) given \( X = x \) is

\[
\Pi_x(C) = \frac{\int_{\mathcal{C}} p(x|\theta) \, d\Pi(\theta)}{\int p(x|\theta) \, d\Pi(\theta)}.
\]

This is also called the posterior probability that \( \Theta \in C \) (given \( X = x \)).
Following Kakutani [3], we define a squared distance \( \delta(m_1, m_2) \) between two probability measures, \( m_1 \) and \( m_2 \) on \( \mathcal{C} \) by

\[
\delta(m_1, m_2) = \int \left[ \sqrt{\frac{dm_1}{dm}} - \sqrt{\frac{dm_2}{dm}} \right]^2 \, dm,
\]

where \( m \) is any measure with respect to which both \( m_1 \) and \( m_2 \) are absolutely continuous, for example \( m=m_1\cdot m_2 \), and \( \frac{dm_1}{dm} \) is the Radon-Nikodym derivative of \( m_1 \) with respect to \( m \). We shall also write (4) in the abbreviated form

\[
\delta(m_1, m_2) = \int \left[ \sqrt{\frac{dm_1}{dm}} - \sqrt{\frac{dm_2}{dm}} \right]^2 = 2 \left( 1 - \int \sqrt{\frac{dm_1}{dm}} \cdot \sqrt{\frac{dm_2}{dm}} \right) = 2(1 - \alpha(m_1, m_2)),
\]

say. For two prior probability measures, \( \Pi_1^{(1)} \) and \( \Pi_2^{(2)} \) on \( \mathcal{C} \), we define the separation \( \delta^*(\Pi_1^{(1)}, \Pi_2^{(2)}) \) between \( \Pi_1^{(1)} \) and \( \Pi_2^{(2)} \) by

\[
\delta^*(\Pi_1^{(1)}, \Pi_2^{(2)}) = E^{(1)} \delta(\Pi_X^{(1)}, \Pi_X^{(2)}),
\]

where the \( \Pi_X^{(1)} \) are the posterior distributions defined by (3), and \( E^{(1)} \) denotes expectation under the assumption that (with \( X \) given \( \mathcal{H} \)) distributed according to (2)) \( \mathcal{H} \) is distributed according to \( \Pi_1^{(1)} \).

Similarly we define

\[
\alpha^*(\Pi_1^{(1)}, \Pi_2^{(2)}) = E^{(1)} \alpha(\Pi_X^{(1)}, \Pi_X^{(2)}),
\]

so that

\[
\delta^*(\Pi_1^{(1)}, \Pi_2^{(2)}) = 2 \left( 1 - \alpha^*(\Pi_1^{(1)}, \Pi_2^{(2)}) \right).
\]
We shall want to extend the definitions (6) and (7) to certain cases where $\prod^{(2)}$ is not a probability measure. If $\prod$ is a non-zero measure on $\mathcal{F}$ for which

$$\int p(x|\theta) \, d\prod(\theta) < \infty \quad \text{a.e.}(\lambda),$$

we shall call $\prod$ a prior measure, improper if $\prod(\mathcal{F}) = \infty$. If $\prod$ is a prior measure, formula (3) defines a probability measure $\prod_x$ for all $x$ except for a set of $\lambda$-measure 0. This has been used by many people, in particular by Jeffreys [2], and will be called the formal posterior measure given $X=x$ under $\prod$. Then (6) and (7) can be applied to the case where $\prod^{(2)}$ is a prior measure, possibly improper, but of course, $\prod^{(1)}$ must be a probability measure. Roughly speaking, $\delta^*(\prod^{(1)}, \prod^{(2)})$ is a measure of how far off the posterior probability computed according to $\prod^{(2)}$ will be on the average if $\prod^{(1)}$ is the true prior distribution.

It can be proved that $\delta^*$ is a strictly convex function of its first argument, that is, for any prior measure $\prod^{(2)}$, distinct prior probability measures $\prod^{(1)}$ and $\prod^{(3)}$ and $0 < \alpha < 1$,

$$\delta^*(\alpha \prod^{(1)} + (1-\alpha) \prod^{(3)}, \prod^{(2)}),$$

$$< \alpha \delta^*(\prod^{(1)}, \prod^{(2)}) + (1-\alpha) \delta^*(\prod^{(1)}, \prod^{(3)}).$$

The proof will be omitted since it is straightforward, though tedious.

Following Kraft [5], let us compute the squared distance $\delta$ between two normal distributions in order to provide a scale for comparison, and also for later use. For $i = 1, 2$ let $m_i$ be a univariate normal distribution with mean $\xi_i$ and variance $\sigma_i^2$. Then
\[ \alpha(m_1, m_2) = \int \sqrt{d\alpha_1 d\alpha_2} \]

\[ = \frac{1}{\sqrt{2\pi} \sigma_1 \sigma_2} \int \exp \left\{ -\frac{1}{4\sigma_1^2} (\theta - \xi_1)^2 - \frac{1}{4\sigma_2^2} (\theta - \xi_2)^2 \right\} d\theta \]

\[ = \sqrt{\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left( -\frac{(\xi_1 - \xi_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right), \]

and

\[ \delta(m_1, m_2) = 2 \left[ 1 - \sqrt{\frac{2\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left( -\frac{(\xi_1 - \xi_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right) \right] \]

\[ \approx \frac{1}{2} \left[ \frac{\sigma_1}{\sigma_2^2} \frac{\sigma_2}{\sigma_1} + 2 \right] + \frac{(\xi_1 - \xi_2)^2}{2(\sigma_1^2 + \sigma_2^2)}, \]

when \[ \frac{\sigma_1}{\sigma_2 - 1} \] and \[ \frac{(\xi_1 - \xi_2)^2}{\sigma_1^2 + \sigma_2^2} \] are small.

Now let us look at an example, which brings out clearly the asymmetry of the separation \( \delta^* \). Suppose we are to observe \( X \), normally distributed with unknown mean \( \theta \) and variance \( 1 \). Let \[ \Pi_1^{(1)} \] and \[ \Pi_2^{(2)} \] be normal prior distributions for \( \theta \) with mean \( 0 \) and variances \( \sigma_1^2 \) and \( \sigma_2^2 \) respectively. Then \[ \Pi_X^{(1)} \] is normal with mean \( \frac{\sigma_1}{\sigma_1^2 + 1} X \) and variance \( \frac{\sigma_2^2}{\sigma_1^2 + 1} \), so that, by (12),

\[ \delta \left( \Pi_X^{(1)}, \Pi_X^{(2)} \right) = 2 \left( 1 - \exp \left( -\frac{b}{2} X^2 \right) \right), \]

where
\[ a = \sqrt{\frac{2}{\sigma_1 \sqrt{\sigma_1^2 + 1} + \sigma_2 \sqrt{\sigma_2^2 + 1}}} \] 

and

\[ b = \frac{\left( \frac{\sigma_1^2}{\sigma_1^2 + 1} - \frac{\sigma_2^2}{\sigma_2^2 + 1} \right)^2}{2 \left( \frac{\sigma_1^2}{\sigma_1^2 + 1} + \frac{\sigma_2^2}{\sigma_2^2 + 1} \right)^2} = \frac{(\sigma_1^2 - \sigma_2^2)^2}{2(\sigma_1^2 + 1)(\sigma_2^2 + 1)(2\sigma_1^2 \sigma_2^2 + \sigma_1^2 + \sigma_2^2)} . \]

But, under \( \prod^{(1)} \), \( X \) is normally distributed with mean 0 and variance \( \sigma_1^2 + 1 \). Thus

\[ \delta^*\left( \prod^{(1)}, \prod^{(2)} \right) = \mathbb{E}^{(1)} \delta\left( \prod_X^{(1)}, \prod_X^{(2)} \right) \]

\[ = \mathbb{E}^{(1)} 2 \left( 1 - a e^{-\frac{b}{2} x^2} \right) \]

\[ = 2 \left( 1 - a \int e^{-\frac{1}{2} y^2 (1 + (\sigma_1^2 + 1)b)} \, dy \right) \]

\[ = 2 \left( 1 - \frac{a}{\sqrt{1 + (\sigma_1^2 + 1)b}} \right) . \]

Now suppose

\[ l \ll \sigma_1 \ll \sqrt{\sigma_2} . \]

Then \( a \approx 1 \), and

\[ (\sigma_1^2 + 1)b \approx \frac{1}{4\sigma_1^2} \ll 1 , \]

so that, by (16)

\[ \delta^*\left( \prod^{(1)}, \prod^{(2)} \right) \ll 1 . \]
However

(20) \[ s^*(\prod^{(2)}, \prod^{(1)}) = 2 \left(1 - \frac{8}{\sqrt{1+(\sigma_2^2+1)b}}\right), \]

and

(21) \[ (\sigma_2^2+1)b \approx \frac{\sigma_2^4}{4\sigma_1} \ll 1 \]

so that

(22) \[ s^*(\prod^{(2)}, \prod^{(1)}) \approx 2. \]

It seems desirable to indicate the interaction of these notions with that of a group of transformations leaving a statistical problem invariant, although it would be inappropriate to attempt a lengthy systematic treatment here. With \((\mathcal{X}, \mathcal{B}, \lambda, T, \ell, p)\) as before, let \(\mathcal{G}\) be a group of \(1\)-\(1\) \((\mathcal{B}, \mathcal{B})\) measurable transformations \(g\) of \(\mathcal{X}\) onto \(\mathcal{X}\) such that (for each such \(g\)) there is given a \((\mathcal{C}, \mathcal{C})\) measurable transformation \(\overline{g}: \mathcal{X} \to \mathcal{X}\) with the property that, for any \(\theta\), if \(X\) is distributed according to the density \(p(\cdot | \theta)\) then \(gX\) is distributed according to \(p(\cdot | \overline{g}\theta)\). We suppose this correspondence \(g \to \overline{g}\) is a homomorphism of \(\mathcal{G}\) onto a group \(\mathcal{G}\) of transformations \(\sigma^2 \mathcal{T}\) (as it must be if \(p(\cdot | \theta_1) \neq p(\cdot | \theta_2)\) for \(\theta_1 \neq \theta_2\)). We say that \(\mathcal{G}\) operates transitively on \(\mathcal{T}\) if, for any \(\theta_1, \theta_2 \in \mathcal{C}\), there exists \(\overline{g} \in \mathcal{G}\) such that \(\overline{g}\theta_1 = \theta_2\). We recall that a confidence procedure may be taken to be a \(\mathcal{B}\)-measurable function \(\tau\) on \(\mathcal{X}\) to \(\ell\), with the interpretation that, after observing \(X\) we assert that \(\theta \in \tau(X)\). This procedure is said to have exact confidence coefficient \(\alpha\) if, for all \(\theta \in \mathcal{T}\),

(23) \[ P(\theta \in \tau(X)) = \alpha. \]
We first want to discuss, in the case where $\mathcal{G}$ operates transitively on $\mathcal{I}$, a relation between a certain formal posterior probability and confidence procedures first pointed out by Pitman [11] in the special case of a group consisting of translations or scale changes or both. The conditions imposed here are unnecessarily restrictive but they cover many interesting applications.

Let us first recall some results concerning invariant measures in separable, locally compact topological groups (see for example, Loomis [8], Chapter 6). Let $\mathbb{A}$ be the $\sigma$-algebra of Borel subsets of such a group $\mathcal{G}$. Then there exists a non-zero, $\sigma$-finite measure $\mu$ on $\mathbb{A}$ that is left-invariant, that is

$$\mu(gA) = \mu(A)$$

for all $g \in \mathcal{G}$ and $A \in \mathbb{A}$, and any other such left invariant measure is a positive multiple of $\mu$. We define

$$\Delta(g_1) = \frac{d\mu(gg_1)}{d\mu(g)},$$

which exists and is independent of $g$ because of the near-uniqueness of $\mu$. Then $\Delta$ is a continuous homomorphism of $\mathcal{G}$ into the multiplicative group of positive real numbers. Let $\nu$ be defined by

$$\nu(Ag) = \mu(A^{-1}).$$

Then $\nu$ is right-invariant, that is

$$\nu(Ag) = \nu(A),$$

and also
(28) \[ \frac{dv}{du}(g) = \Delta(g^{-1}). \]

If \( G \) is abelian or compact \( \Delta = 1 \), and in the compact case, \( \mu \) may be taken to be a probability measure. The invariant measures \( \mu \) and the corresponding functions \( \Delta \) are given below for some groups that arise frequently in statistical problems:

(i) If \( G \) is the group of all translations of a \( k \)-dimensional real linear space \( \mathbb{R}^k \), \( \mu \) is Lebesgue measure and \( \Delta = 1 \).

(ii) If \( G \) is the group of all linear transformations \( g \) of \( \mathbb{R}^k \), then, with \( g \) expressed as a matrix

\[
(29) \quad d\mu(g) = \frac{\prod d_{g_{i,j}}}{|\det g|^k},
\]

and \( \Delta = 1 \).

(iii) If \( G \) is the group of all affine transformations \( g = (h, a) \) of \( \mathbb{R}^k \) operating by

\[
(30) \quad gx = (h, a)x = hx + a,
\]

then

\[
(31) \quad d\mu(g) = \frac{\prod dh_{i,j} \prod da_i}{|\det h|^{k+1}},
\]

and

\[
(32) \quad dv(g) = \frac{\prod dh_{i,j} \prod da_i}{|\det h|^k},
\]

so that

\[
(33) \quad \Delta(g) = \frac{dv}{dv}(g) = \frac{1}{|\det h|}.
\]
(iv) If $\mathcal{G}$ is the multiplicative group of all $k \times k$ lower triangular matrices $g$ (satisfying $g_{ij} = 0$ for $j > i$), then

\begin{equation}
\hat{\mu}(g) = \frac{\prod_{j \leq 1} dg_{ij}}{\prod |g_{11}|^i},
\end{equation}

\begin{equation}
\hat{\nu}(g) = \frac{\prod_{j \leq 1} dg_{ij}}{\prod |g_{11}|^{k+1-i}},
\end{equation}

and

\begin{equation}
\Delta(g) = \frac{\hat{\mu}}{\hat{\nu}}(g) = \prod |g_{11}|^{k+1-2i}.
\end{equation}

Now let $\mathcal{V}$ be a set with a $\sigma$-algebra $\mathcal{B}$ of subsets, and suppose that $\mathcal{X} = \mathcal{G} \times \mathcal{V}$ and $\mathcal{B} = \mathcal{A} \mathcal{B}$ and $\mathcal{G}$ operates on $\mathcal{V}$ by

\begin{equation}
g_1 x = g_1(s, \nu) = (g_1 s, \nu).
\end{equation}

We suppose $\mathcal{G}$ induces a group $\bar{\mathcal{G}}$ of transformations of $\mathcal{X}$ as indicated earlier, and we also suppose that $\bar{\mathcal{G}}$ operates transitively on $\mathcal{V}$. We write $X = (G, V)$ and define the probability measure $\rho$ on $\mathcal{B}$ by

\begin{equation}
\rho(D) = P_0(V \in D).
\end{equation}

Because of the transitivity of $\bar{\mathcal{G}}$, this is independent of $\theta$. Choose $\lambda = \mu_0$. Now let $\mathcal{K}$ be the subgroup of $\mathcal{G}$ consisting of all $h \in \mathcal{G}$ for which $H \theta_0 = \theta_0$ where $\theta_0$ is an arbitrary point of $\mathcal{X}$ fixed for the remainder of this discussion, and suppose $\mathcal{K}$ is compact. Let $\lambda$ be the measure induced in $\mathcal{X}$ by the right-invariant measure $\nu$ in $\mathcal{G}$, that is, with $y: \mathcal{G} \rightarrow \mathcal{X}$ defined by

\begin{equation}
yg = y \theta_0,
\end{equation}

12
let \( \prod \) be defined by

\[
\prod(c) = v(c^{-1}).
\]

Let \( \tau: X \to G \) be chosen so that

\[
\tau(gx) = \tilde{g}(\tau x) \quad \text{for all } g \in G \text{ and } x \in X
\]

and

\[
\prod_x \{ \tau(x) \} = \alpha \quad \text{for all } x.
\]

We shall show that then

\[
P_\theta \{ \theta \in \tau(X) \} = \alpha \quad \text{for all } \theta.
\]

With \( x = (g, v) \),

\[
\alpha = \prod_x \{ \tau(x) \} = \frac{\int_{\theta \in \tau(g, v)} p(g, v|\theta) d\prod(\theta)}{\int p(g, v|\theta) d\prod(\theta)}
\]

\[
= \frac{\int_{g^{-1} g_1 \theta_0 \in \tau(g, v)} p(g, v|g_1 \theta_0) d\nu(g_1)}{\int p(g, v|g_1 \theta_0) d\nu(g_1)}
\]

\[
= \frac{\int_{\theta_0 \in \tau(g_1^{-1} g, v)} p(g_1^{-1} g, v|\theta_0) d\nu(g_1)}{\int p(g_1^{-1} g, v|\theta_0) d\nu(g_1)}
\]

\[
= \int_{\theta_0 \in \tau(g_2, v)} p(g_2, v|\theta) d\mu(g_2),
\]

for all \( v \). It follows that
\[ (45) \quad P_{\theta_0}(\theta_0 \epsilon \tau(X)) = \alpha = \int_{\theta_0 \epsilon \tau(g, v)} p(g, v|\theta) d\mu(g) \]

But, by the transitivity assumption, any \( \theta \) can be expressed as \( g \theta_0 \) for some \( g \in \mathcal{G} \), so that

\[ (46) \quad P_{\theta} (\theta \epsilon \tau(X)) = P_{\theta_0} (g \theta_0 \epsilon \tau(X)) \]

\[ = P_{g \theta_0} (\theta_0 \epsilon \tau(g^{-1}X)) = P_{\theta_0} (\theta_0 \epsilon \tau(X)) = \alpha. \]

We note that the defining conditions (41) and (42), and the conclusion (43) do not depend on the explicit representation \( \mathcal{X} = \mathcal{G} \times \mathcal{N} \), so that we need only be assured of the possibility of such a representation, and need not obtain it explicitly. It is easy to weaken the conditions in certain rather trivial ways. For example it is enough that \( \mathcal{X} = (\mathcal{G} / \mathcal{H}) \times \mathcal{N} \) with \( \mathcal{H} \) compact, or that \( \mathcal{X} \) be a countable union of such spaces.

However, it would be desirable to give a proof valid under conditions not expressed in terms of such an explicit representation. This result, (43), was obtained by Pitman [11] in the special case of a group consisting of translations or scale changes, or both.

At first sight the result is a bit puzzling because the problem of finding a \( \Pi \) such that (41) and (42) imply (43) is invariant under transformation on the left by \( \mathcal{G} \) and \( \mathcal{F} \) so that we would expect the solution \( \Pi \) (assuming it exists and is essentially unique) to be invariant under transformation on the left by \( \mathcal{F} \). However we observe that, in (44)

\[ \Pi_{\mathcal{X}} (\tau(x)) \] is homogeneous of degree 0 in \( \Pi \) so that any relatively invariant \( \Pi \) is effectively invariant for this problem, and which of the relatively invariant measures (if any) is the solution cannot be decided.
by this qualitative argument. A measure $\prod$ is called relatively invariant if there is a function $\delta$ on $\mathcal{F}$ such that

$$\prod(\mathcal{G}C) = \delta(g) \prod(C)$$

for all $g \in \mathcal{G}$ and $C \in \mathcal{C}$.

It is of some interest to ask whether this prior measure $\prod$ induced by the right invariant measure in $\mathcal{G}$ can be approximated by prior probability measures, that is whether, for any $\epsilon > 0$ there exists a prior probability measure $\prod^\epsilon$ for which

$$\delta^*(\prod^\epsilon, \prod) < \epsilon .$$

This question can be answered in a large class of interesting special cases by essentially the methods used by Kudo [6] and Kiefer [4], for the question of whether the corresponding formal Bayes procedures are minimax. The result is that $\prod$ can be approximated by prior probability measures if $\mathcal{G}$ contains a finite sequence of closed subgroups $\mathcal{G}_0 = \{1\} \subset \mathcal{G}_1 \subset \cdots \subset \mathcal{G}_k = \mathcal{G}$ such that each quotient group $\mathcal{G}_k/\mathcal{G}_{k-1}$ is either abelian or compact. A special case of this was considered by M. Stone [13]. He showed that if $X_1, \ldots, X_n$ (with $n \geq 2$) are independently normally distributed with unknown mean $\xi$ and unknown variance $\sigma^2$, then the usual confidence intervals $I(X)$ for $\xi$, based on Student's $t$, are approximately sets of posterior probability equal to the confidence coefficient $\alpha$, in the sense that for any $\epsilon > 0$ there exists a prior probability measure $\prod^\epsilon$ such that

$$\prod^\epsilon \{ x : \int_{-\infty}^x (t \in I(x)) - \alpha > \epsilon \} < \epsilon .$$
Of course, here

\[(50) \quad I(X) = (\overline{x} - a\sqrt{S}, \overline{x} + a\sqrt{S}), \]

where

\[(51) \quad \overline{x} = \frac{1}{n} \sum x_i, \]

and

\[(52) \quad S = \sum (x_i - \overline{x})^2, \]

and

\[(53) \quad \alpha = P_{\xi, \sigma}(\xi \in I(X)). \]

However it is not always true that the prior measure \(\prod\) induced by the right-invariant measure in \(\mathcal{G}\) can be approximated by prior probability measures. In particular such approximation is not in general, possible for problems of multivariate analysis invariant under the full linear group. Counter examples (again in the context of the minimax problem) are given in James and Stein [1], and Lehmann [7]. An earlier counter example of Peisakoff [10] concerning the free group with two generators may also be relevant.

I had hoped to include a computation of \(\inf_{\mathcal{T}^{(1)}} \delta^*(\mathcal{T}^{(1)}, \mathcal{T})\) in a multivariate situation but I have been unable to overcome a major difficulty.

I find that the material on groups in this section has been done in much greater detail by Rajinder Bir Hora in an unpublished Ph.D. thesis at the University of Minnesota.
3. Statistical Decision Problems

In this section, we shall look at a rough argument which indicates that, for sufficiently smooth statistical decision problems, the excess risk incurred by using the formal Bayes procedure corresponding to the prior measure \( \mathbb{P} \) when the true prior distribution is \( \mathbb{P}^{(1)} \) does not exceed a constant multiple of \( \varepsilon^*(\mathbb{P}^{(1)}, \mathbb{P}) \), where the constant depends on the decision problem and on \( \mathbb{P} \). It should not be difficult to turn this rough argument into a reasonably simple proof of a usable precise assertion, but I have not succeeded in doing so.

In addition to the structure \( (\mathcal{X}, \mathcal{B}, \lambda, \mathcal{F}, \mathcal{G}, p) \) considered in section 2, we suppose we are also given a finite-dimensional real coordinate space \( \mathcal{Z} \), the action space, and a non-negative valued function \( L \), the loss function, on \( \mathcal{F} \times \mathcal{Z} \), jointly measurable in both variables and twice continuously differentiable in the second variable \( \mathcal{Z} \) for each value of the first variable \( \theta \). A decision function \( \varphi \) is a measurable function on the sample space \( \mathcal{X} \) to the action space \( \mathcal{Z} \) and its risk function \( \rho(\cdot, \varphi) \) is defined by

\[
\rho(\theta, \varphi) = E_{\theta} L(\theta, \varphi(x)) = \int L(\theta, \varphi(x)) p(x|\theta) d\lambda(x) .
\]

For a prior probability measure \( \mathbb{P} \) on \( \mathcal{B} \), a Bayes procedure is, by definition, a decision function \( \psi \) that minimizes the expected risk under \( \mathbb{P} \), that is

\[
\int \rho(\theta, \psi) d \mathbb{P}(\theta) = \inf_{\varphi} \int \rho(\theta, \varphi) d \mathbb{P}(\theta) < \infty .
\]
We shall suppose that for this $\psi$, $\rho(\theta, \psi)$ is a bounded function of $\theta$.

As explained at the beginning of section 5, this (or even constant $\rho(\theta, \psi)$) can be achieved by an appropriate change of notation provided $\rho(\theta, \psi)$ is everywhere finite (and non zero). The value of the left hand side of (2) is called the Bayes risk under $\underline{\underline{\lambda}}$. It is known that, if a Bayes procedure $\psi$ exists, it satisfies

$$\int L(\theta, \psi(x)) \, \underline{\underline{\lambda}}(\theta) \, d\underline{\underline{x}}(\theta) = \inf_{\lambda} \int L(\theta, \lambda) \, d\underline{\underline{x}}(\theta),$$

except possible on a set $G$ of $x$ for which

$$\int \underline{\underline{d}}(\theta) \int p(x|\theta) d\lambda(x) = 0.$$

Formula (3) can be used to define a formal Bayes procedure with respect to an improper prior measure $\underline{\underline{\lambda}}$. We recall that a decision procedure $\psi$ is said to be admissible if there exist no decision procedure $\phi$ such that

$$\rho(\theta, \phi) \leq \rho(\theta, \psi)$$

for all $\theta \in G$ with strict inequality for some $\theta$. As indicated in section 5, in studying the question of whether the formal Bayes procedure $\psi$ corresponding to a given improper prior measure $\underline{\underline{\lambda}}$ is admissible we shall be interested in an upper bound for $\int \rho(\theta, \psi) \, d\underline{\underline{\lambda}}(\theta) - \inf_{\phi} \int \rho(\theta, \phi) \, d\underline{\underline{\lambda}}(\theta)$ for certain prior probability measures $\underline{\underline{\lambda}}$.

We indicate such a bound by a rough argument, considering only the case where $\dim \lambda = 1$. 

18
Let $\mu$ be an improper prior measure, $\psi$ the corresponding formal Bayes procedure, $\mu^{(1)}$ a prior probability measure and $\varphi_1$ the corresponding Bayes procedure. Denoting differentiation of $L$ with respect to its second argument by prime, we have

\begin{equation}
L(\theta, \psi(x)) \approx L(\theta, \varphi_1(x)) + (\psi(x) - \varphi_1(x)) L'(\theta, \varphi_1(x))
\end{equation}

\begin{equation}
+ \frac{1}{2} (\psi(x) - \varphi_1(x))^2 L''(\theta, \varphi_1(x)),
\end{equation}

so that

\begin{equation}
\int [L(\theta, \psi(x)) - L(\theta, \varphi_1(x))] \, d\mu_x^{(1)}(\theta)
\end{equation}

\begin{equation}
\approx (\psi(x) - \varphi_1(x)) \int L'(\theta, \varphi_1(x)) \, d\mu_x^{(1)}(\theta)
\end{equation}

\begin{equation}
+ \frac{1}{2} (\psi(x) - \varphi_1(x))^2 \int L''(\theta, \varphi_1(x)) \, d\mu_x^{(1)}(\theta)
\end{equation}

\begin{equation}
= \frac{1}{2} (\psi(x) - \varphi_1(x))^2 \int L''(\theta, \varphi_1(x)) \, d\mu_x^{(1)}(\theta).
\end{equation}

Also

\begin{equation}
L'(\theta, \psi(x)) \approx L'(\theta, \varphi_1(x)) + (\psi(x) - \varphi_1(x)) L''(\theta, \varphi_1(x)).
\end{equation}

Integrating with respect to $\mu_x^{(1)}$ we have

\begin{equation}
\psi(x) - \varphi_1(x) \approx \frac{\int L'(\theta, \psi(x)) d\mu_x^{(1)}(\theta)}{\int L''(\theta, \psi(x)) d\mu_x^{(1)}(\theta)}.
\end{equation}

Substituting in (7) we find

19
\[ (10) \quad \int [L(\theta, \psi(x)) - L(\theta, \phi_{1}(x))]d\theta \]

\[ \approx \frac{1}{2} \left[ \frac{\int L'(\theta, \psi(x))d\theta}{\int L''(\theta, \psi(x))d\theta} \right]^{(1)}(\theta) \]

But

\[ (11) \quad \left[ \int L'(\theta, \psi(x))d\theta \right]^{(1)}(\theta) \]

\[ = \left[ \int L'(\theta, \psi(x))(d\theta - d\theta) \right]^{2} \]

\[ = \left[ \int L'(\theta, \psi(x))\sqrt{d\theta} \right]^{(1)}(\theta) \]

\[ \leq \int L'(\theta, \psi(x))d\theta \left[ \sqrt{d\theta} \right]^{(1)}(\theta) + \sqrt{d\theta} \]
\begin{align*}
\leq \int \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta') \int \rho(x|\theta') \text{d} \lambda(x) \frac{\int L''(\theta, \psi(x)) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta) + \int \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta)}{\int L''(\theta, \psi(x)) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta)} \delta(\bar{\bar{\bar{\bar{x}}}}^{(1)}, \bar{\bar{\bar{\bar{x}}}}^{(1)}).
\end{align*}

Thus if

\begin{align*}
\frac{\int L''(\theta, \psi(x)) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta) + \int \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta)}{\int L''(\theta, \psi(x)) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta)}
\end{align*}

is bounded function of $x$ and the approximations leading to (10) are valid, we have

\begin{align*}
(13) \quad \int \rho(\theta, \psi) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta) - \inf_{\varphi} \int \rho(\theta, \varphi) \text{d} \bar{\bar{\bar{\bar{x}}}}^{(1)}(\theta) \\
\leq K \delta^*(\bar{\bar{\bar{\bar{x}}}}^{(1)}, \bar{\bar{\bar{\bar{x}}}}^{(1)}).
\end{align*}
4. An Approximation to $\delta^{*}(\bar{\Gamma}^{(1)}, \bar{\Gamma})$.

We return to the structure $(\kappa, \theta, X, p)$ introduced at the beginning of section 2, where $X$ is now a k-dimensional real coordinate space with $k$ finite. We suppose we are given an improper prior measure $\bar{\Gamma}$ and show that under certain conditions, the separation $\delta^{*}(\bar{\Gamma}^{(1)}, \bar{\Gamma})$ is given approximately by

\[
(1) \quad \delta^{*}(\bar{\Gamma}^{(1)}, \bar{\Gamma}) \approx \frac{1}{4} \int \frac{1}{q(\theta)} \sum \frac{\partial q(\theta)}{\partial \theta_i} \frac{\partial q(\theta)}{\partial \theta^j} g^{ij}(\theta) d\bar{\Gamma}(\theta),
\]

where $\bar{\Gamma}^{(1)}$ is a prior probability measure with

\[
(2) \quad q = \frac{d\bar{\Gamma}^{(1)}}{d\bar{\Gamma}},
\]

and $g(\theta)$ is the covariance matrix, when $X$ is distributed according to $p(\cdot, \theta)$, of the posterior mean $\hat{\theta}(x)$ of $\theta$ given $X$ under $\bar{\Gamma}$. We denote the $i^{th}$ coordinate of $\theta$ by $\theta_i$ and the $(i,j)$ coordinate of $g(\theta)$ by $g^{ij}(\theta)$. In many cases, if $\bar{\Gamma}$ is reasonably smooth, we can expect $g(\theta)$ to be approximately the inverse of the information matrix:

\[
(3) \quad [(g(\theta)^{-1})_{ij}] \approx E_{\theta} \frac{\partial \log p(X|\theta)}{\partial \theta_i} \frac{\partial \log p(X|\theta)}{\partial \theta^j}.
\]

Of course, a precise justification of this could come only from large-sample theory, but, when we come to apply formula (1) in section 5, we shall see that we need only that the ratio of the two sides is bounded away from 0 and $\infty$. We give a precise form of the approximation (1) with remainder in formulas (24), (25), (17), (16), (20), and (21). However the remainder is not expressed in a usable form.
All of our work will be subject to the assumption that the prior probability measure \( \mathcal{L}^{(1)} \) is absolutely continuous with respect to \( \mathcal{L} \) and \( q = d\mathcal{L}^{(1)}/d\mathcal{L} \) is twice continuously differentiable. In addition, roughly speaking, we suppose that \( q \) and its first and second partial derivatives are nearly the same at any two points \( \theta, \theta' \) whose distance is small or moderate in the Riemannian metric determined by \( \theta \to [g(\theta)]^{-1} \). Let

\[
\rho^{(1)}(\mathcal{L}_x^{(1)}, \mathcal{L}_x) = 1 - \frac{1}{2} \delta(\mathcal{L}_x^{(1)}, \mathcal{L}_x) = \int \sqrt{d\mathcal{L}^{(1)}_{x}} \, \mathcal{L}_{x},
\]

and

\[
\rho^{*}(\mathcal{L}_x^{(1)}, \mathcal{L}_x) = 1 - \frac{1}{2} \delta^{*}(\mathcal{L}_x^{(1)}, \mathcal{L}_x) = E^{(1)} \int \sqrt{d\mathcal{L}^{(1)}_{x}} \, \mathcal{L}_{x}
\]

\[
= \int d\mathcal{L}^{(1)}(\theta) \int p(x|\theta) d\lambda(x) \underbrace{\int p(x|\theta') \sqrt{d\mathcal{L}^{(1)}(\theta')} d\lambda(\theta')}_{\sqrt{\int p(x|\theta') \sqrt{d\mathcal{L}^{(1)}(\theta')} d\lambda(\theta')}}
\]

\[
\sqrt{\int p(x|\theta') d\mathcal{L}^{(1)}(\theta')} \int p(x|\theta'') d\mathcal{L}^{(1)}(\theta''),
\]

\[
= \int d\lambda(x) \left( \sqrt{\int p(x|\theta) \sqrt{d\mathcal{L}^{(1)}(\theta)} d\lambda(\theta)} \right) \sqrt{\int p(x|\theta) d\mathcal{L}^{(1)}(\theta)}
\]

We shall apply this with \( \mathcal{L}^{(1)} \) related to \( \mathcal{L} \) by (2). We shall use the summation convention whereby the second term on the right hand side of (7) represents a sum over \( i \), and the third term on the right hand side of (8) a sum over \( i \) and \( j \).
Let

\[ \theta(x) = \frac{\int \theta p(x|\theta) d\theta}{\int p(x|0) d\theta} \]  

Then, by Taylor's formula,

\[ q(\theta) = q(\hat{\theta}(x)) + (\theta^i - \hat{\theta}^i(x)) q_1(\hat{\theta}(x)) + \frac{1}{2} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) q_{ij}(\theta^*(x)) , \]

where \( \theta^*(x) \) is a point of the line segment joining \( \theta \) and \( \theta(x) \) and

\[ q_1(\theta) = \frac{\partial q(\theta)}{\partial \theta^i} , \quad q_{ij}(\theta) = \frac{\partial^2 q(\theta)}{\partial \theta^i \partial \theta^j} . \]

Consequently

\[ \int p(x|\theta) q(\theta) d\theta = \int p(x|\theta) [q(\hat{\theta}(x)) + (\theta^i - \hat{\theta}^i(x)) q_1(\hat{\theta}(x)) + \frac{1}{2} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) q_{ij}(\theta^*(x))] d\theta = q(\hat{\theta}(x)) \int p(x|\theta) d\theta + \frac{1}{2} \int p(x|\theta) (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) q_{ij}(\theta^*(x)) d\theta , \]

by (6). Again, using Taylor's formula, we find that, with \( \theta^{**}(x) \) on the line segment joining \( \theta \) and \( \hat{\theta}(x) \),
\[
\sqrt{q(\theta)} = \sqrt{q(\hat{\theta}(x))} + (\theta^1 - \hat{\theta}^1(x)) \frac{q_{1}(\hat{\theta}(x))}{2\sqrt{q(\theta(x))}} \\
+ \frac{1}{2} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) \left\{ \frac{q_{11}(\theta^{**}(x))}{2\sqrt{q(\theta^{**}(x))}} - \frac{1}{4} \frac{q_{1}(\theta^{**}(x))q_{j}(\theta^{**}(x))}{[q(\theta^{**}(x))]^{3/2}} \right\}.
\]

Thus
\[
\int_{P(x|\theta)} \sqrt{q(\theta)} d|\theta| = \sqrt{q(\hat{\theta}(x))} \int_{P(x|\theta)} d|\theta|
\]
\[
+ \frac{1}{2} \int_{P(x|\theta)} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) \left\{ \frac{q_{11}(\theta^{**}(x))}{2\sqrt{q(\theta^{**}(x))}} - \frac{1}{4} \frac{q_{1}(\theta^{**}(x))q_{j}(\theta^{**}(x))}{[q(\theta^{**}(x))]^{3/2}} \right\} d|\theta|.
\]

Then (5) becomes
\[
\rho^*(|\theta|^{(1)},|\theta|)
\]
\[
= \int_{d\lambda(x)} \left( \int_{P(x|\theta)} \sqrt{q(\theta)} d|\theta| \right) \left( \int_{P(x|\theta)} q(\theta) d|\theta| \right)
\]
\[
= \int_{d\lambda(x)} \sqrt{1 + \frac{1}{2q(\hat{\theta}(x))}} \left( \int_{P(x|\theta)} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) q_{11}(\theta^{**}(x)) p(x|\theta) d|\theta| \right)
\]
\[
\cdot q(\hat{\theta}(x)) \int_{P(x|\theta)} d|\theta|.
\]
\[
\left\{ \frac{1}{2} \int_{P(x|\theta)} (\theta^i - \hat{\theta}^i(x)) (\theta^j - \hat{\theta}^j(x)) \left\{ \frac{q_{11}(\theta^{**}(x))}{2q(\theta^{**}(x))q(\hat{\theta}(x))} - \frac{1}{4} \frac{q_{1}(\theta^{**}(x))q_{j}(\theta^{**}(x))}{[q(\theta^{**}(x))]^{3/2}} \right\} p(x|\theta) d|\theta| \right\}
\]
\[
\right\}
\]

25
However

\[(13) \quad \sqrt{1 + t} = 1 + \frac{t}{2} + O(\min(|t|, t^2)) \]

since

\[(14) \quad \frac{\sqrt{1 + t} - 1 - \frac{t}{2}}{t^2} \]

is a continuous function of \( t \) vanishing as \( \frac{1}{t} \) at \( \pm \infty \). If we apply this and the inequality,

\[(15) \quad |uv| \leq \frac{1}{2} (u^2 + v^2) \]

to (12), we obtain

\[(16) \quad \mu^*(\|\|^{(1)}, \|\|) \]

\[
= \int d\lambda(x)q(\theta(x))\int p(x|\theta')d\|\|^{(1)}(\theta') \]

\[
\cdot \left\{ 1 + \frac{1}{4q(\theta(x))} \left[ \int (\theta^1 - \hat{\theta}^1(x)) (\theta^j - \hat{\theta}^j(x)) q_{ui}(\theta^*(x)) p(x|\theta) d\|\|^{(1)}(\theta) \right] \right. 
\]

\[
\left. \left. \frac{\int p(x|\theta) d\|\|^{(1)}(\theta)}{\int p(x|\theta) d\|\|^{(1)}(\theta)} \right) \right. 
\]

\[
+ \frac{1}{2} \int \left[ \int (\theta^1 - \hat{\theta}^1(x)) (\theta^j - \hat{\theta}^j(x)) \left\{ \frac{q_{ij}(\theta^*(x))}{\sqrt{2q(\theta^*(x))q(\theta(x))}} - \frac{1}{4} \frac{q_{ij}(\theta^*(x))q_{ij}(\theta^*(x))}{q(\theta(x)) [q(\theta^*(x))]^{3/2}} \right\} p(x|\theta) d\|\|^{(1)}(\theta) 
\]

\[
+ R_1(x) \right\} \]

\[(17) \quad |R_1(x)| \leq K(A^2(x) + B^2(x)), \]

with \( K \) an absolute constant, \( A(x) \) the second term in braces in (16) and \( B(x) \) the third term in braces in (16). But, by (7),
Using (18) to evaluate the term in (16) arising from the 1 in braces, we find

\begin{equation}
\rho^*\left(\frac{R_1(x)}{R_2(x)}, \frac{R_2(x)}{R_2(x)}\right) = 1 + \int d\lambda(x) q(\hat{\theta}(x)) R_1(x) \int p(x|\theta) d\frac{R_1(x)}{R_2(x)}
\end{equation}

\begin{equation}
- \frac{1}{8} \int d\lambda(x) \int (\hat{\theta}^i - \hat{\theta}^i(x))(\theta^j - \hat{\theta}^j(x)) \frac{a_{ij}^{(x^*)}(x^*) q_i^{(x^*)}(x^*)}{[q(\theta^*)]^3/2 [q(\hat{\theta}(x))]} p(x|\theta) d\frac{R_1(x)}{R_2(x)}
\end{equation}

\begin{equation}
+ \int d\lambda(x) R_2(x)
\end{equation}

where

\begin{equation}
R_2(x) = \int (\hat{\theta}^i - \hat{\theta}^i(x))(\theta^j - \hat{\theta}^j(x)) \left\{ \frac{a_{ij}^{(x^*)}(x)}{\sqrt{q(\theta^*)} \sqrt{q(\hat{\theta}(x))}} - q_{ij}(\theta^*) \right\}
\end{equation}

\begin{equation}
\cdot p(x|\theta) d\frac{R_1(x)}{R_2(x)}.
\end{equation}

Then, letting

\begin{equation}
R_3(x) = \int (\hat{\theta}^i - \hat{\theta}^i(x))(\theta^j - \hat{\theta}^j(x))
\end{equation}

\begin{equation}
\cdot \left\{ \frac{a_{ij}^{(x^*)}(x) q_i^{(x^*)}(x)}{[q(\theta^*)]^{3/2} [q(\hat{\theta}(x))]^{1/2}} - \frac{a_{ij}(\theta) q_i(\theta)}{q(\theta)} \right\} p(x|\theta) d\frac{R_1(x)}{R_2(x)},
\end{equation}

27
we have

\begin{equation}
\delta^*\left(\left(\left[\frac{1}{\lambda}\right]\left[\theta(1)\right], \left[\right]\right) = 2(1 - \rho^*\left(\left[\frac{1}{\lambda}\right]\left[\theta(1)\right], \left[\right]\right))
\end{equation}

\begin{align*}
&= \frac{1}{4} \int \lambda(x) \int (\theta^1 - \hat{\theta}^1(x))(\theta^j - \hat{\theta}^j(x)) \frac{q_1(\theta)q_4(\theta)}{q(\theta)} p(x|\theta) d\left[\theta\right] \left(\theta\right)
\end{align*}

\begin{align*}
&- 2 \int \lambda(x) \int [q(\hat{\theta}(x))R_1(x)] \int p(x|\theta) d\left[\theta\right] \left(\theta\right) + R_2(x) + R_3(x)
\end{align*}

But

\begin{equation}
\int \lambda(x) \int (\theta^1 - \hat{\theta}^1(x))(\theta^j - \hat{\theta}^j(x)) \frac{q_4(\theta)q_4(\theta)}{q(\theta)} p(x|\theta) d\left[\theta\right] \left(\theta\right)
\end{equation}

\begin{align*}
&= \int \lambda(x) \int p(x|\theta) (\theta^1 - \hat{\theta}^1(x))(\theta^j - \hat{\theta}^j(x)) d\left[\theta\right] \left(\theta\right)
\end{align*}

\begin{align*}
&= \int \lambda(x) \int \frac{q(\theta)q(\theta)}{q(\theta)} g^{ij}(\theta)
\end{align*}

where $g^{ij}(\theta)$ is the covariance matrix of $\hat{\theta}(X)$ when $X$ is distributed according to $p(\cdot|\theta)$. Thus we have, finally

\begin{equation}
\delta^*\left(\left(\left[\frac{1}{\lambda}\right]\left[\theta(1)\right], \left[\right]\right) = \frac{1}{4} \int \frac{q_1(\theta)q_4(\theta)}{q(\theta)} g^{ij}(\theta) d\left[\theta\right] \left(\theta\right) - 2\int R(x) d\lambda(x),
\end{equation}

where

\begin{equation}
R(x) = q(\hat{\theta}(x))R_1(x) \int p(x|\theta) d\left[\theta\right] \left(\theta\right) + R_2(x) + R_3(x).
\end{equation}
5. Rough Derivation of a Condition for Admissibility

We consider a statistical decision problem \((\mathcal{X}, \mathcal{G}, \lambda, \mathcal{F}, \mathcal{G}, p, q, L)\) as formulated in Section 2. In addition, we suppose we are given an improper prior measure \(\mathcal{P}\) and wish to find out whether the formal Bayes procedure \(\psi\) with respect to \(\mathcal{P}\) is admissible. We shall argue that, under certain conditions, we can expect \(\psi\) to be admissible if

\[
\lim_{\lambda \to 0} \inf \int \left\{ \sum g^{ij}(\theta) \frac{\partial f(\theta)}{\partial \theta^i} \frac{\partial f(\theta)}{\partial \theta^j} + \lambda r^2(\theta) \right\} d\mathcal{P}(\theta) = 0,
\]

where

\[
g^{ij}(\theta) = \mathbb{E}_\theta (\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j),
\]

with

\[
\theta^i(x) = \frac{\int \theta^i p(x|\theta) d\mathcal{P}(\theta)}{\int p(x|\theta) d\mathcal{P}(\theta)},
\]

and \(f\) ranges over the set of continuously differentiable functions taking on the value 1 for all points of a given non-empty open set \(S\) with compact closure. We shall not be able to state a precise theorem, but the conditions needed are roughly the following:

(i) The loss function \(L\) is twice continuously differentiable in the action and parameter point together. Both \(\mathcal{G}\) and \(\mathcal{F}\) are differentiable manifolds.

(ii) The formal Bayes procedure \(\psi\) with respect to \(\mathcal{P}\) is unique.

(iii) The risk function of \(\psi\) is bounded.
(iv) The improper prior measure \( \prod \) is absolutely continuous with respect to Lebesgue measure, say with density \( \pi \) with respect to the choice of Lebesgue measure associated with the Riemannian metric determined by \( \theta \rightarrow [g(\theta)]^{-1} \).

(v) The functions \( \log \pi \) and \( \log g \) are uniformly continuous in this Riemannian metric.

(vi) It must be possible to choose the nearly minimizing functions \( \bar{f} \) in (1) so that \( \bar{f}^2 \) satisfies the conditions required in Section 4 to make the remainder in (4.24) smaller than, say, half the leading term.

If, in addition, the decision problem is sufficiently complicated, for example if it requires the complete estimation of \( \theta \) with a loss function equal to the distance between true and estimated values, we can expect (1) to be necessary as well as sufficient, although the argument in this case is even less compelling than for the sufficiency.

First let us observe that the condition (iii) that the risk function of \( \psi_0 \) is bounded is a condition on the form in which the problem is presented rather than on the problem itself. I believe this was first observed (in a different context) by Lindley. We suppose the risk function \( \rho \) of \( \psi_0 \) is everywhere finite and \( \neq 0 \). We define a new loss function \( L' \) and a new prior measure \( \prod' \) by

\[
(4) \quad L'(\theta, \varphi) = \frac{1}{\rho(\theta)} L(\theta, \varphi)
\]

and

\[
(5) \quad d\prod'(\theta) = \rho(\theta) d\prod(\theta).
\]
Then $\psi$ is also the formal Bayes procedure with respect to $\Pi'$ for
the new problem and its risk is the constant $1$. If $\Pi'$ is a bounded
measure, then $\psi$ is admissible because of the uniqueness assumption (ii).

If not, we want

$$
\int p(x|\theta)\,d\Pi'(\theta) < \infty,
$$
in order to be able to consider $\Pi'$ an improper prior measure. Here-
after we suppose this modification has already been made, if necessary.

Next we apply a slight modification of the necessary and sufficient
condition for admissibility given in the author's paper [12]. In order
that $\psi$ be almost admissible with respect to Lebesgue measure it is neces-
sary and sufficient that, for any open set $S \subset \mathcal{J}$ with compact closure, and
any $\epsilon > 0$, there exist $\delta > 0$ and a probability measure $\Pi^{(1)}$ in $\mathcal{J}$,
absolutely continuous with respect to Lebesgue measure such that,

$$
\Pi^{(1)}(S) \geq \delta,
$$
and

$$
\int \rho(\theta, \psi)\,d\Pi^{(1)}(\theta) \leq \inf_{\varphi} \int \rho(\theta, \varphi)\,d\Pi^{(1)}(\theta) + \epsilon \delta,
$$
where

$$
\rho(\theta, \varphi) = \int L(\theta, \varphi(x)) p(x|\theta)\,d\lambda(x),
$$
for any decision function $\varphi: \mathcal{X} \rightarrow \mathcal{A}$. In [ ], the condition is stated for
sets $A$ consisting of a single point, but, by examining the proof, it is
not difficult to see that the change made here is inessential.

The argument given at the end of Section 2 indicates that

$$
\int \rho(\theta, \psi)\,d\Pi^{(1)}(\theta) - \inf_{\varphi} \int \rho(\theta, \varphi)\,d\Pi^{(1)}(\theta)
\leq K \delta^*(\Pi^{(1)}, \Pi),
$$

31
where \( K \) is a constant. Now let us take \( \mathcal{V}^{(1)} \) absolutely continuous with respect to \( \mathcal{V} \), with \( q = d\mathcal{V}^{(1)} \) and apply (4.24). We find that \( \mathcal{V} \) is admissible if, for any open set \( S \subset \mathcal{J} \) with compact closure and any \( \varepsilon > 0 \) there exist \( \delta > 0 \) and a twice continuously differentiable function \( q \) on \( \mathcal{J} \) such that

\[
(11) \quad \int_{\mathcal{J}} q d\mathcal{V} = 1,
\]

\[
(12) \quad \int_{S} q d\mathcal{V} \geq \delta,
\]

\[
(13) \quad \int \sum \frac{q_{1}(\theta)q_{1}(\theta)}{q(\theta)} g^{ij}(\theta) d\mathcal{V}(\theta) \leq \varepsilon \delta,
\]

and the remainder in (4.24) is negligible. Now let

\[
(14) \quad f(\theta) = \sqrt{\frac{q(\theta)}{\delta}}.
\]

Conditions (11), (12), and (13) become

\[
(15) \quad \int_{\mathcal{J}} f^{2} d\mathcal{V} = \frac{1}{\delta},
\]

\[
(16) \quad \int_{S} f^{2} d\mathcal{V} \geq 1,
\]

and

\[
(17) \quad \int \sum f_{1}(\theta)f_{j}(\theta)g^{ij}(\theta) d\mathcal{V}(\theta) \leq \varepsilon.
\]

A simple Lagrange multiplier argument yields the form given at the beginning of this section.
6. Partial solution of the reduced problem.

In section 5 we have been led to ask for conditions on the continuous positive valued function \( \pi \) on \( \mathbb{R}^K \) and the continuous function \( g \) on \( \mathbb{R}^K \) taking positive-definite symmetric \( K \times K \) matrices as values, under which, for any open set \( S \) with compact closure,

\[
\lim_{\lambda \to 0} \inf_{f \in \mathcal{A}(S)} \int \left\{ g^{-ij}(x) \frac{\partial f(x)}{\partial x^i} \frac{\partial f(x)}{\partial x^j} + \lambda f^2(x) \right\} \pi(x) dx = 0 ,
\]

where \( \mathcal{A}(S) \) is the set of all continuously differentiable functions \( f \) for which \( f(x) = 1 \) for all \( x \in S \), and

\[
dx = dx^1 \cdots dx^K,
\]

and we use the summation convention, so that the first term in braces in (1) represents a summation over \( i,j = 1 \cdots K \). The corresponding problem for general differentiable manifolds can also arise, but we shall not try to consider it. Subject to certain conditions that have been indicated, somewhat vaguely, in Section 5, if \( \pi \) is an improper prior density for a given observational situation and \( g \) is the expected formal posterior covariance matrix (computed under \( \pi \)) when the true parameter value is \( x \), condition (1) is sufficient for formal Bayes solutions with respect to \( \pi \) to be admissible for any sufficiently smooth decision problem and, if the decision problem is sufficiently complicated, (1) can also be expected to be necessary for admissibility of these formal Bayes procedures.

We shall solve this problem in two rather trivial special cases, the one-dimensional case and the spherically symmetric case, and also make some remarks about the general problem. Since the question of whether (1) is satisfied remains unchanged when \( \pi \) and \( g \) are replaced by \( \pi' \) and \( g' \) with \( \pi'/\pi \) and the characteristic roots of \( g' \) relative to \( g \)
bounded away from 0 and ∞, the result in the spherically symmetric case is of fairly wide applicability. We observe also that the class of problems considered is invariant under continuously differentiable homeomorphisms of \( \mathbb{R}^K \), transforming as a symmetric contravariant tensor of the second rank and \( \pi \) as a scalar density. More explicitly, let \( \psi \) be a 1-1 continuously differentiable function of \( \mathbb{R}^K \) onto \( \mathbb{R}^K \) with continuously differentiable inverse, and, for \( y \in \mathbb{R}^K \) let

\[
F(y) = f(\psi^{-1}(y)),
\]

\[
\prod(y) = \pi(\psi^{-1}(y)) \det \left( \frac{\partial (\psi^{-1}(y))}{\partial y_j} \right),
\]

and

\[
G^{ij}(y) = g^{k\ell}(\psi^{-1}(y)) \frac{\partial \psi^i}{\partial x^k}(\psi^{-1}(y)) \frac{\partial \psi^j}{\partial x^\ell}(\psi^{-1}(y)).
\]

Then

\[
\int [g^{ij}(x) \frac{\partial f(x)}{\partial x^i} \frac{\partial f(x)}{\partial x^j} + \lambda f^2(x)] \pi(x)dx
\]

\[
= \int [G^{ij}(y) \frac{\partial F(y)}{\partial y^i} \frac{\partial F(y)}{\partial y^j} + \lambda F^2(y)] \prod(y)dy.
\]

Of course, these transformations (3) to (5) are also appropriate to \( f, \)
\( \pi, g \) as they arise from the statistical problem (when \( g^{-1}(x) \), the inverse of \( g(x) \), is the information matrix). It seems likely that a really satisfactory solution of our problem will exploit its tensorial character. We shall see that \( g \) and \( \pi \) seem to enter mainly (but not entirely) through their product, so that, in a way, the geometry of this problem is not that of a Riemannian manifold, but rather that associated with the contravariant tensor density \( \pi g \).

Now let us look at the one-dimensional case, where a complete solu-
tion is almost trivial. Condition (1) reduces to

\[
\lim_{\lambda \to 0} \inf_{f(x) = 1} \int_{x \in [-1,1]} \left\{ g(x) \left( \frac{df(x)}{dx} \right)^2 + \lambda f^2(x) \right\} \pi(x) dx = 0,
\]

which is equivalent to

\[
\lim_{\lambda \to 0} \inf_{f(1) = 1} \int_{1}^{\infty} \left\{ g(x) \left( \frac{df(x)}{dx} \right)^2 + \lambda f^2(x) \right\} \pi(x) dx = 0,
\]

together with the corresponding condition on \((-\infty, 0]\). Since the two problems are completely similar, we consider only

\[
(8). \quad \text{If}
\]

\[
\int_{1}^{\infty} \pi(x) dx < \infty,
\]

condition (8) is trivially satisfied with \(f(x) \equiv 1\). If

\[
\int_{1}^{\infty} \pi(x) dx = \infty,
\]

we shall see that a necessary and sufficient condition for (8) is

\[
\int_{1}^{\infty} \frac{dx}{g(x) \pi(x)} = \infty.
\]

We make a change of variable to

\[
y = \int_{1}^{x} \frac{dt}{g(t) \pi(t)},
\]

and write \(y_\infty\) for the value of \(y\) (infinite if and only if (11) holds) corresponding to \(x = \infty\). Let

\[
(13) \quad F(y) = f(x).
\]

Then}

\[
\int_{1}^{\infty} \left\{ g(x) \left( \frac{df(x)}{dx} \right)^2 + \lambda f^2(x) \right\} \pi(x) dx
\]

\[
35.
\]
\[= \int_1^\infty \left( \frac{dF(y)}{dy} \right)^2 \frac{dx}{g(x) \pi(x)} + \lambda \int_1^\infty F^2(y) \pi(x) dx\]

\[= \int_1^{y_\infty} \left\{ \left( \frac{dF(y)}{dy} \right)^2 + \lambda F^2(y) H(y) \right\} dy\]

where

(15) \[H(y) = g(x) \pi^2(x)\].

If \(y_\infty = \infty\), we can take

(16) \[F(y) = \begin{cases} 1 - \frac{y}{A_y} & \text{for } 0 \leq y \leq A_y, \\ 0 & \text{for } y \geq A_y, \end{cases}\]

where \(A_y\) is chosen so that

(17) \[\int_0^{A_y} H(y) dy = \lambda^{-\frac{1}{2}}.\]

Since \(H\) is continuous,

(18) \[\lim_{y \to 0} A_y = \infty,\]

and thus

(19) \[\int_1^\infty \left\{ \left( \frac{dF(y)}{dy} \right)^2 + \lambda F^2(y) H(y) \right\} dy\]

\[= \frac{1}{A_y} + \lambda \int_1^{A_y} \left( 1 - \frac{y}{A_y} \right)^2 H(y) dy < \frac{1}{A_y} + \sqrt{\lambda},\]

so that (8) is satisfied, by (14). On the other hand, if \(y_\infty < \infty\) and \(f\) is chosen so as to make (14) finite, we must have

(20) \[\lim_{y \to y_\infty} F(y) = \lim_{x \to \infty} f(x) = 0,\]

so that

(21) \[\int_1^{y_\infty} \left( \frac{dF(y)}{dy} \right)^2 dy \geq \frac{1}{y_\infty} \left( \int_1^{y_\infty} \frac{dF(y)}{dy} dy \right)^2 = \frac{1}{y_\infty},\]

and (14) is at least \(\frac{1}{y_\infty}\). Thus (8) cannot hold and we have proved
Proposition 1: In order that (7) hold for given continuous positive-valued functions on the real line it is necessary and sufficient that

(i) if \( \int_{-\infty}^{\infty} \pi(x) \, dx = \infty \), then \( \int_{-\infty}^{\infty} \frac{dx}{g(x) \pi(x)} = \infty \),

and

(ii) if \( \int_{-\infty}^{\infty} \pi(x) \, dx = \infty \), then \( \int_{-\infty}^{\infty} \frac{dx}{g(x) \pi(x)} = \infty \).

Next let us look at the spherically symmetric case. We shall prove

Proposition 2: Let \( \pi \) be a continuous positive-valued function on \( \mathbb{R}^K \) of the form

\[ \pi(x) = \varphi(\|x\|^2), \]  

where

\[ \|x\|^2 = \sum_{i=1}^{K} (x_i)^2, \]  

and let the continuous function \( g \) on \( \mathbb{R}^K \) to the space of positive-definite symmetric \( K \times K \) matrices be given by

\[ g^{ij}(x) = \alpha(\|x\|^2) \delta^{ij} + \beta(\|x\|^2) x_i x_j, \]  

where

\[ \delta^{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \]

Then, in order that (1) hold, that is

\[ 0 = \lim_{\lambda \to 0} \inf_{f(x)=1} \int_{\|x\| \leq 1} \left\{ g^{ij}(x) \frac{\partial f(x)}{\partial x_i} \frac{\partial f(x)}{\partial x_j} + \lambda f^2(x) \right\} \pi(x) \, dx \]

\[ = \lim_{\lambda \to 0} \inf_{f(x)=1} \int_{\|x\| \leq 1} \left\{ \alpha(\|x\|^2) \delta^{ij} + \beta(\|x\|^2) x_i x_j \frac{\partial f(x)}{\partial x_i} \frac{\partial f(x)}{\partial x_j} + \lambda f^2(x) \right\} \varphi(\|x\|^2) \, dx, \]
it is necessary and sufficient that, if

\[ \int_1^\infty \varphi(t) \frac{k}{\varphi(t)^{\frac{k}{2}}} - \frac{1}{t^2} \, dt = \infty, \]

then

\[ \int_1^\infty \frac{dt}{[\alpha(t) + t\beta(t)]\varphi(t)^{\frac{k}{2}}} = \infty. \]

Proof: Because the integral in (26) is a convex function of \( f \) invariant under the compact group of orthogonal transformations \( T : \mathcal{R}^K \to \mathcal{R}^K \) (operating by taking \( f \) into \( \overline{Tf} \) defined by

\[ (\overline{Tf})x = f(T^{-1}x), \]

it follows that the condition (26) is equivalent to the corresponding condition with \( f \) restricted to be invariant under orthogonal transformations, say

\[ f(x) = \xi(\|x\|^2), \text{ with } \xi(1) = 1. \]

Thus (26) is equivalent to

\[ 0 = \lim_{\mathcal{W}_0} \inf_{\xi(1)} \int \left\{ \left[ \alpha(\|x\|^2) x_i x_j + \beta(\|x\|^2) x_i x_j \right] \frac{\partial \xi(\|x\|^2)}{\partial x_i} \frac{\partial \xi(\|x\|^2)}{\partial x_j} \right. \]

\[ + \lambda_2^2(\|x\|^2) \varphi(\|x\|^2) \, dx \]

\[ = 4 \lim_{\mathcal{W}_0} \inf_{\xi(1)} \int \left\{ \left[ \alpha(\|x\|^2) x_i x_j + \beta(\|x\|^2) x_i x_j \right] \xi^2(\|x\|^2) x_i x_j \right. \]

\[ + \lambda_2^2(\|x\|^2) \varphi(\|x\|^2) \, dx \]

\[ = C \lim_{\mathcal{W}_0} \inf_{\xi(1)} \int \left\{ \left[ t\alpha(t) + t^2\beta(t) \right] \xi^2(t) \right. \]

\[ + \lambda_2^2(t) \left\{ t^2 - \frac{1}{t^2} \right\} \varphi(t) \, dt, \]

\[ = \int_1^\infty \varphi(t) \frac{k}{\varphi(t)^{\frac{k}{2}}} - \frac{1}{t^2} \, dt, \]
where \( C \) is a positive constant. By applying Proposition 1, we obtain the conclusion of Proposition 2.

A partial solution of the general problem, which may be useful in special cases, can be obtained by observing that, if the contour surfaces of \( f \) are preassigned, the problem is reduced to the one-dimensional case, which is solved by Proposition 1. Let \( \rho \) be a continuously differentiable, positive valued function on \( \mathcal{R}^K \) such that

\[
(32) \quad \rho(x) = 0 \text{ for all } x \in S.
\]

and, for all real \( r \),

\[
(33) \quad \int \limits_{\rho(x) < r} \pi(x) \, dx < \infty,
\]

and, as \( h \downarrow 0 \),

\[
(34) \quad \int \limits_{r < \rho(x) \leq r + h} \pi(x) \, dx = o(h),
\]

uniformly for \( r \) in any compact set.

Then

\[
(35) \quad \lim_{\lambda \downarrow 0} \inf_{f \in \mathcal{D}(S)} \int \left\{ g^{ij}(x) \frac{\partial f(x)}{\partial x^i} \frac{\partial f(x)}{\partial x^j} + \lambda \varepsilon^2(x) \right\} \pi(x) \, dx
\]

\[
\leq \lim_{\lambda \downarrow 0} \inf_{\xi(0) = 1} \int \left\{ g^{ij}(x) \frac{\partial \xi(x)}{\partial x^i} \frac{\partial \xi(x)}{\partial x^j} + \lambda \varepsilon^2(\rho(x)) \right\} \pi(x) \, dx
\]

\[
= \lim_{\lambda \downarrow 0} \inf_{\xi(0) = 1} \int_0^\infty \left\{ \varepsilon^2(r) \frac{d}{dr} \int \frac{\partial \xi(x)}{\partial x^i} \frac{\partial \xi(x)}{\partial x^j} \pi(x) \, dx \right. \left. \frac{\partial \rho(x)}{\partial x^i} \frac{\partial \rho(x)}{\partial x^j} \right\} \pi(x) \, dx
\]

\[
+ \lambda \varepsilon^2(r) \frac{d}{dr} \int \frac{\partial \xi(x)}{\partial x^i} \frac{\partial \xi(x)}{\partial x^j} \pi(x) \, dx \right. \left. \frac{\partial \rho(x)}{\partial x^i} \frac{\partial \rho(x)}{\partial x^j} \right\} \pi(x) \, dx
\]

By Proposition 1, a necessary and sufficient condition for the right hand side of (35) to be 0 is that either
or \( \pi \) is integrable (a trivial case we exclude in the following discussion). Thus for (1) to hold it is sufficient that, for some choice of \( \rho \), satisfying (32)-(34), condition (36) is satisfied. On the other hand, if (1) holds, we can nearly obtain equality in (35) by determining \( \rho \) from the contour surfaces of a suitably chosen, nearly minimizing \( f \) for small \( \lambda \), so that, again by Proposition 1, the condition stated below (36) is also necessary. Thus we have

Proposition 3: Under the conditions stated at the beginning of this section, with \( \pi \) not integrable, in order that (1) hold it is necessary and sufficient that there exist a continuously differentiable positive valued function \( \rho \) on \( \mathbb{R}^K \) such that (32)-(34) and (36) are satisfied.

Proposition 3 can be specialized to obtain more explicit sufficient conditions. First we observe that \( g \) and \( \pi \) enter (36) only in the combination \( g\pi \), and although \( \pi \) enters separately in (33) a sufficient condition (for (33)), compactness of the sets \( \{x: \rho(x) \leq r\} \) can be given without reference to \( \pi \). Now \( g\pi \) is not a tensor, but rather a contra-variant tensor density. This means that (3)-(6) imply

\[
\delta^{ij}(y) \iota(y) = g^{k\ell}(\psi^{-1}(y)) \pi(\psi^{-1}(y)) \frac{\partial \psi^i}{\partial x^k}(\psi^{-1}(y)) \frac{\partial \psi^i}{\partial y^\ell}(\psi^{-1}(y)) \det \left( \frac{\partial(\psi^{-1}(y))}{\partial y^i} \right),
\]

which differs from the formula (5) for the transformation of a contra-variant tensor by the final factor of a determinant. In the case \( K \geq 3 \), the geometry determined by such a tensor density \( g\pi \) is equivalent to the (Riemannian) geometry determined by the covariant tensor \( h \) defined
by

\[(38) \quad h = (g^\pi)^{-1} \frac{1}{\det(g^\pi)^{1/2} K^{-2}}.\]

It should be possible to obtain more explicit necessary conditions and sufficient conditions in terms of this geometry. For example, it is sufficient that (36) hold with \( \rho(x) \) equal to the distance from a point in \( S \) when this distance is large. The case \( K = 2 \) must be treated separately.

I am indebted to R. Finn, K. Loewner, and A. Novikoff for some helpful conversations in connection with the material of this section. In particular, Finn brought to my attention a paper of Meyers and Serrin [9] dealing with a similar, but perhaps more difficult, problem.
REFERENCES


