The Annals of Statistics
1984, Vol. 12, No. 3, 958–970

DERIVING POSTERIOR DISTRIBUTIONS FOR A LOCATION PARAMETER: A DECISION THEORETIC APPROACH[1]

BY CONSTANTINE A. GATSONIS

University of Massachusetts

In this paper we develop a decision theoretic formulation for the problem of deriving posterior distributions for a parameter θ, when the prior information is vague. Let $\pi(d\theta)$ be the true but unknown prior, $Q_\pi(d\theta \mid X)$ the corresponding posterior and $\delta(d\theta \mid X)$ an estimate of the posterior based on an observation X. The loss function is specified as a measure of distance between $Q_\pi(\cdot \mid X)$ and $\delta(\cdot \mid X)$, and the risk is the expected value of the loss with respect to the marginal distribution of X. When θ is a location parameter, the best invariant procedure (under translations in R^n) specifies the posterior which is obtained from the uniform prior on θ. We show that this procedure is admissible in dimension 1 or 2 but it is inadmissible in all higher dimensions. The results reported here concern a broad class of location families, which includes the normal.

1. Introduction. The problem of assessing posterior probabilities for a parameter when the prior information is vague has led to the development of several criteria for selecting noninformative priors: among others, Jeffrey's rule (see Box and Tiao, 1973), Savage's "principle of precise measurement" (see Edwards et al, 1963), etc. More recently, the problem was also considered with a view towards developing "reference posteriors", which would "approximately describe the inferential content of the data without incorporating any further information" (see Bernardo, 1979). In the case of a location parameter, these approaches recommend the use of a uniform prior, a technique which has been extensively discussed and criticized in recent years (see, for example, Box and Tiao, 1973, and Dawid, Stone and Zidek, 1973). From the point of view of classical decision theory, a major criticism of the use of uniform priors is that, in higher dimensions, they lead to posteriors which provide inadmissible point estimators and confidence procedures (see, for example, Stein, 1956, 1961; Brown, 1966; and Joshi, 1967).

In this paper we develop a decision theoretic formulation for the problem stated at the beginning. Let $\pi(d\theta)$ be the true but unknown prior on the parameter θ, $Q_\pi(d\theta \mid X)$ the corresponding posterior and $\delta(d\theta \mid X)$ an estimate of the posterior based on an observation X. The loss function will be given by the L_2 distance between $Q_\pi(\cdot \mid X)$ and $\delta(\cdot \mid X)$, and the risk of a procedure δ will be the expected value of the loss with respect to the marginal distribution of X. Various properties

Received January 1983; revised December 1983.

[1] The research reported here is part of the author's Ph.D. Thesis (1981), submitted to Cornell University.

AMS 1980 *subject classifications.* Primary 62C10, 62C15; secondary 62A99, 62F15, 62H12.

Key words and phrases. Location parameter, noninformative priors, best invariant procedures, admissibility, Stein phenomenon, normal mean.

of the procedure $\delta(\cdot \mid X)$ will be investigated, in particular admissibility, invariance and minimaxity.

The discussion in this paper is concerned with θ being an arbitrary location parameter. Special attention is given to the case of the normal mean. In Section 2 we give an easy calculation showing that the best invariant procedure (under translations in R^n) can be obtained by assuming a uniform prior on θ, i.e., $\bar{\delta}(\theta \mid X) = f(X - \theta)$, where f denotes the density of X. In Sections 3 and 4 we show that an analog to the classical Stein effect arises in this context. The best invariant procedure is admissible when the dimension n of θ is 1 or 2 but it is inadmissible in all higher dimensions. In fact, if $n \geq 3$ then $\bar{\delta}$ is dominated by procedures of the form $\delta(\theta \mid X) = f(x + \gamma(X) - \theta)$, where $X + \gamma(X)$ is a Stein-type point estimator of θ. In the case of the normal mean, explicit choices of γ are given, while in the general location parameter case δ is shown to dominate $\bar{\delta}$ as a certain parameter of γ tends to its limiting value (an analogous situation appears in Stein, 1956, and Brown, 1966).

The contents of this paper bear some resemblance to those of a recent paper by Eaton (1981). However, Eaton's approach and formulation is quite different than the one presented here.

The Bayesian formulations we mentioned at the outset do not seem to differentiate, within their own framework, between low and high dimensional location parameters. The approach we present enables us to do so and, while it concurs with the suggestions for using uniform priors in lower dimensions, it argues against their use in higher dimensions. The proposed improvements over the best invariant procedure provide point estimates and confidence procedures for θ, which are superior to the usual ones.

2. Definitions and preliminary results. Let X be an observation from an n-dimensional location family with density $f(x - \theta)$. Let $\pi(\theta)$ be the (unknown) prior density of θ (all densities are with respect to Lebesgue measure in R^n). We denote by $Q_\pi(\theta \mid x)$ the posterior density of θ and by $m_\pi(x)$ the marginal density of X. We always consider priors for which $Q_\pi(\cdot \mid x) \in L_2(R^n)$ for each x. The space Π of priors under consideration will be specified in each case.

An estimate of $Q_\pi(\cdot \mid x)$ is a function h in the action space $\mathscr{D}$, where $\mathscr{D} = \{h : h \in L_2(R^n),\ 0 \leq h(\theta)$, for all θ and $0 \leq \int_{R^n} h(\theta)\, d\theta \leq 1\}$. The loss function is given by:

$$L(\pi, h, x) = \| h - Q_\pi(\cdot \mid x) \|^2,$$

where $\| \cdot \|$ denotes the usual L_2 distance.

A (nonrandomized) procedure δ is a measurable map from the sample space into $\mathscr{D}$, i.e., $\delta(\cdot \mid x) \in \mathscr{D}$, for each x. The risk of δ is given by:

$$R(\pi, \delta) = E(L(\pi, \delta(X), X)),$$

where the expectation is taken with respect to m_π. (We will often write $\delta(x)$ and $Q_\pi(x)$ instead of $\delta(\cdot \mid x)$ and $Q_\pi(\cdot \mid x)$.) A procedure δ will be called "admissible" iff there does not exist another procedure δ_1 such that $R(\pi, \delta_1) \leq R(\pi, \delta)$ for all $\pi \in \Pi$, with strict inequality holding for some $\pi_0 \in \Pi$.

Consider now the group G of translations in R^n and the corresponding group $\bar{G}$ of translation operators acting on functions on R^n. $\mathscr{D}$ is invariant under $\bar{G}$. Assume Π is an arbitrary set of densities, invariant under $\bar{G}$. It is easy to verify that the loss function is invariant under these group actions, i.e.

$$L(\bar{g}_c(\pi), \bar{g}_c(h), g_c(x)) = L(\pi, h, x), \quad \text{for all} \quad c \in R^n.$$

A decision procedure δ will be called "invariant" iff $\delta(g_c(x)) = \bar{g}_c(\delta(x))$, i.e., $\delta(\theta \mid x + c) = \delta(\theta - c \mid x)$, for all $x, c, \theta \in R^n$.

PROPOSITION 2.1. *The procedure $\bar{\delta}$ defined by $\bar{\delta}(\theta \mid x) = f(x - \theta)$ is a best invariant procedure.*

PROOF. If δ is any invariant procedure then we can easily show that:

$$R(\pi, \delta) = \int_{R^n} dz \left[\int_{R^n} (\delta(z \mid 0) - Q_\pi(z + x \mid x))^2 m_\pi(x)\, dx \right].$$

The integral inside the brackets is minimized for each z by the choice:

$$\bar{\delta}(z \mid 0) = \int_{R^n} Q_\pi(z + x \mid x) m_\pi(x)\, dx = f(-z). \quad \Box$$

REMARK. (i) It is easy to see that $\bar{\delta}$ can be obtained as the posterior if we assume the uniform prior on θ. For general results of this type see Berger (1980a, pages 259, 263).

(ii) The proof of the proposition shows that $\bar{\delta}$ is unique up to equivalence with respect to Lebesgue measure.

(iii) If $X \sim N(\theta, A)$ then $\bar{\delta}(X)$ is a $N(X, A)$ density. However, if $X_1, \cdots, X_m$, $m \geq 1$, is an random sample from $N(\theta, A)$, we cannot automatically reduce the problem by sufficiency, since $\bar{X}$ is not necessarily a sufficient statistic for the joint marginal distribution of the X_i's. We can transform the problem in the following way which is common in location problems (see Brown, 1966, for example). Let $X = X_1$, $Y_i = X_{i+1} - X_1$, $i = 1, \cdots, m - 1$. The joint density of X, Y is given by: $p(x - \theta, y)\, dx\nu(dy)$ where

$$p(x, y) = \frac{f(x)f(x + y_1) \cdots f(x + y_{m-1})}{\int f(x)f(x + y_1) \cdots f(x + y_{m-1})\, dx}$$

and $\nu(dy) = (\int f(x)f(x + y_1) \cdots f(x + y_{m-1})\, dx) \prod_{i=1}^{m-1} dy_i$. If loss and risk are defined as above and a procedure is "invariant" iff $\delta(\theta \mid x + c, y) = \delta(\theta - c \mid x, y)$, for all $\theta, x, y, c \in R^n$, we can show that the best invariant procedure is $\bar{\delta}(\theta \mid x, y) = p(x - \theta, y)$, i.e., $\theta \sim N(\bar{X}, (1/m)A)$. In fact, the results in the rest of this paper can be shown to hold with $\bar{X}$ instead of X and $(1/m)A$ instead of A.

3. Admissibility.

(I). The normal case. We assume here that $X \sim N(\theta, A)$ with the covariance matrix A known, and Π = set a conjugate priors for θ, i.e., $\theta \sim N(\mu, B)$ for some vector $\mu \in R^n$ and nonsingular matrix B.

THEOREM 3.1. *The best invariant procedure $\bar{\delta}$ is admissible in dimension $n = 1$ or 2.*

PROOF. It suffices to show that $\bar{\delta}$ cannot be dominated over the set Π_B of all conjugate priors with the same covariance matrix B, where B is arbitrary but fixed. The family of marginal distributions of X becomes a location parameter family now, with location vector μ. Thus, our problem fits in the framework developed by Brown and Fox (1974a, b) for proving the admissibility of procedures in one- and two-dimensional location parameter families.

Following the Brown-Fox notation we write

$$L(\mu, h, x) = W(\bar{g}_{-\mu}(h), x - \mu).$$

Let $a = (2\pi)^{-n/2}(\det(A^{-1} + B^{-1})^{-1})^{-1/2}$. It is easy to see that we only need to consider procedures such that $\delta(\theta \mid x) \le a$, for all $\theta, x \in R^n$. Hence, without loss of generality we can restrict our action space to

$$\mathscr{D}' = \left\{h: h \in L_2(R^n);\ 0 \le h(\theta) \le a \text{ for all } \theta \in R^n,\ 0 \le \int_{R^n} h(\theta)\, d\theta \le 1\right\}.$$

Clearly now $W(h, x) \le 4a$, for all $h \in \mathscr{D}'$, $x \in R^n$.

We state here a version of the Brown-Fox sufficient conditions for admissibility of the best invariant procedure and refer the reader to the above mentioned papers, as well as to Brown (1966), for the proof of sufficiency.

(A) The action space is compact in the weak topology of $L_2(R^n)$ and the loss function $L(\mu, \cdot, x)$ is lower semicontinuous for each pair (μ, x).

With condition (A) being satisfied, the following are sufficient conditions for the admissibility of $\bar{\delta}$ in dimension $n = 1$.

(B1) There exists a best invariant procedure $\bar{\delta}$, with risk $\bar{R} < \infty$. $\bar{\delta}$ is essentially uniquely determined, i.e., if δ is any other invariant procedure such that $R(\mu, \delta) = \bar{R}$ then $\delta(x) = \bar{\delta}(x)$ a.e. (dx).

(B2) If $\{\delta_i\}$ is a sequence of invariant procedures such that $R(\mu, \delta_i) \to \bar{R}$, then $\delta_i(0) \to \bar{\delta}(0)$ (weakly). Conversely, if $\{\delta_i\}$ is a sequence of invariant procedures such that $\delta_i(0) \to \bar{\delta}(0)$ then

$$\int [W(\bar{\delta}(x), x) - W(\delta_i(x), x)]^+ m_0(x)\, dx \to 0.$$

(B3) $\int |x| W(\bar{\delta}(x), x) m_0(x)\, dx < \infty$.

(B4) $\int_0^\infty dy\, [\sup_{\delta \text{invariant}} \int_{-y}^{y} (W(\bar{\delta}(x), x) - W(\delta(x), x)) m_0(dx)] < \infty$.

By m_0 we denote here the marginal corresponding to the prior with $\mu = 0$.

In dimension $n = 2$ the sufficient conditions (in addition to (A)) are:

(C1) There exists a best invariant procedure $\bar{\delta}$ with risk $\bar{R} < \infty$.

(C2) The loss W is bounded above and $\int |x|^2 W(\bar{\delta}(x), x) m_0(x)\, dx < \infty$.

(C3) There exists a positive valued nonincreasing function K on R^+ such that $\int_0^\infty K(y)\, dy < \infty$ and, for any invariant procedure δ and $\lambda > 0$ we have

$$\int_{|x|<\lambda} [W(\bar{\delta}(x), x) - W(\delta(x), x)] m_0(x)\, dx \leq K(\lambda) \left[\int_{R^2} (W(\delta(x), x) - W(\bar{\delta}(x), x)) m_0(x)\, dx \right]^{1/2}$$

and

$$\int_{|x|>\lambda} [W(\bar{\delta}(x), x) - W(\delta(x), x)]^+ m_0(x)\, dx \leq K(\lambda) \left[\int_{R^2} (W(\delta(x), x) - W(\bar{\delta}(x), x)) m_0(x)\, dx \right]^{1/2}.$$

REMARK. As Brown and Fox note, (C3) is a condition on the local behavior of the loss near $\bar{\delta}(x)$ and it implies that if δ is another best invariant procedure then

$$W(\bar{\delta}(x), x) = W(\delta(x), x) \quad \text{a.e.} \quad (dx).$$

The proof of Theorem 3.1 will be completed now by establishing the following three lemmas.

LEMMA 3.1. *Condition* (A) *holds for* $\mathscr{D}'$ *and* L.

PROOF. Let S be the ball in $L_2(R^n)$ with center at 0 and radius $\sqrt{a}$. S is weakly compact by the Banach-Alaoglu theorem and, clearly, $\mathscr{D}' \subset S$. For any sequence $\{h_n\} \subset \mathscr{D}'$ there exists a subsequence $\{h_{n_k}\}$ such that $h_{n_k} \to h \in S$. We will show that, in fact, $h \in \mathscr{D}'$.

Let $T = \{x : h(x) < 0\}$ and $g(x) = I_T(x) e^{-|x|^2}$. Then $\int h_{n_k}(x) g(x)\, dx \geq 0$ but $\int h(x) g(x) < 0$ if T has nonzero Lebesgue measure. Hence T has Lebesgue measure zero. Similarly we show that $h(x) \leq a$, a.e. (dx). Finally, for each $M > 0$,

$$\int_{|x|\leq M} h_{n_k}(x)\, dx \to \int_{|x|\leq M} h(x)\, dx,$$

which implies that $\int_{|x|\leq M} h(x)\, dx \leq 1$. This proves that $h \in \mathscr{D}'$.

The lower semicontinuity of $L(\mu, \cdot, x)$ follows from the fact that, if $h_i \to h$ weakly, then $\liminf \| h_i \| \geq \| \liminf h_i \| = \| h \|$ (see Liusternik and Sobolev, 1974, page 148). □

LEMMA 3.2. *Let* $n = 1$. *Then conditions* (B1)–(B4) *are satisfied.*

PROOF. We remarked earlier that $\bar{\delta}$ is essentially unique. Let $\{\delta_i\}$ be a sequence of invariant procedures such that $R(\mu, \delta_i) \to R(\mu, \bar{\delta})$. If $\{\delta_{i_k}(0)\}$ is a (weakly) convergent subsequence of $\{\delta_i(0)\}$ and $h = \lim \delta_{i_k}(0)$, we can define a procedure δ by $\delta(\theta \mid x) = h(x - \theta)$. Note that δ is an invariant procedure and, in fact, $\delta_{i_k}(x) \to \delta(x)$, for all x. The lower semicontinuity of the loss function and Fatou's lemma give:

$$R(\mu, \bar{\delta}) = \lim\inf R(\mu, \delta_{i_k}) \geq R(\mu, \delta).$$

Hence $\delta_i(0) \to \bar{\delta}(0)$.

For the converse, note that the previous argument shows that $[W(\bar{\delta}(x), x) - W(\delta_i(x), x)]^+ \to 0$. Use now the dominated convergence theorem to establish (B2).

Condition (B3) is obvious, having noted that the loss is bounded above. To establish (B4), note that for each invariant procedure δ and $y > 0$ we have:

$$\int_{-y}^{y} [W(\bar{\delta}(x), x) - W(\delta(x), x)] m_0(x)\, dx \leq 4a \int_{|x| \geq y} m_0(x)\, dx. \quad \square$$

LEMMA 3.3. *Let $n = 2$. Then conditions* (C1)–(C3) *are satisfied.*

PROOF. The first two conditions clearly hold. The verification of (C3) is similar to (2.14) in Brown and Fox (1974b). Let

$$K(\lambda) = 2\left[\int_{|x|>\lambda} \| \bar{\delta}(x) - Q_0(x) \|^2 m_0(x)\, dx\right]^{1/2}.$$

The function K is clearly nonincreasing and $\int_0^\infty K(\lambda)\, d\lambda \leq c \int_0^\infty e^{-\lambda^2}\, d\lambda < \infty$ for some positive constant c. Here Q_0 denotes the posterior corresponding to the prior with $\mu = 0$.

If $\langle \cdot, \cdot \rangle$ denotes the L_2 inner product, then for any invariant procedure δ and any $\lambda > 0$ we have:

$$\begin{aligned}
&\int_{|x|<\lambda} [W(\bar{\delta}(x), x) - W(\delta(x), x)] m_0(x)\, dx \\
&\qquad \leq 2 \int_{|x|<\lambda} \langle \bar{\delta}(x) - \delta(x), \bar{\delta}(x) - Q_0(x) \rangle m_0(x)\, dx \\
&\qquad = 2 \int_{|x|\geq\lambda} \langle \delta(x) - \bar{\delta}(x), \bar{\delta}(x) - Q_0(x) \rangle m_0(x)\, dx \\
&\qquad \leq 2K(\lambda)\left[\int \| \delta(x) - \bar{\delta}(x) \|^2 m_0(x)\, dx\right]^{1/2} \\
&\qquad = 2K(\lambda)\left[\int (\| \delta(x) - Q_0(x) \|^2 - \| \bar{\delta}(x) - Q_0(x) \|^2) m_0(x)\, dx\right]^{1/2}.
\end{aligned}$$

The second part of (C3) is verified similarly. $\square$

REMARK. An easy calculation shows that the risk of $\bar{\delta}$ is constant for all μ with fixed B, but depends on B in such a way that $\sup_{(\mu,B)} R(\mu, \bar{\delta}) = \infty$. Hence $\bar{\delta}$ is minimax over each Π_B but not over Π.

II. The general location parameter case. The admissibility results for the normal case can be extended to the case where X is an observation from an arbitrary location parameter family with density $f(x - \theta)$. The space Π is a location parameter family of priors on θ, i.e., $\Pi = \{\pi(\theta - \zeta), \zeta \in R^n\}$. Clearly more general sets Π can be allowed, which are collections of location parameter families, each of which satisfies the regularity conditions we set below.

It can be readily verified that the proof of Theorem 3.1 goes through in this more general setting without any essential changes, if the following conditions hold:

In dimension $n = 1$ we require

(G1) $\exists$ a constant c such that $Q_\zeta(\theta \mid x) \le c$, for all $\theta, \zeta, x \in R^n$, and

(G2) $\int |x| m_0(x)\, dx < \infty$.

In dimension $n = 2$ we require that (G1) hold and also

(G3) $\int |x|^{2+\varepsilon} m_0(x)\, dx < \infty$, for some $\varepsilon > 0$.

REMARKS. (i) It is easy to show that (G2) is implied by

(G2)′ $\int |x| f(x)\, dx < \infty$ and $\int |\theta| \pi(\theta)\, d\theta < \infty$.

(ii) Note that if P_0 denotes the probability measure with density m_0 then, for each $\lambda > 0$, (G3) implies that $P_0(|X| > \lambda) \le d/|\lambda|^{2+\varepsilon}$. Hence

$$\int_0^\infty \sqrt{P_0(|X| > \lambda)}\, d\lambda \le 1 + \int_1^\infty \left(\frac{d}{\lambda^{2+\varepsilon}}\right)^{1/2} d\lambda < \infty.$$

This guarantees that the function K we defined in Lemma 3.3 is integrable.

EXAMPLE. *An unknown scale parameter problem.* Let X be a univariate normal random variable with known mean μ and unknown precision τ. We assume a class of conjugate priors on τ, namely $\pi_\beta \sim$ Gamma (½, β), with β unknown. The problem is to estimate the unknown posterior Q_β when the loss function is given by $L(\beta, h, x) = \int_0^\infty (h(\tau) - Q_\beta(\tau \mid x))^2 \tau\, d\tau$ and risk is calculated with respect to the marginal density $m_\beta(x)$.

The following well-known logarithmic transformation will bring this problem into our location parameter framework. Let $Z = \log(|X - \mu|^2)$, $\theta = -\log \tau$ and $\xi = \log \beta$. A straightforward calculation will show that the density of Z is

$$f(z - \theta) = (1/\sqrt{2\pi})\exp[\tfrac{1}{2}(z - \theta) - \tfrac{1}{2}\exp(z - \theta)]$$

and the prior on θ is

$$\pi(\theta - \xi) = (1/\sqrt{\pi})\exp(\tfrac{1}{2}(\xi - \theta) - \exp(\xi - \theta)).$$

Let Q_ξ be the posterior and m_ξ the marginal density of Z. Denote by L^* the L_2

loss function for estimating Q_ξ and by R^* the corresponding risk. It can be shown that if δ^* is any procedure in the transformed problem and δ is a procedure in the original one defined by $\delta(\tau \mid X) = (1/\tau)\delta^*(-\log \tau \mid 2 \log|X - \mu|)$ then $R^*(\xi, \delta^*) = R(\beta, \delta)$. Hence, if δ^* is admissible in the transformed problem then δ is admissible in the original one.

We can verify that, for the transformed problem (G1) and (G2) hold. Hence $\bar{\delta}^*(\theta \mid Z) = f(Z - \theta)$ is admissible. It follows that the procedure $\bar{\delta}(\tau \mid X) = (1/\sqrt{2\pi\tau})\,|X - \mu|\exp(-\tfrac{1}{2}\tau\,|X - \mu|^2)$ is admissible in the original problem. It is easy to see that $\bar{\delta}$ can be obtained by placing the noninformative prior on τ, namely $\pi(\tau) = 1/\tau$.

4. Inadmissibility

(I) The normal case. We assume here the setting of Section (3I) but allow Π to be of a more general type, namely any set of piecewise continuous prior densities which contains the conjugate priors.

THEOREM 4.1. *The best invariant procedure $\bar{\delta}$ is inadmissible in dimension $n \geq 3$.*

PROOF. Let $f(x - \theta)$ denote the n-dimensional $N(\theta, A)$ density. We will show that $\bar{\delta}$ can be dominated by procedures of the form $\delta(\theta \mid x) = f(x + \gamma(x) - \theta)$, where γ will be explicitly determined as a solution of a differential inequality (in a manner similar to Brown, 1979, or Berger, 1980b).

We write the difference of the risks as follows:

$$R(\pi, \delta) - R(\pi, \bar{\delta})$$

$$= E[\|\delta(X)\|^2 - \|\bar{\delta}(X)\|^2 + 2\langle Q_\pi(X), \bar{\delta}(X) - \delta(X)\rangle]$$

$$= 2\int \pi(\theta)\left[\int f(x - \theta)(f(x - \theta) - f(x + \gamma(x) - \theta))\,dx\right]d\theta.$$

Hence, it suffices to find $\gamma : R^n \to R^n$ such that

$$E[f(X - \theta) - f(X + \gamma(X) - \theta)] < 0, \quad \text{for all} \quad \theta, \tag{4.1}$$

where $X \sim N(\theta, A)$ in the expectation.

Using the inequality $1 - e^{-x} \leq x$ we get:

$$f(x - \theta) - f(x + \gamma(x) - \theta)$$

$$= f(x - \theta)[1 - \exp(-\tfrac{1}{2}\gamma^T(x)A^{-1}\gamma(x) - (x - \theta)^T A^{-1}\gamma(x))]$$

$$\leq f(x - \theta)[\tfrac{1}{2}\gamma^T(x)A^{-1}\gamma(x) + (x - \theta)^T A^{-1}\gamma(x)].$$

Clearly now

$$\text{LHS of (4.1)} \leq cE[\tfrac{1}{2}\gamma^T(X)A^{-1}\gamma(X) + \tfrac{1}{2}(x - \theta)^T(A/2)^{-1}\gamma(X)],$$

where c is positive constant and $X \sim N(\theta, \tfrac{1}{2}A)$ in the expectation. Using Stein's identity (see Stein, 1981) we get that:

$$\text{LHS of (4.1)} \leq (c/2)E[\gamma^T(X)A^{-1}\gamma(X) + \nabla \cdot \gamma(X)].$$

Hence, it suffices to choose γ so that it satisfies the differential inequality:

$$\gamma^T(x)A^{-1}\gamma(x) + \nabla \cdot \gamma(x) < 0. \tag{4.2}$$

Many Stein-type estimators of the multivariate normal mean will satisfy (4.2). In particular, let $\gamma_\lambda(x) = -(\lambda/x^TA^{-1}x)x$. Then (4.2) holds for all $0 < \lambda < n - 2$. The function γ which corresponds to the positive-part Stein estimator also satisfies (4.2). □

REMARKS. (i) Although (4.2) is equality if $\lambda = n - 2$, it is easy to show that (4.1) is still valid for γ_{n-2}. (ii) The procedures which dominate $\bar{\delta}$, if used as posteriors, lead to point estimators which are superior to the sample mean. They also lead to confidence procedures which are superior to the usual confidence sets centered at the sample mean, as was recently demonstrated by Hwang and Casella (1982).

II. The general location parameter case. We assume here the setting of Section (3.II). We will show that the best invariant procedure $\bar{\delta}$ can be dominated by procedures of the form $\delta(\theta \mid x) = f(x + \gamma(x) - \theta)$, for some suitably chosen function γ. The proof will involve taking series expansions (as in the normal case) but the treatment of the error terms is much more involved. In fact, we will only give the form of γ and show that the corresponding procedure δ dominates $\bar{\delta}$ as a certain parameter of γ tends to its limiting value.

The argument given at the beginning of the proof of Theorem 4.1 shows that $R(\pi, \delta) - R(\pi, \bar{\delta}) < 0$ if we can pick a $\gamma: R^n \to R^n$ such that

$$E(f(X - \theta) - f(X + \gamma(X) - \theta)) < 0, \quad \text{for all} \quad \theta \tag{4.3}$$

where $X \sim f(x - \theta)$ in the expectation.

We proceed now to get the form of γ, via the following heuristic argument. A simple change of variables shows that (4.1) is equivalent to

$$E(f(Z) - f(Z + \gamma(Z + \theta))) < 0, \quad \text{for all} \quad \theta \tag{4.4}$$

where $Z \sim f(z)$ in the expectation. By expanding f and γ in their Taylor series around Z and θ respectively we get:

$$\begin{aligned} &E(f(Z) - f(Z + \gamma(Z + \theta))) \\ &\quad = -E(\gamma(\theta) \cdot \nabla f(Z) + (\gamma'(\theta)Z) \cdot \nabla f(Z) + R_2 \cdot \nabla f(Z) \\ &\qquad + \tfrac{1}{2}\gamma^T(\theta)H(Z)\gamma(\theta) + \gamma^T(\theta)H(Z)\gamma'(\theta)Z \\ &\qquad + \gamma^T(\theta)H(Z)R_2 + \tfrac{1}{2}Z^T\gamma'(\theta)H(Z)\gamma'(\theta)Z \\ &\qquad + Z^T\gamma'(\theta)H(Z)R_2 + \tfrac{1}{2}R_2^TH(Z)R_2 + R_1) \end{aligned} \tag{4.5}$$

where $H(z) = (D_iD_jf(z))$ $i, j = 1, \cdots, n$, R_1 is the remainder of f and R_2 is the remainder of γ.

Let $c = \frac{1}{2} \int_{R^n} f^2(z)\, dz$ and $M = (m_{ij})$, $i, j = 1, \cdots, n$ where $m_{ij} = \frac{1}{2} \int_{R^n} D_i f(z) D_j f(z)\, dz$. Integration by parts shows that:

$$E(Z_i D_j f(Z)) = \begin{cases} 0, & i \neq j \\ -c, & i = j \end{cases} \tag{4.6a}$$

$$E(H(Z)) = -2M \tag{4.6b}$$

and

$$E(\nabla f(Z)) = 0. \tag{4.6c}$$

Substituting into (4.5) we get:

$$\begin{aligned} &E[f(Z) - f(Z + \gamma(Z + \theta))] \\ &\quad = c\nabla \cdot \gamma(\theta) + \gamma^T(\theta) M \gamma(\theta) \\ &\qquad - E[R_2 \cdot \nabla f(Z) + \gamma^T(\theta) H(Z) \gamma'(\theta) Z + \gamma^T(\theta) H(Z) R_2 \\ &\qquad + \tfrac{1}{2} Z^T \gamma'(\theta) H(Z) \gamma'(\theta) Z + Z^T \gamma'(\theta) H(Z) R_2 + \tfrac{1}{2} R_2^T H(Z) R_2 + R_1]. \end{aligned} \tag{4.7}$$

If M is nonnegative definite, the first two terms in the RHS of (4.7) suggest the following form of γ:

$$\gamma(x) = -\frac{ax}{b + x^T M x}, \quad \text{where} \quad b > 0 \quad \text{and} \quad 0 < a < c(n-2). \tag{4.8}$$

For such a γ we have:

$$c\nabla \cdot \gamma(\theta) + \gamma^T(\theta) M \gamma(\theta) = \frac{-can}{b + \theta^T M \theta} + \frac{(2ac + a^2)\theta^T M \theta}{(b + \theta^T M \theta)^2} < 0.$$

The following theorem shows that, under regularity conditions, the contribution of the rest of the terms in (4.7) is negligible as $b \to \infty$. ($\| M \|$ will denote the Euclidean norm of a matrix M, i.e., $\| M \|^2 = \sum m_{ij}^2$).

THEOREM 4.2. *Let $n \geq 3$ and f satisfy the following conditions*:

(C1) *There exists a constant $L > 0$ such that, if*

$$g(z) = \sup\{ | \nabla f(z + w(z)) | : | w(z) | < L | z | \},$$

then $E[(| Z |^3 + 1) g(Z)] < \infty$.

(C2) *$H(z)$ is uniformly continuous on R^n and there exists a constant $K > 0$ such that*:

$$E[\sup\{ \| H(Z + w(Z)) \| : | w(Z) | < K | Z | \}] < \infty.$$

If γ is given by (4.8), *then the corresponding procedure δ dominates $\bar{\delta}$ as $b \to \infty$.*

PROOF. Under the above conditions, M exists almost surely, $E(| \nabla f(Z) |) < \infty$ and the derivation of (4.6a–c) is justified. Furthermore, if M exists then it is positive definite and, without loss of generality, we can assume that $M = I$ (in which case f may no longer have total mass 1). Henceforth we assume that $M = I$ and $\gamma(\theta) = -(a\theta/D)$, where $D = D(\theta) = b + |\theta|^2$.

The LHS of (4.4) can be written as follows:

$$\begin{aligned} &E[f(Z) - f(Z + \gamma(\theta + Z))] \\ (4.9)\qquad &= E[f(Z) - f(Z + \gamma(\theta))] + E[f(Z + \gamma(\theta)) - f(Z + \gamma(\theta) + \gamma'(\theta)Z)] \\ &\quad + E[f(Z + \gamma(\theta) + \gamma'(\theta)Z) - f(Z + \gamma(\theta + Z))]. \end{aligned}$$

The first term on the RHS of (4.9) can be written as:

$$E[f(Z) - f(Z + \gamma(\theta))] = -\tfrac{1}{2}\gamma^T(\theta)E[H(Z + \tilde{\gamma}(\theta, Z))]\gamma(\theta),$$

where $\tilde{\gamma}(\theta, Z)$ is a point on the line segment from Z to $Z + \gamma(\theta)$. By (C2) and the dominated convergence theorem we get that $E[H(Z + \tilde{\gamma}(\theta, Z))] \to E[H(Z)] = -2I$ uniformly in θ, as $b \to \infty$. Hence,

$$(4.10)\qquad D\,|E[f(Z) - f(Z + \gamma(\theta))]| < a^2 + o(1).$$

The last term on the RHS of (4.9) is treated similarly. We write it as $E[R_2^T \nabla f(Z + \tilde{\gamma}(\theta, Z))]$, where R_2 is the remainder for the expansion of γ at θ and $\tilde{\gamma}(\theta, Z)$ lies on the line segment between $Z + \gamma(\theta + Z)$ and $Z + \gamma(\theta) + \gamma'(\theta)Z$. A calculation using the well known identity $1/(1 + x) = 1 - x + x^2/(1 + x)$ shows that:

$$R_2 = \frac{a|Z|^2(\theta + Z) + 2aZ^T\theta Z}{D^2} - \frac{a(|Z|^2 + 2\theta^T Z)^2(\theta + Z)}{D^2(b + |Z + \theta|^2)}.$$

It follows that:

$$(4.11)\qquad |R_2| < (1/D^{3/2})(c_1|Z|^3 + c_2|Z|^2)$$

for c_1, c_2 positive constants, independent of θ. If $|Z| > \varepsilon > 0$ and b is large enough we can show that $|\nabla f(Z + \tilde{\gamma}(\theta, Z))| < g(Z)$. If $|Z| < \varepsilon$ then the continuity of ∇f implies $|\nabla f(Z + \tilde{\gamma}(\theta, Z))| <$ constant. It follows now from (4.11), (C1) and the dominated convergence theorem that

$$(4.12)\qquad D\,|E(R_2^T \nabla f(Z + \tilde{\gamma}(\theta, Z)))| \to 0$$

uniformly in θ, as $b \to \infty$.

The remaining term on the RHS of (4.9) is treated as follows. Using the mean value theorem we get:

$$\begin{aligned} &DE[f(Z + \gamma(\theta)) - f(Z + \gamma(\theta) + \gamma'(\theta)Z)] \\ &\qquad = DE[-(\gamma'(\theta)Z) \cdot \nabla f(Z + \gamma(\theta) + \tilde{r}\gamma'(\theta)Z)] \\ (4.13)\qquad &\qquad = aE[Z^T \nabla f(Z + \gamma(\theta) + \tilde{r}\gamma'(\theta)Z)] \\ &\qquad\quad - (2a/D)E[Z^T\theta\theta^T \nabla f(Z + \gamma(\theta) + \tilde{r}\gamma'(\theta)Z)] \end{aligned}$$

for some $\tilde{r} \in (0, 1]$. Note that as $b \to \infty$, $|\gamma(\theta)| \to 0$ and $\|\gamma'(\theta)\| \to 0$ uniformly in θ. It follows now from (4.6), (4.13), (C1) and the dominated convergence theorem that

$$(4.14) \quad DE[f(Z+\gamma(\theta)) - f(Z+\gamma(\theta)+\gamma'(\theta)Z)] \le -(n-2)ac + o(1)$$

uniformly in θ, as $b \to \infty$.

The proof of the theorem follows now from (4.10), (4.12) and (4.14), if we choose $a < c(n-2)$. □

REMARK. The set of conditions below is easier to verify than (C1) and (C2), and implies them:

(C1)′ $E[|Z|^3(1+\nabla f(Z))] < \infty$,
(C2)′ $E[\|H(Z)\|(|Z|^4+1)] < \infty$ and
(C3)′ $\sup_{z\in R^n}\max_{i,j,k}|D_iD_jD_kf(z)| < \infty$.

Acknowledgments. I would like to express my sincere thanks to Professor L. D. Brown for valuable help and guidance during the preparation of the work reported here. My thanks also go to the associate editor and the two referees from whose suggestions this paper has benefitted considerably.

REFERENCES

BERGER, J. O. (1980a). *Statistical Decision Theory.* Springer, New York.

BERGER, J. O. (1980b). Improving on inadmissible estimators in continuous exponential families, with applications to simultaneous estimation of gamma scale parameters. *Ann. Statist.* **8** 545–571.

BERNARDO, J. M. (1979). Reference posterior distributions for Bayesian Inference. *J. Roy. Statist. Soc. B* **41** 113–147.

BOX, G. E. P. and TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, Mass.

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1135.

BROWN, L. D. and FOX, M. (1974a). Admissibility in statistical problems involving a location or scale parameter. *Ann. Statist.* **2** 807–814.

BROWN, L. D. and FOX, M. (1974b). Admissibility of procedures in two-dimensional location parameter problems. *Ann. Statist.* **2** 248–266.

BROWN, L. D. (1979). A heuristic method for determining admissibility of estimators with applications. *Ann. Statist.* **7** 960–994.

DAWID, A. P., STONE, M. and ZIDEK, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. B* **35** 189–233.

EATON, M. (1981). A method for evaluating improper prior distributions. S. S. Gupta and J. Berger (Eds.). *Statistical Decision Theory and Related Topics III.* Academic, New York.

EDWARDS, W., LINDMAN, H. and SAVAGE, L. J. (1963). Bayesian statistical inference for psychological research. *Psychol. Rev.* **70** 193–242.

GATSONIS, C. (1981). Estimation of the posterior density of a location parameter. Ph.D. Thesis, Cornell University, January 1981.

HWANG, J. T. and CASELLA, G. (1982). Minimax confidence sets for the mean of a multivariate normal distribution. *Ann. Statist.* **10** 868–881.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 361–379.

JOSHI, V. M. (1967). Inadmissibility of the usual confidence sets for the mean of a multivariate normal population. *Ann. Math. Statist.* **38** 1868–1875.

LUSTERNIK, L. and SOBOLEV, V. (1974). *Elements of Functional Analysis.* Hindustan, Delhi, India.
STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206.
STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

DEPARTMENT OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003