Stein Estimation for Spherically Symmetric Distributions: Recent Developments

Ann Cohen Brandwein and William E. Strawderman

Abstract. This paper reviews advances in Stein-type shrinkage estimation for spherically symmetric distributions. Some emphasis is placed on developing intuition as to why shrinkage should work in location problems whether the underlying population is normal or not. Considerable attention is devoted to generalizing the "Stein lemma" which underlies much of the theoretical development of improved minimax estimation for spherically symmetric distributions. A main focus is on distributional robustness results in cases where a residual vector is available to estimate an unknown scale parameter, and, in particular, in finding estimators which are simultaneously generalized Bayes and minimax over large classes of spherically symmetric distributions. Some attention is also given to the problem of estimating a location vector restricted to lie in a polyhedral cone.

Key words and phrases: Stein estimation, spherical symmetry, minimaxity, admissibility.

1. INTRODUCTION

We are happy to help celebrate Stein's stunning, deep and significant contribution to the statistical literature. In 1956, Charles Stein (1956) proved a result that astonished many and was the catalyst for an enormous and rich literature of substantial importance in statistical theory and practice. Stein showed that when estimating, under squared error loss, the unknown mean vector θ of a p-dimensional random vector X having a normal distribution with identity covariance matrix, estimators of the form $(1 - a/\{\|X\|^2 + b\})X$ dominate the usual estimator θ, X, for a sufficiently small and b sufficiently large when $p \geq 3$. James and Stein (1961) sharpened the result and gave an explicit class of dominating estimators, $(1 - a/\|X\|^2)X$ for $0 < a < 2(p-2)$, and also showed that the choice of $a = p - 2$

Ann Cohen Brandwein is Professor, Department of Statistics and Computer Information Systems, CUNY Baruch College, One Bernard Baruch Way, New York, New York 10010, USA (e-mail: ann.brandwein@baruch.cuny.edu). William E. Strawderman is Professor, Department of Statistics and Biostatistics, Rutgers University, 110 Frelinghuysen Rd., Piscataway, New Jersey 08854, USA (e-mail: straw@stat.rutgers.edu).

(the James–Stein estimator) is uniformly best. For future reference recall that "the usual estimator," X, is a minimax estimator for the normal model, and more generally for any distribution with finite covariance matrix.

Stein (1974, 1981), considering general estimators of the form $\delta(X) = X + g(X)$, gave an expression for the risk of these estimators based on a key Lemma, which has come to be known as Stein's lemma. Numerous results on shrinkage estimation in the general spherically symmetric case followed based on some generalization of Stein's lemma to handle the cross product term $E_\theta[(X - \theta)'g(X)]$ in the expression for the risk of the estimator.

A substantial number of papers for the multivariate normal and nonnormal distributions have been written over the decades following Stein's monumental results. For an earlier expository development of Stein estimation for nonnormal location models see Brandwein and Strawderman (1990).

This paper covers the development of Stein estimation for spherically symmetric distributions since Brandwein and Strawderman (1990). It is not encyclopedic, but touches on only some of the significant results for the nonnormal case.

Given an observation, X, on a p-dimensional spherically symmetric multivariate distribution with un-

known mean, θ and whose density is $f(\|x - \theta\|^2)$ (for $x, \theta \in R^p$), we will consider the problem of estimating θ subject to the squared error loss function, that is, $\delta(X)$ is a measurable (vector-valued) function, and the loss given by

$$(1.1) \qquad L(\theta, \delta) = \|\delta - \theta\|^2 = \sum_{i=1}^{p} (\delta_i - \theta_i)^2,$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_p)'$ and $\theta = (\theta_1, \theta_2, \ldots, \theta_p)'$. The risk function of δ is defined as

$$R(\theta, \delta) = E_\theta L(\delta(X), \theta).$$

Unless otherwise specified, we will be using the loss defined by (1.1). Other loss functions such as the loss $L(\theta, \delta) = \|\delta - \theta\|^2/\sigma^2$ will be occasionally used, especially when there is also an unknown scale parameter, and minimaxity, as opposed to domination, is the main object of study. We will have relatively little to say about the important case of confidence set loss, or of loss estimation.

In Section 2 we provide some additional intuition as to why the Stein estimator of the mean vector θ makes sense as an approximation to an optimal linear estimator and as an empirical Bayes estimator in a general location problem. The discussion indicates that normality need play no role in the intuitive development of Stein-type shrinkage estimators.

Section 3 is devoted to finding improved estimators of θ for spherically symmetric distributions with a known scale parameter using results of Brandwein and Strawderman (1991) and Berger (1975) to bound the risk of the improved general estimator $\delta(X) = X + \sigma^2 g(X)$.

Section 4 considers estimating the mean vector for a general spherically symmetric distribution in the presence of an unknown scale parameter, and, more particularly, when a residual vector is available to estimate the scale parameter. It extends some of the results from Section 3 to this case as well as presenting new improved estimators for this problem. The results in this section indicate a remarkable robustness property of Stein-type estimators in this setting, namely, that certain of the improved estimators dominate X uniformly for all spherically symmetric distributions simultaneously (subject to risk finiteness).

In Section 5 we consider the restricted parameter space problem, particularly the case where θ is restricted to a polyhedral cane, or more generally a smooth cone. The material in this section is adapted from Fourdrinier, Strawderman and Wells (2003).

In Section 6 we consider some of the advancements in Bayes estimation of location vectors for both the known and unknown scale cases. We present an intriguing result of Maruyama Maruyama (2003b) which is related to the (distributional) robustness of Stein estimators in the unknown scale case treated in Section 4.

Section 7 contains some concluding remarks.

2. SOME FURTHER INTUITION INTO STEIN ESTIMATION

We begin by adding some intuition as to why Stein estimation is both reasonable and compelling, and refer the reader to Brandwein and Strawderman (1990) for some earlier developments. The reader is also referred to Stigler (1990) and to Meng (2005).

2.1 Stein Estimators as an Approximation to the Best Linear Estimator

The following is a very simple intuitive development for optimal linear estimation of the mean vector in R^p that leads to the Stein estimator.

Suppose $E_\theta[X] = \theta$, $\text{Cov}(X) = \sigma^2 I$ (σ^2 known), and consider the linear estimator of the form $\delta_a(X) = (1 - a)X$. What is the optimal value of a? The risk is given by

$$R(\theta, \delta_a) = p(1 - a)^2 \sigma^2 + a^2 \|\theta\|^2$$

and the derivative, with respect to a, is

$$\{d/da\} R(\theta, \delta_a) = 2\{-p(1 - a)\sigma^2 + a\|\theta\|^2\}.$$

Hence, the optimal a is $p\sigma^2/(p\sigma^2 + \|\theta\|^2)$ and the optimal "estimator" is $\delta(X) = (1 - p\sigma^2/\{p\sigma^2 + \|\theta\|^2\})X$, which is, of course, not an estimator because it depends on θ.

However, $E_\theta[\|X\|^2] = p\sigma^2 + \|\theta\|^2$, so $1/\|X\|^2$ is a reasonable estimator of $1/\{p\sigma^2 + \|\theta\|^2\}$. Hence, an approximation to the optimal linear "estimator" is $\delta(X) = (1 - p\sigma^2/\|X\|^2)X$ which is the James–Stein estimator except that p replaces $p - 2$. Note that as p gets larger, $\|X\|^2/p$ is likely to improve as an estimator of $\sigma^2 + \frac{\|\theta\|^2}{p}$ and, hence, we may expect that the dimension, p, plays a role.

2.2 Stein Estimators as Empirical Bayes Estimators for General Location Models

Strawderman (1992) considered the following general location model. Suppose $X|\theta \sim f(x - \theta)$, where $E_\theta[X] = \theta$, $\text{Cov}(X) = \sigma^2 I$ (σ^2 known) but that $f(\cdot)$ is otherwise unspecified. Also assume that the prior distribution for θ is given by $f^{\star n}(\theta)$, the n fold convolution of $f(\cdot)$ with itself. Hence, the prior distribution of

θ can be represented as the distribution of a sum of n i.i.d. variables $u_i, i = 1, \ldots, n$, where each u is distributed as $f(u)$. Also, the distribution of $u_0 = (X - \theta)$ has the same distribution and is independent of the other u's.

The Bayes estimator can therefore be thought of as

$$\delta(X) = E[\theta|X] = E[\theta|X - \theta + \theta]$$

$$= E\left[\sum_{i=1}^{n} u_i \,\Big|\, \sum_{i=0}^{n} u_i\right]$$

and, hence,

$$\delta(X) = n E\left[u_j \,\Big|\, \sum_{i=0}^{n} u_i\right]$$

$$= \frac{n}{n+1} E\left[\sum_{i=0}^{n} u_i \,\Big|\, \sum_{i=0}^{n} u_i\right]$$

$$= \frac{n}{n+1} E[X|X] = \frac{n}{n+1} X$$

or, equivalently, $\delta(X) = E[\theta|X] = (1 - 1/\{n+1\})X$.

Assuming that n is unknown, we may estimate it from the marginal distribution of X, which has the same distribution as $X - \theta + \theta = \sum_{i=0}^{n} u_i$. In particular,

$$E_\theta[\|X\|^2] = E\left[\left\|\sum_{i=0}^{n} u_i\right\|^2\right]$$

$$= \sum_{i=0}^{n} E[\|u_i\|^2] = (n+1)p\sigma^2,$$

since $E[u_i] = 0$ and $\text{Cov}(u_i) = \sigma^2 I$, $E[\|u_i\|^2] = p\sigma^2$. Therefore, $(n+1)$ can be estimated by $(p\sigma^2)^{-1}\|X\|^2$. Substituting this estimator of $(n+1)$ in the expression for the Bayes estimator, we have an empirical Bayes estimator

$$\delta(X) = (1 - p\sigma^2/\|X\|^2)X,$$

which is again the James–Stein estimator, save for the substitution of p for $p-2$.

Note that in both of the above developments, the only assumptions were that $E_\theta(X) = \theta$, and $\text{Cov}(X) = \sigma^2 I$. The Stein-type estimator thus appears intuitively, at least, to be a reasonable estimator in a general location problem.

3. SOME RECENT DEVELOPMENTS FOR THE CASE OF A KNOWN SCALE PARAMETER

Let $X \sim f(\|x - \theta\|^2)$, the loss be $L(\theta, \delta) = \|\delta - \theta\|^2$ so the risk is $R(\theta, \delta) = E_\theta[\|\delta(X) - \theta\|^2]$. Suppose an

estimator has the general form $\delta(X) = X + \sigma^2 g(X)$. Then

$$R(\theta, \delta) = E_\theta[\|\delta(X) - \theta\|^2]$$

$$= E_\theta[\|X + \sigma^2 g(X) - \theta\|^2]$$

$$= E_\theta[\|X - \theta\|^2] + \sigma^4 E_\theta[\|g(X)\|^2]$$

$$+ 2\sigma^2 E_\theta[(X - \theta)' g(X)].$$

In the normal case, Stein's lemma, given loosely as follows, is used to evaluate the last term.

LEMMA 3.1 [Stein (1981)]. *If $X \sim N(\theta, \sigma^2 I)$, then $E_\theta[(X - \theta)' g(X)] = \sigma^2 E_\theta[\nabla' g(X)]$ [where $\nabla' g(\cdot)$ denotes the gradient of $g(\cdot)$], provided, say, that g is continuously differentiable and that all expected values exist.*

PROOF. The proof is particularly easy in one dimension, and is a simple integration by parts. In higher dimensions the proof may just add the one-dimensional components or may be a bit more sophisticated and cover more general functions, g. In the most general version known to us, the proof uses Stokes' theorem and requires $g(\cdot)$ to be weakly differentiable. □

Using the Stein lemma, we immediately have the following result.

PROPOSITION 3.1. *If $X \sim N(\theta, \sigma^2 I)$, then*

$$R(\theta, X + \sigma^2 g(X))$$

$$= E_\theta[\|X - \theta\|^2] + \sigma^4 E_\theta[\|g(X)\|^2 + 2\nabla' g(X)]$$

and, hence, provided the expectations are finite, a sufficient condition for $\delta(X)$ to dominate X is $\|g(x)\|^2 + 2\nabla' g(x) < 0$ a.e. (with strict inequality on a set of positive measures).

The key to most of the literature on shrinkage estimation in the general spherically symmetric case is to find some generalization of (or substitution for) Stein's lemma to evaluate (or bound) the cross product term $E_\theta[(X - \theta)' g(X)]$. We indicate two useful techniques below.

3.1 Generalizations of James–Stein Estimators Under Spherical Symmetry

Brandwein and Strawderman (1991) extended the results of Stein (1974, 1981) to spherically symmetric distributions for estimators of the form $X + ag(X)$. The following two preliminary lemmas are necessary to prove the result in Theorem 3.1.

LEMMA 3.2. *Let X have a distribution that is spherically symmetric about θ. Then*

$$E_\theta[(X-\theta)'g(X)|\|X-\theta\|^2 = R^2]$$
$$= p^{-1}R^2 \text{Ave}_{B(R,\theta)} \nabla' g(X),$$

provided $g(x)$ is weakly differentiable.

PROOF. Notation for this lemma: $S(R,\theta)$ and $B(R,\theta)$ are, respectively, the (surface of the) sphere and (solid) ball, of radius R centered at θ. Note also that $(X-\theta)/R$ is the unit outward normal vector at X on $S(R,\theta)$. Also $d\sigma(X)$ is the area measure on $S(R,\theta)$, while $A(\cdot)$ and $V(\cdot)$ denote area and volume, respectively. Since the conditional distribution of $X-\theta$ given $\|X-\theta\|^2 = R^2$ is uniform on the sphere of radius R, it follows that

$$E_\theta[(X-\theta)'g(X)|\|X-\theta\|^2 = R^2]$$
$$= \text{Ave}_{S(R,\theta)}\{(X-\theta)'g(X)\}$$
$$= \frac{R}{A(S(R,\theta))} \oint_{S(R,\theta)} \frac{(X-\theta)'g(X)}{R} d\sigma(X)$$
$$= \frac{R}{A(S(R,\theta))} \int_{B(R,\theta)} \nabla' g(x)\, dx$$
$$\qquad \left(\text{since } \frac{V(B(R,\theta))}{A(S(R,\theta))} = R/p\right)$$
$$= \frac{R^2}{pV(B(R,\theta))} \int_{B(R,\theta)} \nabla' g(x)\, dx$$
$$\qquad\qquad\qquad \text{(by Stokes' theorem)}$$
$$= p^{-1}R^2 \text{Ave}_{B(R,\theta)} \nabla' g(X). \qquad \square$$

The following result is basic to the study of superharmonic functions and is well known (see, e.g., du Plessis, 1970, page 54).

LEMMA 3.3. *Let $h(x)$ be superharmonic on $S(R)$, [i.e., $\sum_{i=1}^{p}\{\partial^2/\partial x_i^2\}h(x) \leq 0$], then $\text{Ave}_{S(R,\theta)}h(x) \leq \text{Ave}_{B(R,\theta)}h(x)$.*

Consider, now, an estimator of the general form $X + ag(X)$, where a is a scalar, and $g(X)$ maps $R^p \to R^p$.

THEOREM 3.1. *Let X have a distribution that is spherically symmetric about θ. Assume the following:*

1. $\|g(x)\|^2/2 \leq -h(x) \leq -\nabla' g(x)$,
2. $-h(x)$ *is superharmonic, $E_\theta[R^2h(W)]$ is nonincreasing in R for each θ, where W has a uniform distribution on $B(R,\theta)$,*
3. $0 \leq a \leq 1/\{pE_0[1/\|X\|^2]\}$.

Then $X + ag(X)$ is minimax with respect to quadratic loss, provided $g(\cdot)$ is weakly differentiable and all expectations are finite.

PROOF.

$$R(\theta, X + ag(X)) - R(\theta, X)$$
$$= E[E_\theta[a^2\|g(X)\|^2$$
$$\qquad + 2a(X-\theta)'g(X)|\|X-\theta\|^2 = R^2]]$$
$$\leq E[E_\theta[-2a^2h(X)$$
$$\qquad + 2a(X-\theta)'g(X)|\|X-\theta\|^2 = R^2]]$$
$$= E[E_\theta[-2a^2h(X)|\|X-\theta\|^2 = R^2]$$
$$\qquad + 2aE[\{R^2/p\}\text{Ave}_{B(R,\theta)}\nabla'g(X)|R^2]]$$
$$\leq E[E_\theta[-2a^2h(X)|\|X-\theta\|^2 = R^2]$$
$$\qquad + 2aE_\theta[\{R^2/p\}E_\theta h(W)|R^2]]$$
$$\leq E[E_\theta[-2a^2h(W)|R^2]$$
$$\qquad + 2aE_\theta[\{R^2/p\}E_\theta h(W)|R^2]]$$
$$\qquad\qquad\qquad \text{(by Lemma 3.3)}$$
$$= 2aE[E_\theta[R^2h(W)|R^2](-a/R^2 + 1/p)]$$
$$= 2aE[E_\theta[R^2h(W)|R^2]]E[-a/R^2 + 1/p]$$
$$\leq 0$$

by the covariance inequality since $E_\theta[R^2h(W)|R^2]$ is nonincreasing and $-R^{-2}$ is increasing and since $h \leq 0$. $\square$

EXAMPLE 3.1. James–Stein estimators $[g(x) = -2(p-2)x/\|x\|^2]$: In this case both $\|g(x)\|^2/2$ and $-\nabla'g(x)$ are equal to $2(p-2)^2/\|x\|^2$. Conditions 1 and 2 of Theorem 3.1 are satisfied for $h(x) = -2(p-2)^2/\|x\|^2$, provided $p \geq 4$ since $\|x\|^{-2}$ is superharmonic if $p \geq 4$, and since $E_\theta[R^2/\|X\|^2] = E_{\theta/R}[1/\|X\|^2]$ is increasing by Anderson's theorem.

Hence, by condition 3, for any spherically symmetric distribution, the James–Stein estimator $(1 - a2(p-2)/\|X\|^2)X$ is minimax for $0 \leq a \leq 1/\{pE_0[1/\|X\|^2]\}$ and $p \geq 4$. The domination over X is strict for $0 < a < 1/\{pE_0[1/\|X\|^2]\}$, and also for $a = 1/\{pE_0[1/\|X\|^2]\}$, provided the distribution is not normal.

Baranchik (1970), for the normal case, considered estimators of the form $(1 - ar(\|X\|^2)/\|X\|^2)X$ under certain conditions on $r(\cdot)$. Under the assumption that $r(\cdot)$ is monotone nondecreasing, bounded between 0 and 1, and concave, Theorem 3.1 applies to these estimators as well, and establishes minimaxity for $0 \leq a \leq 1/\{pE_0[1/\|X\|^2]\}$ and for $p \geq 4$.

We note in passing that the results in this subsection hold for an arbitrary spherically symmetric distribution with or without a density. The calculations rely only on the distribution of X conditional on $\|X - \theta\|^2 = R^2$, and, of course, finiteness of $E[\|X\|^2]$ and $E[\|g(X)\|^2]$.

3.2 A Useful Expression for the Risk of a James–Stein Estimator

Berger (1975) gave a useful expression for the risk of a James–Stein estimator which is easily generalized to the case of a general estimator, provided the spherically symmetric distribution has a density $f(\|x - \theta\|^2)$.

Some form of this generalization (and extensions to unknown scale case and the elliptically symmetric case) has been used by several authors, including Fourdrinier, Strawderman and Wells (2003), Fourdrinier, Kortbi and Strawderman (2008), Fourdrinier and Strawderman (2008), Maruyama (2003a) and Kubokawa and Srivastava (2001), among others.

LEMMA 3.4. *Suppose* $X \sim f(\|x - \theta\|^2)$, *and let* $F(t) = 2^{-1} \int_t^\infty f(u)\, du$ *and* $Q(t) = F(t)/f(t)$. *Then*

$$R(\theta, X + g(X))$$
$$= E_\theta[\|X - \theta\|^2]$$
$$\quad + E_\theta[\|g(X)\|^2 + 2Q(\|X - \theta\|^2)\nabla' g(X)].$$

PROOF. The lemma follows immediately with the following identity for the cross product term:

$$E[(x - \theta)' g(X)]$$
$$= \int_{R^p} (x - \theta)' g(X) f(\|x - \theta\|^2)\, dx$$
$$= \int_{R^p} g(X)' \nabla F(\|x - \theta\|^2)\, dx$$
$$= \int_{R^p} \nabla' g(X) F(\|x - \theta\|^2)\, dx$$

(by Green's theorem)

$$= E[Q(\|X - \theta\|^2) \nabla' g(X)]. \qquad \square$$

Berger (1975), Maruyama (2003a) and Fourdrinier, Kortbi and Strawderman (2008) used the above result for distributions for which $Q(t)$ is bounded below by a positive constant. In this case, the next result follows immediately from Lemma 3.4.

THEOREM 3.2. *Suppose* $X \sim f(\|x - \theta\|^2)$, *and that* $Q(t) \geq c > 0$. *Then the estimator* $X + g(X)$ *dominates* X *provided* $\|g(x)\|^2 + 2c\nabla' g(x) \leq 0$ *for all* x.

EXAMPLE 3.2. As noted by Berger (1975), if $f(\cdot)$ is a scale mixture of normals, then $Q(t)$ is bounded below. To see this, note that if $X|V \sim N(\theta, VI)$ and $V \sim g(v)$, then $f(t) = \int_0^\infty (2\pi v)^{-p/2} \exp(-t/2v) g(v)\, dv$. Similarly,

$$F(t) = 2^{-1} \int_t^\infty f(u)\, du$$
$$= 2^{-1} \int_0^\infty g(v)(2\pi v)^{-p/2} \int_t^\infty \exp(-u/2v)\, du$$
$$= \int_0^\infty (2\pi v)^{-p/2} v \exp(-t/2v) g(v)\, dv.$$

Hence,

$$Q(t) = \frac{\int_0^\infty v^{(2-p)/2} \exp(-t/2v) g(v)\, dv}{\int_0^\infty v^{-p/2} \exp(-t/2v) g(v)\, dv}$$
$$= E_t[V] \geq E_0[V] = \frac{\int_0^\infty v^{1-p/2} g(v)\, dv}{\int_0^\infty v^{-p/2} g(v)\, dv}$$
$$= \frac{E[V^{1-p/2}]}{E[V^{-p/2}]} = c > 0,$$

where E_t denotes expectation with respect to the density proportional to $v^{-p/2} \exp(-t/2v) g(v)$. The inequality follows since the family has monotone likelihood ratio in t.

Hence, for the James–Stein class $(1 - a/\|X\|^2)X$, this result gives dominance over X for

$$a^2 - 2a(p - 2) \frac{E[V^{1-p/2}]}{E[V^{-p/2}]} \leq 0$$

or

$$0 \leq a \leq 2(p - 2) \frac{E[V^{1-p/2}]}{E[V^{-p/2}]}.$$

This bound on the shrinkage constant, a, compares poorly with that obtained by Strawderman (1974), $0 \leq a \leq 2(p - 2)/E[V^{-1}]$, which may be obtained by using Stein's lemma conditional on V and the fact that $E_\theta[V/\|X\|^2|V]$ is monotone nondecreasing in V. Note that, again by monotone likelihood ratio properties (or the covariance inequality), $(E[V^{-1}])^{-1} > E[V^{1-p/2}]/E[V^{-p/2}]$.

It is therefore somewhat surprising that Maruyama (2003a) and Fourdrinier, Kortbi and Strawderman (2008) were able to use Theorem 3.2, applied to Baranchik-type estimators, to obtain generalized and proper Bayes minimax estimators. Without going into details, the advantage of the cruder bound is that it requires only that $r(t)$ be monotone, while Strawderman's result for mixtures of normal distributions also requires that $r(t)/t$ be monotone decreasing.

Other applications of Lemma 3.4 give refined bounds on the shrinkage constant in the James–Stein or Baranchik estimator depending on monotonicity properties of $Q(t)$. Typically, additional conditions are required on the function $r(t)$ as well. See, for example, Brandwein, Ralescu and Strawderman (1993) (although the calculations in that paper are somewhat different than those in this section, the basic idea is quite similar).

Applications of the risk expression in Lemma 3.4 are complicated relative to those in the normal case using Stein's lemma, in that the mean vector, θ, remains to complicate matters through the function $Q(\|X - \theta\|^2)$. It is both surprising and interesting that matters become essentially simpler (in a certain sense) when the scale parameter is unknown, but a residual vector is available. We investigate this phenomenon in the next section.

4. STEIN ESTIMATION IN THE UNKNOWN SCALE CASE

In this section we study the model $(X, U) \sim f(\|x - \theta\|^2 + \|u\|^2)$, where $\dim X = \dim \theta = p$, and $\dim U = k$. The classical example of this model is, of course, the normal model $f(t) = (\frac{1}{\sqrt{2\pi}\sigma})^{p+k} e^{-t/(2\sigma^2)}$. However, a variety of other models have proven useful. Perhaps the most important alternatives to the normal model in practice and in theory are the generalized multivariate-t distributions

$$f(t) = \frac{c}{\sigma^{p+k}} \left(\frac{1}{a + t/\sigma^2} \right)^b,$$

or, more generally, scale mixture of normals of the form

$$f(t) = \int_0^\infty \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^{p+k} e^{-t/(2\sigma^2)} \, dG(\sigma^2).$$

These models preserve the spherical symmetry about the mean vector and, hence, the covariance matrix is a multiple of the identity. Thus, the coordinates are uncorrelated, but they are not independent except for the case of the normal model. We look (primarily) at estimators of the form $X + \{\|U\|^2/(k+2)\}g(X)$.

The main result may be interpreted as follows: If, when $X \sim N(\theta, \sigma^2 I)$ (σ^2 known), the estimator $X + \sigma^2 g(X)$ dominates X, then, under the model $(X, U) \sim f(\|x - \theta\|^2 + \|u\|^2)$, the estimator $X + \{\|U\|^2/(k+2)\}g(X)$ dominates X. That is, substituting the estimator $\|U\|^2/(k+2)$ for σ^2 preserves domination uniformly for all parameters (θ, σ^2) and (somewhat astonishingly) simultaneously for all distributions, $f(\cdot)$.

Note that, interestingly, $\|U\|^2/(k+2)$ is the minimum risk equivariant estimator of σ^2 in the normal case under the usual invariant loss. This wonderful result is due to Cellier and Fourdrinier (1995). We refer the reader to their paper for the original proof based on Stokes' theorem applied to the distribution of X conditional on $\|X - \theta\|^2 + \|U\|^2 = R^2$. One interesting aspect of that proof is that even if the original distribution has no density, the conditional distribution of X does have a density for all $k > 0$.

We will approach the above result from two different directions. The first approach is essentially an extension of Lemma 3.4. As in that case, the resulting expression for the risk still involves both the data and θ inside the expectation, but the function $Q(\|X - \theta\|^2 + \|U\|^2)$ is a common factor. This allows the treatment of the remaining terms as if they are an unbiased estimate of the risk difference.

The second approach is due to Fourdrinier, Strawderman and Wells (2003), and is attractive because it is essentially statistical in nature, depending on completeness and sufficiency. It may be argued also that this approach is somewhat more general in that it may be useful even when the function $g(x)$ is not necessarily weakly differentiable. In this case an unbiased estimator of the risk difference is obtained which agrees with that in Cellier and Fourdrinier (1995). This is in contrast to the above method whereby the expression for the risk difference still has a factor $Q(\|X - \theta\|^2 + \|U\|^2)$ inside the expectation.

NOTE. Technically, our use of the term "unknown scale" is somewhat misleading in that the scale parameter may, in fact, be known. We typically think of $f(\cdot)$ as being a known density, which implies that the scale is known as well. It may have been preferable to write the density as $(X, U) \sim \{1/\sigma^{p+k}\} f(\{\|x - \theta\|^2 + \|u\|^2\}/\sigma^2)$, emphasizing the unknown scale parameter. This is more in keeping with the usual canonical form of the general linear model with spherically symmetric errors. What is of fundamental importance is the presence of the residual vector, U, in allowing uniform domination over the estimator X simultaneously for the entire class of spherically symmetric distributions. Since the suppression of the scale parameter makes notation a bit simpler, we will, for the most part, use the above notation in this section. Additionally, we continue to use the un-normalized loss, $L(\theta, \delta) = \|\delta - \theta\|^2$, and state results in terms of dominance over X instead of minimaxity, since the minimax risk is infinite. In order to speak meaningfully of minimaxity in the unknown scale case, we should use a normalized version of the loss, such as $L(\theta, \delta) = \|\delta - \theta\|^2/\sigma^2$.

4.1 A Generalization of Lemma 3.4

LEMMA 4.1. *Suppose* $(X, U) \sim f(\|x - \theta\|^2 + \|u\|^2)$, *where* $\dim X = \dim \theta = p$, $\dim U = k$. *Then, provided* $g(x, \|u\|^2)$ *is weakly differentiable in each coordinate*:

1. $E_\theta[\|U\|^2(X - \theta)'g(X, \|U\|^2)] = E_\theta[\|U\|^2 \nabla'_X g(X, \|U\|^2)Q(\|X - \theta\|^2 + \|U\|^2)]$.
2. $E_\theta[\|U\|^4\|g(X, \|U\|^2)\|^2] = E_\theta[h(X, \|U\|^2) \cdot Q(\|X - \theta\|^2 + \|U\|^2)]$, *where* $Q(t) = \{2f(t)\}^{-1} \cdot \int_t^\infty f(s)\,ds$ *and*

$$(4.1) \quad \begin{aligned} &h(x, \|u\|^2) \\ &= (k + 2)\|u\|^2\|g(x)\|^2 \\ &\quad + 2\|u\|^4 \frac{\partial}{\partial \|u\|^2}\|g(x, \|u\|^2)\|^2. \end{aligned}$$

PROOF. The proof of part 1 is essentially the same as the proof of Lemma 3.4, holding U fixed throughout. The same is true of part 2, where the roles of X and U are reversed and one notes that

$$\nabla'_u(\|u\|^2 u) = (k + 2)\|u\|^2,$$

$$\nabla'_u\{(\|u\|^2 u)\|g(x, \|u\|^2)\|^2\} = h(x, \|u\|^2),$$

which is given by (4.1), and, hence,

$$\begin{aligned} &E_\theta[\|U\|^4\|g(X, \|U\|^2)\|^2] \\ &= E_\theta[(\|U\|^2 U)'U\|g(X, \|U\|^2)\|^2] \\ &= E_\theta[\nabla'_U\{(\|U\|^2 U)\|g(X, \|U\|^2)\|^2\} \\ &\qquad\qquad \cdot Q(\|X - \theta\|^2 + \|U\|^2)] \\ &= E_\theta[h(X, \|U\|^2)Q(\|X - \theta\|^2 + \|U\|^2)]. \quad \square \end{aligned}$$

One version of the main result for estimators of the form $X + \{\|U\|^2/(k + 2)\}g(X)$ is the following theorem.

THEOREM 4.1. *Suppose* (X, U) *is as in Lemma 4.1. Then*:

1. *The risk of an estimator* $X + \{\|U\|^2/(k + 2)\}g(X)$ *is given by*

$$\begin{aligned} &R(\theta, X + \{\|U\|^2/(k + 2)\}g(X)) \\ &= E_\theta[\|X - \theta\|^2] \\ &\quad + E_\theta\left[\frac{\|U\|^2}{k + 2}\{\|g(X)\|^2 + 2\nabla'g(X)\} \right. \\ &\qquad\qquad\qquad \left. \cdot Q(\|X - \theta\|^2 + \|U\|^2)\right], \end{aligned}$$

2. $X + \{\|U\|^2/(k + 2)\}g(X)$ *dominates* X *provided* $\|g(x)\| + 2\nabla'g(x) < 0$.

PROOF. Note that

$$\begin{aligned} &R(\theta, X + \{\|U\|^2/(k + 2)\}g(X)) \\ &= E_\theta[\|X - \theta\|^2] \\ &\quad + E_\theta\left[\frac{\|U\|^4}{(k + 2)^2}\|g(X)\|^2 \right. \\ &\qquad\qquad \left. + 2\frac{\|U\|^2}{k + 2}(X - \theta)'g(X)\right] \\ &= E_\theta[\|X - \theta\|^2] \\ &\quad + E_\theta\left[\{\|g(X)\|^2 + 2\nabla'g(X)\} \right. \\ &\qquad\qquad \left. \cdot \frac{\|U\|^2 Q(\|X - \theta\|^2 + \|U\|^2)}{k + 2}\right] \end{aligned}$$

by successive application of parts 1 and 2 of Lemma 4.1. $\square$

EXAMPLE 4.1. Baranchik-type estimators: Suppose the estimator is given by $(1 - \|U\|^2 r(\|X\|^2)/\{(k + 2)\|X\|^2\})X$, where $r(t)$ is nondecreasing, and $0 \le r(t) \le 2(p - 2)$, then for $p \ge 3$ the estimator dominates X simultaneously for all spherically symmetric distributions for which the risk of X is finite. This follows since, if $g(x) = -xr(\|x\|^2)/\|x\|^2$, then

$$\begin{aligned} &\|g(x)\|^2 + 2\nabla'g(x) \\ &= r^2(\|x\|^2)/\|x\|^2 \\ &\quad - 2\{(p - 2)r(\|x\|^2)/\|x\|^2 - 2r'(\|x\|^2)\} \\ &\le r^2(\|x\|^2)/\|x\|^2 - 2(p - 2)r(\|x\|^2)/\|x\|^2 \le 0. \end{aligned}$$

EXAMPLE 4.2. James–Stein estimators: If $r(\|x\|^2) \equiv a$, the Baranchik estimator is a James–Stein estimator, and, since $r'(t) \equiv 0$, the risk is given by

$$\begin{aligned} &E_\theta[\|X - \theta\|^2] + \frac{a^2 - 2a(p - 2)}{k + 2} \\ &\qquad\qquad \cdot E\left[\frac{\|U\|^2}{\|X\|^2}Q(\|X - \theta\|^2 + \|U\|^2)\right]. \end{aligned}$$

Just as in the normal case, $a = p - 2$ is the uniformly best choice to minimize the risk. But here it is the uniformly best choice for every distribution. Hence, the estimator $(1 - (p - 2)\|U\|^2/\{(k + 2)\|X\|^2\})X$ is uniformly best, simultaneously for all spherically symmetric distributions among the class of James–Stein estimators!

A more refined version of Theorem 4.1 which uses the full power of Lemma 4.1 is proved in the same way. We give it for completeness and since it is useful in the study of risks of Bayes estimators.

THEOREM 4.2. *Suppose* (X, U) *is as in Lemma* 4.1. *Then, under suitable smoothness conditions on* $g(\cdot)$:

1. *The risk of an estimator* $X + \{\|U\|^2/(k+2)\}g(X, \|U\|^2)$ *is given by*

$$R(\theta, X + \{\|U\|^2/(k+2)\}g(X, \|U\|^2))$$

$$= E_\theta[\|X - \theta\|^2]$$
$$+ E_\theta[\{(k+2)^{-1}\|U\|^2\|g(X, \|U\|^2)\|^2$$
$$+ 2\nabla'_X g(X, \|U\|^2)$$
$$+ 2(k+2)^{-2}\|U\|^4(\partial/\partial\|U\|^2)$$
$$\cdot \|g(X, \|U\|^2)\|^2\}$$
$$\cdot Q(\|X - \theta\|^2 + \|U\|^2)],$$

2. $X + \{\|U\|^2/(k+2)\}g(X, \|U\|^2)$ *dominates* X *provided*

$$\|g(x, \|u\|^2)\|^2 + 2\nabla'_x g(x, \|u\|^2)$$
$$+ 2\frac{\|u\|^2}{k+2}\frac{\partial}{\partial\|u\|^2}\|g(x, \|u\|^2)\|^2 < 0.$$

COROLLARY 4.1. *Suppose* $\delta(X, \|U\|^2) = (1 - \|U\|^2 r(\|X\|^2/\|U\|^2)/\|X\|^2)X$. *Then* $\delta(X, \|U\|^2)$ *dominates* X *provided*:

1. $0 \le r(\cdot) \le 2(p-2)/(k+2)$ *and*
2. $r(\cdot)$ *is nondecreasing.*

The result follows from Theorem 4.2 by a straightforward calculation.

4.2 A More Statistical Approach Involving Sufficiency and Completeness

We largely follow Fourdrinier, Strawderman and Wells (2003) in this subsection. The nature of the conclusions for estimators is essentially as in Theorem 4.1, but the result is closer in spirit to the result of Cellier and Fourdrinier (1995) in that we obtain an unbiased estimator of risk difference (from X) instead of the expression in Theorem 4.1 where the function $Q(\cdot)$, which depends on θ, intervenes. The following lemma is the key to this development.

LEMMA 4.2. *Let* $(X, U) \sim f(\|x - \theta\|^2 + \|u\|^2)$, *where* $\dim X = \dim\theta = p$ *and* $\dim U = k$. *Suppose* $g(\cdot)$ *and* $h(\cdot)$ *are such that when* $X \sim N_p(\theta, I)$,

$E_\theta[(X - \theta)'g(X)] = E_\theta[h(X)]$. *Then, for* (X, U) *as above,*

$$E_\theta[\|U\|^2(X - \theta)'g(X)]$$
$$= \{1/(k+2)\}E_\theta[\|U\|^4 h(X)],$$

provided the expectations exist.

NOTE. Typically, of course, $h(x)$ is the divergence of $g(x)$, and, in all cases known to us, this remains essentially true. We choose this form of expressing the lemma because in certain instances of restricted parameter spaces the lemma applies even though the function $g(\cdot)$ may not be weakly differentiable, but the equality still holds for $g(x)I_A(g(x))$ and $h(x) = \nabla'g(x)I_A(g(x))$, where $I_A(\cdot)$ is the indicator function of a set A.

PROOF OF LEMMA 4.2. Suppose first, that the distribution of (X, U) is $N_{p+k}(\{\theta, 0\}, \sigma^2 I)$ and that θ is considered known. Then by the independence of X and U we have by assumption that

$$E_\theta[(X - \theta)'g(X)]$$
$$= E_\theta[(1/k)\|U\|^2(X - \theta)'g(X)]$$
$$= E_\theta[\{k(k+2)\}^{-1}\|U\|^4 h(X)].$$

Hence, the claimed result of the theorem is true for the normal case. Now use the fact that in the normal case (for θ known), $\|X - \theta\|^2 + \|U\|^2$ is a complete sufficient statistic. So it must be that

$$E_\theta[\|U\|^2(X - \theta)'g(X)|\|X - \theta\|^2 + \|U\|^2]$$
$$= E_\theta\left[\frac{\|U\|^4 h(X)}{k+2}\bigg|\|X - \theta\|^2 + \|U\|^2\right]$$

for all $\|X - \theta\|^2 + \|U\|^2$ except on a set of measure 0, since each function of $\|X - \theta\|^2 + \|U\|^2$ has the same expected value. Actually, it can be shown that these conditional expectations are continuous in R and, hence, they agree for all R (see Fourdrinier, Strawderman and Wells, 2003).

But the distribution of (X, U) conditional on $\|X - \theta\|^2 + \|U\|^2 = R^2$ is uniform on the sphere centered at $(\theta, 0)$ of radius R, which is the same as the conditional distribution of (X, U) conditional on $\|X - \theta\|^2 + \|U\|^2 = R^2$ for any spherically symmetric distribution. Hence, the equality which holds for the normal distribution holds for all distributions $f(\cdot)$. □

Lemma 4.2 immediately gives the following unbiased estimator of risk difference and a condition for dominating X for estimators of the form $\delta(X) = X + \{\|U\|^2/(k+2)\}g(X)$.

THEOREM 4.3. *Suppose $(X, U), g(x)$ and $h(x)$ are as in Lemma 4.2. Then, for the estimator $\delta(X) = X + \{\|U\|^2/(k+2)\}g(X)$:*

1. *The risk difference is given by*

$$R(\theta, \delta) - E_\theta[\|X - \theta\|^2]$$

$$= E_\theta\left[\frac{\|U\|^4}{(k+2)^2}\{\|g(X)\|^2 + 2\nabla'g(X)\}\right],$$

2. *$\delta(X)$ beats X provided $\|g(x)\|^2 + 2\nabla'g(x) \leq 0$, with strict inequality on a set of positive measure, and provided all expectations are finite.*

5. RESTRICTED PARAMETER SPACES

We consider a simple version of the general restricted parameter space problem which illustrates what types of results can be obtained. Suppose (X, U) is distributed as in Theorem 4.1 but it is known that $\theta_i \geq 0$, $i = 1, \ldots, p$, that is, $\theta \in R_+^p$ the first orthant. What follows can be generalized to the case where θ is restricted to a polyhedral cone, and more generally a smooth cone. The material in this section is adapted from Fourdrinier, Strawderman and Wells (2003).

In the normal case, the MLE of θ subject to the restriction that $\theta \in R_+^p$ is X_+, where the ith component is X_i if $X_i \geq 0$ and 0 otherwise. Here, as in the case of the more general restriction to a convex cone, the MLE is the projection of X onto the restricted cone. Chang (1982) considered domination of the MLE of θ when X has a $N_p(\theta, I)$ distribution and $\theta \in R_+^p$ via certain Stein-type shrinkage estimators. Sengupta and Sen (1991) extended Chang's results to Stein-type shrinkage estimators of the form $\delta(X) = (1 - r_s(\|X_+\|^2)/\|X_+\|^2)X_+$, where $r_s(\cdot)$ is nondecreasing, and $0 \leq r_s(\cdot) \leq 2(s-2)_+$, and where s is the (random) number of positive components of X. Hence, shrinkage occurs only when s, the number of positive components of X, is at least 3 and the amount of shrinkage is governed by the sum of squares of the positive components. A similar result holds if θ is restricted to a general polyhedral cone where X_+ is replaced by the projection of X onto the cone and s is defined to be the dimension of the face onto which X is projected.

We choose the simple polyhedral cone $\theta \in R_+^p$ because it will be reasonably clear that some version of the Stein Lemma 3.1 applies in the normal case. We first indicate a convenient, but complicated looking, alternate representation of an estimator of the above form in this case. Denote the $n = 2^p$ orthants of R^p, by $O_1, \ldots, O_n$, and let O_1 be R_+. Then we may rewrite (a slightly more general version of) the above estimator as

$$\delta(X) = \sum_{i=1}^n \left(1 - \frac{r_i(\|P_i(X)\|^2)}{\|P_i(X)\|^2}\right)P_i(X)I_{O_i}(X),$$

where $P_i(X)$ is the linear projection of X onto F_i, where F_i is the s-dimensional face of $R_+ = O_1$ onto which O_i is projected. Note that if $r_i(\cdot) \equiv 0$, $\forall i$, the estimator is just the MLE.

LEMMA 5.1. *Suppose $X \sim N_p(\theta, I)$, and let each $r_i(\cdot)$ be smooth and bounded. Then:*

1. *For each O_i, $\{r_i(\|P_i(x)\|^2)/\|P_i(x)\|^2\}P_i(x)I_{O_i}(x)$ is weakly differentiable in x.*
2. *Further,*

$$E_\theta\left[(P_i(X) - \theta)'\frac{r_i(\|P_i(X)\|^2)}{\|P_i(X)\|^2}P_i(X)I_{O_i}(X)\right]$$

$$= E_\theta\left[\left\{\frac{(s-2)r_i(\|P_i(X)\|^2)}{\|P_i(X)\|^2}\right.\right.$$

$$\left.\left. + 2r_i'(\|P_i(X)\|^2)\right\}I_{O_i}(X)\right],$$

provided expectations exist.
3. *$\delta(X) = \sum_{i=1}^n\{1 - r_i(\|P_i(X)\|^2)/\|P_i(X)\|^2\} \cdot P_i(X)I_{O_i}(X)$ as given above dominates the MLE X_+, provided r_i is nondecreasing and bounded between 0 and $2(s-2)_+$.*

PROOF. Weak differentiability in part 1 follows since the function is smooth away from the boundary of O_i and is continuous on the boundary except at the origin. Part 2 follows from Stein's Lemma 3.1 and the fact that (essentially) $P_i(X) \sim N_s(\theta, \sigma^2 I)$, since $n - s$ of the coordinates are 0. Part 3 follows by Stein's Lemma 3.1 as in Proposition 3.1 applied to each orthant. We omit the details. The reader is referred to Sengupta and Sen (1991) or Fourdrinier, Strawderman and Wells (2003) for details in the more general case of a polyhedral cone. □

Next, essentially applying Lemma 4.2 to each orthant and using Lemma 5.1 we have the following generalization to the case of a general spherically symmetric distribution.

THEOREM 5.1. *Let $(X, U) \sim f(\|x - \theta\|^2 + \|u\|^2)$ where $\dim X = \dim \theta = p$ and $\dim U = k$ and suppose that $\theta \in R_+^p$. Then*

$$\delta(X) = \sum_{i=1}^n\left\{1 - \frac{\|U\|^2r_i(\|P_i(X)\|^2)}{(k+2)\|P_i(X)\|^2}\right\}P_i(X)I_{O_i}(X)$$

dominates the X_+, provided r_i is nondecreasing and bounded between 0 and $2(s-2)_+$.

6. BAYES ESTIMATION

There have been advancements in Bayes estimation of location vectors in several directions in the past 15 years. Perhaps the most important advancements have come in the computational area, particularly Markov chain Monte Carlo (MCMC) methods. We do not cover these developments in this review.

Admissibility and inadmissibility of (generalized) Bayes estimators in the normal case with known scale parameter was considered in Berger and Strawderman (1996) and in Berger, Strawderman and Tang (2005) where Brown's (1971) condition for admissibility (and inadmissibility) was applied for a variety of hierarchical Bayes models. Maruyama and Takemura (2008) also give admissibility results for the general spherically symmetric case. At least for spherically symmetric priors, the conditions are, essentially, that priors with tails no greater than $O(\|\theta\|^{-(p-2)})$ give admissible procedures.

Fourdrinier, Strawderman and Wells (1998), using Stein's (1981) results (especially Proposition 3.1 above, and its corollaries), give classes of minimax Bayes (and generalized Bayes) estimators which include scaled multivariate-t priors under certain conditions. Berger and Robert (1990) give classes of priors leading to minimax estimators. Kubokawa and Strawderman (2007) give classes of priors in the setup of Berger and Strawderman (1996) that lead to admissible minimax estimators. Maruyama (2003a) and Fourdrinier, Kortbi and Strawderman (2008), in the scale mixture of normal case, find Bayes and generalized Bayes minimax estimators, generalizing results of Strawderman (1974). As mentioned in Section 3, these results use either Berger's (1975) result (a version of which is given in Theorem 3.2) or Strawderman's (1974) result for mixtures of normal distributions. Fourdrinier and Strawderman (2008) proved minimaxity of generalized Bayes estimators corresponding to certain harmonic priors for classes of spherically symmetric sampling distributions which are not necessarily mixtures of normals. The results in this paper are not based directly on the discussion of Section 3 but are somewhat more closely related in spirit to the approach of Stein (1981).

We give below an intriguing result of Maruyama (2003b) for the unknown scale case (see also Maruyama and Strawderman, 2005), which is related to the (distributional) robustness of Stein estimators in the unknown scale case treated in Section 4. First, we give a lemma which will aid in the development of the main result.

LEMMA 6.1. *Suppose $(X, U) \sim \eta^{(p+k)/2} \cdot f(\eta\{\|x-\theta\|^2 + \|u\|^2\})$, the (location-scale invariant) loss is given by $L(\{\theta, \eta\}, \delta) = \eta\|\delta - \theta\|^2$ and the prior distribution on (θ, η) is of the form $\pi(\theta, \eta) = \rho(\theta)\eta^B$. Then provided all integrals exist, the generalized Bayes estimator does not depend on $f(\cdot)$.*

PROOF.

$$\delta(X, U)$$
$$= E[\theta\eta|X, U]/E[\eta|X, U]$$
$$= \left[\int_{R^p}\int_0^\infty \theta\eta^{(p+k)/2+B+1}\right.$$
$$\left. \cdot f(\eta\{\|X-\theta\|^2 + \|U\|^2\})\rho(\theta)\, d\eta\, d\theta\right]$$
$$\cdot \left[\int_{R^p}\int_0^\infty \eta^{(p+k)/2+B+1}\right.$$
$$\cdot f(\eta\{\|X-\theta\|^2$$
$$\left. + \|U\|^2\})\rho(\theta)\, d\eta\, d\theta\right]^{-1}.$$

Making the change of variables $w = \eta(\|X-\theta\|^2 + \|U\|^2)$, we have

$$\delta(X, U)$$
$$= \left[\int_{R^p} \theta(\|X-\theta\|^2 + \|U\|^2)^{-(p+k)/2+B+2}\right.$$
$$\left. \cdot \rho(\theta)\, d\theta \int_0^\infty w^{(p+k)/2+B+1} f(w)\, dw\right]$$
$$\cdot \left[\int_{R^p} (\|X-\theta\|^2 + \|U\|^2)^{-(p+k)/2+B+2}\right.$$
$$\left. \cdot \rho(\theta)\, d\theta \int_0^\infty w^{(p+k)/2+B+1} f(w)\, dw\right]^{-1}$$
$$= \frac{\int_{R^p} \theta(\|X-\theta\|^2 + \|U\|^2)^{-(p+k)/2+B+2}\rho(\theta)\, d\theta}{\int_{R^p} (\|X-\theta\|^2 + \|U\|^2)^{-(p+k)/2+B+2}\rho(\theta)\, d\theta}.$$
□

Hence, for (generalized) priors of the above form, the Bayes estimator is independent of the sampling distribution provided the Bayes estimator exists; thus, they may be calculated for the most convenient density, which is typically the normal. Our next lemma calculates the generalized Bayes estimator for a normal sampling density and for a class of priors for which $\rho(\cdot)$ is a scale mixture of normals.

LEMMA 6.2. *Suppose the distribution of (X, U) is normal with variance $\sigma^2 = 1/\eta$. Suppose also that the conditional distribution of θ given η and λ is normal with mean 0 and covariance $(1 - \lambda)/(\eta\lambda)I$, and the density of (η, λ) is proportional to $\eta^{b/2-p/2+a}$. $\lambda^{b/2-p/2-1}(1 - \lambda)^{-b/2+p/2-1}$, where $0 < \lambda < 1$.*

1. *Then the Bayes estimator is given by $(1 - r(W)/W)X$, where $W = \|X\|^2/\|U\|^2$ and $r(w)$ is given by*

(6.1)
$$r(w) = w\left[\int_0^1 \lambda^{b/2}(1 - \lambda)^{p/2-b/2-1}\right.$$
$$\left. \cdot (1 + w\lambda)^{-k/2-a-b/2-2}\, d\lambda\right]$$
$$\cdot \left[\int_0^1 \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}\right.$$
$$\left. \cdot (1 + w\lambda)^{-k/2-a-b/2-2}\, d\lambda\right]^{-1}.$$

This is well defined for $0 < b < p$, and $k/2 + a + b/2 + 2 > 0$.

2. *Furthermore, this estimator is generalized Bayes corresponding to the generalized prior proportional to $\eta^a\|\theta\|^{-b}$, for any spherically symmetric density $f(\cdot)$ for which $\int_0^\infty t^{(k+p)/2+a+1} f(t)\, dt < \infty$.*

PROOF. Part 1. In the normal case,

$$\delta(X, U) = X + \frac{E[\eta(\theta - X)|X, U]}{E[\eta|X, U]}$$
$$= X - \frac{\nabla_X m(X, U)}{2(\partial/\partial\|U\|^2)m(X, U)},$$

where the marginal $m(x, u)$ is proportional to

$$\int_0^1 \int_0^\infty \int_{R^p} \eta^{b/2+k/2+p/2+a} \lambda^{b/2-1}(1 - \lambda)^{-b/2-1}$$
$$\cdot \exp(-\eta\{\|x - \theta\|^2 + \|u\|^2\}/2)$$
$$\cdot \exp\left(-\frac{\eta\lambda\|\theta\|^2}{2(1 - \lambda)}\right) d\theta\, d\eta\, d\lambda$$
$$= K' \int_0^1 \int_0^\infty \eta^{b/2+k/2+a} \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}$$
$$\cdot \exp(-\eta\{\lambda\|x\|^2 + \|u\|^2\}/2)\, d\eta\, d\lambda$$
$$= K \int_0^1 (\lambda\|x\|^2 + \|u\|^2)^{-b/2-k/2-a-1}\lambda^{b/2-1}$$
$$\cdot (1 - \lambda)^{p/2-b/2-1}\, d\lambda.$$

Hence, we may express the Bayes estimator as $\delta(X, U) = X + g(X, U)$, where

$$g(x, u) = \left[\nabla_x \int_0^1 (\lambda\|x\|^2 + \|u\|^2)^{-b/2-k/2-a-1}\right.$$
$$\left. \cdot \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]$$
$$\cdot \left[-2(d/d\|u\|^2)\right.$$
$$\cdot \int_0^1 (\lambda\|x\|^2 + \|u\|^2)^{-b/2-k/2-a-1}$$
$$\left. \cdot \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]^{-1}$$
$$= -x\left[\int_0^1 (\lambda\|x\|^2 + \|u\|^2)^{-b/2-k/2-a-2}\right.$$
$$\left. \cdot \lambda^{b/2}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]$$
$$\cdot \left[\int_0^1 (\lambda\|x\|^2 + \|u\|^2)^{-b/2-k/2-a-2}\right.$$
$$\left. \cdot \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]^{-1}$$
$$= -x\left[\int_0^1 (\lambda w + 1)^{-b/2-k/2-a-2}\right.$$
$$\left. \cdot \lambda^{b/2}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]$$
$$\cdot \left[\int_0^1 (\lambda w + 1)^{-b/2-k/2-a-2}\right.$$
$$\left. \cdot \lambda^{b/2-1}(1 - \lambda)^{p/2-b/2-1}\, d\lambda\right]^{-1}$$
$$= -\frac{x}{w}r(w).$$

Part 2. A straightforward calculation shows that the unconditional density of (θ, η) is proportional to $\eta^a\|\theta\|^{-b}$. Hence, part 2 follows from Lemma 6.1. □

The following lemma gives properties of $r(w)$.

LEMMA 6.3. *Suppose $0 < b \leq p - 2$ and that $k/2 + a + 1 > 0$. Then, (1) $r(w)$ is nondecreasing, and (2) $0 < r(w) \leq b/(k + 2a + 2)$.*

PROOF. By a change of variables, letting $v = \lambda w$ in (6.1), then

$$r(w) = \left[\int_0^w (v + 1)^{-b/2-k/2-a-2}\right.$$
$$\left. \cdot v^{b/2}(1 - v/w)^{p/2-b/2-1}\, dv\right]$$

$$\cdot \left[\int_0^w (v+1)^{-b/2-k/2-a-2} \right.$$

$$\left. \cdot v^{b/2-1}(1-v/w)^{p/2-b/2-1}\, dv \right]^{-1}.$$

So, we may rewrite $r(w)$ as $E_w[v]$, where v has density proportional to $(1+v)^{-b/2-k/2-a-2}v^{b/2-1}(1-v/w)^{p/2-b/2-1}I_{[0,w]}(v)$. This density has increasing monotone likelihood ratio in w as long as $p/2 - b/2 - 1 \geq 0$. Hence, part 1 follows.

The conditions of the lemma allow interchange of limit and integration in both numerator and denominator of $r(w)$ as $w \to \infty$. Hence,

$$r(w) \leq \frac{\int_0^\infty (1+v)^{-b/2-k/2-a-2}v^{b/2}\, dv}{\int_0^\infty (1+v)^{-b/2-k/2-a-2}v^{b/2-1}\, dv}$$

$$= \frac{\int_0^1 u^{b/2}(1-u)^{k/2+a}\, du}{\int_0^1 u^{b/2-1}(1-u)^{k/2+a+1}\, du}$$

$$\text{[letting } u = v/(v+1)\text{]}$$

$$= \frac{\text{Beta}(b/2+1, k/2+a+1)}{\text{Beta}(b/2, k/2+a+2)}$$

$$= \frac{b/2}{k/2+a+1}. \qquad \square$$

Combining Lemmas 6.1–6.3 with Corollary 4.1 gives as the main result a class of estimators which are generalized Bayes and minimax simultaneously for the entire class of spherically symmetric sampling distributions (subject to integrability conditions).

THEOREM 6.1. *Suppose that the distribution of* (X, U) *and the loss function are as in Lemma* 6.1, *and that the prior distribution is as in Lemmas* 6.2 *and* 6.3 *with a satisfying* $b/(k+2a+2) \leq 2(p-2)/(k+2)$, *and with* $0 < b \leq p-2$. *Then the corresponding generalized Bayes estimator is minimax for all densities* $f(\cdot)$ *such that the* $2(a+2)$*th moment of the distribution of* (X, U) *is finite, that is,* $E(R^{2a+4}) < \infty$.

We note that the above finiteness condition, $E(R^{2a+4}) < \infty$, is equivalent to the finiteness condition, $\int_0^\infty t^{(k+p)/2+a+1}f(t)\, dt < \infty$, in Lemma 6.2.

7. CONCLUDING REMARKS

This paper has reviewed some of the developments in shrinkage estimation of mean vectors for spherically symmetric distributions, mainly since the review paper of Brandwein and Strawderman (1990). Other papers in this volume review other aspects of the enormous literature generated by or associated with Stein's stunning inadmissibility result of 1956.

Most of the developments we have covered are, or can be viewed as, outgrowths of Stein's papers of 1973 and 1981, and, in particular, of Stein's lemma which gives (an incredibly useful) alternative expression for the cross product term in the quadratic risk function.

Among the topics which we have not covered is the closely related literature for elliptically symmetric distributions (see, e.g., Kubokawa and Srivastava, 2001, and Fourdrinier, Strawderman and Wells, 2003, and the references therein). We also have not included a discussion of Hartigan's (2004) beautiful result that the (generalized or proper) Bayes estimator of a normal mean vector with respect to the uniform prior on any convex set in R^p dominates X for squared error loss. Nor have we discussed the very useful and pretty development of the Kubokawa (1994) IERD method for finding improved estimators, and, in particular, for dominating James Stein estimators (see also Marchand and Strawderman, 2004, for some discussion of these last two topics). We nonetheless hope we have provided some intuition for, and given a flavor of the developments and rich literature in the area of improved estimators for spherically symmetric distributions.

The impact of Stein's beautiful 1956 result and his innovative development of the techniques in the 1973 and 1981 papers have inspired many researchers, fueled an enormous literature on the subject, led to a deeper understanding of theoretical and practical aspects of "sharing strength" across related studies, and greatly enriched the field of Statistics. Even some of the early (and later) heated discussions of the theoretical and practical aspects of "sharing strength" across unrelated studies have had an ultimately positive impact on the development of hierarchical models and computational tools for their analysis. We are very pleased to have been asked to contribute to this volume commemorating fifty years of development of one of the most profound results in the Statistical literature in the last half of the 20th century.

REFERENCES

BARANCHIK, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **41** 642–645. MR0253461

BERGER, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Ann. Statist.* **3** 1318–1328. MR0386080

BERGER, J. O. and ROBERT, C. (1990). Subjective hierarchical Bayes estimation of a multivariate normal mean: On the frequentist interface. *Ann. Statist.* **18** 617–651. MR1056330

BERGER, J. O. and STRAWDERMAN, W. E. (1996). Choice of hierarchical priors: Admissibility in estimation of normal means. *Ann. Statist.* **24** 931–951. MR1401831

BERGER, J. O., STRAWDERMAN, W. and TANG, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Ann. Statist.* **33** 606–646. MR2163154

BRANDWEIN, A. C., RALESCU, S. and STRAWDERMAN, W. E. (1993). Shrinkage estimators of the location parameter for certain spherically symmetric distributions. *Ann. Inst. Statist. Math.* **45** 551–565. MR1240354

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1990). Stein estimation: The spherically symmetric case. *Statist. Sci.* **5** 356–369. MR1080957

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1991). Generalizations of James–Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650. MR1126343

BROWN, L. D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *Ann. Math. Statist.* **42** 855–903. MR0286209

CELLIER, D. and FOURDRINIER, D. (1995). Shrinkage estimators under spherical symmetry for the general linear model. *J. Multivariate Anal.* **52** 338–351. MR1323338

CHANG, Y. T. (1982). Stein-type estimators for parameters in truncated spaces. *Keio Sci. Tech. Rep.* **35** 185–193.

DU PLESSIS, N. (1970). *An Introduction to Potential Theory. Conn. Univ. Mathematical Monographs* No. 7. Hafner, Darien, CT. MR0435422

FOURDRINIER, D., KORTBI, O. and STRAWDERMAN, W. E. (2008). Bayes minimax estimators of the mean of a scale mixture of multivariate normal distributions. *J. Multivariate Anal.* **99** 74–93. MR2408444

FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (1998). On the construction of Bayes minimax estimators. *Ann. Statist.* **26** 660–671. MR1626063

FOURDRINIER, D., STRAWDERMAN, W. E. and WELLS, M. T. (2003). Robust shrinkage estimation for elliptically symmetric distributions with unknown covariance matrix. *J. Multivariate Anal.* **85** 24–39. MR1978175

FOURDRINIER, D. and STRAWDERMAN, W. E. (2008). Generalized Bayes minimax estimators of location vectors for spherically symmetric distributions. *J. Multivariate Anal.* **99** 735–750. MR2406080

HARTIGAN, J. A. (2004). Uniform priors on convex sets improve risk. *Statist. Probab. Lett.* **67** 285–288. MR2060127

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. Probab.* **I** 361–379. Univ. California Press, Berkeley. MR0133191

KUBOKAWA, T. (1994). A unified approach to improving equivariant estimators. *Ann. Statist.* **22** 290–299. MR1272084

KUBOKAWA, T. and SRIVASTAVA, M. S. (2001). Robust improvement in estimation of a mean matrix in an elliptically contoured distribution. *J. Multivariate Anal.* **76** 138–152. MR1811829

KUBOKAWA, T. and STRAWDERMAN, W. E. (2007). On minimaxity and admissibility of hierarchical Bayes estimators. *J. Multivariate Anal.* **98** 829–851. MR2322131

MARCHAND, E. and STRAWDERMAN, W. E. (2004). Estimation in restricted parameter spaces: A review. In *A Festschrift for Herman Rubin. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **45** 21–44. IMS, Beachwood, OH. MR2126884

MARUYAMA, Y. (2003a). Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distributions. *J. Multivariate Anal.* **84** 274–283. MR1965222

MARUYAMA, Y. (2003b). A robust generalized Bayes estimator improving on the James–Stein estimator for spherically symmetric distributions. *Statist. Decisions* **21** 69–77. MR1985652

MARUYAMA, Y. and STRAWDERMAN, W. E. (2005). A new class of generalized Bayes minimax ridge regression estimators. *Ann. Statist.* **33** 1753–1770. MR2166561

MARUYAMA, Y. and TAKEMURA, A. (2008). Admissibility and minimaxity of generalized Bayes estimators for spherically symmetric family. *J. Multivariate Anal.* **99** 50–73. MR2408443

MENG, X.-L. (2005). From unit root to Stein's estimator to Fisher's *k* statistics: If you have a moment, I can tell you more. *Statist. Sci.* **20** 141–162. MR2183446

SENGUPTA, D. and SEN, P. K. (1991). Shrinkage estimation in a restricted parameter space. *Sankhyā Ser. A* **53** 389–411. MR1189780

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proc. Third Berkeley Sympos. Math. Statist. Probab.* 1954–1955 **I** 197–206. Univ. California Press, Berkeley. MR0084922

STEIN, C. (1974). Estimation of the mean of a multivariate normal distribution. In *Proceedings of the Prague Symposium on Asymptotic Statistics* (*Charles Univ., Prague*, 1973) **II** 345–381. Charles Univ., Prague. MR0381062

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

STIGLER, S. M. (1990). The 1988 Neyman memorial lecture: A Galtonian perspective on shrinkage estimators. *Statist. Sci.* **5** 147–155. MR1054859

STRAWDERMAN, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* **4** 255–264. MR0362597

STRAWDERMAN, W. E. (1992). The James–Stein estimator as an empirical Bayes estimator for an arbitrary location family. In *Bayesian Statistics* **4** (*Peñíscola*, 1991) 821–824. Oxford Univ. Press, New York. MR1380308