On improved shrinkage estimators for concave loss

Tatsuya Kubokawa[a], Éric Marchand[b], William E. Strawderman[c]
a Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN (e-mail: tatsuya@e.u-tokyo.ac.jp)
b Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)
c Rutgers University, Department of Statistics and Biostatistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

Abstract:
We consider minimax shrinkage estimation of location for spherically symmetric distributions under a concave function of the usual squared error loss. Scale mixtures of normal distributions and losses with completely monotone derivatives are featured.

AMS 2000 subject classifications: Primary 62C20, 62F15; secondary 62A15.
Keywords and phrases: location parameters, minimaxity, Stein estimation, concave loss.

1. Introduction

This paper concerns minimax shrinkage estmation of a location vector of a spherically symmetric distribution under a loss function which is a concave function of the usual squared error loss. The main contribution is an improvement in shrinkage constants for minimax estimators over those in Brandwein and Strawderman (1980, 1991), and Brandwein Ralescu and Strawderman (1994), particularly for variance mixtures of normals (and somewhat more generally), and for concave functions of squared error loss for which the derivative of the concave function is completely monotone (and somewhat more generally). For Baranchik-type estimators and for scale mixtures of multivariate normal distributions, we also show that our minimax improvements hold in dimension 3 which improves over the restriction that $p \geq 4$ in the earlier papers.

Specifically, let X have the p-dimensional spherically symmetric density

$$f(||X-\theta||^2) \tag{1.1}$$

and consider the problem of estimating the unknown location vector θ under the loss function

$$L(\theta, d) = l(||d-\theta)||^2), \tag{1.2}$$

where $l(t)$ is a non-negative, non-decreasing concave function, and $||d-\theta||^2$ is the usual squared error loss function.

For a multivariate normal distribution with squared error loss, James and Stein (1961), Baranchik (1970), Strawderman (1971), Stein (1981) and others gave shrinkage estimators of the mean vector which are minimax and which improve on the usual estimator, X, when the dimension, p, is at least three. Strawderman (1974) gave extensions to variance mixtures of normals. Extensions to wider classes of spherically symmetric distributions were provided by Berger (1975), Brandwein and Strawderman (1978), Brandwein (1979), Brandwein and Strawderman (1991) and others.

Brandwein and Strawderman (1980), Brandwein and Strawderman (1991) and Brandwein, Ralescu, and Strawderman (1993) gave minimax shrinkage estimators which improve on X in higher dimensions for concave functions of squared error loss.

A basic tool in much of the literature on concave loss is the following simple result which will be used extensively in the following.

Lemma 1.1. *Suppose that X is distributed as in (1.1), and that loss is given by (1.2), where $l(t)$ is a non-negative, non-decreasing concave function such that $l'(t)$ exists.*

a. Then the risk, $R(\theta,\delta)$, of an estimator of the form $\delta(X)=X+g(X)$, satisfies the inequality

$$R(\theta,\delta) \leq R(\theta,X) + E_\theta[l'(||X-\theta||^2)(||g(X)||^2 - 2(X-\theta)'g(X))]. \tag{1.3}$$

b. Hence $\delta(X)$ dominates X under loss (1.2) if it dominates X under quadratic loss, $||d-\theta||^2$, for a location family with density $f^*(||x-\theta||^2)$ proportional to $f(||x-\theta||^2)l'(||x-\theta||^2)$.

Proof. Part a. follows easily, on taking expectations, from the concave function inequality $l(t+y) \leq l(t) + yl'(t)$ with $t = ||X-\theta||^2$ and $y = ||g(X)||^2 + 2(X-\theta)'g(X)$.

Part b. follows immediately from part a. □

Remark 1.1. That the usual estimator X is minimax follows fairly easily when the underlying spherically symmetric density (1.1) is unimodal and when the loss function (1.2) is monotone (but not necessarily concave) in $||d-\theta||^2$. This follows because X is the unique minimum risk equivariant (MRE) estimator under these assumptions and is hence minimax by the well known result that if a minimax estimator exists in a location problem, there is an equivariant minimax estimator. For completeness we formalize this in Theorem 1.1.

Theorem 1.1. *Suppose that the density $f(\|X-\theta\|^2)$ is unimodal and that $\ell'(t) \geq 0$ with $\ell'(t) > 0$ on an interval. Then, the usual estimator X is the unique minimum risk equivariant (MRE) estimator, and is hence minimax.*

Proof. Note first that a general equivariant estimator is of the form $X + d$ for some vector d in $\Re^p$, which because of the spherical symmetry we may take to be of the form $(a, 0, \ldots, 0)$. Let $Y = (X_2, \ldots, X_p)$, let $\ell_Y((X_1-a)^2) = \ell((X_1-a)^2)+\|Y\|^2)$. Then, conditioning on Y, it suffices to show that $E[\ell_Y((X_1-a)^2)|Y] > E[\ell_Y(X_1^2)|Y]$ for $a \neq 0$. Let $f_Y((x_1-a)^2)$ be the conditional density of X_1 given Y. The conditional risk difference is expressed as

$$\begin{aligned}
\Delta_Y(a) =& E[\ell_Y((X_1-a)^2)|Y] - E[\ell_Y(X_1^2)|Y] = E\Big[\int_0^a \frac{d}{dt}\ell_Y((X_1-t)^2)dt\Big|Y\Big] \\
=& 2\int_{-\infty}^{\infty}\int_0^a (t-x)\ell_Y'((x-t)^2)f_Y(x^2)dtdx \\
=& 2\int_{-\infty}^{\infty}\int_0^a (-z)\ell_Y'(z^2)f_Y((z+t)^2)dtdz \\
=& 2\int_{-\infty}^{\infty}(-z)\ell_Y'(z^2)\int_z^{a+z} f_Y(u^2)dudz,
\end{aligned}$$

where we used the transformations $z = x - t$ $(dz = dx)$ and $u = t + z$ $(du = dt)$. Let $F_Y(z) = \int_{-\infty}^z f_Y(u^2)du$. Then, we can rewrite the conditional risk difference as

$$\begin{aligned}
\Delta_Y(a) =& 2\int_{-\infty}^{\infty} z\ell_Y'(z^2)\{F_Y(z) - F_Y(z+a)\}dz \\
=& 2\int_0^{\infty} z\ell_Y'(z^2)\Big[\{F_Y(z) - F_Y(z+a)\} - \{F_Y(-z) - F_Y(-z+a)\}\Big]dz \\
=& 2\int_0^{\infty} z\ell_Y'(z^2)\Big[\{F_Y(z) - F_Y(-z)\} - \{F_Y(z+a) - F_Y(-z+a)\}\Big]dz \\
=& 2\int_0^{\infty} z\ell_Y'(z^2)\Big[\int_{-z}^{z} f_Y(u^2)du - \int_{-z+a}^{z+a} f_Y(u^2)du\Big]dz.
\end{aligned}$$

Since $f_Y(u^2)$ is symmetric and unimodal about $u = 0$, it can be seen that $\int_{-z}^{z} f_Y(u^2)du > \int_{-z+a}^{z+a} f_Y(u^2)du$ for $a \neq 0$. Noting that $\ell_Y'(z^2) \geq 0$ with $\ell_Y'(z^2) > 0$ on an interval of z, we can conclude that $\Delta_Y(a) > 0$ for $a \neq 0$. □

In almost all of our examples, the density (1.1) is unimodal and hence both X and our improved estimators will be minimax. It seems likely that minimaxity of X via its uniqueness as an MRE is considerably more general than illutrated herein.

In section 2 of this paper we combine the methods of Strawderman (1974) and Brandwein and Strawderman (1980) to enlarge the range of minimax shrinkage over that in Brandwein and Strawderman (1980) for variance mixtures of normals for James-Stein and Baranchik-type estimators. However we restrict the class of concave loss functions to those for which $l'(t)$ is completely monotone.

In Section 3 we broaden the class of estimators as well as the class of distributions and losses and obtain similar extensions in the range of minimax shrinkage over that in Brandwein and Strawderman (1991).

Our interest in returning to the problem of improved estimation of location parameters under concave loss was generated by our recent research in predictive density estimation for location families under integrated L_1 and L_2 losses. Often, the loss function for a plug in-type density estimator turns out to be equivalent to a concave loss function of the form (1.2) in the associated location estimation problem. Thus improved estimators in the point estimation problem under concave loss lead to improved plug in estimators in the associated predictive density estimation problem. We report on this work in another publication.

2. Scale Mixtures of Normals

A random vector, X, in p -dimensions has a scale mixture of normal distributions if its distribution has the following hierarchical structure: the distribution of X given V is $N_p(\theta, VI_p)$, where V is a non-negative random variable with cdf $H(v)$. Scaled multivariate-t distributions are perhaps the most important of these and form an important alternative error distribution to the multivariate normal in a variety of modelling situations where tails which are thicker than those of the normal distribution seem warranted.

A basic result for estimation of the mean vector θ under quadratic loss, $||d-\theta||^2$, for such mixtures is the following, from Strawderman (1974):

Lemma 2.1. *(Strawderman (1974)) Suppose that X has a scale mixture of normal distribution with mixing distribution $H(v)$. An estimator of θ of the form*

$$\delta_a(X) = (1 - ar(||X||^2)/||X||^2)X \tag{2.1}$$

dominates X under loss $||d-\theta||^2$ provided $E[V]$ and $E[V^{-1}]$ are finite, and
a. $0 \le r(t) \le 1$
b. $r(t)$ is non-decreasing,
c. $r(t)/t$ is non-increasing, and
d. $0 < a \le 2/E_0[1/||X||^2] = 2(p-2)/E[1/V]$,
and $\delta_a(X)$ differs from X on a set of positive measure.

We also require the following preliminary result:

Lemma 2.2. *a. A density of the form (1.1) is a mixture of normals if and only if $f(t)$ is completely monotone, i.e., $(-1)^i(d^i/dt^i f(t)) \ge 0$, for $i = 0, 1, \ldots$.*
b. The product of two completely monotone functions is completely monotone.

Proof. Part a. is well known. See, e.g. Berger (1975) for some discussion and references. A straightforward computation establishes Part b. □

Theorem 2.1 is the main result of this section.

Theorem 2.1. *Suppose that X has a scale mixture of normal distribution with mixing distribution $H(v)$. Consider estimation of θ under the concave loss(1.2), where $l'(t)$ is a completely monotone function.*

Then, an estimator of the form

$\delta_a(X) = (1 - ar(||X||^2)/||X||^2)X$ dominates X provided the conditions of Lemma 2.1 hold with d. replaced by

dd. $0 < a \leq 2/E_0^*[1/||X||^2]$, where E^* denotes expectation with respect to the density of Lemma 1.1 (i.e., $f^*(t)$ is proportional to $f(t)l'(t)$).

Proof. The density $f^*(\cdot)$ defined in Lemma 1.1 is completely monotone and is hence a scale mixture of normals by Lemma 2.2. The result then follows from Lemma 2.1. □

Many, if not most of the concave losses encountered in practice are covered by the theorem, including "L_q" for $0 < q \leq 2$ i.e., ($L(\theta, d) = ||d - \theta||^q$, $l(t) = t^{q/2}$), losses formed from reflecting a multiple of the pdf, f, of a univariate scale mixture of normals ($l(t) = k(f(0) - f(t))$), and also $2F(||d-\theta||) - 1$, where $F(t)$ is a (one-dimensional) cdf for which $F'(t^{1/2})/t^{1/2}$ is completely monotone. These losses play a role in the problem of improved plug in-type predictive density estimation under several different predictive density estimation losses. These include Kullback-Leibler, α-divergence, and integrated L_1 and L_2 differences.

Remark 2.1. The improvement given by Theorem 2.1 over the corresponding result in Brandwein and Strawderman (1980) comes mainly from an extension of the range of the shrinkage constant a from $0 < a \leq 2(p-2)/(pE_0^*[1/||X||^2])$ in the earlier paper to $0 < a \leq 2/E_0^*[1/||X||^2]$ in Theorem 2.1. This represents a large improvement for low dimensional problems. For example, for $p = 4$, Theorem 2.1 gives a 100 percent increase in the shrinkage factor. Additionally, the above result applies for $p \geq 3$, while that of Brandwein and Strawderman (1980) applies for $p \geq 4$. Of course, the result in Brandwein and Strawderman (1980) applies to the more general class of all spherically symmetric unimodal distributions and not just scale mixtures of normals.

Example 1: (Normal Distributions, L_q loss) Suppose X has a multivariate normal distribution with mean vector θ and covariance matrix I_p. Let the loss function be given by $L(\theta, d) = ||d - \theta||^q$ ($l(t) = t^{q/2}$), for some $0 < q \leq 2$. Brandwein and Strawderman (1980) show, for this case, that $E_0^*[1/||X||^2]$ of Theorem 2.1, c. is equal to $1/(p + q - 4)$. Therefore estimators of the form (2.1) satisfying $0 < a \leq 2(p + q - 4)$ dominate X for this loss. The corresponding result in Brandwein and Strawderman (1980) reduces the upper bound by a factor of $(p-2)/p$.

Note that in either case there is a requirement that $p > 4 - q$ in order for the required expectations to exist. Thus for $0 < q \leq 1$, both results require $p \geq 4$. However for $1 < q \leq 2$ only the result of this section applies if $p = 3$.

Example 2: (Scale Mixtures of Normal Distributions, Reflected Normal Loss) Suppose X has a p-variate scale mixture of normal distribution with mean vector θ and mixing variance V. Let the loss function be given by $L(\theta, d) = 1 - \exp(-||d-\theta||^2/w)$ for some $w > 0$ ($l(t) = 1 - \exp(-t/w)$, $l'(t) = \exp(-t/w)/w$). In this case the requirements of Theorem 2.1 are met and it is straightforward to see that the density $f^*(||x - \theta||^2)$ of Lemma 1.1 is a scale mixture of normals with a mixing variance given by $Vw/(V + w)$. Hence the range of values for the shrinkage constant in Theorem 2.1 is $0 < a \leq 2/E_0^*[1/||X||^2] =$

$2(p-2)/E[(Vw/(V+w))^{-1}]$. Once again, the corresponding result in Brandwein and Strawderman (1980) would reduce the upper bound by a factor of $(p-2)/p$. In this example, the dimensions for the applicability of the result differ, being $p \geq 3$ for this paper and $p \geq 4$ for Brandwein and Strawderman (1980). This example can easily be extended to the case where the loss is a mixture (in w) of the above loss.

3. More General Estimators, Distributions, and Losses

This section is devoted to minimaxity of general estimators of the form

$$\delta(X) = X + ag(X), \tag{3.1}$$

and is not restricted to Baranchik-type estimators of the form (2.1).

The family of spherically symmetric distributions will also be enlarged from the class of scale mixtures of normal distributions considered in section 2 to the class of spherically symmetric distributions of the form (1.1) satisfying the following Assumption:

Assumption 1:$f(t) \geq 0$, and $f(t)/(-f'(t))$ is non-decreasing.
We also will refer to this assumption for non-negative functions which are not necessarily a density.

A related assumption is,

Assumption 1a: $F(t)/f(t)$ is monotone non-decreasing, where

$$F(t) = \int_t^\infty f(u)\, du, \tag{3.2}$$

provided $F(0)$ exists and is finite.

Lemma 3.1 below states that Assumption 1 implies Assumption 1a, and also that scale mixtures of normal distributions (and more generally, completely monotone functions) satisfy Assumption 1 (and therefore 1a provided $f(t)$ is integrable). That Assumption 1 implies Assumption 1a is well known (see e.g. Bagnoliane and Bergstrom (2006)) but we provide a simple proof for completeness.

Additionally, the class of concave loss functions will be enlarged from that of Theorem 2.1 to the class of loss functions of the form (1.2) satisfying the following assumption.

Assumption 2: $l(t)$ is non-negative, non-decreasing and concave,
and $l'(t)/(-l''(t))$ is non-decreasing.

Again Lemma 3.1 implies that a concave loss, $l(t)$ such that $l'(t)$ is completely monotone satisfies Assumption 2. Part c gives a useful sufficient (essentially equivalent) condition.

Lemma 3.1. *a. If $h(t)$ is non-negative and satisfies Assumption 1, then it also satisfies Assumption 1a provided $h(t)$ is integrable, i.e $H(0)$ is finite.*

b. If $f(\cdot)$ in (1.1) is the density of a scale mixture of normals, then $f(t)$ satisfies Assumption 1.

c. If $h(t)$ is non-negative, and $\log(h(t))$ is convex, then $h(t)$ satisfies assumption 1.

d. The product of two non-negative functions $g_i(t), i = 1, 2$, such that each satisfies Assumption 1, also satisfies Assumption 1.

Proof. To show part a, note that
$H(t)/h(t)$
$= \int_t^\infty h(u)\ du/(-\int_t^\infty h'(u)\ du)$
$= \int_t^\infty h(u)\ du/(\int_t^\infty (-h'(u)/h(u))h(u)\ du)$
$= 1/E[(-h'(U)/h(u)]$,
where the expectation is with respect to the density proportional to $h(u)$ on the interval $u \geq t$, which has monotone likelihood ratio in t. Hence part a follows by the assumed monotonicity of $h(t)/(-h'(t))$.

Parts b, c, and d follow by straightforward calculations. □

A basic result underlying the main Theorem (3.2) of this section is due to Brandwein, Ralescu, and Strawderman (1993) for quadratic loss, and which is presented next in the notation of this paper.

Theorem 3.1. *(Brandwein, Ralescu, and Strawderman (1993), Theorem 2.2) Let X in $\mathbb{R}^p$ have a a density of the form (1.1) which satisfies Assumption 1a, and let $\delta_a(X) = X + a\ g(X)$ where $g(X)$ satisfies*

a. $-div\ g(X) \geq -h(X)$ where, for each θ, $h(X)$ is such that $E[R^2 h(W)]$ is non-increasing in R, where W has a uniform distribution on the sphere of radius R centered at θ.

b. $||g(X)||^2 + 2\ h(X) \leq 0$, and

c. $0 < a \leq 1/[(p-2)E[1/||X||^2]]$

Then $\delta_a(X)$ is minimax and dominates X under quadratic loss $||d-\theta||^2$ (provided all expectations exist and are finite).

Brandwein, Ralescu, and Strawderman (1993) note, for $p \geq 4$, that for estimators of the Baranchik form (2.1), conditions a. and b. are satisfied provided $0 \leq ar(t) \leq 2(p-2)$, $r(t)$ is non-decreasing and $r(t)/t$ is non-increasing.

Here is the main result of this section.

Theorem 3.2. *Let X in $\mathbb{R}^p$ have a density of the form(1.1) which satisfies Assumption 1, and let $\delta_a(X) = X + a\ g(X)$ where $g(X)$ satisfies a. and b. of Theorem 3.1. Let the loss be of the form (1.2), where $l(t)$ satisfies Assumption 2. Then $\delta_a(X)$ is minimax and dominates X provided $0 < a \leq 1/[(p-2)E^*[1/||X||^2]]$, where E^* represents expectation with respect to the density proportional to $f^*(||X-\theta||^2)$ with $f^*(t) = f(t)l'(t)$ (provided all expectations exist and are finite).*

Proof. The proof follows directly from Theorem 3.1 since f(t)l'(t) satisfies Assumption 1a by Lemma 3.1.

□

Remark 3.1. While Theorem 3.2 applies to wider classes of densities, estimators and losses than Theorem 2.1 it is not strictly more general. In part, this is because Theorem 3.2 requires the dimension, p to be at least 4 when applied to Stein-type estimators (see the comment immediately after theorem 3.1) while Theorem 2.1 applies also for $p = 3$ as long as all expectations are finite. However for James-Stein estimator (and more generally, for Baranchik-type estimators with non-decreasing $r(t)$ and non-increasing $r(t)/t$), the two results give the same range of minimax shrinkage, although this is not immediately obvious from the statements of the two theorems.

To see that both give the same result for James-Stein estimators, note that Theorem 3.2, applied to $\delta_a(X) = (1 - a\ 2(p-2)/||X||^2)X$, with $h(X) = -2(p-2)/||X||^2$, gives minimaxity for $0 \le a \le 1/[(p-2)E^*[1/||X||^2]]$, or equivalently for $0 \le 2a(p-2) \le 2/E^*[1/||X||^2]]$, which is the same range of minimax shrinkage given by Theorem 2.1. Note that $a = 0$ corresponds to the minimax estimator X so we include $a = 0$ in the range of minimax shrinkage (but not domination over X) in each case.

Remark 3.2. The results of this section give essentially the same improvement in the range of minimax shrinkage over corresponding results in Brandwein and Strawderman (1991) as do those observed in the remark of section 2 for Baranchik-type estimators. In addition the requirements on the function $g(X)$ are somewhat less stringent due to the developments in Brandwein, Ralescu, and Strawderman (1993).

Example 3: (Normal Distributions, Reflected Normal Loss) Suppose X has a p-variate normal distribution with mean vector θ and covariance matrix vI_p. Let the loss function be given by $L(\theta, d) = 1 - \exp(-||d-\theta||^2/w)$ for some $w > 0$ ($l(t) = 1 - \exp(-t/w)$, $l'(t) = \exp(-t/w)/w$). In this case Lemma 1.1 applies directly and $f^*(||x-\theta||^2)$ is seen to be a multivariate normal distribution with mean vector θ, and covariance matrix $(vw/(v+w))I_p$. Hence classical normal theory applies and, for example, an estimator of the form $\delta(X) = X + a\ (1/v + 1/w)^{-1}g(X)$ is minimax and dominates X provided $||g(x)||^2 + 2\ divg(x) \le 0$, and $0 < a \le 1$ without recourse to the results of this section (but, instead, directly using the results of Stein (1981) together with Lemma 1.1).

If, however the distribution of X is a scale mixture of normals then, with the additional requirements on $g(X)$, and a given in Theorem 3.2, a class of general minimax estimators dominating X may be obtained. In addition such results may be extended to the case where the loss is a mixture in w of the above loss, e.g., a folded scale mixture of normals loss, such as $L(\theta, d) = 1 - 1/||d-\theta||^2$.

4. Concluding Remarks

In this paper we have studied estimation of the location vector for p-variate spherically symmetric distributions under losses which are concave functions of the usual squared-error loss. For distributions which are scale mixtures of multivariate normals and for concave losses for which the derivative $l'(t)$ is completely monotone we give minimax estimators of the Baranchik form for which the range of minimax shrinkage is considerably extended over that found in Brandwein and Strawderman (1980). For the broader class of distributions and losses satisfying the assumptions of section 3, and for a broader class of estimators (essentially those satisfying a version of Stein's differential inequality plus other monotonicity assumptions), we have similarly broadened the range of minimax shrinkage compared to the results of Brandwein and Strawderman (1991).

Acknowledgements

Tatsuya Kubokawa's research is supported by Grants-in-Aid for Scientific Research Nos. 19200020 and 21540114 from the Japan Society for the Promotion of Science, Erc Marchand's research is supported in part by a grant from the Natural Sciences and Engineering Research of Canada, and William Strawderman's research is partially supported by a grant from the Simomns Foundation (#209035).

References

Bagnoli, M. and Bergstrom, T. (2005). Log-concave probability and its applications. *Economic Theory* **26** 445-469.

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

Berger, J. O. (1975). Minimax estimation of location vectors for a wide class of distributions. *Ann. Statist.* **3** 1318-1328.

Brandwein, A. C. (1979). Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multivariate Anal.* **9** 579-588.

Brandwein, A. C. Ralescu, S. and Strawderman, W. E. (1993). Shrinkage estimation of the location parameters for certain spherically symmetric distributions. *Ann. Inst. Statist. Math.* **45** 551-565.

Brandwein, A. C. and Strawderman, W. E. (1980). Minimax estimation of location parmeters for spherically symmetric distributions with concave loss. *Ann. Statist.* **8** 279-284.

Brandwein, A. C. and Strawderman, W. E. (1978). Minimax estimation of location parmeters for spherically symmetric unimodal distributions. *Ann. Statist.* **6** 377-416.

Brandwein, A. C. and Strawderman, W. E. (1991). Generalizations of James-Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639-1650.

James, W. and Stein, C. (1961). Estimation with quadratic loss. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I* 361–379. Univ. California Press, Berkeley, Calif.

Strawderman, W. E. (1971). Proper Bayes minimax estimators of the multivariate normal mean. *Ann. Math. Statist.* **42** 385-388.

Strawderman, W. E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *J. Multivariate Anal.* **8** 255-264.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.