Chapter 13
Estimating the Location Vector for Spherically Symmetric Distributions

Jian-Lun Xu

Abstract When a $p \times 1$ random vector $\mathbf{X}$ has a spherically symmetric distribution with the location vector $\boldsymbol{\theta}$, Brandwein and Strawderman [7] proved that estimators of the form $\mathbf{X} + a\mathbf{g}(\mathbf{X})$ dominate the $\mathbf{X}$ under quadratic loss if the following conditions hold: (i) $\|\mathbf{g}\|^2/2 \leq -h \leq -\nabla \circ \mathbf{g}$, where $-h$ is superharmonic, (ii) $E[-R^2 h(\mathbf{V})]$ is nondecreasing in R, where $\mathbf{V}$ has a uniform distribution in the sphere centered at $\boldsymbol{\theta}$ with a radius $R = \|\mathbf{X} - \boldsymbol{\theta}\|$, and (iii) $0 < a \leq 1/[pE(R^{-2})]$. In this paper we not only use a weaker condition than their (ii) to show the dominance of $\mathbf{X} + a\mathbf{g}(\mathbf{X})$ over the $\mathbf{X}$, but also obtain a new bound $E(R)/[pE(R^{-1})]$ for a, which is always better than bounds obtained by Brandwein and Strawderman [7] and Xu and Izmirlian [24]. The generalization to concave loss function is also considered. In addition, estimators of the location vector are investigated when the observation contains a residual vector and the scale is unknown.

13.1 Introduction

It is well-known that the normal distribution and its related statistical inference such as estimation of its mean are crucial in application. Ever since Stein [19] discovered the inadmissibility of the best invariant estimator of the p-dimensional ($p \geq 3$) normal mean under quadratic loss, there has been considerable interest in improving upon the best invariant estimator of a location vector by relaxing the normality assumption, studying more general estimators, or considering different loss functions.

Under the quadratic loss, James and Stein [15] presented a class of dominating estimators, $\left(1 - a/\|\mathbf{X}\|^2\right)\mathbf{X}$ for $0 < a < 2(p-2)$ if $\mathbf{X}$ has a normal distribution with the identity covariance matrix I_p. This result remains true if the distribution of $\mathbf{X}$ is spherically symmetric about its location vector and $p \geq 4$ as shown by Brandwein [2],

J.-L. Xu (✉)
Biometry Research Group, National Cancer Institute,
9609 Medical Center Drive, Room 5E128, Bethesda, MD 20892-9789, USA
e-mail: xujia@mail.nih.gov

Brandwein and Strawderman [3, 4, 7], Fan and Fang [12–14], Maruyama [16], and Brown and Zhao [9], Tosh and Dasgupta [21] and others; see the review articles by Brandwein and Strawderman [5, 6]. When the dimension is at least three, Brown [8] also proved that the best invariant estimator of a location vector is inadmissible for a wide class of distributions and loss functions. When the components of $\mathbf{X}$ are independent, identically and symmetrically (iis) distributed about their respective means, Shinozaki [18] studied the dominance conditions of the James-Stein type estimator

$$\delta_{a,b}(\mathbf{X}) = \left(1 - \frac{a}{b + \|\mathbf{X}\|^2}\right)\mathbf{X}, \tag{13.1}$$

over $\mathbf{X}$ and obtained the bounds of a and b in (13.1) that depend on the second and fourth moments of the component distributions. Xu [23] investigated the bounds of a and b in (13.1) when $\mathbf{X}$ has a sign-invariant distribution.

For more general estimators and different loss functions, Miceli and Strawderman [17] restricted the distribution of $\mathbf{X}$ to the subclass of iis distributions called independent component variance mixtures of normals and replaced a in (13.1) by $ar(X_1^2, \ldots, X_p^2)$, where $r(X_1^2, \ldots, X_p^2)$ is a function of $X_1^2, \ldots, X_p^2$. Their loss function is nonquadratic. When $\mathbf{X}$ has a spherically symmetric distribution about its location vector $\boldsymbol{\theta}$ and loss function is a quadratic loss, a concave function of quadratic loss, or the general quadratic loss, Brandwein and Strawderman [7] elegantly used the *divergence* theorem to prove the dominance of the estimators

$$\delta_{a,\mathbf{g}}(\mathbf{X}) = \mathbf{X} + a\mathbf{g}(\mathbf{X}). \tag{13.2}$$

over $\mathbf{X}$ under conditions (i) $\|\mathbf{g}\|^2/2 \leq -h \leq -\nabla \circ \mathbf{g}$, where $-h$ is superharmonic, (ii) $E[-R^2 h(\mathbf{V})]$ is nondecreasing in R, where $\mathbf{V}$ has a uniform distribution in the sphere centered at $\boldsymbol{\theta}$ with a radius $R = \|\mathbf{X} - \boldsymbol{\theta}\|$, and (iii) $0 < a \leq 1/[pE(R^{-2})]$. Clearly, the estimators $\delta_{a,\mathbf{g}}(\mathbf{X})$ given by (13.2), together with conditions (i) and (iii) extend the classical James-Stein estimator to a broader class of estimators, while their condition (ii) is a technical condition. Xu and Izmirlian [24] dropped their technical condition (ii) and obtained a bound $0 < a < [\mu_1/(p^2\mu_{-1})][1 - (p-1)\mu_1/(p\mu_{-1}\mu_2)]^{-1}$ for a, where $\mu_i = E(R^i)$ for $i = -1, 1, 2$. As stated by Xu and Izmirlian [24], their bound of a is sometimes worse than the bound obtained by Brandwein and Strawderman [7]. A question of theoretical interest is raised: Is this possible that bounds of a obtained by Brandwein and Strawderman [7] and Xu and Izmirlian [24] can be improved under a weaker condition than Brandwein and Strawderman's [7] technical condition (ii)? In this paper we provide an affirmative answer to this question. Specifically, we use the fact that the average of $-h$ over the sphere is nonincreasing in the radius to show dominance of $\delta_{a,\mathbf{g}}(\mathbf{X})$ over $\mathbf{X}$ and obtain a new bound $0 < a \leq \mu_1/(p\mu_{-1})$ for a, which is always better than $1/(p\mu_{-2})$ and $[\mu_1/(p^2\mu_{-1})][1 - (p-1)\mu_1/(p\mu_{-1}\mu_2)]^{-1}$.

The paper is organized as follows: In Sect. 13.2 we present the main result that states the dominance conditions of the estimators $\delta_{a,\mathbf{g}}(\mathbf{X})$ with respect to the quadratic loss. To illustrate the construction of the function h and the performance of the new

bound, three examples are also studied in Sect. 13.2. In Sect. 13.3 we extend the main result in Sect. 13.2 to other loss functions that are nondecreasing concave functions of quadratic loss. The estimators of the location vector when the scale is unknown and the observation $(\mathbf{X}^T, \mathbf{Y}^T)^T$ contains a residual vector $\mathbf{Y}$ are also considered in Sect. 13.3. Section 13.4 is devoted to some concluding remarks, while the last section consists of proofs of results in Sects. 13.2 and 13.3.

13.2 Main Results

Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^T$ be an estimator of $\boldsymbol{\theta}$ and let $R(\boldsymbol{\delta}, \boldsymbol{\theta}) = E[L(\boldsymbol{\delta}, \boldsymbol{\theta})]$ be the risk of $\boldsymbol{\delta}$, where the loss function $L(\boldsymbol{\delta}, \boldsymbol{\theta})$ is defined by

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = \|\boldsymbol{\delta} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^{p} (\delta_i - \theta_i)^2. \tag{13.3}$$

That is, the loss function $L(\boldsymbol{\delta}, \boldsymbol{\theta})$ we consider in this section is quadratic. Furthermore, we employ the following notation introduced by Xu and Izmirlian [24]:

$$m(t) = -E_{\mathbf{U}}[h(t\mathbf{U} + \boldsymbol{\theta})],$$
$$M_*(t) = M(t) - M(0) = \int_0^t m(z)dz \tag{13.4}$$

for $t \geq 0$, where $-h$ is a nonnegative and superharmonic function and the random vector $\mathbf{U}$ has a uniform distribution on the surface of the unit sphere. Note that $m(t)$ is a nonincreasing function of t and $M_*(t)$ is a nonnegative and nondecreasing concave function of t (see Du Plessis [[11], p. 54]).

Theorem 13.1 *Suppose that $\mathbf{X} \sim \mathrm{SS}_p(\boldsymbol{\theta}, I_p)$ (spherically symmetric about mean vector $\boldsymbol{\theta}$) and $\boldsymbol{\delta}_{a, \mathbf{g}}(\mathbf{X})$ is defined by (13.2). Then under quadratic loss (13.3), $\boldsymbol{\delta}_{a, \mathbf{g}}(\mathbf{X})$ has a smaller risk than $\boldsymbol{\delta}_{0, \mathbf{g}}(\mathbf{X}) = \mathbf{X}$ if*
  (i) $\|\mathbf{g}\|^2/2 \leq -h \leq -\nabla \circ \mathbf{g}$, where $-h$ is superharmonic,
  (ii) $r \int_0^1 m(rz)pz^{p-1}dz \geq c \int_0^1 M_*(rz)pz^{p-2}dz$ when $r > \sqrt{ap}$, where m and M_* are defined by (13.4) and $1 \leq c \leq p - 1$ is a constant, and
  (iii) $0 < a \leq \mu_1/(p\mu_{-1})$, where $\mu_{-i} = E(R^{-i})$ for $i = -1, 1$ and $R = \|\mathbf{X} - \boldsymbol{\theta}\|$.

Remark 13.1 The condition (ii) of Theorem 13.1 is slightly weaker than the condition (ii) of Brandwein and Strawderman [7]. To see this, we use integration by parts to obtain that

$$r \int_0^1 m(rz)pz^{p-1}dz = p\left(M_*(r) - \int_0^1 M_*(rz)(p-1)z^{p-2}dz\right).$$

Thus, the condition (ii) above is equivalent to $N(r) \geq 0$ when $r > \sqrt{ap}$, where $N(r)$ is defined by

$$N(r) = M_*(r) - (p - 1 - c) \int_0^1 M_*(rz)z^{p-2}dz.$$

Taking the derivative of $N(r)$ gives that

$$N'(r) = m(r) - (p - 1 - c) \int_0^1 m(rz)z^{p-1}dz. \tag{13.5}$$

Since the condition (ii) of Brandwein and Strawderman [7] is equivalent to

$$\int_0^1 m(rz)z^{p-1}dz \leq \frac{1}{p-2}m(r), \quad r > 0. \tag{13.6}$$

Applying (13.6) to (13.5) will yield that

$$N'(r) \geq m(r) - \frac{p-1-c}{p-2}m(r) = \frac{c-1}{p-2}m(r) \geq 0$$

because $c \geq 1$. This shows that $N(r)$ is a nondecreasing function of r. Using the fact that $\lim_{r \to 0^+} N(r) = 0$, we can conclude that $N(r) \geq 0$ when $r > 0$.

It is also worth mentioning that we only require the condition (ii) to be true when $r > \sqrt{ap}$. When $r \leq \sqrt{ap}$, there is no any assumption.

Remark 13.2 Let F denote the distribution function (df) of $R = \|\mathbf{X} - \boldsymbol{\theta}\|$. Then applying Lemma 13.1 in Sect. 13.5 with $f_1(r) = r$, $g_1(r) = 1/r^2$, $f_2(r) = g_2(r) = 1$ and $d\alpha = dF$ yields that

$$\mu_{-1} = E\left(\frac{1}{R}\right) = E\left(R\frac{1}{R^2}\right) \leq E(R)E\left(\frac{1}{R^2}\right) = \mu_1\mu_{-2}.$$

Using this fact, we can conclude that the new bound for a is better than that of Brandwein and Strawderman [7] because

$$\frac{1}{p\mu_{-2}} \leq \frac{\mu_1}{p\mu_{-1}}.$$

Remark 13.3 The new bound for a is also better than that of Xu and Izmirlian [24]. This can be seen from a direct comparison with the fact that $\mu_1 \leq \mu_{-1}\mu_2$, which follows from an application of Lemma 13.1 in Sect. 13.5 with $f_1(r) = 1/r$, $g_1(r) = r^2$, $f_2(r) = g_2(r) = 1$ and $d\alpha = dF$.

Remark 13.4 It needs to be mentioned that the requirement of dimensionality such as $p \geq 4$ usually arises in the condition (i) of Theorem 13.1. Meanwhile, although the

function h used in Theorem 13.1 has many choices, we usually take $h(\mathbf{X}) = \nabla \circ \mathbf{g}(\mathbf{X})$ when $\nabla \circ \mathbf{g}(\mathbf{X})$ is a subharmonic function.

Example 13.1 Consider the James-Stein [15] estimator which is given by

$$\delta_{a,0}(\mathbf{X}) = \left(1 - \frac{a}{\|\mathbf{X}\|^2}\right)\mathbf{X}$$

and discussed by many authors including Brandwein and Strawderman [7] and Fan and Fang [13]. Clearly, taking $\mathbf{g}(\mathbf{X}) = -\mathbf{X}/\|\mathbf{X}\|^2$ in (13.2) will see that $\delta_{a,0}(\mathbf{X})$ is a special case of estimators (13.2). Let

$$h(\mathbf{X}) = \nabla \circ \mathbf{g}(\mathbf{X}) = -\frac{p-2}{\|\mathbf{X}\|^2}.$$

Then $-h$ is superharmonic if $p \geq 4$ because

$$-\sum_{i=1}^{p} \frac{\partial^2 h}{\partial x_i^2} = -\frac{(p-2)(p-4)}{\|\mathbf{X}\|^2} \leq 0.$$

The condition (i) in Theorem 13.1 is clearly satisfied. Meanwhile, condition (ii) in Theorem 13.1 is also true because Brandwein and Strawderman's [7] technical condition (ii) is true, see Lemma 2.2 of Fan and Fang [13].

To illustrate the performance of the new bound of a, we consider two examples below. We use a_{new} to denote the new bound $\mu_1/(p\mu_{-1})$ of a. We denote by $a_{\text{bs}} = 1/(p\mu_{-2})$, the bound of a in Brandwein and Strawderman's [7] Theorem 2.1, and $a_{\text{xi}} = [\mu_1/(p^2\mu_{-1})][1 - (p-1)\mu_1/(p\mu_{-1}\mu_2)]^{-1}$, the bound of a in Xu and Izmirlian's [24] Theorem 1.

Example 13.2 Let $\mathbf{X}$ have a normal distribution with mean $\boldsymbol{\theta}$ and covariance matrix I_p. Then $R^2 = \|\mathbf{X} - \boldsymbol{\theta}\|^2$ has a χ_p^2-distribution, which implies that $\mu_{-2} = 1/(p-2)$, $\mu_{-1} = \Gamma((p-1)/2)/[\sqrt{2}\Gamma(p/2)]$, and $\mu_1 = \sqrt{2}\Gamma((p+1)/2)/\Gamma(p/2)$. Table 13.1 below provides the values of three bounds of a for different p.

One can see from Table 13.1 that the new bound of a is the best, especially, it is much better than other two bounds when the dimensionality is small.

Table 13.1 Bounds of a

p	4	5	6	7	8	9	10	15	20	30	40	50	75	100
a_{new}	0.750	0.800	0.833	0.857	0.875	0.889	0.900	0.933	0.950	0.967	0.975	0.980	0.987	0.990
a_{bs}	0.500	0.600	0.667	0.714	0.750	0.778	0.800	0.867	0.900	0.933	0.950	0.960	0.973	0.980
a_{xi}	0.429	0.444	0.455	0.462	0.467	0.471	0.474	0.483	0.487	0.492	0.494	0.495	0.497	0.497

Table 13.2 Bounds of a

p	4	5	6	7	8	9	10	15	20	30	40	50	75	100
a_{new}	0.150	0.133	0.119	0.107	0.097	0.089	0.082	0.058	0.045	0.031	0.024	0.019	0.013	0.010
a_{bs}	0.125	0.120	0.111	0.102	0.094	0.086	0.080	0.058	0.045	0.031	0.024	0.019	0.013	0.010
a_{xi}	0.115	0.105	0.096	0.088	0.081	0.075	0.070	0.052	0.041	0.029	0.023	0.019	0.013	0.010

Example 13.3 Let $\mathbf{X}$ have a uniform distribution in the unit sphere centered at $\boldsymbol{\theta}$. Then $R = \|\mathbf{X} - \boldsymbol{\theta}\|$ has a probability density function (pdf) pr^{p-1}, $0 \leq r \leq 1$ and $\mu_i = p/(p+i)$, $i = -2, -1, 1, 2$. Thus, $a_{\text{bs}} = (p-2)/p^2$, $a_{\text{xi}} = (p-1)/(p^2 + 3p - 2)$, and $a_{\text{new}} = (p-1)/[p(p+1)]$. Table 13.2 below provides the values of three bounds of a for different p.

One can see from Table 13.2 that the new bound of a is the best. Meanwhile, all three bounds will approach to zero when the dimensionality increases.

13.3 Extensions to Other Loss Functions and the Unknown Scale Case

Similar to Xu and Izmirlian [24], we consider two extensions in this section. The first one is to show that Theorem 13.1 in Sect. 13.2 can be generalized to a larger class of loss functions, while the second one is to estimate the location vector with an unknown scale parameter.

The loss function used in the first extension is

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = W\left(\|\boldsymbol{\delta} - \boldsymbol{\theta}\|^2\right), \tag{13.7}$$

where W is a nonnegative and nondecreasing concave function. The loss function (13.7) has been studied for the spherically symmetric distributions by many investigators including Bock [1], Brandwein and Strawderman [4, 6, 7], Fan and Fang [12–14], Xu [23], and Xu and Izmirlian [24].

Theorem 13.2 *Let F be the df of $R = \|\mathbf{X} - \boldsymbol{\theta}\|$ satisfying*

$$0 < \int_0^\infty W'\left(r^2\right) dF(r) < \infty,$$

where W' is the derivative of W. Suppose that $\mathbf{X}$ is spherically symmetric about $\boldsymbol{\theta}$ and $\boldsymbol{\delta}_{a,\mathbf{g}}(\mathbf{X})$ is defined by (13.2). Then under loss function (13.7), $\boldsymbol{\delta}_{a,\mathbf{g}}(\mathbf{X})$ has a smaller risk than $\mathbf{X}$ if the conditions (i) and (ii) of Theorem 13.1 hold and

(iii) $0 < a < v_1/(pv_{-1})$, *where* $v_i = E_G(R^i)$ *for* $i = -1, 1$ *and G is a weighted df of F with the weight function* $W'\left(r^2\right)$ *defined by*

$$G(t) = \left(\int_0^\infty W'\left(r^2\right) dF(r)\right)^{-1} \int_0^t W'\left(r^2\right) dF(r), \quad t \geq 0.$$

Now we investigate the problem of estimating the location vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ when the observation $(\mathbf{X}^T, \mathbf{Y}^T)^T$ contains an $m \times 1$ residual vector $\mathbf{Y}$ such that $\mathbf{X}_*^T = (1/\sigma)(\mathbf{X}^T, \mathbf{Y}^T)$ follows a spherically symmetric distribution $\text{SS}_{p+m}(\boldsymbol{\theta}_*, \sigma^2 I_{p+m})$, where $\boldsymbol{\theta}_*^T = (\boldsymbol{\theta}^T, \mathbf{0}_m^T)$, $\mathbf{0}_m$ is an $m \times 1$ vector in which all elements are zero, and σ is an unknown scale. The improved estimators we consider is given by

$$\boldsymbol{\delta}_{a,\mathbf{g}}^*(\mathbf{X}_*) = \mathbf{X} + a\mathbf{Y}^T\mathbf{Y}\mathbf{g}(\mathbf{X}). \tag{13.8}$$

Theorem 13.3 *Suppose that* $\mathbf{X}$ *is a* $p \times 1$ *random vector and* $\mathbf{Y}$ *is an* $m \times 1$ *random vector such that* $\mathbf{X}_* = (1/\sigma)\left(\mathbf{X}^T, \mathbf{Y}^T\right)^T \sim \text{SS}_{p+m}(\boldsymbol{\theta}^*, \sigma^2 I_{p+m})$. *Let* $\boldsymbol{\delta}_{a,\mathbf{g}}^*(\mathbf{X}_*)$ *be defined by (13.8). Then under the scaled quadratic loss function*

$$L(\boldsymbol{\delta}, \boldsymbol{\theta}) = \|\boldsymbol{\delta} - \boldsymbol{\theta}\|^2/\sigma^2,$$

$\boldsymbol{\delta}_{a,\mathbf{g}}^*(\mathbf{X}_*)$ *dominates* $\mathbf{X}$ *if conditions (i) and (ii) of Theorem 13.1 hold and*

(iii) $0 < a < (p-1)/[p(m+2)]$.

The bound of a in Theorem 13.3 doesn't depend on the distribution of $\mathbf{X}_*$. Cellier, Fourdrinier and Robert [10] first observed this type of robustness phenomenon for the James-Stein estimator.

13.4 Discussion

If $-h$ is superharmonic, Brandwein and Strawderman [7] used the fact that its average over the ball ("volume") is greater than its average over the sphere ("surface area") to show the dominance of the estimators of the form $\boldsymbol{\delta}_{a,\mathbf{g}}(\mathbf{X})$ over the estimator $\mathbf{X}$. In this paper we use the fact that the average of $-h$ over the sphere is a nonincreasing function of the radius of the sphere. The new approach allows us not only to weaken their technical condition (ii), but also to obtain a better bound for a. The new bound of a is also better than those of Brandwein and Strawderman [7] and Xu and Izmirlian [24]. In addition, we consider two extensions. The first is to extend the quadratic loss (13.3) to the loss function (13.7), while the second is to study the estimators of the location vector when the observation $(\mathbf{X}^T, \mathbf{Y}^T)^T$ contains a residual vector $\mathbf{Y}$ and the scale is unknown. While the bounds of a given by the theorems in Sects. 13.2 and 13.3 are better than those of Brandwein and Straderman [7] and Xu and Izmirlian [24], they are not necessarily optimal and should be considered a guide post. Clearly, one may

be able to obtain better bounds than those given here if the distribution of R is known. Stein [20], for example, used integration by parts to obtain $0 < a \le 1 (= \mu_2/p)$ under normality. Thus, it would be interesting to see if our new bound $0 < a \le \mu_1/(p\mu_{-1})$ can be further improved to $0 < a \le \mu_2/p$ for $\mathbf{X} \sim \mathrm{SS}_p(\boldsymbol{\theta}, I_p)$. As a final point, it would be interesting, but perhaps very difficult, to study the dominance conditions of the estimator $\boldsymbol{\delta}_{a,\mathbf{g}}(\mathbf{X}) = \mathbf{X} + a\mathbf{g}(\mathbf{X})$ over $\mathbf{X}$ for other distributions. As mentioned by Xu and Izmirlian [24], results of estimator (13.1) obtained by Shinozaki [18] for the class of distributions with independently and identically distributed components and by Xu [23] for the sign-invariant distribution are very limited.

13.5   Proofs

In this section we use $f_{c,s}(z)$ to denote the pdf of the Beta distribution $\mathrm{Beta}(c, s)$ given by

$$f_{c,s}(z) = \frac{\Gamma(c+s)}{\Gamma(c)\Gamma(s)} z^{c-1}(1-z)^{s-1}, \quad 0 < z < 1,$$

where $c > 0$ and $s > 0$ are parameters. To shorten the proofs of results in Sects. 13.2 and 13.3, we need the following lemmas in which the first one is taken from Wijsman's [22] Theorem 2.

Lemma 13.1 *Let α be a measure on the real line $\mathbb{R}$ and let f_j, g_j $(j = 1, 2)$ be Borel-measurable functions: $\mathbb{R} \to \mathbb{R}$ such that $f_2 \ge 0$, $g_2 \ge 0$, and $\int |f_i g_j| d\alpha < \infty$ $(i, j = 1, 2)$. If f_1/f_2 and g_1/g_2 are monotonic in the same direction, then*

$$\int f_1 g_1 d\alpha \int f_2 g_2 d\alpha \ge \int f_1 g_2 d\alpha \int f_2 g_1 d\alpha, \qquad (13.9)$$

whereas if f_1/f_2 and g_1/g_2 are monotonic in the opposite directions, then inequality in (13.9) is reversed. The equality in (13.9) holds if and only if $f_2 = 0$ or $g_2 = 0$ or $f_1/f_2 = $ constant or $g_1/g_2 = $ constant almost everywhere with respect to the measure ρ defined by $d\rho = (|f_1| + |f_2|)(|g_1| + |g_2|)d\alpha$.

Lemma 13.2 *Let the function M_* be defined by (13.4). Then*

$$\int_0^1 M_*(rz) f_{c-1,1}(z) dz \le M_*(r) \le \int_0^1 M_*(rz) c z^{c-2} dz,$$

for any $r > 0$, where $c > 1$ is a constant.

Proof Since M_* is a nondecreasing concave function with $M_*(0) = 0$ and the expected value of $\mathrm{Beta}(c - 1, 1)$ distribution is $(c - 1)/c$, using the Jensen's inequality will yield that

$$\int_0^1 M_*(rz) f_{c-1,1}(z) dz \le M_*\left(r\frac{c-1}{c}\right) \le M_*(r).$$

Furthermore, the concavity of M_* implies that $M_*(rz) \ge zM_*(r)$ for $z \in [0, 1]$ and $r > 0$. Thus,

$$\int_0^1 M_*(rz) cz^{c-2} dz \ge \int_0^1 M_*(r) cz^{c-1} dz = M_*(r) \int_0^1 cz^{c-1} dz = M_*(r).$$

Lemma 13.3 *For $z \in [0, 1]$, let*

$$\ell(z) = \frac{\beta(r)}{p} f_{p,1}(z) + 1 - \frac{\beta(r)}{p},$$

where $\beta(r) = r^2/a$ is considered a parameter. Then $\ell(z)$ is a pdf on $[0, 1]$ when $\beta(r) \le p$. Furthermore, when $\beta(r) \le p$, we have

$$m(r) - \beta(r) \int_0^1 m(rz) z^{p-1} dz \le \left(1 - \frac{\beta(r)}{p}\right) \frac{M_*(r)}{r}$$

for $r > 0$, where m and M_* are defined by (13.4).

Proof When $\beta(r) \le p$, $\ell(z)$ is a pdf on $[0, 1]$ because it is a convex combination of pdfs $f_{p,1}(z)$ and $f_{1,1}(z) = 1$ on $[0, 1]$. Furthermore, since m is a nonincreasing function, we have

$$m(r) \le \int_0^1 m(rz)\ell(z) dz,$$

which leads to

$$
\begin{aligned}
m(r) - \beta(r) \int_0^1 m(rz) z^{p-1} dz &\le \int_0^1 m(rz)\ell(z) dz - \frac{\beta(r)}{p} \int_0^1 m(rz) f_{p,1}(z) dz \\
&= \int_0^1 m(rz) \left(\ell(z) - \frac{\beta(r)}{p} f_{p,1}(z)\right) dz \\
&= \left(1 - \frac{\beta(r)}{p}\right) \int_0^1 m(rz) dz \\
&= \left(1 - \frac{\beta(r)}{p}\right) \frac{M_*(r)}{r}.
\end{aligned}
$$

Lemma 13.4 *When $\beta(r) = r^2/a > p$, we have*

$$m(r) - \beta(r) \int_0^1 m(rz) z^{p-1} dz \le \left(1 - \frac{\beta(r)}{p}\right) \frac{c}{r} \int_0^1 M_*(rz) pz^{p-2} dz$$

for $r > 0$, where m and M_* are defined by (13.4) and $c \in [1, p-1]$ is a constant.

Proof Since m is a nonincreasing function, we have

$$m(r) \le \int_0^1 m(rz) f_{p,1}(z) dz,$$

which leads to

$$
\begin{aligned}
m(r) - \beta(r) \int_0^1 m(rz) z^{p-1} dz &\le \left(1 - \frac{\beta(r)}{p}\right) \int_0^1 m(rz) f_{p,1}(z) dz \\
&\le \left(1 - \frac{\beta(r)}{p}\right) \frac{c}{r} \int_0^1 M_*(rz) p z^{p-2} dz.
\end{aligned}
\tag{13.10}
$$

Here the last inequality in (13.10) follows from the condition (ii) of Theorem 13.1 and $\beta(r) > p$.

Remark 13.5 Lemmas 13.3 and 13.4 can be combined below:

$$m(r) - \beta(r) \int_0^1 m(rz) z^{p-1} dz \le N_1(r) N_2(r), \tag{13.11}$$

where $\beta(r) = r^2/a$ and

$$
\begin{aligned}
N_1(r) &= \left(1 - \frac{\beta(r)}{p}\right) \frac{1}{r}, \\
N_2(r) &= I[\beta(r) \le p] M_*(r) + c I[\beta(r) > p] \int_0^1 M_*(rz) p z^{p-2} dz.
\end{aligned}
\tag{13.12}
$$

Here $I[A]$ denotes the indicator function of the event A.

Lemma 13.5 *For $r > 0$, let $N_1(r)$ and $N_2(r)$ be defined by (13.12). Then $N_1(r)$ is strictly decreasing in r and $N_2(r)$ is nondecreasing in r. Furthermore, $E_R[N_1(R) N_2(R)] \le 0$ if $a \le \mu_1/(p\mu_{-1})$.*

Proof Since $N_1(r) = 1/r - r/(ap)$, it is a strictly decreasing function of r. Similarly, since both $M_*(r)$ and $\int_0^1 M_*(rz) p z^{p-2} dz$ are nondecreasing in r and $M_*(r) \le \int_0^1 M_*(rz) p z^{p-2} dz$ from Lemma 13.2, we can conclude that $N_2(r)$ is a nondecreasing function of r. Furthermore, applying Lemma 13.1 with $f_1(r) = N_1(r)$, $g_1(r) = N_2(r)$, $f_2(r) = g_2(r) = 1$, and a probability measure $d\alpha = dF$ will yield that

$$E_R[N_1(R) N_2(R)] \le E_R[N_1(R)] E_R[N_2(R)] = \left(\mu_{-1} - \frac{\mu_1}{ap}\right) E[N_2(R)] \le 0$$

if $a \le \mu_1/(p\mu_{-1})$ because $N_2(r) \ge 0$ for $r > 0$.

Proof of Theorem 13.1. When $\mathbf{X} \sim \mathrm{SS}_p(\boldsymbol{\theta}, I_p)$, we have $\mathbf{X} - \boldsymbol{\theta} = \mathbf{Z} \stackrel{d}{=} R\mathbf{U}$, where R and $\mathbf{U}$ are independent, $R \stackrel{d}{=} \|\mathbf{Z}\|$, and $\mathbf{U}$ has a uniform distribution on the surface of the unit sphere. Using the argument of Xu and Izmirlian [24] with a verbatim copy of their (12), we obtain that the difference between the risks of two estimators $\delta_{a,\mathbf{g}}(\mathbf{X})$ and $\mathbf{X}$ is given by

$$
\begin{aligned}
D_1 &= R\big(\delta_{a,\mathbf{g}}(\mathbf{X}), \boldsymbol{\theta}\big) - R\big(\mathbf{X}, \boldsymbol{\theta}\big) \\
&= a^2 E\left[\|\mathbf{g}(\mathbf{Z}+\boldsymbol{\theta})\|^2\right] + 2aE\left[\mathbf{Z}^T\mathbf{g}(\mathbf{Z}+\boldsymbol{\theta})\right] \\
&= a^2 E\left[\|\mathbf{g}(\mathbf{Z}+\boldsymbol{\theta})\|^2\right] + 2ap^{-1}E\left[R^2 \nabla \circ \mathbf{g}(R\mathbf{V}+\boldsymbol{\theta})\right] \\
&\leq 2a^2 E\left[-h(R\mathbf{U}+\boldsymbol{\theta})\right] + 2ap^{-1}E\left[R^2 h(R\mathbf{V}+\boldsymbol{\theta})\right] \\
&= 2a^2 E_R\left[E_{\mathbf{U}}\left(-h(R\mathbf{U}+\boldsymbol{\theta})\big|R\right) + (ap)^{-1}R^2 E_{\mathbf{V}}\left(h(R\mathbf{V}+\boldsymbol{\theta})\big|R\right)\right] \\
&= 2a^2 E_R\left[m(R) - (ap)^{-1}R^2 E_{\mathbf{V}}\left(h(R\mathbf{V}+\boldsymbol{\theta})\big|R\right)\right] \\
&= 2a^2 E_R\left[m(R) - \beta(R)\int_0^1 m(Rv)v^{p-1}dv\right] \\
&\leq 2a^2 E_R\left[N_1(R)N_2(R)\right] \\
&\leq 0
\end{aligned}
\tag{13.13}
$$

if $a \leq \mu_1/(p\mu_{-1})$. Here the first inequality in (13.13) is based on the condition (i); the fifth equality in (13.13) is from the definition of function m; the last equality in (13.13) follows from the definition of m and the fact that $\mathbf{V} \stackrel{d}{=} V\mathbf{U}$, where the random variable $V \sim \mathrm{Beta}(p, 1)$ and $\mathbf{U}$ having a uniform distribution on the surface of the unit sphere are independent; the second-to-last inequality in (13.13) is based on Lemmas 13.3 and 13.4 or (13.11); the last inequality in (13.13) follows from Lemma 13.5. This completes the proof.

Proof of Theorem 13.2. Using the same approach as in Brandwein and Strawderman [4, 6] or Xu and Izmirlian [24], we obtain that the difference between the risks of two estimators $\delta_{a,\mathbf{g}}(\mathbf{X})$ and $\mathbf{X}$ is given by

$$
\begin{aligned}
D_2 &= R\big(\delta_{a,\mathbf{g}}(\mathbf{X}), \boldsymbol{\theta}\big) - R(\mathbf{X}, \boldsymbol{\theta}) \\
&= E\left[W\left(R^2 + \Delta_a(\mathbf{X})\right)\right] - E\left[W\left(R^2\right)\right],
\end{aligned}
\tag{13.14}
$$

where

$$
\Delta_a(\mathbf{X}) = \|\delta_{a,\mathbf{g}}(\mathbf{X}) - \boldsymbol{\theta}\|^2 - \|\mathbf{X} - \boldsymbol{\theta}\|^2.
$$

Since W is a nondecreasing concave function,

$$
W\left(R^2 + \Delta_a(\mathbf{X})\right) < W\left(R^2\right) + W'\left(R^2\right)\Delta_a(\mathbf{X}).
$$

Then we can conclude from (13.14) that

$$
\begin{aligned}
D_2 &\le E_{\mathbf{X}} \left[W'\left(R^2 \right) \Delta_a(\mathbf{X}) \right] \\
&\le E_R \left\{ W'\left(R^2 \right) E_{\mathbf{U}}[\Delta_a(R\mathbf{U} + \boldsymbol{\theta})|R] \right\} \\
&\le 2a^2 E_R \left[W'(R^2) N_1(R) N_2(R) \right] \\
&= 2a^2 E_{R_*} \left[N_1(R_*) N_2(R_*) \right] E_R \left[W'\left(R^2 \right) \right],
\end{aligned}
$$

where the df G of the random variable R_* is defined by

$$
G(t) = \left(\int_0^\infty W'\left(r^2 \right) dF(r) \right)^{-1} \int_0^t W'\left(r^2 \right) dF(r), \quad t \ge 0,
$$

which is a weighted df of F with the weight function $W'\left(r^2 \right)$. The result follows immediately from the assumption that $0 < E_R \left[W'\left(R^2 \right) \right] < \infty$ and the proof of Theorem 13.1 except for a change from the df F to the df G.

Proof of Theorem 13.3. Like Brandwein and Strawderman [7] and Xu and Izmirlian [24], the difference D_3 between the risks of two estimators $\delta_{a,\mathbf{g}}^*(\mathbf{X}_*)$ and $\mathbf{X}$ is equal to

$$
\begin{aligned}
D_3 &= R\left(\delta_{a,\mathbf{g}}^*(\mathbf{X}_*), \boldsymbol{\theta} \right) - R\left(\mathbf{X}, \boldsymbol{\theta} \right) \\
&= \frac{1}{\sigma^2} E \left[a^2 (\mathbf{Y}^T \mathbf{Y})^2 \|\mathbf{g}(\mathbf{Z} + \boldsymbol{\theta})\|^2 + 2a\, \mathbf{Y}^T \mathbf{Y} \mathbf{Z}^T \mathbf{g}(\mathbf{Z} + \boldsymbol{\theta}) \right] \qquad (13.15) \\
&= \frac{1}{\sigma^2} E \left(a^2 D_{31} + 2a D_{32} \right),
\end{aligned}
$$

where $\mathbf{Z} = \mathbf{X} - \boldsymbol{\theta} \stackrel{d}{=} R\mathbf{U}$, and

$$
\begin{aligned}
D_{31} &= E \left[(\mathbf{Y}^T \mathbf{Y})^2 \|\mathbf{g}(\mathbf{Z} + \boldsymbol{\theta})\|^2 \big| \|\mathbf{Z}\| = R, \|\mathbf{Y}\| = S \right], \\
D_{32} &= E \left[\mathbf{Y}^T \mathbf{Y} \mathbf{Z}^T \mathbf{g}(\mathbf{Z} + \boldsymbol{\theta}) \big| \|\mathbf{Z}\| = R, \|\mathbf{Y}\| = S \right].
\end{aligned}
$$

Using the divergence theorem and condition (i), we obtain that

$$
\begin{aligned}
D_{32} &= E \left[\mathbf{Y}^T \mathbf{Y} \mathbf{Z}^T \mathbf{g}(\mathbf{Z} + \boldsymbol{\theta}) \big| \|\mathbf{Z}\| = R, \|\mathbf{Y}\| = S \right] \\
&= S^2 R E_{\mathbf{U}} \left[\mathbf{U}^T \mathbf{g}(R\mathbf{U} + \boldsymbol{\theta}) \big| \|\mathbf{Z}\| = R, \|\mathbf{Y}\| = S \right] \\
&= \frac{S^2 R^2}{p} E_{\mathbf{V}} \left(\nabla \circ \mathbf{g}(R\mathbf{V} + \boldsymbol{\theta}) \big| \|\mathbf{Z}\| = R, \|\mathbf{Y}\| = S \right) \qquad (13.16) \\
&\le -\frac{S^2 R^2}{p} \int_0^1 m(Rz) f_{p,1}(z) dz,
\end{aligned}
$$

where m is defined by (13.4). Similarly, using the condition (i) will yield that

$$
\begin{aligned}
D_{31} &= E\left[(\mathbf{Y}^T\mathbf{Y})^2\|\mathbf{g}(\mathbf{Z}+\boldsymbol{\theta})\|^2\big|\|\mathbf{Z}\|=R,\|\mathbf{Y}\|=S\right] \\
&\le -2S^4 E\left[h(\mathbf{Z}+\boldsymbol{\theta})\big|\|\mathbf{Z}\|=R,\|\mathbf{Y}\|=S\right] \\
&= -2S^4 E\left[h(R\mathbf{U}+\boldsymbol{\theta})\big|\|\mathbf{Z}\|=R,\|\mathbf{Y}\|=S\right] \\
&= 2S^4 m(R).
\end{aligned}
\tag{13.17}
$$

Combining (13.16), (13.17) with (13.15) and using the same argument as the proof of theorem 13.1 will obtain the following inequality

$$
\begin{aligned}
D_3 &\le \frac{2a}{\sigma^2}E\left(aS^4 m(R) - \frac{S^2 R^2}{p}\int_0^1 m(Rz)\,f_{p,1}(z)\,dz\right) \\
&= \frac{2a}{\sigma^2}E\left[(aS^4)\left(m(R) - \frac{R^2}{aS^2}\int_0^1 m(Rz)z^{p-1}dz\right)\right] \\
&\le \frac{2a}{\sigma^2}E\left[(aS^4)\left(1-\frac{R^2}{apS^2}\right)\frac{1}{R}N_2(R)\right],
\end{aligned}
\tag{13.18}
$$

where the first inequality in (13.18) is based on (13.16) and (13.17), the last inequality of (13.18) follows from (13.11) after replacing a by aS^2, and $N_2(R)$ is defined by (13.12). Let $T^2 = R^2 + S^2$. Then T^2 and $B = R^2/T^2 \sim \text{Beta}(p/2, m/2)$ are independent. Let $C(c,s) = \Gamma(c+s)/[\Gamma(c)\Gamma(s)]$ for $c > 0, s > 0$ and let $C^* = C(p/2, m/2)/C(p/2, (m+2)/2)$. Write $\lambda = a + 1/p$. Then we can see from (13.18) that

$$
\begin{aligned}
\frac{\sigma^2}{2a}D_3 &\le E\left[aS^4\left(1-\frac{R^2}{apS^2}\right)\frac{1}{R}N_2(R)\right] \\
&= E\left[(1-B)(a-\lambda B)T^4\left(N_1\left(TB^{1/2}\right)+\frac{N\left(TB^{1/2}\right)}{TB^{1/2}}\right)\right] \\
&= C^*E\left[(a-\lambda B)T^4\left(N_1\left(TB^{1/2}\right)+\frac{N\left(TB^{1/2}\right)}{TB^{1/2}}\right)\right] \\
&= C^*E\left[(a-\lambda B)T^4 N_1\left(TB^{1/2}\right)\right] \\
&\quad + C^*E\left[\left(aB^{-1/2}-\lambda B^{1/2}\right)T^3 N(TB^{1/2})\right] \\
&\le C^*\left(a-\lambda\frac{p}{p+m+2}\right)E\left[T^4 N_1\left(TB^{1/2}\right)\right] \\
&\quad + C^*\left(a\frac{C(p/2,(m+2)/2)}{C((p-1)/2,(m+2)/2)} - \lambda\frac{C(p/2,(m+2)/2)}{C((p+1)/2,(m+2)/2)}\right) \\
&\quad \times E\left[T^3 N\left(TB^{1/2}\right)\right] \\
&\le 0
\end{aligned}
\tag{13.19}
$$

if

$$a - \lambda \frac{p}{p + m + 2} \leq 0,$$

$$a \frac{C\left(p/2, (m+2)/2\right)}{C\left((p-1)/2, (m+2)/2\right)} - \lambda \frac{C\left(p/2, (m+2)/2\right)}{C\left((p+1)/2, (m+2)/2\right)} \leq 0. \tag{13.20}$$

Here the second-to-last inequality of (13.19) follows from applications of Lemma 13.1 with the measure $d\alpha = f_{p/2,(m+2)/2}(b)db$ on $[0, 1]$ and $f_1(b) = a - \lambda b$, $g_1(b) = T^4 N_1(Tb^{1/2})$, $f_2(b) = g_2(b) = 1$, and $f_1(b) = ab^{-1/2} - \lambda b^{1/2}$, $g_1(b) = T^3 N(Tb^{1/2})$, $f_2(b) = g_2(b) = 1$, respectively. Simple algebra shows that the first inequality in (13.20) is equivalent to $0 < a \leq 1/(m+2)$, while the second inequality in (13.20) is equivalent to $0 < a \leq (p-1)/[p(m+2)]$. Therefore, $D_3 \leq 0$ if $0 < a \leq (p-1)/[p(m+2)]$.

References

1. Bock, M.E.: Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions. J. Multivariate Anal. **17**, 127–147 (1985)
2. Brandwein, A.C.: Minimax estimation of mean of spherically symmetric distributions under general quadratic loss. J. Multivariate Anal. **9**, 579–588 (1979)
3. Brandwein, A.C., Strawderman, W.E.: Minimax estimation of location parameters for spherically symmetric unimodal distributions under quadratic loss. Ann. Statist. **6**, 377–416 (1978)
4. Brandwein, A.C., Strawderman, W.E.: Minimax estimation of location parameters for spherically symmetric distributions with concave loss. Ann. Statist. **8**, 279–284 (1980)
5. Brandwein, A.C., Strawderman, W.E.: Stein estimation, The spherically symmetric case. Statist. Sci. **5**, 356–369 (1990)
6. Brandwein, A.C., Strawderman, W.E.: Stein estimation for spherically symmetric distributions: recent developments. Statist. Sci. **27**, 11–23 (2012)
7. Brandwein, A.C., Strawderman, W.E.: Generalizations of James-Stein estimators under spherical symmetry. Ann. Statist. **19**, 1639–1650 (1991)
8. Brown, L.D.: On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Statist. **37**, 1087–1136 (1966)
9. Brown, L.D., Zhao, L.H.: A geometrical explanation of Stein shrinkage. Statist. Sci. **27**, 24–30 (2012)
10. Cellier, D., Fourdrinier, D., Robert, C.: Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. J. Multivariate Anal. **29**, 39–52 (1988)
11. Du Plessis, N.: An Introduction to Potential Theory. Hafner, Darien, CT (1970)
12. Fan, J., Fang, K.-T.: Inadmissibility of sample mean and sample regression coefficients for elliptically contoured distributions. In: Fang, K.-T., Anderson, T.W. (eds.) Statistical Inference in Elliptically Contoured and Related Distributions, pp. 275–290. Allerton Press Inc., New York (1990)
13. Fan, J., Fang, K.-T.: Inadmissibility of the usual estimator for the location parameters of spherically symmetric distributions. In: Fang, K.-T., Anderson, T.W. (eds.) Statistical Inference in Elliptically Contoured and Related Distributions, pp. 291–297. Allerton Press Inc., New York (1990)
14. Fan, J., Fang, K.-T.: Shrinkage estimators and ridge regression estimators for elliptically contoured distributions. In: Fang, K.-T., Anderson, T.W. (eds.) Statistical Inference in Elliptically Contoured and Related Distributions, pp. 313–326. Allerton Press Inc., New York (1990)

15. James, W., Stein, C.: Estimation with quadratic loss. Proc. Fourth Berkeley Sympos. Math. Statist. Prob. **1**, 361–379 (1961)
16. Maruyama, Y.: Admissible minimax estimators of a mean vector of scale mixtures of multivariate normal distributions. J. Multivariate Anal. **84**, 274–283 (2003)
17. Miceli, R.J., Strawderman, W.E.: Minimax estimation for certain independent component distributions under weighted squared error loss. Comm. Statist. Theory Methods **15**, 2191–2200 (1986)
18. Shinozaki, N.: Simultaneous estimation of location parameters under quadratic loss. Ann. Statist. **12**, 322–335 (1984)
19. Stein, C.: Inadmissibility of the usual estimator for the mean vector of a multivariate normal distribution. Proc. Third Berkeley Sympos. Math. Statist. Prob. **1**, 197–206 (1956)
20. Stein, C.: Estimation of the mean of a multivariate normal distribution. Ann. Statist. **9**, 1135–1151 (1981)
21. Tosh, C., Dasgupta, S.: Maximum likelihood estimation for mixtures of spherical Gaussians is NP-hard. J. Mach. Learn. Res. **18**, 1–11 (2018)
22. Wijsman, R.A.: A useful inequality on ratios of integrals, with application to maximum likelihood estimation. J. Amer. Statist. Assoc. **80**, 472–475 (1985)
23. Xu, J.-L.: Simultaneous estimation of location parameters for sign-invariant distributions. Ann. Statist. **25**, 2259–2272 (1997)
24. Xu, J.-L., Izmirlian, G.: Estimation of location parameters for spherically symmetric distributions. J. Multivariate Anal. **97**, 514–525 (2006)