Chapter 5
Estimation of a Mean Vector for Spherically Symmetric Distributions I: Known Scale

5.1 Introduction

In Chaps. 2 and 3 we studied estimators that improve over the "usual" estimator of the location vector for the case of a normal distribution. In this chapter, we extend the discussion to spherically symmetric distributions discussed in Chap. 4. Section 5.2 is devoted to a discussion of domination results for Baranchik type estimators while Sect. 5.3 examines more general estimators. Section 5.4 discusses Bayes minimax estimation. Finally, Sect. 5.5 discusses estimation with a concave loss.

We close this introductory section by extending the discussion of Sect. 2.2 on the empirical Bayes justification of the James-Stein estimator to the general multivariate (but not necessarily normal) case.

Suppose X has a p-variate distribution with density $f(\|x - \theta\|^2)$, unknown location vector θ and known scale matrix $\sigma^2 I_p$. The problem is to estimate θ under loss $L(\theta, \delta) = \|\delta - \theta\|^2$. Let the prior distribution on θ be given by $\pi(\theta) = f^{\star n}(\theta)$, the n-fold convolution of the density $f(\cdot)$ with itself. Note that the distribution of θ is the same as that of $\sum_{i=1}^{n} Y_i$ where the Y_i are iid with density $f(\cdot)$. Recall from Example 1.3 that the Bayes estimator of θ is given by

$$\delta_n(X) = \frac{n}{n + 1} X = \left(1 - \frac{1}{n + 1}\right) X.$$

Assume now that n is unknown. Since

$$E(X^{\mathrm{T}}X) = E\left(\sum_{i=0}^{n} Y_i^{\mathrm{T}} Y_i\right) = (n + 1)\, E(Y_0^{\mathrm{T}} Y_0) = (n + 1)\, (\mathrm{tr}\, \sigma^2 I) = (n + 1)\, p\, \sigma^2,$$

an unbiased estimator of $n + 1$ is $X^{\mathrm{T}} X / (p \sigma^2)$, and so $p \sigma^2 / (X^{\mathrm{T}} X)$ is a reasonable estimator of $1/(n+1)$. Substituting $p \sigma^2 / (X^{\mathrm{T}} X)$ for $1/(n+1)$ in the Bayes estimator, we have that

$$\delta^{EB}(X) = \left(1 - \frac{p \sigma^2}{X^{\mathrm{T}} X}\right) X$$

can be viewed as an empirical Bayes estimator of θ without any assumption on the form of the density (and in fact there is not even any need to assume there is a density). Hence this Stein-like estimator can be viewed as a reasonable alternative to X from an empirical Bayes perspective regardless of the form of the underlying distribution.

Note that Diaconis and Ylvisaker (1979) introduced the prior $f^{\star n}(\theta)$ as a reasonable conjugate prior for location families since it gives linear Bayes estimators. Strawderman (1992) gave the above empirical Bayes argument. In the normal case the sequence of priors corresponds to that in Sect. 2.2.3 with $\tau^2 = n \sigma^2$. The shrinkage factor $p \sigma^2$ in the present argument differs from $(p - 2) \sigma^2$ in the normal case since in this general case we use a "plug-in" estimator of $1/(n + 1)$ as opposed to the unbiased estimator (in the normal case) of $1/(\sigma^2 + \tau^2)$.

5.2   Baranchik-Type Estimators

In this section, assuming that X has a spherically symmetric distribution with mean vector θ and that loss is $L(\theta, \delta) = \|\delta - \theta\|^2$, we consider estimators of the Baranchik-type, as (2.19) in the normal setting, for different families of densities. In Sect. 5.3, we consider results for general estimators of the form $X + g(X)$.

5.2.1   Variance Mixtures of Normal Distributions

We first consider spherically symmetric densities which are variance mixtures of normal distributions. Suppose

$$f(\|x - \theta\|^2) = \frac{1}{(2\pi)^{p/2}} \int_0^\infty \frac{1}{v^{p/2}} \exp\left\{-\frac{\|x - \theta\|^2}{2v}\right\} dG(v), \qquad (5.1)$$

where $G(\cdot)$ is a probability distribution on $(0, \infty)$, i.e., a mixture of $\mathcal{N}_p(\theta, vI)$ distributions with mixing distribution $G(\cdot)$.

Our first result gives a domination result for Baranchik type estimators for such distributions. This result is analogous to Theorem 2.3 in the normal case.

Theorem 5.1 (Strawderman 1974b) *Let X have density of the form* (5.1) *and let*

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

where the function $r(\cdot)$ is absolutely continuous. Assume the expectations $E[V]$ and $E[V^{-1}]$ are finite where V has distribution G. Then $\delta_{a,r}^B(X)$ is minimax for the loss $L(\theta, \delta) = \|\delta - \theta\|^2$ provided

(1) $0 \le a \le 2(p-2)/E[V^{-1}]$,
(2) $0 \le r(t) \le 1$ *for any* $t \ge 0$,
(3) $r(t)$ *is nondecreasing in t, and*
(4) $r(t)/t$ *is nonincreasing in t.*

Furthermore, $\delta_{a,r}^B(X)$ dominates X provided the inequalities in (1) *or* (2) *(on a set of positive measure) are strict or $r'(t)$ is strictly increasing on a set of positive measure.*

Proof The proof proceeds by calculating the conditional risk given $V = v$, noting that the distribution of $X|V = v$ is normal $N(\theta, vI_p)$. First note that $E[V] < \infty$ is equivalent to $E_0[\|X\|^2] < \infty$ so that the risk of X is finite. Similarly, it can be seen that $E[V^{-1}] < \infty$ if and only if $E_0[\|X\|^{-2}] < \infty$. Then, thanks to (2), we have $E_0[r^2(\|X\|^2)\|X\|^{-2}] < \infty$. Actually, we will see below that, for any θ, $E_\theta[\|X\|^{-2}] \le E_0[\|X\|^{-2}]$, and hence, $E_\theta[r^2(\|X\|^2)\|X\|^{-2}] < \infty$ which guarantees that the risk of $\delta_{a,r}^B(X)$ is finite. Note that, conditionally on V, $\|X\|^2/V$ has a noncentral chi-square distribution with p degrees of freedom and noncentrality parameter $\|\theta\|^2/V$. Hence, since the family of noncentral chi-square distributions have monotone (increasing) likelihood ratios in the noncentrality parameter (and therefore are stochastically increasing), $\|X\|^2/V$ is (conditionally) stochastically decreasing in V and increasing in $\|\theta\|^2$.

Hence,

$$E_\theta\left[\frac{1}{\|X\|^2/V}\right] \le E_0\left[\frac{1}{\|X\|^2/V}\right]$$

and, as a result,

$$\begin{aligned}
E_\theta\left[\frac{1}{\|X\|^2}\right] &= E\left[E_\theta\left[\frac{1}{\|X\|^2}\bigg|V\right]\right]\\
&= E\left[\frac{1}{V}E_\theta\left[\frac{1}{\|X\|^2/V}\bigg|V\right]\right]\\
&\le E\left[\frac{1}{V}E_0\left[\frac{1}{\|X\|^2/V}\right]\right]\\
&= E_0\left[\frac{1}{\|X\|^2}\right].
\end{aligned}$$

This suffices to establish finiteness of the risk of $\delta_{a,r}^B(X)$. We now deal with the main part of the theorem. Using Corollary 2.1 and Theorem 2.3, we have

$$
R(\theta, \delta_{a,r}^B) = E\{E[\|\delta_{a,r}^B(X) - \theta\|^2 \,|V]\}
$$

$$
= E\left\{E\left[\|X - \theta\|^2 + V^2\left(\frac{a^2 r^2(\|X\|^2)}{V^2\|X\|^2} - \frac{2a(p-2)}{V}\frac{r(\|X\|^2)}{\|X\|^2}\right)\right.\right.
$$

$$
\left.\left. -4\,aVr'(\|X\|^2)\,\middle|\,V\right]\right\}
$$

$$
\leq R(\theta, X) + E\left\{aE\left[\frac{r(\|X\|^2)}{\|X\|^2/V}\,\middle|\,V\right]\left(\frac{a}{V} - 2(p-2)\right)\right\}, \qquad (5.2)
$$

since $r^2(\|X\|^2) \leq r(\|X\|^2)$ and $r'(\|X\|^2) \geq 0$. Now, as a consequence of the above monotone likelihood property, $\|X\|^2/V$ is stochastically decreasing in V. It follows that the conditional expectation in (5.2) is nondecreasing in V since, if $v_1 < v_2$, we have

$$
E\left[\frac{r(\|X\|^2)}{\|X\|^2/V}\,\middle|\,V = v_1\right] = E\left[\frac{r\left(v_1\frac{\|X\|^2}{V}\right)}{\|X\|^2/V}\,\middle|\,V = v_1\right]
$$

$$
\leq E\left[\frac{r\left(v_2\frac{\|X\|^2}{V}\right)}{\|X\|^2/V}\,\middle|\,V = v_1\right]
$$

$$
\leq E\left[\frac{r\left(v_2\frac{\|X\|^2}{V}\right)}{\|X\|^2/V}\,\middle|\,V = v_2\right]
$$

$$
= E\left[\frac{r(\|X\|^2)}{\|X\|^2/V}\,\middle|\,V = v_2\right].
$$

The first inequality follows since $r(\|X\|^2)$ is nondecreasing while the second since $r(t)/t$ is nonincreasing and $\|X\|^2/V$ is stochastically decreasing in V. Finally, using the fact that $aV^{-1} - 2(p-2)$ is decreasing in V, and the fact that $E[g(Y)h(Y)] \leq E[g(Y)]E[h(Y)]$ if g and h are monotone in opposite directions, it follows that

$$
R(\theta, \delta_{a,r}^B) \leq R(\theta, X) + aE\left[\frac{Vr(\|X\|^2)}{\|X\|^2}\right]E\left[\frac{a}{V} - 2(p-2)\right]
$$

$$
\leq R(\theta, X) \qquad (5.3)
$$

by assumption (a). Hence $\delta_{a,r}^B(X)$ is minimax, since X is minimax.

The dominance result follows since the inequality in (5.2) is strict if there is strict inequality in (2) or if $r'(\cdot)$ is strictly positive on a set of positive measure and the inequality in (5.3) is strict if the inequalities in (1) are strict. $\square$

Example 5.1 (The multivariate Student-t distribution) The multivariate Student-t distribution: If V has an inverse Gamma $(v/2, v/2)$ distribution (that is, $V \sim v/\chi_v^2$), then the distribution of X is a multivariate Student-t distribution with v degrees of freedom. Since $E[V] = E[v/\chi_v^2] = v/(v-2)$ for $v > 2$ and $E[V^{-1}] = E[\chi_v^2/v] = 1$, the conditions of Theorem 5.1 requires $0 \leq a \leq 2(p-2)$ and $v > 2$.

Example 5.2 (Examples of the function $r(t)$) The James-Stein estimator has $r(t) \equiv 1$ and hence satisfies conditions (2), (3) and (4) of Theorem 5.1. Also $r(t) = t/(t+b)$ satisfies these conditions. Similarly, the positive-part James-Stein estimator $\left(1 - a/X^{\mathrm{T}}X\right)_+ X$ is such that

$$r(t) = \begin{cases} t/a & \text{for } 0 \leq t \leq a \\ 1 & \text{for } t \geq a \end{cases}$$

and

$$\frac{r(t)}{t} = \begin{cases} 1/a & \text{for } 0 \leq t \leq a \\ 1/t & \text{for } t \geq a \end{cases}$$

hence also satisfies the conditions (2), (3) and (4) of Theorem 5.1.

It is worth noting, and easy to see, that if the sampling distribution is $N(\theta, I_p)$ and the prior distribution is any variance mixture of normal distributions as in (3.4), in the Baranchik representation of the Bayes estimator (see Corollary 3.1), the function $r(t)/t$ is always nonincreasing. This fact leads to the following observation on the (sampling distribution) robustness of Bayes minimax estimators for a normal sampling distribution. If $\delta^\pi(X) = \left(1 - a\, r(\|X\|^2)/\|X\|^2\right)X$ is a Bayes minimax estimator with respect to a scale mixture of normal priors for a $N(\theta, I_p)$ sampling distribution, and if $r(t)$ is nondecreasing, this Bayes minimax estimator remains minimax for a multivariate-t sampling distribution in Example 5.1 as long as the degrees of freedom is greater than two.

It is also interesting to note that, in general, there will be no uniformly optimal choice of the shrinkage constant "a" in the James-Stein estimator if the mixing distribution $G(\cdot)$ is nondegenerate. The optimal choice will typically depend on $\|\theta\|^2$. This is in contrast to the normal sampling distribution case, where $G(\cdot)$ is degenerate, and where the optimal choice is $a = (p-2)\sigma^2$.

5.2.2 Densities with Tails Flatter Than the Normal

In this section we consider the subclass of spherically symmetric densities $f(\|x - \theta\|^2)$ such that, for any $t \geq 0$ for which $f(t) > 0$,

$$\frac{F(t)}{f(t)} \geq c > 0 \tag{5.4}$$

for some fixed positive c, where

$$F(t) = \frac{1}{2} \int_t^\infty f(u) du. \tag{5.5}$$

This class was introduced in Berger (1975) (without the constant 1/2 multiplier).

This class of densities contains a large subclass of variance mixtures of normal densities but also many others. The following lemma gives some conditions which guarantee inclusion or exclusion from the class satisfying (5.4) and (5.5).

Lemma 5.1 *Suppose X has density $f(\|x - \theta\|^2)$.*

(1) *(Mixture of normals). If, for some distribution G on $(0, \infty)$,*

$$f(\|x - \theta\|^2) = \left(\frac{1}{\sqrt{2\pi}} \right)^p \int_0^\infty v^{-p/2} \exp\left\{ -\frac{\|x - \theta\|^2}{2v} \right\} dG(v)$$

where $E[V^{-p/2}]$ is finite, E denoting the expectation with respect to G, then $f(\cdot)$ is in the class (5.4) with $c = E[V^{-p/2+1}]/E[V^{-p/2}]$ for $p \geq 3$.
(2) *If $f(t) = h(t)e^{-at}$ with $h(t)$ nondecreasing, then $f(\cdot)$ is in the class (5.4).*
(3) *If $f(t) = e^{-atg(t)}$ where $g(t)$ is nondecreasing and $\lim_{t \to \infty} g(t) = \infty$, then $f(t)$ is not in the class (5.4).*

Proof (1) Applying the definition of F in (5.5) we have

$$F(t) = \frac{1}{2} \int_t^\infty f(u) du$$

$$= \frac{1}{2(\sqrt{2\pi})^p} \int_t^\infty \int_0^\infty v^{-p/2} \exp\{-u/2v\} dG(v) du$$

$$= \frac{1}{(\sqrt{2\pi})^p} \int_0^\infty v^{-p/2+1} \exp\{-t/2v\} dG(v).$$

Hence the ratio in (5.4) equals

$$\frac{F(t)}{f(t)} = \frac{\int_0^\infty v^{-p/2+1} \exp\{-t/2v\} dG(v)}{\int_0^\infty v^{-p/2} \exp\{-t/2v\} dG(v)}$$

$$\geq \frac{\int_0^\infty v^{-p/2+1} dG(v)}{\int_0^\infty v^{-p/2} dG(v)}$$

$$= \frac{E[V^{-p/2+1}]}{E[V^{-p/2}]}. \tag{5.6}$$

The inequality follows since the family of densities proportional to the function $v \mapsto v^{-p/2} \exp\{-t/2v\}$ has monotone (increasing) likelihood ratio in the parameter t. Note that if $p \geq 3$, $E[V^{-p/2}] < \infty$ implies $E[V^{-p/2+1}] < \infty$. This completes the proof of (1).

(2) In this case it follows

$$\frac{F(t)}{f(t)} = \frac{\frac{1}{2}\int_t^\infty h(u)e^{-au}\,du}{h(t)e^{-at}}$$

$$\geq \frac{1}{2}\int_t^\infty e^{-a(u-t)}\,du$$

$$= \frac{1}{2a}.$$

Hence (5.4) is satisfied with $c = 1/2a$, which proves (2).

(3) In this case it follows

$$2\lim_{t\to\infty}\frac{F(t)}{f(t)} = \lim_{t\to\infty}\frac{\int_t^\infty \exp\{-aug(u)\}\,du}{\exp\{-atg(t)\}}$$

$$= \lim_{t\to\infty}\int_t^\infty \exp\{-aug(u) + atg(t)\}\,du$$

$$= \lim_{t\to\infty}\int_0^\infty \exp\{-a(u+t)g(u+t) + atg(t)\}\,du$$

$$\leq \lim_{t\to\infty}\int_0^\infty \exp\{-aug(t)\}\,du$$

$$= \lim_{t\to\infty}\frac{1}{ag(t)}$$

$$= 0.$$

Hence $f(t)$ is not in the class (5.4), which shows (c). □

Part (2) of the lemma shows that densities with tails flatter than the normal (and including the normal) are in the class (5.4), while densities with tails "sufficiently lighter" than the normal are not included. Also the condition in part (3) is stronger than necessary in that it suffices that the condition hold only for all t larger than some positive K. See Berger (1975) for further details and discussion.

Example 5.3 Some specific examples in the class (5.4) include (see Berger 1975 for more details)

$$(1) \qquad f(t) = K/\cosh t \qquad (c \approx 1/2)$$

$$(2) \qquad f(t) = Kt(1+t^2)^{-m} \text{ with } m > p/4 \qquad (c = m/2)$$

$$(3) \qquad f(t) = Ke^{-\alpha t - \beta}/(1 + e^{-\alpha t - \beta})^2 \qquad (c = \alpha/2)$$

$$(4) \qquad f(t) = Kt^n e^{-t/2} \text{ for } n \geq 0 \qquad (c = 1).$$

The latter two distributions are known as the logistic type and Kotz , respectively.

The following lemma plays the role of Stein's lemma (Theorem 2.1) for the family of spherically symmetric densities.

Lemma 5.2 *Let X have density $f(\|x-\theta\|^2)$ and let $g(X)$ be a weakly differentiable function such that $E_\theta[|(X-\theta)^T g(X)|] < \infty$. Then*

$$E_\theta[(X-\theta)^T g(X)] = E_\theta\left[\frac{F(\|X-\theta\|^2)}{f(\|X-\theta\|^2)} \operatorname{div} g(X)\right]$$

$$= C\, E_\theta^*\left[\operatorname{div} g(X)\right]$$

where $F(t)$ is defined as in (5.5) and E_θ^* denotes expectation with respect to the density

$$x \mapsto \frac{1}{C} F(\|x-\theta\|^2)$$

and where it is assumed that

$$C = \int_{\mathbb{R}^p} F(\|x-\theta\|^2)\, dx < \infty.$$

Proof Note that the existence of the expectations in Lemma 5.2 will be guaranteed for any function $g(x)$ such that $E_\theta[\|g(x)\|^2] < \infty$ as soon as $E_0[\|X\|^2] < \infty$. The proof will follow along the lines of Sect. 2.4 making use of Stokes' theorem. It follows that

$$E[(X-\theta)^T g(X)]$$

$$= \int_{R^p} (x-\theta)^T g(x) f(\|x-\theta\|^2)\, dx$$

$$= \int_0^\infty \int_{S_{R,\theta}} (x-\theta)^T g(x)\, f(\|x-\theta\|^2)\, d\sigma_{R,\theta}(x)\, dR \quad \text{(by Lemma 1.4)}$$

$$= \int_0^\infty \int_{S_{R,\theta}} \left(\frac{x-\theta}{\|x-\theta\|}\right)^T d\sigma_{R,\theta}(x)\, R\, f(R^2)\, dR$$

$$= \int_0^\infty \int_{B_{R,\theta}} \operatorname{div} g(x)\, dx\, R\, f(R^2)\, dR \qquad \text{(Stokes' theorem)}$$

$$= \int_{R^p} \operatorname{div} g(x) \int_{\|x-\theta\|}^\infty R f(R^2)\, dR\, dx \qquad \text{(Fubini's theorem)}$$

$$= \int_{R^p} \operatorname{div} g(x)\, F(\|x-\theta\|^2)\, dx$$

$$= E_\theta \left[\text{div } g(x) \frac{F(\|x - \theta\|^2)}{f(\|x - \theta\|^2)} \right]$$

$$= C \, E_\theta^* \left[\text{div } g(X) \right]$$

$\square$

Now, with the important analog of Stein's lemma in hand, we can extend some of the minimaxity results from the Gaussian setting to the case of spherically symmetric distributions. The following result gives conditions for minimaxity of estimators of the Baranchik type.

Theorem 5.2 *Let X have density $f(\|x - \theta\|^2)$ which satisfies* (5.4) *for some $0 < c < \infty$. Assume also that $E_0[\|X\|^2] < \infty$ and $E_0[\|X\|^{-2}] < \infty$. Let*

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

where $r(\cdot)$ is absolutely continuous. Then $\delta_{a,r}^B(X)$ is minimax for $p \geq 3$ provided

(1) $0 < a \leq 2 \, c \, (p - 2)$,
(2) $0 \leq r(t) \leq 1$, *and*
(3) $r(\cdot)$ *is nondecreasing.*

Furthermore $\delta_{a,r}^B(X)$ dominates X provided both inequalities are strict in (1) *or in* (2) *on a set of positive measure or if $r'(\cdot)$ is strictly positive on a set of positive measure.*

Proof We note that the conditions ensure finiteness of the risk so that Lemma 5.2 is applicable. Hence we have

$$R(\theta, \delta_{a,r}^B) = E_\theta \left[\|X - \theta\|^2 + \frac{a^2 r^2(\|X\|^2)}{\|X\|^2} - 2 \frac{a \, r(\|X\|^2) X^{\mathsf{T}}(X - \theta)}{\|X\|^2} \right]$$

$$= R(\theta, X) + a \, E_\theta \left[\frac{ar^2(\|X\|^2)}{\|X\|^2} - 2 \, \text{div} \left(\frac{r(\|X\|^2)X}{\|X\|^2} \right) \frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)} \right]$$

by Lemma 5.2. Therefore the risk difference between $\delta_{a,r}^B(X)$ and X equals

$$\Delta_\theta = a \, E_\theta \left[\frac{ar^2(\|X\|^2)}{\|X\|^2} - \left(\frac{2(p-2)r(\|X\|^2)}{\|X\|^2} + 4 \, r'(\|X\|^2) \right) \frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)} \right]$$

$$\leq a \, E_\theta \left[\frac{r(\|X\|^2)}{\|X\|^2} \left(a - 2(p-2) \frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)} \right) \right]$$

$$\leq a \, E_\theta \left[\frac{r(\|X\|^2)}{\|X\|^2} \left(a - 2(p-2) \, c \right) \right]$$

$$\leq 0.$$

The domination part follows as in Theorem 5.1.

$\square$

Theorem 5.2 applies for certain densities for which Theorem 5.1 is not applicable and additionally lifts the restriction that $r(t)/t$ is nonincreasing. However, if the density is a mixture of normals, and both theorems apply, the shrinkage constant "a" given by Theorem 5.1 (with $a = 2(p-2)/E[V^{-1}]$) is strictly larger than that for Theorem 5.2 (with $a = 2(p-2)c$) whenever the mixing distribution $G(\cdot)$ is not degenerate. To see this note that

$$\frac{1}{E[V^{-1}]} > c = \frac{E[V^{-p/2+1}]}{E[V^{-p/2}]}$$

or equivalently

$$E[V^{-p/2}] > E[V^{-1}]E[V^{-p/2+1}]$$

whenever the positive random variable V is non-degenerate. Note also that $E[V^{-1}] < \infty$ whenever $E[V^{-p/2}] < \infty$ and $p \geq 3$.

Example 5.4 (The multivarite Student-t distribution, continued) Suppose X has a p-variate Student-t distribution with ν degrees of freedom as in Example 5.1, so that V has an inverse $Gamma(\nu/2, \nu/2)$ distribution. In this case

$$E[V^{-p/2}] = \frac{2^{p/2}\Gamma\left(\frac{p+\nu}{2}\right)}{\nu^{p/2}\Gamma\left(\frac{\nu}{2}\right)}$$

which is finite for all $\nu > 0$ and $p > 0$.

The bound on the shrinkage constant, "a", in Theorem 5.1 is $2(p-2)$ as shown in Example 5.1, while the bound on "a", in Theorem 5.2, as indicated above, is given by

$$2(p-2)\frac{E[V^{-p/2+1}]}{E[V^{-p/2}]} = 2(p-2)\left(\frac{\nu}{\nu+p-2}\right) < 2(p-2).$$

Hence, for large p, the bound on the shrinkage factor "a" can be substantially less for Theorem 5.2 than for Theorem 5.1 in the case of a multivariate-t sampling distribution. Note that, for fixed p, as ν tends to infinity the smaller bound tends to the larger one (and the Student-t distribution tends to the normal).

Example 5.5 (Examples 5.3 continued) All of the distributions in Example 5.3 satisfy the assumptions of Theorem 5.2 (under suitable moment conditions for the second density). It is interesting to note that for the Kotz distribution , the value of c ($= 1$), as in (5.4), doesn't depend on the parameter $n > 0$. Hence the bound on the shrinkage factor "a" is $2(p-2)$ and is also independent of n, indicating a certain distributional robustness of the minimaxity property of Baranchik type estimators with $a < 2(p-2)$.

With additional assumptions on the function $F(t)/f(t)$ in (5.4) (i.e. it is either monotone increasing or monotone decreasing), theorems analogous to Theorem 5.2 can be developed which further improve the bounds on the shrinkage factor "a". These typically may involve additional assumptions on the function $r(\cdot)$. We will see examples of this type in the next section.

5.3   More General Minimax Estimators

We now consider minimaxity of general estimators of the form $X + a\,g(X)$. The initial results rely on Lemma 5.2. The first result follows immediately from this lemma and gives an expression for the risk.

Corollary 5.1 *Let X have a density $f(\|x - \theta\|^2)$ such that $E_0[\|X\|^2] < \infty$ and let $g(X)$ be weakly differentiable and be such that $E_\theta[\|g(X)\|^2] < \infty$.*
   Then, for loss $L(\theta, \delta) = \|\delta - \theta\|^2$, the risk of $X + a\,g(X)$ can be expressed as

$$R(\theta, X + a\,g(X)) = R(\theta, X) + E_\theta\big[a^2\,\|g(X)\|^2 + 2\,a\,Q(\|X - \theta\|^2)\,\mathrm{div}\,g(X)\big]$$
(5.7)

where

$$Q(\|X - \theta\|^2) = \frac{F(\|X - \theta\|^2)}{f(\|X - \theta\|^2)}$$
(5.8)

and where $F(\|X - \theta\|^2)$ is defined in (5.5).

An immediate consequence of Corollary 5.1 when the density of f satisfies (5.4), i.e. $Q(t) \geq c > 0$ for some constant c, is the following.

Corollary 5.2 *Under the assumptions of Corollary 5.1, assume that, for some $c > 0$, we have $Q(t) \geq c$ for any $t \geq 0$. Then $X + g(X)$ is minimax and dominates X provided, for any $x \in \mathbb{R}^p$,*

$$\|g(x)\|^2 + 2\,c\,\mathrm{div}\,g(x) \leq 0$$

with strict inequality on a set of positive measure.

The following two theorems establish minimaxity results under the assumption that $Q(t)$ is monotone.

Theorem 5.3 (Brandwein et al. 1993) *Suppose X has density $f(\|x - \theta\|^2)$ such that $E_0[\|X\|^2] < \infty$ and that $Q(t)$ in (5.8) is nonincreasing. Suppose there exists a nonpositive function $h(U)$ such that $E_{R,\theta}[h(U)]$ is nondecreasing where $U \sim U_{R,\theta}$ (the uniform distribution on the sphere of radius R centered at θ) and such that $E_\theta[|h(x)|] < \infty$. Furthermore suppose that $g(X)$ is weakly differentiable and also satisfies*

(1) $\text{div } g(X) \leq h(X)$,
(2) $\|g(X)\|^2 + 2h(X) \leq 0$, and
(3) $0 \leq a \leq E_0(\|X\|^2)/p$.

Then $\delta(X) = X + ag(X)$ is minimax. Also $\delta(X)$ dominates X provided $g(\cdot)$ is nonzero with positive probability and strict inequality holds with positive probability in (1) *or* (2)*, or both inequalities are strict in* (3)*.*

Proof Note that $g(x)$ satisfies the conditions of Corollary 5.1. Then we have

$$R(\theta, \delta) = R(\theta, X) + a\, E[a\, \|g(X)\|^2 + 2\, Q(\|X - \theta\|^2)\, \text{div } g(X)]$$
$$= R(\theta, X) + a\, E[E_{R,\theta}[a\, \|g(X)\|^2 + 2\, Q\, (R^2)\, \text{div } g(X)]]$$

where $E_{R,\theta}$ is as above and E denotes the expectation with respect to the radial distribution. Now, using (1) and (2), we have

$$R(\theta, \delta) \leq R(\theta, X) + a\, E[E_{R,\theta}[-2\, a\, h(X) + 2\, Q(R^2)\, h(X)]]$$
$$= R(\theta, X) + 2\, a\, E[(a - Q(R^2))\, E_{R,\theta}[-h(X)]]$$
$$\leq R(\theta, X) + 2\, a\, E[a - Q(R^2)]\, E_\theta[-h(X)]$$

by the monotonicity assumptions on $E_{R,\theta}[h(\cdot)]$ and $Q(t)$ as well as the covariance inequality.

Hence, since $-h(X) \geq 0$, we have $R(\theta, \delta) \leq R(\theta, X)$, provided $0 \leq a \leq E[Q(R^2)]$. Now $E[Q(R^2)] = E_0[\|X\|^2]/p$ by Lemma 5.3 below, hence δ is minimax. The domination result follows since the additional conditions imply strict inequality between the risks.   □

Lemma 5.3 *For any $k > -p$ such that $E[R^{k+2}] < \infty$,*

$$E[R^k Q(R^2)] = \frac{1}{p+k}\, E[R^{k+2}].$$

In particular, we have

$$E[Q(R^2)] = \frac{1}{p}\, E[R^2] = \frac{1}{p}\, E_0[\|X\|^2]$$

and, for $p \geq 3$,

$$E\left[\frac{Q(R^2)}{R^2}\right] = \frac{1}{p-2}.$$

Proof Recall that the radial density $\varphi(r)$ of $R = \|X - \theta\|$ can be expressed as $\varphi(r) = \sigma(S)r^{p-1}f(r^2)$ where $\sigma(S)$ is the area of the unit sphere S in R^p. By (5.8) and (5.5), we have

$$E[R^k Q(R^2)] = \frac{1}{2} \int_{R^p} \|x\|^k \int_{\|x\|^2}^{\infty} f(t)\, dt\, dx$$

$$= \frac{1}{2} \int_0^{\infty} \int_{B_{\sqrt{t}}} \|x\|^k dx\, f(t)\, dt \quad \text{by Fubini's theorem}$$

$$= \frac{1}{2} \int_0^{\infty} \int_0^{\sqrt{t}} \sigma(S)\, r^{k+p-1} dr\, f(t)\, dt \quad \text{by Lemma 1.4}$$

$$= \frac{1}{2} \int_0^{\infty} \sigma(S) \frac{t^{(k+p)/2}}{k+p} f(t)\, dt$$

$$= \frac{1}{k+p} \int_0^{\infty} r^{k+2} \varphi(r)\, dr \quad \text{by the change of variable } t = r^2$$

$$= \frac{1}{k+p} E[R^{k+2}].$$

Note that positivity of integrands and $E[R^{k+2}] < \infty$ implies $E[R^k Q(R^2)] < \infty$.
□

The next theorem reverses the monotonicity assumption on $Q(\cdot)$ and changes the condition on the function $h(X)$ which, in turn, bounds the divergence of $g(X)$.

Theorem 5.4 (Brandwein et al. 1993)   *Suppose X has a density $f(\|x - \theta\|^2)$ such that $E_0[\|X\|^2] < \infty$ and $E_0[1/\|X\|^2] < \infty$ and such that $Q(t)$ in (5.8) is nondecreasing. Suppose there exists a nonpositive function $h(X)$ such that $E_{R,\theta}\left[R^2 h(U)\right]$ is nonincreasing where $U \sim U_{R,\theta}$ and such that $E_\theta[-h(X)] < \infty$.*
Furthermore suppose that $g(X)$ is weakly differentiable and also satisfies

(1) *div $g(X) \le h(X)$,*
(2) *$\|g(X)\|^2 + 2\,h(X) \le 0$, and*
(3) *$0 \le a \le \frac{1}{(p-2)E_0(1/\|X\|^2)}$.*

Then $\delta(X) = X + a\, g(X)$ is minimax. Also $\delta(X)$ dominates X provided $g(\cdot)$ is nonzero with positive probability and strict inequality holds with positive probability in (1) or (2), or both inequalities are strict in (3).

Proof As in the proof of Theorem 5.3, we have

$$R(\theta, \delta) \le R(\theta, X) + 2\,a\, E[(a - Q(R^2))\, E_{R,\theta}[-h(X)]]$$

$$= R(\theta, X) + 2\,a\, E\left[\left(\frac{a}{R^2} - \frac{Q(R^2)}{R^2}\right) E_{R,\theta}[-R^2\, h(X)]\right]$$

$$\le R(\theta, X) + 2\,a\, E\left[\frac{a}{R^2} - \frac{Q(R^2)}{R^2}\right] E_{R_0,\theta}[-R_0^2\, h(X)]$$

where R_0 is a point such that $a - Q(R_0^2) = 0$, provided such a point exists. Here we have used the version of the covariance inequality that states

$$Ef(X)g(X) \leq Ef(X)g(X_0)$$

provided that $g(X)$ is nondecreasing (respectively, nonincreasing) and $f(X)$ changes sign once from $+$ to $-$ (respectively, $-$ to $+$) at X_0. But such a point R_0 does exist provided

$$E\left[\frac{a}{R^2} - \frac{Q(R^2)}{R^2} \right] \leq 0$$

since $Q(R^2)$ is nondecreasing.

It follows that $R(\theta, \delta) \leq R(\theta, X)$ provided that $a E[\frac{1}{R^2}] \leq E[\frac{Q(R^2)}{R^2}]$. However $E[\frac{Q(R^2)}{R^2}] = \frac{1}{p-2}$ by Lemma 5.3 and hence the result follows as in Theorem 5.3.   □

Note that the bound on "a" in both of these theorems is strictly larger than the bound in Theorem 5.2 provided $Q(R^2)$ is not constant. This is so since the bound in Theorem 5.2 is based on $c = \inf Q(R^2)$ while, in these results, the bound is equal to a (possibly weighted) average of $Q(R^2)$.

We indicate the utility of these two results by applying them to the James-Stein estimator.

Corollary 5.3 *Let $X \sim f(\|x - \theta\|^2)$ for $p \geq 4$ and let $\delta_b^{JS}(X) = (1 - b/\|X\|^2)X$. Assume also that $E_0[\|X\|^2] < \infty$ and $E_0[1/\|X\|^2] < \infty$. Then $\delta_b^{JS}(X)$ is minimax and dominates X provided either*

(1) *$Q(R^2)$ is nonincreasing and*

$$0 < b < 2(p-2)\frac{E_0\|X\|^2}{p}, \quad or$$

(2) *$Q(R^2)$ is nondecreasing and*

$$0 < b < \frac{2}{E_0(1/\|X\|^2)}.$$

Proof We apply Theorems 5.3 and 5.5 with $g(X) = -[2(p-2)/\|X\|^2]X$, $\mathrm{div}\, g(X) = -2(p-2)^2/\|X\|^2 = h(X)$. It follows from Lemma A.5 in Appendix A.10 that when $p \geq 4$, $E_{\theta, R}[h(U)]$ is nondecreasing in R and $E_{\theta, R}[R^2 h(U)]$ is nonincreasing in R. Hence, if $Q(R^2)$ is nonincreasing, Theorem 5.3 implies that

$$\delta_a(X) = X - \frac{2(p-2)a}{\|X\|^2}X = \delta_{2(p-2)a}^{JS}(X)$$

is minimax and dominates X provided $0 < a < E_0[\|X\|^2]/p$ or equivalently $0 < 2(p-2)a < 2(p-2)E_0(\|X\|^2)/p$ which is (1) with $b = 2(p-2)a$. Similarly,

applying Theorem 5.5 when $Q(R^2)$ is nondecreasing, we find that $\delta_a(X)$ is minimax and dominates X if

$$0 < a < \frac{1}{(p-2)E_0(1/\|X\|^2)}$$

which is (2).   □

Example 5.6 (Densities with increasing and decreasing $Q(R^2)$) Note first that variance mixtures of normal distributions have increasing $Q(R^2)$ since, by (5.6) and (5.8), $Q(R^2)$ may be viewed as the expected value of V with respect to a family of distributions with monotone increasing likelihood ratio in $t = R^2$. Note also that the bound for the shrinkage constant "a" in a James-Stein estimator is the same in Corollary 5.3 as it is in Theorem 5.1 for mixtures of normals.

We also note that, if we consider $f(t)$ to be proportional to a density of a positive random variable, then $2\,Q(t)$ is the reciprocal of the hazard rate. There is a large literature on increasing and decreasing hazard rates (see, for example, Barlow and Proschan 1981).

We note that the monotonicity of $Q(t)$ may be determined in many cases by studying the log-convexity or the log-concavity of $f(t)$. In particular, if $\ln f(t)$ is convex (concave), then $Q(t)$ is nondecreasing (nonincreasing). To see this, note that

$$Q(t) = \frac{1}{2} \frac{\int_t^\infty f(u)\,du}{f(t)} = \frac{1}{2} \int_0^\infty \frac{f(s+t)}{f(t)}\,ds$$

and hence $Q(t)$ will be nondecreasing (nonincreasing) if $\frac{f(s+t)}{f(t)}$ is nondecreasing (nonincreasing) in t for each $s > 0$. But, assuming for simplicity that f is differentiable, for any $t \geq 0$ such that $f(t) > 0$,

$$\frac{d}{dt}\left(\frac{f(s+t)}{f(t)} \right) = \frac{f(t)f'(s+t) - f(s+t)f'(t)}{f^2(t)}$$

$$= \frac{f(s+t)}{f(t)}\left[\frac{f'(s+t)}{f(s+t)} - \frac{f'(t)}{f(t)} \right]$$

$$= \frac{f(s+t)}{f(t)}\left[\frac{d}{dt}\ln f(s+t) - d/dt \ln f(t) \right].$$

This is positive or negative when $\ln f(s+t)$ is convex or concave in t, respectively. For example if X has a Kotz distribution with parameter n, $f(t) \propto t^n e^{-t/2}$. Then $\ln f(t) = K + n \ln t - \frac{t}{2}$ which is concave if $n \geq 0$ and convex if $n \leq 0$. Hence $Q(t)$ is decreasing if $n > 0$ and increasing if $n < 0$. Of course the log-convexity (log-concavity) of $f(t)$ is not a necessary condition for the nondecreasing (nonincreasing) monotonicity of $Q(t)$. Thus, it is easy to check that $f(t) \propto \exp(-t^2)\exp[-1/2\int_0^t \exp(-u^2)\,du]$ leads to $Q(t) = \exp(t^2)$, which is increasing. But $\log f(t)$ is not convex.

An important class of distributions is covered by the following corollary.

Corollary 5.4 *Let* $X \sim f(\|x - \theta\|^2)$ *for* $p \geq 4$ *with* $f(t) \propto \exp(-\beta t^\alpha)$ *where* $\alpha > 0$ *and* $\beta > 0$. *Then* $\delta_b^{JS}(X) = (1 - b/\|X\|^2)X$ *is minimax and dominates* X *provided either*

(1) $\alpha \leq 1$ *and* $0 < b < \frac{2}{\beta^{1/\alpha}} \frac{p-2}{p} \frac{\Gamma((p+2)/2\alpha)}{\Gamma(p/2\alpha)}$ *or*

(1) $\alpha > 1$ *and* $0 < b < \frac{2}{\beta^{1/\alpha}} \frac{\Gamma(p/2\alpha)}{\Gamma((p-2)/2\alpha)}$.

Proof By the above discussion, $Q(R^2)$ is nonincreasing (nondecreasing) for $\alpha \geq 1$ ($\alpha \leq 1$). Then the result follows from Corollary 5.3 and the fact that

$$E_0[\|X\|^k] = \frac{1}{\beta^{k/2\alpha}} \frac{\Gamma(\frac{p+k}{2\alpha})}{\Gamma(\frac{p}{2\alpha})}$$

for $k > -p$.   □

The final theorem of this section gives conditions for minimaxity of estimators of the form $X + a\, g(X)$ for general spherically symmetric distributions. Note that no density is needed for this result which relies on the radial distribution defined in Theorem 4.1.

We first need the following lemma which will play the role of the Stein lemma in the proof of the domination and minimaxity results.

Lemma 5.4 *Let* X *have a spherically symmetric distribution around* θ, *and let* $g(X)$ *be a weakly differentiable function such that* $E_\theta[\,|(X - \theta)^\mathsf{T} g(X)|\,] < \infty$. *Then*

$$E_\theta[(X - \theta)^\mathsf{T} g(X)] = \frac{1}{p} E\left[R^2 \int_{B_{R,\theta}} \operatorname{div} g(X)\, d\mathscr{V}_{R,\theta}(X) \right]$$

where E *denotes the expectation with respect to the radial distribution and where* $\mathscr{V}_{R,\theta}(\cdot)$ *is the uniform distribution on* $B_{R,\theta}$, *the ball of radius* R *centered at* θ.

Proof Let ρ be the radial distribution and according to Theorem 4.1, we have

$$E[(X - \theta)^\mathsf{T} g(X)] = \int_{\mathbb{R}_+} \int_{S_{R,\theta}} (x - \theta)^\mathsf{T} g(x)\, d\mathscr{U}_{R,\theta}(x)\, d\rho(R)$$

$$= \int_{\mathbb{R}_+} \frac{R}{\sigma_{R,\theta}(S_{R,\theta})} \int_{S_{R,\theta}} \frac{(x - \theta)^\mathsf{T}}{\|x - \theta\|} g(x)\, d\sigma_{R,\theta}(x)\, d\rho(R)$$

$$= \int_{\mathbb{R}_+} \frac{R}{\sigma_{R,\theta}(S_{R,\theta})} \int_{B_{R,\theta}} \operatorname{div} g(x)\, dx\, d\rho(R) \quad \text{by Stokes' theorem}$$

$$= \frac{1}{p} \int_{\mathbb{R}_+} \int_{B_{R,\theta}} \operatorname{div} g(x)\, d\mathscr{V}_{R,\theta}(x)\, R^2 d\rho(R)$$

since the volume of $B_{R,\theta}$ equals $\lambda(B_{R,\theta}) = R\sigma_{R,\theta}(S_{R,\theta})/p$.   □

Theorem 5.5 (Brandwein and Strawderman 1991a)   *Let X have a spherically symmetric distribution around θ, and suppose $E_0[\|X\|^2] < \infty$ and $E_0[1/\|X\|^2] < \infty$. Suppose there exists a nonpositive function $h(\cdot)$ such that $h(X)$ is subharmonic and $E_{R,\theta}[R^2 h(U)]$ is nonincreasing where $U \sim \mathscr{U}_{R,\theta}$ and such that $E_\theta[|h(x)|] < \infty$. Furthermore suppose that $g(X)$ is weakly differentiable and also satisfies*

(1) *div $g(X) \leq h(X)$,*
(2) *$\|g(X)\|^2 + 2 h(X) \leq 0$, and*
(3) *$0 \leq a \leq \frac{1}{p E_0(1/\|X\|^2)}$.*

Then $\delta(X) = X + a\, g(X)$ is minimax. Also $\delta(X)$ dominates X provided $g(\cdot)$ is nonzero with positive probability and strict inequality holds with positive probability in (1) *or* (2)*, or both inequalities are strict in* (3)*.*

Proof Using Lemma 5.4 and Conditions (1) and (2), we have

$$
\begin{aligned}
R(\theta, \delta) &= R(\theta, X) + a\, E_\theta\big[a\, \|g(X)\|^2 + 2\,(X - \theta)^{\mathrm{T}} g(X)\big] \\
&\leq R(\theta, X) + 2\, a\, E_\theta\big[-a\, h(X) + (X - \theta)^{\mathrm{T}} g(X)\big] \\
&= R(\theta, X) + 2\, a\left\{ E_\theta\big[-a\, h(X)\big] + \frac{1}{p} E\left[R^2 \int_{B_{R,\theta}} \operatorname{div} g(X)\, d\mathscr{V}_{R,\theta}(X)\right]\right\} \\
&\leq R(\theta, X) + 2\, a\left\{ E_\theta\big[-a\, h(X)\big] + \frac{1}{p} E\left[R^2 \int_{B_{R,\theta}} h(X)\, d\mathscr{V}_{R,\theta}(X)\right]\right\}.
\end{aligned}
$$

By subharmonicity of h (see Appendix A.8 and Sections 1.3 and 2.5 in du Plessis 1970),

$$
\int_{B_{R,\theta}} h(X) d\mathscr{V}_{R,\theta}(X) \leq \int_{S_{R,\theta}} h(X) d\mathscr{U}_{R,\theta}(X).
$$

Hence,

$$
\begin{aligned}
R(\theta, \delta) &\leq R(\theta, X) + 2\, a\left\{ E_\theta\big[-a\, h(X)\big] + \frac{1}{p} E\left[R^2 \int_{S_{R,\theta}} h(X)\, d\mathscr{U}_{R,\theta}(X)\right]\right\} \\
&= R(\theta, X) + 2\, a\, E\left[\left(\frac{a}{R^2} - \frac{1}{p}\right) \cdot \left(-R^2 \int_{S_{R,\theta}} h(X) d\mathscr{U}_{R,\theta}(X)\right)\right] \\
&= R(\theta, X) + 2\, a\, E\left[\left(\frac{a}{R^2} - \frac{1}{p}\right)\left(- E_{R,\theta}[R^2 h(X)]\right)\right] \\
&\leq R(\theta, X) + 2\, a\, E\left[\left(\frac{a}{R^2} - \frac{1}{p}\right)\right] E\left[- E_{R,\theta}[R^2 h(X)]\right].
\end{aligned}
$$

The last inequality follows from the monotonicity of $E_{R,\theta}[R^2 h(X)]$ and the covariance inequality. Hence $R(\theta, \delta) \leq R(\theta, X)$ when $E\left[a/R^2 - 1/p\right] \leq 0$ which is equivalent to (3). The domination part follows as before. $\qquad\square$

We note that the shrinkage constant in the above result $1/\{p E_0[1/\|X\|^2]\}$ is somewhat smaller than the constant in Theorem 5.4 ($a = 1/\{(p-2)E_0[1/\|X\|^2]\}$), but Theorem 5.5 has essentially no restrictions on the distribution of X aside from moment conditions (which coincide in Theorems 5.4 and 5.5). In particular we do not even assume that a density exists! However there is an additional assumption of subharmonicity of h.

The following useful corollary gives minimaxity for James-Stein estimators in dimension $p \geq 4$ for all spherically symmetric distributions with finite $E_0[\|X\|^2]$ and $E_0[1/\|X\|^2]$.

Corollary 5.5 *Let X have a spherically symmetric distribution with $p \geq 4$, and suppose $E_0[\|X\|^2] < \infty$ and $E_0[1/\|X\|^2] < \infty$. Then*

$$\delta_a^{JS}(X) = \left(1 - \frac{a}{\|X\|^2}\right) X$$

is minimax and dominates X provided

$$0 < a < \frac{1}{p E_0(1/\|X\|^2)}.$$

Proof Here $g(X) = -X/\|X\|^2$ and is weakly differentiable for $p \geq 3$. Then $\operatorname{div} g(X) = -(p-2)/\|X\|^2$ and $\|g(X)\|^2 = 1/\|X\|^2$ so that Conditions (1) and (2) of Theorem 5.5 are satisfied with $h(X) = -\alpha/\|X\|^2$ where $0 \leq \alpha \leq p - 2$. Now the subharmonicity of $h(X)$ and its monotonicity condition hold since it is shown in the appendix that, for $p \geq 4$, $1/\|X\|^2$ is super-harmonic (so that $E_{R,\theta}[1/\|X\|^2]$ is nonincreasing in R) and that $R^2 E_{R,\theta}[1/\|U\|^2]$ is nondecreasing in R.

Furthermore, it is worth noting that $E_{R,\theta}[1/\|U\|^2]$ is nonincreasing in $\|\theta\|$ (see Lemma A.5 and remark that follows). Hence, for any $\theta \in \mathbb{R}^p$, we have $E_\theta[-h(X)] < \infty$ since

$$E_{R,\theta}[1/\|X\|^2] \leq E_{R,0}[1/\|X^2\|]$$

so that

$$E_\theta[1/\|X\|^2] \leq E_0[1/\|X\|^2] < \infty,$$

by assumption. $\qquad\square$

Example 5.7 (Nonspherical minimax estimators)   In Sect. 2.4.4, we considered estimators which shrink toward a subspace. Theorem 5.5 allows us to show that estimators of this type are minimax for general spherically symmetric distributions.

To be specific, suppose V is a $s < p$ dimensional linear subspace and let

$$\delta_a(X) = P_V X + \left(1 - \frac{a}{\|X - P_V X\|^2}\right)(X - P_V X).$$

As in the proof of Theorem 2.6, it can be shown that the risk of $\delta_a(X)$ equals

$$R(\theta, \delta_a(X)) = E_{\nu_1}[\|Y_1 - \nu_1\|^2] + E_{\nu_2}\left[\left\|\left(1 - \frac{a}{\|Y_2\|^2}\right)Y_2 - \nu_2\right\|^2\right], \qquad (5.9)$$

where Y_1, Y_2, ν_1 and ν_2 are as in Theorem 2.6.

In the present case, Y_2 has a spherically symmetric distribution about ν_2 of dimension $p - s$. Hence, by Theorem 5.5,

$$E(\theta, \delta_a(X)) \leq E_{\nu_1}[\|Y_1 - \nu_1\|^2] + E_{\nu_2}[\|Y_2 - \nu_2\|^2]$$
$$= E_\theta\|X - \theta\|^2$$
$$= R(\theta, X),$$

provided $p - s \geq 4$ and

$$0 < a < \frac{1}{(p - s)\,E_0[1/\|X - P_V X\|^2]}.$$

5.4 Bayes Estimators

In this section, we consider (generalized) Bayes estimators of the location vector $\theta \in \mathbb{R}^p$ of a spherically symmetric distribution. More specifically let X be a random vector in $\mathbb{R}^p$ with density $f(\|x - \theta\|^2)$ and let $\pi(\theta)$ be a prior density. Under quadratic loss $\|\delta - \theta\|^2$, the (generalized) Bayes estimator of θ is the posterior mean given by

$$\delta_\pi(X) = X + \frac{1}{m(X)} \int_{\mathbb{R}^p} (\theta - X)\, f(\|X - \theta\|^2)\, \pi(\theta)\, d\theta \qquad (5.10)$$

where $m(x)$ is the marginal

$$m(x) = \int_{\mathbb{R}^p} f(\|x - \theta\|^2)\pi(\theta)\, d\theta. \qquad (5.11)$$

Recall from Sect. 3.1.1 that, in the normal case (that is, $f(t) \propto \exp(-t/2\sigma^2)$ with σ^2 known) the superharmonicity of $\sqrt{m(x)}$ is a sufficient condition for minimaxity of $\delta_\pi(X)$. This superharmonicity is implied by that of $m(x)$ and in

turn by that of $\pi(\theta)$. While in the nonnormal case minimaxity has been studied by many authors (for example, see Strawderman (1974b); Berger (1975); Brandwein and Strawderman (1978, 1991a)) relatively few results on minimaxity of Bayes estimators are known. The primary technique to establish minimaxity is through a Baranchik representation of the form $(1 - a\, r\, (\|X\|^2)/\|X\|^2)X$. The minimaxity conditions are essentially those developed in Theorems 5.3 and 5.4 and most of the derivations are in the context of variance mixtures of normals. See Strawderman (1974b), Maruyama (2003a) and Fourdrinier et al. (2008) for more discussion and results on Bayes estimation in this setting.

The main difficulty in using Theorem 5.1 with mixtures of normals densities for the sampling distribution is to prove the monotonicity (and boundedness) properties of the function $r(\cdot)$. Maruyama (2003a) and Fourdrinier et al. (2008) consider priors which are mixtures of normals as well. Their main condition for obtaining minimaxity of the corresponding Bayes estimator is that the mixing density g of the sampling distribution has monotone nondecreasing likelihood ratio when considered as a scale parameter family. In Fourdrinier et al. (2008), explicit use is made of that monotone likelihood ratio property for the mixing (possibly generalized) density h of the prior distribution.

The main result of Fourdrinier et al. (2008) is the following. Consult that paper for the somewhat technical proof.

Theorem 5.6 *Let X be a random vector in $\mathbb{R}^p$ ($p \geq 3$) distributed as a variance mixture of multivariate normal distributions with density*

$$f(x) = \int_0^\infty \frac{1}{(2\pi v)^{p/2}} \exp\left(-\frac{1}{2} \frac{\|x - \theta\|^2}{v} \right) g(v)\, dv \qquad (5.12)$$

where g is the density of a known nonnegative random variable V. Let π be a (generalized) prior with density of the form

$$\pi(\theta) = \int_0^\infty \frac{1}{(2\pi t)^{p/2}} \exp\left(-\frac{1}{2} \frac{\|\theta\|^2}{t} \right) h(t)\, dt \qquad (5.13)$$

where h is a function from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that this integral exists.

Assume that the mixing density g is such that

$$E[V] = \int_0^\infty v\, g(v)\, dv < \infty \text{ and } E[V^{-p/2}] = \int_0^\infty v^{-p/2}\, g(v)\, dv < \infty. \qquad (5.14)$$

Assume also that the mixing function h of the (possibly improper) prior density π is absolutely continuous and satisfies

$$\lim_{t \to \infty} \frac{h(t)}{t^\beta} = c \qquad (5.15)$$

*for some $\beta < p/2 - 1$ and some $0 < c < \infty$. Assume, finally, that h and g have
monotone increasing likelihood ratio when considered as a scale parameter family.*
    Then, if there exist $K > 0$, $t_0 > 0$ and $\alpha < 1$ such that

$$h(t) \leq K\, t^{-\alpha} \quad \text{for } 0 < t < t_0, \tag{5.16}$$

*the (generalized or proper) Bayes estimator δ_h with respect to the prior distribution
corresponding to the mixing function h is minimax provided that β satisfies*

$$- (p - 2)\left[\frac{E[V^{-p/2+1}]}{E[V]E[V^{-p/2}]} - \frac{1}{2} \right] \leq \beta. \tag{5.17}$$

For priors with mixing distribution h satisfying (5.16) and (5.17) an argument as
in Maruyama (2003a) using Brown (1979) and a Tauberian theorem suggests that
the resulting generalized Bayes estimator is admissible if $\beta \leq 0$. Maruyama and
Takemura (2008) have verified this under additional conditions which imply, in the
setting of Theorem 5.6, that $E_\theta[\|X\|^3] < \infty$.

As an illustration assume that the sampling distribution is a p-variate Student-
t with n_0 degrees of freedom which corresponds to the inverse gamma mixing
density $(n_0/2, n_0/2)$, that is, to $g(v) \propto v^{-(n_0+2)/2}\exp(-n_0/2v)$. Let the prior be
a Student-t distribution with n degrees of freedom, that is, with mixing density
$h(t) \propto t^{-(n+2)/2}\exp(-n/2t)$. It is clear that Conditions (5.14) and (5.15) are
satisfied with $n_0 \geq 7$. It is also clear that Condition (5.16) holds for any $\alpha < 1$.
Finally a simple calculation shows that

$$\frac{E[V^{-p/2+1}]}{E[V]E[V^{-p/2}]} = \frac{n_0 - 2}{p + n_0 - 2}$$

so that Condition (5.17) reduces to

$$n \leq (p - 2)\left[\frac{2(n_0 - 2)}{p + n_0 - 2} - 1 \right] - 2.$$

Note that, as $n > 0$, this condition holds if and only if $p \geq 5$ and

$$n_0 \geq 3 + p\frac{p}{p - 4}.$$

Other examples (including generalized priors) can be found in Fourdrinier et al.
(2008).

In the following, we consider broader classes of spherically symmetric dis-
tributions which are not restricted to variance mixtures of normals. Minimaxity
of generalized Bayes estimators is obtained for unimodal spherically symmetric
superharmonic priors $\pi(\|\theta\|^2)$ under the additional assumption that the Laplacian
of $\pi(\|\theta\|^2)$ is a nondecreasing function of $\|\theta\|^2$. The results presented below are

derived in Fourdrinier and Strawderman (2008a). An interesting feature is that their approach does not rely on the Baranchik representation used in Maruyama (2003a) and Fourdrinier et al. (2008). Note, however, that the superharmonicity property of the priors implies that the corresponding Bayes estimators cannot be proper (see Theorem 3.2).

First note that, for any prior $\pi(\theta)$, the Bayes estimator in (5.10) can be written as

$$\delta_\pi(X) = X + \frac{\nabla M(X)}{m(X)} \tag{5.18}$$

where, for any $X \in \mathbb{R}^p$,

$$M(x) = \int_{\mathbb{R}^p} F(\|x - \theta\|^2)\, \pi(\theta)\, d\theta$$

with F given in (5.5). Thus $\delta_\pi(X)$ has the general form $\delta_\pi(X) = X + g(X)$ (with $g(X) = \nabla M(X)/m(X)$). If the density $f(\|x - \theta\|^2)$ is as in Sect. 5.2.1, that is, such $F(t)/f(t) \geq c > 0$ for some fixed positive constant c, then Corollary 5.2 applies and $\delta_\pi(X) = X + g(X) = X + \nabla M(X)/m(X)$ is minimax provided, for any $x \in \mathbb{R}^p$,

$$2\, c \,\operatorname{div} g(x) + \|g(x)\|^2 \leq 0.$$

In particular, it follows that if

$$2\, c\, \frac{\Delta M(x)}{m(x)} - 2\, c\, \frac{\nabla M(x) \cdot \nabla m(x)}{m^2(x)} + \frac{\|\nabla M(x)\|^2}{m^2(x)} \leq 0 \tag{5.19}$$

and

$$E_\theta\left[\left\|\frac{\nabla M(X)}{m(X)}\right\|^2\right] < \infty,$$

δ_π is minimax.

For a spherically symmetric prior $\pi(\|\theta\|^2)$, the main result of Fourdrinier and Strawderman (2008a) is the following.

Theorem 5.7 *Assume that X has a spherically symmetric distribution in $\mathbb{R}^p$ with density $f(\|x - \theta\|^2)$. Assume that $\theta \in \mathbb{R}^p$ has a superharmonic prior $\pi(\|\theta\|^2)$ such that $\pi(\|\theta\|^2)$ is nonincreasing and $\Delta\pi(\|\theta\|^2)$ is nondecreasing in $\|\theta\|^2$. Assume also that*

$$E_\theta\left[\left\|\frac{\nabla M(X)}{m(X)}\right\|^2\right] < \infty.$$

Then the Bayes estimator δ_π is minimax under quadratic loss provided that $f(t)$ is log-convex, $c = \frac{F(0)}{f(0)} > 0$ and

$$\int_0^\infty f(t)t^{p/2}dt \le 4c \int_0^\infty -f'(t)t^{p/2}dt < \infty. \tag{5.20}$$

To prove Theorem 5.7 we need some preliminary lemmas whose proofs are given in Appendix A.9. Note first that it follows from the spherical symmetry of π that, for any $x \in \mathbb{R}^p$, $m(x)$ and $M(x)$ are functions of $t = \|x\|^2$. Then, setting

$$m(x) = m(t) \quad \text{and} \quad M(x) = M(t),$$

we have

$$\nabla m(x) = 2m'(t)\,x \quad \text{and} \quad \nabla M(x) = 2\,M'(t)\,x. \tag{5.21}$$

Lemma 5.5 *Assume that $\pi'(t) \le 0$, for any $t \ge 0$. Then we have $M'(t) \le 0$, for any $t \ge 0$.*

Lemma 5.6 *For any $x \in \mathbb{R}^p$,*

$$x \cdot \nabla m(x) = -2 \int_0^\infty H(u,t)\, u^{p/2}\, f'(u)\, du$$

and

$$x \cdot \nabla M(x) = \int_0^\infty H(u,t)\, u^{p/2}\, f(u)\, du$$

where, for $u \ge 0$ and for $t \ge 0$,

$$H(u,t) = \lambda(B) \int_{B_{\sqrt{u},x}} x \cdot \theta\, \pi'(\|\theta\|^2)\, d\mathcal{V}_{\sqrt{u},x}(\theta) \tag{5.22}$$

and $\mathcal{V}_{\sqrt{u},x}$ is the uniform distribution on the ball $B_{\sqrt{u},x}$ of radius $\sqrt{u}$ centered at x and $\lambda(B)$ is the volume of the unit ball.

Lemma 5.7 *For any $t \ge 0$, the function $H(u,t)$ in (5.22) is nondecreasing in u provided that $\Delta\pi(\|\theta\|^2)$ is nondecreasing in $\|\theta\|^2$.*

Lemma 5.8 *Let $h(\|\theta - x\|^2)$ be a unimodal density and let $\psi(\theta)$ be a symmetric function. Then*

$$\int_{\mathbb{R}^p} x \cdot \theta\, \psi(\theta)\, h(\|\theta - x\|^2)\, d\theta \ge 0$$

as soon as ψ is nonnegative.

Proof (Proof of Theorem 5.7) By the superharmonicity of $\pi(\|\theta\|^2)$, we have $\Delta M(x) \leq 0$ for all $x \in \mathbb{R}^p$ so that by (5.19), it suffices to prove that

$$-2\,c\,\nabla M(x) \cdot \nabla m(x) + \|\nabla M(x)\|^2 \leq 0 \qquad (5.23)$$

for all $x \in \mathbb{R}^p$. Since m and M are spherically symmetric, by (5.21), (5.23) reduces to $-2cM'(t)m'(t) + \left(M'(t)\right)^2 \leq 0$ where $t = \|x\|^2$. Since $M'(t) \leq 0$ by Lemma 5.5, (5.23) reduces to $-2cm'(t) + M'(t) \geq 0$ or, by (5.21), to $-2\,c\,x \cdot \nabla m(x) + x \cdot \nabla M(x) \geq 0$ or, by Lemma 5.6, to

$$4cE\left[H(u,t)\frac{f'(u)}{f(u)}\right] + E[h(u,t)] \geq 0, \qquad (5.24)$$

where E denotes the expectation with respect to the density proportional to $u^{p/2}f(u)$. Since, by assumption, $\Delta\pi(\|\theta\|^2)$ is nondecreasing in $\|\theta\|^2$, $H(u,t)$ is nondecreasing in u by Lemma 5.7. Furthermore $f'(u)/f(u)$ is nondecreasing by log-convexity of f so that (5.16) is satisfied as soon as

$$4\,c\,E[H(u,t)]\,E\left[\frac{f'(u)}{f(u)}\right] + E[H(u,t)] \geq 0. \qquad (5.25)$$

Finally, as $\pi'(\|\theta\|^2) \leq 0$ by assumption, Lemma 5.2 guarantees that $H(u,t) \leq 0$ (note that $V_{\sqrt{u},x}$ has a unimodal density) and hence (5.25) reduces to

$$4cE\left[\frac{f'(u)}{f(u)}\right] + 1 \leq 0$$

which is equivalent to (5.20).   □

Several examples of priors and sampling distributions which satisfy the assumptions of Theorem 5.7 are given in Fourdrinier and Strawderman (2008a). We briefly summarize these.

Example 5.8 (Priors related to the fundamental harmonic prior) Let $\pi(\|\theta\|^2) = \left(\dfrac{1}{A + \|\theta\|^2}\right)^c$ with $A \geq 0$ and $0 \leq c \leq \frac{p}{2} - 1$.

Example 5.9 (Mixtures of priors) Let $(\pi_\alpha)_{\alpha \in A}$ be a family of priors such that the assumptions of Theorem 5.7 are satisfied for any $\alpha \in A$. Then any mixture of the form $\displaystyle\int_A \pi_\alpha(\|\theta\|^2)\,dH(\alpha)$ where H is a probability measure on A satisfies these assumptions as well. For instance, Example 5.8 with $c = 1$, $p \geq 4$, $A = \alpha$ and the gamma density $\alpha \longmapsto \dfrac{\beta^{1-v}}{\Gamma(1-v)}\alpha^{-v}e^{-\beta\alpha}$ with $\beta > 0$ and $0 < v < 1$ leads to the prior

$$\|\theta\|^{-2-v}\,e^{\beta\|\theta\|^2}\,\Gamma(v, \beta\|\theta\|^2),$$

where

$$\Gamma(v, y) = \int_y^\infty e^{-x} x^{v-1} \, dx$$

is the complement of the incomplete gamma function.

Example 5.10 (Variance mixtures of normals) Let

$$\pi(\|\theta\|^2) = \int_O^\infty \left(\frac{u}{2\pi}\right)^{p/2} \exp\left(\frac{-u\|\theta\|^2}{2}\right) h(u) du$$

a mixture of normals with respect to the inverse of the variance . As soon as, for any $u > 0$,

$$\frac{uh'(u)}{h(u)} \leq -2,$$

the prior $\pi(\|\theta\|^2)$ satisfies the assumptions of Theorem 5.7. Note that the priors in Example 5.10 arise as such a mixture with $h(u) \propto \alpha u^{k-p/2-1} \exp(-A/2u)$.

Other examples can be given and a constructive approach is proposed in Fourdrinier and Strawderman (2008a).

We now give examples of sampling distributions which satisfy the assumptions of Theorem 5.7.

Example 5.11 (Variance mixtures of normals) Let

$$f(t) = (2\pi)^{-p/2} \int_0^\infty v^{-p/2} \exp\left(-\frac{t}{2v}\right) h(v) \, dv$$

where h is a mixing density and let V be a nonnegative random variable with density proportional to $f(t)$. If $E[V^{-p/2}] < \infty$ and $E[V] E[V^{-p/2}]/E[V^{-p/2+1}] < 2$ then the sampling density f satisfies the assumptions of Theorem 5.7.

Example 5.12 (Densities proportional to $e^{-\alpha t^\beta}$) Let

$$f(t) = K e^{-\alpha t^\beta}$$

where $\alpha > 0$, $\frac{1}{2} < \beta \leq 1$ and K is the normalizing constant. Then the sampling density f satisfies the assumptions of Theorem 5.7 as soon as β is in a neighborhood of the form $]1 - \epsilon, 1]$ with $\epsilon > 0$. However, note that these are not satisfied when $\beta = 1/2$.

Fourdrinier and Strawderman (2008a) give other examples with densities proportional to $e^{-\alpha t + \beta \varphi(t)}$ where φ is a convex function.

5.5   Shrinkage Estimators for Concave Loss

In this section we consider improved shrinkage estimators for loss functions that are concave functions of squared error loss. The basic results are due to Brandwein and Strawderman (1980, 1991b) and we largely follow the method of proof in the later paper. The general nature of the main result is that (under mild conditions) if an estimator can be shown to dominate X under squared error loss then the same estimator, with a suitably altered shrinkage constant, will dominate X for a loss which is a concave function of squared error loss.

Let X have a spherically symmetric distribution around θ, and let $g(X)$ be a weakly differentiable function. The estimators considered are of the form

$$\delta(X) = X + ag(X). \tag{5.26}$$

The loss functions are of the form

$$L(\theta, \delta) = \ell(||\delta - \theta||^2), \tag{5.27}$$

where $\ell(\cdot)$ is a differentiable nonnegative, nondecreasing concave function (so that, in particular $\ell'(\cdot) \geq 0$).

One basic tool needed for the main result is Theorem 5.5, and the other is the basic property of the concave function $\ell(\cdot)$ that $\ell(t + a) \leq \ell(t) + a\ell'(t)$.

The following result shows that shrinkage estimators that improve on X for squared error loss also improve on X for concave loss provided the shrinkage constant is adjusted properly.

Theorem 5.8 (Brandwein and Strawderman 1991a)   *Let X have a spherically symmetric distribution around θ, let $g(X)$ be a weakly differentiable function, and let the loss be given by (5.27).*

Suppose there exists a subharmonic function $h(\cdot)$ such that $E_{\theta,R}[R^2 h(U)]$ is nonincreasing where $U \sim \mathcal{U}_{R,\theta}$. Furthermore suppose that the function $g(\cdot)$ satisfies $E_\theta^*[||g(X)||^2] < \infty$ and also satisfies

(1) div $g(x) \leq h(x)$, *for any $x \in \mathbb{R}^p$,*
(2) $||g(x)||^2 + 2h(x) \leq 0$, *for any $x \in \mathbb{R}^p$, and*
(3) $0 \leq a \leq \frac{1}{pE_0^*(1/||X||^2)}$,

where E_θ^* refers to the expectation with respect to the distribution whose Radon-Nikodyn derivative with respect to the distribution of X is proportional to $\ell'(||X - \theta||^2)$.

Then $\delta(X) = X + ag(X)$ is minimax. Also $\delta(X)$ dominates X provided $g(\cdot)$ is non-zero with positive probability and strict inequality holds with positive probability in (1) or (2), or both inequalities are strict in (3).

Proof Note, by concavity of $\ell(\cdot)$ and the usual identity

$$
\begin{aligned}
R(\theta, \delta) &= E_\theta[\ell(||\delta(X) - \theta||^2)] \\
&\leq E_\theta[\ell(||X - \theta||^2)] \\
&\quad + E_\theta[\ell'(||X - \theta||^2)(a^2||g(X)||^2 + 2a(X - \theta)'g(X))].
\end{aligned}
$$

Hence, the difference in risk, $R(\theta, \delta) - R(\theta, X)$ is bounded by

$$
\begin{aligned}
R(\theta, \delta) - R(\theta, X) &\leq E_\theta[\ell'(||X - \theta||^2)(a^2||g(X)||^2 + 2a(X - \theta)'g(X))] \\
&= E_\theta^*[(a^2||g(X)||^2 + 2a(X - \theta)'g(X))] \\
&\leq 0,
\end{aligned}
$$

by Theorem 5.5 applied to the distribution corresponding to E_θ^*.   □