Chapter 7
Restricted Parameter Spaces

7.1 Introduction

In this chapter, we will consider the problem of estimating a location vector which is constrained to lie in a convex subset of $\mathbb{R}^P$. Estimators that are constrained to a set should be constrasted to the shrinkage estimators discussed in Sect.2.4.4 where one has “vague knowledge” that a location vector is in or near the specified set and consequently wishes to shrink toward the set but does not wish to restrict the estimator to lie in the set. Much of the chapter is devoted to one of two types of constraint sets, balls, and polyhedral cones. However, Sect.7.2 is devoted to general convex constraint sets and more particularly to a striking result of Hartigan (2004) which shows that in the normal case, the Bayes estimator of the mean with respect to the uniform prior over any convex set, $\mathscr{C}$, dominates X for all $\theta \in \mathscr{C}$ under the usual quadratic loss $\|\delta - \theta\|^2$.

Section 7.3 considers the situation where X is normal with a known scale but the constraint set is a ball, B, of known radius centered at a known point in $\mathbb{R}^p$. Here again, a natural estimator to dominate is the projection onto the ball $P_B X$. Hartigan’s result of course applies and shows that the Bayes estimate corresponding to the uniform prior dominates X, but a finer analysis lead to domination over $P_B X$ (provided the radius of the ball is not too large relative to the dimension) by the Bayes estimator corresponding to the uniform prior on the sphere of the same radius.

Section 7.4 will consider estimation of a normal mean vector restricted to a polyhedral cone, $\mathscr{C}$, in the normal case under quadratic loss. Both the cases of known and unknown scale are treated. Special methods need to be developed to handle this restriction because the shrinkage functions considered are not necessarily weakly differentiable. Hence the methods of Chap.4 are not directly applicable. A version of Stein’s lemma is developed for positively homogeneous sets which allows the analysis to proceed.

In general, if the constraint set, $\mathscr{C}$, is convex, a natural alternative to the UMVUE X, is $P_c X$ the projection of X onto $\mathscr{C}$. Our methods lead to Stein type shrinkage

D. Fourdrinier et al., *Shrinkage Estimation*, Springer Series in Statistics,
https://doi.org/10.1007/978-3-030-02185-6_7

estimators that shrink $P_c X$ which dominate $P_c X$, and hence X itself, when $\mathscr{C}$ is a polyhedral cone.

Section 7.5 is devoted to the case of a general spherically symmetric distribution with a residual vector when the mean vector is restricted to a polyhedral cone. As in Sect.7.4, the potential nondifferentiability of the shrinkage factors is a complication. We develop a general method that allows the results of Sect.7.4 for the normal case to be extended to the general spherically symmetric case as long as a residual vector is available. This method also allows for an alternative development of some of the results of Chap.6 that rely on an extension of Stein's lemma to the general spherical case.

7.2 Normal Mean Vector Restricted to a Convex Set

In this section, we treat the case $X \sim \mathscr{N}_p(\theta, \sigma^2 I_p)$ where σ^2 is known and where the unknown mean θ is restricted to lie in a convex set $\mathscr{C} \subseteq \mathbb{R}^p$ (with nonempty interior and sufficiently regular boundary), and where the loss is $L(\theta, \delta) = \|\delta - \theta\|^2$. We show that the (generalized) Bayes estimator with respect to the uniform prior distribution on $\mathscr{C}$, say $\pi(\theta) = \mathbb{1}_{\mathscr{C}}(\theta)$, dominates the usual (unrestricted) estimator $\delta_0(X) = X$. At this level of generality the result is due to Hartigan (2004) although versions of the result (in $\mathbb{R}^1$) date back to Katz (1961). We follow the discussion in Marchand and Strawderman (2004).

Theorem 7.1 (Hartigan 2004) *Let $X \sim \mathscr{N}_p(\theta, \sigma^2 I_p)$ with σ^2 known and $\theta \in \mathscr{C}$, a convex set with nonempty interior and sufficiently regular boundary $\partial\mathscr{C}$ ($\partial\mathscr{C}$ is Lipshitz of order 1 suffices). Then the Bayes estimator, $\delta_U(X)$ with respect to the uniform prior on $\mathscr{C}$, $\pi(\theta) = \mathbb{1}_{\mathscr{C}}(\theta)$, dominates $\delta_0(X) = X$ with respect to quadratic loss.*

Proof Without loss of generality, assume $\sigma^2 = 1$. Recall from (1.20) that the form of the Bayes estimator is $\delta_U(X) = X + \nabla m(X)/m(X)$ where, for any $x \in \mathbb{R}^p$,

$$m(x) \propto \int_{\mathscr{C}} \exp\left(-\frac{1}{2}\|x - \nu\|^2\right) d\nu.$$

The difference in risk between δ_U and δ_0 is $R(\theta, \delta_U) - R(\theta, \delta_0)$

$$\begin{aligned} R(\theta, \delta_U) - R(\theta, \delta_0) &= E_\theta\left[\left\|X + \frac{\nabla m(X)}{m(X)} - \theta\right\|^2 - \|X - \theta\|^2\right] \\ &= E_\theta\left[\frac{\|\nabla m(X)\|^2}{m^2(X)} + 2\frac{\nabla m(X)^{\mathrm{T}}(X - \theta)}{m(X)}\right]. \end{aligned} \tag{7.1}$$

Hartigan's clever development proceeds by applying Stein's Lemma 2.3 to only half of the cross product term in order to cancel the squared norm term in the above.

Indeed, since

$$E_\theta\left[(X-\theta)^{\mathrm{T}}\left(\frac{\nabla m(X)}{m(X)}\right)\right] = E_\theta\left[\operatorname{div}\left(\frac{\nabla m(X)}{m(X)}\right)\right]$$
$$= E_\theta\left[\frac{\Delta m(X)}{m(X)} - \frac{\|\nabla m(X)\|^2}{m^2(X)}\right],$$

(7.1) then becomes

$$R(\theta,\delta_U) - R(\theta,\delta_0) = E_\theta\left[\frac{\Delta m(X) + (X-\theta)^{\mathrm{T}}\nabla m(X)}{m(X)}\right]$$
$$= E_\theta\left[\frac{H(X,\theta)}{m(X)}\right] \tag{7.2}$$

with $H(x,\theta) = \Delta m(x) + (x-\theta)^{\mathrm{T}}\nabla m(x)$. Hence it suffices to show $H(x,\theta) \le 0$ for all $\theta \in \mathscr{C}$ and $x \in \mathbb{R}^p$. Using the facts that

$$\nabla_x \exp\left(-\frac{1}{2}\|x-\nu\|^2\right) = -\nabla_\nu \exp\left(-\frac{1}{2}\|x-\nu\|^2\right)$$

and

$$\Delta_x \exp\left(-\frac{1}{2}\|x-\nu\|^2\right) = \Delta_\nu \exp\left(-\frac{1}{2}\|x-\nu\|^2\right),$$

it follows that

$$H(x,\theta) \propto \Delta_x \int_{\mathscr{C}} \exp\left(-\frac{1}{2}\|x-\nu\|^2\right) d\nu + (x-\theta)'\nabla_x \int_{\mathscr{C}} \exp\left(\frac{1}{2}\|x-\nu\|^2\right) d\nu$$
$$= \int_{\mathscr{C}} \left[\Delta_\nu \exp\left(-\frac{1}{2}\|x-\nu\|^2\right) - (x-\theta)'\nabla_\nu \exp\left(-\frac{1}{2}\|x-\nu\|^2\right)\right] d\nu$$
$$= \int_{\mathscr{C}} \operatorname{div}_\nu \left[\nabla_\nu \exp\left(-\frac{1}{2}\|x-\nu\|^2\right) - (x-\theta)\exp\left(-\frac{1}{2}\|x-\nu\|^2\right)\right] d\nu$$
$$= \int_{\mathscr{C}} \operatorname{div}_\nu \left[(\theta-\nu)\exp\left(-\frac{1}{2}\|x-\nu\|^2\right)\right] d\nu.$$

By Stokes' theorem (see Sect.A.5 of the Appendix) this last expression can be expressed as

$$\int_{\partial\mathscr{C}} \eta^{\mathrm{T}}(\nu)(\theta-\nu)\exp\left(-\frac{1}{2}\|x-\nu\|^2\right) d\sigma(\nu)$$

where $\eta(\nu)$ is the unit outward normal to $\partial\mathscr{C}$ at ν and σ is the surface area Lebesgue measure on $\partial\mathscr{C}$. Finally, since $\mathscr{C}$ is convex and $\theta \in \mathscr{C}$, the angle between $\eta(\nu)$ and $\theta - \nu$ is obtuse for $\nu \in \partial\mathscr{C}$ and so $\eta^{\mathrm{T}}(\nu)(\theta - \nu) \leq 0$, for all $\theta \in \mathscr{C}$ and $\nu \in \partial\mathscr{C}$, which implies the risk difference in (7.2) is nonpositive. □

Note that, if θ is in the interior of $\mathscr{C}$, $\mathscr{C}^0$, then $\eta'(\nu)(\theta - \nu)$ is strictly negative for all $\nu \in \partial\mathscr{C}$, and hence, $R(\theta, \delta_U) - R(\theta\delta_0) < 0$ for all $\theta \in \mathscr{C}^0$. However, if $\mathscr{C}$ is a pointed cone at θ_0, then $\eta^{\mathrm{T}}(\nu)(\theta_0 - \nu) \equiv 0$ for all $\nu \in \partial\mathscr{C}$ and $R(\theta_0, \delta_U) = R(\theta_0, \delta_0)$.

Note also that, if $\mathscr{C}$ is compact, the uniform prior on $\mathscr{C}$ is proper, and hence, $\delta_U(X)$ not only dominates $\delta_0(X)$ (on $\mathscr{C}$) but is also admissible for all p. On the other hand, if $\mathscr{C}$ is not compact, it is often (typically for $p \geq 3$) the case that δ_U is not admissible and alternative shrinkage estimators may be desirable.

Furthermore, it may be argued in general, that a more natural basic estimator which one should seek to dominate is $P_c X$, the projection of X onto $\mathscr{C}$ which is the MLE. We consider this problem for the case where $\mathscr{C}$ is a ball in Sect.7.3 and where $\mathscr{C}$ is a polyhedral cone in Sect.7.4. □

7.3 Normal Mean Vector Restricted to a Ball

When the location parameter $\theta \in \mathbb{R}^p$ is restricted, the most common constraint is a ball, that is, to a set for which $\|\theta\|$ is bounded above by some constant R. In this setting Bickel (1981) noted that, by an invariance argument and analyticity considerations, the minimax estimate is Bayes with respect to a unique spherically symmetric least favorable prior distribution concentrating on a finite number of spherical shells. This result extends what Casella and Strawderman (1981) obtained in the univariate case. Berry (1990) specified that when R is small enough, the corresponding prior is supported by a single spherical shell. In this section we address the issues of minimax estimation under a ball constraint.

Let $X \sim \mathscr{N}_p(\theta, \sigma^2 I_p)$, with unknown mean $\theta = (\theta_1, \ldots, \theta_p)$ and known σ^2, and with the additional information that $\sum_{i=1}^{p}(\theta_i - \tau_i)^2/\sigma^2 \leq R^2$ where $\tau_1, \ldots, \tau_p, \sigma^2, R$ are known. From a practical point of view, a constraint as the one above signifies that the squared standardized deviations $|(\theta_i - \tau_i)/\sigma|^2$ are on average bounded by R^2/p. We are concerned here with estimating θ under quadratic loss $L(\theta, \delta) = \|\delta - \theta\|^2$. Without loss of generality, we proceed by setting $\sigma^2 = 1$ and $\tau_i = 0, i = 1, \ldots, p$, so that the constraint is the ball $B_R = \{\theta \in \mathbb{R}^p \mid \|\theta\| \leq R\}$.

Since the usual minimax estimators take on values outside of B_R with positive probability, they are neither admissible nor minimax when θ is restricted to B_R. The argument given by Berry (1990) is the following. As for inadmissibility, it can be seen that these estimators are dominated by their truncated versions. Thus, the unbiased estimator X is dominated by the MLE $\delta_{MLE}(X) = (R/\|X\| \wedge 1)X$ (which is the truncation of X on B_R). Now, if an estimator which takes on values outside of B_R with positive probability were minimax, its truncated version would be minimax as well, with a strictly smaller risk. This is a contradiction since the risk function is continuous and attains its maximum in B_R. Further $\delta_{MLE}(X)$ is not admissible since, due to its non differentiability, it is not a generalized Bayes estimator. See Sect.3.4. For further discussions on this issue related to inadmissibility of estimators taking values on the boundary of a convex parameter space, see the review paper of Marchand and Strawderman (2004) and the monograph of van Eeden (2006).

As alternative estimators to $\delta_{MLE}(X)$, the Bayes estimators are attractive since they may have good frequentist performances in addition to their Bayesian property. Two natural estimators are the Bayes estimators with respect to the uniform distribution on the ball B_R and the uniform distribution on its boundary, the sphere $S_R = \{\theta \in \mathbb{R}^p \mid \|\theta\| = R\}$. We will see that the latter is particularly interesting.

The model is dominated by the Lebesgue measure on $\mathbb{R}^p$ and has likelihood L given by

$$\forall x \in \mathbb{R}^p \quad \forall \theta \in \mathbb{R}^p \quad L(x, \theta) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\|x - \theta\|^2\right). \tag{7.3}$$

Hence, if the prior distribution is the uniform distribution $\mathcal{U}_R$ on the sphere S_R, the marginal distribution has density m with respect to the Lebesgue measure on $\mathbb{R}^p$ given by

$$\begin{aligned} \forall x \in \mathbb{R}^p \quad m(x) &= \int_{S_R} L(x, \theta) d\mathcal{U}_R(\theta) \\ &= \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{1}{2}\|x\|^2\right) \exp\left(-\frac{1}{2}R^2\right) \int_{S_R} \exp(x^{\mathrm{T}}\theta) d\mathcal{U}_R(\theta), \end{aligned} \tag{7.4}$$

after expanding the likelihood in (7.3). Also, the posterior distribution given $x \in \mathbb{R}^p$ has density $\pi(\theta|x)$ with respect to the prior distribution $\mathcal{U}_R$ given by

$$\forall \theta \in S_R \quad \pi(\theta|x) = \frac{L(x, \theta)}{m(x)} = \frac{\exp(x^{\mathrm{T}}\theta)}{\int_{S_R} \exp(x^{\mathrm{T}}\theta) d\mathcal{U}_R(\theta)}, \tag{7.5}$$

thanks to the second expression of $m(x)$ in (7.4). As the loss is quadratic, the Bayes estimator δ_R is the posterior mean, that is,

$$\forall x \in S_R \quad \delta_R(x) = \int_{S_R} \theta \pi(\theta|x) d\mathcal{U}_R(\theta) = \frac{\int_{S_R} \theta \exp(x^{\mathrm{T}}\theta) d\mathcal{U}_R(\theta)}{\int_{S_R} \exp(x^{\mathrm{T}}\theta) d\mathcal{U}_R(\theta)}. \tag{7.6}$$

The Bayes estimator in (7.6) can be expressed through the modified Bessel function I_ν, solutions of the differential equation $z^2\varphi''(z) + z\varphi'(z) - (z^2 + \nu^2)\varphi(z) = 0$ with $\nu \geq 0$. More precisely, we will use the integral representation of the modified Bessel function

$$I_\nu(z) = \frac{(z/2)^\nu}{\pi^{1/2}\Gamma(\nu+1/2)} \int_0^\pi \exp(z\cos t)\sin^{2\nu} t dt \tag{7.7}$$

from which we may derive the formula

$$I'_\nu(z) = \frac{\nu}{z} I_\nu(z) + I_{\nu+1}(z). \tag{7.8}$$

Using the parametrization in terms of polar coordinates, we can see from the proof of Lemma 1.4 that, for any function h integrable with respect to $\mathscr{U}_R$,

$$\int_{S_R} h(\theta) d\mathscr{U}_R(\theta) = \frac{1}{\sigma(S)} \int_V h(\varphi_R(t_1,\ldots,t_{p-1})) \sin^{p-2} t_1 \ldots \sin t_{p-2} dt_1,\ldots,dt_{p-1}$$

where $\sigma(S)$ is the area measure of the unit sphere and, as in (1.9), where $V = (0,\pi)^{p-2} \times (0,2\pi)$ and for $(t_1,\ldots,t_{n-1}) \in V$, $\varphi_R(t_1,\ldots,t_{p-1}) = (\theta_1,\ldots,\theta_p)$ with

$$\begin{aligned}
\theta_1 &= R\sin t_1 \sin t_2 \ldots \sin t_{p-2} \sin t_{p-1} \\
\theta_2 &= R\sin t_1 \sin t_2 \ldots \sin t_{p-2} \cos t_{p-1} \\
\theta_3 &= R\sin t_1 \sin t_2 \ldots \cos t_{p-2} \\
&\vdots \\
\theta_{p-1} &= R\sin t_1 \cos t_2 \\
\theta_p &= R\cos t_1.
\end{aligned}$$

Setting $h(\theta) = \exp(x^{\mathrm{T}}\theta)$ and choosing the angle between x and $\theta \in S_R$ as the first angle t_1 gives

$$\int_{S_R} \exp(x^{\mathrm{T}}\theta) d\mathscr{U}_R(\theta) = \frac{K}{\sigma(S)} \int_0^\pi \exp(\|x\| R\cos t_1) \sin^{p-2} t_1 dt_1$$

where

$$K = \int_{V^{\mathrm{T}}} \sin^{p-3} t_2 \ldots \sin t_{p-2} dt_2,\ldots,dt_{p-1}$$

with $V^{\mathrm{T}} = (0,\pi)^{p-3} \times (0,2\pi)$.

Therefore, according to (7.7), the marginal in (7.4) is proportional to

$$m_R(\|x\|) = \exp\left(-\frac{1}{2}\|x\|^2\right) \exp\left(-\frac{1}{2}R^2\right) \frac{I_{(p-2)/2}(\|x\|R)}{(\|x\|R)^{(p-2)/2}}, \tag{7.9}$$

the proportionality constant being independent of R. Then the Bayes estimator $\delta_R(X)$ can be obtained thanks to (1.20), that is, for any $x \in \mathbb{R}^p$,

$$\delta_R(x) = x + \nabla \log m(x).$$

As only the quantities depending on x matter, we have using (7.9), for any $x \in \mathbb{R}^p$,

$$\begin{aligned}
\delta_R(x) &= x + \nabla \log m_R(\|x\|) \\
&= -\frac{p-2}{2}\frac{\nabla(\|x\|R)}{\|x\|R} + \frac{\nabla I_{(p-2)/2}(\|x\|R)}{I_{(p-2)/2}(\|x\|R)} \\
&= -\frac{p-2}{2}\frac{x}{\|x\|^2} + \frac{\frac{x}{\|x\|}\left[\frac{p-2}{2\|x\|} I_{(p-2)/2}(\|x\|R) + R I_{p/2}(\|x\|R)\right]}{I_{(p-2)/2}(\|x\|R)} \\
&= \frac{R I_{p/2}(\|x\|R)}{I_{(p-2)/2}(\|x\|R)}\frac{x}{\|x\|},
\end{aligned} \tag{7.10}$$

where (7.8) has been used for the second to last equality.

Thus, according to (7.10), the Bayes estimator is expressed through a ratio of modified Bessel functions, that is, denoting by $\rho_\nu(t) = I_{\nu+1}/I_\nu$ with $t > 0$ and $\nu > -1/2$,

$$\delta_R(x) = \frac{R}{\|x\|}\rho_{p/2-1}(R\|x\|)x. \tag{7.11}$$

Before proceeding, we give two results from Marchand and Perron (2001).

(i) For sufficiently small R, say $R \le c_1(p)$, all Bayes estimators δ_π with respect to an orthogonally invariant prior π (supported on B_R) dominate $\delta_{MLE}(X)$;
(ii) The Bayes estimator $\delta_R(X)$ with respect to the uniform prior on the sphere S_R dominates $\delta_{MLE}(X)$ whenever $R \le \sqrt{p}$.

Note that Marchand and Perron (2002) extend the result in (ii) showing that domination of $\delta_R(X)$ over $\delta_{MLE}(X)$ subsists for some $m_0(p)$ such that $m_0(p) \ge \sqrt{p}$ and for $R \le m_0(p)$.

Various other dominance results, such as those pertaining to a fully uniform prior on B_R and other absolutely continuous priors are also available from Marchand and Perron (2001), but we will focus here on results (i) and (ii) above, following Fourdrinier and Marchand (2010).

With respect to important properties of $\delta_R(X)$, we point out that it is the optimal equivariant estimator for $\theta \in S_R$, and thus necessarily improves upon $\delta_{MLE}(X)$ on S_R. Furthermore, $\delta_R(X)$ also represents the Bayes estimator which expands greatest, or shrinks the least towards the origin (i.e., $\|\delta_\pi\| \le \|\delta_R(X)\|$ for all π supported on B_R; Marchand and Perron 2001). Despite this, as expanded upon below, $\delta_R(X)$

still shrinks more than $\delta_{MLE}(X)$ whenever $R \leq \sqrt{p}$, but not otherwise with the consequence of increased risk at $\theta = 0$ and failure to dominate $\delta_{MLE}(X)$ for large R. With the view of seeking dominance for a wider range of values of R, for potentially modulating these above effects by introducing more (but not too much) shrinkage, we consider the class of uniform priors supported on spheres S_α of radius α; $0 \leq \alpha \leq R$; about the origin, and their corresponding Bayes estimators δ_α. The choice is particularly interesting since the amount of shrinkage is calibrated by the choice of α (as formalized below), with the two extremes $\delta_R(X) \equiv \delta_R(X)$, and $\delta_0 \equiv 0$ (e.g., prior degenerate at 0). Moreover, knowledge about dominance conditions for the estimators δ_α may well lead, through further analytical risk and unbiased estimates of risk comparisons (e.g., Marchand and Perron (2001), Lemma 5 and the Remarks that follow), to implications relative to other Bayes estimators such as the fully uniform on B_R prior Bayes estimator.

Using Stein's unbiased estimator of risk technique, Karlin's sign change arguments, and a conditional risk analysis, Fourdrinier and Marchand (2010) obtain, for a fixed (R, p), necessary and sufficient conditions on α for δ_α to dominate $\delta_{MLE}(X)$.

Theorem 7.2

(a) *An unbiased estimator of the difference in risks*

$$\Delta_\alpha(\|\theta\|) = R(\theta, \delta_\alpha) - R(\theta, \delta_{MLE}(X))$$

is given by $D_\alpha(\|X\|) = D_{\alpha,1}(\|X\|)\ \mathbb{1}[0 \leq \|X\| \leq R] + D_{\alpha,2}(\|X\|)\ \mathbb{1}[\|X\| > R]$*, with*

$$D_{\alpha,1}(r) = 2\alpha^2 + r^2 - 2p - 2\alpha r \rho_{p/2-1}(\alpha r) - \alpha^2 \rho^2_{p/2-1}(\alpha r), \textit{ and}$$

$$D_{\alpha,2}(r) = 2\alpha^2 - m^2 - \alpha^2 \rho^2_{p/2-1}(\alpha r) + 2Rr\{1 - \frac{\alpha}{R}\rho_{p/2-1}(\alpha r)\} - 2(p-1)\frac{R}{r}.$$

(b) *For* $p \geq 3$*, and* $0 \leq \alpha \leq R$*,* $D_\alpha(r)$ *changes signs as a function of* r *according to the order:* (i) $(-,+)$ *whenever* $\alpha \leq \sqrt{p}$*, and* (ii) $(+,-,+)$ *whenever* $\alpha > \sqrt{p}$*.*
(c) *For* $p \geq 3$ *and* $0 \leq \alpha \leq R$*, the estimator* δ_α *dominates* $\delta_{MLE}(X)$ *if and only if*

(i) $\Delta_\alpha(R) \leq 0$ *whenever* $\alpha \leq \sqrt{p}$ *or*
(ii) $\Delta_\alpha(0) \leq 0$ *and* $\Delta_\alpha(R) \leq 0$*, whenever* $\alpha > \sqrt{p}$*.*

Proof

(a) Writing $\delta_{MLE}(x) = x + g_{MLE}(x)$ with $g_{MLE}(x) = (R/r - 1)x\ \mathbb{1}_{[r>R]}$ (with $r = \|x\|$) note that g_{MLE} is weakly differentiable. Then we have

$$2\text{div} g_{MLE}(x) + \|g_{MLE}(x)\|^2 = \left\{2(p-1)\frac{R}{r} - 2p + (R-r)^2\right\} \mathbb{1}_{(R,\infty)}(r)$$

and, by virtue of Stein's identity, $R(\theta, \delta_{MLE}) = E_\theta[\eta_{MLE}(X)]$ with

$$\eta_{MLE}(x) = p\mathbb{1}_{[0,R]}(r) + \left\{2(p-1)\frac{R}{r} - p + (R-r)^2\right\} \mathbb{1}_{(R,\infty)}(r). \qquad (7.12)$$

Analogously, as derived by Berry (1990), the representations of δ_α and $\frac{d}{dt}\rho_\nu(t)$ given in (7.11) and Lemma A.8, along with (2.3), permit us to write $R(\theta, \delta_\alpha) = E_\theta[\lambda_\alpha(X)]$ with

$$\lambda_\alpha(x) = 2\alpha^2 + r^2 - p - 2\alpha r \rho_{p/2-1}(\alpha r) - \alpha^2 \rho^2_{p/2-1}(\alpha r). \qquad (7.13)$$

Finally, the given expression for the unbiased estimator $D_\alpha(\|X\|)$ follows directly from (7.12) and (7.13).

(b) We begin with three intermediate observations which are proven below.

(I) The sign changes of $D_{\alpha,1}(r)$; $r \in [0, R]$; are ordered according to one of the five following combinations: $(+)$, $(-)$, $(-,+)$, $(+,-)$, $(+,-,+)$;

(II) $\lim_{r\to R^+}\{D_\alpha(r)\} = \lim_{r\to R^-}\{D_\alpha(r)\} + 2$;

(III) the function $D_{\alpha,2}(r)$; $r \in (R,\infty)$ is either positive, or changes signs once from $-$ to $+$.

From properties (I), (II) and (III), we deduce that the sign changes of $D_\alpha(r)$ $r \in (0,\infty)$; an everywhere continuous function except for the jump discontinuity at R; are ordered according to one of the three following combinations: $(+)$, $(-,+)$, $(+,-,+)$. Now, recall that δ_α is a Bayes and admissible estimator of θ under squared error loss. Therefore, among the combinations above, $(+)$ is not possible since this would imply that δ_α is dominated by δ_{MLE} in contradiction with its admissibility. Finally, the two remaining cases are distinguished by observing that, $D_\alpha(0) = 2\alpha^2 - 2p \leq 0$ if and only if $\alpha \leq \sqrt{p}$.

Proof of (I): Begin by making use of Lemma A.8 to differentiate $D_{\alpha,1}$ and obtain:

$$r^{-1}\ D'_{\alpha,1}(r) = 2 - 2\frac{\alpha}{r}\rho_{p/2-1}(\alpha r) - 2\alpha^2\rho'_{p/2-1}(\alpha r) - 2\alpha^3\rho'_{p/2-1}(\alpha r)\frac{\rho_{p/2-1}(\alpha r)}{r}.$$

Since, the quantities $r^{-1}\rho_{p/2-1}(\alpha r)$ and $\rho'_{p/2-1}(\alpha r)$ are positive and decreasing in r by virtue of Lemma A.8, $r^{-1}\ D'_{\alpha,1}(r)$ is necessarily increasing in r, $r \in [0, R]$. Hence, $D'_{\alpha,1}(\cdot)$ has, on $[0, R]$, sign changes ordered as either: $(+)$, $(-)$, or $(-,+)$. Finally, observe as a consequence that $D_{\alpha,1}(\cdot)$ has at most two sign changes on $[0, R]$, and furthermore that, among the six possible combinations, the combination $(-,+,-)$ is not consistent with the sign changes of $D'_{\alpha,1}$.

Proof of (II): Follows by a direct evaluation of $D_{\alpha,1}(R)$ and $D_{\alpha,2}(R)$ which are given in part (a) of this lemma.

Proof of (III): First, one verifies from (7.13), part (a) of Lemma A.8, and part (c) of Lemma A.9 that $\lim_{r\to\infty} D_{\alpha,2}(r)$ is $+\infty$, for $\alpha < R$; and equal to $p-1$ if $\alpha = R$. Moreover, part (a) also permits us to express $D_{\alpha,2}(r)$; $r > R$; as $(1 - \frac{\alpha}{R}\ \rho_{p/2-1}(\alpha r))\ \sum_{i=1}^3 H_i(\alpha, r)$ with

$$H_1(\alpha, r) = 2rR\left\{1 - \frac{(1-\frac{\alpha^2}{R^2})R}{r\{1-\frac{\alpha}{R}\rho_{p/2-1}(\alpha r)\}}\right\},$$

$$H_2(\alpha, r) = \frac{-2(p-1)R}{r\,\{\,1-\frac{\alpha}{R}\rho_{p/2-1}(\alpha r)\,\}}$$

and

$$H_3(\alpha, r) = R^2 + \alpha R\rho_{p/2-1}(\alpha r).$$

Hence, to establish property (III), it will suffice to show that each one of the functions $H_i(\alpha, \cdot)$, $i = 1, 2, 3$, is increasing on (R, ∞) under the given conditions on (p, α, R). The properties of Lemma A.8 clearly demonstrate that $H_3(\alpha, \cdot)$ is increasing, and it is the same for $H_2(\alpha, \cdot)$ given also Lemmas A.8 and A.9 since

$$r(1-\frac{\alpha}{R}\rho_{p/2-1}(\alpha r)) = r(1-\rho_{p/2-1}(\alpha r)) + r\;(1-\frac{\alpha}{R})\;\rho_{p/2-1}(\alpha r).$$

Finally, for the analysis of $H_1(\alpha, r), r > R$, begin by differentiation and a rearrangement of terms to obtain

$$\frac{\partial}{\partial r}H_1(\alpha, r) \geq 0 \Leftrightarrow T(R) \geq 0$$

where, for $r > R \geq \alpha$,

$$T(R) = (R - \alpha\rho_{p/2-1}(\alpha r))^2 - \alpha^2(R^2-\alpha^2)\rho'_{p/2-1}(\alpha r).$$

But notice that $T(\alpha) = \alpha^2(1-\rho_{p/2-1}(\alpha r))^2 \geq 0$, and

$$\begin{aligned}\frac{1}{2}\frac{\partial T(R)}{\partial R} &= (R-\alpha\rho_{p/2-1}(\alpha r)) - R\alpha^2\rho'_{p/2-1}(\alpha r)\\ &\geq (\alpha - \alpha\rho_{p/2-1}(\alpha r)) - R\alpha^2\frac{1-\rho_{p/2-1}(\alpha r)}{\alpha r}\\ &= \alpha(1-\rho_{p/2-1}(\alpha r))\left(1-\frac{R}{r}\right)\\ &\geq 0,\end{aligned}$$

by Lemma A.9, part (b), since $r \geq R \geq \alpha$. The above establishes that $T(R) \geq T(\alpha) \geq 0$ for all $R \geq \alpha$, that $H_1(\alpha, r)$ increases in r, and completes the proof of the Theorem.

(c) The probability distribution of $\|X\|^2$ is $\chi^2_p(\lambda^2)$, so that the potential sign changes of $\Delta_\alpha(\lambda) = E_\lambda[D_\alpha(\|X\|)]$ are controlled by the variational properties of $D_\alpha(\cdot)$ in terms of sign changes (e.g., Brown et al. 1981). Therefore, in situation (i) with $\alpha \le \sqrt{p}$, it follows from part (b) of Lemma 7.2 that, as $\Delta_\alpha(\cdot)$ varies on $[0, \infty]$ (or $[0, R]$), the number of sign changes is at most one, and that such a change must be from $-$ to $+$. Therefore, since δ_α is admissible; and that the case $\Delta_\alpha(\lambda) \ge 0$ for all $\lambda \in [0, R]$ is not possible[1]; we must have indeed that $\Delta_\alpha(\cdot) \le 0$ on $[0, R]$ if and only if $\Delta_\alpha(R) \le 0$ establishing (i). A similar line of reasoning implies the result in (ii) as well. □

We refer to Fourdrinier and Marchand (2010) for other results. In particular, large sample determinations of these conditions are provided. Both cases where all such δ_α's, or no such δ_α's dominate δ_{MLE} are elicited. As a particular case, they establish that the boundary uniform Bayes estimator δ_R dominates δ_{MLE} if and only if $R \le k(p)$ with $\lim_{p\to\infty} k(p)/\sqrt{p} = \sqrt{2}$, improving on the previously known sufficient condition of Marchand and Perron (2001) for which $k(p) \ge \sqrt{p}$. Finally, they improve upon a universal dominance condition due to Marchand and Perron, by establishing that all Bayes estimators δ_π with π spherically symmetric and supported on the parameter space dominate δ_{MLE} whenever $R \le c_1(p)$ with $\lim_{p\to\infty} c_1(p)/\sqrt{p} = \sqrt{1/3}$.

See Marchand and Perron (2005) for analogous results for other spherically symmetric distributions including multivariate t.

Other significant contributions to the study of minimax estimation of a normal mean restricted to an interval or a ball of radius R, were given by Bickel (1981) and Levit (1981). These contributions consisted of approximations to the minimax risk and least favourable prior for large R under squared error loss. In particular, Bickel showed that for $p = 1$, as $R \to \infty$, the least favourable distributions rescaled to $[-1, 1]$ converge weakly to a distribution with density $\cos^2(\pi x/2)$, and that the minimax risks behave like $1 - \pi^2/(8R^2) + o(R^{-2})$. There is also a substantial literature on efficiency comparisons of minimax procedures and affine linear minimax estimators for various models, and restricted parameter spaces; see Donoho et al. (1990) and Johnstone and MacGibbon (1992) and the references therein.

Finally, we observe that the loss function plays a critical role. In the case where $p = 1$ and loss is absolute error $|d - \theta|$, $\delta_{MLE}(X)$ is admissible. See Isawa and Moritani (1997) and Kucerovsky et al. (2009).

7.4 Normal Mean Vector Restricted to a Polyhedral Cone

In this section, we consider first the case when $X \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ where σ^2 is known and θ is restricted to a polyhedral cone $\mathscr{C}$ and where the loss is $\|\delta - \theta\|^2$. Later in this Section, we will consider the case where σ^2 is unknown and, in

[1] The risks of δ_α and δ_{MLE} cannot match either, since a linear combination of these two distinct estimators would improve on δ_α.

Sect.7.5, the general spherically symmetric case with a residual vector. The reader is referred to Fourdrinier et al. (2006) for more details.

A natural estimator in this problem is $\delta_{\mathscr{C}}(X) = P_{\mathscr{C}} X$, the projection of X onto the cone $\mathscr{C}$. The estimator $\delta_{\mathscr{C}}$ is the MLE and dominates X which is itself minimax provided $\mathscr{C}$ has a nonempty interior. Our goal will be to dominate $\delta_{\mathscr{C}}$ and therefore also $\delta_0(X) = X$.

We refer the reader to Stoer and Witzgall (1970) and Robertson et al. (1988) for an extended discussion of polyhedral cones. Here is a brief summary. A polyhedral cone $\mathscr{C}$ is defined as the intersection of a finite number of half spaces, that is,

$$\mathscr{C} = \{x \mid a_i^{\mathrm{T}} x \leq 0, i = 1, \ldots, m\} \tag{7.14}$$

for n fixed vectors $a_1, \ldots, a_m$ in $\mathbb{R}^p$.

It is positively homogeneous, closed and convex, and, for each $x \in \mathbb{R}^p$, there exists a unique point $P_{\mathscr{C}} x$ in $\mathscr{C}$ such that $\|P_{\mathscr{C}} x - x\| = \inf_{y \in \mathscr{C}} \|y - x\|$.

We assume throughout that $\mathscr{C}$ has a nonempty interior, $\mathscr{C}^o$ so that $\mathscr{C}$ may be partitioned into $\mathscr{C}_i$, $i = 0, \ldots, m$, where $\mathscr{C}_0 = \mathscr{C}^o$ and $\mathscr{C}_i$, $i = 1, \ldots, m$, are the relative interiors of the proper faces of $\mathscr{C}$. For each set $\mathscr{C}_i$, let $D_i = P_{\mathscr{C}}^{-1} \mathscr{C}_i$ (the pre-image of $\mathscr{C}_i$ under the projection operator $P_{\mathscr{C}}$ and $s_i = \dim \mathscr{C}_i$). Then $D_i, i = 0, \ldots, m$ form a partition of $\mathbb{R}^p$, where $D_0 = \mathscr{C}_0$.

For each $x \in C_i$, we have $P_{\mathscr{C}} x = P_i x$ where P_i is the orthogonal linear projection onto the s_i-dimensional subspace L_i spanned by $\mathscr{C}_i$. Also for each such x, the orthogonal projection onto $L_i^{\perp}$, is equal to $P_{\mathscr{C}^*} x$ where $\mathscr{C}^* = \{y \mid x^{\mathrm{T}} y \leq 0\}$ is the polar cone corresponding to $\mathscr{C}$. Additionally, if $x \in D_i$, then $a P_i x + P_i^{\perp} x \in D_i$ for all $a > 0$, so D_i is positively homogeneous in $P_i x$ for each fixed $P_i^{\perp} x$ (see Robertson et al. 1988, Theorem 8.2.7). Hence we may express

$$\delta_{\mathscr{C}}(X) = \sum_{i=0}^{m} \mathbb{1}_{D_i}(X) P_i X. \tag{7.15}$$

The problem of dominating $\delta_{\mathscr{C}}$ is relatively simple in the case where $\mathscr{C}$ has the form $\mathscr{C} = \mathbb{R}_+^k \oplus \mathbb{R}^{p-k}$ where $\mathbb{R}_+^k = \{(x_1, \ldots, x_k) \mid x_i \geq 0,\ i = 1, \ldots, k\}$. In this case,

$$\delta_{\mathscr{C}}(X)_i = \begin{cases} X_i \text{ if } X_i \geq 0 \\ 0 \quad \text{if } X_i < 0 \text{ for } i = 1, \ldots k \text{ and } \delta_{\mathscr{C}}(X)_i = X_i \text{ for } i = k+1, \ldots, p. \end{cases}$$

Furthermore, $\delta_{\mathscr{C}}(X)$ is weakly differentiable and the techniques of Chap.3 (i.e. Stein's lemma) are available.

As a simple example, suppose $\mathscr{C} = \mathbb{R}_+^p$, i.e. all coordinates of θ are nonnegative. Then $\delta_{\mathscr{C}_i}(X) = X_i + \partial_i(X) \quad i = 1, \ldots, p$ where

$$\partial_i(X) = \begin{cases} -X_i \text{ if } X_i < 0 \\ \quad 0 \text{ if } X_i \geq 0. \end{cases}$$

Also, we may rewrite (7.15) as $X_+ = \sum_{i=1}^{2^p} \mathbb{1}_{O_i}(X) P_i X$ where $O_1 = \mathbb{R}_+^p$, and O_j, for $j > 1$, represent the other $2^p - 1$ orthants and P_i is the projection of X onto the space generated by the face of O_1 adjacent to O_i.

Then a James-Stein type shrinkage estimator that dominates X_+ is given by

$$\delta(X) = \sum_{i=1}^{2^p} \left(1 - \frac{c_i}{\|X_+\|^2}\right) X_+ \mathbb{1}_{O_i}(X)$$

where $c_i = (s_i - 2)_+$ and s_i is the number of positive coordinates in O_i. Note that shrinkage occurs only in those orthants such that $s_i \geq 3$.

The proof of domination follows essentially by the usual argument of Chap.3, Sect.2.4, applied separately to each orthant since X_+ and $X_+/\|X_+\|^2$ are weakly differentiable in O_i and

$$\nabla_X \frac{X_+}{\|X_+\|^2} \mathbb{1}_{O_i}(X) = \frac{s_i - 2}{\|X_+\|^2} \mathbb{1}_{O_i}(X),$$

provided $s_i > 2$. Note also that c_i may be replaced by any value between 0 and $2(s_i - 2)_+$.

Difficulties arise when the cone $\mathscr{C}$ is not of the form $\mathscr{C} = \mathbb{R}_+^k \oplus \mathbb{R}^{p-k}$ because the estimator $P_{\mathscr{C}} X$ may not be weakly differentiable (see Appendix A.1). In this case, a result of Sengupta and Sen (1991) can be used to give an unbiased estimator of the risk. Here is a version of their result.

Lemma 7.1 (Sengupta and Sen 1991) *Let $X \sim \mathscr{N}_p(\theta, \sigma^2 I_p)$ and $\mathscr{C}$ a positively homogeneous set. Then for every absolutely continuous function $h(\cdot)$ from $\mathbb{R}_+$ to $\mathbb{R}$ such that $\lim_{y \to 0,\infty} h(y) y^{k+p/2} e^{-y/2} = 0$ for all $k \geq 0$ and $E_\theta[h^2(\|X\|^2)\|X\|^2] < \infty$ we have*

$$\begin{aligned} E_\theta[h(\|X\|^2) X^{\mathrm{T}} (X - \theta) \mathbb{1}_{\mathscr{C}}(X)] &= \sigma^2 E_\theta[2\|X\|^2 h'(\|X\|^2) + p h(\|X\|^2) \mathbb{1}_{\mathscr{C}}(X)] \\ &= \sigma^2 E_\theta[\mathrm{div}(h(\|X\|^2) X) \mathbb{1}_{\mathscr{C}}(X)]. \end{aligned}$$

Note that for $\mathscr{C} = \mathbb{R}^p$, Lemma 7.1 follows from Stein's lemma with $g(X) = h(\|X\|^2) X$ provided $E[h(\|X\|^2)\|X\|^2] < \infty$. The possible non-weak differentiability of the function $h(\|X\|^2) X \mathbb{1}_{\mathscr{C}}(X)$ prevents a direct use of Stein's lemma for general $\mathscr{C}$.

Proof of Lemma 7.1 Note first that, if, for any θ, $E_\theta[\|g(X)\|] < \infty$, then

$$E_0\left[g(X) e^{X^{\mathrm{T}}\theta/\sigma^2}\right] = \sum_{k=0}^{\infty} E_0\left[\frac{g(X)(X^{\mathrm{T}}\theta/\sigma^2)^k}{k!}\right]$$

by the dominated convergence theorem. Without loss of generality, assume $\sigma^2 = 1$ and let

$$
\begin{aligned}
A_\theta &= E_\theta[h(\|X\|^2)X^{\mathrm{T}}(X-\theta)\mathbb{1}_{\mathscr{C}}(X)] \qquad (7.16)\\
&= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\int_{\mathbb{R}^p} e^{-\|X\|^2/2}e^{X^{\mathrm{T}}\theta}h(\|X\|^2)(\|X\|^2 - X^{\mathrm{T}}\theta)\mathbb{1}_{\mathscr{C}}(X)dx\\
&= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty} E_0\left[h(\|X\|^2)(\|X\|^2 - X^{\mathrm{T}}\theta)\frac{(X^{\mathrm{T}}\theta)^k}{k!}\mathbb{1}_{\mathscr{C}}(X)\right]\\
&= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}E_0\left[h(\|X\|^2)\mathbb{1}_{\mathscr{C}}(X)(X^{\mathrm{T}}\theta)^k(\|X\|^2 - k)\right]\\
&= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}E_0\left[h(\|X\|^2)\mathbb{1}_{\mathscr{C}}(X)\left(\frac{X^{\mathrm{T}}\theta}{\|X\|}\right)^k(\|X\|^{k+2} - k\|X\|^k)\right].
\end{aligned}
$$

By the positive homogeneity of $\mathscr{C}$, we have $\mathbb{1}_{\mathscr{C}}(X) = \mathbb{1}_{\mathscr{C}}(X/\|X\|)$ and, by the independence of $\|X\|$ and $X/\|X\|$ for $\theta = 0$, we have

$$
A_\theta = (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}E_0\left[\left(\frac{X^{\mathrm{T}}\theta}{\|X\|}\right)^k\mathbb{1}_{\mathscr{C}}\left(\frac{X}{\|X\|}\right)\right] \times E_0\left[h(\|X\|^2)\left(\|X\|^{k+2} - k\|X\|^k\right)\right] \qquad (7.17)
$$

When $\theta = 0$, $\|X\|^2$ has a central Chi-square distribution with p degrees of freedom. Thus, with $d = 1/2^{p/2}\Gamma(p/2)$, we have

$$
\begin{aligned}
E_0[h(\|X\|^2)(\|X\|^{k+2} - k\|X\|^k)] &= d\int_0^{\infty} y^{p/2-1}h(y)(y^{(k+2)/2} - ky^{k/2})e^{-y/2}dy\\
&= d\int_0^{\infty} y^{(p+k)/2}h(y)e^{-y/2}dy - dk\int_0^{\infty} y^{(p+k)/2-1}h(y)e^{-y/2}dy
\end{aligned}
$$

Integrating by parts, the first integral gives

$$
\begin{aligned}
&\int_0^{\infty} y^{(p+k)/2}h(y)e^{-y/2}dy\\
&= 2\left[\int_0^{\infty}\frac{p+k}{2}y^{(p+k)/2-1}h(y)e^{-y/2}dy + \int_0^{\infty} y^{(p+k)/2}h'(y)e^{-y/2}dy\right]
\end{aligned}
$$

and thus combining gives

$$E_0[h(\|X\|^2)(\|X\|^{k+2} - k\|X\|^k)] = d\int_0^\infty y^{k/2}[2yh'(y) + ph(y)]y^{(p-2)/2}e^{-y/2}dy$$
$$= E_0[(2\|X\|^2 h'(\|X\|^2) + ph(\|X\|^2))\|X\|^k].$$

Thus (7.17) becomes

$$\begin{aligned}
A_\theta &= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}E_0\left(\frac{X^{\mathrm{T}}\theta}{\|X\|}\right)^k \mathbb{1}_{\mathscr{C}}(X)\left(\frac{X^{\mathrm{T}}\theta}{\|X\|}\right) \\
&\quad \times E_0[(2\|X\|^2h'(\|X\|^2) + ph(\|X\|^2))\|X\|^k] \\
&= (2\pi)^{-p/2}e^{-\|\theta\|^2/2}\sum_{k=0}^{\infty}\frac{1}{k!}E_0[(X^{\mathrm{T}}\theta)^k\{2\|X\|^2h'(\|X\|^2) + ph(\|X\|^2)\}\mathbb{1}_{\mathscr{C}}(X)] \\
&= E_\theta[\{2\|X\|^2h'(\|X\|^2) + ph(\|X\|^2)\}\mathbb{1}_{\mathscr{C}}(X)]
\end{aligned}$$

where the final identity follows by the dominated convergence theorem. □

General dominating estimators will be obtained by shrinking each $P_i X$ in (7.15) on the set D_i. Recall that each D_i has the property that, if $x \in D_i$, then $aP_i x + P_i^{\perp} x \in D_i$ for all $a > 0$. The next result extends Lemma 7.1 to apply to projections P_i onto sets which have this conditional homogeneity property.

Lemma 7.2 *Let $X \sim \mathscr{N}_p(\theta, \sigma^2 I_p)$ and P be a linear orthogonal projection of rank s. Further, let D be a set such that, if $x = Px + P^{\perp}x \in D$, then $aPx + P^{\perp}x \in D$ for all $a > 0$. Then, for any absolutely continuous function $h(\cdot)$ on $\mathbb{R}_+$ into $\mathbb{R}$ such that $\lim_{y\to 0,\infty} h(y)y^{(j+s)/2}e^{-y/2} = 0$ for all $j \geq 0$, we have*

$$\begin{aligned}
&E_\theta[(X-\theta)^{\mathrm{T}}PXh(\|PX\|^2)\mathbb{1}_D(X)] \\
&= \sigma^2 E_\theta[\{2\|PX\|^2h'(\|PX\|^2) + sh(\|PX\|^2)\}\mathbb{1}_D(X)].
\end{aligned}$$

Proof By assumption $(Y_1, Y_2) = (PX, P^{\perp}X) \sim (\mathscr{N}_p(\eta_1, \sigma^2 P), \mathscr{N}_p(\eta_2, \sigma^2 P^{\perp}))$ where $(P\theta, P^{\perp}\theta) = (\eta_1, \eta_2)$. Also

$$\begin{aligned}
A(\theta) &= E_\theta[(X-\theta)^{\mathrm{T}}PXh(\|PX\|^2)\mathbb{1}_D(X)] \\
&= E_\theta[(PX - P\theta)^{\mathrm{T}}PXh(\|PX\|^2)\mathbb{1}_D(X)] \\
&= E_{\eta_1\eta_2}[(Y_1-\eta_1)^{\mathrm{T}}Y_1h(\|Y_1\|^2)\mathbb{1}_{D'}(Y_1, Y_2)]
\end{aligned}$$

where

$$D' = \{(Y_1, Y_2) | (Y_1, Y_2) = (PX, P^{\perp}X) \in D\}.$$

On conditioning on Y_2 (which is independent of Y_1), and applying Lemma 7.1 to Y_1, we have

$$A(\theta) = E_{\eta_2}[E_{\eta_1}[(Y_1 - \eta_1)^{\mathrm{T}} Y_1 h(\|Y_1\|^2) \mathbb{1}_{D'}(Y_1, Y_2)|Y_2]]$$

$$= \sigma^2 E_{\eta_2}[E_{\eta_1}[\{2\|Y_1\|^2 h'(\|Y_1\|^2) + sh(\|Y_1\|^2)\}\mathbb{1}_{D'(Y_1,Y_2)}|Y_2]]$$

$$= \sigma^2 E[\{2\|PX\|^2 h'(\|PX\|^2) + sh(\|PX\|^2)\}\mathbb{1}_D(X)].$$

□

Now we use Lemma 7.2 to obtain the main domination result of this section.

Theorem 7.3 *Let $X \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ where σ^2 is known and θ is restricted to lie in the polyhedral cone $\mathscr{C}$, (7.14), with nonempty interior. Then, under loss $L(\theta, d) = \|d - \theta\|^2/\sigma^2$, the estimator*

$$\delta(X) = \sum_{i=0}^{m} \left(1 - \sigma^2 \frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2}\right) P_i X \mathbb{1}_{D_i}(X) \tag{7.18}$$

dominates the rule $\delta_{\mathscr{C}}(X)$ given by (7.15) provided $0 < r_i(t) < 2$, $r_i(\cdot)$ is absolutely continuous, and $r_i'(t) \geq 0$, for each $i = 0, 1, \ldots, m$.

Proof The difference in risk between δ and $\delta_{\mathscr{C}}$ can be expressed as

$$\begin{aligned}\Delta(\theta) &= R(\theta, \delta) - R(\theta, \delta_{\mathscr{C}}) \\ &= \sum_{i=0}^{m} E_\theta \Bigg[\sigma^2 \frac{r_i^2(\|P_i X\|^2)((s_i - 2)_+)^2}{\|P_i X\|^2} \qquad (7.19)\\ &\qquad -2\frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2}(P_i X)^{\mathrm{T}}(P_i X - \theta)\Bigg] \mathbb{1}_{D_i}(X).\end{aligned}$$

Now apply Lemma 7.2 (noting that $(P_i X)^{\mathrm{T}}(P_i X - \theta) = (P_i X)^{\mathrm{T}}(X - \theta)$) to each summand in the second term to get

$$\begin{aligned}\Delta(\theta) &= \sigma^2 \sum_{i=0}^{m} E_\theta \Bigg[\frac{r_i^2(\|P_i X\|^2)((s_i - 2)_+)^2}{\|P_i X\|^2} \\ &\quad - 2\frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2} - 4r_i'(\|P_i X\|^2)(s_i - 2)_+\Bigg] \mathbb{1}_{D_i}(X) \\ &\leq 0| \qquad (7.20)\end{aligned}$$

since each $r_i'(\cdot) \geq 0$ and $0 < r_i(\cdot) < 2$. □

As noted in Chap.3, the case of an unknown σ^2 is easily handled provided an independent statistic $S \sim \sigma^2 \chi_k^2$ is available. For completeness we give the result for this case in the following theorem.

Theorem 7.4 *Suppose $X \sim \mathcal{N}_p(\theta, \sigma^2 I_p)$ and $S \sim \sigma^2 \chi_k^2$ with X independent of S. Let the loss be $\|d - \theta\|^2/\sigma^2$. Suppose that θ is restricted to the polyhedral cone $\mathscr{C}$, (7.14), with non-empty interior. Then the estimator*

$$\delta(X, S) = \sum_{i=0}^{m} \left(1 - \left(\frac{S}{k+2}\right) \frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2}\right) P_i X \mathbb{1}_{D_i}(X) \tag{7.21}$$

dominates $\delta_{\mathscr{C}}(X)$ given in (7.15) provided $0 < r_i(\cdot) < 2$ and $r_i(\cdot)$ is absolutely continuous with $r_i'(\cdot) \geq 0$, for $i = 0, \ldots, m$.

Many of the classical problems in ordered inference are examples of restrictions to polyhedral cones. Here are a few examples.

Example 7.1 (Orthant Restrictions) Estimation problems where k of the coordinate means are restricted to be greater than (or less than) a given set constants, can be transformed easily into the case where these same components are restricted to be positive. This is essentially the case for $\mathscr{C} = \mathbb{R}_+^k \oplus \mathbb{R}^{p-k}$ mentioned earlier.

Example 7.2 (Ordered Means) The restrictions that $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_p$ (or that a subset are so ordered) is a common example in the literature and corresponds to the finite set of half space restrictions $\theta_2 \geq \theta_1, \theta_3 \geq \theta_2, \ldots, \theta_p \geq \theta_{p-1}$.

Example 7.3 (Umbrella Ordering) The ordering $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k \geq \theta_{k+1} \geq \theta_{k+2}, \ldots, \theta_{p-1} \geq \theta_p$ corresponds to the polyhedral cone generated by the half space restrictions

$$\theta_2 - \theta_1 \geq 0, \theta_3 - \theta_2 \geq 0, \ldots, \theta_k - \theta_{k-1} \geq 0, \theta_{k+1} - \theta_k \leq 0, \ldots, \theta_p - \theta_p \leq 0.$$

In some examples, such as Example 7.1, it is relatively easy to specify P_i and D_i. In others, such as Example 7.2 and 7.3 it is more complicated. The reader is referred to Robertson et al. (1988) and references therein for further discussion of this issue.

7.5 Spherically Symmetric Distribution with a Mean Vector Restricted to a Polyhedral Cone

This Section is devoted to proving the extension of Theorem 7.4 to the case of a spherically symmetric distribution when a residual vector is present. Specifically we assume that $(X, U) \sim SS(\theta, 0)$ where $\dim X = \dim \theta = p, \dim U = \dim 0 = k$ and where θ is restricted to lie in a polyhedral cone, $\mathscr{C}$, with non-empty interior. Recall that the shrinkage functions in the estimator (7.21) are not necessarily weakly differentiable because of the presence of the indicator functions $I_{D_i}(X)$. Hence the methods of Chap.4 are not immediately applicable.

The following theorem develops the required tools for the desired extension of Theorem 7.4. It also allows for an alternative approach to the results in Sect.6.1 as well.

Theorem 7.5 (Fourdrinier et al. 2006) *Let* $(X, U) \sim \mathcal{N}_{p+k}((\theta, 0), \sigma^2 I_{p+k})$ *and assume* $f : \mathbb{R}^p \to \mathbb{R}$ *and* $g : \mathbb{R}^p \to \mathbb{R}^p$ *are such that*

$$E_{\theta,0}[(X-\theta)^{\mathrm{T}} g(X)] = \sigma^2 E_{\theta,0}[f(X)]$$

where both expected values exist for all $\sigma^2 > 0$*. Then, if* $(X, U) \sim SS_{p+k}(\theta, 0)$*, we have*

$$E_{\theta,0}[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)] = \frac{1}{k+2} E_{\theta,0}[\|U\|^4 f(X)]$$

provided either expected value exists.

Proof As (X, U) is normal, $X \sim \mathcal{N}_p(\theta, \sigma^2 I)$ and $\|U\|^2 \sim \sigma^2 \chi^2_k$ are independent, using $E[\|U\|^2] = k\sigma^2$ and $E[\|U\|^4] = \sigma^4 k(k+2)$, we have, for each fixed σ^2,

$$\begin{aligned} E_{\theta,0}[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)] &= k\sigma^2 E_{\theta,0}[(X-\theta)^{\mathrm{T}} g(X)] \\ &= k\sigma^4 E_{\theta,0}[f(X)] \\ &= \frac{1}{k+2} E_{\theta,0}[\|U\|^4 f(X)] \end{aligned} \tag{7.22}$$

For each θ (considered fixed), $\|X-\theta\|^2 + \|U\|^2$ is a complete sufficient statistic for $(X, U) \sim \mathcal{N}_{p+k}((\theta, 0), \sigma^2 I)$. Now noting

$$E_{\sigma^2}[E[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X) \mid \|X-\theta\|^2 + \|U\|^2]] = E_{\theta,0}[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X)]$$

and

$$\frac{1}{k+2} E_{\sigma^2}[\|U\|^4 f(X)] = \frac{1}{k+2} E_{\sigma^2}[E[\|U\|^4 f(X) \mid \|X-\theta\|^2 + \|U\|^2]]$$

it follows from (7.22) and the completeness of $\|X-\theta\|^2 + \|U\|^2$ that

$$\begin{aligned} &E[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X) \mid \|X-\theta\|^2 + \|U\|^2] \\ &= \frac{1}{k+2} E[\|U\|^2 (X-\theta)^{\mathrm{T}} g(X) \mid \|X-\theta\|^2 + \|U\|^2] \end{aligned} \tag{7.23}$$

almost everywhere. We show at the end of this section that the functions in (7.23) are both continuous in $\|X-\theta\|^2 + \|U\|^2$, and hence, they are in fact equal everywhere.

Since the conditional distribution of (X, U) conditional on $\|X-\theta\|^2 + \|U\|^2 = R^2$ is uniform on the sphere centered at $(\theta, 0)$ for all spherically symmetric distributions (including the normal), the result follows on integration of (7.23) with respect to the radial distribution of (X, U). □

The main result of this section results from an application of Theorem 7.3 to the development of the proof of Theorem 7.4.

Theorem 7.6 *Let* $(X, U) \sim SS_{p+k}(\theta, 0)$ *and let* θ *be restricted to the polyhedral cone* $\mathscr{C}$, (7.14)*, with nonempty interior. Then, under loss* $L(\theta, d) = \|d - \theta\|^2$*, the estimator*

$$\delta(X, U) = \sum_{i=0}^{m} \left(1 - \frac{\|U\|^2}{k+2} \frac{(s_i - 2)_+ r_i(\|P_i X\|^2)}{\|P_i X\|^2}\right) P_i X \mathbb{1}_{D_i}(X) \tag{7.24}$$

dominates $P_{\mathscr{C}} X = \delta_{\mathscr{C}}(X)$*, given in* (7.15) *provided,* $0 < r_i(\cdot) < 2$*,* $r_i(\cdot)$ *is absolutely continuous and* $r_i'(\cdot) \geq 0$ *for* $i = 0, \ldots, m$.

Proof The key observation is that, in passing from (7.19) to (7.20) in the proof of Theorem 7.3, we used Lemma 7.2 and the fact that $P_i X^{\mathrm{T}}(P X_i - \theta) = P_i X^{\mathrm{T}}(X - \theta)$ to establish that

$$E\left[\frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2}(P_i X)^{\mathrm{T}}(P_i X - \theta)\mathbb{1}_{D_i}(X)\right]$$
$$= \sigma^2 E\left[\frac{r_i(\|P_i X\|^2)((s_i - 2)_+)^2}{\|P_i X\|^2} + 2r_i'(\|P_i X\|^2)(s_i - 2)_+ \mathbb{1}_{D_i}(X)\right].$$

Hence, by Theorem 7.5,

$$E\left[\frac{\|U\|^2}{k+2}\frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2}(P_i X)^{\mathrm{T}}(P_i X - \theta)\mathbb{1}_{D_i}(X)\right]$$
$$= \sigma^2 E\left[\frac{\|U\|^4}{(k+2)^2}\left\{\frac{r_i(\|P_i X\|^2)((s_i - 2)_+)^2}{\|P_i X\|^2} + 2r_i'(\|P_i X\|^2)(s_i - 2)_+\right\}\mathbb{1}_{D_i}(X)\right].$$

It follows then, as in the proof of Theorem 7.3,

$$R(\theta, \delta(X, U)) - R(\theta, \delta_{\mathscr{C}}) = \sum_{i=0}^{m} E_\theta\left[\frac{\|U\|^4}{(k+2)^2}\left\{\frac{r_i^2(\|P_i X\|^2)((s_i - 2)_+)^2}{\|P_i X\|^2}\right.\right.$$
$$\left.\left. - \left\{\left(2\frac{r_i(\|P_i X\|^2)(s_i - 2)_+}{\|P_i X\|^2} + 4r_i'(\|P_i X\|^2)\right)(s_i - 2)_+\right\}\mathbb{1}_{D_i}(X)\right]\right.$$
$$\leq 0. \tag{7.25}$$

□

Theorem 7.5 is an example of a meta result which follows from Theorem 7.6, and states roughly that, if one can find an estimator $X + \sigma^2 g(X)$ that dominates X for each σ^2 using a Stein-type differential equality in the normal case, then $X + \|U\|^2/(k+2)g(X)$ will dominate X in the general spherically symmetric case, $(X, U) \sim SS_{1+k}(\theta, U)$, under $L(\theta, \delta) = \|\delta - \theta\|^2$. The "proof" goes as follows.

Suppose one can show also that $E[(X-\theta)^{\mathrm{T}} g(X)] = \sigma^2 E[f(X)]$ in the normal case, and also that $\|g(x)\|^2 + 2f(x) \leq 0$, for any $x \in \mathbb{R}^p$. Then, in the normal case,

$$R(\theta, X - \sigma^2 g(X)) - R(\theta, X) = \sigma^4 E[\|g(X)\|^2 + 2f(X)] \leq 0.$$

Using Theorem 7.5 (and assuming finiteness of expectations), it follows in the general case that

$$R\left(\theta, X + \frac{\|U\|^2}{k+2} g(X)\right) - R(\theta, X) = E\left[\frac{\|U\|^4}{(k+2)^2}\{\|g(X)\|^2 + 2f(X)\}\right] \leq 0.$$

In this Section, application of the above meta-result had the additional complication of a separate application (to $P_i X$ instead of X) on each D_i but the basic idea is the same. The results of Chap.6 which rely on extending a version of Stein's lemma to the general spherically symmetric case can be proved in the same way.

We close this Section with a result that implies the claimed continuity of the conditional expectations in (7.23).

Lemma 7.3 *Let* $(X, U) \sim SS_{p+k}(\theta, 0)$ *and let* $\alpha \in N$. *Assume* $\varphi(\cdot)$ *is such that for any* $R > 0$, *the conditional expectation*

$$f(R) = E_{(\theta,0)}[\|U\|^\alpha \varphi(X) \mid \|X - \theta\|^2 + \|U\|^2 = R^2]$$

exists. Then the function f *is continuous on* $\mathbb{R}_+$.

Proof Assume without loss of generality that $\theta = 0$ and $\varphi(\cdot) \geq 0$. Since the conditional distribution or (X, U) conditional on $\|X\|^2 + \|U\|^2 = R^2$ is the uniform distribution U_R on the sphere $S_R = \{y \in R^{p+k}/\|y\| = R\}$ centered at 0 with radius R, we have

$$f(R) = \int_{S_R} \|u\|^\alpha \varphi(x) d\mathscr{U}_R(x, u).$$

Since $\|u\|^2 = R^2 - \|x\|^2$ for any $(x, u) \in S_R$ and X has distribution concentrated on the ball $B_r = \{x \in \mathbb{R}^p | \|x\| \leq R\}$ in $\mathbb{R}^p$ with density proportional to $R^{2-(p+k)}((R^2 - \|x\|^2)^{k/2-1})$ we have that $R^{p+k-2} f(R)$ is proportional to

$$g(R) = \int_{B_R} (R^2 - \|x\|^2)^{(k+\alpha)/2-1} \varphi(x) dx.$$

$$= \int_0^R \int_{S_r} (R^2 - \|x\|^2)^{(k+\alpha)/2-1} \varphi(x) d\sigma_r(x) dr$$

$$= \int_0^R (R^2 - r^2)^{(k+\alpha)/2-1} H(r) dr$$

where

$$H(r) = \int_{S_r} \varphi(x) d\sigma_r(x)$$

and where σ_r is the area measure on the sphere S_r. Since $H(\cdot)$ and $(k+\alpha)/2-1$ are non-negative, the family of integrable functions $r \to K(R,r) = (R^2 - r^2)^{(k+\alpha)/2-1} H(r) I_{[0,R]}(r)$, indexed by R, is nondecreasing in R and bounded above for $R < R_0$ by the integrable function $K(R_0, r)$. Then the continuity of $g(R)$, and hence of $f(R)$, is guaranteed by the dominated convergence theorem. □

Note that the continuity of (7.23) is not necessary for the application to $(X, U) \sim SS_{p+k}(\theta, 0)$ if (X, U) has a density, since then equality a.e. suffices.