Chapter 4
Spherically Symmetric Distributions

4.1 Introduction

In the previous chapters, estimation problems were considered for the normal distribution setting. Stein (1956) showed that the usual estimator of a location vector could be improved upon quite generally for $p \geq 3$ and Brown (1966) substantially extended this conclusion to essentially arbitrary loss functions. Explicit results of the James-Stein type, however, have thus far been restricted to the case of the normal distribution. Recall the geometrical insight from Sect. 2.2.2, the development did not depend on the normality of X or even that θ is a location vector – this suggests that the improvement for Stein-type estimators may hold for more general distributions. Strawderman (1974a) first explored such an extension and considered estimation of the location parameter for scale mixtures of multivariate normal distributions. Other extensions of James-Stein type results to distributions other than scale mixtures of normal distributions are due to Berger (1975), Brandwein and Strawderman (1978), and Bock (1985). In this chapter, we will introduce the general class of spherically symmetric distributions; we will examine point estimation for variants of this general class in subsequent three chapters.

4.2 Spherically Symmetric Distributions

The normal distribution has been generalized in two important directions. First, as a special case of the exponential family and second, as a spherically symmetric distribution. In this chapter, we will consider the latter. There are a variety of equivalent definitions and characterizations of the class of spherically symmetric distributions; a comprehensive review is given in Fang et al. (1990). We now turn our interest to general orthogonally invariant distributions in $\mathbb{R}^n$ and a slightly more general notion of spherically symmetric distributions.

© Springer Nature Switzerland AG 2018

D. Fourdrinier et al., *Shrinkage Estimation*, Springer Series in Statistics,
https://doi.org/10.1007/978-3-030-02185-6_4

Definition 4.1 A random vector $X \in \mathbb{R}^n$ (equivalently the distribution of X) is spherically symmetric about $\theta \in \mathbb{R}^n$ if $X - \theta$ is orthogonally invariant. We denote this by $X \sim SS(\theta)$.

Note that Definition 4.1 states that $X \sim SS(\theta)$ if and only if $X = Z + \theta$ where $Z \sim SS(0)$. As an example, the uniform distribution $\mathscr{U}_{R,\theta}$ (cf. Definition 1.4) on the sphere $S_{R,\theta}$ of radius R and centered at θ is spherically symmetric about θ. Furthermore, if P is a spherically symmetric distribution about θ, then

$$P(HC + \theta) = P(C + \theta),$$

for any Borel set C of $\mathbb{R}^n$ and any orthogonal transformation H.

The following proposition is immediate from the definition.

Proposition 4.1 *If a random vector $X \in \mathbb{R}^n$ is spherically symmetric about $\theta \in \mathbb{R}^n$ then, for any orthogonal transformation H, HX is spherically symmetric about $H\theta$ ($X - \theta$ has the same distribution as $HX - H\theta$).*

The connection between spherical symmetry and uniform distributions on spheres is indicated in the following theorem.

Theorem 4.1 *A distribution P in $\mathbb{R}^n$ is spherically symmetric about $\theta \in \mathbb{R}^n$ if and only if there exists a distribution ρ in $\mathbb{R}_+$ such that $P(A) = \int_{\mathbb{R}_+} \mathscr{U}_{r,\theta}(A)\, d\rho(r)$ for any Borel set A of $\mathbb{R}^n$. Furthermore, if a random vector X has such a distribution P, then the radius $\|X - \theta\|$ has distribution ρ (called the radial distribution) and the conditional distribution of X given $\|X - \theta\| = r$ is the uniform distribution $\mathscr{U}_{r,\theta}$ on the sphere $S_{r,\theta}$ of radius r and centered at θ.*

Proof Sufficiency is immediate since the distribution $\mathscr{U}_{r,\theta}$ is spherically symmetric about θ for any $r \geq 0$.

It is clear that for the necessity it suffices to consider $\theta = 0$. Let X be distributed as P where P is $SS(0)$, $\nu(x) = \|x\|$, and ρ be the distribution of ν. Now, for any Borel sets A in $\mathbb{R}^n$ and B in $\mathbb{R}_+$ and for any orthogonal transformation H, we have (using basic properties of conditional distributions)

$$\begin{aligned}
\int_B P(H^{-1}(A) \mid \nu = r)\, d\rho(r) &= P(H^{-1}(A) \cap \nu^{-1}(B)) \\
&= P(H^{-1}(A \cap H(\nu^{-1}(B)))) \\
&= P(A \cap H(\nu^{-1}(B))) \\
&= P(A \cap \nu^{-1}(B)) \\
&= \int_B P(A \mid \nu = r)\, d\rho(r)
\end{aligned}$$

where we used the orthogonal invariance of the measure P and the function ν. Since the above equality holds for any B, then, almost everywhere with respect to ρ, we have

$$P(H^{-1}(A) \mid \nu = r) = P(A \mid \nu = r).$$

Equivalently, the conditional distribution given ν is orthogonally invariant on S_r. By unicity (see Lemma 1.1), it is the uniform distribution on S_r and the theorem follows. □

Corollary 4.1 *A random vector $X \in \mathbb{R}^n$ has a spherically symmetric distribution about $\theta \in \mathbb{R}^n$ if and only if X has the stochastic representation $X = \theta + R\,U$ where R ($R = \|X - \theta\|$) and U are independent, $R \geq 0$ and $U \sim \mathscr{U}$.*

Proof In the proof of Theorem 4.1, we essentially show that the distribution of $(X - \theta)/\|X - \theta\|$ is $\mathscr{U}$ independently of $\|X - \theta\|$. This is the necessity part of the corollary. The sufficiency part is direct. □

Also, the following corollary is immediate.

Corollary 4.2 *Let X be a random vector in R^n having a spherically symmetric distribution about $\theta \in \mathbb{R}^n$. Let h be a real valued function on R^n such that the expectation $E_\theta[h(X)]$ exists. Then*

$$E_\theta[h(X)] = E[E_{R,\theta}[h(X)]],$$

where $E_{R,\theta}$ is the conditional expectation of X given $\|X - \theta\| = R$ (i.e. the expectation with respect to the uniform distribution $\mathscr{U}_{R,\theta}$ on the sphere $S_{R,\theta}$ of radius R and centered at θ) and E is the expectation with respect to the distribution of the radius $\|X - \theta\|$.

A more general class of distributions where $(X - \theta)/\|X - \theta\| \sim \mathscr{U}$ but not necessarily independently of $\|X - \theta\|$ is known as the isotropic distributions (see Philoche 1977). The class of spherically symmetric distributions with a density with respect to the Lebesgue measure is of particular interest. The form of this density and its connection with the radial distribution are the subject of the following theorem.

Theorem 4.2 *Let $X \in \mathbb{R}^n$ have a spherically symmetric distribution about $\theta \in \mathbb{R}^n$. Then the following two statements are equivalent.*

(1) *X has a density f with respect to the Lebesgue measure on $\mathbb{R}^n$.*
(2) *$\|X - \theta\|$ has a density h with respect to the Lebesgue measure on $\mathbb{R}_+$.*

Further, if (1) *or* (2) *holds, there exists a function g from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that*

$$f(x) = g(\|x - \theta\|^2) \ \ a.e.$$

and

$$h(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1} g(r^2) \ \ a.e.$$

The function g is called the generating function and h the radial density.

Proof The fact that (1) implies (2) follows directly from the representation of X in polar coordinates. We can also argue that (2) implies (1) in a similar fashion using the independence of $\|X - \theta\|$, angles, and the fact that the angles have a density. The following argument shows this directly and, furthermore, gives the relationship between f, g, and h.

It is clear that it suffices to assume that $\theta = 0$. Suppose then that $R = \|X\|$ has a density h. According to Theorem 4.1, for any Borel set A of $\mathbb{R}^n$, we have

$$\begin{aligned}
P(X \in A) &= \int_0^\infty \int_{S_r} \mathbb{1}_A\ (y)\ d\mathscr{U}_r(y)\, h(r)\, dr \\
&= \int_0^\infty \int_{S_r} \mathbb{1}_A(y)\, \frac{d\sigma_r\ (y)}{\sigma_1(S_1) r^{n-1}}\ h(r)\ dr \quad \text{(by (1.4))} \\
&= \int_0^\infty \int_{S_r} \mathbb{1}_A(y)\, \frac{h(\|y\|)}{\sigma_1(S_1)\|y\|^{n-1}}\ d\sigma_r(y)\ dr \\
&= \int_{\mathbb{R}^n} \mathbb{1}_A(y) \frac{h(\|y\|)}{\sigma_1(S_1)\|y\|^{n-1}}\ dy \quad \text{(by Lemma 1.4)} \\
&= \int_A \frac{h(\|y\|)}{\sigma_1(S_1)\|y\|^{n-1}}\ dy.
\end{aligned}$$

This implies that the random vector X has density

$$f(x) = \frac{h(\|x\|)}{\sigma_1(S_1)\|x\|^{n-1}} = g(\|x\|^2)$$

with $h(r) = \sigma_1(S_1)\, r^{n-1}\, g(r^2)$, which is the announced formula for $h(r)$ since $\sigma_1(S_1) = 2\pi^{n/2}/\Gamma(n/2)$ by Corollary 1.1. □

We now turn our attention to the mean and the covariance matrix of a spherically symmetric distribution (when they exist).

Theorem 4.3 *Let $X \in \mathbb{R}^n$ be a random vector with a spherically symmetric distribution about $\theta \in \mathbb{R}^n$. Then, the mean of X exists if and only if the mean of $R = \|X - \theta\|$ exists, in which case $E[X] = \theta$. The covariance matrix of X exists if and only if $E[R^2]$ is finite, in which case*

$$cov(X) = \frac{E[R^2]}{n} I_n.$$

Proof Note that $X = Z + \theta$ where $Z \sim SS(0)$ and it suffices to consider the case $\theta = 0$. By the stochastic representation $X = RU$ in Corollary 4.1 with $R = \|X\|$ independent of U and $U \sim \mathscr{U}$, the expectation $E[X]$ exists if and only if the

expectations $E[R]$ and $E[U]$ exist. However, since U is bounded, $E[U]$ exists and is equal to zero since $E[U] = E[-U]$ by orthogonal invariance.

Similarly, $E[\|X\|^2] = E[R^2]\,E[\|U\|^2] = E[R^2]$ and consequently the covariance matrix of X exists if and only if $E[R^2] < \infty$. Now

$$\operatorname{cov}(RU) = E[R^2]\,E[UU^{\mathrm{T}}] = \frac{E[R^2]}{n} I_n.$$

Indeed $E[U_i^2] = E[U_j^2] = 1/n$ since U_i and U_j have the same distribution by orthogonal invariance and since $\sum_{i=1}^n U_i^2 = 1$. Furthermore, $E[U_i U_j] = 0$, for $i \neq j$, since $U_i U_j$ has the same distribution as $-U_i U_j$ by orthogonal invariance. □

An interesting and useful subclass of spherically symmetric distributions consists of the spherically symmetric unimodal distributions. We only consider absolutely continuous distributions.

Definition 4.2 A random vector $X \in \mathbb{R}^n$ with density f is unimodal if the set $\{x \in \mathbb{R}^n \mid f(x) \geq a\}$ is convex for any $a \geq 0$.

Lemma 4.1 *Let $X \in \mathbb{R}^n$ be a spherically symmetric random vector about θ with generating function g. Then the distribution of X is unimodal if and only if g is nonincreasing.*

Proof Suppose first that the generating function g is nonincreasing. Take the left continuous version of g. For any $a \geq 0$, defining $g^{-1}(a) = \sup\{y \geq 0 \mid g(y) = a\}$ we have

$$\{x \in \mathbb{R}^n \mid g(\|x\|^2) \geq a\} = \{x \in \mathbb{R}^n \mid \|x\|^2 \leq g^{-1}(a)\}$$

which is a ball of radius $\sqrt{g^{-1}(a)}$ and convex. Conversely suppose that the set $\{x \in \mathbb{R}^n \mid g(\|x\|^2) \geq a\}$ is convex for any $a \geq 0$ and let $\|x\| \leq \|y\|$. Then, for $x^{\mathrm{T}} = y/\|y\|\|x\|$, we have $\|x^{\mathrm{T}}\| = \|x\|$ and $x^{\mathrm{T}} \in [-y, y]$ and hence, by the unimodality assumption, $g(\|x\|^2) = g(\|x^{\mathrm{T}}\|^2) \geq g(\|y\|^2)$. □

Theorem 4.1 showed that a spherically symmetric distribution is a mixture of uniform distributions on spheres. It is worth noting that, when the distribution is also unimodal, it is a mixture of uniform distributions on balls.

Theorem 4.4 *Let $X \in \mathbb{R}^n$ be a spherically symmetric random vector about $\theta \in \mathbb{R}^n$ with generating function g. Then the distribution of X is unimodal if and only if there exists a distribution ν in $\mathbb{R}_+$ with no point mass at* 0 *such that*

$$P[X \in A] = \int_{R_+} \mathscr{V}_{r,\theta}(A)\; d\nu(r) \tag{4.1}$$

for any Borel set A of $\mathbb{R}^n$, where $\mathscr{V}_{r,\theta}$ is the uniform distribution on the ball $B_{r,\theta} = \{x \in \mathbb{R}^n \mid \|x - \theta\| \le r\}$.

Proof It is clear that it suffices to consider the case where $\theta = 0$. Suppose first that formula (4.1) is satisfied. Then expressing

$$\mathscr{V}_{r,0}(A) = \frac{1}{\lambda(B_r)} \int_{B_r} \mathbb{1}_A(x)\, dx$$

gives

$$\begin{aligned} P[X \in A] &= \int_{\mathbb{R}_+} \frac{1}{\lambda_(B_r)} \int_{B_r} \mathbb{1}_A(x)\ dx\, d\nu\ (r) \\ &= \int_{\mathbb{R}_+} \frac{1}{\lambda(B_r)} \int_0^r \int_{S_u} \mathbb{1}_A(x)\ d\sigma_u(x)\ du\ d\nu(r) \\ &= \int_{\mathbb{R}_+} \int_{S_u} \mathbb{1}_A(x) \int_u^\infty \frac{1}{\lambda(B_r)}\ d\nu(r)\ d\sigma_u(x)\, du \end{aligned}$$

after applying Lemma 1.4 and Fubini's theorem. Then

$$\begin{aligned} P[X \in A] &= \int_0^\infty \int_{S_u} \mathbb{1}_A(x)\, g(\|x\|^2)\, d\sigma_u(x)\, du \\ &= \int_A g(\|x\|^2)\, dx \end{aligned}$$

again by Lemma 1.4 with the nonincreasing function

$$g(u^2) = \int_u^\infty \frac{1}{\lambda(B_r)}\ d\nu\ (r). \tag{4.2}$$

Hence according to Lemma 4.1, the distribution of X is unimodal.

Conversely, suppose that the distribution of X is unimodal. According to the above, this distribution will be a mixture of uniform distributions on balls if there exists a distribution ν on $\mathbb{R}_+$ with no point mass at 0 such that (4.2) holds. If such a distribution exists, (4.2) implies that ν can be expressed through a Stieltjes integral as

$$\nu(u) = \int_0^u \lambda(B_r)(-dg(r^2)).$$

It suffices therefore to show that ν is a distribution function on $\mathbb{R}_+$ with no point mass at 0. Note that, as g is nonincreasing, ν is the Stieltjes integral of a positive

function with respect to a nondecreasing function and hence ν is nondecreasing. Since $\lambda(B_r) = \lambda(B_1)\, r^n = n\, \sigma_1(S_1)\, r^n$, an integration by parts gives

$$\nu(u) = \sigma_1(S_1) \int_0^u r^{n-1} g(r^2)\, dr - \lambda(B_1)^n\, g(u^2). \tag{4.3}$$

Note that the first term of the right hand side (4.3) is the distribution function of the radial distribution (see Theorem 4.2) and approaches 0 (respectively 1) when u approaches 0 (respectively ∞). Therefore, to complete the proof it suffices to show that

$$\lim_{u\to 0} u^n g(u^2) = \lim_{u\to\infty} u^n g(u^2) = 0\,.$$

Since

$$\int_0^\infty r^{n-1} g(r^2)\, dr < \infty,$$

we have

$$\lim_{r\to\infty} \int_{r/2}^r r^{n-1} g(u^2)\, du = 0.$$

By the monotonicity of g, we have

$$\int_{r/2}^r u^{n-1} g(u^2)\; du \geq\; (r^2) \int_{r/2}^r u^{n-1}\, du = g(r^2)\, r^n\, \frac{1}{n}\left(1 - \frac{1}{2^n}\right).$$

Hence, $\lim\limits_{r\to\infty} r^n g(r^2) = 0$. The limit as r approaches 0 can be treated similarly and the result follows. □

It is possible to allow the possibility of a point mass at 0 for a spherically symmetric unimodal distribution, but we choose to restrict the class to absolutely continuous distributions. For a more general version of unimodality see Section 2.1 of Liese and Miescke (2008).

4.3 Elliptically Symmetric Distributions

By Definition 1.2, a random vector $X \in \mathbb{R}^n$ is orthogonally invariant if, for any orthogonal transformation H, HX has the same distribution as X. The notion of orthogonal transformation is relative to the classical scalar product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$. It is natural to investigate orthogonal invariance with respect to orthogonal transformations relative to a general scalar product $\langle x, y\rangle_\Gamma = x^{\mathrm{T}} \Gamma y =$

$\sum_{1\le i,j\le n} x_i \Gamma_{ij} y_j$ where Γ is a symmetric positive definite $n \times n$ matrix. We define a transformation H to be Γ-orthogonal if it preserves the scalar product in the sense that, for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$, $\langle Hx, Hy\rangle_\Gamma = \langle x, y\rangle_\Gamma$ or, equivalently, if it preserves the associated norm $\|x\|_\Gamma = \sqrt{\langle x, x\rangle_\Gamma}$, that is, if $\|Hx\|_\Gamma = \|x\|_\Gamma$. Note that H is necessarily invertible since

$$\ker H = \{x \in \mathbb{R}^n \mid Hx = 0\} = \{x \in \mathbb{R}^n \mid \|Hx\|_\Gamma = 0\} = \{x \in \mathbb{R}^n \mid \|x\|_\Gamma = 0\} = \{0\}\,.$$

Then it can be seen that H is Γ-orthogonal if and only if $\langle Hx, y\rangle_\Gamma = \langle x, H^{-1}y\rangle_\Gamma$, for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ or, equivalently, if $H^{\mathrm{T}}\Gamma H = \Gamma$.

In this context, the Γ-sphere of radius $r \ge 0$ is defined as

$$S_r^\Gamma = \{x \in \mathbb{R}^n \mid x^{\mathrm{T}}\Gamma x = r^2\}\,.$$

Definition 4.3 A random vector $X \in \mathbb{R}^n$ (equivalently the distribution of X) is Γ-orthogonally invariant if, for any Γ-orthogonal transformation H, the distribution of $Y = HX$ is the same as that of X.

We can define a uniform measure on the ellipse S_r^Γ in a manner analogous to (1.3) and the resulting measure is indeed Γ-orthogonally invariant. It is not however the superficial measure mentioned at the end of Sect. 1.3, but is, in fact, a constant multiple of this measure where the constant of proportionality depends on Γ and reflects the shape of the ellipse. Whatever the constant of proportionality is, it allows the construction of a unique uniform distribution on S_r^Γ as in (1.4). The uniqueness follows from the fact that the Γ-orthogonal transformations form a compact group. We can then adapt the material from Sects. 1.3 and 4.2 to the case of a general positive definite matrix Γ. However, we present an alternative development.

The following discussion indicates a direct connection between the usual orthogonal invariance and Γ-orthogonal invariance. Suppose, for the moment, that $X \in \mathbb{R}^n$ has a spherically symmetric density given by $g(\|x\|^2)$. Let Σ be a positive definite matrix and A be a nonsingular matrix such that $AA^{\mathrm{T}} = \Gamma$. Standard change of variables gives the density of $Y = AX$ as $|\Sigma|^{-1/2} g(y^{\mathrm{T}}\Sigma^{-1}y)$. Let H be any Σ^{-1} orthogonal transformation and let $Z = HY$. The density of Z is $|\Sigma|^{-1/2} g(z^{\mathrm{T}}\Sigma^{-1}z)$ since H^{-1} is also Σ^{-1} -orthogonal and hence, $(H^{-1})^{\mathrm{T}}\Sigma^{-1}H^{-1} = \Sigma^{-1}$. This suggests that, in general, $Y = \Sigma^{\frac{1}{2}}X$ is Σ^{-1}-orthogonally invariant if and only if X is orthogonally invariant. The following result establishes this general fact.

Theorem 4.5 *Let Σ be a positive definite $n \times n$ matrix. A random vector $Y \in \mathbb{R}^n$ is Σ^{-1}-orthogonally invariant if and only if $Y = \Sigma^{1/2}X$ with X orthogonally invariant.*

Proof First note that, for any Σ^{-1}-orthogonal matrix H, $\Sigma^{-1/2}H\Sigma^{-1/2}$ is an I_n -orthogonal matrix since

$$
\begin{aligned}
(\Sigma^{-1/2} H \Sigma^{1/2})^{\mathrm{T}} (\Sigma^{-1/2} H \Sigma^{1/2}) &= \Sigma^{1/2} H^{\mathrm{T}} \Sigma^{-1} H \Sigma^{1/2} \\
&= \Sigma^{1/2} \Sigma^{-1} \Sigma^{1/2} \\
&= I_n.
\end{aligned}
$$

Then, if X is orthogonally invariant, for any Borel set C, of $\mathbb{R}^n$ we have

$$
\begin{aligned}
P[H \Sigma^{1/2} X \in C] &= P[\Sigma^{-1/2} H \Sigma^{1/2} X \in \Sigma^{-1/2} C] \\
&= P[X \in \Sigma^{-1/2} C] \\
&= P[\Sigma^{1/2} X \in C].
\end{aligned}
$$

Hence $Y = \Sigma^{1/2} X$ is Σ^{-1}-orthogonally invariant.

Similarly, for any orthogonal matrix G, $\Sigma^{1/2} G \Sigma^{-1/2}$ is a Σ^{-1}-orthogonal matrix. So, if $Y = \Sigma^{1/2} X$ is Σ^{-1}-orthogonally invariant, then X is orthogonally invariant. □

Note that, if X is orthogonally invariant and its covariance matrix exists, it is of the form $\sigma^2 I_n$ by Theorem 4.3. Therefore, if $Y = \Sigma^{1/2} X$, the covariance matrix of Y is $\sigma^2 \Sigma$, while, by Theorem 4.5, Y is Σ^{-1}-orthogonal invariant. In statistical models, it is often more natural to parametrize through a covariance matrix Σ than through its inverse (graphical models are the exception) and this motivates the following definition of elliptically symmetric distributions.

Definition 4.4 Let Σ be a positive definite $n \times n$ matrix. A random vector X (equivalently the distribution of X) is elliptically symmetric about $\theta \in \mathbb{R}^n$ if $X - \theta$ is Σ^{-1}-orthogonally invariant. We denote this by $X \sim ES(\theta, \Sigma)$.

Note that, if $X \sim SS(\theta)$, then $X \sim ES(\theta, I_n)$. If $Y \sim ES(\theta, \Sigma)$, then $\Sigma^{-1/2} Y \sim SS(\Sigma^{-1/2}\theta)$.

In the following, we briefly present some results for elliptically symmetric distributions that follow from Theorem 4.5 and are the analogues of those in Sects. 1.3 and 4.2. The proofs are left to the reader.

For the rest of this section, let Σ be a fixed positive definite $n \times n$ matrix and denote by $S_R^{\Sigma^{-1}} = \{x \in \mathbb{R}^n \mid x^{\mathrm{T}} \Sigma^{-1} x = R^2\}$ the Σ^{-1}- ellipse of radius R and by $\mathscr{U}_R^{\Sigma}$ the uniform distributions on $S_R^{\Sigma^{-1}}$.

Lemma 4.2

(1) *The uniform distribution $\mathscr{U}_R^{\Sigma}$ on $S_R^{\Sigma^{-1}}$ is the image under the transformation $Y = \Sigma^{\frac{1}{2}} X$ of the uniform distribution $\mathscr{U}_R$ on the sphere S_R, that is,*

$$
\mathscr{U}_R^{\Sigma}(\Omega) = \mathscr{U}_R(\Sigma^{-\frac{1}{2}} \Omega)
$$

for any Borel set Ω of $S_R^{\Sigma^{-1}}$.

(2) *If X is distributed as $\mathscr{U}_R^{\Sigma}$, then*

(a) $\Sigma^{-1/2}X/(X^{\mathrm{T}}\Sigma^{-1}X)^{1/2}$ *is distributed as* $\mathscr{U}$ *and*
(b) $X/(X^{\mathrm{T}}\Sigma^{-1}X)^{1/2}$ *is distributed as* $\mathscr{U}_1^{\Sigma}$.

Theorem 4.6 *A random vector $X \in \mathbb{R}^n$ is distributed as $ES(\theta, \Sigma)$ if and only if there exists a distribution $\rho \in \mathbb{R}_+$ such that*

$$P[X \in A] = \int_{\mathbb{R}_+} \mathscr{U}_{r,\theta}^{\Sigma}(A)\, d\rho(r)$$

for any Borel set A on $\mathbb{R}^n$, where $\mathscr{U}_{r,\theta}^{\Sigma}$ is the uniform distribution $\mathscr{U}_r^{\Sigma}$ translated by θ. Equivalently X has the stochastic representation $X = RU$ where $R = \|X - \theta\|_{\Sigma^{-1}} = ((x-\theta)^{\mathrm{T}}\Sigma^{-1}(x-\theta))^{1/2}$ and U are independent, $R \geq 0$ and $U \sim \mathscr{U}_1^{\Sigma}$. For such X, the radius R has distribution ρ (called the radial distribution).

Theorem 4.7 *Let $X \in \mathbb{R}^n$ be distributed as $ES(\theta, \Sigma)$. Then the following two statements are equivalent:*

(1) *X has a density f with respect to the Lebesgue measure on $\mathbb{R}^n$; and*
(2) *$\|X - \theta\|_{\Sigma^{-1}}$ has a density h with respect to Lebesgue measure on $\mathbb{R}_+$.*

Further, if (1) *or* (2) *holds, there exists a function g from $\mathbb{R}_+$ into $\mathbb{R}_+$ such that*

$$f(x) = g(\|x - \theta\|_{\Sigma^{-1}}^2)$$

and

$$h(r) = \frac{2\pi^{n/2}}{\Gamma(n/2)}|\Sigma|^{-1/2} r^{n-1} g(r^2).$$

Theorem 4.8 *Let $X \in \mathbb{R}^n$ be distributed as $ES(\theta, \Sigma)$. Then the mean of X exists if and only if the mean of $R = \|X - \theta\|_{\Sigma^{-1}}$ exists, in which case $E[X] = \theta$. The covariance matrix exists if and only if $E[R^2]$ is finite, in which case $cov(X) = E[R^2]\,\Sigma/n$.*

Theorem 4.9 *Let $X \in \mathbb{R}^n$ be distributed as $ES(\theta, \Sigma)$ with generating function g. Then the distribution of X is unimodal if and only if g is nonincreasing. Equivalently there exists a distribution $\nu \in \mathbb{R}_+$ with no point mass at* 0 *such that*

$$P[X \in A] = \int_{\mathbb{R}_+} \mathscr{V}_{r,\theta}^{\Sigma}(A)\, d\nu(r)$$

for any Borel set A of $\mathbb{R}^n$, where $\mathscr{V}_{r,\theta}^{\Sigma}$ is the uniform distribution on the ball (solid ellipse)

$$B_{r,\theta}^{\Sigma} = \{x \in \mathbb{R}^n \mid \|x - \theta\|_{\Sigma^{-1}} \leq r\}.$$

4.4 Marginal and Conditional Distributions for Spherically Symmetric Distributions

In this section, we study marginal and conditional distributions of spherically symmetric distributions. We first consider the marginal distributions for a uniform distribution on S_R.

Theorem 4.10 *Let $X = (X_1^{\mathrm{T}}, X_2^{\mathrm{T}})^{\mathrm{T}} \sim \mathscr{U}_R$ in $\mathbb{R}^n$ where dim $X_1 = p$ and dim $X_2 = n - p$. Then, for $1 \le p < n$, X_1 has an absolutely continuous spherically symmetric distribution with generating function g_R given by*

$$g_R(\|x_1\|^2) = \frac{\Gamma(n/2)\, R^{2-n}}{\Gamma((n-p)/2)\, \pi^{p/2}} \, (R^2 - \|x_1\|^2)^{(n-p)/2-1} \, \mathbb{1}_{B_R}(x_1). \tag{4.4}$$

Proof The proof is based on the fact that $RY/\|Y\| \sim \mathscr{U}_R$, for any random variable Y with a spherically symmetric distribution (see Lemma 1.2), in particular $\mathscr{N}_n(0, I_n)$, and on the fact that X_1 has an orthogonally invariant distribution in $\mathbb{R}^p$. To see this invariance, note that, for any $p \times p$ orthogonal matrix H_1 and any $(n-p) \times (n-p)$ orthogonal matrix H_2, the matrix

$$H = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix},$$

is a block diagonal $n \times n$ orthogonal matrix. Hence

$$H \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} H_1 X_1 \\ H_2 X_2 \end{pmatrix} \tag{4.5}$$

is distributed as $(x_1^{\mathrm{T}}, x_2^{\mathrm{T}})^{\mathrm{T}}$ and it follows that $H_1 X_1 \sim X_1$ and so X_1 is orthogonally invariant.

Therefore, if $Y = (Y_1^{\mathrm{T}}, Y_2^{\mathrm{T}})^{\mathrm{T}} \sim \mathscr{N}_n(0, I_n)$, then $\|Y_1\|^2$ is independent of $\|Y_2\|^2$ and, according to standard results, $Z = \|Y_1\|^2/\|Y\|^2$ has a beta distribution, that is $Beta(p/2, (n-p)/2)$. It follows that $Z' = \|X_1\|^2/\|X\|^2 = \|X_1\|^2/R^2$ has the same distribution since both $X/\|X\|$ and $Y/\|Y\|$ have distribution $\mathscr{U}_R$.

Thus $\|X_1\|^2 = R^2 Z'$ has a $Beta(p/2, (n-p)/2)$ density scaled by R^2. By a change of variable, the density of $\|X_1\|$ is equal to

$$h_R(r) = \frac{2}{B(p/2, (n-p)/2)} \, \frac{r^{p-1}(R^2 - r^2)^{(n-p)/2-1}}{R^{n-2}} \, \mathbb{1}_{(0,R)}(r).$$

Hence, by Theorem 4.2, X_1 has the density given by (4.4). □

Corollary 4.3 *Let $X = (X_1^{\mathrm{T}}, X_2^{\mathrm{T}})^{\mathrm{T}} \sim SS(\theta)$ in $\mathbb{R}^n$ where dim $X_1 = p$ and dim $X_2 = n - p$ and where $\theta = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})^{\mathrm{T}}$.*

Then, for $1 \leq p < n$, *the distribution of* X_1 *is an absolutely continuous spherically symmetric distribution* $SS(\theta_1)$ *on* $\mathbb{R}^p$ *with generating function given by* $\int g_R(\|X_1 - \theta_1\|^2)\, d\nu(R)$ *where* ν *is the radial distribution of* X *and* g_R *is given by* (4.4).

Unimodality properties of the densities of projections are given in the following result.

Corollary 4.4 *For the setup of Corollary* 4.3, *the density of* X_1 *is unimodal whenever* $n - p \geq 2$. *Furthermore, if* $p = n - 2$ *and* $X =\sim \mathscr{U}_{R,\theta}$, *then* X_1 *has the uniform distribution on* B_{R,θ_1} *in* $\mathbb{R}^{n-2}$.

In this book, we will have more need for the marginal distributions than the conditional distributions of spherically symmetric distributions. For results on conditional distributions, we refer the reader to Fang and Zhang (1990) and to Fang et al. (1990). We will however have use for the following result.

Theorem 4.11 *Let* $X = (X_1^{\mathrm{T}}, X_2^{\mathrm{T}})^{\mathrm{T}} \sim \mathscr{U}_{R,\theta}$ *in* $\mathbb{R}^n$ *where dim* $X_1 = p$ *and dim* $X_2 = n - p$ *and where* $\theta = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})^{\mathrm{T}}$. *Then the conditional distribution of* X_1 *given* X_2 *is the uniform distribution on the sphere in* $\mathbb{R}^p$ *of radius* $(R^2 - \|X_2 - \theta_2\|^2)^{1/2}$ *centered at* θ_1.

Proof First, it is clear that the support of the conditional distribution of X_1 given X_2 is the sphere in $\mathbb{R}^p$ of radius $(R^2 - \|X_2 - \theta_2\|^2)^{1/2}$ centered at θ_1. It suffices to show that the translated distribution centered at 0 is orthogonally invariant. To this end, note that, for any orthogonal transformation H on $\mathbb{R}^p$, the block diagonal transformation with blocks H and I_{n-p}, denoted by $\tilde{H}$, is orthogonal in $\mathbb{R}^n$. Then

$$\tilde{H}\big((X_1 - \theta_1)^{\mathrm{T}}, (X_2 - \theta_2)^{\mathrm{T}}\big)^{\mathrm{T}} \sim \big((X_1 - \theta_1)^{\mathrm{T}}, (X_2 - \theta_2)^{\mathrm{T}}\big)^{\mathrm{T}} \sim \mathscr{U}_{R,\theta}$$

that is,

$$\big((H(X_1 - \theta_1))^{\mathrm{T}}, (X_2 - \theta_2)^{\mathrm{T}}\big)^{\mathrm{T}} \sim \big((X_1 - \theta_1)^{\mathrm{T}}, (X_2 - \theta_2)^{\mathrm{T}}\big)^{\mathrm{T}} \sim \mathscr{U}_{R,\theta}\,.$$

Hence

$$H(X_1 - \theta_1)|(X_2 - \theta_2) \sim (X_1 - \theta_1)|(X_2 - \theta_2)\,,$$

and therefore, the distribution of X_1 given X_2 is orthogonally invariant, since θ_2 is fixed. The lemma follows. □

When properly interpreted, Corollaries 4.3 and 4.4 and Theorem 4.11 continue to hold for a general orthogonal projection π from $\mathbb{R}^n$ onto any subspace V of dimension p. See also Sect. 2.4.4 where the distribution is assumed to be normal.

4.5 The General Linear Model

This section is devoted to the general linear model, its canonical form and the issues of estimation, sufficiency and completeness.

4.5.1 The Canonical Form of the General Linear Model

Much of this book is devoted to some form of the following general problem. Let $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ be a partitioned random vector in $\mathbb{R}^n$ with a spherically symmetric distribution around a vector partitioned as $(\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$ where $\dim X = \dim \theta = p$ and $\dim U = \dim 0 = k$ with $p + k = n$. Such a distribution arises from a fixed orthogonally invariant random vector $(X_0^{\mathrm{T}}, U_0^{\mathrm{T}})^{\mathrm{T}}$ and a fixed scale parameter σ through the transformation

$$(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}} = \sigma\, (X_0^{\mathrm{T}}, U_0^{\mathrm{T}})^{\mathrm{T}} + (\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}\,, \tag{4.6}$$

so that the distribution of $((X - \theta)^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ is orthogonally invariant. We also refer to θ as a location parameter.

We will assume that the covariance matrix of $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ exists, which is equivalent to the finiteness of the expectation $E[R^2]$ where $R = (\|X - \theta\|^2 + \|U\|^2)^{1/2}$ is its radius (in this case, we have $\mathrm{cov}(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}} = E[R^2]\, I_n/n$). Then it will be convenient to assume that the radius $R_0 = (\|X_0\|^2 + \|U_0\|^2)^{1/2}$ of $(X_0^{\mathrm{T}}, U_0^{\mathrm{T}})^{\mathrm{T}}$ satisfies $E[R_0^2] = n$ since we have

$$\mathrm{cov}(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}} = \sigma^2\, \mathrm{cov}(X_0^{\mathrm{T}}, U_0^{\mathrm{T}})^{\mathrm{T}} = \sigma^2\, I_n\,.$$

Note that when it is assumed that the distribution in (4.6) is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$, the corresponding density may be represented as

$$\frac{1}{\sigma^n}\, g\left(\frac{\|z - \theta\|^2 + \|u\|^2}{\sigma^2}\right) \tag{4.7}$$

where g is the generating function.

This model also arises as the canonical form of the following seemingly more general model, the general linear model. For an $n \times p$ matrix V (often referred to as the design matrix and assumed here to be full rank p), suppose that an $n \times 1$ vector Y is observed such that

$$Y = V\beta + \varepsilon\,, \tag{4.8}$$

where β is a $p \times 1$ vector of (unknown) regression coefficients and ε is an $n \times 1$ vector with a spherically symmetric error distribution about 0. A common alternative representation of this model is $Y = \eta + \varepsilon$ where ε is as above and η is in the column space of V.

Using partitioned matrices, let $G = (G_1^{\mathrm{T}} \; G_2^{\mathrm{T}})^{\mathrm{T}}$ be an $n \times n$ orthogonal matrix partitioned such that the first p rows of G (*i.e.* the rows of G_1 considered as column vectors) span the column space of V. Now let

$$\begin{pmatrix} X \\ U \end{pmatrix} = G\,Y = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix} V\,\beta + G\,\varepsilon = \begin{pmatrix} \theta \\ 0 \end{pmatrix} + G\,\varepsilon \tag{4.9}$$

with $\theta = G_1 V\beta$ and $G_2 V\beta = 0$ since the rows of G_2 are orthogonal to the columns of V. It follows from the definition that $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ has a spherically symmetric distribution about $(\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$. In this sense, the model given in the first paragraph is the canonical form of the above general linear model.

This model has been considered by various authors such as Cellier et al. (1989), Cellier and Fourdrinier (1995), Maruyama (2003b), Maruyama and Strawderman (2005), and Fourdrinier and Strawderman (2010). Also, Kubokawa and Srivastava in (2001) addressed the multivariate case where θ is a mean matrix (in this case where X and U are matrices as well).

4.5.2 Least Squares, Unbiased and Shrinkage Estimation

Consider the model in (4.9). Since the columns of G_1^{T} (the rows of G_1) and the columns of V span the same space, there exists a nonsingular $p \times p$ matrix A such that

$$V = G_1^{\mathrm{T}} A, \text{ which implies } A = G_1 V, \tag{4.10}$$

since $G_1 G_1^{\mathrm{T}} = I_p$. So

$$\theta = A\beta, \text{ that is, } \beta = A^{-1}\theta\,. \tag{4.11}$$

Noting that $V^{\mathrm{T}} V = A^{\mathrm{T}}\, G_1\, G_1^{\mathrm{T}}\, A = A^{\mathrm{T}} A$, it follows that the estimation of θ by $\hat{\theta}(X, U)$ under the loss

$$L(\theta, \hat{\theta}) = (\hat{\theta} - \theta)^{\mathrm{T}}(\hat{\theta} - \theta) = \|\hat{\theta} - \theta\|^2 \tag{4.12}$$

is equivalent to the estimation of β by

$$\hat{\beta}(Y) = A^{-1}\,\hat{\theta}(G_1 Y, G_2 Y) = (G_1 V)^{-1}\,\hat{\theta}(G_1 Y, G_2 Y) \tag{4.13}$$

under the loss

$$L^*(\beta, \hat{\beta}) = (\hat{\beta} - \beta)^{\mathrm{T}} A^{\mathrm{T}} A(\hat{\beta} - \beta) = (\hat{\beta} - \beta)^{\mathrm{T}} V^{\mathrm{T}} V(\hat{\beta} - \beta) \tag{4.14}$$

in the sense that the resulting risk functions are equal,

$$R^*(\beta, \hat{\beta}) = E[L^*(\beta, \hat{\beta}(Y))] = E[L(\theta, \hat{\theta})] = R(\theta, \hat{\theta})\,.$$

Actually, the corresponding loss functions are equal. To see this, note that

$$\begin{aligned} L^*(\beta, \hat{\beta}(Y)) &= (\hat{\beta}(Y) - \beta)^{\mathrm{T}} A^{\mathrm{T}} A(\hat{\beta}(Y) - \beta) \\ &= (A(\hat{\beta}(Y) - \beta))^{\mathrm{T}} (A(\hat{\beta}(Y) - \beta)) \\ &= (\hat{\theta}(X, U) - \theta)^{\mathrm{T}} (\hat{\theta}(X, U) - \theta) \\ &= L(\theta, \hat{\theta}(X, U))\,, \end{aligned}$$

where (4.13) and (4.11) were used for the third equality.

Note that the above equivalence between the estimation of θ, the mean vector of X, and the estimation of the regression coefficients β also holds for the respective invariant losses

$$L(\theta, \hat{\theta}, \sigma^2) = \frac{1}{\sigma^2}\,(\hat{\theta} - \theta)^{\mathrm{T}} (\hat{\theta} - \theta) = \frac{1}{\sigma^2}\,\|\hat{\theta} - \theta\|^2 \tag{4.15}$$

and

$$L^*(\beta, \hat{\beta}, \sigma^2) = \frac{1}{\sigma^2}\,(\hat{\beta} - \beta)^{\mathrm{T}} A^{\mathrm{T}} A(\hat{\beta} - \beta) = \frac{1}{\sigma^2}\,(\hat{\beta} - \beta)^{\mathrm{T}} V^{\mathrm{T}} V(\hat{\beta} - \beta)\,. \tag{4.16}$$

Additionally, the correspondence (4.13) can be reversed as

$$\hat{\theta}(X, U) = A\,\hat{\beta}(G_1^{\mathrm{T}} X + G_2^{\mathrm{T}} U) = G_1 X\,\hat{\beta}(G_1^{\mathrm{T}} X + G_2^{\mathrm{T}} U) \tag{4.17}$$

since, according to (4.9),

$$Y = G^{\mathrm{T}} \begin{pmatrix} X \\ U \end{pmatrix} = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} X \\ U \end{pmatrix} = (G_1^{\mathrm{T}}\, G_2^{\mathrm{T}}) \begin{pmatrix} X \\ U \end{pmatrix} = G_1^{\mathrm{T}} X + G_2^{\mathrm{T}} U\,. \tag{4.18}$$

There is also a correspondence between the estimation of θ and the estimation of η in the following alternative representation of the general linear model. Here

$$\eta = G^{\mathrm{T}} \begin{pmatrix} \theta \\ 0 \end{pmatrix} = \begin{pmatrix} G_1 \\ G_2 \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \theta \\ 0 \end{pmatrix} = (G_1^{\mathrm{T}}\, G_2^{\mathrm{T}}) \begin{pmatrix} \theta \\ 0 \end{pmatrix} = G_1^{\mathrm{T}} \theta + G_2^{\mathrm{T}} 0 = G_1^{\mathrm{T}} \theta$$

and

$$G_1\eta = G_1\, G_1^{\mathrm{T}}\,\theta = \theta\,.$$

It follows that the estimation of $\theta \in \mathbb{R}^p$ by $\hat{\theta}(X, U)$ under the loss $\|\hat{\theta} - \theta\|^2$ (the loss (4.12)) is equivalent to the estimation of η in the column space of V under the loss $\|\hat{\eta} - \eta\|^2$ by

$$\hat{\eta}(Y) = G_1^{\mathrm{T}}\,\hat{\theta}(G_1 Y, G_2 Y) \tag{4.19}$$

in the sense that the risks functions are equal. The easy demonstration is left to the reader.

Consider the first correspondence expressed in (4.13) and (4.17) between estimators in Models (4.8) and (4.9). We will see that it can be made completely explicit for a wide class of estimators. First, note that the matrix G_1 can be easily obtained by the Gram-Schmidt orthonormalization process or by the QR decomposition of the design matrix X, where Q is an orthogonal matrix such that $Q^{\mathrm{T}}V = R$ and R is an $n \times p$ upper triangular matrix (so that $G_1 = Q_1^{\mathrm{T}}$ and $G_2 = Q_2^{\mathrm{T}}$). Second, a particular choice of A can be made that gives rise to a closed form of G_1.

To see this, let

$$A = (V^{\mathrm{T}}V)^{1/2} \tag{4.20}$$

(a square root of $V^{\mathrm{T}}V$, which is invertible since V has full rank) and set

$$G_1 = A\,(V^{\mathrm{T}}V)^{-1}V^{\mathrm{T}} = (V^{\mathrm{T}}V)^{-1/2}V^{\mathrm{T}}\,. \tag{4.21}$$

Then we have

$$G_1\,V = A, \qquad V = G_1^{\mathrm{T}}A, \tag{4.22}$$

and

$$G_1\,G_1^{\mathrm{T}} = (V^{\mathrm{T}}V)^{-1/2}V^{\mathrm{T}}V(V^{\mathrm{T}}V)^{-1/2} = I_p\,. \tag{4.23}$$

Hence, as in (4.10), (4.22) expresses that the columns of G_1^{T} (the rows of G_1) span the same space as the columns of V, noticing that (4.23) means that these vectors are orthogonal. Therefore, completing G_1^{T} through the Gram-Schmidt orthonormalization process, we obtain an orthogonal matrix $G = (G_1^{\mathrm{T}}G_2^{\mathrm{T}})^{\mathrm{T}}$, with G_1 in (4.21), such that

$$G\,V = \begin{pmatrix} (V^{\mathrm{T}}V)^{1/2} \\ 0 \end{pmatrix}\,. \tag{4.24}$$

The relationship linking A and G_1 in (4.21) is an alternative to (4.10) and is true in general; that is,

$$G_1 = A\,(V^{\mathrm{T}}V)^{-1}\,V^{\mathrm{T}} \text{ or equivalently } A = (V^{\mathrm{T}}G_1^{\mathrm{T}})^{-1}V^{\mathrm{T}}V\,.$$

Indeed, we have $V^{\mathrm{T}}V = A^{\mathrm{T}}A$ so that $(V^{\mathrm{T}}V)^{-1}\,(A^{\mathrm{T}}A) = I_p$. Hence, $(V^{\mathrm{T}}V)^{-1}\,A^{\mathrm{T}} = A^{-1}$, which implies $V\,(V^{\mathrm{T}}V)^{-1}\,A^{\mathrm{T}} = V\,A^{-1} = G_1^{\mathrm{T}}A\,A^{-1} = G_1^{\mathrm{T}}$, according to (4.10).

As a consequence, if $\hat{\beta}_{ls}$ is the least squares estimator of β, we have

$$\hat{\beta}_{ls}(Y) = (V^{\mathrm{T}}V)^{-1}\,V^{\mathrm{T}}\,Y \tag{4.25}$$

so that the corresponding estimator $\hat{\theta}_0$ of θ is the projection $\hat{\theta}_0(X, U) = X$ since

$$\hat{\theta}_0(X, U) = A\,\hat{\beta}_{ls}(Y) = A\,(V^{\mathrm{T}}V)^{-1}V^{\mathrm{T}}\,Y = G_1\,Y = X\,. \tag{4.26}$$

From this correspondence, the estimator $\hat{\theta}_0(X, U) = X$ of θ is often viewed as the standard estimator. Note that, with the choice of A in (4.20), we have the closed form

$$\hat{\beta}_{ls}(Y) = (V^{\mathrm{T}}V)^{-1/2}X\,. \tag{4.27}$$

Furthermore, the correspondence between $\hat{\theta}(X, U)$ and $\hat{\beta}_{ls}(Y)$ can be specified when $\hat{\theta}(X, U)$ depends on U only through $\|U\|^2$, in which case, with a slight abuse of notation, we write $\hat{\theta}(X, U) = \hat{\theta}(X, \|U\|^2)$. Indeed, first note that

$$\begin{aligned}
\|X\|^2 &= (A\,\hat{\beta}_{ls}(Y))^{\mathrm{T}}(A\,\hat{\beta}_{ls}(Y)) \\
&= (\hat{\beta}_{ls}(Y))^{\mathrm{T}}\,A^{\mathrm{T}}A\,(\hat{\beta}_{ls}(Y)) \\
&= (\hat{\beta}_{ls}(Y))^{\mathrm{T}}\,V^{\mathrm{T}}V\,(\hat{\beta}_{ls}(Y)) \\
&= (V\,\hat{\beta}_{ls}(Y))^{\mathrm{T}}V\,(\hat{\beta}_{ls}(Y)) \\
&= \|V\,\hat{\beta}_{ls}(Y)\|^2\,.
\end{aligned} \tag{4.28}$$

On the other hand, according to (4.9), we have $\|X\|^2 + \|U\|^2 = \|G\,Y\|^2 = \|Y\|^2$. Hence,

$$\begin{aligned}
\|U\|^2 &= \|Y\|^2 - \|X\|^2 \\
&= \|Y\|^2 - \|V\,\hat{\beta}_{ls}(Y)\|^2 \\
&= \|Y - V\,\hat{\beta}_{ls}(Y)\|^2
\end{aligned} \tag{4.29}$$

since $Y - V\,\hat{\beta}_{ls}(Y)$ is orthogonal to $V\,\hat{\beta}_{ls}(Y)$. Consequently, according to (4.13) and (4.10), Equations (4.29) and (4.26) give that the estimator $\hat{\beta}(Y)$ of β corresponding to the estimator $\hat{\theta}(X, \|U\|^2)$ of θ is

$$\hat{\beta}(Y) = (G_1\, V)^{-1}\, \hat{\theta}(G_1\, V\, \hat{\beta}_{ls}(Y), \|Y - V\, \hat{\beta}_{ls}(Y)\|^2)\,. \tag{4.30}$$

Note that, when ones chooses G_1 as in (4.21), $\hat{\beta}(Y)$ in (4.30) has the closed form

$$\begin{aligned}\hat{\beta}(Y) &= (V^{\mathrm{T}}\, V)^{-1/2}\, \hat{\theta}\Big((V^{\mathrm{T}}\, V)^{1/2}\, \hat{\beta}_{ls}(Y), \|Y - V\, \hat{\beta}_{ls}(Y)\|^2\Big) \\ &= (V^{\mathrm{T}}\, V)^{-1/2}\, \hat{\theta}\Big((V^{\mathrm{T}}\, V)^{-1/2} V^{\mathrm{T}} Y, \|Y - V\, \hat{\beta}_{ls}(Y)\|^2\Big)\,. \end{aligned} \tag{4.31}$$

In particular, we can see through (4.28), that the "robust" Stein-type estimators of θ,

$$\hat{\theta}_r(X, \|U\|^2) = \left(1 - a\, \frac{\|U\|^2}{\|X\|^2}\right) X \tag{4.32}$$

have as a correspondence the "robust" estimators of β

$$\begin{aligned}\hat{\beta}_r(Y) &= (G_1\, V)^{-1} \left(1 - a\, \frac{\|Y - V\, \hat{\beta}_{ls}(Y)\|^2}{\|V\, \hat{\beta}_{ls}(Y)\|^2}\right) G_1\, V\, \hat{\beta}_{ls}(Y) \\ &= \left(1 - a\, \frac{\|Y - V\, \hat{\beta}_{ls}(Y)\|^2}{\|V\, \hat{\beta}_{ls}(Y)\|^2}\right) \hat{\beta}_{ls}(Y) \end{aligned} \tag{4.33}$$

(note that the two $G_1\, V$ terms simplify). We use the term "robust" since, for appropriate values of the positive constant a, they dominate X whatever the spherically symmetric distribution, as we will see in Chap. 5 (see also Cellier et al. 1989; Cellier and Fourdrinier 1995).

According to the correspondence seen above between the risk functions of the estimators of θ and the estimators of β, using these estimators in (4.33) is then a good alternative to the least squares estimator: they dominate the least squares estimator of β simultaneously for all spherically symmetric error distributions with a finite second moment (see Fourdrinier and Strawderman (1996) for the use of these robust estimators when σ^2 is known and also Sect. 5.2).

4.5.3 Sufficiency in the General Linear Model

Suppose $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ has a spherically symmetric distribution about $(\theta^{\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$ with $\dim X = \dim \theta = p > 0$ and $\dim U = \dim 0 = k > 0$. Furthermore, suppose that the distribution is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ for $n = p + k$. The corresponding density may be represented as in (4.7). We refer to θ as a location vector and to σ as a scale parameter. As seen in the previous section, such a distribution arises from a fixed orthogonally invariant random vector $(X_0^{\mathrm{T}}, U_0^{\mathrm{T}})^{\mathrm{T}}$ with generating function g through the transformation

$$\begin{pmatrix} X \\ U \end{pmatrix} = \sigma \begin{pmatrix} X_0 \\ U_0 \end{pmatrix} + \begin{pmatrix} \theta \\ 0 \end{pmatrix}.$$

Each of θ, σ^2 and $g(\cdot)$ may be known or unknown, but perhaps the most interesting case from a statistical standpoint is the following.

Suppose θ and σ^2 are unknown and $g(\cdot)$ is known. It follows immediately from the factorization theorem that $(X, \|U\|^2)$ is sufficient . It is intuitively clear that this statistic is also minimal sufficient since $\dim(X, \|U\|^2) = \dim(\theta, \sigma^2)$. Here is a proof of that fact.

Theorem 4.12 *Suppose that* $(X^{\mathrm{T}}, U^{\mathrm{T}})^{\mathrm{T}}$ *is distributed as* (4.7). *Then the statistic* $(X, \|U\|^2)$ *is minimal sufficient for* (θ, σ^2) *when* g *is known.*

Proof By Theorem 6.14 of Casella and Berger (2001), it suffices to show that if, for all $(\theta, \sigma^2) \in \mathbb{R}^p \times \mathbb{R}_+$,

$$\frac{g\left(\frac{\|x_1-\theta\|^2+\|u_1\|}{\sigma^2}\right)}{g\left(\frac{\|x_2-\theta\|^2+\|u_2\|^2}{\sigma^2}\right)} = c \tag{4.34}$$

where c is a constant then $x_1 = x_2$ and $\|u_1\|^2 = \|u_2\|^2$. Note that $0 < c < \infty$ since otherwise (4.7) cannot be a density.

Letting $\tau^2 = 1/\sigma^2$, (4.34) can be written, for all $\tau > 0$, as

$$g(\tau^2 v_1^2) = c g(\tau^2 v_2^2) \tag{4.35}$$

where $v_1^2 = \|x_1 - \theta\|^2 + \|u_1\|^2$ and $v_2^2 = \|x_2 - \theta\|^2 + \|u_2\|^2$ for each fixed $\theta \in \mathbb{R}^p$. First, we will show that $v_1^2 = v_2^2$. Note that

$$\begin{aligned} 1 &= \int_{\mathbb{R}^p \times \mathbb{R}^k} g(\|x\|^2 + \|u\|^2)\, dx\, du \\ &= K \int_0^\infty r^{p+k-1} g(r^2)\, dr \qquad \text{(by Theorem 4.2)} \\ &= K \upsilon^{p+k} \int_0^\infty \tau^{p+k-1} g(\upsilon^2 \tau^2)\, d\tau \end{aligned} \tag{4.36}$$

for any $\upsilon > 0$. Then it follows from (4.35) and (4.36) that

$$\begin{aligned} 1 &= K v_1^{p+k} \int_0^\infty \tau^{p+k-1} g(v_1^2 \tau^2)\, d\tau \\ &= c K v_1^{p+k} \int_0^\infty \tau^{p+k-1} g(v_2^2 \tau^2)\, d\tau \\ &= c \frac{v_1^{p+k}}{v_2^{p+k}}. \end{aligned} \tag{4.37}$$

Let $F(b) = \int_0^b \tau^{p+k-1} g(\tau^2)\, d\tau$ and choose b such that F is strictly increasing at b. Suppose $v_1 > v_2$. Then, for any $v > 0$,

$$F(b) = v^{p+k} \int_0^{b/v} \tau^{p+k-1}\, g(v^2\tau^2)\, d\tau$$

and consequently

$$\begin{aligned}
\int_0^{b/v_1} \tau^{p+k-1}\, g(v_1^2\tau^2) d\tau &= \frac{F(b)}{v_1^n} \\
&= c \int_0^{b/v_1} \tau^{p+k-1}\, g(v_2^2\tau^2) d\tau \\
&< c \int_0^{b/v_2} \tau^{p+h-1}\, g(v_2^2\tau^2) d\tau \\
&= c\, \frac{F(b)}{v_2^n}.
\end{aligned}$$

It follows that $c\, v_1^{p+k}/v_2^{p+h} > 1$, which contradicts (4.37). A similar argument would give $c\, v_1^{p+k}/v_2^{p+h} < 1$ for $v_1 < v_2$ and $v_1 = v_2$. Now, setting $\theta = \frac{x_1+x_2}{2}$ in the expressions for v_1 and v_2 implies $\|u_1\|^2 = \|u_2^2\|$. It then follows that $\|x_1 - \theta\|^2 = \|x_2 - \theta\|^2$ for all $\theta \in \mathbb{R}^p$, which implies $x_1 = x_2$ by setting $\theta = x_2$ (or x_1). □

In the case where θ is unknown, σ^2 is known, and the distribution is multivariate normal, X is minimal sufficient (and complete). However, in the non-normal case, $(X, \|U\|^2)$ is typically minimally sufficient, and may or may not be complete, which is the subject of the next section.

4.5.4 Completeness for the General Linear Model

The section largely follows the development in Fourdrinier et al. (2014). In the case where both θ and σ^2 are unknown and g is known, the minimal sufficient statistic $(X, \|U\|^2)$ can be either complete or incomplete depending on g. If g corresponds to a normal distribution, the statistic is complete by standard results for exponential families. However, when the generating function is of the form $K\, g(t) \mathbb{1}_{(r_1,r_2)}(t)$ with $0 < r_1 < r_2 < \infty$ and K is the normalizing constant, $(X, \|U\|^2)$ is not complete. In fact incompleteness of $(X, \|U\|^2)$ follows from the fact that the minimal sufficient statistic, when θ is known, σ^2 is unknown and g is known, is incomplete.

Theorem 4.13

(1) *If $X \sim f(x-\theta)$ with $\theta \in \mathbb{R}^p$ where f has compact support, then X is not complete for θ.*
(2) *If $X \sim 1/\sigma f(x/\sigma)$, where f has support contained in an interval $[a, b]$ with $0 < a < b < \infty$, then X is not complete for σ.*

Before giving the proof of Theorem 4.13, note that if the generating function is of the form $K\, g(t)\mathbb{1}_{[r_1,r_2]}(t)$ for $0 < r_1 < r_2 < \infty$ and the value of θ is assumed to be known and equal to θ_0, then $T = \|X-\theta_0\|^2 + \|U\|^2$ is minimal sufficient and has density of the form $K/\sigma^{p+k}\, t^{(p+k)/2}\, g(t/\sigma^2)\mathbb{1}_{[r_1\sigma^2, r_2\sigma^2]}(T)$.

Therefore, T is not a complete statistic for σ^2 by Lemma 4.13 (2). It follows that there exists a function $h(\cdot)$ not equal to zero a.e. such that $E_\sigma[h(T)] = 0$ for all $\sigma > 0$. Since $E_{\sigma^2}[h(\beta T)] = E_{\beta\sigma^2}[h(T)]$, it follows that $E_{\sigma^2}[h(\beta T)] = 0$ for all $\sigma^2 > 0, \beta > 0$, and also that $M(t) = \int_0^1 E_{\sigma^2}[h(\beta t)]\, m(\beta)\, d\beta = 0$ for any function $m(\cdot)$ for which the integral exists. In particular, this holds when $m(\cdot)$ is the density of a Beta$(k/2, p/2)$ random variable (where finiteness of the integral is guaranteed since $E_{\sigma^2}[h(\beta T)]$ is continuous in β). Now, since $B = \|U\|^2/T$ has a Beta$(k/2, p/2)$ distribution, $\|U\|^2 = BT$, and $M(\sigma^2) = E_{\sigma^2}[h(BT) = E_{\sigma^2}[h(\|U\|^2)] \equiv 0$.

Since the distribution of $\|U\|^2$ does not depend on θ, it follows that when both θ and σ^2 are unknown, $E_{\theta,\sigma^2}[h(\|U\|^2)] \equiv 0$. Hence, $(X, \|U\|^2)$, while minimal sufficient, is not complete for the case of a generating function of the form $g(t)\mathbb{1}_{[r_1,r_2]}(T)$ with $0 < r_1 < r_2 < \infty$.

Note that whenever θ is unknown, σ^2 is known, and $(X, \|U\|^2)$ is minimal sufficient (so the distribution is not normal, since then X would be minimal sufficient) $\|U\|^2$ is ancillary and the minimal sufficient statistic is not complete.

Proof of Theorem 4.13 First, note that part (2) follows from part (1) by the standard technique of transforming a scale family to a location family by taking logs.

We will show the incompleteness of a location family in $\mathbb{R}$ when F has bounded support. We show first that, if F is a cdf with bounded support contained inside $[a, b]$, the characteristic function (c.f.) $\hat{f}$ is analytic in $\mathbb{C}$ (the entire complex plane) and is of order 1 (i.e., $|\hat{f}(\eta)|$ is $O\exp(|\eta|^{1+\varepsilon})$ for all $\epsilon > 0$ and is not $O(\exp(|\eta|^{1-\varepsilon})$ for any $\epsilon > 0$).

To see this, without loss of generality assume $0 < a < b < \infty$. Then

$$\begin{aligned}
|\hat{f}(\eta)| &\le \int_a^b \exp(|\eta|X)\, dF(x) \\
&\le \exp(b|\eta|) \int_a^b dF(x) \\
&= \exp(b|\eta|) \\
&= O(\exp(|\eta|^{1+\varepsilon})).
\end{aligned}$$

for all $\varepsilon > 0$. Also, if $\eta = -iv$ for $v > 0$, then

$$|\hat{f}(\eta)| = \int_a^b \exp(vx)\, dF(x) \geq \exp(av) \int_a^b dF(x) = \exp(av)\,.$$

However, $\exp(av)$ is not $O(\exp(v^{1-\varepsilon})$ for any $\varepsilon > 0$. Hence $\hat{f}(\eta)$ is of order 1.

In the step above, we used $0 < a < b < \infty$. Note that if either a and/or b is negative then the distribution of X is equal to the distribution of $\mathrm{z} + \theta_0$ where θ_0 is negative and where the distribution of z satisfies the assumptions of the theorem. Hence $E\exp(i\eta x) = E\exp(i\eta z)e^{i\eta\theta_0}$, so $|E\exp(i\eta x)| \leq \exp(|\eta|b)\exp(|i\eta||\theta_0|)$ which is $O\exp(|\eta|^{1+\epsilon})$ for all $\epsilon > 0$.

Similarly, for $\eta = -iv$ (recall $\theta_0 < 0$),

$$\begin{aligned} |E\exp(i\eta x)| &= E\exp(tvz)\ \exp(-v\theta_0) \\ &\geq e^{v|\theta_0|}\exp(av) \\ &= \exp(v(a + |\theta_0|) \end{aligned}$$

and this is not $O(\exp^{v^{1-\varepsilon}})$ for any $\varepsilon > 0$. □

Note that $\hat{f}(\eta)$ exists in all of $\mathbb{C}$ since F has bounded support and is analytic by standard results in complex analysis (See e.g. Rudin 1966). To complete the proof of Theorem 4.13 we need the following lemma.

Lemma 4.3 *If $X \sim F(x)$ where the cdf F has bounded support in $\mathbb{R}$ and F is not degenerate, then the characteristic function $\hat{f}(\eta)$ has at least one zero in $\mathbb{C}$.*

Proof This follows almost directly from the Hadamard factorization theorem which implies that a function $\hat{f}(z)$ that is analytic in all of $\mathbb{C}$ and of order 1 is of the form $\hat{f}(z) = \exp(az + b)P(z)$. $P(z)$ is the so called canonical product formed from the zeros of $\hat{f}(z)$, where $P(0) = 1$ and $P(z) = 0$ for each such root. (See e.g., Titchmarsh (1932) for an extended discussion of the form of $P(z)$). Therefore, either $\hat{f}(z)$ has no zeros, in which case $\hat{f}(z) = \exp(az)$ (since $\hat{f}(0) = 1 = e^b \Rightarrow b = 0$) and $P(z) \equiv 1$, or $\hat{f}(z)$ has at least one zero. The case where $\hat{f}(z) = \exp(az)$ corresponds to the degenerate case where $\exp(az) = \hat{f}(z) = E\exp(izx)$ with $P[X = -ia] = 1$. Since F is assumed to not be degenerate, $\hat{f}(z)$ must have at least one zero by the uniqueness of the Fourier transform.

To finish the proof of Theorem 4.13 note that, by By Lemma 4.3, there exists an η_0 such that

$$\hat{f}(\eta_0) = \int_{-\infty}^{\infty} \exp(i\eta_0 x) f(x)\, dx = 0.$$

This implies that for any $\theta \in \mathbb{R}$,

$$
\begin{aligned}
0 &= \left(\int_{-\infty}^{\infty} \exp(i\eta_0 x) f(x)\, dx\right) \exp^{i\eta_0\theta} \\
&= \int_{-\infty}^{\infty} \exp(ix(\eta_0+\theta)) f(x)\, dx \\
&= \int_{-\infty}^{\infty} \exp(i\eta_0 x) f(x-\theta)\, dx \\
&= E_\theta[\exp(i\eta_0 X) = E_\theta[\exp(i(a_0+b_0 i))X] \\
&= E_\theta \exp(i\eta_0 X)\exp(-b_0 X)] \\
&= E_\theta[\exp(-b_0 X)\{\cos a_0 x + i \sin a_0 x\}].
\end{aligned}
$$

Hence, for any $\theta \in \mathbb{R}$, we have $E_\theta[\exp(-b_0 x)\cos(a_0 x)] \equiv 0$.

Additionally, $E_\theta[|\exp(-b_0 x)\cos(a_0 x)|] < \infty$ for all θ since $f(\cdot)$ has bounded support. The theorem then follows, since $h(X) = e^{-b_0 X}\cos a_0 X$ is an unbiased estimator of 0, which is not equal to 0 almost surely for each θ. This proves the result for $p = 1$. The extension from $\mathbb{R}$ to $\mathbb{R}^p$ is straightforward since the marginal distribution of each coordinate has compact support. □

4.6 Characterizations of the Normal Distribution

There is a large literature on characterizations of the normal distribution that has had a long history. A classical reference that covers a number of characterizations of the normal distribution is Kagan et al. (1973). We give only a small sample of these characterizations. The first result gives a characterization in terms of the normality of linear transformations.

Theorem 4.14 *Let $X \sim ES(\theta)$ in $\mathbb{R}^n$. If A is any fixed linear transformation of positive rank such that AX has a normal distribution then X has a normal distribution.*

Proof First note that it suffices to consider the case $\theta = 0$. Furthermore it suffices to prove the result for $X \sim SS(0)$ since an elliptically symmetric distribution is the image of a spherically symmetric distribution by a nonsingular transformation. Note also that, if $X \sim SS(0)$, its characteristic function $\varphi_X(t) = \Psi(t^{\mathrm{T}}t)$ since, for any orthogonal transformation H, the characteristic function φ_{HX} of HX satisfies

$$\varphi_{HX}(t) = \varphi_X(H^{\mathrm{T}}t) = \varphi_X(t).$$

Now the characteristic function φ_{AX} of AX equals

$$\varphi_{AX}(t) = E[\exp\{it^{\mathrm{T}}AX\}] = E[\exp\{i(A^{\mathrm{T}}t)^{\mathrm{T}}X\}] = \Psi(t^{\mathrm{T}}AA^{\mathrm{T}}t). \tag{4.38}$$

Also, by Theorem 4.3, $\mathrm{Cov}(X) = E[R^2]/nI_n$. Hence $\mathrm{Cov}(AX) = (E[R^2]/n)\,AA^{\mathrm{T}}$ and the fact that AX is normal implies that $E[R^2] < \infty$ and that $\mathrm{Cov}(AX) = \alpha\, AA^{\mathrm{T}}$ for $\alpha \geq 0$. This implies that $\varphi_{AX}(t) = \exp\{-\alpha\, t^{\mathrm{T}} AA^{\mathrm{T}} t/2\}$. Therefore, by (4.38), $\Psi(z) = \exp\{-\alpha z/2\}$ and $\varphi_X(t) = \exp\{-\alpha t^{\mathrm{T}} t/2\}$, so X is normal. □

Corollary 4.5 *Let $X \sim ES(\theta)$ in $\mathbb{R}^n$. If any orthogonal projection Π has a normal distribution (and, in particular, any marginal), then X has a normal distribution.*

The next theorem gives a characterization in terms of the independence of linear projections.

Theorem 4.15 *Let $X \sim ES(\theta)$ in $\mathbb{R}^n$. If A and B are any two fixed linear transformations of positive rank such that AX and BX are independent, then X has a normal distribution.*

Proof As in the proof of Theorem 4.14, we can assume that $X \sim SS(0)$. Then the characteristic function φ_X of X is $\varphi_X(t) = \Psi(t^{\mathrm{T}} t)$. Hence, the characteristic functions φ_{AX} and φ_{BX} of AX and BX are $\varphi_{AX}(t_1) = \Psi(t_1^{\mathrm{T}} AA^{\mathrm{T}} t_1)$ and $\varphi_{BX}(t_2) = \Psi(t_2^{\mathrm{T}} BB^{\mathrm{T}} t_2)$, respectively. By the independence of AX and BX, we have

$$\Psi(t_1^{\mathrm{T}} AA^{\mathrm{T}} t_1 + t_2^{\mathrm{T}} BB^{\mathrm{T}} t_2) = \Psi(t_1^{\mathrm{T}} AA^{\mathrm{T}} t_1)\Psi(t_2^{\mathrm{T}} BB^{\mathrm{T}} t_2).$$

Since A and B are of positive rank this implies that, for any $u \geq 0$ and $v \geq 0$,

$$\Psi(u + v) = \Psi(u)\Psi(v).$$

This equation is known as Hamel's equation and its only continuous solution is $\Psi(u) = e^{\alpha u}$ for some $\alpha \in \mathbb{R}$ (see for instance Feller 1971, page 305). Hence, $\varphi_X(t) = e^{\alpha t^{\mathrm{T}} t}$ for some $\alpha \leq 0$ since φ_X is a characteristic function. It follows that X has a normal distribution. □

Corollary 4.6 *Let $X \sim ES(\theta)$ in $\mathbb{R}^n$. If any two projections (in particular, any two marginals) are independent, then X has a normal distribution.*