STEIN ESTIMATION IN HIGH DIMENSIONS AND THE BOOTSTRAP[1]

Rudolf Beran
University of California, Berkeley

The Stein estimator $\hat{\xi}_S$ and the better positive-part Stein estimator $\hat{\xi}_{PS}$ both dominate the sample mean, under quadratic loss, in the $N(\xi, I)$ model of dimension $q \geq 3$. Standard large sample theory does not explain this phenomenon well. Plausible bootstrap estimators for the risk of $\hat{\xi}_S$ do not converge correctly at the shrinkage point as sample size n increases. By analyzing a submodel exactly, with the help of results from directional statistics, and then letting dimension $q \to \infty$, we find:

- In high dimensions, $\hat{\xi}_S$ and $\hat{\xi}_{PS}$ are approximately admissible and approximately minimax on large compact balls about the shrinkage point. The sample mean is neither.
- A new estimator of ξ, asymptotically equivalent to $\hat{\xi}_{PS}$ as $q \to \infty$, appears to dominate $\hat{\xi}_{PS}$ slightly.
- Resampling from a $N(\hat{\xi}, I)$ distribution, where $|\hat{\xi}|^2$ estimates $|\xi|^2$ well, is the key to consistent bootstrap risk estimation for orthogonally equivariant estimators of ξ. Choosing $\hat{\xi}$ to be the Stein estimator or the positive-part Stein estimator or the sample mean does not work.
- Estimators of $|\xi|$ are subject to a sharp local asymptotic minimax bound as q increases.

AMS 1991 subject classification: Primary 62J07; secondary 62C15.

Key words and phrases: Admissible, minimax, high dimension, orthogonal group, equivariant, directional statistics.

[1] This research was supported in part by NSF grant DMS 9001710. Some of the author's work was done as a guest of Sonderforschungsbereich 123 at Universität Heidelberg.

1. Introduction.

For the mean vector ξ of a q-variate $N(\xi, I)$ distribution, the sample mean is not an admissible estimator, under squared-error loss, when dimension $q \geq 3$. First proved by C. Stein (1956) and subsequently sharpened in James and Stein (1961), this remarkable result came as a surprise to the statistical community. Notable contributions to our understanding of the Stein phenomenon include Stein (1962), Brown (1966), Baranchik (1970), Strawderman (1972), Efron and Morris (1973), Stein (1981), Berger and Wolpert (1983). A valuable survey article is Brandwein and Strawderman (1990).

Large sample theory has difficulty in explaining what Stein estimation is about. Suppose $Y_1, Y_2, \ldots, Y_n$ are i.i.d. random q-vectors, each having a $N(\xi, I)$ distribution with ξ unknown and $q \geq 3$. Let $|\cdot|$ denote euclidean norm on R^q and let $\bar{Y}_n$ denote the sample mean vector. The basic Stein estimator

$$\hat{\xi}_S = \left[1 - \frac{q-2}{n|\bar{Y}_n|^2}\right]\bar{Y}_n \tag{1.1}$$

has risk

$$R_{q,n}(\hat{\xi}_S, \xi) = q^{-1} n E_\xi |\hat{\xi}_S - \xi|^2 = 1 - E_\xi\left[\frac{(q-2)^2/q}{n|\bar{Y}_n|^2}\right], \tag{1.2}$$

which is strictly less than the risk of $\bar{Y}_n$ at every ξ and equals $2/q$ at the shrinkage point $\xi = 0$ (James and Stein 1961).

Standard large sample theory tells us that, as $n \to \infty$ with q fixed:

(a) Both $\hat{\xi}_S$ and $\bar{Y}_n$ are locally asymptotically minimax estimators at every ξ, in the sense of Hájek (1972) and LeCam (1972).

(b) Both $\hat{\xi}_S$ and $\bar{Y}_n$ are Hájek regular at every $\xi \neq 0$ and are asymptotically least dispersed among such regular estimators of ξ, by virtue of Hájek's (1970) convolution theorem.

(c) At the shrinkage point $\xi = 0$, $\bar{Y}_n$ is still Hájek regular but $\hat{\xi}_S$ is not (van der Vaart 1988).

(d) At every $\xi \neq 0$, the risk of $\hat{\xi}_S$ improves upon the risk of $\bar{Y}_n$ by $O(n^{-1})$ (Ibragimov and Has'minskii 1981).

None of these results explain the form of $\hat{\xi}_S$ or whether significant improvement on $\hat{\xi}_S$ is possible.

Related to the lack of Hájek regularity in point (c) is the inconsistency at the shrinkage point, as $n \to \infty$, of plausible bootstrap estimators for the risk $R_{q,n}(\hat{\xi}_S, \xi)$. Writing $R_n(\xi)$ for this risk, two natural parametric bootstrap estimators are $R_n(\hat{\xi}_S)$ and $R_n(\bar{Y}_n)$. These correspond to resampling from the $N(\hat{\xi}_S, I)$ and $N(\bar{Y}_n, I)$ distributions respectively. Consider an arbitrary sequence $\{\xi_n \in R^q\}$ such that $n^{1/2}(\xi_n - \xi) \to h$, a fixed finite q-vector. Let Z denote a standard normal random q-vector. Then

$$\lim_{n\to\infty} R_n(\xi_n) = \begin{cases} 1 \text{ if } \xi \neq 0 \\ w(h) \text{ if } \xi = 0 \end{cases} \tag{1.3}$$

where

$$w(h) = 1 - E\left[\frac{(q-2)^2/q}{|Z+h|^2}\right] \tag{1.4}$$

Thus, when $\xi = 0$ and q is fixed as $n \to \infty$, the bootstrap risk estimator $R_n(\hat{\xi}_S)$ converges *in distribution* to the non-degenerate random variable $w[(1-(q-2)/|Z|^2)Z]$, rather than to the correct risk $w(0)$. Similarly, the alternative bootstrap risk estimator $R_n(\bar{Y}_n)$ converges in distribution to the non-degenerate random variable $w(Z)$.

This paper pursues the theme that dimensional asymptotics, in which $q \to \infty$, help to clarify the Stein phenomenon. The basis for this approach is in Stein (1956) and (1962). Asymptotics in q have received little attention in the subsequent literature on Stein estimation, but are common in the logically related nonparametric regression literature, where ξ_i is assumed to depend smoothly upon i.

Our results are organized as follows. Section 2 studies the best orthogonally equivariant estimator $\hat{\xi}_E(\rho_0)$ of ξ in the $N(\xi, I)$ submodel where $\rho_0 = |\xi|$ is fixed. The orthogonal group is transitive on the parameter space of the submodel. We find an explicit analytical formula for $\hat{\xi}_E(\rho_0)$ by using theory from directional statistics for the Langevin (or Fisher-von Mises) distribution on the unit sphere. A mathematically simpler equivariant estimator,

$$\hat{\xi}_{AE}(\rho_0) = (\rho_0^2/|\bar{Y}_n|^2)\bar{Y}_n \tag{1.5}$$

is shown to approximate $\hat{\xi}_E(\rho_0)$ well in high dimensions. It was the perception of this approximation that guided the treatment in Stein (1956) and (1962).

Section 3 develops good estimates $\hat{\rho}$ of $|\xi|$ in the full $N(\xi, I)$ model and then analyzes the adaptive estimators $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ of ξ. In particular, Section 3.1 establishes that the estimators

$$\hat{\rho}^2 = |\bar{Y}_n|^2 - (q-d)/n, \quad \hat{\rho}^2 = [|\bar{Y}_n|^2 - (q-d)/n]_+, \tag{1.6}$$

d being any constant, are locally asymptotically minimax for $|\xi|^2$ as $q \to \infty$. Here $[x]_+$ is the positive-part function, equal to the larger of x and 0. When $\hat{\rho}^2$ is taken to be $|\bar{Y}_n|^2 - (q-2)/n$, then $\hat{\xi}_{AE}(\hat{\rho})$ is the Stein estimator $\hat{\xi}_S$. On the other hand, when $\hat{\rho}^2$ is $[|\bar{Y}_n|^2 - (q-2)/n]_+$, then $\hat{\xi}_{AE}(\hat{\rho})$ becomes the positive-part Stein estimator

$$\hat{\xi}_{PS} = \left[1 - \frac{(q-2)}{n|\bar{Y}_n|^2}\right]_+ \bar{Y}_n. \tag{1.7}$$

Section 3.2 develops asymptotic optimality results concerning the estimation of ξ. We prove that as $q \to \infty$, with $\hat{\rho}$ given by (1.6), the estimators $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ are asymptotically minimax and asymptotically ϵ-admissible on large compact balls about the origin. The estimator $\bar{Y}_n$ has neither optimality property. The best choice of the constant d in the estimator $\hat{\xi}_E(\hat{\rho})$ is not entirely clear. However, a numerical experiment strongly suggests that, for every q, there exist values of d such that the estimator $\hat{\xi}_E(\hat{\rho})$ dominates the positive-part Stein estimator $\hat{\xi}_{PS}$.

Section 3.3 returns to the question of bootstrapping orthogonally equivariant estimators such as Stein's. We prove that resampling from the $N(\hat{\xi}, I)$ distribution, where $|\hat{\xi}|^2$ is an asymptotically efficient estimator of $|\xi|^2$ in the sense of Section 3.1, yields consistent risk estimators as $q \to \infty$ for regular orthogonally equivariant estimators of ξ. The argument also shows why resampling from the $N(\hat{\xi}_S, I)$ or $N(\hat{\xi}_{PS}, I)$ or $N(\bar{Y}, I)$ distributions fails. A simple adjustment reduces the bias of the proposed risk estimators.

2. The Fixed-Length Submodel.

Without any loss of generality, we fix the sample size n at 1. Observed is the random q-vector $X = (X_1, \ldots, X_q)'$ whose distribution is $N(\xi, I)$, the vector $\xi \in R^q$ being unknown. The risk of an estimator $\hat{\xi} = \hat{\xi}(X)$ is

$$R_q(\hat{\xi}, \xi) = q^{-1} E_\xi |\hat{\xi} - \xi|^2. \tag{2.1}$$

Of special interest in this paper are estimators $\hat{\xi}$ that are equivariant under the orthogonal group: $\hat{\xi}(OX) = O\hat{\xi}(X)$ for every $q \times q$ orthogonal matrix O. Every such estimator can be written in the form

$$\hat{\xi}(X) = h(|X|)X \tag{2.2}$$

for some real-valued function h (Stein 1956, Section 3).

2.1. Exact theory. Consider the estimation of ξ when $|\xi|$ is fixed at a known value ρ_0 and only the direction vector $\mu = \xi/|\xi|$ is unknown. In this submodel, we derive the minimum risk equivariant estimator of ξ and the minimum risk equivariant estimator of ξ among estimators whose length is ρ_0.

The conditional risk, given $|X|$, of any equivariant estimator (2.2) is

$$q^{-1}[h^2(|X|)|X|^2 - 2h(|X|)E_\xi(\xi' X \,|\, |X|) + \rho_0^2]. \tag{2.3}$$

Let $\hat{\mu} = X/|X|$ denote the direction vector of X. The choice of h that minimizes (2.3) is

$$h_0(|X|) = |X|^{-2} E_\xi(\xi' X \,|\, |X|) = \rho_0 |X|^{-1} E_\xi(\mu' \hat{\mu} \,|\, |X|). \tag{2.4}$$

The conditional expectation in (2.4) may be evaluated as follows. When $q \geq 2$, the conditional distribution of $\hat{\mu}$ given $|X|$ is Langevin on the unit sphere in R^q, with mean direction $\mu = \xi/|\xi|$ and dispersion parameter $\kappa = \rho_0|X|$ (cf. Watson 1986). The density of this distribution, relative to spherical surface measure, is $a_q(\kappa)\exp(\kappa \mu' x)$, where

$$a_q(\kappa) = (2\pi)^{-q/2} \kappa^{q/2-1} I_{q/2-1}^{-1}(\kappa) \tag{2.5}$$

and $I_\nu(\kappa)$ is the modified Bessel function of the first kind and order ν (cf. Schou 1978). When $q = 1$, the conditional distribution of $\hat{\mu}$ is discrete, supported on the two points ± 1 with

$$\begin{aligned} P_\xi(\hat{\mu} = 1 \,|\, |X|) &= [\exp(\xi|X|) + \exp(-\xi|X|)]^{-1} \exp(\xi|X|) \\ P_\xi(\hat{\mu} = -1 \,|\, |X|) &= [\exp(\xi|X|) + \exp(-\xi|X|)]^{-1} \exp(-\xi|X|). \end{aligned} \tag{2.6}$$

From the analysis in Appendix A of Watson (1986), it follows that for every integer $q \geq 2$,

$$E_\xi(\mu'\hat{\mu}||X|) = A_q(\rho_0|X|), \tag{2.7}$$

where

$$A_q(z) = I_{q/2}(z)/I_{q/2-1},(z), \quad z \geq 0. \tag{2.8}$$

For $q = 1$, the conditional distribution (2.6) yields

$$E_\xi(\mu'\hat{\mu}||X|) = \tanh(\rho_0|X|). \tag{2.9}$$

This calculation agrees conveniently with formula (2.7) for $q = 1$. Thus, by (2.4), (2.7) and (2.9), the minimum risk orthogonally equivariant estimator of ξ in the fixed length submodel is

$$\hat{\xi}_E(\rho_0) = \rho_0 A_q(\rho_0|X|)\hat{\mu}, \quad q \geq 1. \tag{2.10}$$

If we restrict attention to equivariant estimators $\hat{\xi}$ such that $|\hat{\xi}| = \rho_0$, the only possibilities, according to (2.2), are $\hat{\xi} = \pm\rho_0\hat{\mu}$. The positive sign minimizes the conditional risk (2.3). Consequently, the best constrained length equivariant estimator of ξ is

$$\hat{\xi}_{CE}(\rho_0) = \rho_0\hat{\mu}, \tag{2.11}$$

in agreement with intuition.

These considerations, the compactness of the orthogonal group on R^q, and the Hunt-Stein theorem prove the following result.

THEOREM 2.1. *In the fixed length submodel where $|\xi| = \rho_0$, the minimum risk orthogonally equivariant estimator of ξ is $\hat{\xi}_E(\rho_0)$, defined in (2.10). This estimator is minimax and admissible among all estimators of ξ. Among estimators of ξ whose length is constrained to be ρ_0, the minimum risk orthogonally equivariant estimator is $\hat{\xi}_{CE}(\rho_0)$, defined in (2.11). This alternative estimator is minimax and admissible among all estimators whose length is ρ_0.*

It is of interest to compare $\hat{\xi}_E$ with $\hat{\xi}_{CE}$ and with two other orthogonally equivariant estimators: X and

$$\hat{\xi}_{AE}(\rho_0) = (\rho_0^2/|X|)\hat{\mu}. \tag{2.12}$$

The latter estimator will be seen to approximate $\hat{\xi}_E(\rho_0)$ for large values of q (Theorem 2.3). While $\hat{\xi}_E(\rho_0)$ strictly dominates every orthogonally equivariant estimator in the fixed length submodel, the improvement is large in the case of X and is much smaller in the case of $\hat{\xi}_{AE}(\rho_0)$ or $\hat{\xi}_{CE}(\rho_0)$. These points will be clarified through the next two theorems.

From (2.3) and (2.7), the risk of the general orthogonally equivariant estimator $\hat{\xi} = h(|X|)X$ is

$$R_q(\hat{\xi}, \xi) = q^{-1}E_\xi[h^2(|X|)|X|^2 - 2\rho_0 h(|X|)|X|A_q(\rho_0|X|) + \rho_0^2]. \tag{2.13}$$

Substituting the appropriate values of $h(|X|)$ into (2.13) yields

THEOREM 2.2. *In the fixed length submodel where* $|\xi| = \rho_0$,

$$R_q(\hat{\xi}_E(\rho_0), \xi) = q^{-1}E_\xi[\rho_0^2 - \rho_0^2 A_q^2(\rho_0|X|)] \tag{2.14}$$

$$R_q(\hat{\xi}_{AE}(\rho_0), \xi) = q^{-1}E_\xi[\rho_0^2 - 2\rho_0^3|X|^{-1}A_q(\rho_0|X|) + \rho_0^4|X|^{-2}] \tag{2.15}$$

$$R_q(\hat{\xi}_{CE}(\rho_0), \xi) = q^{-1}E_\xi[2\rho_0^2 - 2\rho_0^2 A_q(\rho_0|X|)]. \tag{2.16}$$

The risks of the three estimators in Theorem 2.2 can also be computed by Stein's (1981) method for estimators of the form $\hat{\xi} = X + g(X)$:

$$R_q(\hat{\xi}, \xi) = 1 + q^{-1}E_\xi[|g(X)|^2 + 2\sum_{i=1}^{q} \partial g_i(X_i)/\partial X_i] \tag{2.17}$$

where g_i is the ith component of g. This approach yields strikingly different, though necessarily equivalent, expressions for the risks of the three estimators. For example, Stein's formula gives

$$R_q(\hat{\xi}_E(\rho_0), \xi) = q^{-1}E_\xi[\{|X| - \rho_0 A_q(\rho_0|X|)\}^2 - 2\rho_0^2\{1 - A_q^2(\rho_0|X|)\}] - 1 \tag{2.18}$$

From (2.18) and (2.14), we see that the sufficient statistic X is not complete in the fixed length submodel.

2.2. Properties of $A_q(z)$. Further developments and the calculation of $\hat{\xi}_E(\rho_0)$ rely on the following results. The function A_q satisfies the recursion

$$A_q(z) = 1/A_{q-2}(z) - (q-2)/z, \quad q \geq 3 \tag{2.19}$$

by Schou (1978, Appendix A). In particular,

$$\begin{aligned} A_1(z) &= \tanh(z) \\ A_3(z) &= \coth(z) - 1/z \end{aligned} \tag{2.20}$$

and so forth for odd orders q. The function A_q also satisfies the differential equation

$$A_q'(z) = 1 - (q-1)A_q(z)/z - A_q^2(z) \tag{2.21}$$

as in Schou (1978, Section 2). For every integer $q \geq 1$, $A_q(z)$ is strictly monotone increasing and concave on $z \geq 0$, with $A_q''(z) \leq 0$,

$$\begin{aligned} A_q(0) = 0, &\quad \lim_{z\to\infty} A_q(z) = 1 \\ A_q'(0) = 1/q, &\quad \lim_{z\to\infty} A_q'(z) = 0 \end{aligned} \tag{2.22}$$

by Watson (1986, Appendix A). Finally, for every $z \geq 0$,

$$\lim_{q \to \infty} z A_q(qz) = (z^2 + 1/4)^{1/2} - 1/2. \tag{2.23}$$

To verify (2.23), let $B_q(z) = zA_q(qz)$ and write $B(z)$ for the limit of a convergent subsequence in $\{B_q(z) : q \geq 1\}$. Equation (2.21) and the second line in (2.22) give

$$0 = z^2 - B(z) - B^2(z). \tag{2.24}$$

The positive root of (2.24) is $B(z) = (z^2 + 1/4)^{1/2} - 1/2$, implying (2.23).

2.3. Asymptotic risks. For $t \geq 0$, let

$$r_E(t) = t/(1+t) \tag{2.25}$$

and

$$r_{CE}(t) = [1 + \{(1+t)^{1/2} - t^{1/2}\}^2]t/(1+t). \tag{2.26}$$

Evidently $r_{CE}(t) > r_E(t)$ whenever $t > 0$. The maximum difference between $r_{CE}(t)$ and $r_E(t)$ is only .091. As the next theorem shows, this figure is the maximum difference between the asymptotic risk of $\hat{\xi}_{CE}(\rho_0)$ and the asymptotic risk of the best orthogonally equivariant estimator $\hat{\xi}_E(\rho_0)$.

THEOREM 2.3. *In the fixed length submodel where $|\xi| = \rho_0$, the following uniform risk approximations hold for every finite $c > 0$:*

$$\lim_{q \to \infty} \sup_{\rho_0^2 \leq qc} |R_q(\hat{\xi}_E(\rho_0), \xi) - r_E(\rho_0^2/q)| = 0 \tag{2.27}$$

and likewise for $R_q(\hat{\xi}_{AE}(\rho_0), \xi)$ while

$$\lim_{q \to \infty} \sup_{\rho_0^2 \leq qc} |R_q(\hat{\xi}_{CE}(\rho_0), \xi) - r_{CE}(\rho_0^2/q)| = 0 \tag{2.28}$$

Moreover, the estimators $\hat{\xi}_E(\rho_0)$ and $\hat{\xi}_{AE}(\rho_0)$ are asymptotically equivalent in the sense that

$$\lim_{q \to \infty} \sup_{\rho_0^2 \leq qc} q^{-1} E_\xi |\hat{\xi}_E(\rho_0) - \hat{\xi}_{AE}(\rho_0)|^2 = 0. \tag{2.29}$$

PROOF. Let $\{\rho_q : q \geq 1\}$ be any sequence of positive numbers such that $\rho_q^2/q \to a$, where a is finite. Let $\{\xi_q \in R^q\}$ be any sequence such that $|\xi_q| = \rho_q$. To prove (2.27) it suffices to show that

$$\lim_{q \to \infty} R_q(\hat{\xi}_E(\rho_q), \xi_q) = r_E(a). \tag{2.30}$$

Let X_q be a random q-vector with $N(\xi_q, I)$ distribution. As $q \to \infty$,

$$(2.31) \qquad |X_q|^2/q \to 1 + a \text{ in probability}$$

and therefore, by (2.23),

$$(2.32) \qquad q^{-1}\rho_q|X_q|A_q(\rho_q|X_q|) \to a \text{ in probability.}$$

Limit (2.27) now follows from (2.14) in Theorem 2.2.

Similar reasoning handles the risks of $\hat{\xi}_{AE}(\rho_0)$ and $\hat{\xi}_{CE}(\rho_0)$.

Let

$$(2.33) \qquad C_q = q^{-1}(\hat{\xi}_E - \xi_q)'(\hat{\xi}_{AE} - \xi_q)$$

and observe that

$$(2.34) \qquad q^{-1}E_{\xi_q}|\hat{\xi}_E - \hat{\xi}_{AE}|^2 = R_q(\hat{\xi}_E, \xi_q) + R_q(\hat{\xi}_{AE}, \xi_q) - 2E_{\xi_q}(C_q).$$

From (2.31) and (2.32), $C_q \to a/(1+a)$ in probability. Hence

$$(2.35) \qquad \liminf_{q\to\infty} E_{\xi_q}(C_q) \geq r_E(a).$$

On the other hand, by Cauchy-Schwarz and (2.27) and its counterpart for $\hat{\xi}_{AE}(\rho_0)$,

$$(2.36) \qquad \limsup_{q\to\infty} E_q(C_q) \leq r_E(a).$$

Conclusion (2.29) now follows from (2.34) and these considerations.

2.4. Geometry of the asymptotics. Figure 1, which is suggested by the figures in Stein (1962) and Brandwein and Strawderman (1990), exhibits the geometry of the limits in Theorem 2.3. Under the triangular array asymptotics used to prove the theorem, the following relations are very nearly true with high probability when dimension q is large:

$$(2.37) \qquad |q^{-1/2}\xi_q|^2 = a, \quad |q^{-1/2}X - q^{-1/2}\xi_q|^2 = 1, \quad |q^{-1/2}X|^2 = 1 + a.$$

Consequently, the large triangle in Figure 1 is nearly right-angled, with

$$(2.38) \qquad \cos^2(\theta) = a/(1+a).$$

The circle in Figure 1 represents the parameter space of the fixed-length submodel in which $|q^{-1/2}\xi_q|^2 = a$.

As was noted in (2.2), orthogonally equivariant estimators lie along the vector X. The scaled equivariant estimator $q^{-1/2}\hat{\xi}$ that minimizes the loss $|q^{-1/2}\hat{\xi} - q^{-1/2}\xi|^2$ is the orthogonal projection of $q^{-1/2}\xi$ onto X. For large q, the minimizing $\hat{\xi}$ approximately satisfies

$$(2.32) \qquad q^{-1/2}\hat{\xi} = |q^{-1/2}\xi_q|\cos(\theta)\hat{\mu} = [a/(1+a)]q^{-1/2}X$$

with high probability. Algebraically, $q^{-1/2}\hat{\xi}$ coincides asymptotically with $q^{-1/2}\hat{\xi}_{AE}$. On the other hand, since minimizing loss also minimizes risk, $q^{-1/2}\hat{\xi}$ coincides asymptotically with $q^{-1/2}\hat{\xi}_E$.

Thus, from the geometry,

$$q^{-1}|\hat{\xi}_E - \xi|^2 = q^{-1}|\hat{\xi}_{AE} - \xi|^2 = a\sin^2(\theta) = r_E(a) \tag{2.40}$$

is very nearly true with high probability for large q. This conclusion agrees with limit (2.28) in Theorem 2.3. Applying Pythagoras' theorem to the smallest right-angled triangle in Figure 1 yields the asymptotic approximation

$$q^{-1}|\hat{\xi}_{CE} - \xi|^2 = r_E(a) + [a^{1/2} - a^{1/2}\cos(\theta)]^2 = r_{CE}(a) \tag{2.41}$$

in agreement with limit (2.29).

[Figure 1 goes here]

3. The Full Model.

In the full $N(\xi, I)$ model, with $\xi \in R^q$, we can pursue an adaptive strategy for estimating ξ: first devise a good estimator $\hat{\rho}$ of $|\xi|$ and then form

$$\begin{aligned} \hat{\xi}_E(\hat{\rho}) &= \hat{\rho}A_q(\hat{\rho}|X|)\hat{\mu} \\ \hat{\xi}_{AE}(\hat{\rho}) &= (\hat{\rho}^2/|X|)\hat{\mu} \\ \hat{\xi}_{CE}(\hat{\rho}) &= \hat{\rho}\hat{\mu}. \end{aligned} \tag{3.1}$$

When $\hat{\rho}^2$ is taken to be $|X|^2 - q + 2$ or $[|X|^2 - q + 2]_+$, then $\xi_{AE}(\hat{\rho})$ becomes the Stein estimator $\hat{\xi}_S$ or the positive-part Stein estimator $\hat{\xi}_{PS}$, respectively. The proper choice of $\hat{\rho}$ and the performance of the estimators (3.1) for ξ are the main themes of this section.

3.1. Estimation of $|\xi|^2$. The following triangular array central limit theorem suggests good estimators for $|\xi|^2$:

LEMMA 3.1. *Let* $\{\xi_q \in R^q\}$ *be any sequence such that* $|\xi_q|^2/q \to a < \infty$ *as* $q \to \infty$. *Then*

$$\mathcal{L}[q^{-1/2}(|X|^2 - q - |\xi_q|^2)|\xi_q] \Rightarrow N(0, 2 + 4a). \tag{3.2}$$

The weak convergence in (3.2) is implied by the algebraic representation

$$|X|^2 = q + |\xi_q|^2 + \{|X - \xi_q|^2 - q\} + 2\xi_q'(X - \xi_q) \tag{3.3}$$

and the Lindeberg-Feller theorem. To apply the latter, note that $\mathcal{L}[|X||\xi_q\}$ depends on ξ_q only through $|\xi_q|$. Hence, there is no loss of generality in taking each component of ξ_q to be $q^{-1/2}|\xi_q|$.

The next theorem gives a local asymptotic minimax bound on the mean squared error of estimators of $|\xi|^2$ in high dimension. The proof is in Section 4.

THEOREM 3.2. *In the full $N(\xi, I)$ model, for every finite $a > 0$,*

$$\lim_{c\to\infty} \liminf_{q\to\infty} \inf_{\hat\rho} \sup_{||\xi|^2/q-a|\le q^{-1/2}c} q^{-1}E_\xi(\hat\rho^2 - |\xi|^2)^2 \ge 2 + 4a, \tag{3.4}$$

the infimum being taken over all estimators $\hat\rho$.

The lower bound (3.4) is sharp in the following sense: if $\hat\rho^2 = |X|^2 - q + d$ or $[|X|^2 - q + d]_+$, where d is a constant, then

$$\lim_{q\to\infty} \sup_{||\xi|^2/q-a|\le q^{-1/2}c} q^{-1}E_\xi(\hat\rho^2 - |\xi|^2)^2 = 2 + 4a \tag{3.5}$$

for every finite $c > 0$. This assertion is immediate from Lemma 3.1. In particular, the uniformly minimum variance unbiased estimator of $|\xi|^2$, which is $\hat\rho^2 = |X|^2 - q$, is locally asymptotically minimax among all estimators of $|\xi|^2$ as dimension q increases. The UMVU has the unfortunate property of being negative with positive probability.

An analogous lower bound for estimators of $|\xi|$ is

$$\lim_{c\to\infty} \liminf_{q\to\infty} \inf_{\hat\rho} \sup_{||\xi|^2/q-a|\le q^{-1/2}c} E_\xi(\hat\rho - |\xi|)^2 \ge (1+2a)/(2a). \tag{3.6}$$

It is attained asymptotically by the estimator

$$\hat\rho = [|X|^2 - q + d]_+^{1/2}, \tag{3.7}$$

where d is any constant.

3.2. Estimation of ξ. We begin by computing the risks of the adaptive estimators (3.1) when $\hat\rho$ is given by (3.7). Since these adaptive estimators are orthogonally equivariant, it follows from (2.3) that

$$R_q(\hat\xi_E(\hat\rho), \xi) = q^{-1}E_\xi[\hat\rho^2 A_q^2(\hat\rho|X|) - 2|\xi|A_q(|\xi||X|)\hat\rho A_q(\hat\rho|X|) + |\xi|^2] \tag{3.8}$$

$$R_q(\hat\xi_{AE}(\hat\rho), \xi) = q^{-1}E_\xi[\hat\rho^4|X|^{-2} - 2\hat\rho^2|\xi||X|^{-1}A_q(|\xi||X|) + |\xi|^2] \tag{3.9}$$

$$R_q(\hat\xi_{CE}(\hat\rho), \xi) = q^{-1}E_\xi[\hat\rho^2 - 2\hat\rho|\xi|A_q(|\xi||X|) + |\xi|^2]. \tag{3.10}$$

Formula (3.9) is also valid for $\hat\rho^2 = |X|^2 - q + d$. Thus, it applies to both the Stein estimator and the positive-part Stein estimator, which arise when $d = 2$. This choice of d is known to minimize the risk of $\hat\xi_{AE}(\hat\rho)$.

The selection of d to minimize the risk of $\hat\xi_E(\hat\rho)$ is less clear. A numerical study based on 40,000 Monte Carlo samples and double-precision arithmetic suggests that, when $q = 3$ and $d = 2.85$ or when $q = 5$ and $d = 2.7$, the estimator $\hat\xi_E(\hat\rho)$ dominates the positive-part Stein estimator. Figure 2 exhibits the risk function difference computed in the experiment for $q = 3$. We conjecture that, for every q, there exist choices of d, depending on q, such that $\hat{\xi}_E(\hat{\rho})$ dominates the positive-part Stein estimator. Any such improvement in risk must tend to zero as $q \to \infty$, because of the next result.

[Figure 2 goes here]

THEOREM 3.3. *In the full $N(\xi, I)$ model with $\hat{\rho}^2 = [|X|^2 - q + d]_+$, the following risk approximations hold for every finite $c > 0$:*

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} |R_q(\hat{\xi}_E(\hat{\rho}), \xi) - r_E(|\xi|^2/q)| = 0 \tag{3.11}$$

and likewise for $R_q(\hat{\xi}_{AE}(\hat{\rho}), \xi)$, while

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} |R_q(\hat{\xi}_{CE}(\hat{\rho}), \xi) - r_{CE}(\rho_0^2/q)| = 0. \tag{3.12}$$

Moreover, the adaptive estimators $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ are asymptotically equivalent in the sense that

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} q^{-1} E_\xi |\hat{\xi}_E(\hat{\rho}) - \hat{\xi}_{AE}(\hat{\rho})|^2 = 0. \tag{3.13}$$

The proof of this theorem is similar to that for Theorem 2.3, relying on an asymptotic analysis of the exact risks given in (3.8) to (3.10). Equations (3.11) and (3.12) indicate that the exact risks are better plotted against $|\xi|^2/q$ rather than $|\xi|^2$. Figure 3 displays in this fashion the risk functions of the positive-part Stein estimator when $q = 3, 5, 9, 19$ (solid curves), computed from 40,000 Monte Carlo samples. The dotted curve in Figure 3 is the limiting risk function as $q \to \infty$, given by (3.11) and (2.25). The rate of convergence seems quick.

[Figure 3 goes here]

We turn now to the main result of this section—the asymptotic optimality of $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ as dimension q increases. An estimator $\hat{\xi}$ is said to be *ϵ-admissible on* $B_q(c) = \{\xi \in R^q : |\xi|^2 \le qc\}$ if there does not exist another estimator $\tilde{\xi}$ such that

$$R_q(\tilde{\xi}, \xi) < R_q(\hat{\xi}, \xi) - \epsilon \tag{3.14}$$

for every $\xi \in B_q(c)$.

THEOREM 3.4. *In the full $N(\xi, I)$ model, for every finite $c > 0$,*

$$\liminf_{q\to\infty} \inf_{\hat{\xi}} \sup_{|\xi|^2 \le qc} R_q(\hat{\xi}, \xi) \ge r_E(c) \tag{3.15}$$

the infimum being taken over all estimators $\hat{\xi}$. If $\hat{\rho}^2 = [|X|^2 - q + d]_+$ then for every finite $c > 0$,

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} R_q(\hat{\xi}_E(\hat{\rho}), \xi) = r_E(c) \tag{3.16}$$

and $\hat{\xi}_E(\hat{\rho})$ is ϵ-admissible on $B_q(c)$ for all sufficiently large q. The same assertions hold for $\hat{\xi}_{AE}(\hat{\rho})$, in which case $\hat{\rho}^2 = |X|^2 - q + d$ also works.

The theorem is proved in Section 4. It entails that both the positive-part Stein estimator and the Stein estimator, as well as the new estimator $\hat{\xi}_E(\hat{\rho})$, are asymptotically minimax and asymptotically ϵ-admissible on large compact balls about the origin. The estimator X has neither property, because

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} R_q(X, \xi) = 1 \tag{3.17}$$

in contrast to (3.16). Similarly, $\hat{\xi}_{CE}(\hat{\rho})$ lacks both asymptotic optimality properties.

3.3. Estimation of risk. Stein's formula (2.17) generates unbiased estimators for the risks of $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ that are consistent as $q \to \infty$. Because this approach requires considerable algebra, at least in the case of $\hat{\xi}_E(\hat{\rho})$, it seems worth looking for simpler bootstrap or asymptotic risk estimators. Three such risk estimators are discussed in what follows.

Let $\hat{\xi}_I$ be any orthogonally equivariant estimator of ξ, with risk

$$R_q(\hat{\xi}_I, \xi) = r_q(|\xi|^2). \tag{3.18}$$

We say that $\hat{\xi}_I$ is *regular* if, for every sequence $\{\xi_q \in R^q\}$ such that $|\xi_q|^2/q \to a$ as q increases, we have

$$\lim_{q\to\infty} r_q(|\xi_q|^2) = r(a) \tag{3.19}$$

for some function r that does not depend on the sequence $\{\xi_q\}$. By Theorem 3.3, the estimators $\hat{\xi}_{CE}(\hat{\rho})$, $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ are each regular in this sense, the function r being respectively r_{CE}, r_E and r_E.

The estimator $\hat{\xi}_{CE}(\hat{\rho})$, defined by (2.11) and (3.7) is just

$$\hat{\xi}_{CE}(\hat{\rho}) = [1 - (q-d)/|X|^2]_+^{1/2} X. \tag{3.20}$$

As estimators of the risk $R_q(\hat{\xi}_I, \xi)$, let us consider the parametric *bootstrap* estimator

$$\hat{R}_B = r_q(|\hat{\xi}_{CE}(\hat{\rho})|^2) \tag{3.21}$$

and the *asymptotic* estimator

$$\hat{R}_A = r(|\hat{\xi}_{CE}(\hat{\rho})|^2/q). \tag{3.22}$$

Both $\hat{R}_B$ and $\hat{R}_A$ are always non-negative, a property not intrinsic to Stein's unbiased estimator of risk.

THEOREM 3.5. *Suppose that $\hat{\xi}_I$ is a regular orthogonally equivariant estimator of ξ. Then*

$$\lim_{q\to\infty} \sup_{|\xi|^2 \le qc} P_\xi[|\hat{R}_B - R_q(\hat{\xi}_I, \xi)| > \epsilon] \ge 0 \tag{3.23}$$

for every $\epsilon > 0$ and every finite $c > 0$. If r is a continuous function, then the same limit also holds for $\hat{R}_A$.

PROOF. The argument by contradiction for (3.23) rests on the following fact: the convergence $|\xi_q|^2/q \to a$ implies that

$$|\hat{\xi}_{CE}(\hat{\rho})|^2/q = \hat{\rho}^2/q \to a \tag{3.24}$$

in probability, because of Lemma 3.1. Consequently, by the regularity of $\hat{\xi}_I$,

$$\hat{R}_B \to r(a) \tag{3.25}$$

in probability. If r is continuous, then also

$$\hat{R}_A \to r(a) \tag{3.26}$$

in probability. Since $R_q(\hat{\xi}_I, \xi_q) \to r(a)$ by regularity, the theorem follows.

The calculation of $\hat{R}_B$ and $\hat{R}_A$ is straightforward in the case of the estimators $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$. On the one hand, it follows from (3.22) and Theorem 2.3 that

$$\hat{R}_A = |\hat{\xi}_{CE}|^2/(q + |\hat{\xi}_{CE}|^2) \tag{3.27}$$

for $\hat{\xi}_E(\hat{\rho})$ or $\hat{\xi}_{AE}(\hat{\rho})$. On the other hand, let X^* be a random q-vector whose conditional distribution given X is $N(\hat{\xi}_{CE}(\hat{\rho}), I)$. Let E^* denote conditional expectation with respect to this distribution and let ξ_I^* stand for the bootstrap recalculation $\hat{\xi}_I(X^*)$. Then

$$\hat{R}_B = q^{-1}E^*|\xi_I^* - \hat{\xi}_I|^2 \tag{3.28}$$

for the regular orthogonally equivariant estimator $\hat{\xi}_I$. In the case of $\hat{\xi}_E(\hat{\rho})$ or $\hat{\xi}_{AE}(\hat{\rho})$, representation (3.28) provides the basis for a simple Monte Carlo approximation to $\hat{R}_B$.

The proof of Theorem 3.4 makes it clear why resampling from the $N(\hat{\xi}_E(\hat{\rho}), I)$ or $N(\hat{\xi}_{AE}(\hat{\rho}), I)$ or $N(X, I)$ distribution does not give consistent bootstrap estimators of the risk $R_q(\hat{\xi}_I, \xi)$. The difficulty is that $|\xi_q|^2/q \to a$ implies that

$$\begin{gathered} |\hat{\xi}_E(\hat{\rho})|^2/q \to a^2/(1+a) \\ |\hat{\xi}_{AE}(\hat{\rho})|^2/q \to a^2/(1+a) \\ |X|^2/q \to 1+a \end{gathered} \tag{3.29}$$

in probability. None of these limits has the value a that is required for consistency of the corresponding bootstrap risks. As is illustrated in Figure 2, both $\hat{\xi}_E(\hat{\rho})$ and $\hat{\xi}_{AE}(\hat{\rho})$ are too short while X is too long. The length of $\hat{\xi}_{CE}$ is just right.

A possible drawback to the risk estimator $\hat{R}_A$ and $\hat{R}_B$ is non-negligible bias when q is small. To address the question of reducing bias, let

$$b_q(|\xi|^2) = E_\xi \hat{R}_A - R_q(\hat{\xi}_I, \xi) \tag{3.30}$$

and define the *bias-adjusted* risk estimator to be

$$\hat{R}_{BA} = \hat{R}_A - b_q(|\hat{\xi}_{CE}(\hat{\rho})|^2). \tag{3.31}$$

The second term on the right-side of (3.31) is a bootstrap estimator for the bias of $\hat{R}_A$. Let

$$R_A^* = r(|\xi_I^*|^2/q) \tag{3.32}$$

denote the asymptotic risk estimate recalculated on the bootstrap sample X^*. Then

$$\hat{R}_{BA} = \hat{R}_B + \hat{R}_A - E^* \hat{R}_A^*. \tag{3.33}$$

Thus, $\hat{R}_{BA}$ can be viewed as a modification of either $\hat{R}_A$ or of $\hat{R}_B$. A Monte Carlo algorithm for approximating $\hat{R}_{BA}$ follows immediately from the representation (3.33).

The effectiveness of the bias adjustment may be assessed by considering a special case: estimating the risk of the Stein estimator $\hat{\xi}_S$. In this instance,

$$\begin{aligned} \hat{R}_A &= |\hat{\xi}_{CE}|^2/(q + |\hat{\xi}_{CE}|^2) \\ \hat{R}_B &= 1 - E^* \left[\frac{(q-2)^2/q}{|X^*|^2} \right] \\ \hat{R}_{BA} &= \hat{R}_B + \hat{R}_A - E^*[|\xi_{CE}^*|^2/(q + |\xi_{CE}^*|^2)]. \end{aligned} \tag{3.34}$$

On the other hand, Stein's unbiased estimator for the risk of $\hat{\xi}_S$ is just

$$\hat{R}_U = 1 - \frac{(q-2)^2/q}{|X|^2}. \tag{3.35}$$

Taylor expansions based on (3.3) show that the biases of $\hat{R}_A$ and $\hat{R}_B$ are $O(q^{-1})$ while the bias of $\hat{R}_{BA}$ is $O(q^{-2})$. All four estimators are consistent for the risk of $\hat{\xi}_S$ and differ only by terms of order $O_p(q^{-1})$.

More precisely, suppose that $|\xi_q|^2/q \to a$ as q increases. Then, by (1.2), (3.3) and Taylor expansion,

$$R_q(\hat{\xi}_S, \xi_q) = r_E(|\xi_q|^2/q) + q^{-1} r_{E,1}(|\xi_q|^2/q) + O(q^{-2}), \tag{3.36}$$

where r_E is given by (2.25) and

$$r_{E,1}(t) = (1+t)^{-3}[2(1+t)^2 + 2t^2]. \tag{3.37}$$

The function $r_{E,1}(t)$ is positive for every non-negative t, in agreement with Figure 3. Moreover,

$$\hat{R}_B = r_E(|\hat{\xi}_{CE}|^2/q) + q^{-1} r_{E,1}(|\hat{\xi}_{CE}|^2/q) + O_p(q^{-2}). \tag{3.38}$$

From (3.34) and (3.3),

$$E_{\xi_q} \hat{R}_A = r_E(|\xi_q|^2/q) - q^{-1} r_{E,2}(|\xi_q|^2/q) + O(q^{-2}), \tag{3.39}$$

where

$$r_{E,2}(t) = (1+t)^{-3}[2 - d + (4-d)t]. \tag{3.40}$$

When $d = 2$, $r_{E,2}$ is positive for every positive t. On the other hand, from (3.38), (3.36) and (3.3),

$$E_{\xi_q} \hat{R}_B = R_q(\hat{\xi}_S, \xi_q) - q^{-1} r_{E,2}(|\xi_q|^2/q) + O(q^{-2}). \tag{3.41}$$

Thus

$$\begin{aligned} \text{Bias}(\hat{R}_A, \xi_q) &= -q^{-1}[r_{E,2}(|\xi_q|^2/q) + r_{E,1}(|\xi_q|^2/q)] + O(q^{-2}) \\ \text{Bias}(\hat{R}_B, \xi_q) &= -q^{-1} r_{E,2}(|\xi_q|^2/q) + O(q^{-2}) \end{aligned} \tag{3.42}$$

The extra term in the bias of $\hat{R}_A$ reflects the amount by which the asymptotic risk $r_E(|\xi_q|^2/q)$ understates the actual risk.

From (3.39),

$$E^*(\hat{R}_A^*) = \hat{R}_A - q^{-1} r_{E,2}(|\hat{\xi}_{CE}|^2/q) + O_p(q^{-2}). \tag{3.43}$$

Combining (3.42) and (3.43) with expression (3.33) for $\hat{R}_{BA}$ establishes

$$\text{Bias}(\hat{R}_{BA}, \xi_q) = O(q^{-2}). \tag{3.44}$$

A numerical example. Measurements of thickness were made on four samples of green "one-inch" redwood boards produced at a lumber mill in northern California. Each sample of boards was the outcome of a different sequence of sawing operations. The data was gathered as part of a study on how resawing errors accumulate in the cutting of a log into boards. The measured average thicknesses for the four samples were, in inches,

$$X = (0.91, 0.89, 0.91, 0.89). \tag{3.45}$$

Previous data indicated that the components of X could be treated as realizations of independent normal random variables having different means but a common variance $\tau^2 = 0.00022$. The target thickness for the boards in the four samples was 0.87 inch.

Taking this target thickness as shrinkage point and rescaling appropriately for the variance yields the estimated mean thicknesses

$$\hat{\xi}_S = (0.905, 0.879, 0.905, 0.879). \tag{3.46}$$

Set $d = 2$ in the definition (3.20) of $\hat{\xi}_{CE}$ and consider the normalized risk $\tau^{-2} R_q(\hat{\xi}_S, \xi)$. The asymptotic and bootstrap risk estimates associated with (3.46) are then

$$\hat{R}_A = 0.00771, \quad \hat{R}_B = 0.00925, \quad \hat{R}_{BA} = 0.00941 \tag{3.47}$$

while Stein's unbiased risk estimate is

$$\hat{R}_U = 0.00935. \tag{3.48}$$

These numbers agree qualitatively with the preceding asymptotic theory in (3.42) and (3.44). As expected, $\hat{R}_A$ shows a substantial downward bias, $\hat{R}_B$ shows a smaller downward bias, and the adjusted bootstrap estimator $\hat{R}_{BA}$ is the closest in value to the unbiased risk estimate $\hat{R}_U$.

4. Proofs.

PROOF OF THEOREM 3.2. To simplify notation, let

$$S_q(\hat{\rho}, \xi) = q^{-1} E_q(\hat{\rho}^2 - |\xi|^2)^2, \quad \sigma^2 = 2 + 4a. \tag{4.1}$$

Suppose that the theorem is false. Then there exists $\epsilon > 0$ such that

$$\liminf_{q\to\infty} \inf_{\hat{\rho}} \sup_{||\xi|^2/q - a| \le q^{-1/2}c} S_q(\hat{\rho}, \xi) \le \sigma^2 - \epsilon \tag{4.2}$$

for every $c > 0$. Fix c. By going to a subsequence, we may assume without loss of generality that

$$\inf_{\hat{\rho}} \sup_{||\xi|^2/q - a| \le q^{-1/2}c} S_q(\hat{\rho}, \xi) \le \sigma^2 - \epsilon/2 \tag{4.3}$$

for every q. Hence, there exists an estimator sequence $\{\hat{\rho}_q\}$ such that

$$S_q(\hat{\rho}_q, \xi_q) \le \sigma^2 - \epsilon/4, \quad q \ge 1, \tag{4.4}$$

for every ξ_q such that $||\xi_q|^2/q - a| \le q^{-1/2}c$.

For each q, the estimation problem in (3.4) is invariant under the orthogonal group on R^q. The induced group on the decision space consists solely of the identity map. By the Hunt-Stein theorem and (4.4), there exist orthogonally equivariant estimators $\{\hat{\rho}_{I,q}\}$ such that

$$S_q(\hat{\rho}_{I,q}, \xi_q) \le \sigma^2 - \epsilon/4, \quad q \ge 1, \tag{4.5}$$

for every ξ_q such that $||\xi_q|^2/q - a| \le q^{-1/2}c$. Moreover

$$\hat{\rho}^2_{I,q} = q g_q(|X|^2) \tag{4.6}$$

for some function g_q; that is, $\hat{\rho}_{I,q}$ depends on X only through $|X|^2$.

For any $|h| \le c$, take ξ_q such that $|\xi_q|^2/q = a + q^{-1/2}h$. Let

$$W_q = q^{-1/2}\sigma^{-1}[|X|^2 - q - |\xi_q|^2]. \tag{4.7}$$

Then

$$\begin{aligned} S_q(\hat{\rho}_{I,q}, \xi_q) &= E_{\xi_q}\{q^{1/2}[g_q(|X|^2) - a] - h\}^2 \\ &= \int \{q^{1/2}[\tilde{g}_q(u) - a] - h\}^2 \sigma^{-1}\varphi_q[(u-h)/\sigma]du \end{aligned} \tag{4.8}$$

where φ_q is the density of W_q and

$$\tilde{g}_q(u) = g_q[q^{1/2}u + qa + a]. \tag{4.9}$$

Let $w(x) = x^2 \wedge A$, where $A > 0$ is finite. By Lemma 3.1, the local central limit theorem, and (4.8),

$$\begin{aligned} S_q(\hat{\rho}_{I,q}, \xi_q) &\ge \int w\{q^{1/2}[\tilde{g}_q(u) - a] - h\}\sigma^{-1}\varphi_q[(u-h)/\sigma]du \\ &= E\{w(V_q - h)\sigma^{-1}\varphi[(U-h)/\sigma]/\varphi(U/\sigma)\} + o(1) \end{aligned} \tag{4.10}$$

where U has the standard normal distribution with density φ and $V_q = q^{1/2}[\tilde{g}_q(U) - a]$. By going to a subsequence, we may assume that the $\{(V_q, U)\}$ converge weakly, as random elements of $\bar{R}^q \times R$, to (V, U). Hence, by Fatou's lemma,

$$\begin{aligned} \liminf_{q\to\infty} S_q(\hat{\rho}_{I,q}, \xi_q) &\ge E\{w(V-h)\sigma^{-1}\varphi[(U-h)/\sigma]/\varphi(U/\sigma)\} \\ &= \int E[w(V-h)|U=u]\,\sigma^{-1}\varphi[(u-h)/\sigma]du \\ &= \int\int w(v-h)K(dv,u)\,\sigma^{-1}\varphi[(u-h)/\sigma]du \end{aligned} \tag{4.11}$$

where $K(dv, u)$ is the probability element of the conditional distribution of V given $U = u$.

Let the constant A in the definition of u tend to infinity. Inequalities (4.5) and (4.11) then yield

$$\int\int |v-h|^2 K(dv,u)\sigma^{-1}\varphi[(u-h)/\sigma]du \le \sigma^2 - \epsilon/4 \tag{4.12}$$

for every $|h| \le c$. Since the choice of $c > 0$ was arbitrary, (4.12) contradicts the classical minimax bound for randomized estimators of h in the $N(h, \sigma^2 I)$ model under squared error loss.

PROOF OF THEOREM 3.4. The estimation problem is invariant under the orthogonal group. By the Hunt-Stein theorem,

$$\inf_{\hat{\xi}} \sup_{|\xi|^2 \le qc} R_q(\hat{\xi}, \xi) \ge \inf_{\hat{\xi}_I} \sup_{|\xi|^2 \le qc} R_q(\hat{\xi}_I, \xi) \tag{4.13}$$

the infimum on the right side being taken only over orthogonally equivariant estimators $\hat{\xi}_I$. By Theorem 2.1,

$$\inf_{\hat{\xi}_I} \sup_{|\xi|^2 \le qc} R_q(\hat{\xi}_I, \xi) \ge \inf_{\hat{\xi}_I} \sup_{|\xi|^2 = qc} R_q(\hat{\xi}_I, \xi) = \sup_{|\xi|^2 = qc} R_q[\hat{\xi}_E(q^{1/2}c^{1/2}), \xi]. \tag{4.14}$$

In view of Theorem 2.3, equation (2.27), the right side of (4.14) converges to $r_E(c)$ as $q \to \infty$, proving (3.15).

Conclusion (3.16) is immediate from Theorem 3.3.

Suppose $\hat{\xi}_E(\hat{\rho})$ is not ϵ-admissible on $B_q(c)$ for all sufficiently large q. Then there exists a sequence $\{q_j\}$ tending to infinity and estimators $\{\tilde{\xi}_{q_j}\}$ such that

$$R_{q_j}(\tilde{\xi}_{q_j}, \xi) < R_{q_j}(\hat{\xi}_E(\hat{\rho}), \xi) - \epsilon, \quad j \ge 1 \tag{4.15}$$

for every $\xi \in B_{q_j}(c)$. In view of (3.16),

$$\liminf_{q \to \infty} \sup_{\xi \in B_{q_j}(c)} R_{q_j}(\tilde{\xi}_{q_j}, \xi) \le r_E(c) - \epsilon \tag{4.16}$$

contradicting (3.15).

The last claim in Theorem 3.4 now follows from Theorem 3.3.

REFERENCES

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

BERGER, J. and WOLPERT, R. (1983). Estimating the mean function of a Gaussian process and the Stein effect. *J. Multivariate Anal.* **13** 401–424.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1990). Stein estimation: the spherically symmetric case. *Statistical Science* **5** 356–369.

BROWN, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. *Ann. Math. Statist.* **37** 1087–1135.

EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimators. *Z. Wahrsch. verw. Gebiete* **14** 323–339.

HÁJEK, J. (1972). Local asymptotic minimax and admissibility in estimation. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 175–194. Univ. California Press.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 361–380. Univ. California Press.

LECAM, L. (1972). Limits of experiments. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **1** 245–261. Univ. California Press.

SCHOU, G. (1978). Estimation of the concentration parameter in von Mises-Fisher distributions. *Biometrika* **65** 369–377.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 197–206. Univ. California Press.

STEIN, C. (1962). Confidence sets for the mean of a multivariate normal distribution. *J. Roy. Statist. Soc., Ser. B* 265–296.

STEIN, C. (1981). Estimation of the mean of a normal distribution. *Ann. Statist.* **9** 1135–1151.

VAN DER VAART, A. W. (1988). *Statistical estimation in large parameter spaces.* C.W.I. Tract **44**, Amsterdam.

WATSON, G. S. (1983). *Statistics on Spheres.* Wiley-Interscience, New York.