

# Reversing the Stein Effect

Michael D. Perlman and Sanjay Chaudhuri

*Abstract.* The *Reverse Stein Effect* is identified and illustrated: A statistician who shrinks his/her data toward a point chosen without reliable knowledge about the underlying value of the parameter to be estimated but based instead upon the observed data will not be protected by the minimax property of shrinkage estimators such as that of James and Stein, but instead will likely incur a greater error than if shrinkage were not used.

*Key words and phrases:* James–Stein estimator, shrinkage estimator, Bayes and empirical Bayes estimators, multivariate normal distribution.

## 1. THE CASE FOR SHRINKAGE: THE STEIN EFFECT

Suppose that  $X$  is an observed random vector in  $p$ -dimensional Euclidean space  $\mathbb{R}^p$  such that  $X = Y + \delta$ , where  $\delta$  is an unknown location parameter and  $Y$  is an unobserved absolutely continuous random vector. Under the mild assumption that  $Y \equiv X - \delta$  is *directionally symmetric*,<sup>1</sup> it is easy to heuristically justify “shrinkage” estimators for  $\delta$  of the form

$$(1) \quad \hat{\delta}_\gamma \equiv \hat{\delta}_\gamma(X; \delta_0) = \gamma(X - \delta_0) \cdot (X - \delta_0) + \delta_0,$$

where  $\gamma \equiv \gamma(X - \delta_0) \in [0, 1)$  and  $\delta_0$  is any fixed shrinkage target point in  $\mathbb{R}^p$ . The improvement offered by such shrinkage estimators is often referred to as *the Stein Effect*.

First, for fixed  $\delta$  and  $\delta_0$ , let  $B_1 \equiv B_1(\|\delta - \delta_0\|; \delta_0) \subset \mathbb{R}^p$  denote the ball of radius  $\|\delta - \delta_0\|$  centered at  $\delta_0$  and let  $H$  be the halfspace bounded by a hyperplane  $\partial H$  tangent to  $B_1$  at  $\delta$  (see Figure 1). Then

$$(2) \quad \{X \mid \|X - \delta_0\| > \|\delta - \delta_0\|\} = B_1^c,$$

$$\Pr_\delta[\|X - \delta_0\| > \|\delta - \delta_0\| \mid \delta_0] \\ = \Pr_\delta[X \in B_1^c \mid \delta_0]$$

$$(3) \quad > \Pr_\delta[X \in H \mid \delta_0] \\ = \frac{1}{2},$$

Michael D. Perlman is Professor, Department of Statistics, University of Washington, Seattle, Washington 98195, USA (e-mail: michael@stat.washington.edu). Sanjay Chaudhuri is Assistant Professor, Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546 (e-mail: sanjay@stat.nus.edu.sg).

<sup>1</sup> $\bar{Y} \stackrel{d}{=} -\bar{Y}$ , where  $\bar{Y} := Y/\|Y\|$  is the unit vector in the direction of  $Y$  (see Appendix A).

where (3) follows from directional symmetry by Proposition 1(c) in Appendix A. Furthermore, under somewhat stronger but still general assumptions (see Proposition 2 in Appendix A),

$$(4) \quad \lim_{p \rightarrow \infty} \Pr_\delta[\|X - \delta_0\| > \|\delta - \delta_0\|] \\ \equiv \lim_{p \rightarrow \infty} \Pr_\delta[X \in B_1^c] = 1.$$

Thus,  $\|X - \delta_0\|$  is usually an overestimate of  $\|\delta - \delta_0\|$ , so an estimator of the form  $\gamma(X - \delta_0) \cdot (X - \delta_0)$  for  $\delta - \delta_0$  should be preferable to  $X - \delta_0$  itself. Writing  $\delta$  as  $(\delta - \delta_0) + \delta_0$  immediately leads to estimators for  $\delta$  of the form (1).

Second (see Appendix B),

$$(5) \quad \{X \mid \exists \tilde{\gamma} \in [0, 1) \ni \|\hat{\delta}_{\tilde{\gamma}} - \delta\| < \|X - \delta\|\} = B_2^c,$$

where  $\tilde{\gamma} \equiv \tilde{\gamma}(X - \delta_0, \delta - \delta_0)$  is allowed to depend on  $\delta$  and  $B_2 \equiv B_2(\|\delta - \delta_0\|; \bar{\delta})$  is the ball of radius  $\frac{1}{2}\|\delta - \delta_0\|$  centered at  $\frac{1}{2}(\delta_0 + \delta) \equiv \bar{\delta}$ . Since  $B_2^c \supset B_1^c$ , also

$$(6) \quad \Pr_\delta[X \in B_2^c \mid \delta_0] > \frac{1}{2}$$

and, under the assumptions of Proposition 2 in Appendix A,

$$(7) \quad \lim_{p \rightarrow \infty} \Pr_\delta[X \in B_2^c] = 1.$$

This shows that if  $\delta$  were known, then usually *some* shrinkage factor  $\tilde{\gamma}$  applied to  $X - \delta_0$  will move  $X$  closer to  $\delta$ , again suggesting a search for estimators of the form (1).

## 2. THE STEIN PARADOX

Assume now that  $Y \sim N_p(0, \sigma^2 1_p)$ , the multivariate normal distribution with mean 0 and covariance matrix

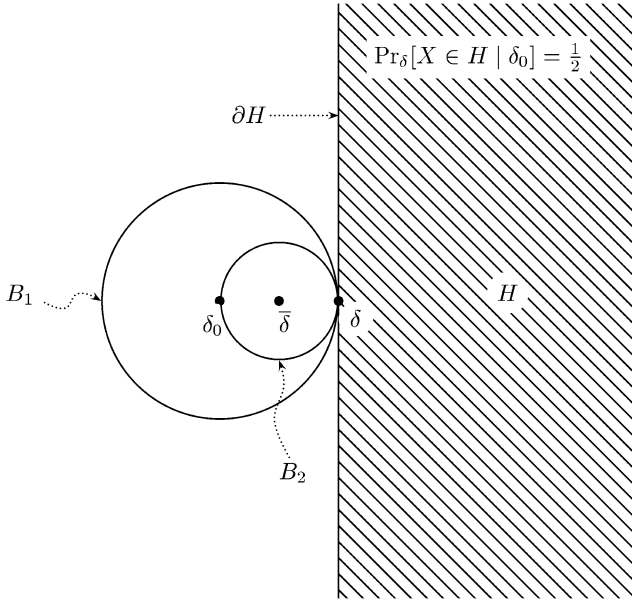


FIG. 1. The balls  $B_1$  and  $B_2$  in (2) and (5).

$\sigma^2 \mathbf{1}_p$ , where  $\sigma^2 > 0$  is known, so  $X \sim N_p(\delta, \sigma^2 \mathbf{1}_p)$ . In this simple case, the James–Stein (JS) estimator for  $\delta$  is given by

$$(8) \quad \begin{aligned} \hat{\delta}_{\text{JS}} &\equiv \hat{\delta}_{\text{JS}}(X; \delta_0) \\ &= \left(1 - \frac{\sigma^2(p-2)}{\|X - \delta_0\|^2}\right)(X - \delta_0) + \delta_0, \end{aligned}$$

where  $\delta_0$  is a fixed but *arbitrary* point in  $\mathbb{R}^p$ . The truncated  $\equiv$  “plus-rule” JS estimator

$$(9) \quad \begin{aligned} \hat{\delta}_{\text{JS}}^+ &\equiv \hat{\delta}_{\text{JS}}^+(X; \delta_0) \\ &= \left(1 - \frac{\sigma^2(p-2)}{\|X - \delta_0\|^2}\right)^+(X - \delta_0) + \delta_0 \end{aligned}$$

is a shrinkage estimator of the form (1). These renowned estimators have the property that when  $p \geq 3$ , they dominate  $X$  under both the mean square error (MSE) and Pitman closeness (PC) criteria:<sup>2</sup> for every fixed  $\delta, \delta_0 \in \mathbb{R}^p$ ,

$$(10) \quad \begin{aligned} E_\delta[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\|^2 | \delta_0] \\ < E_\delta[\|\hat{\delta}_{\text{JS}}(X; \delta_0) - \delta\|^2 | \delta_0] \end{aligned}$$

$$(11) \quad < E_\delta[\|X - \delta\|^2] \equiv p\sigma^2,$$

<sup>2</sup>See Baranchik (1964) or Efron and Morris (1973) for (10), James and Stein (1961), Efron and Morris (1973), Arnold (1981), Anderson (1984), Berger (1985), or Lehmann and Casella (1998) for (11), our Appendix C for (12), and Efron (1975) or Sen, Kubokawa and Saleh (1989) for (13). In Efron’s equation (2.11), page 265, the second inequality should be reversed.

$$(12) \quad \begin{aligned} &Pr_\delta[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| < \|X - \delta\| | \delta_0] \\ &> Pr_\delta[\|\hat{\delta}_{\text{JS}}(X; \delta_0) - \delta\| < \|X - \delta\| | \delta_0] \end{aligned}$$

$$(13) \quad = Pr\left[\chi_p^2\left(\frac{\|\delta - \delta_0\|^2}{4\sigma^2}\right) \geq \frac{\|\delta - \delta_0\|^2}{4\sigma^2} + \frac{p-2}{2}\right]$$

$$(14) \quad > \frac{1}{2}$$

and approaches 1 as  $p \rightarrow \infty$  if  $\frac{\|\delta - \delta_0\|}{\sigma} = o(p)$  (apply Chebyshev’s inequality), where  $\chi_p^2(\eta)$  denotes a noncentral chi-square random variate with  $p$  degrees of freedom and noncentrality parameter  $\eta$ . Note especially that:

- (A) the improvements offered by the JS estimators can be great, especially when  $p$  is large: if  $\delta = \delta_0$ , then  $\text{MSE}(\hat{\delta}_{\text{JS}}^+) < \text{MSE}(\hat{\delta}_{\text{JS}}) = 2\sigma^2 \ll p\sigma^2$ , and if  $\|\delta - \delta_0\| = o(p)$  with  $\sigma^2$  fixed, then  $Pr_\delta[\|\hat{\delta} - \delta\| < \|X - \delta\|] \rightarrow 1$  as  $p \rightarrow \infty$  for both  $\hat{\delta} = \hat{\delta}_{\text{JS}}$  and  $\hat{\delta}_{\text{JS}}^+$ ;
- (B) the MSE and PC dominances of  $X$  by  $\hat{\delta}_{\text{JS}}$  and  $\hat{\delta}_{\text{JS}}^+$  hold even if the true mean  $\delta$  is arbitrarily far from the shrinkage target  $\delta_0$ .

Of the two properties (A) and (B), it is (B) that is most surprising, since it is not difficult to construct estimators that satisfy (A), for example, a Bayes estimator w.r. to a normal prior centered at  $\delta_0$ . However, such a Bayes estimator will not satisfy (B), the difference stemming from the fact that the Bayes estimator will have a constant shrinkage factor, while the shrinkage factors in (8) and (9) are adaptive.<sup>3</sup>

When first discovered, the domination of  $X$  by the JS estimators was highly surprising, because the estimator  $X$  itself is as follows:<sup>4</sup>

- (a) the best unbiased estimator of  $\delta$ ,
- (b) the best translation-invariant estimator of  $\delta$ ,
- (c) the maximum likelihood estimator (MLE) of  $\delta$ ,
- (d) a minimax estimator of  $\delta$ , and
- (e) an admissible estimator of  $\delta$  when  $p = 1$  or 2.

So compelling were these properties of  $X$  that its domination by the JS estimators came to be known as *the Stein Paradox*.<sup>5</sup>

<sup>3</sup>In fact, the JS estimator can be derived via an empirical Bayes argument based on such priors—see Stein [(1966), page 356], Efron and Morris [(1973), pages 117–118], Arnold [(1981), Section 11.4].

<sup>4</sup>Cf. Berger (1985), Lehmann and Casella (1998).

<sup>5</sup>Cf. Efron and Morris (1977).

**3. LOST IN SPACE: THE REVERSE STEIN EFFECT**

*Star Trek, Stardate 4598.0:* The Federation Starship U.S.S. Enterprise, about to rendezvous with interstellar space station Delta, was struck by a mysterious distortion of the space-time continuum that disrupted all its power systems, including navigation, communications, and computers. Out of control, the Enterprise careened wildly and randomly through interstellar space at maximum warp for three days until, equally mysteriously, its warp drive went off-line and the ship came to a full stop. Captain Kirk knew that, without power and communication, their only hope for rescue was to launch a probe that would come close enough to Delta to be detected and convey their present location.

By means of stellar charts, Lieutenant Ohura determined the present location  $X$  of the Enterprise, but because all computer records had been lost, the location  $\delta$  of station Delta was unknown. Mr. Chekov, fresh out of the Space Academy where he studied multivariate statistical analysis under Admiral Emeritus Stein, immediately suggested a solution:

“We can utilize the Stein Effect! Because the Enterprise essentially followed a random walk while out of control we know that  $X \sim N_3(\delta, \sigma^2 1_3)$ , while from the duration of the disruption and the characteristics of our warp engines we know that  $\sigma = 2400$  light-years. If we use the truncated James–Stein estimator  $\hat{\delta}_{JS}^+(X; \delta_0)$  with  $p = 3$  to estimate  $\delta$  by shrinking  $X$  toward a fixed point  $\delta_0$ , then by (11) and (14),  $\hat{\delta}_{JS}^+(X; \delta_0)$  is more likely to be closer to Delta than our present location  $X$  is, no matter where Delta is! And what’s more, we can shrink  $X$  toward any  $\delta_0$  that we like!”

“Amazing!” Kirk said. “Now I wish I had paid more attention in my stats class,” (*smiling to himself: but that’s not how one makes Admiral!*) “But what about  $\delta_0$ ? To what shrinkage target point should we actually send our probe?”

“Why, toward Earth, of course,” Scotty<sup>6</sup> said in his thick Scottish brogue. “The Scotch there is the best in the galaxy.”

“No, toward Qo’noS<sup>7</sup>” Lt. Worf<sup>8</sup> exclaimed. “Perhaps they will send us some fresh qagh<sup>9</sup>—I am *so* tired of this replicated stuff.”

<sup>6</sup>A.k.a. James Doohan, who, during the writing of this paper, beamed out of this universe on July 20, 2005, the 36th anniversary of the first human landing on an extraterrestrial body.

<sup>7</sup>The Capitol of the Klingon Empire.

<sup>8</sup>Yes, we know, Worf didn’t appear until *Star Trek: The Next Generation*—some slack, please.

<sup>9</sup>A Klingon dish of serpent worms, best when served live.

“Permit me to suggest Denobula,” Dr. Phlox<sup>10</sup> offered. “Tomorrow is the tenth wedding anniversary of my third wife and her fourth husband—perhaps the probe might convey my congratulations to them.”

Suggestions for the shrinkage target point  $\delta_0$  were soon received from every member of the 400-person crew, all except Mr. Spock. After several minutes he raised his left eyebrow and said “This is not logical. Please accompany me to the holodeck.<sup>11</sup>”

When the officers were assembled on the holodeck, Spock commanded: “Computer,<sup>12</sup> construct a three-dimensional star chart showing the distribution in the galaxy of the homeworlds  $\delta_0$  of our crew members. What if any statistical properties does this distribution possess?”

“The dis-tri-bu-tion of home-worlds is such that  $\delta_0$  is di-rec-tion-al-ly sym-me-tric a-bout our pre-sent lo-ca-tion  $X$ ,” the computer intoned monotonically.

“Computer, display the following set:

$$(15) \quad \{\delta_0 \mid \exists \check{\gamma} \in [0, 1) \ni \|\hat{\delta}_{\check{\gamma}} - \delta\| < \|X - \delta\|\},$$

where  $\check{\gamma} \equiv \check{\gamma}(X - \delta_0, X - \delta)$  may depend on  $\delta$ .”

“This set is ex-act-ly  $H^c$ , the com-ple-ment of the closed half-space  $H$  in Fig-ure 2 on my mon-i-tor.”

“Then, since  $\Pr[\delta_0 \in H^c \mid X] = \frac{1}{2}$  by directional symmetry, this shows that shrinkage toward a randomly

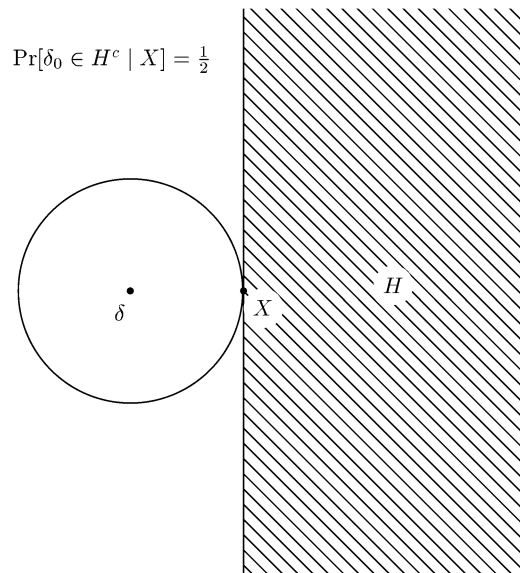


FIG. 2. The complement of  $H$  is the set (15).

<sup>10</sup>Okay, he appeared a century earlier on *Star Trek: Enterprise*—more slack please.

<sup>11</sup>And still more slack.

<sup>12</sup>Ok, let’s suppose that the computer power has been restored, but only momentarily.

chosen  $\delta_0$  would have at most a 50–50 chance of moving  $X$  closer to  $\delta$  even when the shrinkage factor is chosen optimally for  $\delta$ .<sup>13</sup>

“As for James–Stein shrinkage,” Spock continued, “Computer, for representative values of  $\delta$ , display the set of all  $\delta_0$  such that the James–Stein shrinkage estimator  $\hat{\delta}_{\text{JS}}^+(X; \delta_0)$  lies closer to  $\delta$  than does our present location  $X$ .”

“The two re-pres-ent-a-tive ca-ses are now displayed in Fig-ures 3a and 3b<sup>13</sup> on my mon-i-tor.”

“Thank you, Computer. It is apparent from these two displays,” Spock said to the assembled officers, “that the set<sup>14</sup> of  $\delta_0$  such that James–Stein shrinkage toward  $\delta_0$  does more harm than good is quite extensive. Furthermore, since Mr. Chekov assures us that this choice

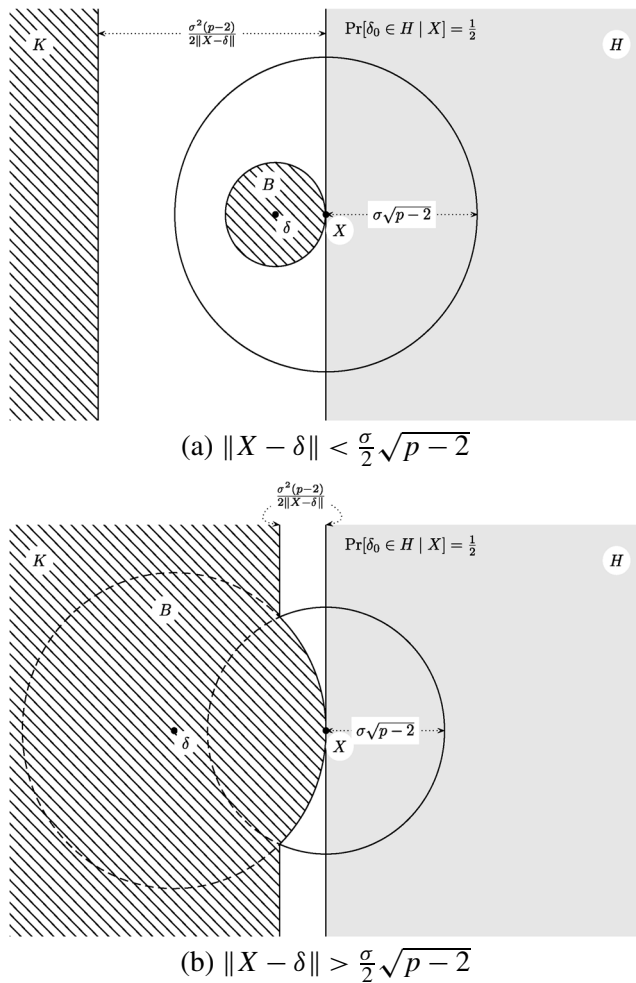


FIG. 3. The cross-hatched region is the set  $\{\delta_0 \mid \|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| < \|X - \delta\|\}$ .

<sup>13</sup>See Appendix D for their derivation.

<sup>14</sup>This set is the complement of the cross-hatched region in Figure 3a or 3b

can be made arbitrarily, in the interest of fairness, we may as well choose  $\delta_0$  at random from our crew members’ homeworlds. But then, contrary to Mr. Chekov’s assertion,  $\hat{\delta}_{\text{JS}}^+(X; \delta_0)$  is *less likely* to be closer to  $\delta$  than is our present location  $X$ .<sup>15</sup>

“More precisely, by the directional symmetry of  $\delta_0$  about  $X$ , it follows from Figures 3a and 3b that

$$(16) \quad \Pr[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| > \|X - \delta\| \mid X] > \Pr[\delta_0 \in H \mid X] = \frac{1}{2}.$$

If  $\delta_0$  is actually symmetrically distributed about  $X$ , then it is easy to see that

$$(17) \quad E[\hat{\delta}_{\text{JS}}^+(X; \delta_0) \mid X] = X,$$

so by Jensen’s inequality,

$$(18) \quad E[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\|^2 \mid X] > E[\|X - \delta\|^2 \mid X] \equiv p\sigma^2 \quad \forall \delta \in \mathbb{R}^p.$$

Furthermore, under additional but still general assumptions,<sup>15</sup>

$$(19) \quad \lim_{p \rightarrow \infty} \Pr_{\delta}[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| > \|X - \delta\|] = 1.$$

Thus, it is likely that James–Stein shrinkage will actually move us farther away from  $\delta$ . I conclude, therefore, that we should simply tether the probe to the Enterprise and hope that Delta can detect our present location  $X$ .<sup>16</sup>

“Boy, Spock, you *are* a party pooper,” Bones<sup>16</sup> said. “I sure hope we don’t shrink toward Vulcan.”

“Resistance is futile,” said Seven-of-Nine.<sup>17</sup>

“But, but,—I don’t understand this,” Chekov stammered. “How can the James–Stein estimator be inferior to  $X$  after all? Don’t (16) and (18) contradict (14) and (11)? For example, under *any* probability distribution for  $\delta_0$ , (11) yields

$$(20) \quad E_{\delta}[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\|^2] < E_{\delta}[\|X - \delta\|^2] \equiv p\sigma^2 \quad \forall \delta \in \mathbb{R}^p,$$

while (18) yields

$$(21) \quad E_{\delta}[\|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\|^2] > E_{\delta}[\|X - \delta\|^2] \equiv p\sigma^2 \quad \forall \delta \in \mathbb{R}^p.$$

I am so confused!”

“Beam me to the bar, Scotty,” Kirk finally mumbled. “Maybe I can figure this out after I belt down a few.”

<sup>15</sup>See Proposition 3 in Appendix D.

<sup>16</sup>Dr. McCoy.

<sup>17</sup>Right again, but how could we leave her out?



#### 4. TO SHRINK OR NOT TO SHRINK—THAT IS THE QUESTION

Mr. Spock quickly assured Mr. Chekov that no formal contradiction had occurred: the probabilities and expectations appearing in (11), (14), (16), and (18) are conditional probabilities and conditional expectations with different conditioning variables. Furthermore, the joint distributions of  $(X, \delta_0)$  in (20) and (21) are different, having joint pdfs of the forms  $f_\delta(X)f(\delta_0)$  and  $f_\delta(X)f(\delta_0|X)$ , respectively. In the former,  $X$  and  $\delta_0$  are independent, whereas in the latter,  $\delta_0$  is dependent on  $X$ .

However, Captain Kirk's dilemma<sup>18</sup> remains: to shrink or not to shrink? If, according to property (B), the shrinkage target  $\delta_0$  can be chosen arbitrarily and still reduce the MSE and PC, can choosing  $\delta_0$  at random in some symmetric manner actually increase the MSE and PC?

The short answer is yes, the Reverse Stein Effect is just as real as the original Stein Effect itself—both are simply manifestations of the strong curvature of spheres in multi-dimensional Euclidean space. Figures 3a, 3b, and the results (16), (18), and (19) show that, without some prior knowledge of the location  $\delta$ , *Captain Kirk should not shrink X*. If the shrinkage target  $\delta_0$  is chosen without reliable prior information but instead is based upon the data  $X$ , the minimax/Bayesian robustness property (B) of the JS estimator is lost and no longer guarantees that shrinking is not harmful on average.

The implications for statistical practice are apparent. A shrinkage estimator is only as good as, *but no better than*, the prior information upon which it is based. Without reliable *prior*, as opposed to *posterior*,<sup>19</sup> information, *shrinkage is likely to decrease the accuracy of estimation*. As Barnard<sup>20</sup> concluded, if the statistical estimation problem is truly invariant under translation, then the best invariant estimator should be used, namely,  $X$  itself.

<sup>18</sup>Captain Kirk is “exactly in the position of Buridan’s ass,” as described in Barnard’s discussion of the noninvariant nature of the James–Stein estimator in Stein [(1962), page 288]. The ass, when presented with two bales of hay, equidistant to his right and left, refused to move, seeing no reason to prefer one direction over the other. Like Barnard, we maintain that, in the absence of additional influences, such as prior information about the delectability of dextral vs. sinistral hay (or a loss function reflecting a negative effect of starvation), the ass’s refusal to budge was correct.

<sup>19</sup>As represented, for example, by “data-dependent” priors.

<sup>20</sup>Cf. Stein [(1962), page 288].

#### APPENDIX A: DIRECTIONAL AND SPHERICAL SYMMETRY; VERIFICATION OF (4)

DEFINITION 1.  $Y \in \mathbb{R}^p$  is *directionally symmetric* if  $\vec{Y} \stackrel{d}{=} -\vec{Y}$ , where  $\vec{Y} := \frac{Y}{\|Y\|}$  is the unit vector in the direction of  $Y$ .  $Y$  is *directionally symmetric about  $y_0$*  if  $Y - y_0$  is directionally symmetric.

Clearly  $Y$  is directionally symmetric if  $Y$  is *symmetric*:  $Y \stackrel{d}{=} -Y$ . Thus, any multivariate normal or elliptically contoured random vector  $Y$  centered at 0 is directionally symmetric. Directional symmetry is much weaker than symmetry, as seen from the following result.

PROPOSITION 1. *Let  $Y$  be an absolutely continuous random vector in  $\mathbb{R}^p$ . The following are equivalent:*

- (a)  $Y$  is directionally symmetric.
- (b)  $\Pr[Y \in C] = \Pr[-Y \in C]$  for every closed convex cone  $C \subseteq \mathbb{R}^p$ .
- (c)  $\Pr[Y \in H] = \frac{1}{2}$  for every central (i.e.,  $0 \in \partial H$ ) halfspace  $H \subseteq \mathbb{R}^p$ .

PROOF. The implications (a)  $\Leftrightarrow$  (b)  $\Rightarrow$  (c) are straightforward. We will show that (c)  $\Rightarrow$  (a). Let  $P$  (resp.,  $Q$ ) denote the probability distribution of  $Y$  (resp.,  $\vec{Y}$ ). First note that since  $P[\partial H] = 0$ ,  $P[H] = \frac{1}{2}$  is equivalent to

$$(22) \quad P[H] = P[H^c] = P[-H].$$

Thus, for any two central halfspaces  $H$  and  $H_0$ ,

$$\begin{aligned} P[H \cap H_0] - P[H^c \cap H_0^c] &= P[H] - P[H_0^c] \\ &= P[H^c] - P[H_0] \\ &= P[H^c \cap H_0^c] - P[H \cap H_0], \end{aligned}$$

hence,

$$(23) \quad \begin{aligned} P[H \cap H_0] &= P[H^c \cap H_0^c] \\ &= P[(-H) \cap (-H_0)]. \end{aligned}$$

It follows from Lemma 1 below that

$$(24) \quad Q[A \cap S_0] = Q[(-A) \cap (-S_0)]$$

for every Borel set  $A \subseteq S^p$  (the unit sphere in  $\mathbb{R}^p$ ), where  $S_0 = H_0 \cap S^p$ . Thus,

$$(25) \quad \begin{aligned} Q[A] &= Q[A \cap S_0] + Q[A \cap (-S_0)] \\ &= Q[(-A) \cap (-S_0)] + Q[(-A) \cap (S_0)] \\ &= Q[-A] \end{aligned}$$

for every such  $A$ , hence, (a) holds.  $\square$

LEMMA 1. *Let  $Y$  be an absolutely continuous random vector in  $\mathbb{R}^p$  and let  $P$  and  $Q$  be as defined above. Suppose that  $H_0 \subset \mathbb{R}^p$  is a central halfspace such that  $P[H_0] = \frac{1}{2}$ , so also  $Q[S_0] = \frac{1}{2}$  where  $S_0 = H_0 \cap S^p$ . If*

$$(26) \quad Q[S | S_0] = Q[-S | -S_0]$$

for every hemisphere  $S \subset S^p$ , then

$$(27) \quad Q[A | S_0] = Q[-A | -S_0]$$

for every Borel set  $A \subseteq S^p$ , which is equivalent to (24) because  $Q[\pm S_0] = P[\pm H_0] = \frac{1}{2}$ . Since every hemisphere  $S$  has the form  $H \cap S^p$  for some central halfspace  $H$ , (26) is equivalent to

$$(28) \quad P[H | H_0] = P[-H | -H_0]$$

for every central halfspace  $H \subset \mathbb{R}^p$ , which in turn is equivalent to (23).

PROOF. Without loss of generality, set  $H_0 = \{(y_1, \dots, y_{p-1}, y_p) \mid y_p > 0\}$  so

$$(29) \quad S_0 = \left\{ (y_1, \dots, y_{p-1}, y_p) \mid \sum_{i=1}^p y_i^2 = 1, y_p > 0 \right\},$$

and let  $\pi$  denote the stereographic projection<sup>21</sup> of  $S_0$  onto its tangent hyperplane  $L_0 \equiv \{(y_1, \dots, y_{p-1}, 1)\}$ . Then the relation

$$(30) \quad \pi(S \cap S_0) = K$$

determines a bijection between the sets of all hemispheres  $S \subset S^p$  and all (not necessarily central) halfspaces  $K \subset L_0$ .

Let  $\tilde{Q}$  denote the probability measure on  $S^p$  given by

$$(31) \quad \tilde{Q}[A] = Q[-A | -S_0],$$

so (26) states that

$$(32) \quad Q[S | S_0] = \tilde{Q}[S]$$

for every Borel set  $A \subseteq S^p$ . Let  $R$  and  $\tilde{R}$  denote the probability measures induced on  $L_0$  by  $Q[\cdot | S_0]$  and  $\tilde{Q}$ , respectively, under the mapping  $\pi$ , that is,

$$(33) \quad R[B] = Q[\pi^{-1}(B) | S_0],$$

$$(34) \quad \tilde{R}[B] = \tilde{Q}[\pi^{-1}(B)]$$

for every Borel set  $B \subseteq L_0$ . Then for each halfspace  $K \subset L_0$ ,

$$(35) \quad \begin{aligned} R[K] &= Q[\pi^{-1}(K) | S_0] = Q[S \cap S_0 | S_0] \\ &\equiv Q[S | S_0], \end{aligned}$$

$$(36) \quad \begin{aligned} \tilde{R}[K] &= \tilde{Q}[\pi^{-1}(K)] \\ &= \tilde{Q}[S \cap S_0] \equiv Q[-(S \cap S_0) | -S_0] \end{aligned}$$

$$(37) \quad = Q[-S | -S_0] \equiv \tilde{Q}[S],$$

hence,  $R[K] = \tilde{R}[K] \forall K$  by (32). Thus by the Cramér–Wold device [cf. Billingsley (1979), page 334],  $R[B] = \tilde{R}[B] \forall B$ , hence, setting  $A = \pi^{-1}(B)$  in (33) and (34),  $Q[A | S_0] = \tilde{Q}[A] \forall A$ , which establishes (27).  $\square$

DEFINITION 2.  $Y \in \mathbb{R}^p$  is *spherically symmetric*  $\equiv$  *orthogonally invariant* if  $Y \stackrel{d}{=} \Gamma Y$  for every orthogonal transformation  $\Gamma$  of  $\mathbb{R}^p$ .  $Y$  is *spherically symmetric about*  $y_0$  if  $Y - y_0$  is spherically symmetric.

For example,  $Y \sim N_p(0, \sigma^2 1_p)$  is spherically symmetric. Clearly spherical symmetry implies symmetry. It is well known that  $Y$  is spherically symmetric iff  $\tilde{Y}$  is uniformly distributed on the unit sphere  $S_p$  and is independent of  $\|Y\|$ . We now use this fact to verify (4) by the following proposition, where  $\delta, \delta_0, X, Y, \sigma, \psi$ , and  $\tau$  all depend on  $p$ .

PROPOSITION 2. *Assume that:*

- (i)  $\delta_0$  is (fixed or) random is independent of  $X$ ;
- (ii)  $Y \equiv X - \delta$  is spherically symmetric;
- (iii)  $\frac{\|\delta_0 - \delta\|}{\|Y\|} \equiv \frac{\|\delta_0 - \delta\|}{\|X - \delta\|} = o(p^{1/2})$  in probability as  $p \rightarrow \infty$ . Then [cf. (4)]

$$(38) \quad \begin{aligned} &\lim_{p \rightarrow \infty} \Pr_\delta[\|X - \delta_0\| > \|\delta - \delta_0\|] \\ &\equiv \lim_{p \rightarrow \infty} \Pr_\delta[X \in B_1^c] = 1. \end{aligned}$$

The boundedness assumption (iii) is satisfied, for example, if  $X \sim N_p(\delta, \sigma^2 1_p)$  and  $\delta_0 \sim N_p(\psi, \tau^2 1_p)$  with  $\|\psi - \delta\|/\sigma = o(p)$  and  $\tau/\sigma = o(p^{1/2})$ .

PROOF. Let  $\mu_p$  denote the uniform probability measure on  $S_p$ . By (i) and (ii),  $\Pr_\delta[X \in B_1 | \delta_0]$  depends on  $\delta_0$  only via  $\|\delta_0 - \delta\|$  (the radius of  $B_1$ ), and

$$(39) \quad \begin{aligned} &\Pr_\delta[X \in B_1 | \|\delta_0 - \delta\|] \\ &= \Pr[Y \in B_1 - \delta | \|\delta_0 - \delta\|] \\ &= E\{\Pr[\tilde{Y} \in \|Y\|^{-1}(B_1 - \delta) | \|Y\|, \|\delta_0 - \delta\|] \\ &\quad \| \delta_0 - \delta \| \} \\ &= E\{\mu_p(\|Y\|^{-1}(B_1 - \delta)) | \|\delta_0 - \delta\| \}. \end{aligned} \tag{40}$$

<sup>21</sup>Cf. Ambartzumian [(1982), page 26], Watson [(1983), page 23].

Because  $B_1 - \delta$  is a ball with  $0 \in \partial(B_1 - \delta)$ , the set  $(\|Y\|^{-1}(B_1 - \delta)) \cap \mathcal{S}_p$  is a spherical cap on  $\mathcal{S}_p$  which, after some geometry, can be expressed as

$$(41) \quad \left\{ (z_1, \dots, z_p) \mid \frac{z_1}{(z_1^2 + \dots + z_p^2)^{1/2}} \geq \frac{\|Y\|}{2\|\delta_0 - \delta\|} \right\}$$

when  $\|Y\| \leq 2\|\delta_0 - \delta\|$ , and is empty otherwise. Furthermore,  $\mu_p$  can be represented as the distribution of  $\vec{Z}$ , where  $Z \equiv (Z_1, \dots, Z_p) \sim N_p(0, 1_p)$ . Therefore,

$$(42) \quad \begin{aligned} & \mu_p(\|Y\|^{-1}(B_1 - \delta)) \\ &= \frac{1}{2} \Pr \left[ \frac{Z_1^2}{Z_1^2 + \dots + Z_p^2} \geq \frac{\|Y\|^2}{2\|\delta_0 - \delta\|^2} \mid \right. \\ & \quad \left. \|Y\|, \|\delta_0 - \delta\| \right] \end{aligned}$$

when  $\|Y\| \leq 2\|\delta_0 - \delta\|$ , and = 0 otherwise. Thus, by (40),

$$(43) \quad \begin{aligned} & \Pr_\delta[X \in B_1 \mid \|\delta_0 - \delta\|] \\ & \leq \frac{1}{2} \Pr \left[ \frac{Z_1^2}{Z_1^2 + \dots + Z_p^2} \geq \frac{\|Y\|^2}{2\|\delta_0 - \delta\|^2} \mid \right. \\ & \quad \left. \|\delta_0 - \delta\| \right], \end{aligned}$$

hence,

$$(44) \quad \begin{aligned} & \Pr_\delta[X \in B_1] \\ & \leq \frac{1}{2} \Pr \left[ \frac{Z_1^2}{Z_1^2 + \dots + Z_p^2} \geq \frac{\|Y\|^2}{2\|\delta_0 - \delta\|^2} \right]. \end{aligned}$$

But  $Z_1^2 + \dots + Z_p^2 = O(p)$  in probability by the Law of Large Numbers, so by (iii), the right-hand side of (44) approaches 0 as  $p \rightarrow \infty$ , which yields (38).  $\square$

#### APPENDIX B: VERIFICATION OF (5)

If we set  $h(\gamma) = \|\hat{\delta}_\gamma(X; \delta_0) - \delta\|^2$  and  $\bar{\delta} = \frac{1}{2}(\delta_0 + \delta)$ , then

$$(45) \quad \begin{aligned} h'(1) &= 2(X - \delta_0)^t(X - \delta) \\ &= 2[\|X - \bar{\delta}\|^2 - \|\delta - \bar{\delta}\|^2]. \end{aligned}$$

Since the right-hand side of (5) is the set  $\{X \mid h'(1) > 0\}$ , (5) follows.

#### APPENDIX C: VERIFICATION OF (12)

First note that  $\hat{\delta}_{\text{JS}}(X; \delta_0) \neq \hat{\delta}_{\text{JS}}^+(X; \delta_0)$  iff  $\|X - \delta_0\| < \sigma\sqrt{p-2}$  (the ball  $B_1$  of radius  $\sigma\sqrt{p-2}$  centered at

$\delta_0$ ; see Figures 4a, 4b), in which case  $\hat{\delta}_{\text{JS}}^+(X; \delta_0) = \delta_0$ . Define

$$(46) \quad C := \{X \mid \|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| < \|X - \delta\|\},$$

$$(47) \quad D := \{X \mid \|\hat{\delta}_{\text{JS}}(X; \delta_0) - \delta\| < \|X - \delta\|\},$$

so (12) is equivalent to

$$(48) \quad \Pr[C \mid \delta_0] > \Pr[D \mid \delta_0].$$

Since  $C \setminus B_1 = D \setminus B_1$ , this is equivalent to

$$(49) \quad \Pr[C \cap B_1 \mid \delta_0] > \Pr[D \cap B_1 \mid \delta_0].$$

But (see Figures 4a, 4b)

$$(50) \quad X \in B_2 \Rightarrow \|\hat{\delta}_{\text{JS}}(X; \delta_0) - \delta\| > \|X - \delta\|,$$

where  $B_2$  is the ball of radius  $\|\delta_0 - \delta\|$  centered at  $\delta$ , so  $D \cap B_2 = \emptyset$ . Thus,

$$(51) \quad C \cap B_1 = (B_1 \setminus B_2) \supseteq D \cap B_1,$$

hence, (49) holds. (Note that no distributional assumption on  $X$  is needed.)

#### APPENDIX D: VERIFICATION OF FIGURES 3A AND 3B; VERIFICATION OF (19)

First we verify that Figures 3a and 3b accurately depict the region

$$(52) \quad \{\delta_0 \mid \|\hat{\delta}_{\text{JS}}^+(X; \delta_0) - \delta\| < \|X - \delta\|\}.$$

Let  $\gamma = (1 - \frac{\sigma^2(p-2)}{\|X - \delta_0\|^2})^+$ , so  $0 \leq \gamma < 1$  and  $\hat{\delta}_{\text{JS}}^+(X; \delta_0) = \gamma(X - \delta_0) + \delta_0$ . Each of the following inequalities is equivalent to that in (52):

$$\begin{aligned} & \|(1 - \gamma)(\delta_0 - \delta) + \gamma(X - \delta)\|^2 < \|X - \delta\|^2, \\ & (1 - \gamma)^2\|\delta_0 - \delta\|^2 + 2\gamma(1 - \gamma)(\delta_0 - \delta)^t(X - \delta) \\ & \quad < (1 - \gamma^2)\|X - \delta\|^2, \\ & (1 - \gamma)\|\delta_0 - \delta\|^2 + 2\gamma(\delta_0 - \delta)^t(X - \delta) \\ & \quad < (1 + \gamma)\|X - \delta\|^2, \\ & \|\delta_0 - \delta\|^2 < \gamma[\|X - \delta_0\|^2] + \|X - \delta\|^2. \end{aligned}$$

If  $\|\delta_0 - X\|^2 < \sigma^2(p-2)$ , that is,  $\delta_0$  lies inside the ball of radius  $\sigma\sqrt{p-2}$  (see Figures 3a, 3b), then  $\gamma = 0$  and the last inequality becomes  $\|\delta_0 - \delta\|^2 < \|X - \delta\|^2$ , which holds iff  $\delta_0$  lies inside the ball  $B$  of radius  $\|X - \delta\|$  centered at  $\delta$ . If  $\|\delta_0 - X\|^2 < \sigma^2(p-2)$ , that is,  $\delta_0$  lies outside this ball, then  $\gamma = (1 - \frac{\sigma^2(p-2)}{\|X - \delta_0\|^2})$  and

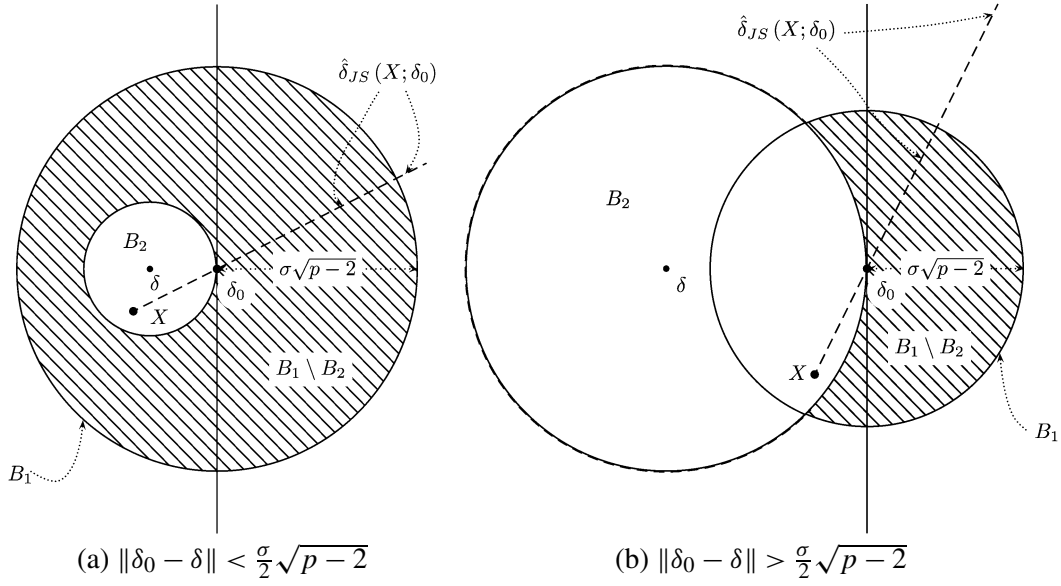


FIG. 4. Illustrating the implication (50).

the last inequality instead is equivalent to each of the following:

$$\|\delta_0 - \delta\|^2 < \|X - \delta_0\|^2 + \|X - \delta\|^2 - \sigma^2(p-2),$$

$$\sigma^2(p-2) < 2\|X - \delta\|^2 + 2(X - \delta)^t(\delta - \delta_0),$$

$$\sigma^2(p-2) < 2(X - \delta)^t(X - \delta_0),$$

$$\frac{\sigma^2(p-2)}{2\|X - \delta\|} < \overrightarrow{(X - \delta)}^t(X - \delta_0),$$

which holds exactly in the open halfspace \$K\$ shown in Figures 3a and 3b. Thus, the region (52) is the union \$B \cup K\$ of the cross-hatched regions in these figures.

Finally, we verify (19), which now can be written equivalently as

$$(53) \quad \lim_{p \rightarrow \infty} \Pr_\delta[\delta_0 \in B \cup K] = 0,$$

by the following proposition, in which \$\delta\$, \$\delta\_0\$, \$X\$, \$V\$, \$\sigma\$, and \$\tau\$ now depend on \$p\$.

**PROPOSITION 3.** Assume the following:

(i') \$V \equiv \delta\_0 - X\$ is independent of \$X\$;

(ii') \$V\$ is spherically symmetric;

(iii') \$\frac{\|X - \delta\|}{\|V\|} \equiv \frac{\|X - \delta\|}{\|\delta\_0 - X\|} = o(p^{1/2})\$ in probability as \$p \to \infty\$;

(iv') \$\sigma^{-2}\|X - \delta\| \cdot \|V\| \equiv \sigma^{-2}\|X - \delta\| \cdot \|\delta\_0 - X\| = o(p^{3/2})\$ in probability as \$p \to \infty\$. Then [cf. (19)]

$$(54) \quad \lim_{p \rightarrow \infty} \Pr_\delta[\|\hat{\delta}_{JS}^+(X; \delta_0) - \delta\| > \|X - \delta\|] = 1.$$

The boundedness assumption (iii') [resp., (iv')] is satisfied, for example, if \$X \sim N\_p(\delta, \sigma^2 1\_p)\$ and \$\delta\_0 \sim\$

\$N\_p(X, \tau^2 1\_p)\$ with \$\tau/\sigma = o(p^{1/2})\$ [resp., \$\sigma/\tau = o(p^{1/2})\$], so both are satisfied if \$\tau/\sigma \sim p^\epsilon\$ with \$0 \le \epsilon < 1/2\$.

**PROOF.** By the argument that yielded (38) in Appendix A [with (i)–(iii) and \$X\$, \$Y\$, \$B\_1 - \delta\$, and \$\|\delta\_0 - \delta\|\$ replaced by (i')–(iii') and \$\delta\_0\$, \$V\$, \$B - X\$, and \$\|X - \delta\|\$, we obtain

$$(55) \quad \lim_{p \rightarrow \infty} \Pr_\delta[\delta_0 \in B] = \lim_{p \rightarrow \infty} \Pr_\delta[V \in B - X] = 0.$$

Next, again by the argument in Appendix A but with \$B\_1 - \delta\$ replaced by \$K - X\$,

$$(56) \quad \Pr_\delta[\delta_0 \in K \mid \|X - \delta\|]$$

$$(57) \quad = E\{\mu_p(\|V\|^{-1}(K - X)) \mid \|X - \delta\|\}$$

$$(58) \quad = \frac{1}{2} \Pr \left[ \frac{Z_1^2}{Z_1^2 + \dots + Z_p^2} \geq \frac{\sigma^4(p-2)^2}{4\|X - \delta\|^2 \|V\|^2} \mid \|X - \delta\| \right].$$

Thus, by (iv'),

$$(59) \quad \lim_{p \rightarrow \infty} \Pr_\delta[\delta_0 \in K] = \frac{1}{2} \lim_{p \rightarrow \infty} \Pr \left[ \frac{Z_1^2}{Z_1^2 + \dots + Z_p^2} \geq \frac{\sigma^4(p-2)^2}{4\|X - \delta\|^2 \|V\|^2} \right] = 0,$$

so (53) and (54) are confirmed. \$\square\$



## ACKNOWLEDGMENTS

We gratefully acknowledge the contributions of T. W. Anderson, Steen Andersson, Morris Eaton, Charlie Geyer, Erich Lehmann, Ingram Olkin, and Charles Stein to our understanding of the role of invariance in statistical analysis. We warmly thank Mathias Drton, Brad Efron, Carl Morris, and Jon Wellner for helpful comments and suggestions. This research was supported in part by NSA Grant MSPF-05G-014 and Grant R-155-000-081-112 from the National University of Singapore.

## REFERENCES

- AMBARTZUMIAN, R. V. (1982). *Combinatorial Integral Geometry*. Wiley, New York. [MR0679133](#)
- ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York. [MR0771294](#)
- ARNOLD, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York. [MR0606011](#)
- BARANCHIK, A. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Unpublished Ph.D. thesis, Technical Report 51, Dept. Statistics, Stanford Univ. [MR2614375](#)
- BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag, New York. [MR0804611](#)
- BILLINGSLEY, P. (1979). *Probability and Measure*. Wiley, New York. [MR0534323](#)
- EFRON, B. (1975). Biased versus unbiased estimation. *Adv. in Math.* **16** 259–277. [MR0375555](#)
- EFRON, B. and MORRIS, C. (1973). Stein's estimation rule and its competitors—an empirical Bayes approach. *J. Amer. Statist. Assoc.* **68** 117–130. [MR0388597](#)
- EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Scientific American* **236** 119–127.
- JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* **1** 311–319. Univ. California Press, Berkeley. [MR0133191](#)
- LEHMANN, E. and CASELLA, G. (1998). *Theory of Point Estimation*, 2nd ed. Springer, New York. [MR1639875](#)
- SEN, P. K., KUBOKAWA, T. and SALEH, A. K. E. (1989). The Stein paradox in the sense of the Pitman measure of closeness. *Ann. Statist.* **17** 1375–1386. [MR1015158](#)
- STEIN, C. M. (1962). Confidence sets for the mean of a multivariate normal distribution (with discussion). *J. Roy. Statist. Soc. Ser. B* **24** 265–296. [MR0148184](#)
- STEIN, C. M. (1966). An approach to the recovery of inter-block information. In *Festschrift for J. Neymann* (F. N. David, ed.). Wiley, New York. [MR0210232](#)
- WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York. [MR0709262](#)