Reversing the Stein Effect

Michael D. Perlman and Sanjay Chaudhuri

Department of Statistics, University of Washington, Seattle, WA 98195, USA
Department of Statistics and Applied Probability, National University of Singapore, Singapore 117546

michael@ms.washington.edu
sanjay@stat.nus.edu.sg

17 November 2006

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 δ 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*[1], it is easy to heuristically justify "shrinkage" estimators for δ of the form

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

where $\gamma \equiv \gamma(X-\delta_0) \in [0,1)$ and δ_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 δ and δ_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 δ_0 and let H be the halfspace bounded by a hyperplane ∂H tangent to B_1 at δ (see Figure 1). Then

$$\left\{ X \;\middle|\; \|X-\delta_0\| > \|\delta-\delta_0\| \right\} = B_1^c, \tag{2}$$

$$\begin{aligned}
\Pr_\delta\left[\, \|X-\delta_0\| > \|\delta-\delta_0\| \;\middle|\; \delta_0\right] &= \Pr_\delta\left[\, X \in B_1^c \;\middle|\; \delta_0\right] \\
&> \Pr_\delta[\, X \in H \mid \delta_0] \\
&= \frac{1}{2},
\end{aligned} \tag{3}$$

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

$$\lim_{p\to\infty} \Pr_\delta\left[\, \|X-\delta_0\| > \|\delta-\delta_0\|\,\right] \equiv \lim_{p\to\infty} \Pr_\delta\left[\, X \in B_1^c\right] = 1. \tag{4}$$

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 δ as $(\delta-\delta_0)+\delta_0$ immediately leads to estimators for δ of the form (1).

Second (see Appendix 2),

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

where $\tilde{\gamma} \equiv \tilde{\gamma}(X-\delta_0, \delta-\delta_0)$ is allowed to depend on δ 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

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

and, under the assumptions of Proposition 2 in Appendix 1,

$$\lim_{p\to\infty} \Pr_\delta[X \in B_2^c] = 1. \tag{7}$$

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

---

[1] $\vec{Y} \stackrel{d}{=} -\vec{Y}$, where $\vec{Y} := Y/\|Y\|$ is the unit vector in the direction of Y (see Appendix 1).

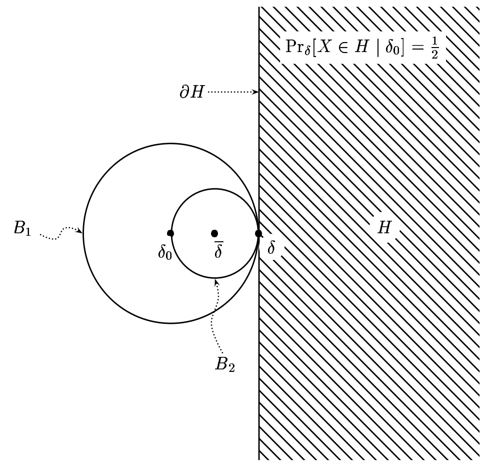


Figure 1: The balls B_1 and B_2 in (2) and (5).

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

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

where δ_0 is a fixed but *arbitrary* point in $\mathbb{R}^p$. The truncated $\equiv$ "plus-rule" JS estimator

$$\hat{\delta}_{JS}^{+} \equiv \hat{\delta}_{JS}^{+}(X; \delta_0) = \left(1 - \frac{\sigma^2(p-2)}{\|X - \delta_0\|^2}\right)^{+}(X - \delta_0) + \delta_0 \tag{9}$$

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[2]: for every fixed $\delta, \delta_0 \in \mathbb{R}^p$,

$$\mathrm{E}_\delta[\|\hat{\delta}_{JS}^{+}(X; \delta_0) - \delta\|^2 \mid \delta_0] < \mathrm{E}_\delta[\|\hat{\delta}_{JS}(X; \delta_0) - \delta\|^2 \mid \delta_0] \tag{10}$$

$$< \mathrm{E}_\delta[\|X - \delta\|^2] \equiv p\sigma^2, \tag{11}$$

[2]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 3 for (12), and Efron (1975) or Sen *et al.* (1989) for (13). In Efron's eqn. (2.11), p.265, the second inequality should be reversed.

$$
\begin{aligned}
&\Pr_\delta[\,\|\hat{\delta}^+_{JS}(X;\delta_0)-\delta\| < \|X-\delta\| \bigm| \delta_0] \\
&\qquad > \Pr_\delta[\,\|\hat{\delta}_{JS}(X;\delta_0)-\delta\| < \|X-\delta\| \bigm| \delta_0] && (12)\\
&\qquad = \Pr\left[\chi^2_p\left(\frac{\|\delta-\delta_0\|^2}{4\sigma^2}\right) \geq \frac{\|\delta-\delta_0\|^2}{4\sigma^2} + \frac{p-2}{2}\right] && (13)\\
&\qquad > \frac{1}{2} && (14)
\end{aligned}
$$

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

(A) the improvements offered by the JS estimators can be great, especially when p is large: if $\delta = \delta_0$ then MSE($\hat{\delta}^+_{JS}$) <MSE($\hat{\delta}_{JS}$) $= 2\sigma^2 \ll p\sigma^2$, and if $\|\delta - \delta_0\| = o(p)$ with σ^2 fixed then $\Pr_\delta[\,\|\hat{\delta}-\delta\| < \|X-\delta\|] \to 1$ as $p \to \infty$ for both $\hat{\delta} = \hat{\delta}_{JS}$ and $\hat{\delta}^+_{JS}$;

(B) the MSE and PC dominances of X by $\hat{\delta}_{JS}$ and $\hat{\delta}^+_{JS}$ hold even if the true mean δ is arbitrarily far from the shrinkage target δ_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 δ_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[3].

When first discovered, the domination of X by the JS estimators was highly surprising, because the estimator X itself is[4]:

(a) the best unbiased estimator of δ,

(b) the best translation-invariant estimator of δ,

(c) the maximum likelihood estimator (MLE) of δ,

(d) a minimax estimator of δ, and

(e) an admissible estimator of δ 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*[5].

[3]In fact, the JS estimator can be derived via an empirical Bayes argument based on such priors – see Stein (1966, p.356), Efron and Morris (1973, pp.117-118), Arnold (1981, §11.4).

[4]cf. Berger (1985), Lehmann and Casella (1998).

[5]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 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 δ of station Delta was unknown. Mr. Chekov, fresh out of 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 δ by shrinking X toward a fixed point δ_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 δ_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 δ_0? To what shrinkage target point should we actually send our probe?

"Why, toward Earth, of course," Scotty[6] said in his thick Scottish brogue. "The Scotch there is the best in the galaxy."

"No, toward Qo'noS[7]" Lt. Worf[8] exclaimed. "Perhaps they will send us some fresh qagh[9] – I am *so* tired of this replicated stuff."

"Permit me to suggest Denobula", Dr. Phlox[10] 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 δ_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[11]".

[6] 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.

[7] The Capitol of the Klingon Empire.

[8] Yes, we know, Worf didn't appear until *Star Trek: The Next Generation* – some slack, please.

[9] A Klingon dish of serpent worms, best when served live.

[10] Okay, he appeared a century earlier on *Star Trek: Enterprise* – more slack please.

[11] And still more slack.

When the officers were assembled on the holodeck, Spock commanded: "Computer[12], construct a three-dimensional star chart showing the distribution in the galaxy of the homeworlds δ_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 δ_0 is di-rec-tion-al-ly symme-tric a-bout our pre-sent lo-ca-tion X." the computer intoned monotonically.

"Computer, display the following set:

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

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

"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 $\Pr[\,\delta_0 \in H^c|X\,] = \frac{1}{2}$ by directional symmetry, this shows that shrinkage toward a randomly chosen δ_0 would have at most a 50-50 chance of moving X closer to δ even when the shrinkage factor is chosen optimally for δ."

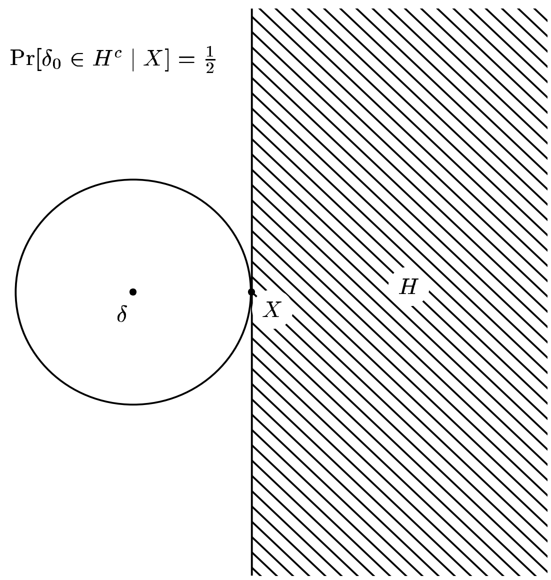


Figure 2: The complement of H is the set (15).

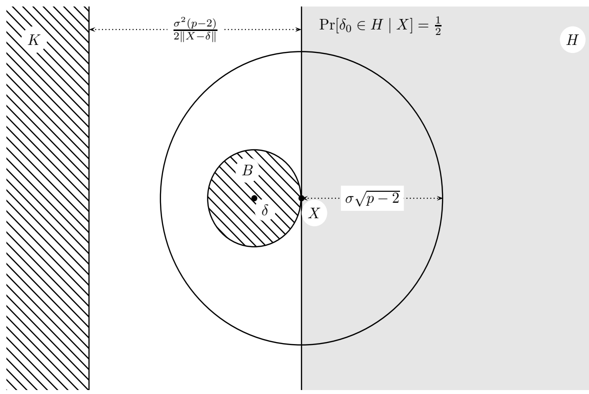
"As for James-Stein shrinkage," Spock continued, "Computer, for representative values of δ, display the set of all δ_0 such that the James-Stein shrinkage estimator $\hat{\delta}^+_{JS}(X;\delta_0)$ lies closer to δ than does our present location X."

"The two re-pres-ent-a-tive ca-ses are now dis-played in Fig-ures 3a and 3b[13] on my mon-i-tor."

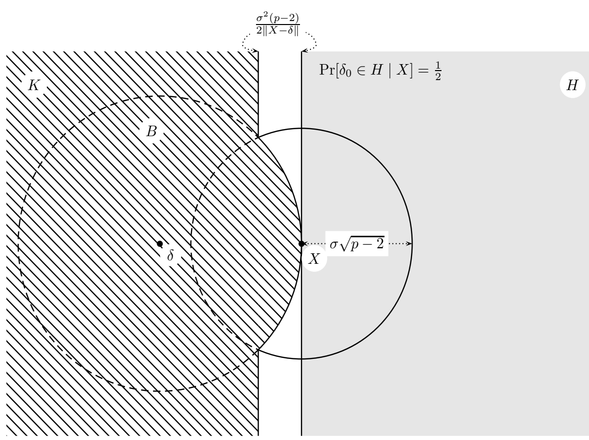
[12] Ok, let's suppose that the computer power has been restored, but only momentarily.

[13] See Appendix 4 for their derivation.

"Thank you, Computer. It is apparent from these two displays," Spock said to the assembled officers, "that the set[14] of δ_0 such that James-Stein shrinkage toward δ_0 does more harm than good is quite extensive. Furthermore, since Mr. Chekov assures us that this choice can be made arbitrarily, in the interest of fairness we may as well choose δ_0 at random from our crew members' home-worlds. But then, contrary to Mr. Chekov's assertion, $\hat{\delta}^{+}_{JS}(X;\delta_0)$ is *less likely* to be closer to δ than is our present location X.



(a) $\|X-\delta\| < \frac{\sigma}{2}\sqrt{p-2}$



(b) $\|X-\delta\| > \frac{\sigma}{2}\sqrt{p-2}$

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

"More precisely, by the directional symmetry of δ_0 about X, it follows from Figures 3a and 3b that

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

[14] This set is the complement of the cross-hatched region in Figure 3a or 3b

If δ_0 is actually symmetrically distributed about X then it is easy to see that

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

so by Jensen's inequality,

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

Furthermore, under additional but still general assumptions[15],

$$\lim_{p\to\infty} \mathrm{Pr}_\delta[\,\|\hat{\delta}^{+}_{JS}(X;\delta_0) - \delta\| > \|X - \delta\|\,] = 1. \tag{19}$$

Thus it is likely that James-Stein shrinkage will actually move us farther away from δ. I conclude, therefore, that we should simply tether the probe to Enterprise and hope that Delta can detect our present location X."

"Boy, Spock, you *are* a party pooper," Bones[16] said. "I sure hope we don't shrink toward Vulcan."

"Resistance is futile" said Seven-of-Nine[17].

"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 δ_0, (11) yields

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

while (18) yields

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

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."

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, δ_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 δ_0 are independent whereas in the latter, δ_0 is dependent on X.

[15] See Proposition 3 in Appendix 4.

[16] Dr. McCoy.

[17] Right again, but how could we leave her out?

However, Captain Kirk's dilemma[18] remains: to shrink or not to shrink? If, according to property (B), the shrinkage target δ_0 can be chosen arbitrarily and still reduce the MSE and PC, can choosing δ_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 δ, *Captain Kirk should not shrink* X. If the shrinkage target δ_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*[19], information, *shrinkage is likely to decrease the accuracy of estimation.* As Barnard[20] concluded, if the statistical estimation problem is truly invariant under translation then the best invariant estimator should be used, namely X itself.

Acknowledgement: 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 National University of Singapore.

[18] Captain Kirk is "exactly in the position of Buridan's ass," as described in Barnard's discussion of the non-invariant nature of the James-Stein estimator in Stein (1962, p.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.

[19] As represented, for example, by "data-dependent" priors.

[20] cf. Stein (1962, p.288).

Appendix 1: Directional and Spherical Symmetry; Verification of (4)

Definition 1. *$Y \in \mathbb{R}^p$ is* **directionally symmetric** *if $\vec{Y} \overset{d}{=} -\vec{Y}$, 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:* $Y \overset{d}{=} -Y$. 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) $\iff$ (b) $\implies$ (c) are straightforward. We will show that (c) $\implies$ (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

$$P[\, H] = P[\, H^c] = P[\, -H]. \tag{22}$$

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

$$P[\, H \cap H_0] = P[\, H^c \cap H_0^c] = P[\, (-H) \cap (-H_0)]. \tag{23}$$

It follows from Lemma 1 below that

$$Q[\, A \cap S_0] = Q[\, (-A) \cap (-S_0)] \tag{24}$$

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

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

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 $S_0 = H_0 \cap \mathcal{S}^p$. If*

$$Q[\,S \mid S_0] = Q[\,-S \mid -S_0] \tag{26}$$

for every hemisphere $S \subset \mathcal{S}^p$, then

$$Q[\,A \mid S_0] = Q[\,-A \mid -S_0] \tag{27}$$

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

$$P[\,H \mid H_0] = P[\,-H \mid -H_0] \tag{28}$$

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

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

$$S_0 = \{(y_1, \ldots, y_{p-1}, y_p) \mid \sum\nolimits_{i=1}^{p} y_i^2 = 1,\, y_p > 0\}, \tag{29}$$

and let π denote the stereographic projection[21] of S_0 onto its tangent hyperplane $L_0 \equiv \{(y_1, \ldots, y_{p-1}, 1)\}$. Then the relation

$$\pi(S \cap S_0) = K. \tag{30}$$

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

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

$$\tilde{Q}[\,A\,] = Q[\,-A \mid -S_0\,], \tag{31}$$

so (26) states that

$$Q[\,S \mid S_0] = \tilde{Q}[\,S\,] \tag{32}$$

for every Borel set $A \subseteq \mathcal{S}^p$. Let R and $\tilde{R}$ denote the probability measures induced on L_0 by $Q[\,\cdot \mid S_0]$ and $\tilde{Q}$, respectively, under the mapping π, i.e.

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

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

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

$$R[\,K\,] = Q[\,\pi^{-1}(K) \mid S_0] = Q[\,S \cap S_0 \mid S_0] \equiv Q[\,S \mid S_0], \tag{35}$$

$$\tilde{R}[\,K\,] = \tilde{Q}[\,\pi^{-1}(K)] = \tilde{Q}[\,S \cap S_0] \equiv Q[\,-(S \cap S_0) \mid -S_0] \tag{36}$$

$$= Q[\,-S \mid -S_0] \equiv \tilde{Q}[\,S\,], \tag{37}$$

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

[21] cf. Ambartzumian (1982, p.26), Watson (1983, p.23).

Definition 2. *$Y \in \mathbb{R}^p$ is* **spherically symmetric $\equiv$ orthogonally invariant** *if $Y \overset{d}{=} \Gamma Y$ for every orthogonal transformation Γ 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 $\vec{Y}$ is uniformly distributed on the unit sphere $\mathcal{S}_p$ and is independent of $\|Y\|$. We now use this fact to verify (4) by the following proposition, where δ, δ_0, X, Y, σ, ψ, and τ all depend on p.

Proposition 2. *Assume that*

(i) δ_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 $p \to \infty$.

Then (cf. (4)*)*

$$\lim_{p\to\infty} \Pr_\delta [\, \|X - \delta_0\| > \|\delta - \delta_0\| \,] \equiv \lim_{p\to\infty} \Pr_\delta [\, X \in B_1^c] = 1. \tag{38}$$

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 μ_p denote the uniform probability measure on $\mathcal{S}_p$. By (i) and (ii), $\Pr_\delta[\, X \in B_1 \mid \delta_0]$ depends on δ_0 only via $\|\delta_0 - \delta\|$ (the radius of B_1), and

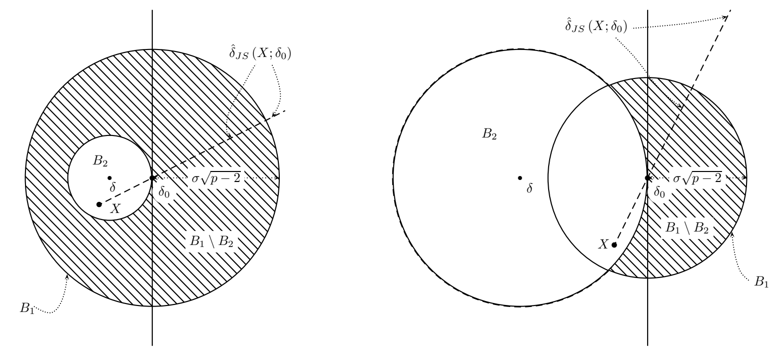
$$\begin{aligned} & \Pr_\delta[\, X \in B_1 \bigm| \|\delta_0 - \delta\| \,] & (39)\\ = \;& \Pr[\, Y \in B_1 - \delta \bigm| \|\delta_0 - \delta\| \,] \\ = \;& \mathrm{E}\left\{ \Pr\left[\vec{Y} \in \|Y\|^{-1}(B_1 - \delta) \bigm| \|Y\|,\ \|\delta_0 - \delta\| \right] \Bigm| \|\delta_0 - \delta\| \right\} \\ = \;& \mathrm{E}\left\{ \mu_p\left(\|Y\|^{-1}(B_1 - \delta)\right) \bigm| \|\delta_0 - \delta\| \right\}. & (40) \end{aligned}$$

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

$$\left\{ (z_1, \ldots, z_p) \Bigm| \frac{z_1}{(z_1^2 + \cdots + z_p^2)^{1/2}} \geq \frac{\|Y\|}{2\|\delta_0 - \delta\|} \right\} \tag{41}$$

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

$$\mu_p\left(\|Y\|^{-1}(B_1 - \delta)\right) = \frac{1}{2}\Pr\left[\frac{Z_1^2}{Z_1^2 + \cdots + Z_p^2} \geq \frac{\|Y\|^2}{2\|\delta_0 - \delta\|^2} \Bigm| \|Y\|,\ \|\delta_0 - \delta\| \right] \tag{42}$$



(a) $\|\delta_0 - \delta\| < \frac{\sigma}{2}\sqrt{p-2}$ (b) $\|\delta_0 - \delta\| > \frac{\sigma}{2}\sqrt{p-2}$

Figure 4: Illustrating the implication (50).

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

$$\Pr_\delta[\, X \in B_1 \bigm| \|\delta_0 - \delta\|\,] \leq \frac{1}{2}\Pr\left[\frac{Z_1^2}{Z_1^2 + \cdots + Z_p^2} \geq \frac{\|Y\|^2}{2\|\delta_0 - \delta\|^2} \biggm| \|\delta_0 - \delta\|\right], \quad (43)$$

hence

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

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

Appendix 2: 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

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

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

Appendix 3: Verification of (12)

First note that $\hat{\delta}_{JS}(X;\delta_0) \neq \hat{\delta}^+_{JS}(X;\delta_0)$ iff $\|X - \delta_0\| < \sigma\sqrt{p-2}$ (the ball B_1 of radius $\sigma\sqrt{p-2}$ centered at δ_0; see Figures 4a, 4b), in which case $\hat{\delta}^+_{JS}(X;\delta_0) = \delta_0$. Define

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

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

so (12) is equivalent to

$$\Pr[C \mid \delta_0] > \Pr[D \mid \delta_0]. \tag{48}$$

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

$$\Pr[C \cap B_1 \mid \delta_0] > \Pr[D \cap B_1 \mid \delta_0]. \tag{49}$$

But (see Figures 4a, 4b)

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

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

$$C \cap B_1 = (B_1 \setminus B_2) \supsetneq D \cap B_1, \tag{51}$$

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

Appendix 4: Verification of Figures 3a and 3b; Verification of (19)

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

$$\left\{\delta_0 \;\middle|\; \|\hat{\delta}^{+}_{JS}(X;\delta_0) - \delta\| < \|X - \delta\|\right\}. \tag{52}$$

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

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

If $\|\delta_0 - X\|^2 < \sigma^2(p-2)$, i.e., δ_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 δ_0 lies inside the ball B of radius $\|X - \delta\|$ centered at δ. If $\|\delta_0 - X\|^2 < \sigma^2(p-2)$, i.e., δ_0 lies outside this ball, then $\gamma = \left(1 - \frac{\sigma^2(p-2)}{\|X-\delta_0\|^2}\right)$ and the last inequality instead is equivalent to each of the following:

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

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

$$\lim_{p\to\infty} \Pr_\delta[\,\delta_0 \in B \cup K] = 0, \tag{53}$$

by the following proposition, in which δ, δ_0, X, V, σ, and τ now depend on p.

Proposition 3. *Assume that*

(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)*)*

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

The boundedness assumption (iii′) (resp., (iv′) is satisfied, e.g., 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 1 (with (i)-(iii) and X, Y, $B_1 - \delta$, and $\|\delta_0 - \delta\|$ replaced by (i′)-(iii′) and δ_0, V, $B - X$, and $\|X - \delta\|$), we obtain

$$\lim_{p\to\infty} \Pr_\delta[\,\delta_0 \in B\,] = \lim_{p\to\infty} \Pr_\delta[\,V \in B - X\,] = 0. \tag{55}$$

Next, again by the argument in Appendix 1 but with $B_1 - \delta$ replaced by $K - X$,

$$
\begin{aligned}
& \Pr_\delta[\,\delta_0 \in K \;\big|\; \|X-\delta\|\,] && (56)\\
= \;& \mathrm{E}\left\{\,\mu_p\left(\|V\|^{-1}(K - X)\right) \;\big|\; \|X-\delta\|\,\right\} && (57)\\
= \;& \frac{1}{2}\Pr\left[\frac{Z_1^2}{Z_1^2+\cdots+Z_p^2} \ge \frac{\sigma^4(p-2)^2}{4\|X-\delta\|^2\|V\|^2} \;\Big|\; \|\,X-\delta\|\right]. && (58)
\end{aligned}
$$

Thus by (iv′),

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

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

References

Ambartzumian, R. V. (1982). *Combinatorial Integral Geometry.* Wiley, New York.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis, 2nd Ed.* Wiley, New York.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* Wiley, New York.

Baranchik, A. (1964). Multiple regression and estimation of the mean of a multivariate normal distribution. Unpublished Ph. D. Thesis, Tech. Report 51, Department of Statistics, Stanford University.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd Ed.* Springer-Verlag, New York.

Billingsley, P. (1982). *Probability and Measure.* Wiley, New York.

Efron, B. (1975). Biased versus unbiased estimation. *Adv. in Math.* **16** 259-277.

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

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. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1** 311-319. University of California Press, Berkeley.

Lehmann, E. and Casella, G. (1998). *Theory of Point Estimation 2nd Ed.* Springer, New York.

Sen, P. K., Kubokawa, T., and Ehsanes Saleh, A. K. (1989). The Stein paradox in the sense of the Pitman measure of closeness. *Ann. Statist.* **17** 1375-1386.

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.

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.

Watson, G. S. (1983). *Statistics on Spheres.* Wiley, New York.