The Annals of Statistics
1996, Vol. 24, No. 2, 809–815

SHRINKAGE ESTIMATORS, SKOROKHOD'S PROBLEM AND STOCHASTIC INTEGRATION BY PARTS

BY STEVEN N. EVANS[1] AND PHILIP B. STARK[2]

University of California, Berkeley

For a broad class of error distributions that includes the spherically symmetric ones, we give a short proof that the usual estimator of the mean in a d-dimensional shift model is inadmissible under quadratic loss when $d \geq 3$. Our proof involves representing the error distribution as that of a stopped Brownian motion and using elementary stochastic analysis to obtain a generalization of an integration by parts lemma due to Stein in the Gaussian case.

Let X have a d-variate, spherically symmetric normal distribution with mean vector θ. Stein (1956) showed that, for $d \geq 3$, X is an inadmissible estimator of θ under quadratic loss. Stein observed that X is dominated by "shrinkage" estimators $(1 - a/(b + \|X\|^2))X$ for $a > 0$ sufficiently small and b sufficiently large. James and Stein (1961) showed that $0 < a < 2(d-2)$ and $b = 0$ suffice. Using the results of rather lengthy explicit calculations in Brandwein and Strawderman (1978), Brandwein (1979) extended James and Stein's (1961) result to d-variate shift models for $d \geq 4$ that have general spherically symmetric error distributions with finite second moments. (In the d-variate shift model, the distribution of the observation vector X is, under $\mathbf{P}_\theta$, the same as $Z + \theta$, where Z is a zero-mean random variable whose distribution does not depend on θ.) A somewhat shorter proof was given in Brandwein and Strawderman (1990) for the case $d \geq 5$. Meanwhile, the proof in the normal case had been simplified considerably by the "unbiased estimation of risk" technique, which depends on a simple integration by parts identity for the normal distribution [cf. Stein (1981)].

The following result, proved using elementary stochastic analysis to generalize Stein's lemma, establishes in a simple way that certain shrinkage estimators dominate X when $d \geq 3$ for a large class of error distributions, including, as a special case, spherically symmetric ones. The result substantially extends the class of distributions for which shrinkage estimators were known to dominate X. Our majorizing estimators are of the form $(1 - a/(1 + \|X\|^2))X$ for $a > 0$ sufficiently small, but, with extra moment assumptions, a similar proof works for estimators of the form $(1 - a/\|X\|^2)X$.

Received April 1995; revised October 1995.

[1]Research supported by a Presidential Young Investigator Award and an Alfred P. Sloan Foundation Fellowship.

[2]Research supported by a Presidential Young Investigator Award and grants from the NSF and NASA.

AMS 1991 *subject classifications*. Primary 62C15, 62F10.

Key words and phrases. Admissibility, balayage, Brownian motion, location parameter, quadratic loss, shift model.

Generalizations of Stein's lemma for spherically symmetric distributions have been established using classical calculus methods in Brandwein and Strawderman (1991) and Cellier and Fourdrinier (1995), but it appears that such methods cannot be extended to the class of error distributions considered here, which contains singular distributions supported on fractal sets of nonintegral dimension.

THEOREM. *Let Z be a random variable taking values in $\mathbf{R}^d$, $d \geq 3$. Suppose that Z is not almost surely* 0, $\mathbf{E}[Z] = 0$, $\mathbf{E}[\|Z\|^2] < \infty$ *and*

$$\mathbf{E}[\|Z+\theta\|^{2-d}] \leq \|\theta\|^{2-d}, \qquad \forall\, \theta \in \mathbf{R}^d. \tag{1}$$

(i) *For sufficiently small $a > 0$,*

$$\mathbf{E}[\|Z\|^2] > \mathbf{E}\left[\left\|\left(1 - \frac{a}{1+\|Z+\theta\|^2}\right)(Z+\theta) - \theta\right\|^2\right] \tag{2}$$

for all $\theta \in \mathbf{R}^d$.

(ii) *If the support of the distribution of Z is contained by the ball $\{z \in \mathbf{R}^d\colon \|z\| \leq A\}$, then any*

$$a \in \left(0, 2\frac{(d-2)}{d}\left(\frac{\alpha^*}{2+\alpha^*}\right)^6 \mathbf{E}[\|Z\|^2]\right)$$

satisfies (2), *where α^* is the unique positive root of*

$$(d-2)\alpha^6 A^2 (1 + (2+\alpha)^2 A^2)^2 - d(2+\alpha)^4 = 0.$$

(iii) *If the support of the distribution of Z does not intersect the ball $\{z \in \mathbf{R}^d\colon \|z\| \leq A\}$, then any*

$$a \in \left(0, 2\frac{(d-2)}{d} A^2\right)$$

satisfies (2).

PROOF. (i) Inequality (2) is equivalent to the inequality

$$a\mathbf{E}\left[\frac{\|Z+\theta\|^2}{(1+\|Z+\theta\|^2)^2}\right] < 2\mathbf{E}\left[\frac{Z\cdot(Z+\theta)}{1+\|Z+\theta\|^2}\right]$$

for all $\theta \in \mathbf{R}^d$.

Let B be a standard d-dimensional Brownian motion starting at $0 \in \mathbf{R}^d$. Recall that the potential kernel of B is given by

$$U(x, dy) = \Gamma(d/2 - 1)(2\pi)^{-d/2}\|y - x\|^{2-d}\, dy$$

[see, e.g., Proposition II.3.1 of Bass (1995)]. Rost's (1971) solution of Skorokhod's problem for transient Markov processes establishes that condition (1) is equivalent to the existence of a (possibly randomized) stopping time T such that the distribution of B_T is that of Z.

Put $S_r = \inf\{s \geq 0: \|B_s\| = r\}$, and let $T_n = T \wedge S_n \wedge n$. By bounded convergence,

$$\mathbf{E}\left[\frac{\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right] = \lim_{n\to\infty} \mathbf{E}\left[\frac{\|B_{T_n}+\theta\|^2}{(1+\|B_{T_n}+\theta\|^2)^2}\right]$$

and

$$\mathbf{E}\left[\frac{B_T\cdot(B_T+\theta)}{1+\|B_T+\theta\|^2}\right] = \lim_{n\to\infty} \mathbf{E}\left[\frac{B_{T_n}\cdot(B_{T_n}+\theta)}{1+\|B_{T_n}+\theta\|^2}\right].$$

For $y \in \mathbf{R}^d$, $\varepsilon \in \mathbf{R}$ and F: $\mathbf{R}^d \to \mathbf{R}$ any bounded Borel function, Girsanov's formula [e.g., Theorem IV.38.5 of Rogers and Williams (1987)] gives that

$$\mathbf{E}[F(B_{T_n}+\varepsilon y T_n)] = \mathbf{E}\big[\exp(\varepsilon y\cdot B_{T_n} - \tfrac{1}{2}\varepsilon^2\|y\|^2 T_n)F(B_{T_n})\big].$$

If F is bounded and continuous with bounded and continuous first-order partial derivatives, we can differentiate both sides of this equality with respect to ε at 0 to get

$$\mathbf{E}[y\cdot B_{T_n}F(B_{T_n})] = \mathbf{E}[y\cdot\nabla F(B_{T_n})T_n]. \tag{3}$$

Equation (3) generalizes Stein's lemma beyond the Gaussian case and is an instance of "stochastic integration by parts." See Section IV.41 of Rogers and Williams (1987) for a discussion and references to the relevant literature.

Applying (3) repeatedly, with $F(B_{T_n}) = [(B_{T_n})_i + \theta_i]/[1+\|B_{T_n}+\theta\|^2]$ and y the ith coordinate vector, then summing the results over i yields

$$\mathbf{E}\left[\frac{B_{T_n}\cdot(B_{T_n}+\theta)}{1+\|B_{T_n}+\theta\|^2}\right] = \mathbf{E}\left[T_n\frac{d+(d-2)\|B_{T_n}+\theta\|^2}{(1+\|B_{T_n}+\theta\|^2)^2}\right]$$

and so, by applying Fatou's lemma to the right-hand side,

$$\mathbf{E}\left[\frac{B_T\cdot(B_T+\theta)}{1+\|B_T+\theta\|^2}\right] \geq \mathbf{E}\left[T\frac{d+(d-2)\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right].$$

We therefore need to find $a > 0$ such that

$$\begin{aligned} a\mathbf{E}\left[\frac{\|Z+\theta\|^2}{(1+\|Z+\theta\|^2)^2}\right] &= a\mathbf{E}\left[\frac{\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right] \\ &< 2\mathbf{E}\left[T\left(\frac{d+(d-2)\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right)\right] \\ &\leq 2\mathbf{E}\left[\frac{B_T\cdot(B_T+\theta)}{1+\|B_T+\theta\|^2}\right] \\ &= 2\mathbf{E}\left[\frac{Z\cdot(Z+\theta)}{1+\|Z+\theta\|^2}\right] \end{aligned} \tag{4}$$

for all $\theta \in \mathbf{R}^d$.

By bounded convergence,

$$\theta \mapsto \mathbf{E}\left[\frac{\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right]$$

is continuous, and

$$\lim_{\theta\to\infty} \|\theta\|^2 \mathbf{E}\left[\frac{\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right] = 1.$$

By Fatou's lemma,

$$\theta \mapsto \mathbf{E}\left[T\frac{d+(d-2)\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right]$$

is lower semicontinuous. As this function is strictly positive at each point, it is bounded away from 0 on compacts. Again, by Fatou,

$$\liminf_{\theta\to\infty} \|\theta\|^2 \mathbf{E}\left[T\left(\frac{d+(d-2)\|B_T+\theta\|^2}{(1+\|B_T+\theta\|^2)^2}\right)\right] \geq (d-2)\mathbf{E}[T].$$

We may therefore find $a > 0$ such that (4) holds.

(ii) Fix $\alpha > 0$. Observe that, for $\|z\| \leq A$ and $\|\theta\| \geq (1+\alpha)A$, we have

$$\frac{\alpha}{1+\alpha}\|\theta\| \leq \|z+\theta\| \leq \frac{2+\alpha}{1+\alpha}\|\theta\|,$$

and hence, for $\|\theta\| \geq \alpha A$,

$$\begin{aligned}\inf_{\|z\|\leq A} & \frac{d+(d-2)\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2} \\ & \geq \frac{(d-2)(\alpha/(1+\alpha))^2\|\theta\|^2}{(1+((2+\alpha)/(1+\alpha))^2\|\theta\|^2)^2} \\ & \geq (d-2)\left(\frac{\alpha}{2+\alpha}\right)^6 \frac{((2+\alpha)/(1+\alpha))^2\|\theta\|^2}{(1+(\alpha/(1+\alpha))^2\|\theta\|^2)^2} \\ & \geq (d-2)\left(\frac{\alpha}{2+\alpha}\right)^6 \sup_{\|z\|\leq A} \frac{\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2}.\end{aligned}$$

On the other hand, for $\|z\| \leq A$ and $\|\theta\| < (1+\alpha)A$, we have $0 \leq \|z+\theta\| \leq (2+\alpha)A$ and so, for $\|\theta\| < (1-\alpha)A$,

$$\begin{aligned}\inf_{\|z\|\leq A} & \frac{d+(d-2)\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2} \\ & \geq \frac{d}{(1+(2+\alpha)^2A^2)^2} \\ & \geq \frac{d}{(2+\alpha)^2A^2(1+(2+\alpha)^2A^2)^2}(2+\alpha)^2A^2 \\ & \geq \frac{d}{(2+\alpha)^2A^2(1+(2+\alpha)^2A^2)^2} \sup_{\|z\|\leq A} \frac{\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2}.\end{aligned}$$

Therefore,

$$\inf_{\|z\|\le A} \frac{d+(d-2)\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2}$$
$$\ge \left((d-2)\left(\frac{\alpha}{2+\alpha}\right)^6 \wedge \frac{d}{(2+\alpha)^2A^2(1+(2+\alpha)^2A^2)^2}\right) \sup_{\|z\|\le A} \frac{\|z+\theta\|^2}{(1+\|z+\theta\|^2)^2}.$$

Note that the function

$$\alpha \mapsto (d-2)\left(\frac{\alpha}{2+\alpha}\right)^6 \wedge \frac{d}{(2+\alpha)^2A^2(1+(2+\alpha)^2A^2)^2}$$

is maximized at $\alpha = \alpha^*$ with the corresponding value being $(d-2)\times(\alpha^*/(2+\alpha^*))^6$.

It only remains to note that since $\{\|B_t\|^2 - td\}_{t\ge 0}$ is a martingale, it follows from the monotone convergence theorem and Fatou's lemma that

$$\mathbf{E}[T] = \lim_n \mathbf{E}[T_n] = \frac{1}{d}\lim_n \mathbf{E}[\|B_{T_n}\|^2] \ge \frac{1}{d}\mathbf{E}[\|B_T\|^2],$$

and so (4) holds for the stated values of a.

(iii) From Fitzsimmons (1991) [see also Heath (1974)], we may suppose that the stopping time T is of the form $\tau_{C(U)}$, where $\{C(u)\colon 0\le u\le 1\}$ is a decreasing family of finely closed sets such that $C(u)\subset\{z\in\mathbf{R}^d\colon \|z\|\ge A\}$, $\tau_{C(u)} = \inf\{t\ge 0\colon B_t\in C(u)\}$ and U is a random variable independent of B that is uniformly distributed on $[0,1]$. (A finely open set is one that a Brownian motion requires strictly positive time to exit, when started at any point in the set. A finely closed set is the complement of a finely open set.) From the strong Markov property and the rotational invariance of B, we have

$$\mathbf{E}[T|B_T] \ge \mathbf{E}[S_A|B_T] = \mathbf{E}[S_A] = \frac{A^2}{d},$$

and so (4) holds for the stated values of a. □

REMARK 1. Condition (1) certainly holds if the distribution of Z is spherically symmetric, since if σ_r is normalized surface measure on the sphere of radius r, then

$$\int \|x+\theta\|^{2-d}\,\sigma_r(dx) = r^{2-d} \wedge \|\theta\|^{2-d}$$

[see, e.g., equation II.3.6 of Bass (1995)]. However, spherically symmetric distributions are but few of the instances. As mentioned in the proof, condition (1) is equivalent to Z having the same distribution as B_T for some (possibly randomized) stopping time T, so Z could have the exit distribution of Brownian motion from a finely open domain containing 0, or mixtures of such distributions. It follows from Fitzsimmons (1991) [see also Heath (1974)] that the distribution of Z is necessarily of this form. Spherically symmetric distributions arise just when the domains are balls centered at 0.

REMARK 2. If $Z_1, \dots, Z_n$ are independent random variables, each of which satisfies condition (1), and $a_1, \dots, a_n \in \mathbf{R}$, then the linear combination $a_1 Z_1 + \cdots + a_n Z_n$ also satisfies condition (1), as a simple induction shows. This establishes that, in the d-variate shift model with repeated i.i.d. observations $\{X_j\}_{j=1}^n$ and error distribution satisfying (1), the sample mean as an estimator of the mean vector θ is dominated by estimators that "shrink" the sample mean appropriately. For many distributions, maximum-likelihood estimators do not make sense; for example, when the family of distributions generated as θ varies is not dominated. Similarly, minimum-risk equivariant estimators are often impossible to compute. Shrinking the sample mean is thus an easy way to improve on the "naive" sample mean in some cases where more sophisticated approaches fail to provide an "optimal" alternative. Of course, when we have only one observation X, that observation is the minimum-risk equivariant estimator of θ.

REMARK 3. If Z_1 and Z_2 respectively satisfy the conditions in parts (ii) and (iii) of the theorem and Z is a random variable whose distribution is a mixture of those of Z_1 and Z_2, then we may take a to be the minimum of any pair of values that "work" for Z_1 and Z_2, respectively. In particular, this approach yields an explicit value of a for any spherically symmetric distribution, although it does not lead to results as refined as those previously obtained for that special case. In the case that Z has the uniform distribution on the surface of a sphere, the bound given in part (iii) of the theorem asymptotically cannot be improved as $d \to \infty$.

REFERENCES

BASS, R. F. (1995). *Probabilistic Techniques in Analysis.* Springer, New York.

BRANDWEIN, A. C. (1979). Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multivariate Anal.* **9** 579–588.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1978). Minimax estimation of location parameters for spherically symmetric unimodal distributions. *Ann. Statist.* **6** 377–416.

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

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1991). Generalizations of James–Stein estimators under spherical symmetry. *Ann. Statist.* **19** 1639–1650.

CELLIER, D. and FOURDRINIER, D. (1995). Shrinkage estimators under spherical symmetry for the general linear model. *J. Multivariate Anal.* **52** 338–351.

FITZSIMMONS, P. J. (1991). Skorokhod embedding by randomized hitting times. In *Seminar on Stochastic Processes, 1990* (E. Çinlar, ed.) 183–192. Birkhäuser, Boston.

HEATH, D. (1974). Skorohod stopping via potential theory. *Séminaire de Probabilités VIII. Lecture Notes in Math.* **381** 150–154. Springer, New York.

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

ROGERS, L. C. G. and WILLIAMS, D. (1987). *Diffusions, Markov Processes, and Martingales* **2**. *Itô Calculus.* Wiley, New York.

ROST, H. (1971). The stopping distributions of a Markov process. *Invent. Math.* **14** 1–16.

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

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720-3860