The Annals of Statistics
1992, Vol. 20, No. 1, 351–369

ROBUST DIRECTION ESTIMATION

BY XUMING HE AND DOUGLAS G. SIMPSON[1]

National University of Singapore and University of Illinois

We relate various measures of the stability of estimates in general parametric families and consider their application to direction estimates on spheres. We show that constructions such as the SB-robustness of Ko and Guttorp and the information-standardized gross-error sensitivity of Hampel, Ronchetti, Rousseeuw and Stahel fit into a general framework in which one measures the effect of model contamination by the Kullback–Leibler discrepancy. We also define a breakdown point appropriate for a compact parameter space. Specific results concerning direction estimation include the optimal robustness of the circular median, the optimal breakdown point of the least median of squares on the sphere, the SB-robustness of certain scale-adjusted M-estimators and the SB-robustness in arbitrary dimensions of a class of estimators including the L_1-estimator and the hyperspherical median. The latter estimators avoid the need for simultaneous scale estimates, and they have breakdown points approaching 1/2 as the model becomes concentrated. A slight modification in their definition yields the same theoretical breakdown point as the least median of squares.

1. Introduction. The study of the stability of parameter estimates for directional data dates back at least to Watson (1967). Wehrly and Shine (1981) and Watson (1983) noted that the influence function of the normalized directional mean is bounded, which indicates that this estimate is robust to a certain extent. However, Lenth (1981) presented simulations indicating that for heavy-tailed distributions around the circle the mean estimate loses efficiency relative to, say, the directional median if most of the data are concentrated. Fisher (1985) discussed the use of median-type estimators on the sphere. Ko and Guttorp (1988) pointed out that the influence should be expressed relative to the concentration of the distribution. They introduced a notion of scale-standardized-bias (or SB) robustness to adjust for the concentration of the data on the sphere, and they showed that the directional mean estimate is not SB-robust with their scaling. Outliers in the circular and spherical data are treated in Barnett and Lewis (1984).

Ko and Guttorp (1988) relied on the construction of a scale-standardized gross-error sensitivity. With an appropriate choice of scaling, this is a special case of the information-standardized gross-error sensitivity of Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 229, which applies to general

Received August 1989; revised January 1991.

[1]Research partially supported by an NSF Mathematical Sciences Postdoctoral Research Fellowship and grant AFOSR-87-0041.

AMS 1980 *subject classifications*. Primary 62F35; secondary 62F12.

Key words and phrases. Breakdown function, breakdown point, directional mean, least median of squares, SB-robustness, spherical median, von Mises distribution.

parametric families. The literature contains a variety of other measures of stability such as the breakdown point and the sup-bias [see, e.g., Huber (1981)]. In another context, He, Simpson and Portnoy (1990) used a breakdown function to analyze the stability of tests. Certain connections among these quantities are known. In one-dimensional estimation, for instance, regularity conditions imply that the derivative of the sup-bias at 0, the bias sensitivity, coincides with the unstandardized gross-error sensitivity [Huber (1981), page 15, and He (1989)]. The breakdown point is the amount of contamination of the model such that the sup-bias becomes infinite. The purpose of this article is to establish further connections among the various global and local robustness measures and to fill in some of the gaps in the current state of knowledge about robust direction estimation.

We observe in Section 2 that the information-standardized gross-error sensitivity coincides, under sufficient regularity conditions, with a more general measure based on the Kullback–Leibler distance between fitted models. If one generalizes the definition of the gross-error sensitivity slightly, then it turns out that the gross-error sensitivity is always bounded by the bias sensitivity with respect to the same metric for the bias, for example, the Kullback–Leibler distance between parameter values. The bias sensitivity is in turn bounded by the reciprocal of the breakdown slope, the slope at 0 of the breakdown function with respect to the same metric. These results hold without regularity conditions on the class of estimators, and they suggest the following scheme for analyzing the local stability of an estimator. To establish the nonrobustness of an estimator in the local sense, it is sufficient to establish that it has an unbounded influence function. It then follows automatically that it has unbounded bias sensitivity and breakdown slope 0. Conversely, to establish the local robustness of an estimator, it is sufficient to establish a positive lower bound on its breakdown slope. Bounds on the bias sensitivity and gross-error sensitivity follow automatically. Replacing the usual Euclidean metric by the Kullback–Leibler discrepancy in these calculations provides an automatic standardization with respect to the concentration of the fitted model, which is locally equivalent to the standardization by Fisher information. See, for instance, Kass (1989) for further uses of the Kullback–Leibler discrepancy between parameters and the connection with Fisher information. A distinct advantage of the Kullback–Leibler discrepancy in the case of directional data is its parameterization invariance, which means that it automatically accounts for the constraints on the parameters. The breakdown function machinery described in Section 2 provides a way to define a breakdown point appropriate for compact parameter spaces such as the unit sphere.

The general discussion in Section 2 forms the basis for the more specific developments in subsequent sections on direction estimation. We consider rotationally symmetric models for directional data with axis of rotation μ and concentration parameter κ. The prototype is the von Mises distribution, which has a density on the unit sphere of the form

$$f(x;\mu,\kappa) = c_\kappa e^{\kappa x'\mu}, \qquad \kappa > 0,\ \mu \in S_p,\ x \in S_p. \tag{1.1}$$

Here the normalizing constant is $c_\kappa = \kappa^{(p-1)/2}/\{(2\pi)^{p/2}I_{(p-1)/2}(\kappa)\}$, where $I_\tau(\kappa)$ denotes the modified Bessel function of the first kind of order τ. Watson (1983) has provided extensive information about this model. We focus primarily on the stability of estimates of μ. Many of our results adapt readily to the more general class of models of the form $f_0(\kappa x'\mu)$ on the sphere, where $f_0(t)$ is decreasing on $[0, \infty)$.

Section 3 is concerned with the notion of standardized bias robustness for direction estimates. We suggest standardizing with respect to the Kullback–Leibler distance. An estimate is said to be SB-robust if its KL-standardized breakdown slope is bounded away from 0 uniformly in $\kappa \in (0, \infty)$. Hampel, Ronchetti, Rousseeuw and Stahel (1986), among others, discussed information-standardized gross-error sensitivity, and our suggestion is also an adaptation of that idea. A technical benefit of this standardization is that we can treat the dispersed case ($\kappa \to 0$) as well as the concentrated case ($\kappa \to \infty$). In other settings it can occur that no Fisher consistent estimate is SB-robust with respect to the information standardization [see He and Simpson (1990) for examples].

In Section 4 we obtain a sharp upper bound on the breakdown point of any estimator that is Fisher consistent for rotationally symmetric distributions. This bound implies that the breakdown point must go to 0 as the data become more dispersed over the sphere. On the other hand, it is possible to have a breakdown point near 1/2 if the data are concentrated. We give a sufficient condition for a direction estimate to achieve the bound on the breakdown point. The directional least median of squares is a leading example.

Section 5 focuses on the circular case and shows, for instance, that the circular median is the most robust among Fisher consistent estimators of the central direction. Section 6 allows arbitrary dimensions. We establish the SB-robustness of a class of concentration-adjusted M-estimators proposed by Lenth (1981), as well as another class of estimators that includes the hyperspherical median and L_1-estimator as special cases. The latter estimators have the advantage that there is no need for a simultaneous estimate of concentration. Indeed, they provide a robust direction on which one might project the data in order to obtain a robust estimate of the concentration. This automatic scale adjustment is analogous to the behavior of the median in a univariate location model. All the proofs are given in Section 7.

2. Generalities: Breakdown functions, bias and influence. First, consider a general parametric family of distributions $\{F_\mu, \mu \in \Theta\}$. In later sections μ will be the central direction vector on the circle or hypersphere. Suppose that $T(F)$ is a Fisher consistent estimating functional for μ, that is, $T(F_\mu) = \mu$ for any $\mu \in \Theta$. Starting at $F = F_{\mu_0}$, we ask how much contamination of F is needed to drive the estimate to $\mu \neq \mu_0$. Hence we define the breakdown function [cf. He, Simpson and Portnoy (1990)]:

$$\varepsilon_\mu^* = \inf\{\varepsilon > 0: T((1-\varepsilon)F + \varepsilon G) = \mu \text{ for some } G\}. \tag{2.1}$$

For fixed μ_0 and μ, larger values of ε_μ^* correspond to greater stability of the estimator T. The breakdown function (2.1) has the following invariance property with respect to reparameterizations: If g is a one-to-one transformation, then $\varepsilon^*_{g(\mu)}(g(T), F) = \varepsilon^*_\mu(T, F)$.

We can assess both the local stability and the global stability of an estimator via the breakdown function. One global measure is $\varepsilon^*_{\sup} = \sup_{\mu \in \Theta} \varepsilon^*_\mu$, the amount of contamination required to drive the estimate to any point in the parameter space, which might be taken as an invariant breakdown point. It can be an unduly optimistic measure of stability, however. Scale estimation provides an example. The standard deviation functional has $\varepsilon^*_{\sup} = 1$ because the estimator cannot be driven to 0. A better definition of the breakdown point is

$$\varepsilon^* = \sup_{\pi \in \Pi} \inf_{\mu \in \pi^c} \varepsilon^*_\mu, \tag{2.2}$$

where Π is the collection of all nontrivial compact subsets of Θ, and π^c is the complement of π in Θ. This definition is similar to that of Hampel [see Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 97]. Specializations of (2.2) to location and scale estimation agree with the usual definition of the breakdown point. (2.2) also applies to compact parameter spaces like the unit sphere in R^p. It is clear that $\varepsilon^* \le \varepsilon^*_{\sup}$ in general. Equality holds if Θ is compact and ε^*_μ is continuous, which is true of the direction estimates we shall consider. The inequality can be strict if Θ is not compact, as illustrated in Example 4.1 below.

The breakdown point (2.2) does not require us to specify a metric for the change in T. For intermediate values of μ, we make use of a distance measure $d(\mu_0, \mu)$ in order to interpret what it means to drive an estimate from μ_0 to μ and to discuss the local stability of T. On the sphere $d(\mu_0, \mu)$ might be the Euclidean distance, $\|\mu - \mu_0\| = \{2 - 2\mu'\mu_0)^{1/2}$, or the angular distance, $\cos^{-1}(\mu'\mu_0)$. An invariant measure is provided by the information distance

$$d_J(\mu_0, \mu) = \left\{(\mu - \mu_0)' J(\mu_0)(\mu - \mu_0)\right\}^{1/2}, \tag{2.3}$$

where $J(\mu_0)$ is the Fisher information matrix evaluated at μ_0. This distance is unaffected by full-rank differentiable and one-to-one transformations of the parameters, and it adapts to the concentration and shape of the model at μ_0. Rao (1973), page 331, called the information distance "a measure of the intrinsic accuracy of a distribution." Hampel, Ronchetti, Rousseeuw and Stahel (1986) made use of the corresponding norm in their definition of the information-standardized gross-error sensitivity, an invariant measure of local stability.

Although (2.3) is an appealing measure, it imposes regularity conditions on the model [see, e.g., Rao (1973), pages 329–331, or Ibragimov and Has'minskii (1981), pages 62–64]. Moreover, the standard definition of $J(\mu)$ fails to account for the constraint that μ is a unit vector in the von Mises model. Of course, we can avoid this difficulty by reparameterizing in polar coordinates as in Mardia (1975). However, we prefer to work with the more transparent (in higher dimensions) rectangular parameterization in which we can easily express the projection on the axial direction and the projection on the space orthogonal to the axis [cf. Watson (1983)].

A simpler way to define an invariant distance is to set $d(\mu_0, \mu)$ equal to $D(f_{\mu_0}, f_\mu)$, where D is a distance between densities. The idea is to view the fitted model as the functional estimate of interest. If a parameterization with full-rank Fisher information matrix exists, then certain choices of D are locally equivalent to the information distance in that parameterization [Rao (1973), page 332]. Examples are the scaled Kullback–Leibler discrepancy

$$d_{\mathrm{KL}}(\mu_0, \mu) = \left\{2\int f_{\mu_0} \log(f_{\mu_0}/f_\mu)\, d\lambda\right\}^{1/2} \tag{2.4}$$

and the Hellinger distance

$$d_{\mathrm{H}}(\mu_0, \mu) = \left\{\int \left(f_{\mu_0}^{1/2} - f_\mu^{1/2}\right)^2 d\lambda\right\}^{1/2}, \tag{2.5}$$

where λ is a dominating measure for $\{F_\mu,\ \mu \in \Theta\}$. The density-based distance takes care of the bookkeeping, so we need not transform to an unconstrained parameterization. Moreover, the density-based distance adapts to the concentration in an automatic way.

Given a distance measure d, we define the distance-based breakdown function

$$\varepsilon^*(\delta) = \inf\{\varepsilon_\mu^*: d(\mu_0, \mu) \ge \delta\} \tag{2.6}$$

for $\delta \in [0, \delta^*)$ with $\delta^* = \sup_{\mu \in \Theta} d(\mu_0, \mu)$. It is clear that $\varepsilon^*(\delta)$ is increasing in δ. A local summary measure, the breakdown slope, is given by

$$\beta^* = \lim_{\delta \downarrow 0} \varepsilon^*(\delta)/\delta.$$

The global summary

$$\varepsilon^* = \sup\{\varepsilon^*(\delta): \delta \in [0, \delta^*)\} \tag{2.7}$$

is the distance-based breakdown point of T. Through an appropriate choice of the distance, we can recover the breakdown point of (2.2).

PROPOSITION 2.1. *If* $\{\mu \in \Theta: d(\mu_0, \mu) \le \delta\}$ *is a compact subset of* Θ *for any* $\delta \in [0, \delta^*)$ *and increases to* Θ *as* δ *tends to* δ^*, *then* (2.7) *agrees with* (2.2).

For scale estimation, the choice $d(\sigma_0, \sigma) = |\sigma - \sigma_0|$ for $\Theta = (0, \infty)$ does not satisfy the condition of Proposition 2.1, but $d(\sigma_0, \sigma) = |\log(\sigma/\sigma_0)|$ does. The latter yields the usual breakdown point for scale estimators.

The breakdown function $\varepsilon^*(\delta)$ is closely related to the supremum bias function of the estimate T with contamination neighborhoods. Following Huber (1981), page 11, define the contamination bias of T with respect to a distance measure $d(\cdot,\cdot)$ by

$$b(\varepsilon) = \sup_G \{d(\mu_0, T((1-\delta)F + \delta G)) \colon \delta \le \varepsilon\}.$$

From (2.6) it follows that

$$\varepsilon^*(\delta) = \inf\{\varepsilon \colon b(\varepsilon) \ge \delta\}, \tag{2.8}$$

which implies that $b(\varepsilon^*(\delta)) \le \delta$. Hence, if $\varepsilon^*(\delta) > 0$ for some $\delta > 0$, then T has a nonzero breakdown point, and moreover, the change in T is bounded by δ if the amount of contamination is less than $\varepsilon^*(\delta)$.

The quantity $\gamma^* = b'(0)$, called the bias sensitivity, has been used as a measure of local stability; it indicates the effect that small contaminations can have on the estimator [see Donoho and Liu (1988) and He and Simpson (1990)]. If $b(\varepsilon)$ is continuous, then $b(\varepsilon^*(\delta)) = \delta$. Furthermore, if either $b(\varepsilon)$ or $\varepsilon^*(\delta)$ is differentiable at 0, then $\gamma^* = 1/\beta^*$, the reciprocal of the breakdown slope. In general, however, $b(\varepsilon)$ and $\varepsilon^*(\delta)$ might not be differentiable at 0. Nevertheless, we can measure the local stability of T via

$$\gamma^+ := \limsup_{\varepsilon \downarrow 0} b(\varepsilon)/\varepsilon \quad \text{and} \quad \beta^- := \liminf_{\delta \downarrow 0} \varepsilon^*(\delta)/\delta,$$

the upper sensitivity and lower breakdown slope respectively. They are related as follows.

LEMMA 2.1. (i) $\gamma^+ \le 1/\beta^-$. (ii) *If γ^* exists, then β^* exists, and $\beta^* = 1/\gamma^*$.*

The bias sensitivity and breakdown slope are connected with Hampel's (1974) gross-error sensitivity, which is defined in terms of the influence function. Suppose $T(F)$ is a vector-valued functional of the distribution F. Let

$$F_{\varepsilon,x} = (1-\varepsilon)F + \varepsilon\Delta_x, \qquad 0 \le \varepsilon \le 1,$$

where Δ_x is the distribution of a point mass at x. The influence function (IF) [Hampel (1974)] is the directional derivative

$$\mathrm{IF}(x; T, F) = \lim_{\varepsilon \downarrow 0} \{T(F_{\varepsilon,x}) - T(F)\}/\varepsilon.$$

The existence of an influence function for T is a regularity condition; by itself it does not imply robustness of the estimator. T is said to be B-robust at F if the gross-error sensitivity

$$\gamma^*_{\mathrm{GE}} = \sup_x \|\mathrm{IF}(x; T, F)\| \tag{2.9}$$

is finite [Hampel, Ronchetti, Rousseeuw and Stahel (1986)]. Replacing the Euclidean norm in (2.9) by the information distance (2.3) yields the information-standardized gross-error sensitivity discussed by Hampel, Ronchetti, Rousseeuw and Stahel (1986), page 229.

In fact, one can define gross-error sensitivity with respect to any distance measure without requiring an influence function by setting

$$\gamma_{\mathrm{GE}}^{+} = \sup_{x} \limsup_{\varepsilon \downarrow 0} d(T(F), T(F_{\varepsilon,x}))/\varepsilon. \tag{2.10}$$

Under sufficient regularity conditions, $\gamma_{\mathrm{GE}}^{*} = \gamma^{*}$ [see, e.g., Huber (1981), page 15]. Here we report the simple fact that if we work with upper sensitivities, the gross-error sensitivity is always bounded by the bias sensitivity.

LEMMA 2.2. $\gamma_{\mathrm{GE}}^{+} \le \gamma^{+}$.

Taken together Lemmas 2.1 and 2.2 imply $\gamma_{\mathrm{GE}}^{+} \le \gamma^{+} \le 1/\beta^{-}$. Hence, if a particular estimator has an unbounded influence function with respect to $d(\cdot, \cdot)$, then it also has infinite bias sensitivity and breakdown slope 0. On the other hand, if the lower breakdown slope is positive, then the bias and gross-error sensitivities are bounded.

3. Stability measures for direction estimates. Now consider the estimation of μ, the central direction of the von Mises distribution on S_p. We study the stability of the estimates with respect to the Kullback–Leibler (KL) discrepancy. As

$$\log\{f(x;\mu_0,\kappa)/f(x;\mu,\kappa)\} = \kappa(\mu_0 - \mu)'x,$$

it follows from (2.4) that

$$d_{\mathrm{KL}}^{2}(\mu_0, \mu) = 2\kappa A(\kappa)(\mu_0 - \mu)'\mu_0 = \kappa A(\kappa)\|\mu - \mu_0\|^2, \tag{3.1}$$

where $A(\kappa) = -d\log(c_\kappa)/d\kappa = I_{p/2+1}(\kappa)/I_{p/2-1}(\kappa)$. The KL discrepancy simply rescales the Euclidean distance to adapt to the concentration about μ_0. Watson (1983) summarized various properties of $A(\kappa)$ including the limits, $\lim_{\kappa\to\infty} A(\kappa) = 1$ and $\lim_{\kappa\to 0}\kappa/A(\kappa) = p$.

In the sequel let $\varepsilon_{\mathrm{KL}}^{*}(\cdot)$ be the breakdown function with respect to the KL discrepancy, and let $\varepsilon^{*}(\cdot)$ be the breakdown function with respect to Euclidean distance on S_p. Using (2.6) and (3.1),

$$\varepsilon_{\mathrm{KL}}^{*}\big(\delta\sqrt{\kappa A(\kappa)}\big) = \varepsilon^{*}(\delta), \qquad 0 \le \delta \le 2.$$

The global analysis (breakdown point) using the KL discrepancy is equivalent to that using Euclidean distance, which also satisfies the conditions of Proposition 2.1. Section 4 pursues this further. To measure local stability, we use the KL breakdown slope, which is a rescaling of the Euclidean slope

$$\beta_{\mathrm{KL}}^{-} = \liminf_{\delta\downarrow 0} \frac{\varepsilon_{\mathrm{KL}}^{*}\big(\delta\sqrt{\kappa A(\kappa)}\big)}{\delta\sqrt{\kappa A(\kappa)}} = \liminf_{\delta\downarrow 0} \frac{\varepsilon^{*}(\delta)}{\delta\sqrt{\kappa A(\kappa)}} = \frac{\beta^{-}}{\sqrt{\kappa A(\kappa)}}.$$

Adapting terminology of Ko and Guttorp (1988), we call a direction estimator

SB-robust if

$$\inf_{0<\kappa<\infty} \beta^-_{\mathrm{KL}} > 0, \tag{3.2}$$

which also implies that the KL bias sensitivity and gross-error sensitivity are bounded uniformly in κ. In Sections 5 and 6 we show that the directional median and certain other estimators satisfy (3.2). This entails showing

$$\liminf_{\kappa\to\infty} \beta^-/\kappa^{1/2} > 0 \tag{3.3}$$

and

$$\liminf_{\kappa\to 0} \beta^-/\kappa > 0 \tag{3.4}$$

by the discussion following (3.1). An easy consequence of (4.1) below is that the directional mean has $\beta^* = A(\kappa)$ which fails to satisfy (3.3), so it is not SB-robust. This fact also follows from an influence calculation of Ko and Guttorp (1988) via Lemmas 2.1 and 2.2.

Wehrly and Shine (1981) observed that the unstandardized influence function of the directional mean is bounded. On the other hand, Ko and Guttorp (1988) argued that the influence function should be standardized by a dispersion measure, because a given amount of change in the direction estimate is deemed more serious if the data are more concentrated. They defined the standardized influence function (SIF),

$$\mathrm{SIF}(x; T, F, S) = \mathrm{IF}(x; T, F)/S(F),$$

where $S(F)$ is a measure of dispersion on S_p, and they called T an SB-robust estimator over a family of distributions F if

$$\sup_F \sup_x \| \mathrm{SIF}(x; T, F, S) \| < \infty. \tag{3.5}$$

They chose S to standardize the influence as the distribution becomes concentrated and restricted F to a class $\{F\colon S(F) > s > 0\}$, ruling out $\kappa \to 0$.

The standardized influence function has an appealing nonparametric flavor, but the results depend on the choice of $S(F)$. However, we can fit SIF into the present framework, because, with an appropriate choice of S, (3.2) implies that the information-standardized gross-error sensitivity is bounded in κ, allowing also $\kappa \to 0$. Essentially the same local criterion arises in an automatic way from the KL sensitivity; if T has an influence function, then its KL gross-error sensitivity is $\gamma^*_{\mathrm{KL}} = \sup_x \sqrt{\kappa A(\kappa)}\, \|\mathrm{IF}(x; T, F)\|$.

Krasker and Welsch (1982), Ruppert (1985), Hampel, Ronchetti, Rousseeuw and Stahel (1986) and others have discussed standardization of the influence function by the asymptotic covariance of the estimate itself. The resulting self-standardized sensitivity favors less efficient estimators, however, so we prefer the information standardization, more generally the KL sensitivity, so that the bias is expressed relative to the best asymptotic standard deviation of regular estimators.

4. Global stability of direction estimates. For the analysis of global stability, we now consider the breakdown function (2.6) with respect to the

Euclidean distance $d = \|\mu - \mu_0\|$. The breakdown point ε^* in this case is equal to ε^*_μ evaluated at $\mu = -\mu_0$, the breakdown for a direction reversal.

As a first example, we give the breakdown function of the directional mean estimator, which rephrases Theorem 3 of Ko and Guttorp (1988).

THEOREM 4.1. *The breakdown function of the directional mean estimator is*

$$\varepsilon^*(d) = 1 - \left\{1 + \rho\left(1 - \{\max(0, 1 - d^2/2)\}^2\right)^{1/2}\right\}^{-1}, \tag{4.1}$$

where $\rho = A(\kappa)$ *is the length of the mean vector.*

This result implies that it takes the same amount of contamination to drive the estimate to the opposite direction of μ_0 as to any direction μ more than 90 degrees apart from μ_0. It also implies that $\varepsilon^* = \rho/(1 + \rho)$ decreases from $1/2$ to 0 as ρ ranges from 1 to 0, that is, the estimator is easier to break down if the data are less concentrated.

EXAMPLE 4.1. The maximum likelihood estimator (MLE) of the concentration parameter κ has a one-to-one correspondence with the length of the resultant vector (sample mean vector). Because of the invariance property of the breakdown function, the breakdown point of the concentration estimate can be computed from the breakdown function of $\|EX\|$, the length of the resultant vector. The parameter space Θ corresponding to it is an open interval $(0, 1)$. Since $\|(1 - \varepsilon)E_F X + \varepsilon E_G X\| = |(1 - \varepsilon)\rho + b\varepsilon|$, by choosing G to be such that $E_G X = b\rho^{-1}E_F X$ for $b \in [-1, 1]$, it can be driven to 0 with $\varepsilon = \rho/(1 + \rho)$. So its breakdown point is $\varepsilon^* = \rho/(1 + \rho)$. In contrast, one cannot obtain a unit resultant vector using any fraction of contamination less than one for a nondegenerated von Mises distribution, so the supremum of the breakdown function $\varepsilon^*_{\sup} = 1$.

For linear data, it is well known that the breakdown point of any equivariant location estimator is bounded above by $1/2$. What about the direction estimators? This question is partially answered in Theorem 4.2.

Let $Q(\mu)$ be the set of all axially symmetric distributions about the axis μ whose density at x depends only on the value of $x'\mu$, and with $P\{x'\mu > c\} \geq P\{x'\mu < -c\}$ for all $c > 0$. It can be shown that for any distribution in $Q(\mu)$, μ is the mean direction.

It sometimes occurs that an estimate is not uniquely defined for certain distributions. In that case, a selection rule or combination rule is needed for the multiple solutions in order to uniquely define the functional T. For estimators that are equivariant under reflection about an axis, one naturally requires that if F is axially symmetric, then $T(F)$ must lie on this axis of symmetry. Therefore, it is reasonable to require that a direction estimator be Fisher consistent for $\mathbf{Q} = \{F \in Q(\mu),\ \mu \in \Theta\}$, that is, $T(F) = \mu$ for any $F \in Q(\mu)$ and $\mu \in \Theta$.

THEOREM 4.2. *If a direction estimate is Fisher consistent for* $\mathbf{Q}$, *and if the underlying distribution* $F \in Q(\mu_0)$, *then the breakdown point of the estimate is bounded from above by*

$$\bar{\varepsilon} = 1 - (2q_0)^{-1}, \quad \text{where } q_0 = P\{x: x'\mu_0 \geq 0\}. \tag{4.2}$$

We show how to achieve the upper bound on the breakdown point. Define $\hat{\mu}_c$ to be the maximizer of $P(V_c(\mu))$, where $V_c(\mu) = \{x: x'\mu \geq c\}$ for $|c| < 1$. Then $\hat{\mu}_c$ is Fisher consistent for $\mathbf{Q}$. For symmetric and unimodal distributions, it is easy to show that its breakdown point for directional reversal is

$$\varepsilon^* \geq 1 - \{1 + P\{x'\mu_0 \geq c\} - P\{x'\mu_0 \leq -c\}\}^{-1}.$$

If $c = 0$, then $\varepsilon^* = 1 - (2q_0)^{-1}$ which achieves the upper bound $\bar{\varepsilon}$. In fact, we have the following theorem.

THEOREM 4.3. *If an estimate* $T(F)$ *is such that* $P_F\{X'T(F) > 0\} \geq 1/2$ *for any* F, *then its breakdown point achieves the upper bound* (4.2).

However, the estimator $\hat{\mu}_0$ as defined above is very unstable in finite samples. It can mainly be explained by its poor local breakdown property; it is not B-robust. This is one example where the breakdown point alone does not capture the robustness of an estimator. The directional least median of squares (LMS) [e.g., the midpoint of the shortest arc containing at least half of the data points on the circle; see Rousseeuw (1984)] differs from $\hat{\mu}_c$ in that it adjusts the "tuning constant" c according to the data; it can be written as $\arg\min_\mu c(\mu)$, where $c(\mu) = \inf\{c > 0: P\{V_c(\mu)\} \geq 1/2\}$. The LMS outperforms $\hat{\mu}_c$ in finite sample examples, and also attains the upper bound $\bar{\varepsilon}$ for its breakdown point, as it satisfies the condition of Theorem 4.3.

5. SB-robust estimators on the circle. In this section, we focus on the circular data on the unit circle. We take angular observations $\theta_1, \theta_2, \dots, \theta_n$ instead of unit vectors. We shall briefly consider the M- and L-estimators of the mean direction for the von Mises distribution with the probability density function

$$f(\theta; \kappa, \theta_0) = c_\kappa e^{\kappa \cos(\theta - \theta_0)}, \qquad \theta \in (\theta_0 - \pi, \theta_0 + \pi). \tag{5.1}$$

We restrict ourselves to the estimators which are equivariant under rotation on the circle. We always take the true mean $\theta_0 = 0$, unless specified otherwise.

An M-estimator of the mean direction can be defined as the root to

$$\sum_{i=1}^{n} \psi(\theta_i - T) = 0, \tag{5.2}$$

where ψ satisfies the following two conditions:

(A1) ψ is odd, bounded and piecewise differentiable with $\psi(0) = 0$.
(A2) 2π is the smallest period for $|\psi(t)|$.

The estimate is clearly Fisher consistent and can be shown to be asymptotically normal [see Lenth (1981)]. It is also easy to show that the breakdown function ε_δ^*, the smallest fraction of contamination needed to drive the estimator to δ, is given by

$$\varepsilon_\delta^* = \frac{|\int\psi(\theta-\delta)\,dF(\theta)|}{\sup|\psi| + |\int\psi(\theta-\delta)\,dF(\theta)|} \tag{5.3}$$

and the breakdown slope is $\beta^* = E_F\psi'(\theta)/\sup|\psi|$. These results are valid for all distributions that are symmetric and unimodal at mean 0. If the density $f(\theta)$ is differentiable, then the differentiability of ψ can be removed, and $\beta^* = -\int\psi(\theta)f'(\theta)\,d\theta/\sup|\psi|$.

REMARKS. (1) Condition (A2) excludes the angular mean estimate whose breakdown function is given by Theorem 4.1. (2) The breakdown slope of the M-estimators is maximized at $\psi(t) = \operatorname{sign}(t)$, which corresponds to the directional median. The directional median has the breakdown slope $\beta^* = 2(f(0) - f(\pi))$, which is in the order of $\kappa^{1/2}$ as $\kappa \to \infty$ and κ as $\kappa \to 0$ for the von Mises distributions, so it is SB-robust.

The M-estimators with smooth score functions satisfy $\lim_{\kappa\to\infty}\kappa^{-1/2}\beta^* = 0$, so they are not SB-robust. But with a proper scaling, they can be SB-robust. For example, Lenth (1981) considered the score function $\psi_\kappa(\theta) = \psi_1(t(\theta,\kappa))\sin\theta/t(\theta,\kappa)$, where $t(\theta,\kappa) = \{2\kappa(1-\cos\theta)\}^{1/2}$ and ψ_1 can be any monotone score function satisfying (A1). It follows as a special case of Theorem 6.1 that this estimator is SB-robust.

Between the mean and median, one may naturally ask how the trimmed mean behaves. The α-trimmed mean averages over the $[(1-2\alpha)n]$ points between the αth and $(1-\alpha)$th quantiles [see Mardia (1972), page 33]. We show that the symmetrically trimmed mean on the circle is SB-robust for any $\alpha > 0$. Intuitively, the trimmed mean automatically scales itself to the concentration of the data.

THEOREM 5.1. *Suppose that F is unimodal and symmetric about $\theta = 0$ on $(-\pi,\pi)$. The breakdown slope (with respect to θ) of the α-trimmed mean estimate is given by*

$$\beta^* = (f(0) - f(\pi))(f(0)\sin\theta_\alpha)^{-1}\int_{-\theta_\alpha}^{\theta_\alpha}\cos(\theta)\,dF(\theta), \tag{5.4}$$

where θ_α is determined by $\int_0^{\theta_\alpha} dF(\theta) = 1/2 - \alpha$. Moreover,

$$\lim_{\kappa\to\infty}\kappa^{-1/2}\beta^* = (1-2\alpha)/\Phi^{-1}(1-\alpha) > 0 \quad \text{and} \quad \lim_{\kappa\to 0}\kappa^{-1}\beta^* = (2/\pi)^{1/2} \tag{5.5}$$

with $\Phi^{-1}(1-\alpha)$ denoting the αth upper quantile of the standard normal distribution, so the estimate is SB-robust.

Analogous to the median for the linear data, the directional median turns out to be the most SB-robust estimator for the circular mean.

THEOREM 5.2. *Suppose that the density function $f(\theta)$ on $(0, 2\pi)$ of the underlying distribution is symmetric and unimodal about θ_0, then among all estimators that are Fisher consistent for distributions that are symmetric and unimodal about θ_0, the directional median achieves the highest breakdown slope. It also attains the breakdown point bound* (4.2).

6. SB-robust estimators on hyperspheres.

We have seen SB-robust estimators for the circular data in Section 5. In this section we first show that the Lenth M-estimators are SB-robust in any dimension. We also give a class of SB-robust M-estimators that does not rely on the concentration parameter κ explicitly and has a built-in scaling. This class includes not only the directional L_1-estimator but also the spherical median [Fisher (1985)]. In the case of $p = 2$, Ko and Guttorp (1988) established that the L_1-estimate (circular median) is SB-robust in the sense of the standardized influence function, but they did not handle the higher-dimensional cases. We establish the stronger property of SB-robustness in the sense of the breakdown slope, and allow $\kappa \to 0$ as well as $\kappa \to \infty$. One way to combine the SB-robustness with high breakdown point of the direction estimates is provided at the end.

Generalizing Lenth (1981), an M-estimator of the direction is taken to be a solution to

$$\sum_{i=1}^{n} \rho\left(\kappa^{1/2}\|x_i - \mu\|\right) = \text{minimum}, \tag{6.1}$$

where ρ is so chosen to provide an estimator with the desired robustness property. We shall call such an estimator a Lenth M-estimator, in contrast to a more general formulation of M-estimators which in functional form minimizes $E\eta(\|x_i - \mu\|; \kappa)$ for a two-variate η function.

THEOREM 6.1. *Assume that $\rho(r)$ is increasing and differentiable on $(0, \infty)$ with a bounded derivative $\rho'(r) = \psi(r)$, then the Lenth M-estimate is SB-robust at the von Mises distributions.*

As argued in Lenth (1981), the factor $\kappa^{1/2}$ provides the right standardization in the estimate. However, the parameter κ is usually unknown in practice and needs to be estimated separately or simultaneously. Some suggestions made in Lenth (1981) can be directly generalized to higher dimensions.

It is possible, however, to obtain SB-robust estimators of μ_0 without knowledge of κ, nor even estimating it. One example is the popular L_1-estimator by taking $\rho(r) = r$ in (6.1). The following theorem shows that estimators that minimize a criterion locally equivalent to the L_1 are SB-robust. One might use these as preliminary estimates of the direction on which to project the data to get a robust estimate of κ.

THEOREM 6.2. *Any estimator that minimizes $E\eta(\|X - \mu\|)$ over $\mu \in S_P$ is SB-robust at the von Mises distributions if η satisfies the following conditions*:

(B1) $\eta(r)$ *is increasing on* $(0, 2)$.
(B2) $|\eta(\|x - y\|) - \eta(\|x - z\|)| \leq \eta(\|y - z\|)$ *for any* $x, y, z \in S_P$, *or*
(B2′) η *is differentiable with bounded derivative on* $(0, 2)$.
(B3) $\lim_{r \to 0} \eta(r)/r = b$ *for some constant* $b > 0$.

(B2) requires that $\eta(\|x - y\|)$ be a metric on S_P. (B3) simply specifies that the η function must be locally equivalent to $\eta(r) = r$ which gives the well-known L_1-estimator. The hyperspherical median [see Fisher (1985)] which minimizes $E\cos^{-1}(X'\mu)$ is an example of an SB-robust estimator by Theorem 6.2, as $\eta(r) = \cos^{-1}(1 - r^2/2)$ satisfies (B1), (B2) and (B3).

As κ becomes larger, the distribution concentrates more on the true direction μ_0; therefore, it is intuitively clear that it is the local behavior of the η function that matters most. The proof of Theorem 6.2 is given in the Section 7. The proof of Theorem 6.1 is completely parallel.

Now, what about the global breakdown point for these estimators? Take the hyperspherical median for example. It can be shown that its breakdown point is

$$\varepsilon^* \geq 1 - \tfrac{1}{2}\pi\left\{\pi - E\left[\cos^{-1}(x'\mu_0)\right]\right\}^{-1}.$$

This lower bound tends to $1/2$ as $\kappa \to \infty$. As $\kappa \to 0$, it expands to $4\kappa/\pi^2 + o(\kappa)$, compared to $4\kappa + o(\kappa)$ for the upper bound $\bar{\varepsilon}$ of (4.2). For the L_1-estimate, one can show that it has breakdown point approaching $1/2$ as $\kappa \to \infty$, but a suboptimal breakdown point for finite κ. A minor modification of either of these estimators can yield the optimal breakdown point.

Let $M = \{\mu: P(X' > 0) \geq 1/2\}$. Basically, M is the half-space that contains favorable directions. Define $\mu(F)$ to be the minimizer of $E_F\eta(\|X - \mu\|)$ among $\{\mu \in M\}$, where η satisfies the conditions of Theorem 6.2. Then it has the breakdown point $\varepsilon^* = \bar{\varepsilon}$, but retains the SB-robustness established by Theorem 6.2.

7. Proofs.

PROOF OF PROPOSITION 2.1. We write $\varepsilon^*(\delta^*)$ for (2.7). For any $\pi \in \Pi$, there exists $\delta_\pi \in (0, \delta^*)$ such that $\pi \subset \{\mu: d(\mu, \mu_0) \leq \delta_\pi\}$. Thus

$$\inf_{\mu \in \pi^c} \varepsilon_\mu^* \leq \varepsilon^*(\delta_\pi) \leq \varepsilon^*(\delta^*).$$

On the other hand, $\varepsilon^*(\delta) \leq \inf\{\varepsilon_\mu^*: d(\mu_0, \mu) > \delta\}$, and $\{\mu: d(\mu_0, \mu) \leq \delta\}$ is an element of Π, so $\varepsilon^*(\delta) \leq \sup_{\pi \in \Pi} \inf_{\mu \in \pi^c}\{\varepsilon_\mu^*\}$ for every $\delta \in (0, \delta^*)$, which completes the proof. □

PROOF OF LEMMA 2.1. (i) It is sufficient to show that $\gamma^+ > M \geq 0$ implies $\beta^- \leq M^{-1}$. The case $M = 0$ is trivial. Suppose $\gamma^+ > M > 0$. Then there is a sequence $\{\varepsilon_\nu\}$ with $\varepsilon_\nu \downarrow 0$ such that $b(\varepsilon_\nu) \geq M\varepsilon_\nu$ for ν sufficiently large. Hence, by (2.8), and for ν sufficiently large,

$$\varepsilon^*(M\varepsilon_\nu) = \inf\{\varepsilon: b(\varepsilon) \geq M\varepsilon_\nu\} \leq \varepsilon_\nu.$$

As $\varepsilon_\nu \downarrow 0$,

$$\liminf_{\delta \downarrow 0} \frac{\varepsilon^*(\delta)}{\delta} \leq \liminf_{\nu \to \infty} \frac{\varepsilon^*(M\varepsilon_\nu)}{M\varepsilon_\nu} \leq \frac{1}{M}.$$

(ii) If γ^* is infinite, it is clear that β^* has to be 0 by definition. Now suppose that $\gamma^* < \infty$. Observe that $\varepsilon^*(\delta)/\delta = \inf\{a: b(\delta a) \geq \delta\}$. Since $b(\varepsilon)$ has derivative at 0, we have for any $c > 0$ and $z \in (0, (1-c)/\gamma^*)$,

$$b(\delta z) = (1-c)\delta + o(\delta) < \delta$$

for sufficiently small δ. Hence $\liminf_{\delta \downarrow 0} \varepsilon^*(\delta)/\delta \geq (1-c)/\gamma^*$. Letting $c \downarrow 0$, we obtain $\beta^- \geq 1/\gamma^*$.

On the other hand, for $z = (1+c)/\gamma^*$ and $c > 0$,

$$b(\delta z) = (1+c)\delta + o(\delta) \geq \delta$$

for sufficiently small δ. Therefore, $\limsup_{\delta \downarrow 0} \varepsilon^*(\delta)/\delta \leq (1+c)/\gamma^*$ for any $c > 0$. So $\limsup_{\delta \downarrow 0} \varepsilon^*(\delta)/\delta \leq 1/\gamma^*$.

Putting things together, we get $\gamma^* = 1/\beta^*$. □

PROOF OF LEMMA 2.2. Let $T_\varepsilon(G) = T((1-\varepsilon)F + \varepsilon G)$. Then

$$\sup_x \limsup_{\varepsilon \downarrow 0} d(T(F), T_\varepsilon(\Delta_x))/\varepsilon \leq \limsup_{\varepsilon \downarrow 0} \sup_x d(T(F), T_\varepsilon(\Delta_x))/\varepsilon$$

$$\leq \limsup_{\varepsilon \downarrow 0} \sup_G d(T(F), T_\varepsilon(G))/\varepsilon. \qquad \square$$

PROOF OF THEOREM 4.2. Consider $G = (1-\varepsilon)F + \varepsilon H$, where H puts point mass at the opposite direction of μ_0. As $\varepsilon > \bar{\varepsilon}$, $G \in Q(-\mu_0)$. By Fisher consistency, $T(G) = -\mu_0$, a direction reversal occurs. So $\varepsilon^* \leq \bar{\varepsilon}$. □

PROOF OF THEOREM 4.3. A direction reversal requires

$$(1-\varepsilon)P\{X'(-\mu_0) > 0\} + \varepsilon \geq 1/2 \quad \text{or} \quad \varepsilon \geq 1 - (2q_0)^{-1}. \qquad \square$$

PROOF OF THEOREM 5.1. We compute the breakdown slope from the local contamination bias function. Let $H = (1-\varepsilon)F + \varepsilon G$ and θ_0 be the median of H. Also let θ_1 and θ_2 be the lower and upper trimming points for computing the α-trimmed mean estimate. Their dependence on ε and α is suppressed. We shall only consider the case with $\theta_0 > 0$.

The least favorable contaminant G should have mass on $(\theta_0, \theta_0 + \pi)$ to drive the median θ_0 as far from 0 as possible. Furthermore, it should have no mass in (θ_0, θ_2) to make the trimmed mean the largest. So we can choose G to have point mass at θ_2, as moving all mass in $(\theta_2, \theta_0 + \pi)$ to the point θ_2 will have no effect on the trimmed mean. The sup-bias is attained at such contamination distribution G with θ_i's determined by

$$(1 - \varepsilon)\int_{\theta_0}^{\theta_0 + \pi} dF(\theta) + \varepsilon = \tfrac{1}{2},$$

$$(1 - \varepsilon)\int_{\theta_1}^{\theta_0} dF(\theta) = \tfrac{1}{2} - \alpha$$

and

$$(1 - \varepsilon)\int_{\theta_0}^{\theta_2} dF(\theta) = \tfrac{1}{2} - \alpha.$$

Direct calculation gives

$$\lim_{\varepsilon \downarrow 0} \theta_0/\varepsilon = \{2(f(0) - f(\pi))\}^{-1},$$

$$\lim_{\varepsilon \downarrow 0} (\theta_1 + \theta_\alpha)/\varepsilon = (f(\theta_\alpha))^{-1}\left[\tfrac{1}{2}f(0)/(f(0) - f(\pi)) - \tfrac{1}{2} + \alpha\right]$$

and

$$\lim_{\varepsilon \downarrow 0} (\theta_2 - \theta_\alpha)/\varepsilon = (f(\theta_\alpha))^{-1}\left[\tfrac{1}{2}f(0)/(f(0) - f(\pi)) + \tfrac{1}{2} - \alpha\right].$$

The sub-bias $b(\varepsilon)$ is then given by

$$b(\varepsilon) = \tan^{-1} \frac{\int_D \sin\theta \, dF(\theta)}{\int_D \cos\theta \, dF(\theta)}, \tag{7.1}$$

where $D = (\theta_1, \theta_2)$.

By direct calculations,

$$\gamma^* = \lim_{\varepsilon \downarrow 0} b(\varepsilon)/\varepsilon = \left\{(f(0) - f(\pi))\int_{(-\theta_\alpha, \theta_\alpha)} \cos\theta \, dF(\theta)\right\}^{-1} f(0)\sin\theta_\alpha.$$

By Lemma 2.1(ii), we have $\beta^* = /\gamma^*$. The rest of the theorem is straightforward. □

Proof of Theorem 5.2. Assume $\theta_0 = 0$. The optimality in breakdown point follows immediately from Theorem 4.3. For the breakdown slope part, we show that for any Fisher consistent estimator the breakdown function satisfies

$$\varepsilon_\alpha^* \le 1 - \left\{1 + 2\alpha f(0) - 2\int_{\pi - \alpha}^{\pi} f(\theta)\, d\theta\right\}^{-1}.$$

It then follows that $\beta^* \le 2(f(0) - f(\pi))$, the breakdown slope for the directional median.

Let the density function f on $(0, 2\pi)$ be symmetric and unimodal about the direction $\theta = 0$. Contaminate f by

$$g(\theta) = \varepsilon^{-1}(1-\varepsilon)(f(0) - f(\theta)), \quad \text{for } \theta \in [0, 2\alpha]$$

and

$$g(\theta) = \varepsilon^{-1}(1-\varepsilon)(f(\theta - 2\alpha) - f(\theta)), \quad \text{for } \theta \in [2\alpha, \pi + \alpha],$$

but $g(\theta) = 0$ elsewhere. Note that

$$\int_0^{2\pi} g(\theta)\, d\theta = \varepsilon^{-1}(1-\varepsilon)\left\{2\alpha f(0) - 2\int_{\pi-\alpha}^{\pi} f(\theta)\, d\theta\right\}.$$

So, for g to be a density, ε must be $1 - \{1 + 2\alpha f(0) - 2\int_{\pi-\alpha}^{\pi} f(\theta)\, d\theta\}^{-1}$. Clearly, $h = (1-\varepsilon) f + \varepsilon g$ is axially symmetric and unimodal about $\theta = \alpha$. Therefore, any Fisher consistent estimator can be driven to α under this ε-contamination. This completes the proof. □

Proof of Theorem 6.2. Take $b = 1$ in (B3). Breakdown to μ under ε-contamination requires that for some distribution G on S_p,

$$\begin{aligned}(1-\varepsilon) E_F \eta(\|X - \mu\|) &+ \varepsilon E_G \eta(\|X - \mu\|)\\ &\le (1-\varepsilon) E_F \eta(\|X - \mu_0\|) + \varepsilon E_G \eta(\|X - \mu_0\|).\end{aligned}$$

Therefore, the breakdown function is

$$\varepsilon_\mu^* \ge \frac{E\eta(\|X - \eta\|) - E\eta(\|X - \mu_0\|)}{\max_y |\eta(\|y - \mu\|) - \eta(\|y - \mu_0\|)| + E\eta(\|X - \mu\|) - E\eta(\|X - \mu_0\|)},$$

where the expectations are under $F(x; \mu_0, \kappa)$ unless specified otherwise.

Under the conditions (B2) [or (B2′)] and (B3),

$$\limsup_{\mu \to \mu_0} \frac{\max_y |\eta(\|y - \mu\|) - \eta(\|y - \mu_0\|)|}{\|\mu - \mu_0\|} \le c$$

for some constant c. It therefore suffices to show that

$$\liminf_{\kappa \to \infty} \liminf_{\mu \to \mu_0} \frac{E\eta(\|X - \mu\|) - E\eta(\|X - \mu_0\|)}{\kappa^{1/2}\|\mu - \mu_0\|^2} \ge c_0 \tag{7.2}$$

and

$$\liminf_{\kappa \to 0} \liminf_{\mu \to \mu_0} \frac{E\eta(\|X - \mu\|) - E\eta(\|X - \mu_0\|)}{\kappa\|\mu - \mu_0\|^2} \ge c_0 \tag{7.3}$$

for some constant $c_0 > 0$.

Let $H = I - \mu_0 \mu_0'$. By Lemma 7.1, (7.2) is equivalent to

$$\begin{aligned} L \equiv \lim_{k \to \infty} &\left[v' E\{\kappa^{3/2} \eta(\|X - \mu_0\|) H X X' H\} v - E\{\kappa^{1/2} \eta(\|X - \mu_0\|) X' \mu_0\}\right]\\ &\ge c_0 > 0\end{aligned}$$

for $v \in S_p$ such that $\|Hv\| = 1$ or equivalently $v'\mu_0 = 0$.

By Lemma 7.2,

$$v'E\{\kappa^{3/2}\eta(\|X-\mu_0\|)HXX'H\}v$$
$$= \frac{1}{p-1}E\Big\{\kappa^{3/2}\eta\Big(\sqrt{2(1-X'\mu_0)}\Big)\big(1-(X'\mu_0)^2\big)\Big\}.$$

Since $2\kappa(1-X'\mu_0)\to\chi^2_{p-1}$ as $\kappa\to\infty$, with (B3) one obtains by direct verification

$$E\Big\{\kappa^{3/2}\eta\Big(\sqrt{2(1-X'\mu_0)}\Big)\big(1-(X'\mu_0)^2\big)\Big\}\to E(Z^{3/2})=\frac{2\sqrt{2}\,\Gamma(p/2+1)}{\Gamma((p-1)/2)}$$

where $Z\sim\chi^2_{p-1}$. Similarly,

$$E\{\kappa^{1/2}\eta(\|X-\mu_0\|)X'\mu_0\}\to E(Z^{1/2})=\frac{\sqrt{2}\,\Gamma(p/2)}{\Gamma((p-1)/2)}.$$

Therefore,

$$L=\frac{2\sqrt{2}\,\Gamma(p/2+1)}{(p-1)\Gamma((p-1)/2)}-\frac{\sqrt{2}\,\Gamma(p/2)}{\Gamma((p-1)/2)}=\frac{1}{p-1}\frac{\sqrt{2}\,\Gamma(p/2)}{\Gamma((p-1)/2)}>0,$$

which proves (7.2).

Again from Lemma 7.1. (7.3) is equivalent to $E\eta(\|U-\mu_0\|)U'\mu_0<0$ for a uniform variable U on S_p, which is immediate by the monotonicity of η function. □

Lemma 7.1.

$$E\eta(\|X-\mu\|)-E\eta(\|X-\mu_0\|)$$
$$=\tfrac{1}{2}\kappa^2 v'E\{\eta(\|X-\mu_0\|)HX'XH\}v\|\mu-\mu_0\|^2$$
$$-\tfrac{1}{2}\kappa E\{\eta(\|X-\mu_0\|)X'\mu_0\}\|\mu-\mu_0\|^2+o\big(\kappa^2\|\mu-\mu_0\|^2\big)$$

for any $v\in S_p$ *with* $v'\mu_0=0$.

Proof. By rotation and reflection, one has

$$\int\eta(\|x-\mu\|)\,dF(x;\mu_0,\kappa)=\int\eta(\|x-\mu_0\|)\,dF(x;\mu,\kappa),\tag{7.4}$$

where $dF(x;\mu,\kappa)=c_\kappa e^{\kappa x'\mu}\,dS(\kappa)$ with $dS(x)$ being the uniform measure on S_p. Then

$$E\eta(\|X-\mu\|)-E\eta(\|X-\mu_0\|)$$
$$=\int\eta(\|x-\mu_0\|)(e^{\kappa x'(\mu-\mu_0)}-1)\,dF(x;\mu_0,\kappa)$$
$$=\sum_{m=1}^{\infty}\frac{\kappa^m}{m!}E\{\eta(\|x-\mu_0\|)(x'(\mu-\mu_0))^m\}\tag{7.5}$$
$$=\kappa E\eta(\|X-\mu_0\|)X'(\mu-\mu_0)+\tfrac{1}{2}\kappa^2E\eta(\|X-\mu_0\|)\{X'(\mu-\mu_0)\}^2$$
$$+o\big(\kappa^2\|\mu-\mu_0\|^2\big),$$

where the change of sum and integral can be justified by Fubini's theorem.

Write $x = Hx + (x'\mu_0)\mu_0$ with $H = I - \mu_0\mu_0'$ being the projection onto the space orthogonal to μ_0. Also note that

$$x'(\mu - \mu_0) = (Hx)'(\mu - \mu_0) - \tfrac{1}{2}(x'\mu_0)\|\mu - \mu_0\|^2.$$

After some simplification, (7.5) becomes

$$\begin{aligned}&\kappa E\{\eta(\|X - \mu_0\|)(HX)'\}(\mu - \mu_0)\\ &+\tfrac{1}{2}\kappa^2(\mu - \mu_0)' E\{\eta(\|X - \mu_0\|)HXX'H\}(\mu - \mu_0)\\ &-\tfrac{1}{2}\kappa E\{\eta(\|X - \mu_0\|)X'\mu_0\}\|\mu - \mu_0\|^2 + o(\kappa^2\|\mu - \mu_0\|^2).\end{aligned} \tag{7.6}$$

By symmetry, the first term vanishes. Observe that

$$H(\mu - \mu_0) = \mu - \mu_0 + \tfrac{1}{2}\|\mu - \mu_0\|^2,$$
$$\|H(\mu - \mu_0)\| = \|\mu - \mu_0\| + O(\|\mu - \mu_0\|^2),$$

we have

$$H(\mu - \mu_0) \equiv v\|H(\mu - \mu_0)\| = v\|\mu - \mu_0\| + O(\|\mu - \mu_0\|^2).$$

Therefore, (7.6) reduces to

$$\begin{aligned}&\tfrac{1}{2}\kappa^2 v'E\{\eta(\|X - \mu_0\|)HXX'H\}v\|\mu - \mu_0\|^2\\ &\quad-\tfrac{1}{2}\kappa E\{\eta(\|X - \mu_0\|)X'\mu_0\}\|\mu - \mu_0\|^2 + o(\kappa^2\|\mu - \mu_0\|^2).\end{aligned}$$

By Lemma 7.2, the first term here is invariant in v. □

LEMMA 7.2. *If F is axially symmetric about μ_0, then for any bounded function $\tau(r)$,*

$$v'E_F\{\tau(X'\mu_0)HXX'H\}v = \frac{1}{p - 1}E_F\{\tau(X'\mu_0)(1 - (X'\mu_0)^2)\}$$

for any $v \in S_p$ with $\|Hv\| = 1$ or equivalently $v'\mu_0 = 0$.

PROOF. By symmetry, $U = \|HX\|^{-1}HX$ is independent of $\|HX\|$. Furthermore, U is uniformly distributed on the $(p - 1)$-dimensional unit sphere orthogonal to μ_0. So $v'(EUU')v$ is invariant in v and equal to $Ez^2 = (p - 1)^{-1}$, where z is any component of a uniform variable on S_{p-1}. Therefore,

$$v'E_F\{\tau(X'\mu_0)HXX'H\}v = E_F\{\tau(X'\mu_0)\|HX\|^2\}v'(EUU')v.$$

The lemma follows by writing $\|HX\|^2 = 1 - (X'\mu_0)^2$. □

Acknowledgments. We are grateful to two referees and an Associate Editor for their valuable comments which led to a substantial improvement of the article and to elimination of an error in an earlier draft.

REFERENCES

BARNETT, V. and LEWIS, T. (1984). *Outliers in Statistical Data*. Wiley, New York.

DONOHO, D. L. and LIU, E. L. (1988). The "automatic" robustness of minimum distance functionals. *Ann. Statist.* **16** 552–586.

FISHER, N. I. (1985). Spherical medians. *J. Roy. Statist. Soc. Ser. B* **47** 342–348.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HE, X. (1989). Contributions to the theory of statistical breakdown. Ph.D. dissertation, Univ. Illinois.

HE, Z. and SIMPSON, D. G. (1990). Bias-robust estimation for parametric models. Technical Report 42, Dept. Statistics, Univ. Illinois.

HE, X., SIMPSON, D. G. and PORTNOY, S. L. (1990). Breakdown robustness of tests. *J. Amer. Statist. Assoc.* **85** 446–452.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

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

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4** 188–234.

KO, D. and GUTTORP, P. (1988). Robustness of estimators for directional data. *Ann. Statist.* **16** 609–618.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

LENTH, R. V. (1981). Robust measure of location for directional data. *Technometrics* **23** 77–81.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic, New York.

MARDIA, K. V. (1975). Statistics of directional data. *J. Roy. Statist. Soc. Ser. B* **37** 349–393.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

ROUSSEEUW, P. J. (1984). Least median of squares regression. *J. Amer. Statist. Assoc.* **79** 871–880.

RUPPERT, D. (1985). On the bounded-influence regression estimator of Krasker and Welsch. *J. Amer. Statist. Assoc.* **80** 205–208.

WATSON, G. S. (1967). Some problems in the statistics of directions. *Bull. 36th Session, ISI.* Sidney, Australia.

WATSON, G. S. (1983). *Statistics on Spheres. Univ. Arkansas Lecture Notes Math. Sci.* Wiley, New York.

WEHRLY, T. E. and SHINE, E. P. (1981). Influence curves of estimators for directional data. *Biometrika* **68** 334–335.

DEPARTMENT OF MATHEMATICS
NATIONAL UNIVERSITY OF SINGAPORE
10 KENT RIDGE CRESCENT
SINGAPORE 0511
REPUBLIC OF SINGAPORE

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820