18.465, April 12, 2005

Some notes on location and scatter functionals

Recall that a sequence Q_k of laws (probability measures), here on $\mathbb{R}^d$, is said to converge *weakly* to a law Q if $\int f dQ_k \to \int f dQ$ for every bounded continuous function f. There exists a metric ρ on the set of all laws on $\mathbb{R}^d$ which metrizes weak convergence, in other words $Q_k \to Q$ weakly if and only if $\rho(Q_k, Q) \to 0$, e.g. Dudley (2002, Sec. 11.3). A set U of laws is called *weakly open* if and only if whenever $Q \in U$ and $Q_k \to Q$ weakly we have $Q_k \in U$ for all k large enough. Equivalently, for each $Q \in U$, there is an $r > 0$ such that whenever $\rho(Q, P) < r$ we have $P \in U$.

Much of robustness theory emphasizes mixture laws

$$P = (1 - \lambda)F_0 + \lambda Q \tag{1}$$

where Q is an arbitrary "contaminating" distribution, F_0 is a special distribution with a density, say for definiteness a normal, and $0 \leq \lambda < 1/2$, e.g. Huber [20, pp. 86, 89]. Despite the generality of Q, the contamination model (1) doesn't include some, perhaps the majority, of laws P treated as normal to an acceptable approximation in practice, such as laws P on $\mathbb{R}$ with $P([0, \infty)) = 1$, and laws discretized by rounding to finitely many decimal places. The latter laws also cannot be obtained by replacement of up to half a normal or other continuous law, but can be quite close to normal laws in metrics for the weak topology. Huber [20, p. 3] says that "in the physical sciences typical 'good data' samples appear to be well modeled by an error law" (1) with $0.01 \leq \lambda \leq 0.1$. But, "modeled" seems to allow a further approximation and "error" seems to exclude many, perhaps most, data sets.

Another basic notion in robustness theory is that of breakdown point. Before giving some definitions of them, here are remarks on *Notations with* δ: below, "δ" is used in the following three ways: δ_x (without any superscript) denotes the law which is a point mass at x; δ^* with varying subscripts will be breakdown points, to be defined; and δ with neither subscript nor superscript will be a (small) number, usually introduced by "for any $\delta > 0$" or the like.

Some definitions of breakdown points are for estimators T_n defined on a finite sample of size n under replacement of a fraction k/n of the observations by arbitrary values, or by adjoining k new such values to the data. Then the asymptotics of the breakdown point (largest k/n for replacement, or

$k/(n+k)$ for adjunction, such that T_n doesn't escape from all compact sets) as $n \to \infty$ are considered. Another type of definition is for functionals T defined on laws P, which yield estimators when applied to empirical measures P_n. In a functional definition one has a set of neighborhoods $N_\varepsilon(P)$ of P indexed by $\varepsilon > 0$. These may be defined by a metric d on laws through $N_\varepsilon(P) := N_{\varepsilon,d}(P) := \{Q : \ d(Q,P) < \varepsilon\}$, or in most of the literature, as contamination neighborhoods, for $0 < \varepsilon \le 1$ (nearly always $\varepsilon \le 1/2$),

$$N_\varepsilon^C(P) \ := \ \{Q = (1-\lambda)P + \lambda\rho : \ 0 \le \lambda \le \varepsilon, \ \rho \ \text{ any law}\}.$$

The *total variation* distance between two laws P and Q on a sample space $(X, \mathcal{B})$ is

$$d_1(P,Q) \ := \ \sup_{A\in\mathcal{B}} |(P-Q)(A)| = \sup_{A\in\mathcal{B}} (P-Q)(A) = \sup_{B\in\mathcal{B}} (Q-P)(B)$$

by the Hahn-Jordan decomposition, e.g. Dudley (2002, Theorem 5.6.1). Total variation for laws corresponds approximately to replacement for finite samples. If $\mathcal{P} \ := \ \mathcal{P}(X,\mathcal{B})$ is the set of all laws on a sample space $(X,\mathcal{B})$, and for each $P \in \mathcal{P}$, $\{N_\varepsilon(P)\}_{0\le\varepsilon<\infty}$ is a collection of subsets of $\mathcal{P}$, then $\{N_\varepsilon(P) : \ 0 \le \varepsilon < \infty, \ P \in \mathcal{P}\}$ will be called a *suitable set of neighborhoods* iff for all $P \in \mathcal{P}$, (a) $N_0(P) = \{P\}$, (b) For $0 \le \varepsilon < \varepsilon'$, $N_\varepsilon(P) \subset N_{\varepsilon'}(P)$, and (c) For $\varepsilon > 0$, $\varepsilon' > 0$, $Q \in N_\varepsilon(P)$, and $\rho \in N_{\varepsilon'}(Q)$, we have $\rho \in N_{\varepsilon+\varepsilon'}(P)$.

These conditions clearly hold for neighborhoods defined by metrics. They also hold for contamination neighborhoods if we define $N_\varepsilon^C(P) \ := \ N_1^C(P) = \mathcal{P}$ for $\varepsilon > 1$.

In most definitions found in the literature, T takes values in a parameter space Θ with a topology and for $\varepsilon > \varepsilon^*(T,P)$, the breakdown point at P, there is no proper compact subset $K \subset \Theta$ such that T on $N_\varepsilon(P)$ takes values in K. (For non-compact parameter spaces, as in these notes, "proper" is redundant.)

A set of n points in $\mathbb{R}^d$ are said to be in *general position* if for $k = 0, 1, \ldots, d-1$, no k-dimensional hyperplane contains $k+2$ or more of the points.

Many authors consider breakdown points of functionals in the contamination sense at laws P on $\mathbb{R}^d$ such that $P(H) = 0$ for any hyperplane H of dimension $d-1$. I.i.d. samples from such laws are almost surely in general position. On finite samples, breakdown (in the replacement or contamination sense) is usually considered for samples in general position. At other laws or samples, the breakdown points may be lower.

Another issue is: for $0 < \varepsilon < \varepsilon^*$, is T required to be uniquely defined at all $Q \in N_\varepsilon(P)$? Different answers might be deduced from the literature. On the "no" side, the Rousseeuw minimum-volume-ellipsoid (MVE) functional, to be defined after Proposition 2, has been generally agreed to have breakdown point 1/2 at suitable P although it had only been shown to be uniquely defined at symmetric, unimodal distributions satisfying further restrictions, as in Tatsuoka and Tyler [35]; this set is not dense in $N_\varepsilon(P)$ for any $\varepsilon > 0$, for any of the families of neighborhoods mentioned so far. On the "yes" side, proofs of upper bounds for the breakdown points of some M-functionals (Hampel, Ronchetti, Rousseeuw and Stahel [17, §5.5(a) p. 298]; Tyler [36]) assume that the functionals are defined on contamination neighborhoods of a normal law, or of finite samples in general position, respectively. Since each answer is of independent interest, separate definitions will be given.

Definition. Let Θ be a topological space, $(X, \mathcal{B})$ a sample space, and $\{N_\varepsilon(P),\ 0 \le \varepsilon < \infty,\ P \in \mathcal{P} := \mathcal{P}(X, \mathcal{B})\}$ a suitable set of neighborhoods. Let T be a functional defined uniquely on a domain $\mathcal{D} \subset \mathcal{P}$ with values in Θ. Then for each $P \in \mathcal{D}$, the *explosion breakdown point* of T at P is

$$\varepsilon^*(T, P) \ := \ \varepsilon^*(T, P, \{N_\varepsilon\}_{0 \le \varepsilon < \infty}, \Theta) \ := \ \inf\{\varepsilon \in [0, 1] :$$
$$\text{for each compact } \ K \subset \Theta,\ T(Q) \notin K \ \text{ for some } \ Q \in \mathcal{D} \cap N_\varepsilon(P)\}.$$

If there is no such ε, set $\varepsilon^* := 1$. Let $\varepsilon^*_C(T, P)$ denote the explosion breakdown point for contamination neighborhoods, and $\varepsilon^*_d(T, P)$ the one for d-neighborhoods.

The next definition, of δ^*, requires T to be uniquely defined on some neighborhoods. Sometimes T becomes undefined only just after escaping from compact sets, so that $\delta^* = \varepsilon^*$.

Definition. Let $\delta^*(T, P)$, the *definition-explosion breakdown point* of T at P, be defined as the supremum of ε with $0 < \varepsilon < \varepsilon^*(T, P)$ such that $N_\varepsilon(P) \subset \mathcal{D}$, or 0 if there is no such ε. Define δ^*_C and δ^*_d by analogy with ε^*_C and ε^*_d.

Here is a further definition, of r^*. It will not be called a breakdown point since discontinuity has not generally been considered as breakdown.

Definition. Let $r^*(T,P)$, the *radius of continuity* of T at P, be defined as $\delta^*(T,P)$ with the additional requirement that $T(\cdot)$ is weakly continuous at Q for all $Q \in N_\varepsilon(P)$. Define r^*_C and r^*_d again analogously.

If neighborhoods N_ε are defined by the total variation (replacement) distance d_1 then the corresponding breakdown points and radii will be written as $\varepsilon^*_R := \varepsilon^*_{d_1}$. If $Q = (1-\lambda)P + \lambda\rho$ for any law ρ then clearly $d_1(P,Q) \le \lambda$, with equality if ρ is singular with respect to P. Thus $\varepsilon^*_R \le \varepsilon^*_C$ and likewise for δ^* and r^*.

Notions of "location" and "scale" or multidimensional "scatter" functional will be defined in terms of equivariance, as follows.

Definitions. Let $\mathcal{N}_d$ be the set of symmetric nonnegative definite $d \times d$ matrices and $\mathcal{P}_d$ its subset of strictly positive definite matrices. Let $Q \mapsto \mu(Q) \in \mathbb{R}^d$, resp. $\Sigma(Q) \in \mathcal{N}_d$, be a functional defined on a set $\mathcal{D}$ of laws Q on $\mathbb{R}^d$. Then μ (resp. Σ) is called an *affinely equivariant location* (resp. *scatter*) *functional* iff for any nonsingular $d \times d$ matrix A and $v \in \mathbb{R}^d$, with $f(x) := Ax + v$, and any law $Q \in \mathcal{D}$, the image measure $P := Q \circ f^{-1} \in \mathcal{D}$ also, with $\mu(P) = A\mu(Q) + v$ or, respectively, $\Sigma(P) = A\Sigma(Q)A'$. For $d = 1$, $\sigma(\cdot)$ with $0 \le \sigma < \infty$ will be called an *affinely equivariant scale functional* iff σ^2 satisfies the definition of affinely equivariant scatter functional. If we have affinely equivariant location and scatter functionals μ and Σ on the same domain $\mathcal{D}$ then (μ,Σ) will be called an affinely equivariant location-scatter functional on $\mathcal{D}$, and likewise for a location-scale functional (μ,σ).

Dispersion often occurs in the literature as a synonym for "scatter." Clearly, for laws Q with finite second moments, the mean $\mu(Q)$ and covariance matrix $\Sigma(Q)$ give affinely equivariant location and scatter functionals.

The median is an affinely equivariant location functional with $\delta^*_C = 1/2$ at any law. The MAD is an affinely equivariant scale functional with $\delta^*_C(\mathrm{MAD},P) \equiv 1/2$ also if the scale parameter space is taken as $0 \le \sigma < \infty$. If $\sigma > 0$ is required, however, the MAD is not defined at laws P with $p := \sup\{P(\{t\}) : t \in \mathbb{R}\} > 1/2$ and at other laws P has $\delta^*_C(\mathrm{MAD},P) = \beta = (\frac{1}{2} - p)/(1-p)$, with $\beta = 1/2$ only for continuous laws P. Such a dependence on Θ naturally also occurs for other scale functionals, e.g. the interquartile range.

Let T be an affinely equivariant location or scatter functional. Then ε^*_C, δ^*_C, and r^*_C are all affinely invariant and $1/2$ as a target maximal value for

ε_C^* has been much emphasized in the literature. As will be seen, however, striving for $\varepsilon_C^* = 1/2$ has led to some functionals for which $\delta_C^* = 0$ or r_d^* may be 0 (even at laws with smooth densities).

For metrics d that metrize that weak topology and so are not affinely invariant (e.g. the Prohorov metric, Dudley, 2002, Sec. 11.3), ε_d^*, δ_d^* and r_d^* may still be affinely invariant (if they are constant!), e.g. for T the median and d the Prohorov metric, $\varepsilon_d^*(T,P) = \delta_d^*(T,P) \equiv 1/2 > 0 \equiv r_d^*(T,P)$ for all P. But e.g. for $T = \text{MAD}$ and $\Theta = (0,\infty)$, $\varepsilon_d^* = \delta_d^*$ is not affinely invariant. On the other hand the sets where $\varepsilon_d^* > 0$, $\delta_d^* > 0$ and $r_d^* > 0$ are affinely invariant. Thus, one may seek T for which these sets are as large as possible, rather than making the values of ε^* as large as possible.

Location functionals which in some respects improve on the median and still have $\delta_C^* = 1/2$ at all laws have been proposed, especially by Huber, e.g. [20, pp. 52-53, (5.22) p. 86]. Such functionals can be adjusted for scale, e.g. using the MAD, to make them equivariant [20, §§6.4-6.7], and can be defined when the scale functional $\sigma = 0$, as we saw in earlier handouts.

The requirement of affine equivariance seems to be especially natural for laws on $\mathbb{R}$. In $\mathbb{R}^d$ for $d \geq 1$, the *spatial median* for a random vector X or its law is an m that minimizes $E(|X-m|-|X|)$. For $d = 1$, m satisfies this iff it is a median of X. For $d \geq 2$ the spatial median is unique except for distributions concentrated in lines with non-unique medians there [28], as also shown in a handout. The spatial median is equivariant under Euclidean transformations where A is an orthogonal transformation, or a constant multiple of one, but not under general affine transformations for $d > 1$.

The following easy fact gives consequences of affine equivariance without any further assumptions.

Theorem 1. *Let $\mu(\cdot)$ be an affinely equivariant location functional defined on a class $\mathcal{D}$ of laws on $\mathbb{R}^d$, and let $\mathcal{A}$ be a set of non-singular affine transformations of $\mathbb{R}^d$. Let $P \in \mathcal{D}$ be such that $P \circ A^{-1} = P$ for each $A \in \mathcal{A}$. Then*
(a) $\mu(P) \in S_{\mathcal{A}} := \{x \in \mathbb{R}^d : \ Ax = x \text{ for all } A \in \mathcal{A}\}$.
(b) If $S_{\mathcal{A}}$ is a singleton $\{x_{\mathcal{A}}\}$, then $\mu(P) = x_{\mathcal{A}}$.
(c) If for some $v \in \mathbb{R}^d$, $\mathcal{A}$ consists of the one map $x \mapsto 2v - x$, then $\mu(P) = v$.
(d) Let $2 \leq n \leq d+1$. Let V be a set of n points of $\mathbb{R}^d$ in general position. Then for any of the $n!$ permutations π of the points of V, there is a non-singular affine A_π, uniquely determined on the unique $(n-1)$-dimensional

hyperplane H including V, with $A_\pi(v) = \pi(v)$ for each $v \in V$. If the hypotheses on P hold for $\mathcal{A}$ equal to the set of all these A_π, and $P(H) = 1$, then $\mu(P) = n^{-1} \sum_{v \in V} v$.
(e) In part (d), suppose $n = d + 1$ and the points of V are the vertices of a regular simplex. Let Σ be an affinely equivariant scatter functional. Then $\Sigma(P) = cI$ for some $c \geq 0$ where I is the $d \times d$ identity matrix.

Proof. Part (a) follows directly from the definition of equivariant location functional. Then part (b) follows from part (a). For part (c), note that $x \mapsto 2v - x$ has a unique fixed point v, so (c) follows from (b); here P is symmetric around v.

For part (d), let $x_1, ..., x_n$ be the points of V. Since they are in general position, the vectors $v_j = x_j - x_1$ for $j = 2, ..., n$ are linearly independent. Let $y_1, ..., y_n$ be the vertices of a regular $(n-1)$-dimensional simplex with all edges of equal length. Then clearly $y_1, ..., y_n$ are also in general position and $w_j = y_j - y_1$ for $j = 2, ..., n$ are linearly independent. Thus there is a non-singular linear transformation (matrix) B with $Bv_j = w_j$ for $j = 2, ..., n$. Defining a non-singular affine transformation by $Ax = B(x - x_1) + y_1 = Bx + (y_1 - Bx_1)$ we have $Ax_j = y_j$ for $j = 1, ..., n$, so we can assume that x_j are the vertices of a regular simplex.

Recall from group theory that any permutation can be obtained by composing transpositions, so given any two points u, v of V we need to find an affine A with $Au = v$, $Av = u$, and $Aw = w$ for all $w \in V$ other than u and v. For V the set of vertices of a regular simplex, we can take A as reflection in the $(d-1)$-dimensional hyperplane perpendicular to $u - v$ and through the midpoint of the line segment from u to v, so the affine transformations A_π exist.

Let $W := \{\sum_{i=2}^n s_i(x_i - x_1) : \ s_i \in \mathbb{R}, \ i = 2, ..., n\}$, an $(n-1)$-dimensional linear subspace of $\mathbb{R}^d$. Then $W = \{\sum_{i=1}^n t_i x_i : \ t_i \in \mathbb{R}, \ i = 1, ..., n, \sum_{j=1}^n t_j = 0\}$, as is seen by the relations $t_i = s_i$ for $i = 2, ..., n$ and $t_1 = -\sum_{j=2}^n s_j$. For a given point of W, the numbers s_i or t_i are uniquely determined. It's easily seen that $H = x_1 + W = \{x_1 + w : \ w \in W\}$. Then $H = \{\sum_{j=1}^n \lambda_j x_j : \ \lambda_j \in \mathbb{R}, \ j = 1, ..., n, \ \sum_{j=1}^n \lambda_j = 1\}$, where $\lambda_1 = 1 + t_1$ and $\lambda_j = t_j$ for $j = 2, ..., n$, and the λ_j are uniquely determined. If A is any affine transformation of $\mathbb{R}^d$, then for any $\{\lambda_j\}_{j=1}^n \in \mathbb{R}^n$ with $\sum_{j=1}^n \lambda_j = 1$, $A\left(\sum_{j=1}^n \lambda_j x_j\right) = \sum_{j=1}^n \lambda_j A(x_j)$. If A leaves each x_j fixed, it follows that A leaves fixed each point of H, so A is uniquely determined on H as stated.

There is an affine transformation A_H of $\mathbb{R}^d$ such that $A_H x = x$ for all x in H and $A_H y \neq y$ for all y not in H. Then A_H induces the identity permutatation of V, and we can assume it is the affine transformation chosen to do so since $P(H) = 1$ and so $P \circ A_H^{-1} = P$. Thus by part (a), $\mu(P) \in H$.

Let π be a permutation interchanging x_i and x_j for some $i \neq j$ and A_π the corresponding affine transformation. We have $\mu(P) = \sum_{j=1}^n \lambda_j x_j$ for some real λ_j with sum 1, and $\mu(P) = A_\pi \mu(P) = \mu(P) + (\lambda_i - \lambda_j)(x_j - x_i)$, so $\lambda_i = \lambda_j$, and $\lambda_i = 1/n$ for all i, so $\mu(P) = n^{-1} \sum_{v \in V} v$, proving (d).

For part (e), we can assume $\sum_{v \in V} v = 0$. Let $v \neq w$ in V and let $S := V \setminus \{v, w\}$. Let A be the linear transformation interchanging v and w and leaving each $s \in S$ fixed. Then A is the reflection in the linear subspace spanned by S, which contains $(v + w)/2$. By affine equivariance it follows that $v - w$ is an eigenvector of $\Sigma(P) \in \mathcal{N}_d$. Eigenvectors with distinct eigenvalues are orthogonal, but for $v \neq u \neq w$ in V, $v - w$ and $v - u$ are not orthogonal, so they must have the same eigenvalue. Iterating, we find that all such eigenvectors have the same eigenvalue $c \geq 0$. Since they span $\mathbb{R}^d$, $\Sigma(P) = cI$ follows. $\Box$

In part (c), of course, not all symmetric distributions P are necessarily in the domain $\mathcal{D}$ on which $\mu(\cdot)$ is (uniquely) defined. To put all symmetric distributions in $\mathcal{D}$ could violate some other useful property of $\mu(\cdot)$, as Tatsuoka and Tyler [35, p. 1235]) note. One can look for $\mu(\cdot)$ with good properties defined on as many symmetric laws as possible.

The simplest special case of Theorem 1 part (d) is that P puts mass $1/n$ in each point of V. That case is natural in that any simplex is affinely equivalent to a regular simplex with all vertices equidistant, whose centroid is the obvious location. Yet, if $n - 1$ observations are close together and the nth moves far away, it retains its non-robust influence. By nesting multiple such simplices for $n = d + 1$, Donoho and Gasko [10] illustrate why the breakdown point of a purportedly robust estimator is as low as $1/(d + 1)$, a bound which, apparently for different reasons, they also found for another class of estimators, as Maronna [26] did earlier for equivariant M-estimators of location and scatter.

Here is a related consequence of Theorem 1, not directly about breakdown points:

Proposition 2. *For $d = 1, 2, \ldots$, there is a sequence $\{Q_m\}_{m \geq 3}$ of laws on $\mathbb{R}^d$ having densities such that for a compact set $K \subset \mathbb{R}^d$, for all $m \geq 3$,*

$Q_m(K) = d/(d+1)$ *and there exist* $\mu_m \in \mathbb{R}^d$ *such that for every affinely equivariant location functional* $\mu(\cdot)$ *defined at* Q_m*,* $\mu(Q_m) = \mu_m$*, and* $|\mu_m| \to \infty$ *as* $m \to \infty$*.*

Proof. Let V be the set of $d+1$ vertices of a regular simplex S such that $e_1 := (1, 0, \ldots, 0)' \in V$, the other d vertices all are in the subspace where $x_1 = 0$, and the centroid of S is $e_1/(d+1)$. All points of V are within a distance 1 of 0. Let $P_{d+1} := \frac{1}{d+1}\sum_{v\in V}\delta_v$. For any $r > 0$ let U_r be the uniform distribution on the ball in $\mathbb{R}^d$ with center 0 and radius r. Let $\rho_r := P_{d+1} * U_r$, which has a density. For $\mathcal{A}$ as in Theorem 1(e) and (d), since each $A \in \mathcal{A}$ is an orthogonal transformation preserving U_r, we have $\rho_r \circ A^{-1} = \rho_r$. Let $\mu(\cdot)$ be an affinely equivariant location functional defined at ρ_r. Then $\mu(\rho_r) = e_1/(d+1)$. Let $M_a((x_1, \ldots, x_d)') := (ax_1, x_2, \ldots, x_d)'$ for any $a > 0$ and $\tau_r := \rho_r \circ M_{1/r}^{-1}$. Then for $r \le 1/3$, τ_r has probability $1/(d+1)$ in the half-space $x_1 \ge (1/r) - 1$ and $d/(d+1)$ in the ball $K := \{x : |x| \le 2\}$, with $|\mu(\tau_r)| = 1/[r(d+1)]$ if $\mu(\tau_r)$ is defined. Letting $Q_m := \tau_{1/m}$ for $m \ge 3$ gives the conclusion. $\Box$

Rousseeuw [30] defined the minimum-volume ellipsoid (MVE) location-scatter estimator whose functional form is as follows. Given a law P on $\mathbb{R}^d$, suppose there is a unique ellipsoid $E = \{x : (x-\mu)'C^{-1}(x-\mu) \le 1\}$ of smallest d-dimensional volume with $P(E) \ge 1/2$, where x and μ are column vectors in $\mathbb{R}^d$ and C is a positive definite symmetric $d \times d$ matrix. Dividing C by a constant $c_d > 0$ depending on d we can write $E = \{x : (x-\mu)'\Sigma^{-1}(x-\mu) \le c_d\}$, where c_d is chosen so that if P is a normal distribution, Σ is its covariance matrix. Then μ and Σ are affinely equivariant location and scatter functionals of P respectively, because any affine transformation A with $Ax \equiv Bx + v$ takes ellipsoids to ellipsoids and multiplies all volumes by the same (Jacobian) factor $\det B$.

For a finite sample of size n, if $\lfloor x \rfloor$ denotes the largest integer $\le x$, the MVE was originally defined in terms of the ellipsoid of smallest volume containing $\lfloor n/2 \rfloor + 1$ of the sample points. Later, this was adjusted to require E to contain $\lfloor (n+d+1)/2 \rfloor$ points, with the aim of maximizing the finite-sample breakdown point. In either case, asymptotically as $n \to \infty$, one gets the minimum-volume ellipsoid with probability $\ge 1/2$, if it is unique.

Location-scatter functionals with $\varepsilon_C^* = 1/2$ (at continuous distributions) for all d have been proposed, including the Rousseeuw minimum-volume-ellipsoid estimator just defined ([30], [5]), but δ_C^* for it is 0 as shown in Section 3 below.

Proposition 2 showed that mass $1/(d+1)$ escaping to ∞ can cause breakdown of quite general equivariant location functionals provided that the remaining $d/(d+1)$ of the mass, while remaining in a compact set, approaches some restricted limit (the limit apparently cannot have a density). In the given proof, the limit is concentrated in a union of d line segments parallel to the x_1 axis and so gives mass $k/(d+1)$ to some k-dimensional hyperplanes for $k = 1, \ldots, d-1$.

Proposition 6 will show that any affinely equivariant location functional $\mu(\cdot)$ on $\mathbb{R}$, if it has $\delta_C^* = 1/2$ at one or more laws, cannot be extended to be weakly continuous at the law $Q = \frac{1}{2}(\delta_0 + \delta_1)$. For a nonparametric location functional this is a drawback since by Theorem 1(c), $\mu(Q)$ naturally would be defined as 1/2. On the other hand there do exist location and scale functionals μ and σ, defined and weakly continuous on all laws on $\mathbb{R}$ and affinely equivariant, with $\delta_C^* = \alpha$ at every law, for any α with $0 < \alpha < 1/2$, via trimming (Section 3). Thus the notion of 1/2 as "optimal" breakdown point, often stated in the literature, may not apply from a nonparametric viewpoint.

1 Nonexistence facts in dimension 2 or higher

Call a location functional $\mu(\cdot)$ or a scatter functional $\Sigma(\cdot)$ *singularly affine equivariant* if in the definition of affine equivariance A can be any matrix, possibly singular. It's easily seen that if a functional is defined on all laws, affinely equivariant, and weakly continuous, then it is singularly affine equivariant. For empirical measures $P_n = n^{-1}(\delta_{X_1} + \cdots + \delta_{X_n})$, the classical sample mean and covariance are evidently singularly affine equivariant. It turns out that in dimension $d \geq 2$, there are essentially no other singularly affine equivariant location and scatter functionals, and so weak continuity at all laws is not possible. First the known fact for location will be recalled, then an at least partially known fact for scatter will be stated and proved.

Let X be a $d \times n$ data matrix whose jth column is $X_j \in \mathbb{R}^d$. Let X^i be the ith row of X. Let 1_n be the $n \times 1$ vector with all components 1. Let $\overline{X} = \int x dP_n$ be the sample mean vector in $\mathbb{R}^d$, so that $X - \overline{X}1_n'$ is the centered data matrix. Note that P_n, and thus $\overline{X}$ and $\Sigma(X)$, are preserved by any permutation of the columns of X. The next fact was proved in detail in the handout "Non-existence of some affinely equivariant functionals in dimension $d \geq 2$."

Theorem 3. *(a) If $\mu(\cdot)$ is a singularly affine equivariant location functional (estimator) defined for all P_n on $\mathbb{R}^d$ for $d \geq 2$ and a fixed n, then $\mu(P_n) \equiv \int x dP_n$, the sample mean.*
(b) If in addition $\mu(\cdot)$ is defined for all n and all P_n on $\mathbb{R}^d$, then as n varies, $\mu(\cdot)$ is not weakly continuous. Thus, there is no affinely equivariant, weakly continuous location functional defined on all laws on $\mathbb{R}^d$ for $d \geq 2$.

Proof. Part (a) follows from a result and proof of Obenchain [29, Lemma 1] and permutation invariance, as noted in an unpublished paper of Donoho and by Rousseeuw [30], [31, Proposition 2]. Then (b) follows directly, for $x_1 = n$, $x_2 = \cdots = x_n = 0$, $n \to \infty$. $\square$

Next is a related fact about scatter functionals. Davies [7, p. 1879] made a statement closely related to part (b), strong but not quite in the same generality, and very briefly indicated a proof by saying that the fact "corresponds" to one for location functionals, as in the preceding theorem. I don't know a reference for part (a), so a proof will be given.

Theorem 4. *(a) Let $\Sigma(\cdot)$ be a singularly affine equivariant scatter functional defined on all empirical measures P_n on $\mathbb{R}^d$ for $d \geq 2$ and some fixed $n \geq 2$. Write $\Sigma(X) := \Sigma(P_n)$. Then there is a constant $c_n \geq 0$, depending on $\Sigma(\cdot)$, such that for any X, $\Sigma(X - \overline{X}1_n') = c_n(X - \overline{X}1_n')(X - \overline{X}1_n')'$. In other words, applied to centered data matrices, Σ is proportional to the sample covariance matrix.*
(b) If $\Sigma(\cdot)$ is an affinely equivariant scatter functional defined for all n and P_n on $\mathbb{R}^d$ for $d \geq 2$, weakly continuous as a function of P_n, then $\Sigma \equiv 0$.

Proof. (a) We have $\Sigma(BX) = B\Sigma(X)B'$ for any $d \times d$ matrix B. For any $U, V \in \mathbb{R}^n$ let $X^1 = U'$, $X^2 = V'$, and $(U, V) := \Sigma_{12}(X)$. Then $(\cdot, \cdot)$ is well-defined, letting $B_{11} = B_{22} = 1$ and $B_{ij} = 0$ otherwise. It will be shown that $(\cdot, \cdot)$ is a semi-inner product. We have $(U, V) \equiv (V, U)$ via B with $B_{12} = B_{21} = 1$ and $B_{ij} = 0$ otherwise, since Σ is symmetric. For $B_{11} = B_{21} = 1$ and $B_{ij} = 0$ otherwise we get for any $U \in \mathbb{R}^n$ that

$$(U, U) = \Sigma_{12}(BX) = (B\Sigma(X)B')_{12} = \Sigma_{11}(X) \geq 0. \tag{2}$$

For constants a and b, $(aU, bV) \equiv ab(U,V)$ follows for $B_{11} = a$, $B_{22} = b$, and $B_{ij} = 0$ otherwise. It remains to prove biadditivity $(U, V + W) \equiv (U,V)+(U,W)$. For $d \geq 3$ this is easy, letting $X^3 = W$, $B_{11} = B_{22} = B_{23} = 1$, and $B_{ij} = 0$ otherwise. For $d = 2$, we first get $(U + V, V) = (U, V) + (V, V)$ from $B = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Symmetrically, $(U, U + V) = (U, U) + (U, V)$. Then from $B = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ we get

$$(U + V, U + V) \;=\; (U, U) + 2(U, V) + (V, V). \tag{3}$$

Letting $\|W\|^2 := (W, W)$ for any $W \in \mathbb{R}^n$ we get the parallelogram law $\|U + V\|^2 + \|U - V\|^2 \equiv 2\|U\|^2 + 2\|V\|^2$. Applying this repeatedly we get for any W, Y, and $Z \in \mathbb{R}^n$ that

$$\|W+Y+Z\|^2-\|W-Y-Z\|^2 = \|W+Y\|^2-\|W-Y\|^2+\|W+Z\|^2-\|W-Z\|^2,$$

letting first $U = W + Y$, $V = Z$, then $U = W - Z$, $V = Y$, then $U = W$, $V = Z$, and lastly $U = W$, $V = Y$. Applying (3) and dividing by 4 gives $(W, Y + Z) \equiv (W, Y) + (W, Z)$, the desired biadditivity. So $(\cdot, \cdot)$ is indeed a semi-inner product, in other words there is a $C(n) \in \mathcal{N}_n$ such that $(U, V) \equiv U'C(n)V$. By the permutation invariance, there are numbers $a_n \geq 0$ and b_n such that $C(n)_{ii} = a_n$ for all $i = 1, \ldots, n$ and $C(n)_{ij} = b_n$ for all $i \neq j$. Let $c_n := a_n - b_n$.

Let $e_i \in \mathbb{R}^n$ be the ith standard unit vector. For each $y \in \mathbb{R}^n$ let $y = \sum_{i=1}^n y_i e_i$. Let $\overline{y} := \frac{1}{n}\sum_{i=1}^n y_i$, so that $y - \overline{y}1_n = \sum_{i=1}^n (y_i - \overline{y})e_i$. Then for any $z \in \mathbb{R}^n$,

$$(y - \overline{y}1_n, z - \overline{z}1_n) \;=\; \sum_{i,j=1}^n C(n)_{ij}(y_i - \overline{y})(z_j - \overline{z}) \;=\; c_n(y - \overline{y}1_n)'(z - \overline{z}1_n).$$

For $1 \leq j \leq k \leq d$, let $B_{ir} := \delta_{r\pi(i)}$ for a function π from $\{1, 2, \ldots, d\}$ into itself with $\pi(1) = j$ and $\pi(2) = k$. Then $(BX)^1 = X^j$ and $(BX)^2 = X^k$. Thus $(X^j, X^k) = \Sigma_{12}(BX) = \Sigma_{jk}(X)$, recalling (2) for $j = k$.

Let $\overline{X} \in \mathbb{R}^d$ have ith component $\overline{X}^i$ and $Y^j := (X^j)'$. Then

$$\Sigma_{jk}(X - \overline{X}1_n') \;=\; (Y^j - \overline{X}^j 1_n, Y^k - \overline{X}^k 1_n) \;=\; c_n(Y^j - \overline{X}^j 1_n)'(Y^k - \overline{X}^k 1_n),$$

where $c_n \geq 0$ is seen when $j = k$ and the coefficient of c_n is strictly positive, as it can be since $n \geq 2$. Thus part (a) is proved.

For part (b), consider empirical measures $P_n = P_{mn}$, so that each X_j in P_n is repeated m times in P_{mn}. Since the $\overline{X}$'s and Σs for P_n and P_{mn} must be the same, we get that $c_{mn} = c_n/m$ which likewise equals c_m/n. Thus there is a constant c_1 such that $c_n = c_1/n$ for all n.

Let $X_{11} := -X_{12} := \sqrt{n}$, let $X_{ij} = 0$ for all other i, j and let $n \to \infty$. Then $\overline{X} \equiv 0$, $P_n \to \delta_0$ weakly, and $\Sigma(\delta_0)$ is the 0 matrix by singular affine equivariance with $B = 0$, but $\Sigma(P_n)$ don't converge to 0 unless $c_1 = 0$ and so $c_n = 0$ for all n, proving (b). $\square$

So, the three properties of T: (a) affine equivariance, (b) weak continuity on its domain $\mathcal{D}$, and (c) being everywhere defined, cannot all hold for location or scatter functionals on $\mathbb{R}^d$ for $d \geq 2$ although they can for $d = 1$. Which one(s) should be given up? Some functionals, such as the median and MAD, fail (b), but for $d > 1$, it seems that known functionals tend to fail (c). Specifically, if Σ is required to be strictly positive definite, then at a law concentrated in a proper hyperplane, Σ cannot be defined and affinely equivariant.

One can then ask: on how large a domain $\mathcal{D}$ of laws can (a) and (b) hold? Consider replacing (c) by:

(c$'$) $\mathcal{D}$ is open and dense for the weak topology in the set of all laws on $\mathbb{R}^d$.

If (c$'$) holds then the functional is undefined only on some nowhere dense and thus topologically small set. Let d metrize weak convergence. Then both $\mathcal{D}$ is open and (b) holds if and only if for each $P \in \mathcal{D}$, $r_d^*(T, P) > 0$. Then, almost surely the empirical measures P_n will also be in $\mathcal{D}$ for n large enough and $T(P_n) \to T(P)$.

An open domain $\mathcal{D}$ offers the possibility that continuity can be improved to Fréchet differentiability of some order or all orders with respect to some norm metrizing weak convergence.

For some location and scatter functionals or estimators T on $\mathbb{R}^d$ for $d \geq 2$, there are η_k with $0 \leq \eta_0 \leq \eta_1 \leq \cdots \leq \eta_{d-1} < 1$ and $\eta_{d-1} > 0$ such that $T(P)$ is undefined only for some P such that there is a hyperplane H of dimension $k = 0, 1, \ldots,$ or $d - 1$, with $P(H) \geq \eta_k$. Such P form a closed, nowhere dense set F for any η_k as described, so T restricted to the complement of F satisfies (c$'$). For example, this holds for the Stahel-Donoho functional based on the median and MAD, cf. e.g. [27], with $\eta_{d-1} = 1/2 > \eta_{d-2} = 0$, and for the M-functionals based on t_ν distributions for $d \geq 2$ and $\nu > 1$ (Kent and

Tyler [21], for finite samples), with $\eta_k = (\nu + k)/(\nu + d)$.

On the other hand the median and MAD (for $d = 1$) are discontinuous on weakly dense sets and so do not satisfy (b) on any open (still less dense open) domain. This makes it hard, perhaps impossible, to verify (b) on open domains for other functionals based on the MAD or other scale functionals with the same discontinuity property, for example in scale-adjusted M-estimates of location for $d = 1$ (Huber [20, §§6.5,6.6], Rousseeuw and Croux [32]) or for $d > 1$ in the Stahel-Donoho functional, where univariate functionals μ, σ with more continuity can be used, specifically, t_ν-functionals (Tyler [38], Maronna and Yohai [27]).

Rousseeuw's minimum-volume-ellipsoid (MVE) functional can be defined for laws with $P(H)$ close to or even equal to 1 for a hyperplane H of dimension $k < d$, by restricting to H and using k-dimensional volume, as Lopuhaä and Rousseeuw [25, p. 235] suggested. But, for any $d \geq 1$, $\delta_C^*(\mathrm{MVE}, P) = 0$ at laws P with densities, by Proposition 8.

2 Collapse points

The following notion of "collapse point" is specific to scatter functionals. It and the "implosion breakdown point" defined e.g. by Rousseeuw and Croux [32], both involve mass converging toward lower-dimensional hyperplanes. But the collapse point is not defined in terms of neighborhoods N_ε (contamination or other).

Definition. If a functional $\Sigma(\cdot)$ defined on a non-empty set $\mathcal{D}$ of laws on $\mathbb{R}^d$ has values in $\mathcal{N}_d$, the *collapse point* $\kappa(\Sigma)$ is the infimum of all $y \in [0,1]$ such that there is a law Q on $\mathbb{R}^d$ with $Q(H) \leq y$ for every $(d-1)$-dimensional hyperplane H, and there exist laws $Q_k \in \mathcal{D}$ converging to Q weakly with $\det \Sigma(Q_k) \to 0$. If there is no such y set $\kappa(\Sigma) := 1$. For $d = 1$, the collapse and breakdown points of a scale functional $\sigma(\cdot)$ are defined as those of the scatter functional $\sigma^2(\cdot)$.

Remarks. For an affinely equivariant scatter functional on a non-empty domain $\mathcal{D}$, the "no such y" case cannot occur. Hampel, Ronchetti, Rousseeuw, and Stahel [17, §5.5 (a) p. 298] gave a proof that suggested those of the present section. For a comparison of statements, see the paragraph before Theorem 7.

For $d = 1$ and the classical standard deviation functional $\sigma(Q) := (\int x^2 dQ - (\int x dQ)^2)^{1/2}$, defined on the set $\mathcal{D}$ of laws Q with $\int x^2 dQ < \infty$, it's well known and easy to check that $\varepsilon_C^* \equiv 0$. It's also easy to see that the collapse point of σ is 1. For the MAD, with parameter space $[0, \infty)$, recall that $\delta_C^* = \varepsilon_C^* = 1/2$ at any law; the collapse point is also $1/2$.

It is well known that if a law Q puts high probability p in a hyperplane of dimension $< d$, and a scatter functional Σ is required to take values in $\mathcal{P}_d$, so that $\det \Sigma > 0$, then Σ can be undefined at such Q, e.g. Kent and Tyler [21], and thus have low breakdown point at laws with somewhat smaller values of p. The following shows that even allowing $\det \Sigma = 0$, there is still a tradeoff between (definition-explosion) breakdown and collapse points.

Theorem 5. *Let Σ be any affinely equivariant scatter functional with values in $\Theta = \mathcal{N}_d$ defined on a non-empty family $\mathcal{D}$ of laws on $\mathbb{R}^d$. Define its (maximum explosion-definition) breakdown point as $\delta_C^*(\Sigma) := \sup\{\delta_C^*(\Sigma, P) : P \in \mathcal{D}\}$. Then $\delta_C^*(\Sigma) + \kappa(\Sigma) \leq 1$. Moreover, for any λ with $0 < \lambda < \delta_C^*(\Sigma)$, there is a law ζ with $\zeta(H) = 1 - \lambda$ where H is a $(d-1)$-dimensional vector subspace and a sequence of laws $\zeta_k \to \zeta$ weakly with $\Sigma(\zeta_k)$ converging to a matrix with range included in H and $\det \Sigma(\zeta_k) \to 0$.*

Proof. If $\delta_C^*(\Sigma) = 0$ the conclusion holds since $\kappa(\Sigma) \leq 1$ by definition. So we can assume that for some law P and $0 < \varepsilon < \delta_C^*(\Sigma, P) \leq 1$, for $0 < \lambda < \varepsilon$ and any law Q, we have $\rho := (1-\lambda)P + \lambda Q \in \mathcal{D}$ and $\Sigma(\rho)$ remains bounded as Q varies. For any $a > 0$ let $M_a(x) := (ax_1, x_2, \ldots, x_d)'$. For any law G on $\mathbb{R}^d$ and $k = 1, 2, \ldots$, let $\rho_k := (1-\lambda)P + \lambda(G \circ M_k^{-1})$, so $\rho_k \in \mathcal{D}$, and

$$\zeta_k := \rho_k \circ M_{1/k}^{-1} = (1-\lambda)(P \circ M_{1/k}^{-1}) + \lambda G.$$

Then by affine equivariance, $\zeta_k \in \mathcal{D}$ and $\det \Sigma(\zeta_k) = \det \Sigma(\rho_k)/k^2 \to 0$. Also, ζ_k converge weakly to $\zeta := (1-\lambda)\tau + \lambda G$ where τ is a law concentrated in the hyperplane $H := \{x_1 = 0\}$. Since G is arbitrary, now let it have a density. Then clearly $\zeta(J) \leq 1 - \lambda$ for every $(d-1)$-dimensional hyperplane J. It follows that $\kappa(\Sigma) \leq 1 - \lambda$. Letting $\lambda \uparrow \varepsilon \uparrow \delta_C^*(\Sigma)$, we get $(\kappa + \delta_C^*)(\Sigma) \leq 1$. In $\Sigma(\zeta_k)$, the entries in the first row and first column go to 0 and the rest remain bounded. Thus, taking a subsequence, we can get convergence of $\Sigma(\zeta_k)$ to a limit as claimed. $\square$

By a similar proof we get a conclusion about location functionals:

Proposition 6. *Let $\mu(\cdot)$ be an affinely equivariant location functional defined on a non-empty family of laws on $\mathbb{R}$ and suppose that $\delta_C^*(\mu(\cdot)) = 1/2$. Then the domain of $\mu(\cdot)$ cannot be extended to contain any law $\frac{1}{2}(\delta_a + \delta_b)$ with $a \neq b$ and be weakly continuous at such a law.*

Proof. We can take $a = 0$ and $b = 1$. By part of the proof of Theorem 5 with $G := \delta_1$, for any $m = 1, 2, \ldots$ and $\lambda_m := \frac{1}{2} - \frac{1}{m}$, there is a sequence $\zeta_{m,k}$ of laws converging weakly as $k \to \infty$ to $(1 - \lambda_m)\delta_0 + \lambda_m\delta_1$ with $\mu(\zeta_{m,k}) \to 0$. Since weak convergence is metrizable, e.g. [11, Theorem 11.3.3], there exist $k(m) \to \infty$ as $m \to \infty$ such that as $m \to \infty$, $\zeta_{m,k(m)} \to \frac{1}{2}(\delta_0 + \delta_1)$ weakly and $\mu(\zeta_{m,k(m)}) \to 0$. But symmetrically and by affine equivariance, there also exist laws $\eta_m \to \frac{1}{2}(\delta_0 + \delta_1)$ weakly with $\mu(\eta_m) \to 1$. The conclusion follows. $\square$

A law on $\mathbb{R}^d$ will be called α-*degenerate* for $\alpha > 0$ if it puts mass at least α on some $(d-1)$-dimensional hyperplane. The first conclusion of the next theorem bounds the less-studied replacement (total variation) breakdown point at a $(1-\gamma)$-degenerate law where $0 < \gamma < 1$. The second conclusion bounds the usual contamination breakdown point at a general law F_0, e.g. a normal law, assuming the functional T is defined and continuous at a related $(1-\gamma)$-degenerate law. Such an assumption seems not to hold for many location and scatter functionals given in the literature for $\gamma \leq 1/2$, although only then is the conclusion $\varepsilon^* \leq \gamma$ of any interest. The assumption holds for M-functionals defined by t distributions with ν degrees of freedom where ν is large if γ is small, see Kent and Tyler [21], Dümbgen and Tyler [14]. Hampel et al. [17, §5.5] made such an assumption about a $((d-1)/d)$-degenerate law in proving an upper bound $1/d$ for the breakdown point of M-functionals of location and scatter. For M-functionals of scatter Maronna [26] stated an upper bound $1/(d+1)$; Tyler [36], using results in Tyler [37], gave a proof, without any assumption about α-degenerate laws. The following statement extends that of Hampel et al. (but not those of Maronna and Tyler) in that it holds for any γ, $0 < \gamma < 1$, does not use any M-functional property, and has a form applying to functionals of location alone.

Theorem 7. *Let T be an affinely equivariant location functional μ or scatter functional Σ defined on a domain $\mathcal{D}$ of laws on $\mathbb{R}^d$. For $0 \leq a < \infty$ let M_a map $\mathbb{R}^d$ into itself via $x \mapsto (ax_1, x_2, \ldots, x_d)'$. Let F_0 be any law on $\mathbb{R}^d$ and $\tilde{F}_0$ its projection into the linear subspace $H := \{x : \ x_1 = 0\}$ via M_0. Suppose* *that for some $\gamma \in (0,1)$ and law ρ on $\mathbb{R}^d$, the law $P := (1-\gamma)\tilde{F}_0+\gamma\rho \in \mathcal{D}$ and if $T = \Sigma$, $\Sigma(P)$ is non-singular, or if $T = \mu$, $\mu(P) \notin H$. Then $\varepsilon^*_R(T,P) \leq \gamma$.*

If in addition, $T(\cdot)$ is weakly continuous at P on $\mathcal{D}$, and if for all $a > 0$, $(1-\gamma)F_0 + \gamma\rho \circ M_a^{-1} \in \mathcal{D}$, then $\varepsilon^*_C(T,F_0) \leq \gamma$.

Proof. By affine equivariance, $P_a := P \circ M_a^{-1} = (1-\gamma)\tilde{F}_0 + \gamma\rho \circ M_a^{-1} \in \mathcal{D}$ for each $a > 0$, and if $T = \Sigma$, $\det\Sigma(P_a) = a^2 \det\Sigma(P) \to +\infty$ or if $T = \mu$, $|\mu(P_a)| \to +\infty$ as $a \to +\infty$. Thus, we get breakdown of T at P by replacing $\gamma\rho$ by $\gamma\rho \circ M_a^{-1}$, remaining in $\mathcal{D}$, so the first conclusion follows.

Under the further hypotheses, we have $Q_a := (1-\gamma)F_0 \circ M_{1/a}^{-1} + \gamma\rho \to P$ weakly as $a \to +\infty$, so $T(Q_a) \to T(P)$. Thus, for $T = \Sigma$,

$$\det\left(\Sigma\left((1-\gamma)F_0 + \gamma\rho \circ M_a^{-1}\right)\right) = \det\Sigma(Q_a \circ M_a^{-1}) \to +\infty.$$

For $T = \mu$, $|\mu(Q_a \circ M_a^{-1})| \to +\infty$, and the second conclusion follows. $\square$

For $\gamma = 1/(d+1)$, the hypothesis $\mu(P) \notin H$ of Theorem 7 holds by Theorem 1(d) with $n = d+1$ and $P = P_{d+1}$ an empirical measure if $P \in \mathcal{D}$.

3 Univariate trimming and the shorth

Let J be a probability density function on $[0,1]$ such that $J(y) \equiv J(1-y)$ for $0 \leq y \leq 1$ and for some $\alpha > 0$, $J(y) > 0$ if and only if $\alpha < y < 1-\alpha$. Let $\mathcal{J}$ be the law on $[\alpha, 1-\alpha]$ with density J. Most often $\mathcal{J}$ has been taken as the uniform distribution $U[\alpha, 1-\alpha]$, but J can be taken to be continuous (and so continuous a.e. for each law, e.g. Stigler [34]) or as smooth as desired (e.g. Helmers [18]).

Let Q be any law on $\mathbb{R}$ and F its distribution function. For $0 < y < 1$ let $F^{\leftarrow}(y) := \inf\{x : F(x) \geq y\}$. Let Q_J be the image measure $\mathcal{J} \circ (F^{\leftarrow})^{-1}$. Then Q_J has support in the bounded interval $[F^{\leftarrow}(\alpha), F^{\leftarrow}(1-\alpha)]$, so the *$J$-trimmed mean* of Q, i.e. the mean of Q_J,

$$\mu_J(Q) := \int_{-\infty}^{\infty} x dQ_J(x) = \int_0^1 F^{\leftarrow}(y)J(y)dy$$

and the *J-trimmed variance* of Q, i.e. the variance of Q_J,

$$\sigma_J^2(Q) := \int_{-\infty}^{\infty} x^2 dQ_J(x) - \mu_J(Q)^2$$

exist and are finite. One may multiply $\sigma_J^2(Q)$ if desired by a constant $c > 1$ so that if Q is the standard normal distribution then $c\sigma_J^2(Q) = 1$. It is straightforward to verify that with or without such a multiplication, the two functionals are defined for an arbitrary law Q on $\mathbb{R}$ and are affinely equivariant location and scatter functionals respectively, weakly continuous at all laws. The median of Q_J is always the same as the median of Q.

For the t-functionals to be considered beginning in the next section, weak continuity at all laws also holds (after an extension to allow $\sigma = 0$) but seems considerably harder to prove. So let's consider some other properties.

The functionals μ_J and σ_J^2 both have $\delta_C^* = \alpha$. The collapse point of σ_J^2 is easily seen to be $1 - 2\alpha$. For α close to $1/2$, Q_J is determined by Q in a small interval around its median (if the median is unique), which seems undesirable, as evidenced by the collapse point being close to 0. If $\alpha > 1/4$, the collapse point is less than $1/2$, which still seems unfortunate.

In some related methods of trimming, $U[\alpha, 1-\alpha]$ is replaced by $U[\beta, 1-\gamma]$ where $\beta, \gamma \geq 0$ and $\beta + \gamma = 2\alpha$. Here β and γ can be chosen to minimize the variance of the resulting distribution [3], the distance of points in its support from the median [23], or otherwise. Location and scale functionals based on such trimming will still give collapse point $1 - 2\alpha$ and can increase δ_C^* to 2α, for $\alpha < 1/4$. Asymmetric trimming works well to prune asymmetric outlier contamination from a symmetric true distribution [23], but apparently not so well for an asymmetric true distribution.

There are various multivariate extensions of trimming, e.g. Donoho and Gasko [10] and Liu, Parelius and Singh [24]. But, by Theorem 1(d) and Theorem 3(b), no method can have the same complete success in defining a location functional as in one dimension.

The shorth and LMS functionals. Let $0 < \alpha < 1$. Let P be a law on $\mathbb{R}$ and $\sigma_{\mathrm{Sh},\alpha}(P) := \inf\{b - a : P([a, b]) \geq \alpha\}$. Then there are always some a, b with $P([a, b]) \geq \alpha$ and $b - a = h := \sigma_{\mathrm{Sh},\alpha}(P)$. Let $I_\alpha(P)$ be the set of such intervals $[a, b]$. Note that for each such $[a, b]$ and $\varepsilon > 0$, $P([a, a+\varepsilon]) > 0$ and $P([b - \varepsilon, b]) > 0$. Let $K_\alpha(P)$ denote the set of all conditional means $\int_a^b x dP/P([a, b])$ for $[a, b] \in I_\alpha(P)$ and $M_\alpha(P)$ the set of midpoints $(a + b)/2$. Then $K_\alpha(P)$ and $M_\alpha(P)$ are compact, nonempty sets. If $I_\alpha(P)$ consists of just one interval $[a, b]$, let $\mu_{\mathrm{Sh},\alpha}(P) := \mu$ and $m_{\mathrm{Sh},\alpha}(P) := (a + b)/2$, which for $\alpha = 1/2$ Davies [7, p. 1856] calls the "middle of the shortest half" functional; Rousseeuw and Leroy [33, p. 169] call it the "least median of squares" (LMS) functional, specializing a form of regression. Also, $\mu_{\mathrm{Sh},\alpha}(P)$ is called the α-*shorth* of P and $\mu_{\mathrm{Sh}}(P) := \mu_{\mathrm{Sh},1/2}(P)$ is the *shorth* of P.

The LMS location functional $m_{\mathrm{Sh},1/2}$ is the location part of the functional limit of the univariate case of Rousseeuw's [30] minimum-volume ellipsoid (MVE) location-scatter estimator (μ, Σ), which for a finite sample of size n in $\mathbb{R}^d$, finds an ellipsoid $\{x : (x-\mu)'\Sigma^{-1}(x-\mu) \le c\}$ of smallest volume containing $[n/2]+1$ of the observations (or, $[(n+d+1)/2]$ observations, to maximize the finite-sample breakdown point, e.g. Rousseeuw and Leroy [33, p. 264]). Davies [5] briefly notes that although an MVE can be selected uniquely from its set of possible values, there is a dense set of laws for which the MVE is not unique and "no affine equivariant choice can be made." The following fact, then, is to some degree known, but it gives strong forms of denseness. Parts (a) and (c) of the following give contamination neighborhoods $N_\varepsilon^C(P)$, which are included in total variation neighborhoods and in turn in neighborhoods for weak convergence.

Proposition 8. *(a) For any law P on $\mathbb{R}$ having a continuous density f and any $\varepsilon > 0$, there is a law $\zeta \in N_\varepsilon^C(P)$, also with a continuous density, for which $I_{1/2}(\zeta)$ contains more than one interval and so $\mu_{Sh}(\zeta)$ and $m_{Sh,1/2}(\zeta)$ are not defined. Thus $\delta_C^*(m_{Sh,1/2}, P) = 0$, also if contamination neighborhoods are replaced by any larger neighborhoods.*
(b) For any α with $0 < \alpha < 1$ there exist laws P symmetric about a point $m \notin K_\alpha(P)$. For such P there is no way to select $\mu^{(\alpha)} \in K_\alpha(P)$, nor as a midpoint of any $[a,b] \in I_\alpha(P)$, to get an affine equivariant location functional $\mu^{(\alpha)}(\cdot)$.
(c) Let F_0 be any distribution on $\mathbb{R}$ having a strictly unimodal density f_0 with $f_0(-x) \equiv f_0(x)$. Then for any $\lambda > 0$ there is a $P \in N_\lambda^C(F_0)$ satisfying (b) for $\alpha = 1/2$.

Proof. (a) Let P have a continuous density f. If $I_{1/2}(P)$ contains more than one interval we are done, so suppose $I_{1/2}(P)$ contains just one interval $[a,b]$, which we can assume is $[0,1]$. Thus $\int_x^{x+1} f(u)du < 1/2$ for $x \neq 0$. Take any δ with $0 < \delta < 1$. Another continuous density g will be defined as follows. Let $g(x) = 1/2$ for $0 \le x \le 1$ and $h_\delta := (1-\delta)f + \delta g$. Then $\int_0^1 h_\delta(x)dx = 1/2$. For $x > 0$ let

$$g_\delta(x) := \frac{1-\delta}{\delta}[f(x) - f(x+1)] + \frac{1}{2}.$$

We have $(d/dx)\int_x^{x+1} f(u)du = f(x+1) - f(x) = 0$ when $x = 0$, so $f(0) = f(1)$. There is a $\gamma > 0$ such that $\gamma < 1/2$ and $(1-\delta)[f(u+1) - f(u)] \le \delta/2$ for $0 \le u \le \gamma$. Choose a $\beta > 0$ small enough so that $g_\delta(u) > 0$ for $0 \le u \le \beta$ and $\int_1^{1+\beta} g_\delta(u-1)du < \gamma/2$. Define $g(1+x) := g_\delta(x)$ for $0 < x \le \beta/2$, $g(1+x) := 2(\beta - x)\beta^{-1}g_\delta(x) < g_\delta(x)$ for $\beta/2 < x \le \beta$, and $g(1+x) := 0$ for $x > \beta$. For $0 < x < \beta/2$ we have

$$\frac{d}{dx}\int_x^{1+x} h_\delta(u)du \;=\; (1-\delta)[f(x+1) - f(x)] + \delta\left[g_\delta(x) - \frac{1}{2}\right] \;=\; 0,$$

so $\int_x^{1+x} h_\delta(u)du = 1/2$ for $0 \le x \le \beta/2$. For $\beta/2 < x \le \beta$ we then have by definition of $g(1+x)$

$$\int_x^{1+x} h_\delta(u)du < 1/2. \tag{4}$$

For $x > \gamma$, $\int_x^{1+x} g(u)du < 1/2$ (for $x \le 1$ or $x > 1$), so (4) holds. To prove (4) for $\beta < x \le \gamma$ it suffices to show $\int_x^{1+x} h_\delta(u)du \le \int_\beta^{1+\beta} h_\delta(u)du$, or $[\int_\beta^x - \int_{1+\beta}^{1+x}]h_\delta(u) \ge 0$, or $\int_\beta^x (1-\delta)[f(u) - f(1+u)] + \frac{\delta}{2}du \ge 0$, which holds by choice of γ.

Since $\int_0^\infty g(u)du = \int_0^{1+\beta} g(u)du < 3/4$ by choice of β and γ, we can and do define $g(u)$ for $u < 0$ to be nonnegative and continuous at 0 with $g(u) < 1/2$ for all $u < 0$ such that $\int_{-\infty}^\infty g(u)du = 1$. Then g and h_δ are both probability densities. For any $x < 0$, $\int_x^{1+x} h_\delta(u)du < \frac{1}{2}(1 - \delta + \delta)$. So (4) holds for all $x \notin [0, \beta/2]$ while $\int_x^{1+x} h_\delta(u)du = 1/2$ for $0 \le x \le \beta/2$. Thus for ζ with density h_δ, we have $\sigma_{\mathrm{Sh},1/2}(\zeta) = 1$ and $I_{1/2}(\zeta) = \{[x, x+1] :\ 0 \le x \le \beta/2\}$. Now $\int_x^{1+x} uh_\delta(u)du/\int_x^{1+x} h_\delta(u)du = 2\int_x^{1+x} uh_\delta(u)du$ is a strictly increasing function of x for $0 \le x \le \beta/2$, so (a) is proved.

For (b), if P exists, the conclusions follow from Theorem 1(c). To show P exists, for $0 < \alpha \le 1/2$, let $P = \frac{1}{2}(U[0,1] + U[3,4])$. For $1/2 < \alpha < 1$, let $P = (2\alpha - 1)\delta_0 + (1-\alpha)(\delta_{-1} + \delta_1)$. Then $m = 0$, $\sigma_{\mathrm{Sh},\alpha}(P) = 1$, and $0 \notin K_\alpha(P) = \{\pm(1-\alpha)\}$, proving (b).

For (c), $I_{1/2}(F_0)$ contains just one interval, namely $[-\xi, \xi]$, where ξ is the upper quartile of F_0. For any $\delta > 0$ let Q_δ be the law with density f_δ where $f_\delta(-x) \equiv f_\delta(x)$, $f_\delta(\xi + t) = t/\delta^2$ for $0 \le t \le \delta$ and $f_\delta(x) = 0$ for all other $x > 0$. For fixed $\lambda \in (0,1)$ and $P_\delta := (1-\lambda)F_0 + \lambda Q_\delta$, the unique interval $[-\eta, \eta]$ with $P_\delta([-\eta, \eta]) = 1/2$ has length $2\xi + \sqrt{2}\delta + o(\delta)$ as $\delta \downarrow 0$. But $P_\delta([-\xi, \xi + \delta]) > 1/2$ for a shorter interval, proving (c) and the proposition. $\square$

References

[1] Bassett, G. W. (1991). Equivariant, monotonic, 50% breakdown estimators. *American Statistician* **45**, 135-137.

[2] Bickel, P. J., and Lehmann, E. L. (1975). Descriptive statistics for nonparametric models. I, Introduction; II, location. *Ann. Statist.* **3**, 1038-1044, 1045-1069.

[3] Butler, R. W. (1982). Nonparametric interval and point prediction using data trimmed by a Grubbs-type outlier rule. *Ann. Statist.* **10**, 197-204.

[4] Clarke, B. R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist.* **11**, 1196-1205.

[5] Davies, [P.] L. (1992a). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20**, 1828-1843.

[6] Davies, [P.] L. (1992b). An efficient Fréchet differentiable high breakdown location and dispersion estimator. *J. Multivariate Analysis* **40**, 311-327.

[7] Davies, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21**, 1843-1899.

[8] Davies, [P.] L. (1994). Desirable properties, breakdown and efficiency in the linear regression model. *Statistics and Probability Letters* **19**, 361-370.

[9] Davies, P. L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *Ann. Statist.* **26**, 1103-1125.

[10] Donoho, D. L., and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20**, 1803-1827.

[11] Dudley, R. M. (2002). *Real Analysis and Probability*, 2d ed. Cambridge University Press.

[12] Dümbgen, L. (1997). The asymptotic behavior of Tyler's M-estimator of scatter in high dimension. Preprint.

[13] Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Ann. Inst. Statist. Math.* **50**, 471-491.

[14] Dümbgen, L., and Tyler, D. (2004). On the breakdown properties of some multivariate M-functionals. Preprint.

[15] Ferguson, T. S. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. *J. Amer. Statist. Assoc.* **73**, 211-213.

[16] Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42**, 1887-1896.

[17] 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.

[18] Helmers, R. (1980). Edgeworth expansions for linear combinations of order statistics with smooth weight functions. *Ann. Statist.* **8**, 1361-1374.

[19] Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probability* **1**, 221-233. University of California Press, Berkeley and Los Angeles.

[20] Huber, P. J. (1981). *Robust Statistics.* Wiley, New York. Reprinted, 2004.

[21] Kent, J. T., and Tyler, D. E. (1991). Redescending M-estimates of multivariate location and scatter. *Ann. Statist.* **19**, 2102-2119.

[22] Kent, J. T., Tyler, D. E., and Vardi, Y. (1994). A curious likelihood identity for the multivariate t-distribution. *Commun. Statist.–Simula.* **23**, 441-453.

[23] Kim, Seong-Ju (1992). The metrically trimmed mean as a robust estimator of location. *Ann. Statist.* **20**, 1534-1547.

[24] Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **27**, 783-858.

[25] Lopuhaä, H. P., and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19**, 229-248.

[26] Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4**, 51-67.

[27] Maronna, R. A., and Yohai, V. J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.* **90**, 330-341.

[28] Milasevic, P. and Ducharme, G. R. (1987). Uniqueness of the spatial median. *Ann. Statist.* **15**, 1332-1333.

[29] Obenchain, R. L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42**, 1569-1578.

[30] Rousseeuw, P. J. (1986). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications*, Vol. B (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, eds.). Dordrecht, Reidel.*

[31] Rousseeuw, P. J. (1994). Unconventional features of positive-breakdown estimators. *Statistics and Probability Letters* **19**, 417-431.

[32] Rousseeuw, P. J., and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc.* **88**, 1273-1283.

[33] Rousseeuw, P. J., and Leroy, A. *Robust Regression and Outlier Detection.* Wiley, New York.

[34] Stigler, S. M. (1974). Linear functions of order statistics with smooth weight functions. *Ann. Statist.* **2**, 676-693; corr. *ibid.* **7** (1979), 466.

[35] Tatsuoka, K. S., and Tyler, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.* **28**, 1219-1243.

[36] Tyler, D. E. (1986). Breakdown properties of the M-estimators of multivariate scatter. Technical Report, Rutgers University.

[37] Tyler, D. E. (1988). Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter. *SIAM J. on Scientific and Statistical Computing* **9**, 354-362.

[38] Tyler, D. E. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22**, 1024-1044.

* The author has seen the Rousseeuw (1986) item cited in secondary sources but not in the original.