

DIFFERENTIABILITY OF t -FUNCTIONALS OF LOCATION AND SCATTER

BY R. M. DUDLEY,¹ SERGIY SIDENKO² AND ZUOQIN WANG²

Massachusetts Institute of Technology

The paper aims at finding widely and smoothly defined nonparametric location and scatter functionals. As a convenient vehicle, maximum likelihood estimation of the location vector μ and scatter matrix Σ of an elliptically symmetric t distribution on \mathbb{R}^d with degrees of freedom $\nu > 1$ extends to an M -functional defined on all probability distributions P in a weakly open, weakly dense domain U . Here U consists of P putting not too much mass in hyperplanes of dimension $< d$, as shown for empirical measures by Kent and Tyler [*Ann. Statist.* **19** (1991) 2102–2119]. It will be seen here that (μ, Σ) is analytic on U for the bounded Lipschitz norm, or for $d = 1$ for the sup norm on distribution functions. For $k = 1, 2, \dots$, and other norms, depending on k and more directly adapted to t functionals, one has continuous differentiability of order k , allowing the delta-method to be applied to (μ, Σ) for any P in U , which can be arbitrarily heavy-tailed. These results imply asymptotic normality of the corresponding M -estimators (μ_n, Σ_n) . In dimension $d = 1$ only, the t_ν functional (μ, σ) extends to be defined and weakly continuous at all P .

1. Introduction. This paper aims at developing some nonparametric location and scatter functionals, defined and smooth on large (weakly dense and open) sets of distributions. The nonparametric view is much as in the work of Bickel and Lehmann (1975) (but not adopting, e.g., their monotonicity axiom) and to a somewhat lesser extent, that of Davies (1998). Although there are relations to robustness, that is not the main aim here: there is no focus on neighborhoods of model distributions with densities such as the normal. It happens that the parametric family of ellipsoidally symmetric t densities provides an avenue toward nonparametric location and scatter functionals, somewhat as maximum likelihood estimation of location for the double-exponential distribution in one dimension gives the median, generally viewed as a nonparametric functional.

Given observations X_1, \dots, X_n in \mathbb{R}^d , let $P_n := \frac{1}{n} \sum_{j=1}^n \delta_{X_j}$. Given P_n , and the location-scatter family of elliptically symmetric t_ν distributions on \mathbb{R}^d with $\nu > 1$, maximum likelihood estimates of the location vector μ and scatter matrix Σ exist and are unique for “most” P_n . Namely, it suffices that $P_n(J) < (\nu + q)/(\nu + d)$

Received April 2007.

¹Supported in part by NSF Grants DMS-01-03821 and DMS-05-04859.

²Supported in part by NSF Grant DMS-05-04859.

AMS 2000 subject classifications. Primary 62G05, 62GH20; secondary 62G35.

Key words and phrases. Affinely equivariant, Fréchet differentiable, weakly continuous.

for each affine hyperplane J of dimension $q < d$, as shown by Kent and Tyler (1991). The estimates extend to M-functionals defined at all probability measures P on \mathbb{R}^d satisfying the same condition; that is shown for integer ν and in the sense of unique critical points by Dümbgen and Tyler (2005) and for general $\nu > 0$ and M-functionals in the sense of unique absolute minima in Theorem 3, using Theorem 6(a), for pure scatter and then in Theorem 6(e) for location and scatter with $\nu > 1$. A method of reducing location and scatter functionals in dimension d to pure scatter functionals in dimension $d + 1$ was shown to work for t distributions by Kent and Tyler (1991) and only for such distributions by Kent, Tyler and Vardi (1994) as will be recalled after Theorem 6.

So the t functionals are defined on a weakly open and weakly dense domain, whose complement is thus weakly nowhere dense. One of the main results of the present paper gives analyticity (defined in the Appendix) of the functionals on this domain, with respect to the bounded Lipschitz norm [Theorem 11(d)]. An adaptation gives differentiability of any given finite order k with respect to norms, depending on k , chosen to give asymptotic normality of the t location and scatter functionals (Theorem 14), for arbitrarily heavy-tailed P (for such P , the central limit fails in the bounded Lipschitz norm). In turn, this yields delta-method conclusions [Theorem 7(b)], uniformly over suitable families of distributions (Proposition 8); these statements do not include any norms, although their proofs do. It follows in Corollary 15 that continuous Fréchet differentiability of the t_ν location and scatter functionals of order k also holds with respect to affinely invariant norms defined via suprema over positivity sets of polynomials of degree at most $2k + 4$.

For the delta-method, one needs at least differentiability of first order. To get first derivatives with respect to probability measures P via an implicit function theorem we use second-order derivatives with respect to matrices. Moreover, second-order derivatives with respect to P (or in the classical case, with respect to an unknown parameter) can improve the accuracy of the delta-method and the speed of convergence of approximations. It turns out that derivatives of arbitrarily high order are obtainable with little additional difficulty.

For norms in which the central limit theorem for empirical measures holds for all probability measures, such as those just mentioned, bootstrap central limit theorems also hold [Giné and Zinn (1990)], which then via the delta-method can give bootstrap confidence sets for the t location and scatter functionals.

In dimension $d = 1$, the domain on which differentiability is proved is the class of distributions having no atom of size $\nu/(\nu + 1)$ or larger. On this domain, analyticity holds, Theorem 11(e), with respect to the usual supremum norm for distribution functions. Also, only for $d = 1$, the t_ν location and scatter (scale) functionals extend to be defined and weakly continuous at arbitrary distributions (having arbitrarily large atoms, Theorem 12).

Weak continuity on a dense open set implies that for distributions in that set, estimators (functionals of empirical measures) eventually exist almost surely and converge to the functional of the distribution. Weak continuity, where it holds, also

is a robustness property in itself and implies a strictly positive (not necessarily large) breakdown point. The t_ν functionals, as re-descending M-functionals, down-weight outliers. Among such M-functionals, only the t_ν functionals are known to be uniquely defined on a satisfactorily large domain.

A probability measure P on \mathbb{R}^d is said to be *symmetric* around a point m if the map $x \mapsto 2m - x$ preserves P . The t_ν estimators are \sqrt{n} -consistent estimators of t_ν functionals where each t_ν location functional, at any distribution in its domain and symmetric around a point, (by equivariance) equals the center of symmetry.

It seems that few other known location and scatter functionals exist and are unique and continuous, let alone differentiable, on a dense open domain. For example, the median is discontinuous on a dense set. Smoothly trimmed means and variances are defined and differentiable at all distributions in one dimension; see for example, Boos (1979) for the means.

In higher dimensions there are analogues of trimming, called peeling or depth weighting; see for example, the recent work Zuo and Cui (2005). Location-scatter functionals differentiable on a dense domain apparently have not been found by depth weighting thus far (in dimension $d > 1$).

The t location and scatter functionals, on their domain, can be effectively computed via EM algorithms [cf. Kent, Tyler and Vardi (1994), Section 4; Arslan, Constable and Kent (1995); Liu, Rubin and Wu (1998)].

Proofs of most facts stated in the paper are given in the longer version Dudley, Sidenko and Wang (2008). Some are given in Dudley et al. (2007).

2. Definitions and preliminaries. In this paper the sample space will be a finite-dimensional Euclidean space \mathbb{R}^d with its usual topological and Borel structure. A *law* will mean a probability measure on \mathbb{R}^d . Let \mathcal{S}_d be the collection of all $d \times d$ symmetric real matrices, \mathcal{N}_d the subset of nonnegative definite symmetric matrices and $\mathcal{P}_d \subset \mathcal{N}_d$ the further subset of strictly positive definite symmetric matrices. The parameter spaces Θ considered will be \mathcal{P}_d , \mathcal{N}_d (pure scatter matrices), $\mathbb{R}^d \times \mathcal{P}_d$, or $\mathbb{R}^d \times \mathcal{N}_d$. For $(\mu, \Sigma) \in \mathbb{R}^d \times \mathcal{N}_d$, μ will be viewed as a location parameter and Σ as a scatter parameter, extending the notions of mean vector and covariance matrix to arbitrarily heavy-tailed distributions. Matrices in \mathcal{N}_d but not in \mathcal{P}_d will only be considered in one dimension, in Section 6, where the scale parameter $\sigma \geq 0$ corresponds to $\sigma^2 \in \mathcal{N}_1$.

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

DEFINITIONS. 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* if and only if 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 $0 \leq \sigma < \infty$ will be called an *affinely*

equivariant scale functional if and only if 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} .

To define M-functionals, suppose we have a function $(x, \theta) \mapsto \rho(x, \theta)$ defined for $x \in \mathbb{R}^d$ and $\theta \in \Theta$, Borel measurable in x and lower semicontinuous in θ , that is, $\rho(x, \theta) \leq \liminf_{\phi \rightarrow \theta} \rho(x, \phi)$ for all θ . For a law Q , let $Q\rho(\phi) := \int \rho(x, \phi) dQ(x)$ if the integral is defined (not $\infty - \infty$), as it always will be if $Q = P_n$. An *M-estimate* of θ for a given n and P_n will be a $\hat{\theta}_n$ such that $P_n\rho(\theta)$ is minimized at $\theta = \hat{\theta}_n$, if it exists and is unique. A measurable function, not necessarily defined a.s., whose values are M-estimates is called an M-estimator.

For a law P on \mathbb{R}^d and a given $\rho(\cdot, \cdot)$, a $\theta_1 = \theta_1(P)$ is called the *M-functional* of P for ρ if and only if there exists a measurable function $a(x)$, called an *adjustment function*, such that for $h(x, \theta) = \rho(x, \theta) - a(x)$, $Ph(\theta)$ is defined and satisfies $-\infty < Ph(\theta) \leq +\infty$ for all $\theta \in \Theta$, and is minimized uniquely at $\theta = \theta_1(P)$; see, for example, Huber (1967, 1981). As Huber showed, $\theta_1(P)$ does not depend on the choice of $a(\cdot)$, which can moreover be taken as $a(x) \equiv \rho(x, \theta_2)$ for a suitable θ_2 .

The following definition will be used for $d = 1$. Suppose we have a parameter space Θ , specifically \mathcal{P}_d or $\mathcal{P}_d \times \mathbb{R}^d$, which has a closure $\overline{\Theta}$, specifically \mathcal{N}_d or $\mathcal{N}_d \times \mathbb{R}^d$, respectively. The *boundary* of Θ is then $\overline{\Theta} \setminus \Theta$. The functions ρ and h are not necessarily defined for θ in the boundary, but M-functionals may have values anywhere in $\overline{\Theta}$ according to the following.

DEFINITION. A $\theta_0 = \theta_0(P) \in \overline{\Theta}$ will be called the (extended) *M-functional* of P for ρ or h if and only if for every neighborhood U of θ_0 ,

$$(1) \quad -\infty \leq \liminf_{\phi \rightarrow \theta_0, \phi \in \Theta} Ph(\phi) < \inf_{\phi \in \Theta, \phi \notin U} Ph(\phi).$$

The above definition extends that of M-functional given by Huber (1967) in that if θ_0 is on the boundary of $\overline{\Theta}$, then $h(x, \theta_0)$ is not defined, $Ph(\theta_0)$ is defined only in a lim inf sense, and at θ_0 (but only there), the lim inf may be $-\infty$.

From the definition, an M-functional, if it exists, must be unique. If P is an empirical measure P_n , then the M-functional $\hat{\theta}_n := \theta_0(P_n)$, if it exists, is the maximum likelihood estimate of θ , in a lim sup sense if $\hat{\theta}_n$ is on the boundary. Clearly, an M-estimate $\hat{\theta}_n$ is the M-functional $\theta_1(P_n)$ if either exists.

For a differentiable function f , recall that a *critical point* of f is a point where the gradient of f is 0. For example, on \mathbb{R}^2 let $f(x, y) = x^2(1 + y)^3 + y^2$. Then f has a unique critical point $(0, 0)$, which is a strict relative minimum, but these conditions do not suffice for an absolute minimum since $f(1, y) \rightarrow -\infty$ as $y \rightarrow -\infty$. This example appeared in Durfee et al. (1993). Thus we will need to check global as well as local conditions.

3. Multivariate scatter. This section will treat the pure scatter problem in \mathbb{R}^d , with parameter space $\Theta = \mathcal{P}_d$. The results here are extensions of those of Kent and Tyler (1991), Theorems 2.1 and 2.2, on unique maximum likelihood estimates for finite samples, to the case of M-functionals for general laws on \mathbb{R}^d .

For $A \in \mathcal{P}_d$ and a function ρ from $[0, \infty)$ into itself, consider the function

$$(2) \quad L(y, A) := \frac{1}{2} \log \det A + \rho(y' A^{-1} y), \quad y \in \mathbb{R}^d.$$

For adjustment, let

$$(3) \quad h(y, A) := L(y, A) - L(y, I),$$

where I is the identity matrix. Then

$$(4) \quad Qh(A) = \frac{1}{2} \log \det A + \int (\rho(y' A^{-1} y) - \rho(y' y)) dQ(y)$$

if the integral is defined.

As a referee suggested, one can differentiate functions of matrices in a coordinate-free way, as follows. The d^2 -dimensional vector space of all $d \times d$ real matrices becomes a Hilbert space (Euclidean space) under the inner product $\langle A, B \rangle := \text{trace}(A'B)$. It is easy to verify that this is indeed an inner product and is invariant under orthogonal changes of coordinates in the underlying d -dimensional vector space. The corresponding norm $\|A\|_F := \langle A, A \rangle^{1/2}$ is called the Frobenius norm. Here $\|A\|_F^2$ is simply the sum of squares of all elements of A , and $\|\cdot\|_F$ is the specialization of the (Hilbert)–Schmidt norm for Hilbert–Schmidt operators on a general Hilbert space to the case of (all) linear operators on a finite-dimensional Hilbert space. Let $\|\cdot\|$ be the usual matrix or operator norm, $\|A\| := \sup_{|x|=1} |Ax|$. Then

$$(5) \quad \|A\| \leq \|A\|_F \leq \sqrt{d} \|A\|,$$

with equality in the latter for $A = I$ and the former when $A = \text{diag}(1, 0, \dots, 0)$. In statements such as $\|A\| \rightarrow 0$ or expressions such as $O(\|A\|)$ the particular norm does not matter for fixed d .

We have the following, stated for $Q = Q_n$ an empirical measure in Kent and Tyler (1991), equation (1.3). Here (6) is a redescending condition.

PROPOSITION 1. *Let $\rho : [0, \infty) \rightarrow [0, \infty)$ be continuous and have a bounded continuous derivative on $[0, \infty)$, where $\rho'(0) := \rho'(0+) := \lim_{x \downarrow 0} [\rho(x) - \rho(0)]/x$. Let $0 \leq u(x) := 2\rho'(x)$ for $x \geq 0$ and suppose that*

$$(6) \quad \sup_{0 \leq x < \infty} x u(x) < \infty.$$

Then for any law Q on \mathbb{R}^d , Qh in (4) is a well-defined and C^1 function of A , which has a critical point at $A = B$ if and only if

$$(7) \quad B = \int u(y' B^{-1} y) y y' dQ(y).$$

In the proof of Proposition 1, it is shown that

$$(8) \quad \text{For any compact } \mathcal{K} \subset \mathcal{P}_d, \quad \sup\{|h(y, A)| : y \in \mathbb{R}^d, A \in \mathcal{K}\} < \infty.$$

The following extends to any law Q the uniqueness part of Kent and Tyler (1991), Theorem 2.2.

PROPOSITION 2. *Under the hypotheses of Proposition 1 on ρ and $u(\cdot)$, if in addition $u(\cdot)$ is nonincreasing and $s \mapsto su(s)$ is strictly increasing on $[0, \infty)$, then for any law Q on \mathbb{R}^d , Qh has at most one critical point $A \in \mathcal{P}_d$.*

Our proof of the preceding [Dudley, Sidenko and Wang (2008)] shows that existence of critical points implies that Q must not be concentrated in any proper linear subspace. More precisely, a sufficient condition for existence of a minimum (which will then be unique by Proposition 2) will include the following assumption from Kent and Tyler (1991), equation (2.4). For a given function $u(\cdot)$ as in Proposition 2, let $a_0 := a_0(u(\cdot)) := \sup_{s>0} su(s)$. Since $s \mapsto su(s)$ is increasing, we will have

$$(9) \quad su(s) \uparrow a_0 \quad \text{as } s \uparrow +\infty.$$

Kent and Tyler (1991) gave the following conditions for empirical measures.

DEFINITION. For a given number $a_0 := a(0) > 0$ let $\mathcal{U}_{d,a(0)}$ be the set of all probability measures Q on \mathbb{R}^d such that for every linear subspace H of dimension $q \leq d - 1$, $Q(H) < 1 - (d - q)/a_0$, so that $Q(H^c) > (d - q)/a_0$.

If $Q \in \mathcal{U}_{d,a(0)}$, then $Q(\{0\}) < 1 - (d/a_0)$, which is impossible if $a_0 \leq d$. So we will need $a_0 > d$ and assume it, for example, in the following theorem. In the t_ν case later we will have $a_0 = \nu + d > d$ for any $\nu > 0$. For $a(0) > d$, $\mathcal{U}_{d,a(0)}$ is weakly open and dense and contains all laws with densities. For part (a), Tyler (1988) gave a proof that extends to the present case. In part (b), Kent and Tyler (1991), Theorems 2.1 and 2.2, proved that there is a unique $B(Q_n)$ minimizing $Q_n h$ for an empirical $Q_n \in \mathcal{U}_{d,a(0)}$.

THEOREM 3. *Let $u(\cdot) \geq 0$ be a bounded continuous function on $[0, \infty)$ satisfying (6), with $u(\cdot)$ nonincreasing and $s \mapsto su(s)$ strictly increasing. Then for $a(0) = a_0$ as in (9):*

- (a) *If $Q \notin \mathcal{U}_{d,a(0)}$, then Qh has no critical points.*
- (b) *If $a_0 > d$ and $Q \in \mathcal{U}_{d,a(0)}$, then Qh attains its minimum at a unique $B = B(Q) \in \mathcal{P}_d$ and has no other critical points.*

4. Location and scatter t functionals. The main result of this section, Theorem 6, is an extension of results of Kent and Tyler [(1991), Theorem 3.1] who found maximum likelihood estimates for finite samples, and Dümbgen and Tyler (2005) for M-functionals, defined as unique critical points, for integer ν , to the case of M-functionals in the sense of absolute minima, and any $\nu > 0$.

Kent and Tyler [(1991), Section 3] and Kent, Tyler and Vardi (1994) showed that location-scatter problems in \mathbb{R}^d can be treated by way of pure scatter problems in \mathbb{R}^{d+1} , specifically for functionals based on t log likelihoods. The two papers prove the following:

PROPOSITION 4. (i) For any $d = 1, 2, \dots$, there is a 1–1 correspondence between matrices $A \in \mathcal{P}_{d+1}$ and triples (Σ, μ, γ) where $\Sigma \in \mathcal{P}_d$, $\mu \in \mathbb{R}^d$, and $\gamma > 0$, given by $A = A(\Sigma, \mu, \gamma)$ where

$$(10) \quad A(\Sigma, \mu, \gamma) = \gamma \begin{bmatrix} \Sigma + \mu\mu' & \mu \\ \mu' & 1 \end{bmatrix}.$$

The correspondence is C^∞ in either direction.

(ii) For $A = A(\Sigma, \mu, \gamma)$, we have

$$(11) \quad A^{-1} = \gamma^{-1} \begin{bmatrix} \Sigma^{-1} & -\Sigma^{-1}\mu \\ -\mu'\Sigma^{-1} & 1 + \mu'\Sigma^{-1}\mu \end{bmatrix}.$$

(iii) If (10) holds, then for any $y \in \mathbb{R}^d$ (a column vector),

$$(12) \quad (y', 1)A^{-1}(y', 1)' = \gamma^{-1}(1 + (y - \mu)'\Sigma^{-1}(y - \mu)).$$

For M-estimation of location and scatter in \mathbb{R}^d , we will have a function $\rho: [0, \infty) \mapsto [0, \infty)$ as in the previous section. The parameter space is now the set of pairs (μ, Σ) for $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$, and we have a multivariate ρ function (the two meanings of ρ should not cause confusion)

$$\rho(y, (\mu, \Sigma)) := \frac{1}{2} \log \det \Sigma + \rho((y - \mu)'\Sigma^{-1}(y - \mu)).$$

For any $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$ let $A_0 := A_0(\mu, \Sigma) := A(\Sigma, \mu, 1) \in \mathcal{P}_{d+1}$ by (10) with $\gamma = 1$, noting that $\det A_0 = \det \Sigma$. Now ρ can be adjusted, in light of (8) and (12), by defining

$$(13) \quad h(y, (\mu, \Sigma)) := \rho(y, (\mu, \Sigma)) - \rho(y'y).$$

Laws P on \mathbb{R}^d correspond to laws $Q := P \circ T_1^{-1}$ on \mathbb{R}^{d+1} concentrated in $\{y: y_{d+1} = 1\}$, where $T_1(y) := (y', 1)' \in \mathbb{R}^{d+1}$, $y \in \mathbb{R}^d$. We will need a hypothesis on P corresponding to $Q \in \mathcal{U}_{d+1, a(0)}$. Kent and Tyler (1991) gave these conditions for empirical measures.

DEFINITION. For any $a_0 := a(0) > 0$ let $\mathcal{V}_{d, a(0)}$ be the set of all laws P on \mathbb{R}^d such that for every affine hyperplane J of dimension $q \leq d - 1$, $P(J) < 1 - (d - q)/a_0$, so that $P(J^c) > (d - q)/a_0$.

The next fact is rather straightforward to prove.

PROPOSITION 5. *For any law P on \mathbb{R}^d , $a > d + 1$, and $Q := P \circ T_1^{-1}$ on \mathbb{R}^{d+1} , we have $P \in \mathcal{V}_{d,a}$ if and only if $Q \in \mathcal{U}_{d+1,a}$.*

For laws $P \in \mathcal{V}_{d,a(0)}$ with $a(0) > d + 1$, one can prove that there exist $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$ at which $Ph(\mu, \Sigma)$ is minimized, as [Kent and Tyler \(1991\)](#) did for empirical measures, by applying part of the proof of [Theorem 3](#) restricted to the closed set where $\gamma = A_{d+1,d+1} = 1$ in (12). But the proof of uniqueness ([Proposition 2](#)) does not apply in general under the constraint $A_{d+1,d+1} = 1$. For minimization under a constraint the notion of critical point changes, for example, for a Lagrange multiplier λ one would seek critical points of $Qh(A) + \lambda(A_{d+1,d+1} - 1)$, so [Propositions 1 and 2](#) no longer apply. Uniqueness will hold under an additional condition. A family of ρ functions that will satisfy the condition, as pointed out by [Kent and Tyler \(1991\)](#), equations (1.5) and (1.6), comes from elliptically symmetric multivariate t densities with ν degrees of freedom as follows: for $0 < \nu < \infty$ and $0 \leq s < \infty$ let

$$(14) \quad \rho_\nu(s) := \rho_{\nu,d}(s) := \frac{\nu + d}{2} \log\left(\frac{\nu + s}{\nu}\right).$$

For this ρ , u is $u_\nu(s) := u_{\nu,d}(s) := (\nu + d)/(\nu + s)$, which is decreasing, and $s \mapsto su_{\nu,d}(s)$ is strictly increasing and bounded, so that (6) holds, with supremum and limit at $+\infty$ equal to $a_{0,\nu} := a_0(u_\nu(\cdot)) = \nu + d$ for any $\nu > 0$.

The following fact was shown in part by [Kent and Tyler \(1991\)](#) and further by [Kent, Tyler and Vardi \(1994\)](#) for empirical measures, with a short proof, and with equation (15) only implicit. The relation that ν degrees of freedom in dimension d correspond to $\nu' = \nu - 1$ in dimension $d + 1$, due to [Kent, Tyler and Vardi \(1994\)](#), is implemented more thoroughly in the following theorem and the proof in [Dudley \(2006\)](#). The extension from empirical to general laws follows from [Theorem 3](#), specifically for part (a) of the next theorem since $a_0 = \nu + d > d$.

THEOREM 6. *For any $d = 1, 2, \dots$:*

(a) *For any $\nu > 0$ and $Q \in \mathcal{U}_{d,\nu+d}$, the map $A \mapsto Qh(A)$ defined by (4) for $\rho = \rho_{\nu,d}$ has a unique critical point $A(\nu) := A_\nu(Q)$ which is an absolute minimum.*

In parts (b) through (f) let $\nu > 1$, let P be a law on \mathbb{R}^d , $Q = P \circ T_1^{-1}$ on \mathbb{R}^{d+1} , and $\nu' := \nu - 1$. Assume $P \in \mathcal{V}_{d,\nu+d}$ in parts (b) through (e). We have:

(b) $A(\nu')_{d+1,d+1} = \int u_{\nu',d+1}(z' A(\nu')^{-1} z) dQ(z) = 1.$

(c) *For any $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathcal{P}_d$ let $A = A(\Sigma, \mu, 1) \in \mathcal{P}_{d+1}$ in (10). Then for any $y \in \mathbb{R}^d$ and $z := (y', 1)'$, we have*

$$(15) \quad u_{\nu',d+1}(z' A^{-1} z) \equiv u_{\nu,d}((y - \mu)' \Sigma^{-1} (y - \mu)).$$

In particular, this holds for $A = A(\nu')$ and its corresponding $\mu = \mu_\nu \in \mathbb{R}^d$ and $\Sigma = \Sigma_\nu \in \mathcal{P}_d$.

(d)

$$(16) \quad \int u_{v,d}((y - \mu_v)' \Sigma_v^{-1} (y - \mu_v)) dP(y) = 1.$$

(e) For $h := h_v := h_{v,d}$ defined by (13) with $\rho = \rho_{v,d}$, (μ_v, Σ_v) is an M-functional for P .

(f) If, on the other hand, $P \notin \mathcal{V}_{d,v+d}$, then $(\mu, \Sigma) \mapsto Ph(\mu, \Sigma)$ for h as in part (e) has no critical points.

Kent, Tyler and Vardi [(1994), Theorem 3.1] showed that if $u(s) \geq 0$, $u(0) < +\infty$, $u(\cdot)$ is continuous and nonincreasing for $s \geq 0$, and $su(s)$ is nondecreasing for $s \geq 0$, with $a_0 := \lim_{s \rightarrow +\infty} su(s) > d$, and if equation (16) holds with u in place of $u_{v,d}$ at each critical point (μ, Σ) of $Q_n h$ for any Q_n , then u must be of the form $u(s) = u_{v,d}(s) = (v + d)/(v + s)$ for some $v > 0$. Thus, the method of relating pure scatter functionals in \mathbb{R}^{d+1} to location-scatter functionals in \mathbb{R}^d given by Theorem 6 for t functionals defined by functions $u_{v,d}$ does not extend directly to other functions u . For $0 < v < 1$, we would get $v' < 0$, so the methods of Section 3 do not apply. In fact, (unique) t_v location and scatter M-functionals may not exist, as Gabrielsen (1982) and Kent and Tyler (1991) noted. For example, if $d = 1$, $0 < v < 1$, and P is symmetric around 0 and nonatomic but concentrated near ± 1 , then for $-\infty < \mu < \infty$, there is a unique $\sigma_v(\mu) > 0$ where the minimum of $Ph_v(\mu, \sigma)$ with respect to σ is attained. Then $\sigma_v(0) \doteq 1$ and $(0, \sigma_v(0))$ is a saddle point of Ph_v . Minima occur at some $\mu \neq 0, \sigma > 0$, and at (μ, σ) if and only if at $(-\mu, \sigma)$. The Cauchy case $v = 1$ can be treated separately, see Kent, Tyler and Vardi (1994), Section 5, and references there.

When $d = 1$, $P \in \mathcal{V}_{1,v+1}$ requires that $P(\{x\}) < v/(1 + v)$ for each point x . Then Σ reduces to a number σ^2 with $\sigma > 0$. If $v > 1$ and $P \notin \mathcal{V}_{1,v+1}$, then for some unique x , $P(\{x\}) \geq v/(v + 1)$. One can extend (μ_v, σ_v) by setting $\mu_v(P) := x$ and $\sigma_v(P) := 0$, with (μ_v, σ_v) then being weakly continuous at all P , as will be seen in Section 6.

For $d > 1$ there is no weakly continuous extension to all P , because such an extension of μ_v would give a weakly continuous affinely equivariant location functional defined for all laws, which is known to be impossible [Obenchain (1971)].

Here is a delta-method fact.

THEOREM 7. (a) For any $d = 1, 2, 3, \dots$, $v > 0$, and $Q \in \mathcal{U}_{d,v+d}$ with empirical measures Q_n , we have $Q_n \in \mathcal{U}_{d,v+d}$ with probability $\rightarrow 1$ as $n \rightarrow \infty$ and $\sqrt{n}(A_v(Q_n) - A_v(Q))$ converges in distribution to a normal distribution $N(0, S)$ on \mathcal{S}_d . The covariance matrix S has full rank $d(d + 1)/2$ if Q is not concentrated in any set where a non-zero second-degree polynomial vanishes, for example, if Q has a density. For general $Q \in \mathcal{U}_{d,v+d}$, if $d = 1$ the rank is exactly 1, and for $d \geq 2$, the smallest possible rank of S is $d - 1$.

(b) For any $d = 1, 2, \dots$, $1 < \nu < \infty$ and $P \in \mathcal{V}_{d,\nu+d}$ with empirical measures P_n , we have $P_n \in \mathcal{V}_{d,\nu+d}$ with probability $\rightarrow 1$ as $n \rightarrow \infty$ and the functionals μ_ν and Σ_ν are such that as $n \rightarrow \infty$,

$$\sqrt{n}[(\mu_\nu, \Sigma_\nu)(P_n) - (\mu_\nu, \Sigma_\nu)(P)]$$

converges in distribution to some normal distribution with mean 0 on $\mathbb{R}^d \times \mathbb{R}^{d^2}$, whose marginal on \mathbb{R}^{d^2} is concentrated on \mathcal{S}_d . The covariance of the asymptotic normal distribution for $\mu_\nu(P_n)$ has full rank d . The rank of the covariance for $\Sigma_\nu(P_n)$ has the same behavior as the rank of S in part (a).

The proof is based on some differentiability and an implicit function theorem. Although A is finite-dimensional, P is infinite-dimensional, so an infinite-dimensional implicit function theorem is needed, specifically the Hildebrandt–Graves theorem, for example, Deimling (1985), pages 148–150. For this we need to choose suitable norms defined on the space of all probability measures, which is not a simple matter; it is treated in Sections 5 and 7 and the related part of the Appendix.

For $0 < \delta < 1$ and $d = 1, 2, \dots$, define an open subset of $\mathcal{P}_d \subset \mathcal{S}_d$ by

$$(17) \quad \mathcal{W}_\delta := \mathcal{W}_{\delta,d} := \{A \in \mathcal{P}_d : \max(\|A\|, \|A^{-1}\|) < 1/\delta\}.$$

Now, here is a statement on uniformity as P and Q vary. Its proof is partly based on the results of Giné and Zinn (1991) and Bousquet, Koltchinskii and Panchenko (2002).

PROPOSITION 8. For any $\delta > 0$ and $M < \infty$, the rate of convergence to normality in Theorem 7(a) is uniform over the set $\mathcal{Q} := \mathcal{Q}(\delta, M, \nu)$ of all $Q \in \mathcal{U}_{d,\nu+d}$ such that $A_\nu(Q) \in \mathcal{W}_\delta$ and

$$(18) \quad Q(\{|y| > M\}) \leq (1 - \delta)/(\nu + d),$$

or in part (b), over all $P \in \mathcal{V}_{d,\nu+d}$ such that $\Sigma_\nu(P) \in \mathcal{W}_\delta$ and (18) holds for P in place of Q .

REMARK. The example after Lemma 10 will show that $A = A_\nu(Q)$ itself does not control Q well enough to keep it away from the boundary of $\mathcal{U}_{d,\nu+d}$ or give uniformity in the limit theorem. For a class \mathcal{Q} of laws to have the uniform asymptotic normality of A_ν , uniform tightness is not necessary, but a special case (18) of uniform tightness is assumed.

5. Differentiability of t functionals. To prove the delta-method Theorem 7 for a general P , for example, for the Cauchy distribution on \mathbb{R} , we use differentiability for special norms to be introduced in Section 7. First let us consider a familiar norm that metrizes weak convergence. For a bounded function f

from \mathbb{R}^d into a normed space, the sup norm is $\|f\|_{\text{sup}} := \sup_{x \in \mathbb{R}^d} \|f(x)\|$. Let V be a k -dimensional real vector space with a norm $\|\cdot\|$, where $1 \leq k < \infty$. Let $BL(\mathbb{R}^d, V)$ be the vector space of all functions f from \mathbb{R}^d into V such that the norm

$$\|f\|_{BL} := \|f\|_{\text{sup}} + \sup_{x \neq y} \|f(x) - f(y)\|/|x - y| < \infty,$$

that is bounded Lipschitz functions. The space $BL(\mathbb{R}^d, V)$ does not depend on $\|\cdot\|$, although $\|\cdot\|_{BL}$ does. Take any basis $\{v_j\}_{j=1}^k$ of V . Then $f(x) \equiv \sum_{j=1}^k f_j(x)v_j$ for some $f_j \in BL(\mathbb{R}^d) := BL(\mathbb{R}^d, \mathbb{R})$ where \mathbb{R} has its usual norm $|\cdot|$. Let $X := BL^*(\mathbb{R}^d)$ be the dual Banach space. For $\phi \in X$, let

$$\phi^* f := \sum_{j=1}^k \phi(f_j)v_j \in V.$$

Then because ϕ is linear, $\phi^* f$ does not depend on the choice of basis.

For a given domain U , let $C_b^k(U)$ denote the space of real-valued functions on U having continuous derivatives (equivalently, partial derivatives) through order k bounded on U . Then clearly $C_b^1(\mathbb{R}^d) \subset BL(\mathbb{R}^d)$.

Let $\mathcal{P}(\mathbb{R}^d)$ be the set of all probability measures on the Borel sets of \mathbb{R}^d . Then each $Q \in \mathcal{P}(\mathbb{R}^d)$ defines a $\phi_Q \in BL^*(\mathbb{R}^d)$ via $\phi_Q(f) := \int f dQ$. For any $P, Q \in \mathcal{P}(\mathbb{R}^d)$ let $\beta(P, Q) := \|P - Q\|_{BL}^* := \|\phi_P - \phi_Q\|_{BL}^*$. Then β is a metric on $\mathcal{P}(\mathbb{R}^d)$ which metrizes the weak topology, for example, Dudley (2002), Theorem 11.3.3.

Substituting $\rho_{v,d}$ from (14) into (2) gives for $y \in \mathbb{R}^d$ and $A \in \mathcal{P}_d$,

$$(19) \quad L_{v,d}(y, A) := \frac{1}{2} \log \det A + \frac{v+d}{2} \log[1 + v^{-1}y'A^{-1}y].$$

Then, reserving $h_v := h_{v,d}$ for the location-scatter case as in Theorem 6(e), we get in (3) for the pure scatter case

$$(20) \quad H_v(y, A) := H_{v,d}(y, A) := L_{v,d}(y, A) - L_{v,d}(y, I).$$

We have for fixed $A \in \mathcal{P}_d$ as $\Delta \rightarrow 0$ in \mathcal{S}_d that

$$(21) \quad \begin{aligned} & \log \det(A + \Delta) - \log \det A \\ &= \text{trace}(A^{-1}\Delta) - \|A^{-1/2}\Delta A^{-1/2}\|_F^2/2 + O(\|\Delta\|^3). \end{aligned}$$

It follows from (21) and (19) that for $A \in \mathcal{P}_d$ and $C = A^{-1}$, gradients with respect to C are given by

$$(22) \quad G_{(v)}(y, A) := \nabla_C H_{v,d}(y, A) = \nabla_C L_{v,d}(y, A) = -\frac{A}{2} + \frac{(v+d)yy'}{2(v+y'Cy)} \in \mathcal{S}_d.$$

For any $A \in \mathcal{P}_d$, $C = A^{-1}$, and $L_\nu := L_{\nu,d}$, let

$$I(C, Q) := QH_\nu(A) = \int (L_\nu(y, A) - L_\nu(y, I)) dQ(y),$$

$$J(C, Q) := \frac{1}{2} \log \det C + I(C, Q) = \frac{\nu + d}{2} \int \log \left[\frac{\nu + y'Cy}{\nu + y'y} \right] dQ(y).$$

PROPOSITION 9. (a) *The function $C \mapsto I(C, Q)$ is an analytic function of C on the open subset \mathcal{P}_d of \mathcal{S}_d .*

(b) *Its gradient is*

$$(23) \quad \nabla_C I(C, Q) \equiv \frac{1}{2} \left((\nu + d) \int \frac{yy'}{\nu + y'Cy} dQ(y) - A \right).$$

(c) *The functional $C \mapsto J(C, Q)$ has the Taylor expansion around any $C \in \mathcal{P}_d$*

$$(24) \quad J(C + \Delta, Q) - J(C, Q) = \frac{\nu + d}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \int \frac{(y' \Delta y)^k}{(\nu + y'Cy)^k} dQ(y),$$

convergent for $\|\Delta\| < 1/\|A\|$.

(d) *For any $\delta \in (0, 1)$, $\nu \geq 1$ and $j = 1, 2, \dots$, the function $C \mapsto I(C, Q)$ is in $C_b^j(\mathcal{W}_{\delta,d})$.*

Usually, one might show that the Hessian (matrix of second partial derivatives) is positive definite at a critical point to show it is a strict relative minimum. Here we already know from Theorem 6(a) that we have a unique critical point which is a strict absolute minimum. The following lemma is useful instead in showing differentiability of t functionals via the Hildebrandt–Graves implicit function theorem, for example, Deimling (1985), pages 148–150, in that it implies that the derivative of the gradient (the Hessian) is nonsingular. To treat t functionals of location and scatter in any dimension p we will need functionals of pure scatter in dimension $p + 1$, so here we only need dimension $d \geq 2$.

LEMMA 10. *For each $\nu > 0$, $d = 2, 3, \dots$, and $Q \in \mathcal{U}_{d,\nu+d}$, at $A(\nu) = A_\nu(Q)$ given by Theorem 6(a), the Hessian of QH_ν on \mathcal{S}_d with respect to $C = A^{-1}$ is positive definite.*

EXAMPLE. For Q such that $A_\nu(Q) = C_\nu = I_d$, the $d \times d$ identity matrix, a large part of the mass of Q can escape to infinity, Q can approach the boundary of $\mathcal{U}_{d,\nu+d}$, and some eigenvalues of the Hessian can approach 0, as follows. Let e_j be the standard basis vectors of \mathbb{R}^d . For $c > 0$ and p such that $1/[2(\nu + d)] < p \leq 1/(2d)$, let

$$Q := (1 - 2dp)\delta_0 + p \sum_{j=1}^d \delta_{-ce_j} + \delta_{ce_j}.$$

To get $A_\nu(Q) = I_d$, by (7) we need $(\nu + d) \cdot 2pc^2 = \nu + c^2$, or $\nu = c^2[2p(\nu + d) - 1]$. There is a unique solution for $c > 0$ but as $p \downarrow 1/[2(\nu + d)]$, we have $c \uparrow +\infty$. Then, for each $q = 0, 1, \dots, d - 1$, for each q -dimensional subspace H where $d - q$ of the coordinates are 0, $Q(H) \uparrow 1 - \frac{d-q}{\nu+d}$, the critical value for which $Q \notin \mathcal{U}_{d,\nu+d}$. Also, an amount of probability for Q converging to $d/(\nu + d)$ is escaping to infinity. The Hessian has d arbitrarily small eigenvalues $\nu/(\nu + c^2)$.

The second-order term in the Taylor expansion of $C \mapsto I(C, Q)$, for example, (36) in the Appendix, using also (21), is the quadratic form, for $\Delta \in \mathcal{S}_d$,

$$(25) \quad \Delta \mapsto \frac{1}{2} \left(\|A^{1/2} \Delta A^{1/2}\|_F^2 - (\nu + d) \int \frac{(y' \Delta y)^2}{(\nu + y' C y)^2} dQ(y) \right).$$

For the relatively open set $\mathcal{P}_d \subset \mathcal{S}_d$ and $G_{(\nu)}$ from (22), define the function $F := F_\nu$ from $X \times \mathcal{P}_d$ into \mathcal{S}_d by

$$(26) \quad F(\phi, A) := \phi^*(G_{(\nu)}(\cdot, A)).$$

Then F is well defined because $G_{(\nu)}(\cdot, A)$ is a bounded and Lipschitz \mathcal{S}_d -valued function of x for each $A \in \mathcal{P}_d$; in fact, each entry is C^1 with bounded derivative, as is straightforward to check.

For $d = 1$, and a finite signed Borel measure τ , let

$$(27) \quad \|\tau\|_{\mathcal{X}} := \sup_x |\tau((-\infty, x])|.$$

Let P and Q be two laws with distribution functions F_P and F_Q . Then $\|P - Q\|_{\mathcal{X}}$ is the usual sup (Kolmogorov) norm distance $\sup_x |(F_Q - F_P)(x)|$.

The next statement and its proof in Dudley, Sidenko and Wang (2008) call on some basic notions and facts from infinite-dimensional calculus, reviewed in the Appendix.

THEOREM 11. *Let $\nu > 0$ in parts (a) through (c), $\nu > 1$ in parts (d), (e).*

- (a) *The function $F = F_\nu$ is analytic from $X \times \mathcal{P}_d$ into \mathcal{S}_d where $X = BL^*(\mathbb{R}^d)$.*
- (b) *For any law $Q \in \mathcal{U}_{d,\nu+d}$, and the corresponding $\phi_Q \in X$, at $A_\nu(Q)$ given by Theorem 6(a), the partial derivative linear map $\partial_C F(\phi_Q, A)/\partial C := \nabla_C F(\phi_Q, A)$ from \mathcal{S}_d into \mathcal{S}_d is invertible.*
- (c) *Still for $Q \in \mathcal{U}_{d,\nu+d}$, the functional $Q \mapsto A_\nu(Q)$ is analytic for the BL^* norm.*
- (d) *For each $P \in \mathcal{V}_{d,\nu+d}$, the t_ν location-scatter functional $P \mapsto (\mu_\nu, \Sigma_\nu)(P)$ given by Theorems 3 and 6 is also analytic for the norm on X .*
- (e) *For $d = 1$, the t_ν location and scatter functionals μ_ν, σ_ν^2 are on $\mathcal{V}_{1,\nu+1}$ with respect to the sup norm $\|\cdot\|_{\mathcal{X}}$.*

If a functional T is differentiable at P for a suitable norm, with a nonzero derivative, then one can look for asymptotic normality of $\sqrt{n}(T(P_n) - T(P))$ by way

of some central limit theorem and the delta-method. For this purpose, the dual-bounded-Lipschitz norm $\|\cdot\|_{BL}^*$, although it works for large classes of distributions, is still too strong for some heavy-tailed distributions. For $d = 1$, let P be a law concentrated in the positive integers with $\sum_{k=1}^\infty \sqrt{P(\{k\})} = +\infty$. Then a short calculation shows that as $n \rightarrow \infty$, $\sqrt{n} \sum_{k=1}^\infty |(P_n - P)(\{k\})| \rightarrow +\infty$ in probability. For any numbers a_k there is an $f \in BL(\mathbb{R})$ with usual metric such that $f(k)a_k = |a_k|$ for all k and $\|f\|_{BL} \leq 3$. Thus $\sqrt{n}\|P_n - P\|_{BL}^* \rightarrow +\infty$ in probability. Giné and Zinn (1986) proved equivalence of the related condition $\sum_{j=1}^\infty \Pr(j - 1 < |X| \leq j)^{1/2} < \infty$ for X with general distribution P on \mathbb{R} to the Donsker property [defined in Dudley (1999), Section 3.1] of $\{f : \|f\|_{BL} \leq 1\}$. But norms more directly adapted to the functions needed will be defined in Section 7.

6. The one-dimensional case. In dimension $d = 1$, the scatter matrix Σ reduces to a number σ^2 . The ρ and h functions in this case become, for $\theta := (\mu, \sigma)$ with $\sigma > 0$,

$$(28) \quad \rho_\nu(x, \theta) := \log \sigma + \frac{\nu + 1}{2} \log \left(1 + \frac{(x - \mu)^2}{\nu \sigma^2} \right),$$

$$(29) \quad h_\nu(x, \theta) := \log \sigma + \frac{\nu + 1}{2} \log \left(\frac{1 + [(x - \mu)^2 / (\nu \sigma^2)]}{1 + x^2 / \nu} \right).$$

The function h_ν is bounded uniformly in x and for $|\mu|$ bounded and σ bounded away from 0 and ∞ . Thus it is integrable for any probability distribution P on \mathbb{R} . Let $Ph_\nu(\theta) := \int h_\nu(x, \theta) dP(x)$. In the next theorem, extended M-functionals are defined by (1) with $\theta := (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$ and $\overline{\Theta} = \mathbb{R} \times [0, \infty)$.

THEOREM 12. *Let $d = 1$ and $1 < \nu < \infty$. Then:*

(a) *For any law Q on \mathbb{R} satisfying*

$$(30) \quad \max_t Q(\{t\}) < \nu / (\nu + 1),$$

the M-functional $(\mu, \sigma) = (\mu_\nu, \sigma_\nu)(Q)$ exists with $\sigma_\nu(Q) > 0$ and is the unique critical point with $\partial Q h_\nu / \partial \mu = \partial Q h_\nu / \partial \sigma = 0$. On the set of laws satisfying (30), (μ_ν, σ_ν) is analytic with respect to the dual-bounded-Lipschitz norm and thus weakly continuous.

(b) *For any law Q on \mathbb{R} , the extended M-functional $\theta_0(Q) := (\mu_\nu, \sigma_\nu)(Q) \in \overline{\Theta}$ exists for h_ν from (29).*

(c) *If $Q(\{s\}) \geq \nu / (\nu + 1)$ for some (unique) s , then $\mu_\nu(Q) = s$ and $\sigma_\nu(Q) = 0$.*

(d) *The map $Q \mapsto \theta_0(Q)$ is weakly continuous at every law Q . For X_1, X_2, \dots i.i.d. (Q) and empirical measures $Q_n := n^{-1} \sum_{j=1}^n \delta_{X_j}$, we thus have maximum likelihood estimates $\hat{\theta}_n = \theta_0(Q_n)$ existing for all n and converging to $\theta_0(Q)$ almost surely.*

REMARKS. For $Q_p := (1-p)\delta_0 + p\delta_1$, $d\sigma_v^2(Q_p)/dp$ has different left and right limits at $p = 1/(v+1)$. Thus σ_v^2 is not differentiable, and σ_v is not Lipschitz, at Q_p , with respect to p or any norm. So in part (d), continuity cannot be improved to Lipschitz.

The theorem does not extend to $0 < v \leq 1$. For some Q , points s in part (c) are not unique. For example if $v = 1$ (the Cauchy case) and $Q = \frac{1}{2}(\delta_{-1} + \delta_1)$, the likelihood is maximized on the semicircle $\mu^2 + \sigma^2 = 1$, as Copas (1975) noted.

7. Some Banach spaces generated by rational functions. The classes of functions and norms defined in this section are used in the proof (although not the statement) of the main theorem of the paper, the delta-method Theorem 7. Some further facts on these norms are stated at the end of the Appendix. Details and proofs are given in Dudley et al. (2007). Throughout this section let $0 < \delta < 1$, $d = 1, 2, \dots$ and $r = 1, 2, \dots$ be arbitrary unless further specified. Let \mathcal{MM}_r be the set of monic monomials g from \mathbb{R}^d into \mathbb{R} of degree r , in other words $g(x) = \prod_{i=1}^d x_i^{n_i}$ for some $n_i \in \mathbb{N}$ with $\sum_{i=1}^d n_i = r$. For \mathcal{W}_δ defined in (17), let

$$\mathcal{F}_{\delta,r} := \mathcal{F}_{\delta,r,d} := \left\{ f : \mathbb{R}^d \rightarrow \mathbb{R}, f(x) \equiv g(x) / \prod_{s=1}^r (1 + x' C_s x), \right. \\ \left. \text{where } g \in \mathcal{MM}_{2r}, \text{ and for } s = 1, \dots, r, C_s \in \mathcal{W}_\delta \right\}.$$

For $1 \leq j \leq r$, let $\mathcal{F}_{\delta,r}^{(j)} := \mathcal{F}_{\delta,r,d}^{(j)}$ be the set of $f \in \mathcal{F}_{\delta,r}$ such that for $s = 1, \dots, r$, C_s ranges over at most j different matrices. Such functions relate to our purposes as follows: the coordinates of derivatives (partial derivatives) of t_v log likelihood functions with respect to C , evaluated at a $C \in \mathcal{W}_\delta$, are constants plus constants times functions in $\mathcal{F}_{\delta,r}^{(1)}$ (here all $C_s = C$ so $j = 1$). To show that Fréchet differentiability of some order holds in some norm, or to show that the derivatives with respect to C can be interchanged with integrals, we need to consider difference-quotients and thus to take $j = 2$, as we do in the definition $X_{\delta,r,v}$ below and so in Theorem 14 and other facts based on it.

We have $\mathcal{F}_{\delta,r} = \mathcal{F}_{\delta,r}^{(r)}$. Let $\mathcal{G}_{\delta,r}^{(j)} := \mathcal{G}_{\delta,r,d}^{(j)} := \bigcup_{v=1}^r \mathcal{F}_{\delta,v}^{(j)}$. For the reasons mentioned above we will be interested in $j = 1$ and 2. Clearly $\mathcal{F}_{\delta,r}^{(1)} \subset \mathcal{F}_{\delta,r}^{(2)} \subset \dots \subset \mathcal{F}_{\delta,r}$ for each δ and r . The next lemma is straightforward:

LEMMA 13. For any $f \in \mathcal{G}_{\delta,r}^{(r)}$ we have $(\delta/d)^r \leq \|f\|_{\text{sup}} \leq \delta^{-r}$.

For any $f : \mathbb{R}^d \rightarrow \mathbb{R}$, define

$$(31) \quad \|f\|_{\delta,r}^{*,j} := \|f\|_{\delta,r,d}^{*,j} := \inf \left\{ \sum_{s=1}^{\infty} |\lambda_s| : \exists g_s \in \mathcal{G}_{\delta,r}^{(j)}, s \geq 1, f \equiv \sum_{s=1}^{\infty} \lambda_s g_s \right\},$$

or $+\infty$ if no such λ_s, g_s with $\sum_s |\lambda_s| < \infty$ exist. Lemma 13 implies that for $\sum_s |\lambda_s| < \infty$ and $g_s \in \mathcal{G}_{\delta,r}^{(r)}$, $\sum_s \lambda_s g_s$ converges absolutely and uniformly on \mathbb{R}^d . Let $Y_{\delta,r}^j := Y_{\delta,r,d}^j := \{f : \mathbb{R}^d \rightarrow \mathbb{R}, \|f\|_{\delta,r}^{*,j} < \infty\}$. It's easy to see that each $Y_{\delta,r}^j$ is a real vector space of functions on \mathbb{R}^d and $\|\cdot\|_{\delta,r}^{*,j}$ is a seminorm on it.

Let $\mathbb{R} \oplus Y_{\delta,r}^j$ be the set of all functions $c + g$ on \mathbb{R}^d for any $c \in \mathbb{R}$ and $g \in Y_{\delta,r}^j$. Then c and g are uniquely determined since $g(0) = 0$. Let $\|c + g\|_{\delta,r,d}^{**,j} := |c| + \|g\|_{\delta,r,d}^{*,j}$. Let $X_{\delta,r,v}$ be the dual Banach space of $\mathbb{R} \oplus Y_{\delta/v,r,d}^2$, that is, the set of all real-valued linear functionals ϕ on it for which the norm

$$\|\phi\|_{\delta,r,v} := \sup\{|\phi(f)| : \|f\|_{\delta/v,r,d}^{**,2} \leq 1\} < \infty.$$

Let $X_{\delta,r,v}^0 := \{\phi \in X_{\delta,r,v} : \phi(c) = 0 \text{ for all } c \in \mathbb{R}\}$. For $\phi \in X_{\delta,r,v}^0$, by (31)

$$\begin{aligned} \|\phi\|_{\delta,r,v} &\equiv \|\phi\|_{\delta,r,v}^0 := \sup\{|\phi(0, g)| : \|g\|_{\delta/v,r,d}^{*,2} \leq 1\} \\ (32) \quad &\leq \sup\{|\phi(0, g)| : g \in \mathcal{G}_{\delta/v,r}^{(2)}\} \leq \sup\{|\phi(0, g)| : g \in \mathcal{G}_{\delta/v,r}^{(r)}\}. \end{aligned}$$

For $A \in \mathcal{W}_{\delta,d}$ and $\phi \in X_{\delta,r,v}$, define $F(\phi, A)$ again by (26), which makes sense since for any $r = 1, 2, \dots$, $G_{(v)}$ has entries in $Y_{\delta/v,1,d}^1 \subset Y_{\delta/v,r,d}^2$. Proposition 19, closely related to Theorem 18, implies that in the following theorem $k + 2$ cannot be replaced by $k + 1$.

THEOREM 14. *For any $d = 1, 2, \dots$, $k = 1, 2, \dots$, $0 < v < \infty$, and $Q \in \mathcal{U}_{d,v+d}$, there is a δ with $0 < \delta < 1$ such that the conclusions of Theorem 11 hold for $X = X_{\delta,k+2,v}$ in place of $BL^*(\mathbb{R}^d)$, $\mathcal{W}_{\delta,d}$ in place of \mathcal{P}_d , $v > 1$ in part (d), and analyticity replaced by C^k in parts (a), (c) and (d).*

8. Norms based on classes of sets. Suppose $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on a vector space V such that for some $K < \infty$, $\|x\|_2 \leq K\|x\|_1$ for all $x \in V$. Let $U \subset V$ be open for $\|\cdot\|_2$ and so also for $\|\cdot\|_1$. Let $v \in U$ and suppose a functional T from U into some other normed space is Fréchet differentiable at v for $\|\cdot\|_2$. Then the same holds for $\|\cdot\|_1$ since the identity from V to V is a bounded linear operator from $(V, \|\cdot\|_1)$ to $(V, \|\cdot\|_2)$ and so equals its own Fréchet derivative everywhere on V , and we can apply a chain rule, for example, Dieudonné (1960), equation (8.12.10).

If \mathcal{F} is a class of bounded real-valued functions on a set χ , measurable for a σ -algebra \mathcal{A} of subsets of χ , and ϕ is a finite signed measure on \mathcal{A} , (e.g., $P_n - P$) let $\|\phi\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |\int f d\phi|$. For $\mathcal{C} \subset \mathcal{A}$ let $\|\phi\|_{\mathcal{C}} := \|\phi\|_{\mathcal{G}}$, where $\mathcal{G} := \{1_C : C \in \mathcal{C}\}$.

Let \mathcal{F} be a VC major class of functions, as defined in Dudley (1999), Section 4.7, for the VC class \mathcal{E} of sets where $\mathcal{E} \subset \mathcal{A}$ and suppose for some $M < \infty$,

$|f(x)| \leq M$ for all $f \in \mathcal{F}$ and $x \in \chi$. Then for any finite signed measure ϕ on \mathcal{A} having total mass $\phi(\chi) = 0$ (e.g., $\phi = P - Q$ for any two laws P and Q), we have

$$(33) \quad \|\phi\|_{\mathcal{F}} \leq 2M\|\phi\|_{\mathcal{E}},$$

by the rescaling $f \mapsto (f + M)/(2M)$ to get functions with values in $[0, 1]$ and then a convex hull representation [Dudley (1987), Theorem 2.1(a) or Dudley (1999), Theorem 4.7.1(b)]; additive constants make no difference since $\phi(\chi) = 0$.

Let $\Gamma_{\delta, \nu}^{r+2, d} := \mathcal{G}_{\delta/\nu, r+2, d}^{(r+2)}$. Each $\Gamma_{\delta, \nu}^{r+2, d}$ is a uniformly bounded VC major class for the VC class $\mathcal{E}(2r + 4, d)$ of sets, namely positivity sets of polynomials of degree $\leq 2r + 4$. So by (32) and (33), for some $M < \infty$ depending on r, δ, ν and d , we have

$$(34) \quad \|\phi\|_{\delta, r+2, \nu} \leq 2M\|\phi\|_{\mathcal{E}(2r+4, d)}$$

for all finite signed measures ϕ on \mathbb{R}^d with $\phi(\mathbb{R}^d) = 0$.

COROLLARY 15. *For each $d = 1, 2, \dots$, and $\nu > 1$, the Fréchet C^k differentiability property of the t_ν location and scatter functionals at each P in $\mathcal{V}_{d, \nu+d}$, as in Theorem 14 with respect to $\|\cdot\|_{\delta, k+2, \nu}$, also holds with respect to $\|\cdot\|_{\mathcal{E}(2k+4, d)}$.*

Each class $\mathcal{E}(r, d)$ for $r = 1, 2, \dots$ is invariant under all nonsingular affine transformations of \mathbb{R}^d , and hence so is the norm $\|\cdot\|_{\mathcal{E}(r, d)}$. Davies (1993), pages 1851–1852, defines norms $\|\cdot\|_{\mathcal{L}}$ based on suitable VC classes \mathcal{L} of subsets of \mathbb{R}^d and points out Donsker and affine invariance properties. The norms $\|\cdot\|_{\delta, r, \nu}$ are not affinely invariant.

On the other hand, note that M in (34) depends on δ , and there is no corresponding inequality in the opposite direction. Thus, Fréchet differentiability (of any order) is a strictly stronger property for $\|\cdot\|_{\delta, k+2, \nu}$ than it is for $\|\cdot\|_{\mathcal{E}(2k+4, d)}$.

APPENDIX

Derivatives in Banach spaces. The standard mathematicians’ definition of Fréchet differentiability requires that a function(al) be defined on an open subset of a normed vector space. Statisticians often find it natural to consider functionals defined on sets of probability measures and adapt the definition of Fréchet differentiability, although no set of probability measures is open in any normed space of signed measures. In this paper we need the mathematicians’ definition, to be recalled, in order to apply implicit function theorems.

Let X and Y be Banach spaces over the real numbers. Let $B(X, Y)$ be the space of bounded, that is continuous, linear operators A from X into Y , with the norm $\|A\| := \sup\{\|Ax\| : \|x\| = 1\}$. Let U be an open subset of X , $x \in U$, and f a function from U into Y . Then f is said to be *Fréchet differentiable* at x iff there is an $A \in B(X, Y)$ such that

$$f(u) = f(x) + A(u - x) + o(\|u - x\|)$$

as $u \rightarrow x$. If so let $(Df)(x) := A$. Then f is said to be C^1 on U if it is Fréchet differentiable at each $x \in U$ and $x \mapsto Df(x)$ is continuous from U into $B(X, Y)$. Iterating the definition, the second derivative $D^2 f(x) = D(Df)(x)$, if it exists for a given x , is in $B(X, B(X, Y))$, and the k th derivative $D^k f(x)$ will be in $B(X, B(X, \dots, B(X, Y))) \dots$ with k B 's. Then f is called C^k on U if its k th derivative exists and is continuous on U . If f is C^k on U for all $k = 1, 2, \dots$, it is called C^∞ on U . In some cases, higher-order derivatives will be seen to simplify or to reduce to more familiar notions.

Let X and Y be real vector spaces. For $k \geq 1$, a mapping $T : (x_1, \dots, x_k) \mapsto T(x_1, \dots, x_k)$ from X^k into Y is called k -linear iff for each $j = 1, \dots, k$, T is linear in X_j if x_i for $i \neq j$ are fixed. T is called symmetric if and only if for each $\pi \in S_k$, the set of all permutations of $\{1, \dots, k\}$, we have $T(x_{\pi(1)}, \dots, x_{\pi(k)}) \equiv T(x_1, \dots, x_k)$. Any k -linear mapping T has a symmetrization T_s , which is symmetric, defined by

$$T_s(x_1, \dots, x_k) := \frac{1}{k!} \sum_{\pi \in S_k} T(x_{\pi(1)}, \dots, x_{\pi(k)}).$$

A function g from X into Y is called a k -homogeneous polynomial if and only if for some k -linear $T : X^k \rightarrow Y$, we have $g(x) \equiv g_T(x) := T x^{\otimes k} := T(x, x, \dots, x)$ for all $x \in X$. Since $g_{T_s} \equiv g_T$ one can assume that T is symmetric. For the following, one can obtain T from g by the ‘‘polarization identity,’’ for example, Chae (1985), Theorem 4.6.

PROPOSITION 16. *For any two real vector spaces X and Y and $k = 1, 2, \dots$, there is a 1–1 correspondence between symmetric k -linear mappings T from X^k into Y and k -homogeneous polynomials $g = g_T$ from X into Y .*

Now suppose $(X, \|\cdot\|)$ and $(Y, |\cdot|)$ are normed vector spaces. It is known and not hard to show that a k -linear mapping T from X^k into Y is jointly continuous if and only if

$$\|T\| := \sup\{|T(x_1, \dots, x_k)| : \|x_1\| = \dots = \|x_k\| = 1\} < \infty,$$

and that a k -homogeneous polynomial g from X into Y is continuous if and only if $\|g\| := \sup\{|g(x)| : \|x\| = 1\} < \infty$. In general, for a symmetric k -linear T with $\|T\| < \infty$ we have $\|g_T\| \leq \|T\| \leq k^k \|g_T\| / k!$, for example, Chae (1985), Theorem 4.13. The bounds are sharp in general Banach spaces, but if X and Y are Hilbert spaces we have $\|g_T\| \equiv \|T\|$.

If f is a C^k function from an open set $U \subset X$ into Y then at each $x \in U$, $D^k f(x)$ defines a k -linear mapping $d^k f(x)$ from X^k into Y ,

$$(35) \quad d^k f(x)(x_1, \dots, x_k) := (\dots((D^k f(x))(x_1))(x_2) \cdots (x_k)).$$

Then $d^k f(x)$ is symmetric, for example Chae (1985), Theorem 7.9. The corresponding k -homogeneous polynomial $u \mapsto g_{d^k f(x)}(u)$ is $u \mapsto d^k f(x)u^{\otimes k}$.

Also, f will be called *analytic* from U into Y if and only if it is C^∞ and for each $x \in U$ there is an $r > 0$ and k -homogeneous polynomials V_k from X into Y for each $k \geq 1$ such that for any $v \in X$ with $\|v - x\| < r$, we have $v \in U$ and

$$(36) \quad f(v) = f(x) + \sum_{k=1}^{\infty} V_k(v - x).$$

It is known that then necessarily for each $k \geq 1$ and $u \in X$

$$(37) \quad V_k(u) = d^k f(x) u^{\otimes k} / k!.$$

For any Banach space X let $(X', \|\cdot\|')$ be the dual Banach space $B(X, \mathbb{R})$. The product $X' \times X$ with coordinatewise operations is a vector space and a Banach space with the norm $\|(\phi, x)\| := \|\phi\|' + \|x\|$. The mapping $\gamma : (\phi, x) \mapsto \phi(x)$ is C^∞ from $X' \times X$ into \mathbb{R} (it is analytic and a 2-homogeneous polynomial): for $\psi, \phi \in X'$ and $x, y \in X$ we have

$$\gamma(\psi, y) = \psi(y) = \phi(x) + (\psi - \phi)(x) + \phi(y - x) + (\psi - \phi)(y - x).$$

As $(\psi, y) \rightarrow (\phi, x)$, clearly $(\psi - \phi)(x) + \phi(y - x)$ give first-derivative terms and $(\psi - \phi)(y - x)$ a second-derivative term. We have that $D\gamma$ is continuous (linear) and $D^2\gamma$ has a fixed value $((\eta, u), (\zeta, v)) \mapsto \eta(v) + \zeta(u)$ in $B(X' \times X, B(X' \times X, \mathbb{R}))$, so $D^3\gamma \equiv 0$.

If U is an open subset of a Banach space Y and f is a C^k function from U into X , then

$$(38) \quad (\phi, u) \mapsto \phi(f(u))$$

is C^k on $X' \times U$ by a chain rule, for example Dieudonné (1960), equation (8.12.10).

More about Banach spaces based on rational functions. Recall the notations of Section 7. Let $h_C(x) := 1 + x'Cx$ for $C \in \mathcal{P}_d$ and $x \in \mathbb{R}^d$. Then clearly $f \in \mathcal{F}_{\delta,r}^{(1)}$ if and only if for some $P \in \mathcal{MM}_{2r}$ and $C \in \mathcal{W}_\delta$, $f(x) \equiv f_{P,C,r}(x) := P(x)h_C(x)^{-r}$.

PROPOSITION 17. For any $P \in \mathcal{MM}_{2r}$, let $\psi(C, x) := f_{P,C,r}(x) = P(x)/h_C(x)^r$ from $\mathcal{W}_\delta \times \mathbb{R}^d$ into \mathbb{R} . Then:

- (a) For each fixed $C \in \mathcal{W}_\delta$, $\psi(C, \cdot) \in \mathcal{F}_{\delta,r}^{(1)}$.
- (b) For each x , $\psi(\cdot, x)$ has the partial derivative $\nabla_C \psi(C, x) = -rP(x)xx' / h_C(x)^{r+1}$.
- (c) The map $C \mapsto \nabla_C \psi(C, \cdot) \in \mathcal{S}_d$ on \mathcal{W}_δ has entries Lipschitz into $Y_{\delta,r+2}^2$.
- (d) The map $C \mapsto \psi(C, \cdot)$ from \mathcal{W}_δ into $\mathcal{F}_{\delta,r}^{(1)} \subset Y_{\delta,r}^1$, viewed as a map into the larger space $Y_{\delta,r+2}^2$, is Fréchet C^1 .

THEOREM 18. *Let $r = 1, 2, \dots, d = 1, 2, \dots, 0 < \delta < 1$, and $f \in Y_{\delta,r}^1$, so that for some a_s with $\sum_s |a_s| < \infty$ we have $f(x) \equiv \sum_s a_s P_s(x)/(1 + x' C_s x)^{k_s}$ for $x \in \mathbb{R}^d$, where each $P_s \in \mathcal{MM}_{2k_s}$, $k_s = 1, \dots, r$, and $C_s \in \mathcal{W}_\delta$. Then f can be written as a sum of the same form in which the triples (P_s, C_s, k_s) are all distinct. In that case, the C_s, P_s, k_s and the coefficients a_s are uniquely determined by f .*

For any $P \in \mathcal{MM}_{2r}$ and any $C \neq D$ in \mathcal{W}_δ , let

$$f_{P,C,D,r}(x) := f_{P,C,D,r,d}(x) := \frac{P(x)}{(1 + x' C x)^r} - \frac{P(x)}{(1 + x' D x)^r}.$$

For C fixed and $D \rightarrow C$ it is not hard to show that $\|f_{P,C,D,r}\|_{\delta,r+1}^{*,2} \rightarrow 0$. The following shows this is not true if $r + 1$ in the norm is replaced by r , even if the number of different C_s 's in the denominator is allowed to be as large as possible, namely r :

PROPOSITION 19. *For any $r = 1, 2, \dots, d = 1, 2, \dots$, and $C \neq D$ in \mathcal{W}_δ , we have $\|f_{P,C,D,r}\|_{\delta,r}^{*,r} = 2$.*

Proposition 17 can be adapted, replacing $h_C(x)$ by $h_{C,v}(x) := v + x' C x$ and making suitable other constant multiples, replacing δ by δ/v in $\mathcal{F}_{\delta,r}^{(1)}$ and each $Y_{\delta,s}^j$ (but not \mathcal{W}_δ) and in part (a) only, $\psi(C, \cdot)$ by $v^r \psi(C, \cdot)$.

By (24), (36) and (37), for any $0 < \delta < 1, C \in \mathcal{W}_\delta, \Delta \in \mathcal{S}_d$, and $k = 0, 1, 2, \dots$, the k th differential of $G_{(v)}$ from (22) with respect to C , cf. (35), is given by

$$(39) \quad d_C^k G_{(v)}(y, A) \Delta^{\otimes k} = K_k(A) \Delta^{\otimes k} + g_k(y, A, \Delta),$$

where

$$g_k(y, A, \Delta) = \frac{v + d}{2} (-1)^k k! \frac{(y' \Delta y)^{k+1}}{(v + y' C y)^{k+1}}$$

for some k -homogeneous polynomial $K_k(A)$ not depending on y . For $\Delta \in \mathcal{S}_d$, by the Cauchy inequality, $\sum_{i,j=1}^d |\Delta_{ij}| \leq \|\Delta\|_F d$, so $g_k(\cdot, A, \Delta) \in Y_{\delta/v,k+1,d}^1$ with

$$(40) \quad \|g_k(\cdot, A, \Delta)\|_{\delta/v,k+1,d}^{*,1} \leq (v + d) k! (\|\Delta\|_F d/v)^{k+1}.$$

Thus $d_C^k G_{(v)}(\cdot, A) \Delta^{\otimes k} \in \mathbb{R} \oplus Y_{\delta/v,k+1,d}^1$.

Acknowledgments. We thank Lutz Dümbgen and David Tyler very much for kindly providing copies of their preprints Dümbgen (1997) and Tyler (1986), and their paper Dümbgen and Tyler (2005) in preprint versions. We also thank four anonymous referees (of different versions) for helpful comments, Xiongjiu Liao for useful literature searches, and Fangyun Yang for assistance with Section 7 and the related part of the Appendix.

REFERENCES

- ARSLAN, O., CONSTABLE, P. D. L. and KENT, J. T. (1995). Convergence behavior of the EM algorithm for the multivariate t -distribution. *Comm. Statist.—Theory Methods* **24** 2981–3000. [MR1364707](#)
- BICKEL, P. J. and LEHMANN, E. L. (1975). Descriptive statistics for nonparametric models. I. Introduction. II. Location. *Ann. Statist.* **3** 1038–1044, 1045–1069. [MR0395020](#)
- BOOS, D. D. (1979). A differential for L -statistics. *Ann. Statist.* **7** 955–959.
- BOUSQUET, O., KOLTCHINSKII, V. and PANCHENKO, D. (2002). Some local measures of complexity of convex hulls and generalization bounds. In *Computational Learning Theory (Proc. Conf. Sydney, 2002). Lecture Notes in Comput. Sci.* **2375** 59–73. Springer, New York. [MR0536500](#)
- CHAE, S. B. (1985). *Holomorphy and Calculus in Normed Spaces*. Dekker, New York. [MR0788158](#)
- COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704. [MR0388627](#)
- DAVIES, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899. [MR1245772](#)
- DAVIES, P. L. (1998). On locally uniformly linearizable high breakdown location and scale functionals. *Ann. Statist.* **26** 1103–1125. [MR1635450](#)
- DEIMLING, K. (1985). *Nonlinear Functional Analysis*. Springer, Berlin. [MR0787404](#)
- DIEUDONNÉ, J. (1960). *Foundations of Modern Analysis*. Academic Press, New York. [MR0120319](#)
- DUDLEY, R. M. (1987). Universal Donsker classes and metric entropy. *Ann. Probab.* **15** 1306–1326. [MR0905333](#)
- DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge Univ. Press. [MR1720712](#)
- DUDLEY, R. M. (2002). *Real Analysis and Probability*, 2nd ed. Cambridge Univ. Press. [MR1932358](#)
- DUDLEY, R. M. (2006). Some facts about functionals of location and scatter. In *High Dimensional Probability Proc. 4th Internat. Conf.* (E. Giné, V. Koltchinskii, W. Li and J. Zinn, eds.). *IMS Lect. Notes Monograph Ser.* **51** 207–219. IMS, Beachwood, OH. [MR2387771](#)
- DUDLEY, R. M., SIDENKO, S. and WANG, Z. (2008). Differentiability of M -functionals of location and scatter based on t likelihoods. Available at [arXiv:0801.3052](#).
- DUDLEY, R. M., SIDENKO, S., WANG, Z. and YANG, F. (2007). Some classes of rational functions and related Banach spaces. Preprint. Available at [arXiv:math/0709.2449](#).
- DÜMBGEN, L. (1997). The asymptotic behavior of Tyler’s M -estimator of scatter in high dimension. Preprint.
- DÜMBGEN, L. and TYLER, D. E. (2005). On the breakdown properties of some multivariate M -functionals. *Scand. J. Statist.* **32** 247–264. [MR1664575](#)
- DURFEE, A., KRONENFELD, N., MUNSON, H., ROY, J. and WESTBY, I. (1993). Counting critical points of real polynomials in two variables. *Amer. Math. Monthly* **100** 255–271. [MR1212833](#)
- GABRIELSEN, G. (1982). On the unimodality of the likelihood for the Cauchy distribution: Some comments. *Biometrika* **69** 677–678. [MR0695217](#)
- GINÉ, E. and ZINN, J. (1986). Empirical processes indexed by Lipschitz functions. *Ann. Probab.* **14** 1329–1338. [MR0866353](#)
- GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. [MR1055437](#)
- GINÉ, E. and ZINN, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19** 758–782. [MR1106285](#)
- HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proc. Fifth Berkeley Sympos. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley. [MR0216620](#)
- HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York. [MR0606374](#)
- KENT, J. T. and TYLER, D. E. (1991). Redescending M -estimates of multivariate location and scatter. *Ann. Statist.* **19** 2102–2119. [MR1135166](#)

- KENT, J. T., TYLER, D. E. and VARDI, Y. (1994). A curious likelihood identity for the multivariate T-distribution. *Comm. Statist.—Simul.* **23** 441–453. [MR1279675](#)
- LIU, C., RUBIN, D. B. and WU, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika* **85** 755–770. [MR1666758](#)
- OBENCHAIN, R. L. (1971). Multivariate procedures invariant under linear transformations. *Ann. Math. Statist.* **42** 1569–1578. [MR0343463](#)
- TYLER, D. E. (1986). Breakdown properties of the M -estimators of multivariate scatter. Technical report, Rutgers Univ.
- TYLER, D. E. (1988). Some results on the existence, uniqueness and computation of the M -estimates of multivariate location and scatter. *SIAM J. Sci. Statist. Comput.* **9** 354–362. [MR0930051](#)
- ZUO, Y. and CUI, H. (2005). Depth weighted scatter estimators. *Ann. Statist.* **33** 381–413. [MR2157807](#)

R. M. DUDLEY
ROOM 2-245
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139
USA
E-MAIL: rmd@math.mit.edu

S. SIDENKO
400 ARGYLE RD., APT. LF4
BROOKLYN, NEW YORK 11218
USA
E-MAIL: sidenko@alum.mit.edu

Z. WANG
DEPARTMENT OF MATHEMATICS
JOHNS HOPKINS UNIVERSITY
3400 N. CHARLES ST.
BALTIMORE, MARYLAND 21218
USA
E-MAIL: zwang@math.jhu.edu