Bernoulli **10**(3), 2004, 397–419

Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean

JEAN-CLAUDE MASSÉ

Département de Mathématiques et de Statistique, Université Laval, Sainte-Foy, Québec, Canada, G1K 7P4. E-mail: jcmasse@mat.ulaval.ca

We describe the asymptotic behaviour of the empirical Tukey depth process. It is seen that the latter may not converge weakly, even though its marginals always do. Closed subsets of the index set where weak convergence does occur are identified and a necessary and a sufficient condition for the asymptotic normality of the marginals is given. As an application, asymptotic normality of a Tukey depth-based multivariate trimmed mean is obtained for smooth distributions.

Keywords: Brownian bridge; empirical process; multidimensional trimmed mean; Tukey depth

1. Introduction

Let F be a probability distribution on $\mathbb{R}^d$, and let $\mathcal{H}$ denote the class of closed half-spaces H in $\mathbb{R}^d$. The *Tukey depth* (or *half-space depth*) of a point $x \in \mathbb{R}^d$ (with respect to F) is defined as

$$D(x) := \inf\{FH : H \in \mathcal{H}, x \in H\}.$$

In the following, when necessary the notation $D(x, F)$ will be used to indicate the dependence on F.

Tukey (1975) introduced the notion of the depth of a point in a multivariate data set. The Tukey depth is but one of several depth functions that have been proposed to measure the degree of centrality of a multidimensional point with respect to a probability distribution. Thus, Liu (1990) defines the simplicial depth $SD(x)$ of a point x by taking

$$SD(x) := \int \dots \int I(x \in S[y_1, \dots, y_{d+1}])\mathrm{d}F^{d+1}(y_1, \dots, y_{d+1}),$$

where $S[y_1, \dots, y_{d+1}]$ is the closed d-dimensional simplex with vertices $y_1, \dots, y_{d+1}$ (see also Liu and Singh 1993; Liu *et al.* 1999). Two other well-known depth functions are the so-called Mahalanobis and Oja depths. The Mahalanobis depth of a point x is defined to be

$$MD(x) := [1 + (x - \mu_F)'\Sigma_F^{-1}(x - \mu_F)]^{-1},$$

where μ_F and Σ_F denote the mean and covariance matrix of F respectively, and the latter functionals are assumed to exist; as for the Oja depth of x, it is defined as

1350–7265

$$OD(x) := \left[1 + \int \ldots \int \mathrm{Vol}(S[x, y_1, \ldots, y_d])\mathrm{d}F^d(y_1, \ldots, y_d)\right]^{-1},$$

where $\mathrm{Vol}(S[x, y_1, \ldots, y_d])$ denotes the volume of the simplex $S[x, y_1, \ldots, y_d]$. A systematic study of the concept of the depth function can be found in Zuo and Serfling (2000), where other examples are also introduced.

For any depth function, points with high depth are viewed as being close to the 'centre' of the distribution, whereas those with low depth are seen as belonging to the tails of the distribution. The Tukey depth has been used by Donoho and Gasko (1992) to define multivariate location estimators, and by Yeh and Singh (1997) to construct bootstrap confidence regions for multivariate parameters. For general depth functions, location and scale statistics have been studied by Liu *et al*. (1999) and Serfling (2000). Depth functions have also been applied to discriminant and cluster analysis by Ruts and Rousseeuw (1996), and to quality control by Liu and Singh (1993).

Depth functions provide a convenient tool for ranking multivariate observations. Given a sample $X_1, \ldots, X_n$ from F, the empirical distribution function $\hat{F}_n$ determines a centre-inward ranking through the ordering

$$D_n(X_{[1]}) \leqslant D_n(X_{[2]}) \leqslant \ldots \leqslant D_n(X_{[n]}),$$

where D_n is the empirical Tukey depth, that is, the Tukey depth with respect to $\hat{F}_n$. Such a ranking makes it possible to extend to a multivariate setting the concept of the L-statistic. Indeed, for a suitable weight function $W : [0, 1] \to [0, 1]$, a multidimensional L-statistic can be defined as the weighted average

$$L(\hat{F}_n) := \sum_i X_i W(D_n(X_i)) / \sum_i W(D_n(X_i)). \tag{1}$$

Such statistics include multidimensional versions of familiar location estimators such as trimmed means and Winsorized means.

In a Monte Carlo experiment, Massé and Plante (2003) have compared ten bivariate location estimators from the standpoint of accuracy and robustness. The comparison was made under various sampling situations determined by three sample sizes (10, 50, 200) and 14 centrally symmetric distributions ranging from the uniform and normal to heavy-tailed distributions such as the Cauchy and various mixtures distributions. Besides the arithmetic mean and five bivariate medians, two depth-based trimmed means were investigated, including one based on Tukey's depth. The simulation showed that, for moderate and large sample sizes, the accuracy of the Tukey depth-based trimmed means is overall as good as or better than any other of the estimators retained for the study. In addition, assessing robustness according to the performance on heavy-tailed distributions, it was seen that, for moderate and high sample sizes, the Tukey depth-based trimmed means are as good as if not better than any other estimator in the study, including the spatial median, the Tukey median and the Oja median.

This paper is organized as follows. Section 2 covers the limit theorems describing the asymptotic behaviour of the empirical Tukey depth process. Section 3 applies the foregoing limit theorems by deriving the asymptotic normality of a multidimensional trimmed mean of the form (1) for sufficiently smooth distributions. Section 4 examines properties of the Tukey depth that are needed to understand the asymptotics of its empirical depth process; examples illustrating these properties are also presented. Proofs of the main theorems are contained in Section 5, the main tool used being empirical processes theory.

In the following, we shall assume that all our random variables and vectors live on a fixed complete probability space $(\Omega, \mathcal{F}, \mathcal{P})$.

2. Limit theorems for the depth process

Clearly, $D(x)$ is determined by the class of closed half-spaces H such that $x \in \partial H$, the topological boundary of H. Let $U := \{u \in \mathbb{R}^d : |u| = 1\}$. Writing $H[x, u] := \{y \in \mathbb{R}^d : u' \cdot y \geqslant u' \cdot x\}$ for $u \in U$, it is seen that

$$D(x) = \inf_{u \in U} FH[x, u].$$

Simple examples show that the above infimum is not necessarily attained. For instance, let $F := (F_1 + F_2)/2$, where F_1 is uniform over one side of an equilateral triangle and F_2 is the unit mass on the vertex opposite. Then, the depth of the centroid is $1/4$, which is strictly less than FH for any closed half-plane H whose border goes through the centroid.

The probability distribution F is said to satisfy *smoothness condition* (S) if:

(S) $F(\partial H) = 0$ for all H in $\mathcal{H}$.

Condition (S) is useful to obtain stronger properties for the Tukey depth and its applications. It is satisfied, for instance, if F is absolutely continuous with respect to the Lebesgue measure. The two conditions are not, however, equivalent. Indeed, consider $F = F_1 \times F_2$, where F_1 and F_2 are both continuous distributions on $\mathbb{R}$ and F_2 is singular. Obviously F is not absolutely continuous, but Fubini's theorem implies that F satisfies (S).

According to Proposition 4.5(i) below, if F satisfies (S), then the Tukey depth of a point x can be expressed as the probability of some closed half-space whose boundary goes through x. If $D(x) = FH[x, v]$, we shall say that $H[x, v]$ is a *minimal half-space* at x and v a *minimal direction* at the same point.

A minimal half-space at a point is not necessarily unique. In what follows, points in $\mathbb{R}^d$ are classified into two types according to their multiplicity. Point x is said to be (F)-*smooth* if x has depth 0 or if there exists a unique minimal half-space at x; otherwise x is said to be (F)-*rough*. Let $\mathcal{S}_{\mathcal{F}}$ $(\mathcal{R}_{\mathcal{F}})$ denote the set of points of $\mathbb{R}^d$ that are smooth (rough). Examples 4.1 and 4.2 below illustrate the distinction between smooth and rough points.

The empirical Tukey depth process is defined to be

$$L_n(x) := n^{1/2}[D_n(x) - D(x)], \qquad x \in \mathbb{R}^d.$$

Let $\ell^\infty(T)$ denote the space of bounded real functions on T equipped with the uniform norm:

$$\|a\|_\infty := \sup_{t \in T} |a(t)|.$$

It is easily seen that $\{L_n(x), x \in \mathbb{R}^d\}$ is a stochastic process with bounded sample paths, allowing us henceforth to view L_n as a map into $\ell^\infty(\mathbb{R}^d)$.

Weak convergence of a sequence of maps such as $\{L_n\}$ will be understood in the Hoffmann–Jørgensen sense (see van der Vaart and Wellner 1996, Definition 1.3.3). Given that X, X_n, $n = 1, 2, \ldots,$ are maps from $(\Omega, \mathcal{F}, \mathcal{P})$ into a metric space S and that X is Borel measurable, $\{X_n\}$ is said to converge weakly to X if $\mathrm{E}^* f(X_n) \to \mathrm{E} f(X)$ for every bounded continuous real-valued function f defined on S, where E^* denotes outer expectation in the event that X_n may not be Borel measurable. In what follows, weak convergence is denoted by $X_n \rightsquigarrow X$.

The empirical process associated with F is defined to be $\nu_n := n^{1/2}[\hat{F}_n - F]$, viewed here as a map into the space $\ell^\infty(\mathcal{H})$. It is known (see Pollard 1984, Theorem VII.21) that $\nu_n \rightsquigarrow \nu_F$, where ν_F is an F-Brownian bridge, that is, a centred Gaussian tight Borel measurable map into $\ell^\infty(\mathcal{H})$ with the covariance function $P[\nu_F(H)\nu_F(H')] = F(H \cap H') - F(H)F(H')$; furthermore, a version of ν_F can be chosen such that each sample path is uniformly continuous with respect to the $L^2(F)$ seminorm on $\mathcal{H}$.

In addition to being smooth, probability distributions for which limit theorems for the empirical depth process are obtained have to satisfy a condition involving the multiplicity of minimal directions at each point. For every $x \in \mathbb{R}^d$, let $V(x)$ denote the collection of all minimal directions at x. The condition of *local regularity* (LR) is said to hold for F if:

(LR) F satisfies (S) and, for every x of positive depth, either $V(x)$ has a finite number of elements or $V(x) = U$.

The following limit theorem is the first part of our description of the asymptotic behaviour of the empirical Tukey depth process L_n. In the statement, we use the notation

$$J(\nu_F)(x) := \inf_{v \in V(x)} \nu_F H[x, v], \qquad x \in \mathbb{R}^d.$$

Theorem 2.1. *Let F satisfy condition* (LR). *Then, for any closed subset A of $\mathbb{R}^d$ having no accumulation point in $\mathcal{R}_{\mathcal{F}}$, $L_n \rightsquigarrow J(\nu_F)$ in $\ell^\infty(A)$. The limit process $J(\nu_F)$ is a tight Borel measurable map into $\ell^\infty(A)$.*

The following two corollaries are immediate.

Corollary 2.2. *Let F satisfy* (LR). *Then, for fixed x, $\{L_n(x)\}$ is asymptotically distributed as:*

(i) *$N(0, \alpha(1-\alpha))$ if x is F-smooth and $D(x) = \alpha > 0$;*
(ii) *$J(\nu_F)(x) \equiv \inf_{v \in V(x)} \nu_F H[x, v]$ if x is F-rough or $D(x) = 0$.*

Corollary 2.3. *Let F satisfy* (LR) *and suppose A is a closed subset of $\mathcal{S}_{\mathcal{F}}$. For each $x \in \mathcal{S}_{\mathcal{F}}$, let $H[x]$ be a minimal closed half-space at x – where $H[x]$ is uniquely defined if $D(x) > 0$. Then, in $\ell^\infty(A)$, $\{L_n\}$ converges weakly to the centred Gaussian process $J(\nu_F)$ with covariance kernel*

$$P[J(\nu_F)(x)J(\nu_F)(x')] = F(H[x] \cap H[x']) - FH[x]FH[x'], \qquad x, x' \in A.$$

Under the hypotheses of Corollary 2.3, consider the pseudometric $\rho(x, x') := F(H[x]\Delta H[x'])$ defined on A. Then there exists a version of $J(\nu_F)$ whose sample paths all belong to $C_\rho(A)$, the set of all bounded uniformly continuous real functions on A with respect to ρ. Viewed as a subspace of the metric space $\ell^\infty(A)$, $C_\rho(A)$ is separable and complete (van der Vaart and Wellner 1996, Chapter 1.5).

Thus, under regularity conditions, weak convergence of the empirical Tukey depth process holds if the index set is restricted to be a closed set containing smooth points and/or isolated rough points. The limit process is not necessarily Gaussian (see Remark 2.1), even though some of its marginals may be so; furthermore, the next theorem shows that weak convergence may not hold on the whole index set. For a full picture of the asymptotic behaviour of the sequence $\{L_n\}$, we need to take a closer look at the effect of roughness on tightness. We do so by describing what may happen when there exists one rough point that can be approached by arbitrarily close smooth points. For the sake of simplicity, the result is stated and proved in the case where $d = 2$.

Theorem 2.4. *Let F be a probability distribution on $\mathbb{R}^2$ which satisfies condition* (LR). *Suppose there exists $x_0 \in \mathcal{R}_{\mathcal{F}}$ such that $x_n \to x_0$ for some sequence $\{x_n\}$ in $\mathcal{S}_{\mathcal{F}}$ and such that $J(\nu_F)(x_0)$ is non-Gaussian or non-zero almost surely. Then, as a sequence of processes indexed by $\mathbb{R}^d$, $\{L_n\}$ is not asymptotically tight, hence $L_n \rightsquigarrow J(\nu_F)$ does not hold.*

Remark 2.1. The conditions on F appearing in Theorem 2.4 are satisfied whenever x_0 is the point of symmetry of an absolutely continuous elliptically symmetric distribution. Indeed, as seen in Example 4.4, $J(\nu_F)(x_0)$ is then a non-positive random variable, and hence is non-Gaussian. Thus in this case the empirical Tukey depth process does not converge weakly on the whole of $\mathbb{R}^d$, but may converge in the same sense on some smaller index sets. It is conjectured that $J(\nu_F)(x_0)$ is non-Gaussian whenever x_0 is rough or, in other words, $\inf\{\nu_F H[x_0, v] : v \in V(x_0)\}$ is non-Gaussian when $V(x_0)$ is not a singleton.

Remark 2.2. There is a sharp contrast between the asymptotic behaviour of the empirical Tukey depth process and that of the empirical simplicial depth process. For any F, it is known that the latter is asymptotically Gaussian (Arcones and Giné 1993, Corollary 6.8; see also Dümbgen 1992, Theorem 2).

3. Asymptotic normality of a multivariate trimmed mean

It is known that the Tukey depth attains a strictly positive supremum $\alpha^* \equiv \alpha^*(F) := \sup_x D(x)$ and that trimmed regions $Q_\alpha := \{x : D(x) \geqslant \alpha\}$ are non-empty compact convex subsets of $\mathbb{R}^d$ for every $\alpha \in (0, \alpha^*]$ (see Section 4).

Let $W : [0, 1] \to \mathbb{R}$ be a Borel measurable weight function such that, for some α_0 such that $0 < \alpha_0 < \alpha^*$, $W(a) = 0$ if $a \leqslant \alpha_0$. Then provided $0 < |\int W(D(x))F(\mathrm{d}x)| \leqslant \int |W(D(x))|F(\mathrm{d}x) < \infty$, an $\mathbb{R}^d$-valued functional is well defined through

$$L(F) := \int xW(D(x))F(\mathrm{d}x) \Big/ \int W(D(x))F(\mathrm{d}x).$$

Given an independent and identically distributed sample $X_1, \ldots, X_n$ from F, and $\hat{F}_n$ the corresponding empirical distribution, $L(\hat{F}_n)$ is a measurable statistic which can be seen as a multivariate trimmed mean. Here we investigate the properties of $L(\hat{F}_n)$ as an estimator of $L(F)$.

Proposition 3.1. *Assume that W is continuous. Then $L(\hat{F}_n)$ is strongly consistent, that is, $L(\hat{F}_n) \to L(F)$ almost surely.*

Proof. First we show that the difference between numerators converges to 0. Observe that

$$\begin{aligned} &\left| \int xW(D_n(x))\hat{F}_n(\mathrm{d}x) - \int xW(D(x))F(\mathrm{d}x) \right| \\ &\quad \leqslant \left| \int x[W(D_n(x)) - W(D(x))]\hat{F}_n(\mathrm{d}x) \right| + \left| \int xW(D(x))(\hat{F}_n - F)(\mathrm{d}x) \right|, \end{aligned}$$

and let I_{1n}, I_{2n} denote the integrals on the right-hand side. For any $\epsilon > 0$, take η such that

$$\eta \int_{Q_{\alpha_0/2}} |x| F(\mathrm{d}x) < \epsilon.$$

Strong uniform consistency of $\{D_n\}$ (Proposition 4.4) and uniform continuity of W then imply that there exists $0 < \delta < \alpha_0/2$ such that, almost surely for n large enough, $\|D_n - D\|_\infty < \delta$, so that $\|W(D_n(\cdot)) - W(D(\cdot))\|_\infty < \eta$. Furthermore, for n large enough,

$$|I_{1n}| \leqslant \eta \int_{Q_{\alpha_0/2}} |x| \hat{F}_n(\mathrm{d}x);$$

indeed, if $x \notin Q_{\alpha_0/2}$ and n is large enough, then $W(D_n(x)) = W(D(x)) = 0$. Applying the strong law of large numbers, it follows that almost surely

$$\limsup_n |I_{1n}| < \epsilon,$$

hence $I_{1n} \to 0$ almost surely. Convergence of I_{2n} to 0 as well as convergence of the sequence of denominators follow in the same way, which completes the proof. □

A central limit theorem for the multivariate trimmed mean is now obtained. If x is F-smooth, let $H[x]$ again denote any minimal half-space at x.

Theorem 3.2. *Let F satisfy condition* (LR) *and $W : [0, 1] \to \mathbb{R}$ be continuously differentiable. Assume that there exists $\alpha_1 \in (\alpha_0, 1]$ such that $\mathcal{A}_F := \{x : D_F(x) \leqslant \alpha_1\} \subset \mathcal{S}_F$ and W is constant on $[\alpha_1, 1]$. Then*

$$n^{1/2}[L(\hat{F}_n) - L(F)] \rightsquigarrow N(0, \Sigma(F)),$$

where $\Sigma(F)$ is the covariance matrix of

$$[(X - L(F))W(D(X)) + B(X)]/\int W(D(x))F(\mathrm{d}x),$$

$\mathcal{L}(X) = F$ *and*

$$B(y) = \int (x - L(F))W'(D(x))1_{H[x]}(y)F(\mathrm{d}x).$$

Example 3.1. Recall that a probability distribution F is said to be *centrosymmetric* about μ (Donoho and Gasko 1992, p. 1806) if $F(\mu + B) = F(\mu - B)$ for any Borel set B in $\mathbb{R}^d$. For such a distribution, affine equivariance of D (Proposition 4.1) implies that $L(F) = \mu$. Clearly, the centre of symmetry μ is F-rough with depth at least $1/2$. Within the class of centrosymmetric distributions, some, like the elliptically symmetric ones, have no other rough point. In view of Theorem 3.2, such distributions give rise to asymptotically normal trimmed means provided W is taken constant on $[\alpha_1, 1]$, where $\alpha_0 < \alpha_1 < 1/2$. The multivariate trimmed mean $L(\hat{F}_n)$ can be regarded as an estimator of the centre of symmetry or, equivalently, the point of maximal depth. Any point of maximal depth is sometimes called a *half-space median* of F (see Chen 1995). Observe that if F is centrosymmetric about μ, it follows that $L(\hat{F}_n)$ is itself centrosymmetric about μ, therefore median unbiased in the sense of the half-space median.

Remark 3.1. The above estimator of the centre of symmetry is expected to be useful in situations where the mean does not exist. A similar trimmed mean has been proposed by Dümbgen (1992) for the simplicial depth. Both estimators are asymptotically normal affine equivariant multivariate trimmed means, but by virtue of the greater regularity of the empirical simplicial depth, Dümbgen's estimator remains applicable to weight functions that are not necessarily constant for large values of the depth. We are aware of another asymptotically normal estimator of the centre of symmetry based on the notion of depth, that of Arcones *et al*. (1994), which uses the simplicial median. Arcones (1995), Kim (1995) and van der Vaart and Wellner (1996, p. 395) have also obtained asymptotically normal multivariate trimmed means, but those do not explicitly use a notion of depth in their definition. There is no doubt that it would be interesting to compare the efficiencies of these estimators of location. As noted in Section 1, through a Monte Carlo study Massé and Plante (2003) have shown that in the bivariate case Tukey depth-based trimmed means compare most favourably with several alternative location estimators.

4. Basic properties of Tukey's depth and examples

The next four propositions are well known. The first one states that in a sense the Tukey depth function is coordinate-free.

Proposition 4.1 (Affine invariance property; Rousseeuw and Ruts 1999, P. 215). *Suppose*

$T : \mathbb{R}^d \to \mathbb{R}^d$ is an affine transformation of full rank, that is, $T = L + c$, where L is linear of full rank and c is a constant in $\mathbb{R}^d$. Then

$$D(x, F) = D(Tx, F \circ T^{-1}), \qquad x \in \mathbb{R}^d.$$

A real-valued function f defined on $\mathbb{R}^d$ is said to be *quasi-concave* (Dharmadikhari and Joag-Dev 1988, pp. 84–85) if $\{x : f(x) \geqslant c\}$ is convex for each real c.

Proposition 4.2 (Quasi-concavity property; Rousseeuw and Ruts 1999, Proposition 1). *As a function of x, D is quasi-concave.*

Proposition 4.3 (Maximality property; Rousseeuw and Ruts 1999, Proposition 7). *D attains its supremum.*

Proposition 4.4 (Strong uniform consistency property). *Almost surely,*

$$\lim_{n\to\infty} \|D_n - D\|_\infty = 0.$$

Proof. See Donoho and Gasko (1992, pp. 1816–1817). □

In the following, for any two sets A and B, let $A\Delta B$ denote the symmetric difference of A and B, that is $(A \backslash B) \cup (B \backslash A)$.

Proposition 4.5. *Let F satisfy condition* (S).

(i) *The function $(x, u) \mapsto FH[x, u]$ is continuous on $\mathbb{R}^d \times U$.*
(ii) *The function $x \mapsto D(x)$ is continuous.*

Proof. (i) Indeed, let $\{x_n\}$, $\{u_n\}$ be sequences in $\mathbb{R}^d$ and U respectively, such that $x_n \to x$ and $u_n \to u$. Then

$$|FH[x_n, u_n] - FH[x, u]| \leqslant F(H[x_n, u_n]\Delta H[x, u]);$$

hence continuity follows by condition (S) and dominated convergence.

(ii) This is a straightforward extension of the proof of Lemma 6.1 in Donoho and Gasko (1992) for absolutely continuous distributions. □

Remark 4.1. Clearly D is not continuous at x_0 if x_0 is an atom. If F is non-atomic but such that $F(\partial H) > 0$ for some $H \in \mathcal{H}$, then: (i) for some $x \in \partial H$, $u \mapsto FH[x, u]$ is not continuous; (ii) for $u \perp \partial H$, $x \mapsto FH[x, u]$ is not continuous on ∂H.

According to Proposition 4.3, we may set

$$\alpha^* \equiv \alpha^*(F) := \max_{\mathbb{R}^d} D(x).$$

For any $\alpha \in (0, \alpha^*]$, let

$$Q_\alpha := \{x : D(x) \geqslant \alpha\}.$$

As α increases from 0 to α^*, Propositions 4.2 and 4.5 above and Proposition 5 in Rousseeuw and Ruts (1999) imply that the Q_αs form a decreasing family of non-empty compact convex subsets of $\mathbb{R}^d$.

The next proposition points to a close relationship between the minimal half-spaces and the compact convex Q_αs. Recall that a hyperplane h is said to support a set C at a point $x \in \partial C$ if $x \in h$ and C is contained in one of the two closed half-spaces having h as their boundary (Valentine 1964, Definition 2.8). Let $S(F)$ denote the support of F, that is, the set of points whose neighbourhoods all have positive F-probability.

Proposition 4.6. *Let F satisfy* (S)*, and assume $S(F)$ is connected. If $H[x, v]$ is a minimal half-space at x such that $0 < D(x) = FH[x, v] = \alpha < 1$, then $x \in \partial Q_\alpha$ and $\partial H[x, v]$ is a hyperplane of support of Q_α at x.*

Proof. To prove that $x \in \partial Q_\alpha$, take any neighbourhood W of x. Let $y \in W \cap H[x, v]$ such that $v' \cdot y > v' \cdot x$. Then by the connectedness of $S(F)$, $FH[x, v] - FH[y, v] > 0$. Indeed, if not so, part of the support is contained in $H[y, v]$ (a set of probability α) and part of it in $H[x, v]^c$ (of probability $1 - \alpha$). This clearly implies that $S(F) \subseteq A \cup B$, where A is the interior of $H[(x + y)/2, v]$ and $B = H[(x + y)/2, v]^c$, which contradicts connectedness. Thus $D(y) < \alpha$, and so $x \in \partial Q_\alpha$. To show that $\partial H[x, v]$ is a hyperplane of support of Q_α at x, take any $z \in Q_\alpha$. If $D(z) > \alpha$, then clearly $v' \cdot z < v' \cdot x$. If $D(z) = \alpha$ and $v' \cdot z > v' \cdot x$, then the connectedness of $S(F)$ implies that $D(z) \leqslant FH[z, v] < FH[x, v] = \alpha$, a contradiction. This shows that $Q_\alpha \subset \{y : v' \cdot y \leqslant v' \cdot x\}$, hence $\partial H[x, v]$ is a hyperplane of support of Q_α at x. □

Remark 4.2. If $S(F)$ is not connected, it is not necessarily true that $D(x) = \alpha$ implies $x \in \partial Q_\alpha$. Indeed, let F be the mixture distribution on $\mathbb{R}^2$ given by $0.25\mathcal{U}_1 + 0.75\mathcal{U}_2$, where $\mathcal{U}_1, \mathcal{U}_2$ are uniform on the disks of radius 1/2 centred at $(-1, 0)$ and $(1, 0)$, respectively. Then $S(F)$ is not connected and $\{x : D(x) = 0.25\}$ strictly includes $\partial Q_{0.25}$.

Recall that a convex subset of $\mathbb{R}^d$ is said to be smooth at a point on its boundary if there is a unique hyperplane of support at that point (Valentine 1964, p. 134); otherwise the convex set is said to be rough at the point. For a compact convex set with a non-empty interior, it is known that the set of smooth boundary points is dense within the boundary (Eggleston 1969, p. 32). In the plane, a more precise result says that the set of rough boundary points is countable (Valentine 1964, Theorem 10.7). Under the hypotheses of Proposition 4.6, if x is of positive depth α, then the boundary of each of its minimal half-spaces is a hyperplane of support of Q_α at x. Thus, uniqueness of a minimal half-space at a point $x \in \partial Q_\alpha$ depends on the smoothness of ∂Q_α at x.

Examples 4.1 and 4.2 below show that some points of positive depth may possess several minimal half-spaces. Furthermore, Example 4.2 exhibits a bidimensional distribution such that Q_α $(0 < \alpha < \alpha^*)$ has three boundary points through which infinitely many hyperplanes (i.e. straight lines) of support pass while each such point has exactly two minimal half-planes.

Example 4.1. Let F be the elliptically symmetric distribution on $\mathbb{R}^d$ with density

$$f_{\mu,\Sigma}(x) \propto (\det \Sigma)^{-1/2} g([(x-\mu)'\Sigma^{-1}(x-\mu)]^{1/2}),$$

where $\mu \in \mathbb{R}^d$, Σ is a positive definite matrix and g is a non-negative function on $[0, \infty)$ whose support is an interval. Then Q_α is an ellipsoid given by

$$Q_\alpha = \{x : (x-\mu)'\Sigma^{-1}(x-\mu) \leqslant k(\alpha)\}$$

for some strictly decreasing continuous function $k : (0, 1/2] \to [0, \infty)$ such that $k(1/2) = 0$. Here, for any $0 < \alpha < 1/2$, each point x of depth α has a unique minimal half-space whose border is a tangent to the boundary ∂Q_α. Indeed, Proposition 4.6 applies and there exists a unique hyperplane of support of Q_α at x. Moreover, the point of symmetry μ is the unique point of maximal depth $1/2$, and each half-space $H[\mu, u]$, $u \in U$, is minimal at μ. In the terminology of Section 2, all points of depth less than $1/2$ are smooth, whereas the point of symmetry is rough.

Example 4.2. Let F be the uniform distribution on the solid equilateral triangle in $\mathbb{R}^2$. Then a simple geometric argument (Caplin and Nalebuff 1988, p. 794) shows that the unique point of maximal depth $\alpha^* = 4/9$ is the centroid. For $0 < \alpha < 4/9$, each point of the contour ∂Q_α belongs to a tangent line bounding a half-plane of probability α. Clearly, ∂Q_α is determined by the envelope of the family of all such bounding lines. Figure 1, drawn with Mathematica, illustrates some of the contours thus obtained. Each point of a contour has either exactly one hyperplane (i.e. straight line) of support, except for three points located on the axes of symmetry which have infinitely many such straight lines. Furthermore, each of the points on the axes of symmetry has exactly two minimal half-planes and all the other points on the contours have only one. Finally, among points of positive depth it can be seen that the centroid is alone in having exactly three minimal half-planes. In the terminology of Section 2,

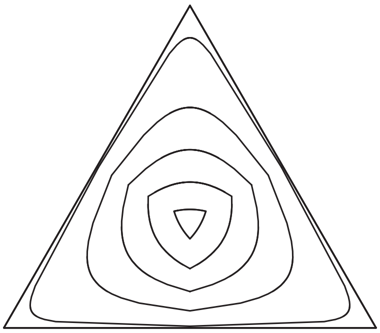

Figure 1. Contours of depths 0.01, 0.1, 0.2, 0.3, 0.4 for the uniform distribution

all points are smooth except for the centroid and three points on each contour of depth $< 4/9$.

If F satisfies (S) and $S(F)$ is connected, Proposition 4.6 shows that, for a point belonging to a contour, smoothness or roughness is determined by the shape of the contour at the point, specifically through the existence of exactly one or several hyperplanes of support at the point. To a great extent, the shape of the contours is governed by the degree of symmetry in the distribution. Exhibiting the highest level of symmetry, the elliptical distributions of Example 4.1 determine points of positive depth that are always smooth, except for the centre. Another instance of high symmetry is provided by the uniform equilateral triangle of Example 4.2. More generally, for each positive integer $n \geqslant 3$ one could easily show that all points of the solid uniform regular polygon with n sides are smooth except for the centroid and n rough points on each contour. Furthermore, maximal depth is attained at the centroid and the latter depth is 1/2 if and only if n is even (Donoho and Gasko 1992, Lemma 6.2). Rousseeuw and Ruts (1999) have studied the contours of solid uniform regular polygons, obtaining in particular that, for odd m,

$$\alpha^{\star} = \frac{1}{2} - \frac{(\tan(\pi/2m))^2}{2m}.$$

It has been seen that the centre of symmetry of an elliptically symmetric distribution is always rough. More generally, if F is the distribution of a random variable X such that, for some point μ, $(X-\mu)/\|X-\mu\|$ and $-(X-\mu)/\|X-\mu\|$ are identically distributed, then μ is a rough point. Indeed, $FH[\mu, u]$ is then constant as u varies in U. Such a distribution is said to be angularly symmetric about μ (Liu 1990, p. 409).

It is believed that all distributions have at least one rough point. The following proposition gives a sufficient condition under which a point of maximal depth is rough. If X has the distribution F and $u \in U$, let F_u denote the one-dimensional distribution of the projection $u' \cdot X$. According to Proposition 4.1, there is no loss of generality in assuming that 0 is a point of maximal depth.

Proposition 4.7. *For every $u \in U$, assume F_u has a density f_u such that, for some $R > 0$,*

$$\inf_{u \in U} \inf_{|t| \leqslant R} f_u(t) = \beta > 0.$$

Then $0 \in \mathcal{R}_F$.

Proof. As in Donoho and Gasko (1992, pp. 1818–1820), the conditions imply that there is a set of minimal directions $\{u_i, i \in J\}$ at 0, where J has cardinality between 2 and $d+1$. □

Remark 4.3. The condition in Proposition 4.7 is satisfied for instance if F has a density which is positive in the neighbourhood of 0.

Example 4.3. As the following example shows, a point of maximal depth is not necessarily rough, even if the maximal depth is greater than 1/2. Let F be the mixture distribution

$0.75\delta_0 + 0.25\mathcal{U}_0$, where $\mathcal{U}_0$ is uniform on a unit half-disk with diameter centred at 0. Then $D(0) = \alpha^* = 0.75$ and 0 has a unique minimal half-plane. Of course, F does not satisfy (S).

Example 4.4. To illustrate some of the results of Section 2, we return to Examples 4.1 and 4.2 and note that in each of them F satisfies (LR). In Example 4.1, Corollary 2.2 implies that, for each $x \neq \mu$ with $D(x) > 0$, $\{L_n(x)\}$ has a Gaussian limit distribution. This is no longer the case at the centre of symmetry μ. Indeed, since each closed half-space $H[\mu, u]$, $u \in U$, is minimal at μ, Corollary 2.2 implies that $L_n(\mu) = n^{1/2}[D_n(\mu) - 0.5]$ converges in distribution to $J(\nu_F)(\mu) \equiv \inf_{u \in U} \nu_F H[\mu, u]$, where, for all $u, v \in U$, $\nu_F H[\mu, u]$ is $N(0, 1/4)$, $P(\nu_F H[\mu, u]\nu_F H[\mu, v]) = F(H[\mu, u] \cap H[\mu, v]) - FH[\mu, u]FH[\mu, v]$. Since $\nu_F H[\mu, u]$ and $\nu_F H[\mu, -u]$ have a correlation of -1 for every u, it follows that the limit distribution at μ is non-positive, and therefore non-Gaussian.

In Example 4.2, all smooth points of positive depth determine a Gaussian limit. Each rough point x of depth $\alpha < 4/9$ located on the axes of symmetry has exactly two minimal half-planes. Denoting these half-planes by $H[x, v_1(x)]$ and $H[x, v_2(x)]$, the limit distribution is, according to Corollary 2.2, that of $\min\{\nu_F H[x, v_1(x)], \nu_F H[x, v_2(x)]\}$. As for the centroid C, it has three minimal half-planes, say $H[C, v_1(C)]$, $H[C, v_2(C)]$ and $H[C, v_3(C)]$. Acccording to Corollary 2, it follows that $L_n(C)$ converges in distribution to $\min\{\nu_F H[C, v_1(C)], \nu_F H[C, v_2(C)], \nu_F H[C, v_3(C)]\}$.

5. Proofs

5.1. Weak convergence in a not necessarily separable metric space

As indicated earlier, the empirical depth process L_n is viewed as a map into the metric space of bounded functions $\ell^\infty(\mathbb{R}^d)$. In what follows, the study of the asymptotic behaviour of L_n intimately involves that of the usual empirical process, also seen as a map into the non-separable space $\ell^\infty(\mathbb{R}^d)$. It is well known that in general the latter map is not Borel measurable (van der Vaart and Wellner 1996, p. 3), which led us in Section 2 to understand weak convergence in the Hoffman–Jørgensen sense. The corresponding notion of tightness is now presented.

Let $\{X_n\}$ be a sequence of not necessarily Borel measurable maps from $(\Omega, \mathcal{F}, P)$ into a metric space (S, d_0). Then $\{X_n\}$ is *asymptotically tight* (van der Vaart and Wellner 1996, Definition 1.3.7) if for every $\epsilon > 0$ there exists a compact subset K of S such that

$$\inf_{\delta > 0} \liminf_n P_*(X_n \in K^\delta) \geqslant 1 - \epsilon,$$

where P_* denotes inner probability and $K^\delta = \{y \in S : d_0(y, K) < \delta\}$. In general, asymptotic tightness and the usual notion of (uniform) tightness do not coincide; however, both notions are the same when (S, d_0) is separable (van der Vaart and Wellner 1996, p. 27).

All processes used in this paper can be regarded as maps into some space of bounded functions. Weak convergence of such maps can be studied with the help of the following two criteria, also taken from van der Vaart and Wellner (1996).

Proposition 5.1 (van der Vaart and Wellner 1996, Theorem 1.5.4). *Let X_n, $n = 1, 2, \ldots, X$, be maps from* $(\Omega, \mathcal{F}, P)$ *into* $\ell^\infty(T)$*, where* $\ell^\infty(T)$ *is the space of bounded functions on T equipped with the uniform norm and X is supposed to be Borel measurable and tight. Then $X_n \rightsquigarrow X$ if and only if $\{X_n\}$ is asymptotically tight and the marginals of X_n converge weakly to the marginals of X.*

Proposition 5.2 (van der Vaart and Wellner 1996, Theorem 1.5.6). *A sequence $\{X_n : (\Omega, \mathcal{F}, P) \to \ell^\infty(T)\}$ is asymptotically tight if and only if $\{X_n(t)\}$ is tight for every $t \in T$ and, for all $\epsilon, \eta > 0$, there exists a finite partition $T = \bigcup_{i=1}^k T_i$ such that*

$$\limsup_n P^* \left(\sup_i \sup_{s,t \in T_i} |X_n(s) - X_n(t)| > \epsilon \right) < \eta.$$

5.2. Proofs for Sections 2 and 3

If F satisfies (S), then it can be seen that $V(x)$ is a closed subset of U. If $v \in V(x)$, we say that a closed subset $U(v)$ of U is an *isolating set* for v if $U(v) \cap V(x) = \{v\}$. Existence of appropriate coverings of U by isolating sets for the minimal directions at points of positive depth is needed in most of the limit theorems that follow. Clearly, if condition (LR) holds, then such a covering exists: for each x such that $D(x) > 0$,

$$U = \bigcup\nolimits_{v \in V(x)} U_x(v)$$

for some isolating sets $U_x(v)$. In what follows, non-uniqueness of such coverings has no impact on the limit theorems.

The next lemma relates uniform convergence of the Tukey empirical depth with the strong uniform law of large numbers (Pollard 1984, Theorem II.14).

Lemma 5.3. *Suppose that F satisfies condition* (LR). *For each x of positive depth, let $\{U_x(v), v \in V(x)\}$ denote a covering of U by isolating sets. Then*

$$\sup_{\{x : D(x) > 0\}} \sup_{v \in V(x)} \left| \inf_{u \in U_x(v)} \hat{F}_n H[x, u] - D(x) \right| \leqslant \sup_{H \in \mathcal{H}} |(\hat{F}_n - F)H|.$$

Proof. This clearly follows from the fact that, for every $v \in V(x)$,

$$\begin{aligned}\left| \inf_{u \in U_x(v)} \hat{F}_n H[x, u] - D(x) \right| &= \left| \inf_{u \in U_x(v)} \hat{F}_n H[x, u] - \inf_{u \in U_x(v)} FH[x, u] \right| \\ &\leqslant \sup_{u \in U_x(v)} \left| (\hat{F}_n - F) H[x, u] \right| \\ &\leqslant \sup_{H \in \mathcal{H}} |(\hat{F}_n - F)H|.\end{aligned}$$

□

The following lemma is critical for the understanding of the asymptotic distribution of the empirical Tukey depth. In its proof, convergence of a sequence $\{H[x_n, u_n]\}$ to $H[x_0, u_0]$ in $\mathcal{H}$ means that $x_n \to x_0$ and $u_n \to u_0$. For any $A \subseteq \mathbb{R}^d$, write

$$\mathcal{H}_{\min}(A) := \{H[x, v] : x \in A, v \in V(x)\}.$$

If $H[x, v] \in \mathcal{H}_{\min}(A)$, let

$$\lambda_n H[x, v] := \begin{cases} n^{1/2}\Big(\inf\limits_{u \in U_x(v)} \hat{F}_n H[x, u] - FH[x, v]\Big), & D(x) > 0, \\ 0, & D(x) = 0, \end{cases}$$

thus defining a map λ_n into $\ell^\infty(\mathcal{H}_{\min}(A))$.

Lemma 5.4. *Let F satisfy condition* (LR). *Let A be a closed subset of $\mathbb{R}^d$ having no accumulation point in $\mathcal{R}_F$. Then $\{\lambda_n\}$ converges weakly to ν_F restricted to $\mathcal{H}_{\min}(A)$.*

Proof. Clearly, without loss of generality one can assume that all elements of A are of positive depth. For every x, let $\{U_x(v), v \in V(x)\}$ be a covering of U by isolating sets. For every $x \in A$ and $v \in V(x)$, choose $u_n(x, v) \in U_x(v)$ depending on $\omega \in \Omega$ such that

$$\hat{F}_n H[x, u_n(x, v)] = \inf_{u \in U_x(v)} \hat{F}_n H[x, u].$$

Then

$$\begin{aligned} \lambda_n H[x, v] &= n^{1/2}\big(\hat{F}_n H[x, u_n(x, v)] - \hat{F}_n H[x, v]\big) + \nu_n H[x, v] \\ &= \nu_n H[x, u_n(x, v)] + n^{1/2}(FH[x, u_n(x, v)] - FH[x, v]), \end{aligned}$$

and so

$$\nu_n H[x, u_n(x, v)] \leqslant \lambda_n H[x, v] \leqslant \nu_n H[x, v].$$

Since the restriction of ν_n to $\mathcal{H}_{\min}(A)$ converges weakly to the limit appearing in the statement, it is enough to show that $\nu_n(H[x, v] - H[x, u_n(x, v)]) = o_{P^*}(1)$ uniformly in $\mathcal{H}_{\min}(A)$ as $n \to \infty$, that is, for every $\epsilon > 0$,

$$\limsup_{n\to\infty} P^*(\sup\{|\nu_n(H[x, v] - H[x, u_n(x, v)])| : x \in A, v \in V(x)\} > \epsilon) = 0.$$

Note that outer probability is used here to guard against non-measurability problems.

For $m = 1, 2, \ldots,$ put

$$\mathcal{F}_m := \{H_1 - H_2 : H_1, H_2 \in \mathcal{H}, F(H_1 \Delta H_2) < 1/m\}.$$

It follows that, for all m,

$$P^*(\sup\{|\nu_n(H[x, v] - H[x, u_n(x, v)])| : x \in A, v \in V(x)\} > \epsilon)$$

$$\leqslant P\left(\sup_{f \in \mathcal{F}_m} |\nu_n f| > \epsilon\right) + P^*(\sup\{F(H[x, v]\Delta H[x, u_n(x, v)]) : x \in A, v \in V(x)\} \geqslant 1/m),$$

where measurability of the supremum on $\mathcal{F}_m$ is guaranteed by the existence of a countable dense subset of $\mathcal{F}_m$. Since $\mathcal{H}$ is a Vapnik–Chervonenkis class, hence a Donsker class, the asymptotic uniform equicontinuity of the empirical process entails that the first term on the right-hand side of the last inequality is less than $\epsilon/2$ if m and n are large enough (van der Vaart and Wellner 1996, Prob. 14, p. 152). The proof will conclude by showing that

$$\sup\{F(H[x, v]\Delta H[x, u_n(x, v)(\omega)]) : x \in A, v \in V(x)\}$$

converges almost uniformly to 0 in ω (van der Vaart and Wellner 1996, Definition 1.9.1), that is, for each $\delta > 0$ there is some $B \in \mathcal{F}$ with $P(B) > 1 - \delta$ such that

$$\sup_{\omega \in B} \sup\{F(H[x, v]\Delta H[x, u_n(x, v)(\omega)]) : x \in A, v \in V(x)\} \to 0$$

as $n \to \infty$. It is known that this implies convergence in outer probability (van der Vaart and Wellner 1996, Lemma 1.9.2), therefore the second term in the last inequality will be less than $\epsilon/2$ if n is large enough.

Suppose that the asserted almost uniform convergence does not hold. Then there exist $\delta_0 > 0$, $\epsilon_0 > 0$ and a sequence of positive integers $\{n_k\}$ such that, for any $B \in \mathcal{F}$ with $P(B) > 1 - \delta_0$,

$$\sup_{\omega \in B} \sup\{F(H[x, v]\Delta H[x, u_{n_k}(x, v)(\omega)]) : x \in A, v \in V(x)\} > \epsilon_0 \tag{2}$$

for $k = 1, 2, \ldots$. Now, according to the strong uniform law of large numbers, there exists a null set $N \in \mathcal{F}$ such that $\sup\{|(F - \hat{F}_n(\omega))H| : H \in \mathcal{H}\} < \epsilon_0/4$ if $\omega \in \Omega \backslash N$ and n is large enough. Egoroff's theorem (Dudley 1989, Theorem 7.5.1) says that this holds true uniformly on some measurable set with probability greater than $1 - \delta_0$. Choose such a B in (2). For $\omega \in B$ and n large enough, it follows from Lemma 5.3 that

$$\begin{aligned}
&\sup\{F(H[x, v]\Delta H[x, u_n(x, v)(\omega)]) : x \in A\backslash Q_{\epsilon_0/4}, v \in V(x)\} \\
&\quad\leqslant \sup\{FH[x, v] : x \in A\backslash Q_{\epsilon_0/4}, v \in V(x)\} \\
&\qquad + \sup\{FH[x, u_n(x, v)(\omega)] : x \in A\backslash Q_{\epsilon_0/4}, v \in V(x)\} \\
&\quad\leqslant \epsilon_0/2 + \sup\{\hat{F}_n(\omega)H[x, u_n(x, v)(\omega)] : x \in A\backslash Q_{\epsilon_0/4}, v \in V(x)\} \\
&\quad\leqslant \epsilon_0/2 + \sup\{D(x) : x \in A\backslash Q_{\epsilon_0/4}\} + \epsilon_0/4 \\
&\quad\leqslant \epsilon_0.
\end{aligned}$$

The foregoing implies that we may replace (2) by

$$\sup_{\omega\in B}\sup\{F(H[x,v]\Delta H[x,u_{n_k}(x,v)(\omega)]) : x \in A\cap Q_{\epsilon_0/4}, v\in V(x)\} > \epsilon_0$$

for $k = 1, 2, \ldots$. Hence we can find ω_k in B, x_k in $A\cap Q_{\epsilon_0/4}$, $v_k \in V(x_k)$ such that

$$F(H[x_k, v_k]\Delta H[x_k, u_{n_k}(x_k, v_k)(\omega_k)]) > \epsilon_0 \tag{3}$$

for $k = 1, 2, \ldots$.

By compactness, there exist subsequences $\{x_{k_j}\}$ in $A\cap Q_{\epsilon_0/4}$, $\{v_{k_j}\}$ and $\{u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j})\}$ in U, converging to some $x_0\in A\cap Q_{\epsilon_0/4}$, $v_0\in U$ and $\tilde{v}_0\in U$, respectively. Proposition 4.5 then implies that

$$D(x_0) = \lim_{j\to\infty} D(x_{k_j}) = \lim_{j\to\infty} FH[x_{k_j}, v_{k_j}] = FH[x_0, v_0],$$

so that $v_0\in V(x_0)$. Let us show that $\tilde{v}_0$ is also minimal at x_0. Since

$$\left|D(x_{k_j}) - \inf_{u\in U_{x_{k_j}}(v_{k_j})} \hat{F}_{n_{k_j}}(\omega_{k_j})H[x_{k_j}, u]\right| = |D(x_{k_j}) - \hat{F}_{n_{k_j}}(\omega_{k_j})H[x_{k_j}, u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j})]|$$

and

$$|(F-\hat{F}_{n_{k_j}}(\omega_{k_j}))H[x_{k_j}, u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j})]| \leqslant \sup_{\omega\in B}\sup_{H\in\mathcal{H}} |(F-\hat{F}_{n_{k_j}}(\omega))H|,$$

Proposition 4.5 and Lemma 5.3 imply that

$$\begin{aligned} FH[x_0, \tilde{v}_0] &= \lim_{j\to\infty} FH[x_{k_j}, u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j})] \\ &= \lim_{j\to\infty} \hat{F}_{n_{k_j}}(\omega_{k_j})H[x_{k_j}, u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j})] \\ &= D(x_0), \end{aligned}$$

hence $\tilde{v}_0\in V(x_0)$.

If $\tilde{v}_0 = v_0$, dominated convergence implies that (3) is violated for $k = k_j$ large enough, therefore we may assume that $\tilde{v}_0 \neq v_0$, so that $x_0\in\mathcal{R}_F$. Since A has no accumulation point in $\mathcal{R}_F$, x_0 has to be an isolated point and so $x_{k_j} = x_0$ for j large enough. Now, according to our assumptions, $V(x_0)$ either has isolated points only or is identical to U. In the first case, it is seen that $v_{k_j} = v_0 \in V(x_0)$ if j is large enough. Since

$$u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j}) \in U_{x_{k_j}}(v_{k_j}) = U_{x_0}(v_0)$$

for j large enough, it follows that $u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j}) \to \tilde{v}_0 = v_0$, a contradiction. Finally, if $V(x_0) = U$, $U_{x_0}(v) = \{v\}$ for all $v\in V(x_0)$, therefore $u_{n_{k_j}}(x_{k_j}, v_{k_j})(\omega_{k_j}) = v_{k_j}$ if j is large enough, leading again to $\tilde{v}_0 = v_0$. □

Write $\mathcal{H}_{\min}\equiv\mathcal{H}_{\min}(\mathbb{R}^d)$. If $\phi\in\ell^\infty(\mathcal{H}_{\min})$, let

$$J(\phi)(x) := \inf_{v\in V(x)} \phi H[x, v], \qquad x\in\mathbb{R}^d,$$

thus defining $J : \ell^\infty(\mathcal{H}_{\min}) \to \ell^\infty(\mathbb{R}^d)$. The proof of T\heorem 2.1 below will use the fact that

$L_n = J(\lambda_n)$. For any $A \subseteq \mathbb{R}^d$, the above obviously induces a map J_A from $\ell^\infty(\mathcal{H}_{\min}(A))$ into $\ell^\infty(A)$, where the two spaces of bounded functions are equipped with the uniform norm.

Lemma 5.5. *For any $A \subseteq \mathbb{R}^d$, J_A is a uniformly continuous map.*

Proof. Given $\phi_1, \phi_2 \in \ell^\infty(\mathcal{H}_{\min}(A))$,

$$\begin{aligned}\|J(\phi_1) - J(\phi_2)\|_A &= \sup_{x\in A}\left|\inf_{v\in V(x)} \phi_1 H[x, v] - \inf_{v\in V(x)} \phi_2 H[x, v]\right| \\ &\leqslant \sup_{x\in A}\sup_{v\in V(x)} |\phi_1 H[x, v] - \phi_2 H[x, v]| \\ &= \|\phi_1 - \phi_2\|_{\mathcal{H}_{\min}(A)},\end{aligned}$$

hence the lemma follows. □

Proof of Theorem 2.1. Since ν_F is tight Borel measurable, Lemma 5.5 implies that, as a map into $\ell^\infty(A)$, $J(\nu_F)$ is tight Borel measurable. Furthermore, because $\{x : D_n(x) = 0\}$ is almost surely open, one can see that the paths of L_n are almost surely 0 on $\{x : D(x) = 0\}$. This implies that almost surely $L_n = J(\lambda_n)$, therefore the first statement follows from Lemma 5.4 and the continuous mapping theorem (van der Vaart and Wellner 1996, Theorem 1.3.6). □

Let $\mathrm{aff}\{x_i, 1 \leqslant i \leqslant n\}$ ($\mathrm{span}\{x_i, 1 \leqslant i \leqslant n\}$) denote the affine (linear) space generated by $x_1, \ldots, x_n$ in $\mathbb{R}^d$. Recall that the affine space generated by $x_1, \ldots, x_n$ is defined by

$$\mathrm{aff}\{x_i, 1 \leqslant i \leqslant n\} := \left\{\sum_{i=1}^n \lambda_i x_i : \text{for all } \lambda_1, \ldots, \lambda_n \in \mathbb{R} \text{ such that} \sum_{i=1}^n \lambda_i = 1\right\}.$$

Furthermore, $x_1, \ldots, x_n$ are said to be affinely independent if none of these points belongs to the affine space generated by the remaining points.

Lemma 5.6. *Let $g : \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \mathbb{R}$ be the map defined by*

$$g(x_1, x_2, x_3) = 1_{\mathrm{aff}(x_2, x_3)}(x_1).$$

Then g is a Borel function.

Proof. Put $\underline{x} := (x_1, x_2, x_3)$ and $C := \{\underline{x} : x_2 \neq x_3\}$. Let g_C and g_{C^c} denote the restrictions of g to C and C^c, respectively. Since C is an open set, it suffices to check that both g_C and g_{C^c} are Borel functions (Dudley 1989, Lemma 4.2.4).

To prove that g_C is a Borel function, it is enough to show that it is upper semicontinuous or, equivalently, that $\{\underline{x} \in C : g_C(\underline{x}) = 0\}$ is open. Assume the contrary. Then for some $\underline{x}_0 \in C$ there exists a sequence $\{\underline{x}^n := \{(x_1^n, x_2^n, x_3^n)\}$ in C such that $\underline{x}^n \to \underline{x}^0$ and $g_C(\underline{x}^0) = 0 < g_C(\underline{x}^n) = 1$ for every n. Let $\{\lambda^n\}$ be a sequence in $\mathbb{R}$ such that $x_1^n = \lambda^n x_2^n + (1 - \lambda^n)x_3^n$ for every n. We claim that $\{\lambda^n\}$ is bounded. If not, one can assume without loss of generality that $|\lambda^n| \to \infty$. Then since

$$\lim_{n\to\infty} x_2^n + (\lambda^n)^{-1}(1-\lambda^n)x_3^n = x_2^0 - x_3^0 \neq 0,$$

one has

$$|x_1^0| = \lim_{n\to\infty} |\lambda^n||x_2^n + (\lambda^n)^{-1}(1-\lambda^n)x_3^n| = \infty,$$

a contradiction. Thus $\{\lambda^n\}$ is bounded and there exists some subsequence which converges to a real number λ^0. This implies that $x_1^0 = \lambda^0 x_2^0 + (1-\lambda^0)x_3^0$, which contradicts the hypothesis that $x_1^0 \notin \mathrm{aff}\{x_2^0, x_3^0\}$.

Finally, Borel measurability of g_{C^c} follows from the fact that

$$\{\underline{x} \in C^c : g_{C^c}(\underline{x}) = 1\} = \{\underline{x} \in C^c : \underline{x} = (x_1, x_1, x_1)\}$$

is a closed subset of C^c. □

Lemma 5.7. *Assume F is a probability distribution on $\mathbb{R}^2$ which satisfies condition* (S). *Let $I \subseteq \{1, 2, \ldots, m\}$ where m is fixed. Then for any I such that $1 \leqslant |I| \leqslant 3$, the X_i such that $i \in I$ are affinely independent almost surely. In particular:*

(i) *if $|I| = 2$, then $\mathrm{aff}\{X_i, i \in I\}$ is a straight line almost surely;*
(ii) *if $|I| = 3$, then $\mathrm{aff}\{X_i, i \in I\} = \mathbb{R}^2$ almost surely.*

Proof. To avoid trivialities, assume $|I| > 1$. For any $l \in I$, write $I_l := I\backslash\{l\}$. According to Lemma 5.6, $g(X_1, X_2, X_3)$ is measurable, hence condition (S) and Fubini's theorem yield

$$\begin{aligned} P(X_l \in \mathrm{aff}\{X_i, \in I_l\}) &= \int P(X_l \in \mathrm{aff}\{x_i, \in I_l\})\mathrm{d}F^{(|I|-1}(x_i, \in I_l) \\ &= 0, \end{aligned}$$

since any affine space $\mathrm{aff}\{x_i, \in I_l\}$ is contained in a straight line. □

Lemma 5.8. *Assume F is a probability distribution on $\mathbb{R}^2$ which satisfies condition* (S). *Let y be any fixed point in $\mathbb{R}^2$. Then, for every m, D_m is almost surely continuous at y.*

Proof. Put $\max_{x\in\mathbb{R}^2} D_m(x) = k_m^*/m$. For $k = 0, 1, \ldots, k_m^*$, define the empirical trimmed region

$$Q_k^m := \{x : D_m(x) \geqslant k/m\}$$

and let $\tilde{Q}_k^m$ denote the interior of Q_k^m. By upper semicontinuity of D_m (Donoho and Gasko 1992, Lemma 6.1), $(Q_k^m)^c$ is open, hence D_m is continuous in the open set

$$\left(\bigcup_{k=0}^{k_m^*-1} \tilde{Q}_k^m \backslash Q_{k+1}^m\right) \cup \tilde{Q}_{k_m^*}^m.$$

If y is a discontinuity point of D_m, then y belongs to some boundary ∂Q_k^m, where

$D_m(y) = k \geqslant 1$ because $\tilde{Q}_0^m = \mathbb{R}^2$. It will now be shown that this event occurs with probability 0.

Given that y belongs to ∂Q_k^m, $k \geqslant 1$, let (y_n) in $(Q_k^m)^c$ be such that $y_n \to y$. For every n, let $H[y_n, u_n]$ be a minimal closed half-plane at y_n. By compactness, it can be assumed without loss of generality that $H[y_n, u_n] \to H[y, u_0]$ for some u_0. For every n and $x \in Q_k^m$, $u_n' \cdot y_n > u_n' \cdot x$, hence $\partial H[y, u_0]$ is a straight line of support of Q_k^m at y; we next show that this line contains at least one data point.

Because y has depth k, $\hat{F}_m H[y, u_0] = (k + p)/m$ for some $p \geqslant 0$. Put

$$I := \{i \in \{1, 2, \ldots, m\} : u_0' \cdot X_i = u_0' \cdot y\},$$
$$J := \{i \in \{1, 2, \ldots, m\} : u_0' \cdot X_i > u_0' \cdot y\}$$

and

$$K := \{i \in \{1, 2, \ldots, m\} : u_0' \cdot X_i < u_0' \cdot y\}.$$

We need to prove that I is non-empty. This is obvious if J is empty, hence assume $|J| \geqslant 1$. Then by continuity, $u_n' \cdot X_i > u_n' \cdot y_n$ for every $i \in J$ if n is large enough. Since $H[y_n, u_n]$ is minimal at $y_n \in (Q_k^m)^c$, it follows that $|J| = k + p - |I| \leqslant \hat{F}_m H[y_n, u_n] < k$, hence $|I| > p$.

If $y \in \partial Q_k^m$, the above shows that, for some non-empty I, $\{y\} \cup \operatorname{aff}\{X_i, i \in I\}$ is contained in a straight line of support of Q_k^m at y. Two events may then occur: either $y \in \operatorname{aff}\{X_i, i \in I\}$, or $y \notin \operatorname{aff}\{X_i, i \in I\}$. Since $\operatorname{aff}\{X_i, i \in I\}$ is contained in a line, Lemma 5.7 implies that the first (second) event has probability 0 if $|I| > 2$ ($|I| > 1$). For smaller values of $|I|$, it is next shown that both events again have probability 0. Since there is are finite number of equally likely choices for the random set I, without loss of generality we assume in the following that I is fixed.

Let $y \in \operatorname{aff}\{X_i, i \in I\}$ with $1 \leqslant |I| \leqslant 2$. For any $i \in I$, $P(X_i = y) = 0$, hence we may assume $|I| = 2$. Note that

$$y \in \operatorname{aff}\{X_i, i \in I\} \Longleftrightarrow X_l \in \operatorname{aff}\{y, X_i, i \in I_l\} \text{ for some } l \in I.$$

The argument used in the proof of Lemma 5.6 implies that the indicator function of the event $X_l \in \operatorname{aff}\{y, X_i, i \in I_l\}$ is measurable. Thus, another application of (S) and Fubini's theorem yields

$$\begin{aligned} P(y \in \operatorname{aff}\{X_i, i \in I\}) &= P(X_l \in \operatorname{aff}\{y, X_i, i \in I_l\} \text{ for some } l \in I) \\ &= \sum_{l \in I} P(X_l \in \operatorname{aff}\{y, X_i, i \in I_l\}) \\ &= 0. \end{aligned}$$

Now let $I = \{l\}$ and assume $y \notin \operatorname{aff}\{X_l\}$ (i.e. $y \neq X_l$), where y and X_l both belong to a straight line of support of Q_k^m at y. Suppose this event occurs with positive probability. As above, one has

$$P(y \in \operatorname{span}\{X_l\}) = P(y \in \operatorname{aff}\{0, X_l\}) = 0,$$

so that y and X_l are linearly independent almost surely. Thus, with positive probability there exists $w \in \mathbb{R}^2$ such that $w' \cdot y > 0$ and $w' \cdot X_l < 0$. Again, let $\partial H[y, u_0]$ denote the straight line of support at y defining I, J and K. For every n large enough, put

$$v_n := \frac{u_0 + w/n}{|u_0 + w/n|} \in U,$$

so that $v_n \to u_0$ as $n \to \infty$. Then, by continuity, for n large enough, $v_n' \cdot X_i > v_n' \cdot y$ for every $i \in J$, and $v_n' \cdot X_i < v_n' \cdot y$ for every $i \in K$; furthermore, by the choice of w, for all n, $v_n' \cdot X_l < v_n' \cdot y$. As seen above, $|J| < k$, therefore, for n large enough, $D_m(y) \leqslant \hat{F}_m H[y, v_n] = |J|/m < k/m$. This contradiction establishes that the second event holds with probability 0, which concludes the proof. □

Proof of Theorem 2.4. By Proposition 5.1, it suffices to check that $\{L_n\}$ is not asymptotically tight when its index set is restricted to some neighbourhood W of x_0. For every $x \in \mathbb{R}^d$, $\{L_n(x)\}$ is a convergent sequence of real-valued Borel measurable maps, hence is tight. According to Proposition 5.2, the assertion follows if we can find $\epsilon_0 > 0$ and $\eta_0 > 0$ such that, for any partition $W = \bigcup_{i=1}^k T_i$,

$$\limsup_n P^* \left(\sup_i \sup_{x,y \in T_i} |L_n(x) - L_n(y)| > \epsilon_0 \right) \geqslant \eta_0.$$

Let $\{x_n\}$ be any sequence in $\mathcal{S}_F$ such that $x_n \to x_0$. Then $J(\nu_F)(x_n)$ is Gaussian for every n, hence $J(\nu_F)(x_n) \not\to J(\nu_F)(x_0)$ in probability, since otherwise $J(\nu_F)(x_0)$ would have to be Gaussian or 0 almost surely. Thus there exist $\epsilon_0, \eta_0 > 0$ and sequences of increasing integers $\{n_k\}$ and $\{p_k\}$ such that

$$P(|J(\nu_F)(x_{n_k}) - J(\nu_F)(x_{p_k})| > \epsilon_0) \geqslant \eta_0, \qquad k = 1, 2, \ldots.$$

Without loss of generality, it can be assumed that for all k, $x_{n_k}, x_{p_k} \in T_i$ for some fixed i. For any fixed m and $y = x_0$, Lemma 5.2 shows that $L_m(x_{n_j}) \to L_m(x_0)$ almost surely as $j \to \infty$. For every k, weak convergence of marginals thus implies

$$\begin{aligned}\eta_0 &\leqslant \lim_{m\to\infty} P(|L_m(x_{n_k}) - L_m(x_{p_k})| > \epsilon_0) \\ &\leqslant \liminf_{m\to\infty} P^* \left(\sup_i \sup_{x,y \in T_i} |L_m(x) - L_m(y)| > \epsilon_0 \right),\end{aligned}$$

which concludes the proof. □

Proof of Theorem 3.2. The proof follows closely that of Dümbgen (1992, Theorem 3). For some $\theta_n(x)$ between $D(x)$ and $D_n(x)$, Taylor's theorem gives

$$\begin{aligned}&n^{1/2} \int (x - L(F)) W(D_n(x)) \hat{F}_n(\mathrm{d}x) \\ &= \int (x - L(F)) W'(\theta_n(x)) L_n(x) \hat{F}_n(\mathrm{d}x) + \int (x - L(F)) W(D(x)) \nu_n(\mathrm{d}x),\end{aligned}$$

where ν_n is the normalized empirical measure and

$$\|\theta_n - D\|_\infty \leqslant R_n := \|D_n - D\|_\infty. \tag{4}$$

We shall prove that

$$\int (x - L(F))W'(\theta_n(x))L_n(x)\hat{F}_n(\mathrm{d}x) = \int B(y)\nu_n(\mathrm{d}y) + o_P^*(1).$$

First, for $0 < \delta \leqslant 1$, let

$$\omega(\delta) := \sup\{|W'(t_1) - W'(t_2)| : t_1, t_2 \in [0, 1], |t_1 - t_2| \leqslant \delta\}$$

denote the modulus of continuity of W'. Almost surely, for any x such that $D(x) \leqslant \alpha_0/2$ and n large enough, it follows from (4) and Proposition 4.4 that $\theta_n(x) \leqslant \alpha_0$, so that

$$\begin{aligned}
&\left|\int (x - L(F))[W'(\theta_n(x)) - W'(D(x))]L_n(x)\hat{F}_n(\mathrm{d}x)\right| \\
&\leqslant \omega(R_n)\|L_n\|_\infty \int_{Q_{\alpha_0/2}} |x - L(F)|\hat{F}_n(\mathrm{d}x) \\
&= o_P^*(1).
\end{aligned}$$

Indeed, this follows from the uniform continuity of W', the strong law of large numbers and the boundedness in probability of $\|L_n\|_\infty$:

$$\inf_{H \in \mathcal{H}} \nu_n H \leqslant L_n(x) \leqslant \sup_{H \in \mathcal{H}} \nu_n H, \qquad x \in \mathbb{R}^d.$$

Next, since F satisfies (LR), Corollary 2.3 implies that $\{L_n\}$ is asymptotically tight on $\mathcal{A}_F := \{x : D(x) \leqslant \alpha_1\}$; therefore, for any $\epsilon > 0$, there exists a compact subset K of $C_\rho(\mathcal{A}_F)$ such that

$$\liminf_n P_*(L_n \in K^{\epsilon/2}) \geqslant 1 - \epsilon.$$

By total boundedness, K can be covered by a finite union of balls of radius $\epsilon/2$ centred at $f_1, \ldots, f_p \in K$. This means that with inner probability at least $1 - \epsilon$ and n large enough, $\min_i \|L_n - f_i\|_{\mathcal{A}_F} < \epsilon$. Since W is constant on $[\alpha_1, 1]$, it follows that

$$\begin{aligned}
&\left|\int (x - L(F))W'(D(x))L_n(x)(\hat{F}_n - F)(\mathrm{d}x)\right| \\
&\leqslant 2\epsilon \max_{x \in \mathcal{A}_F} |(x - L(F))W'(D(x))| + \max_i \left|\int_{\mathcal{A}_F} (x - L(F))W'(D(x))f_i(x)(\hat{F}_n - F)(\mathrm{d}x)\right|,
\end{aligned}$$

hence we conclude that the left-hand term of the last inequality is $o_P^*(1)$.

Now, since $\|L_n - \nu_n H[\cdot]\|_{\mathcal{A}_F} = o_P^*(1)$, the above and Fubini's theorem imply that

$$\int (x - L(F))W'(\theta_n(x))L_n(x)\hat{F}_n(\mathrm{d}x)$$

$$= \int (x - L(F))W'(D(x))L_n(x)F(\mathrm{d}x) + o_P^*(1)$$

$$= \int (x - L(F))W'(D(x))\nu_n H[x]F(\mathrm{d}x) + o_P^*(1)$$

$$= \int B(y)\nu_n(dy) + o_P^*(1).$$

Combining the foregoing results, we obtain

$$n^{1/2}\int (x - L(F))W(D_n(x))\hat{F}_n(\mathrm{d}x)$$

$$= \int [B(x) + (x - L(F))W(D(x))]\nu_n(\mathrm{d}x) + o_P^*(1).$$

Finally, it can be shown in the same way that

$$\int W(D_n(x))\hat{F}_n(\mathrm{d}x) \to \int W(D(x))F(\mathrm{d}x)$$

almost surely, hence the assertion follows by the multivariate central limit theorem. □

Acknowledgements

The author greatly appreciates the careful reading, constructive remarks and suggestions made by an Associate Editor and two referees, which led to improvements to the paper.

This research was supported by grants from the National Sciences and Engineering Research Council of Canada and the Fonds Québécois de la Recherche sur la nature et les Technologies.

References

Arcones, M.A. (1995) Asymptotic normality of multivariate trimmed means. *Statist. Probab. Lett.*, **25**, 43–53.

Arcones, M.A. and Giné, E. (1993) Limit theorems for U-processes. *Ann. Probab.*, **21**, 1494–1542.

Arcones, M.A., Chen, Z. and Giné, E. (1994) Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.*, **22**, 1460–1477.

Caplin, A. and Nalebuff, B. (1988) On 64%-majority rule. *Econometrica*, **56**, 786–814.

Chen, Z. (1995) Robustness of the half-space median. *J. Statist. Plann. Inference*, **46**, 175–181.

Dharmadhikari, S. and Joag-Dev K. (1988) *Unimodality, Convexity, and Applications*. New York: Academic Press.

Donoho, D. and Gasko, M. (1992) Breakdown properties of location estimates based on half-space depth and projected outlyingness. *Ann. Statist.*, **20**, 1803–1827.

Dudley, R.M. (1989) *Real Analysis and Probability.* Pacific Grove, CA: Wadsworth & Brooks/Cole.

Dümbgen, L. (1992) Limit theorems for the simplicial depth. *Statist. Probab. Lett.*, **14**, 119–128.

Eggleston, H.G. (1969) *Convexity.* Cambridge: Cambridge University Press.

Kim, J. (1995) An asymptotic theory of the multivariate metrically trimmed mean. Technical Report, Dept of Statistics, Inha University, Inchon, Korea.

Liu, R.Y. (1990) On a notion of data depth based on random simplices. *Ann. Statist.*, **18**, 405–414.

Liu, R.Y. and Singh, K. (1993) A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, **88**, 252–260.

Liu, R.Y., Parelius, J.M. and Singh, K. (1999) Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.*, **27**, 783–840.

Massé, J.-C. and Plante, J.-F. (2003) A Monte Carlo study of the accuracy and robustness of ten bivariate location estimators. *Comput. Statist. Data Anal.*, **42**, 1–26.

Pollard, D. (1984) *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Rousseeuw, P.J. and Ruts, I. (1999) The depth function of a population distribution. *Metrika*, **49**, 213–244.

Ruts, I. and Rousseeuw, P.J. (1996) Computing depth contours of bivariate point clouds. *Comput. Statist. Data Anal.*, **23**, 153–168.

Serfling, R. (2000) Quantile functions for multivariate analysis: approaches and applications. Preprint.

Tukey, J.W. (1975) Mathematics and picturing data. In R.D. James (ed.), *Proceedings of the 1974 International Congress of Mathematicians*, Vol. 2, pp. 523–531. Vancouver: Canadian Mathematical Congress.

Valentine, F.A. (1964) *Convex Sets*. New York: McGraw-Hill.

van der Vaart A.W. and Wellner J.A. (1996) *Weak Convergence and Empirical Processes.* New York: Springer-Verlag.

Yeh, A.B. and Singh, K. (1997) Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B*, **59**, 639–652.

Zuo, Y. and Serfling, R. (2000) General notions of statistical depth functions. *Ann. Statist.*, **28**, 461–482.

Received January 2000 and revised December 2003