The Annals of Statistics
1994, Vol. 22, No. 3, 1460–1477

ESTIMATORS RELATED TO U-PROCESSES WITH APPLICATIONS TO MULTIVARIATE MEDIANS: ASYMPTOTIC NORMALITY

BY MIGUEL A. ARCONES,[1] ZHIQIANG CHEN[2] AND EVARIST GINÉ[3]

University of Utah, William Paterson College and University of Connecticut

If a criterion function $g(x_1, \dots, x_m; \theta)$ depends on $m > 1$ samples, then a natural estimator of $\arg\max P^m g(x_1, \dots, x_m; \theta)$ is the arg max of a U-process. It is observed that, under suitable conditions, these estimators are asymptotically normal. This is then applied to prove asymptotic normality of Liu's simplicial median and of Oja's medians in $\mathbb{R}^d$.

1. Introduction. If a criterion function $g(x_1, \dots, x_m; \theta)$ depends on $m > 1$ samples, then a natural estimator of $\theta_0 = \theta(P) = \arg\max P^m g(x_1, \dots, x_m; \theta)$ is an arg max of the U-process

$$U_n(\theta) = \frac{(n-m)!}{n!} \sum_{\substack{i_1, \dots, i_m \le n \\ \text{and distinct}}} g\left(X_{i_1}, \dots, X_{i_m}; \theta\right),$$

say, θ_n [here X_i are i.i.d. (P)]; θ_n is a generalization of an M-estimator in the sense of Huber (1967). Relatively recent techniques for the study of rates of convergence of M-estimators [Huber (1967), Pollard (1985), Kim and Pollard (1990), Kolčinskii (1992)], combined with U-process theory [Arcones and Giné (1993)] give a general result on asymptotic normality of $n^{1/2}(\theta_n - \theta_0)$. We were originally interested in such a result in order to prove that Liu's empirical simplicial median [Liu (1990)] is asymptotically normal, at least under regularity and symmetry conditions. Oja's spatial medians [Oja (1983)] fall as well into the same pattern and therefore their asymptotic normality is also a consequence of the general result on M-estimators based on U-processes.

In Section 2 we adapt arguments from Pollard (1985) to obtain asymptotic normality of the arg max of a U-process over a class of functions $\{f_\theta : \theta \in \Theta\}$. See Kolčinskii (1992) for another approach to M-estimators and generalizations via convergence of processes. See also Oja (1984) on asymptotic normality of estimators which are (approximate) zeros of systems of U-statistics, under conditions similar to those of Huber (1967).

We follow Liu (1990) for the following definitions, with a slight variation. Given a probability measure P on $\mathbb{R}^d$, the *depth* $D(\theta)$ of a point $\theta \in \mathbb{R}^d$ is defined as the probability that the simplex whose vertices are $d+1$ independent

Received August 1992; revised August 1993.

[1]Research supported by NSF Grant DMS-85-05550 at M.S.R.I.

[2]Research supported in part by a University of Connecticut Research Foundation Grant.

[3]Research supported in part by NSF Grants DMS-90-00132 and DMS-93-00725 and by the Center de Recerca Matemàtica of the Institut d'Estudis Catalans, Barcelona.

AMS 1991 *subject classifications.* Primary 62F12; secondary 62E20.

Key words and phrases. M-estimators, U-processes, Liu's simplicial median, Oja's medians.

observations from P contains θ, that is,

$$D(\theta)\big(=D(P;\theta)\big) = P^{d+1}I\big(\theta \in S(x_1,\dots,x_{d+1})\big), \tag{1.1}$$

where $S(x_1,\dots,x_{d+1})$ is the *open* simplex of $\mathbb{R}^d$ with vertices $x_1,\dots,x_{d+1}$. If $\{X_i\}_{i=1}^{\infty}$ are independent identically distributed random vectors in $\mathbb{R}^d$ with common law P, then, for each $n \in \mathbb{N}$, the (nth) *empirical depth process* is

$$D_n(\theta) = \frac{1}{\binom{n}{d+1}} \sum_{i_1<\cdots<i_{d+1}\le n} I\Big(\theta \in S\big(X_{i_1},\dots,X_{i_{d+1}}\big)\Big), \qquad \theta \in \mathbb{R}^d. \tag{1.2}$$

The *simplicial median* θ_0 of P is defined as the argument of the supremum of $D(\theta)$ if it exists and is unique, or as any point in $\{\arg\max D(\theta)\}$ if this set consists of more than one point. The (nth) *empirical simplicial median* is any random variable θ_n almost surely in the set $\{\arg\max D_n(\theta)\}$. In fact, Liu's definitions of the depth function and the simplicial median of P, as well as of their empirical counterparts, are in terms of *closed* simplices $\overline{S}(x_1,\dots,x_{d+1})$. We will denote by $\overline{D}(\theta)$, $\overline{\theta}_0$, $\overline{D}_n(\theta)$ and $\overline{\theta}_n$ the analogues of the above variables defined via *closed* simplices. If P assigns mass zero to hyperplanes, then $D(\theta) \equiv \overline{D}(\theta)$ and the two definitions of simplicial median coincide; however, their empirical counterparts are not the same although θ_n and $\overline{\theta}_n$ are both consistent estimators of θ_0 [Arcones and Giné (1993)] and both have an appealing geometric meaning.

The empirical simplicial median is very similar to an M-estimator: the only difference is that, since the criterion function is defined on the product of the parameter space and $d+1$ copies of the space where P lies ($\mathbb{R}^d$), the estimator is naturally the maximizer of a U-process (as opposed to the usual sum of independent random processes—an empirical process). Empirical process theory, or at least techniques, have proved very useful for the asymptotic theory of M-estimators [Huber (1967), Pollard (1985)] and therefore it is only natural that U-processes should be valuable tools for proving asymptotic normality of the empirical simplicial median. Arcones and Giné (1993) proved

$$\Big\{n^{1/2}\big(\overline{D}_n(\theta) - \overline{D}(\theta)\big) \colon \theta \in \mathbb{R}^d\Big\} \to_{\mathcal{L}} \big\{G_P(\theta) \colon \theta \in \mathbb{R}^d\big\} \tag{1.3}$$

in the space $\ell^{\infty}(\mathbb{R}^d)$, where G_P is a sample continuous Gaussian process, and their proof also gives the analogous limit theorem for $n^{1/2}(D_n - D)$. We will use this fact as well as weak convergence of the normalized degenerate U-processes in the Hoeffding decomposition of D_n together with the result from Section 2 to prove that, under certain (perhaps too restrictive) conditions, namely, smoothness and angular symmetry of P,

$$n^{1/2}(\theta_n - \theta_0) \to_d N(0, \Gamma), \tag{1.4}$$

for some covariance Γ to be specified later. The same result is also proved for $\overline{\theta}_n$ when $d \ge 3$, but we only obtain that $n^{1/2}\overline{\theta}_n = O_P(1)$ for $d = 2$. This is done in Section 3.

Oja (1983) proposes the following family of multidimensional location parameters. Letting $V(x_1, \dots, x_d, \theta)$ denote the volume of the simplex $S(x_1, \dots, x_d, \theta)$ (where, as above, x_i and θ are points in $\mathbb{R}^d$) Oja's *α-median*, $1 \leq \alpha < \infty$, is $\theta_0 = \arg\min P^d V^\alpha(x_1, \dots, x_d, \theta)$ if it exists and is unique, and the empirical Oja's α-median is any random variable θ_n which minimizes the process

$$U_n(\theta) = \frac{1}{\binom{n}{d}} \sum_{i_1 < \cdots < i_d \leq n} V^\alpha\big(X_{i_1}, \dots, X_{i_d}, \theta\big), \qquad \theta \in \mathbb{R}^d, \tag{1.5}$$

almost surely. In Section 4 we obtain the limit result (1.4) for Oja's median in some generality. Oja (1984) obtains the same result under less explicit conditions, and Oja and Niinimaa (1985) explicitly compute the covariance of the limit for full multivariate normals P.

We refer, for example, to Giné and Zinn (1986) or to Dudley (1984) for the definitions of convergence in law in $l^\infty(\mathcal{F})$ of processes indexed by classes $\mathcal{F}$ of measurable functions, as well as for the associated asymptotic equicontinuity condition; of Vapnik–Červonenkis subgraph classes of functions (previously called VC-graph classes); of P-Donsker classes $\mathcal{F}$ [also denoted by $\mathcal{F} \in CLT(P)$]; and of $L_2(P)$-entropy and $L_2(P)$-entropy with bracketing, denoted, respectively, by $\log N_2(\varepsilon, \mathcal{F}, P)$ and $\log N_2^{[\,]}(\varepsilon, \mathcal{F}, P)$. See, for example, Arcones and Giné (1993) for notation on U-statistics but, for the reader's convenience, we recall the definition of the projection operators π_k and Hoeffding's decomposition. Let $(S, \mathcal{S}, P)$ be a probability space, and let X_i: $S^{\mathbb{N}} \to S$ be the coordinate functions [which, of course, are i.i.d. (P)]. If f: $S^m \to S$ is a P^m-integrable function, then the U-statistic based on f and the sample $\{X_i\}$ is

$$U_n(f)\big(=U_n^m(f)\big) = \frac{(n-m)!}{n!} \sum_{\substack{i_1, \dots, i_m \leq n \\ \text{and distinct}}} f\big(X_{i_1}, \dots, X_{i_m}\big), \qquad n \in \mathbb{N}.$$

There is no loss of generality in assuming f symmetric in its arguments, and we will assume this in all that follows. The Hoeffding projections π_k of f, $k \leq m$, are defined as

$$\pi_k f(x_1, \dots, x_k)\big(=\pi_k^P f(x_1, \dots, x_k)\big) = \big(\delta_{x_1} - P\big) \cdots \big(\delta_{x_k} - P\big) P^{m-k} f,$$

where, as usual $Qf := \int f\,dQ$ and δ_x is point mass at $x \in S$. Then Hoeffding's decomposition is just

$$U_n(f)\big(= U_n^{(m)}(f)\big) = P^m f + m(P_n - P)\big(P^{m-1} f\big) + \sum_{k=2}^{m} \binom{m}{k} U_n^{(k)}(\pi_k f),$$

where P_n is the empirical measure, that is, $P_n = (1/n)\sum_{i=1}^n \delta_{X_i}$.

2. The argmax of a *U*-process. In this section we use the framework set up in the last few lines. Let $\mathcal{F} = \{f_\theta\colon \theta \in \Theta\}$ be a measurable class of functions on S^m indexed by a set Θ of $\mathbb{R}^d$ containing a neighborhood of the origin, and assume each f_θ is P^m-integrable and symmetric in its arguments. Here is the extension of Pollard's theorem to M-estimators based on criterion functions of several variables; we sketch its proof, which completely follows the proof of Theorem 8 in Pollard (1985), in order to clarify the crucial Remark 2.2. See Pollard (1985) for the details.

THEOREM 2.1. *Assume that the following hold*:

(i) $U(\theta) := P^m f_\theta, \theta \in \Theta$, *has a unique maximum, attained at* $\theta = 0$, *and admits the development*

$$U(\theta) = U(0) - \tfrac{1}{2}\theta A \theta^t + o(|\theta|^2) \tag{2.1}$$

near zero, where A is a symmetric, (strictly) positive definite $d \times d$ *matrix.*

(ii) *For* $2 \le k \le m$ *and for all* $\varepsilon > 0$,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr\left\{ \sup_{|\theta| \le \delta} \left| nU_n\big(\pi_k(f_\theta - f_0)\big)\right| > \varepsilon \right\} = 0. \tag{2.2}$$

(iii) *There exists a measurable function* $\Delta: S \to \mathbb{R}^d$ *satisfying* $E\Delta(X) = 0$ *and* $E|\Delta|^2 < \infty$, *and such that the functions defined by* $r(x, 0) = 0$ *for* $x \in \mathbb{R}^d$ *and*

$$r(x, \theta) = \frac{\pi_1 f_\theta(x) - \pi_1 f_0(x) - \theta \cdot \Delta(x)}{|\theta|}, \qquad \theta \in \Theta \backslash \{0\},\ x \in \mathbb{R}^d,$$

satisfy

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr\left\{ \sup_{|\theta| \le \delta} |n^{1/2} P_n r(\cdot, \theta)| > \varepsilon \right\} = 0, \tag{2.3}$$

for all $\varepsilon > 0$.

(iv) *Let* $\{\theta_n\}$ *be a sequence of random variables such that*

$$\theta_n \to 0 \quad \textit{in probability} \tag{2.4}$$

and

$$n\Big(\sup_{\theta \in \Theta} U_n(\theta) - U_n(\theta_n)\Big) \to 0 \quad \textit{in probability.} \tag{2.5}$$

Then

$$n^{1/2}\theta_n \to_d Z,$$

where Z is $N(0, \Gamma)$ *and* $\Gamma = m^2 A^{-1}(\operatorname{cov} \Delta(X)) A^{-1}$. [*In* (iii), r *can be replaced by* r_n *obtained from r by setting the denominator equal to* $|\theta| \vee n^{-1/2}$.]

PROOF (Sketch). First we indicate how to show that the sequence $\{n^{1/2}\theta_n\}$ is stochastically bounded. By (2.1), there exist $\varepsilon > 0$ and $c > 0$ such that $U(0) - U(\theta) \ge c|\theta|^2$, for all $|\theta| < \varepsilon$. So, since $\theta_n \to 0$ in probability, (2.1) and (2.5) give, using Hoeffding's expansion,

$$
\begin{aligned}
cn|\theta_n|^2 &\le n\big(U(0) - U(\theta_n)\big) + o_P(1)\\
&\le n(P^m - U_n + U_n)\big(f_0 - f_{\theta_n}\big) + o_P(1)\\
&= mn^{1/2}\theta_n \cdot \frac{\sum_{i=1}^n \Delta(X_i)}{n^{1/2}} + mn^{1/2}|\theta_n|\frac{\sum_{i=1}^n r(X_i,\theta_n)}{n^{1/2}}\\
&\quad + n\sum_{k=2}^m \binom{m}{k} U_n\Big(\pi_k\big(f_{\theta_n} - f_0\big)\Big) + o_P(1).
\end{aligned}
$$

So, by (2.2) and (2.3), $n|\theta_n|^2 \le \xi_n n^{1/2}|\theta_n| + \zeta_n$, where $\zeta_n \to 0$ in probability and the sequence $\{\xi_n\}$ is stochastically bounded. Hence, $\{n^{1/2}\theta_n\}$ is stochastically bounded. [Note that the same conclusion obtains if r is replaced by r_n in condition (iii).]

Now let $\Delta_n = mP_n\Delta$ and let B be a nondegenerate matrix such that $BAB^t = I$. Since $n^{1/2}|\Delta_n| = O_P(1)$ by the central limit theorem, using the hypotheses (i)–(iii) on $U_n = U_n - U + U$ and developing as above, we obtain

$$n\Big(U_n\big(\Delta_n A^{-1}\big) - U_n(0)\Big) = \tfrac{1}{2} n\Delta_n A^{-1}\Delta_n^t + o_P(1).$$

Similarly, since $n^{1/2}\theta_n = O_P(1)$ by the above computation, we also have

$$n\big(U_n(\theta_n) - U_n(0)\big) = -\tfrac{1}{2} n\theta_n A\theta_n^t + n\theta_n\Delta_n^t + o_P(1).$$

Since, by (2.5), $nU_n(\Delta_n A^{-1}) \le nU_n(\theta_n) + o_P(1)$, the last two developments give $n|\Delta_n B^t - \theta_n B^{-1}|^2 \to 0$ in probability, and the result follows. □

REMARK 2.2. A close look at the previous proof shows that the full strength of condition (2.3) is not needed in Theorem 2.1 but rather only

$$n^{-1/2}\sum_{i=1}^n r(X_i,\theta_n) \to 0 \quad \text{in probability} \tag{2.3$'$}$$

and

$$n^{-1/2}\sum_{i=1}^n r(X_i,V_n) \to 0 \quad \text{in probability,} \tag{2.3$''$}$$

where $V_n = \Delta_n A^{-1} + \delta_n$ for some sequence of random variables δ_n such that $n^{1/2}\delta_n \to 0$ in probability. Therefore, condition (iii) in Theorem 2.1 can be replaced by:

(iii$'$) There is a class of functions $\tilde{r}(x,\theta), x \in \mathbb{R}^d, \theta \in \Theta$, satisfying condition (2.3) and such that

$$
\begin{aligned}
&n^{1/2}P_n\big(r(\cdot,\theta_n) - \tilde{r}(\cdot,\theta_n)\big) \to 0 \quad \text{in probability,}\\
&n^{1/2}P_n\big(r(\cdot,V_n) - \tilde{r}(\cdot,V_n)\big) \to 0 \quad \text{in probability,}
\end{aligned}
\tag{2.6}
$$

for $i \le n, n \in \mathbb{N}$.

Note also that if the quantities in (2.6) are $O_P(1)$ instead of $o_P(1)$, then we can still conclude $n^{1/2}\theta_n = O_P(1)$ by the first part of the proof of Theorem 2.1. It is conceivable that the class of functions r does not satisfy (2.3) but that there exists a class $\tilde{r}$ satisfying (2.3) and (2.6). This remark is used in our treatment of Liu's median.

REMARK 2.3. The usual way to verify condition (2.2) is to show that the processes $\{n^{k/2}U_n(\pi_k f_\theta): |\theta| < \delta\}$ converge in law to Radon measures in $l^\infty(\{\theta: |\theta| < \delta\})$ for some $\delta > 0$ and that $P^m(f_\theta - f_0)^2 \to 0$ as $\theta \to 0$ since these two conditions imply an asymptotic equicontinuity property even stronger than (2.2) [Arcones (1994); see also (1.7) and its paragraph in Arcones and Giné (1993)]. Likewise, to verify (2.3) for $\tilde{r}$ (or for r) it suffices to check that the set of functions $\{\tilde{r}(\cdot, \theta): |\theta| < \delta\}$ (or $\{r(\cdot, \theta): |\theta| < \delta\}$) is P-Donsker for some $\delta > 0$ and that $P\tilde{r}^2(\cdot, \theta) \to 0 [Pr^2(\cdot, \theta) \to 0]$ as $\theta \to 0$ [e.g., Giné and Zinn (1986), Theorem 1.1.3].

REMARK 2.4. In fact the above theorem follows as a corollary to the following statement, which has a similar proof. Let $U(\theta) = U(f_\theta)$ be a linear function of f_θ (not necessarily an integral) and let, for each $n \in \mathbb{N}$, $U_n(\theta) = U_n(f_\theta)$ be a statistic linear in f_θ, satisfying the following:

(i) $U(0) \geq U(\theta)$ for all θ;
(ii) $U(\theta) = U(0) - \frac{1}{2}\theta A\theta^t + o(|\theta|^2)$ near 0, with A as in Theorem 2.1;
(iii) $$\lim_{\delta \to 0} \limsup_{n \to \infty} \Pr\left\{ \sup_{|\theta| \leq \delta} \frac{|n^{1/2}(U_n - U)(f_\theta - f_0 - \theta \cdot \Delta)|}{|\theta| + n^{-1/2}} > \varepsilon \right\} = 0$$

for some function $\Delta: S^m \to \mathbb{R}^d$ such that $P^m|\Delta|^2 < \infty$, and for all $\varepsilon > 0$.

Suppose $\{\theta_n\}$ is a sequence of random variables such that the following hold:

(iv) $\theta_n \to 0$;
(v) $n(\sup_{\theta \in \Theta} U_n(\theta) - U_n(\theta_n)) \to 0$ in probability.

Then

$$n^{1/2}\left(\theta_n - \Delta_n A^{-1}\right) \to 0 \quad \text{in probability},$$

where $\Delta_n = (U_n - U)(\Delta)$.

3. The empirical simplicial median. We begin with some notation. Let P be a Borel probability measure on $\mathbb{R}^d$ with a smooth density f (smoothness conditions to be specified later). Let f_i and f_{ij} denote the first and second partial derivatives of f. For $\theta \in \Theta, \mathbf{x} = (x_1, \ldots, x_k) \in (\mathbb{R}^d)^k, k \in \mathbb{N}$ (only $k = d$ and $k = d+1$ are used in what follows), we define the joint density

$$\phi_k(\mathbf{x}, \theta) = f(x_1 + \theta) \cdots f(x_k + \theta)$$

and let $\phi_k'(\mathbf{x}, 0)$ and $\phi_k''(\mathbf{x}, 0)$ denote, respectively, the first and second derivatives of ϕ_k with respect to θ at $\theta = 0$, that is,

$$\phi'_k(\mathbf{x}, 0) = \left(\frac{\partial \phi_k}{\partial \theta_j}(\mathbf{x}, 0) : j = 1, \dots, d\right), \qquad \phi''_k(\mathbf{x}, 0) = \left(\frac{\partial^2 \phi_k}{\partial \theta_i \, \partial \theta_j}(\mathbf{x}, 0)\right)_{i, j = 1, \dots, d}$$

Note that these partial derivatives are sums of products of $f(z_r), f_i(z_s)$ and $f_{ij}(z_t)$. The following vector and matrix will appear in the statement of the next theorem:

$$\Delta(x)\big(= \Delta_f(x)\big) := \int_{(\mathbb{R}^d)^d} I\big(0 \in S(\mathbf{x}, x)\big) \phi'_d(\mathbf{x}, 0) \, d\mathbf{x} \tag{3.1}$$

and

$$A(= A_f) := -\int_{(\mathbb{R}^d)^{(d+1)}} I\big(0 \in S(\mathbf{x})\big) \phi''_{d+1}(\mathbf{x}, 0) \, d\mathbf{x}. \tag{3.2}$$

From this point on, whenever the domain of integration is the whole space, it will be omitted. If $x \neq 0$, we let $\gamma(x) = x/|x|$ and $r(x) = |x|$ be its polar coordinates. The *angular density* a_θ of P about $\theta \in \mathbb{R}^d$ is defined as

$$a_\theta(\gamma) = \int_0^\infty f_\theta(r\gamma) r^{d-1} \, dr, \qquad \gamma \in S^{d-1},$$

where $f_\theta(x) = f(x + \theta)$; P is defined to be *angularly symmetric* about a point $\theta \in \mathbb{R}^d$ if the random vector $(X - \theta)/|X - \theta|$ is symmetric, where $\mathcal{L}(X) = P$. If P is angularly symmetric about θ, then $a_\theta(\gamma) = a_\theta(-\gamma), \gamma \in S^{d-1}$.

We will make the following assumptions on P:

(P1) P is angularly symmetric about zero and has a density f with $f(0) \neq 0$;

(P2) f is twice differentiable with continuous second partial derivatives and there exist $\delta > 0$ and $h \in L_1(\mathbb{R}^d, \lambda)$ such that

$$|f_i(x)| \leq h(x), \qquad \sup_{|\theta| \leq \delta} |f_{ij}(x + \theta)| \leq h(x), \qquad i, j \leq d; \tag{3.3}$$

(P3) P has angular densities $a_\theta(\gamma)$ bounded uniformly in $\gamma \in S^{d-1}$ and $|\theta| \leq \delta$ for some $\delta > 0$;

(P4) $\det A_f \neq 0$.

As mentioned in the Introduction, Liu [(1990), Theorem 4] observes that if P is angularly symmetric about θ_0 and has a positive density at θ_0, then θ_0 is its unique simplicial median. It also follows from the proof of her theorem that for such P's we have

$$\Pr\{\theta_0 \in S(X_1, \dots, X_d, x)\} \equiv 2^{-d}, \tag{3.4}$$

for $x \neq \theta_0$, a fact that we use below. [Liu's results apply because if P is absolutely continuous, then $\Pr\{\theta \in \partial S(X_1, \dots, X_{d+1})\} = 0$ and therefore it is irrelevant whether open or closed simplices are used in the definition of the simplicial depth of P; the difference is in the empirical simplicial depth.] There is no loss of generality in assuming $\theta_0 = 0$, and we do so in all that follows. We are now ready to prove asymptotic normality of the empirical simplicial median.

THEOREM 3.1. *Let P be a Borel probability measure on $\mathbb{R}^d$ with the properties* (P1)–(P4). *Let $\theta_n, n \in \mathbb{N}$, be a sequence of empirical simplicial medians (defined by open simplices) corresponding to a random sample $X_i, i \in \mathbb{N}$, from P. Then*

$$n^{1/2}\theta_n \to_d N\Big(0, (d+1)^2 A^{-1}\big(\mathrm{Cov}_P(\Delta)\big)A^{-1}\Big). \tag{3.5}$$

PROOF. We apply Theorem 2.1 as modified by Remark 2.2, with $U(\theta) = D(\theta)$ and $U_n(\theta) = D_n(\theta)$, and with θ_n as an empirical simplicial median. Condition (2.5) for θ_n holds by definition, and condition (2.3), that is, consistency of θ_n, is proved in Arcones and Giné [(1993), Theorem 6.9 and comments following it]. So, condition (iv) in Theorem 2.1 holds. We now verify the expansion (2.1) of $D(\theta)$ at zero, that is, (i) in Theorem 2.1. We note that condition (P2) implies that the functions f and f_i are continuous and, moreover,

$$\sup_{|\theta|\le\delta} |f_i(x+\theta)| \le h(x), \qquad \sup_{|\theta|\le\delta} |f(x+\theta)| \le h(x), \qquad i \le d, \tag{3.3$'$}$$

for some h integrable. Note also that $D(\theta)$ has gradient zero at zero: this follows from Liu [(1990), Remark C], or by the following computation using (3.4) [(3.3) is also used in order to justify differentiation under the integral sign]:

$$\begin{aligned}\frac{\partial D}{\partial\theta_j}\bigg|_{\theta=0} &= \frac{\partial}{\partial\theta_j}\bigg(\int I\big(0 \in S(\mathbf{x})\big)\phi_{d+1}(\mathbf{x},\theta)\,d\mathbf{x}\bigg)\bigg|_{\theta=0}\\ &= (d+1)\int I\big(0 \in S(\mathbf{x})\big)f(x_1)\cdots f(x_d)f_j(x_{d+1})\,dx_1\cdots dx_{d+1}\\ &= 2^{-d}(d+1)\int f_j(x)\,dx = 0.\end{aligned}$$

Then, (3.3) and (3.3$'$) provide sufficient integrability to give, for θ in a neighborhood of zero,

$$\begin{aligned}D(\theta) - D(0) &= \int I\big(0 \in S(\mathbf{x})\big)\big[\phi_{d+1}(\mathbf{x},\theta) - \phi_{d+1}(\mathbf{x},0)\big]\,d\mathbf{x}\\ &= -\tfrac{1}{2}\theta A\theta^t + \tfrac{1}{2}\theta\bigg[\int I\big(0 \in S(\mathbf{x})\big)\big[\phi''_{d+1}(\mathbf{x},\eta\theta) - \phi''_{d+1}(\mathbf{x},0)\big]d\mathbf{x}\bigg]\theta^t\\ &= -\tfrac{1}{2}\theta A\theta^t + o\big(|\theta|^2\big),\end{aligned}$$

where $0 \le \eta \le 1$ depends on θ and $\mathbf{x}$. Since D has a maximum at zero, the above equation and (P4) show that the matrix A is strictly positive definite and therefore, (2.1) holds for D.

Next we check condition (ii) of Theorem 2.1. Arcones and Giné [(1993), Corollary 6.7] show that the collection of sets $\{\{\mathbf{x} \in (\mathbb{R}^d)^{d+1}: \theta \in S(\mathbf{x})\}: \theta \in \mathbb{R}^d\}$ is an image admissible Suslin Vapnik–Červonenkis class and therefore, by Arcones and Giné [(1993), Corollary 5.7], the sequences of processes $\{n^{k/2}U_n(\pi_k f_\theta): \theta \in \mathbb{R}^d\}, k = 1, \dots, d+1$, converge in law in $l^\infty(\mathbb{R}^d)$. So, according to Theorem 2.1 and Remark 2.3, in order to prove condition (2.2) it suffices to check the continuity of $\theta \to f_\theta$ in L_2 at $\theta = 0$. For $\mathbf{x} = (x_1, \dots, x_d)$ fixed, different and different from θ, the set $C_\theta(\mathbf{x}) = \{x: \theta \in S(\mathbf{x}, x)\}$ is a pyramidal convex cone with vertex θ and edges $\theta - \lambda(x_i - \theta)$, $\lambda > 0$; hence its boundary is contained in the union of at most d hyperplanes. So, $P\,\partial C_0(\mathbf{x}) = 0$ for almost every $\mathbf{x}$. Since $I_{C_\theta(\mathbf{x})}(x) \to I_{C_0(\mathbf{x})}(x)$ as $\theta \to 0$ except for $x \in \partial C_0(\mathbf{x})$, we obtain that, by bounded convergence,

$$\int \Big[I\big(\theta \in S(\mathbf{x}, x_{d+1})\big) - I\big(0 \in S(\mathbf{x}, x_{d+1})\big)\Big]^2 f(x_{d+1})\,dx_{d+1} \to 0 \quad \text{a.s.},$$

as $\theta \to 0$. Applying bounded convergence once more, we get

$$P^{d+1}\Big[I\big(\theta \in S(x_1, \dots, x_{d+1})\big) - I\big(0 \in S(x_1, \dots, x_{d+1})\big)\Big]^2 \to 0$$

as $\theta \to 0$.

It remains only to prove condition (iii′) of Remark 2.2. Let $\Delta(x)$ be as given in (3.1). Then the same computation showing that $\mathrm{Grad}|D(\theta)|_{\theta=0} = 0$ gives $P\Delta = 0$. Also, $P|\Delta|^2 < \infty$ since Δ is bounded. Let

$$g(x, \theta) := P^d I\big(\theta \in S(x_1, \dots, x_d, x)\big), \qquad x \in \mathbb{R}^d,\ \theta \in \Theta. \tag{3.6}$$

Then,

$$r(x, \theta) = \frac{g(x, \theta) - g(x, 0) - \theta \cdot \Delta(x)}{|\theta|} - \frac{D(\theta) - D(0)}{|\theta|} \quad \text{for } \theta \neq 0$$

(and 0 for $\theta = 0$). It can be checked that r does not satisfy (2.3), but we will show that, letting

$$\widehat{r}(x, \theta) = \frac{g(x, \theta) - g(x - \theta, 0) - \theta \cdot \Delta(x)}{|\theta|}, \qquad \theta \neq 0,$$

and $\widehat{r}(x, 0) = 0$, the class of functions

$$\big\{\widetilde{r}(x, \theta) = \widehat{r}(x, \theta) - P\widehat{r}(x, \theta): \theta \in \Theta\big\} \tag{3.7}$$

satisfies (iii′) in Remark 2.2. These functions satisfy (2.6) because the identities $g(X_i - \theta_n, 0) = g(X_i, 0)$ and $g(X_i - V_n, 0) = g(X_i, 0)$ hold a.s. To see that these identities hold a.s. (for a permissible definition of V_n), we note that $X_i \neq 0$ a.s. and that the random set $\{\arg\max D_n\}$ is open and does not contain any $X_i, i \leq n$, as is easy to see; hence also $X_i - \theta_n \neq 0$ a.s., and the first identity follows from (3.4); for the second, take δ_n in the definition of $V_n = \Delta_n A^{-1} + \delta_n$ as $\delta_n = 0$ if $\Delta_n A^{-1} \neq X_i$ for all $i \leq n$, and $\delta_n = \min\{n^{-1} \wedge |X_i - X_j|: 1 \leq i \neq j \leq n\}$ otherwise. These observations are the reason why we work first with open simplices and why we take P to be angularly symmetric. So, it suffices to prove that the functions in (3.7) satisfy condition (2.3). For this we will use empirical process theory, as indicated in Remark 2.3.

We have

$$\begin{aligned}\widehat{r}(x,\theta) &= \frac{g(x,\theta)-g(x-\theta,0)-\theta\cdot\Delta(x-\theta)}{|\theta|} + \frac{\theta}{|\theta|}\cdot\big(\Delta(x-\theta)-\Delta(x)\big)\\ &:= \widehat{r}_1(x,\theta)+\widehat{r}_2(x,\theta),\end{aligned} \tag{3.8}$$

and will prove that (2.3) holds for $\widehat{r}_i - P\widehat{r}_i, i = 1,2$. Note that

$$\widehat{r}_1(x,\theta) = \int I\big(0\in S(\mathbf{x},x-\theta)\big)\left[\frac{\phi_d(\mathbf{x},\theta)-\phi_d(\mathbf{x},0)-\theta\cdot\phi_d'(\mathbf{x},0)}{\theta}\right]d\mathbf{x}.$$

By (P2) and Taylor's formula in $\mathbb{R}^d$, there is a $c < \infty$ such that, if g denotes f, f_i or $f_{ij}, i,j \le d$, then $\sup_{|\theta|\le\delta}|g(x+\theta)| \le ch(x)$. Therefore, also by Taylor's formula,

$$\int \sup_{|\theta|\le\delta}\left|\frac{\phi_d(\mathbf{x},\theta)-\phi_d(\mathbf{x},0)-\theta\cdot\phi_d'(\mathbf{x},0)}{\theta}\right| d\mathbf{x} \le \left(c\int h(x)\,dx\right)^d \delta := K\delta.$$

So, letting $C_0^\theta(\mathbf{x})$ denote the translation by θ of $C_0(\mathbf{x})$ that is, the set of points x in $\mathbb{R}^d$ such that $0\in S(\mathbf{x},x-\theta)$, we obtain

$$\sup_{|\theta|\le\delta}\left|n^{1/2}(P_n-P)\widehat{r}_1(\,\cdot\,,\theta)\right| \le K\delta\sup_{\mathbf{x},\theta}\left|n^{1/2}(P_n-P)\big(C_0^\theta(\mathbf{x})\big)\right|. \tag{3.9}$$

The family of sets $\{C_0^\theta(\mathbf{x})\colon \mathbf{x}\in(\mathbb{R}^d)^d, \theta\in\mathbb{R}^d\}$ is a Vapnik–Červonenkis image admissible Suslin class and therefore is P-Donsker for all P in $\mathbb{R}^d$. [It is VC because each of these sets is a convex cone with d edges, hence the intersection of d half-spaces—see 9.2.1 and 9.2.3 Dudley (1984)—and it is image admissible Suslin because the evaluation map $(\mathbf{x},\theta,x)\to I_{C_0^\theta(\mathbf{x})}(x)$ is jointly measurable—Dudley (1984). Section 10.3; and therefore it is Donsker by, e.g., Dudley (1984), Section 11]. In particular,

$$\lim_{M\to\infty}\sup_n \Pr\left\{\sup_{\mathbf{x},\theta}\left|n^{1/2}(P_n-P)\big(C_0^\theta(x)\big)\right| > M\right\} = 0,$$

which, combined with (3.9), gives

$$\lim_{\delta\to 0}\limsup_{n\to\infty}\Pr\left\{\sup_{|\theta|\le\delta}\left|n^{1/2}(P_n-P)\widehat{r}_1(\,\cdot\,,\theta)\right| > \varepsilon\right\} = 0, \tag{3.10}$$

for all $\varepsilon > 0$, that is, (2.3) for $\widehat{r}_1$.

We do not know how to apply Vapnik–Červonenkis theory for $\widehat{r}_2$ unless f and f_j are all Riemann integrable. However, we can use Ossiander's (1987) central limit theorem for empirical processes satisfying a bracketing entropy condition, with essentially the same computations needed to prove that

$$E\widehat{r}_2^2(X,\theta)^2 \to 0 \quad \text{as } \theta\to 0, \tag{3.11}$$

a condition that is also required. In (3.11) and in all that follows, X will denote a random vector with law P. Let δ be as in hypotheses (P2) and (P3). We want to prove that the class of functions $\mathcal{R}_\delta := \{\widehat{r}_2(\cdot,\theta)\colon 0 \le |\theta| \le \delta\}$ is P-Donsker and that (3.11) holds. Then $\widehat{r}_2 - P\widehat{r}_2$ will satisfy (2.3) by Remark 2.3, concluding the proof of the theorem by Theorem 2.1 and Remark 2.2. According to Ossiander's theorem, $\mathcal{R}_\delta$ is P-Donsker if

$$\int_0^\infty \left[\log N_2^{[\,]}(\varepsilon,\mathcal{R}_\delta,P)\right]^{1/2} d\varepsilon < \infty. \tag{3.12}$$

Inequality (3.11) will become proved on the way to proving (3.12). In order to prove (3.12) it suffices to find, for each $0 < \varepsilon \le \delta$, a partition $A_1(\varepsilon),\dots,A_{n(\varepsilon)}(\varepsilon)$ of $\{\theta\colon 0 \le |\theta| \le \delta\}$ and P-square integrable functions $h_1,\dots,h_{n(\varepsilon)}$ with

$$n(\varepsilon) \le \frac{c_1}{\varepsilon^{v_1}} \tag{3.13}$$

and such that

$$E \sup_{\theta \in A_j(\varepsilon)} \big(\widehat{r}_2(X,\theta) - h_j(X)\big)^2 \le c_2 \varepsilon^{v_2}, \tag{3.14}$$

for some $c_i, v_i \in (0,\infty)$. If this is the case, $N_2^{[\,]}(\varepsilon,\mathcal{R}_\delta,P)$ is bounded by a power of $1/\varepsilon$ and therefore the integral in (3.12) is finite. For the rest of this proof, c will denote a finite positive constant independent of ε that may vary from line to line. We begin by showing that $A_1(\varepsilon)$ can be taken to be a neighborhood of zero, and $g_1 \equiv 0$. This automatically gives (3.11). We have, letting, as usual, $\mathbf{x} = (x_1,\dots,x_d)$,

$$\begin{aligned}
E \sup_{|\theta| \le \varepsilon} \widehat{r}_2^{\,2}(X,\theta) &\le E \sup_{|\theta|\le\varepsilon} |\Delta(X-\theta) - \Delta(X)|^2 \\
&\le E \sup_{|\theta|\le\varepsilon} \sum_{i=1}^d d^2 \bigg[\int \big|I\big(0 \in S(\mathbf{x}, X-\theta)\big) - I\big(0 \in S(\mathbf{x}, X)\big)\big| \\
&\qquad\qquad \times \big|f_i(x_1)\big| f(x_2)\cdots f(x_d)\, d\mathbf{x}\bigg]^2 \\
&\le d^2\Big(c\int h(x)\,dx\Big) \sum_{i=1}^d \int \sup_{|\theta|\le\varepsilon} \big|I\big(0 \in S(\mathbf{x}, x-\theta)\big) - I\big(0 \in S(\mathbf{x},x)\big)\big| \\
&\qquad\qquad \times |f_i(x_1)| f(x_2)\cdots f(x_d) f(x)\, d\mathbf{x}\, dx \\
&\le K \sup_{\substack{x_i \ne 0 \text{ and} \\ \text{distinct}}} \int_{|x|>\varepsilon} \sup_{|\theta|\le\varepsilon} \big|I\big(0 \in S(\mathbf{x},x-\theta)\big) - I\big(0\in S(\mathbf{x},x)\big)\big| f(x)\,dx \\
&\qquad\qquad + KP\big\{|x| \le \varepsilon\big\} \\
&\le K \sup_{\substack{x_i \ne 0 \text{ and} \\ \text{distinct}}} P\Big\{\big(\partial C_0(x)\big)_{\varepsilon]}\Big\} + K'\varepsilon_d,
\end{aligned}$$

where $K = d^3(c\int h(x)\,dx)^2$, and $K' = K(\sup_{|x|\le\varepsilon} f(x))\mathrm{Vol}\{|x| \le 1\}$, and where, for any set $F, F_{\varepsilon]} = \{y: \inf_{x\in F}|x-y| \le \varepsilon\}$. Since $\partial C_0(\mathbf{x})$ is contained in the union of at most d hyperplanes and f is continuous (hence bounded on $\{|x| \le 1\}$), we have

$$P\Big\{\big(\partial C_0(x)\big)_{\varepsilon]} \cap \{|x| \le 1\}\Big\} \le c\varepsilon$$

for some $c < \infty$. On the other hand, $(\partial C_0(\mathbf{x}))_{\varepsilon]} \cap \{|x| > 1\}$ is contained in a cone of vertex zero whose intersection with the sphere S^{d-1} has area $O(\varepsilon)$. Since the angular density $a_0(\gamma)$ of P is bounded [(P3)] it follows that

$$P\Big\{\big(\partial C_0(\mathbf{x})\big)_{\varepsilon]} \cap \{|x| > 1\}\Big\} \le c\varepsilon.$$

So, we conclude that

$$E \sup_{|\theta|\le\varepsilon} \widehat{r}_2^{\,2}(X,\theta) \le c\varepsilon. \tag{3.15}$$

To construct the "brackets" away from zero, given $0 < \varepsilon \le \frac{1}{2}\wedge\delta$, we fix $|\theta| > \varepsilon$, let θ' be any point such that $|\theta'| \le \delta$ and $|\theta-\theta'| \le \varepsilon^2$ and decompose $\widehat{r}_2(x,\theta') - \widehat{r}_2(x,\theta)$ as

$$\widehat{r}_2(x,\theta') - \widehat{r}_2(x,\theta) = \left(\frac{\theta'}{|\theta'|} - \frac{\theta}{|\theta|}\right)\big(\Delta(x-\theta') - \Delta(x)\big) + \frac{\theta}{|\theta|}\big(\Delta(x-\theta') - \Delta(x-\theta)\big).$$

Since Δ is bounded, the first term is dominated by a constant times ε. The second term can be analyzed as we just have done for $\theta = 0$ on account of (P3). The conclusion is

$$E \sup_{\theta': |\theta-\theta'|\le\varepsilon^2,\, |\theta'|\le\delta} \big[\widehat{r}_2(X,\theta) - \widehat{r}_2(X,\theta')\big]^2 \le c\varepsilon^2. \tag{3.16}$$

From (3.15) and (3.16) we conclude that we can take $A_1(\varepsilon)$ to be the neighborhood of zero of radius $\varepsilon, h_0 = 0, A_j(\varepsilon), j \ge 2$, to be balls of radius ε^2 around points $\theta_j, \varepsilon < |\theta_j| \le \delta$, and $h_j(x) = \widehat{r}_2(x,\theta_j)$, with $v_2 = 1$ in (3.14) and $v_1 = 2d$ in (3.15) [since the number of A_j needed to cover the region $\{|\theta| \le \delta\}\backslash A_1(\varepsilon)$ is of the order of ε^{-2d}]. So, the integral in (3.12) is finite and therefore, by Ossiander's theorem, $\mathcal{R}_\delta$ is P-Donsker; also, by (3.15), (3.11) holds. Then Remark 2.3 gives that the functions $\widehat{r}_2$ satisfy (2.3). Hence, by (3.8) and (3.10), so do the functions $\widehat{r}$, and the conclusion of the theorem follows from Theorem 2.1 and Remark 2.2. [It is worth mentioning that Remark 2.2 is basic in this proof: as is easily verified, the functions

$$\big(g(x,0) - g(x-\theta,0)\big)/\big(|\theta| \vee n^{-1/2}\big)$$

do not satisfy condition (2.3); hence, by the above, neither do the r_n's.] □

REMARK 3.2 (Relaxing the smoothness conditions in $\mathbb{R}^2$). Although many densities satisfy the hypotheses (P1)–(P4), such as, for example, nondegenerate multivariate normals, they are quite restrictive, particularly condition (P2). For densities in the plane these hypotheses can be somewhat relaxed using Green's theorem (or, in $\mathbb{R}^d$, using the divergence theorem). [See Kim and Pollard (1990), where they treat similar problems.] Here we only state a set of weaker conditions under which asymptotic normality of the empirical simplicial median still holds, but omit the proof (a sketch of which can be obtained from the authors).

(Replacing hypotheses). The hypotheses (P1) and (P4) are kept, and (P2) and (P3) are replaced by the following:

(P2′) (i) f is differentiable and its partial derivatives are continuous and (Lebesgue) integrable; (ii) there is a $\delta > 0$ such that, for $|\theta|, |\eta| \leq \delta/2$, there exists $h(\theta, x)$ such that $\int h(0,x)\,dx < \infty$, $\sup_{|\theta| \leq \delta/2} \int_0^\infty h(\theta, ru) u_2\,dr < \infty$ and $|f(\theta + \eta + ru)|, |f_j(\theta + \eta + ru)| \leq h(\theta, ru)$ for all $u = (u_1, u_2) \in S^1$ and $j = 1, 2$; (iii) $|f(x)|$ and $|f_j(x)|, j = 1, 2$, are $o(|x|^{-1})$ as $|x| \to \infty$;

(P3′) $a_\theta(\xi)$ is continuous in (θ, ξ).

EXAMPLE 3.3. Computing the covariance of the limit in Theorem 3.1 can become quite involved. If P is $N(0, I)$ in $\mathbb{R}^2$, the computations are straightforward, and we have

$$\Delta(x) = \left(\frac{2}{\pi^3}\right)^{1/2} \frac{x}{|x|} \quad \text{for } x \neq 0,$$

and

$$A = \frac{3}{2\pi} I.$$

Then $\operatorname{Cov} \Delta(X) = (1/\pi^3) I$ and it follows that $n^{1/2}\theta_n$ is asymptotically $N(0, (4/\pi)I)$.

REMARK 3.4 (Closed simplices). For closed simplices, we can prove the following (which is not satisfactory for $d = 2$): Under the hypotheses of Theorem 3.1, if $\overline{\theta}_n, n \in \mathbb{N}$, is a sequence of empirical simplicial medians defined via closed simplices, then the limit (3.5) also holds for $\{n^{1/2}\overline{\theta}_n\}$ for $d \geq 3$ and this sequence is stochastically bounded for $d = 2$. The proof only differs from the one for open simplices in the argument allowing the replacement of r by $\tilde{r}$. In this case, the part of condition (2.6) relative to V_n is proved as above, but the part of (2.6) relative to $\overline{\theta}_n$ requires a different argument as follows: with g defined using closed simplices we have, as above, $g(X_i, 0) = g(X_i - \overline{\theta}_n, 0) = 2^{-d}$ if $X_i \neq 0$ (which happens a.s.) and $X_i \neq \overline{\theta}_n$ (which may not happen a.s.), whereas $g(X_i - \overline{\theta}_n, 0) = 1$ if $X_i = \overline{\theta}_n$. Since $g(x - \theta, 0) = g(x, 0)$ a.s., we can ignore centerings in the proof of (2.6) and must therefore show that

$$n^{1/2} \sum_{i=1}^{n} \left(g(X_i, 0) - g(X_i - \overline{\theta}_n, 0)\right) / |\overline{\theta}_n| = o_P(1)$$

if $d \geq 3$ and $O_P(1)$ for $d = 2$. However, this quantity is a.s. 0 if $\overline{\theta}_n \neq X_i, i = 1, \ldots, n$, and its absolute value is a.s. dominated by $(n^{1/2} \min_{i \leq n} |X_i|)^{-1}$ otherwise (the

X_i's being a.s. different, at most one summand is not zero). Now,

$$\Pr\Big\{\big(n^{1/2}\min_{i\le n}|X_i|\big)^{-1}>\varepsilon\Big\}=1-\bigg[1-\Pr\Big(|X|<\frac{1}{\varepsilon n^{1/2}}\Big)\bigg]^n\le 1-\bigg[1-\frac{c(d,f)}{\varepsilon^d n^{d/2}}\bigg]^n,$$

for some constant $c(d,f)$. This tends to zero for all $\varepsilon>0$ as $n\to\infty$ if $d>2$ but not if $d=2$, in which case we still have, however, that the sup over n tends to zero as $\varepsilon\to\infty$. Now, the theorem for closed simplices follows as above. It would be surprising if $n^{1/2}\overline{\theta}_n$ did not converge for $d=2$, but we do not know how to prove this at present.

4. Oja's medians. We have defined $V(\mathbf{x})=V(x_1,\dots,x_d,\theta), x_i,\theta\in\mathbb{R}^d$, in the Introduction. Note that

$$V(\mathbf{x},\theta)=|\det(x_i-\theta)|.$$

[In all that follows, we write $\det(x_i-\theta)$ to mean $\det(x_1-\theta,\dots,x_d-\theta)$ in the canonical basis of $\mathbb{R}^d$.] In this section we let $f_\theta(\mathbf{x})=V^\alpha(\mathbf{x};\theta)$ for some $1\le\alpha<\infty$ fixed, and then

$$U(\theta)=P^d f_\theta(\mathbf{x}),\qquad U_n(\theta)=\frac{1}{\binom{n}{d}}\sum_{i_1<\cdots<i_d\le n} f_\theta\big(X_{i_1},\dots,X_{i_d}\big).$$

In the proof of asymptotic normality of Oja's medians, smoothness of the distribution only plays a role in the limited expansion at θ_0 of $U(\theta)$. We will assume in the next theorem that this expansion holds and then will discuss the smoothness conditions in a proposition below. As usual, there is no loss of generality in assuming Oja's median of P, θ_0, to be 0.

THEOREM 4.1. *Suppose the following hold*:

(i) $E|X|^{2\alpha}<\infty$;
(ii) $U(\theta)$ *has a unique minimum at* $\theta=0$;
(iii) *the distribution P of X is not supported by any subspace of $\mathbb{R}^d$ of dimension* $d-1$;
(iv) $U(\theta)=U(0)+\frac{1}{2}\theta A\theta^t+o(|\theta|^2)$ *near zero, with A positive definite*;
(v) $\Pr\{\det(X_i)=0\}=0$.

Then if θ_n are random variables that almost surely minimize $U_n(\theta), n\in\mathbb{N}$, the sequence

$$\{n^{1/2}\theta_n\}_{n=1}^{\infty}$$

converges in distribution to a normal law.

PROOF. First we observe that

$$\lim_{|\theta|\to\infty}U(\theta)=\infty. \tag{4.1}$$

For simplicity of notation and since the general case is similar, we restrict the proof of (4.1) to the two-dimensional case. Since P is not concentrated on a line, there are three disjoint balls B_1, B_2 and B_3 such that $P(B_i) = \alpha_i > 0$. Using these three balls, we can divide the space into three parts C_1, C_2 and C_3 such that if $x_1 \in B_i, x_2 \in B_j, \theta \in C_k, i, j, k$ all different, and if $|\theta|$ is large, then $f_\theta(x_1, x_2)$ is large. So, if $\theta \in C_3, E(f_\theta(X_1, X_2)I_{X_1 \in B_1, X_2 \in B_2})^2$ is large if $|\theta|$ is large, and the same happens for all possible combinations of x_1, x_2, θ and B and C. Relation (4.1) follows.

Since for n large we have that $P_n(B_i) > 2^{-1}\alpha_i > 0$, the last argument shows that there is a finite M such that

$$\liminf_{n\to\infty} \inf_{|\theta|>M} U_n(\theta) > U(0) + 1 \quad \text{a.s.} \tag{4.2}$$

We fix this M and consider the set $\mathcal{G}$ of graphs of $\mathcal{F} := \{f_\theta : |\theta| \le M\}$,

$$\begin{aligned}\mathcal{G} &= \{(x, t) : f_\theta(\mathbf{x}) \ge t \ge 0, |\theta| \le M\} \\ &= \Big\{(\mathbf{x}, t) : \big(\det(x_i - \theta)\big)^2 - t^{2/\alpha} \ge 0, t \ge 0, |\theta| \le M\Big\}.\end{aligned}$$

With $\theta = (\theta^{(1)}, \dots, \theta^{(d)})$, we have that

$$\det(x_i - \theta) = c_0(\mathbf{x}) + \sum_{j=1}^{d} \theta^{(j)} c_j(\mathbf{x}),$$

for certain polynomials $c_j(\mathbf{x}), j = 0, \dots, d$ (these are polynomials in the coordinates of $x_1, \dots, x_d$). Then the set of functions $(\det(x_i - \theta))^2 - t^{2/\alpha}, \theta \in \mathbb{R}^d$, is contained in a finite-dimensional vector space of functions in the variables $x_1, \dots, x_d, t$. This implies, by Dudley [(1984). Theorem 9.2.1], that $\mathcal{F}$ is a VC-subgraph class of functions. Also, observing that the monomials of the polynomials c_j have degree at most 1 in the coordinates of each vector x_i, condition (i) implies that the envelope of this class is square integrable (for P^d). So, the law of large numbers for U-processes [Arcones and Giné (1993)] gives

$$\sup_{|\theta|\le M} |U_n(\theta) - U(\theta)| \to 0 \quad \text{a.s.} \tag{4.3}$$

By the dominated convergence theorem, $U(\theta)$ is a continuous function. Hence, by (4.1) and condition (ii),

$$\sup_{|\theta|\ge\delta} U(\theta) > U(0), \tag{4.4}$$

for each $\delta > 0$. Relations (4.1)–(4.4) yield, for example, by the argument in the proof of Theorem 6.9 in Arcones and Giné (1993),

$$\theta_n \to 0 \quad \text{in probability}$$

(in fact, a.s.). This gives condition (2.4) of Theorem 2.1.

By Corollary 5.7 in Arcones and Giné (1993), for $1 \le k \le d$, the processes $\{n^{k/2}U_n(\pi_k f_\theta): |\theta| \le M\}_{n=1}^{\infty}$ converge in law in $\ell^\infty(|\theta| \le M)$, and moreover, by dominated convergence,

$$\lim_{\theta \to 0} E(f_\theta - f_0)^2 = 0.$$

Hence, (2.2) in Theorem 2.1 holds [by, e.g., (1.7) in Arcones and Giné (1993)]. So, only condition (iii) of Theorem 2.1 remains to be verified.

Let

$$\Delta(\mathbf{x}) = \alpha |c_0(\mathbf{x})|^{\alpha - 1} (\operatorname{sign} c_0(\mathbf{x})) (c_1(\mathbf{x}), \dots, c_d(\mathbf{x})),$$

and let

$$R_\theta(\mathbf{x}) = \frac{f_\theta(\mathbf{x}) - f_0(\mathbf{x}) - \theta \cdot \Delta(\mathbf{x})}{|\theta|}.$$

The functions r_θ and Δ of Theorem 2.1 will be $r_\theta(x) = \pi_1 s(R_\theta)(x)$ and $\Delta(x) = \pi_1 s(\Delta)(x)$, respectively, where $sf(x_1, \dots, x_d) = (d!)^{-1} \sum_\sigma f(x_{\sigma(1)}, \dots, x_{\sigma(d)})$ is the symmetrization of f (σ runs over all the permutations of $1, \dots, d$). By a previous argument, the class of functions

$$\{R_\theta(\mathbf{x}, \theta): \theta \in \mathbb{R}^d\}$$

is a VC-subgraph class. Using $||x|^\alpha - |y|^\alpha| \le 2^{\alpha - 1}|x - y|^\alpha$, we have

$$|R_\theta| \le |\theta|^{-1} 2^{\alpha - 1} \left| \sum_{j=1}^{d} \theta^{(j)} c_j \right|^\alpha + |\theta|^{-1} \alpha |c_0|^{\alpha - 1} \sum_{j=1}^{d} |\theta^{(j)}||c_j|;$$

hence, for all θ, $|R_\theta| \le 2^{\alpha - 1} |\theta|^{\alpha - 1} (\sum_{j=1}^d c_j^2)^{\alpha/2} + \alpha |c_0|^{\alpha - 1} (\sum_{j=1}^d c_j^2)^{1/2}$, a function in $L_2(P^d)$ for θ bounded. So, Corollary 5.7 in Arcones and Giné (1993) shows that

$$\{n^{1/2}(U_n - P^d)(R_\theta): |\theta| \le \delta\}$$

converges weakly to a Gaussian process for every $\delta > 0$. Condition (v) implies that $\det(X_i - \theta)$ is a.s. differentiable at 0; hence $R_\theta \to 0$ a.s. as $\theta \to 0$, and, by dominated convergence,

$$\lim_{\theta \to 0} E R_\theta^2(X_1, \dots, X_d) = 0.$$

Therefore, it follows from Arcones and Giné [(1993), Corollary 4.2] that

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{|\theta| \le \delta} \left| n^{1/2}(P_n - P)(\pi_1 s(R_\theta)) \right| = 0.$$

This implies that $r_\theta(x)$ and $\Delta(x)$ verify condition (iii) of Theorem 2.1. The result now follows from this theorem. [In fact, if we use the more general result stated in Remark 3.4 instead of Theorem 2.1, then there is no need to invoke Corollary 4.2 of Arcones and Giné (1993)]. □

REMARK 4.2. (a) Condition (iii) in Theorem 4.1 is very natural: if the law of X lives in a proper subspace of $\mathbb{R}^d$, then $U(\theta) = 0$ for all θ in the subspace. (b) condition (v) is not very restrictive; for instance, if the law of X gives mass zero to every proper subspace, then condition (v) holds. (c) If $\alpha = 2k$ for some positive integer, then $U(\theta)$ is a polynomial in θ; hence it is automatically differentiable and therefore condition (v) is not needed to prove $\lim_{\theta \to 0} ER_\theta^2 = 0$. So if α *is even*, the following conditions suffice for asymptotic normality of Oja's median:

(i) $E|X|^{2\alpha} < \infty$;
(ii) $U(\theta)$ attains its minimum at only one point θ_0;
(iii) the distribution of X does not live in a proper subspace of $\mathbb{R}^d$;
(iv) the matrix of second derivatives of U at θ_0 is nondegenerate.

Condition (iv) of Theorem 4.1 requires some smoothness of the law of X if α is not even. Here we give a set of sufficient conditions in the case $\alpha = 1$, which seems to be the most interesting, and is also the least smooth. In the next proposition, f' and f'' denote, respectively, the vector of partial derivatives and the matrix of second partial derivatives of f. The proof, which is standard, is omitted.

PROPOSITION 4.3. *If the following hold*:

(i) *P has a density $f(x)$ which is twice differentiable and its second partial derivatives are continuous*;

(ii)

$$\int |x| f(x)\,dx < \infty, \int |x||f'(x)|\,dx < \infty, \int |x| \|f''(x)\| dx < \infty$$

and, for some $\delta > 0$ and some function g such that $\int |x| g(x)\,dx < \infty$,

$$\sup_{|\theta| \le \delta} \|f''(x+\theta) - f''(x)\| \le g(x);$$

then

$$U(\theta) = U(0) + \theta U'(0) + \theta \cdot A \cdot \theta^t + o(|\theta|^2),$$

where $A = \frac{1}{2} \int |\det(x_i)| \phi_d''(\mathbf{x}, 0)\,d\mathbf{x}$.

It is obvious that full multivariate normal laws satisfy the hypotheses of Theorem 4.1 and Proposition 4.3; the covariance of the limiting normal distribution of Oja's median in this case can be found in Oja and Niinimaa (1985).

REFERENCES

ARCONES, M. A. (1994). Limits of canonical U-processes and B-valued U-statistics. *J. Theoret. Probab.* **7** 339–349.

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

DUDLEY, R. M. (1984). A course on empirical processes. *Ecole d'Eté de Probabilitiés de Saint-Flour XII. Lecture Notes in Math.* **1097** 1–142. Springer, New York.

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

GINÉ, E. and ZINN, J. (1986). Lectures on the central limit theorem for empirical processes. *Probability and Banach Spaces. Lecture Notes in Math.* **1221** 50–113. Springer, Berlin.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–223. Univ. California Press, Berkeley.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191–219.

KOLČINSKII, V. I. (1992). Asymptotic properties of maximal points of stochastic processes with applications to statistical functionals. Unpublished manuscript.

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

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–332.

OJA, H. (1984). Asymptotic properties of estimators based on U-statistics. Unpublished manuscript.

OJA, H. and NIINIMAA, A. (1985). Asymptotic properties of the generalized median in the case of multivariate normality. *J. Roy. Statist. Soc. Ser. B* **47** 372–377.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.* **15** 897–919.

POLLARD, D. (1985). New ways to prove central limit theorems. *Econometric Theory* **1** 295–314.

MIGUEL A. ARCONES
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112

ZHIQIANG CHEN
DEPARTMENT OF MATHEMATICS
WILLIAM PATERSON COLLEGE
WAYNE, NEW JERSEY 07470

EVARIST GINÉ
DEPARTMENT OF MATHEMATICS AND
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
STORRS, CONNECTICUT 06269