The Annals of Statistics
1990, Vol. 18, No. 1, 405–414

ON A NOTION OF DATA DEPTH BASED ON RANDOM SIMPLICES[1]

BY REGINA Y. LIU

Rutgers University

To the memory of my teacher and friend John Van Ryzin

For a distribution F on $\mathbb{R}^p$ and a point x in $\mathbb{R}^p$, the simplicial depth $D(x)$ is introduced, which is the probability that the point x is contained inside a random simplex whose vertices are $p+1$ independent observations from F. Mathematically and heuristically it is argued that $D(x)$ indeed can be viewed as a measure of depth of the point x with respect to F. An empirical version of $D(\cdot)$ gives rise to a natural ordering of the data points from the center outward. The ordering thus obtained leads to the introduction of multivariate generalizations of the univariate sample median and L-statistics. This generalized sample median and L-statistics are affine equivariant.

1. Introduction. The main goal of this paper is to introduce a new notion of data depth. This notion emerges naturally out of a fundamental concept underlying affine geometry, namely that of a *simplex*, and it satisfies the requirements one would expect from a notion of data depth. Thus it leads to an affine invariant, center-outward ranking of the data points. We now turn to a detailed description.

Let $X_1, \dots, X_n$ be a bivariate data set. Given any three data points X_i, X_j and X_k, we can form the closed triangle with vertices X_i, X_j and X_k which we denote by $\Delta(X_i, X_j, X_k)$. From the n data points, we generate in this way $\binom{n}{3}$ triangles. To any point x in $\mathbb{R}^2$ we can associate then the number of those triangles which contain x inside. This number should be larger if x is "deep" inside or near the "center" of the data cloud, and smaller if x is relatively near its outskirts. This suggests the following notion of depth measure, which we shall call *simplicial depth* since it is based on triangles and their p-dimensional generalizations, which are simplices. Denote by $x \in \Delta(X_i, X_j, X_k)$ the event that x falls inside the closed random triangle $\Delta(X_i, X_j, X_k)$ and by $I(A)$ the indicator function of an event A, i.e., $I(A) = 1$ if A occurs and $I(A) = 0$ otherwise. Then

$$(1.1) \qquad D_n(x) \equiv \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} I\big(x \in \Delta(X_i, X_j, X_k)\big)$$

expresses the proportion of triangles containing x. To visualize the situation, we may imagine placing a layer of clay with thickness $\binom{n}{3}^{-1}$ on the region corre-

Received July 1987; revised March 1989.
[1]Research supported by NSF grants DMS-85-02945 and DMS-88-02558.
AMS 1980 *subject classifications*. Primary 62H05, 62H12; secondary 60D05.
Key words and phrases. Simplex, simplicial depth, multivariate median, L-statistics, angularly symmetric distributions, location estimators, consistency, affine equivariance.

sponding to each triangle $\Delta(X_i, X_j, X_k)$, one by one until all $\binom{n}{3}$ triangles are exhausted. The resulting solid will represent the exact shape of $D_n(\cdot)$.

It is clear that $D_n(x)$ defined in (1.1) is an empirical version of the probability

$$(1.2) \qquad D(x) \equiv P_F(x \in \Delta(X_1, X_2, X_3)),$$

if X_i's are i.i.d. with a common distribution function F. The quantity $D(x)$ should assume higher values when x is near the center of the distribution and should tend to 0 as x moves away from the center. We shall refer to $D(x)$ in (1.2) as the *simplicial depth* (SD) of x with respect to F in $\mathbb{R}^2$ and to $D_n(x)$ in (1.1) as the *sample simplicial depth* of x with respect to the data cloud $X_1, \dots, X_n$.

It may be instructive to consider the univariate analog of SD, namely,

$$(1.3) \qquad D(x) = P(x \in \overline{X_1 X_2}).$$

Here x is in $\mathbb{R}^1$, X_1 and X_2 are two independent observations from a univariate c.d.f. F and $\overline{X_1 X_2}$ represents the closed line segment connecting X_1 and X_2. When F is continuous,

$$(1.4) \qquad D(x) = 2F(x)[1 - F(x)].$$

It follows immediately that any point which maximizes $D(x)$ is a population median. The maximum value of $D(\cdot)$ is $\frac{1}{2}$ in this case, and $D(x)$ decreases monotonically to 0 as x is pulled away from the median.

The above observation suggests that we call a point in $\mathbb{R}^2$ which maximizes $D(\cdot)$ a *bivariate simplicial median*. We denote such a point by μ and will also call it center when geometric understanding is emphasized. The sample version of the bivariate median is then

$$(1.5) \qquad \hat{\mu}_n = \text{the data point } X_{i_0} \text{ attaining highest sample SD}.$$

If the maximum is achieved at more than one data point, we can define $\hat{\mu}_n$ as the average of those data points which maximize $D_n(\cdot)$. The heuristic motivation for $\hat{\mu}_n$ as the sample median is the following: If $D(\cdot)$ is continuous and μ is the unique maximizer for $D(\cdot)$ in $\mathbb{R}^2$, an estimator for μ would be a point x_0 in the plane which maximizes $D_n(\cdot)$. If F has a nonzero density in the neighborhood of μ, we would expect the data point X_{i_0} which maximizes $D_n(\cdot)$ among all the data points to be close to x_0 and, hence, to μ. These arguments can actually be made rigorous, as we shall see in Section 3.

A major task here is to show that $D(x)$ defined in (1.2) can indeed be viewed as a measure of depth; that is, to show formally that it possesses some kind of monotonicity property similar to the one that $D(\cdot)$ possesses in the univariate analog. This is established in Theorem 3 of Section 2. To be more precise, the theorem asserts that when the underlying distribution is angularly symmetric (see Section 2 for the definition) about a point μ, then $D(x)$ decreases monotonically as x moves away from μ along any fixed ray.

All the concepts introduced so far can be easily extended to higher dimensions. For a distribution F on $\mathbb{R}^p$, a random triangle in the definition of SD is now replaced by a random simplex whose vertices are $p + 1$ independent observations from F. Consequently, we define:

1. the *simplicial depth* (SD) *function* $D(\cdot)$ on $\mathbb{R}^p$ with respect to F to be

$$D(x) \equiv P_F\big(x \in S[X_1, \dots, X_{p+1}]\big), \tag{1.6}$$

   where $X_1, \dots, X_{p+1}$ are independent observations from F and $S[X_1, \dots, X_{p+1}]$ is the simplex with vertices $X_1, \dots, X_{p+1}$ (in other words, $S[X_1, \dots, X_{p+1}]$ is the set of all points in $\mathbb{R}^p$ which are convex combinations of $X_1, \dots, X_{p+1}$);
2. a (*multivariate*) *simplicial median of* F, μ, to be a point which maximizes $D(\cdot)$;
3. the *sample simplicial depth function* $D_n(\cdot)$ to be

$$D_n(x) \equiv \binom{n}{p+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{p+1} \le n} I\big(x \in S[X_{i_1}, \dots, X_{i_{p+1}}]\big) \tag{1.7}$$

   if $X_1, \dots, X_n$ is a random sample from F;
4. the (*multivariate*) *sample simplicial median* $\hat{\mu}_n$ to be the sample point which maximizes $D_n(\cdot)$, or the average of such points if there are many.

We observe that it is straightforward to check whether or not a point x in $\mathbb{R}^p$ is inside the simplex $S[x_1, \dots, x_{p+1}]$. It actually amounts to solving the following system of linear equations:

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_{p+1} x_{p+1}; \qquad \alpha_1 + \alpha_2 + \cdots + \alpha_{p+1} = 1. \tag{1.8}$$

For a nondegenerate simplex, this system of $p+1$ equations with $p+1$ unknowns $\alpha_1, \alpha_2, \dots, \alpha_{p+1}$ has a unique solution, and x is inside the simplex if and only if $\alpha_1, \alpha_2, \dots, \alpha_{p+1}$ are all positive.

Let A be a $p \times p$ matrix and $b \in \mathbb{R}^p$. Then (1.8) immediately implies that

$$D_{A,b}(Ax + b) = D(x), \tag{1.9}$$

where $D_{A,b}(y)$ is the probability that $y (\in \mathbb{R}^p)$ is contained inside the simplex with vertices $AX_i + b$, $i = 1, \dots, p+1$. In other words, the function $D(\cdot)$ is invariant under affine transformations. Such invariance property clearly holds for the sample SD $D_n(\cdot)$ as well. This property of $D_n(\cdot)$ is sufficient to assert the *affine equivariance property* of all location estimators proposed in this article.

Some applications of simplicial depth.

(A) *Multivariate L-statistics.* In addition to giving the above generalized median, the notion of SD leads to a new way of ordering data points and, consequently, a generalization of the so-called L-statistics (linear combinations of order statistics) in the multivariate setting. Let $X_{[i]}$ be the data point associated with the ith highest sample SD value. Then $X_{[1]}, X_{[2]}, \dots, X_{[n]}$ are the order statistics of X_j's with an ordering *from the center outward*. Let $w(\cdot)$ be a nonincreasing weight function of $[0, 1]$. We define a class of multivariate

L-statistics as

$$L_w = \sum_{i=1}^{n} X_{[i]} w(i/n) / \sum_{j=1}^{n} w(j/n). \tag{1.10}$$

When $w(t) = I(t \leq 1/n)$, L_w is the same as the sample median $\hat{\mu}_n$ if $D_n(\cdot)$ is uniquely maximized among the sample points.

When $w(t) = I(t \leq 1 - \alpha)$, L_w is a $100\alpha\%$ trimmed mean. In practice, $\alpha = 0.05$ or 0.1 are the commonly used values. We would like to mention that the trimmed mean with $\alpha = 0.95$ (or so) should be an appealing alternative to $\hat{\mu}_n$ when the population SD is not uniquely maximized.

(B) *Directional data and simplicial depth.* A direction in the plane can be viewed as a point on a unit circle, while a direction in three-dimensional space can be similarly viewed as a point on a unit sphere. The study of directional data leads to situations where the ambient space is not a p-dimensional Euclidean space, but rather a sphere in $(p - 1)$ dimensions. The notion of simplicial depth can be adapted by using geodesic simplices instead of simplices. For example, in the case of a circle, the short arc connecting two observations is to replace the random line segment used to define SD in $\mathbb{R}^1$. This is investigated in Liu and Singh (1988).

(C) *Testing the center of (angular) symmetry.* We are often required to determine the center of a symmetric population. A class of distributions somewhat broader than the usual symmetric distributions is the class of angularly symmetric distributions. Roughly speaking, a distribution is angularly symmetric about a point x if every hyperplane passing through x divides the whole space into two half-spaces with equal probabilities. (For the precise definition and further discussions, see Section 2.) In the present paper, it is shown (cf. Theorems 3 and 4) that the SD is maximized at the center of angular symmetry and takes there the value 2^{-p} in $\mathbb{R}^p$. Thus, if b_0 is a hypothesized center of angular symmetry, then a large value of $(2^{-p} - D_n(b_0))$ is an indication of the null hypothesis being false. The observation in Remark B of Section 2 says that the test statistic $(2^{-p} - D_n(b_0))$ is a degenerate U-statistic. This fact leads us to conclude that $n(2^{-p} - D_n(b_0))$ has as its weak limit a linear combination of χ^2-distributions [cf. Gregory (1977)]. A detailed study of this testing procedure will appear separately.

Other applications of SD include deriving a class of multivariate scales and a multivariate classification rule. In fact, a measure of scale can be derived by considering how far away one has to move from the center (i.e., the maximum point of the sample SD) in order to reduce the SD value to a fraction of its maximum. As for classification, the idea there is roughly the following [see Gross and Liu (1988)]: Suppose that two training samples from two different populations are given. A classification rule is a way of assigning any new data point Z to one of these two populations. Such a rule can be obtained by comparing the relative center-outward ranks of Z w.r.t. the training samples. Z should be assigned to the population whose training sample leads to a smaller relative rank for Z.

General remarks. An earlier concept of data depth was introduced by Tukey (1975). Tukey's data depth and the related sample median studied in Stahel (1981), Donoho (1982), and Donoho and Gasko (1988) are based on the inspection of "every" one-dimensional projection of the sample data. In a different direction, Oja (1983) defined a sample median in $\mathbb{R}^p$ as a point x which yields the minimum total volume of all simplices formed by x and p of the data points. As far as the "generalized" multivariate median is concerned, there is an extensive literature and a thorough coverage can be found in Rousseeuw and Leroy (1987).

2. Main properties of the simplicial depth function $D(\cdot)$. The main properties of $D(\cdot)$ are summarized in Theorems 1–4.

THEOREM 1. *For any F on $\mathbb{R}^p$ and $x \in \mathbb{R}^p$, $\sup_{\|x\| \geq M} D(x) \to 0$ as $M \to \infty$.*

THEOREM 2 [Continuity of $D(\cdot)$]. *If F is an absolutely continuous distribution on $\mathbb{R}^p$, then $D(\cdot)$ is continuous.*

The next two theorems are stated for angularly symmetric distributions. The reason we focus on these distributions is that they form a large class of distributions possessing an obvious center, and we shall show that this center agrees with the one predicted by the simplicial depth function.

DEFINITION. *A random variable X in $\mathbb{R}^p$ or its distribution F is said to be angularly symmetric about the point b (in $\mathbb{R}^p$) if and only if the random variables $(X - b)/\|X - b\|$ and $-(X - b)/\|X - b\|$ are identically distributed, where $\|\cdot\|$ stands for the Euclidean norm.*

For $p = 2$, F is angularly symmetric about b simply means $a_b(\theta) = a_b(\theta + \pi)$ for all θ, $0 \leq \theta < \pi$, where $a_b(\cdot)$ is the angular density around the point b induced by F provided that such angular density exists. It is easy to see that if F is symmetric about b, then F is angularly symmetric about b. It is also easy to see that if F is angularly symmetric about b, then any hyperplane passing through b will divide $\mathbb{R}^p$ into two open half-spaces with equal probabilities. This probability is $\frac{1}{2}$ if the distribution is absolutely continuous. Thus the center of angular symmetry is what one would want as a (multivariate) median. In view of this and Theorem 3, it is only natural to define a median by the maximal point of $D(\cdot)$. Finally, we note that the center of angular symmetry is unique when it exists, except in the case when the distribution F has its whole probability mass concentrated on a line and its probability distribution along that line has more than one median. In fact, if b_1 and b_2 are two different centers of angular symmetry, then the region between any two parallel hyperplanes passing through b_1 and b_2, respectively, would have zero probability. Rotating these two hyperplanes, it follows that the entire $\mathbb{R}^p$ except for the line passing through b_1 and b_2 has zero probability.

THEOREM 3 [Monotonicity of $D(\cdot)$]. *If F is absolutely continuous and angularly symmetric about the origin, then $D(\alpha x)$ is a monotone nonincreasing in $\alpha \geq 0$ for all $x \in \mathbb{R}^p$.*

THEOREM 4. *If F is an absolutely continuous distribution on $\mathbb{R}^p$ and it is angularly symmetric about $b \in \mathbb{R}^p$, then $D(b) = 2^{-p}$.*

In particular, Theorems 3 and 4 imply that for any point a in $\mathbb{R}^p$, $D(a) \leq 2^{-p}$ if F is an angularly symmetric distribution.

Before discussing the proofs of Theorems 1–4, we pause to make two remarks.

REMARK A. Theorem 3 is equivalent to saying that the contours defined by $\{x \in \mathbb{R}^p\colon D(x) = c\}$ for positive numbers $c \leq 2^{-p}$ are nested within one another. As c decreases, they move further and further away from the center. Their geometry should contain useful information about the distribution F. In the special case when F is spherical, each contour is a circle and $D(x)$ is a monotonic function of $\|x\|$. In the case of an elliptical distribution, i.e., when the density at x is a function of $(x - \mu)'V^{-1}(x - \mu)$, it is not hard to show that $D(x)$ is also a function of $(x - \mu)'V^{-1}(x - \mu)$. In other words, the contours of $D(\cdot)$ resemble the contours of the underlying density in the elliptical case. This observation again confirms that $D(\cdot)$ indeed provides us with an appropriate notion of ordering.

REMARK B. The Proof of Theorem 4 will further imply the following fact: Under the assumption of angular symmetry at a center b_0, the conditional SD value at b_0 given one of the random vertices is the same as the unconditional one. In other words,

$$P\big(b_0 \in S[X_1, \dots, X_{p+1}] | X_i\big) = 2^{-p} \tag{2.1}$$

for each $i = 1, \dots, p + 1$. Evidently, (2.1) implies that $(2^{-p} - D_n(b_0))$ is a degenerate U-statistic, that is $E[(D_n(b_0) - 2^{-p})|X_i] = 0$ for all i, $1 \leq i \leq n$.

PROOF OF THEOREM 1. Given x in $\mathbb{R}^p$, we observe that the event $\{x \in S[X_1, \dots, X_{p+1}]\}$ is contained in the event $\bigcup_{i=1}^{p+1}\{\|X_i\| \geq \|x\|\}$. The theorem follows. □

PROOF OF THEOREM 2. Let $p = 2$ for simplicity. Let x and y be two distinct points. A random triangle can contribute to the difference $D(x) - D(y)$ only if it contains one point but not the other. This however implies that there must be a line segment joining two data points which intersects the line segment $\overline{xy}$. It follows that if x_n is a sequence in $\mathbb{R}^2$ which converges to x, then

$$|D(x) - D(x_n)| \leq 3P(A_n),$$

where $A_n = \{(X_1, X_2): \overline{X_1X_2} \text{ intersects } \overline{xx_n}\}$. Note that $\limsup_{n\to\infty} P(A_n) \le P(\limsup_{n\to\infty} A_n)$. Note also that $\limsup_{n\to\infty} A_n = \{(X_1, X_2): x \in \overline{X_1X_2}\}$. Since $P(\limsup_{n\to\infty} A_n) = 0$, the theorem follows. □

PROOF OF THEOREM 3. Let $p = 2$ for simplicity. The idea is to examine two different events which contribute to the difference $D(x) - D(\alpha x)$, $\alpha \ge 1$. They are the events that the arrow from x to αx enters or leaves the random triangle $\Delta(X_1, X_2, X_3)$. We shall call them A_{in} and A_{out}, respectively.

To make the argument precise, we need some notation. For any two distinct points a, b, the line which contains $\overline{ab}$ divides the plane into two half-planes. If that line does not contain the origin, we call the half-plane which contains the origin the "inner side" $I(a, b)$. Let $x, \alpha \ge 1$ be fixed and $C = \{(a, b): \overline{ab} \cap \overline{x, \alpha x} \ne \varnothing\}$ be the set of all segments which intersect with $\overline{x, \alpha x}$. Then, neglecting null sets, $A_{\text{in}} = A_{\text{in}}^{12} \cup A_{\text{in}}^{23} \cup A_{\text{in}}^{31}$, where $A_{\text{in}}^{12} = \{(X_1, X_2) \in C\} \cap \{X_3 \notin I(X_1, X_2)\}$ and the three events A_{in}^{ij} are disjoint and equally probable. Similar remarks hold for A_{out}.

If $B_\alpha = \{\alpha x \in \Delta(X_1, X_2, X_3)\}$, then, according to the definition of $D(\cdot)$, $D(x) - D(\alpha x) = P(B_1 \setminus B_\alpha) - P(B_\alpha \setminus B_1)$. Now,

$$B_1 \setminus B_\alpha = A_{\text{out}} \setminus A_{\text{in}},\ B_\alpha \setminus B_1 = A_{\text{in}} \setminus A_{\text{out}}$$

and

$$\begin{aligned} D(x) - D(\alpha x) &= P(A_{\text{out}}) - P(A_{\text{out}} \cap A_{\text{in}}) - [P(A_{\text{in}}) - P(A_{\text{in}} \cap A_{\text{out}})] \\ &= 3P(A_{\text{out}}^{12}) - 3P(A_{\text{in}}^{12}) \\ &= 3\int_{(x_1, x_2)\in C} [P(X_3 \in I(x_1, x_2)) \\ &\qquad - P(X_3 \notin I(x_1, x_2))]\, dF(x_1)\, dF(x_2). \end{aligned} \tag{2.2}$$

Because of the angular symmetry, $P(X_3 \in I(x_1, x_2)) \ge \frac{1}{2}$ and the integrand is nonnegative. This proves the assertion. □

REMARK C. From (2.2) in the Proof of Theorem 3, we may deduce that $D(x) - D(\alpha x) > 0$ for any $\alpha > 1$ if the following two additional conditions hold: (i) f is positive in a neighborhood of the origin, and (ii) f is positive in a neighborhood of βx for some β such that $1 < \beta < \alpha$, where f denotes the density of F. Clearly, (i) implies that the integrand in (2.2) is positive almost surely, and (ii) implies that the domain of the integral in (2.2) has positive probability. In particular, $D(\cdot)$ is *uniquely maximized* at the origin under condition (i).

PROOF OF THEOREM 4. W.l.o.g. we may assume that F is angularly symmetric about the origin 0. In this case $X_i^* \equiv X_i/\|X_i\|$ and $(-X_i^*)$ are identically distributed, and the following four events are equivalent except for a null set:

(i) $\{(X_1, \dots, X_p, X_{p+1}): 0 \in S[X_1, \dots, X_p, X_{p+1}]\}$;
(ii) $\{(X_1, \dots, X_p, X_{p+1}): 0 \in S[X_1^*, \dots, X_p^*, X_{p+1}^*]\}$;
(iii) $\{(X_1, \dots, X_p, X_{p+1}): 0 \in S[e_1, \dots, e_p, [[X_1^*, \dots, X_p^*]]^{-1} X_{p+1}^*]\}$, where e_i

is the ith unit vector in $\mathbb{R}^p$ and $[[X_1^*, \ldots, X_p^*]]$ is the matrix with columns $X_1^*, \ldots, X_p^*$;

(iv) $\{(X_1, \ldots, X_p, X_{p+1})\colon W_1 < 0, \ldots, W_p < 0\}$, where W_i is the ith component of the vector $[[X_1^*, \ldots, X_p^*]]^{-1} X_{p+1}^*$.

By exchanging X_i^* with $-X_i^*$, we can show that the random vector $[[X_1^*, \ldots, X_p^*]]^{-1} X_{p+1}^*$ is coordinatewise symmetric about the origin. This implies that each "orthant" determined by $(W_1, \ldots, W_p)'$ has an equal probability, which must be 2^{-p}. Therefore the event (iv) has the probability 2^{-p} and Theorem 4 follows. □

3. Consistency of the sample simplicial depth $D_n(\cdot)$.

THEOREM 5. *Let F be an absolutely continuous distribution on $\mathbb{R}^p$ with bounded density f. Then*:

(a) *The uniform consistency of $D_n(\cdot)$ holds, i.e.*,

$$\sup_{x \in \mathbb{R}^p} |D_n(x) - D(x)| \to 0 \quad a.s. \text{ as } n \to \infty.$$

(b) *Furthermore, if f does not vanish in a neighborhood of μ and if $D(\cdot)$ is uniquely maximized at μ, then $\hat{\mu}_n \to \mu$ a.s. as $n \to \infty$.*

The proof of Theorem 5 is based on the following three lemmas.

LEMMA 1. *For any F on $\mathbb{R}^p$ and $x \in \mathbb{R}^p$,*

$$\sup_{\|x\| \geq M} D_n(x) \to 0 \quad a.s. \text{ as } M \to \infty.$$

LEMMA 2. *Suppose that F is absolutely continuous. Let δ and c be arbitrary but fixed positive constants. Then, for any positive ε, we have*

$$\sup_{\{x,\, y \in \mathrm{Ball}(\mu, c),\, \|x - y\| < \varepsilon\}} |D_n(x) - D_n(y)| \leq \gamma(\varepsilon) + \delta + R_n,$$

where $\gamma(\varepsilon)$ is nonrandom, $\gamma(\varepsilon) \to 0$ as $\varepsilon \to 0$ and $R_n \to 0$ a.s. as $n \to \infty$.

LEMMA 3. *Let F be a distribution on $\mathbb{R}^p$ and $X_1, \ldots, X_n$ be a random sample from F. Let $U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} h(X_{i_1}, \ldots, X_{i_m})$ be a U-statistic with the kernel $h(\cdot)$ of degree m. If h is bounded, say by c, then for any $r \geq 2$,*

$$E(U_n - EU_n)^r \leq \frac{K}{n^{r/2}},$$

where K depends on c.

Note that Lemmas 1 and 2 are close in spirit to Theorems 1 and 2, respectively. Since Lemma 2 concerns the sample version of $D(\cdot)$, we need to invoke the crucial fact that the class of all convex Borel measurable sets in $\mathbb{R}^p$ form a

Glivenko–Cantelli class if F has a density w.r.t. Lebesgue measure [cf. Gaenssler and Stute (1979)]. In other words,

$$\sup_{A \in \mathscr{C}} |F_n(A) - F(A)| \to 0 \quad \text{a.s.},$$

where $\mathscr{C}$ is the class of all convex Borel measurable sets. We refer to Liu (1987) for the details. Lemma 3 is needed because $D_n(x)$ is a U-statistic. In fact, Lemma 3 is essentially Lemma A on page 185 of Serfling (1980).

Finally, we come to the Proof of Theorem 5.

PROOF OF THEOREM 5. By Theorem 1 and Lemma 1, part (a) will follow if we can show that, for a chosen $M > 0$,

$$\sup_{x \in Q(\mu, M)} |D_n(x) - D(x)| \to 0 \quad \text{a.s. as } n \to \infty, \tag{3.1}$$

where $Q(\mu, M)$ is the hypercube with μ as its center and M as the length of its sides.

Divide each side of $Q(\mu, M)$ into N equal pieces to form N^p subhypercubess. In view of Theorem 2 and Lemma 2, since N can be arbitrarily large, we only need to show that

$$\max_{x \in C(\mu, M)} |D_n(x) - D(x)| \to 0 \quad \text{a.s. as } n \to \infty, \tag{3.2}$$

where $C(\mu, M)$ is the set of all corner points of the subhypercubes.

Using Lemma 3 with $m = p + 1$, $r = 4$ and $c = 1$, we obtain

$$P\Big(\max_{x \in C(\mu, M)} |D_n(x) - D(x)| > \varepsilon\Big)$$
$$\le N^p \max_{x \in C(\mu, M)} P(|D_n(x) - D(x)| > \varepsilon) = O(n^{-2}).$$

The claim (3.2) therefore follows from the Borel–Cantelli lemma.

The idea of the proof of part (b) can be outlined as follows: We begin with two balls, each centered at μ. The radius of the bigger ball is arbitrarily small but fixed. Then it is shown that the D_n value at any point inside the inner ball is larger than that at any point outside the bigger ball, for all large n. Since, for all large n, at least one data point will fall inside the inner ball, the possibility of $\hat{\mu}_n$ lying outside the bigger ball is ruled out.

Assuming that $D(\cdot)$ is uniquely maximized at μ, we see that, for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $D(x) \le D(\mu) - \delta$ for all $x \notin \text{Ball}(\mu, \varepsilon)$. By the continuity of $D(\cdot)$ (cf. Theorem 2), we may choose $\varepsilon_1 < \varepsilon$ such that $|D(y) - D(\mu)| < \delta/2$ for all $y \in \text{Ball}(\mu, \varepsilon_1)$. Thus, $D(x) \le D(y) - \delta/2$ for all $x \notin \text{Ball}(\mu, \varepsilon)$ and $y \in \text{Ball}(\mu, \varepsilon_1)$. The uniform convergence of D_n to D given in part (a) of Theorem 5 guarantees that, starting from a certain n, $D_n(x) \le D_n(y) - \delta/4$ for all $x \notin \text{Ball}(\mu, \varepsilon)$ and $y \in \text{Ball}(\mu, \varepsilon_1)$.

Now we claim that there is at least one sample point inside the smaller ball $\text{Ball}(\mu, \varepsilon_1)$ for all large n, almost surely. Since f does not vanish in a neighborhood of μ, we have

$$\rho \equiv P(X_1 \in \text{Ball}(\mu, \varepsilon_1)) > 0.$$

Consequently,

$$P(\text{Ball}(\mu, \varepsilon_1) \text{ does not contain any of } X_1, \dots, X_n) = (1 - \rho)^n.$$

Therefore, almost surely, after certain n, there exists some sample point, say X_k, inside $\text{Ball}(\mu, \varepsilon_1)$. By the definition of $\hat{\mu}_n$, $D_n(\hat{\mu}_n) \geq D_n(X_k)$ and, hence, $\hat{\mu}_n \in \text{Ball}(\mu, \varepsilon)$. Since ε can be chosen arbitrarily small, part (b) follows. □

Acknowledgments. I would like to thank Joop Kemperman for his encouragement and many helpful discussions. I am grateful to the referee and the Editor for their kind suggestions, which helped improve the article greatly.

REFERENCES

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Qualifying paper, Harvard Univ.

DONOHO, D. and GASKO, M. (1988). Multivariate generalization of the median and trimmed means. I. Unpublished.

GAENSSLER, P. and STUTE, W. (1979). Empirical processes: A survey of results for independent and identically distributed random variables. *Ann. Probab.* **7** 193–243.

GREGORY, G. G. (1977). Large sample theory for U-statistics and tests of fit. *Ann. Statist.* **5** 110–123.

GROSS, S. and LIU, R. (1988). Classification rules based on concepts of data depth. Unpublished.

LIU, R. (1987). Simplicial depth and the related location estimators. Technical report, Dept. Statist., Rutgers Univ.

LIU, R. and SINGH, K. (1988). On the ordering of directional data: Concepts of data depth on circles and spheres. Unpublished.

OJA, H. (1983). Descriptive statistics for multivariate distribution. *Statist. Probab. Lett.* **1** 327–333.

ROUSSEEUW, P. J. and LEROY, A. (1987). *Robust Regression and Outliers Detection.* Wiley, New York.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

STAHEL, W. A. (1981). Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen. Ph.D. dissertation, ETH, Zürich.

TUKEY, J. W. (1975). Mathematics and picturing data. *Proceedings of International Congress of Mathematics, Vancouver,* **2** 523–531.

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903