L_1-*Statistical Procedures and Related Topics*
IMS Lecture Notes – Monograph Series (1997) Volume 31

Modes, caps and concentration: A geometric approach to estimation on the sphere

Philip Milasevic [1]

University of Lausanne, Switzerland

Deborah Nolan

University of California at Berkeley, USA

Abstract: This paper surveys some recent results on mode and concentration estimation in multidimensions, including excess-mass sets and multidimensional quantiles. Extensions of these estimators to the hypersphere are developed here. In particular, the modal direction is measured by the center of a minimal cap, and concentration is measured by a function of the opening of that cap. For samples from a distribution for which the minimal cap is unique, it is shown that the center and the opening of the empirical cap are strongly consistent estimators for their respective parameters. Rates of convergence and limiting distributions of the estimators are established by means of empirical process theory.

Key words: Multidimensional mode estimation, directional data, cube-root rates, empirical processes.

AMS subject classification: Primary 62H12; secondary 62H11, 62H10, 60G99.

1 Background

A variety of location estimators for multidimensional data have been recently proposed and investigated. Examples of extensions of the median to higher dimensions include the L_1 median (Brown, 1983, and Ducharme and Milasevic, 1987), Oja's simplex (Oja, 1983), the halfspace median (Donoho

[1] Milasevic died while this manuscript was in preparation. Nolan would like to dedicate it to his memory.

and Gasko, 1992), and the simplicial depth median (Liu, 1990). For an overview of recent results in this area see Small (1990). Small (1990) also discusses extensions of these notions of the median to directional data, and Liu and Singh (1992) investigate them in greater detail. Asymptotic properties of the simplicial depth were studied by Arcones et. al. (1994) and Dümbgen (1992), those of the halfspace median have been studied by Nolan (1992), and Chaudhuri (1996) has presented a general approach to studying quantiles in multidimensions.

Parallel to the development of the multidimensional median there have been investigations into the properties of estimators of the mode and concentration. Chernoff (1964) and Venter (1967) estimate the mode of a density function in one dimension by the center of the interval of fixed length to contain the greatest number of observations and by the center of the shortest interval to contain at least half of the observations, respectively. Sager (1979) generalized these univariate set statistics to the multidimensional case. He estimates the contours of a unimodal density by a sequence of nested convex sets. The first and largest set is the smallest convex set to contain a fixed proportion q of the observations; the second set is the smallest convex set that contains proportion q of the observations within the first set, and so forth. Eddy and Hartigan (1977) proposed a similar multidimensional estimator. The asymptotic properties of the shorth, the shortest interval to contain at least half of the observations, were investigated by Grübel (1988) and Kim and Pollard (1990). Grübel (1988) handled the length of the shorth, and Kim and Pollard developed theory for cube-root rates of convergence to address the center of the shorth. Einmahl and Mason (1992) produced asymptotic theory for generalizations of the length of the shorth.

Close relatives to these estimators of the mode are contour estimators based on excess-mass, proposed independently by Hartigan (1987) and Müller and Sawitzki (1991). An excess mass set for a distribution is the set that maximizes the difference between the probability content of the set and a multiple of its Lebesgue measure.

Nolan (1991) considered the properties of these sets when restricted to ellipsoids, and found the parameters of the ellipsoid have cube-root rates of convergence. Polonik (1995a,b) provides a comprehensive investigation into the properties of these excess mass sets, their connections to maximum likelihood estimation under shape restrictions, and their use in tests of multimodality. He (Polonik, 1995b) shows that the excess mass sets can be used to form a density estimator. The estimator coincides with Grenander's estimator in one dimension when the sets are restricted to intervals with left endpoint 0, and with Sager's estimator in higher dimensions.

2 The hypersphere

Here, we consider the extension to directional data of the shorth's intuitive geometric approach to measuring location and concentration. Location and concentration of the distribution are derived from the smallest cap on the sphere that has probability content at least α, for $0 < \alpha < 1$. When unique, the center of this minimal cap can be interpreted as a *modal* direction of the distribution and the cosine of the half-angle of this cap is a measure of *concentration*. Given a sample of size n on the sphere, the modal direction and concentration can be estimated from the empirical minimal cap, the smallest cap that contains at least $n\alpha$ observations.

The method proposed here is in some sense a geometric counterpart to Watson's estimator (Watson, 1983, Chapter 5) because it too can provide information on location for axial distributions. For example, with the Scheidegger-Watson, Arnold and uniform distributions, there is no unique minimal cap, but the collection of centers of the caps provides meaningful modal directions. For the Scheidegger-Watson distribution the collection of centers is the axis of symmetry; for the Arnold distribution it is the plane of the girdle and for the uniform distribution it is the entire sphere. The minimal cap differs from Watson's estimator and other current estimators of location and concentration on the hypersphere (Ducharme and Milasevic, 1987, Fisher et. al., 1987, Watson, 1983) in its geometric rather than metric nature. As with other mode estimates, the convergence of the center of the sample minimal cap is at a cube–root rate. The concentration of the cap however has a square–root rate of convergence, as it is similar in behavior to a quantile estimator.

To formally define the minimal cap, we introduce some notation. Let S denote the Euclidean unit sphere S^{p-1} in R^p and $C(u,t)$ the (hyper)spherical cap with center u and half-angle $\arccos(t)$. That is, $C(u,t) = \{x \in S : u'x \geq t\}$, for $t \in [-1,1]$. To simplify notation, given a distribution F on S we shall write $F(u,t)$ for $F(C(u,t))$.

Definition 1 *Let $0 < \alpha < 1$ and F be a probability measure on S. A cap $C(u_0,t_0)$ is called a minimal α-cap of F if $F(u_0,t_0) \geq \alpha$ and if for each $t > t_0$, $\sup_{u\in S} F(u,t) < \alpha$.*

Definition 2 *Given a set $X^{(n)}$ of n points on S, $C_n = C(u_n,t_n)$ is called a minimal α-cap of $X^{(n)}$ if it is a minimal α-cap of the empirical measure F_n based on $X^{(n)}$.*

Note that $t_0 = \sup\{t : \sup_{u\in S} F(u,t) \geq \alpha\}$, for $0 < \alpha < 1$. We shall call this value the *α-concentration coefficient* of F. If $C(u_0,t_0)$ is the unique minimal α-cap of F then t_0 is the $(1-\alpha)$-quantile of the variable $u_0'X$, where X has distribution F.

The minimal α-cap of a distribution F is unique with center $u_0 \in S$ if and only if F satisfies the following property:

(A) *For every $u \neq u_0$ and every t such that $F(u,t) \geq \alpha$, there exists $t' > t$ such that $F(u_0, t') \geq \alpha$.*

We shall denote by $\mathcal{M}_\alpha(u_0)$ the class of absolutely continuous distributions on S verifying property (A) and by $\mathcal{M}(u_0)$ the class $\bigcap_\alpha \mathcal{M}_\alpha(u_0)$. An absolutely continuous distribution F belongs to the latter class if and only if it satisfies the following "unimodality" property:

(B) *For every $u \neq u_0$ and every t, there exists $t' > t$ such that $F(u_0, t') \geq F(u,t)$.*

An absolutely continuous distribution with unimodal density does generally not satisfy property (B). It is however the case if moreover F is rotationally symmetric about the mode of its density. The Langevin distribution is such an example. Its density is proportional to $\exp(ku_0'x)$, and so is both unimodal at u_0 and rotationally symmetric about u_0. Not surprisingly, the α-concentration coefficient is a strictly increasing function of the concentration parameter k appearing in the density. For $p = 3$ we have $t_0 = k^{-1}\log[e^k - 2\alpha\sinh(k)]$. On the other hand, if f is rotationally symmetric and bimodal, then there are two minimal α-caps $C(u_0, t_0)$ and $C(-u_0, t_0)$, which provide a unique axis of rotation. Finally, the normalized mean (see Watson, 1983) and the normalized median (see Ducharme and Milasevic, 1987) both coincide with u_0 if F belongs to $\mathcal{M}(u_0)$.

The following algorithm will generally allow us to find a minimal α-cap of a set $X^{(n)}$ of points on S. It is stated for the case $p = 3$ but can be generalized to any dimension. It points out the similarity between the minimal cap and the minimum-volume sphere.

S1. *For each triple of points of $X^{(n)}$ consider the circle on S^2 determined by them. Sort the circles according to the number of elements of $X^{(n)}$ in them, and for each $[\alpha n]$, choose among the caps with $[\alpha n]$ elements those with minimal opening.*

S2. *For each pair of points of $X^{(n)}$ consider the smallest circle on S^2 containing the two points and then proceed as in S1.*

S3. *Let C_n^* be the smallest of the two caps obtained in S1 and S2.*

Proposition 1 *If $X^{(n)}$ is in general position, i.e. no more than p points of $X^{(n)}$ lie on an affine hyperplane, then C_n^* is a minimal α-cap of $X^{(n)}$.*

The proposition states that every minimal cap can be found via the above algorithm. This follows from the fact that the boundary of a minimal cap must contain two or three points of $X^{(n)}$. Otherwise, there would exist a smaller cap with the same number of elements, which contradicts the minimality property.

To reduce computation time, the algorithm proposed by Rousseeuw and Leroy (1987) can be adapted to this problem. Rather than searching over all caps determined by the $\binom{n}{3} + \binom{n}{2}$ subsets of observations, a set of m triples of observations can be selected at random. For each triple, the corresponding cap is computed; then the cap is shrunk or expanded until it contains $[\alpha n]$ points. The minimal cap is chosen from among these m α–caps, and the order of operations is now reduced to $O(mn)$.

3 Asymptotic properties

Let $X^{(n)}$ be an i.i.d. sample of size n from $F \in \mathcal{M}_\alpha(u_0)$, and let $C(u_n, t_n)$ be a minimal α-cap of $X^{(n)}$. Note that $X^{(n)}$ is almost surely in general position. We then have the following

Proposition 2 *Assume that $F(u_0, t_0 - \delta) > \alpha$ for each (admissible) $\delta > 0$. Then i) $t_n \longrightarrow t_0$ almost surely as $n \to \infty$ and ii) $u_n \longrightarrow u_0$ almost surely as $n \to \infty$.*

In the case where F is rotationally symmetric about u_0, its density is of the form $f(u_0'x)$ for some suitable function on $[-1, 1]$. The assumption of the proposition is then satisfied if, for example, f is continuous at t_0 and $f(t_0) > 0$.

Note that u_n can be defined as any unit vector maximizing the $(1-\alpha)$-quantile of the set $\{u'X_1, .., u'X_n\}$ and that t_n is the value of this maximal quantile. We shall prove that the asymptotic distribution of t_n is the same as that of the empirical $(1-\alpha)$-quantile of the variable $u_0'X$. More precisely, the following result holds.

Proposition 3 *Assume that $F \in \bigcap_{|\beta-\alpha|<\epsilon} \mathcal{M}_\beta(u_0)$ for some $\epsilon > 0$ and that the density p_0 of $u_0'X$ is positive and continuous at t_0. Then*

$$\sqrt{n}(t_n - t_0) \longrightarrow \mathcal{N}(0, \frac{\alpha(1-\alpha)}{p_0^2(t_0)}).$$

Now turn to the behavior of u_n. For the next result, we assume that $F \in \mathcal{M}_\alpha(u_0)$ and that F has a rotationally symmetric density of the form $f(u_0'x)$ for some f defined on $[-1, 1]$. Moreover, we only consider the cases where $p \geq 3$ and $\alpha \leq 1/2$. The case $p = 2$ amounts to the one-dimensional context of the shorth, which is treated in Kim and Pollard (1990). The restriction on α ensures $t_0 \geq 0$, and greatly simplifies the covariance structure of the limit process.

Denote by I_q the q-dimensional identity matrix and by B_{p-1} the unit closed ball $\{u^* \in R^{p-1}; |u^*| \leq 1\}$. For a measurable subset $E \subset S^q$, $Area(E)$ represents the q-dimensional volume of E. Given a function h and a distribution P, write Ph for the expectation of h under P. Let $Z(x)$ be a Gaussian process indexed by R^{p-1} with continuous sample paths, zero expectation and covariance kernel

$$\Gamma(x,y) = (p-2)^{-1}\ Area(S^{p-3})\ f(t_0)(1-t_0^2)^{\frac{p-2}{2}}(|x|+|y|-|x-y|).$$

Proposition 4 *Assume that f is twice differentiable on $(-1,1)$ and that $f'(t_0) > 0$. Then*

$$n^{1/3}(I_p - u_0 u_0')u_n \xrightarrow{\mathcal{L}} x_{max},$$

where x_{max} is the almost surely unique vector maximizing $Z(x) + \frac{1}{2}x'Ux$, with

$$U = -\frac{1}{p-1}\ Area(S^{p-2})\ f'(t_0)(1-t_0^2)^{\frac{p-1}{2}} I_{p-1}.$$

Corollary 1 *Under the hypotheses of Proposition 4,*

$$2n^{2/3}(1-u_0'u_n) \xrightarrow{\mathcal{L}} |x_{max}|^2.$$

The two different rates of convergence, cube-root for the direction and square-root for the concentration, parallel those of the center and length of the shorth. The minimal caps can be extended to the excess-mass approach by finding the cap that maximizes

$$F_n(u,t) - \alpha Area(C(u,t)).$$

In this case, both the direction and opening of the excess-mass cap would have cube-root rates of convergence. Properties of the excess-mass cap are not addressed here.

4 Proofs

Proof of Proposition 2

i) The class of sets $\mathcal{C} = \{C(u,t); u \in S,\ -1 \leq t \leq 1\}$ is a Vapnik-Cervonenkis class and thus (e.g. Pollard, 1984)

$$\gamma_n = \sup_{u,t} |F_n(u,t) - F(u,t)| \xrightarrow{a.s.} 0.$$

It follows that for each $\delta > 0$, $F_n(u_0, t_0 - \delta) > \alpha$ eventually, almost surely. This implies that, for n large enough, $t_n \geq t_0 - \delta$ except on a set of measure zero. Moreover, the almost sure convergence of γ_n to 0 implies that

$$\sup_u |F_n(u, t_0 + \delta) - F(u, t_0 + \delta)| \xrightarrow{a.s.} 0,$$

which in turn gives the almost sure upper bound $t_n < t_0 + \delta$ for n large enough.

ii) Note first that for each $\eta > 0$ there exists $\epsilon > 0$ such that

$$\sup_{\|u-u_0\| \geq \eta} F(u, t_0) = \alpha - \epsilon.$$

It then follows from continuity of $F(.,.)$ that there exists $\delta > 0$ such that

$$\sup_{\|u-u_0\| \geq \eta, |t-t_0| \leq \delta} F(u, t) < \alpha - \epsilon/2.$$

Consequently, from i,

$$\sup_{\|u-u_0\| \geq \eta} F(u, t_n) < \alpha - \epsilon/2$$

eventually, almost surely. The latter inequality and $\gamma_n \xrightarrow{a.s.} 0$ imply that $\|u_n - u_0\| < \eta$, almost surely. □

Proof of Proposition 3

The idea is to wedge t_n between two random variables having the same asymptotic behavior. Let s_n be the $(1 - \alpha)$ sample quantile of $\{u_0' X_1, .., u_0' X_n\}$. Then by definition

$$F_n(u_0, s_n) = \alpha + O_P(1/n)$$

and it follows from the minimality of $C(u_n, t_n)$ that $s_n \leq t_n$.

The hypotheses of the proposition entail (e.g. see Serfling, 1980) that

$$\sqrt{n}(s_n - t_0) \longrightarrow \mathcal{N}(0, \frac{\alpha(1-\alpha)}{p_0^2(t_0)}).$$

Now we find an upper bound for $\sqrt{n}(t_n - t_0)$ with the same asymptotic distribution as $\sqrt{n}(s_n - t_0)$. We have

$$\begin{aligned}
\alpha + O_P(1/n) &= F_n(u_n, t_n) \\
&= F(u_n, t_n) + (F_n - F)(u_n, t_n) \\
&\leq F(u_0, t_n) + (F_n - F)(u_n, t_n) \\
&= \alpha - p_0(t_0)(t_n - t_0) + o_P(t_n - t_0) + (F_n - F)(u_0, t_0) \\
&\quad + o_P(n^{-1/2}).
\end{aligned}$$

The inequality follows from the hypothesis on F and consistency of t_n. It holds on a set A_n of probability tending to one. Note here that the hypothesis of Proposition 2 is satisfied. The last o_P term follows from consistency of u_n and t_n and the fact that the process $\sqrt{n}(F_n - F)$ indexed by the class $\mathcal{C}$ is stochastically equicontinuous. This sequence of inequalities imply that on A_n

$$\sqrt{n}(t_n - t_0) \leq \frac{1}{p_0(t_0)}\sqrt{n}(F_n - F)(u_0, t_0) + o_P(1).$$

Therefore, the random variable on the right hand side converges weakly to the desired normal distribution (Pollard, 1984, Theorem VII.21). □

Proof of Proposition 4

We may assume without loss of generality that $u_0 = (0, ..., 0, 1)'$. For $u^* \in B_{p-1}$, let $u = u(u^*)$ be the point on S^{p-1} above u^*, i.e. $u = (u^{*'}, \sqrt{1 - |u^*|^2})'$. For n large enough, u_n can be uniquely represented as $u_n = u(u_n^*)$ for some $u_n^* \in B_{p-1}$. Note that the north pole u_0 corresponds to $u(0)$. Next, define

$$W(., u^*, \delta) = C(u(u^*), t_0 + \delta) - C(u_0, t_0 + \delta)$$

which is to be understood as the difference of the indicator functions of the corresponding sets.

Note that u_n^* is a solution of the maximization $\sup_{u^*} F_n W(., u^*, t_n - t_0)$. To prove this proposition we shall use the main theorem of Kim and Pollard (1990). To do so, we need to establish two results. The first is that u_n also maximizes $F_n W(., u^*, 0)$, i.e.

$$F_n W(., u_n^*, 0) \geq \sup_{u^*} F_n W(., u^*, 0) - o_\mathrm{p}(n^{-2/3}).$$

To show this we need the following lemma.

Lemma 1 *Define the function* $M : B_{p-1} \times [-1, 1] \to [0, 1]$ *by* $M(u^*, t) = F(u, t)$ *and let*

$$\gamma(t) = -\frac{1}{p-1} Area(S^{p-2}) f'(t)(1 - t^2)^{(p-1)/2}.$$

Then we have the expansion

$$\begin{aligned} M(u^*, t_0 + \delta) = \alpha + \frac{\partial}{\partial t}|_{t_0} M(0, t)\ \delta + \frac{1}{2}\frac{\partial^2}{\partial t^2}|_{t_0} M(0, t)\ \delta^2 \\ + \frac{1}{2}\gamma(t_0)|u^*|^2 + o(\delta^2) + o(|u^*|^2). \end{aligned}$$

Proof: We have $M(u^*,t) = \int_{C(u_0,t)} f(u_0' T_u x) dS(x)$ where T_u is the rotation mapping u_0 to u. Note that $T_{u_0} = I_p$. Taking into account the fact that u_0 is the north-pole, that the p-th row of the matrix T_u is $(-u^{*\prime}, \sqrt{1-|u^*|^2})$, and using the measure decomposition $dS^{p-1} = (1-s^2)^{(p-3)/2} ds \otimes dS^{p-2}$ (see e.g. Watson, 1983), we obtain after some calculations that for each t and each $j,k \in \{1,...,p-1\}$,

$$\nabla|_{u^*=0}\ M(u^*,t) = 0 = \frac{\partial^2}{\partial u_j^* \partial u_k^*}|_{u^*=0} M(u^*,t)$$

and

$$\frac{\partial^2}{\partial u_j^{*2}}|_{u^*=0} M(u^*,t) = \gamma(t). \ \square$$

As a consequence of Lemma 1,

$$FW(.,u^*,\delta) = \frac{1}{2}\gamma(t_0)|u^*|^2 + o(\delta^2) + o(|u^*|^2).$$

Use this equation, the fact that $\gamma(t_0) < 0$, and stochastic equicontinuity of $n^{2/3}(F_n - F)$ (which follows Kim and Pollard) in order to obtain the inequality

$$F_n W(.,u_n^*,0) \geq \sup_{u^*} F_n W(.,u^*,0) - o_p(n^{-2/3}).$$

The second result that needs to be established is that the limiting covariance is, for each $x, y \in R^{p-1}$,

$$\Gamma(x,y) = \lim_{\beta \downarrow 0}\ \beta^{-1}\ FW(.,\beta x,0) W(.,\beta y,0).$$

Let $t \geq 0$ and, given two vectors $u, v \in S$, consider the set

$$J_t(u,v) = C(u,t) \setminus (C(u,t) \cap C(v,t)).$$

Write $A_t(\xi)$ for its area, where ξ denotes the angle between u and v. To obtain this result, we need the following lemma, which has interest in its own right.

Lemma 2 *The value of the right hand derivative of A_t at $\xi = 0$ is given by*

$$A_t'(0) = (p-2)^{-1}\ Area(S^{p-3})\ (1-t^2)^{\frac{p-2}{2}}.$$

Proof: We shall compute the right hand derivative of $B_t(\xi) = Area(C(u,t) \cap C(v,t))$. Note that $A_t'(0) = -B_t'(0)$. The cap $C(u,t)$ can be viewed as a $(p-1)$-dimensional spherical disc on S with radius $r = \arccos(t)$. Its boundary is a $(p-2)$-dimensional sphere on S with center $u \in S$ and (spherical) radius r. We shall denote it by $S^{p-2}(u,r)$ and we shall denote by $S^{p-2}(u,r,\phi)$ a cap on $S^{p-2}(u,r)$ with half-angle ϕ (there is no need for our purposes to specify its center on $S^{p-2}(u,r)$). We have

$$Area(S^{p-2}(u,r,\phi)) = \sin^{p-2}(r)\ Area(S^{p-3}) \int_0^{\phi} \sin^{p-3}(\theta)\ d\theta.$$

Now let S^+ be the hemisphere of S determined by $S^{p-2}(u,r) \cap S^{p-2}(v,r)$ which contains v. The Riemannian distance from u to S^+ is given by $a = \xi/2$. Denote by $L^{p-1}(a,r)$ the intersection of $C(u,t)$ and S^+. Then $B_t(\xi) = 2\ Area(L^{p-1}(a,r))$.

Now, $L^{p-1}(a,r)$ is composed of caps of the form $S^{p-2}(u,r,\phi(\rho))$, where $a \le \rho \le r$ and where $\phi(\rho)$ is given by the spherical trigonometry formula

$$\tan(\phi(\rho)) = \frac{\sqrt{\cos^2(a) - \cos^2(\rho)}}{\sin(a)\cos(\rho)}.$$

It follows that

$$B_t(\xi) = 2\ Area(S^{p-3}) \int_a^r \sin^{p-2}(\rho)(\int_0^{\phi(\rho)} \sin^{p-3}(\theta)\ d\theta)d\rho$$

and we obtain after some computations that

$$B_t'(0) = -\frac{1}{p-2} Area(S^{p-3}) \sin^{p-2}(r). \ \square$$

Now, for β small enough, denote by x_β, y_β the vectors on S^{p-1} which are above $\beta x, \beta y$. Then, from continuity of f and Lemma 2, we obtain

$$\begin{aligned}
&\lim_{\beta \downarrow 0} \beta^{-1}\ FW(.,\beta x, 0)W(.,\beta y, 0) \\
&= \lim_{\beta \downarrow 0} \beta^{-1}(F(J_{t_0}(x_\beta, u_0)) + F(J_{t_0}(u_0, y_\beta)) - F(J_{t_0}(x_\beta, y_\beta))) \\
&= f(t_0) A_{t_0}'(0) \lim_{\beta \downarrow 0} \beta^{-1}[\xi(x_\beta, u_0) + \xi(u_0, y_\beta) - \xi(x_\beta, y_\beta)] \\
&= \Gamma(x,y).\ \square
\end{aligned}$$

References

[1] Andrews, D.F, Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W. (1972). *Robust Estimation of Location: Survey and Advances.* Princeton University Press.

[2] Arcones, M.A., Chen, Z., and Gine, E. (1994). Estimators related to U-processes with applications to multivariate medians: Asymptotic normality. *Ann. Statist* **22**, 1460-1477.

[3] Brown, B.M. (1983). Statistical uses of the spatial median. *J.R. Statist. Soc.* B **45** 25-35.

[4] Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J.R. Statist. Soc* **91**, 862-872.

[5] Chernoff, H. (1964). Estimation of the mode. *Ann. Math. Statist* **16**, 31-41.

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

[7] Ducharme, G.R., Milasevic, P. (1987). Spatial median and directional data. *Biometrika* **74**, 212-215.

[8] Dümbgen, L. (1992) Limit theorems of the simplicial depth. *Statist. Prob. Lett.* **14**, 119-128.

[9] Eddy, W.F., and Hartigan, J.A. (1977). Uniform convergence of the empirical distribution function over convex sets. *Ann. Statist.* **5**, 370-374.

[10] Einmahl, J.H.J., and Mason, D.M. (1992). Generalized quantile processes. *Ann. Statist.* **20**, 1062-1078.

[11] Fisher, N.I., Lewis, T., Embleton, B.J.J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press.

[12] Grübel, R. (1988). The length of the shorth. *Ann. Statist.* **16**, 619-629.

[13] Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *J. Am. Statist. Soc.* **82**, 267-270.

[14] Kim, J., and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.

[15] Liu, R.Y. (1990). On a notion of data depth based upon random simplices. *Ann. Statist.* **18**, 405-414.

[16] Liu, R.Y., and Singh, K. (1992) Ordering directional data: Concepts of data depth on circles and spheres. *Ann. Statist.* **20**, 1468-1484.

[17] Müller, D.W. and Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *J. Am. Statist. Soc.* **86**, 738-746.

[18] Nolan, D. (1991). The excess-mass ellipsoid. *J. Multiv. Anal.* **39**, 348-371.

[19] Nolan, D. (1992). Asymptotics for multivariate trimming. *Stoch. Proc. Appl.* **42**, 157-169.

[20] Pollard, D. (1984). *Convergence of Stochastic Processes.* New York: Springer.

[21] Polonik, W. (1995a). Measuring mass concentrations and estimating

density contour clusters - An excess mass approach. *Ann. Statist.* **23**, 855-881.
[22] Polonik, W. (1995b). Density estimation under qualitative assumptions in higher dimensions. *J. Multiv. Anal.* **55**, 61-81.
[23] Rousseeuw, P.J., and Leroy, A.M. (1987). *Robust Regression and Outlier Detection.* New York: Wiley.
[24] Sager, T.W. (1979). An iterative method for estimating a multivariate mode and isopleth. *J. Am. Statist. Soc.* **74** 329-339.
[25] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.
[26] Small, C.G. (1990) A survey of multidimensional medians. *Int. Statist. Rev.* **58**, 263-277.
[27] Venter, J.H. (1967). On estimation of the mode. *Ann. Math. Statist.* **38**, 1446-1455.
[28] Watson, G.S.(1983). *Statistics on Spheres.* New York: Wiley.