Journal of Multivariate Analysis 101 (2010) 2222–2226

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Smooth depth contours characterize the underlying distribution

Linglong Kong *, Yijun Zuo

Department of Statistics and Probability, Michigan State University, East Lansing, MI, 48824, United States

ARTICLE INFO

Article history:
Received 6 December 2009
Available online 16 June 2010

AMS 2001 subject classifications:
primary 62H05
secondary 62H10
62H12

Keywords:
Halfspace depth
Depth contour
Characterization
Smooth contour

ABSTRACT

The Tukey depth is an innovative concept in multivariate data analysis. It can be utilized to extend the univariate order concept and advantages to a multivariate setting. While it is still an open question as to whether the depth contours uniquely determine the underlying distribution, some positive answers have been provided. We extend these results to distributions with smooth depth contours, with elliptically symmetric distributions as special cases. The key ingredient of our proofs is the well-known Cramér–Wold theorem.

1. Introduction

In one dimension, the univariate quantile function uniquely characterizes the underlying distribution [16] and has permeated into many applied statistical fields. In high dimensions, due to the lack of natural ordering, there is no universally preferred definition of multivariate quantiles. Among the various notions of multivariate quantile (see [5] for a review), that of Tukey's depth [18] may be the most successful representative. The Tukey depth of a point is the minimal proportion of data points whose removal makes it lie outside the convex hull of the remaining data points. Struyf and Rousseeuw [17] proved that the Tukey depth completely determines empirical distributions by actually reconstructing the data points from the depth contours. More generally, atomic distributions have been proved to be uniquely determined by the Tukey depth as well; see [7,2]. Koshevoy [8] also proved that the Tukey depth determines absolutely continuous distributions satisfying certain integrability conditions.

We study the properties of the Tukey depth contours and look into the probabilistic information of the underlying distribution carried by the contours and show that any distribution with smooth depth contours is completely determined by its Tukey depth. Proofs are based on the well-known Cramér–Wold theorem. As a special case, elliptically symmetric distributions are proved to have smooth depth contours and thus are determined by their Tukey depth. Note that it is not necessary that distributions with smooth depth contours are absolutely continuous; our results therefore are not covered by that of Koshevoy [8].

2. Definitions and notation

Tukey [18] and Donoho and Gasko [1] defined the *halfspace depth* of a point $x \in \mathbb{R}^d$ with respect to an empirical distribution P_n on $\mathbb{R}^d$ based on data $\{y_1, y_2, \ldots, y_n\}$ as the smallest proportion of data points in any closed halfspace with

* Corresponding author.
E-mail addresses: lkong@ualberta.ca (L. Kong), zuo@msu.edu (Y. Zuo).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.06.007

x on the boundary. The commonly used name "Tukey depth" reflects that Tukey [18] proposed using depth contours for plotting bivariate data, although Hodges [4] first introduced this. In detail, let u be a vector on unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$; then the Tukey depth of a point x can be written as

$$d(x, P_n) = \min_{u \in \mathbb{S}^{d-1}} \#\{i : u^T y_i \geq u^T x\}/n = \min_{u \in \mathbb{S}^{d-1}} \#\{i : (y_i - x) \in H_u\}/n,$$

where P_n is the empirical distribution based on data $\{y_1, y_2, \ldots, y_n\}$, $\#\{\cdot\}$ denotes the number of data points in $\{\cdot\}$ and $H_u = \{x : u^T x \geq 0\}$ is the closed halfspace containing 0 on its boundary with u pointing inside the halfspace and orthogonal to the boundary. Let P be a probability distribution on $\mathbb{R}^d$; then the Tukey depth of a point $x \in \mathbb{R}^d$ with respect to P is

$$d(x, P) = \inf_{u \in \mathbb{S}^{d-1}} P\{y : u^T y \geq u^T x\} = \inf_{u \in \mathbb{S}^{d-1}} P\{y : (y - x) \in H_u\},$$

where H_u is defined as the above. In what follows we will use $d(x)$ as an abbreviation of $d(x, P)$ or $d(x, P_n)$ if no confusion arises.

The Tukey depth is independent of the coordinate system; that is, it is affine invariant; see [19,20,11,17]. In other words, if we have two distributions P_1 and P_2 on $\mathbb{R}^d$ such that, for any data set $\mathbb{A} \subset \mathbb{R}^d$, $P_1(\mathbb{A}) = P_2(B\mathbb{A} + b)$, where $B\mathbb{A} + b = \{Bx + b : x \in \mathbb{A}\}$, B is a $d \times d$ non-singular matrix and b is a vector in $\mathbb{R}^d$, then $d(Bx + b, P_2) = d(x, P_1)$, for any $x \in \mathbb{R}^d$. The point(s) with the maximum Tukey depth provides a measure of centrality known as the *Tukey median*, denoted by μ_T; it is an alternative to the mean but more resistant to outliers; and it can be considered as a generalization of the median, considering that it is actually the median in one dimension. The minimum value of the depth of μ_T on $\mathbb{R}^d$ is $1/(d + 1)$; the breakdown point of μ_T is at most $1/(d + 1)$, and can be as high as $1/3$ for a centrally symmetric distribution; see Donoho and Gasko [1].

For $p \in (0, 1)$, the p-th Tukey depth contour $D(p)$ is the collection of x in $\mathbb{R}^d$ such that $d(x) \geq p$; that is,

$$D(p) = \{x : d(x) \geq p \text{ and } x \in \mathbb{R}^d\},$$

although a stricter usage might reserve this phrase for the boundary of D_p; see [3]. Immediately, we have that the depth contours form a sequence of nested convex sets: each $D(p)$ is convex and $D(p_1) \subset D(p_2)$ for any $p_2 \leq p_1$; see [21,1,10,12]. The innermost depth contour, which might be a singleton, is the Tukey median. The outmost depth contour for an empirical distribution is the convex hull of the data set. As the Tukey depth is a generalization of univariate quantiles, the depth contours are essentially multivariate quantile contours [5].

3. Retrieval theorem and characterization theorem

Denote the boundary of a bounded convex set E in $\mathbb{R}^d$ by ∂E. Let e be a point on ∂E. A *tangent hyperplane* of E at the point e is any hyperplane passing through the point e and having no intersection with the interior of E. Such a hyperplane determines the corresponding *tangent halfspace* at the point e, the halfspace that has the tangent hyperplane as its boundary and whose interior does not contain any point of E. In other words, a tangent halfspace of E at the point e can only have intersection with E on the tangent hyperplane of E at e. Obviously, the tangent hyperplane and hence the tangent halfspace at a point are not unique; for example, any vertex of a polygon has infinite tangent halfplanes and tangent halfspaces.

Let X be a random variable in $\mathbb{R}^d$ with distribution P. The *maximal mass* of P at a hyperplane, denoted by $\Delta(P)$, is defined as

$$\Delta(P) = \sup\left\{\mathbb{P}[s^T X = c] : s \in \mathbb{S}^{d-1}, c \in \mathbb{R}\right\}.$$

In the univariate case, $\Delta(P) = 0$ is equivalent to P being continuous, no point having positive probability mass. If P is an empirical distribution with finite sample size n, then $\Delta(P)$ is at least $1/n$. For more details, see [6,5]. The following theorem is essential in interpreting the depth contours when retrieving probabilistic information from them. In this sense, we call it the retrieval theorem.

Theorem 3.1. *Let P be the distribution of a random variable X in $\mathbb{R}^d$ and $0 < p < d(\mu_T)$. If H is a tangent halfspace of the p-th depth contour $D(p)$, then $p \leq P(H) \leq 2p + \Delta(P)$. Moreover, $p \leq P(H) \leq p + \Delta(P)$, if the boundary of H is the unique tangent hyperplane of $D(p)$ at some point on $D(p)$; in particular, $P(H) = p$ if $\Delta(P) = 0$.*

Theorem 3.1 provides a practical guideline for the recovery of the probabilistic information from the Tukey depth contours. When the tangent hyperplane to the p-th depth contour $D(p)$ is unique, we can uniquely identify the probability mass, which is exactly p, of the tangent halfspace whose direction is perpendicular to the tangent hyperplane pointing outside of $D(p)$ for distributions with $\Delta(P) = 0$. This is a long-standing open problem: does the depth function uniquely characterize the probability distribution? Positive answers have been established for partial cases: the Tukey depth uniquely characterizes empirical distributions [14], and more generally atomic distributions [7,2], and also absolutely continuous distributions satisfying certain integrability conditions [8]. The condition in [8] is that $\exp(v^T x)$ for any $v \in \mathbb{R}^d$ is integrable with respect to that distribution, which includes many usual distributions, for example, the multivariate normal distribution. A special case is absolutely continuous distributions with compact support. Certain progress in this line has been made by our following results regarding distributions with smooth depth contours. Under some mild conditions, if the depth contours are smooth, they can uniquely determine the underlying distribution. The conditions here are not covered either by Koshevoy [8] or us. Our conditions here are not covered by Koshevoy [8].

Before going to the characterization theorem, we need to recall and introduce some preliminary concepts. A boundary point of a convex set that admits more than one tangent hyperplane is called a *rough* (singular) point, also known as a corner point. It is known (see Theorem 2.2.4 of Schneider [15]) that such points are quite exceptional; in particular, for any closed convex set in $\mathbb{R}^d$, rough points are at most countable. Convex, closed subsets of $\mathbb{R}^d$ without rough points are called *smooth*, consistent with the natural geometric perception of the boundary in this case. Let P be a probability distribution in $\mathbb{R}^d$ of a random variable X. We say that P has *contiguous support* if there is no intersection of any two halfspaces with parallel boundaries that has nonempty interior but zero probability and divides the support of P into two parts.

Theorem 3.2. *If the p-th Tukey depth contour $D(p)$ of the probability distribution P of a random variable X in $\mathbb{R}^d$ with contiguous support and $\Delta(P) = 0$ is smooth for every $p \in (0, 1/2)$, then there is no other probability distribution with the same depth contours.*

Although the assumption of smoothness may sound optimistically mild, the examples in Rousseeuw and Ruts [14] showed that distributions with depth contours having a few rough points are not that uncommon. It may be argued that all these examples have a somewhat contrived flavor, especially when the support of the distribution is some regular geometric figure. It is not impossible that typical population distributions have smooth depth contours; however, we were not able to find a suitable formal condition reinforcing this belief, beyond the somewhat restricted realm of elliptically contoured distributions. Recall that the distribution P of a random variable X is called *elliptically symmetric* if it can be transformed by an affine transformation to a circularly symmetric distribution. If the elliptically symmetric distribution P has density, then the density function is in the form of

$$f(x) = c|\Sigma|^{-1/2} h\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right),$$

where h is a nonnegative scalar function, μ is the location parameter and Σ is a $d \times d$ positive definitive matrix, denoted by $X \sim E_d(h; \mu.\Sigma)$ following the notion of Liu and Singh [9]. Zuo and Serfling [21] proved that, if $X \sim E_d(h; \mu, \Sigma)$, then the p-th depth contour can be written as

$$D(p) = \{x : \ (x-\mu)^T \Sigma^{-1}(x-\mu) \le c_p^2\},$$

where c_p is a constant. Note that the results were formulated for more general depth functions. In the case of the Tukey depth, the constant c_p satisfies $\mathbb{P}[u^T \Sigma^{-1/2}(X-\mu) \le c_p] = p$ for any unit direction $u \in \mathbb{S}^{d-1}$; see [5] for details.

Corollary 3.1. *If $X \sim E_d(h; \mu, \Sigma)$, then, for any $p \in (0, 1/2)$, the p-th depth contour $D(p)$ is smooth; and thus the underlying distribution of X is uniquely determined by its depth contours.*

The proof of Corollary 3.1 is trivial in the view of the results of Zuo and Serfling [21] and is omitted here. As a consequence, we are able to determine the underlying elliptically symmetric distributions by their depth contours. In particular, this confirms the following fact: normal contours uniquely determine the underlying normal distribution, as a normal distribution is elliptically symmetric. For other distributions, this may or may not be true. However, we do have Theorem 3.2, if the depth contours are smooth.

Acknowledgments

This research partially originates in the doctoral dissertation of the first author defended at the University of Alberta. The authors are grateful to Ivan Mizera, the former advisor of the first author, for encouraging Linglong Kong. The authors greatly appreciate the constructive comments of the two referees and associate editor. The research of the second author was partially supported by NSF grants DMS-0234078 and DMS-0501174.

Appendix. Proofs of the theorems

A.1. Proof of Theorem 3.1

The inequality $P(H) \ge p$ follows from the fact that the Tukey depth of any point in $D(p)$ is at least p. As H is a tangent halfspace of $D(p)$, there exists a point, denoted by y, in H such that y belongs to $D(p)$ as well. More specifically, $y \in \partial H$, the boundary of H and $y \in \partial D(p)$, the boundary of $D(p)$. Note that such a point y may not be unique. The fact that $y \in D(p)$ implies that $d(y) \ge p$. Therefore, any halfspace containing y has at least probability mass p. As a result, $P(H) \ge p$.

Suppose that ∂H is the unique tangent hyperplane of $D(p)$ at some point $y \in \partial D(p)$. Let $Q(p, s) = Q(p, s, X) = \inf\{u : \mathbb{P}[s^T X \le u] \ge p\}$, where $\|s\| = 1$. Hereafter we always assume that s is a unit direction; i.e. $\|s\| = 1$. We claim that there is a direction s such that $s^T y = Q(p, s)$. Otherwise, there must exist an $\epsilon > 0$ such that $s^T y - Q(p, s) \ge \epsilon$ for any direction s as the situation $s^T y - Q(p, s) \le \epsilon$ for some direction s obviously does not exist. The ball centered at y with radius $\epsilon/2$ would then belong to $\{x : s^T x \ge Q(p, s)\}$ for any s and thus belong to $D(p)$ as well. The definition of $D(p)$, any point of $D(p)$ has depth at least p, implies that $D(p) = \cap_{\|s\|=1}\{x : s^T x \geq Q(p,s)\}$; for a detailed proof, see [5]. Therefore, y is an interior point of $D(p)$, which is a contradiction with the fact that $y \in \partial D(p)$.

Actually such a direction s satisfies $H = \{x : s^T x \leq Q(p,s)\}$ and hence is unique. As $y \in \partial H$, H can be rewritten as $\{x : t^T x \leq t^T y\}$ for some direction t. We need to show that $t = s$. The uniqueness of the tangent hyperplane at point y implies that any other hyperplane $\{x : u^T x = u^T y\}$ passing through y in a direction u, $u \neq t$, has common points with the interior of $D(p)$. In particular, if $t \neq s$, let $u = s$ and x be one of the common points. An ϵ-ball centered at such a point x is contained in $\{x : s^T x \leq s^T y\}$ and $D(p)$ when ϵ is small enough. The hyperplane $\{x : s^T x = Q(p,s)\}$ is parallel to the hyperplane $\{x : s^T x = s^T y\}$ with a perpendicular distance at least $2\epsilon > 0$, which is contradictory to the fact that $s^T y = Q(p,s)$. Therefore, $H = \{x : s^T x \leq Q(p,s)\}$. On the other hand, if X is a random variable with the distribution P, we have

$$P(H) = \mathbb{P}[s^T X \leq Q(p,s) - \epsilon_n] + \mathbb{P}[Q(p,s) - \epsilon_n \leq s^T X \leq Q(p,s)].$$

Note that $\mathbb{P}[s^T X \leq Q(p,s) - \epsilon_n] < p$ considering the definition of $Q(p,s)$, and as $\epsilon_n \to 0$, $\mathbb{P}[Q(p,s) - \epsilon_n \leq s^T X \leq Q(p,s)] \to \Delta(P)$. Let $\epsilon_n \to 0$; then $P(H) \leq p + \Delta(P)$.

It remains to prove the inequality $P(H) \leq 2p + \Delta(P)$ when the tangent halfspace H is not unique. For simplicity, only the bivariate case, where $d = 2$, is proved here. The argument for $d > 2$ is a little bit more difficult but quite similar. The key is to use projection to reduce the dimension of the distribution. Let $T_{D(p)}(y)$ be the directions of all tangent hyperplanes of the convex set $D(p)$ at the point y, where the direction of a tangent hyperplane is defined as a direction perpendicular to the hyperplane and pointed outside of $D(p)$. It is trivial that $T_{D(p)}(y)$ is an infinitely large pie slice, a convex cone in high dimension. If we only includes unit directions in $T_{D(p)}(y)$, then it is a closed arc. Hereafter, we assume that $T_{D(p)}(y)$ is the closed arc. Suppose u is one endpoint of $T_{D(p)}(y)$. Theorem 24.1 of [13] implies that there is a sequence $y_n \in \partial D_A(p)$ such that $y_n \to y$ and y_n has a unique tangent hyperplane with direction s_n such that $s_n \to u$. The convergence $s_n^T X \to u^T X$ and $s_n^T y_n \to u^T y$ make that $\liminf_{n\to\infty} \mathbb{P}[s_n^T X < s_n^T y_n] \geq \mathbb{P}[u^T X < u^T y]$. Using the fact proved in the first part, we obtain $s_n^T y_n = Q(p, s_n)$, because each y_n has a unique tangent hyperplane, which means that $\mathbb{P}[s_n^T X < s_n^T y_n] < p$. Thus $\mathbb{P}[u^T X < u^T y] \leq p$. Meanwhile, if v is the other endpoint of the arc $T_{D(p)}(y)$, then $\mathbb{P}[v^T X < v^T y] \leq p$ by similar derivation. Finally, $t \in T_{D(p)}(y)$ implies that $H \subseteq \{x : u^T x < u^T y\} \cap \{x : v^T x < v^T y\} \cap \{y\}$, and thus

$$P(H) \leq P(\{x : u^T x < u^T y\}) + P(\{x : v^T x < v^T y\}) + P(\{y\}) \leq 2p + \Delta(P).$$

A.2. Proof of Theorem 3.2

For simplicity, we assume that the origin is the point with maximal depth, the Tukey median. In a general case, the affine equivariance of depth makes it viable to transform the origin to be the Tukey median. We claim that, for any direction $s \in \mathbb{S}^{d-1}$ and constant $c \leq 0$,

$$P(\{x : s^T x \leq c\}) = \mathbb{P}[s^T X \leq c] = p^*,$$

where $p^* = \sup\{d(x) : s^T x = c\}$, the maximal depth on the hyperplane $\{x : s^T x = c\}$.

To show the claim, we will look into $p^* = 0$ and $p^* > 0$ respectively. If the maximum depth $p^* = 0$, then $d(x) = 0$ for any point $x \in \{x : s^T x = c\}$. Apparently, in this case $c \neq 0$ as $\mu_T = 0$. The convexity of the set with positive depth $\{x : d(x) > 0\}$ makes it that $d(x) = 0$ for any point belonging to the halfspace $\{x : s^T x \leq c\}$. Meanwhile, the set $\{x : d(x) > 0\}$ can be rewritten as the intersection of all halfspaces with positive probability mass and hence is a subset of any such halfspaces; see [5]. The halfspace $\{x : s^T x \leq c\}$ has no intersection with the set $\{x : d(x) > 0\}$, and thus has non-positive probability mass. That is, $P(\{x : s^T x \leq c\}) = p^*$. In the case of $p^* > 0$, to verify the claim it is sufficient to show that $\{x : s^T x = c\}$ is tangent to the p^*-th depth contour $D(p^*)$. In other words, the intersection $\{x : s^T x = c\} \cap D(p^*)$ is a singleton, only a point, as $D(p^*)$ is smooth. Otherwise, the intersection $\{x : s^T x = c\} \cap D(p^*)$ would be a compact convex set on the hyperplane $\{x : s^T x = c\}$ with nonempty interior; and so would the intersection $\{x : s^T x \leq c\} \cap D(p^*)$. Let y be one of the interior points on $\{x : s^T x = c\} \cap D(p^*)$. Then $d(y) = p^*$, that is, $P(\{x : t^T x \leq t^T y\}) = p^*$ for some direction $t \in \mathbb{S}^{d-1}$. Actually, for any $x \in \{x : s^T x \leq c\} \cap D(p^*)$, $d(x) = p^*$. As y is also an interior point of $D(p^*)$, there exists a point $z \in D(p^*) \cap \{x : t^T x \leq t^T y\}$ such that $t^T(z - y) < 0$. Thus $P(\{x : t^T x \leq t^T z\}) \geq p^*$, which implies that $P(\{x : s^T z < t^T x < t^T y\}) = 0$, a contradiction with the condition that P is contiguous. Therefore, $\{x : s^T x = c\} \cap D(p^*)$ is a singleton; and equivalently $\{x : s^T x = c\}$ is tangent to $D(p^*)$, denoting the tangent point by y. Obviously $d(y) = p^*$. From Theorem 3.1, we have $P(\{x : s^T x \leq c\}) = p^*$.

Our claim was proved, which means that, for any $s \in \mathbb{S}^{d-1}$, the distribution of $s^T X$ is uniquely determined by $D(p)$. The Cramér–Wold theorem allows the theorem to hold true.

References

[1] D.L. Donoho, M. Gasko, Breakdown properties of location estimates based on halfspace depth and projected outlyingness, Annals of Statistics 20 (1992) 1803–1827.
[2] A. Hassairi, O. Regaieh, On the Tukey depth of an atomic measure, Statistical Methodology 4 (2007) 244–249.
[3] X. He, G. Wang, Convergence of depth contours for multivariate datasets, Annals of Statistics 25 (1997) 495–504.
[4] J.L. Hodges Jr., A bivariate sign test, The Annals of Mathematical Statistics 26 (1955) 523–527.

[5] L. Kong, On multivariate quantile regression: directional approach and application with growth charts, Ph.D. Thesis, University of Alberta (Advisor: Ivan Mizera), Edmonton, Canada, 2009.
[6] L. Kong, I. Mizera, Quantile tomography: using quantiles with multivariate data, 2010 (submitted for publication) arXiv:0805.0056v1.
[7] G.A. Koshevoy, The Tukey depth characterizes the atomic measure, Journal of Multivariate Analysis 83 (2002) 360–364.
[8] G.A. Koshevoy, Lift-zonoid and multivariate depths, in: R. Dutter, P. Filzmoser, U. Gather, P.J. Rousseeuw (Eds.), Developments in Robust Statistics: International Conference on Robust Statistics 2001, Springer-Physica, Heidelberg, 2003, pp. 194–202.
[9] R.Y. Liu, K. Singh, A quality index based on data depth and multivariate rank tests, Journal of the American Statistical Association 88 (1993) 252–260.
[10] J.-C. Massé, R. Theodorescu, Halfplane trimming for bivariate distributions, Journal of Multivariate Analysis 48 (1994) 188–202.
[11] I. Mizera, C.H. Müller, Location-scale depth (with discussion), Journal of the American Statistical Association 99 (2004) 949–966.
[12] I. Mizera, M. Volauf, Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods, Journal of Multivariate Analysis 83 (2002) 365–388.
[13] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1996.
[14] P.J. Rousseeuw, I. Ruts, The depth function of a population distribution, Metrika 49 (1999) 213–244.
[15] R. Schneider, Convex Bodies: The Brunn–Minkowski Theory, Cambridge University Press, Cambridge, 1993.
[16] G.R. Shorack, Probability for Statisticians, Springer-Verlag, New York, 2000.
[17] A. Struyf, P.J. Rousseeuw, Halfspace depth and regression depth characterize the empirical distribution, Journal of Multivariate Analysis 69 (1999) 135–153.
[18] J.W. Tukey, Mathematics and the picturing of data, in: Proceedings of the International Congress of Mathematicians (Vancouver, BC, 1974), in: Canadian Mathematical Congress, vol. 2, Quebec, 1975, pp. 523–531.
[19] Y. Zuo, Contributions to the theory and applications of statistical depth functions, Ph.D. Thesis, University of Texas at Dallas (Advisor: Robert Serfling), Richardson, Texas, 1998.
[20] Y. Zuo, R. Serfling, General notions of statistical depth function, Annals of Statistics 28 (2000) 461–482.
[21] Y. Zuo, R. Serfling, Structural properties and convergence results for contours of sample statistical depth functions, Annals of Statistics 28 (2000) 483–499.