

Jensen's inequality for multivariate medians

MILAN MERKLE¹

Abstract. Given a probability measure μ on Borel sigma-field of \mathbb{R}^d , and a function $f : \mathbb{R}^d \mapsto \mathbb{R}$, the main issue of this work is to establish inequalities of the type $f(m) \leq M$, where m is a median (or a deepest point in the sense explained in the paper) of μ and M is a median (or an appropriate quantile) of the measure $\mu_f = \mu \circ f^{-1}$. For a most popular choice of halfspace depth, we prove that the Jensen's inequality holds for the class of quasi-convex and lower semi-continuous functions f .

To accomplish the task, we give a sequence of results regarding the "type D depth functions" according to classification in Y. Zuo and R. Serfling, Ann. Stat. **28** (2000), 461-482, and prove several structural properties of medians, deepest points and depth functions. We introduce a notion of a median with respect to a partial order in \mathbb{R}^d and we present a version of Jensen's inequality for such medians. Replacing means in classical Jensen's inequality with medians gives rise to applications in the framework of Pitman's estimation.

2000 Mathematics Subject Classification. 26B25, 62H11

Key words and phrases. Tukey's median, depth function, halfspace depth, partial order, convexity.

¹Faculty of Electrical Engineering, Bulevar Kralja Aleksandra 73, 11120 Belgrade, Serbia, emerkle@etf.rs.

Jensen's inequality for multivariate medians

MILAN MERKLE

1. INTRODUCTION

Let μ be a probability measure on Borel sets of \mathbb{R}^d , $d \geq 1$, and let f be a real valued convex function defined on \mathbb{R}^d . The Jensen's inequality states that

$$(1.1) \quad f(m) \leq M$$

where

$$m = \int_{\mathbb{R}^d} \mathbf{x} \, d\mu(\mathbf{x}) \quad \text{and} \quad M = \int_{\mathbb{R}^d} f(\mathbf{x}) \, d\mu(\mathbf{x}).$$

It is natural to ask if the means m and M can be replaced with some other kind of mean values. For $d = 1$, it is shown in [5] that a version of (1.1) holds for medians. Let us recall that a median of a probability distribution μ on \mathbb{R} is any real number m (need not be unique) such that

$$(1.2) \quad \mu((-\infty, m]) \geq \frac{1}{2}, \quad \mu([m, +\infty)) \geq \frac{1}{2}.$$

Given the measure μ and a measurable real valued function f , let μ_f be a measure defined by $\mu_f(B) = \mu(\{x \mid f(x) \in B\})$, and let M be its median. The result of [5] can be stated as follows.

Theorem 1.1. *Let μ be a probability measure on \mathbb{R} and let f be a quasi-convex lower semi-continuous function defined on \mathbb{R} . Then for every median M of μ_f there exists a median m of μ such that (1.1) holds, i.e.,*

$$(1.3) \quad \min\{f(\{\text{Med } \mu\})\} \leq \min\{\text{Med } \mu_f\}.$$

A final Internet search after a definitive draft of the present paper was written, revealed apparently forgotten paper [12] from the year 1975, with a similar result, a version of which can be stated as follows.

Theorem 1.2. *Let μ be a probability measure on \mathbb{R} and let f be a convex function defined on \mathbb{R} . Then for every median m of μ there exists a median M of μ_f such that (1.1) holds, i.e.,*

$$(1.4) \quad \max\{f(\{\text{Med } \mu\})\} \leq \max\{\text{Med } \mu_f\}.$$

In this paper we generalize both Theorem 1.1 and Theorem 1.2 to $d > 1$. To accomplish that task, we had to choose among many possible notions of multivariate medians. The Jensen's type inequality that we offer in this

paper holds for a very general case of *type D depth based median*, a term introduced by Zuo and Serfling in [15] (and conceptually initiated by C. G. Small in [10]). The depth function (to be strictly defined in Section 2) grasps a property of the one dimensional median that can be best understood from (1.2) for uniform distribution across a finite set of points (data set): the median is the *deepest point* of the data set, because to reach each median point m from either left or right you have to pass at least half of points.

This paper contains also a sequence of results related to the type D depth function and corresponding medians, which appear here as tools for our main purpose. We give several structural properties of medians and depth functions, that generalize or improve the results obtained in [8, 10, 15], and we also discuss a new class of depth functions based on partial orders in \mathbb{R}^d .

In the next section we give the precise definition of the notion of depth function and describe some of its properties, with a special attention to the halfspace depth. The Section 3 is devoted to depth functions based on a partial order, and the relation to the one dimensional case.

Let us note that there are several other concepts of the "center" or "median", which are different than those considered in this paper (see [11] for a survey) and it can be a direction of further research to make suitable versions of Jensen's inequality.

A comment on notations is now in order. A probability measure μ on Borel sigma field \mathcal{B}^d of \mathbb{R}^d can be always thought of being a probability distribution for some random vector $\mathbf{X} = (X_1, \dots, X_d)$ on an abstract probability space (Ω, \mathcal{F}, P) . That is, $\mu(B) = P(\mathbf{X} \in B)$ for any $B \in \mathcal{B}^d$. Notations $\text{Med } \mu$ and $\text{Med } \mathbf{X}$ will be used interchangeably to denote medians, and the set of all medians will be denoted by $\{\text{Med } \mu\}$ and $\{\text{Med } \mathbf{X}\}$ respectively.

2. DEPTH FUNCTIONS AND DEEPEST POINTS

Let \mathcal{U} be a specified collection of sets in \mathbb{R}^d , $d \geq 1$, and let μ be a probability measure on Borel sets of \mathbb{R}^d . For each $x \in \mathbb{R}^d$, define a depth function

$$(2.1) \quad D(x; \mu, \mathcal{U}) = \inf\{\mu(U) \mid x \in U \in \mathcal{U}\}.$$

The motivating example for this definition is the case $d = 1$, with \mathcal{U} being the set of intervals of the form $[a, +\infty)$ and $(-\infty, b]$; here (2.1) reduces to

$$D(x; \mu, \mathcal{U}) = \min\{\mu((-\infty, x]), \mu([x, +\infty))\}.$$

In this (one dimensional) case it is easy to see that the set of median points defined by (1.2) has the following three properties: (i) it is the set where the depth function reaches its maximum; (ii) it is a compact interval; (iii) it is the set of all points x with the property that $D(x; \mu, \mathcal{U}) \geq \frac{1}{2}$.

We will see that under mild regularity conditions imposed on the family \mathcal{U} , we can ensure that the set of deepest points (i.e, the set where the depth function reaches its global maximum) is compact in \mathbb{R}^d for $d > 1$, but only

in very special cases the property (iii) can be preserved. Although the set of deepest points is called "median set" by many authors, we will here use the term "center" (of a distribution or of a data set) for the set of deepest points in general, and reserve the term "median" only for the cases in which the deepest points have the depth $\frac{1}{2}$ or greater. This terminology is in accordance with the attitude expressed in [10] and [15].

The conditions that will be assumed are the following:

$$(C_1) \quad \text{for every } x \in \mathbb{R}^d \text{ there is a } U \in \mathcal{U} \text{ so that } x \in U.$$

$$(C'_2) \quad D(x; P, \mathcal{U}) > 0 \text{ for at least one } x \in \mathbb{R}^d \quad \text{and}$$

$$(C''_2) \quad \lim_{\|x\| \rightarrow +\infty} D(x; P, \mathcal{U}) = 0$$

If both conditions (C'_2) and (C''_2) are meant to be satisfied, we will refer to them as (C_2) . Note that (C_1) prevents the set on the right hand side of (2.1) of becoming empty (and hence of D becoming $-\infty$); whereas under the conditions (C_2) the depth function can not be constant in \mathbb{R} . The condition (C''_2) was also singled out in [15], as a requirement for any reasonable depth function.

Before proceeding further, let us see some simple examples. In some of these examples, and in results that we are going to prove, the family

$$(2.2) \quad \mathcal{V} = \{U^c \mid U \in \mathcal{U}\}$$

will play a special role. In the rest of the paper, the notation \mathcal{V} will be reserved for the collection of complements of the sets in \mathcal{U} , where \mathcal{U} is the family that determines the depth function via (2.1). Of course, it suffices to specify either \mathcal{U} or \mathcal{V} .

Example 2.1. 1° The simplest family \mathcal{U} that satisfies (C_1) contains only one set - the whole space \mathbb{R}^d . Here $D(x; P, \mathcal{U}) = 1$ for all x ; condition (C''_2) does not hold. The examples below satisfy (C_1) and both conditions in (C_2) .

2° The one dimensional depth function based on

$$\mathcal{U} = \{[a, +\infty) \mid a \in \mathbb{R}\} \cup \{(-\infty, b] \mid b \in \mathbb{R}\}$$

yields the classical median in dimension $d = 1$.

3° Let us take $V \in \mathcal{V}$ to be arbitrary convex and compact sets with a property that the collection \mathcal{V} is closed under translations, and that every ball in \mathbb{R}^d should be contained in some $V \in \mathcal{V}$. Because of compactness and the translation property, for each $x \in \mathbb{R}^d$ there exists a $V \in \mathcal{V}$ that does not contain x , and (C_1) follows. The conditions (C_2) follow from Lemma 4.1.

4° Consider now the class \mathcal{U} of all closed halfspaces. The corresponding depth function is known as Tukey's (or halfspace) depth (after Tukey's paper [13] of 1974). Note that the closed halfspaces with $d = 1$ are of the form $(-\infty, b]$ or $[a, +\infty)$; hence, this is a direct generalization of one dimensional

depth that leads to the usual one dimensional median. In fact, we may take \mathcal{U} to be the class of all open halfspaces (by Theorem 2.1 below). \square

In the next theorem we give a sufficient condition for equivalence of depth functions based on different sets \mathcal{U} .

Theorem 2.1. *Let \mathcal{A} and \mathcal{B} be families of subsets of \mathbb{R}^d . Suppose that the condition (C_1) holds for at least one of these families, and, in addition, the following condition (E) :*

(E') For each $A \in \mathcal{A}$, $A = \bigcup_{B \in \mathcal{B}, B \subset A} B$, and

(E'') For each $B \in \mathcal{B}$, there exists at most countable collection of sets $A_i \in \mathcal{A}$, such that $A_1 \supseteq A_2 \supseteq \dots$ and $B = \bigcap_i A_i$.

Then the condition (C_1) holds for both \mathcal{A} and \mathcal{B} and depth function with respect to both families are equal, with any probability distribution μ :

$$D(x; \mu, \mathcal{A}) = \inf\{\mu(A) \mid x \in A \in \mathcal{A}\} = \inf\{\mu(B) \mid x \in B \in \mathcal{B}\} = D(x; \mu, \mathcal{B})$$

An important application of Theorem 2.1 is to establish the equivalence of depth functions defined by a family of open sets $A \in \mathcal{A}$ and their topological closures $\bar{A} \in \mathcal{B}$. In this setup, we note that (E) holds whenever \mathcal{A} is any family of open halfspaces which is invariant with respect to translations. In particular, this implies that (i) one-dimensional median of μ can be defined as being any m such that

$$\mu((-\infty, b)) \geq \frac{1}{2}, \quad \text{and} \quad \mu(a, +\infty) \geq \frac{1}{2}.$$

for all intervals $(-\infty, b)$ and $(a, +\infty)$ that contain m , and (ii) that Tukey's depth can be defined via the family \mathcal{U} of all *open* halfspaces.

We are here interested chiefly in finding the set where the function D attains its global maximum, or, more generally, the sets of the form

$$(2.3) \quad S_\alpha = S_\alpha(\mu, \mathcal{U}) := \{x \in \mathbb{R}^d \mid D(x; \mu, \mathcal{U}) \geq \alpha\},$$

The next Lemma gives a way to find S_α without evaluation of the depth function.

Lemma 2.1. *Let \mathcal{U} be any collection of non-empty sets in \mathbb{R}^d , such that the condition (C_1) holds; let \mathcal{V} be the collection of complements of sets in \mathcal{U} . Then, for any probability measure μ ,*

$$(2.4) \quad S_\alpha(\mu, \mathcal{U}) = \bigcap_{V \in \mathcal{V}, \mu(V) > 1-\alpha} V,$$

for any $\alpha \in (0, 1]$ such that there exists a set $U \in \mathcal{U}$ with $\mu(U) < \alpha$; otherwise $S_\alpha(\mu, \mathcal{U}) = \mathbb{R}^d$.

If α_m is the maximum value of $D(x; \mu, \mathcal{U})$ for a given distribution μ , the set S_{α_m} , i.e., the set of deepest points with respect to μ , is called the center of the distribution μ , and will be denoted by $C(\mu, \mathcal{U})$.

In the next theorem, we discuss some properties of the center, in the case when sets in \mathcal{U} are open. A similar result for the family \mathcal{U} of closed sets was obtained in [15, Theorem 2.11], but under more restrictive assumptions.

Theorem 2.2. *Let \mathcal{V} be a collection of closed subsets of \mathbb{R}^d , and let \mathcal{U} be the collection of complements of sets in \mathcal{V} , such that the condition (C_1) holds. Then, for arbitrary probability measure μ , the function $x \mapsto D(x; \mu, \mathcal{U})$ is upper semi-continuous. In addition, under conditions (C_2) , the set $C(\mu, \mathcal{U})$ on which D reaches its maximum is equal to the minimal nonempty set S_α , that is,*

$$C(\mu, \mathcal{U}) = \bigcap_{\alpha: S_\alpha \neq \emptyset} S_\alpha(\mu, \mathcal{U}).$$

The set $C(\mu, \mathcal{U})$ is a non-empty compact set and it has the following representation:

$$(2.5) \quad C(\mu, \mathcal{U}) = \bigcap_{V \in \mathcal{V}, \mu(V) > 1 - \alpha_m} V, \quad \text{where } \alpha_m = \max_{x \in \mathbb{R}^d} D(x; \mu, \mathcal{U}).$$

It is instructive first to observe S_α in $d = 1$, as in the next example.

Example 2.2. Let \mathcal{V} be the family of all closed intervals in \mathbb{R} , and \mathcal{U} the family of their complements. Then for any probability measure μ , $S_\alpha = [q_\alpha, Q_{1-\alpha}]$, where q_α is the smallest quantile of μ of order α , and $Q_{1-\alpha}$ is the largest quantile of μ of order $1 - \alpha$:

$$(2.6) \quad \begin{aligned} q_\alpha &= \min\{t \in \mathbb{R} \mid \mu((-\infty, t]) \geq \alpha\} \quad \text{and} \\ Q_{1-\alpha} &= \max\{t \in \mathbb{R} \mid \mu([t, +\infty)) \geq \alpha\}. \end{aligned}$$

For $\alpha = \frac{1}{2}$, $[q_{\frac{1}{2}}, Q_{\frac{1}{2}}]$ is the median interval. □

Example 2.3. 1° Consider the halfspace depth, as in Example 4° of 2.1, in \mathbb{R}^2 , with the probability measure μ which assigns mass $1/3$ to points $A(0, 1)$, $B(-1, 0)$ and $C(1, 0)$ in the plane. Each point x in the closed triangle ABC has $D(x) = \frac{1}{3}$; points outside of the triangle have $D(x) = 0$. So, the function D reaches its maximum value $\frac{1}{3}$.

2° Let us now observe the same distribution, but with depth function defined with the family \mathcal{V} of closed disks. The intersection of *all* closed disks V with $\mu(V) > 2/3$ is, in fact, the intersection of all disks that contain all three points A, B, C , and that is the closed triangle ABC . For any $\varepsilon > 0$, a disc V with $\mu(V) > 2/3 - \varepsilon$ may contain only two of points A, B, C , but then it is easy to see that the family of all such discs has the empty intersection. Therefore, S_α is non-empty for $\alpha \leq 1/3$, and again, the function D attains its maximum value $1/3$ at the points of closed triangle ABC . We will see in Section 4 that, in fact, depth functions in cases 1° and 2° are equivalent regardless of the dimension. We will also see that $1/3$ is the maximal depth that can be generally expected in the two dimensional plane.

3° If \mathcal{V} is the family of rectangles with sides parallel to coordinate axes, then the maximum depth is $2/3$ and it is attained at $(0, 0)$. Families \mathcal{V} that are generalizations of intervals and rectangles are considered in the next section. We will show that the maximal depth with alike families is always at least $1/2$, regardless of dimension. \square

3. PARTIAL ORDERS, INTERVALS AND MULTIVARIATE MEDIAN SETS

We start with a characteristic property of univariate median set. Let μ be a probability distribution and let J be any closed interval with $\mu(J) > 1/2$. We will show that J contains every median of μ . Indeed, if m is a median of μ and $m \notin J$, then one of the intervals $(-\infty, m]$ or $[m, +\infty)$ is disjoint with J , which is not possible, since the sum of probabilities in both cases is greater than 1. Therefore, the intersection

$$\bigcap_{J=[a,b]: \mu(J) > 1/2} J$$

is nonempty, and it contains the median set $[u, v]$ of μ . Now, observe that for $J_{2n-1} = (-\infty, v + \frac{1}{(2n-1)})]$ and $J_{2n} = [u - \frac{1}{2n}, +\infty)$, $n = 1, 2, \dots$ we have that $\mu(J_n) > 1/2$ and so

$$\{\text{Med } \mu\} = [u, v] = \bigcap_{n=1}^{+\infty} J_n \supset \bigcap_{J=[a,b]: \mu(J) > 1/2} J,$$

which together with the previous part, shows that

$$(3.1) \quad \{\text{Med } \mu\} = \bigcap_{J=[a,b]: \mu(J) > 1/2} J$$

The relation (3.1) can be as well taken as a definition of the univariate median set for a given distribution, and this definition can be extended in a multidimensional environment if we choose one of many possible extensions of the concept of one-dimensional interval. Out of several ones that we may think of (convex sets, star-shaped sets, balls and other special convex sets), only intervals with respect to a partial order can do the work, to ascertain non-emptiness of the intersection at the right hand side of (3.1).

Let \preceq be a partial order in $\overline{\mathbb{R}^d}$ and let \mathbf{a}, \mathbf{b} be arbitrary points in $\overline{\mathbb{R}^d}$. We define a d -dimensional interval $[\mathbf{a}, \mathbf{b}]$ as the set of points in \mathbb{R}^d that are between \mathbf{a} and \mathbf{b} :

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^d \mid \mathbf{a} \preceq \mathbf{x} \preceq \mathbf{b}\}$$

Note that the interval can be an empty set, or a singleton. For the sake of simplicity, we want all intervals to be topologically closed. The interval can be norm bounded or norm unbounded; it would be reasonable to expect intervals with finite "endpoints" to be norm bounded, hence compact. Further, we would expect that intervals can be "big" as we wish, to contain

any ball or any compact set. Finally, we expect that bounded (with respect to partial order) sets has the least upper bound and greatest lower bound. To summarize, we assume the following three technical conditions:

- (I1) Any interval $[\mathbf{a}, \mathbf{b}]$ is topologically closed, and for any $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ (i.e., with finite coordinates), the interval $[\mathbf{a}, \mathbf{b}]$ is a compact set.
- (I2) For any ball $B \subset \mathbb{R}^d$, there exist $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$ such that $B \subset [\mathbf{a}, \mathbf{b}]$.
- (I3) For any set S which is bounded from above with a finite point, there exists a finite $\sup S$. For any set S which is bounded from below with a finite point, there exists a finite $\inf S$.

Example 3.1. Let K be a closed convex cone in \mathbb{R}^d , with vertex at origin, and suppose that there exists a closed hyperplane π , such that $\pi \cap K = \{0\}$ (that is, $K \setminus \{0\}$ is a subset of one of open halfspaces determined by π). Define the relation \preceq by $\mathbf{x} \preceq \mathbf{y} \iff \mathbf{y} - \mathbf{x} \in K$. The interval is then

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \mid \mathbf{x} - \mathbf{a} \in K \wedge \mathbf{b} - \mathbf{x} \in K\} = (\mathbf{a} + K) \cap (\mathbf{b} - K).$$

If the endpoints have some coordinates infinite, then the interval is either $\mathbf{a} + K$ (if $\mathbf{b} \notin \mathbb{R}^d$) or $\mathbf{b} - K$ (if $\mathbf{a} \notin \mathbb{R}^d$) or \mathbb{R}^d (if neither endpoint is in \mathbb{R}^d).

It is not difficult to show that \preceq is a partial directed order, and that it satisfies conditions (I1)–(I3). Note that the intervals defined with convex cone partial order are convex sets.

The simplest, coordinate-wise ordering, can be obtained with K chosen to be the orthant with $x_i \geq 0, i = 1, \dots, d$. Then

$$(3.2) \quad \mathbf{x} \preceq \mathbf{y} \iff x_i \leq y_i, \quad i = 1, \dots, d.$$

For the sake of illustration, let us note that possible kinds of intervals with respect to the relation (3.2) in \mathbb{R}^2 include:

$$\begin{aligned} & [(a_1, a_2), (b_1, b_2)], [(a_1, a_2), (b_1, +\infty)], [(a_1, a_2), (+\infty, b_2)], \\ & [(a_1, a_2), (+\infty, +\infty)], [(-\infty, -\infty), (b_1, b_2)], [(a_1, -\infty), (+\infty, b_2)], \end{aligned}$$

where a_1, a_2, b_1, b_2 are real numbers. For infinite endpoints we use strict inequalities, for example the last interval above is the set of $(x, y) \in \mathbb{R}^2$ such that $a_1 \leq x < +\infty$ and $-\infty < y \leq b_2$. The intervals may be empty; for example, the first listed interval is empty if $a_1 > b_1$ or if $a_2 > b_2$. □

The next theorem fully extends the one-dimensional property discussed in the beginning of this section.

Theorem 3.1. *Let \preceq be a partial order in $\overline{\mathbb{R}^d}$ such that conditions (I1)–(I3) hold. Let μ be a probability measure on \mathbb{R}^d and let \mathcal{J} be a family of intervals with respect to a partial order \preceq , with the property that*

$$(3.3) \quad \mu(J) > \frac{1}{2}, \quad \text{for each } J \in \mathcal{J}.$$

Then the intersection of all intervals from \mathcal{J} is a non-empty compact interval.

The compact interval claimed in the Theorem 3.1 can be, in analogy to (3.1), taken as a definition of the median induced by the partial order \preceq :

$$(3.4) \quad \{\text{Med } \mu\}_{\preceq} := \bigcap_{J=[\mathbf{a}, \mathbf{b}]: \mu(J) > 1/2} J,$$

In what follows, we will omit the subscript if the underlying relation \preceq is obvious.

In the case of coordinate-wise partial order, the median set is the Cartesian product of coordinate-wise median sets [10, Example 2.3.2].

Let \mathcal{V} be the family of all closed intervals with respect to some partial order \preceq that satisfies conditions (I1)–(I3) and let \mathcal{U} be the family of their complements. Assuming that the condition (C_1) holds, we find, via Lemma 2.1, that the level sets S_α with respect to the depth function $D(x; \mu, \mathcal{U})$ can be expressed as in (2.4). Hence, under the condition (C_1) , $D(x; \mu, \mathcal{U}) \geq 1/2$ for all $x \in \{\text{Med } \mu\}_{\preceq}$.

4. CONVEX SETS AND HALFSACES.

It is natural to have a convex center of distribution, which is achieved (via Theorem 2.2) if sets in \mathcal{V} are convex. For more arguments in favor of convex sets see [1].

A prototype of depth functions that we discuss in this section is a depth function defined with respect to families \mathcal{U} of complements of compact convex sets. These requirements are natural and they are not too restrictive (see also Lemma 4.1). Although it may look that by these requirements we are excluding the halfspace depth from consideration, it is not so, as we will see after Theorem 4.2.

From the material of Section 3, it follows that the depth function based on a family \mathcal{V} of intervals, attains the maximal value of at least $1/2$, regardless of the dimension d . In general, the maximum depth with a family \mathcal{V} of convex sets, can not be smaller than $\frac{1}{d+1}$. This conclusion follows from the next theorem, which is an extension of results in [2] and [8].

Theorem 4.1. *Let μ be any probability measure on Borel sets of \mathbb{R}^d . Let \mathcal{V} be any family of closed convex sets in \mathbb{R}^d , and let \mathcal{U} be the family of their complements. Assume that conditions (C_1) and (C_2'') hold. Then the condition (C_2') also holds, and there exists a point $x \in \mathbb{R}^d$ with $D(x; \mu, \mathcal{U}) \geq \frac{1}{d+1}$.*

The lower bound for D in Theorem 4.1 is the greatest generally possible. As the next example shows, for the halfspace depth, in any dimension $d \geq 1$, there exist a probability measure μ such that $D(x; \mu, \mathcal{U}) \leq \frac{1}{d+1}$ for all $x \in \mathbb{R}^d$.

Example 4.1. This is an extension of the example 2.3. Let A_1, \dots, A_{d+1} be points in \mathbb{R}^d such that they do not belong to the same hyperplane (i.e. to any affine subspace of dimension less than d), and suppose that $\mu(\{A_i\}) = \frac{1}{d+1}$ for each $i = 1, 2, \dots, d+1$. Let S be a closed d -dimensional simplex with

vertices at A_1, \dots, A_{d+1} , and let $x \in S$. If x is a vertex of S , then there exists a closed halfspace H such that $x \in H$ and other vertices do not belong to H ; then $D(x) = \mu(H) = 1/(d+1)$. Otherwise, let S_x be a d -dimensional simplex with vertices in x and d points among A_1, \dots, A_{d+1} that make together an affinely independent set. Then for S_x and the remaining vertex, say A_1 , there exists a separating hyperplane π such that $\pi \cap S_x = \{x\}$ and $A_1 \notin \pi$ (see [7, Section 11]). Let H be a halfspace with boundary π , that contains A_1 . Then also $D(x) = \mu(H) = 1/(d+1)$. So, all points $x \in S$ have $D(x) = 1/(d+1)$. Points x outside of S have $D(x) = 0$, which is easy to see. So, the maximal depth in this example is exactly $1/(d+1)$. \square

In fact, if we have a family of compact convex sets \mathcal{V} that contain arbitrary large sets (in the sense of the following lemma), then it is sufficient to assume only condition (C_1) , and then (C_2) will automatically hold.

Lemma 4.1. *Let \mathcal{V} be a family of compact convex sets in \mathbb{R}^d , and let \mathcal{U} be the family of complements of sets in \mathcal{V} , such that the condition (C_1) holds. Suppose that for every closed ball $B \in \mathbb{R}^d$ there exist a set $V \in \mathcal{V}$, such that $B \subset V$. Then the family \mathcal{U} and the depth function $D(\cdot; \mu, \mathcal{U})$ satisfy conditions (C'_2) and (C''_2) , with any probability measure μ on \mathbb{R}^d .*

In the next theorem, we use the fact that every closed convex set can be represented as an intersection of closed halfspaces (see, for example, [7, Theorem 11.5]). This representation is not unique (and we do not need uniqueness neither in the statement nor in the proof); however, there is a unique minimal representation of a convex set as the intersection of all its tangent halfspaces [7, Theorem 18.8], which is an intuitive model for the representation (4.1) below.

Theorem 4.2. *Let \mathcal{V} be a collection of closed convex sets and \mathcal{U} the collection of complements of all sets in \mathcal{V} . For each $V \in \mathcal{V}$, consider a representation*

$$(4.1) \quad V = \bigcap_{\alpha \in A_V} H_\alpha,$$

where H_α are closed subspaces and A_V is an index set. Let

$$\mathcal{H}^V = \{\overline{H_\alpha^c} + x \mid \alpha \in A_V, x \in \mathbb{R}^d\}$$

be the collection of closures of complements of halfspaces H_α and their translations. Further, let

$$\mathcal{H} = \bigcup_{V \in \mathcal{V}} \mathcal{H}^V.$$

If for any $H \in \mathcal{H}$ there exists at most countable collection of sets $V_i \in \mathcal{V}$, such that

$$(4.2) \quad V_1 \subseteq V_2 \subseteq \dots \quad \text{and} \quad \overset{\circ}{H} = \bigcup V_i,$$

then

$$D(x; \mu, \mathcal{U}) = D(x; \mu, \mathcal{H}) = D(x; \mu, \overset{\circ}{\mathcal{H}}), \quad \text{for every } x \in \mathbb{R}^d,$$

where $\overset{\circ}{\mathcal{H}}$ is the family of open halfspaces from \mathcal{H} .

As a corollary to Theorem 4.2, we can single out two important particular cases. Conditions (4.1) and (4.2) in both cases can be easily proved.

Corollary 4.1. a) *Let \mathcal{V} be the family of closed intervals with respect to the partial order defined with a convex cone K , as in the Section 3. Then for any probability distribution and any $x \in \mathbb{R}^d$,*

$$D(x; \mu, \mathcal{U}) = D(x; \mu, \mathcal{H}),$$

where \mathcal{U} is the family of complements of sets in \mathcal{V} and \mathcal{H} is the family of all tangent halfspaces to K , and their translations.

In particular, if \mathcal{V} is the family of intervals with respect to the coordinate-wise partial order, then the corresponding depth function is the same as the depth function generated by halfspaces with borders parallel to the coordinate hyperplanes.

b) *Let \mathcal{H} be the family of all closed halfspaces, and let $\mathcal{U}_c, \mathcal{U}_k$ and \mathcal{U}_b be families of complements of all closed convex sets, compact convex sets and closed balls, respectively. Then*

$$D(x; \mu, \mathcal{H}) = D(x; \mu, \mathcal{U}_c) = D(x; \mu, \mathcal{U}_k) = D(x; \mu, \mathcal{U}_b).$$

The second part of Corollary 4.1 implies, via Lemma 2.1, that for the halfspace depth function D , we have

$$(4.3) \quad S_\alpha = \{x \mid D(x) \geq \alpha\} = \bigcap_{C: \mu(C) > 1-\alpha} B = \bigcap_{K: \mu(K) > 1-\alpha} K = \bigcap_{B: \mu(B) > 1-\alpha} B,$$

where C are convex sets, K are compact convex sets, and B are closed balls. Hence, the center of distribution with respect to halfspace depth (commonly referred to as "Tukey's median") can be found as the intersection of compact convex sets, or balls.

5. JENSEN'S INEQUALITY FOR MEDIANS, DEEPEST POINTS AND LEVEL SETS

Given a family \mathcal{V} of closed subsets of \mathbb{R}^d , a probability measure μ and the depth function $D(x; \mu, \mathcal{U})$, where \mathcal{U} is the family of complements of the sets in \mathcal{V} , we are seeking upper bounds for $f(m)$, where f is a function, and m is a median of μ , or a point where D reaches its maximum, or, most generally, a point in the level set $S_\alpha(\mu, \mathcal{U})$ for any α such that $S_\alpha \neq \emptyset$. The upper bounds are expressed as medians or appropriate quantiles of μ_f . A class of functions for which these results hold is defined bellow.

Definition 5.1. A function $f : \mathbb{R}^d \mapsto \mathbb{R}$ will be called a *C-function* with respect to a given family \mathcal{V} of closed subsets of \mathbb{R}^d , if for every $t \in \mathbb{R}$, $f^{-1}((-\infty, t]) \in \mathcal{V}$ or is empty set. .

Example 5.1. 1° If \mathcal{V} is the family of all closed convex sets in \mathbb{R}^d , then the class of corresponding C-functions is precisely the class of lower semi-continuous quasi-convex functions, i.e., functions f that have the property that $f^{-1}((-\infty, t])$ is a closed set for any $t \in \mathbb{R}$ and

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}, \quad \lambda \in [0, 1], \quad x, y \in \mathbb{R}^d.$$

This is easy to see, starting from the definition 5.1. In particular, every convex function on \mathbb{R}^d is a C-function with respect to the class of all convex sets.

2° In general, since we require sets in \mathcal{V} to be closed, every C-function is lower semi-continuous, and if these sets are convex, then a C-function is quasi-convex, so for a common choice of sets in \mathcal{V} to be closed convex sets, the family of C-functions is a subset of the class of lower semi-continuous and quasi-convex functions.

3° A function f is a C-functions with respect to a family of closed intervals (with respect to some partial order in \mathbb{R}^d), if and only if

$$\{\mathbf{x} \in \mathbb{R}^d \mid f(\mathbf{x}) \leq t\} = [\mathbf{a}, \mathbf{b}], \quad \text{for some } \mathbf{a}, \mathbf{b} \in \overline{\mathbb{R}^d}.$$

It is not clear if this condition can be replaced with some other, easier to check, as it was done in the case 1°.

4° By Theorem 2.2, under condition (C_1) the depth function $D(x; \mu, \mathcal{U})$ is upper semi-continuous with respect to x . Therefore, the function $x \mapsto 1 - D(x; \mu, \mathcal{U})$ is lower semi-continuous, and to be a C-function, it suffices, via (2.1), that any non-empty intersection of sets in \mathcal{V} is also in \mathcal{V} . Hence, for the halfspace depth, and for the depth based on (complements of) closed intervals that satisfy (I1)–(I3), the function $x \mapsto 1 - D(x; \mu, \mathcal{U})$ is a C-function.

5° In \mathbb{R}^2 , with coordinate-wise intervals, the function f defined by

$$f(x, y) = \max\{|x - a_1| - |x - b_1|, |y - a_2| - |y - b_2|\}$$

is a C-function, where $\mathbf{a}(a_1, a_2)$ and $\mathbf{b}(b_1, b_2)$ are given points in \mathbb{R}^2 . □

We start with the most general theorem about the level sets.

Theorem 5.1. *Let \mathcal{V} be a family of closed subsets of \mathbb{R}^d , and let \mathcal{U} be the family of their complements. Assume that conditions (C_1) and (C_2) hold with a given probability measure μ . Let $\alpha > 0$ be such that the level set $S_\alpha = S_\alpha(\mu, \mathcal{U})$ as defined by (2.3) is nonempty, and let f be a C-function with respect to \mathcal{V} .*

Then for every $m \in S_\alpha$ we have that

$$(5.1) \quad f(m) \leq Q_{1-\alpha},$$

where $Q_{1-\alpha}$ is the largest quantile of order $1 - \alpha$ for μ_f . □

If α_m is the maximum value of the depth function, then S_{α_m} becomes the center of the distribution, $C(\mu, \mathcal{U})$, and one can rephrase the above theorem for m being a deepest point of the distribution μ . For the most popular choice of halfspace depth, from materials of Section 4 and Example 5.1, 1° , it follows that the relevant class of C -functions is the class of quasi-convex and lower semi-continuous functions on \mathbb{R}^d . We separate this case in the following corollary.

Corollary 5.1. (*Jensen's inequality for "Tukey's median"*). *Let f be a lower semi-continuous and quasi-convex function on \mathbb{R}^d , and let μ be an arbitrary probability measure on Borel sets of \mathbb{R}^d . Suppose that the depth function with respect to halfspaces reaches its maximum α_m on the set $C(\mu)$ ("Tukey's median set"). Then for every $m \in C(\mu)$,*

$$(5.2) \quad f(m) \leq Q_{1-\alpha_m},$$

where $Q_{1-\alpha_m}$ is the largest quantile of order $1 - \alpha_m$ for μ_f .

To show that the upper bound in (5.2) can not be globally improved, consider the following example:

Example 5.2. Let A, B, C be non-collinear points in the two dimensional plane, and let \mathcal{H} be the collection of open halfplanes. Let $l(AB)$ be the line determined by A and B . Let H_1 be the closed halfspace that does not contain the interior of the triangle ABC and has $l(AB)$ for its boundary, and let H_2 be its complement. Define a function f by

$$f(x) = e^{-d(x, l(AB))} \quad \text{if } x \in H_1, \quad f(x) = e^{d(x, l(AB))} \quad \text{if } x \in H_2,$$

where $d(\cdot, \cdot)$ is euclidean distance. Then $f(A) = 1$, $f(B) = 1$ and $f(C) > 1$, and f is a convex function. Now suppose that μ assigns mass $1/3$ to each of the points A, B, C . Then, by example 2.3, we know that the center $C(\mu, \mathcal{H})$ of this distribution is the set of points of the triangle ABC , with $\alpha_m = 1/3$. Hence, for $m \in C(\mu, \mathcal{H})$, $f(m)$ takes all values in $[1, f(C)]$. On the other hand, quantiles for μ_f of the order $2/3$ are points in the closed interval $[1, f(C)]$; hence the most we can state is that $f(m) \leq f(C)$, with $f(C)$ being the largest quantile of order $2/3$. \square

If the depth function is based on (complements of) intervals, we know, after Section 3, that there exist a "true" median set, i.e., that the depth function reaches values $\geq 1/2$. For this case, we have a direct generalization of Theorems 1.1 and 1.2 from Introduction.

Theorem 5.2. *Let \mathcal{V} be a family of closed intervals with respect to a partial order in \mathbb{R}^d , such that conditions (I1)–(I3) are satisfied. Let $\{\text{Med } \mu\}$ be the median set of a probability measure μ with respect to the chosen partial order, and let f be a C -function with respect to the family \mathcal{V} . Then for every $M \in \text{Med } \{\mu_f\}$, there exists an $m \in \{\text{Med } \mu\}$, such that*

$$(5.3) \quad f(m) \leq M$$

Further, for every $m \in \{\text{Med } \mu\}$,

$$(5.4) \quad f(m) \leq \max\{\text{Med } \mu_f\}.$$

□

Since every C -function is lower semi-continuous, and $\{\text{Med } \mu\}$ is compact set, the set $f(\{\text{Med } \mu\})$ has its minimum. Hence, the equivalent form of the first part of Theorem 5.2 is

$$(5.5) \quad \min f(\{\text{Med } \mu\}) \leq \min\{\text{Med } \mu_f\},$$

and the second part is clearly equivalent to

$$(5.6) \quad \sup f(\{\text{Med } \mu\}) \leq \max\{\text{Med } \mu_f\}.$$

In next two examples we present some applications.

Example 5.3. For a d -dimensional random variable X with expectation $\text{E}X$ and $\text{Med } X = \text{E}X$, we may use both classical Jensen's inequality $f(\text{E}X) \leq \text{E}f(X)$ or one of inequalities derived above, provided that f is a convex C -function and that $\text{E}f(X)$ exists. It can happen that the upper bound in terms of medians or quantiles is lower than $\text{E}f(X)$. To illustrate the point, consider univariate case, with $X \sim \mathcal{N}(0, 1)$ and $f(x) = (x - 2)^2$. Then the classical Jensen's inequality with means gives $4 \leq 5$. Since here $\text{Med } (X - 2)^2 = 4.00032$ (numerically evaluated), the inequality $f(\text{E}X) \leq \text{Med } f(X)$ is sharper. Of course, if $\text{E}f(X)$ does not exist, the median alternative is the only choice.

Example 5.4. Let a and b are points in \mathbb{R}^d , and let $\|\cdot\|$ be usual euclidean norm. Since the function $x \mapsto \|x - a\|^2 - \|x - b\|^2$ is affine, it is a C -function for the halfspace depth. Let m be a point in the center of a distribution μ , and let α_m be the value of the depth function in the center. Let X be a d -dimensional random variable on some probability space (Ω, \mathcal{F}, P) with the distribution μ . Consider the function $f(x) = \|x - a\|^2 - \|x - m\|^2$. Then by Corollary 5.1 we have that $0 \leq \|m - a\|^2 \leq Q_{1-\alpha_m}$, which implies that $P(f(X) \geq 0) \geq \alpha_m$, or, equivalently,

$$(5.7) \quad P(\|X - m\| \leq \|X - a\|) \geq \alpha_m \quad \text{for any } a \in \mathbb{R}^d.$$

The expression on the left hand side of (5.7) is known as Pitman's measure of nearness; in this case it measures the probability that X is closer to m than to any other chosen point a . For distributions with $\alpha_m = \frac{1}{2}$, (5.7) means that each point in "Tukey's median set" is a best non-random estimate of X (or, a most representative value) in the sense of Pitman's criterion, with the euclidean distance as a loss function. The analogous result in one dimensional case is well known. For Pitman's nearness see, for example [3], [4], or [6] and references therein.

6. PROOFS AND AUXILIARY RESULTS

PROOF OF THEOREM 2.1. Suppose that the stated conditions hold. If (C_1) holds for \mathcal{A} , then (E') implies that it holds for \mathcal{B} . If (C_1) holds for \mathcal{B} , then it clearly holds for \mathcal{A} by (E'') .

Let $x \in \mathbb{R}^d$ be fixed. Then by (E') , for each $A \in \mathcal{A}$ that contains x , there exists a $B_A \in \mathcal{B}$ such that $x \in B_A \subset A$, and, consequently, $\mu(A) \geq \mu(B_A)$. Therefore,

$$\begin{aligned} D(x; \mu, \mathcal{A}) &\geq \inf\{\mu(B_A) \mid x \in B_A \in \mathcal{B}, A \in \mathcal{A}\} \\ &\geq \inf\{\mu(B) \mid x \in B \in \mathcal{B}\} = D(x; \mu, \mathcal{B}) \end{aligned}$$

as the class of all B_A is a subset of the class of all $B \in \mathcal{B}$ that may contain x . On the other hand, by (E'') , for each $\varepsilon > 0$ and for each $B \in \mathcal{B}$ that contains x , there exists $A_B \in \mathcal{A}$, such that $\mu(B) \geq \mu(A_B) - \varepsilon$. Then

$$\begin{aligned} \inf\{\mu(B) \mid x \in B \in \mathcal{B}\} &\geq \inf\{\mu(A_B) \mid x \in A_B, A_B \in \mathcal{A}, B \in \mathcal{B}\} - \varepsilon \\ &\geq \inf\{\mu(A) \mid x \in A \in \mathcal{A}\} - \varepsilon = D(x; \mu, \mathcal{A}) - \varepsilon, \end{aligned}$$

and since $\varepsilon > 0$ is arbitrary, we conclude that

$$D(x; \mu, \mathcal{B}) = \inf\{\mu(B) \mid x \in B \in \mathcal{B}\} \geq D(x; \mu, \mathcal{A}),$$

which ends the proof.

PROOF OF LEMMA 2.1. Evidently, $x \in S_\alpha^c$ if and only if $D(x) < \alpha$, i.e., if and only if there exists a set $U \in \mathcal{U}$ such that $x \in U$ and $\mu(U) < \alpha$. Therefore, if there are $U \in \mathcal{U}$ with $\mu(U) < \alpha$, then

$$S_\alpha^c = \bigcup_{U \in \mathcal{U}, \mu(U) < \alpha} U, \quad \text{and so,} \quad S_\alpha = \bigcap_{U \in \mathcal{U}, \mu(U) < \alpha} U^c,$$

which is equivalent to the assertion that we wanted to prove.

PROOF OF THEOREM 2.2. Under (C_1) and if all sets in \mathcal{V} are closed, the set S_α is closed for every α , via (2.4), and hence, the function D is upper semi-continuous. Under additional conditions (C_2) , we will show that there exists at least one α such that S_α is a nonempty compact set. Indeed, by the assumption, there is $x \in \mathbb{R}^d$ so that $D(x) = \alpha_0 > 0$. On the other hand, by assumption of convergence of $D(x)$ to zero as $\|x\| \rightarrow +\infty$, there exists an $R > 0$ so that $D(x) < \alpha_0$ for $\|x\| > R$. Therefore, the set S_{α_0} is nonempty and norm bounded, and being closed, it is compact. Then all sets S_α with $\alpha \geq \alpha_0$ are compact, because $S_\alpha \subset S_{\alpha_0}$ for $\alpha \geq \alpha_0$. The intersection of non-empty compact nested sets S_α is a non-empty compact set, and it is clearly the set on which D reaches its maximum. The representation (2.5) follows from Lemma 2.1. \square

In order to prove Theorem 3.1, we need the following lemma.

Lemma 6.1. *Let \preceq be a partial order in \mathbb{R}^d such that the conditions (I1) and (I3) hold. Let*

$$\mathcal{J} = \{J_\alpha \mid J_\alpha = [\mathbf{a}^\alpha, \mathbf{b}^\alpha], \quad \alpha \in A\}$$

be a collection of closed intervals, where A is an index set. Assume that there is at least one α such that $\mathbf{a}^\alpha \in \mathbb{R}^d$ (i.e., have all coordinates finite) and at least one β such that $\mathbf{b}^\beta \in \mathbb{R}^d$. Suppose that $J_\alpha \cap J_\beta \neq \emptyset$ for all α, β . Then

- (i) $\mathbf{a}^\alpha \preceq \mathbf{b}^\beta$, for any $\alpha, \beta \in A$;
- (ii) The intersection of all sets in \mathcal{J} is a non-empty compact interval $[\mathbf{a}, \mathbf{b}]$, with $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$.

Proof. If intervals $[\mathbf{a}, \mathbf{b}]$ and $[\mathbf{c}, \mathbf{d}]$ have a common point \mathbf{x} , then $\mathbf{a} \preceq \mathbf{x} \preceq \mathbf{b}$ and $\mathbf{c} \preceq \mathbf{x} \preceq \mathbf{d}$; hence $\mathbf{a} \preceq \mathbf{d}$ and $\mathbf{c} \preceq \mathbf{b}$. This shows (i). Further, to show (ii), note that by assumptions and (i), the set $\{\mathbf{a}^\alpha, \alpha \in A\}$ is bounded from above with a finite point, and so by (I3), there exists $\mathbf{a} = \sup_{\alpha \in A} \mathbf{a}^\alpha$. In an analogous way we conclude that there exists $\mathbf{b} = \inf_{\beta \in A} \mathbf{b}^\beta$. By properties of the infimum and supremum, we have that $\mathbf{a}^\alpha \preceq \mathbf{a} \preceq \mathbf{b} \preceq \mathbf{b}^\alpha$, for all $\alpha \in A$, so the interval $[\mathbf{a}, \mathbf{b}]$ is non-empty; it is compact by assumption (II), and it is contained in all intervals of the family \mathcal{J} . On the other hand, any point \mathbf{c} that is common for all intervals J_α must be an upper bound for $\{\mathbf{a}^\alpha\}$ and a lower bound for $\{\mathbf{b}^\alpha\}$; hence $\mathbf{a} \preceq \mathbf{c} \preceq \mathbf{b}$, that is, $\mathbf{c} \in [\mathbf{a}, \mathbf{b}]$, and (ii) is proved. \square

PROOF OF THEOREM 3.1. It is clear that any two intervals in \mathcal{J} have a non-empty intersection; besides, by (I2), at least one of the intervals has finite endpoints. Then the assertion follows by Lemma 6.1. \square

The next Lemma is technical, and we need it for the proof of Theorem 4.1.

Lemma 6.2. *Let μ be any probability measure on Borel sets of \mathbb{R}^d . Let K be a compact set in \mathbb{R}^d and let \mathcal{A} be a family of closed convex subsets of K , with $\mu(A) > \frac{d}{d+1}$ for every $A \in \mathcal{A}$. Then the intersection of all sets $A \in \mathcal{A}$ is a non-empty compact set.*

Proof. If $\mu(A_i) > 1 - \varepsilon$, $i = 1, 2, \dots$, then it is easy to prove by induction that $\mu(A_1 \cdots A_n) > 1 - n\varepsilon$ for $n \geq 2$. Therefore, under given assumptions, for any $d+1$ sets $A_1, \dots, A_n \in \mathcal{A}$, it holds that $\mu(A_1 \cdots A_{d+1}) > 1 - (d+1) \cdot \frac{1}{d+1} = 0$. Hence, every $d+1$ sets of the family \mathcal{A} have a non-empty intersection. By Helly's intersection theorem ([9, 12.12.]), every finite number of convex sets in \mathcal{A} have a non-empty intersection. Since K is compact, then all sets in \mathcal{A} have a non-empty intersection (see e.g. [14, Theorem 17.4]). The intersection is compact since all sets in \mathcal{A} are compact. \square

PROOF OF THEOREM 4.1. Let $\delta \in (0, 1)$ be fixed. Assuming that (C_1) holds, we will first prove that every compact convex set $K \subset \mathbb{R}^d$ with $\mu(K) = 1 - \delta > 0$ contains a point x with $D(x; \mu, \mathcal{U}) \geq \frac{1-\delta}{d+1}$. Indeed, let $\varepsilon = \frac{1-\delta}{d+1}$ and suppose, contrary to the statement, that $D(x; \mu, \mathcal{U}) < \varepsilon$ for every $x \in K$,

where K is a compact set with $\mu(K) = 1 - \delta > 0$. Then (by (C_1)), for every $x \in K$ there exists a $U_x \in \mathcal{U}$, such that $\mu(U_x) < \varepsilon$. Clearly,

$$(6.1) \quad \bigcup_{x \in K} U_x \supset K.$$

Let $U_x^c = V_x$. Then $V_x \in \mathcal{V}$, and by (6.1) it follows that

$$(6.2) \quad \bigcap_{x \in K} (V_x \cap K) = \emptyset$$

Let us now define a new probability measure μ^* on \mathbb{R}^d , by

$$\mu^*(B) = \frac{\mu(B \cap K)}{1 - \delta}, \quad \text{where } B \subset \mathbb{R}^d \text{ is a Borel set.}$$

For each $x \in K$, we have that $\mu(V_x) > 1 - \varepsilon$, and

$$\mu(V_x \cap K) > \mu(V_x) + \mu(K) - 1 > 1 - \varepsilon - \delta = \frac{d(1 - \delta)}{d + 1},$$

hence $\mu^*(V_x \cap K) > \frac{d}{d+1}$. Now by Lemma 6.2, we conclude that the family of sets $V_x \cap K$ have non-empty intersection, which contradicts (6.2). So, the statement about compact convex sets is proved.

To prove the statement of the Theorem 4.1, note that the statement that we already proved yields the condition (C'_2) , and, with additional assumption (C''_2) , Theorem 2.2 is applicable. By the first part of the proof, each of the sets

$$S_n = \{x \in \mathbb{R}^d \mid D(x; \mu, \mathcal{U}) \geq \frac{1 - \frac{1}{n}}{d + 1}\}, \quad n = 1, 2, \dots$$

is non-empty; then their intersection.

$$\bigcap_{n=1}^{+\infty} S_n = \{x \in \mathbb{R}^d \mid D(x; \mu, \mathcal{U}) \geq \frac{1}{d + 1}\},$$

is also non-empty, by Theorem 2.2. This ends the proof. \square

PROOF OF LEMMA 4.1. We first prove that (C''_2) holds. For a fixed $\varepsilon > 0$, and a given probability measure μ , let $B_{1-\varepsilon}$ be a closed ball centered at origin, with $\mu(B_{1-\varepsilon}) > 1 - \varepsilon$. Then, by assumptions, there exists a set $V \in \mathcal{V}$ such that $B_{1-\varepsilon} \subset V$. By compactness, there exists $r > 0$ such that all points $x \in V$ satisfy $\|x\| \leq r$. Therefore, all points x with $\|x\| > r$ are in $U = V^c$, and, since $\mu(U) = 1 - \mu(V) < \varepsilon$, we conclude that for a given $\varepsilon > 0$ there exists $r > 0$ so that $D(x; \mu, \mathcal{U}) < \varepsilon$ for all x with $\|x\| > r$, which proves (C''_2) . Then by Theorem 4.1, the condition (C'_2) also holds. \square

PROOF OF THEOREM 4.2. Let $\overset{\circ}{H}$ be an open halfspace from $\overset{\circ}{\mathcal{H}}$, and let H be its closure. Given any $x \in \overset{\circ}{H}$, there exists a closed halfspace H_x that

can be obtained by translation of H in such a way that the border of H_x contains x . Then $H_x \in \mathcal{H}$ and, clearly,

$$\overset{\circ}{H} = \bigcup_{x \in \overset{\circ}{H}} H_x,$$

which implies condition (E') of Theorem 2.1 with $\mathcal{A} = \overset{\circ}{\mathcal{H}}$ and $\mathcal{B} = \mathcal{H}$. On the other hand, for any given closed halfspace $H \in \mathcal{H}$, there exists a sequence of halfspaces H_i , obtained from H by translation, such that

$$\overset{\circ}{H}_1 \supset \overset{\circ}{H}_2 \supset \dots \quad \text{and} \quad H = \bigcap_i \overset{\circ}{H}_i,$$

which is the condition (E'') . Therefore, by Theorem 2.1,

$$(6.3) \quad D(x; \mu, \mathcal{H}) = D(x; \mu, \overset{\circ}{\mathcal{H}}).$$

Now note that (4.1) gives condition (E') for $\mathcal{A} = \mathcal{U}$ and $\mathcal{B} = \overset{\circ}{\mathcal{H}}$ (by taking complements on both sides); then, as in the proof of Theorem 2.1, we find that

$$(6.4) \quad D(x; \mu, \mathcal{U}) \geq D(x; \mu, \overset{\circ}{\mathcal{H}}),$$

for every $x \in \mathbb{R}^d$. In the same way, (4.2) gives condition (E'') for $\mathcal{A} = \mathcal{U}$ and $\mathcal{B} = \mathcal{H}$, and so, again as in the proof of 2.1,

$$(6.5) \quad D(x; \mu, \mathcal{U}) \leq D(x; \mu, \mathcal{H}).$$

The statement of the theorem now follows from (6.3), (6.4) and (6.5).

PROOF OF THEOREM 5.1. Let $Q = Q_{1-\alpha}$. Then for every $\varepsilon > 0$, $\mu_f((-\infty, Q + \varepsilon]) = \mu(f^{-1}((-\infty, Q + \varepsilon])) > 1 - \alpha$, and, therefore, the set

$$V_\varepsilon = f^{-1}((-\infty, Q + \varepsilon]) \in \mathcal{V}$$

contains the level set $S_\alpha(\mu, \mathcal{U})$. This implies that

$$f(m) \leq Q + \varepsilon, \quad \text{for every } m \in S_\alpha(\mu, \mathcal{U}) \text{ and every } \varepsilon > 0.$$

Letting here $\varepsilon \rightarrow 0$, we get (5.1). \square

PROOF OF THEOREM 5.2. Starting from (3.4) and proceeding in the same way as in the proof of Theorem 5.1, we find that $f(m) \leq Q_{1/2} = \max\{\text{Med } \mu_f\}$, which proves (5.4). If, besides $Q_{1/2}$, any other median M of μ_f exists, then we have that $\mu_f((-\infty, M]) = 1/2$, hence the set $V_M = \{x \in \mathbb{R}^d \mid f(x) \leq M\}$ has the probability $\mu(V_M) = 1/2$. Therefore, $V_M = [\mathbf{a}, \mathbf{b}]$ has a non empty intersection with any interval $V_\alpha = [\mathbf{a}^\alpha, \mathbf{b}^\alpha]$ with $\mu(V_\alpha) > \frac{1}{2}$. Then, as in the proof of Lemma 6.1, it follows that $\mathbf{a}^\alpha \preceq \mathbf{b}$ and $\mathbf{a} \preceq \mathbf{b}^\alpha$ for all α , which implies, via relations $\mathbf{a}_0 = \sup_\alpha \mathbf{a}^\alpha$ and $\mathbf{b}_0 = \inf_\alpha \mathbf{b}^\alpha$, that

$$\mathbf{a}_0 \preceq \mathbf{b} \quad \text{and} \quad \mathbf{a} \preceq \mathbf{b}_0,$$

hence, $[\mathbf{a}, \mathbf{b}] \cap [\mathbf{a}_0, \mathbf{b}_0] = [\mathbf{a}_0, \mathbf{b}_0] \neq \emptyset$. Then the inequality (5.3) holds with any $m \in [\mathbf{a}_0, \mathbf{b}_0]$.

ACKNOWLEDGEMENT

This work is supported by the Ministry of Science and Technological Development of Republic of Serbia, project number 144021.

REFERENCES

- [1] Đ. Baljžović, M. Merkle, *Spatial medians, depth functions and multivariate Jensen's inequality*, 2007, <http://arxiv.org/abs/math/0701922>, preprint.
- [2] D.L. Donoho and M. Gasko, *Breakdown properties of location estimates based on halfspace depth and projected outlyingness*, *Ann. Stat.* **20** (1992), 1803–1827.
- [3] M. Ghosh, P. K. Sen, *Median Unbiasedness and Pitman Closeness*, *J. Amer. Statist. Assoc.* **84** (1989), 1089–1091.
- [4] B. Mendes, M. Merkle, *Some remarks regarding Pitman closeness*, *Univ. Beograd Publ. Elektrotehn. Fak. Ser. Mat.* **16** (2005), 1–11.
- [5] M. Merkle, *Jensen's inequality for medians*, *Stat. Prob. Letters* **71** (2005), 277–281.
- [6] T. K. Nayak, *Estimation of location and scale parameters using generalized Pitman nearness criterion*, *J. Statist. Plann. Inference*, 24 (1990), 259268.
- [7] T. R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton (1972).
- [8] P. J. Rousseeuw, I. Ruts, *The depth function of a population distribution*, *Metrika* **49**(1999), 213–244.
- [9] E. Schechter, *Handbook of Analysis and Its Foundations*, Academic Press, New York (1997).
- [10] C. G. Small, *Measures of centrality for multivariate and directional distributions*, *Canadian J. Statistics*, **15** (1987), 31–39.
- [11] C. G. Small, *A survey of multidimensional medians*, *Internat. Stat. Inst. Rev.* **58** (1990), 263–277.
- [12] R. J. Tomkins, *On conditional medians*, *Ann. Probab.* **3** (1975), 375–379.
- [13] J.W. Tukey, *Mathematics and picturing of data*, *Proc. International Congress of Mathematicians, Vancouver, 1974 Vol. 2* (1975), 523–531, ISBN 0-919558-04-6.
- [14] S. Willard, *General Topology*, Addison-Weseley (1970), republished by Dover (2004).
- [15] Y. Zuo and R. Serfling, *General notions of statistical depth function*, *Ann. Stat.* **28** (2000), 461–482.