Concentration of the empirical level sets of Tukey's halfspace depth

As Published: https://doi.org/10.1007/s00440-018-0850-0

Publisher: Springer Berlin Heidelberg

Persistent URL: https://hdl.handle.net/1721.1/131425

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Probability Theory and Related Fields manuscript No.
(will be inserted by the editor)

Concentration of the empirical level sets of Tukey's halfspace depth

Victor-Emmanuel Brunel

Received: date / Accepted: date

Abstract Tukey's halfspace depth has attracted much interest in data analysis, because it is a natural way of measuring the notion of depth relative to a cloud of points or, more generally, to a probability measure. Given an i.i.d. sample, we investigate the concentration of upper level sets of the Tukey depth relative to that sample around their population version. We show that under some mild assumptions on the underlying probability measure, concentration occurs at a parametric rate and we deduce moment inequalities at that same rate. In a computational prospective, we study the concentration of a discretized version of the empirical upper level sets.

Keywords Tukey depth · level set · multivariate quantiles · support function · semi-infinite linear programming

Mathematics Subject Classification (2010) MSC 62H11

1 Preliminaries and notation

1.1 Preliminary

Tukey's halfspace depth or, in short, Tukey depth, introduced by Tukey [46], has attracted much attention in multivariate data analysis, as a tool for understanding and describing which data are relevant in a given cloud of points. For a finite multivariate sample, Tukey depth at any given point x is the minimum proportion of points of the sample enclosed in a closed halfspace containing x. Tukey depth, together with other notions of statistical depths (see [48] for general definitions) has been studied and used extensively especially for

V.-E. Brunel
Department of Mathematics
Massachusetts Institute of Technology
77 Massachusetts Avenue,
Cambridge, MA 02139-4307, USA E-mail: vebrunel@mit.edu

description or graphical representation of data [28], robust [1, 12] or nonparametric (e.g., [29]) inference, bootstrap [47], supervised classification [16, 17], etc. When the sample consists of i.i.d. random points, we call it empirical Tukey depth and it has a population analog (one can find formal definitions of the population Tukey depth in Euclidean spaces in [42] and extensions to infinite dimensional Banach spaces in [13]). Consistency and limit theorems for the empirical Tukey depth are well-known (see [34], for instance, where the author tackles the asymptotic properties of the empirical Tukey depth seen as a stochastic process).

In this work, we are interested in the upper level sets of Tukey depth (we drop the qualifying *upper* in the sequel). These sets are nested and the center of gravity of the deepest one is called the *Tukey median*. On the opposite, the convex hull of a sample of n points is the largest bounded empirical level set. Convergence and concentration of this random polytope has attracted a lot of attention in convex and stochastic geometry (see [4, 6, 15] and the references therein). Intermediate level sets of Tukey depth for a sample of n points are extensions of the convex hull, called k-hulls, for $k \geqslant 1$. The k-hull of a sample of n points is the intersection of all closed halfspaces that contain at least $n-k$ sample points (see [9] for a similar definition) and the convex hull corresponds to the 0-hull of the sample.

As Tukey mentioned in his seminal work on statistical depth [46], the contours of Tukey depth are an informative tool for exploratory statistics and data visualization. Moreover, like univariate quantiles, they provide a statistical tool that is robust to outliers. In [40] and [22], Tukey depth level sets are used for classification. Let μ and ν be two probability measures and let X be a random variable distributed either according to μ or to ν. Consider the task of determining which of μ or ν is the distribution of X. If μ and ν are known, one way of solving this task consists of computing the Tukey depth level sets of order α of both μ and ν and determining to which X is the closest. Here, $\alpha \in (0, 1)$ is a tuning parameter. If μ and ν are unknown and only independent samples from μ and ν are available, a solution is to use the corresponding empirical level sets. In that case, one needs to control how close they are to their population versions.

We consider the level sets of the the empirical Tukey depth function with a given and fixed level, i.e., that does not depend on the sample size n. We show that they concentrate about their population version at the parametric speed of convergence, i.e., $n^{-1/2}$. These empirical level sets correspond to the k-hulls of the sample, when k is linear in n, i.e., of order αn for some $\alpha \in (0, 1)$. In the sublinear regime, i.e., when $k/n \to 0$, the concentration rates can be much worse (see [15] when the underlying distribution is log-concave and [4] when it is supported on a convex body). Estimation of level sets is a general problem in statistics (e.g., [11, 36]. An important case is that of estimating the level sets of a density [38, 39, 45]. For the level sets of depth functions, a very natural estimator is available: The level sets of the empirical depth function. The asymptotics of these empirical level sets have already been tackled in previous works for several depth functions, including Tukey depth. Consistency was proven in [20, 49]. In [24], the author shows that for all $\varepsilon \in (0,1)$, with probability $1-\varepsilon$, the empirical depth level set is sandwiched between two population level sets whose levels are at a distance of order $n^{-1/2}$ from each other. However, the constants are not explicit and the way they depend on ε cannot be derived from the results, which, as a consequence, do not yield moment inequalities. In [20, 24, 49], the proofs are based on the global behavior of the stochastic process defined by the empirical depth, indexed by the ambient Euclidean space. Hence, the results in these works are based on global and strong assumptions on the underlying probability measure. For instance, in [24] it is assumed that the underlying distribution has a unimodal density and that the depth function does not oscillate and is not too flat. In our work, we focus on Tukey depth and only make assumptions that guarantee some local continuity properties of the directional marginals of the underlying distribution. We show that these assumptions are very weak, in the sense that they are satisfied by a broad class of distributions, including most commonly used ones. Not only we achieve the same (parametric) rate as obtained in [24], but our main result is nonasymptotic with explicit constants and it yields moment inequalities with a parametric rate.

Our approach is based on a polyhedral representation of the level sets of the population and empirical Tukey depths. As we will see in Lemma 1, which is a refinement of Theorem 2 in [26], these level sets can also be written as multivariate quantile sets, defined as convex regions that satisfy infinitely many linear constraints. It is because of such a multivariate quantile representation that the level sets of Tukey depth have also attracted attention in multivariate quantile regression (see [7, 19] and the references therein). With this approach, we reduce the problem to that of estimating the support function of the population level sets. We believe that the techniques we use in our proofs could be useful in other problems related to support function estimation. For instance, in [18], the support function of an unknown convex set is observed up to some noise; We believe that our proof method could be used in order to bound from above the risk for estimation of the unknown convex set in Hausdorff distance, whereas the measure of the risk used in [18] does not have a natural, geometric interpretation.

Computation of the empirical Tukey depth level sets for samples of n points is a challenging problem. In dimension 2, they can be computed in $O(n^2)$ (see [35]). A naive computation of the Tukey depth at one point would require to explore infinitely many halfspaces, which is not feasible. In higher dimensions, there is no practical and efficient way to compute the level sets of the Tukey depth. A random version of the Tukey depth was defined in [10] where only M independent random directions are considered. In [10], the choice of M is only based on empirical evidence. We use this random version of the Tukey depth in order to approximate the empirical level sets. We show that for a specific choice of M, they are consistent and still concentrate at the same parametric speed as the original empirical sets. The choice that we use for M grows exponentially with the dimension of the ambient space, but have no lower bound on the minimal number of random directions that are necessary to still achieve $n^{-1/2}$-consistency.

Before going further into details, we introduce some notation. In this paper, $d \geqslant 2$ and $n \geqslant 1$ are fixed integers, unless stated otherwise. The Euclidean norm in R^d is denoted by $|\cdot|$ and the dot product between two vectors x and y is denoted by $\langle x, y\rangle$. The $(d-1)$-dimensional unit sphere is $\mathcal{S}^{d-1} = \{u \in \mathbb{R}^d : |u| = 1\}$. For $u \in \mathcal{S}^{d-1}$, $u^\perp$ stands for the hyperplane in $\mathbb{R}^d$ that is orthogonal to u. If k is a positive integer, $a \in \mathbb{R}^k$ and $R \geqslant 0$, $B_k(a, R)$ (resp. $B'_k(a, R)$) stands for the closed (resp. open) Euclidean ball in $\mathbb{R}^k$ with center a and radius R. When $k = d$, we drop the subscript k.

The complement of a set A is denoted by $A^\complement$. The symmetric difference between two sets A and B in $\mathbb{R}^d$ is denoted by $A \triangle B$. For $k \geqslant 1$, if A is a measurable set in $\mathbb{R}^k$ (equipped with the Lebesgue measure), we denote by $\mathsf{Vol}_k(A)$ its k-dimensional volume, i.e., its Lebesgue measure in $\mathbb{R}^k$.

For $A \subseteq \mathbb{R}^d$, the interior of A is denoted by $\overset{\circ}{A}$: this is the largest open set included in A. The collection of closed halfspaces in $\mathbb{R}^d$ is denoted by $\mathcal{H}$. For $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$, we define the closed halfspace $H_{u,t} = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant t\}$.

The Hausdorff distance between two sets $K, K' \subseteq \mathbb{R}^d$ is

$$d_{\mathsf{H}}(K, K') = \inf\{\varepsilon > 0 : K \subseteq K' + \varepsilon B(0,1) \text{ and } K' \subseteq K + \varepsilon B(0,1)\},$$

where we set $\inf(\varnothing) = \infty$. If K is a convex set, its support function h_K is defined as $h_K(u) = \sup_{x \in K} \langle u, x\rangle$, $u \in \mathbb{R}^d$: If u is a unit vector, $h_K(u)$ is the signed distance from the origin to the farthest tangent hyperplane of K in the direction of u.

The cardinality of a finite set I is denoted by $\#I$. For $x \in \mathbb{R}$, we denote by $\lceil x\rceil$ (resp. $\lfloor x\rfloor$) the smallest integer larger (resp. largest integer smaller) or equal to x.

Throughout the paper, $X, X_1, X_2, \ldots$ are independent and identically distributed (i.i.d.) random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in $\mathbb{R}^d$. Their common probability distribution is denoted by μ and is defined on the Borel σ-algebra of $\mathbb{R}^d$. The empirical distribution μ_n is defined by $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, where δ_a is the Dirac measure at the point $a \in \mathbb{R}^d$.

For two positive sequences $(a_n)_{n\geqslant 1}$ and $(b_n)_{n\geqslant 1}$, we write $a_n = O(b_n)$ when the ratio a_n/b_n is bounded uniformly in $n \geqslant 1$. For two positive sequences of random variables $(A_n)_{n\geqslant 1}$ and $(B_n)_{n\geqslant 1}$, we write $A_n = O_{\mathbb{P}}(B_n)$ when for all $\delta > 0$, there exists $M_\delta > 0$ such that $\mathbb{P}[A_n > M_\delta B_n] \leqslant \delta, \forall n \geqslant 1$.

Section 2 is devoted to general results about Tukey depth level sets. Our main theorems are given in Section 3 and the proofs are deferred to Section 4. The rest of this section is dedicated to important definitions.

1.2 Definitions

The Tukey depth associated with a probability measure μ in $\mathbb{R}^d$ is the function

$$D_\mu(x) = \inf_{H\in\mathcal{H}:x\in H} \mu(H), \quad \forall x \in \mathbb{R}^d.$$

We refer to D_μ as the *population Tukey depth* and to D_{μ_n} as the *empirical Tukey depth*.

In this work, we are interested in comparing the level sets of D_μ and D_{μ_n}. Let $\alpha \in (0,1)$ be fixed. The α-level set of D_μ is defined as $G_\mu = \{x \in \mathbb{R}^d : D_\mu(x) \geqslant \alpha\}$ and we denote by $\hat{G}$ the α-level set of D_{μ_n}: $\hat{G} = \left\{x \in \mathbb{R}^d : D_{\mu_n}(x) \geqslant \alpha\right\}$. We study how fast $\hat{G}$ concentrates around G_μ, i.e., how fast the stochastic convergence of $d_{\mathsf{H}}(\hat{G}, G_\mu)$ to zero is. As intermediate tools and for independent interest, we introduce the following sets associated with μ:

1. *The multidimensional* $(1-\alpha)$*-quantile set of* μ*:*
   Let X be a random variable with probability distribution μ. For $u \in \mathbb{R}^d$, let $q_u^\flat$ and $q_u^\sharp$ be the lower and upper $(1-\alpha)$-quantile of $\langle u, X\rangle$, respectively:

   $$q_u^\flat = \inf\{t \in \mathbb{R} : \mathbb{P}[\langle u, X\rangle \leqslant t] \geqslant 1-\alpha\} \quad \text{and}$$

   $$q_u^\sharp = \sup\{t \in \mathbb{R} : \mathbb{P}[\langle u, X\rangle \geqslant t] \geqslant \alpha\}.$$

   The corresponding lower and upper multidimensional $(1-\alpha)$-quantile sets of μ are defined as

   $$G_{\mathsf{MQ}}^\eta = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant q_u^\eta, \forall u \in \mathcal{S}^{d-1}\}, \quad \eta \in \{\flat, \sharp\}. \tag{1}$$

2. *The* α*-floating body of* μ*:* $G_{\mathsf{FB}} = \displaystyle\bigcap_{H\in\mathcal{H}:\mu(H)\geqslant 1-\alpha} H.$

As we will see in Lemma 1 below, these sets are other representations of the Tukey depth level sets. The representation in terms of multidimensional quantile sets is particularly convenient for our purposes because it characterizes the Tukey depth level sets through linear constraints. We make the floating body part of our analysis because it plays an important role for random polytopes. Barany and Larman [2] proved that if μ is the uniform distribution in a convex and compact set of volume 1, then the expected missing volume of the convex hull of $X_1, \dots, X_n$ behaves aymptotically as the missing volume of the $(1/n)$-floating body of μ. Fresen [15] proved that if μ is log-concave, the convex hull of $X_1, \dots, X_n$ approximates the $(1/n)$-floating body of μ with high probability. For very small values of α, even smaller than $1/n$, when the empirical level set would be a very poor estimator of G_μ, [21] defines and studies an estimator that extends univariate estimators from extreme value theory.

1.3 The one dimensional case

Tukey depth is a very simple and intuitive concept in dimension one. Let μ be a probability measure in $\mathbb{R}$ and $X_1, \ldots, X_n$ be i.i.d. random points distributed according to μ. The Tukey depth of a given real number x is simply given by

$$D_\mu(x) = \min\left(\mu\left((-\infty, x]\right), \mu\left([x, \infty)\right)\right).$$

In particular, if the cumulative distribution function F of μ is continuous, then $D_\mu(x) = \min\left(F(x), 1 - F(x)\right)$. In that case, the α-level set of D_μ is the interval $[q_\alpha, q_{1-\alpha}]$, where q_β is the β-quantile of μ, for all $\beta \in (0,1)$. Then, the corresponding empirical level set $\hat{G}$ is the interval $[X_{(\lceil n\alpha \rceil)}, X_{(\lfloor n(1-\alpha) \rfloor + 1)}]$, where $(X_{(1)}, \ldots, X_{(n)})$ is the increasing reordering of the list $(X_1, \ldots, X_n)$. Then, convergence and concentration of $\hat{G}$ to G_μ reduces to convergence and concentration of the empirical quantiles of $X_1, \ldots, X_n$. For all $\beta \in (0,1)$, q_β can be estimated consistently if F is not too flat around β. In particular, if ν has a density f such that $f(q_\beta) > 0$, then the empirical β-quantile is asymptotically normal and in order to show concentration with rate $n^{-1/2}$, it is sufficient to assume that f is bounded away from zero in a vicinity of q_β, i.e., that F increases at least linearly around q_β.

Our results are built upon that fact: We will reduce the multivariate problem to one-dimensional problems, by considering projections of the measure μ along all possible directions, in which we will prove concentration of empirical quantiles around their population versions.

2 General results on Tukey depth level sets

We start with a simple lemma that shows the relationships between the sets defined above: The Tukey depth level sets, the lower and upper multidimensional quantile sets and the floating bodies. This lemma is a refinement of Theorem 2 in [26] but we include its proof at the end for the sake of completeness.

Lemma 1 $G_{FB} = G^\flat_{MQ} \subseteq G^\sharp_{MQ} = G_\mu$.

In particular, if μ satisfies some continuity property, e.g., Assumption 1 below, then $q^\flat_u = q^\sharp_u$ for all unit vectors u, so the inclusion becomes an equality and all four sets are equal.

[26] provides an interesting discussion about the multivariate quantile representation of G_μ: In brief, the knowledge of G_μ does not imply the knowledge of all univariate quantiles $q^\sharp_u, u \in \mathcal{S}^{d-1}$. Indeed, some of the linear constraints that define $G^\sharp_{MQ}$ may not be active, i.e., there may be some unit vectors u for which $\langle u, x \rangle < q^\sharp_u, \forall x \in G^\sharp_{MQ}$. This fact constitutes the main difficulty in the proof of Theorem 2 below, where we use the support function of $G^\sharp_{MQ}$. For $u \in \mathcal{S}^{d-1}$, it is clear that the linear constraint "$\langle u, x \rangle \leqslant q^\sharp_u$" is active if and only if $h_{G_\mu}(u) = q^\sharp_u$. If that constraint is not active, then $h_{G_\mu}(u) < q^\sharp_u$. In that

case, not only G_μ provides no information about $q_u^\sharp$, as discussed in [26], but $q_u^\sharp$ alone does not give any information about $h_{G_\mu}(u)$, and we need to understand how $h_{G_\mu}(u)$ depends on the $q_v^\sharp$'s that correspond to active constraints.

For its independent interest, we may ask the following question: For which distributions μ are all the linear constraints that determine $G_{\mathsf{MQ}}^\sharp$ active ? First, we have the following proposition about polyhedral representations of convex sets.

Proposition 1 *Let $(t_u)_{u\in\mathbb{R}^d} \subseteq \mathbb{R}$ be positively homogeneous, i.e., $t_{\lambda u} = \lambda t_u$, $\forall \lambda \geqslant 0, u \in \mathbb{R}^d$ and define the convex set $G = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant t_u, \forall u \in \mathcal{S}^{d-1}\}$. Assume that $\overset{\circ}{G} \neq \varnothing$. Then, the following statements are equivalent:*

(i) All the linear constraints that define G are active;
(ii) For all $u \in \mathcal{S}^{d-1}$, $h_G(u) = t_u$;
(iii) The family $(t_u)_{u\in\mathbb{R}^d}$ is subadditive, i.e., $t_{u+v} \leqslant t_u + t_v, \forall u, v \in \mathbb{R}^d$.

As a consequence of this lemma, the upper quantiles $q_u^\sharp, u \in \mathcal{S}^{d-1}$, are completely determined by G_μ if and only if the family $(q_u^\sharp)_{u\in\mathbb{R}^d}$ is sublinear, i.e., subadditive and positively homogeneous.

Open question 1 *For what distributions μ are the upper quantiles $q_u^\sharp, u \in \mathbb{R}^d$, sublinear, no matter the value of $\alpha \in (0,1)$?*

A Gaussian distribution has sublinear upper quantiles, as a consequence of the triangle inequality for symmetric positive semidefinite matrices. If μ is the Gaussian distribution with centroid m and covariance matrix Σ, then for all $u \in \mathcal{S}^{d-1}$, $q_u^\sharp = \langle u, m\rangle + \Phi^{-1}(1-\alpha)\sqrt{\Sigma(u,u)}$, where Φ is the cumulative distribution function of the univariate standard Gaussian distribution. The triangle inequality ensures that the map $u \in \mathbb{R}^d \mapsto \sqrt{\Sigma(u,u)}$ is sublinear, yielding sublinearity of $(q_u^\sharp)_{u\in\mathbb{R}^d}$.

As a generalization of Gaussian distributions, and because they are known to be rigid (see [31] for examples of this rigidity), we may ask if a log-concave probability measure have sublinear upper quantiles.

Open question 2 *Assume that μ is log-concave. Is it true that the upper quantiles $q_u^\sharp, u \in \mathbb{R}^d$ are sublinear, no matter the value of $\alpha \in (0,1)$?*

Remark 1 The multidimensional quantile sets are convex sets. Thus, they fail to capture the structure of complex probability measures, such as mixtures. The floating body (also called *convex floating body* in the convex geometry literature, see [44]) is defined as an intersection of closed halfspaces, i.e., the complement of the union of open halfspaces. Instead, one could think of an r-convex floating body, using the notion of r-convexity (see [32]): $G_{\mathsf{FB}}^{(r)} = \left(\bigcup_{a\in\mathbb{R}^d:\mu(B'(a,r))<\alpha} B'(a,r)\right)^{\complement}$ and its empirical analog $\hat{G}_{\mathsf{FB}}^{(r)}$ can be defined similarly, by replacing μ with μ_n. When $r = \infty$, $G_{\mathsf{FB}}^{(r)} = G_{\mathsf{FB}}$. An asymptotic analysis of $\hat{G}^{(r)}_{\mathsf{FB}}$ would require a different approach than ours, but seems to be relevant in order to describe more complex probability measures. In [37], r-convexity is exploited to estimate the support of probability distributions while relaxing convexity and even connectivity assumptions. We leave this question for further work.

The next result shows that unless μ has atoms, the level set G_μ is empty when α is too large.

Theorem 1 *Let $\alpha > 1/2$. Then, either G_μ is empty or it contains exactly one point. In the latter case, i.e., if $G_\mu = \{x\}$ for some $x \in \mathbb{R}^d$, then x is an atom of μ: $\mu(\{x\}) \geqslant 2\alpha - 1 > 0$.*

On the one hand, if μ has an atom x with $\mu(\{x\}) > 1/2$, then $D_\mu(x) \geqslant \mu(\{x\}) > 1/2$, hence, $G_\mu \neq \varnothing$ for $\alpha = \mu(\{x\}) > 1/2$. On the other hand, it is known ([12], Lemma 6.3) that G_μ is always nonempty when $\alpha \geqslant 1/(d+1)$. The following two examples show that very general probability measures μ can satisfy $G_\mu \neq \varnothing$ for large values of $\alpha \leqslant 1/2$, independent of the dimension d, and yet have no atoms:

- If μ is centrally symmetric, i.e., satisfies $\mu(x + A) = \mu(x - A)$ for all Borel set $A \subseteq \mathbb{R}^d$, where x is the center of symmetry of μ, then $D_\mu(x) \geqslant 1/2$, hence, G_μ is nonempty for all $\alpha \in [0, 1/2]$.
- If μ is log-concave, then any closed halfspace H containing the centroid of μ satisfies $\mu(H) \geqslant e^{-1}$ (see Lemma 5.12 in [31]). Hence, the depth of the centroid of μ is at least e^{-1}, which implies that G_μ is non empty for α as large as $e^{-1} \approx .37$.

3 Concentration of the empirical Tukey depth level sets

Consider the following assumptions, where we let ε, L, r, R be fixed positive numbers satisfying $\varepsilon < r \leqslant R$.

Assumption 1
- *For all $u \in \mathcal{S}^{d-1}$, the cumulative distribution function F_u of $\langle u, X\rangle$ is continuous on $[q_u^\sharp - \varepsilon, q_u^\sharp + \varepsilon]$.*
- *$\left|F_u(t) - F_u(q_u^\sharp)\right| \geqslant L\left|t - q_u^\sharp\right|$, for all $u \in \mathcal{S}^{d-1}$ and all $t \in [q_u^\sharp - \varepsilon, q_u^\sharp + \varepsilon]$.*

Assumption 2 *There exists $a \in \mathbb{R}^d$ such that $B(a, r) \subseteq G_\mu \subseteq B(a, R)$.*

Assumption 1 ensures that $q_u^\flat = q_u^\sharp$ for all $u \in \mathcal{S}^{d-1}$, hence, that $G^\flat_{\mathsf{MQ}} = G^\sharp_{\mathsf{MQ}}$ and that the cumulative distribution functions F_u are not too flat around their quantiles $q_u^\flat = q_u^\sharp$. Note that this assumption is not global in nature, however, the local control of F_u is required to hold uniformly in all directions $u \in \mathbb{S}^{d-1}$. Yet, we introduce two more assumptions in the sequel (Assumptions 3 and 4), each of which is shown to be stronger than Assumption 1 for some values of the parameters ε and L that depend on μ, and which are satisfied by most commonly used distributions.

By Lemma 1, $\hat{G}$ can also be written as the empirical upper multidimensional $(1-\alpha)$-quantile set associated with $X_1, \ldots, X_n$:

$$\hat{G} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leqslant \hat{q}_u^\sharp, \forall u \in \mathcal{S}^{d-1}\}, \tag{2}$$

where, for $u \in \mathbb{R}^d$, $\hat{q}_u^\sharp = \sup\left\{t \in \mathbb{R} : \#\{i = 1, \ldots, n : \langle u, X_i \rangle \geqslant t\} \geqslant n\alpha\right\}$ is the upper empirical $(1-\alpha)$-quantile of $\langle u, X_1 \rangle, \ldots, \langle u, X_n \rangle$. For the sake of notation, we will write $\hat{q}_u$ instead of $\hat{q}_u^\sharp$ in the sequel.

As a consequence of Lemma 1, in order to show concentration of $\hat{G}$ around G_μ, one can compare their polyhedral representations given by (2) and $G^\sharp_{\mathsf{MQ}}$, which are written in terms of linear constraints. This is essential in the proof of our next theorem, which uses semi-infinite linear programming as one of its main ingredients.

Next theorem asserts that if Assumptions 1 and 2 are both satisfied, then $\hat{G}$ concentrates around G_μ at a parametric speed. In particular, that speed depends on the dimension d only through multiplicative constants.

Theorem 2 *Let μ satisfy Assumptions 1 and 2. Then, the random set $\hat{G}$ satisfies the following deviation inequality:*

$$\mathbb{P}\left[d_{\mathsf{H}}(\hat{G}, G_\mu) > \frac{Cx}{\sqrt{n}}\right] \leqslant Ae^{-L^2x^2/2+10\sqrt{5(d+1)}x},$$

for all $x \geqslant 0$ with $\frac{10\sqrt{5(d+1)}}{L} \leqslant x < \varepsilon\sqrt{n}$, where $C = \frac{R}{r}\frac{1+\varepsilon/r}{1-\varepsilon/r}$ and $A = e^{-250(d+1)}$.

Note that in Theorem 2, if n is not large enough, the domain for x will be empty. Moreover, the upper bound in Theorem 2 is nontrivial as soon as $x \geqslant \frac{10\sqrt{5(d+1)}}{L}\left(1+\sqrt{\max(0, 1-L)}\right)$. The constants depend on d and the parameters ε, r, R, L. These parameters are hard to compute in practice, for a given distribution μ. However, Theorem 2 provides a concentration inequality that is uniform over all distributions μ that satisfy Assumption 1 and 2. Moreover, we give simple asymptotic consequences of Theorem 2 below.

First, a truncated version of $\hat{G}$ has its expected error converging to zero at the speed $n^{-1/2}$:

Corollary 1 *Define the random set*

$$\hat{G}^* = \begin{cases} \hat{G} \cap B'(0, \log n) & \textit{if } \hat{G} \neq \varnothing \\ \{0\} & \textit{otherwise.} \end{cases}$$

Let μ satisfy Assumptions 1 and 2 and assume, in addition, that $|a| \leqslant \tau$ for some $\tau > 0$. Then, for all $k > 0$, $\mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)^k\right] = O\left(n^{-k/2}\right)$. The multiplicative constants in these asymptotic comparisons depend on $d, r, R, \varepsilon, L, \tau$ and k only.

Remark 2 – In Corollary 1, the upper bounds are uniform on the class of probability measures μ that satisfy both Assumptions 1 and 2 with $|a| \leqslant \tau$. Hence, Corollary 1 gives an upper bound for the rate of the minimax risk in estimation of G_μ on that class of probability measures, and this rate is parametric. Note that the assumption $|a| \leqslant \tau$ could be dropped in Corollary 1, but then the multiplicative constants in the asymptotic comparisons would also depend on a and we would loose uniformity of the upper bounds.
– The threshold $\log n$ in the definition of $\hat{G}^*$ is arbitrary and could be replaced with any sequence that grows to infinity at most polynomially in n.

Define the maximal depth α_μ^* of μ as $\max_{x\in\mathbb{R}^d} D_\mu(x)$. Consider the two following assumptions:

Assumption 3 *The probability measure μ is absolutely continuous with respect to the Lebesgue measure, its density f is continuous and positive everywhere and there exist $C > 0$ and $\nu > d-1$ such that $f(x) \leqslant C(1+|x|)^{-\nu}, \forall x \in \mathbb{R}^d$.*

In the sequel, if μ has a density f with respect to the Lebesgue measure, we call the support of μ the set of vectors $x \in \mathbb{R}^d$ for which $f(x) > 0$.

Assumption 4 *The probability measure μ is absolutely continuous with respect to the Lebesgue measure, its support is bounded and convex and its density is uniformly continuous on its support.*

Assumptions 3 and 4 are sufficient but not necessary for next corollary. However, they include a lot of useful distributions. For example, any log-concave distribution in $\mathbb{R}^d$ with positive density satisfies Assumption 3: A log-concave density is continous on its support and decays exponentially fast when $|x| \to \infty$. If μ has a density of the form $f(x) = h(\langle x, \Sigma x\rangle)$, where Σ is a $d \times d$ symmetric positive definite matrix and h is a positive continuous function that satisfies $h(t) \leqslant C(1+|t|)^{-\nu}$ for all $t \in \mathbb{R}$, with $\nu > d-1$, then μ satisfies Assumption 3 as well. If μ is the uniform distribution on a compact, convex set in $\mathbb{R}^d$, then it satisfies Assumption 4.

Corollary 2 *Let μ satisfy either Assumption 3 or Assumption 4. Suppose that $\alpha \in (0, \alpha_\mu^*)$, independently of n. Then, $d_H(\hat{G}, G_\mu) = O_{\mathbb{P}}\left(n^{-1/2}\right)$.*

Remark 3 – Corollary 2 shows that the rate of convergence of the empirical level sets is parametric.
– Surprisingly, if μ is the uniform distribution on a compact, convex set K in $\mathbb{R}^d$, the rate does not depend on the smoothness of the boundary of K. This seems paradoxical, since it is known that if $\alpha = 1/n$, $\hat{G}$ is the convex hull of $X_1, \ldots, X_n$, which converges to K at a rate that depends on the smoothness of the boundary of K (see [2]). However, in [2]:

  - $\alpha = 1/n$ depends on n. In our work, α does not depend on n and hence, the floating body $G_{\mathsf{FB}} = G_\mu$ is bounded away from the boundary of K, which attenuates the effect of its smoothness.
  - Convergence is towards the support K itself, not towards the floating body of μ. When $\alpha = 1/n$, it is not clear whether the convergence of the distance between the empirical and the population $(1/n)$-floating bodies depends on the smoothness of the boundary of K. By the triangle inequality, $d_{\mathsf{H}}(\hat{G}, K) \leqslant d_{\mathsf{H}}(\hat{G}, G_\mu) + d_{\mathsf{H}}(G_\mu, K)$. The $(1/n)$-floating body G_μ converges to K at a speed that depends on the smoothness of the boundary of K [2, 44], but to the best of our knowledge, it is not known whether the speed of convergence of $d_{\mathsf{H}}(\hat{G}, G_\mu)$ depends on the smoothness of K too.
- [24] obtained the parametric rate $n^{-1/2}$ for general measures of statistical depth, under quite strong assumptions on μ which rule out many important distributions, as compared to ours (e.g., compactly supported densities). In addition, they do not compare $\hat{G}$ to G_μ directly, but to level sets of D_μ with levels $\alpha \pm Mn^{-1/2}$, for some $M > 0$, leaving out a deterministic bias. Yet, we believe that they could achieve the same rate as ours. However, unlike Theorem 2, their result is not informative about the tail of the distribution of $d_{\mathsf{H}}(\hat{G}, G_\mu)$, because of implicit dependency of the constant M on the probability level (see [24], Theorem 1).

Computation of $\hat{G}$ is a hard problem. Its concentration around G_μ is a question of its own geometric and probabilistic interest, but it also has important statistical implications. For instance, as we saw in Corollary 1, it provides a benchmark for the minimax risk for estimation of G_μ based on an i.i.d. sample. However, if $\hat{G}$ is too hard to compute, this does not have much of a practical interest. Computation of the Tukey depth D_{μ_n} at a single point is equivalent to the problem of finding a hemisphere that contains the largest number of points positioned on the unit sphere, which is NP hard in high dimension [23]. However, in fixed dimension, some deterministic and random algorithms to compute an approximate or exact value of the Tukey depth have been suggested (see [14, 40, 41] and the references therein). For the actual computation of the Tukey depth level sets relative to a point cloud in dimension 2, we refer to [35]. These sets are polygons, hence, their computation reduces to finding either their vertices or their faces. To our knowledge, there are no algorithms to compute these sets exactly when $d \geqslant 3$. Here, we define a random approximation of $\hat{G}$ that can be computed exactly, yet in an exponential time in d. Lemma 1 gives a representation of $\hat{G}$ through infinitely many linear constraints. By selecting a finite number of these constraints, using a collection of unit vectors that are well spread on the unit sphere, one can obtain a suitable approximation of $\hat{G}$.

Our random approximation is obtained by sampling random vectors on the unit sphere. For $M \geqslant 1$, set $\tilde{G}_M = \{x \in \mathbb{R}^d : \langle U_j, x\rangle \leqslant \hat{q}_{U_j}, \quad \forall j = 1, \dots, M\}$, where $U_1, \dots, U_M$ are i.i.d. uniform random variables on $\mathcal{S}^{d-1}$, independent of $X_1, \ldots, X_n$. The following theorem shows that a certain choice of M leads to an estimator that of G_μ that concentrates as fast as $\hat{G}$.

Theorem 3 *Let μ satisfy Assumptions 1 and 2 and assume that the quantiles $(q_u^\sharp)_{u\in\mathbb{R}^d}$ are subadditive. Then, for all $M \geqslant 1$, the random set $\tilde{G}_M$ satisfies the following deviation inequality:*

$$\mathbb{P}\left[d_H(\tilde{G}_M, G_\mu) > \frac{Cx+4R}{\sqrt{n}}\right] \leqslant Ae^{-L^2x^2/2+10\sqrt{5(d+1)}Lx} + 6^d \exp\left(-\frac{M}{2d8^{(d-1)/2}n^{d-1}} + (d/2)\log n\right),$$

for all real numbers x with $10\sqrt{5(d+1)} \leqslant x < \varepsilon\sqrt{n}$, where $C = \dfrac{R}{r}\dfrac{1+\varepsilon/r}{1-\varepsilon/r}$.

Remark 4 In Theorem 3, we assume that the population quantiles are subadditive, which, by Proposition 1, ensures that they are completely characterized by the knowledge of G_μ. How strong this assumption is is an open question (see Open questions 1 and 2).

Theorem 3 yields the following asymptotic upper bound for a truncated version of $\tilde{G}_M$, if M is chosen large enough.

Corollary 3 *Define the random set $\tilde{G}_M^*$ as*

$$\tilde{G}_M^* = \begin{cases} \tilde{G}_M \cap B(0, \log n) & \text{if } \tilde{G}_M \neq \varnothing, \\ \{0\} & \text{otherwise.} \end{cases}$$

Let $k > 0$. Recall the notation and assumptions of Theorem 3. If, in addition, $|a| \leqslant \tau$ for some $\tau > 0$, then for $M > d8^{(d-1)/2}(d+k)n^{d-1}\log n$, $\tilde{G}_M^*$ satisfies $\mathbb{E}\left[d_H(\tilde{G}_M^*, G_\mu)^k\right] = O\left(n^{-k/2}\right)$. The multiplicative constants in this asymptotic comparison depend only on d, r, R, ε, L and k.

In addition, the following stochastic upper bound holds under subadditivity of the population quantiles and either Assumption 3 or Assumption 4 :

Corollary 4 *Let $\alpha \in (0, \alpha_\mu^*)$ and μ satisfy either Assumption 3 or Assumption 4. Let $M = d^2 8^{(d-1)/2}n^{d-1}\log n$. Then, if the quantiles $(q_u^\sharp)_{u\in\mathbb{R}^d}$ are subadditive, $d_H(\tilde{G}_M, G_\mu) = O_{\mathbb{P}}\left(1/\sqrt{n}\right)$.*

Note that the prescribed value of M rapidly becomes very large as d grows. By the way $\tilde{G}_M$ is defined, there is a membership oracle (i.e., a systematic way of concluding whether a given point is in $\tilde{G}_M$) that only needs to check nM linear inequalities, where nM is of order $d^2 8^{(d-1)/2}n^d \log n$. This number is exponentially large in d. However, compare this to the optimal cost for outputing the list of vertices of the convex hull of n points in dimension d, which is of order $n^{\lfloor d/2\rfloor}$ [8]. Therefore, it is not clear whether the prescribed value of M is optimal, but it seems that an exponential dependency in d is unavoidable in general.

Finally, we note that Theorem 2 can be extended to more classes of distributions. Namely, consider the following assumption, that is an alternative to Assumption 1. We still let ε and L be fixed positive numbers and we add one additional parameter $\gamma > 0$.

Assumption 5 – *For all $u \in \mathcal{S}^{d-1}$, the cumulative distribution function F_u of $\langle u, X\rangle$ is continuous on $[q_u^\sharp - \varepsilon, q_u^\sharp + \varepsilon]$.*
– *$F_u(t) - F_u(q_u^\sharp) \geqslant L(t - q_u^\sharp)^\gamma$, for all $u \in \mathcal{S}^{d-1}$ and all $t \in [q_u^\sharp - \varepsilon, q_u^\sharp + \varepsilon]$.*

If $\gamma < 1$, Assumption 5 allows for less smooth cumulative distributions than Assumption 1. Then, it is easy to adapt the proofs of Theorem 2 and its intermediate lemmas in order to show the following theorem.

Theorem 4 *Let μ satisfy Assumptions 5 and 2. Then, the random set $\hat{G}$ satisfies the following deviation inequality:*

$$\mathbb{P}\left[d_H(\hat{G}, G_\mu) > \frac{Cx}{n^{1/(2\gamma)}}\right] \leqslant Ae^{-L^2x^{2\gamma}/2+10\sqrt{5(d+1)}x^\gamma},$$

for all $x \geqslant 0$ with $\left(\frac{10\sqrt{5(d+1)}}{L}\right)^{1/\gamma} \leqslant x < \varepsilon n^{-1/\gamma}$, where C and A are the same constants as in Theorem 2.

As a consequence, under Assumptions 5 and 2, the rate of convergence of the empirical level set towards its population version is $n^{-1/(2\gamma)}$.

4 Proofs

4.1 Preliminary lemmas in convex geometry and semi-infinite linear programming

Lemma 2 *Let K and L be two convex sets and assume that L is closed. Then, $K \subseteq L \iff h_K(u) \leqslant h_L(u), \forall u \in \mathcal{S}^{d-1}$. In particular, K is bounded if and only if the restriction of its support function to the unit sphere is bounded.*

Proof In the first part of the lemma, the left-to-right direction directly follows from the definition of the support function. The right-to-left direction is a consequence of [43, Theorem 1.3.7]. Assume that $h_K(u) \leqslant h_L(u), \forall u \in \mathcal{S}^{d-1}$ and that there exists $x \in K\backslash L$. Then, by [43, Theorem 1.3.7], $\{x\}$ and L can be strongly separated, i.e., there exist $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$ such that $\langle u, x\rangle > t$ and $\langle u, y\rangle \leqslant t$ for all $y \in L$. As a consequence, $h_L(u) \leqslant t < \langle u, x\rangle \leqslant h_K(u)$, which is false.

For the second part of the lemma, note that the support function of a ball centered at the origin with radius $R \geqslant 0$ is constant, equal to R on the unit sphere. Hence, $h_K(u) \leqslant R, \forall u \in \mathcal{S}^{d-1} \iff K \subseteq B(0, R)$, which proves the second part of the lemma.

The following lemma, borrowed from [43, Theorem 1.8.11], conveniently connects the Hausdorff distance between two convex bodies (i.e., convex and compact sets) to their support functions.

Lemma 3 *Let K and L be two convex bodies. Then,*

$$d_{H}(K,L) = \max_{u\in\mathcal{S}^{d-1}} |h_K(u) - h_L(u)|.$$

In the next two lemmas, we let $\phi : \mathcal{S}^{d-1} \to \mathbb{R}$ and $K = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant \phi(u), \quad \forall u \in \mathcal{S}^{d-1}\}$.

Lemma 4 *The set K is convex and compact.*

Proof If K is empty, then it is convex and compact. Assume that K is nonempty. It is closed and convex, as the intersection of closed halfspaces. Let us show that K is bounded, which will end the proof. Since h_K is convex, it is continuous on the interior of its domain (i.e., $\{u \in \mathbb{R}^d : h_K(u) < \infty\}$). Moreover, $h_K(u) \leqslant |u|\phi(u) < \infty$ for all $u \in \mathbb{R}^d$, yielding that h_K is continuous on $\mathbb{R}^d$. Hence, since $\mathcal{S}^{d-1}$ is compact, the restriction of h_K on the $\mathcal{S}^{d-1}$ is bounded. Hence, by Lemma 2, K is bounded.

Lemma 5 *If ϕ is continuous and $x \in \mathbb{R}^d$, then $x \in \mathring{K} \iff \langle u, x\rangle < \phi(u), \forall u \in \mathcal{S}^{d-1}$.*

Proof Let $x \in \mathring{K}$. Then, $B'(x,\eta) \subseteq K$ for some $\eta > 0$. Let $u \in \mathcal{S}^{d-1}$. Then, $x + \eta u \in K$, yielding $\langle u, x + \eta u\rangle \leqslant \phi(u)$. Hence, $\langle u, x\rangle \leqslant \phi(u) - \eta < \phi(u)$ and this has to be true for all $u \in \mathcal{S}^{d-1}$. Now, let $x \in \mathbb{R}^d$ satisfying $\langle u, x\rangle < \phi(u), \forall u \in \mathcal{S}^{d-1}$. The map $u \in \mathcal{S}^{d-1} \mapsto \phi(u) - \langle u, x\rangle$ is continuous and positive on the compact $\mathcal{S}^{d-1}$, hence, there exists $\eta > 0$ such that for all $u \in \mathcal{S}^{d-1}$, $\phi(u) - \langle u, x\rangle \geqslant \eta$. Then, it is easy to verify that $B'(x,\eta) \subseteq K$, yielding $x \in \mathring{K}$.

When a convex set is defined through a collection of linear inequalities indexed by the unit sphere, the support function at a given unit u_0 vector can be interpreted as the value of a semi-infinite linear program. The following lemma states that under a continuity assumption, u_0 needs to lie in the convex cone spanned by the constraints that are active at a point x^* that is a solution of that linear program. Note that when the number of linear constraints is infinite, the existence of active constraints is not granted, as the following example shows.

Let $u_0 \in \mathcal{S}^{d-1}$ and $G = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant 1, \forall u \in \mathcal{S}^{d-1}\backslash\{u_0\}, \langle u_0, x\rangle \leqslant 2\}$. Then, since it is also true that $G = B(0,1)$, the value of the semi-infinite linear program $\max\{\langle u_0, x\rangle : x \in G\}$ is 1, uniquely attained at $x^* = u_0$. Yet, no constraint is active at x^*.

Lemma 6 *Let ϕ be a continuous function on $\mathcal{S}^{d-1}$ and let $K = \{x \in \mathbb{R}^d : \langle u, x\rangle \leqslant \phi(u), \forall u \in \mathcal{S}^{d-1}\}$. Assume that $\mathring{K} \neq \emptyset$. For all $u_0 \in \mathcal{S}^{d-1}$, there exists $x^* \in K$ such that $h_K(u_0) = \langle u_0, x^*\rangle$. Moreover, there exists $I \subseteq \mathcal{S}^{d-1}$ such that*

– $\#I \leqslant d$,
– $\langle u, x^* \rangle = \phi(u), \forall u \in I$,
– $u_0 = \sum_{u \in I} \lambda_u u$, *for some nonnegative numbers* $\lambda_u, u \in I$.

Proof By Lemma 4, K is compact, which grants the existence of x^*, since $K \neq \varnothing$. Let $I^* = \{u \in \mathcal{S}^{d-1} : \langle u, x^* \rangle = \phi(u)\}$ be the set of active constraints at x^* and let us prove that I^* is not empty. The rest will follow using Theorem 2 in [30] (Slater's condition is satisfied since we assume that K has nonempty interior).

If I^* was empty, then

$$\forall u \in \mathcal{S}^{d-1}, \langle u, x^* \rangle < \phi(u). \tag{3}$$

Hence, by Lemma 5, $x^* \in \mathring{K}$: $B(x^*, \eta) \subseteq K$ for some $\eta > 0$. This yields

$$\begin{aligned} h_K(u_0) &\geqslant \langle u_0, x^* + \eta u_0 \rangle \\ &= \langle u_0, x^* \rangle + \eta \\ &> h_K(u_0), \end{aligned}$$

which is a contradiction.

In the next two lemmas, for any map $\zeta : \mathcal{S}^{d-1} \to \mathbb{R}$ and any subset $\mathcal{N} \subseteq \mathcal{S}^{d-1}$, we define $G_\zeta = \{x \in \mathbb{R}^d : \langle u, x \rangle \leqslant \zeta(u), \forall u \in \mathcal{S}^{d-1}\}$ and $G_\zeta^{\mathcal{N}} = \{x \in \mathbb{R}^d : \langle u, x \rangle \leqslant \zeta(u), \forall u \in \mathcal{N}\}$.

Lemma 7 *Let ϕ and $\hat{\phi}$ be two continuous functions on $\mathcal{S}^{d-1}$. Assume that G_ϕ and $G_{\hat{\phi}}$ have nonempty interiors. Let $R > r > 0$ and assume that $B'(0, r) \subseteq G_\phi \subseteq B'(0, R)$. Let $\eta = \max_{u \in \mathcal{S}^{d-1}} |\hat{\phi}(u) - \phi(u)|$. If $\eta < r$, then $d_H(G_{\hat{\phi}}, G_\phi) \leqslant \frac{\eta R}{r} \frac{1 + \eta/r}{1 - \eta/r}$.*

Proof Let $u_0 \in \mathcal{S}^{d-1}$. By Lemma 6, there exist $x \in G_\phi, \hat{x} \in G_{\hat{\phi}}$, $I, \hat{I} \subseteq \mathcal{S}^{d-1}$ with $\#I \leqslant d, \#\hat{I} \leqslant d$, such that $h_{G_\phi}(u_0) = \langle u_0, x \rangle$, $h_{G_{\hat{\phi}}}(u_0) = \langle u_0, \hat{x} \rangle$, $\langle u, x \rangle = \phi(u), \forall u \in I$, $\langle v, \hat{x} \rangle = \hat{\phi}(v), \forall v \in \hat{I}$ and $u_0 = \sum_{u \in I} \lambda_u u = \sum_{v \in \hat{I}} \hat{\lambda}_v v$, for some nonnegative families $(\lambda_u)_{u \in I}, \left(\hat{\lambda}_v\right)_{v \in \hat{I}} \geqslant 0$. Note that necessarily, for all $u \in I$ and $v \in \hat{I}$, $\phi(u) = h_{G_\phi}(u)$ and $\hat{\phi}(v) = h_{G_{\hat{\phi}}}(v)$. Then,

$$\begin{aligned} h_{G_{\hat{\phi}}}(u_0) = h_{G_{\hat{\phi}}}\left(\sum_{u \in I} \lambda_u u\right) &\leqslant \sum_{u \in I} \lambda_u h_{G_{\hat{\phi}}}(u) \leqslant \sum_{u \in I} \lambda_u \hat{\phi}(u) \leqslant \sum_{u \in I} \lambda_u (\phi(u) + \eta) \\ &= \sum_{u \in I} \lambda_u \langle u, x \rangle + \eta \sum_{u \in I} \lambda_u = \langle u_0, x \rangle + \eta \sum_{u \in I} \lambda_u = h_{G_\phi}(u_0) + \eta \sum_{u \in I} \lambda_u. \end{aligned} \tag{4}$$

In a similar fashion, we have that

$$h_{G_\phi}(u_0) \leqslant h_{G_{\hat\phi}}(u_0) + \eta \sum_{v\in\hat I} \hat\lambda_v. \tag{5}$$

By Lemma 2 and since $B'(0,r) \subseteq G_\phi \subseteq B'(0,R)$, $r \leqslant h_{G_\phi}(u) \leqslant R$, for all $u \in \mathcal{S}^{d-1}$, yielding $R \geqslant \langle u_0, x\rangle = \sum_{u\in I} \lambda_u \langle u, x\rangle = \sum_{u\in I} \lambda_u h_{G_\phi}(u) \geqslant r \sum_{u\in I} \lambda_u$. Hence, $\sum_{u\in I} \lambda_u \leqslant \frac{R}{r}$ and by (4),

$$h_{G_{\hat\phi}}(u_0) \leqslant h_{G_\phi}(u_0) + \frac{\eta R}{r}. \tag{6}$$

On the other hand,

$$\sum_{v\in\hat I} \hat\lambda_v \langle v, \hat x\rangle = \langle u_0, \hat x\rangle = h_{G_{\hat\phi}}(u_0) \leqslant h_{G_\phi}(u_0) + \frac{\eta R}{r} \leqslant R + \frac{\eta R}{r}, \tag{7}$$

where the first inequality comes from (6). In addition,

$$\sum_{v\in\hat I} \hat\lambda_v \langle v, \hat x\rangle = \sum_{v\in\hat I} \hat\lambda_v \hat\phi(v) \geqslant \sum_{v\in\hat I} \hat\lambda_v (\phi(v) - \eta) \geqslant \sum_{v\in\hat I} \hat\lambda_v (r - \eta)$$

yielding, together with (7),

$$\sum_{v\in\hat I} \hat\lambda_v \leqslant \frac{R}{r}\frac{1+\eta/r}{1-\eta/r}. \tag{8}$$

Finally, (4), (5) and (8) yield

$$|h_{G_\phi}(u_0) - h_{G_{\hat\phi}}(u_0)| \leqslant \frac{\eta R}{r}\frac{1+\eta/r}{1-\eta/r}. \tag{9}$$

Since (9) is true for any arbitrary $u_0 \in \mathcal{S}^{d-1}$, Lemma 7 is proven, using Lemma 3.

Definition 1 Let $\delta > 0$. A δ-net of the sphere S^{d-1} is a subset $\mathcal{N} \subseteq \mathcal{S}^{d-1}$ such that $\sup_{u\in\mathcal{S}^{d-1}} \inf_{v\in\mathcal{N}} |u - v| \leqslant \delta$.

Lemma 8 *Let $\delta \in (0,1)$ and $\mathcal{N}$ be a δ-net of $\mathcal{S}^{d-1}$. Let ϕ and $\hat\phi : \mathbb{R}^d \to \mathbb{R}$, and assume that ϕ is sublinear. Let $r < R$ be two positive numbers and assume that $B'(0,r) \subseteq G_\phi \subseteq B'(0,R)$. Let $\eta = \max_{u\in\mathcal{N}} |\phi(u) - \hat\phi(u)|$. If $\eta < r$, then $d_H(G_\phi, G_{\hat\phi}^{\mathcal{N}}) \leqslant \frac{\eta R}{r}\frac{1+\eta/r}{1-\eta/r} + \frac{2R\delta}{1-\delta}$.*

Proof Before starting the proof, let us recall the following important property for support functions. If $K \subseteq B'(0, M)$ is a convex set, with $M > 0$, then its support function is M-Lipschitz. By the triangle inequality,

$$d_{\mathsf{H}}(G_\phi, G_{\hat{\phi}}^{\mathcal{N}}) \leqslant d_{\mathsf{H}}(G_\phi, G_\phi^{\mathcal{N}}) + d_{\mathsf{H}}(G_\phi^{\mathcal{N}}, G_{\hat{\phi}}^{\mathcal{N}}). \tag{10}$$

By Proposition 1, $\phi(u) = h_{G_\phi}(u), \forall u \in \mathcal{S}^{d-1}$. Hence, $\phi(u) \leqslant R, \forall u \in \mathcal{S}^{d-1}$. Let $x \in G_\phi^{\mathcal{N}}$ with $x \neq 0$ and let $u = x/|x|$. Then, $|u - u^*| \leqslant \delta$ for some $u^* \in \mathcal{N}$, yielding $|x| = \langle u, x\rangle = \langle u^*, x\rangle + \langle u - u^*, x\rangle \leqslant \phi(u^*) + \delta|x| \leqslant R + \delta|x|$. Hence, $|x| \leqslant \dfrac{R}{1-\delta}$ and $G_\phi^{\mathcal{N}} \subseteq B'(0, R/(1-\delta))$. This entails that $h_{G_\phi^{\mathcal{N}}}$ is $R/(1-\delta)$-Lipschitz. Now, let $u_0 \in \mathcal{S}^{d-1}$. On the one hand, since $G_\phi \subseteq G_\phi^{\mathcal{N}}$, $h_{G_\phi}(u_0) \leqslant h_{G_\phi^{\mathcal{N}}}(u_0)$. On the other hand, if $u^* \in \mathcal{N}$ satisfies $|u_0 - u^*| \leqslant \delta$, then

$$\begin{aligned} h_{G_\phi^{\mathcal{N}}}(u_0) &\leqslant h_{G_\phi^{\mathcal{N}}}(u^*) + \frac{R\delta}{1-\delta} \leqslant \phi(u^*) + \frac{R\delta}{1-\delta} = h_{G_\phi}(u^*) + \frac{R\delta}{1-\delta} \\ &\leqslant h_{G_\phi}(u_0) + R|u_0 - u^*| + \frac{R\delta}{1-\delta} \leqslant h_{G_\phi}(u_0) + R\delta + \frac{R\delta}{1-\delta} \\ &= h_{G_\phi}(u_0) + \frac{R\delta(2-\delta)}{1-\delta} \leqslant h_{G_\phi}(u_0) + \frac{2R\delta}{1-\delta}, \end{aligned}$$

where we used the fact that h_{G_ϕ} is R-Lipschitz. Therefore,

$$d_{\mathsf{H}}(G_\phi, G_{\hat{\phi}}^{\mathcal{N}}) \leqslant \frac{2R\delta}{1-\delta}. \tag{11}$$

Since $B'(0, r) \subseteq G_\phi \subseteq G_\phi^{\mathcal{N}}$, $G_\phi^{\mathcal{N}}$ has nonempty interior. So does $G_{\hat{\phi}}^{\mathcal{N}}$, since it is clear that $B'(0, r-\eta) \subseteq G_{\hat{\phi}}^{\mathcal{N}}$, using the facts that $\phi(u) \geqslant r, \forall u \in \mathcal{S}^{d-1}$, by Lemma 2 and that $\eta < r$. Hence, using similar arguments as in the proof of Lemma 7,

$$d_{\mathsf{H}}(G_\phi^{\mathcal{N}}, G_{\hat{\phi}}^{\mathcal{N}}) \leqslant \frac{\eta R}{r} \frac{1 + \eta/r}{1 - \eta/r}. \tag{12}$$

Thus, (10), (11) and (12) yield the desired result.

Lemma 9 *Let $K \subseteq \mathbb{R}^d$ be a convex set with nonempty interior. Let $A = \{(u, t) \in \mathcal{S}^{d-1} \times \mathbb{R} : (tu + u^\perp) \cap K \neq \varnothing\}$. Then, a pair $(u, t) \in \mathcal{S}^{d-1} \times \mathbb{R}$ is in $\mathring{A}$ if and only if there exists $\eta > 0$ satisfying*

$$(su + u^\perp) \cap \mathring{K} \neq \varnothing, \quad \forall s \in [t - \eta, t + \eta]. \tag{13}$$

Proof Let $(u, t) \in \mathcal{S}^{d-1} \times \mathbb{R}$.

Assume that $(u, t) \in \mathring{A}$. Then, there exists $\eta > 0$ such that $(u, s) \in A$, for all $s \in [t - 2\eta, t + 2\eta]$. Let $s \in [t - 2\eta, t + 2\eta]$. Since $(u, s) \in A$, the affine hyperplane $su + u^\perp$ intersects K. It actually needs to intersect $\mathring{K}$. Indeed, $\mathring{K}$ is also the relative interior of K, since K has nonempty interior. Hence, for the affine hyperplane $su + u^\perp$ to intersect K but not its interior, it has to be a supporting hyperplane of K. This contradicts the fact that K has elements on both sides of $su + u^\perp$.

Now, assume that (u, t) satisfies (13) for some $\eta > 0$. Then, there exists $x \in (tu + u^\perp) \cap \mathring{K} \neq \emptyset$. In particular, $x \in \mathring{K}$. Without log of generality, assume that η is small enough so $B(x, \eta) \subseteq \mathring{K}$. Let $\delta = \eta/(1 + |x|)$ and $(v, s) \in \mathcal{S}^{d-1} \times \mathbb{R}$ with both $|v - u| \leqslant \delta$ and $|s - t| \leqslant \delta$. Since $x \in \langle v, x\rangle v + v^\perp$, the affine hyperplane $sv + v^\perp$ intersects $B'(x, \eta)$ if and only if $|s - \langle v, x\rangle| \leqslant \eta$, which holds by our choice of δ.

Lemma 10 *Let k be a positive integer and K be a compact and convex set in $\mathbb{R}^k$ such that $0 \in \mathring{K}$. Let $u \in \mathcal{S}^{k-1}$ and let $(u_n)_{n\geqslant 1}$ a sequence of unit vectors in $\mathbb{R}^k$ that converges to u. Let $(x_n)_{n\geqslant 1}$ be a sequence in $\mathbb{R}^k$ that converges to zero and $(U_n)_{n\geqslant 1}$ be a sequence of isometries in $\mathbb{R}^k$ that converges to the identity. Then, as $n \to 0$,*

1. $\mathsf{Vol}_{k-1}\left(\left((K + x_n) \cap u_n^\perp\right) \triangle (K \cap u_n^\perp)\right) \longrightarrow 0$;
2. $\mathsf{Vol}_{k-1}\left(\left(U_n(K) \cap u_n^\perp\right) \triangle (K \cap u_n^\perp)\right) \longrightarrow 0$.

Proof For $u \in \mathcal{S}^{k-1}$, set $p_K(u) = \max\{\lambda \geqslant 0 : \lambda u \in K\}$. This is the (multiplicative) inverse of the gauge of K. By [43, Section 1.7], the gauge function g_K (defined on $\mathbb{R}^d$) of K satisfies the following properties:

- Since $0 \in \mathring{K}$, there exists $M > 0$ with $g_K(u) \leqslant M$, for all $u \in \mathcal{S}^{d-1}$;
- Since K is bounded, there exists $m > 0$ with $g_K(u) \geqslant m$, for all $u \in \mathcal{S}^{d-1}$;
- g_K is subadditive and positively homogeneous, so it is M-Lipschitz on $\mathcal{S}^{d-1}$:

$$|g_K(u) - g_K(v)| \leqslant M|u - v|, \forall u, v \in \mathcal{S}^{d-1}.$$

As a consequence, $p_K(u) \geqslant M^{-1}$ for all $u \in \mathcal{S}^{k-1}$ and p_K is Lipschitz on $\mathcal{S}^{d-1}$, with Lipschitz constant $L = M/m^2$.

First statement of the lemma: For $n \geqslant 1$, write $x_n = \lambda_n u_n + v_n$, with $\lambda_n \in \mathbb{R}$ and $v_n \in u_n^\perp$ and denote by $K_n = K + \lambda_n u_n$. Recall that $(A, B) \mapsto \mathsf{Vol}_{k-1}(A \triangle B)$ is a pseudo-metric on the class of compact subsets of $u_n^\perp$, so it satisfies the triangle inequality:

$$\begin{aligned}\mathsf{Vol}_{k-1}&\left(\left((K + x_n) \cap u_n^\perp\right) \triangle (K \cap u_n^\perp)\right) \leqslant \\ &\mathsf{Vol}_{k-1}\left(\left((K_n + v_n) \cap u_n^\perp\right) \triangle (K_n \cap u_n^\perp)\right) \\ &\quad + \mathsf{Vol}_{k-1}\left(\left((K + \lambda_n u_n) \cap u_n^\perp\right) \triangle (K \cap u_n^\perp)\right). \end{aligned} \tag{14}$$

Since $v_n \in u_n^\perp$, the first term on the right hand side of (14) is equal to

$$\mathsf{Vol}_{k-1}\left(\left((K_n \cap u_n^\perp) + v_n\right) \triangle (K_n \cap u_n^\perp)\right). \tag{15}$$

It is easy to see that $d_\mathsf{H}\left((K_n \cap u_n^\perp) + v_n, K_n \cap u_n^\perp\right) \leqslant |v_n|$, which is less than one if n is large enough. Hence, using the same argument as in the proof of Lemma 1 in [5], there is a positive constant C that does not depend on

n such that (15) is bounded from above by $Cd_{\mathsf{H}}\left((K_n \cap u_n^{\perp}) + v_n, K_n \cap u_n^{\perp}\right)$. Therefore, the first term of the right hand side of (14) goes to zero as n goes to infinity. Let $n \geqslant 1$ be large enough so $\lambda_n < m$. Set $\alpha_n = p_K(u_n)$ and $\beta_n = p_K(-u_n)$. Suppose that $\lambda_n \geqslant 0$ (the case $\lambda_n < 0$ would be handled similarly). Then, by convexity of K,

$$\frac{\alpha_n}{\alpha_n + \lambda_n}(K + \lambda_n u_n) \subseteq K \subseteq \frac{\beta_n}{\beta_n - \lambda_n}(K + \lambda_n u_n). \tag{16}$$

Since $0 \in K$, it is true that for all $\lambda \in \mathbb{R}$ and $v \in \mathcal{S}^{k-1}$, $(\lambda K) \cap v^{\perp} = \lambda(K \cap v^{\perp})$. Using this fact together with (16) yields

$$\begin{aligned}\left((K + \lambda_n u_n) \cap u_n^{\perp}\right) \triangle (K \cap u_n^{\perp}) \subseteq \\ \left(\left(\frac{\alpha_n + \lambda_n}{\alpha_n}(K \cap u_n^{\perp})\right) \setminus (K \cap u_n^{\perp})\right) \\ \cup \left((K \cap u_n^{\perp}) \setminus \left(\frac{\beta_n - \lambda_n}{\beta_n}(K \cap u_n^{\perp})\right)\right).\end{aligned} \tag{17}$$

Since $0 \in K$, the volume of the set in the right hand side of (17) is bounded from above by

$$\left(\left(\frac{\alpha_n + \lambda_n}{\alpha_n}\right)^{k-1} - 1 + \left(\frac{\beta_n - \lambda_n}{\beta_n}\right)^{k-1} - 1\right) \mathsf{Vol}_{k-1}(K \cap u_n^{\perp}),$$

which goes to zero as n goes to infinity, since K is bounded and α_n and β_n are bounded away from zero (they are at not smaller than m). This ends the proof of the first statement of the lemma.

Second statement of the lemma: Let L be the corresponding Lipschitz constant. Let $t_n = \|U_n - I_k\|$, where I_k is the identity map in $\mathbb{R}^k$ and we define the norm of any linear map $A : \mathbb{R}^k \to \mathbb{R}^k$ by $\|A\| = \max_{v \in \mathcal{S}^{k-1}} |A(v)|$. Then, since U_n converges to the identity, t_n goes to zero as n goes to infinity. Define $c_n = \dfrac{M^{-1}}{M^{-1} + Lt_n}$. Note that $0 \leqslant c_n \leqslant 1$. Then, let us show that for all $n \geqslant 1$,

$$c_n K \subseteq U_n(K). \tag{18}$$

Let $x \in K$ and set $y = U_n^{-1}(c_n x)$. If $x = 0$, then $y = 0$ yielding $y \in K$ by assumption, which proves (18). If $x \neq 0$, then $y \neq 0$ and let $v = y/|y|$. In order to prove that $y \in K$, it is enough to show that

$$|y| \leqslant p_K(v). \tag{19}$$

Since U_n is an isometry, $|y| = c_n|x|$ and since $x \in K$, $|x| \leqslant p_K(x/|x|)$. Therefore,

$$\begin{aligned}|y| &= c_n|x| \leqslant c_n p_K(x/|x|) = c_n p_K(U_n(v)) \leqslant c_n p_K(v) + c_n L|U_n(v) - v| \\ &\leqslant c_n p_K(v) + c_n L t_n = p_K(v) + c_n L t_n - (1 - c_n)p_K(v) \\ &\leqslant p_K(v) + c_n t_n L - (1 - c_n)M^{-1} \\ &= p_K(v),\end{aligned}$$

by definition of c_n. This proves (19) and hence, (18). As a consequence, since $0 \in K$, $c_n(K \cap u_n^\perp) = (c_n K) \cap u_n^\perp \subseteq U_n(K) \cap u_n^\perp$, yielding

$$\begin{aligned}\mathsf{Vol}_{k-1}\left((K \cap u_n^\perp)\setminus\left(U_n(K) \cap u_n^\perp\right)\right) &\leqslant \mathsf{Vol}_{k-1}\left((K \cap u_n^\perp)\setminus\left(c_n(K \cap u_n^\perp)\right)\right)\\ &\leqslant (1-c_n^{k-1})\mathsf{Vol}_{k-1}(K \cap u_n^\perp),\end{aligned}$$

which goes to zero as $n \to \infty$, since K is bounded and $c_n \to 1$. In a similar fashion, we prove that $\mathsf{Vol}_{k-1}\left(\left(U_n(K) \cap u_n^\perp\right)\setminus(K \cap u_n^\perp)\right)$ also goes to zero as $n \to \infty$, which ends the proof of the second statement of the lemma.

Lemma 11 *Let M be a positive integer and let $U_1, \dots, U_M$ be i.i.d. uniform random variables on $\mathcal{S}^{d-1}$. Let $\delta \in (0,1]$ and let $\mathcal{C}$ be the event satisfied when the collection $\{U_1, \dots, U_M\}$ is a δ-net of the sphere (see Definition 1). Then, the complement $\mathcal{C}^\complement$ of $\mathcal{C}$ satisfies*

$$\mathbb{P}[\mathcal{C}^\complement] \leqslant M\left(1-\left(\frac{\delta}{4}\right)^{(d-1)/2}\right)^M \leqslant 6^d \exp\left(-\frac{M\delta^{d-1}}{2d8^{(d-1)/2}} + d\log\left(\frac{1}{\delta}\right)\right).$$

Proof Let $\mathcal{N}$ be a $(\delta/2)$-net of $\mathcal{S}^{d-1}$. By a simple volume argument, it is possible to choose $\mathcal{N}$ satisfying $\#\mathcal{N} \leqslant (6/\delta)^d$, which we assume in the sequel. If $\mathcal{C}$ is not satisfied, there exists $u \in \mathcal{S}^{d-1}$ for which $|u - U_j| > \delta$, for all $j = 1, \dots, M$. Hence, if $v \in \mathcal{N}$ is such that $|u - v| \leqslant \delta/2$, one has, for all $j = 1, \dots, M$, by the triangle inequality, $|v-U_j| \geqslant |u-U_j|-|u-v| \geqslant \delta-\delta/2 \geqslant \delta/2$. Therefore, using the union bound and mutual independence of the U_j's,

$$\mathbb{P}[\mathcal{C}^\complement] \leqslant \mathbb{P}\left[\exists v \in \mathcal{N}, |v-U_j| > \frac{\delta}{2}, \forall j = 1, \dots, M\right] \leqslant \sum_{v\in\mathcal{N}} \mathbb{P}\left[|v-U_1| > \frac{\delta}{2}\right]^M. \tag{20}$$

For any $v \in \mathcal{S}^{d-1}$, $\mathbb{P}\left[|v-U_1| \leqslant \frac{\delta}{2}\right]$ is the ratio of the surface area of a spherical cap of the unit sphere and the total surface area of the unit sphere. The height of this cap is $h = \delta^2/8 < 1$. Then,

$$\mathbb{P}\left[|v-U_1| \leqslant \frac{\delta}{2}\right] = \frac{1}{2} I_{2h-h^2}\left(\frac{d-1}{2}, \frac{1}{2}\right), \tag{21}$$

where $I_x(a,b) = \dfrac{\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t}$, for $x \in [0,1]$ and $a, b > 0$ (see, e.g., [27]). If $b \leqslant 1$, one has $\displaystyle\int_0^x t^{a-1}(1-t)^{b-1}\,\mathrm{d}t \geqslant \int_0^x t^{a-1}\,\mathrm{d}t = \frac{x^a}{a}$ and $a\displaystyle\int_0^1 t^{a-1}(1-t)^{b-1}\,\mathrm{d}t = (a+b)\int_0^1 t^a(1-t)^{b-1}\,\mathrm{d}t \leqslant (a+b)\int_0^1 (1-t)^{b-1}\,\mathrm{d}t = \frac{a+b}{b}$. Hence, $I_x(a,b) \geqslant \dfrac{b}{a+b}x^a$ and (21) yields, with $x = 2h - h^2$, $a = \frac{d-1}{2}$ and $b = 1/2$, that $\mathbb{P}\left[|v-U_1|\leqslant\frac{\delta}{2}\right]\geqslant\frac{1}{2d}(2h-h^2)^{(d-1)/2}$. Since $h<1$, $2h-h^2\geqslant h=\delta^2/8$, hence,

$$\mathbb{P}\left[|v-U_1|\leqslant\frac{\delta}{2}\right]\geqslant\frac{\delta^{d-1}}{2d8^{(d-1)/2}}. \tag{22}$$

Together with (22), (20) implies

$$\mathbb{P}[\mathcal{C}^{\complement}]\leqslant\#\mathcal{N}\left(1-\frac{\delta^{d-1}}{2d8^{(d-1)/2}}\right)^M\leqslant 6^d\exp\left(-\frac{M\delta^{d-1}}{2d8^{(d-1)/2}}+d\log\left(\frac{1}{\delta}\right)\right),$$

which ends the proof of Lemma 11.

4.2 Preliminary lemmas for empirical and population quantiles

Lemma 12 *Let μ satisfy Assumption 1. Then, the map $u\in\mathcal{S}^{d-1}\mapsto q_u^\sharp$ is continuous.*

Proof For notation's sake, we write q_u instead of $q_u^\sharp$ in the sequel of the proof.

Step 1: Denote by $\Phi(u,t)=\mathbb{P}[\langle u,X\rangle\leqslant t], u\in\mathcal{S}^{d-1}, t\in\mathbb{R}$. We first show that Φ is continuous $A=\left\{(u,t)\in\mathcal{S}^{d-1}\times\mathbb{R}: q_u-\varepsilon<t<q_u+\varepsilon\right\}$.

Let $(u,t)\in A$ and $(u_p,t_p)_{p\geqslant 1}$ be a sequence in A that converges to (u,t) as p goes to infinity. Let η be an arbitrary positive number. We show that if p is large enough, then $|\Phi(u_p,t_p)-\Phi(u,t)|\leqslant 2\eta$, which will prove our statement. First, note that $|\Phi(u_p,t_p)-\Phi(u,t)|\leqslant\mu\left(H_{u,t}\triangle H_{u_p,t_p}\right)$. Let $R>0$ satisfy $\mathbb{P}[|X|>R]\leqslant\eta$. Then,

$$\begin{aligned}\mu\left(H_{u,t}\triangle H_{u_p,t_p}\right)&\leqslant\mu\left(B(0,R)\cap(H_{u,t}\triangle H_{u_p,t_p})\right)+\mu(R^d\backslash B(0,R))\\&\leqslant\mu\left(B(0,R)\cap(H_{u,t}\triangle H_{u_p,t_p})\right)+\eta.\end{aligned}$$

It is easy to check that

$$B(0,R)\cap(H_{u,t}\triangle H_{u_p,t_p})\subseteq\left(H_{u,t_p+R|u_p-u|}\backslash H_{u,t}\right)\cup\left(H_{u,t}\backslash H_{u,t_p-R|u_p-u|}\right),$$

which entails

$$\begin{aligned}&\mu\left(B(0,R)\cap(H_{u,t}\triangle H_{u_p,t_p})\right)\\&\qquad\leqslant|F_u\left(t_p+R|u_p-u|\right)-F_u(t)|+|F_u(t)-F_u\left(t_p-R|u_p-u|\right)|.\end{aligned} \tag{23}$$

Since $(u_p,t_p)\xrightarrow[p\to\infty]{}(u,t)$ and $q_u-\varepsilon<t<q_u+\varepsilon$, one has $q_u-\varepsilon\leqslant t_p-R|u_p-u|\leqslant t_p+R|u_p-u|\leqslant q_u+\varepsilon$ for all large enough p. Hence, since F_u is continuous on $[q_u-\varepsilon,q_u+\varepsilon]$, (23) implies that $\mu\left(B(0,R)\cap(H_{u,t}\triangle H_{u_p,t_p})\right)\leqslant\eta$ if p is large enough, which ends the the proof of the continuity of Φ on A.

Step 2: Let $u \in \mathcal{S}^{d-1}$ and $(u_p)_{p \geqslant 1}$ be a sequence of unit vectors converging to u as p goes to infinity. Let us show that q_{u_p} converges to q_u. If this was not the case, there would be a positive number η and an increasing sequence of positive integers $(p_k)_{k \geqslant 1}$ satisfying $|q_{u_{p_k}} - q_u| \geqslant \eta, \forall k \geqslant 1$. Let us assume that $q_{u_{p_k}} \geqslant q_u + \eta$ for an infinite number of indices $k \geqslant 1$. The case when $q_{u_{p_k}} \leqslant q_u - \eta$ for an infinite number of indices $k \geqslant 1$ would be handled similarly. For the sake of notation, we renumber the sequence and assume that for $k \geqslant 1$, $q_{u_k} \geqslant q_u + \eta$. Without loss of generality, assume that $\eta < \varepsilon$. Hence, for all $k \geqslant 1$,

$$\begin{aligned} 1 - \alpha = F_{u_k}(q_{u_k}) &\geqslant F_{u_k}(q_u + \eta) \\ &= F_u(q_u + \eta) + \Phi(u_k, q_u + \eta) - \Phi(u, q_u + \eta) \\ &\geqslant F_u(q_u) + L\eta + \Phi(u_k, q_u + \eta) - \Phi(u, q_u + \eta) \\ &= 1 - \alpha + L\eta + \Phi(u_k, q_u + \eta) - \Phi(u, q_u + \eta). \end{aligned} \tag{24}$$

The fact that $F_v(q_v) = 1 - \alpha, \forall v \in \mathcal{S}^{d-1}$, is a consequence of the continuity and strict monotony of F_v in a neighborhood of q_v, for all $v \in \mathcal{S}^{d-1}$. Since $\eta < \varepsilon$, $(u, q_u + \eta) \in A$, so by the first part of the proof, $\Phi(u_k, q_u + \eta) - \Phi(u, q_u + \eta) \xrightarrow[k \to \infty]{} 0$. Thus, by letting k grow to infinity in (24), we get that $L\eta \leqslant 0$, which is a contradiction. Hence, we have proved that $q_{u_p} \xrightarrow[p \to \infty]{} q_u$, which ends the proof.

Lemma 13 *Let μ be a probability measure on $\mathbb{R}^d$ that satisfies either Assumption 3 or 4 and let K be its support. For $u \in \mathcal{S}^{d-1}$, let f_u and F_u be, respectively, the density and the cumulative distribution function of $\langle u, X \rangle$, where X is a random variable with distribution μ. Let $A = \{(u, t) \in \mathcal{S}^{d-1} \times \mathbb{R} : (tu + u^{\perp}) \cap K \neq \varnothing\}$. Define $\phi(u, t) = f_u(t)$ and $\Phi(u, t) = F_u(t)$, for all $(u, t) \in A$. Then,*

- *ϕ and Φ are continuous on $\mathring{A}$;*
- *$\forall (u, t) \in \mathring{A}$, $\phi(u, t) > 0$ and $0 < \Phi(u, t) < 1$.*

Proof Note $\Phi(u, t) = \int_{-\infty}^{t} \phi(u, s)\, \mathrm{d}s$, for all $(u, t) \in A$, where we set $\phi(u, s)$ to zero if $(u, s) \notin A$. Hence, by dominated convergence, continuity of ϕ will automatically yield that of Φ. Let $(u, t) \in A$ and consider an arbitrary sequence $(u_n, t_n)_{n \geqslant 1}$ of elements of A that converges to (u, t).

Let μ satisfy Assumption 3. In this case, the second statement is trivial since f is continuous and positive everywhere. Hence, we only prove the first statement. Let $\epsilon > 0$ and $R > 0$. For all $n \geqslant 1$,

$$\begin{aligned}
|\phi(u_n, t_n) - \phi(u, t)| &= \left| \int_{u_n^\perp} f(t_n u_n + v)\,\mathrm{d}v - \int_{u^\perp} f(tu + v)\,\mathrm{d}v \right| \\
&\leqslant \left| \int_{\substack{v \in u_n^\perp: \\ |v| \leqslant R}} f(t_n u_n + v)\,\mathrm{d}v - \int_{\substack{v \in u^\perp: \\ |v| \leqslant R}} f(tu + v)\,\mathrm{d}v \right| \\
&\qquad + \int_{\substack{v \in u_n^\perp: \\ |v| > R}} f(t_n u_n + v)\,\mathrm{d}v + \int_{\substack{v \in u^\perp: \\ |v| > R}} f(tu + v)\,\mathrm{d}v \\
&\leqslant \left| \int_{\substack{v \in u_n^\perp: \\ |v| \leqslant R}} f(t_n u_n + v)\,\mathrm{d}v - \int_{\substack{v \in u^\perp: \\ |v| \leqslant R}} f(tu + v)\,\mathrm{d}v \right| \\
&\qquad + C \int_{\substack{v \in u_n^\perp: \\ |v| > R}} (1 + |v|)^{-\nu}\,\mathrm{d}v + \int_{\substack{v \in u^\perp: \\ |v| > R}} (1 + |v|)^{-\nu}\,\mathrm{d}v.
\end{aligned} \tag{25}$$

For $n \geqslant 1$, let U_n be an isometry in $\mathbb{R}^d$ such that $U_n(u_n) = u$. Then, the first term in (25) can also be written as $\left| \int_{\substack{v \in u^\perp: \\ |v| \leqslant R}} \left(f(t_n u_n + U_n^{-1}(v)) - f(tu + v) \right) \mathrm{d}v \right|$, which converges to zero by dominated convergence. Hence, for large n, the first term in (25) is smaller than ϵ.

Using polar coordinates, both the second and third terms in (25) can be rewritten as $C' \int_R^\infty x^{d-2}(1 + x)^{-\nu}\,\mathrm{d}x$, for some positive constant C' that does not depend on n or R. Hence, both the second and third terms in (25) are bounded from above by $C'' R^{-(\nu - d + 1)}$, for some positive constant C'' that does not depend on R or n. Hence, if R was chosen large enough, both these terms are smaller than ϵ. Finally, we have proved that $\phi(u_n, t_n) \to \phi(u, t)$, as $n \to \infty$.

Let μ satisfy Assumption 4. Let $\epsilon > 0$. For $n \geqslant 1$, write

$$\begin{aligned}
|\phi(u_n, t_n) - \phi(u, t)| \leqslant & \int_{u_n^\perp} |f(t_n u_n + v) - f(tu + v)|\,\mathrm{d}v \\
& + \left| \int_{u_n^\perp} f(tu + v)\,\mathrm{d}v - \int_{u^\perp} f(tu + v)\,\mathrm{d}v \right|.
\end{aligned} \tag{26}$$

Let $B_n = \{v \in u_n^\perp : t_n u_n + v \in K\}$ and $D_n = \{v \in u_n^\perp : tu + v \in K\}$. The first integral in (26) can be decomposed as

$$\begin{aligned}
& \int_{B_n \cap D_n} |f(t_n u_n + v) - f(tu + v)|\,\mathrm{d}v \\
& \qquad + \int_{B_n \triangle D_n} |f(t_n u_n + v) - f(tu + v)|\,\mathrm{d}v.
\end{aligned} \tag{27}$$

Recall that f is uniformly continuous on K and $\mathsf{Vol}_{d-1}(B_n \cap D_n)$ is bounded uniformly in n, by boundedness of K. Hence, if n is large enough, the first integral in (27) is smaller than ϵ. For the second integral, since f is uniformly continuous on the bounded set K and vanishes everywhere else, it is bounded and the integral is bounded from above by $(\sup_K f)\mathsf{Vol}_{d-1}(B_n \triangle D_n)$. The latter converges to zero as n goes to infinity, thanks to Lemma 10. Hence, it becomes smaller than ϵ if n is large enough, so the first term in (26) is at most 2ε for large values of n. For $n \geqslant 1$, let U_n be an isometry in $\mathbb{R}^d$ such that $U_n(u_n) = u$ and such that U_n converges to the identity, as n goes to infinity. Then, the second term in the right hand side of (26) can be written as

$$\left| \int_{u^\perp} \left(f(tu + U_n^{-1}(v)) - f(tu+v) \right) \mathrm{d}v \right|. \tag{28}$$

Let $K_u = (K - tu) \cap u^\perp$ and $K_u^n = \left(U_n^{-1}(K - tu)\right) \cap u^\perp$. Since the integrand vanishes outside of $K_u \cup K_u^{(n)}$, the integral inside the absolute value in (28) can be decomposed as the sum of two integrals: One on $K_u \cap K_u^n$ and the other on $K_u \triangle K_u^n$. Since U_n converges to the identity as n goes to infinity, $U_n^{-1}(v) \to v$ as $n \to \infty$, for all $v \in \mathbb{R}^d$. Since f is uniformly continuous on K and K is bounded, f is bounded. Hence, by dominated convergence, uniform continuity of f on K together with the fact that $\mathsf{Vol}_{d-1}(K_u \cap K_u^n)$ is bounded uniformly in n implies that the first term goes to zero as $n \to \infty$. Since $f(x) = 0$ for $x \notin K$, f is bounded on $\mathbb{R}^d$. Hence, by Lemma 10, the second term goes to zero as $n \to \infty$, since U_n converges to the identity. This ends the proof of the first statement of the lemma.

For the second statement, first note that K needs to have a nonempty interior. Otherwise, since it is convex, it would be included in a hyperplane, i.e., there would exist $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$ such that $\langle u, x \rangle = t, \forall x \in K$. Hence, $\langle u, X \rangle = t$ almost surely, which contradicts the fact that X has a density with respect to the Lebesgue measure in $\mathbb{R}^d$. Let $(u,t) \in \mathring{A}$. By Lemma 9, there exists $\eta > 0$ such that both $(t+\eta)u + u^\perp$ and $(t-\eta)u + u^\perp$ intersect $\mathring{K}$. Hence, by convexity of $\mathring{K}$, $(su + u^\perp) \cap \mathring{K} \neq \varnothing$, yielding that the $(d-1)$-dimensional Lebesgue measure of $(su + u^\perp) \cap K$ needs to be positive, for all $s \in [t - \eta/2, t + \eta/2]$. Therefore, f_u is positive on this interval, yielding $\phi(u,t) > 0$ and $0 < \Phi(u,t) < 1$.

Lemma 14 *Let μ be a probability measure on $\mathbb{R}^d$ that satisfies either Assumption 3 or 4 and let X be a random variable with distribution μ. Denote by F_u the cumulative distribution function of $\langle u, X \rangle$. Let $\beta \in (0,1)$. For $u \in \mathcal{S}^{d-1}$, let q_u be the β-quantile of $\langle u, X \rangle$, defined as in Lemma 12 (with $\beta = 1 - \alpha$). Then,*

- *For all $u \in \mathcal{S}^{d-1}$, q_u is the unique real number t that satisfies $F_u(t) = \beta$;*
- *The map $u \in \mathcal{S}^{d-1} \mapsto q_u$ is continuous.*

Proof Let K be the support of μ and let $A = \{(u,t) \in \mathcal{S}^{d-1} \times \mathbb{R} : (tu + u^\perp) \cap K \neq \varnothing\}$.

Let $u \in \mathcal{S}^{d-1}$. Since μ is absolutely continuous with respect to the Lebesgue measure, so is the distribution of $\langle u, X\rangle$. Hence, F_u is continuous on $\mathbb{R}$, which yields that $F_u(q_u) = \beta$. In addition, if f_u is the density of $\langle u, X\rangle$, then f_u is positive in a neighborhood of q_u. Indeed, since K is convex, the support of f_u is an interval. Since $F_u(q_u) = \beta \in (0,1)$ and F_u is continuous, there is a neighborhood of q_u on which $F_u(t) \in (0,1)$, i.e., there is a neighborhood of q_u that is included in the support of f_u. In particular, F_u is strictly increasing on this neighborhood, which shows the uniqueness of q_u.

Let $u \in \mathcal{S}^{d-1}$ and let $(u_n)_{n\geqslant 1}$ be an arbitrary sequence of unit vectors that converges to u. Suppose that q_{u_n} does not converge to q_u. Then, there exists $\eta > 0$ and a subsequence of u_n (renamed u_n after renumbering) such that $|q_{u_n} - q_u| \geqslant \eta$, for all $n \geqslant 1$. Assume that for an infinite number of indices n, $q_{u_n} \geqslant q_u + \eta$. The case when $q_{u_n} \leqslant q_u - \eta$ for an infinite number of indices n would be handled similarly. Thus, up to renumbering the sequence again, assume that $q_{u_n} \geqslant q_u + \eta$, for all $n \geqslant 1$. By a similar argument as in the end of the proof of Lemma 13, for all $(u,t) \in A$, $(u,t) \in \mathring{A}$ if and only if $0 < F_u(t) < 1$. Hence, $(u, q_u) \in \mathring{A}$. Hence, there exists $\xi > 0$ such that $(v,t) \in \mathring{A}$ for all $v \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$ with $|v - u| \leqslant \xi$ and $|q_u - t| \leqslant \xi$. By Lemma 13, since ϕ is continuous and positive on $\mathring{A}$, there is a positive constant c such that $\phi(v,t) \geqslant c > 0$ for all $(v,t) \in \mathcal{S}^{d-1} \times \mathbb{R}$ with $|v - u| \leqslant \xi$ and $|q_u - t| \leqslant \xi$. Assume that $\xi \leqslant \eta$, without loss of generality. Then,

$$\begin{aligned}
\beta = F_{u_n}(q_{u_n}) &= \Phi(u_n, q_{u_n}) \geqslant \Phi(u_n, q_u + \eta) \geqslant \Phi(u_n, q_u + \xi) \\
&= \Phi(u_n, q_u) + \int_0^\xi \phi(u_n, t)\,\mathrm{d}t \\
&\geqslant \Phi(u_n, q_u) + c\xi \to \beta + c\xi,
\end{aligned}$$

as n goes to infinity. This is a contradiction, since $\beta + c\xi > \beta$. Hence, q_{u_n} needs to converge to q_u as $n \to \infty$ and Lemma 14 is proven.

Lemma 15 *Let μ satisfy Assumption 1. Then, for all $n \geqslant 1$ and $z \in \mathbb{R}$ with $\frac{10\sqrt{5(d+1)}}{L\sqrt{n}} \leqslant z < \varepsilon$,*

$$\mathbb{P}\left[\sup_{u\in\mathcal{S}^{d-1}} |\hat{q}_u - q_u^\sharp| \leqslant z\right] \geqslant 1 - A\exp\left(-L^2z^2n/2 + 10\sqrt{5(d+1)}Lz\sqrt{n}\right),$$

where $A = e^{-250(d+1)}$.

Proof Let $\mathcal{C}_0 = \{(u,t) \in \mathcal{S}^{d-1} \times \mathbb{R} : q_u^\sharp - \varepsilon \leqslant t \leqslant q_u^\sharp + \varepsilon\}$ and $\widetilde{\mathcal{C}}_0 = \{(u,t) \in \mathcal{C}_0 : u \in \mathbb{Q}^{d-1} \times \mathbb{R}, t \in \mathbb{Q}\}$. Denote by $\mathcal{H}_0 = \{H_{u,t} : (u,t) \in \mathcal{C}_0\}$ and $\widetilde{\mathcal{H}}_0 = \{H_{u,t} : (u,t) \in \widetilde{\mathcal{C}}_0\}$.

Step 1: We first show that

$$\sup_{H\in\mathcal{H}_0}|\mu_n(H)-\mu(H)| = \sup_{H\in\widetilde{\mathcal{H}}_0}|\mu_n(H)-\mu(H)| \quad \text{almost surely.} \tag{29}$$

If $(u,t)\in\mathcal{S}^{d-1}$, denote by $\hat{F}_u(t)=\mu_n(H_{u,t})$, i.e., the empirical cumulative distribution function of $\langle u,X\rangle$. Then,

$$\begin{aligned}\sup_{H\in\mathcal{H}_0}|\mu_n(H)-\mu(H)| &= \sup_{(u,t)\in\mathcal{C}_0}|\hat{F}_u(t)-F_u(t)| \\ &= \max\left(\sup_{(u,t)\in\mathcal{C}_0}(\hat{F}_u(t)-F_u(t)), \sup_{(u,t)\in\mathcal{C}_0}(F_u(t)-\hat{F}_u(t))\right).\end{aligned}$$

Hence, it suffices to prove that $\sup_{(u,t)\in\mathcal{C}_0}(F_u(t)-\hat{F}_u(t)) = \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}(F_u(t)-\hat{F}_u(t))$ and that $\sup_{(u,t)\in\mathcal{C}_0}(\hat{F}_u(t)-F_u(t)) = \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}(\hat{F}_u(t)-F_u(t))$. These two statements follow from the fact that the supremum of a lower semicontinuous function on a given set coincides with its supremum on a dense subset. First, note that $\widetilde{\mathcal{C}}_0$ is dense in $\mathcal{C}_0$. Second, $(u,t)\mapsto\hat{F}_u(t)$ is the average of indicator functions of closed sets, which are all upper semicontinuous. Hence, $(u,t)\mapsto\hat{F}_u(t)$ is upper semicontinuous, yielding that $(u,t)\mapsto-\hat{F}_u(t)$ is lower semicontinuous. Since $(u,t)\in\mathcal{C}_0\mapsto\hat{F}_u(t)$ is continuous on $\mathcal{C}_0$, as proved in Step 1 of the proof of Lemma 12, $(u,t)\in\mathcal{C}_0\mapsto F_u(t)-\hat{F}_u(t)$ is lower semicontinuous on $\mathcal{C}_0$. Hence, $\sup_{(u,t)\in\mathcal{C}_0}(F_u(t)-\hat{F}_u(t)) = \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}(F_u(t)-\hat{F}_u(t))$. Now, note that for all $u\in\mathcal{S}^{d-1}$, continuity of F_u on the segment $[q_u^\sharp-\varepsilon, q_u^\sharp+\varepsilon]$ ensures that $\sup_{q_u^\sharp-\varepsilon<t<q_u^\sharp+\varepsilon}\hat{F}_u(t)-F_u(t) = \sup_{q_u^\sharp-\varepsilon<t<q_u^\sharp+\varepsilon}\hat{G}_u(t)-F_u(t)$, where $\hat{G}_u(t)=\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\langle u,X_i\rangle<t}$. Now, the function $(u,t)\in\mathcal{C}_0\mapsto\hat{G}_u(t)-F_u(t)$ is lower semicontinuous on $\mathcal{C}_0$, yielding $\sup_{(u,t)\in\mathcal{C}_0}\hat{G}_u(t)-F_u(t) = \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}\hat{G}_u(t)-F_u(t)$. As a consequence,

$$\begin{aligned}\sup_{(u,t)\in\mathcal{C}_0}\hat{F}_u(t)-F_u(t) &= \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}\hat{G}_u(t)-F_u(t) \\ &\leqslant \sup_{(u,t)\in\widetilde{\mathcal{C}}_0}\hat{F}_u(t)-F_u(t) \\ &\leqslant \sup_{(u,t)\in\mathcal{C}_0}\hat{F}_u(t)-F_u(t),\end{aligned}$$

where the second inequality comes from the fact that $G_u(t)\leqslant F_u(t)$, for all $u\in\mathcal{S}^{d-1}$ and $t\in\mathbb{R}$. This yields the second statement and proves (29). In particular, the random variable $\sup_{(u,t)\in\mathcal{C}_0}|\hat{F}_u(t)-F_u(t)|$ is measurable and the probability term in the statement of the lemma is well defined.

Step 2: Let $u \in \mathcal{S}^{d-1}$. By definition of $\hat{q}_u$, the following holds for all $t \in \mathbb{R}$, where, as we recall, $H_{-u,-t}$ is the halfspace $H = \{x \in \mathbb{R}^d : \langle u, x\rangle \geqslant t\}$:

– If $t < \hat{q}_u$, then $\mu_n(H_{-u,-t}) \geqslant \alpha$,
– If $t > \hat{q}_u$, then $\mu_n(H_{-u,-t}) < \alpha$.

Assume that for some $u \in \mathcal{S}^{d-1}$, $|\hat{q}_u - q_u^\sharp| > z$. Then, either $\hat{q}_u > q_u^\sharp + z$ or $\hat{q}_u < q_u^\sharp - z$. If $\hat{q}_u > q_u^\sharp + z$, let $H = H_{-u,-(q_u^\sharp+z)} \in \mathcal{H}_0$. Then, $\mu_n(H) \geqslant \alpha$. Hence, by Assumption 1, $\mu(H) = \mathbb{P}[\langle u, X\rangle \geqslant q_u^\sharp + z] = 1 - F_u(q_u^\sharp + z) \leqslant 1 - F_u(q_u^\sharp) - Lz = \alpha - Lz$, yielding that $\mu_n(H) - \mu(H) \geqslant Lz$. If $\hat{q}_u < q_u^\sharp - z$ a similar reasoning yields $|\mu(H) - \mu_n(H)| \geqslant Lz$ for $H = H_{-u,-(q_u^\sharp-z)} \in \mathcal{H}_0$. Hence, using (29), it follows that

$$\mathbb{P}\left[\sup_{u\in\mathcal{S}^{d-1}} |\hat{q}_u - q_u^\sharp| > z\right] \leqslant \mathbb{P}\left[\sup_{H\in\widetilde{\mathcal{H}}_0} |\mu_n(H) - \mu(H)| \geqslant Lz\right]. \tag{30}$$

Now, denote by $S = \sup_{H\in\widetilde{\mathcal{H}}_0} |\mu_n(H) - \mu(H)|$. Since $\widetilde{\mathcal{H}}_0 \subseteq \mathcal{H}$, it has Vapnik-Chervonenkis dimension at most $d+1$. Moreover, it is a countable class of sets, so Proposition 3.1 in [3] yields $\mathbb{E}[S] \leqslant \frac{10\sqrt{5(d+1)}}{\sqrt{n}}$. Therefore, by Theorem 2.5 in [25], if $Lz \geqslant \frac{10\sqrt{5(d+1)}}{\sqrt{n}}$,

$$\begin{aligned}\mathbb{P}[S \geqslant Lz] &\leqslant \mathbb{P}\left[S - \mathbb{E}[S] \geqslant Lz - \frac{10\sqrt{5(d+1)}}{\sqrt{n}}\right] \\ &\leqslant A\exp\left(-L^2z^2n/2 + 10\sqrt{5(d+1)}Lz\sqrt{n}\right),\end{aligned} \tag{31}$$

where $A = e^{-250(d+1)}$. Lemma 15 follows from (30) and (31).

Lemma 16 *Let $f_1, \ldots, f_n$ be n real valued continuous functions defined on a topological space E and $k \in \{1, \ldots, n\}$. For $x \in E$, denote by $f_{(k)}(x)$ the k-th smaller number in the list $f_1(x), \ldots, f_n(x)$. Then, $f_{(k)}$ is continuous.*

Proof Write $f_{(k)}(x) = \min_{J\in\mathcal{P}_k} \max_{j\in J} f_j(x)$, where $\mathcal{P}_k$ is the collection of all subsets of $\{1, \ldots, n\}$ of size k. Continuity of $f_{(k)}$ follows from continuity of the maximum and minimum of finitely many continuous functions.

4.3 Proofs of the main results

Proof of Lemma 1: Let us first show that $G_{\mathsf{MQ}}^\flat = G_{\mathsf{FB}}$. Let $x \in G_{\mathsf{MQ}}^\flat$ and $H \in \mathcal{H}$ satisfying $\mu(H) \geqslant 1-\alpha$. Write $H = H_{u,t}$, for some $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$. Then, $\mu(H) = \mathbb{P}[\langle u, X\rangle \leqslant t] \geqslant 1-\alpha$, which yields $t \geqslant q_u^\flat$. Since $x \in G_{\mathsf{MQ}}^\flat$, $\langle u, x\rangle \leqslant q_u^\flat$ and, hence, $x \in H$. Therefore, $G_{\mathsf{MQ}}^\flat \subseteq G_{\mathsf{FB}}$. Now, let $x \in G_{\mathsf{FB}}$ and $u \in \mathcal{S}^{d-1}$. Let $H = H_{u,q_u^\flat}$. By definition of $q_u^\flat$ and since F_u is right continuous, $\mu(H) = F_u(q_u^\flat) \geqslant 1-\alpha$, so $x \in H$. Hence, $x \in G_{\mathsf{MQ}}^\flat$ and thus, $G_{\mathsf{FB}} \subseteq G_{\mathsf{MQ}}^\flat$.

This ends the proof of the equality $G^{\flat}_{\mathsf{MQ}} = G_{\mathsf{FB}}$.
Inclusion $G^{\flat}_{\mathsf{MQ}} \subseteq G^{\sharp}_{\mathsf{MQ}}$ follows from the inequalities $q^{\flat}_u \leqslant q^{\sharp}_u$, for all $u \in \mathcal{S}^{d-1}$.
Now, let us prove that $G^{\sharp}_{\mathsf{MQ}} = G_\mu$. For $x \in G^{\sharp}_{\mathsf{MQ}}$, we show that $D_\mu(x) \geqslant \alpha$, i.e., that any closed halfspace H containing x needs to satisfy $\mu(H) \geqslant \alpha$. Let H be such a halfspace and write $H = H_{u,t}$ for some $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$. Then, $\langle u, x\rangle \leqslant t$,so $\langle -u, x\rangle \geqslant -t$. Since $x \in G^{\sharp}_{\mathsf{MQ}}$, $\langle -u, x\rangle \leqslant q^{\sharp}_{-u}$, hence, $-t \leqslant q^{\sharp}_{-u}$. Therefore,

$$\begin{aligned}\mu(H) &= \mathbb{P}[\langle u, X\rangle \leqslant t] = 1 - \mathbb{P}[\langle u, X\rangle > t] = 1 - \mathbb{P}[\langle -u, X\rangle < -t] \\ &\geqslant 1 - \mathbb{P}[\langle -u, X\rangle < q^{\sharp}_{-u}] \geqslant 1 - (1-\alpha) = \alpha.\end{aligned}$$

Thus, $x \in G_\mu$, and hence, $G^{\sharp}_{\mathsf{MQ}} \subseteq G_\mu$. Now, let $x \in G_\mu$ and $u \in \mathcal{S}^{d-1}$. Since $x \in H_{-u,\langle -u,x\rangle}$ and $D_\mu(x) \geqslant \alpha$, $\mu(H_{-u,\langle -u,x\rangle}) \geqslant \alpha$, i.e., $\mathbb{P}[\langle -u, X\rangle \leqslant \langle -u, x\rangle] \geqslant \alpha$. Hence, $\mathbb{P}[\langle u, X\rangle < \langle u, x\rangle] \leqslant 1-\alpha$,which, by definition of $q^{\sharp}_u$, implies that $\langle u, x\rangle \leqslant q^{\sharp}_u$. So, $x \in G^{\sharp}_{\mathsf{MQ}}$. Therefore, $G^{\sharp}_{\mathsf{MQ}} = G_\mu$.

Proof of Proposition 1

- *(i)* $\Rightarrow$ *(ii)*: Assume that all the constraints are active and let $u \in \mathcal{S}^{d-1}$. First, by definition of the support function, $h_G(u) \leqslant t_u$. Second, since the constraint corresponding to u is active, there exists $x^* \in G$ such that $\langle u, x^*\rangle = t_u$, yielding $t_u \leqslant h_G(u)$, hence, $t_u = h_G(u)$.
- *(ii)* $\Rightarrow$ *(i)*: Let $u \in \mathcal{S}^{d-1}$. By Lemma 4, G is compact, yielding the existence of $x^* \in G$ satisfying $h_G(u) = \langle u, x^*\rangle$. Hence, the constraint corresponding to u is active.
- *(ii)* $\Rightarrow$ *(iii)* is a direct consequence of the sublinearity of support functions.
- *(iii)* $\Rightarrow$ *(ii)*: Assume that the family $(t_u)_{u\in\mathbb{R}^d}$ is subadditive and let $u_0 \in \mathcal{S}^{d-1}$. Since $u \in \mathbb{R}^d \mapsto t_u$ is subadditive and positively homogeneous, it is convex. Hence, it is continuous on the interior of its domain, here, $\mathbb{R}^d$. Since $\overset{\circ}{G} \neq \varnothing$, Lemma 6 yields the existence of $x^* \in G$ satisfying $h_G(u_0) = \langle u_0, x^*\rangle$ and of $u_1, \dots, u_d \in \mathcal{S}^{d-1}$, $\lambda_1, \dots, \lambda_d \geqslant 0$ satisfying $u_0 = \sum_{i=1}^d \lambda_i u_i$ and, for $i = 1, \dots, d$, $\langle u_i, x^*\rangle = t_{u_i}$. Hence, $h_G(u_0) = \langle u_0, x^*\rangle = \displaystyle\sum_{i=1}^d \lambda_i \langle u_i, x^*\rangle = \sum_{i=1}^d \lambda_i t_{u_i} \geqslant t_{u_0}$, by positive homogeneity and subadditivity of $v \mapsto t_v$. Since, in addition, $h_G(u_0) \leqslant t_{u_0}$ by definition of the support function, we obtain $h_G(u_0) = t_{u_0}$.

Proof of Theorem 1 Let α be greater than $1/2$ and assume that G_μ is nonempty. Let $x \in G_\mu$: We prove that $\mu(\{x\}) \geqslant 2\alpha - 1$.

Let E be an affine hyperplane passing through x. Let H_1 and H_2 be the two distinct halfspaces whose common boundary is E. Since $x \in G_\mu$, $D_\mu(x) \geqslant \alpha$. In particular, since both H_1 and H_2 contain x, $\mu(H_j) \geqslant \alpha, j = 1, 2$. Hence, $1 \geqslant \mu(H_1 \cup H_2) = \mu(H_1) + \mu(H_2) - \mu(E) \geqslant 2\alpha - \mu(E)$, which implies that $\mu(E) \geqslant 2\alpha - 1$.

Let $k \in \{1, \ldots, d-1\}$. Assume it is known that any affine subspace E of dimension k, containing x, satisfies $\mu(E) \geqslant 2\alpha - 1$. Let F be an affine subspace of dimension $k-1$, containing x. Let G be the linear subspace of vectors that are orthogonal to F. Let $p \geqslant 2$ be an integer and let $u_1, \ldots, u_p$ be unit vectors in G, such that no two of them are collinear. For $i = 1, \ldots, p$, set $E_i = F + \mathbb{R}u_i = \{f + \lambda u_i : f \in F, \lambda \in \mathbb{R}\}$. Then, for all $I \subseteq \{1, \ldots, p\}$ with $\#I \geqslant 2$, $\bigcap_{i \in I} E_i = F$ and as a consequence of the inclusion-exclusion principle,

$$1 \geqslant \mu\left(\bigcup_{i=1}^{p} E_i\right) = \sum_{i=1}^{p} \mu(E_i) - \sum_{j=2}^{p} (-1)^j \binom{p}{j} \mu(F) \geqslant p(2\alpha - 1) - (p-1)\mu(F),$$

yielding $\mu(F) \geqslant \frac{p}{p-1}(2\alpha - 1) - \frac{1}{p-1}$. Since p is an arbitrary integer, we can let it go to infinity and we get $\mu(F) \geqslant 2\alpha - 1$.

By induction, this proves that $\mu(\{x\}) \geqslant 2\alpha - 1 > 0$ and this must hold for all $x \in G_\mu$. Since G_μ is convex, it cannot contain more than one point. Indeed, if $x, y \in G_\mu$, then $[x, y] \subseteq G_\mu$, yielding $\mu(\{z\}) \geqslant 2\alpha - 1$, for all $z \in [x, y]$. Hence, if $x \neq y$, then $\mu([x, y]) = \infty$, which is impossible.

Proof of Theorem 2 Without loss of generality, let us assume that $a = 0$ in Assumption 2: translating the measure μ and the sample points does not affect the Haussdorf distance between G_μ and $\hat{G}$. For the sake of notation, we write $q_u = q_u^\flat = q_u^\sharp$ for all $u \in \mathcal{S}^{d-1}$.

Let $z \in [10\sqrt{5(d+1)}/(L\sqrt{n}), \varepsilon)$ and let the event $\mathcal{A} = \{|\hat{q}_u - q_u| \leqslant z, \forall u \in \mathcal{S}^{d-1}\}$ hold. Since $B(0, r) \subseteq G_\mu$, $ru \in G_\mu$ for all $u \in \mathcal{S}^{d-1}$. By Lemma 1, this implies that $r = \langle u, ru \rangle \leqslant q_u, \forall u \in \mathcal{S}^{d-1}$. Hence, for all $u \in \mathcal{S}^{d-1}$, $\hat{q}_u \geqslant q_u - z \geqslant r - \varepsilon > 0$, yielding that $B'(0, r-\varepsilon) \subseteq \hat{G}$, hence, that $\hat{G}$ has a nonempty interior. So does G_μ, since it contains $B'(0, r)$.

By Lemmas 12 and 16, the maps $u \mapsto q_u$ and $u \mapsto \hat{q}_u$ are continuous. Indeed, $\hat{q}_u$ is the $\lceil n(1-\alpha) + 1 \rceil$-th order function of $\langle u, X_1 \rangle, \ldots, \langle u, X_n \rangle$. Note that the map $t \in [0, 1) \mapsto \frac{1+t}{1-t}$ is nondecreasing. Thus, by Lemma 7, $d_{\mathsf{H}}(\hat{G}, G_\mu) \leqslant \frac{zR}{r} \frac{1 + z/r}{1 - z/r} \leqslant Cz$, where $C = \frac{R}{r} \frac{1 + \varepsilon/r}{1 - \varepsilon/r}$. Hence, if $\mathcal{A}^\complement$ stands for the complement of the event $\mathcal{A}$, then

$$\mathbb{P}[d_{\mathsf{H}}(\hat{G}, G_\mu) > Cz] \leqslant \mathbb{P}\left[\mathcal{A}^\complement\right]. \tag{32}$$

Write $z = x/\sqrt{n}$, for some real number x satisfying $\frac{10\sqrt{5(d+1)}}{L} \leqslant x < \varepsilon\sqrt{n}$. By Lemma 15,

$$\mathbb{P}\left[\mathcal{A}^\complement\right] \leqslant A \exp\left(-L^2 x^2/2 + 10\sqrt{5(d+1)} L x\right). \tag{33}$$

The desired result is a consequence of (32) and (33).

Proof of Corollary 1 The proof is based on a consequence of Fubini's theorem which ensures that if Z is a nonnegative random variable, then

$$\mathbb{E}[Z^k] = k\int_0^\infty t^{k-1}\mathbb{P}\left[Z > t\right]\mathrm{d}t, \tag{34}$$

for all positive number k.

First, note that for all $k > 0$,

$$\mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)^k\right] = \mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)\mathbb{1}_{\hat{G}\neq\varnothing}\right] + \mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)^k\mathbb{1}_{\hat{G}=\varnothing}\right], \tag{35}$$

where $\mathbb{1}$ stands for the indicator function.

By definition of $\hat{G}^*$, the second term in the right hand side of (35) is equal to $d_{\mathsf{H}}(\{0\}, G_\mu)^k\mathbb{P}\left[\hat{G} = \varnothing\right]$. First, it is clear that $d_{\mathsf{H}}(\{0\}, G_\mu) \leqslant |a|+R \leqslant \tau+R$. Second, as we saw in the proof of Theorem 2, $\mathbb{P}\left[\hat{G} = \varnothing\right] \leqslant \mathbb{P}\left[\mathcal{A}^\complement\right]$ where we set $z = \varepsilon/2$. Hence, by (33),

$$\mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)^k\mathbb{1}_{\hat{G}=\varnothing}\right] = O\left(n^{-k/2}\right), \tag{36}$$

with multiplicative constants that depend on d, ε, R, L and τ only.

For the first term of (35), note that if $\hat{G} \neq \varnothing$, then, since $\hat{G}^* \subseteq B'(0, \log n)$ and $G_\mu \subseteq B'(a, R)$, $d_{\mathsf{H}}(\hat{G}^*, G_\mu) \leqslant |a| + \log n + R \leqslant \tau + \log n + R$. Denote by $B = \tau + \log n + R$. Then, if we set $Z = d_{\mathsf{H}}(\hat{G}, G_\mu)$,

$$\mathbb{E}\left[d_{\mathsf{H}}(\hat{G}^*, G_\mu)^k\mathbb{1}_{\hat{G}\neq\varnothing}\right] \leqslant \mathbb{E}\left[Z^k\mathbb{1}_{Z\leqslant B}\right]. \tag{37}$$

In the following, we set $k = 1$. General values of k would be handled similarly, using (34). Using (37) and (34) with $k = 1$, $\mathbb{E}[d_{\mathsf{H}}(\hat{G}^*, G_\mu)\mathbb{1}_{\hat{G}\neq\varnothing}] \leqslant \int_0^B \mathbb{P}\left[d_{\mathsf{H}}(\hat{G}, G_\mu) > t\right]\mathrm{d}t$. Split the integral in three integrals. First, from 0 to $\frac{10C\sqrt{5(d+1)}}{L\sqrt{n}}$, where we bound the integrand by 1. Then, from $\frac{10C\sqrt{5(d+1)}}{L\sqrt{n}}$ to ε, where we use the bound provided by Theorem 2. Third, in the remaining interval, where, using monotonicity, we bound the integrand using the upper bound given in Theorem 2 with $x = \varepsilon\sqrt{n}$. Then,

$$\mathbb{E}[d_{\mathsf{H}}(\hat{G}^*, G_\mu)\mathbb{1}_{\hat{G}\neq\varnothing}] = O\left(n^{-1/2}\right), \tag{38}$$

with multiplicative constants that depend on d, ε, r, R and L only. Together with (36), (38) yields the desired result.

Proof of Corollary 2 It is enough to prove that if μ satisfies either Assumption 3 or 4, then it satisfies both Assumptions 1 and 2, for some values of ε, L, r and R. Hence, Theorem 2 will apply and yield the desired result.

Let X be a random variable in $\mathbb{R}^d$ with probability measure μ. If $u \in \mathcal{S}^{d-1}$, denote by f_u the density of $\langle u, X\rangle$ and by F_u its cumulative distribution function. For $u \in \mathcal{S}^{d-1}$ and $t \in \mathbb{R}$, let $\phi(u,t) = f_u(t) = \int_{u^\perp} f(tu+v)\,\mathrm{d}v$, where the integral is evaluated with respect to the $(d-1)$-dimensional Lebesgue measure on $u^\perp$.

Let K be the support of μ and let $A = \{(u,t) \in \mathcal{S}^{d-1} \times \mathbb{R} : (tu + u^\perp) \cap K \neq \varnothing\}$. Note that $\mathring{A}$ is included in the support of ϕ. Thus, by Lemma 13, ϕ is continuous on $\mathring{A}$.

From now on, we assume that μ satisfies either Assumption 3 or 4. For $u \in \mathcal{S}^{d-1}$, since F_u is continuous, $q_u^\sharp = q_u^\flat$: Denote this value by q_u. Let $\alpha_{\mathsf{max}} = \max_{x\in\mathbb{R}^d} D_\mu(x)$. This quantity is well defined, since D_μ is upper semicontinuous and quasi-concave (see [34]). Let $T \in \mathbb{R}^d$ satisfy $D_\mu(T) = \alpha_{\mathsf{max}}$. Since μ has a connected support and is absolutely continuous with respect to the Lebesgue measure, such a point exists and is unique (see [34] or Prop. 3.5 in [33]). Let α_1 and α_2 be positive numbers such that $\alpha_1 < \alpha < \alpha_2 < \alpha_{\mathsf{max}}$. For $u \in \mathcal{S}^{d-1}$, denote by $q_u^{(1)}$ the $(1-\alpha_1)$-quantile of F_u and by $q_u^{(2)}$ the $(1-\alpha_2)$-quantile of F_u. By Lemma 14, $q_u, q_u^{(1)}$ and $q_u^{(2)}$ are continuous functions of u. In addition, for all $u \in \mathcal{S}^{d-1}$, $\langle u, T\rangle < q_u^{(2)} < q_u < q_u^{(1)}$, by definition of the quantiles and by the first part of Lemma 14. Hence, since $\mathcal{S}^{d-1}$ is compact, there exist positive numbers r, R and ε with $\varepsilon < r < R$ and such that for all $u \in \mathcal{S}^{d-1}$,

$$\langle u, T\rangle + r \leqslant q_u^{(1)} \leqslant q_u - \varepsilon \leqslant q_u + \varepsilon \leqslant q_u^{(2)} \leqslant \langle u, T\rangle + R. \tag{39}$$

In particular, the first and last inclusions of (39) imply that $B(T,r) \subseteq G_\mu \subseteq B(T,R)$. Hence, μ satisfies Assumption 2. In addition, by a similar argument as in the proof of Lemma 14, the intermediate inclusions show that the compact set $B = \{(u,t) : u \in \mathcal{S}^{d-1}, q_u - \varepsilon \leqslant t \leqslant q_u + \varepsilon\}$ is included in the interior of A. Hence, by Lemma 13, ϕ is continuous on and positive on B, thus, it is bounded from below by a positive constant L on B, yielding $F_u(t') - F_u(t) \geqslant L(t'-t)$, for all $u \in \mathcal{S}^{d-1}$ and $t, t' \in \mathbb{R}$ such that $q_u - \varepsilon \leqslant t \leqslant t' \leqslant q_u + \varepsilon$. This, together with continuity of F_u, for all $u \in \mathcal{S}^{d-1}$, shows that μ satisfies Assumption 1, which finally ends the proof of Corollary 2.

Proof of Theorem 3: Let $M \geqslant 1$, $\dfrac{10\sqrt{5(d+1)}}{L\sqrt{n}} \leqslant z < \varepsilon$ and $\delta = 1/\sqrt{n}$. For simplicity, we denote by $q_j = q_{U_j}$ and $\hat{q}_j = \hat{q}_{U_j}$, for $j = 1, \ldots, M$. Let $\mathcal{A} = \{|\hat{q}_j - q_j| \leqslant z, \forall j = 1, \ldots, M\}$ and $\mathcal{C} = \big\{\{U_1, \ldots, U_M\}$ is a δ-net of $\mathcal{S}^{d-1}\big\}$. Let both $\mathcal{A}$ and $\mathcal{C}$ hold. Then, by Lemma 8, $d_{\mathsf{H}}(\tilde{G}_M, G_\mu) \leqslant Cz + 4R\delta$, where

$C = \frac{R}{r}\frac{1+\varepsilon/r}{1-\varepsilon/r}$. Therefore, by Lemmas 15 and 11, setting $x = z\sqrt{n}$,

$$\mathbb{P}\left[d_{\mathsf{H}}(\tilde{G}_M, G_\mu) > Cz + \frac{4R}{\sqrt{n}}\right]$$
$$\leqslant Ae^{-L^2x^2/2+10\sqrt{5(d+1)}Lx} + 6^d \exp\left(-\frac{M}{2d8^{(d-1)/2}n^{d-1}} + (d/2)\log n\right),$$

for any $x \in \mathbb{R}$ satisfying $\frac{10\sqrt{5(d+1)}}{L} \leqslant x < \varepsilon\sqrt{n}$.

References

1. Miguel A. Arcones, Zhiqiang Chen, and Evarist Giné. Estimators related to U-processes with applications to multivariate medians: asymptotic normality. *Ann. Statist.*, 22(3):1460–1477, 1994.
2. I. Bárány and D. G. Larman. Convex bodies, economic cap coverings, random polytopes. *Mathematika*, 35(2):274–291, 1988.
3. Yannick Baraud. Bounding the expectation of the supremum of an empirical process over a (weak) VC-major class. *Electron. J. Stat.*, 10(2):1709–1728, 2016.
4. V.-E. Brunel. Uniform behaviors of random polytopes in hausdorff metric. *Submitted*, 2017.
5. Victor-Emmanuel Brunel. Adaptive estimation of convex polytopes and convex sets from noisy data. *Electron. J. Stat.*, 7:1301–1327, 2013.
6. Victor-Emmanuel Brunel. A universal deviation inequality for random polytopes. *submitted, arXiv:1311.2902*, 2014.
7. Probal Chaudhuri. On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc.*, 91(434):862–872, 1996.
8. B. Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete Computational Geometry*, 10:377–409, 1993.
9. Richard Cole, Micha Sharir, and Chee-K. Yap. On k-hulls and related problems. *SIAM J. Comput.*, 16(1):61–77, 1987.
10. J. A. Cuesta-Albertos and A. Nieto-Reyes. The random Tukey depth. *Comput. Statist. Data Anal.*, 52(11):4979–4988, 2008.
11. Antonio Cuevas, Wenceslao González-Manteiga, and Alberto Rodríguez-Casal. Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.*, 48(1):7–19, 2006.
12. David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827, 1992.
13. Subhajit Dutta, Anil K. Ghosh, and Probal Chaudhuri. Some intriguing properties of Tukey's half-space depth. *Bernoulli*, 17(4):1420–1434, 2011.
14. Rainer Dyckerhoff and Pavlo Mozharovskyi. Exact computation of the halfspace depth. *Comput. Statist. Data Anal.*, 98:19–30, 2016.

15. Daniel Fresen. A multivariate Gnedenko law of large numbers. *Ann. Probab.*, 41(5):3051–3080, 2013.
16. Anil K. Ghosh and Probal Chaudhuri. On data depth and distribution-free discriminant analysis using separating surfaces. *Bernoulli*, 11(1):1–27, 2005.
17. Anil K. Ghosh and Probal Chaudhuri. On maximum depth and related classifiers. *Scand. J. Statist.*, 32(2):327–350, 2005.
18. Adityanand Guntuboyina. Optimal rates of convergence for convex set estimation from support functions. *Ann. Statist.*, 40(1):385–411, 2012.
19. Marc Hallin, Davy Paindaveine, and Miroslav Šiman. Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist.*, 38(2):635–669, 2010.
20. Xuming He and Gang Wang. Convergence of depth contours for multivariate datasets. *Ann. Statist.*, 25(2):495–504, 1997.
21. Y. He. Multivariate extreme value statistics for risk assessment. *PhD thesis*, 2016.
22. Mia Hubert, Peter Rousseeuw, and Pieter Segaert. Multivariate and functional classification using depth and distance. *Adv. Data Anal. Classif.*, 11(3):445–466, 2017.
23. D. S. Johnson and F. P. Preparata. The densest hemisphere problem. *Theoret. Comput. Sci.*, 6(1):93–107, 1978.
24. Jeankyung Kim. Rate of convergence of depth contours: with application to a multivariate metrically trimmed mean. *Statist. Probab. Lett.*, 49(4):393–400, 2000.
25. Vladimir Koltchinskii. *Oracle inequalities in empirical risk minimization and sparse recovery problems*, volume 2033 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 38th Probability Summer School held in Saint-Flour, 2008, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].
26. Linglong Kong and Ivan Mizera. Quantile tomography: using quantiles with multivariate data. *Statist. Sinica*, 22(4):1589–1610, 2012.
27. S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian J. Math. Stat.*, 4(1):66–70, 2011.
28. Regina Y. Liu, Jesse M. Parelius, and Kesar Singh. Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.*, 27(3):783–858, 1999. With discussion and a rejoinder by Liu and Singh.
29. Regina Y. Liu and Kesar Singh. A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.*, 88(421):252–260, 1993.
30. R. Lòpez and G. Still. Semi-infinite programming. *European J. Oper. Res.*, 180(2):491–518, 2007.
31. László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Structures Algorithms*, 30(3):307–358, 2007.
32. Peter Mani-Levitska. Characterizations of convex sets. pages 19–41, 1993.
33. J.-C. Massé and R. Theodorescu. Halfplane trimming for bivariate distributions. *J. Multivariate Anal.*, 48(2):188–202, 1994.

34. Jean-Claude Massé. Asymptotics for the Tukey depth process, with an application to a multivariate trimmed mean. *Bernoulli*, 10(3):397–419, 2004.
35. Kim Miller, Suneeta Ramaswami, Peter Rousseeuw, J. Antoni Sellarès, Diane Souvaine, Ileana Streinu, and Anja Struyf. Efficient computation of location depth contours by methods of computational geometry. *Stat. Comput.*, 13(2):153–162, 2003.
36. Ilya S. Molchanov. A limit theorem for solutions of inequalities. *Scand. J. Statist.*, 25(1):235–242, 1998.
37. B. Pateiro-Lopez. Set estimation under convexity type restrictions. *PhD Thesis*, 2008.
38. W. Polonik. Measuring mass concentrations and estimating density contour clusters - an excess mass approach. *The Annals of Statistics*, 23:855–881, 1995.
39. Philippe Rigollet and Régis Vert. Optimal rates for plug-in estimators of density level sets. *Bernoulli*, 15(4):1154–1178, 2009.
40. P. J. Rousseeuw and I. Ruts. Computing depth contours of bivariate point clouds. *Computational statistics and data analysis*, 23:153–168, 1996.
41. P. J. Rousseeuw and A. Struyf. Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8:193–203, 1998.
42. Peter J. Rousseeuw and Ida Ruts. The depth function of a population distribution. *Metrika*, 49(3):213–244, 1999.
43. Rolf Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014.
44. C. Schtt and E. Werner. The convex floating body. *Math. Scand*, 66:275–290, 1990.
45. A. Tsybakov. On nonparametric estimation of density level sets. *Ann. Stat.*, 25:948–969, 1997.
46. John W. Tukey. Mathematics and the picturing of data. pages 523–531, 1975.
47. Arthur B. Yeh and Kesar Singh. Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B*, 59(3):639–652, 1997.
48. Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482, 2000.
49. Yijun Zuo and Robert Serfling. Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.*, 28(2):483–499, 2000.