Efficiency of the floating body as a robust measure of dispersion

Joseph Anderson Luis Rademacher

February 11, 2019

Abstract

Among robust notions of shape, depth and dispersion of a distribution or dataset we have Tukey depth and depth curves, which are essentially the same as the convex floating body in convex geometry. These notions are important because they play the role of multidimensional quantiles and rank statistics. At the same time, they can be difficult to use because they are computationally intractable in general. We develop a theory of algorithmic efficiency for these notions for several broad and relevant families of distributions: symmetric log-concave distributions and certain multivariate stable distributions and power-law distributions. As an example of the power of these results, we show how to solve the Independent Component Analysis problem for power-law distributions, even when the first moment is infinite.

1 Introduction

One of the aims of robust statistics is to infer information from data in a way that is not too sensitive to perturbations of the input. One of the basic tools in this context are the quantiles of a distribution. Informally speaking, a quantile of level $q \in (0,1)$ of a set of real numbers is the least value that is above a q fraction of the set. While quantiles are well-understood, it is an active area of research to understand generalizations to multi-dimensional data sets. A natural starting point is to consider the quantile function of a multi-dimensional data set, namely the function that maps q and a direction θ to the quantile of level q of the marginal along θ.

How helpful this is depends partly on whether it is computationally efficient to answer questions about this function given the data. Even without a complete answer to the issue of computational efficiency, notions closely related to the quantile function have been central to robust and descriptive statistics: Tukey (or halfspace) depth, Tukey median, depth contours (depth curves) [Tuk75], [DG92] as well as centerpoints. In particular, the notion of "halfspace depth" was proposed in [Tuk75] as a way to generalize the notion of quantiles to multivariate samples. In this formulation, for a probability measure $\mathbb{P}$ on $\mathbb{R}^n$, the halfspace depth of a point $x \in \mathbb{R}^n$ is defined as $\mathrm{HD}(x; \mathbb{P}) = \inf\{\mathbb{P}(H) : H \text{ is a halfspace containing } x\}$. From a notion of depth (of a sample or distribution), one can consider *depth curves*: the set of points that have at least a particular depth. For Tukey's depth function and many others, these curves form convex sets in the ambient space and can be used to get information about the "shape" of the data.

There are several reasons to study the quantile function and related notions such as depth: They are robust against outliers [DG92] and they are defined even for heavy tailed distributions (without any moment assumption). They can therefore be used for outlier detection and as a way of understanding the dispersion and shape of a distribution or dataset (as a multivariate analog of rank statistics).

Notions such as depth and depth curves are known to be computationally intractable in general: testing whether the depth of a point with respect to a sample is at least some fixed bound is coNP-complete [JP78]. Given the issue of computational efficiency of questions about the quantile function, in this paper we argue that the quantile function and many natural questions about it are computationally tractable in a broad family of situations. The main idea is to consider the quantile function of a multivariate distribution and use known results that show that it has a convexity property for several relevant families of distributions. The convexity property is that, as a function of θ and for fixed q, the quantile function is the support function of a convex set when the distribution is either (1) centrally symmetric and log-concave, or (2) a product distribution with coordinates following a symmetric stable or power-law distribution. A consequence of the conditions that guarantee efficiency is the following: *While questions about the empirical quantile function of a sample (equivalently, a discrete distribution) are in general computationally hard, if the sample comes from a distribution corresponding to the convex case, then the empirical quantile function is close to the convex case, and this leads to computational efficiency.* This can be proven rigorously by means of the ellipsoid algorithm, which has polynomial time guarantees even when it is only given an approximation to its input function or set.

The floating body. The notion of a depth curve coincides with a central concept in convex geometry, the *convex floating body*. The *convex floating body* was introduced in [SW90] and independently in [BL88]. It is a variation of a much older notion, Dupin's floating body (see [Dup22, Lei98]). To avoid confusion, we use the term "floating body" to refer exclusively to the one studied in [SW90]. However, we define this body not just for the uniform distribution in a convex body, but for random vectors in $\mathbb{R}^n$. See [NSW18] for an in-depth discussion. We use a standard notion of quantiles of a random variable, and the quantiles in each direction then determine the floating body for random vector X:

Definition 1. *The q-quantile of a random variable X at level $q \in (0, 1)$ is given by $Q_q(X) := \inf\{t : P(X \leq t) \geq q\}$.*

Definition 2. *The floating body of a random vector $X \in \mathbb{R}^n$ at level $q \in (0, 1)$ is $\Phi_q X = \bigcap_{\theta \in \mathbb{S}^{n-1}} \{x \in \mathbb{R}^n : \langle x, \theta \rangle \leq Q_q(\langle X, \theta \rangle)\}$.*

The primary appeal of the floating body is that it exists for any distribution over $\mathbb{R}^n$, specifically without the existence of any moments, and is also affine equivariant (that is, an affine transformation of X transforms the floating body in the same way).

A natural question arises: what is the statistical behavior of a sample restricted to the floating body of its underlying distribution? Which properties of the underlying distribution are preserved? A similar question is addressed in [AGNR15] and [AGNR17], but about the *centroid body* — a depth measure dependent on directional means of the underlying distribution.

We present Independent Component Analysis as a motivational application of this question and our algorithmic results. Specifically, we extend the results of [AGNR15] by using the floating body as an algorithmically efficient summary of the geometry of the given data that comes from a multidimensional heavy-tailed distribution.

Our contributions. We study the computational efficiency of the floating body of a distribution or sample. In this context, we show that in a broad family of relevant situations one can efficiently answer direct questions about the floating body. The situations include symmetric log-concave distributions (Section 5), power-law (Section 8) and stable distributions (Section 7), as well as datasets that behave like samples from any of those distributions families. The main question we handle about the floating body is efficient membership, both in terms of samples and computation. From membership one can directly answer more specific questions such as the halfspace depth of a point and the covariance matrix of the floating body as a measure of dispersion.

Our results include approximation guarantees of the floating body from a sample, uniformly (in the Hausdorff metric, via VC-theory). Related results can be found in [Bru18, Theorem 2]. Our results are specialized to our case and in this way we achieve simpler assumptions and a more elementary proof: our work only uses the results in [VC71] while [Bru18] uses much newer refinements.

As a proof of concept, we show how efficient access to the floating body can be used to efficiently solve the Independent Component Analysis problem for power-law distributions with tail decay no slower than the Cauchy distribution (with infinite first moment). This is based on the idea of reducing the power-law case to the stable distribution case by taking sums of power-law samples and using the fact that such a sum converges to a stable distribution. This is formally based on known results that provide an efficient rate of convergence in the generalized central limit theorem. The stable distribution case of ICA can then be solved based on the fact that the resulting floating body is an l_p-ball and known results to learn an l_p-ball [AGR13].

The ICA problem has been widely studied in both theoretical and applied settings; see, for example, [CJ10], [HO00]. Our extension of the ICA framework to stable and power-law distributions can easily be adapted to other settings where robust estimation is needed.

2 Preliminaries

For a convex set $K \subset \mathbb{R}^n$, the *support function* of K is $h_K(x) = \sup\{\langle x, k\rangle : k \in K\}$. The *radial function* of K is $\rho_K(x) = \sup\{a > 0 : ax \in K\}$. The *polar* of K is the convex set $K^\circ = \{x \in \mathbb{R}^n : \langle x, k\rangle \leq 1 \ \forall k \in K\}$. Then, a key relationship between the two is that the radial function of K° is the reciprocal of the support function of K, i.e. $\rho_{K^\circ} = 1/h_K$. This, along with the fact that $(K^\circ)^\circ = K$ under mild assumptions, will provide the basis of our membership algorithms in later sections. For a matrix A, we denote the smallest and largest singular values by $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$. We also denote the smallest and largest eigenvalues by $\lambda_{\min}$ and $\lambda_{\max}(A)$.

For point $x \in \mathbb{R}^n$, $p > 0$, and $r > 0$, we denote the l_p ball $B_p(x, r) = \{y \in \mathbb{R}^n : \|y\|_p \leq r\}$. We denote the unit l_p ball in $\mathbb{R}^n$ as $B_p^n := B_p(0, 1)$. For a convex set $K \subset \mathbb{R}^n$ and $\epsilon > 0$, define the outer parallel body $K_\epsilon = \{x \in \mathbb{R}^n : \|x - k\|_2 \leq \epsilon \text{ for some } k \in K\}$. Similarly, the inner parallel body is $K_{-\epsilon} = \{x \in K : B(x, \epsilon) \subset K\}$. Given a vector x, we denote the normalized version as $\hat{x} := x/\|x\|_2$. For sets $A, B \subseteq \mathbb{R}^n$, we denote the Hausdorff distance as $d_H(A, B) = \inf\{\epsilon > 0 : A \subseteq B_\epsilon \text{ and } B \subseteq A_\epsilon\}$.

Algorithmic convexity. One main result of this paper is to establish the existence of efficient algorithms which give a "probably approximate" membership decision in the following sense:

Definition 3 ([GLS88])**.** *The* ϵ-weak membership problem *for K is the following: Given a point $y \in \mathbb{Q}^n$ and a rational number $\epsilon > 0$, either (i) assert that $y \in K_\epsilon$, or (ii) assert that $y \notin K_{-\epsilon}$. An* ϵ-weak membership oracle *for K is an oracle that solves the weak membership problem for K. For $\delta \in [0, 1]$, an* (ϵ, δ)-weak membership oracle *for K acts as follows: Given a point $y \in \mathbb{Q}^n$, with probability at least $1 - \delta$ it solves the ϵ-weak membership problem for y, K, and otherwise its output can be arbitrary.*

The algorithmic efficiency of using a support function for membership in the polar is known, for instance in [AGNR15], but we provide a specific result to be used in our analysis.

Lemma 4. *Let $K \subset \mathbb{R}^n$ be a compact convex set with support function h and such that $rB_2^n \subseteq K \subseteq RB_2^n$ for $r, R > 0$. Suppose there exists an algorithm $\mathcal{A}$ to compute $\tilde{h}$ such that $P(|h - \tilde{h}| > \epsilon_1) < \delta$, and the time and sample complexity of that algorithm is $T(n, \epsilon_1, \delta, \Theta)$, where Θ denotes the other parameters of the algorithm. Then there exists an algorithm to give an (ϵ, δ)-weak membership decision for K° with using time and sample complexity $T(n, \min\{r^2\epsilon/2, r/2\}, \delta, \Theta)$.*

Proof. From the convexity of K, we know that the radial function of K° is $\rho(\theta) = 1/h(\theta)$. Then, to estimate membership, the natural algorithm is, given $x \in \mathbb{R}^n$, to obtain the estimate $\tilde{h} := \tilde{h}(x/\|x\|)$ from $\mathcal{A}$ and returning TRUE iff $\|x\| \leq 1/\tilde{h}$.

Suppose now that the estimate is accurate, so that $|h - \tilde{h}| \leq \epsilon_1 \leq r/2$. Then, to control the radial function error we compute

$$\left|\frac{1}{h} - \frac{1}{\tilde{h}}\right| = \frac{|h - \tilde{h}|}{h\tilde{h}} \leq \frac{\epsilon_1}{r(r - \epsilon_1)} \leq \frac{2\epsilon_1}{r^2}.$$

So if we choose $\epsilon_1 = \min\{r^2\epsilon/2, r/2\}$, then the result follows. □

Probabilistic tools. A random variable X (and its distribution) is centrally-symmetric if $-X$ has the same distribution as X; this applies similarly to densities and measures in the proper context. Furthermore, we say a random vector $X \in \mathbb{R}^n$ (or its density or corresponding measure) is isotropic if its mean is zero and its covariance matrix is the identity: $E(X) = 0$ and $E(XX^T) = I_n$. Unless otherwise noted, an empirical estimate of a quantity f is denoted as $\tilde{f}$.

One of the motivations of this paper is to be able to handle heavy-tailed distributions. Several of our results involve stable distributions, a family of heavy-tailed distributions.

Definition 5. *For $\alpha \in (0, 2)$, the* standard symmetric α-stable distribution *is the one dimensional probability distribution with characteristic function $t \mapsto e^{-|t|^\alpha}$. The standard symmetric* 1*-stable distribution is the* standard Cauchy distribution.

The Cauchy distribution is denoted $\mathrm{Cauchy}(x_0, \gamma)$ and has pdf $x \mapsto \frac{1}{\pi\gamma(1+(\frac{x-x_0}{\gamma})^2)}$ and cdf $x \mapsto \frac{1}{\pi}\arctan(\frac{x-x_0}{\gamma}) + \frac{1}{2}$. The standard Cauchy distribution corresponds to $\mathrm{Cauchy}(0, 1)$

Our analysis of stable distributions also applies to the case $\alpha = 2$ (the Gaussian distribution), but we do not include it in the definition of standard stable because characteristic function e^{-t^2} does not have the right scale to be the standard Gaussian distribution. In any case, the Gaussian distribution is also log-concave and better handled by that part of our paper.

The standard symmetric α-stable distribution does not have moments of order less than or equal to α, which causes problems for algorithms or results which rely on estimating the mean, variance, or higher-order moments from data.

Uniform convergence of empirical distributions. For the analysis of our algorithms, we use some well-known ideas from learning theory. Let $Z_1, \dots, Z_N$ be iid random variables in $\mathbb{R}^n$. For a class of sets $\mathcal{A}$, we consider a probability measure on $\mathcal{A}$ as $\mu(A) = P(Z_1 \in A)$. The empirical measure of μ is then $\mu_N(A) = (1/n)\sum_{j=1}^N \mathbb{1}\{Z_j \in A\}$, where I is the indicator random variable. The shatter coefficients measure a type of "complexity" of the class $\mathcal{A}$:

Definition 6. *Let $\mathcal{A}$ be a collection of measurable sets. For $\{z_1, \dots, z_N\} \in \{\mathbb{R}^n\}^N$, let $N_{\mathcal{A}}(z_1, \dots, z_N)$ be the number of different sets in $\{\{z_1, \dots, z_N\} \in A : A \in \mathcal{A}\}$. The n-th shatter coefficient of $\mathcal{A}$ is $s(\mathcal{A}, N) = \max_{(z_1,\dots,z_N)\in\{\mathbb{R}^n\}^N} N_{\mathcal{A}}(z_1, \dots, z_N)$. That is, the shatter coefficient is the maximal number of different subsets of N points that can be identified by the class of sets $\mathcal{A}$.*

In our case, we will be working primarily with half-spaces in $\mathbb{R}^n$, which have shatter coefficient bounded by $2(N-1)^n + 2$. In this work, the shatter function is used primarily in the well-known "VC Theorem" from [VC71], adapted here as Theorem 7. For reference on the wider application of this theorem in statistical learning theory, refer to [KVV94] and [SSBD14].

Theorem 7. *If $s(\mathcal{A}, N) \leq N^n + 1$ then $P\big(\sup_{A\in\mathcal{A}}|\mu_N(A) - \mu(A)| > \epsilon\big) \leq \delta$ when $N \geq \frac{16}{\epsilon^2}\left(n \log\frac{16n}{\epsilon^2} + \log\frac{4}{\delta}\right)$.*

3 The role of convexity and subadditivity

The use of the floating body (or, equivalently, depth curves) algorithmically poses the challenge of answering questions about it efficiently. One of the simplest questions is *membership*, that is, given a query point $x \in \mathbb{R}^n$, determine whether it is in $\Phi_q X$. By definition, this means to check whether x is contained in the intersection of an infinite number of halfspaces. There could be two situations, depending on whether all halfspaces in the definition are "active" constraints (using optimization language). In the language of convex geometry, this can be rephrased as: is the quantile function $\theta \mapsto Q_q(\langle X, \theta\rangle)$ the support function of $\Phi_q X$? If the answer is yes, then using standard results in algorithmic convexity one can decide membership in polynomial time [GLS88]. Moreover, standard results in convex geometry (see for example [Sch14, Theorem 1.7.1 and proof]) imply the following characterization: Let $f : \mathbb{R}^n \to \mathbb{R}$ be the positively homogeneous extension from S^{n-1} to $\mathbb{R}^n$ of the quantile function, namely $f(\theta) = Q_q(\langle X, \theta\rangle)$. Then f is the support function of $\Phi_q X$ iff f is subadditive (that is, $f(x + y) \leq f(x) + f(y)$ for all $x, y \in \mathbb{R}^n$). Note that if f is a positively homogeneous function then f is subadditive iff f is convex.

The quantile function is know to be subadditive when the distribution of X is centrally symmetric and log-concave (by Meyer and Reisner [MR91] partly based on ideas by Ball). More generally, subadditivity holds when the measure of X is κ-concave for $\kappa \geq -1$ ([Bob10]), a class that includes log-concave distributions and some heavy-tailed distributions, but we do not exploit this more general result in our paper. Finally, subadditivity also holds for a product distribution with iid standard α-stable components for $\alpha \in [1, 2)$. This follows immediately from the definition of stable distributions via characteristic functions:

By the convolution property of characteristic functions, if $X = (X_1, \dots, X_n)$ is a random vector with iid coordiantes following a standard symmetric α-stable distribution, then $\langle X, \theta \rangle / \|\theta\|_\alpha$ follows a standard symmetric α-stable distribution for all $\theta \in \mathbb{R}^n \setminus \{0\}$. Thus, $Q_q(\langle X, \theta \rangle) = \|\theta\|_\alpha Q_q(X_1)$ (which is a subadditive in θ) and $\Phi_q X$ is a scaled $l^n_{\alpha^*}$-ball with $1 = 1/\alpha^* + 1/\alpha$.

4 Quantile estimation

In this section we establish basic bounds on the error of quantile estimation from an approximation to the desired distribution.

Lemma 8. *Let F, G be distribution functions. Assume $\sup_x |F(x) - G(x)| \leq \delta$. Let $q \in (0, 1)$, $\epsilon > 0$ and $x_0 = Q_q(G)$. Assume*

1. *$G(x_0 + \epsilon) - G(x_0) \geq \delta$ and*
2. *$G(x_0) - G(x_0 - \epsilon - \alpha) > \delta$ for all $\alpha > 0$.*

Then $|Q_q(F) - Q_q(G)| \leq \epsilon$.

Proof. We get $F(x_0 + \epsilon) \geq q$, which implies $Q_q(F) \leq x_0 + \epsilon$.

For any $\alpha > 0$ we have $G(x_0) - G(x_0 - \epsilon - \alpha) > \delta$. We get $F(x_0 - \epsilon - \alpha) < q$, which implies $Q_q(F) > x_0 - \epsilon - \alpha$. The claim follows by letting $\alpha \to 0$. □

Note that assumption 2 of Lemma 8 is implied by either of the following conditions:

1. $G(x_0) - G(x_0 - \epsilon) > \delta$, or
2. $G(x_0) - G(x_0 - \epsilon) \geq \delta$ and G is strictly increasing.

Lemma 9. *Let $q \geq 1/2$. Let $\delta < 1 - q$. Let G be a distribution function with symmetric density function g supported on a (possibly infinite) interval $(-a, a)$. Let F be a distribution function satisfying $\sup_x |F(x) - G(x)| \leq \delta$. Then*

$$|Q_q(F) - Q_q(G)| \leq \frac{\delta}{g(G^{-1}(q + \delta))}.$$

Proof. We will use Lemma 8. Let x_0 be such that $q = G(x_0)$. As $x_0 \geq 0$, we have $G(x_0) - G(x_0 - \epsilon) \geq G(x_0 + \epsilon) - G(x_0)$.

We will first find ϵ such that assumption 1 holds. This happens whenever $x_0 + \epsilon \geq G^{-1}(q + \delta)$. The monotonicity of G implies $G^{-1}(q + \delta) \leq G^{-1}(q) + \delta (G^{-1})'(q + \delta) = x_0 + \delta (G^{-1})'(q + \delta)$. It is therefore sufficient to have $\epsilon \leq \delta \frac{1}{g(G^{-1}(q+\delta))}$ to satisfy assumption 1.

Let ϵ_0 be the smallest ϵ that satisfies assumption 1. We just showed that $\epsilon_0 \leq \delta \frac{1}{g(G^{-1}(q+\delta))}$. We will show that ϵ_0 also satisfies assumption 2. To see this, notice that $\epsilon_0 < a$: assumption 1 gives that ϵ_0 is smallest ϵ such that $G(x_0 + \epsilon) \geq q + \delta$ with $q + \delta < 1$. Thus, $x_0 + \epsilon_0 < a$ and therefore $\epsilon_0 < a$ (as $x_0 \geq 0$ by assumption). This implies $x_0 - \epsilon_0 > -a$, i.e. $x_0 - \epsilon_0$ is in the interior of the support of g and assumption 2 is satisfied. We can then apply Lemma 8 and the claim follows. □

5 Log-concave floating body estimation

In this section we use concentration of the empirical quantiles to study convergence of an empirical floating body to the true one. Our approach is to take a fixed sample of random vector X and consider random variable Y uniformly drawn from that sample. We then show that the floating body of Y is close to that of X with high probability after only polynomially many samples. This behavior translates membership $\Phi_q Y$ to approximate membership in $\Phi_q X$ — in the sense of Defintion 3. When the distribution of X is log-concave, the marginals given by $\langle X, \theta \rangle$, for $\theta \in \mathbb{S}^{n-1}$, are also log-concave. As discussed in Section 3, it is known that directional quantiles of log-concave distributions are subadditive, and have been studied in finance and measures of portfolio risk, for example in [Ibr09]. This means that for our algorithm, the directional quantiles coincide with the support function of $\Phi_q X$, and we can proceed by applying tools from convex geometry to recover a membership algorithm. This sidesteps the computational challenge posed by the floating body by focusing not on capturing the whole body at once, but instead being able to efficiently study whether single points are inside, opening the door, e.g., for efficient algorithms which rely on sampling.

Theorem 10 (Log-concave floating body concentration)**.** *Let $X \in \mathbb{R}^n$ be a random vector with centrally-symmetric, isotropic, log-concave distribution. Let $q \in (1/2, 1)$, $\delta > 0$, and $\epsilon \in (0, 1)$. Take $X_1, X_2, \ldots, X_N$ iid copies of X and define Y as a random vector uniform in $\{X_1, X_2, \ldots, X_N\}$. Then*

$$P\Big(d_H(\Phi_q Y, \Phi_q X) \leq \epsilon\Big) \geq 1 - \delta$$

when

$$N \geq \frac{1024}{\epsilon^2(1-q)^2}\left(n \log \frac{1024n}{\epsilon^2(1-q)^2} + \log \frac{4}{\delta}\right).$$

Proof. For random variable Z, let f_Z, F_Z, and μ_Z respectively be the associated pdf, cdf, and measure. Then we have $\mu_Y(A) = (1/N)\sum_{i=1}^{N} \mathbb{1}\{X_j \in A\}$ be the measure associated with Y. Let $\mathcal{H}$ denote the set of (affine) halfspaces in $\mathbb{R}^n$. From [VC71, Example 3] we have that the shatter function $m^{\mathcal{H}}$ of the class $\mathcal{H}$ of (affine) halfspaces in $\mathbb{R}^n$ satisfies $m^{\mathcal{H}}(l) \leq l^n + 1$. It follows from Theorem 7 that

$$P\left(\sup_{H \in \mathcal{H}} |\mu_Y(H) - \mu_X(H)| < \epsilon_1\right) \geq 1 - \delta \tag{1}$$

when

$$N \geq \frac{16}{\epsilon_1^2}\left(n \log \frac{16n}{\epsilon_1^2} + \log \frac{4}{\delta}\right). \tag{2}$$

When the event in (1) happens, we have that $\sup_t |F_{\langle Y,\theta\rangle}(t) - F_{\langle X,\theta\rangle}(t)| < \epsilon_1$ for any choice of $\theta \in \mathbb{S}^{n-1}$. From the log-concavity of X, we have that $f_{\langle X,\theta\rangle}(F^{-1}_{\langle X,\theta\rangle}(t))$ is concave and positive on $(0, 1)$, see [Bob96], Proposition A.1. This concavity together with the fact that $f_{\langle X,\theta\rangle}(0) \geq 1/8$ and $f_{\langle X,\theta\rangle}(x) \leq 1$ for all x (by isotropy, [LV07, Lemma 5.5]) implies $f_{\langle X,\theta\rangle}(F^{-1}_{\langle X,\theta\rangle}(t) \geq \min\{t, 1-t\}/4$ for all $t \in (0, 1)$. Then with $\epsilon_1 < 1 - q$, we can apply Lemma 9 to get

$$\sup_{\theta \in \mathbb{S}^{n-1}} |Q_q(\langle Y, \theta\rangle) - Q_q(\langle X, \theta\rangle)| \leq \frac{4\epsilon_1}{1 - q - \epsilon_1}.$$

Choosing $\epsilon_1 = \epsilon(1-q)/8$ gives

$$\frac{4\epsilon_1}{1 - q - \epsilon_1} \leq \frac{(1-q)/2}{1 - q - \frac{(1-q)}{8}}\epsilon = \frac{8\epsilon}{14} \leq \epsilon.$$

Finally, since $Q_q(\langle Y, \theta\rangle)$ and $Q_q(\langle X, \theta\rangle)$ are the support functions of $\Phi_q Y$ and $\Phi_q X$, a uniform bound on their difference ensures that the corresponding floating bodies have the same bound in the Hausdorff metric, see [Sch14] Theorem 1.8.11. Substituting $\epsilon_1 = \epsilon(1-q)/8$ in (2) completes the proof. □

Next, the floating body of a random vector is equivariant under invertible affine transformations:

Lemma 11. *Let $X \in \mathbb{R}^n$ be a random vector. Then for any invertible affine transformation $A : \mathbb{R}^n \to \mathbb{R}^n$ we have $\Phi_q(AX) = A\,\Phi_q(X)$.*

Proof. This is well-known, see for example [NSW18, Section 3.2.1] □

Now we can generalize Theorem 10 for linear transformations of X by using Lemma 11.

Corollary 12. *Let $X \in \mathbb{R}^n$ be a random vector with centrally-symmetric log-concave measure μ with invertible covariance matrix Σ. Let $q \in (1/2, 1)$, with $\delta > 0$, and $\epsilon \in \left(0, \sqrt{\lambda_{\max}(\Sigma)}\right)$. Take $X_1, X_2, \ldots, X_N$ iid copies of X and define Y as a random vector uniform in $\{X_1, X_2, \ldots, X_N\}$. Then*

$$P\left(d_H(\Phi_q Y, \Phi_q X) \leq \epsilon\right) \geq 1 - \delta$$

when

$$N \geq \frac{1024\lambda_{\max}(\Sigma)}{\epsilon^2(1-q)^2}\left(n \log \frac{1024n\lambda_{\max}(\Sigma)}{\epsilon^2(1-q)^2} + \log\frac{4}{\delta}\right)$$

where $\lambda_{\max}(\Sigma)$ is the largest eigenvalue of Σ.

Proof. From the linear equivariance of Φ (Lemma 11), we have that for any invertible n-by-n matrix A,

$$\begin{aligned}
d_H(\Phi_q(AY), \Phi_q(AX)) &= \sup_{\theta \in \mathbb{S}^{n-1}} \left| h_{\Phi_q(AY)}(\theta) - h_{\Phi_q(AX)}(\theta) \right| \\
&= \sup_{\theta \in \mathbb{S}^{n-1}} \|A^T\theta\|_2 \left| h_{\Phi_q Y}\left(\frac{A^T\theta}{\|A^T\theta\|_2}\right) - h_{\Phi_q X}\left(\frac{A^T\theta}{\|A^T\theta\|_2}\right)\right| \\
&= \sigma_{\max}(A) d_H(\Phi_q Y, \Phi_q X)
\end{aligned}$$

where $\sigma_{\max}(A)$ is the largest singular value of A. To see the above equalities, we refer to Section 1.7 of [Sch14] and Section 0.6 in [Gar95].

We can write $\Sigma = AA^T$, so that $X = AZ$ where Z satisfies the assumptions of Theorem 10 and Y is uniform in $\{AZ_1, AZ_2, \ldots, AZ_N\}$. Then we have

$$d_H(\Phi_q Y, \Phi_q X) = \sigma_{\max}(A) d_H(\Phi_q(A^{-1}Y), \Phi_q Z) \leq \epsilon_1 \sigma_{\max}(A)$$

with probability at least $1 - \delta$ when

$$N \geq \frac{1024}{\epsilon_1^2(1-q)^2}\left(n \log \frac{1024n}{\epsilon_1^2(1-q)^2} + \log\frac{4}{\delta}\right).$$

Choosing $\epsilon_1 = \epsilon/\sigma_{\max}(A) = \epsilon/\sqrt{\lambda_{\max}(\Sigma)}$ completes the proof. □

To estimate membership, Theorem 10 will imply efficient estimation of the support function of the floating body, which can be used together with the ellipsoid algorithm to estimate the radial function in the direction of each new sample.

Lemma 13. *Let X be a random variable with centrally-symmetric, isotropic, log-concave density. Then, for $q \in (1/2, 1)$, we have*

$$q - \frac{1}{2} < Q_q(X) \leq 1 + \log\frac{1}{2(1-q)}.$$

Proof. Let f be the density of X, supported on $(-a, a)$. Note that for all $t \in (-a, a)$,

- $P(X \leq t) \geq q \implies Q_q(X) \leq t$ and
- $P(X \leq t) \leq q \implies Q_q(X) \geq t$.

By Lemma 5.7 in [LV07] we have that $P(|X| > t) < \exp(-t + 1)$ for all $t \geq 0$. This gives

$$P(X \leq t) = 1 - P(X \geq t) \geq 1 - e^{-t+1}/2$$

when $t \geq 0$. So we get $Q_q(X) \leq t$ whenever $t \geq 1 - \ln\left(2(1-q)\right)$. For the lower bound, we use a simple bound of $P(X \in [0, t]) \leq tf(0) \leq t$ by log-concavity (see, e.g. [LV07]) so that $P(X \leq t) < q$ when $t < q - 1/2$. □

In particular, Lemma 13 implies that for a n-dimensional, centrally-symmetric, isotropic, log-concave distribution, the floating body is contained in the ball of radius $1 - \ln\big(2(1-q)\big)$ and contains the ball of radius $q - 1/2$. This generalizes to the following result:

Corollary 14. *Let X be a random variable with centrally-symmetric log-concave density f. Let F be the distribution function of X and let Σ be the covariance matrix of X. Then for $q \in (1/2, 1)$,*

$$\sqrt{\lambda_{\min}(\Sigma)}(q-1/2)B_2^n \subseteq \big(\Phi_q(X)\big)^\circ \subseteq \sqrt{\lambda_{\max}(\Sigma)}\big(1-\ln\big(2(1-q)\big)\big)B_2^n.$$

With bounds on the size of the floating body, we are ready to construct a membership oracle for the floating body. However, we only have access to the support function; to check a single query point x directly, one would have to verify that $\langle x, \theta\rangle \leq h_{\Phi_q(X)}(\theta)$ for all $\theta \in \mathbb{S}^{n-1}$. To get around this infinite set of constraints, we use the same approach as [AGNR15] and go through the polar body, where an estimate of the support serves as an estimate of the radial function, meaning only a single constraint to check.

Subroutine 1 Weak Membership oracle for $(\Phi_q\, X)^\circ$

Input: Query point $x \in \mathbb{R}^n$, samples of centrally-symmetric and log-concave random variable X, quantile q, bounds $s_M^2 \geq \lambda_{\max}(\mathrm{Cov}(X))$, $s_m^2 \leq \lambda_{\min}(\mathrm{Cov}(X))$, accuracy ϵ, confidence δ.

Output: ϵ-weak-membership decision for $x \in (\Phi_q\, X)^\circ$.

1: Generate iid samples $x^{(1)}, x^{(2)}, \ldots, x^{(N)}$ of X, for $N = \mathrm{poly}(n, 1/(1-q), s_M, 1/\epsilon, \log(1/\delta))$, as given by Corollary 12.

2: Compute the qth-quantile along $\hat{x}$, call this $\tilde{h} = \inf\{h : |j : \langle x^{(j)}, \hat{x}\rangle \leq h| \geq qN\}$.

3: Return TRUE ("Feasible") if $\|x\| \leq 1/\tilde{h}$ and FALSE otherwise.

Lemma 15 (Correctness of Subroutine 1)**.** *Let X be a random variable with centrally-symmetric, log-concave density. Let $\Sigma = \mathrm{Cov}(X)$ and suppose we have bounds $s_M^2 \geq \lambda_{\max}(\Sigma)$ and $s_m^2 \leq \lambda_{\min}(\Sigma)$. Then, given query point x, $q \in (1/2, 1)$, s_M, s_m, and $\epsilon, \delta > 0$, Subroutine 1 returns an ϵ-weak membership decision with probability $1-\delta$ with time and sample complexity* $\mathrm{poly}(n, 1/(1-q), s_M, 1/s_m, 1/\epsilon, \log(1/\delta))$.

Proof. Corollary 12 gives that the estimate $\tilde{h}$ is within ϵ_1 of the true support function h with probability at least $1-\delta$, when the number of samples is $N > \mathrm{poly}(n, 1/(1-q), s_M, 1/\epsilon_1, \log(1/\delta))$. Using the bounds provide by Corollary 14, we can invoke Lemma 4 to complete the proof. □

To construct an (ϵ, δ)-weak membership algorithm for $\Phi_q(X)$, we mimic the approach of [AGNR15]. The appropriate values of r and R come from Corollary 14 so that $rB_2^n \subseteq (\Phi_q(X))^\circ \subseteq RB_2^n$.

1. Use Subroutine 1 as an (ϵ_2, δ)-weak membership oracle $\mathsf{WMEM}_{(\Phi_q\, X)^\circ}(\epsilon_2, \delta, R, r)$. Theorem 4.3.2 of [GLS88] — reproduced as Theorem 32 here — is used to implement an (ϵ_1, δ)-weak validity oracle $\mathsf{WVAL}_{(\Phi_q\, X)^\circ}(\epsilon_1, \delta, R, r)$ which runs in oracle-polynomial time; it invokes the membership oracle $\mathsf{WMEM}_{(\Phi_q\, X)^\circ}(\epsilon_2, \delta/Q, R, r)$ a polynomial number of times for $Q = \mathrm{poly}(n, \log R)$. Then the proof of Theorem 4.3.2 in [GLS88] can be modified so that $\epsilon_2 \geq 1/\mathrm{poly}(1/\epsilon_1, R, 1/2)$.

2. Lemma 4.4.1 of [GLS88] shows how to construct an (ϵ, δ)-weak membership oracle $\mathsf{WMEM}_{\Phi_q\, X}(\epsilon, \delta, 1/r, 1/R)$ using $\mathsf{WVAL}_{(\Phi_q\, X)^\circ}(\epsilon_1, \delta, R, r)$ as a subroutine. The proof in [GLS88] shows that the membership oracle invokes the validity oracle only once, with $\epsilon_1 \geq 1/\mathrm{poly}(1/\epsilon, \|x\|, 1/r)$ where x is the query point.

This approach is accurate and efficient as needed to estimate memberhip in the floating body of a sample. The efficiency and correctness of this method follows the same analysis as Lemma 17 in [AGNR15], when combined with the correctness and efficiency of Subroutine 1.

Theorem 16. *Let X be a random vector with centrally-symmetric, log-concave density. Let $\Sigma = \mathrm{Cov}(X)$. Then, given a query point $x \in \mathbb{R}^n$, $\epsilon, \delta > 0$, and bounds $s_M^2 \geq \lambda_{\max}(\Sigma)$, $s_m^2 \leq \lambda_{\min}(\Sigma)$, there exists an algorithm which gives an ϵ-weak membership decision for $\Phi_q(X)$ with probability $1-\delta$ using time and sample complexity* $\mathrm{poly}(n, 1/(1-q), 1/s_m, s_M, 1/\epsilon, \log(1/\delta))$.

6 Stable distribution quantile estimation

The class of stable distributions is a four-parameter family which includes important distributions such as the Cauchy and Gaussian distributions. The former is of particular interest here because it is heavy-tailed in the sense that it does not have a well-defined mean, and hence no covariance or higher moments. In this section we consider the problem of estimating the quantiles of a symmetric α-stable distribution. We state a bound on the error if the quantiles are estimated from a distribution that is close to the true distribution, Lemma 18.

Lemma 17. *Let G be a standard α-stable distribution function for $0 < \alpha < 2$ and with density function g. Then there exists a constant $a = a(\alpha) > 0$ such that, for $1/2 \le q \le 1$, we have*

$$g(G^{-1}(q)) \ge a(1-q)^{1+1/\alpha}.$$

Proof. We have the following tail behavior for $0 < \alpha < 2$ as $x \to \infty$ [Nol18, Theorem 1.12]:

$$\begin{aligned} 1 - G(x) &\sim c_\alpha x^{-\alpha} \\ g(x) &\sim \alpha c_\alpha x^{-(\alpha+1)} \end{aligned}$$

(where $f(x) \sim g(x)$ as $x \to \infty$ means $f(x)/g(x) \to 1$). That is, for any $\epsilon > 0$ there exists $x_0 > 0$ such that for all $x \ge x_0$ we have $1-G(x) \le (1+\epsilon)c_\alpha x^{-\alpha}$ and $g(x) \ge (1-\epsilon)\alpha c_\alpha x^{-(\alpha+1)}$. This implies $x \ge G^{-1}(1-(1+\epsilon)c_\alpha x^{-\alpha})$. We write this inequality in terms of $q = 1-(1+\epsilon)c_\alpha x^{-\alpha}$ to get $G^{-1}(q) \le ((1+\epsilon)c_\alpha)^{1/\alpha}(1-q)^{-1/\alpha}$. This inequality holds for $q \in [q_0, 1)$, where $q_0 = G(x_0)$. We use the tail behavior of g and the fact that it is monotone decreasing in $[0,\infty)$ to get

$$\begin{aligned} g(G^{-1}(q)) &\ge g(((1+\epsilon)c_\alpha)^{1/\alpha}(1-q)^{-1/\alpha}) \\ &\ge (1-\epsilon)\alpha c_\alpha((1+\epsilon)c_\alpha)^{-(\alpha+1)/\alpha}(1-q)^{1+1/\alpha}. \end{aligned}$$

for $q \in [q_0, 1)$. Now set ϵ to any small constant (say, $\epsilon = 1/2$), and notice that $g \circ G^{-1}$ is continuous, positive and monotone decreasing in $[1/2, q_0]$. Thus, one can choose $a > 0$ so that $g(G^{-1}(q)) \ge a(1-q)^{1+1/\alpha}$ for $1/2 \le q \le 1$. □

Lemma 18. *Let G be a standard α-stable distribution function for $0 < \alpha < 2$. Let F be a distribution function satisfying $\sup_x|F(x) - G(x)| \le \delta$. Then there exists a constant $c = c(\alpha) > 0$ such that, for $q \ge 1/2$, $\delta < 1-q$, we have*

$$|Q_q(F) - Q_q(G)| \le c\delta/(1-q-\delta)^{1+1/\alpha}.$$

Proof. Combine Lemma 9 with Lemma 17. □

7 Floating body of the product of stable distributions

Here we build upon the results in Section 6 to construct a membership algorithm for the floating body when the data comes from a stable distribution. We first suppose X is a random vector with iid α-stable coordinates, and then further generalize to situations where X undergoes an unknown invertible linear transformation. From the discussion in Section 6, the support function of $\Phi_q X$ is given by $h_{\Phi_q(X)}(\theta) = \|\theta\|_\alpha Q_q(X_1)$, which implies that $(\Phi_q X)^\circ = B^n_\alpha/Q_q(X_1)$. This gives us the radial function of $(\Phi_q X)^\circ$ and efficient membership in it and, via the ellipsoid algorithm, efficient membership in $\Phi_q X$.

We start with properties that will be helpful for the ICA example.

Lemma 19. *Let $X \in \mathbb{R}^n$ be a random vector so that the coordinates have iid standard α-stable distributions. Then, for $q \in (1/2, 1)$,*

1. *$\Phi_q(X)$ is absolutely symmetric (that is, invariant under reflection through all coordinate hyperplanes).*

2. *If Y is drawn uniformly from $\Phi_q(X)$, then $\mathbb{E}\, YY^T$ is positive definite.*

3. *(Linearly equivariant) $\Phi_q(AX) = A\,\Phi_q(X)$ for any invertible $A \in \mathbb{R}^{n\times n}$.*

Proof. Parts 1 and 2 are immediate from the definitions. Part 3 is a special case of Lemma 11. □

Lemma 19 illustrates a way that the floating body still captures the "shape" of data in even the stable setting, which includes many heavy-tailed distributions. As we show in Sections 9 and 10, one application of this idea is that one can now recover an unknown linear transformation of data with stable distributions, including ones which may be (very) heavy-tailed. In general, the floating body might serve as a surrogate for data from the underlying distribution in situations that rely on estimating the moments of a sample but which are sensitive to outliers present in heavy-tailed data.

The primary result of this section is that given only polynomially-many observations of X, the empirical quantiles will provide an accurate estimate of the support function for the floating body:

Theorem 20. *Let $\alpha \in [1, 2)$. Let $X \in \mathbb{R}^n$ be a random vector with iid standard α-stable coordinates. Let $q \in (1/2, 1)$, $\delta > 0$, $\epsilon \in (0, 1]$. Let $X_1, \ldots, X_N$ be iid copies of X. Let Y be a random vector distributed uniformly in $\{X_1, \ldots, X_N\}$. Then there exists a constant $c = c(\alpha) > 0$ such that*

$$\mathbb{P}\big(d_H(\Phi_q Y, \Phi_q X) \le \epsilon\big) \ge 1 - \delta \text{ when } N \ge \frac{cn}{\epsilon^2(1-q)^4}\left(n \log \frac{n}{\epsilon(1-q)} + \log \frac{4}{\delta}\right).$$

Proof. From [VC71, Example 3] we have that the shatter function $m^{\mathcal{H}}$ of the class $\mathcal{H}$ of (affine) halfspaces in $\mathbb{R}^n$ satisfies $m^{\mathcal{H}}(l) \le l^n + 1$. Thus, from Theorem 7 we get

$$\mathbb{P}\left(\sup_{H\in\mathcal{H}} |\mu_N(H) - \mu(H)| > \epsilon_1\right) \le \delta \tag{3}$$

when

$$N \ge \frac{16}{\epsilon_1^2}\left(n \log \frac{16n}{\epsilon_1^2} + \log \frac{4}{\delta}\right). \tag{4}$$

Let $\theta \in S^{n-1}$. From the additivity property of the α-stable distribution (as in the discussion in Section 3), we have that $\langle X, \theta\rangle / \|\theta\|_\alpha$ is distributed as a standard α-stable distribution. Let G be the cdf of $\langle X, \theta\rangle / \|\theta\|_\alpha$. Let F be the cdf of $\langle Y, \theta\rangle / \|\theta\|_\alpha$. If the event in (3) does not hold, then we have $\sup_x |F(x) - G(x)| \le \epsilon_1$ (using that this distance between distributions is invariant under scaling along the x-axis). Using Lemma 18 we conclude

$$|Q_q(F) - Q_q(G)| \le \frac{c\epsilon_1}{(1 - q - \epsilon_1)^{1+1/\alpha}}.$$

for $\epsilon_1 < 1 - q$. That is,

$$|Q_q(\langle Y, \theta\rangle) - Q_q(\langle X, \theta\rangle)| \le \frac{c\epsilon_1 \|\theta\|_\alpha}{(1 - q - \epsilon_1)^{1+1/\alpha}} \le \frac{c\epsilon_1\sqrt{n}}{(1 - q - \epsilon_1)^2}.$$

We want this quantile error to be at most ϵ. We take $\epsilon_1 = c'(1-q)^2\epsilon/\sqrt{n}$ for $c' > 0$ to be determined. Using that $\epsilon \le 1$ and if we choose $c' \le 1/(2c)$, the error satisfies:

$$\frac{c\epsilon_1\sqrt{n}}{(1 - q - \epsilon_1)^2} \le \frac{cc'(1-q)^2\epsilon}{(1 - q - \frac{c'(1-q)^2}{\sqrt{n}})^2} \le \frac{\frac{1}{2}(1-q)^2}{(1 - q - \frac{c'(1-q)^2}{\sqrt{n}})^2}\epsilon.$$

If $c' \le 1/4$ then the fraction in front of ϵ is at most 1. Thus, we set $c' = \min\{1/(2c), 1/4\}$.

Since $\tilde{Q}_q$ and Q_q are the support functions of $\Phi_q Y$ and $\Phi_q X$, a uniform bound on their difference ensures that the corresponding floating bodies have the same bound in the Hausdorff metric, see [Sch14] Theorem 1.8.11. Substituting our choice of ϵ_1 in (4) completes the proof. □

Corollary 21. *Let $\alpha \in [1,2)$. Let $X = AS \in \mathbb{R}^n$ where S is a random vector with iid standard α-stable coordinates. Let $q \in (1/2, 1)$, $\delta > 0$, $\epsilon \in (0, 1]$. Let $X_1, \dots, X_N$ be iid copies of X. Let Y be a random vector distributed uniformly in $\{X_1, \dots, X_N\}$. Then there exists a constant $c = c(\alpha) > 0$ such that*

$$\mathbb{P}\big(d_H(\Phi_q Y, \Phi_q X) \le \epsilon\big) \ge 1 - \delta \text{ when } N \ge \frac{cn\sigma_{\max}^2(A)}{\epsilon^2(1-q)^4}\left(n \log \frac{n\sigma_{\max}(A)}{\epsilon(1-q)} + \log \frac{4}{\delta}\right).$$

Proof. The proof follows an identical procedure as the proof of Corollary 12, where we replace ϵ in the conclusion of Theorem 20 by $\epsilon/\sigma_{\max}(A)$. □

For the membership algorithm, Subroutine 2, we will need bounds on the size of $\big(\Phi_q X\big)^\circ$:

Lemma 22. *Let $q \in (1/2, 1)$. Let $X = AS \in \mathbb{R}^n$ where S is a random vector with each coordinate iid as a standard α-stable distribution for $\alpha \in [1,2]$, and A is an n-by-n invertible matrix. Then $(Q_q(X_1)\sigma_{\max}(A)\sqrt{n})^{-1} B_2^n \subseteq (\Phi_q(X))^\circ \subseteq (Q_q(X_1)\sigma_{\min}(A))^{-1} B_2^n$ and $Q_q(X_1)\sigma_{\min}(A) B_2^n \subseteq \Phi_q(X) \subseteq Q_q(X_1)\sigma_{\max}(A)\sqrt{n} B_2^n$.*

Proof. From the discussion above, $(\Phi_q S)^\circ$ is the l_α ball, scaled by $1/Q_q(S_1)$. Letting α vary from 1 to 2 and using linear equivariance gives the result. The second part of the conclusion follows from polarity. □

With bounds on the size of $(\Phi_q(X))^\circ$, we construct the membership algorithm, Subroutine 2:

Subroutine 2 Weak Membership oracle for $(\Phi_q X)^\circ$

Input: Query point x, samples of random vector $X = AS$ where S has iid standard α-stable coordinates quantile q, bounds $s_M \ge \sigma_{\max}(A)$, $s_m \le \sigma_{\min}(A)$, accuracy ϵ, confidence δ.

Output: ϵ-weak-membership decision for $x \in (\Phi_q X)^\circ$.

1: Generate iid samples $x^{(1)}, x^{(2)}, \dots, x^{(N)}$ of X, for $N = \mathrm{poly}(n, 1/(1-q), s_M, 1/\epsilon, \log(1/\delta))$ as needed in Corollary 21.

2: Estimate $Q_q(\langle X, \hat{x}\rangle)$ as $\tilde{h} = \inf\{h : |j : \langle x^{(j)}, \hat{x}\rangle \le h| \ge qN\}$.

3: Return TRUE ("Feasible") if $\|x\| \le 1/\tilde{h}$ and FALSE ("Infeasible") otherwise.

Lemma 23 (Correctness of Subroutine 2)**.** *Let $X = AS \in \mathbb{R}^n$ where S has a standard α-stable distribution with $\alpha \ge 1$ and A is an invertible n-by-n matrix. Let $\epsilon \in (0,1]$ and $\delta > 0$. Then, given $q \in (1/2, 1)$, $s_M \ge \sigma_{\max}(A)$, $s_m \le \sigma_{\min}(A)$, Subroutine 2 is an (ϵ, δ)-weak membership oracle for $(\Phi_q X)^\circ$ with time and sample complexity* $\mathrm{poly}(n, 1/(1-q), s_M, 1/s_m, 1/\epsilon, \log(1/\delta))$.

Proof. Corollary 21 gives that $|h_{\Phi_q(X)}(\theta) - h_{\Phi_q(Y)}(\theta)| < \epsilon_1$ for all θ with probability at least $1 - \delta$ when

$$N > \mathrm{poly}(n, 1/(1-q), s_M, 1/\epsilon_1, \log(1/\delta)).$$

So we can use Lemma 4 with Lemma 22 to complete the proof. □

Finally, from membership in the polar of the floating body, one can weakly query membership in the primal, through the same process as in the log-concave setting. That is, using Subroutine 2, we can now construct a membership oracle for the floating body of X. The algorithm is the same as in the log-concave case, see Section 5, specifically the discussion around Theorem 16.

Theorem 24. *Let $X = AS \in \mathbb{R}^n$ satisfy the requirements of Lemma 23. Let $\epsilon \in (0,1]$ and $\delta > 0$. Then, given $q \in (1/2, 1)$, $s_M \ge \sigma_{\max}(A)$, $s_m \le \sigma_{\min}(A)$, there exists an (ϵ, δ)-weak membership oracle for $\Phi_q X$ which uses and sample complexity* $\mathrm{poly}(n, 1/(1-q), s_M, 1/s_m, 1/\epsilon, \log(1/\delta))$.

8 Approximation for power-law distributions

In this section we use our understanding of the floating body of a product of stable distributions to more general distributions with power-law distributions (sometimes known as Paretian, whose tails decay approximately like $\Pr(X \geq x) \sim 1/x^\alpha$, $\alpha > 0$). The main idea is the following. Given a random vector X whose coordinates have power-law distributions of fixed exponent α, consider the floating body of $S_k = (X^{(1)} + \cdots + X^{(k)})/b_k - a_k$, where $X^{(1)}, \ldots, X^{(k)}$ are iid copies of X and a_k, b_k are constants. By the generalized central limit theorem (GCLT) as k grows, there are constants a_k, b_k so that the distribution of S_k converges to a stable distribution with the given parameter α [GK68, Pet75]. Algorithmic versions of our idea follow easily from known results about the rate of convergence in the GCLT [Sat73].

A natural application of this idea is ICA with power-law distributions and $\alpha \geq 1$ (discussed in Section 10). The restriction to $\alpha \geq 1$ is because the quantile function is convex only in this case (Section 7), and convexity leads to efficient algorithms. The case $\alpha > 1$ is already known to be efficiently solvable [AGNR15], so for clarity of presentation we only state formal results for the case $\alpha = 1$.

8.1 Approximation for product of power-law distributions with $\alpha = 1$

We first bound the error of estimating quantiles of a symmetric stable distribution from an approximation to the distribution (Lemma 25) Then we use this to show that a sum of iid copies of a random vector distributed as a product of power-law distributions with a certain tail decay behaves like a product of Cauchy distributions (Theorem 27).

Lemma 25. *Let $q \geq 1/2$. Let $\delta < 1 - q$. Let G be the standard Cauchy distribution function. Let F be a distribution with $\sup_x |F(x) - G(x)| \leq \delta$. Then*

$$|Q_q(F) - Q_q(G)| \leq \frac{\pi\delta}{4(1-q)(1-q-\delta)}.$$

Proof. We will use Lemma 8. Let x_0 be such that $q = G(x_0)$. This implies $x_0 = \cot(\pi(1-q))$. As $x_0 \geq 0$, we have $G(x_0) - G(x_0 - \epsilon) \geq G(x_0 + \epsilon) - G(x_0) = \frac{\arctan(x_0+\epsilon) - \arctan(x_0)}{\pi} = \frac{1}{\pi}\arctan(\frac{\epsilon}{1+x_0^2+x_0\epsilon})$. Given that G is strictly increasing and in view of the observation after Lemma 8, it is enough to require that this last expression be greater than or equal to δ. Solving for ϵ, we get that we can apply Lemma 8 whenever

$$\epsilon \geq \frac{(1+x_0^2)\tan(\pi\delta)}{1 - x_0\tan(\pi\delta)}$$

and $1 - x_0 \tan(\pi\delta) > 0$. This last condition is equivalent to $\delta < 1 - q$. Lemma 8 gives

$$\begin{aligned}
|Q_q(F) - x_0| &\leq (1+x_0^2)\frac{\tan(\pi\delta)}{1 - x_0\tan(\pi\delta)} \\
&= \frac{1}{\sin^2(\pi(1-q))}\frac{\tan(\pi(1-q))\tan(\pi\delta)}{\tan(\pi(1-q)) - \tan(\pi\delta)} \\
&= \frac{1}{\sin(\pi(1-q))}\frac{\sin(\pi\delta)}{\sin(\pi(1-q-\delta))}.
\end{aligned}$$

The claim follows from this and inequality $2x \leq \sin(\pi x) \leq \pi x$ for $0 \leq x \leq 1/2$. □

Given random variable X, let F_X denote the cdf of X. Given cdfs F, G, let $d(F,G) := \sup_{x\in\mathbb{R}} |F(x) - G(x)|$. The following property holds for r.v.s X, Y and any real number $c \neq 0$: $d(F_{cX}, F_{cY} = d(F_X, F_Y)$.

Lemma 26. *Let X, Y, Z be mutually independent. Then $d(F_{X+Z}, F_{Y+Z}) \leq d(F_X, F_Y)$.*

Proof. Let $t \in \mathbb{R}$. We have

$$F_{X+Z}(t) = \Pr(X + Z \leq t) = \int \Pr(X \leq t - z)\,\mathrm{d}F_Z(z)$$

and a similar inequality for Y, and

$$\left| \int \Pr(X \leq t - z)\,\mathrm{d}F_Z(z) - \int \Pr(Y \leq t - z)\,\mathrm{d}F_Z(z) \right| \leq d(F_X, F_Y).$$

Thus, $|F_{X+Z}(t) - F_{Y+Z}(t)| \leq d(F_X, F_Y)$. The claim follows. $\square$

Theorem 27. *Let X be a random vector with independent coordinates $X_1, \ldots, X_n$ with cdfs $F_1, \ldots, F_n$, respectively. Suppose that all F_is satisfy the assumptions of Theorem 33. Let $X^{(1)}, \ldots, X^{(k)}$ be iid copies of X. Let $S_k = (X^{(1)} + \cdots + X^{(k)})/k$. Let Y be a random vector with iid standard Cauchy coordinates. Let $q \in [1/2, 1]$. Then, there exists a constant $c > 0$ (that depends only on κ from Theorem 33 among F_is) such that*

$$\sup_{\theta \in S^{n-1}} |Q_q(S_k, \theta) - Q_q(Y, \theta)| \leq \frac{c}{(1-q)(1-q-c(n/k))} \frac{n}{k}.$$

(Note that the conclusion simplifies to a more interpretable $\sup_{\theta \in S^{n-1}} |Q_q(S_k, \theta) - Q_q(Y, \theta)| \leq c'' n/k$ for some c'' whenever $1 - q \geq c'$ and $cn/k \leq c'/2$ for some $c' > 0$).

Proof. Let $\theta \in S^{n-1}$. We have $Q_q(S_k, \theta) = Q_q(\langle S_k, \theta \rangle)$. Let $F = \langle S_k, \theta \rangle$, $G = \langle Y, \theta \rangle$. We have $(S_k)_i = \sum_{j=1}^k X_i^{(j)}$ so that $d((S_k)_i, \mathrm{Cauchy}(0,1)) \leq c/k$, for $i = 1, \ldots, n$ (Theorem 33). Also, $\langle S_k, \theta \rangle = \sum_{i=1}^n (S_k)_i \theta_i$ and $d((S_k)_i \theta_i, \mathrm{Cauchy}(0, \theta_i)) = d((S_k)_i, \mathrm{Cauchy}(0,1)) \leq c/k$. Lemma 26 inductively with triangle inequality implies $d(\langle S_k, \theta \rangle, G) \leq cn/k$. Lemma 25 implies $|Q_q(S_k, \theta) - Q_q(Y, \theta)| \leq c \frac{1}{(1-q)(1-q-c(n/k))} n/k$. The claim follows. $\square$

9 ICA with stable distributions for $\alpha \geq 1$

Finally, having access to the floating body means that we can generate a sample approximately uniform inside it. In this section, we give an application which uses the floating body in a broader algorithmic context. We show that one can recover an unknown linear transformation of an α-stable random vector efficiently in high dimemsions when $\alpha \geq 1$. This extends the result in [AGNR15] which requires $\alpha > 1$. We formalize this setting as follows:

Definition 28 (α-Stable ICA Model)**.** *We say that random vector X follows an α-*stable ICA model *if $X = AS$ where A is an unknown invertible matrix and the coordinates of S are iid and drawn from an α-stable distribution.*

If $X = AS$ is drawn form an α-Stable ICA model, then we general samples from its corresponding floating body, those samples will be ultimately from a scaled l_α ball. Then, we can take advantage of a reduction given in [AGR13] specifically developed to relate the ICA problem with that of learning a linearly transformed l_α ball. The procedure of this would be as follows:

1. Construct a weak membership oracle for $\Phi_q(X)$.

2. Use a random walk, e.g. [DFK91], for (approximately) uniform samples from $\Phi_q(X)$.

3. Provide samples from step 2 to Algorithm 3 in [AGR13], along with α, and get estimated matrix $\tilde{A}$. Return $\tilde{A}$.

This reduction, though convenient, suffers slightly from the need to know α *a priori*. However, if α is unknown, one can estimate it first from samples using known techniques, see, e.g. [UZ11]. We omit an analysis of the effect this would have on the total error of our algorithm. Furthermore, the choice of q here would play a role in the practical efficiency, but not its asymptotic behavior.

For a more general algorithm (that does not need to know α), one can use the approach in [AGNR15] where the samples from the centroid body of X are replaced by samples from the floating body. This will have more precise guarantees during the orthogonalization stage than the reduction to the algorithm in [AGR13], but is slightly complicated by the second, "damping", phase to recover the rotation. Combining Lemma 19 with [AGNR15, Lemma 18], we now know that if $X = AS$ satisfies Definition 28 then $\mathrm{Cov}(\Phi_q(X))^{-1/2}$ is an orthogonalizer of X. The high-level adaptation of the algorithm in [AGNR15] is shown below as Algorithm 1.

Algorithm 1 ICA For α-stable signals

Input: Samples of random vector $X = AS$ with iid standard α-stable coordinates, quantile parameter $q \in (1/2, 1)$, bounds $s_M \geq \sigma_{\max}(A)$, $s_m \leq \sigma_{\min}(A)$, accuracy ϵ, confidence δ.

Output: Matrix $\tilde{A}$

1: Use a random walk such as the one constructed by [DFK91] (provided access to a weak oracle from Subroutine 2) to generate N samples uniformly distributed in $\Phi_q(X)$.

2: Apply Algorithm 3 from [AGR13] to the samples with parameter α, to get $\tilde{A}$. Return $\tilde{A}$.

10 ICA with Cauchy tails

In Section 9 we demonstrated a solution for ICA with stable distributions when $\alpha \geq 1$. This result is based on reducing ICA to learning a linear transformation of an l_p-ball (solvable by Algorithm 3 from [AGR13]). The reduction is based on the idea that the floating body of an ICA model $X = AS$ in this case is an l_p-ball, which transforms in the same way as the underlying random vector X so that recovering the transformation of the floating body recovers A.

In this section we extend these ideas to distributions whose tails decay approximately like stable distributions with $\alpha \geq 1$ (Paretian tails whose densities decays like $1/x^{\alpha+1}$ as $x \to \infty$), without the need to be *exactly* a stable distribution. The results are an application of the rate of convergence to a stable distribution in the GCLT. The approximation of distributions is in the sense of the domain of attraction of a stable distribution in that theorem (the set of distributions such that sums of iid random variables according to any such distribution would converge to a stable distribution).

The idea of our algorithm for ICA model $X = AS$ with Paretian tails is the following: For k large enough, generate samples of $S_k = (X^{(1)} + \cdots X^{(k)})/k$. By Theorem 27, the quantile function of S_k, $Q_q(S_k, \cdot)$, is close to the quantile function of A applied to a vector with iid stable coordinates. This quantile function is convex for $\alpha \geq 1$, and therefore $Q_q(S_k, \cdot)$ is approximately convex which leads to and efficient implementation of an ϵ-membership oracle in the floating body of the transformed stable vector. This floating body is a linearly transformed l_p-ball. Algorithm 3 from [AGR13]) recovers the transformation given the oracle. For the reasons explained in Section 8, formal results are only stated for the case $\alpha = 1$.

Theorem 29. *Let X be a random vector drawn from an ICA model $X = AS$ such that coordinates S_i are iid and follow distributions F_i satisfying the assumptions of Theorem 33.Let $T_k = (X^{(1)} + \cdots + X^{(k)})/k$. Let $\bar{S}$ be a random vector with iid standard Cauchy coordinates. Let $\bar{X} = A\bar{S}$. Let Y be uniformly random in $\Phi_q(\bar{X})$. Then*

1. $\mathrm{Cov}(Y)^{-1/2}$ *is an orthogonalizer of $\bar{X}$ and X (in the sense of [AGNR15]).*
2. $\sup_{\theta \in S^{n-1}} |Q_q(T_k, \theta) - Q_q(\bar{X}, \theta)| \leq \frac{c_{33}}{(1-q)(1-q-c_{33}(n/k))} \frac{n}{k} \|A\|$.
3. *Subroutine 3 is an ϵ-weak membership oracle for $(\Phi_q \bar{X})^\circ$.*

4. Algorithm 1 with Subroutine 3 approximately recovers A.

Proof. 1. ICA models $\bar{X}$ and X use the same matrix A and therefore an orthogonalizer for $\bar{X}$ is also an orthogonalizer for X. From Lemma 19 and [AGNR15, Lemma 33], we have that $\mathrm{Cov}(Y)^{-1/2}$ is an orthogonalizer for $\bar{X}$. The claim follows.

2. This follows from Theorem 27 and an analysis of how the error changes under a linear transformation. Let $U_k = A^{-1}T_k$. Note that $U_k = (S^{(1)} + \cdots + S^{(k)})/k$ where $S^{(i)} = A^{-1}X^{(i)}$ are iid copies of S. We have

$$\begin{aligned} Q_q(T_k, \theta) - Q_q(\bar{X}, \theta) &= Q_q(\langle AU_k, \theta\rangle) - Q_q(\langle A\bar{S}, \theta\rangle) \\ &= Q_q(\langle U_k, A^T\theta\rangle) - Q_q(\langle \bar{S}, A^T\theta\rangle) \\ &= \|A^T\theta\|[Q_q(\langle U_k, (\widehat{A^T\theta})\rangle) - Q_q(\langle \bar{S}, (\widehat{A^T\theta})\rangle)]. \end{aligned}$$

Thus,

$$\sup_{\theta \in S^{n-1}} |Q_q(T_k, \theta) - Q_q(\bar{X}, \theta)| = \sup_{\theta \in S^{n-1}} \|A^T\theta\| \cdot |Q_q(\langle U_k, (\widehat{A^T\theta})\rangle) - Q_q(\langle \bar{S}, (\widehat{A^T\theta})\rangle)|.$$

The claim follows from this and Theorem 27 applied to U_k and $\bar{S}$.

3. The choice of k in Subroutine 3 implies $k \geq 2c_{33}n/(1-q)$ so that $(1 - q - c_{33}(n/k)) \geq (1-q)/2$. This and the choice of k again in part 2 imply

$$\sup_{\theta \in S^{n-1}} |Q_q(T_k, \theta) - Q_q(\bar{X}, \theta)| \leq \frac{1}{2\epsilon}.$$

With this guarantee, the rest of the analysis is the same the analysis of Subroutine 2 (with $\epsilon/2$ in place of ϵ).

4. Given samples from model $X = AS$, Subroutine 3 implements a weak membership oracle for $\Phi_q(\bar{X})^\circ$ (part 3). $\bar{X} = A\bar{S}$ is an ICA model with Cauchy components. Therefore, Algorithm 1, when given access to a weak membership oracle for $\Phi_q(\bar{X})^\circ$, recovers A.

□

Subroutine 3 Weak Membership oracle for $(\Phi_q \, \bar{X})^\circ$

Input: Query point x, samples of random vector $X = AS$ as in Theorem 29, quantile parameter q, bounds $s_M \geq \sigma_{\max}(A)$, $s_m \leq \sigma_{\min}(A)$, accuracy parameter ϵ, confidence parameter δ.

Output: ϵ-weak-membership decision for $x \in (\Phi_q \, \bar{X})^\circ$.

1: Generate iid samples $x^{(i,j)}$, $i = 1, \dots, k$, $j = 1 \dots, N$ of X, for $N = \mathrm{poly}(n, 1/(1-q), s_M, 1/\epsilon, \log(1/\delta))$ as in Subroutine 2 and $k = 2c_{33}\frac{n}{1-q}\max\{(1-q)^{-1}, 2s_M/\epsilon\}$.

2: Let $t_k^{(j)} = \sum_{i=1}^k x^{(i,j)}$, $j = 1, \dots, N$.

3: Estimate $Q_q(T_k, \hat{x})$ as $\tilde{h} = \inf\{h : |j : \langle t_k^{(j)}, \hat{x}\rangle \leq h| \geq qN\}$

4: Return TRUE if $\|x\| \leq 1/\tilde{h}$ and FALSE otherwise.

References

[AGNR15] Joseph Anderson, Navin Goyal, Anupama Nandi, and Luis Rademacher. Heavy-tailed independent component analysis. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 290–309. IEEE, 2015.

[AGNR17] Joseph Anderson, Navin Goyal, Anupama Nandi, and Luis Rademacher. Heavy-tailed analogues of the covariance matrix for ICA. In *AAAI*, pages 1712–1718, 2017.

[AGR13] Joseph Anderson, Navin Goyal, and Luis Rademacher. Efficient learning of simplices. In *Conference on Learning Theory*, pages 1020–1045, 2013.

[BL88] I. Bárány and D. G. Larman. Convex bodies, economic cap coverings, random polytopes. *Mathematika*, 35(2):274–291, 1988.

[Bob96] Sergei Bobkov. Extremal properties of half-spaces for log-concave distributions. *The Annals of Probability*, 24(1):35–48, 1996.

[Bob10] Sergey G. Bobkov. Convex bodies and norms associated to convex measures. *Probability Theory and Related Fields*, 147(1):303–332, May 2010.

[Bru18] Victor-Emmanuel Brunel. Concentration of the empirical level sets of tukey's halfspace depth. *Probability Theory and Related Fields*, May 2018.

[CJ10] Pierre Comon and Christian Jutten. *Handbook of Blind Source Separation: Independent component analysis and applications.* Academic press, 2010.

[DFK91] Martin E. Dyer, Alan M. Frieze, and Ravi Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991.

[DG92] David L. Donoho and Miriam Gasko. Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827, 1992.

[Dup22] Charles Dupin. *Applications de géométrie de mécanique à la marine, aux ponts et chaussées; pour faire suite aux développemens de géométrie.* Paris, 1822.

[Gar95] Richard J Gardner. *Geometric tomography*, volume 58. Cambridge University Press Cambridge, 1995.

[GK68] B. V. Gnedenko and A. N. Kolmogorov. *Limit distributions for sums of independent random variables.* Translated from the Russian, annotated, and revised by K. L. Chung. With appendices by J. L. Doob and P. L. Hsu. Revised edition. Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills., Ont., 1968.

[GLS88] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer, 1988.

[HO00] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4-5):411–430, 2000.

[Ibr09] Rustam Ibragimov. Portfolio diversification and value at risk under thick-tailedness. *Quantitative Finance*, 9(5):565–580, 2009.

[JP78] D.S. Johnson and F.P. Preparata. The densest hemisphere problem. *Theoretical Computer Science*, 6(1):93 – 107, 1978.

[KVV94] Michael J Kearns, Umesh Virkumar Vazirani, and Umesh Vazirani. *An introduction to computational learning theory.* MIT press, 1994.

[Lei98] Kurt Leichtweiß. *Affine geometry of convex bodies.* Wiley-VCH, 1998.

[LV07] László Lovász and Santosh Vempala. The geometry of logconcave functions and sampling algorithms. *Random Struct. Algorithms*, 30(3):307–358, 2007.

[MR91] M. Meyer and S. Reisner. Characterizations of affinely-rotation-invariant log-concave measures by section-centroid location. In *Geometric aspects of functional analysis (1989–90)*, volume 1469 of *Lecture Notes in Math.*, pages 145–152. Springer, Berlin, 1991.

[Nol18] J. P. Nolan. *Stable Distributions - Models for Heavy Tailed Data*. Birkhauser, Boston, 2018. In progress, Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html.

[NSW18] Stanislav Nagy, Carsten Schuett, and Elisabeth M. Werner. Data depth and floating body, 2018.

[Pet75] V. V. Petrov. *Sums of independent random variables*. Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[Sat73] K. Satybaldina. Absolute estimates of the rate of convergence to stable laws. *Theory of Probability & Its Applications*, 17(4):726–728, 1973.

[Sch14] Rolf Schneider. *Convex bodies: the Brunn–Minkowski theory*. Number 151. Cambridge university press, 2014.

[SSBD14] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[SW90] Carsten Schütt and Elisabeth Werner. The convex floating body. *Mathematica Scandinavica*, pages 275–290, 1990.

[Tuk75] John W Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531, 1975.

[UZ11] Vladimir V Uchaikin and Vladimir M Zolotarev. *Chance and stability: stable distributions and their applications*. Walter de Gruyter, 2011.

[VC71] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264 – 280, 1971.

A Algorithmic convexity

We use the following result: Given a membership oracle for a convex body K one can implement efficiently a membership oracle for K°, the polar of K. This follows from applications of the ellipsoid method from [GLS88]. Specifically, we use the following facts: (1) a validity oracle for K can be constructed from a membership oracle for K [GLS88, Theorem 4.3.2]; (2) a membership oracle for K° can be constructed from a validity oracle for K [GLS88, Theorem 4.4.1]. The following definitions and theorems come (occasionally with slight rephrasing) from [GLS88] except for the notion of (ϵ, δ)-weak oracle.

Definition 30 ([GLS88])**.** *The* ϵ-weak validity problem *for K is the following: Given a vector $c \in \mathbb{Q}^n$, a rational number γ, and a rational number $\epsilon > 0$, either (i) assert that $c^T x \leq \gamma + \epsilon$ for all $x \in K_{-\epsilon}$, or (ii) assert that $c^T x \geq \gamma - \epsilon$ for some $x \in K_\epsilon$. The notion of ϵ-weak validity oracle and (ϵ, δ)-weak validity oracle can be defined similarly to Definition 3.*

Definition 31 ([GLS88, Section 2.1])**.** *We say that an oracle algorithm is an* oracle-polynomial time algorithm *for a certain problem defined on a class of convex sets if the running time of the algorithm is bounded by a polynomial in the encoding length of K and in the encoding length of the possibly existing further input, for every convex set K in the given class.*

The *encoding length* of a convex set depends on how the convex set is presented; further discussion can also be found in [GLS88].

Theorem 32 (Theorem 4.3.2 in [GLS88]). *Let $R > r > 0$ and $a_0 \in \mathbb{R}^n$. There exists an oracle-polynomial time algorithm that solves the weak validity problem for every convex body $K \subseteq \mathbb{R}^n$ contained in the ball of radius R and containing a ball of radius r centered at a_0 given by a weak membership oracle. The encoding length of K is n plus the length of the binary encoding of R, r, and a_0.*

We remark that Theorem 4.3.2 as stated in [GLS88] is stronger than the above statement in that it constructs a weak violation oracle (not defined here) which gives a weak validity oracle which suffices for us. The algorithm given by Theorem 4.3.2 makes a polynomial (in the encoding length of K) number of queries to the weak membership oracle.

B Rate of convergence to Cauchy distribution

The following result gives the rate of convergence of a sum of symmetric iid random variables, whose distribution function decays at a rate of $1/x$, to the Cauchy distribution. It is a specialization of a characterization of distributions converging to a stable distribution [Pet75, Chapter IV, Section 3, Theorem 14] and a result giving the rate [Sat73].

A positive function $h(x)$ for $x > 0$ is *slowly varying* if for $c > 0$ we have $\frac{h(cx)}{h(x)} \to 0$ as $x \to \infty$.

Theorem 33 (Rate of convergence to Cauchy distribution). *Let $\xi_1, \xi_2, \ldots$ be symmetric iid random variables with cdf F. Suppose that $1 - F(x) = \frac{1+o(1)}{x} h(x)$ as $x \to \infty$, where h is slowly varying (in words, F is in the domain of attraction of symmetric stable distribution with $\alpha = 1$). Let G be the standard Cauchy cdf. Suppose $\int x \, \mathrm{d}(F(x) - G(x)) = 0$. Let $\kappa = 2 \int |x||F(x) - G(x)| \, \mathrm{d}x$. Let F_k be the cdf of $(\xi_1 + \cdots + \xi_k)/k$. Assume $\kappa < \infty$. Then there exist an absolute constant $c > 0$ such that for all $k \geq 1$, $\sup_x |F_k(x) - G(x)| \leq c \max(\kappa, \kappa^{k/(2k+1)}) k^{-1}$.*

Proof. The condition on the domain of attraction is well known in this context, see for example [Pet75, Chapter IV, Section 3, Theorem 14]. The rate is a specialization of a result from [Sat73] to the symmetric case with $\alpha = 1$. $\square$