Stereological determination of particle size distributions for similar convex bodies

Jagt, Thomas van der; Jongbloed, Geurt; Vittorietti, Martina

DOI
10.1214/24-EJS2215

Publication date
2024

Document Version
Final published version

Published in
Electronic Journal of Statistics

Electronic Journal of Statistics
Vol. 18 (2024) 742–774
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2215

Stereological determination of particle size distributions for similar convex bodies

Thomas van der Jagt[1], **Geurt Jongbloed**[1], **and Martina Vittorietti**[2,1]

[1] *Delft Institute of Applied Mathematics, Delft University of Technology*
e-mail: T.F.W.vanderJagt@tudelft.nl; G.Jongbloed@tudelft.nl; M.Vittorietti@tudelft.nl

[2] *Scienze Economiche, Aziendali e Statistiche, Università degli studi di Palermo*
e-mail: martina.vittorietti@unipa.it

Abstract: Consider an opaque medium that contains 3D particles. All particles are convex bodies of the same shape, but they vary in size. The particles are randomly positioned and oriented within the medium and cannot be observed directly. Taking a planar section of the medium we obtain a sample of observed 2D section profile areas of the intersected particles. In this paper, the distribution of interest is the underlying 3D particle size distribution for which an identifiability result is obtained. Moreover, a nonparametric estimator is proposed for this size distribution. The estimator is proven to be consistent and its performance is assessed in a simulation study.

MSC2020 subject classifications: Primary 62G05, 60D05; secondary, 60E10.
Keywords and phrases: Stereology, consistency, Mellin transform, particle system, iterative convex minorant, EM.

Received May 2023.

1. Introduction

In the classical Wicksell corpuscle problem [24] spheres are randomly positioned in an opaque body. The problem is to estimate the size distribution of the spheres using the circular profiles observed in a planar section. The motivation for the problem originated from anatomy as well as astronomy. In the anatomical setting, it is of interest as so-called follicles may be observed in slices of organs during post-mortem studies. Such follicles are approximately spherical, resulting in approximately circular section profiles from the intersected follicles. We may then wonder what is the distribution of the radii of the follicles. Questions of similar nature appear in the field of materials science. An important feature of the so-called microstructure of a steel are the grains. Knowing the size distribution of the 3D grains allows for studying the relationship between grain

arXiv: 2305.02856

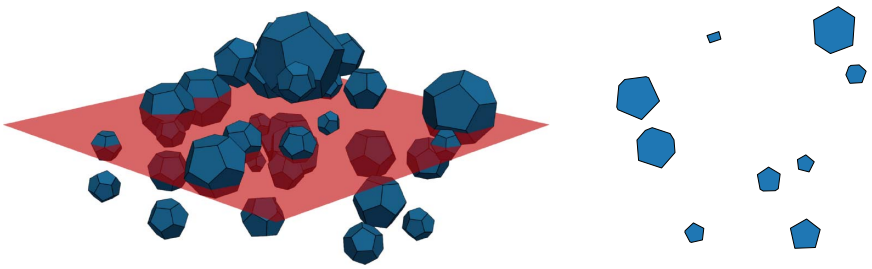

FIG 1. *Left: Random spatial system of convex dodecahedra intersected with a plane. Right: Observed section profiles.*

size distribution and the mechanical properties of the metal. It is much simpler to obtain 2D information by observing a planar cross section of the metal, compared to obtaining 3D information of the steel's microstructure. Hence, it is of interest to use the 2D observations for estimating 3D information. These problems belong to the field of stereology, which deals with the estimation of higher dimensional information from lower dimensional samples.

We study a generalization of the Wicksell problem. Consider 3D particles, convex bodies to be precise, which are randomly positioned in an opaque body and randomly oriented. A convex body is a compact and convex set with a non-empty interior. These particles all have the same known shape, but they do not have the same size. The particles cannot be observed directly, we only observe 2D section profiles of these particles in a planar section. We address the statistical problem of estimating the size distribution of the particles, using a sample of observed areas of the section profiles. A visualization of the problem setting is given in Figure 1. In this particular example, each particle is a convex dodecahedron.

An overview of estimators for the size distribution in the spherical setting is presented in [7]. The problem has been studied for shapes other than spheres as well. In [19] the case of cubic particles is considered. In [17] a variation of the problem is studied: the particles are random polyhedra, and therefore not all particles have the same shape in this setting. A system of oriented cylinders is considered in [16].

More generally, the problem we consider in this paper is a statistical inverse problem. For an overview of the key concepts in statistical inverse problems, illustrated with the deconvolution problem, we refer to [4]. Further aspects of statistical inverse problems are discussed in [6]. In this paper we also make a connection with deconvolution problems, and the proposed estimator may also be applicable in more general settings. In particular, it can be used for what may be described as a multiplicative deconvolution problem. In the classical deconvolution setting the observations are contaminated with independent additive noise, in the multiplicative setting the (positive) observations are contaminated with independent (positive) multiplicative noise. A well-known instance of this problem is multiplicative censoring. In multiplicative censoring the mul-

tiplicative noise is uniformly distributed, this model was first introduced in [22].

The main contributions of this paper are as follows. A key insight in our instance of the problem highlights that we can separate the shape of the particles from their sizes in the sense that an observed area may be interpreted as the product of two independent random variables, one related to the particle size and the other related to the known particle shape. The density function of the shape-related random variable is explicitly known only in exceptional cases, therefore we rely on the simulation procedure proposed in [21] that can be used to approximate it arbitrarily well.

Using that shape-related distribution as an ingredient, we design a maximum likelihood procedure to estimate the size distribution of the particles, a procedure that can be used for a large class of possible shapes. Furthermore, we show the consistency of the resulting estimator and provide algorithms that can be used to compute it. Additionally, we assess the proposed estimator in a small simulation study in which we focus on convex polyhedra for the shape of the particles.

The paper is organized as follows. In section 2 we introduce necessary notation and definitions. In section 3 an integral equation is derived which describes the problem. Via this equation, we obtain an identifiability result in section 4 stating that the profile area distribution uniquely determines the 3D size distribution. We define an estimator for the so-called biased size distribution in section 5. In section 6 we prove the consistency of this estimator. Algorithms for computing the proposed estimator are discussed in section 7. In section 8 we describe how to estimate the particle size distribution via the biased size distribution. In section 9 some simulations are performed and at the end of the paper we provide some conclusions in section 10.

2. Preliminaries

In this section we introduce necessary notation and collect some preliminary results which are needed in the rest of this paper. We consider a system of randomly positioned particles, and each particle is a convex body. Formally, a convex body is a convex and compact set with non-empty interior. Given a $\lambda > 0$ and a set $A \subset \mathbb{R}^3$ the scalar multiplication of λ with A is defined as: $\lambda A = \{\lambda x : x \in A\}$. Let $\mathrm{SO}(3)$ denote the rotation group of degree 3. It contains all 3×3 rotation matrices, which are orthogonal matrices of determinant 1. Some definitions are necessary to precisely describe what is meant by a random plane section of a particle. The sphere in $\mathbb{R}^3$ is given by: $S^2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\}$. The upper hemisphere in $\mathbb{R}^3$ is given by: $S^2_+ = \{(x_1, x_2, x_3) \in S^2 : x_3 \geq 0\}$. Let σ_2 denote the spherical measure on S^2, also known as the spherical Lebesgue measure on S^2. In integrals over (a subset of) S^2 the notation $\mathrm{d}\theta$ should be interpreted as $\mathrm{d}\sigma_2(\theta)$. A plane in $\mathbb{R}^3$ may be parameterized via a unit normal vector $\theta \in S^2_+$ and its signed distance $s \in \mathbb{R}$ to the origin:

$$T_{\theta,s} = \{x \in \mathbb{R}^3 : \langle x, \theta \rangle = s\}, \tag{1}$$

with $\langle \cdot, \cdot \rangle$ being the usual inner product in $\mathbb{R}^3$.

We define what is meant by an Isotropic Uniformly Random (IUR) plane hitting a given convex body in $\mathbb{R}^3$. The notion of IUR planes was introduced in [8], we follow the definition as in section 5.6 in [2].

Definition 2.1 (IUR plane)**.** An IUR plane T hitting a given convex body $K \subset \mathbb{R}^3$, is defined as $T = T_{\Theta,S}$ where (Θ, S) has joint probability density, $f_K : S^2_+ \times \mathbb{R} \to [0, \infty)$ given by:

$$f_K(\theta, s) = \begin{cases} \frac{1}{2\pi \bar{b}(K)} & \text{if } K \cap T_{\theta,s} \neq \emptyset \\ 0 & \text{otherwise,} \end{cases}$$

with $T_{\theta,s}$ as in (1) and $\bar{b}(K)$ is the mean caliper diameter, or mean width of K:

$$\bar{b}(K) := \frac{1}{2\pi} \int_{S^2_+} L(p_\theta(K)) \mathrm{d}\theta.$$

Here, $p_\theta(K)$ represents the orthogonal projection of K on the line through the origin with direction θ. $L(p_\theta(K))$ is the length of this orthogonal projection, and may also be called the width of K in direction θ. Convexity of K ensures $L(p_\theta(K))$ is the length of an interval.

Loosely speaking this means that for an IUR plane through K, every plane which has a non-empty intersection with K has equal probability of occurring. We need the following lemma, which appears as proposition 1 in [8]:

Lemma 2.2. *Suppose that $Q \subset \mathbb{R}^3$ is a convex body and $K \subset Q$ is another convex body. Let T be an IUR plane hitting Q, then:*

1. *Hitting probability:*
$$\mathbb{P}(T \cap K \neq \emptyset) = \frac{\bar{b}(K)}{\bar{b}(Q)}.$$
2. *Conditional property: Given that T hits K, i.e. $T \cap K \neq \emptyset$, T is an IUR plane hitting K.*

When a convex body K is hit by an IUR plane, we obtain a section with a random area. Let G_K denote the cumulative distribution function (CDF) associated with such a random area. It is sometimes referred to as cross section area distribution. The CDF G_K is studied in [21], and we collect the following properties:

Theorem 2.3. *Let $K \subset \mathbb{R}^3$ be a convex body and let T be an IUR plane hitting K. The random variable $Z := \mathrm{area}(K \cap T)$ has distribution function G_K. Let G_K^S denote the distribution function of $\sqrt{Z}$. The following properties hold:*

1. *Motion invariance: G_K and G_K^S are invariant under translations and rotations of K.*
2. *Scaling of convex bodies: $G_{\lambda K}(z) = G_K\left(\frac{z}{\lambda^2}\right)$ for all $\lambda > 0$, $z \in \mathbb{R}$.*

3. *Absolute continuity: If K is strictly convex or if it is a polyhedron then G_K and G_K^S have a Lebesgue density.*
4. *Initial monotonicity: If G_K^S has Lebesgue density g_K^S, then g_K^S is non-decreasing on $(0, \tau_K)$ for some $\tau_K > 0$.*

Note in particular that for a large class of convex bodies, G_K has a Lebesgue density. Whether G_K is absolutely continuous with respect to Lebesgue measure for all convex bodies is an open problem. In section 5 we define an estimator for the particle size distribution. It will then become clear why the square root transformation in Theorem 2.3 is relevant.

3. Derivation of the stereological integral equation

In this section we give a formal description of the model and derive a stereological integral equation. As mentioned in the introduction, stereology deals with estimating higher dimensional information from lower dimensional samples. The stereological equation in this section directly relates the distribution of the 3D particle sizes to the distribution of observed 2D section profile areas. We derive an expression for the density f_A of the observed section areas. Another derivation of this density appears in chapter 16 of [20]. The derivation has two purposes, it provides an intuitive understanding of the problem and the equation is used for defining an estimator.

Let $Q \subset \mathbb{R}^3$ be the opaque convex body containing the randomly positioned particles. The intersection of Q with a random plane yields a sample of observed section profile areas. For now, assume that Q contains just one particle, a convex body K_1. Assume that K_1 is similar to a known convex body $K \subset \mathbb{R}^3$, which we refer to as the reference particle. This means that there exists a rotation $M \in \mathrm{SO}(3)$, a point $x \in \mathbb{R}^3$ and a scalar $\Lambda > 0$ such that $K_1 = \Lambda MK + x := \{\Lambda Mk + x : k \in K\}$. We refer to the scalar Λ as the size of K_1, which is distributed according to an unknown size distribution on $(0, \infty)$ with CDF H and PDF h. As such the size is the scaling with respect to the reference particle, which has size 1. The mean size is denoted by:

$$\mathbb{E}(\Lambda) = \int_0^\infty \lambda h(\lambda) \mathrm{d}\lambda,$$

and we assume $0 < \mathbb{E}(\Lambda) < \infty$ throughout. Let T be an IUR plane hitting Q. Let $B := \{T \cap K_1 \neq \emptyset\}$ be the event that K_1 is hit by T. By Lemma 2.2, the probability that K_1 is hit by T given that it has size λ is given by:

$$\mathbb{P}(B|\Lambda = \lambda) = \frac{\bar{b}(\lambda K)}{\bar{b}(Q)} = \lambda \frac{\bar{b}(K)}{\bar{b}(Q)}. \tag{2}$$

Here, we use the fact that $\bar{b}(\lambda K) = \lambda \bar{b}(K)$ and $\bar{b}(K)$ is invariant under rotations and translations of K. While Λ is drawn from H, the size of a particle which appears in the plane section follows a different distribution from H. By this we mean that $\Lambda|B$ is not distributed according to H. Note that the probability in

(2) is proportional to λ, via Bayes' rule the density of $\Lambda|B$, denoted by h^b is computed as:

$$h^b(\lambda) := f_{\Lambda|B}(\lambda) = \frac{\mathbb{P}(B|\Lambda=\lambda)h(\lambda)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|\Lambda=\lambda)h(\lambda)}{\int_0^\infty \mathbb{P}(B|\Lambda=\lambda)h(\lambda)\mathrm{d}\lambda} = \frac{\lambda h(\lambda)}{\mathbb{E}(\Lambda)}.$$

Throughout this paper we refer to h^b as the density of the length-biased size distribution associated with h. Let H^b be the CDF corresponding to h^b and note that H and H^b are related via:

$$H^b(\lambda) = \frac{\int_0^\lambda x\mathrm{d}H(x)}{\int_0^\infty x\mathrm{d}H(x)} \quad \text{and} \quad H(\lambda) = \frac{\int_0^\lambda \frac{1}{x}\mathrm{d}H^b(x)}{\int_0^\infty \frac{1}{x}\mathrm{d}H^b(x)}, \quad \lambda \geq 0. \tag{3}$$

We refer to H^b as the length-biased size distribution function, or the length-biased version of H. For an elaborate overview of length-biased and more generally size-biased distributions we refer to [1]. The authors also prove the following general property of length-biased distributions: if $\Lambda_b \sim H^b$ and $\Lambda \sim H$, then: $\mathbb{P}(\Lambda_b \geq \lambda) \geq \mathbb{P}(\Lambda \geq \lambda)$. Hence, as H^b is the size distribution of the particles which appear in the plane section, this means that larger particles are more likely to appear in the cross section.

We can now derive the distribution of an observed section area, resulting from K_1 being hit by the section plane. Conditional on K_1 being hit let $A := \text{area}(K_1 \cap T)$. By the conditional property of IUR planes in Lemma 2.2, given that T hits K_1 it is an IUR plane hitting K_1. Therefore, if K_1 with size λ appears in the section plane, its section area is distributed according to $G_{\lambda K}$. Recall that the definition and some properties of $G_{\lambda K}$ are given in Theorem 2.3. Using the rules of conditional probability we find:

$$\begin{aligned} F_A(a) := \mathbb{P}(A \leq a|B) &= \int_0^\infty \mathbb{P}(A \leq a|B, \Lambda=\lambda) f_{\Lambda|B}(\lambda)\mathrm{d}\lambda \\ &= \int_0^\infty G_{\lambda K}(a)\mathrm{d}H^b(\lambda). \end{aligned}$$

Using point 2 of Theorem 2.3, F_A may be written as:

$$F_A(a) = \int_0^\infty G_K\left(\frac{a}{\lambda^2}\right)\mathrm{d}H^b(\lambda) = \frac{1}{\mathbb{E}(\Lambda)}\int_0^\infty G_K\left(\frac{a}{\lambda^2}\right)\lambda\mathrm{d}H(\lambda). \tag{4}$$

Suppose now that we randomly position and orient non-overlapping particles $K_1, K_2, \ldots$ in Q, each similar to K. More specifically, the centers of the particles are distributed according to a homogeneous Poisson point process. As for the orientations, all orientations of the particles are equally likely and independent. The sizes of the particles $\Lambda_1, \Lambda_2, \ldots$ are independent and identically distributed (iid) according to H. Intersecting Q with an IUR plane yields an iid sample $A_1, \ldots, A_n$ from F_A of observed section areas, for some random n. Let K be a convex body such that G_K has a density g_K, recall Theorem 2.3. Then, F_A has a density given by:

$$f_A(a) = \frac{1}{\mathbb{E}(\Lambda)}\int_0^\infty g_K\left(\frac{a}{\lambda^2}\right)\frac{1}{\lambda}\mathrm{d}H(\lambda). \tag{5}$$

Let $a_{\max}$ be the largest possible section area of K, such that g_K has support $(0, a_{\max})$. Hence, $g_K(a/\lambda^2) = 0$ whenever $a/\lambda^2 > a_{\max} \iff \lambda < \sqrt{a/a_{\max}}$. As a result, the lower bound of the integration region in (5) is effectively equal to $\sqrt{a/a_{\max}}$. The stereological equation (5) directly relates the sizes of the 3D particles to the areas of the observed 2D section profiles.

Example (Wicksell's corpuscle problem)**.** Choose for the reference particle $K = \bar{B}(0,1) = \{x \in \mathbb{R}^3 : \|x\| \leq 1\}$, the ball with radius 1. Then: $g_K(z) = 1/(2\pi\sqrt{1 - z/\pi})$, $0 < z < \pi$. We may interpret H as the distribution function of the radii of the 3D balls. Note that any plane section of a ball yields a circular disc. Given $A \sim f_A$ set $A = \pi R^2$, the density of the observed circle radii satisfies: $f_R(r) = f_A(\pi r^2)2\pi r$, $r > 0$. Combining this with (5) yields:

$$f_R(r) = \frac{1}{\mathbb{E}(\Lambda)} \int_r^\infty \frac{1}{2\pi\sqrt{1 - \frac{r^2}{\lambda^2}}} \frac{2\pi r}{\lambda} \mathrm{d}H(\lambda) = \frac{r}{\mathbb{E}(\Lambda)} \int_r^\infty \frac{1}{\sqrt{\lambda^2 - r^2}} \mathrm{d}H(\lambda),$$

which corresponds to the well-known Wicksell's integral equation [24].

Remark 1. By taking an appropriate choice for the reference particle, the size distribution may be directly related to a more convenient distribution. For example, if the reference particle has a diameter 1, then the size distribution corresponds to the distribution of the diameters of the particles. When choosing a reference particle with volume 1, then a particle with size λ has volume λ^3. The volume distribution function is then given by $F_V(x) = \mathbb{P}(\Lambda^3 \leq x) = H(x^{\frac{1}{3}})$.

The derived stereological equation also holds under different assumptions. The random system of particles may be defined by choosing an isotropic typical particle, and then positioning the particles using a stationary point process on $\mathbb{R}^3$. This model is also known as a germ-grain process. Relevant references are sections 6.5 and 10.5 in [7], as well as [18] and [19]. Hence, there is no need to restrict the particles to an opaque body or to position the particles via a Poisson point process.

In this setting, let N_V denote the expected number of 3D particles per unit volume, which corresponds to the intensity parameter of the point process. Intersecting the system of particles with a plane, let N_A denote the expected number of observed 2D section profiles per unit area. Here, F_A is the CDF associated with the typical section profile area. By combining the well known stereological equation (Theorem 10.1 in [7]):

$$N_A = N_V \bar{\bar{b}} \quad \text{with} \quad \bar{\bar{b}} := \bar{b}(K) \int_0^\infty \lambda \mathrm{d}H(\lambda),$$

and (4), yields:

$$N_A(1 - F_A(a)) = N_V \bar{b}(K) \int_0^\infty \lambda(1 - G_{\lambda K}(a))\mathrm{d}H(\lambda). \tag{6}$$

A derivation of a slightly more general version of (6) may be found in chapter 6 of [3]. We specifically mention (6) since it appears more frequently in the literature than (5).

In order to obtain a better understanding of the problem it helps to apply a transformation. We apply a square root transformation to (4). For $A \sim F_A$, set $S = \sqrt{A}$ such that $S \sim F_S$ and $F_S(s) = F_A(s^2)$ for $s \in \mathbb{R}$. As in Theorem 2.3, let $Z \sim G_K$ then $\sqrt{Z} \sim G_K^S$ and $G_K^S(z) = G_K(z^2)$ for $z \in \mathbb{R}$. Let $s \in \mathbb{R}$, then the following holds:

$$F_S(s) = \int_0^\infty G_K^S\left(\frac{s}{\lambda}\right) \mathrm{d}H^b(\lambda). \tag{7}$$

This expression may be recognized as the distribution function corresponding to a product of two independent random variables. This is a key insight which is made precise in the following lemma.

Lemma 3.1. *Consider a distribution function H with length-biased version H^b. Suppose $Z \sim G_K$ and $\Lambda_b \sim H^b$ with Z and Λ_b^2 independent. Set $A = Z\Lambda_b^2$. Then, $A \sim F_A$, and F_A, G_K and H^b are related via (4).*

Proof. Let X, Y, Z be non-negative random variables, with CDF F_X, F_Y and F_Z respectively. If $X = YZ$ with Y and Z independent, then their distribution functions are related via:

$$F_X(x) = \int_0^\infty F_Y\left(\frac{x}{z}\right) \mathrm{d}F_Z(z).$$

Comparing this with (7), the result is immediate. □

Let us provide some further intuition for Lemma 3.1. Note that point 2 of Theorem 2.3 means that for a given size $\lambda > 0$, if $Z \sim G_K$ then $Z\lambda^2 \sim G_{\lambda K}$. As the sizes of the particles in the section plane are distributed according to H^b, this hints towards the relationship given in Lemma 3.1.

Therefore, there are two main considerations in this problem. First, the size distribution of particles appearing in the cross section is a length-biased version of the actual size distribution. Second, we can separate the common shape of the particles and their sizes in some sense. Taking a random size from H^b, and independently taking an IUR section of the reference particle yields a sample from F_A via the relationship given in Lemma 3.1.

4. Identifiability of the particle size distribution

In this section, we present a general identifiability result for our model. This means that under appropriate conditions, given a known reference particle, there are no two size distributions which yield the same distribution of observed section areas. For this result we need the Mellin-Stieltjes transform, which we will also refer to as the Mellin transform. While characteristic functions appear naturally when studying sums of independent random variables, the Mellin transform is appropriate when studying products of independent random variables. We collect some properties of the Mellin transform, for details we refer to section 7.8 in [14] and [25]. We note that the use of the Mellin transform for this problem was already considered in [15]. The authors obtain a slightly different expression due to the fact that an inversion formula for the density h was

derived and because the density f_A in (5) was studied up to a normalization constant. The identifiability result in this section is new, a sufficient condition for identifiability in this context has not been derived before.

Definition 4.1 (Mellin-Stieltjes transform)**.** Given a non-negative random variable X, with CDF F, the Mellin-Stieltjes transform of X is defined as:

$$\mathcal{M}_X(s) = \mathbb{E}(X^{s-1}) = \int_0^\infty x^{s-1} \mathrm{d}F(x),$$

for $s \in \mathbb{C}$, whenever the integral is absolutely convergent.

Note in particular, that whenever $\int_0^\infty x^{c-1}\mathrm{d}F(x) < \infty$ for some $c \in \mathbb{R}$, then the Mellin transform exists for all $s = c + it$, $t \in \mathbb{R}$. Hence, existence of the Mellin transform corresponds to the existence of moments of a distribution. Let $\mathrm{St}(\alpha, \beta) := \{s \in \mathbb{C} : \alpha < \mathrm{Re}(s) < \beta\}$ denote the open strip parallel to the imaginary axis. Analogously, $\mathrm{St}[\alpha, \beta] := \{s \in \mathbb{C} : \alpha \leq \mathrm{Re}(s) \leq \beta\}$ denotes the closed strip. If we find $\alpha < \beta$ such that the Mellin transform of X converges absolutely on $\mathrm{St}[\alpha, \beta]$, then $\mathcal{M}_X$ is analytic on $\mathrm{St}(\alpha, \beta)$. Taking α as small as possible and β as large as possible, this open strip is referred to as the strip of analyticity of $\mathcal{M}_X$. A Mellin transform uniquely determines a distribution in the following sense:

Lemma 4.2 (Uniqueness of the Mellin transform)**.** *Let $X \sim F_1$ and $Y \sim F_2$. Assume the integrals in $\mathcal{M}_X$ and $\mathcal{M}_Y$ converge absolutely on $\mathrm{St}[\alpha, \beta]$, $0 \leq \alpha < \beta$. If $c \in (\alpha, \beta)$ and $\mathcal{M}_X(c + it) = \mathcal{M}_Y(c + it)$ for all $t \in \mathbb{R}$ then $F_1 = F_2$.*

The proof is given in Appendix A. A similar statement is proven in Theorem 8 in [5] for the case that the CDF has a Lebesgue density. Finally, we recall the Mellin convolution theorem. Let X, Y, Z be non-negative random variables, such that $X = YZ$ with Y and Z independent. For any $s \in \mathbb{C}$ such that $\mathcal{M}_Y(s)$ and $\mathcal{M}_Z(s)$ are finite:

$$\mathcal{M}_X(s) = \mathbb{E}\left(X^{s-1}\right) = \mathbb{E}\left((YZ)^{s-1}\right) = \mathbb{E}\left(Y^{s-1}\right)\mathbb{E}\left(Z^{s-1}\right) = \mathcal{M}_Y(s)\mathcal{M}_Z(s).$$

Having collected these properties we now state the identifiability result.

Theorem 4.3 (Identifiability)**.** *Suppose we are given densities f_A, g_K such that f_A can be expressed as in (5) for some CDF H.*

1. *If $\int_0^\infty z^{-\alpha} g_K(z)\mathrm{d}z < \infty$ for some $\alpha > 0$, then there is only one distribution function H on $(0, \infty)$ satisfying (5).*
2. *Assume $\int_0^\infty x^{1+\delta}\mathrm{d}H(x) < \infty$, for some $\delta > 0$. Then, there is only one such distribution function H on $(0, \infty)$ satisfying (5).*

Proof. We first consider statement 1 of the theorem. Let $\Lambda_b \sim H^b$, with H^b as in (3), and let $\Lambda \sim H$. Let $Z \sim g_K$ and $A \sim f_A$. We first determine on which strips the Mellin transforms of the random variables of interest are analytic. Since $\mathbb{E}(\Lambda_b^{-1}) = 1/\mathbb{E}(\Lambda)$ and $\mathbb{E}(\Lambda_b^0) = 1$ we obtain that $\mathcal{M}_{\Lambda_b}$ is analytic on $\mathrm{St}(0, 1)$. Note that $1/2 < \mathrm{Re}(s) < 1 \iff 0 < \mathrm{Re}(2s - 1) < 1$. As a result:

$$\mathcal{M}_{\Lambda_b^2}(s) = \mathbb{E}\left(\Lambda_b^{2s-2}\right) = \mathcal{M}_{\Lambda_b}(2s - 1),$$

for all $s \in \mathrm{St}(1/2,1)$ and $\mathcal{M}_{\Lambda_b^2}$ is analytic on $\mathrm{St}(1/2,1)$. Choose $\alpha > 0$ such that $\mathbb{E}(Z^{-\alpha}) < \infty$. Because g_K has bounded support, all non-negative moments of Z exist and therefore $\mathcal{M}_Z$ is analytic on $\mathrm{St}(1-\alpha,\infty)$. By Lemma 3.1 and the Mellin convolution theorem we obtain:

$$\mathcal{M}_A(s) = \mathcal{M}_Z(s)\mathcal{M}_{\Lambda_b^2}(s),$$

for all $s \in \mathrm{St}(1-\alpha,\infty) \cap \mathrm{St}(1/2,1) = \mathrm{St}(\max\{1-\alpha,1/2\},1)$. Moreover, this also means that $\mathcal{M}_A$ is analytic on $\mathrm{St}(\max\{1-\alpha,1/2\},1)$. Let $c \in (\max\{1-\alpha,1/2\},1)$. Define:

$$L_Z := \{c+it : t \in \mathbb{R},\ \mathcal{M}_Z(c+it) \neq 0\}, \quad \text{and} \quad L := \{c+it : t \in \mathbb{R}\}.$$

For all $s \in L_Z$ we find: $\mathcal{M}_A(s)/\mathcal{M}_Z(s) = \mathcal{M}_{\Lambda_b^2}(s)$. Define $f : L_Z \to \mathbb{C}$ by $f(s) = \mathcal{M}_A(s)/\mathcal{M}_Z(s)$. Note that f is analytic on L_Z, because $s \mapsto \mathcal{M}_{\Lambda_b^2}(s)$ is analytic on the line L. As a result there is a unique analytic continuation of f to L. The uniqueness of this analytic continuation implies: $f(s) = \mathcal{M}_{\Lambda_b^2}(s)$, for all $s \in L$. Suppose $\bar{H}$ also satisfies (5), with $\bar{H}^b$ denoting its length-biased version and $\bar{\Lambda}_b \sim \bar{H}^b$. Then, following the same steps as before, we obtain: $f(s) = \mathcal{M}_{\bar{\Lambda}_b^2}(s)$ for all $s \in L$. By Lemma 4.2, Λ_b^2 and $\bar{\Lambda}_b^2$ have the same CDF. Therefore, for all $x \in \mathbb{R}$:

$$H^b(x) = \mathbb{P}(\Lambda_b^2 \leq x^2) = \mathbb{P}(\bar{\Lambda}_b^2 \leq x^2) = \bar{H}^b(x).$$

By (3) this also implies $H = \bar{H}$.

The proof of the second statement of the theorem is analogous, we simply highlight the differences. Let $\delta > 0$ be such that $\mathbb{E}(\Lambda^{1+\delta}) < \infty$. Note that $\mathbb{E}(\Lambda_b^{-1}) = 1/\mathbb{E}(\Lambda)$ and $\mathbb{E}(\Lambda_b^{\delta}) = \mathbb{E}(\Lambda^{\delta+1})/\mathbb{E}(\Lambda)$. It then follows that $\mathcal{M}_{\Lambda_b}$ is analytic on $\mathrm{St}(0,1+\delta)$ and $\mathcal{M}_{\Lambda_b^2}$ is analytic on $\mathrm{St}(1/2,1+\delta/2)$. Clearly, $\mathcal{M}_Z$ is analytic on $\mathrm{St}(1,\infty)$. Hence, $\mathcal{M}_A$ is analytic on $\mathrm{St}(1/2,1+\delta/2) \cap \mathrm{St}(1,\infty) = \mathrm{St}(1,1+\delta/2)$. In this case we take $c \in (1,1+\delta/2)$ and the remainder of the proof is as before. □

We obtain as a consequence:

Corollary 1. *In the following cases the distribution function H is identifiable:*

1. *The Wicksell corpuscle problem.*
2. *G_K^S has a bounded density g_K^S.*

Proof. Recall the expression for g_K in Example 3. Then:

$$\int_0^{\pi} z^{-\frac{1}{2}} g_K(z)\mathrm{d}z = \int_0^{\pi} \frac{1}{2\pi\sqrt{z}\sqrt{1-\frac{z}{\pi}}}\mathrm{d}z = \frac{\sqrt{\pi}}{2}.$$

This integral may be computed by substituting $t = z/\pi$ and recognizing the resulting integral as an integral of a constant times the density of a Beta distribution. Hence, condition 1 of Theorem 4.3 is satisfied. If G_K^S has a density g_K^S with $g_K^S \leq m$, then:

$$\int_0^{a_{\max}} z^{-\frac{1}{4}} g_k(z)\mathrm{d}z = \int_0^{\sqrt{a_{\max}}} z^{-\frac{1}{2}} g_k^S(z)\mathrm{d}z \leq m \int_0^{\sqrt{a_{\max}}} z^{-\frac{1}{2}}\mathrm{d}z = 2m(a_{\max})^{\frac{1}{4}}.$$

Therefore, in this case condition 1 of Theorem 4.3 is also satisfied. □

Identifiability for the Wicksell problem is a classical result, in this case there is also a well-known explicit inverse relation.

Remark 2. The condition in Theorem 4.3: $\int_0^\infty x^{1+\delta} \mathrm{d}H(x) < \infty$, for some $\delta > 0$, is also implied by the assumption $H(M) = 1$ for some $M > 0$. Recall the derivation of (5) in section 3, a maximum size of the particles is clearly enforced by that fact that they are contained within the body Q. This is a typical assumption in stereological problems.

Note that the proof of Theorem 4.3 also presents (a rather implicit) inversion formula for H^b. Assume H^b is continuous. Let c be as in the proof of Theorem 4.3. Since analytic functions only have isolated zeros, $\mathcal{M}_Z(c + it) \neq 0$ for almost all $t \in \mathbb{R}$. By using the Mellin inversion formula as in the proof of Lemma 4.2:

$$H^b\left(\sqrt{x}\right) = \mathbb{P}\left(\Lambda_b^2 \leq x\right) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} -\frac{\mathcal{M}_A(s)}{\mathcal{M}_Z(s)} \frac{x^{-s+1}}{s} \mathrm{d}s, \quad x \geq 0. \tag{8}$$

H can then be retrieved via (3). We note that in [15] another expression for h was derived in terms of (inverse) Mellin transforms.

5. Estimator for the length-biased particle size distribution

In this section, we propose an estimator for the length-biased size distribution H^b. The proposed estimator is inspired by the approach taken in [13], for Wicksell's corpuscle problem. Given the random fraction interpretation of Lemma 3.1, first estimating H^b seems a natural intermediate step. We note that biased or weighted distributions frequently appear in stereology, see also section 7.5 in [18]. For the remainder of the paper we consider the number of observed section profiles n to be deterministic. This is a common convention for problems of this type. This means that we condition on a fixed number of particles being hit by the section plane. All further analysis is based on an iid sample of size n from f_A.

The reference particle K is considered to be known and we assume that it satisfies one of the conditions in Theorem 2.3 such that G_K^S has a density g_K^S. We stress that this also means that we consider g_K^S to be known. While there are very few shapes for which an explicit expression is known for g_K^S, in [21] a Monte Carlo simulation scheme is proposed which can be used to approximate such a density arbitrarily closely. To give some insight in how these densities look, see Figure 2 for approximations of these densities for the cube, dodecahedron and tetrahedron. These approximations are obtained by computing a kernel density estimator with boundary correction, based on a sample of size $N = 10^7$.

Recall the square root transformation and the resulting expression (7). Because G_K^S has density g_K^S, F_S has a density f_S given by:

$$f_S(s) = \int_0^\infty g_K^S\left(\frac{s}{\lambda}\right) \frac{1}{\lambda} \mathrm{d}H^b(\lambda). \tag{9}$$

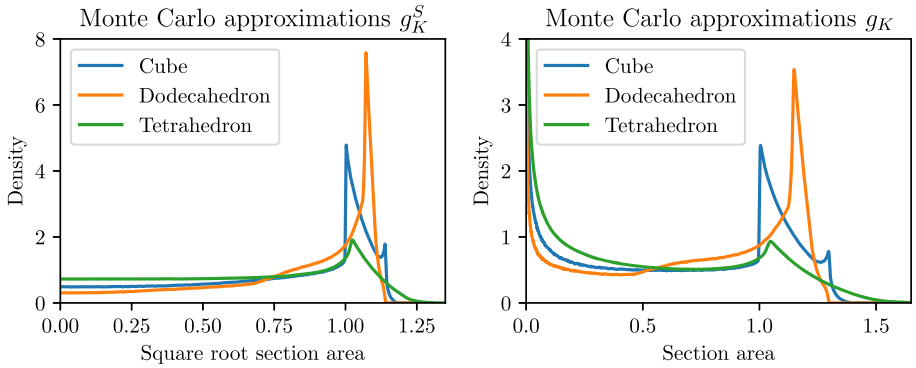

FIG 2. *Left: Monte Carlo approximations of g_K^S, for various shapes K of unit volume. Right: Approximations of g_K, obtained via $g_K(z) = g_K^S(\sqrt{z})/(2\sqrt{z})$.*

Keep in mind that g_K^S is supported on $(0, \sqrt{a_{\max}})$, such that the lower bound of the integration region is effectively $s/\sqrt{a_{\max}}$.

Suppose we have a sample of observed section areas: $A_1, \ldots, A_n \overset{\text{iid}}{\sim} f_A$. Let $S_i = \sqrt{A_i}$, then $S_1, \ldots, S_n \overset{\text{iid}}{\sim} f_S$, with f_S as in (9). Now, let $s_1 < s_2 < \cdots < s_n$ be a realization of the order statistics of $S_1, \ldots, S_n$. We use (9) to implicitly define an estimator for H^b via nonparametric maximum likelihood. This is achieved by considering a large class of distribution functions for H^b. Let $\mathcal{F}^+$ be the class of all distribution functions on $(0, \infty)$. Define:

$$\mathcal{F}_n^+ = \{F \in \mathcal{F}^+ : F \text{ is constant on } [s_{i-1}, s_i), i \in \{1, \ldots, n\}, \text{with } F(s_0) = 0\},$$

for some $0 < s_0 < s_1$. This means that $\mathcal{F}_n^+$ contains all piece-wise constant distribution functions with jump locations restricted to the set of observations, the s_i's. Note that as $n \to \infty$ the set of observed s_i's becomes dense in the support of f_S and the class $\mathcal{F}_n^+$ grows to the class of all distribution functions with the same support as f_S.

Remark 3. If $H^b(M) = 1$ for some $M > 0$, then f_S is supported on the interval $(0, M\sqrt{a_{\max}})$. When choosing the size of the reference particle K, it is important that $a_{\max} \geq 1$. This is due to the choice of the sieve $\mathcal{F}_n^+$. Then, as n tends to infinity $\mathcal{F}_n^+$ grows to the class of distribution functions which also contains the true CDF H^b, since $M\sqrt{a_{\max}} \geq M$. Taking a very large K means $a_{\max}$ is large, such that $M\sqrt{a_{\max}}$ is much larger than M. Then, the s_i's will be quite sparse in $[0, M]$, which is also undesirable. For the sake of interpretability of H, recall Remark 1, we choose a K with volume 1. For the shapes considered in simulations we observed $a_{\max} \geq 1$.

For $H^b \in \mathcal{F}_n^+$ we define the (scaled by $\frac{1}{n}$) log-likelihood:

$$L(H^b) := \frac{1}{n} \sum_{i=1}^{n} \log(f_S(s_i)) = \frac{1}{n} \sum_{i=1}^{n} \log \left(\int_0^\infty g_K^S \left(\frac{s_i}{\lambda} \right) \frac{1}{\lambda} \mathrm{d}H^b(\lambda) \right). \tag{10}$$

A maximum likelihood estimator (MLE) $\hat{H}_n^b$ for H^b is defined as a maximizer of the log-likelihood L, which may be written as:

$$\hat{H}_n^b \in \underset{H^b \in \mathcal{F}_n^+}{\arg\max} \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} g_K^S \left(\frac{s_i}{s_j} \right) \frac{1}{s_j} (H^b(s_j) - H^b(s_{j-1})) \right). \tag{11}$$

The following theorem shows that this estimator is well-defined, and provides a sufficient condition for uniqueness:

Theorem 5.1 (Existence and uniqueness of $\hat{H}_n^b$)**.** *A maximizer of the log-likelihood L in $\mathcal{F}_n^+$ always exists. The maximizer is unique if the matrix $A = (\alpha_{i,j})$, with $\alpha_{i,j} = g_K^S(s_i/s_j)/s_j$, $i, j \in \{1, \ldots, n\}$, is full-rank.*

Proof. For $H^b \in \mathcal{F}_n^+$ define: $\beta_j = H^b(s_j)$ and write $\beta = (\beta_1, \beta_2, \ldots, \beta_n)^\intercal$. Consider the closed convex set:

$$\mathcal{C} := \{\beta \in \mathbb{R}^n : 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_n \leq 1\}. \tag{12}$$

The maximization problem (11) is equivalent to maximizing $l : \mathcal{C} \to \mathbb{R} \cup \{-\infty\}$ with l given by:

$$l(\beta) = \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \alpha_{i,j} (\beta_j - \beta_{j-1}) \right), \tag{13}$$

where $\alpha_{i,j} = g_K^S(s_i/s_j)/s_j$ and $\beta_0 = 0$. The set $\mathcal{C}$ is closed and bounded, and therefore compact. Because of the continuity of l on $\mathcal{C}$, it has a maximum. We now show that l is strictly concave if and only if $A = (\alpha_{i,j})$ is full-rank. Strict concavity implies uniqueness of the maximum as well as the maximizer. Fix $\beta \in \mathcal{C}$ such that $l(\beta) > -\infty$. Let $j, k \in \{1, \ldots, n\}$, computing the partial derivatives and Hessian of l yields:

$$\frac{\partial}{\partial \beta_j} l(\beta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_{i,j} - \alpha_{i,j+1}}{\sum_{q=1}^{n} \alpha_{i,q} (\beta_q - \beta_{q-1})} \tag{14}$$

$$\frac{\partial^2}{\partial \beta_j \partial \beta_k} l(\beta) = -\frac{1}{n} \sum_{i=1}^{n} \frac{(\alpha_{i,j} - \alpha_{i,j+1})(\alpha_{i,k} - \alpha_{i,k+1})}{\left(\sum_{q=1}^{n} \alpha_{i,q} (\beta_q - \beta_{q-1}) \right)^2} =: H_{j,k}(\beta). \tag{15}$$

Here, we have set $\alpha_{i,n+1} = 0$ for all $i \in \{1, \ldots, n\}$. Since $l(\beta) > -\infty$, there are no divisions by zero in (14) and (15). Note that the following holds for $k \in \{1, \ldots, n\}$:

$$\sum_{j=1}^{n} \alpha_{k,j} (\beta_j - \beta_{j-1}) = \sum_{j=1}^{n} \alpha_{k,j} \beta_j - \sum_{j=0}^{n-1} \alpha_{k,j+1} \beta_j = \sum_{j=1}^{n} (\alpha_{k,j} - \alpha_{k,j+1}) \beta_j.$$

Using this fact we show that the Hessian of l is negative definite if and only if

A is full-rank. Let $\gamma \in \mathbb{R}^n$ and set $\gamma_0 = 0$, then:

$$\begin{aligned}\gamma^{\mathsf{T}} H(\beta)\gamma &= \sum_{j=1}^{n}\sum_{k=1}^{n} H_{j,k}(\beta)\gamma_j\gamma_k \\ &= -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{(\alpha_{i,j}-\alpha_{i,j+1})(\alpha_{i,k}-\alpha_{i,k+1})\gamma_j\gamma_k}{\left(\sum_{q=1}^{n}\alpha_{i,q}(\beta_q-\beta_{q-1})\right)^2} \\ &= -\frac{1}{n}\sum_{i=1}^{n} \frac{\left(\sum_{j=1}^{n}\alpha_{i,j}(\gamma_j-\gamma_{j-1})\right)^2}{\left(\sum_{q=1}^{n}\alpha_{i,q}(\beta_q-\beta_{q-1})\right)^2}\end{aligned}$$

Clearly, $H(\beta)$ is negative semidefinite. Note that the following holds:

$$\gamma^{\mathsf{T}} H(\beta)\gamma = 0 \iff \sum_{j=1}^{n}\alpha_{i,j}(\gamma_j-\gamma_{j-1}) = 0, \quad \text{for all } i \in \{1,\dots,n\}. \tag{16}$$

Define $x \in \mathbb{R}^n$ via $x_j = \gamma_j - \gamma_{j-1}$, $j \in \{1,\dots,n\}$. Consider the matrix $A = (\alpha_{i,j})$, then the RHS of (16) may be written as $Ax = 0$. Since $\gamma_0 = 0$: $\gamma = 0 \iff x = 0$. Therefore, the Hessian is negative definite if and only if $Ax = 0 \iff x = 0$, which corresponds to A being full-rank. □

Recall that g_K^S is supported on $(0, \sqrt{a_{\max}})$. Suppose we choose the reference particle such that $\sqrt{a_{\max}} = 1+\epsilon$ for some small $\epsilon > 0$. Then, $g_K^S(1) > 0$ ensuring that the diagonal of A contains positive entries. Whenever $s_i/s_j > 1+\epsilon$ for all $i > j$, A is an upper triangular matrix, because all entries below the diagonal are zero. It is well-known that such matrices are of full-rank. If $\epsilon > 0$ is chosen sufficiently small, then with high probability $s_{j+1}/s_j > 1+\epsilon$ for all $j \in \{1,\dots,n\}$, such that the MLE is unique with high probability. For the sake of convenience, we will refer to $\hat{H}_n^b$ as the MLE, even though we cannot always guarantee uniqueness. Note especially for the consistency result in the next section that consistency of the MLE should be interpreted as consistency of any sequence of MLE's.

From the proof of Theorem 5.1 it is clear that $\hat{H}_n^b$ may be computed by maximizing l. Because l is a concave function, computing $\hat{H}_n^b$ can be done efficiently as we will discuss in section 7.

6. Consistency of the maximum likelihood estimator

In this section, we show that the MLE $\hat{H}_n^b$ (11) for H^b is uniformly strongly consistent. In order to prove this, we transform the problem into a deconvolution problem. Deconvolution problems have been studied before quite extensively, see for example [10] and [11]. We use some results on estimators in deconvolution problems to show consistency of the MLE. In deconvolution problems, it is typical that assumptions are made on the so-called noise kernel to ensure

consistency. For this problem, this translates into assumptions on the density g_K^S.

We start by rewriting the problem of estimating H^b into a deconvolution problem. Recall Lemma 3.1, for $S \sim f_S$, $\sqrt{Z} \sim g_K^S$ and $\Lambda_b \sim H^b$ we have: $S \overset{d}{=} \sqrt{Z}\Lambda_b$, with $\sqrt{Z}$ and Λ_b independent. Let us now perform a log-transformation, define: $Y = \log(S)$, $\epsilon = \log(\sqrt{Z})$ and $X = \log(\Lambda_b)$. The densities of Y and ϵ are related to those of S and $\sqrt{Z}$ by: $f_Y(y) = f_S(e^y)e^y$, $f_\epsilon(z) = g_K^S(e^z)e^z$. The distribution function of X is given by $F_X(x) = H^b(e^x)$. We then obtain:

$$Y \overset{d}{=} X + \epsilon,$$

with X and ϵ independent. Note that f_Y is the convolution of f_ε and F_X:

$$f_Y(y) = \int_{-\infty}^{\infty} f_\varepsilon(y - x)\mathrm{d}F_X(x) =: (f_\varepsilon * \mathrm{d}F_X)(y). \tag{17}$$

In this setting, F_X is the distribution function of interest. We do not have direct observations from F_X, there is additive noise from the known distribution of ϵ. Let $\mathcal{F}$ be the class of all distribution functions on $\mathbb{R}$. Define:

$$\mathcal{F}_n = \{F \in \mathcal{F} : F \text{ is constant on } [y_{i-1}, y_i), \text{ for } i \in \{1, \ldots, n\}, \text{with } F(y_0) = 0\}.$$

The observed order statistics $s_1, \ldots, s_n$ are transformed as well: $y_i = \log(s_i)$, $i \in \{0, 1, \ldots, n\}$. We proceed similarly as before, the log-likelihood may be written as:

$$\tilde{L}(F) = \frac{1}{n}\sum_{i=1}^{n} \log\left(f_Y(y_i)\right) = \frac{1}{n}\sum_{i=1}^{n} \log\left(\int_{-\infty}^{\infty} f_\epsilon\left(y_i - x\right)\mathrm{d}F_X(x)\right).$$

A maximum likelihood estimator $\hat{F}_n$ for F_X is defined as:

$$\hat{F}_n \in \underset{F_X \in \mathcal{F}_n}{\arg\max}\, \tilde{L}(F_X). \tag{18}$$

We now show that we may assume $\hat{H}_n^b(x) = \hat{F}_n(\log(x))$. The likelihoods of the two problems are related as follows. Let $F_X \in \mathcal{F}_n$, define $H^b(x) = F_X(\log(x))$ such that $H^b \in \mathcal{F}_n^+$. Then:

$$\begin{aligned}
\tilde{L}(F_X) &= \frac{1}{n}\sum_{i=1}^{n} \log\left(\sum_{j=1}^{n} f_\epsilon\left(\log(s_i) - \log(s_j)\right)\left(F_X(\log(s_j)) - F_X(\log(s_{j-1}))\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \log\left(\sum_{j=1}^{n} g_K^S\left(\frac{s_i}{s_j}\right)\frac{s_i}{s_j}(H^b(s_j) - H^b(s_{j-1}))\right) \\
&= L(H^b) + \frac{1}{n}\sum_{i=1}^{n} \log(s_i).
\end{aligned}$$

If we find a distribution function which provides a better likelihood in one of the problems, we immediately obtain a distribution function which provides the same improvement in likelihood in the other problem. So indeed, there exist MLE's which are related via: $\hat{H}_n^b(x) = \hat{F}_n(\log(x))$. The estimator $\hat{F}_n$ was studied in [10] and shown to be strongly uniformly consistent under some conditions on f_ε. This result may be used to show strong uniform consistency of $\hat{H}_n^b$, since:

$$\sup_{x>0} |\hat{H}_n^b(x) - H^b(x)| = \sup_{x>0} |\hat{F}_n(\log(x)) - F_X(\log(x))| = \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F_X(x)|.$$

Let us now specify the assumptions we require for f_ϵ. We assume it belongs to the class $\mathcal{G}$ of upper semicontinuous functions that are of bounded variation on some compact interval and monotone outside this interval. Let $V_a^b(f)$ denote the total variation of the function f on the interval $[a, b]$, $a < b$. The class $\mathcal{G}$ may be written as:

$$\mathcal{G} = \Big\{ g : \mathbb{R} \to [0,\infty) : g \text{ is an upper semicontinuous density such that } \exists M > 0 \\ \text{with } V_{-M}^{M}(g) < \infty \text{ and } g \text{ is monotone on } (-\infty, -M] \text{ and } [M,\infty) \Big\}.$$

This corresponds with the following assumptions on g_K^S:

Lemma 6.1. *Assume that g_K^S is upper semicontinuous and of bounded variation on its support. Then, the density $f_\epsilon : \mathbb{R} \to [0,\infty)$ given by $f_\epsilon(z) = g_K^S(e^z)e^z$ belongs to $\mathcal{G}$.*

The proof of this lemma can be found in Appendix A. We now collect some lemmas to obtain a consistency result for $\hat{F}_n$. The following result can be found in [10], as Corollary 1:

Lemma 6.2. *Let $\hat{F}_n$ be the MLE for F_X defined in (18). Assume $f_\epsilon \in \mathcal{G}$. Set $\hat{f}_n := f_\epsilon * \mathrm{d}\hat{F}_n$, $f_Y := f_\epsilon * \mathrm{d}F_X$, then almost surely:*

$$\lim_{n\to\infty} \|\hat{f}_n - f_Y\|_{L_1} = \lim_{n\to\infty} \int \left|\hat{f}_n(s) - f_Y(s)\right| \mathrm{d}s = 0.$$

The following lemma is a generalization of Lemma 3 in [10]. The proof is given in Appendix A.

Lemma 6.3. *Let f_ε be a Lebesgue density on $\mathbb{R}$. Let $(F_n)_{n\geq 1}$ be a sequence of distribution functions on $\mathbb{R}$, converging weakly to a distribution function F_X. Then, for $f_n := f_\epsilon * \mathrm{d}F_n$ and $f_Y := f_\epsilon * \mathrm{d}F_X$:*

$$\lim_{n\to\infty} \|f_n - f_Y\|_{L_1} = 0.$$

We now state the following theorem, which closely follows the proof of Theorem 3 in [10].

Theorem 6.4 (Consistency of $\hat{F}_n$)**.** *Let $f_\epsilon \in \mathcal{G}$. Assume that the deconvolution problem with this f_ϵ is identifiable. Then, with probability one, $\lim_{n\to\infty} \hat{F}_n(x) = F_X(x)$ for each x where F_X is continuous. If F_X is continuous then with probability one:*

$$\lim_{n\to\infty} \left\| \hat{F}_n - F_X \right\|_\infty = 0.$$

Proof. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space supporting a sequence $Y_1, Y_2, \ldots$ of iid random variables, distributed according to f_Y as in (17). Set $\hat{f}_n := f_\epsilon * \mathrm{d}\hat{F}_n$. By Lemma 6.2 we know there exists a set $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ we have $\|\hat{f}_n(\cdot, \omega) - f_Y(\cdot)\|_{L_1} \to 0$ as $n \to \infty$. Fix $\omega \in \Omega_0$ and choose an arbitrary subsequence $(n_l)_{l\geq 1} \subset (n)_{n\geq 1}$. By Helly's selection principle, there exists a further subsequence $(n_k)_{k\geq 1} \subset (n_l)_{l\geq 1}$ such that $\hat{F}_{n_k}(\cdot, \omega)$ converges weakly to a distribution function F. By Lemma 6.3 this implies that $\hat{f}_{n_k}$ converges to $f_\epsilon * \mathrm{d}F$ in L_1. Because the whole sequence $\hat{f}_n$ converges to $f_\epsilon * \mathrm{d}F_X$ in L_1 this implies $F = F_X$ by identifiability of the deconvolution problem. Therefore, every subsequence of MLE's contains a further subsequence converging weakly to F_X. This implies weak convergence of the whole sequence to F_X. Finally, the uniform result follows from the monotonicity of all distribution functions in the sequence and F_X, and continuity of F_X. □

Turning to a consistency result for $\hat{H}_n^b$ as in (11), we need to make sure that g_K^S satisfies the conditions in Lemma 6.1. If g_k^S satisfies these conditions, its boundedness implies the problem is identifiable by Corollary 1. Note that identifiability in the original problem implies identifiability in the corresponding deconvolution problem.

Corollary 2 (Consistency of $\hat{H}_n^b$)**.** *Assume g_K^S is upper semicontinuous and of bounded variation on its support. Then, with probability one, $\lim_{n\to\infty} \hat{H}_n^b(\lambda) = H^b(\lambda)$ for each λ where H^b is continuous. If H is continuous, then so is H^b, and with probability one:*

$$\lim_{n\to\infty} \left\| \hat{H}_n^b - H^b \right\|_\infty = 0.$$

7. Algorithms

In this section we describe some algorithms for computing the maximum likelihood estimator $\hat{H}_n^b$ (11). Since a distribution function in $\mathcal{F}_n^+$ is discrete, it may be described by a probability vector. Let $\mathcal{P}_n$ be the class of probability vectors in $\mathbb{R}^n$:

$$\mathcal{P}_n = \left\{ (p_1, \ldots, p_n) \in \mathbb{R}^n : \sum_{i=1}^n p_i = 1 \text{ and } p_i \geq 0 \text{ for all } i \in \{1, \ldots, n\} \right\}.$$

A distribution function $H^b \in \mathcal{F}_n^+$ may be associated with the probability vector $p \in \mathcal{P}_n$ defined as $p_j = H^b(s_j) - H^b(s_{j-1})$ (recall: $H^b(s_0) = 0$). We can switch

between probability vectors and distribution functions via:

$$H^b(s_j) = \sum_{i=1}^{j} p_i, \quad \text{and} \quad p_j = H^b(s_j) - H^b(s_{j-1}). \tag{19}$$

7.1. Expectation Maximization (EM)

The EM algorithm was first thoroughly studied in [9]. While it is typically used in parametric settings, it may also be used for non-parametric estimation. It is especially appealing due to its ease of implementation and its interpretation of incomplete data models. For the application of EM to our problem, we follow the description of EM in [23]. The authors describe the EM algorithm for problems similar to the one we are facing. The class of problems they consider is the following.

Suppose we aim to estimate a distribution function F. We cannot directly observe a sample X from F. Instead, we observe $Y = T(X, C)$, with $X \sim F$ and C some random variable independent of X. Clearly, given Lemma 3.1 the problem of estimating H^b belongs to this class of problems with $T(x, c) = xc$ and $C \sim g_K^S$. Suppose we have an initial estimate $H_0^b \in \mathcal{F}_n^+$ of the CDF H^b. Let $p^{(0)}$ be the associated probability vector as in (19). Let $X_1, \ldots, X_n \sim H^b$. In [23] it is shown that in their general context the EM algorithm yields the following update rule:

$$p_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{P}_{p^{(k)}} \left(X_i = s_j | s_1, \ldots, s_n \right). \tag{20}$$

We use the notation $\mathbb{P}_p$ to indicate the probability measure associated with the probability vector p. For our 'random fraction' setting, we use Bayes' rule to obtain:

$$\mathbb{P}_{p^{(k)}} \left(X_i = s_j | s_1, \ldots, s_n \right) = \frac{g_K^S \left(\frac{s_i}{s_j} \right) \frac{1}{s_j} p_j^{(k)}}{\sum_{q=1}^{n} g_K^S \left(\frac{s_i}{s_q} \right) \frac{1}{s_q} p_q^{(k)}}.$$

Plugging this into (20) yields:

$$p_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{\alpha_{i,j}}{\sum_{q=1}^{n} \alpha_{i,q} p_q^{(k)}} p_j^{(k)}, \quad \text{with: } \alpha_{i,j} = g_K^S \left(\frac{s_i}{s_j} \right) \frac{1}{s_j}. \tag{21}$$

When terminating the EM algorithm after an appropriate number of iterations we obtain $\hat{H}_n^b$ from $p^{(k)}$ via (19). The EM algorithm may for example be terminated when successive iterations do not meaningfully change the log-likelihood anymore. We do not provide a specific stopping criterion for EM, since we do not use it directly. We only use it in hybrid form with the Iterative Convex Minorant algorithm (ICM) which is described in the next section. For the ICM algorithm and the hybrid ICM-EM algorithm we do provide explicit termination conditions.

7.2. Iterative Convex Minorant (ICM)

The ICM algorithm was first introduced in [11]. The version of ICM we discuss is described in [12] and is sometimes called the modified ICM algorithm. This modification of ICM ensures convergence under fairly general conditions. The algorithm is designed to minimize a convex function ϕ over the closed convex cone:

$$\mathcal{C}_+ := \{\beta \in \mathbb{R}^n : 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_n\}.$$

Recall equation (13), computing the estimator $\hat{H}_n^b$ is equivalent to solving the following optimization problem:

$$\hat{\beta} \in \arg\max_{\beta \in \mathcal{C}} l(\beta) = \arg\max_{\beta \in \mathcal{C}} \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \alpha_{i,j} (\beta_j - \beta_{j-1}) \right). \tag{22}$$

With $\beta_j = H^b(s_j)$ and $\hat{\beta} = (\hat{H}_n^b(s_1), \ldots, \hat{H}_n^b(s_n))$. From (22) it is clear that for any $\beta \in \mathcal{C}_+$ with $\beta_n < 1$, the likelihood can be increased by setting $\beta_n = 1$, since $\alpha_{i,j} \geq 0$. Hence, we may incorporate the constraint $\beta_n = 1$ instead of $\beta_n \leq 1$. We achieve this via a Lagrange multiplier. Define the convex function $\phi : \mathcal{C}_+ \to \mathbb{R} \cup \{\infty\}$ as:

$$\phi(\beta) = -l(\beta) + \beta_n = -\frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{n} \alpha_{i,j} (\beta_j - \beta_{j-1}) \right) + \beta_n.$$

Hereby we have incorporated the constraint, with a Lagrange multiplier equal to one. Also, the problem is now written as a convex minimization problem since $\hat{\beta} \in \arg\min_{\beta \in \mathcal{C}_+} \phi(\beta)$. Therefore, the ICM algorithm may be used to compute the MLE. Suppose we have some initial estimate $\beta^{(0)}$. The idea of ICM is to locally approximate ϕ with the following quadratic form in iteration k:

$$\begin{aligned} \phi_{(k)}(\beta) = \left(\beta - \beta^{(k)} + W\left(\beta^{(k)}\right)^{-1} \nabla\phi\left(\beta^{(k)}\right) \right)^{\mathsf{T}} W\left(\beta^{(k)}\right) \cdot \\ \cdot \left(\beta - \beta^{(k)} + W\left(\beta^{(k)}\right)^{-1} \nabla\phi\left(\beta^{(k)}\right) \right). \end{aligned} \tag{23}$$

The notation $\nabla\phi$ is used for the gradient of ϕ, the vector of partial derivatives of ϕ. The matrix W is a diagonal matrix, its diagonal is often chosen equal to the diagonal of the Hessian matrix of ϕ:

$$W\left(\beta^{(k)}\right) = \operatorname{diag}\left(\frac{\partial^2}{\partial \beta_j^2} \phi\left(\beta^{(k)}\right) \right).$$

In the ICM algorithm, $\phi_{(k)}$ is minimized over $\mathcal{C}_+$ instead of ϕ to obtain a candidate β for $\beta^{(k+1)}$. If this candidate β sufficiently decreases ϕ it is accepted, and we set $\beta^{(k+1)} = \beta$. Otherwise, a line-search is performed to obtain $\beta^{(k+1)}$,

which is then given by a convex combination of β and $\beta^{(k)}$. A precise description of the algorithm is given in Appendix B. We remark that minimizing (23) is equivalent to computing the weighted least-squares estimator of a monotone regression function. This can be done efficiently, for more details see [12]. The partial derivatives of ϕ are related to those of l as in (14) and (15), via:

$$\frac{\partial}{\partial\beta_j}\phi(\beta) = -\frac{\partial}{\partial\beta_j}l(\beta) + \mathbb{1}\{j = n\} \quad \text{and} \quad \frac{\partial^2}{\partial\beta_j^2}\phi(\beta) = -\frac{\partial^2}{\partial\beta_j^2}l(\beta).$$

For ICM we use the following stopping criterion, stop whenever:

$$\max_{j\in\{1,\ldots,n\}} \left|\beta_j^{(k)} - \beta_j^{(k-1)}\right| < \varepsilon, \tag{24}$$

for 10 successive iterations. In simulations we set $\varepsilon = 10^{-4}$. The interpretation of this criterion is that we stop whenever the largest change in probability mass is below ε for 10 successive iterations. We note that this criterion could be inappropriate if ICM approaches the optimum very slowly. In simulations (section 9) this was not an issue.

7.3. Hybrid ICM-EM

In [23] it was proposed to combine ICM and EM into a hybrid algorithm. The idea is that a single iteration of this hybrid algorithm consists of first performing one iteration of the ICM algorithm followed by one iteration of the EM algorithm. The ICM algorithm appears somewhat slow initially, if it is started far from the MLE, whereas close to the optimal value it converges quickly. On the contrary, the EM algorithm seems quicker at the start but has trouble converging when close to the optimum. Moreover, when performing an EM step after an ICM step, ICM ensures that many of the p_j's are zero. From (21) we see that EM will never set such a p_j to a positive value, hence EM only needs to operate in a lower dimensional space. In practice, it seems that the hybrid ICM-EM algorithm inherits the strengths of both algorithms and is quicker than both ICM and EM. This was for example observed in simulations in [13] and [23]. As with ICM, the same termination condition (24) is used.

8. Regularization of the maximum likelihood estimator

In this section, we describe how the MLE $\hat{H}_n^b$ may be used to estimate H, the distribution function of interest. At first glance, it seems reasonable to plug in $\hat{H}_n^b$ for H^b in equation (3). Unfortunately, simulations indicate that this yields a poor estimate of H. In section 9 we describe in detail how simulations are performed. For now, Figure 3 shows the result of a single simulation run. This simulation corresponds to the case where each particle is a dodecahedron, $n = 1000$, and H corresponds to a standard exponential distribution. For this H, H^b corresponds to a gamma distribution. The left panel of Figure 3 shows

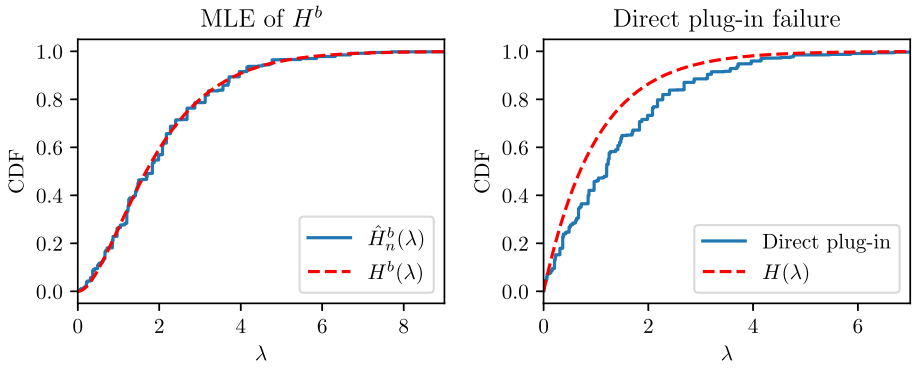

FIG 3. *Left: MLE of H^b. Right: Direct plug-in estimate of H.*

that $\hat{H}_n^b$ closely resembles H^b. Meanwhile, in the right panel of Figure 3 we observe that plugging in $\hat{H}_n^b$ for H^b in equation (3) yields a poor estimate of H. This is due to the influence of the behavior of $\hat{H}_n^b$ near zero.

We propose a regularization technique to resolve this issue. Let $t_n > 0$, truncating $\hat{H}_n^b$ at t_n yields:

$$\hat{H}_n^b(\lambda; t_n) := \begin{cases} \frac{\hat{H}_n^b(\lambda) - \hat{H}_n^b(t_n)}{1 - \hat{H}_n^b(t_n)} & \text{if } \lambda \geq t_n \\ 0 & \text{otherwise} \end{cases}.$$

Plugging this truncated version of $\hat{H}_n^b$ into (3) we obtain:

$$\hat{H}_n(\lambda; t_n) := \frac{\int_0^\lambda \frac{1}{x} \mathrm{d}\hat{H}_n^b(x; t_n)}{\int_0^\infty \frac{1}{x} \mathrm{d}\hat{H}_n^b(x; t_n)} = \begin{cases} \frac{\int_{t_n}^\lambda \frac{1}{x} \mathrm{d}\hat{H}_n^b(x)}{\int_{t_n}^\infty \frac{1}{x} \mathrm{d}\hat{H}_n^b(x)} & \text{if } \lambda \geq t_n \\ 0 & \text{if } 0 \leq \lambda < t_n \end{cases}. \tag{25}$$

Therefore, we introduce a new parameter t_n, which we refer to as the truncation parameter. In the following lemma, we show that for an appropriate choice of the truncation parameter t_n, a sequence of approximating CDFs converging to H^b may be de-biased to obtain a close approximation of H.

Lemma 8.1. *Let H be a continuous CDF on $(0, \infty)$, with finite first moment and length-biased version H^b. Let $(t_n)_{n \geq 1}$, $t_n > 0$ be a sequence such that $\lim_{n \to \infty} t_n = 0$. Let $(H_n^b)_{n \geq 1}$ be a sequence of CDFs. Assume H_n^b converges uniformly to H^b with rate at least t_n, that is: $||H_n^b - H^b||_\infty = o(t_n)$. Define:*

$$H_n(\lambda) = \begin{cases} \frac{\int_{t_n}^\lambda \frac{1}{x} \mathrm{d}H_n^b(x)}{\int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x)} & \textit{if } \lambda \geq t_n \\ 0 & \textit{if } 0 \leq \lambda < t_n \end{cases},$$

then: $\lim_{n \to \infty} ||H_n - H||_\infty = 0$.

The proof is given in Appendix A. Lemma 8.1 shows that truncation is a viable approach for consistent estimation of H. Note that in Lemma 8.1, we

may take $t_n = \sqrt{\|H_n^b - H^b\|_\infty}$. In practice the result cannot directly be applied to $\hat{H}_n^b$ since the quantity $\|\hat{H}_n^b - H^b\|_\infty$ is unknown. We propose a rule of thumb for t_n. Let $s \in \mathbb{R}$ and define:

$$\hat{F}_n^S(s;t) := \int_0^\infty G_K^S\left(\frac{s}{\lambda}\right) \mathrm{d}\hat{H}_n^b(\lambda;t) \tag{26}$$
$$\bar{F}_n^S(s) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}\{s_i \leq s\}.$$

Note that $\hat{F}_n^S(\cdot;t)$ is the distribution function of observed square root section areas induced by the biased size distribution $\hat{H}_n^b(\cdot;t)$. That is, if $\hat{H}_n^b(\cdot;t)$ is the true biased size distribution, then $\hat{F}_n^S(\cdot;t)$ is the corresponding distribution of observed square root section areas. We propose the following choice for t_n:

$$\hat{t}_n := \underset{t\in\{s_1,\ldots,s_n\}}{\arg\min} \int_0^\infty |\hat{F}_n^S(s;t) - \bar{F}_n^S(s)|\mathrm{d}s. \tag{27}$$

Hence, $\hat{t}_n$ minimizes the L^1-distance between the CDF of the observed square root section areas, induced by the estimated (biased) size distribution, and the empirical CDF of observed square root section areas. We minimize over $\{s_1, \ldots, s_n\}$ for computational convenience. In practice, the integral in (27) can be computed via numerical integration.

9. Simulations

In the previous sections, we have introduced the MLE $\hat{H}_n^b$, and shown that under reasonable assumptions it is a consistent estimator of H^b. Also, a regularization technique was introduced to consistently estimate the size distribution function H using the MLE. In this section, some simulation results are presented to assess the performance of these estimators for H^b and H. The code used for the simulations may be found at `https://github.com/thomasvdj/pysizeunfolder`. Using this code the simulation and estimation procedure can be carried out in principle for any choice of convex polyhedron for the reference particle K.

Let us start by describing how to generate an iid sample of observed section areas, for a given H and a chosen reference particle K. Lemma 3.1 shows that it is sufficient to draw $Z \sim G_K$ and independently draw $\Lambda_b \sim H^b$, followed by setting $A := Z\Lambda_b^2$. A may be considered a random section area, and repeating these steps n times yields an iid sample $A_1, \ldots, A_n$ distributed according to f_A. Taking the square root yields a sample of observed square root section areas. A sampling scheme for generating IUR planes through K is described in [8], see [21] for sampling from G_K. Finally, we consider some well-known parametric distributions for H, for these choices H^b corresponds to some other well-known parametric distribution. Hence, drawing from H^b is straightforward. The following choices for H are considered, with the corresponding H^b:

1. *Exponential distribution*: For H we consider a standard exponential distribution, such that H^b corresponds to a gamma distribution.

$$H(\lambda) = 1 - e^{-\lambda}, \quad \text{and} \quad H^b(\lambda) = 1 - (\lambda + 1)e^{-\lambda}, \ \lambda \geq 0.$$

2. *Lognormal distribution*: For H we consider a lognormal distribution with parameters μ and σ. For this H, H^b corresponds to a lognormal distribution with parameters $\mu + \sigma^2$ and σ. We set $\mu = 2$, $\sigma = 1/2$.

$$H(\lambda) = \Phi\left(\frac{\log(\lambda) - \mu}{\sigma}\right), \quad \text{and} \quad H^b(\lambda) = \Phi\left(\frac{\log(\lambda) - \mu - \sigma^2}{\sigma}\right), \ \lambda > 0.$$

Here, Φ denotes the CDF of a standard normal distribution.

For the simulations, we consider the following shapes for the particles: the dodecahedron, cube, and tetrahedron. As for the specific choice of the reference particle K, each of the shapes is scaled such that they have volume 1. Because these shapes are polyhedra, each of the corresponding distribution functions G_K has a Lebesgue density by Theorem 2.3.

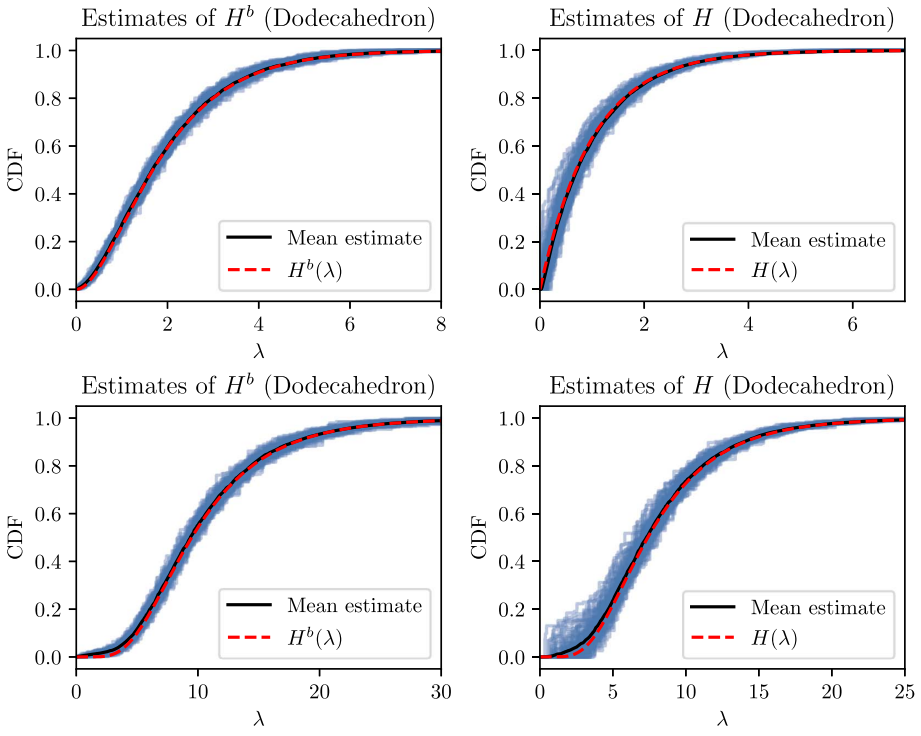

FIG 4. *Simulation results for the dodecahedron, $n = 1000$. Top left: H^b is a gamma distribution. Top right: H is an exponential distribution. Bottom left: H is a lognormal distribution. Bottom right: H is a lognormal distribution.*

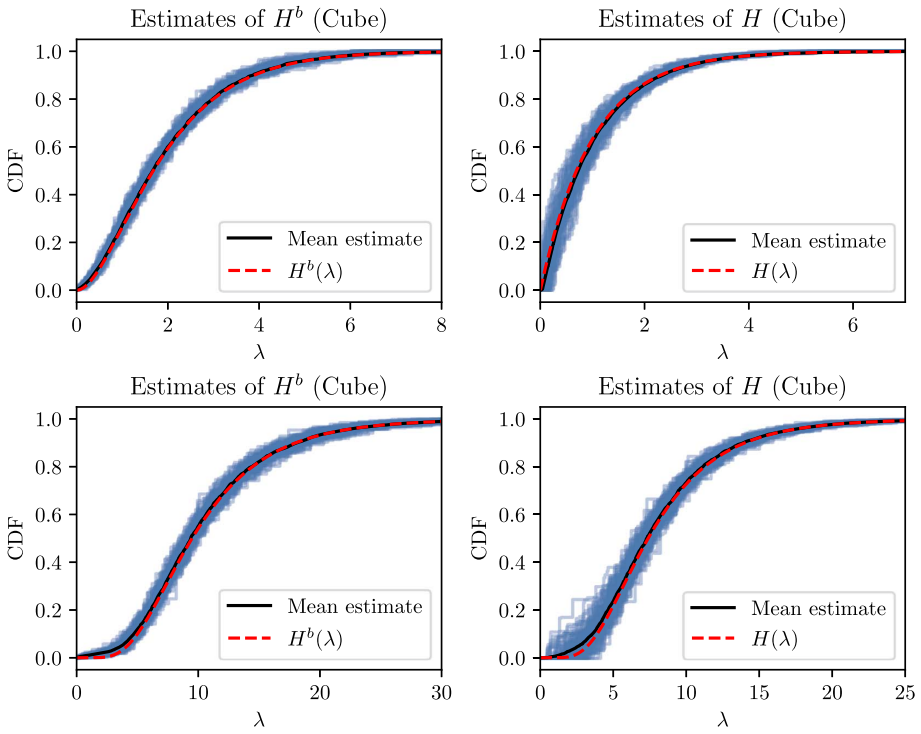

FIG 5. *Simulation results for the cube, $n = 1000$. Top left: H^b is a gamma distribution. Top right: H is an exponential distribution. Bottom left: H is a lognormal distribution. Bottom right: H is a lognormal distribution.*

Now that we covered the simulation of iid samples we discuss the computation of estimators. For a given choice of n, H, and shape for the particles we generate a sample of n observed (square root) section areas. The MLE $\hat{H}_n^b$ is computed using the hybrid ICM-EM algorithm. The computation of the MLE requires that we can evaluate g_K^S in given points. As mentioned before, there is typically no explicit expression for g_K^S and we use the Monte Carlo simulation scheme described in [21] for approximating g_K^S (recall Figure 2). For estimating H we compute $\hat{H}_n(\cdot, \hat{t}_n)$ as in (25), with $\hat{t}_n$ as in (27). Throughout this section, we refer to this estimator simply as $\hat{H}_n$. Note that for the computation of $\hat{t}_n$ we require G_K^S, which is also not explicitly known. Hence, similarly to g_K^S we use a Monte-Carlo approximation of G_K^S. In this case, we use an empirical distribution function based on the same sample used for approximating g_K^S.

We perform repeated simulations as follows. For various choices of n we generate a sample of n observed section areas. This is repeated 100 times for each choice of n, H, and shape for the particles. Simulation results for the dodecahedron and cube are shown in Figures 4 and 5 respectively. These results correspond to $n = 1000$. Each of the blue lines corresponds to one of the 100 estimates, each estimate based on a different sample of size $n = 1000$. The black

line is the point-wise average of all estimates. Further simulation results for the other shapes are summarized in Tables 1, 2 and 3. We quantify the error of the estimate as the supremum distance between the true H^b and $\hat{H}^b_n$, and similarly for the error of the estimates of H. The mean error is then the mean taken over the 100 resulting errors of the estimates. For these 100 resulting errors, the 2.5% and 97.5% quantiles are also shown.

Let us discuss the content of Tables 1, 2, and 3. As expected, as n increases the average error decreases, for all chosen shapes and size distributions, both for

TABLE 1
Simulation results for the dodecahedron.

		$\|\hat{H}^b_n - H^b\|_\infty$		$\|\hat{H}_n - H\|_\infty$	
n	H	mean error	(2.5%, 97.5%)	mean error	(2.5%, 97.5%)
1000	Exponential	0.0577	(0.038, 0.076)	0.118	(0.064, 0.20)
1000	Lognormal	0.0657	(0.045, 0.10)	0.0924	(0.057, 0.18)
2000	Exponential	0.0452	(0.032, 0.063)	0.0972	(0.054, 0.17)
2000	Lognormal	0.0528	(0.034, 0.081)	0.0783	(0.044, 0.13)
5000	Exponential	0.0318	(0.024, 0.045)	0.0687	(0.036, 0.14)
5000	Lognormal	0.0380	(0.027, 0.054)	0.0586	(0.037, 0.097)
10000	Exponential	0.0260	(0.019, 0.035)	0.0578	(0.029, 0.12)
10000	Lognormal	0.0298	(0.023, 0.040)	0.0478	(0.028, 0.086)

TABLE 2
Simulation results for the cube.

		$\|\hat{H}^b_n - H^b\|_\infty$		$\|\hat{H}_n - H\|_\infty$	
n	H	mean error	(2.5%, 97.5%)	mean error	(2.5%, 97.5%)
1000	Exponential	0.0647	(0.045, 0.094)	0.134	(0.073, 0.23)
1000	Lognormal	0.0794	(0.055, 0.12)	0.107	(0.062, 0.17)
2000	Exponential	0.0509	(0.036, 0.070)	0.107	(0.059, 0.19)
2000	Lognormal	0.0630	(0.043, 0.087)	0.0911	(0.050, 0.17)
5000	Exponential	0.0394	(0.029, 0.053)	0.0787	(0.043, 0.13)
5000	Lognormal	0.0460	(0.033, 0.059)	0.0672	(0.040, 0.11)
10000	Exponential	0.0308	(0.022, 0.042)	0.0620	(0.036, 0.096)
10000	Lognormal	0.0368	(0.028, 0.047)	0.0544	(0.031, 0.091)

TABLE 3
Simulation results for the tetrahedron.

		$\|\hat{H}^b_n - H^b\|_\infty$		$\|\hat{H}_n - H\|_\infty$	
n	H	mean error	(2.5%, 97.5%)	mean error	(2.5%, 97.5%)
1000	Exponential	0.0948	(0.062, 0.15)	0.197	(0.090, 0.39)
1000	Lognormal	0.110	(0.078, 0.15)	0.163	(0.091, 0.30)
2000	Exponential	0.0792	(0.058, 0.10)	0.153	(0.085, 0.28)
2000	Lognormal	0.0930	(0.069, 0.13)	0.134	(0.082, 0.24)
5000	Exponential	0.0602	(0.046, 0.080)	0.120	(0.061, 0.25)
5000	Lognormal	0.0761	(0.058, 0.093)	0.0997	(0.068, 0.15)
10000	Exponential	0.0514	(0.038, 0.064)	0.101	(0.054, 0.20)
10000	Lognormal	0.0643	(0.048, 0.083)	0.0805	(0.059, 0.11)

TABLE 4
Algorithms mean run-times and mean number of iterations.

	ICM		ICM-EM	
n	time (s)	# iterations	time (s)	# iterations
1000	3.99	314	0.511	26.4
2000	27.6	502	2.42	30.7
5000	415	784	25.8	62.5

the estimates of H and H^b. Comparing the average supremum error for a fixed n, and a fixed size distribution, it is clear that the errors are smallest for the dodecahedron, followed by the cube and finally, the average error is largest for the tetrahedron. This is the case for both the average errors for estimating H as well as H^b. Note that estimating H instead of H^b increases the supremum error, and the corresponding mean supremum errors are also larger. These larger errors are also evident in Figures 4 and 5. We note that for some practical applications, an estimate of H^b may be sufficient.

Finally, we briefly touch upon the computational efficiency of the algorithms for computing $\hat{H}_n^b$. We take for the shape of the particles the dodecahedron and for H the previously introduced lognormal distribution. In Table 4 the average run-times and iteration counts of the ICM and ICM-EM algorithms are shown, averaged over 10 simulation runs. The EM algorithm is not included in the table, in simulations it was several orders of magnitude slower than the other algorithms. Clearly, ICM-EM is considerably faster than ICM.

10. Concluding remarks

In this paper we have studied a generalization of the classical Wicksell corpuscle problem, considering an arbitrary convex shape for the particles instead of spheres. In particular, for the problem of estimating the CDF H of the particle size distribution an identifiability result is derived. We also obtain an inversion formula via the Mellin transform. A nonparametric maximum likelihood estimator is proposed for the biased size distribution H^b and it is proven to be uniformly strongly consistent. Moreover, this estimator can be computed efficiently in practice. In a simulation study the proposed estimators for H^b and H perform well for various choices of particle shapes and particle size distributions.

Appendix A: Proofs

Proof of Lemma 4.2. The result follows almost immediately from theorem 7.8.2. in [14], which is a Mellin inversion theorem. Suppose $X \sim F_1$ and $Y \sim F_2$. By assumption $\mathcal{M}_X$ and $\mathcal{M}_Y$ are analytic on $\mathrm{St}(\alpha, \beta)$, $0 \leq \alpha < \beta$. Let $c \in (\alpha, \beta)$, and assume $\mathcal{M}_X(c+it) = \mathcal{M}_Y(c+it)$ for all $t \in \mathbb{R}$. Let $x > 0$, by theorem 7.8.2.

from [14] we obtain:

$$\bar{F}_1(x) := \frac{1}{2}(F_1(x+) + F_1(x-)) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} -\mathcal{M}_X(s)\frac{x^{-s+1}}{s}\mathrm{d}s$$
$$\bar{F}_2(x) := \frac{1}{2}(F_2(x+) + F_2(x-)) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} -\mathcal{M}_Y(s)\frac{x^{-s+1}}{s}\mathrm{d}s.$$

Here: $F(x+) := \lim_{h\downarrow 0} F(x+h)$ and $F(x-) := \lim_{h\uparrow 0} F(x+h)$. Note that for a continuity point x of F_1, $\bar{F}_1(x) = F_1(x)$. Because CDFs are right continuous we obtain: $F_1(x) = \bar{F}_1(x+)$ and $F_2(x) = \bar{F}_2(x+)$. Hence, it is sufficient to show $\bar{F}_1 = \bar{F}_2$. Because $\mathcal{M}_X(c+it) = \mathcal{M}_Y(c+it)$ for all $t \in \mathbb{R}$:

$$\bar{F}_1(x) - \bar{F}_2(x) = \lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} -(\mathcal{M}_X(s) - \mathcal{M}_Y(s))\frac{x^{-s+1}}{s}\mathrm{d}s = 0,$$

which finishes the proof. □

Proof of Lemma 6.1. f_ϵ is upper semicontinuous, as it is given by a product, and a composition of an upper semicontinuous function and a continuous function. By Theorem 2.3, g_K^S is non-decreasing on $(0, \tau_K)$ for some $0 < \tau_K \leq \sqrt{a_{\max}}$. Choose $M > \sqrt{a_{\max}}$ large enough such that $e^{-M} < \tau_K$. It now immediately follows that $f_\epsilon(z) = 0$ for $z \in [M, \infty)$ and f_ϵ is monotonically increasing on $(-\infty, -M]$. It remains to show that f_ϵ is of bounded variation on $[-M, M]$. Let $-M < z_0 < z_1 < \cdots < z_m < M$ be an arbitrary partition of $[-M, M]$. Then it follows:

$$\begin{aligned}
&\sum_{i=1}^{m} |f_\epsilon(z_i) - f_\epsilon(z_{i-1})| = \\
&\quad = \sum_{i=1}^{m} |g_K^S(e^{z_i})e^{z_i} - g_K^S(e^{z_i})e^{z_{i-1}} + g_K^S(e^{z_i})e^{z_{i-1}} - g_K^S(e^{z_{i-1}})e^{z_{i-1}}| \\
&\quad \leq ||g_K^S||_\infty \sum_{i=1}^{m} |e^{z_i} - e^{z_{i-1}}| + e^M \sum_{i=1}^{m} |g_S(e^{z_i}) - g_S(e^{z_{i-1}})| \qquad (28) \\
&\quad \leq ||g_K^S||_\infty e^M + e^M V_0^{\sqrt{a_{\max}}}\left(g_K^S\right) < \infty.
\end{aligned}$$

Note that the first sum in (28) telescopes. In the final step we use the fact that g_K^S is bounded and is of bounded variation on its support. Because the above computation holds for arbitrary partitions of $[-M, M]$ we find: $V_{-M}^{M}(f_\epsilon) < \infty$, which finishes the proof. □

Proof of Lemma 6.3. Because $f_\epsilon \geq 0$ is a Lebesgue density, for every $m \in \mathbb{N}$ there exists a bounded continuous probability density function f_ϵ^m such that

$||f_\epsilon^m - f_\epsilon||_{L^1} \leq 1/m$ (see Lemma A.1 in Appendix A). Let $m \in \mathbb{N}$, then:

$$\begin{aligned} ||f_n - f_Y||_{L_1} &= \int \left| \int f_\epsilon(z-x) - f_\epsilon^m(z-x) + f_\epsilon^m(z-x) \mathrm{d}(F_n - F_X)(x) \right| \mathrm{d}z \\ &\leq \int \left| \int f_\epsilon(z-x) - f_\epsilon^m(z-x) \mathrm{d}(F_n - F_X)(x) \right| \mathrm{d}z \\ &\quad + \int \left| \int f_\epsilon^m(z-x) \mathrm{d}(F_n - F_X)(x) \right| \mathrm{d}z. \end{aligned} \tag{29}$$

Via the triangle inequality and Fubini, the first term in (29) is bounded by:

$$\begin{aligned} &\int \left| \int f_\epsilon(z-x) - f_\epsilon^m(z-x) \mathrm{d}F_n(x) \right| \mathrm{d}z \\ &\quad + \int \left| \int f_\epsilon(z-x) - f_\epsilon^m(z-x) \mathrm{d}F_X(x) \right| \mathrm{d}z \\ &\leq \int \int |f_\epsilon(z-x) - f_\epsilon^m(z-x)| \, \mathrm{d}z \mathrm{d}F_n(x) \\ &\quad + \int \int |f_\epsilon(z-x) - f_\epsilon^m(z-x)| \, \mathrm{d}z \mathrm{d}F_X(x) \leq 2||f_\epsilon^m - f_\epsilon||_{L^1} \leq \frac{2}{m}. \end{aligned}$$

The second term in (29) may be written as:

$$\int \left| \int f_\epsilon^m(z-x) \mathrm{d}(F_n - F_X)(x) \right| \mathrm{d}z = ||\varphi_{n,m} - \varphi_m||_{L^1},$$

with $\varphi_{n,m}$ and φ_m defined as:

$$\varphi_{n,m}(z) = \int f_\epsilon^m(z-x) \mathrm{d}F_n(x), \quad \text{and} \quad \varphi_m(z) = \int f_\epsilon^m(z-x) \mathrm{d}F_X(x).$$

Because f_ϵ^m is a probability density, so are φ_m and $\varphi_{n,m}$ for all $n \in \mathbb{N}$. By the continuity of f_ϵ^m and the weak convergence of F_n to F_X we obtain that $\varphi_{n,m}$ converges pointwise to φ_m as $n \to \infty$. By Scheffé's Theorem pointwise convergence of probability densities to another probability density implies that these densities also converge in L^1. Combining all results yields:

$$\lim_{n\to\infty} ||f_n - f_Y||_{L_1} \leq \lim_{n\to\infty} \frac{2}{m} + ||\varphi_{n,m} - \varphi_m||_{L^1} = \frac{2}{m}.$$

Letting $m \to \infty$ we obtain the desired result. □

Lemma A.1. *Let f be a Lebesgue density on $\mathbb{R}$, for every $\varepsilon > 0$ there exists a bounded continuous probability density function g such that $\|g - f\|_{L^1} < \varepsilon$.*

Proof. Recall that the space of compactly supported continuous functions is dense in L^1. For $n \in \mathbb{N}$ choose a continuous, compactly supported and non-negative function g_n such that $\|g_n - f\|_{L^1} \leq 1/(n+1)$. By the reverse triangle inequality:

$$|\|g_n\|_{L^1} - 1| = |\|g_n\|_{L^1} - \|f\|_{L^1}| \leq \|g_n - f\|_{L^1} \leq \frac{1}{n+1}. \tag{30}$$

Define: $\tilde{g}_n = g_n / \|g_n\|_{L^1}$. Note that by (30), $\|g_n\|_{L^1} > 0$. Hence, $\tilde{g}_n$ is a bounded and continuous probability density function. Combining all results:

$$\begin{aligned}\|\tilde{g}_n - f\|_{L^1} &= \frac{1}{\|g_n\|_{L^1}} \int |g_n(x) - f(x) + f(x) - \|g_n\|_{L^1} f(x)| \, \mathrm{d}x \\ &\leq \frac{1}{\|g_n\|_{L^1}} \left(\|g_n - f\|_{L^1} + |\|g_n\|_{L^1} - 1| \cdot \|f\|_{L^1} \right) \\ &\leq \frac{\frac{1}{n+1} + \frac{1}{n+1}}{1 - \frac{1}{n+1}} = \frac{2}{n}.\end{aligned}$$

Because this holds for all $n \in \mathbb{N}$ we obtain the desired result. □

Proof of Lemma 8.1. We first note the following:

$$\sup_{0 \leq \lambda < t_n} |H_n(\lambda) - H(\lambda)| = \sup_{0 \leq \lambda < t_n} H(\lambda) = H(t_n). \tag{31}$$

Let us now assume $\lambda \geq t_n$. By definition:

$$H(\lambda) - H_n(\lambda) = \frac{\int_0^\lambda \frac{1}{x} \mathrm{d}H^b(x) \int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x) - \int_{t_n}^\lambda \frac{1}{x} \mathrm{d}H_n^b(x) \int_0^\infty \frac{1}{x} \mathrm{d}H^b(x)}{\int_0^\infty \frac{1}{x} \mathrm{d}H^b(x) \int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x)}. \tag{32}$$

The numerator of (32) may be written as:

$$\begin{aligned}&\int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x) \left(\int_0^\lambda \frac{1}{x} \mathrm{d}H^b(x) - \int_{t_n}^\lambda \frac{1}{x} \mathrm{d}H_n^b(x) \right) \\ &\quad - \int_{t_n}^\lambda \frac{1}{x} \mathrm{d}H_n^b(x) \left(\int_0^\infty \frac{1}{x} \mathrm{d}H^b(x) - \int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x) \right) \\ &= \int_{t_n}^\infty \frac{1}{x} \mathrm{d}H_n^b(x) \left(\int_{t_n}^\lambda \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) + \int_0^{t_n} \frac{1}{x} \mathrm{d}H^b(x) \right) \\ &\quad - \int_{t_n}^\lambda \frac{1}{x} \mathrm{d}H_n^b(x) \left(\int_{t_n}^\infty \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) + \int_0^{t_n} \frac{1}{x} \mathrm{d}H^b(x) \right).\end{aligned} \tag{33}$$

Recall: $\mathbb{E}(\Lambda) = \int_0^\infty \lambda \mathrm{d}H(\lambda) = 1 / \int_0^\infty (1/x) \mathrm{d}H^b(x)$. Plugging (33) back into (32) yields:

$$\begin{aligned}H(\lambda) - H_n(\lambda) = \mathbb{E}(\Lambda) &\left(\int_{t_n}^\lambda \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) \right) + H(t_n) \\ &- \mathbb{E}(\Lambda) H_n(\lambda) \left(\int_{t_n}^\infty \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) \right) - H_n(\lambda) H(t_n).\end{aligned}$$

Therefore, we obtain the following bound:

$$\sup_{\lambda \geq t_n} |H(\lambda) - H_n(\lambda)| \leq 2\mathbb{E}(\Lambda) \sup_{\lambda \geq t_n} \left| \int_{t_n}^\lambda \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) \right| + H(t_n). \tag{34}$$

The integral in (34) may be computed via integration by parts:

$$\begin{aligned}
&\sup_{\lambda \geq t_n} \left| \int_{t_n}^{\lambda} \frac{1}{x} \mathrm{d}(H^b - H_n^b)(x) \right| = \\
&\quad = \sup_{\lambda \geq t_n} \left| \frac{H_n^b(\lambda) - H^b(\lambda)}{\lambda} - \frac{H_n^b(t_n) - H^b(t_n)}{t_n} - \int_{t_n}^{\lambda} \left(H_n^b(x) - H^b(x) \right) \mathrm{d}\frac{1}{x} \right| \\
&\quad \leq 2 \frac{\sup_{\lambda \geq t_n} \left| H_n^b(\lambda) - H^b(\lambda) \right|}{t_n} + \sup_{\lambda \geq t_n} \left| H_n^b(\lambda) - H^b(\lambda) \right| \cdot \left| \int_{t_n}^{\lambda} \mathrm{d}\frac{1}{x} \right| \\
&\quad \leq 3 \frac{\| H_n^b - H^b \|_\infty}{t_n}. \qquad (35)
\end{aligned}$$

Note that the bound in (34) is greater than $H(t_n)$, by (31) this means that the bound also holds when taking the supremum over $\lambda \geq 0$ instead. Combining (34) and (35) we finally obtain:

$$\| H_n - H \|_\infty \leq 6 \mathbb{E}(\Lambda) \frac{\| H_n^b - H^b \|_\infty}{t_n} + H(t_n). \qquad (36)$$

Letting n go to infinity, $H(t_n)$ converges to zero by the continuity of H. Using this and the fact that $(H_n^b)_{n \geq 1}$ converges uniformly to H^b with rate t_n (by assumption) the RHS of (36) converges to zero. □

Appendix B: Pseudo-code of algorithms

Algorithm 1 Iterative Convex Minorant (ICM)

Input: A convex function $\phi : \mathcal{C}_+ \to \mathbb{R} \cup \{\infty\}$. $\beta^{(0)}$ with $\phi(\beta^{(0)}) < \infty$.
Output: A minimizer of ϕ.

$k := 0$
$\beta^{(0)} := (\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}) \in \mathcal{C}$
while Stopping criterion is not met **do**
  $\beta := \arg\min_{y \in \mathcal{C}_+} \phi_{(k)}(y)$ ▷ With $\phi_{(k)}$ as in (23)
  if $\phi(\beta) < \phi(\beta^{(k)}) + \epsilon \nabla\phi(\beta^{(k)})^\intercal (\beta - \beta^{(k)})$ **then**
    $\beta^{(k+1)} := \beta$
  else
    $\lambda := 1$, $s := \frac{1}{2}$, $z := \beta$.
    while $\phi(z) < \phi(\beta^{(k)}) + (1-\epsilon) \nabla\phi(\beta^{(k)})^\intercal (z - \beta^{(k)})$ (I) **or**
        $\phi(z) > \phi(\beta^{(k)}) + \epsilon \nabla\phi(\beta^{(k)})^\intercal (z - \beta^{(k)})$ (II) **do**
      if (I) **then** $\lambda := \lambda + s$
      if (II) **then** $\lambda := \lambda - s$
      $z := \beta^{(k)} + \lambda(\beta - \beta^{(k)})$
      $s := \frac{s}{2}$
    end while
    $\beta^{(k+1)} := z$
  $k := k + 1$
end while
return $\beta^{(k)}$

Algorithm 2 Expectation Maximization (EM)

Input: Observed order statistics: $s_1 < s_2 < \cdots < s_n$.
Output: The MLE $\hat{H}_n^b$.
1: $k := 0$
2: $p^{(0)} := (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}) \in \mathcal{P}_n$
3: **while** Stopping criterion is not met **do**
4: $\quad p_j^{(k+1)} := \frac{1}{n} \sum_{i=1}^n \frac{\alpha_{i,j}}{\sum_{q=1}^n \alpha_{i,q} p_q^{(k)}} p_j^{(k)}, \quad$ with: $\alpha_{i,j} = g_K^S\left(\frac{s_i}{s_j}\right) \frac{1}{s_j}$
5: $\quad k := k + 1$
6: **end while**
7: $\hat{H}_n^b(s_j) := \sum_{i=1}^j p_i^{(k)}$ for $j \in \{1, \ldots, n\}$.
8: **return** $\hat{H}_n^b$

Acknowledgments

We thank Kees Bos, Jilt Sietsma and Karo Sedighiani for fruitful discussions. We thank two reviewers and an associate editor for their constructive remarks.

References

[1] Arratia, R., Goldstein, L. and Kochman, F. (2019). Size bias for one and all. *Probab. Surv.* **16** 1–61. https://doi.org/10.1214/13-PS221. MR3896143
[2] Baddeley, A. and Jensen, E. B. V. (2004). *Stereology for Statisticians*. Chapman and Hall/CRC. https://doi.org/10.1201/9780203496817. MR2107000
[3] Beneš, V. and Rataj, J. (2004). *Stochastic Geometry: Selected Topics*. Kluwer Academic Publishers. https://doi.org/10.1007/b129604. MR2068607
[4] Bissantz, N. and Holzmann, H. (2008). Statistical inference for inverse problems. *Inverse Probl.* **24** 034009. https://doi.org/10.1088/0266-5611/24/3/034009. MR2421946
[5] Butzer, P. L. and Jansche, S. (1997). A direct approach to the Mellin transform. *J. Fourier Anal. Appl.* **3** 325–376. https://doi.org/10.1007/BF02649101. MR1468369
[6] Cavalier, L. (2008). Nonparametric statistical inverse problems. *Inverse Probl.* **24** 034004. https://doi.org/10.1088/0266-5611/24/3/034004. MR2421941
[7] Chiu, S. N., Stoyan, D., Kendall, W. S. and Mecke, J. (2013). *Stochastic Geometry and its Applications*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118658222. MR3236788
[8] Davy, P. and Miles, R. E. (1977). Sampling theory for opaque spatial specimens. *J. R. Stat. Soc. Ser. B Methodol.* **39** 56–65. https://doi.org/10.1111/j.2517-6161.1977.tb01605.x. MR0501482
[9] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser.*

B Methodol. **39** 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x. MR0501537

[10] Groeneboom, P., Jongbloed, G. and Michael, S. (2013). Consistency of maximum likelihood estimators in a large class of deconvolution models. *Can. J. Stat.* **41** 98–110. https://doi.org/10.1002/cjs.11152. MR3030787

[11] Groeneboom, P. and Wellner, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Birkhäuser Basel. https://doi.org/10.1007/978-3-0348-8621-5. MR1180321

[12] Jongbloed, G. (1998). The iterative convex minorant algorithm for nonparametric estimation. *J. Comput. Graph. Stat.* **7** 310–321. https://doi.org/10.2307/1390706. MR1646718

[13] Jongbloed, G. (2001). Sieved maximum likelihood estimation in Wicksell's problem and related deconvolution problems. *Scand. J. Stat.* **28** 161–183. https://doi.org/10.1111/1467-9469.00230. MR1844355

[14] Kawata, T. (1972). *Fourier Analysis in Probability Theory*. Academic Press. https://doi.org/10.1016/C2013-0-07405-1. MR0464353

[15] Kisel'ák, J. and Baluchová, G. (2021). On particle-size distribution of convex similar bodies in $\mathbb{R}^3$. *J. Math. Imaging Vision* **63** 108–119. https://doi.org/10.1007/s10851-020-00997-y. MR4205280

[16] McGarrity, K. S., Sietsma, J. and Jongbloed, G. (2014). Nonparametric inference in a stereological model with oriented cylinders applied to dual phase steel. *The Annals of Applied Statistics* **8** 2538–2566. https://doi.org/10.1214/14-AOAS787. MR3292508

[17] Ohser, J. and Mücklich, F. (1995). Stereology for some classes of polyhedrons. *Adv. Appl. Prob.* **27** 384–396. https://doi.org/10.2307/1427832. MR1334820

[18] Ohser, J. and Mücklich, F. (2000). *Statistical Analysis of Microstructures in Materials Science*. John Wiley & Sons, Inc.

[19] Ohser, J. and Nippe, M. (1997). Stereology of cubic particles: various estimators for the size distribution. *J. Microsc.* **187** 22–30. https://doi.org/10.1046/j.1365-2818.1997.2020762.x

[20] Santaló, L. A. and Kac, M. (2004). *Integral Geometry and Geometric Probability*. Cambridge University Press. https://doi.org/10.1017/CBO9780511617331. MR2162874

[21] van der Jagt, T., Jongbloed, G. and Vittorietti, M. (2023). Existence and approximation of densities of chord length- and cross section area distributions. *Image Analysis and Stereology* **42** 171–184. https://doi.org/10.5566/ias.2923

[22] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* **76** 751–761. https://doi.org/10.2307/2336635. MR1041420

[23] Wellner, J. A. and Zhan, Y. (1997). A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J. Am. Stat. Assoc.* **92** 945–959. https://doi.org/10.1080/01621459.1997.10474049. MR1482125

[24] Wicksell, S. D. (1925). The corpuscle problem. A mathematical study of a biometric problem. *Biometrika* **17** 84–99. https://doi.org/10.1093/biomet/17.1-2.84

[25] Zolotarev, V. M. (1957). Mellin-Stieltjes transforms in probability theory. *Theory Probab. Appl.* **2** 433–460. https://doi.org/10.1137/1102031. MR0108843