DEFINING AND APPLYING A NEW APPROXIMATION FOR THE PARAMETRIC PROBABILITY DENSITIES OF SPHERICAL PARTICLE PROFILE SIZES

Ekaterina Poliakova[1,2]

[1]Arctic and Antarctic Research Institute, Climate and Environment Research Laboratory, St. Petersburg, Russia,
[2]Norwegian University of Science and Technology, Department of Mathematical Sciences, Trondheim, Norway
e-mail: ekaterina.poliakova@ntnu.no

ABSTRACT

A novel approximation for the parametric probability densities of spherical particle profile diameters is suggested. The spheres are approximated from solids obtained by rotating regular polygons around their axes of symmetry. This approximation facilitates using the maximum likelihood method in samples that have too few profiles for applying classical stereological approaches.

Keywords: distribution parameters, maximum likelihood, probability density function, small samples, spherical particles, statistics.

INTRODUCTION

While stereological approaches for large samples (containing thousands of profiles) are well-developed (Cruz-Orive, 2017), treating small samples (ranging from 20 to several hundred profiles) remains challenging. However, such small samples are common in the geosciences. This study was specifically motivated by examining the fabric of the Antarctic ice core, where the diameter of the borehole and the thickness of the homogeneous ice layers were comparable to the crystal size and where additional samples were unavailable (Lipenkov *et al.*, 2016). Complications caused by lack of material in the geosciences often occur along with the following simplifications. The distribution of the 3D sizes belongs to either one or a few parametric families; it is acceptable to approximate particles as spheres; and the presence of a small bias is much less important than minimising the variance of the parameter estimation (e.g. Gulbin, 2008; Durand *et al.*, 2006; Lopez-Sanchez and Llana-Fúnez, 2016).

Even for parameter estimation, non-parametric stereological methods are typically applied (e.g. Gulbin, 2008). However, minimum distance estimation (MDE) was recently suggested as a powerful technique for modelling a distribution function of particle profiles, which is useful for parameter estimation in medium and large samples of spherical particle profiles (Depriester and Kubler, 2019). Also, the method of moments (MoM) has been described for spherical particle profiles (Goldsmith, 1967), although used very seldom. Still, the maximum likelihood

(ML) method provides point estimators that are asymptotically unbiased and of the lowest variance under mild assumptions. It is also a flexible and powerful instrument for estimating intervals and testing hypotheses (Lehmann and Romano, 2006). ML has become increasingly popular (Schweder and Hjort, 2016) as computational methods have developed. However, the stereological use of ML is rare (Hobolth and Jensen, 2002; Keiding and Jensen, 1972), and, in practice, it is typically applied to grouped data and empirical densities (e.g. Gulbin, 2008).

This paper suggests an alternative approach for parametric distribution modelling that is useful for the ML method. We approximate the profile length distribution as a mixture of uniform distributions, but the coefficients depend on the angular approximation shape of the particles, and not on preliminary chosen size classes.

MATERIALS AND METHODS

We approximate the probability density $g(y)$ of spherical particle profile sizes y, given the probability density of sample sizes as follows. We consider spherical particles (Fig. 1) with centres that are uniformly distributed in a very large piece of volume and the diameters D having some probability distribution with cumulative distribution function F and probability density f.

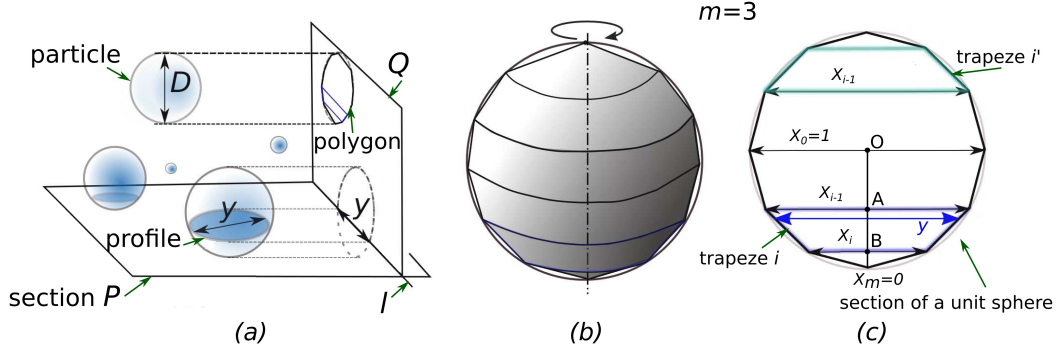We call the section plane P (Fig. 1, a). We choose an arbitrary line l on P and draw a plane Q, which

Fig. 1. *Approximating spherical particles by solids of revolution: (a) the spherical particles, the notation and the 4m-sided polygon on the plane Q; (b) the approximating solid obtained by rotating the 4m-sided polygon; (c) the denotations for the sizes, used for deriving Eq. 7 and Eq. 11.*

is orthogonal to P and contains l. We project all the particles orthogonally onto Q. Their projections are of the same diameter as the particles; a particle is intersected by P if and only if its projection is intersected by l. We choose a positive integer m and assign to any particle a regular $4m$-sided polygon inscribed into the particle's projection and oriented in such a way that its longest diagonal is perpendicular to l (and, hence, perpendicular to P).

We will first introduce a biased intermediate approximation g^* for the density of the diameters y of the profiles of the particles. This density will be the density $g^*(y_s)$ of the profiles y_s of the solids. Thereafter, we will obtain the main resulting approximation $g(y)$ via scaling transformation. To derive $g^*(y_s)$, we approximate any spherical particle by a solid of revolution obtained by rotating the $4m$-sided polygon around one of its diagonals that is perpendicular to P (and moving back the solid so that its centre coincides with the centre of the particle in Fig. 1, b).

All the profiles of the solids are circles. The y_s have the same distribution as the lengths of the profiles of the $4m$-sided regular polygons intersected by l. The diameters D of the intersected particles follow the size-weighted distribution (Baddeley and Jensen, 2004, p. 34) with density

$$\phi(t) = \frac{t f(t)}{E(D)}, \qquad (1)$$

where $E(D)$ is the expected value of the particle diameter. Each of the $4m$-sided polygons is a union of $2m$ trapezia with bases parallel to l. We denote by p_i the probability that given that a particle is intersected, the corresponding profile of the polygon belongs to either the i-th trapezium or the trapezium which is congruent to the i-th trapezium (denoted as

i'-th trapezium in Fig. 1, c). The p_i is proportional to the height of the i-th trapezium. Denoting the particle centre by O and the distances from O to the trapezium bases by OA and OB,

$$p_i = \frac{2AB}{D}, \qquad (2)$$

and from the geometry of the regular polygon we calculate

$$p_i = sin\frac{i\pi}{2m} - sin\frac{(i-1)\pi}{2m}, \quad for \quad i = 1,...,m. \quad (3)$$

When a particle diameter D is fixed to equal some number t and the number of the intersected trapeze is given, the distances h from particle centres to the intersections between diameters of the profiles and the axis of rotation are distributed uniformly between lengths of OA and OB, because of the initial assumption on the uniform distribution of the centres in the volume and the fixed $D = t$. Hence the diameters of the profiles are also distributed uniformly between tx_{i-1} and tx_i where the dimensionless constants x_{i-1} and x_i are bases of the i-th trapezium which is inscribed in the sphere with the unit diameter (Fig. 1, c). The x_i are computed from the geometry of the regular polygon and equal:

$$x_i = cos\frac{i\pi}{2m}, \quad for \quad i = 0,...,m. \qquad (4)$$

The diameters of the profiles of the solid have density

$$g^*(y_s|D=t,i) = \begin{cases} \frac{1}{t \cdot (x_{i-1}-x_i)}, y_s \in (tx_i, tx_{i-1}) \\ 0, \quad otherwise. \end{cases} \qquad (5)$$

When a particle diameter D is fixed to equal t but a trapezium may be arbitrary, the density of the profiles is

$$g^*(y_s|D=t) = \sum_{i=1}^{m} p_i g^*(y_s|D=t,i), \qquad (6)$$

2

where p_i is the probability to belong to the i-th trapeze, computed by the Eq. 3. When D is not fixed, we observe that the condition $y_s \in (Dx_i, Dx_{i-1})$ is equivalent to $D \in (\frac{y_s}{x_{i-1}}, \frac{y_s}{x_i})$, and therefore the probability density of the profiles becomes

$$g^*(y_s) = \int_{y_s}^{\infty} g^*(y_s|D=t)\frac{tf(t)}{E(D)}dt =$$

$$\frac{1}{E(D)}\sum_{i=1}^{m}\frac{p_i\left(F\left(\frac{y_s}{x_i}\right) - F\left(\frac{y_s}{x_{i-1}}\right)\right)}{x_{i-1} - x_i}. \quad (7)$$

The coefficients p_i and x_i are computed from Eqs. 3 and 4.

This intermediate approximation g^* is applicable but inexact if the number of summands m is small, because the approximating profiles of the polygon are shorter than in reality and hence its graph is deformed as compared to the graph of the true probability density. In order to reduce the bias of the mean diameter estimates, the approximation may be improved by applying a scaled transformation, so that the mean of the approximation $g(x)$ is scaled to equal the mean of the actual density. For any fixed $D = t$, $E(y_s)$ in the intermediate approximation equals

$$E(y_s) = \int_{y_s=0}^{t} y_s g^*(y_s|D=t)dy_s =$$

$$\sum_{i=1}^{m}\frac{x_{i-1} + x_i}{2}tp_i =$$

$$\sum_{i=1}^{m}\frac{Area(trapeze_i) + Area(trapeze_{i'})}{t} = \quad (8)$$

$$\frac{Area(regular\ polygon)}{t} = \frac{mt}{2}sin\frac{\pi}{2m},$$

but

$$E(y) = \frac{Area(profile)}{t} = \frac{\pi t}{4} \quad (9)$$

actually (as the sum for infinitely many narrow trapezia). So we choose the scaling coefficient

$$a = \frac{\pi}{2m\,sin\frac{\pi}{2m}}. \quad (10)$$

The final form of approximation becomes

$$g(y) = \frac{1}{aE(D)}\sum_{i=1}^{m}\frac{p_i\left(F\left(\frac{y}{ax_i}\right) - F\left(\frac{y}{ax_{i-1}}\right)\right)}{x_{i-1} - x_i}, \quad (11)$$

where p_i, x_i, and a are defined by Eqs. 3, 4, 10 respectively, and $E(D)$ is computed from the known distribution of three-dimensional particles and depends on the parameters.

[1]Sometimes spelled 'Saltikov' (Depriester and Kubler, 2019)

In the following, the ML estimates were computed via numerical maximisation of the likelihood. Following Casella and Berger (2002, p. 290), Shao (2003, p. 273), or a more elementary description (Larsen and Marx, 2012, p. 284), the *likelihood* for each individual profile measurement x_i is a probability density computed in this measurement, as a function of the unknown parameters. When profile sizes are independent, the likelihood of the sample is a product of the likelihoods for individual measurements, also treated as a function of the parameters. The ML estimates are the values of the parameters that maximise the likelihood of the sample, given the measurements. The densities involved were approximated by Eq. 11.

The samples of profile diameters were simulated as sections of loose spheres. The diameters of the spheres intersecting the section were sampled from the diameter-weighted distribution with given parameters and have been element-wise multiplied with the diameters for random profiles of unit spheres.

RESULTS

We observed that a single probability density computation (Eq. 11) works two to three times faster than the equally exact density computation, integrating Wicksell's equation via the trapezoidal rule and unequal partition length. Additionally, using this approximation, ML functioned six to 30 times faster than MDE when applied to the same samples. The effect of the number of summands m is illustrated in Fig. 2 on three distributions that are especially typical in petrology, where our approach may be applied. The number of summands is not observed to affect the standard deviation of the estimates. In most cases, the bias component added by using too few summands was indistinguishable as well. However, when using eight summands or less in the approximation, and for some shape parameters, this bias component reached 1/5 of the standard deviation. When using 15 summands, the bias component was practically negligible.

Table 1 illustrates a typical example of parameter estimation precision by ML with our approximation and also by other methods. The table shows that, although MDE is far more efficient than the usual Saltykov [1] method for such sample sizes, ML gives even more precise estimates. For smaller sample sizes, the difference between the precision of ML and of MDE was even larger, and the amount of totally wrong estimates by MDE increased more than that by ML. Since the speed of the numerical estimation decreases

Sample size		ML	MoM	MDE	Saltykov, q=20
n = 200	bias	$2.5 \cdot 10^{-3}$	$4.2 \cdot 10^{-3}$	$-14.5 \cdot 10^{-3}$	$228 \cdot 10^{-3}$
	st. dev.	$6.1 \cdot 10^{-2}$	$6.7 \cdot 10^{-2}$	$6.4 \cdot 10^{-2}$	$66 \cdot 10^{-2}$
n = 2000	bias	$0.12 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$	$-5.3 \cdot 10^{-3}$	$310 \cdot 10^{-3}$
	st. dev.	$1.9 \cdot 10^{-2}$	$2.2 \cdot 10^{-2}$	$2.0 \cdot 10^{-2}$	$71 \cdot 10^{-2}$

Table 1. *The biases and standard deviations (st. dev.) of the estimates of the median diameter of log-normally$(0, 0.5)$-distributed spheres. The true value is 1. Each line corresponds 1000 simulated samples, resulting in 1000 estimates for each entry. ML estimates were computed using our approximation, MoM by using the known moments and (Baddeley and Jensen, 2004, p. 37), MDE by (Depriester and Kubler, 2019), Saltykov method by (Gulbin, 2008) using $q = 20$ size classes.*
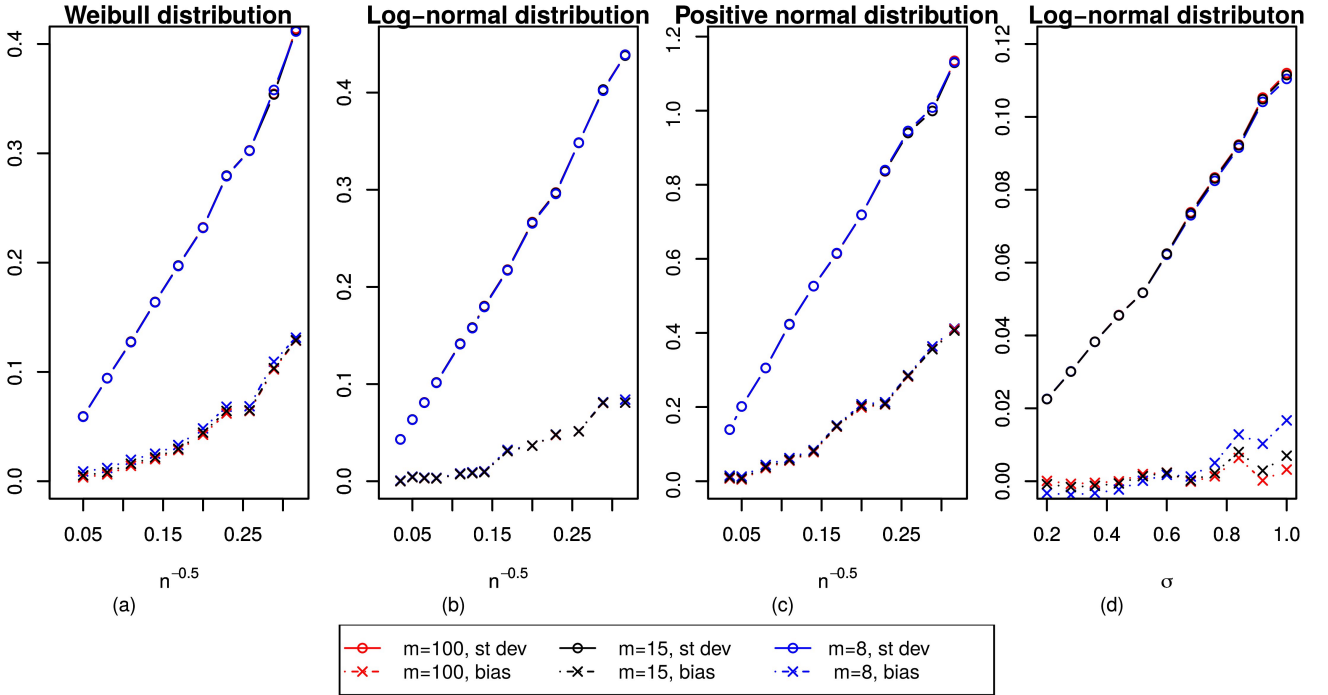

Fig. 2. *The standard deviation and the bias of the mean diameter estimates in three typical distributions as functions of $1/\sqrt{n}$ (where n is the sample size): (a) Weibull distribution: $\lambda = 1$, $k = 1.2$, $1/\sqrt{n}$ varies; (b) log-normal distribution: $\mu = 0$, $\sigma = 0.7$, $1/\sqrt{n}$ varies; (c) positive normal distribution: $\mu = 3$, $\sigma = 3$, $1/\sqrt{n}$ varies; and as functions of the shape parameter: (d) in log-normal distribution: $\mu = 0$, $n = 300$, σ varies. Each point on the plot corresponds 5000 simulated samples of sphere profiles, of fixed size n and fixed parameters, and consequent 5000 parameter estimates for each value of m.*

when an initial guess for the values becomes further from an estimate (Depriester and Kubler, 2019), the estimation using ML and our approximation was also numerically more efficient, as was the single density computation.

A possible technique for applying our approach, as well as a more detailed numerical comparison of the precision of ML (using the approximation) with that of MDE and the MoM, are described in more detail in preprint of Poliakova (2020). For samples with ≥ 70 profiles ML has always been observed to give more precise estimates. However, the MoM is a reasonable small-sample alternative to ML, having a bit smaller standard deviation of scale parameter estimates for samples with fewer than $20 - 70$ profiles, though sometimes still larger bias. The minimal number of profiles for which ML is preferable depends on the distribution family and the value of the shape parameter.

DISCUSSION

Our approach has a different application area than those of Depriester and Kubler (2019). We

approximate the probability density which facilitates ML, while Depriester and Kubler (2019) approximated the distribution function which facilitates MDE. In principle, our approach can be used for MDE via numerical integration, but the computational advantages would be lost.

The parametric methods of ML and MDE are both especially applicable when the available materials are limited. While MDE is especially usable for medium-sized samples and is known for its robustness with respect to mixtures of probability distributions, the ML is usable for small and very small samples, which is why we needed to develop the approximation. ML was observed to be most exact for the samples from 70 profiles, and in many cases even smaller samples, when the underlying assumptions hold. Moreover, ML is also flexible with respect to truncation and censoring, especially in contrast to the MoM which is an alternative technique for parameter estimation using very small samples. ML also allows for studying mixtures by incorporating the parameters and weights of the mixture components into the overall parameter vector. It also allows for estimating the confidence interval using the likelihood ratio method by Wilks' theorem (e.g. Schweder and Hjort, 2016, chap. 2) or choosing the model by Akaike information criterion (Akaike, 1974). In general, ML is 'the most popular technique for deriving estimators' (Casella and Berger, 2002, p. 315).

Our approximation is based on a natural approximation of spheres via similar shapes. Hence, varying between a few and many summands in our approximation may give an understanding of how imprecise models and estimates may be when the particles are angular or of irregular shape rather than spherical. Nevertheless, if the particles are not absolutely ideal spheres, it is reasonable to assume that increasing the number of summands in our approximation will not increase its precision, and, at most, $10 - 15$ summands will provide both a satisfactory computation speed and a reasonably precise estimation for approximately spherical particles.

ACKNOWLEDGEMENTS

REFERENCES

Akaike H (1974). A new look at the statistical model identification. IEEE T Automat Contr 19:716–23.

Baddeley A, Jensen EBV (2004). Stereology for statisticians. CRC Press.

Casella G, Berger RL (2002). Statistical inference. Duxbury Pacific Grove, CA.

Cruz-Orive LM (2017). Stereology: a historical survey. Image Anal Stereol 36:153–77, DOI: 10.5566/ias.1767.

Depriester D, Kubler R (2019). Resolution of the Wicksell's equation by minimum distance estimation. Image Anal Stereol 38:213–26, DOI: 10.5566/ias.2133.

Durand G, Gagliardini O, Thorsteinsson T, Svensson A, Kipfstuhl S, Dahl-Jensen D (2006). Ice microstructure and fabric: an up-to-date approach for measuring textures. J Glaciol 52:619–30, doi:10.3189/172756506781828377.

Goldsmith P (1967). The calculation of true particle size distributions from the sizes observed in a thin slice. Brit J Appl Phys 18:813.

Gulbin Y (2008). On estimation and hypothesis testing of the grain size distribution by the Saltykov method. Image Anal Stereol 27:163–74, DOI: 10.5566/ias.v27.p163–74.

Hobolth A, Jensen EBV (2002). A note on design-based versus model-based variance estimation in stereology. Adv Appl Probab 34:484–90, https://www.jstor.org/stable/1428242.

Keiding N, Jensen ST (1972). Maximum likelihood estimation of the size distribution of liver cell nuclei from the observed distribution in a plane section. Biometrics 28:813–29.

Larsen RJ, Marx ML (2012). Introduction to Mathematical Statistics and Its Applications: Pearson New International Edition. Pearson Prentice Hall.

Lehmann EL, Romano JP (2006). Testing statistical hypotheses. Springer Science & Business Media.

Lipenkov VY, Ekaykin AA, Polyakova EV, Raynaud D (2016). Characterization of subglacial Lake Vostok as seen from physical and isotope properties of accreted ice. Philos T Roy Soc A 374:1 – 15, https://doi.org/10.1098/rsta.2014.0303.

Lopez-Sanchez MA, Llana-Fúnez S (2016). An extension of the Saltykov method to quantify 3D grain size distributions in mylonites. J Struct Geol 93:149–61, https://doi.org/10.1016/j.jsg.2016.10.008.

Poliakova E (2020). Maximum likelihood estimation of parameters of spherical particle

size distributions from profile size measurements and its application for small samples. ArXiv preprint : https://arxiv.org/abs/2008.09091.

Schweder T, Hjort NL (2016). Confidence, likelihood, probability, vol. 41. Cambridge University Press.

Shao J (2003). Mathematical Statistics. Springer-Verlag New York.