ESTIMATING VOLUME DISTRIBUTIONS USING MAXIMUM ENTROPY

JONATHAN M KEITH AND STEPHEN L GAY

Julius Kruttschnitt Mineral Research Centre, Isles Rd., Indooroopilly Qld. 4068
e-mail: jonathan@maths.uq.edu.au

ABSTRACT

This paper describes a new method for estimating particle volume distributions using isotropic, uniform, random (IUR) sections. An attractive feature of the method is that it makes no assumptions about the shape of the particles. Since IUR sections do not uniquely determine the volume distribution for particles of general shape, the method takes a probabilistic approach. It uses the principle of maximum entropy to determine the most likely distribution given the available information. This information is extracted from the sections using geometric probability. The algorithm was tested on simulated ellipsoids, with excellent results.

Keywords: volume distribution, geometric probability, maximum entropy, IUR sections.

INTRODUCTION

The problem of estimating volume distributions using IUR sections is known to be indeterminate. This fact was demonstrated by Cruz-Orive (1976). Cruz-Orive's proof consisted of constructing two sets of ellipsoids with different volume distributions but identical IUR sections. Although this result precludes exact determination, methods have been developed to approximate volume distributions using IUR sections. Some of these methods solve the problem of indeterminacy by making additional assumptions about the shape of the particles. Two examples are the methods developed by Wicksell (1925a, b) and King (1982). More recently, Gay (1996) developed a method that makes no assumptions about particle shape. The method is called the allocation method, and it solves the problem of indeterminacy by finding a volume distribution that is consistent with the laws of geometric probability and also optimal with respect to certain regularisation criteria.

In this paper, we present a new method for estimating volume distributions using IUR sections. Like the allocation method, the new method is based on geometric probability and is applicable to particles of general shape. The essential new feature of the method is the use of the principle of maximum entropy to determine the most likely distribution consistent with the laws of geometric probability.

METHOD

In what follows, the symbol V represents particle volume, A represents section area, d represents the average distance between points of a section, Δ represents the average triangular area defined by three points of a section, B represents the perimeter of a section and s represents a surface parameter described by Jensen and Gundersen (1985, 1987) and later developed by Gay (1997). The five section properties are collectively represented by a vector **s**.

The first step of the method estimates a distribution p(V, **s**). This distribution is the proportion of sections which have section properties **s** and which came from particles with volume V.

For the sake of simplicity, the algorithm allows only discrete values for volume. For example, the results shown in Fig. 1 used 70 volumes ranging from 1 unit to 139 units in increments of two. In general, there are N_p volumes indexed by $i = 1, \ldots, N_p$.

The algorithm also allows only discrete values of **s**. The allowed values are the properties of the observed sections. There are N_s sections indexed by $j = 1, \ldots, N_s$.

In this discrete approximation, the distribution p(V, **s**) is replaced by p_{ij} – the proportion of sections which have section properties $\mathbf{s_j}$ and which came from particles with volume V_i. This is a proportion of the total area (not number) of sections.

The p_{ij}'s satisfy a number of conditions. Firstly, the p_{ij}'s are related to the volume distribution f_i by equation (1).

$$f_i = \sum_j p_{ij} \qquad (1)$$

The p_{ij}'s are related to the distribution of sections g_j by equation (2). The value of g_j is the area of

section j divided by the total area of the sections.

$$g_j = \sum_i p_{ij} \qquad (2)$$

The p_{ij}'s also satisfy the geometric probability equations (3) to (5). The symbol z represents any function of volume. We used ten polynomial functions of the form $z_m(V) = V^m$ for m = 0, 1, …, 9.

$$\sum_{i,j} p_{ij} \, z(V_i) \, (V_i - 2A_j d_j) = 0 \qquad (3)$$

$$\sum_{i,j} p_{ij} \, z(V_i) \, (V_i^2 - 4\pi A_j^2 \Delta_j) = 0 \qquad (4)$$

$$\sum_{i,j} p_{ij} \, z(V_i) \, \frac{B_j}{A_j} (V_i - \frac{\pi}{4} \frac{A_j}{B_j} s_j) = 0 \qquad (5)$$

These equations were derived by Gay based on Jensen's first volume equation (1991), Jensen and Gundersen's second volume equation (1985) and Jensen and Gundersen's surface area equation (1987, 1989). In outline, the proof consists of rearranging Jensen and Gundersen's equations as they apply to particles with a single volume, multiplying this expression by a suitable function z(V), and then integrating with respect to volume. Details of the proofs are found in Gay (1994). The above equations are discrete approximations to Gay's equations.

Our method is to find p_{ij}'s to maximise the entropy function shown in (6) subject to the constraints (2) to (5). The method is based on an explanation of maximum entropy given by Jaynes (1968).

$$-\sum_{i,j} p_{ij} \log(p_{ij}) \qquad (6)$$

Once the p_{ij}'s have been determined, equation (1) is used to calculate the volume distribution f_i.

RESULTS

To test the algorithm, three sets of computer-generated particles were created. The particles were ellipsoids with the cumulative volume distributions shown in Fig. 1 to Fig. 3. The horizontal axis label 'size' refers to the equivalent spherical diameter.

Sectioning of these particles was simulated and our method applied to estimate the volume distributions. The estimates are also shown in the figures. In all three tests, the estimated volume distribution is a good approximation to the true distribution. The level of accuracy is acceptable for many practical purposes.

The method smooths corners in all three cases. This might be a weakness of our method. Alternatively, it could be that the sections do not contain enough information to accurately detect corners, either because there are not enough sections, or because the problem is indeterminate.

The example in Fig. 2 shows that our method has difficulty distinguishing two nearby peaks. Again, it might not be possible to remove this problem. However, the method does perform well when the peaks are well separated, as Fig. 3 shows.
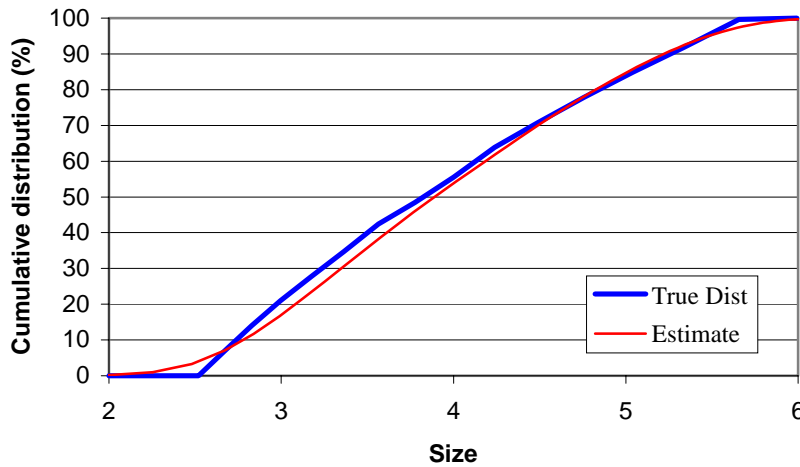

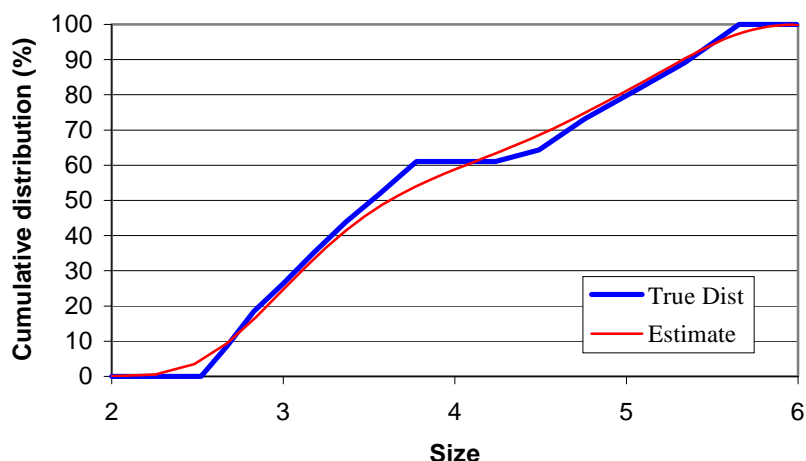Fig. 1. *Correction for first set of ellipsoids.*

Fig. 2. *Correction for second set of ellipsoids. There are two peaks close together.*
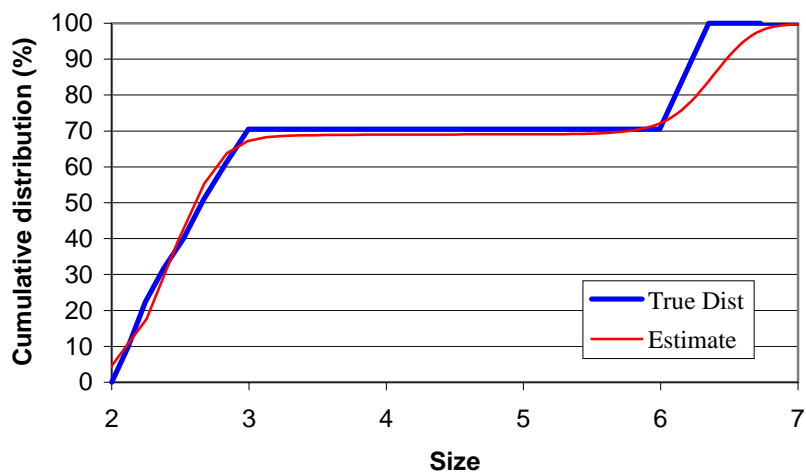

Fig. 3. *Correction for third set of ellipsoids. There are two well separated peaks.*

DISCUSSION

The data used to test the method were obtained as follows. A computer program generated ellipsoids of various shapes and sizes. Exactly one IUR section was simulated for each ellipsoid. For each section, the volume of the associated particle was recorded. The true distributions were obtained by calculating the area-weighted distributions of the recorded volumes.

The computer-generated particles represent particles that have been intercepted by an IUR section. Consequently, large particles are over-represented in the sample. It is therefore not valid to obtain the true distribution by calculating the volume-weighted distribution of the recorded volumes. However, the area-weighted distribution of the recorded volumes is a valid reconstruction of the true distribution.

The estimated distributions were obtained by discarding the recorded volumes and then executing the algorithm described above.

The authors have developed a similar method for estimating distributions of composition. The method has been extended for particles containing more than two phases. The authors believe that the method can be further extended to estimate distributions of many geometric properties simultaneously.

These features are being incorporated into the commercial software package JKStereo. The authors are interested in persons wishing to use the algorithms or collaborate on further development of the methods.

CONCLUSION

Volume distributions can be accurately estimated using IUR sections. Geometric probability provides information about the volume distribution and maximum entropy provides a technique for processing this information. In the examples tested so far, the distribution estimated using the method is close to the true distribution. This indicates that very little information has been lost in the sectioning process.

A preliminary report of some of the data was presented at the Xth International Congress for Stereology, Melbourne, Australia, 1-4 November 1999.

REFERENCES

Cruz-Orive L-M (1976). Particle size-shape distributions: the general spheroid problem. J Microsc 107:235-54.

Gay SL (1994). Liberation modelling using particle sections. University of Queensland PhD thesis.

Gay SL (1996). A geometric probability approach to estimating particle volume distribution. Acta Stereol 15/1:25-30.

Gay SL (1997). Stereological estimation of the surface area of phases within particles. Acta Stereol 16/1:7-18.

Jaynes ET (1968). Prior Probabilities. IEEE Transactions on Systems Science and Cybernetics 4/3:227-41.

Jensen EB (1991). Recent developments in the stereological analysis of particles. Ann Inst Statist Math 43:455-68.

Jensen EB, Gundersen HJG (1985). The stereological estimation of moments of particle volume. J Appl Prob 22:82-98.

Jensen EB, Gundersen HJG (1987). Stereological estimation of surface area of arbitrary particles. Acta Stereol 6:25-30.

Jensen EB, Gundersen HJG (1989). Fundamental stereological formulae based on isotropically oriented probes through fixed points with application to particle analysis. J Microsc 153:249-67.

King RP (1982). Determination of the distribution of size of irregularly shaped particles from measurements on sections of projected areas. Powder Tech 32:87-100.

Wicksell SD (1925). The corpuscle problem I. Biometrika 17:84.

Wicksell SD (1925). The corpuscle problem II. Biometrika 18:152.