doi: 10.1255/tosf.124

Material intrinsic heterogeneity: statistical derivation

Geoff Lyman
Principal, Materials Sampling & Consulting, materials-sampling-and-consulting.com

The value of a fully statistical sampling theory is that it is possible to quantify a measure of material intrinsic heterogeneity and, on this basis, provide the entire distribution of the analyte content of potential samples to be extracted from the lot. The analyte content of a sample of a given mass is a random quantity as samples of nominally equal masses taken from a lot in a given state of comminution will not have exactly the sample analyte content. The analyte content of a sample is correctly described as a random variable and to characterise a random variable completely it is necessary to know either the probability density function or distribution function for the random variable, or all of the moments of the random variable (mean, variance and all the higher moments). The following discussion derives the fundamental sampling variance from a purely mathematical statistics basis, relying on the *assumption* that the number of particles of any one type (size and analyte content) that fall into a sample taken in a mechanically correct manner (following the principle of equiprobable sampling) follows a Poisson distribution. In addition, the Poisson distributions of particle numbers are statistically independent. A more fully argued substantiation of this fundamental assumption, partial experimental evidence and standard statistical introduction to the definition and properties of the Poisson distribution, and reasons for its use, can be found at the end of this article.

Material heterogeneity

The intrinsic heterogeneity (IH) of a particulate material with respect to a particular analyte or component of the material reflects the extent to which the size and compositions of the particles differ. A jar of identical white marbles (same size, mass and chemical composition) possesses zero IH. A jar containing both black and white marbles of the same size possesses some positive IH (IH cannot be negative as it is a variance). Similarly, a sack of wheat will possess some intrinsic heterogeneity, but this will differ depending on what aspect of the wheat is under scrutiny. There will be one level for the IH with respect to nitrogen, another due to concentration of mycotoxins and another with respect to the content of small mineral or stone particles of the sack of grain. Likewise, a sample of a broken base metal ore will have different IH measures with respect to Cu, Ag, Au, Fe, Zn, Pb and S. IH is essentially a measure of the variability of a material with respect to the mass of an analyte carried within a particle. If all particles are the same, as the case of white marbles, the IH is zero.

Derivation

In a general mixture of particles there will be a range of particle sizes and, within each size fraction, there will be a range of particle compositions with respect to a target analyte. The material can be conceptually broken down into particle classes within which all particles are assumed to have the same nominal volume v_k. Then within each volume class, particles can be put into composition classes of average analyte content a_{kp} and average density ρ_{kp}. The k^{th} volume class represents a mass fraction x_k within the mixture and the p^{th} composition class represents a mass fraction y_{kp} within the k^{th} size class. Table 1 summarises the notation for this model of the particles. The model can be made arbitrarily accurate by expanding the number of volume and composition classes. For fine particles (<5 mm), it is possible to collect information on individual particle sections using an automated scanning electron microscope (Qemscan or MLA or other machine) from which this information can be calculated directly.

This form of a material model can be adapted to most particulate mixtures, but low grade gold ores are a possible exception, especially when the gold is present as relatively large grains within particles (in general, the term "grain" can be used to refer to a contiguous volume of a phase within a larger "particle" which is a single contiguous and distinct volume of a material). In such a case it is the size distribution of the gold

Table 1. Descriptors of a particulate material—the material model.

Symbol	Description	Units	Convenient unit
N_s	Number of size classes into which the material is divided	—	—
N_c	Number of composition (analyte content) classes into which the material is divided	—	—
x_k	Mass fraction of the total lot or sample falling in the k^{th} size class	—	—
y_{kp}	Mass fraction of the p^{th} composition class within the k^{th} size class	—	—
v_k	Volume of the average fragment within the k^{th} size class	L^3	cm^3
ρ_{kp}	Density of the average fragment in the p^{th} composition and k^{th} size class	$m\,L^{-3}$	$\text{g}\,\text{cm}^{-3}$
a_{kp}	Average concentration of the analyte in the p^{th} composition and k^{th} size class—may be a mass fraction or other (w/w) concentration unit such as ppm	—	—

grains that is of primary importance. Special methods have to be adopted to determine the size distribution of the gold.

The usual definition of the volume of the average fragment in the k^{th} size class is

$$v_k = f\,\frac{d_{k-1}^3 + d_k^3}{2} \tag{1}$$

where d_{k-1} and d_k are the square screen apertures defining the size fraction and f is a shape factor usually taken to be 0.5. Other definitions may be taken such as the geometric mean size which is the square root of the product of the defining sieve apertures with the inclusion of a shape factor.

Note that sorting particles into density classes is not the same as sorting into composition classes. A density class can have a substantial range of compositions with respect to a particular analyte if there are more than two mineral phases in the particle mixture. Similarly, a composition class with respect to a particular analyte can contain particles of a range of densities in a multiphase mineral mixture.

The general equation for the composition of a sample, based on conservation of mass, can be written as

$$A_S = \frac{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} a_{kp} m_{kp} N_{kp}}{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} m_{kp} N_{kp}} \tag{2}$$

where the N_{kp} are the numbers of particles and the m_{kp} are the masses from each size-composition class appearing in the sample. The numbers, N_{kp} are taken to be independent Poisson random variables. A_s is a random variable as it is a function of other random variables.

The variance of the sample composition can be derived by the well-known rules of propagation of variance by which the composition is expanded in a Taylor series about the expected value of the number of particles. The derivatives are

$$\left.\frac{\partial A_S}{\partial N_{kp}}\right|_{\tilde{\mathbf{N}}} = \frac{a_{kp} m_{kp}}{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} m_{kp}\tilde{N}_{kp}} - \frac{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} a_{kp} m_{kp}\tilde{N}_{kp}}{\left(\sum_{i=1}^{N_S}\sum_{j=1}^{N_C} m_{kp}\tilde{N}_{kp}\right)^2} m_{kp} \tag{3}$$

and are evaluated at the expected number of particles in the sample. Expected values of a quantity are denoted by a tilde, for example, $\tilde{N}_{kp}$.

The expected sample mass is identified as

$$\tilde{M}_S = \sum_{k=1}^{N_S}\sum_{p=1}^{N_C} m_{kp}\tilde{N}_{kp} \tag{4}$$

and the expected sample analyte content is

$$\tilde{A}_S = \frac{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} a_{kp} m_{kp}\tilde{N}_{kp}}{\sum_{k=1}^{N_S}\sum_{p=1}^{N_C} m_{kp}\tilde{N}_{kp}} \tag{5}$$

so the derivatives can be simplified as

$$\left.\frac{\partial A_S}{\partial N_{kp}}\right|_{\tilde{\mathbf{N}}} = \frac{a_{kp} m_{kp}}{\tilde{M}_S} - \frac{\tilde{A}_S}{\tilde{M}_S} m_{kp} = \frac{m_{kp}}{\tilde{M}_S}\left(a_{kp} - \tilde{A}_S\right) \tag{6}$$

Using only the first term in the expansion, the variance is by definition

$$\begin{aligned} \operatorname{var}\{A_S\} &= \sum_{k=1}^{N_S}\sum_{p=1}^{N_C}\left(\left.\frac{\partial A_S}{\partial N_{kp}}\right|_{\tilde{\mathbf{N}}}\right)^2 \operatorname{var}\{N_{kp}\} \\ &= \sum_{k=1}^{N_S}\sum_{p=1}^{N_C}\left(\frac{m_{kp}}{\tilde{M}_S}\left(a_{kp} - \tilde{A}_S\right)\right)^2 \operatorname{var}\{N_{kp}\} \end{aligned} \tag{7}$$

Since the particle numbers are Poisson random variables for which their variance is equal to their expected value,

$$\operatorname{var}\{N_{kp}\} = \tilde{N}_{kp} \tag{8}$$

so

$$\begin{aligned} \operatorname{var}\{A_S\} &= \sum_{k=1}^{N_S}\sum_{p=1}^{N_C}\left(\frac{m_{kp}}{\tilde{M}_S}\left(a_{kp} - \tilde{A}_S\right)\right)^2 \tilde{N}_{kp} \\ &= \frac{1}{\tilde{M}_S}\sum_{k=1}^{N_S}\sum_{p=1}^{N_C}\frac{m_{kp}\tilde{N}_{kp}}{\tilde{M}_S} m_{kp}\left(a_{kp} - \tilde{A}_S\right)^2 \end{aligned} \tag{9}$$

The usual variables in the model of the material can be identified as follows

$$\begin{aligned} x_k y_{kp} &= \frac{m_{kp}\tilde{N}_{kp}}{\tilde{M}_S} \\ m_{kp} &= v_k \rho_{kp} \end{aligned} \tag{10}$$

Then

$$\operatorname{var}\{A_S\} = \frac{1}{\tilde{M}_S}\sum_{k=1}^{N_S} x_k v_k \sum_{p=1}^{N_C} y_{kp}\rho_{kp}\left(a_{kp} - \tilde{A}_S\right)^2 \tag{11}$$

or

$$\frac{\operatorname{var}\{A_S\}}{\tilde{A}_S^2} = \frac{1}{\tilde{M}_S}\sum_{k=1}^{N_S} x_k v_k \sum_{p=1}^{N_C} y_{kp}\rho_{kp}\left(\frac{a_{kp} - \tilde{A}_S}{\tilde{A}_S}\right)^2 \tag{12}$$

The result of the derivation is the following expression for the sampling constant

$$K_S = \sum_{i=1}^{N_S} x_i v_i \sum_{j=1}^{N_C} y_{ij}\rho_{ij}\left(\frac{\tilde{A}_S - a_{ij}}{\tilde{A}_S}\right)^2 \tag{13}$$

The sampling constant is an intrinsic property of the material; hence the term “Intrinsic Heterogeneity”. An intrinsic property of a material is one that does not depend on the mass or amount of the material present. In his work, Pierre Gy refers to a quantity he calls the “Constitution Heterogeneity” (CH). There is a difference between these two quantities, as Gy’s CH is a dimensionless measure. Pitard makes the same definition of the intrinsic heterogeneity as above.

The IH is more appropriate and in line with the naming of properties in thermodynamics and the engineering literature. It is a direct measure of the heterogeneity of the material with respect to a given analyte in a given state of comminution. It has the units of mass.

The relative variance of the sample analyte content when taking a sample of nominal mass $\tilde{M}_S$ is

$$\frac{\operatorname{var}\{A_S\}}{\tilde{A}_S^2} = \frac{K_S}{\tilde{M}_S} \tag{14}$$

This expression can also be written in terms of particle numbers and masses but Equation 13 is the most convenient.

This expression for the sampling constant is entirely consistent with Gy’s work. It is also consistent with an expansion of the denominator in the expression for the sample assay, which provides another means of making the derivation.

It must be stressed that the sampling constant will differ for each chemical element of interest in the mixture unless the elements of interest all occur only in a single mineral phase. The sampling constant may also be calculated to apply to a mineral phase in the mixture. If an element is carried by only one phase in the mineral mixture, then the sampling constant for the element will be equal to the sampling constant for the phase.

It is convenient to define a heterogeneity for a size fraction as

$$IH_k = \sum_{p=1}^{N_C} y_{kp}\rho_{kp}\left(\frac{a_{kp} - \tilde{A}_S}{\tilde{A}_S}\right)^2 \tag{15}$$

Note that this quantity has units of density rather than mass.

This quantity is a mass-weighted second moment of the particle composition about the sample analyte content (not the analyte content of the size fraction itself) and generally (but not always) increases as one moves from one size fraction to a finer one.

For a distribution of particle composition, as shown in Figure 1, the quantities involved in the summation, except for the density term, are shown on the Figure.

With the above definition, the sampling constant can be written as

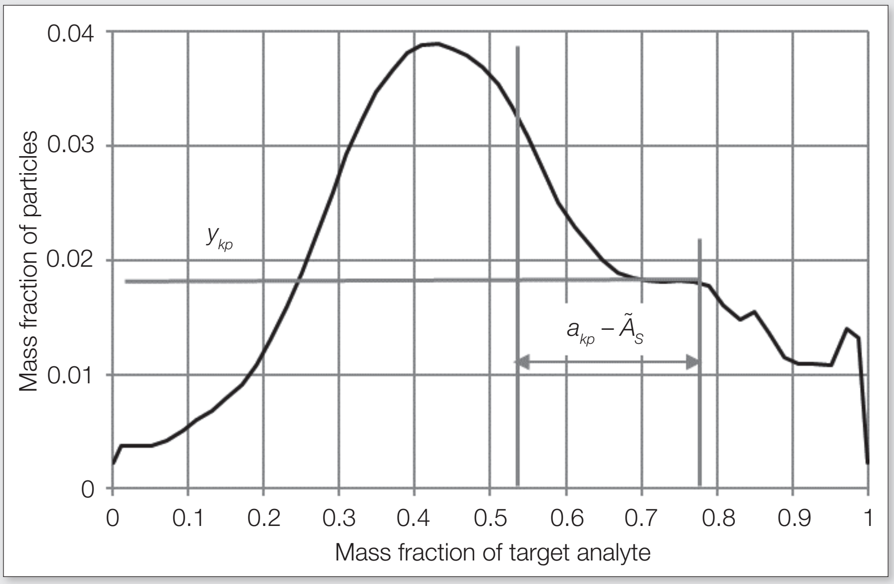

Figure 1. Illustration of terms involved in calculation of the IH value for a size fraction.

$$K_S = \sum_{i=1}^{N_S} x_i v_i IH_i \tag{16}$$

This form of the equation emphasises that the sampling constant is a weighted sum of the IH values for the size fractions where the weighting factor is the product of the mass fraction of the i^{th} size fraction in the material and the nominal volume of the particles in that size fraction. Even if the IH values for the size fractions increase modestly as finer particles are considered, the volume weighting tends to be a dominant factor, so it is the IH values for the coarser size fractions that dominate the value of K_s. The exception to this is nuggetty gold ores and diamonds.

This derivation demonstrates that it is possible to derive the expression for the constitutional heterogeneity, or the intrinsic heterogeneity, which is an important component of the fundamental sampling variance (error), from a purely statistical basis without reliance on a concept of material heterogeneity.

The derivation is based on Poisson distributions of particles numbers that fall into "correctly extracted" samples. This basis of Poisson distributions of particle numbers can also be used to derive the full sampling distribution using the method of characteristic functions. When the analyte content of the sample defined in Equation (2), is expanded as a Taylor series in particle numbers, it becomes a weighted sum of independent Poisson random variables for which the characteristic function can be written explicitly. Inversion of this function then provides the probability density function for the analyte content of potential samples from a lot. Knowing the form of the characteristic function also permits calculation of all the moments of the sampling distribution of which the sampling variance is simply the second central moment. These expressions are explicit and require no inversion calculation.

The Poisson Distribution: brief tutorial

When events occur at random but with a particular average rate, λ, the number of events occurring in a defined period of time follows a Poisson distribution. If the period of time or the width of an interval is denoted as t, the expected number of events in t is $n = \lambda t$. The probability that J events will take place in this interval is

$$p(J \mid n) = \frac{n^J e^{-n}}{J!} \tag{17}$$

The expected value of the number of events and the variance of the number of events are equal, so

$$\mathrm{E}\{J\} = \mathrm{var}\{J\} = n \tag{18}$$

The characteristic function of the Poisson density is

$$\varphi(u) = \exp\left[n\left(e^{iu} - 1\right)\right] \tag{19}$$

The time or distance between consecutive events in a Poisson process is exponentially distributed.

The probability density function is illustrated in Figure 2. The function is substantially skewed for low expected values, but becomes more symmetrical as the expected value increases.

Why is the assumption of Poisson distributed particle numbers a sound and reasonable one? There are several arguments that can be marshalled to support the use of the Poisson distribution for particle numbers in correctly taken samples.

First, from a mathematical viewpoint, the rules of propagation of variance used to arrive at the results in Equations (11) or (12) are known to be a correct methodology. If the variance of the number of particles were not taken to be equal to the expected value of the number of particles, Gy's result for the sampling variance would not result from the present analysis. The sample variance would be either smaller or larger. The assumption of the property of the Poisson distribution is the only one that leads to Gy's result, because there is no other discrete distribution having a variance equal to the expected value.

Second, there is experimental evidence. It is possible to take a batch of small (about 2.5 mm) coloured plastic particles, combined in known proportions, and pass a sample of this material through a small rotary sample divider to create a number of subsamples. The particles in each of the subsamples can then be counted and the variance of the numbers over the subsamples can be compared with the average or the expected number. In courses given by the author, participants and students have carried out this exercise with results that are statistically compatible with the Poisson assumption.

Third, from the point of view of ways of describing randomness, picture a set of particles laid out along a line in the most random possible manner, in such a way that there is some average number of particles per unit length when a long line segment is considered. If particles are placed at positions along the line in the most random manner possible, it is natural to think that the probability of finding a particle in any one short line segment "*dx*" is constant; no part of the line is favoured. This uniform distribution of the particles along the line leads to the mathematical conclusion that the number of particles in any finite line

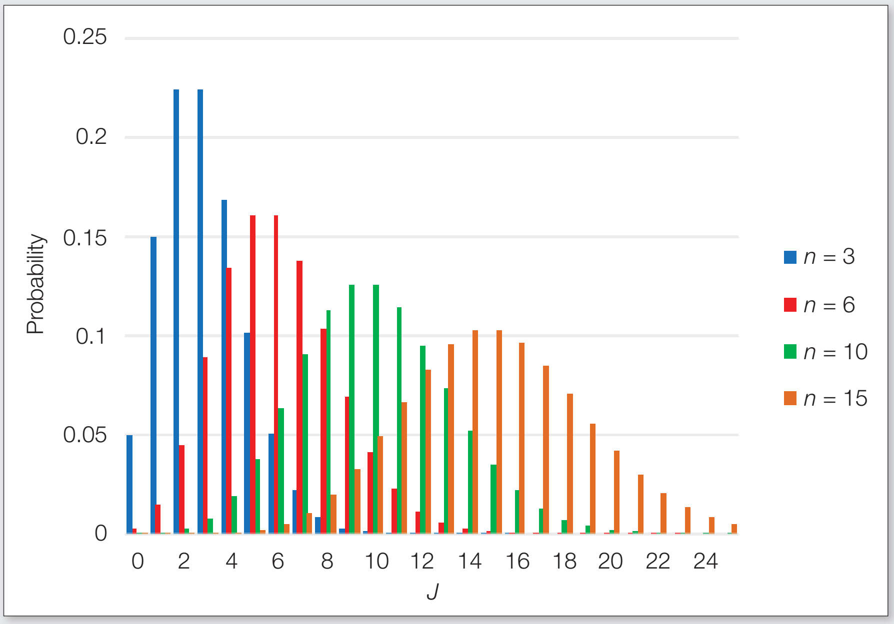

Figure 2. Examples of the Poisson distribution for various expected values.

segment of length L will follow a Poisson distribution. The nature of placement of the particles is known to statistical mathematicians as a Poisson point process. So, if one were to take a “stopped belt” sample over a particular length of the line, the number of particles in that sample would follow a Poisson distribution with an expected value proportional to the particular length considered. The Poisson point process for distributing particles in space is the most random and statistically uniform of all point processes and is, therefore, the most reasonable assumption to make regarding the distribution of the numbers of particles falling into a correctly taken sample.

Finally, a fourth argument can be made on the basis of equiprobable sampling. If particles are selected from a lot one by one and at random and the entire set is accessible for selection (correct sampling), the number of particles of any one type that are selected will follow a hypergeometric distribution (drawing elements at random from a finite set without replacement). However, if the size of the set is large compared to the number of particles selected, the hypergeometric distribution converges to a binomial distribution. And, when the size of the set is very large, the binomial distribution can be represented by a Poisson distribution. These limiting cases are well-known in the statistical literature.