Complete sampling distribution for primary sampling, sample preparation and analysis

Geoffrey J. Lyman

Materials Sampling & Consulting, Southport, Queensland, Australia 4215. E-mail: glyman@iprimus.com.au

Following from the author's recent paper at Sampling 2014 which presented a method for calculation of the sampling probability density function due to the particulate heterogeneity (density function of the fundamental sampling uncertainty), it is possible to apply the same characteristic function method to arrive at the overall sampling distribution for any sampling protocol and analysis method. This paper develops the application of the method of characteristic functions to the overall sampling problem including the uncertainty which derives from the primary sampling from a process stream. The assay distribution in a process stream or of impurities in the flow of a final product can be governed by non-Gaussian, serially correlated distributions. The paper shows how such circumstances can be dealt with to arrive at robust solutions. The paper represents an end-point in the theory of sampling as it provides a means of determining the entire distribution function for a sampling system. Such a determination has not previously been possible and having determined the entire distribution function, the statistics of the sampling process are completely determined.

Introduction

One cannot say too much about the theory of sampling put forward by Pierre Gy. Gy consolidated the elements of sampling theory that had been proposed over the years leading up to his definitive works in the 50s and 60s. For the English-speaking world, the theory of sampling arrived in 1979 with his book published by Elsevier.

This was close to the time when I first began to take an interest in sampling theory so that I could design plant tests intended to reveal the performance of unit operations in coal processing, with which I was then involved. There was always in mind that party A claims an increase in yield of X percent while party B claims an increase of Y percent. Who was correct? How were the trials carried out? Is it possible to assess the uncertainty involved in the claims of improved performance? If everyone who carried out trials was correct, the yield of product would be 120% of the feed content.

The resolution of this conundrum is found in the provision of estimates of uncertainty to be attached to each of the quantities measured in the test work and to have those uncertainties propagated through to the final figure for recovery, yield or whatever performance indicator is preferred. Unlike the physicists who were working at the time when I was an undergraduate and graduate engineer learning my trade, I found that the mineral processors never provided error bars on their results in the same way that the physicists and chemists did. I found this to be an unscientific approach and to be rather political in nature. An engineering or physical quantity has no validity until there is a reasonably accurate estimate of its uncertainty that is stated along with the figure.

Pierre Gy waged a campaign to bring the mineral processors into the world of modern science by focussing on the uncertainties that we experience when doing test work or running a plant. While a number of investigators had made estimates of sampling variance due to the particulate nature of a mineral mixture, Gy created a mathematical structure that could be used in a coherent fashion to describe the variance of sampling due not just to the particulate nature of the mineral but also to the process variance in the flow in the plant that was being sampled. His recognition that he could borrow from the nascent theory of geostatistics to describe the variance due to grade variation in a process stream was a unique and brilliant step forward. This very important component of sampling variance had been ignored up to that point in time.

The mineral processing world is still struggling to come to terms with the power of Gy's work. The full power of his theory is often neglected in the design of sampling systems. We have new analytical tools to look at fine particle compositions that permit the implementation of the detail of Gy's work; we don't have to guess at a liberation constant any more or postulate how that value may vary with the top size of the sample.

I have recently been lucky enough to come across some work by a skilled statistician that lead me to develop a means of estimating the entire sampling distribution, due to all factors. This paper presents the outline of how these calculations are made.

I respectfully dedicate this presentation to Pierre whose work has been a constant inspiration since I learned of it and met him many years ago in Sydney.

The paper will briefly recall the mathematical method by which the calculations can be made and will then provide an example of the outcome of the calculations, for a gold ore.

Mathematical background

The method of calculation of the entire sampling distribution is based on the fact that given a set of random variables that are statistically independent and each have arbitrary probability density functions, the characteristic function for the probability density of a weighted sum of the random variables is determined from the product of the characteristic functions for each of the random variables. This is a fundamental relationship of mathematical statistics. In fact, the probability density function and the characteristic function are Fourier transform pairs. Knowledge of the characteristic function for a random variable is equivalent to knowledge of the probability density function.

For every probability density function, $p(x)$, the characteristic function is defined as

$$\varphi_X(u) = E\{e^{iuX}\} = \int_{-\infty}^{\infty} e^{iux} p(x) dx = \int_{-\infty}^{\infty} [\cos(ux) + i\sin(ux)] p(x) dx$$

(1)

where, $i = \sqrt{-1}$ so it is a complex-valued function and is a Fourier transform of the density function. Given the characteristic function

of a density function, the density function can be recovered as an inverse transform yielding a real-valued function

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} \varphi(u) \, du \qquad (2)$$

The characteristic function has the following properties

$$\varphi(0) = 1$$
$$|\varphi(u)| \leq 1 \qquad (3)$$
$$\varphi(-u) = \overline{\varphi}(u)$$

where $\overline{\varphi}(u)$ is the complex conjugate of $\varphi(u)$.

It is also possible to calculate the non-central moments of the density function directly from the characteristic function without making the inversion, as

$$\mu_r' = i^{-r} \left. \frac{\partial^r \varphi(u)}{\partial u^r} \right|_{u=0} \qquad (4)$$

This last relationship is very useful as one can find the sampling variance without having to make an inversion.

In sampling a process stream for a particular critical content, when the sampling is carried out in a mechanically correct manner, there are only three sources of variance:

■ variance due to the time variation of the critical content of the process stream
■ variance due to the intrinsic (particulate) heterogeneity of the primary increments and the subsamples retained in the sample preparation protocol, including that of the analytical aliquot
■ variance due to the final analysis of the aliquot by some appropriate means

It is usual to base the calculation of the variance due to intrinsic heterogeneity on an average composition of the material being sampled, although this is not mandatory. It is then implicit that the variance due to intrinsic heterogeneity depends only on the particular state of comminution of the material being sampled. Indeed, the distribution of the uncertainty due to intrinsic heterogeneity is taken to be dependent only on the average grade and the state of comminution of the material.

In such a case, it is possible to state that the three sources of uncertainty are statistically independent. Consequently, if it is possible to determine the probability density functions for each of the three sources of uncertainty, it will be possible to calculate the probability density function for the sampling protocol as a whole by finding the characteristic functions for each of the sources of uncertainty, taking their product and inverting this product. Even if the probability density function varies with the state of comminution of the subsample within the sampling protocol, that change can be accommodated within the procedure by introducing additional independent density functions into the calculation.

The first source of variance, due to time variation of grade in the stream is something about which we know very little in practical terms as it is very rare to undertake sampling campaigns in which frequent increments are taken with preparation and analysis of individual increments. Such information as we have usually comes from on-line analysis systems. For the purpose of this analysis, the grade variation in the process stream under examination will be simulated on a very fine time scale (all potential increments will be created) and the stream will be sampled at an appropriate frequency over an 8 hour shift.

We know more about the second source of variance if an appropriate investigation of the ore is undertaken to determine the size distribution of the gold grains. This information is absolutely fundamental to the sampling of the material and development of sampling protocols. The required information can be developed in conjunction with gravity recoverable gold studies.

The final source of uncertainty, the analytical uncertainty can be determined from laboratory duplicate assay information. It is important that such information be uncensored (all assays made must be captured by the laboratory information management system). This distribution should be Gaussian if the protocol and method is correct.

Example

The calculations will be illustrated by considering a variable feed to a gold plant that is treating ore from two different sources, one of
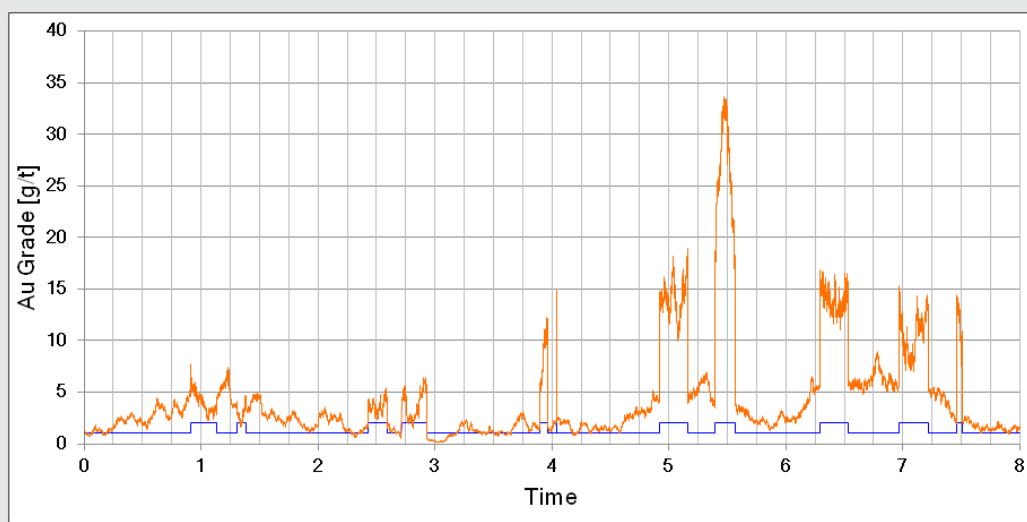

Figure 1. Typical simulated variation of feed gold grade, showing the switching between the two feed types (bottom trace).

which has a higher grade than the other and a different size distribution of gold in the ore. We take the following case:

■ Ore A is treated 25% of the time and is the higher grade material
■ The switching of the feed between the two sources is random in that the duration for which each ore is treated follows an exponential distribution with an expected duration of 10 minutes (high grade) and 30 minutes (low grade).
■ Both ores carry gold with two distributions of grain sizes to explore the impact of 'coarse gold'.

Feed variation and primary sampling density function

A typical trace of the feed gold grade is shown in Figure 1. Both the low grade and the high grade material are taken to follow gamma distributions of grade. The low grade material has an average of 2 g/t and the high grade material, 10 g/t. Both distributions have an order of two. Their density functions are shown in Figure 2.

The time variations of the grade for both ore types are taken to follow random functions with an exponential covariance functions with a range of about 70 minutes and are used in such a way that the variation in an ore type remains correlated even when interrupted by feeding of the other ore type. This simulates feeding alternately from one of the two ore sources.

The simulated trace of grade as a function of time is sampled at a 15 minute period (32 increments per shift) and the grade of the accumulated sample is compared to the true grade for the time period. The sampling is of course unbiased. The quantity of interest is the distribution of the difference between the true unknown grade and the grade of the sample as that is the sampling uncertainty. It is essentially impossible to calculate this distribution *a priori*, so the simulation method must suffice.

The simulations were run 5000 times, simulating 5000, 8 hour periods on a time base of 1 second (28800 points in the simulation) Such large simulations require special methods to ensure that the simulation is exact. Methods such as sequential Gaussian simulations are not exact. The histogram of differences was extracted and was recognised to follow a Laplace distribution very closely (double sided exponential distribution). The distribution parameters were extracted by a fitting method. The result is shown in Figure 3 and the Laplacian model is very good indeed (could this be a general result for the sampling distribution?).
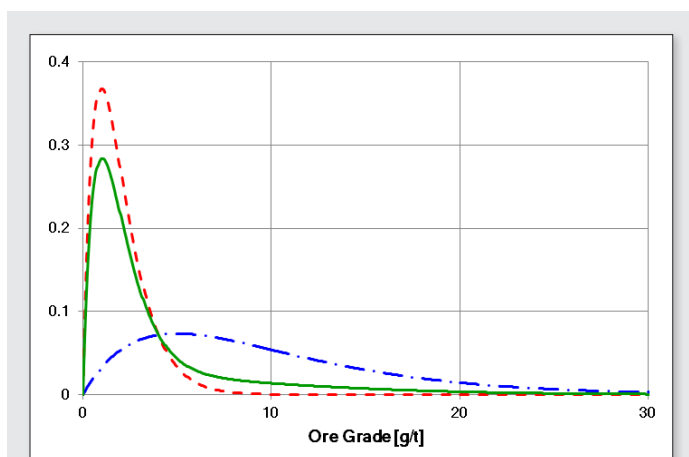
Figure 2. Probability density functions for the low grade (- - -), high grade (- . - .) and average grade (solid line).
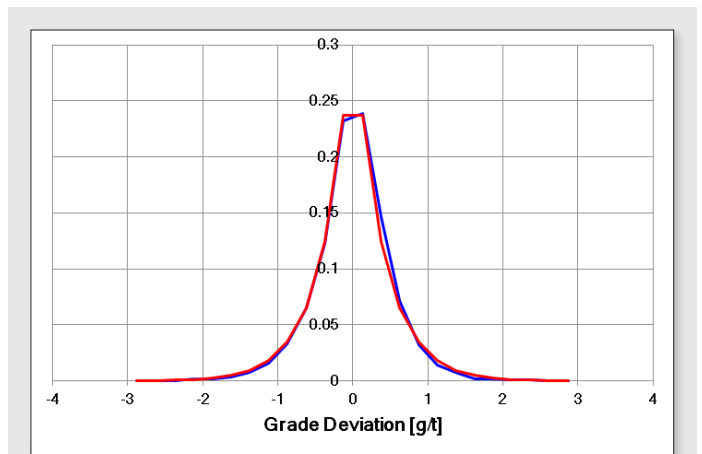
Figure 3. Fit of the Laplacian density function to the histogram on an interval of 0.1 g/t for the difference between true and sampled shift grade. 5000 simulations. Density function parameter = 0.389 g/t.

The Laplace density function centred on zero is given by

$$p(x) = \frac{1}{2\alpha} e^{-\frac{|x|}{\alpha}} \quad ; \quad -\infty < x < \infty \tag{5}$$

The tails of the Laplacian distribution are much heavier than a normal distribution. The variance of the distribution is $2\alpha^2$, so the standard deviation of the sampling distribution is 0.550 g/t and a 95% confidence interval is ±1.17 g/t (±2.12σ).

Intrinsic heterogeneity of the sampled ore and density function

To correctly evaluate the sampling distribution for the plant feed, it is necessary to consider the two ore types and possible dilution by barren material. The methodology for doing this was provided by Lyman.[1]

For the case at hand, the intrinsic heterogeneity of the ore will be based on a 75:25 mix of the two ores and dilution will be ignored (dilution would further broaden the distribution described above).

The plant feed is considered to have a 95% passing size of 200 mm and to be fed at 1000 tph. The cutter has an aperture of 0.6 m and a speed of 0.6 m/s, leading to a primary increment mass of 277 kg, which is numerically equal to the feed rate in kg/s. The increment therefore represents 1 second of plant feed (hence the simulation of the feed on a 1 second basis). The sampling interval is 15 minutes, providing 32 increments, as noted above.

The two ores are taken to carry both fine and coarse gold. For convenience, the mass distributions of the gold particles are taken to follow Weibull (Rosin-Rammler) distributions. The mass fraction passing at a sieve size *d* is given by

$$w(d) = 1 - \exp\left[-\left(\frac{d}{\tilde{d}}\right)^n\right] \tag{6}$$

where $\tilde{d}$ is the size at which 0.632 of the sample passes.

The distributions are shown in Figure 4.

In sampling gold ores, it is very important to recognise that the objective of sampling is to collect an adequate number of gold particles in order to control the variance and distribution of sampling. When there is no dilution, the uncertainty in the grade of the sample derives only from the Poisson distribution of the number of
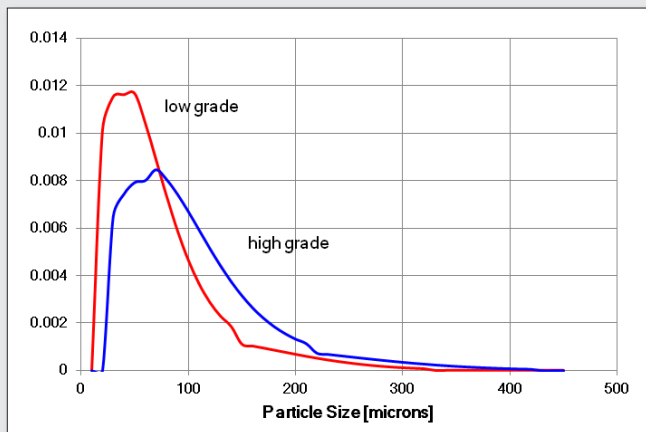
Figure 4. Gold particle size distributions, each a weighted sum of Weibull distributions, for low and high grade ore.

Figure 6. Total sampling distribution, with distribution due to sample intrinsic heterogeneity for comparison.

gold grains arriving in the sample. It does not matter whether these grains are contained in large or small particles of the host rock. The link between the number of particles of gold and the grade of the sample is the distribution by mass of the gold particles: their size distribution. Once the size distribution (or better still, the mass distribution if the gold is present in particles of a complex structure) is defined, the sampling distribution can be calculated.

If a sampling constant, K_s, for the gold is defined such that

$$\frac{\sigma}{\overline{x}} = \sqrt{\frac{K_S}{M_S}} \tag{7}$$

where σ is the variance of sampling and $\overline{x}$ is the mean grade and M_S is the mass of the sample, the value of the sampling constant for the gold will not change until the comminution of the ore is such that the number of gold particles increases due to the breaking of the particles. It is not enough to 'liberate' or flatten the particles, they must break. When a subsample is taken at any stage in the sample preparation protocol, the only impact is the expected number of gold particles in the subsample. With the smaller mass being retained, the sampling variance will increase, but the sampling constant will not change unless the gold particles have been broken.
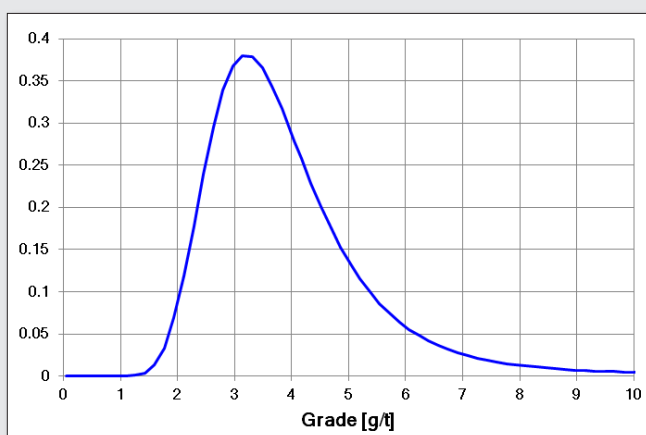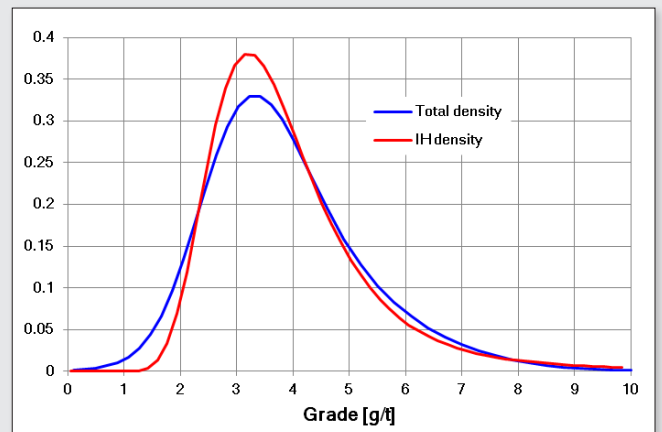

Figure 5. Grade distribution of 30 g aliquots of the ore.

The situation differs when there are totally barren fragments as these will impact on the sampling constant, but this is not the case in this example.

For the purpose of illustration of the calculation of the total sampling distribution, we will take the sampling to the stage at which we have 30 g of comminuted material for fire assay. We will assume that the gold particles have not been broken, but simple flattened by the sample pulverisation. The grade distribution of 30 g aliquots is shown in Figure 5 and has an expected value of 4 g/t.

Analytical density function
Given that the expected grade is 4 g/t, it is not unreasonable to assume that the uncertainty attached to the fire assay procedure itself is normally distributed. Note that this uncertainty does not include the intrinsic heterogeneity of the analytical aliquot. If this factor were to be included, the distribution might be skewed. The SD is taken to be 0.2 g/t, for the purpose of illustration.

Total sampling distribution
To determine the total sampling distribution, it is necessary to take the total uncertainty to be the sum of three random variables: the uncertainty due to distributional heterogeneity and a fixed number of increments per shift (a Laplace distribution), the uncertainty due to the intrinsic heterogeneity of the ore (a skewed distribution) and the uncertainty due to the analytical procedure (a normal distribution). The total distribution is shown in Figure 6.

Discussion
Up to this point in time, assessment of sampling uncertainty has been limited to knowledge of the variance of sampling without being able to assess whether the sampling distribution is skewed or, worse, bimodal.

In the realm of geostatistics, some practitioners have indulged in the practice of 'top cutting', that is, the discarding or reduction of grade values that seem to be too high. These high values destroy the 'normality' of the data and have adverse effects on the estimation of the variogram. The methods presented here, in combination with a full knowledge of the mineralogy of the ore under scrutiny, have the ability to determine precisely the distribution of the grades of samples, permitting assessment of the probability of observing

high grade results. At the start of evaluation of a potential orebody, one of the first steps should be the determination of the heterogeneity of the ore and ore samples so that the exploration samples collected are of sufficient mass to yield assay results of a precision that are fit for purpose, that is to say, representative. The tools presented here permit a full assessment for the first time.

In process sampling, especially in sampling a feed or product material where metallurgical accounting or product valuation is involved, the assumption that grade values follow a Gaussian (normal) distribution carries significant financial risk. In the metallurgical accounting setting, skewed grade distributions for assays can invalidate the reconciliation of the material balance. The classical example of this is the weighted least squares adjustment of assays, to arrive at a material balance that closes exactly, that produces adjusted tailings grades that are negative. Failure to take non-normality of tailings (or other process flows) grades into account distorts the entire material balance leading to bias in the adjusted balance.

In the sale of any commodity, or the valuation of a raw ore for toll treatment, knowledge of the sampling distribution is vital to the construction of the contract of sale and the development of quality reconciliation agreements between the seller and buyer. It is not uncommon to see agreements that involve unrealistically narrow splitting limits let alone ones that do not take possible skewness of grades into account. Responsible development of contracts must be based on knowledge of the full sampling distribution of a product or ore, so that an operational characteristic curve can be drawn up that reveals the financial risk to both seller and buyer.[2,3]

Conclusion

The foregoing example demonstrates that it is possible to develop a full statistical model of all sources of variance impacting the sampling of a process stream. The method also has potential application in geostatistical studies. The method for dealing with ore heterogeneity has been dealt with in detail by Lyman.[1] A key to the advances made here is the recognition that the distribution of the uncertainty due to the extraction of primary increments from a process flow can be estimated by simulation and that that distribution can be modelled by any suitable means.

The combination of the three sources of uncertainty can be combined using the characteristic function method to arrive at a total sampling distributions.

The value of the method is particularly evident when sampling ores of precious metals where the material may be nuggetty leading to skewed distributions due to the intrinsic heterogeneity of the ore. Knowledge of the full sampling distribution is of great value in the case wherein an operation is toll treating a gold ore for a client. The contract and sampling protocol can be developed with full knowledge of the financial risk involved through the use of an operational characteristic curve which depends entirely and directly on the full sampling distribution. The fact that the full sampling distribution is known is also reassuring to the client.

The statistical basis of the theory of sampling can now be considered to be complete.

References

1. G. J. Lyman, "Determination of the Complete Sampling Distribution for a Particulate Material", Proceedings of Sampling 2014, 29-30 July 2014, Perth, Australia, AusIMM, Publication series No 5/2014, pp. 17-24.

2. F. S. Bourgeois and G. J. Lyman, "Getting high added-value from sampling", in *Proceedings of the 7th International Conference on Sampling and Blending*, Ed by K.H. Esbensen and C. Wagner, *TOS forum* **Issue 5,** 95–100 (2015). doi: 10.1255/tosf.46

3. G.J. Lyman and F. S. Bourgeois, "Sampling, Corporate Governance and Risk Analysis", to be presented at MetPlant 2015, Perth, Australia, in press.