Theory of Sampling (TOS) – Up for Debate?

By Dominique François-Bongarçon[1]

DOI: 10.62178/sst.002.004

1. Introduction

Scientific debates are useful. The World Conferences on Sampling and Blending (WCSB), besides being the biannual highpoints of social interaction for the International Pierre Gy Sampling Association (IPGSA) community, are very much also about presenting, debating and discussing the latest research results. And here there are healthy debates between ‚experts' about a number of Theory of Sampling (TOS) issues. But these have by and large rendered the understanding of the general theory and its direction of development somewhat cloudy in the minds of the general audience beyond the IPGSA boundaries.

Scientific research does not progress based on certainties, but instead on systematic skepticism. But undisciplined casting doubt about the next steps of development are not a good way to disseminate the Theory of Sampling (TOS), which is otherwise very well established, and very useful in practice - nor is this a good way to gain new adepts, nor to convince students they should be interested in pursuing it.

Here a partial selection of these issues will be reviewed, if only in a few sentences or paragraphs each, and in loose ordering, hoping to clarify the real concepts behind them, irrespective of the amount of debate they are currently triggering. No precise references are given, the reader is referred to the abundant literature on each subject, particularly in the series of WCSB conference proceedings.

2. The Legacy from Pierre Gy

The first comments will be about clarifying the WCSB conferences and their raison d'être. The first WCSB conferences were not initially conceptualized and designed to be a debating forum. The inaugural conference was specifically designed to honor Pierre Gy and his legacy. But from there, the conference concept developed itself along the way, very much without specific guidance.

But our biannual conferences have been very useful over the last twenty years for disseminating Gy's ideas and the details about his admirable work. Gy was not a promoter, he worked alone, with no associates, his circle of followers was scarce, he never read other people's works on TOS issues – so over the last 30 years much work was needed by his followers, to promote TOS, and to motivate the industrial world to use it, and universities to teach it.

Gy's work was dual: he discovered and designed the first principles of sampling, relating to how samples should be taken physically, creating the concepts of sampling correctness and segregation - and subsequently he worked out the mathematical modeling of the sampling variance for randomly taken samples, resulting in an elegant equation, famously now known as "Gy's Formula". It is important to observe that Pierre Gy was often distinctly dissatisfied with the way 'his formula' was misused, often grossly, based on a far too superficial understanding of the basic assumptions behind its derivation.

He also addressed how a new tool, the variogram, developed by G. Matheron, could be used to characterize one-dimensional estimation problems that were in fact improperly likened to sampling by users at large. Indeed, the distinction between sampling s.s. on the one hand, i.e., extracting a small mass intending to represent the whole lot, and on the other hand, measuring a concentration of interest at specific points with coordinates in some 1D (e.g. time), 2D or 3D space, over a measurement support (not a ‚sample' per se) with the aim of performing a geostatistical estimation has been very indistinct and blurred, even up to this day in many users' minds, courtesy of our relaxed day-to-day vocabulary, alas often misleading. This important distinction should hopefully clarify matters, especially in the minds of new students of TOS.

1 Agoratek International Consultants Inc., Canada.

3. Some Particular Applications of TOS First Principles

3.1 Sampling Methods

A debate exists, for instance in surface mining, about comparing blast hole (BH) and reverse circulation (RC) rig sampling. Articles have been published on a false debate that never really existed. The economics, the practicalities/applicability and the performance of each of these two sampling methods are for the users to compare in each particular case. Making it a matter of religious preference is contrary to science and detrimental to TOS dissemination. The rest is a matter of proper implementation, avoiding extraction errors, and addressing the major issues as best as possible with the right priorities. A particularly illuminating contribution to this 'debate' can be found in a paper published from a recent Ph.D. by Karin Engström (2017).

3.2 Sampling Equipment Design

There are issues about sampling equipment that are also confusing, due to the sometimes misleading terminology used by some OEM manufacturers. The only quantitative samplers that we know how to use with good results to sample a flow are i) linear and ii) circular (Vezin) cross-stream devices, provided they are used within certain limit conditions. Apart from these, we are in the unknown, with no guarantees of unbiasedness. In the case of two families of 'samplers', which have a certainty for size distribution biases: a) ‚rotary distribution' samplers (sometimes called "rotary dividers', or sometimes improperly called Vezin subsamplers), and b) classical cross-belt samplers (which are in fact mere ‚material pushers', with no reasonable chance of ever being unbiased). These are harsh judgements on some OEMs, but somebody has to state them.

In particular, all 'process control samplers' used in metallurgical plants are non-quantitative, without exception, and almost always biased. The base metal industry would be well advised to imitate the precious metals industry by adding TOS-compliant quantitative samplers to plants in order to achieve objective metal balancing.

3.3 Segregation

The propensity for segregation to be omnipresent in aggregate mixtures of minerals and similar mixtures of unit elements with different density, surface roughness, etc., cannot be effectively combated using mechanical mixing.

Within bulk materials, a variety of bed-blending methods turning the segregation to the advantage of better sampling, on the one hand, and multi-incremental sampling on the other, are, conversely, fully effective. A well-used riffle splitter is effective in removing most of the effects of segregation, but the method with the highest score is rotary splitting over a rotating carrousel fed by a vibrated feeder of the proper length, i.e., long enough for the migrating layer of material to fall over the carrousel as a fragment-fine layer.

Segregation is likely to be the next large field of research in TOS. Gy's demonstrations have included the segregation term in the theoretical variance formula, but it was deemed to be non-quantifiable and was therefore not pursued further, while Visman, using the same formalism, actually proposed some very useful quantification experiments in some particular cases.

4. Numerical Control and Mathematical Modeling

4.1 Is TOS Mathematically Complex?

The statistical model on which Gy based his demonstration of a variance formula is heavy and cryptic. The only fully rigorous demonstration, that vindicated Gy's own, was purely mathematical, though even much more cryptic and complex and it was published in French by Matheron (2015). That demonstration, more recently translated into English, augmented and commented by François-Bongarçon and Pitard is extremely complex, with heavy use of integral and differential calculus and at the end it reaches a formula that is only a first order approximation.

It is always possible to establish more didactic, simplified demonstrations, that better show the underlying theoretical structures, but the price to pay for these useful derivations, is a lack of rigor. There is not such a thing as a demonstration both simple and rigorous of the variance of a sample of particulate material when the particles have different physical properties. Matheron's demonstration shows this beyond the shadow of a doubt.

In essence, yes, TOS mathematics is complex, because corners cannot be cut while rigorously establishing a theory, but, after it has been established, there are various options for clearer derivations.

4.2 Liberation Factor

An additional difficulty that cannot be overstressed, is that the first order approximation reached by both Gy and Matheron independently of each other, can only be calculated explicitly in the case of fully liberated units (mineral grains, fragments). There is no such a thing as a general variance formula for sampling of non-liberated materials, unless the formula is added a diminishing factor between 0 and 1, called the liberation factor, l, for which no practical and generally valid model was initially proposed by Gy.

Gy knew that factor to be correlated with the degree of liberation, i.e. the proportion of liberated component of interest, and to depend on the liberation and comminution sizes dL and dN. The model $l = \sqrt{dL/dN}$ had once been proposed (but later rescinded) by Gy, because it invariably resulted in erroneous variances. In precious metals dealings, the erroneous nature of that model was directly obvious (for monetary reasons). An example from Gy published in Pitard (1993) was published, which, pushed to the limit, could be proved to be absurd by simply eliciting the liberation size.

To date only one model has been offered, which is a generalization of the above formula for l, with a variable exponent, b, between 0 and 3 (instead of the fixed approximation 0.5 used by Gy). That model has been used successfully for 30 years now. One should note though, that a model of liberation factor is strictly required only if the variance formula is to be used for predictive purposes, i.e. for predictions in which the concentration, or the comminution size, will vary. As an example, for an existing sample preparation protocol, with fixed comminution sizes, characterizing the variance of each sampling stage and optimizing only the required sample masses, can be done from duplicate samples generated at each stage.

4.3 Heterogeneity, heterogeneity testing

This brings up the issue of heterogeneity characterization and heterogeneity testing. Full heterogeneity characterization makes use of the liberation factor - explicitly, when the developed version of the variance formula is used, or implicitly, when it is replaced using a ‚Heterogeneity Factor' divided by the sample mass. Indeed, if the material is not liberated, the liberation factor is always present but embedded in said heterogeneity factor, and changes in concentration (which may trigger changes in liberation size) or in comminution size, require an explicit model of the liberation factor, lest calculations become completely illusory. The experimental calibration of that model is often called heterogeneity testing.

Even though only one model so far has been proposed for the liberation factor, many experimental methods have been proposed for heterogeneity testing. If they are well performed, their choice largely is a matter of preference. However, some common errors are often seen in heterogeneity testing studies that unfortunately invalidates them. The most common error consists of equating the one-stage formula to be calibrated to multi-stage sampling variances that have parasitic, non-primary components. The other common mistake is to ignore the variations of the liberation factor with dL an dN, by simply not using a model for it, also resulting in illusory results.

4.4 Bad Sampling Consequences

While the first principles of TOS dictate how samples or measurements should be extracted, even when they are properly applied, difficulties can still arise. In particular when sampling variances grow too large (e.g. if sample masses are too small). What happens then is the distribution of possible sample values, or its translated distribution of sampling errors, may then become overly skewed. As a result, the empirical median is much lower than the mean of the real-world distribution (which, for an unbiased sample, is the true, unknown concentration value). Consequently, more than 50% of the samples will return a value lower than the true value. Note that this is not a mathematical bias, as occasional very high values will also be returned so that, on average, all converges towards the true value. When taking a single sample, however, this is not a real consolation, especially as, on top of this, the occasionally compensating high-flier will often be capped or suppressed (the fallacy of considering all outliers faulty). Additionally, the most probable sample value generally is the mode, which is then even lower than the median.

This phenomenon is sometimes illustrated as the infamous Poisson effect (due to the Poissonian nature of the distribution of the grains of the component of interest). A graph is then built that shows the most probable sample value (MPSV) as a function of the sample size. This very didactic graph, however, needs to be carefully interpreted. It does not represent a bias (again there is no real bias), and the most probable sample value is not the one that generally will be obtained - it is only the value with the highest frequency of occurrence, if the sampling would be repeated an infinite number of times. On this graph, the average sample result is simply the horizontal line on which the graph is centered.

The difference between the MPSV curve and this average simply illustrates the skewness of distribution mentioned above - and nothing more.

If the relative sampling variability standard deviation (RSD) reaches or exceeds 32% or above, then in the binomial/Poissonian case of fully liberated material, the distribution will become noticeably asymmetrical. In practice, it commonly accepted that that 32% limit is a good and safe criterion to apply, for liberated or non-liberated materials alike, not only to each sampling stage, but also to the overall, combined RSD in the case of a multi-stage protocol.

It is fair to say however, that there are "healthy debates" as to what constitutes a reasonable upper threshold for acceptable sampling variance (both 16% and 20% have been suggested). This evergreen debate is perhaps a good example of ..... "considering the issue case-by-case."

5. Conclusion

Based on these brief reflections, the author welcomes one or more "healthy debates", preferentially in this journal.

References

Engström, K. (2017) A comprehensive literature review reflecting fifteen years of debate regarding the representativity of reverse circulation vs blast hole drill sampling. TOS Forum, issue 7, pp.36-46 doi: 10.1255/tosf.99

Matheron, G. (2015) [translated by Bongarçon, D-F. and Pitard, F.] Comparison between samples with constant mass and samples with constant fragment population size (and calculations of their sampling variances). TOS Forum, Issue 7, pp. 231-238 doi: 10.1255/tosf.80

Pitard, F.F. (1993) Pierre Gy´s sampling theory and sampling practice, 2nd edn, CRC Press, Boca Raton.