Theoretical, practical, and economic difficulties in sampling for trace constituents

by F.F. Pitard*

Synopsis

Many industries base their decisions on the assaying of tiny analytical sub-samples. The problem is that most of the time several sampling and sub-sampling stages are required before the laboratory provides its ultimate assays using advanced chemical and physical methods of analysis. As long as each sampling and sub-sampling stage is the object of due diligence using the theory of sampling it is likely that the integrity of the sought after information has not been altered and the generated database is still capable to fulfil its informative mission. Unfortunately, more often than not, unawareness of the basic properties of heterogeneous materials combined with the unawareness of stringent requirements listed in the theory of sampling, lead to the conclusion that massive discrepancies may be observed between the expensive outcome of a long chain of sampling and analytical custody, and reality. There are no areas that are more vulnerable to such misfortune than sampling and assaying for trace amounts of constituents of interest in the environment, in high purity materials, in precious metals exploration, food chain, chemicals, and pharmaceutical products. Without the preventive suggestions of the theory of sampling serious difficulties may arise when making Gaussian approximations or even lognormal manipulations in the subsequent interpretations. A complementary understanding of Poisson processes injected in the theory of sampling may greatly help the practitioner understand structural sampling problems and prevent unfortunate mistakes from being repeated over and over until a crisis is reached. This paper presents an overview of the theoretical, practical and economic difficulties often vastly underestimated in the search for quantifying trace amounts of valuable or unwelcome components.

Scope

The heterogeneity of trace constituents in lots to be sampled for the determination of their contents has been the object of extensive work by many authors in the past. The scope of this paper is to focus attention on the works done by Gy[1–3], Ingamells[4–7] and Pitard[4, 8–11]. Links between the works of these authors are investigated, and an up-to-date strategy to resolve sampling difficulties is suggested. The challenge is to provide adequate, realistic sample and sub-sample mass at all sampling and sub-sampling stages, all the way to the balance room at the assaying laboratory. More often than not, meeting theory of sampling (TOS) basic requirements to keep the variance of the fundamental sampling error (FSE) within reasonable limits are beyond economic reach, or at least in appearance. Therefore, when these difficulties are ignored for practical reasons, awareness becomes the only tool at our disposal to show the possible consequences. Such awareness must be properly managed, which is the primary objective of this paper. For the unaware reader, TOS refers to Gy's work combined with compatible and positive contributions made by others. TOS is a dynamic knowledge that should be complemented by existing and future contributions, which is the mission of WCSB in many ways.

Definitions and notations

The length of this paper being limited, the reader is referred to textbooks for some definitions and notations (Gy[1–3]; Pitard[8]; Ingamells and Pitard[4]). Only the essential ones are listed below.

Latin letters

a	content of a constituent of interest
FSE	fundamental sampling error
GSE	grouping and segregation error
IDE	increment delimitation error
IEE	increment extraction error
IH	invariant of heterogeneity
IPE	increment preparation error
IWE	increment weighting error
M	mass or weight of a sample or lot to be sampled
r	number of low frequency isolated grains of a given constituent of interest
s	experimental estimate of a standard deviation
Y	a grouping factor
Z	a segregation factor

---

* *Francis Pitard Sampling Consultants, USA.*

© *The Southern African Institute of Mining and Metallurgy, 2010. SA ISSN 0038–223X/3.00 + 0.00. This paper was first published at the SAIMM Conference, Fourth World Conference on Sampling & Blending, 21–23 October 2009.*

Greek letters

θ average number of constituent of interest grains per sample

μ average number of constituent of interest grains per sample in a primary sampling stage when two consecutive sampling stages introduce a Poisson process

σ true unknown value of a standard deviation

γ a most probable result

Industries that should be concerned

Regardless of what the constituent of interest is in a material to be sampled, it always carries a certain amount of heterogeneity. Many industries are concerned about such a structural property. Some industries using materials of mineral origin such as metallurgy, cement, coal, glass, ceramics, uranium, and so on, are challenged every day to quantify contents of critically important elements. These difficulties reach a paroxysm when these elements are present in trace amounts. There are many other similar examples in the agricultural, food, paper, chemical, and pharmaceutical industries. There is another stunning example in sampling for trace constituents in the environment; companies struggling to meet regulatory requirements have great concerns about the capability to collect representative samples that will be assayed for trace constituents. All these examples are just the tip of the iceberg.

A logical approach suggested by the theory of sampling

The theory of sampling is by definition a preventive tool for people working in the industry to find ways to minimize the negative effects of the heterogeneity carried by critically important components. Such heterogeneity generates variability in samples, therefore variability in data that are later created. The following steps are essential for the definition of a logical and successful sampling protocol. The discussion is limited to the sampling of zero-dimensional, movable lots. For one-dimensional lots the reader is referred to Chronostatistics (Pitard[12]); for two- and three-dimensional lots the reader is referred to more in-depth reading of the TOS (Gy[1–3]; Pitard[8]; Esbensen and Minkkinen[13]; Petersen[14]; David[15]).

Mineralogical and microscopic observations

At the early stage of any sampling project it is mandatory to proceed with a thorough mineralogical study or microscopic study that may show how a given trace constituent behaves in the material to be sampled. The conclusions of such study may not be stationary in distance or time, nevertheless they give an idea about the direction that one may go when reaching the point when an experiment must be designed to measure the typical heterogeneity of the constituent of interest. These important studies must remain well focused. For example, in the gold industry it is not rare to see a mineralogical study of the gold performed for a given ore for a given mining project. Then, the final report may consist of 49 pages elaborating on the many minerals present in the ore, and only one page for gold which is by far the most relevant constituent; well-focused substance should be the essence.

Heterogeneity tests

Many versions of heterogeneity tests have been suggested by various authors. For example, Gy suggested about three versions, François-Bongarçon suggested at least two, Pitard suggested several, Visman suggested one, and Ingamells suggested several. They all have something in common: they are usually tailored to a well-focused objective and they all have their merits within that context. It is important to refer to François-Bongarçon's works[16–19] because of his well-documented approaches. It is the view of this author that for trace constituents, experiments suggested by Visman[20] and Ingamells provide the necessary information to make important decisions, about sampling protocols, the interpretation of the experimental results, and the interpretation of future data collected in similar materials; this is especially true to find methods to overcome nearly unsolvable sampling problems because of the unpopular economic impact of ideal sampling protocols.

Respecting the cardinal rules of sampling correctness

Let us be very clear on a critically important issue: if any sampling protocol or any sampling system does not obey the cardinal rules of sampling correctness listed in the Theory of Sampling, then minimized sampling errors leading to an acceptable level of uncertainty no longer exist within a reachable domain. In other words, if increment delimitation errors (IDE), increment extraction errors (IEE), increment weighting errors (IWE) and increment preparation errors (IPE) are not addressed in such a way that their mean is no longer close to zero, we slowly leave the domain of sampling and enter the domain of gambling. In this paper the assumption is made that the mean of these bias generator errors is zero. In the eventuality anyone bypasses sampling correctness for some practical reason, solutions no longer reside in the world of wisdom and generated data are simply invalid and unethical. It is rather baffling that many standards committees on sampling are still at odds with the rules of sampling correctness.

Quantifying the fundamental sampling error

Enormous amounts of work have been done by Gy, François-Bongarçon, and Pitard on the many ways to calculate the variance of the fundamental sampling error. For the record the theory of sampling offers very different approaches and formulas for the following cases:

- ➤ The old, classic parametric approach[1] where shape factor, particle size distribution factor, mineralogical factor, and liberation factor must be estimated
- ➤ A more scientific approach[3] involves the global determination of the constant factor of constitution heterogeneity (i.e., *IHL*)
- ➤ A totally different approach[22] focuses on the size, shape, and size distribution of the liberated, non-liberated, or even *in situ* grains of a certain constituent of interest
- ➤ A special case when the emphasis of sampling is on the determination of the size distribution of a material[2,22].

The careful combination of cases 3 and 4 can actually provide a very simple, practical and economical strategy that may have been overlooked by many sampling practitioners.

Minimizing the grouping and segregation error

The grouping and segregation error GSE is characterized by the following properties of its mean and variance:

$$mean\left(GSE\right) \approx 0$$

$$s_{GSE}^{2} = YZs_{FSE}^{2}$$

If the variance of GSE is the product of three factors, this would suggest that the cancellation of only one factor could eliminate GSE.

- It is not possible to cancel the variance of FSE unless the sample is the entire lot, which is not the objective of sampling. However, it should be minimized and we know how to do this.
- It is not possible to cancel Y unless we collect a sample by collecting one-fragment increments at random one at a time. This is not practical; however, it is done in recommended methods by Gy and Pitard for the experimental determination of IH_L. In a routine sampling protocol, the right strategy is to collect as many small increments as practically possible so the factor Y can be drastically minimized; this proves to be by far the most effective way to minimize the variance of GSE.
- It is not possible to cancel the factor Z which is the result of transient segregation. All homogenizing processes have their weaknesses and are often wishful thinking processes; this proves to be the most ineffective way to minimize the variance of GSE.

The challenges of reality

Reality often shows that between what is suggested by Gy's theory and what the actual implemented protocols are, there is an abysmal difference and we should understand the reasons for such unfortunate shortcoming; there could be several reasons:

- Requirements from Gy's theory are dismissed as impractical and too expensive.
- The TOS is not understood, leading to the impression the TOS does not cover some peculiar problems when it most certainly does.
- The practitioner does not know how to go around some assumptions made in some parts of the TOS when limitations of these assumptions have been well addressed and cured where necessary.
- Protocols are based on past experience from somebody else.
- Top management does not understand the link between hidden cost and sampling.
- Normal or lognormal statistics are applied within domains where they do not belong.
- Poisson processes are vastly misunderstood and ignored.
- People have a naïve definition of what an outlier is, etc.

Ingamells's work to the rescue

Clearly, we need a different approach in order to make TOS more palatable to many practitioners and this is where the work of Ingamells can greatly help. Ingamells's approach can help sampling practitioners to better understand the behaviour of bad data, so management can better be convinced that after all, Gy's preventive approach is the way to go, even if it seems expensive at first glance; in this statement there is a political and psychological subtlety that has created barriers for the TOS for many years, and breaking this barrier was the entire essence of Pitard's thesis[22].

From Visman to Ingamells

Most of the valuable work of Ingamells is based on Visman's sampling theory. It is not the intention of this paper to inject Visman's work in the TOS. What is most relevant is Ingamells's work on Poisson distributions that can be used as a convenient tool to show the risks involved when the variance of FSE is going out of control: it cannot be emphasized strongly enough that the invasion of any database by Poisson processes can truly have catastrophic economic consequences in any project such as exploration, feasibility, processing, environmental, and legal assessments. Again, let us make it very clear, any database invaded by a Poisson process because of the sampling and sub-sampling procedures that were used is a direct, flagrant departure from the due diligence practices in any project. Yet, sometimes we do not have the luxury of a choice, such as the sampling of diamonds; then awareness is the essence.

Limitations of normal and lognormal statistical models

At one time, scientists became convinced that the Gaussian distribution was universally applicable, and an overwhelming majority of applications of statistical theory are based on this distribution.

A common error has been to reject 'outliers' that cannot be made to fit the Gaussian model or some modification of it as the popular lognormal model. The tendency, used by some geostatisticians, has been to make the data fit a preconceived model instead of searching for a model that fits the data. On this issue, a Whittle quote[21] later on used and modified by Michel David[15], was superb: 'there are no mathematical models that can claim a divine right to represent a variogram.'

It is now apparent that outliers are often the most important data points in a given data-set, and a good understanding of Poisson processes is a convenient way of understanding how and why they are created.

Poisson's processes

Poisson and double Poisson processes[6–8,22] explain why highly skewed distribution of assay values can occur. The grade and location of an individual point assay, which follows a single or double-Poisson distribution, will have virtually no relationship, and it will be impossible to assign a grade other than the mean value to mineable small-size blocks. Similar difficulties can occur with the assessment of impurity contents in valuable commodities. Now, there is a little subtlety about Poisson processes, as someone may say the position of a grain or cluster of grains is never completely random as explained in geostatistics; this is not the point. The point is that, more often than not, the volume of observation we use may itself generate the Poisson process; there is a difference.

The single Poisson process

The Poisson model is a limit case of the binomial model where the proportion p of the constituent of interest is very small (e.g., fraction of 1%, ppm or ppb), while the proportion $q = 1–p$ of the material surrounding the constituent of interest is practically 1. Experience shows that such constituent may occur as rare, tiny grains, relatively pure at times, and they may or may not be liberated; they may even be *in situ*. Sampling practitioners must exit the paradigm of looking at liberated grains exclusively; the problem is much wider than that. As the sample becomes too small, the probability of having one grain or a sufficient amount of them in one selected sample diminishes drastically. For *in situ* material the sample can be replaced by an imaginary volume of observation at any given place. When one grain or one cluster is present, the estimator a_S of a_L becomes so high that it is often considered as an outlier by the inexperienced practitioner whereas it is the most important finding that should indeed raise attention. All this is a well-known problem for those involved with the sampling of diamonds.

Let us call $P(x = r)$ the random probability x of r low-frequency isolated coarse grains appearing in a sample, and θ is the average number of these grains per sample; see derivation of the following formula in appropriate literature[8,22].

$$P(x=r)=\frac{\theta^n}{r!}e^{-\theta} \qquad [1]$$

with $r = 0, 1, 2, 3,\ldots$

If m is the number of trials (i.e., selected, replicate samples), the variance of the Poisson distribution is $\theta = mpq \approx mp$ since q is close to 1. For all practical purposes, the mean value of the Poisson distribution is $\theta \approx mp$. As clearly shown in the derivation of Equation [1] we could assume in a first order approximation that $\theta \approx mp$.

The double Poisson process

When primary samples taken from the deposit contain the constituent of interest in a limited (e.g., less than 6)[4,15] average number μ of discrete grains or clusters of such grains (i.e., $P[y=n]$), and they are sub-sampled in such a way that the sub-samples also contain discrete grains of reduced size in a limited (e.g., less than 6)[4,15] average number θ (i.e., $P[x=r]$), a double Poisson distribution of the assay values is likely.

The probability P of r grains of mineral appearing in any sub-sample is determined by the sum of the probabilities of r grains being generated from samples with n grains.

Let us define the ratio f:

$$f=\frac{\theta}{\mu} \qquad [2]$$

With $\theta = \mu \cdot f$ or $\theta = n \cdot f$ for each possibility, the equation for the resulting, compounded probability of the double Poisson distribution is:

$$P(x=r)=\sum P(y=n)\cdot P(x=r)= \sum\frac{\mu^n e^{-\mu}}{n!}\cdot\frac{(nf)^r e^{-nf}}{r!}=\frac{f^r e^{-\mu}}{r!}\sum_{n=0}^{n=\infty}\frac{\mu^n e^{-nf}n^r}{n!} \qquad [3]$$

for $r = 0, 1, 2, 3,\ldots$

An example of this case is given in Pitard's thesis[22].

This is the probability of obtaining a sample with r grains of the constituent of interest. The equation could be modified using improved Stirling approximations for factorials, for example:

$$n!=\left[\frac{n}{e}\right]^n\sqrt{\frac{\pi(6n+1)}{3}} \qquad [4]$$

In practice, one does not usually count grains; concentrations are measured. The conversion factor from number of grains to, per cent X for example, is C, the contribution of a single average grain. Since the variance of a single Poisson distribution is equal to the mean:

$$s=C\sqrt{\mu} \qquad [5]$$

Therefore:

$$s^2=C^2\mu \qquad [6]$$

But variances of random variables are additive, then for a double Poisson distribution we would have:

$$s^2=C^2\mu+c^2\theta \qquad [7]$$

The data available are usually assays in % metal, gram/ton, ppm or ppb. They are related by the equation:

$$x_i=a_H+cr_i \qquad [8]$$

where x_i is the assay value of a particular sample, in % for example; a_H is the low more homogeneous background concentration in % for example, which is easier to sample; r_i is the number of mineral grains in the sample; c is the contribution of one grain to the assay in % for example:

$$c=\left[\frac{\text{weight of grain}}{\text{weight of sample}}\right]100\% \qquad [9]$$

Thus the probability of a sample having an assay value of x_i equals the probability of the sample having r_i grains when a_H is relatively constant.

The mean value of a set of assays can be shown to be:

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i=a_H+c\theta \qquad [10]$$

For a single Poisson distribution this equation would be:

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i=a_H+C\mu \qquad [11]$$

where $\bar{x}$ is an estimator of the unknown average content a_L of the constituent of interest. Assuming sampling is correct, and for the sake of simplicity, in the following part of this paper we should substitute $\bar{x}$ with a_L. Then:

$$a_L=a_H+C\mu=a_H+c\theta \qquad [12]$$

then:

$$a_L-a_H=C\mu=c\theta \qquad [13]$$

$$\frac{\theta}{\mu}=\frac{C}{c}=f \qquad [14]$$

Substituting Equation [14] in Equation [7]:

$$s^2=C(a_L-a_H)+c(a_L-a_H)= fc(a_L-a_H)+c(a_L-a_H) \qquad [15]$$

whence:

$$f = \frac{s^2 - c(a_L - a_H)}{c(a_L - a_H)} \qquad [16]$$

The probability that there will be no difficult-to-sample grains of the constituent of interest in a randomly taken sub-sample is found by substituting $r = 0$ in Equation [3]:

$$P(x = 0) = e^{-\mu} \sum \frac{\mu^n e^{-nf}}{n!} = e^{\mu\left(e^{-f} - 1\right)} \qquad [17]$$

If a data-set fits a double Poisson distribution, the parameters μ and θ of this distribution may be found from a reiterative process, as follows:

Make a preliminary low estimate of a_H. Give c an arbitrary low value. Calculate a preliminary value for f from Equation [16], and for μ by rearranging Equation [12]:

$$\mu = \frac{(a_L - a_H)}{C} = \frac{(a_L - a_H)}{fc} \qquad [18]$$

Substitute these preliminary estimates in Equation [17]; averaging the lowest $P(x=0)$ of the data to obtain a new estimate of a_H. Increment c and repeat until a best fit is found. If a Poisson process is involved, which is not necessarily the case, this incremental process for c will indeed converge very well.

Notion of minimum sample weight

There is a necessary minimum sample mass M_{Smin} in order to include at least one particle of the constituent of interest about 50% of the time in the collected sample which happens when $r = 1$ in Equation [1] or when $n = 1$ in Equation [3]; Ingamells shows that it can be calculated as follows:

$$M_{S\min} = \frac{IH_L a_L^2}{\left[a_L - a_H\right]^2} \qquad [19]$$

For replicate samples to provide a normally distributed population the recommended sample mass M_{Srec} should be at least 6 times larger than M_{Smin}. As shown by Ingamells and Pitard[4] and David[15] it takes at least r or n=6 to minimize the Poisson process to the point that a more normally distributed data will appear. Of course there is no magical number and r or n should actually be much larger than 6 to bring back the variance of the fundamental sampling error (FSE) into an acceptable level[22]. At this point there is an important issue to address: all equations suggested by Gy to estimate the appropriate sample mass when the heterogeneity IH_L carried by the constituent of interest is roughly estimated, should be used in such a way that we know we are reasonably within a domain that does not carry any Poisson skewness. Ingamells suggested that the needed sample mass M_{Srec} is about 6 times larger than M_{Smin}. The recommended limit suggested in Gy's early work is a $\%s_{FSE} = \pm 16\%$ relative which happens to be even more stringent than the Ingamells's suggestion.

Notion of optimum sample weight

In a logical sampling protocol a compromise must be found between the necessary sample mass required for minimizing the variance of the FSE and the number of samples that are needed to have an idea about the lot variability due to either small-scale or large-scale segregation. Such optimum sample mass M_{Sopt} was found by Ingamells and translated in the appropriate TOS notations in Pitard's thesis[22], and can be written as follows:

$$M_{Sopt} = \frac{IH_L}{s_{SE}^2} \qquad [20]$$

Where s_{se}^2 is a local variance due to the segregation of the constituent of interest in the lot to be sampled.

Case study: estimation of the iron content in high-purity ammonium paratungstate

The following case study involves a single stage Poisson process and the economic consequences can already be staggering because of the non-representative assessment of the impurity content of an extremely valuable high purity material. It should be emphasized that the analytical protocol that was used was categorized as fast, cheap, and convenient. In other words, it was called a cost-effective analytical method.

A shipment of valuable high-purity ammonium paratungstate used in the fabrication of tungsten coils in light bulbs was assayed by an unspecified supplier to contain about 10 ppm iron. The contractual limit was that no shipment should contain more than 15 ppm iron. The client's estimates using large assay samples were much higher than the supplier's estimates using tiny 1-gram assay samples. The maximum particle size of the product was 150-μm. To resolve the dispute a carefully prepared 5000-gram sample, representative of the shipment, was assayed 80 times using the standard 1-gram assay sample weight used at the supplier's laboratory. Table I shows all the assay values generated for this experiment.

A summary of results is as follows:

- The estimated average $x \approx a_L$ of the 80 assays was 21 ppm.
- The absolute variance $s^2 = 378$ ppm^2
- The relative, dimensionless variance $s_R^2 = 0.86$
- The absolute standard deviation $s = 19$ ppm
- The relative, dimensionless standard deviation $s_R = \pm 0.93$ or $\pm 93\%$.

From the TOS the following relationship can be written:

$$\frac{s_1^2}{a_L^2} = s_{FSE1}^2 + s_{GSE1}^2 = \left[\frac{1}{M_{S1}} - \frac{1}{M_L}\right] IH_L + s_{FSE1}^2 ZY \qquad [21]$$

All terms are well defined in the TOS. The subscript 1 refers to the information that is available from a small sample weighing 1 gram; it is in that case only a reference relative to the described experiment. The effect of M_L is negligible since it is very large relative to M_S.

The value of the variance s_{GSE1}^2 of the grouping and segregation error is not known; however, the material is well calibrated and there are no reasons for a lot of segregation to take place because the isolated grains containing high iron content have about the same density as the other grains since their composition is mainly ammonium paratungstate. Therefore it can be assumed in this particular case that $s_{FSE1}^2 \geq s_{GSE}^2$ if each 1-gram sample is made of several random increments, so the value of IH_L that is calculated is only slightly pessimistic. The nearly perfect fit to a Poisson model as shown in Figure 1 was at the time sufficient proof that the grouping and segregation error was not the problem, and was further confirmed latter on by the good reproducibility obtained by collecting much larger 34-gram samples. The following equation can therefore be written:

Table I

Summary of 80 replicate iron assays in high-purity ammonium paratungstate

Sample number	ppm Fe	Sample number	ppm Fe	Sample number	ppm Fe	Sample number	ppm Fe
1	**4**	21	**44**	41	**5**	61	**28**
2	**20**	22	**21**	42	**31**	62	**4**
3	**21**	23	**21**	43	**19**	63	**21**
4	**31**	24	**18**	44	**6**	64	**29**
5	**16**	25	**21**	45	**18**	65	**20**
6	**16**	26	**4**	46	**18**	66	**35**
7	**14**	27	**17**	47	**4**	67	**19**
8	**12**	28	**32**	48	**4**	68	**48**
9	**4**	29	**7**	49	**5**	69	**4**
10	**9**	30	**18**	50	**4**	70	**14**
11	**36**	31	**20**	51	**19**	71	**8**
12	**32**	32	**21**	52	**6**	72	**6**
13	**31**	33	**4**	53	**44**	73	**115**
14	**4**	34	**19**	54	**74**	74	**4**
15	**22**	35	**32**	55	**16**	75	**9**
16	**4**	36	**4**	56	**4**	76	**13**
17	**4**	37	**64**	57	**33**	77	**26**
18	**19**	38	**7**	58	**4**	78	**32**
19	**48**	39	**48**	59	**34**	79	**4**
20	**68**	40	**18**	60	**64**	80	**12**

$$0.86 - s^2_{GSE1} \approx IH_L \qquad [22]$$

Therefore, it can be assumed that $IH_L \leq 0.86$ g, since the mass $M_S = 1$ gram, justifying the way Equation [22] is written. If the tolerated standard deviation of the FSE is ±16% relative, the optimum necessary sample mass M_S can be calculated as follows:

$$M_S = \frac{0.86}{(0.16)^2} = 34\,\text{g} \qquad [23]$$

Obviously, it is a long way from the 1-gram that was used for practical reasons. This mass of 34 grams is the minimum sample mass that will make the generation of normally distributed assays results possible. Another parameter that can be obtained is the low background content a_H, which is likely around 4 ppm by looking at the histogram in Figure 1. This high-frequency low value may sometimes represent only the lowest possible detection of the analytical method; therefore caution is recommended when the true low background content of a product for a given impurity is calculated.

Investigation of the histogram

Figure 1 illustrates the histogram of 80 assays shown in Table I. In this histogram it is clear that the frequency of a given result reaches a maximum at regular intervals, suggesting that the classification of the data in various zones is possible; zone A with 27 samples showing zero grain of the iron impurity; zone B with 29 samples showing 1 grain; zone C with 13 samples showing 2 grains; zone D with 5 samples showing 3 grains; zone E with 3 samples showing 4 grains; zone F with 1 sample showing 5 grains; Zone G with 6 grains shows no event; finally zone H with 7 grains shows one event, which may be an anomaly in the model of the distribution. The set of results appears Poisson distributed, and a characteristic of the Poisson distribution is that the variance is equal to the mean. The following equivalences can be written:

$$s^2 = a_L \Leftrightarrow 21\,\text{ppm} \Leftrightarrow 1.18\ \text{grain/sample} \Leftrightarrow \theta \qquad [24]$$

The assumption that $a_H = 4$ ppm needs to be checked. The probability that the lowest assay value represents a_H can be calculated. If the average number of grains of impurity per sample θ is small, there is a probability that the lowest assays represent a_H. The probability that a single collected sample will have zero grain is:

$$P(x=0) = \frac{\theta^0 e^{-\theta}}{0!} = e^{-\theta} = e^{-1.18} \qquad [25]$$

If we call $P(x = 0)$ the probability for a success, then the probability P_x of n successes in N trials is given by the binomial model:

$$P_x = \frac{N!}{n!(N-n)!} P^n (1-P)^{N-n} \qquad [26]$$

Where P is the probability of having a sample with no grain when only one sample is selected, and (1-P) is the

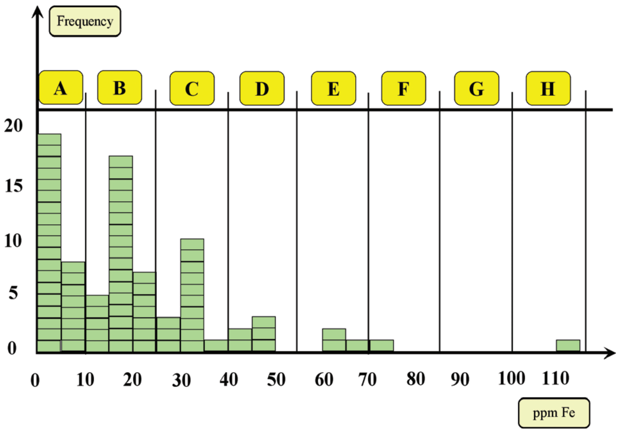


Figure 1—Histogram of eighty 1-gram assays for iron in ammonium paratungstate

probability of having at least one grain when only one sample is collected; then the probability of no success $P(x \neq 0)$ with N samples is:

$$P(x \neq 0) = \frac{N!}{0!(N-0)!}\left(e^{-\theta}\right)^{0}\left(1-e^{-\theta}\right)^{N-0} = \left(1-e^{-\theta}\right)^{N} \quad [27]$$

Equation [27] shows the probability that none of N samples is free from low-frequency impurity grains. The probability that the lowest assay value represents a_H is:

$$P(x = 0) = 1 - \left(1-e^{-\theta}\right)^{N} = 0.9999 \quad [28]$$

Assuming that a_H is not the analytical detection limit, we can be sure that the lowest assay represents a_H. Having found that the value $\theta = 1.18$, we may calculate the Poisson probabilities for samples located in each zone illustrated in Figure 1. Thus, by multiplying each probability by 80, we may compare the calculated distribution with the observed distribution. Results are summarized in Table II.

The observed distribution is very close to the calculated distribution if we exclude the very high result showing 115 ppm, which should not have appeared with only 80 samples. A characteristic of the Poisson distribution is that the variance s^2 of the assays is equal to the average a_L.

$$s^2 = \bar{x} \approx a_L = \theta \quad [29]$$

or

$$s = \sqrt{\theta} \quad [30]$$

But, in practice the number of grains is not used; instead concentrations are used such as %, g/t, ppm, or ppb. Let us call C the conversion factor and rewrite Equation [30] properly:

$$s = C\sqrt{\theta} \quad [31]$$

Thus, we may calculate the contribution C of a single average impurity grain to a single iron assay:

$$C = \frac{19}{\sqrt{1.18}} = 17 \text{ ppm} \quad [32]$$

Discussion of acceptable maximum for the standard deviation of the FSE

Ingamells suggested that a minimum of six of the largest grains, or clusters of tinier grains into a single fragment, of impurity should be present in a sample for the analysis of this sample to be meaningful. The objective of such statement is to eliminate the Poisson process from damaging the database. If a 1-gram sample contains an average $\theta = 1.18$ grains, then the minimum recommended sample mass is around 5 grams. Using this mass and the value of IHL obtained earlier we may write:

$$s_{SFE}^2 \leq \frac{0.86}{5} = 0.17 \quad [33]$$

$$s_{SFE} \leq 0.41 \quad \text{or} \quad \pm 41\% \text{ relative} \quad [34]$$

But, following Gy's recommendations[23] a 34-gram sample is recommended in order to achieve a ±16% relative standard deviation, which would contain about 41 grains; it is understood that for certain applications, such as sampling for material balance or for commercial settlements, Gy's recommendations in his publications were far more stringent (5% or even 1%). In order to further discuss this difference, let us construct the useful Ingamells's sampling diagram.

With the set of data given in Table I a set of artificial, large 10-gram samples made of Q=10 small one-gram samples can be created, and they are shown in Table III.

Visman sampling equation

With this information Visman sampling constants A and B can be calculated; it is understood that Visman would have suggested the collection of larger samples as well explained by Pitard in his thesis[22]:

$$S^2 = \frac{s^2}{N} = \frac{A}{M_S} + \frac{B}{N} \quad [35]$$

where S is the uncertainty in the average of N=80 assays on samples of individual mass

$$M_S^* = \frac{M_S}{N} = 1 \text{ gram}$$

A is the Visman homogeneity constant. It is the Gy's Constant Factor of Constitution Heterogeneity IH_L multiplied by the square of the average content of the lot.

From the variances and Visman's equation we obtain:

$$A = IH_L \cdot a_L^2 = \frac{\left(s_1^2 - s_2^2\right)M_{S1}}{a_L^2\left(1 - \dfrac{1}{Q}\right)} = 379773 \text{ mg/ppm}^2$$

From Gy we suggested earlier:

$$A = IH_L a_L^2 = 0.86(21.475)^2 = 396611 \text{ mg/ppm}^2$$

Those numbers are very close and within the variances precision, therefore this would suggest there is no room to calculate the amount of segregation for iron in the lot. It is wise to assume that B, the Visman segregation constant is:

Table II

Comparison of the calculated distribution with the observed distribution

r	Poisson probability for θ = 1.18	Calculated distribution	Observed distribution
0	0.307	25	27
1	0.363	29	29
2	0.213	17	14
3	0.084	7	5
4	0.025	2	3
5	0.006	0	1
6	0.001	0	0
7	0.0002	0	1
Total	**0.999**	**80**	**80**

Table III

Iron content of artificial large samples of mass equal to 10 grams

N sample number	Composited small samples	Iron content in large samples
1	1–10	**15**
2	11–20	**27**
3	21–30	**20**
4	31–40	**24**
5	41–50	**11**
6	51–60	**30**
7	61–70	**22**
8	71–80	**23**

$$B = s^2 - \frac{A}{M_S^*} = s_{GSE}^2 a_L^2 N = s_{SE}^2 a_L^2 \approx 0$$

This confirms the opinion that iron in calibrated ammonium paratungstate grains has no reason whatsoever to segregate in a substantial way, and all the observed variability is due to the variance of FSE, as suggested by a nearly perfect Poisson's fit illustrated in Figure 1 showing the problem was a sample mass too small by more than one order of magnitude.

The most probable result

The most probable result γ for the assaying of iron as a function of analytical sample mass M_S is calculated with the following Ingamells Equation[1–3, 22]:

$$\gamma = \frac{2a_L(a_L - a_H)^2 M_S + a_H IH_L a_L^2}{2(a_L - a_H)^2 M_S + IH_L a_L^2} \qquad [36]$$

Values of γ are illustrated in Figure 3, and it basically represents the location of the mode of the probability distribution relative to the expected arithmetic average a_L. Notations in Figure 3 are old Ingamells notations and the author apologizes for this inconvenience due to the use of an old software program (i.e., $\gamma = Y$, a_L = x, $L = a_H$).

A careful study of the γ curve in Figure 3 (i.e., light blue curve) is the key to complete our discussion of the difference between Ingamells's recommendation and Gy's recommendation for a suggested maximum value for the standard deviation of FSE. It can be observed that the recommended mass by Ingamells (i.e., 6 grains in the sample or a $\%s_{SFE}$ = ±41%) leads to a location of the mode still substantially away from the expected arithmetic average a_L. It is not the case with the necessary sample mass of 34 grams in order to obtain a $\%s_{SFE}$ = ±16% as recommended by Gy.

Ingamells's gangue concentration

The low background iron content a_H estimated earlier to be 4 ppm can be calculated using the mode from results from small samples and the mode γ_2 from results from large samples. Modes can be calculated using the harmonic means h_1 and h_2 of the data distribution of the small and large samples. The harmonic mean is calculated as follows:

$$h_i = \frac{N}{\sum_i \frac{1}{x_i}} \qquad [37]$$

where N is the number of samples.

From Equation [36] we may write:

$$h_1 = \gamma_1 = \frac{2a_L(a_L - a_H)^2 M_{S1} + a_H IH_L a_L^2}{2(a_L - a_H)^2 M_{S1} + IH_L a_L^2} \qquad [38]$$

$$h_2 = \gamma_2 = \frac{2a_L(a_L - a_H)^2 M_{S2} + a_H IH_L a_L^2}{2(a_L - a_H)^2 M_{S2} + IH_L a_L^2} \qquad [39]$$

Then from Equations [1], [2], and [3] the low background content a_H can be calculated as follows:

$$a_H = \frac{\gamma_1 M_{S2}(a_L - \gamma_2) - \gamma_2 M_{S1}(a_L - \gamma_1)}{M_{S2}(a_L - \gamma_2) - M_{S1}(a_L - \gamma_1)} \qquad [40]$$

Results are shown in Figure 2 and confirm the earlier estimate of 4 ppm.

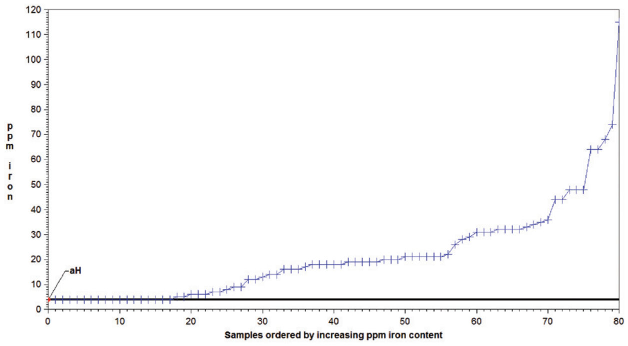
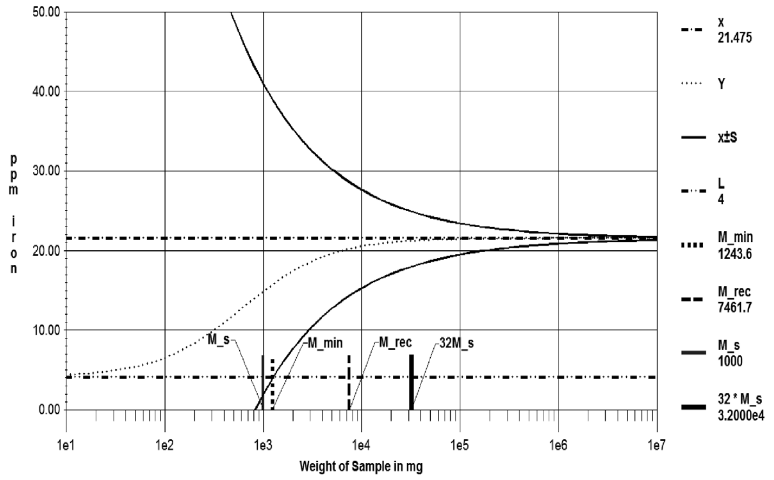


Figure 2—Search for a value of the low background iron content a_H



```
Field variability for very large samples = 2.6064
N_samples = 32 and M_s = 1000 mg
provides overall precision S(3.2000e4) = 3.436 ppm and S_VR = 16.041%
M_s1 = 1000  S_1 = 19.444  S_1R = 90.543%
```

Figure 3—Illustration of the Ingamells's sampling diagram for iron traces in pure ammonium paratungstate (γ=Y, a_L = x, L= a_H)

If the reader is interested by the full derivation of the above formulas, refer to Pitard's doctoral thesis[22].

Conclusions

The key to sampling for trace amounts of a given constituent of interest is a thorough microscopic investigation of the ways such constituent is distributed on a small scale in the material to be sampled. Liberated or not, the coarsest grains of such constituent must be measured and placed into the context of their average, local expected grade. The use of Poisson statistics and the calculation of an Ingamells' sampling diagram can lead to defining a minimum sampling effort during the early phase of a project. Equipped with such valuable preliminary information, someone can proceed with a feasibility study to implement a necessary sampling protocol as suggested by the TOS, with a full understanding of what the consequences would be if no due diligence is exercised.

Recommendations

As the sampling requirements necessary to minimize the variance of FSE as suggested in the TOS may become cumbersome to many economists, it becomes important to proceed with preliminary tests in order to provide the necessary information to create valuable risk assessments. The following steps, in chronological order, are recommended:

- ➤ Carefully define data quality objectives.
- ➤ Always respect the fundamental rules of sampling correctness as explained in the TOS. This step is not negotiable.
- ➤ Perform a thorough microscopic investigation of the material to be sampled in order to quantify the grains size of the constituent of interest, liberated or not. Emphasize the size of clusters of such grains if such a thing can be observed.
- ➤ Proceed with a Visman's experiment, calculate the Ingamells's parameters, and draw an informative Ingamells's sampling diagram.
- ➤ Show the information to executive managers who must make a feasibility study to justify more funds to perform a wiser and necessary approach using Gy's requirement to minimize the variance of FSE.

References

1. Gy, P.M. *L'Echantillonage des Minerais en Vrac* (Sampling of particulate materials). Volume 1. Revue de l'Industrie Minerale, St. Etienne, France. Numero Special (Special issue, January 15, 1967).
2. Gy, P.M. *Sampling of Particulate Materials, Theory and Practice*. Elsevier Scientific Publishing Company. Developments in Geomathematics 4. 1979 and 1983.
3. Gy, P.M. *Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling and Homogenizing*. Amsterdam, Elsevier. 1992.
4. Ingamells, C.O, and Pitard F. *Applied Geochemical Analysis*. Wiley Interscience Division, John Wiley and Sons, Inc., New York, 1986. 733 pp.
5. Ingamells, C.O. and Switzer, P. A Proposed Sampling Constant for Use in Geochemical Analysis, *Talanta*, vol. 20, 1973. pp. 547–568.
6. Ingamells, C.O. Control of geochemical error through sampling and subsampling diagrams. *Geochim. Cosmochim. Acta,* vol. 38, 1974. pp. 1225–1237.
7. Ingamells, C.O. Evaluation of skewed exploration data—the nugget effect. *Geochim. Cosmochim. Acta,* vol. 45, 1981. pp. 1209–1216.
8. Pitard, F. *Pierre Gy's Sampling Theory and Sampling Practice*. CRC Press, Inc., 2000 Corporate Blvd., N.W. Boca Raton, Florida 33431. Second edition, July 1993.
9. Pitard, F. Effects of Residual Variances on the Estimation of the Variance of the Fundamental Error', Published in a special issue by Chemometrics and Intelligent Laboratory Systems, Elsevier: 50 years of Pierre Gy's Theory of Sampling, *Proceedings: First World Conference on Sampling and Blending* -(WCSB1), 2003. pp. 149–164.
10. Pitard, F. Blasthole sampling for grade control: the many problems and solutions. *Sampling* 2008. 27–28 May 2008, Perth, Australia. The Australasian Institute of Mining and Metallurgy. 2008.
11. Pitard, F. Sampling Correctness—A Comprehensive Guideline. *Second World Conference on Sampling and Blending 2005*. The Australian Institute of Mining and Metallurgy. Publication Series No 4/2005. 2005.
12. Pitard, F. Chronostatistics—A powerful, pragmatic, new science for metallurgists. *Metallurgical Plant Design and Operating strategies (MetPlant 2006)* 18–19 September 2006 Perth, WA. The Australasian Institute of Mining and Metallurgy. 2006.
13. Esbensen, K.H. and Minkkinen, P. (guest eds.). Chemometrics and Intelligent Laboratory Systems. volume 74, 1 2004. Special issue: 50 years of Pierre Gy's Theory of Sampling. *Proceedings: First World Conference on Sampling and Blending* (WCSB1). Tutorials on Sampling: Theory and Practice.
14. Petersen, L. Pierre Gy's Theory of Sampling (TOS)—in Practice: Laboratory and Industrial Didactics including a first foray into image analytical sampling. ACABS Research Group. Aalborg University Esbjerg, Niels Bohrs Vej 8, Dk-67000 Esbjerg, Denmark. Thesis submitted to the International Doctoral School of Science and Technology. January 2005.
15. David, M. *Geostatistical Ore Reserve Estimation. Developments in Geomathematics 2*. Elsevier Scientific Publishing Company. 1977.
16. François-Bongarçon, D. Theory of sampling and geostatistics: an intimate link. Published in a special issue by Chemometrics and Intelligent Laboratory Systems, Elsevier: 50 years of Pierre Gy's Theory of Sampling, *Proceedings: First World Conference on Sampling and Blending* (WCSB1), 2003. pp. 143–148.
17. François-Bongarçon, D. The modeling of the Liberation Factor and its Calibration. *Proceedings: Second World Conference on Sampling and Blending* (WCSB2). Published by The AusIMM. 10-12 May 2005, Sunshine Coast, Quennsland, Australia. 2005. pp. 11–13.
18. François-Bongarçon, D. The most common error in applying Gy's formula in the Theory of Mineral Sampling, and the history of the liberation factor. *Monograph 'Toward 2000'*, AusImm. 2000.
19. François-Bongarçon, D and Gy, P. The most common error in applying 'Gy's Formula' in the theory of mineral sampling and the history of the Liberation factor, in Mineral Resource and Ore Reserve Estimation. *The AusIMM Guide to Good Practice*. The Australasian Institute of Mining and Metallurgy: Melbourne. 2001. pp. 67–72
20. Visman, J. A General Theory of Sampling, Discussion 3, *Journal of Materials, JMLSA*, vol. 7, no. 3, September 1972, pp. 345–350.
21. Whittle, P. On stationary processes in the plane. *Biometrika*, vol. 41. 1954. pp. 431–449.
22. Pitard, F. Pierre Gy's Theory of Sampling and C.O. Ingamells' Poisson Process Approach— Pathways to Representative Sampling and Appropriate Industrial Standards. Doctoral thesis: Aalborg University, campus Esbjerg, 2009. ISBN: 978-87- 7606-032-9.
23. Gy, P.M. *Nomogramme d'echantillonnage*, Sampling Nomogram, Probeanhme Nomogramm. Minerais et Metaux, Paris (Ed.). 1956.

◆