The accuracy of estimation from samples of ore in bulk

J. H. VENTER*

SYNOPSIS

This paper contributes towards the solution of the problem of characterizing the accuracy of estimation of ore quality based on a random sample from the ore bulk.

SINOPSIS

Hierdie artikel lewer 'n bydrae tot die oplossing van die probleem van karakterisering van die noukeurigheid van skatting van ertsgehalte gebaseer op 'n steekproef uit die ertsmassa.

INTRODUCTION

Consider a population of m ore particles (m will usually be very large). In principle the particles can be indexed in some way; denote by w_i the mass of the i-th particle. Also, let $t_w=\Sigma w_i, \mu_w=t_w/\mathrm{m}$ and $\sigma_w^2=\Sigma(w_i-\mu_w)^2/\mathrm{m}$ denote the total mass of all particles, the average mass per particles and the variance of mass per particle respectively.

Suppose a certain substance R is of interest and the i-th particle contains mass q_i of R. Let $t_q=\Sigma q_i$ denote the total mass of this substance in the population and let $\mu_q=t_q/\mathrm{m}$ and $\sigma_q^2=\Sigma(q_i-\mu_q)^2/\mathrm{m}$ denote the average mass of substance R per particle and the variance of R-mass respectively. Further denote by $\sigma_{qw}=\Sigma(q_i-\mu_q)(w_i-\mu_w)/\mathrm{m}$ the covariance and by $\rho=\sigma_{qw}/\sigma_q\sigma_w$ the correlation coefficient between mass and R-mass. Finally let π be the proportion by mass of the population consisting of substance R, i.e.

$$\hat{\pi}=t_q/t_w=\mu_q/\mu_w. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

This is the parameter of main interest.

Now suppose that n subsamples (increments), each of mass W, are taken from the ore population, making up a total sample of mass nW; suppose also the mass Q of substance R in the total sample is determined; then a reasonable estimate of π is

$$\hat{\pi}=Q/nW. \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

What characteristics of the population influence the precision of this estimate? This is the question considered in this paper.

A standard method for obtaining confidence bounds for π is this:[1] suppose that the R-masses $Q_1,\ldots,Q_n$ of the individual subsamples are determined; then estimates $\hat{\pi}_1,\ldots,\hat{\pi}_n$ of π are available; if these are approximately and independently $N(\pi,\tau^2)$-distributed, then since $\hat{\pi}=\Sigma\hat{\pi}_i/n$, approximate $(1-\alpha)$-confidence bounds for π are $\hat{\pi}\pm t_{\alpha,n-1}s/n^{1/2}$ where $t_{\alpha,n-1}$ is the $(1-\alpha/2)$-th percentile of the t_{n-1}-distribution and $s^2=\Sigma(\hat{\pi}_i-\hat{\pi})^2/(n-1)$ is the standard estimate of τ^2.

This procedure is not applicable if $n=1$; moreover, whatever the value of n, it provides no information as to in what way the characteristics of the population influence τ^2. Our result contributes towards solving this problem.

*Potchefstroom University

MAIN RESULT

In this section we assume $n=1$ so that a sample of mass W is taken; we also make the following:

Assumption A. The sample is taken in such a way that it may effectively be thought of as having been drawn particle by particle, completely at random, with replacement, until the desired mass W is reached.

It will be shown later that under this assumption the distribution of $\hat{\pi}$ approaches the $N(\pi,\tau^2)$-distribution as W increases, where

$$\tau^2=\sigma^2/W\mu_w$$

$$\sigma^2=\sigma_q^2+2\pi\rho\sigma_q\sigma_w+\pi^2\sigma_w^2. \quad \ldots\ldots\ldots\ldots \quad (3)$$

Let us remark that the actual mass of the sample may overshoot the target W slightly; this will be of no practical consequence if particle masses are small relative to W, but for purposes of the discussion to follow let us denote by $\hat{\pi}$ the estimate obtained by dividing in (2) not by W but by the actual sample mass obtained.

DISCUSSION

In extreme situations the result obtained reduces to what one would expect. For example, if all particles had exactly the same proportion of substance R then one would expect $\bar{\pi}$ to be exactly correct, i.e. $\bar{\pi}=\pi$. Indeed, we then have $q_i=\pi w_i$ for all i; hence $\sigma_q^2=\pi^2\sigma_w^2, \sigma_{wq}=\pi\sigma_w^2$, $\rho=1$ and hence $\tau^2=0$. Again, if all particles had the same mass, i.e. $w_i=\mu_w$ for all i, then $\sigma_w^2=0$ and $\sigma_{qw}=0$ so that $\tau^2=\sigma_q^2/W\mu_w$; in this case N is non-random and for all practical purposes equal to W/μ_w; also π (or $\bar{\pi}$) equals $\Sigma_1^N q^i/N\mu_w$; hence from standard sampling theory we expect $\tau^2=\sigma_q^2/N\mu_w^2=\sigma_q^2/W\mu_w$ as found above.

In order to form some idea of how large W should be before the asymptotic results become applicable a small Monte Carlo experiment was done. It was assumed that particle masses effectively followed an exponential distribution with parameter 1 (so $\mu_w=1, \sigma_w^2=1$); for each

	Target Sample Mass W					
	10		20		50	
	$\hat{\pi}$	$\bar{\pi}$	$\hat{\pi}$	$\bar{\pi}$	$\hat{\pi}$	$\bar{\pi}$
π	.5		.5		.5	
Average estimate	.545	.497	.529	.503	.510	.499
τ^2	.0167		.0083		.0033	
Average variance	.0212	.0153	.0090	.0077	.0032	.0030
Average No. particles	10.84		21.04		50.97	

particle drawn the proportion of substance R was assumed to be uniformly distributed over [0, 1] independent of the particle mass; from this it can readily be calculated that $\pi=.5$, $\sigma_q^2=5/12, \sigma_{qw}=.5$ and hence $\tau^2=1/6W$. The following table shows the average values of $\hat{\pi}$ and $\bar{\pi}$ and their average variances together with the average numbers of particles for various values of W based on 500 repetitions of the Monte Carlo simulation.

It is clear that there is (at least in this case) good agreement between the theoretical values of τ^2 and the Monte Carlo values even for small W, especially for $\bar{\pi}$.

In practice samples from materials in bulk are never drawn particle by particle; at first sight this seems to imply that assumption A is unrealistic as far as applications are concerned. However, the following argument shows that this is not necessarily the case. Suppose the sample is small compared to the bulk of the ore mass. If the sample is selected by scooping a suitable container full from the population then the population must have been mixed thoroughly beforehand if a representative sample is to be obtained; we can imagine that this mixing process has deposited a random selection of particles in any given neighbourhood within the population; if this is true then the sample obtained by taking all particles at once in a particular neighbourhood should conform well with the requirements of assumption A.

So far we have assumed that a given mass W is to be included in the sample; it may happen that instead a given volume V is to be taken. If the proportion of substance R by volume is to be estimated, then reinterpretation of all masses as volumes makes all the above results applicable. If it is still the proportion by mass that is of interest, then only slight changes in the analysis are required; one would then use $\bar{\pi}$ as estimate and it is easy to show that its asymptotic distribution as $V\to\infty$ is again $N(\pi,\tau^2)$ with now

$$\tau^2=\sigma^2\mu_v/V\mu_w^2$$

where σ^2 and μ_w are as before while μ_v is the average volume per particle.

If the ore is subjected to a crushing process then the characteristics will usually change in such a way that the numerator of τ^2 decreases faster than the denominator for fixed W, i.e. crushing will usually reduce the sample mass needed to attain the same accuracy. For example, suppose that crushing splits up each particle into k particles, each having a k-th part of the mass and the R-mass of the original particle (admittedly not a very realistic assumption, but it will serve to illustrate the point). Then one readily shows that the average mass and R-mass per particle is $1/k$-th that of the original particles while the variances of mass and R-mass and the covariance are only $1/k^2$-th of the original values; hence τ^2 is reduced by a factor of $1/k$, or equivalent, only $1/k$-th the original sample mass is required to attain the same accuracy.

In an application of the above theory it was desired that the sample mass be so large that the estimate be within $\frac{1}{2}$ per cent of the actual value with 99 per cent certainty. The ore in this case was that remaining after all particles under a certain size had been removed. A preliminary sample of particles yielded the following estimates of the characteristics of this population:

$\hat{\mu}_w=2248\ gm$ $\quad$ $\hat{\mu}_q=1854\ gm$

$\hat{\sigma}_w^2=486607\ gm^2$ $\quad$ $\hat{\sigma}_q^2=251938\ gm^2$

$\hat{\sigma}_{qw}=342743\ gm^2$ $\quad$ $\hat{\rho}=0{,}979$

$\hat{\pi}=0{,}825$

From these we estimate τ^2 as

$\hat{\tau}^2=7{,}833/W$

Now, if $\hat{\pi}$ is $N(\pi,\tau^2)$ distributed then 99 per cent confidence bounds for $\hat{\pi}$ are $\hat{\pi}\pm 2{,}58\hat{\tau}$; W must therefore be determined by requiring that $2{,}58\ \hat{\tau}=0{,}005$; this yields $W=2\,086$ Kgm, i.e. a sample mass of approximately 2 metric tons. As the expected number of particles in such a sample will be more than 900 one can be reasonably confident that the large sample theory used is applicable.

DERIVATION OF MAIN RESULT

In this section we show how the main result can be derived.

Let $J_1, J_2,\ldots$ denote the indices of the particles selected for inclusion in the sample. Then assumption A implies that $J_1, J_2,\ldots$ are independent random variables each uniformly distributed over the integers $1,2,\ldots, m$. The number $N=N(W)$ of particles included in the sample is also random, being determined by the condition

$$\Sigma_{i=1}^{N-1} w_{J_i} < W \leqq \Sigma_{i=1}^{N} w_{J_i} \qquad \ldots\ldots\ldots \quad (4)$$

The estimate (2) can now be expressed as

$$\hat{\pi}=\Sigma_1^N q_{J_i}/W. \quad \ldots\ldots\ldots\ldots\ldots (5)$$

and the estimate $\bar{\pi}$ by

$$\bar{\pi}=\Sigma_1^N q_{J_i}/\Sigma_1^N w_{J_i}. \quad \ldots\ldots\ldots\ldots\ldots (6)$$

As $W\to\infty$ the asymptotic properties of these two estimates are similar as will be clear from the following analysis.

First we show that $N(W)\to\infty$ a.s. as $W\to\infty$. It is clear that $N(W)$ increases with W; suppose it tends to a finite limit; then (4) implies that some particle must be of infinite mass which we assume is not the case.

Define random variables

$$X_i=(q_{J_i}-\pi w_{J_i})/\sigma \quad \ldots\ldots\ldots\ldots\ldots (7)$$

with σ as defined in (3). Then

$E(X_i)=(\mu_q-\pi\mu_w)/\sigma=0$

by (1), while

$$\text{Var}\,(X_i)=(\sigma_q^2-2\pi\sigma_{qw}+\pi^2\sigma_w^2)/\sigma^2=1.$$

Also

$$\Sigma_1^N X_i=(\bar{\pi}-\pi)\Sigma_1^N w_{J_i}/\sigma=(\hat{\pi}-\pi)W/\sigma=(\hat{\pi}-\pi)W/\sigma + \pi(W-\Sigma_1^N w_{J_i})/\sigma. \quad \ldots\ldots\ldots (8)$$

From (4) it follows that

$$\Sigma_1^N w_{J_i}/W\to 1 \text{ a.s.} \quad \ldots\ldots\ldots\ldots\ldots (9)$$

Indeed even more is true, viz.

$$|\, W-\Sigma_1^N w_{J_i}\,| \leqq w_{J_N} \leqq \max_i w_i.$$

Hence

$$|\, W-\Sigma_1^N w_{J_i}\,|/N^{1/2}\to 0 \text{ a.s. as } W\to\infty. \quad \ldots\ldots (10)$$

Also (4) and the strong law of large numbers show that

$$W/N\to\mu_w \text{ a.s. as } W\to\infty. \quad \ldots\ldots\ldots\ldots (11)$$

Applying the central limit theorem for a random number of summands[2], we have

$$D(\Sigma_1^N X_i/N^{1/2})\to N(0,1).$$

Combining this with (8)-(11) we find

$$D((\bar{\pi}-\pi)(W\mu_w)^{1/2}/\sigma)\to N(0,1)$$

$$D((\pi-\pi)(W\mu_w)^{1/2}/\sigma)\to N(0,1)$$

from which the result follows.

REFERENCES

1 Bicking, C. A. (1967), 'The Sampling of bulk materials.' Materials Research and Standards. Vol. 7. No. 3. p. 95.

2 Blum, Hanson and Rosenblatt, (1963), 'On the central limit theorem for the sum of a random number of independent random variables.' *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte gebiete.* Vol. 11, p. 389.

The accuracy of estimation from samples of ore in bulk

By J. H. Venter

Contribution by P. R. Janisch B.Sc. (Sur) (Natal)

An Assessment of the practical effects of the Venter analysis.

The kernel of Professor Venter's contribution to the problem of characterising the precision of a bulk sample lies in his equation (3). The effect of this equation is that estimates of mineral values from bulk sampling can, subject to his assumption A, be regarded as normally distributed with parameters depending only on the distribution of size and mineral content of individual particles and on the mass of the sample. In demonstrating that this is the case, Venter has provided a basis upon which further important theoretical and practical work can be constructed.

With a few exceptions, in which Venter's technique may well have an immediate and direct application, certain problems still require to be resolved before any general extension of the technique can be contemplated.

The first difficulty — a perennial one in sampling — is concerned with the actual drawing of the sample. Professor Venter has argued that assumption A may be accepted as valid if the bulk population is thoroughly mixed prior to sampling. Except possibly for bulk sampling from ore in a reduction works, this will rarely be the case. The passage of ore drawn from underground mining operations through orepasses and storage bins produces a degree of mixing but the question remains as to how thorough this will be. Further mixing takes place in the primary and secondary crushing processes, to the extent that Venter's analysis will in all probability be applicable to the determination of sorted waste and mill head values by bulk sampling methods. There is scope, then, for experimental work in the mill which, when subjected to the Venter analysis, would yield information as to whether excessive or inadequate samples are being taken at various stages in the reduction process. This is vital if calculated Mine Call Factors are to be at all meaningful.

There are, of course, many instances of bulk sampling of ore which cannot be regarded as having been thoroughly mixed; these would include broken ore sampling in stopes and at tips, and sampling of ore blasted from open pit faces. Despite genuine attempts at obtaining suitably representative samples the danger of bias is always there and doubt must be cast on the validity of assumption A, and hence on the determi-