

# Methods of Computation for Estimating Geochemical Abundance

---

GEOLOGICAL SURVEY PROFESSIONAL PAPER 574-B



# Methods of Computation for Estimating Geochemical Abundance

By A. T. MIESCH

STATISTICAL STUDIES IN FIELD GEOCHEMISTRY

---

GEOLOGICAL SURVEY PROFESSIONAL PAPER 574-B

*A review of statistically efficient procedures  
for estimating the population arithmetic mean  
where the data are censored at the lower limit  
of analytical sensitivity or positively skewed*



---

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1967

**UNITED STATES DEPARTMENT OF THE INTERIOR**

**STEWART L. UDALL, *Secretary***

**GEOLOGICAL SURVEY**

**William T. Pecora, *Director***

---

For sale by the Superintendent of Documents, U.S. Government Printing Office  
Washington, D.C. 20402 - Price 20 cents (paper cover)

## CONTENTS

---

	Page
Abstract.....	B1
Introduction.....	1
Problems in estimation of geochemical abundance.....	2
Analytical sensitivity.....	3
Small groups of data.....	3
Geometric classes.....	3
Analytical discrimination.....	4
Recommended techniques.....	4
Transformations.....	5
Cohen's method for censored distributions.....	5
$t$ and $t_a$ estimators of Sichel and Krige.....	7
Recommended techniques—Continued	B8
Confidence intervals.....	8
Summary of methods.....	10
Examples.....	10
Molybdenum sulfide in drill core.....	10
Iron in sandstone.....	11
Uranium in granite.....	12
Arsenic in basalts and diabases.....	13
Conclusions.....	14
Literature cited.....	14

---

## ILLUSTRATIONS

---

	Page
FIGURE 1. Histograms showing frequency distributions.....	B6
2. Curves for estimating $\lambda$ for equations 5 and 6.....	7
3. Graphs of $\tau$ as a function of the number of analyses, $n$ , and the antilog of $s$ or $\hat{\sigma}$ .....	8
4. Flow chart showing methods of computation for estimating geochemical abundance.....	9
5. Probability graphs of frequency distributions A-G in figure 1.....	11

---

## TABLE

---

	Page
TABLE 1. Summary of computations for estimating geochemical abundance from data represented in histograms in figure 1....	B12

III

## STATISTICAL STUDIES IN FIELD GEOCHEMISTRY

### METHODS OF COMPUTATION FOR ESTIMATING GEOCHEMICAL ABUNDANCE

By A. T. MIESCH

#### ABSTRACT

Geochemical abundance of an element is regarded as the proportion of a total rock body, or group of rock bodies, that is made up of the element and is equivalent to the population arithmetic mean of sample analyses, generally in units of percentage or parts per million. Computational problems encountered in estimation of abundances can arise from having limited ranges of analytical sensitivity, from having only small groups of data, and from reporting of data in broad geometric classes. These problems can be partly resolved by use of a combination of techniques described by Sichel (1952), Cohen (1959, 1961), and Krige (1960). The combination of techniques allows efficient estimation of geochemical abundances from a wide variety of frequency distribution types if the analytical discrimination is sufficient to allow effective use of data transformations.

Examples of data from the literature were used to demonstrate the techniques; up to 76 percent of the data values were arbitrarily placed in the "not detected" class, and estimated arithmetic means agreed with those estimated from the complete data set to two significant figures. The types of frequency distributions displayed by the data examples are generally representative of those commonly encountered in geochemical problems.

#### INTRODUCTION

The estimation of abundances of constituents in rock bodies, or in groups of rock bodies, is necessary in a broad range of geologic problems. In mining and ore reserve estimation, for example, unbiased and precise estimates of abundance of ore constituents are prerequisite to efficient operational planning, and become mandatory as the grade of the ore declines toward the minimum grade that can be mined profitably. As lower grade deposits of broad extent are being appraised as potential ores of the future, the need for accurate and precise abundance estimates will increase. Aside from these economic problems, the estimation of element abundances is necessary in geochemical problems related to the origin of rock bodies and in both small- and large-scale assessments of geochemical balance among different rock units and other material at the earth's surface.

Estimates of abundance are generally expressed in terms of weight percent or parts per million and are regarded as estimates of the proportion of the rock body that is made up of the constituent of concern. An abundance of 2 percent iron in a 500-million-ton rock body, for example, implies that 10 million tons of iron are present.

The estimation of reliable geochemical abundances is met with serious difficulties in both sampling and in computation. Errors due to sampling are undoubtedly the more serious in a majority of problems, but where sampling has been well planned and successfully executed, correct computational procedures are increasingly desirable. Where sampling has been inadequate (often unavoidably so owing to limited outcrop, for example), reliable abundance estimates are difficult to obtain. Where the sampling has been unbiased but of limited extent, the problem faced is to obtain the best estimate of abundance possible; computational methods are important for this purpose.

Many types of averages (such as medians, modes, geometric means, mean logs, and arithmetic means) are suitable for specific geochemical problems, but the arithmetic mean is the only one of these that is a correct expression of abundance—or an unbiased estimate of abundance—under all circumstances. This was discussed and verified by Sichel (1947, 1952). If the frequency distribution of values used in computing the average is unimodal and symmetrical, the median, mode, and arithmetic mean will be the same; but if the distribution is skewed (asymmetrical), they may differ widely. The geometric mean will always be less than the arithmetic mean regardless of the type of frequency distribution, except where they are equal owing to zero variance in the data. Any type of average may be appropriate for a specific geochemical problem, depending on the purpose for which the average is estimated. The geometric mean, for example, is useful for representing the typical concentration of an element in a

group of rock specimens where the concentrations exhibit a symmetrical frequency distribution on a log scale. The median is a useful expression of average concentration among specimens regardless of the frequency distribution and will be the same as the geometric mean where the distribution is symmetrical on a log scale.

The equivalence of the arithmetic mean of the total population of samples in an entire rock body to geochemical abundance follows from the definition of abundance,  $A$  (in weight percent), as:

$$A = \frac{W}{T} \times 100,$$

where  $W$  is the mass of the constituent in the rock body, and  $T$  is the mass of the rock body. If there are  $N$  potential discrete samples in the total rock body, the mass,  $W$ , of the constituent is given by:

$$W = \frac{1}{100} \sum_{j=1}^N x_j S_j,$$

where  $x_j$  is the concentration (in weight percent) of the constituent in the  $j$ th sample and  $S_j$  is the weight of the sample. If all samples are of the same weight, then

$$W = \frac{T}{100N} \sum_{j=1}^N x_j,$$

and

$$A = \frac{\sum_{j=1}^N x_j}{N}.$$

The right side of this equation is the population arithmetic mean which can be estimated from analytical data by a number of different techniques.

In some problems involving estimation of geochemical abundance, the weighted arithmetic mean is necessary to estimate the population arithmetic mean and has been known to provide accurate abundance estimates. This is well illustrated in many papers on, for example, polygonal methods of ore-reserve estimation. In some other types of problems, weighted arithmetic means are required to correct for highly uneven (biased) sampling, such as in estimation of the abundance of elements in large parts of the earth's crust. Some of the problems met in this work were reviewed by Fleischer and Chao (1960). The methods described in this paper are not intended for problems where weighted arithmetic means are more suitably used.

I am grateful to N. C. Matalas and L. B. Riley of the U.S. Geological Survey for technical criticism and much helpful discussion.

### PROBLEMS IN ESTIMATION OF GEOCHEMICAL ABUNDANCE

Problems in the estimation of abundance (population arithmetic means) arise in both sampling and computation, but only the computational problems are considered in this paper. The more common circumstances leading to computational problems result from (1) limited ranges of analytical sensitivity, (2) small groups of data (small  $n$ ), and (3) data reported in geometric classes. Where none of these circumstances cause difficulty, abundance estimates derived from the ordinary expression for the arithmetic mean,

$$\bar{x} = \frac{1}{n} \sum x_i, \quad (1)$$

will tend to be correct and will be at least relatively efficient. An efficient estimate, in the statistical sense, is one that has a small variance (Fisher, 1950, p. 12), or one that would be expected to change little with the addition of new data. Concern for the efficiency of statistical estimates gave rise to the "maximum-likelihood method" used in devising estimation methods (Fisher, 1950, p. 14). Estimators derived in this manner are said to yield statistics with minimum variance (maximum efficiency). The expression for the arithmetic mean in equation 1 is a maximum-likelihood estimator where the data are derived from a normally distributed population; it is not a maximum-likelihood estimator for many of the problems in geochemical studies.

In problems where data from a normally distributed population are censored (some concentration values fall beyond the range of sensitivity for the analytical method), a maximum-likelihood method described by Cohen (1959, 1961) may be used. Where the data are from a lognormal population, a maximum-likelihood method given by Sichel (1952) will be applicable. A modification of the method given by Sichel will be useful where the data indicate certain kinds of departure from the lognormal (Krige, 1960).

Some estimation methods derived by the method of maximum likelihood are slightly biased for small  $n$ . Unbiased likelihood estimators are those that have been corrected for the bias, where such correction is needed.

Abundance estimates derived by maximum-likelihood techniques will not differ greatly, in many instances, from estimates derived by other reasonable though less precise methods. In much geochemical work, especially that conducted on a large scale, errors due to sampling and analysis are overwhelmingly dominant in determination of the total estimation error; computational errors are relatively small. Nevertheless, the possibility of large errors from other sources cannot serve to justify additional errors where they can be easily avoided.

#### ANALYTICAL SENSITIVITY

Every analytical method has limits of sensitivity beyond which it is ineffective for determination of concentration values. These limits may occur at both the lower and the upper bounds for a concentration range. The spectrographic method described by Myers, Havens, and Dunton (1961), for example, may be used to determine silicon within the range from 0.002 to 10 percent. Concentrations judged to be lower or higher than this range are reported as <0.002 percent or >10 percent, respectively. Thus, the spectrographic data may be either left or right censored. The term "censored" is applicable here because the number of values beyond each sensitivity limit is known for any set of analyzed rock samples. In other types of problems, where the number of values beyond certain limits is unknown, the data are said to be truncated (Cohen, 1959, p. 217).

A number of methods have been used by individual geologists for computing means from censored data. (See, for example, Miesch, 1963, p. 21-23). Other methods that are available, however, are a great deal more satisfactory (Cohen, 1959, 1961; Hald, 1952, p. 30; Hubaux and Smiriga-Snoeck, 1964). The maximum-likelihood method described by Cohen will be applied to abundance estimation problems in later sections.

Abundance estimates can be obtained directly from censored normal distributions, using Cohen's method, but not from other types of censored distributions. Where the analytical data do not indicate a normal distribution, some transformation of the data can be used, as will be demonstrated in a following section. The mean and variance of the transformed values, in many problems, can then be used to derive abundance estimates.

#### SMALL GROUPS OF DATA

The precision, or reproducibility, of any statistical estimate is largely dependent on the number of values on which the estimate is based. Those estimates derived from large data sets will be relatively precise, even where the statistical method is not the most efficient one that could be used (where the statistical method has not been derived by the method of maximum likelihood). Where the data set is small, however, an unbiased maximum-likelihood technique should be used wherever possible if the precision of the statistical estimate is important. It is not possible here to state a value of  $n$  which will serve to distinguish between large and small data sets because this value will vary according to the variation in the data. A better approach may be to accept the maximum-likelihood estimate, if obtainable,

wherever it differs from an estimate derived by other techniques.

Where the data are derived from a normal distribution, an estimate of abundance derived from equation 1 is based on the method of maximum likelihood and is the most efficient estimate of abundance possible for a given number of samples. However, most underlying frequency distributions in geochemical studies are not normal; more commonly they are (1) asymmetrical with a long tail toward high values (positive skewness), (2) asymmetrical with a long tail toward low values (negative skewness), or (3) multimodal with more than one peak in the frequency distribution curve. Where skewness or the presence of more than one mode is sufficient to indicate a departure of the underlying frequency distribution from the normal form, abundance estimates derived from equation 1 will not be as efficient as maximum-likelihood estimates, if the latter are available. The difference in efficiency may be large if  $n$  is small.

The most common departure of geochemical data from the normal distribution is a positive skewness. In many frequency distributions the skewness, along with other properties of the distribution, may reflect an underlying lognormal distribution. Where this is true, an unbiased maximum-likelihood method for deriving efficient estimates of abundance (the population arithmetic mean) can be applied (Sichel, 1952). Where the positive skewness in a particular data set is less or greater than that which can be ascribed to a lognormal distribution, a modification of this method may be used (Krige, 1960). The methods of Sichel and Krige involve data transformations.

Unbiased maximum-likelihood methods of abundance estimation applicable to frequency distributions which are negatively skewed or multimodal are unknown to the writer.

#### GEOMETRIC CLASSES

Any quantitative analytical determination may be regarded as a reported range or class; a concentration reported as 78.62 percent, for example, signifies the range from 78.615 to 78.625 percent as the analyst's best estimate of the true value. The classes are broader where fewer significant figures are reported. Where the values are reported to an equal number of decimal places, the class widths are equal; the class boundaries are arithmetic, because they increase by a constant increment.

In much spectrographic and colorimetric work the analyst reports concentration values in broad classes, without specifying single values (except where they are meant only to identify a class). Classes are used for

reporting because of relatively poor discriminatory capacity of the analytical methods. Generally, the boundaries of the classes increase geometrically, each being higher than the previous boundary by a constant multiplier. Geometric classes are used because the error variance is at least approximately proportional to the amount of the constituent present.

An example of the geometric classes used in spectrographic work was given by Myers, Havens, and Dunton (1961, p. 217), who formerly reported the concentrations of 68 elements in classes having the boundaries—0.00010, 0.00022, 0.00046, 0.0010, 0.0022, . . .—increasing by a factor equal to the cube root of 10, up to 10 percent. Myers and other U.S. Geological Survey spectrographers currently report values in classes with boundaries increasing by a factor equal to the 6th root of 10 (0.00012, 0.00018, 0.00026, 0.00038, 0.00056, 0.00083, 0.0012, . . .). Other similar systems of reporting, using various other factors for the generation of class boundaries, were described by Barnett (1961, p. 184).

Although the reporting of spectrographic analyses in geometric classes has allowed spectrographers to produce a vast amount of data useful in geochemical problems, the practice has presented some difficulties in statistical analysis that have not, so far, been satisfactorily resolved. Because the class boundaries form a geometric progression, the class widths are unequal in size, and conventional methods of treating grouped data are awkward and less efficient than other methods that might be used. There is also the problem of choosing the best midpoint to represent a frequency class in the grouped-data computations. Where a number of concentration values are reported to occur within the range from 0.0046 to 0.010 percent, for example, it would seem proper to use the arithmetic midpoint of 0.0073, rather than the geometric midpoint of 0.0068, for computing the sample arithmetic mean for the whole data set. The use of the geometric midpoint, however, nearly always gives better answers. (An example is given later in this paper.) The reason for this is that the use of the lower value partly compensates for a positive bias caused by the grouping technique, but the compensation is without any sound theoretical basis and should not be relied on.

When a logarithmic transformation of the class boundaries is made, the class widths (in log units) become equal, thus allowing straightforward use of grouped-data computational methods for deriving the mean logarithm and standard deviation of the logs. By using these values, the abundance (population arithmetic mean) can be estimated according to the maximum-likelihood method of Sichel (1952), if the form

of the underlying frequency distribution is at least approximately lognormal.

#### ANALYTICAL DISCRIMINATION

For many types of spectrographic or colorimetric analytical techniques the analyst is unable to estimate concentrations to more than one significant figure. If a number of concentrations of the constituent sought are similar among the samples analyzed, the suitability of the technique for discriminating among the samples is poor. Commonly, a large proportion of the determinations are reported as the same value. For example, Huff (1955, p. 111) gave copper determinations from a chromatographic field technique on 23 samples of Tapeats Sandstone (Cambrian) from near Jerome, Ariz. Seventeen of the 23 determinations are given as 10 ppm (parts per million); the remaining 6 are either 50, 100, or 150 ppm. Many other examples of poor analytical discrimination could be given involving data reported in geometric classes. In many instances only three or four adjacent classes are used in reporting, and commonly 20–50 percent of the concentrations occur within one class. Discrimination among values in the same class, of course, is impossible. The proportion of values reported in a single class is dependent on the class widths, the variation in the concentrations, and the form of the frequency distribution.

As has been pointed out, some of the computational problems encountered in abundance estimation can be wholly or partly resolved by means of data transformations. Where analytical discrimination is poor, however, some of the more useful transformations are impossible or ineffective. No transformation would be useful, for example, in normalizing Huff's data referred to above, and transformations other than the logarithmic transformation would be ineffective in treating most data reported in broad geometric classes.

#### RECOMMENDED TECHNIQUES

Where large sets of uncensored data are available, abundances may be estimated directly by using the conventional formula for arithmetic means in equation 1. If the data are part of an underlying normal distribution of values, the abundance estimate will be the most efficient possible. If the underlying distribution is not normal, it may be possible to obtain a more efficient estimate by other techniques. However, if  $n$ , the number of values, is large, the increase in efficiency may be small.

Where small sets of uncensored data are used, abundance estimates derived from the conventional formula for the arithmetic mean will be as efficient as possible if the underlying frequency distribution is normal.

Where the distribution is positively skewed, abundance estimates that are significantly more efficient may be obtained by other methods. These methods begin with data transformations.

Where the data are censored, abundances may be estimated using the techniques available for estimating the population arithmetic mean, if the total underlying distribution is normal. Where it is not normal, data transformations may be employed. The mean and standard deviation of the transformed values may then be used to derive abundance estimates.

#### TRANSFORMATIONS

Many types of data transformations have been used in geologic and geochemical problems to normalize observed frequency distributions, but the types that appear more useful in problems of estimating geochemical abundance are the  $\log y$  and  $\log(y+\alpha)$  transformations, where  $y$  is the concentration of the element, in percent or in parts per million, and  $\alpha$  is a constant. (All logarithms used in this paper are to the base 10.) Data which can be transformed to the normal form by these functions are referred to as 2- and 3-parameter lognormal, respectively (Aitchison and Brown, 1957, p. 7, 14).

These two transformations generally appear to be effective where the distribution of the original analytical data is unimodal and positively skewed. The simple log transformation of some positively skewed data will lead to a log distribution with negative skewness. For other data (fig. 1E) the log transformation will lead to a distribution with some degree of positive skewness remaining (fig. 1F). In either case the constant,  $\alpha$ , can be estimated using techniques described by Krige (1960, p. 236) and used in the transformation  $\log(y+\alpha)$ . Where a negative skewness of logs is to be corrected,  $\alpha$  will be a positive value; where a positive skewness in the logs is to be corrected,  $\alpha$  will be negative. In some cases the absolute quantity of the derived negative value of  $\alpha$  will equal or exceed some of the concentration values,  $y$ , so that the quantity  $(y+\alpha)$  is zero or negative and  $\log(y+\alpha)$  is undefined for some analytical values. Transformed distributions that result are, in effect, censored (fig. 1G), and abundance estimates may be obtained by using techniques for censored distributions.

The choice of a transformation that may be effective in producing a normal, or at least symmetrical, frequency distribution can be governed by examination and testing of the data available or by previous experience with similar data. Selections based on the data avail-

able are generally preferred, though this is not always possible when a group of data is small or when a large proportion of the data is censored. Many data pertaining to the concentration of minor elements in rock samples collected over large areas correspond more closely to the lognormal form than to the normal, and the log transformation is appropriate in a large number of problems. The  $\log(y+\alpha)$  transformation can frequently be used to improve on the simple log transformation if a satisfactory estimate of  $\alpha$  can be obtained.

When the data have been transformed by  $x = \log y$  or  $x = \log(y+\alpha)$ , the mean of  $x$  is estimated from equation 1 and the standard deviation is estimated from equation 2. The mean and standard deviation of these transformed values are of little value themselves in problems of abundance estimation, but they may be used to derive estimates of arithmetic means by techniques developed and described by Sichel (1952) and Krige (1960).

$$s = \left[ \frac{\sum x^2 - n\bar{x}^2}{n} \right]^{1/2} \quad (2)$$

Equation 2 is a biased estimator of the population standard deviation but is used in this form in computations described later in the paper. Where unbiased estimates of standard deviation are needed,  $n$  in the denominator of equation 2 is replaced by  $n-1$ .

#### COHEN'S METHOD FOR CENSORED DISTRIBUTIONS

Cohen (1959, 1961) presented maximum-likelihood techniques for estimation of the mean and standard deviation from either censored or truncated normal distributions. These techniques may be used whether the unknown part of the distribution is in the low- or the high-value region (left- or right-censored, respectively). Because left-censored distributions are by far the more common in geochemical problems, only the part of Cohen's techniques applicable to these distributions will be discussed here. Although Cohen's methods are strictly applicable to normal distributions only, it can easily be demonstrated that the method for left-censored distributions provides satisfactory estimates of arithmetic means wherever the total distributions are symmetrical about one mode. The method given by Cohen has not been corrected for a small bias, and other methods for treating censored distributions may be more accurate when  $n$  is less than about 10 (Cohen, 1959, p. 218).

The methods given by Cohen (1959, 1961) are preferred to others given by Hald (1952) and Hubaux and Smiriga-Snoeck (1964) only because of the simplicity of computation.

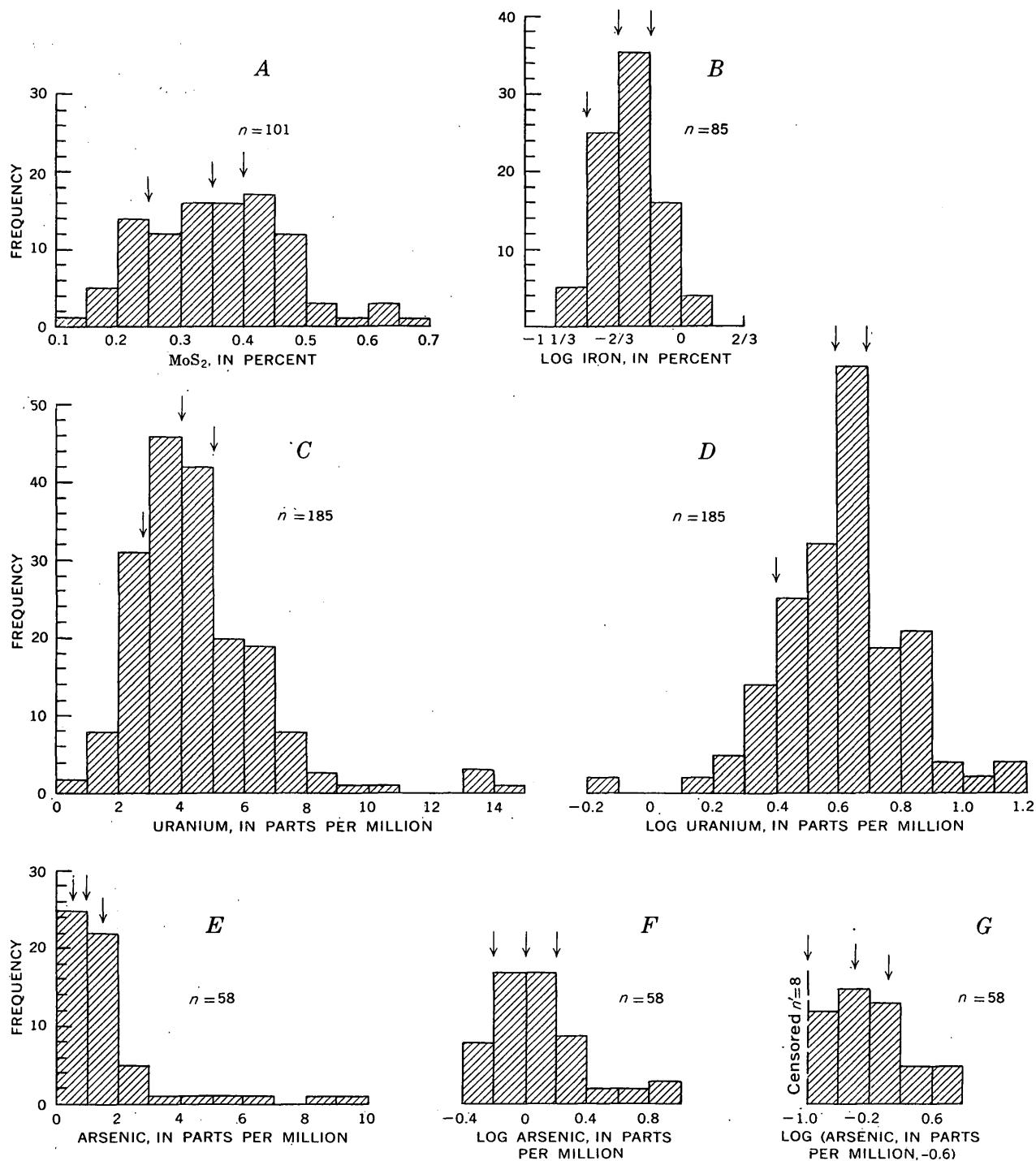


FIGURE 1.—Frequency distributions. *A*, MoS<sub>2</sub> values for 300-degree samples at 2-foot intervals in drill core from Climax, Colo. (from Hazen and Berkenhotter, 1962, p. 84–86). *B*, Iron in sandstones (from Miesch, 1963, pl. 3). *C* and *D*, “Uranium in granite à deux micas de Pontivy” (from Hubaux and Smiriga-Snoeck, 1964, p. 1207). *E*, *F*, and *G*, Arsenic in basalts and diabases (from Onishi and Sandell, 1955, tables 5, 10).

In Cohen's technique  $x_o$  is taken as either the transformed value of the lower limit of analytical sensitivity or as the limit of sensitivity (in percent or in parts per million) if the data transformation is not required. The quantity  $n'$  is the number of concentration values that are below the limit of sensitivity; such concentrations are generally reported by the analyst as "not detected" or zero. The ratio  $h=n'/n$  is the fraction of the total number of analyzed specimens in which the element was not detected. The mean and standard deviation of the analytical values above the limit of sensitivity are computed as:

$$\bar{x}' = \frac{\sum x}{n-n'} \quad (3)$$

and

$$s' = \left[ \frac{\sum x^2}{n-n'} - (\bar{x}')^2 \right]^{1/2}. \quad (4)$$

The mean and standard deviation of the entire distribution are then estimated from:

$$\hat{\mu} = \bar{x}' - \lambda(\bar{x}' - x_o) \quad (5)$$

and

$$\hat{\sigma} = [(s')^2 + \lambda(\bar{x}' - x_o)^2]^{1/2}. \quad (6)$$

The value  $\lambda$  in equations 5 and 6 is a function of  $h$  and the quantity  $(s')^2/(\bar{x}' - x_o)^2$ . Values of  $\lambda$  for  $0.01 \leq h \leq 0.90$  and  $0.00 \leq (s')^2/(\bar{x}' - x_o)^2 \leq 1.00$  were tabulated by Cohen (1961, p. 538), and graphs of  $\lambda$  were given in an earlier paper (Cohen, 1959, p. 231). The graphs are reproduced here in figure 2, with Dr. Cohen's permission.

#### *t AND t<sub>α</sub> ESTIMATORS OF SICHEL AND KRIGE*

Where the original analytical data have been used in computation without transformations, the derived values of  $\alpha$  (eq 1) or  $\check{\mu}$  (eq 5) may be taken directly as estimates of geochemical abundance. However, when  $\bar{x}$  or  $\mu$  are derived for  $\alpha = \log y$  or  $\alpha = \log(y + \alpha)$ , abundances can be estimated using methods described by Sichel (1952) and Krige (1960). The method by Sichel is an unbiased maximum-likelihood technique; that by Krige is a modification of Sichel's method, to be used where the  $\log(y + \alpha)$  transformation is required. The estimators, where the  $\alpha = \log y$  and  $\alpha = \log(y + \alpha)$  transformations are used, are referred to as  $t$  and  $t_\alpha$ , respectively, and should not be confused with Student's  $t$ , which has wide application in statistical procedures.

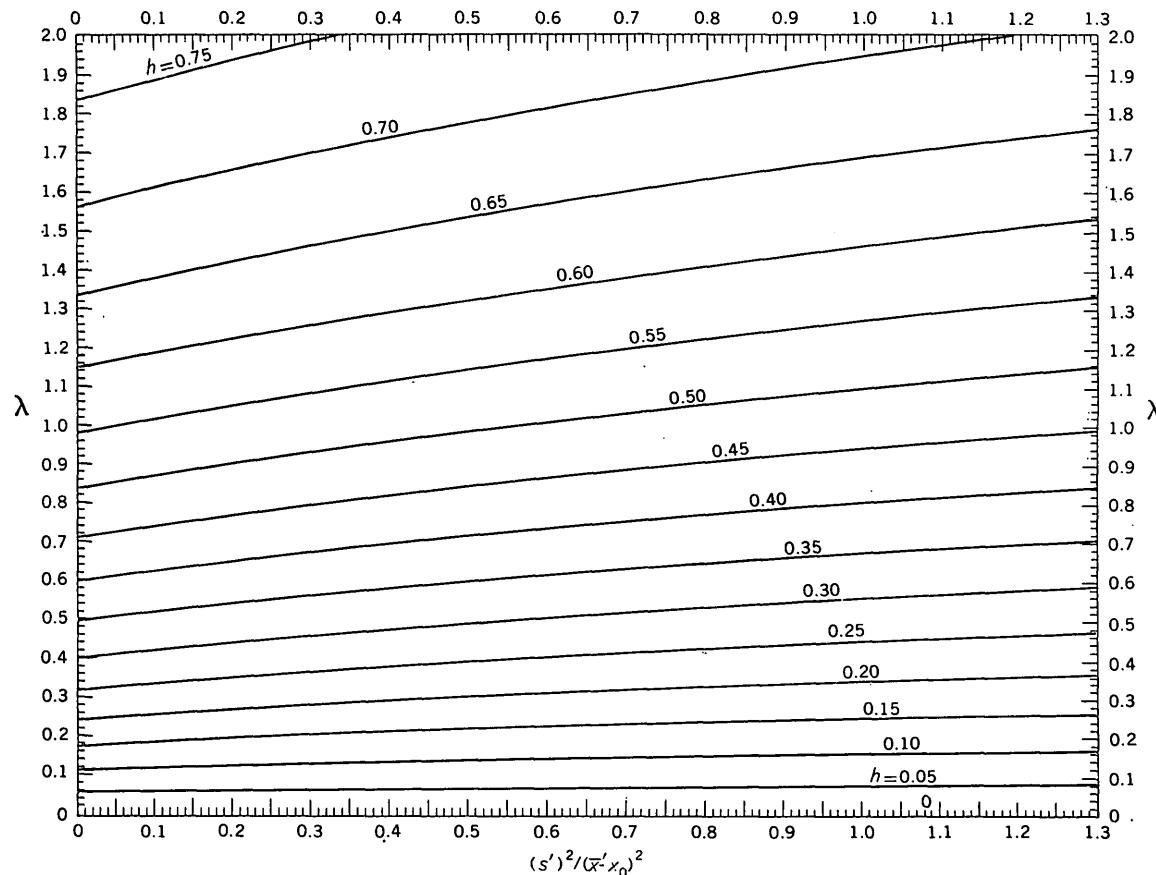


FIGURE 2.—Curves for estimating  $\lambda$  for equations 5 and 6 (from Cohen, 1959, p. 231; reproduced with author's permission).

The  $t$  and  $t_\alpha$  statistics provide useful estimates of the arithmetic mean and geochemical abundance when the distribution of  $\log y$  or  $\log(y+\alpha)$ , respectively, approximates the normal distribution form.

Where  $\log y$  is normally distributed:

$$\left. \begin{aligned} t &= \tau \times 10^{\bar{x}} \\ \text{or} \\ t &= \tau \times 10^{\hat{\mu}} \end{aligned} \right\}, \quad (7)$$

where  $\bar{x}$  and  $\hat{\mu}$  are the arithmetic means of  $\log y$  estimated from complete (eq 1) and censored (eq 5) distributions, respectively. The factor  $\tau$  is a function of  $n$ , the total number of specimens analyzed and either the antilog of  $s$  (eq 2) or of  $\hat{\sigma}$  (eq 6), depending on whether the distribution is complete or censored. Values of  $\tau$  may be read from graphs in figure 3; for more exact work the reader is referred to the original equation and tables (Sichel, 1952, p. 275, 284-288) from which equation 7 and the graphs in figure 3 were derived. Simplified equations and tables were given recently by Sichel (1966). Other equations and tables that are mathematically equivalent to those of Sichel were given by Aitchison and Brown (1957, p. 45, 156-158) and were based on the work of Finney (1941).

Where  $\log(y+\alpha)$  is normally distributed:

$$\left. \begin{aligned} t_\alpha &= (\tau \times 10^{\bar{x}}) - \alpha \\ \text{or} \\ t_\alpha &= (\tau \times 10^{\hat{\mu}}) - \alpha \end{aligned} \right\}, \quad (8)$$

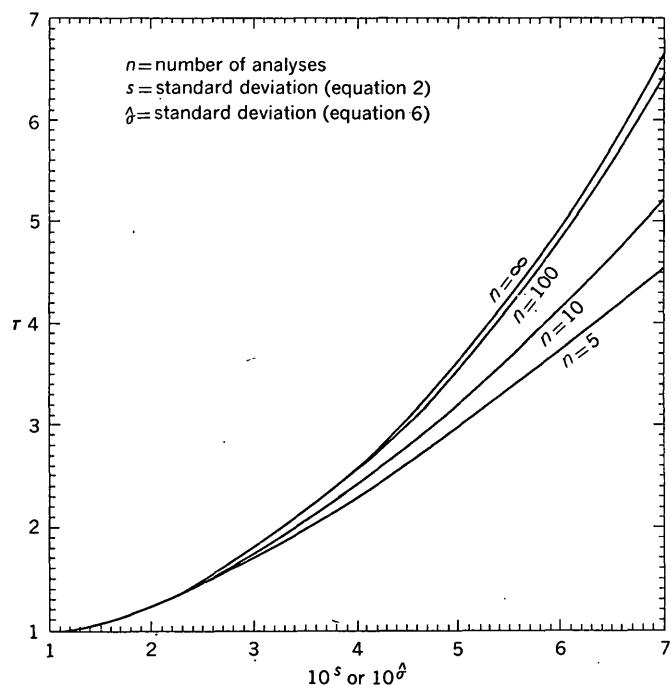


FIGURE 3.—Graphs of  $\tau$  as a function of the number of analyses,  $n$ , and the antilog of  $s$  or  $\hat{\sigma}$  (based on tables by Sichel, 1952).

where  $\bar{x}$  and  $\hat{\mu}$  are the arithmetic means of  $\log(y+\alpha)$  estimated from complete (eq 1) and censored (eq 5) distributions, respectively. The equations in 8 are modified forms of one given by Krige (1960, p. 239). The factor  $\tau$  is a function of  $n$ , the total number of specimens analyzed, and of the standard deviation of the transformed values,  $\log(y+\alpha)$ . If complete distributions are used, the standard deviation is estimated by  $s$  (eq 2); if censored distributions are used, it is estimated by  $\hat{\sigma}$  (eq 6). The value of  $\tau$  can be read from figure 3 by using  $n$  and the antilog of either  $s$  or  $\hat{\sigma}$  for the transformed values. For more exact computation the reader is referred to Krige (1960, p. 239) and tables given by Sichel (1952, p. 284-288), or to Sichel (1966).

#### CONFIDENCE INTERVALS

Estimates of geochemical abundance should be accompanied by estimates of the confidence intervals wherever possible. The method for estimating this interval about the arithmetic mean derived from equation 1 is well known (see Dixon and Massey, 1957, p. 128) and leads to approximately correct intervals even where the distributions are not normal. Cohen (1959, p. 231-232) gave methods for estimating confidence intervals of  $\mu$  derived from censored distributions. Approximate confidence intervals for the  $t$  and  $t_\alpha$ , may be estimated by using equations given by Aitchison and Brown (1957, p. 46); more exact confidence intervals can be obtained by using methods given recently by Sichel (1966).

#### SUMMARY OF METHODS

A flow chart indicating recommended computational procedures for estimating geochemical abundances is given in figure 4. With large groups of uncensored data, abundances may be estimated as arithmetic means in the conventional manner, using equation 1. However, where means obtained from equation 1 differ from those obtained from procedures indicated in figure 4, the means obtained from the latter procedures should be preferred. The higher precision of the  $t$  and  $t_\alpha$  estimators have been adequately demonstrated by Sichel (1952, p. 276-278) and Krige (1960, p. 242-244), respectively, and the advantages of treating censored distributions by the statistical method of Cohen (1959, 1961) over the approximate methods used in the past will be obvious.

The methods of Sichel (1952), Krige (1960), and Cohen (1959, 1961) are individually inadequate for a large number of computational problems met in analyzing geochemical data. Sichel's  $t$  estimator is useful where the data correspond to the lognormal frequency distribution and are uncensored. Krige's  $t_\alpha$  estimator

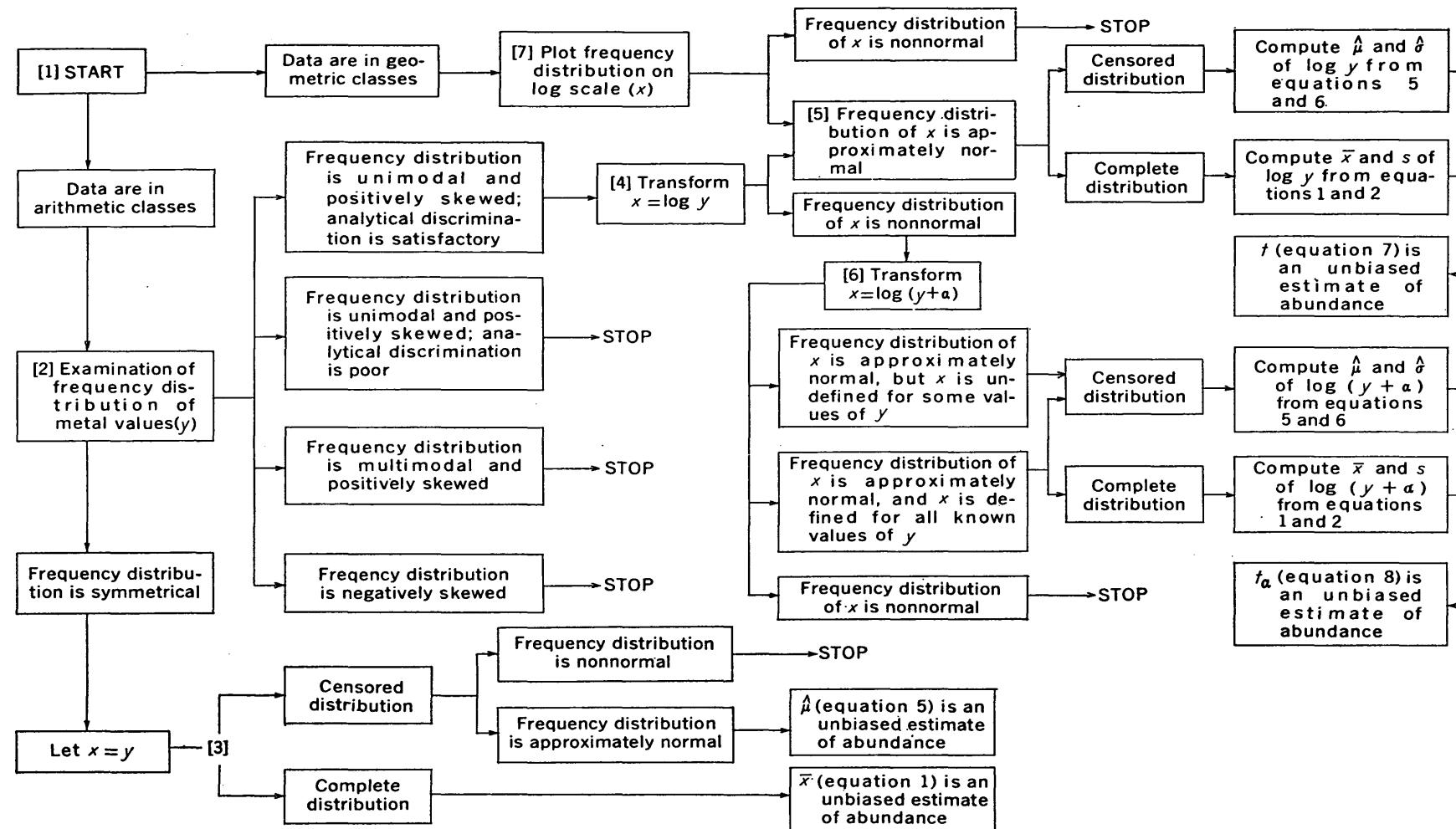


FIGURE 4.—Methods of computation for estimating geochemical abundance from small groups of data.

is applicable to a wider variety of distribution types, but, as with the  $t$  estimator, the data must be uncensored. Moreover, with highly skewed frequency distributions, the  $\log(y+\alpha)$  transformation may impose censoring of data which are completely above the lower sensitivity limit of the analytical method. Cohen's method is directly applicable to data which approximate a normal distribution, but data of this type are not common in minor-element studies pertaining to areas larger than a single mine or small outcrop. Generally the required normal distribution can be achieved only through a data transformation; estimates of the mean of the transformed variate are not valid estimates of abundance but may be transformed to valid abundance estimates using the methods of Sichel or Krige, if the necessary data transformation is accomplished by  $\log y$  or  $\log(y+\alpha)$ . The transformations are, indeed, sufficient in a wide variety of geochemical studies.

Some properties of small data sets prohibit the estimation of precise arithmetic means, or abundances. These are indicated by STOP signs on the flow chart in figure 4, and are as follows:

1. A censored frequency distribution which, though believed to be part of a symmetrical distribution, departs widely from the normal form (for example, some censored multimodal distributions).
2. A frequency distribution which is markedly asymmetrical and cannot be transformed to an approximate normal distribution by  $\log y$  or  $\log(y+\alpha)$ . This particularly includes multimodal skewed distributions and all negatively skewed distributions.
3. A frequency distribution that is unimodal and positively skewed but is based on data that cannot be normalized by the  $\log y$  or  $\log(y+\alpha)$  transformations owing to poor analytical discrimination or other factors.
4. Data reported in broad geometric classes, but not approximately normal on a log scale.

Where these properties are present, the conventional method of estimating the population arithmetic mean (eq 1) may be the only way readily available to the geologist for estimating geochemical abundance.

The use of the flow chart, figure 4, is demonstrated in the following section by employing examples of data from the literature.

#### EXAMPLES

Four data sets were selected from the literature to illustrate use of the recommended methods for widely differing types of data. One data set is approximately normally distributed, two are approximately lognormal (one of these is reported in geometric classes), and a fourth set is approximately lognormal after adjustment

by a constant,  $\alpha$ . However, except for the data in geometric classes, none of the data sets indicate a close correspondence to the normal or lognormal form; this selection has been intentional to demonstrate that the distribution requirement is not rigid.

As shown by these and other examples, the principal requirement is that the distributions be unimodal and approximately symmetrical on any of the three scales—the scale of original measurement,  $y$ , or one of the two transformed scales,  $\log y$  or  $\log(y+\alpha)$ .

The number of values in each of the four data sets used as examples (fig. 1) is large compared with the number available in many geochemical problems. Most of the estimated abundances, therefore, agree fairly well with the abundances derived using the conventional method for estimating the population arithmetic mean. Large data sets were used so that this comparison could be made to verify the accuracy of the techniques for use in actual abundance estimation problems where  $n$  is small or where the data are censored.

#### MOLYBDENUM SULFIDE IN DRILL CORE

The first set of data is from Hazen and Berkenhotter (1962, p. 84-86). The assays, of percent  $\text{MoS}_2$ , were made on samples of drill core (300-degree segments) from the Climax Molybdenum mine, Lake County, Colo. We shall assume, for purposes of illustration, that the assays are an objective and unbiased sample of those that might have been obtained from the block of ore penetrated by the drill hole (that is, there is no sampling problem). The frequency distribution of the original assays is represented by histogram  $A$  in figure 1 and by the probability graph for distribution  $A$  in figure 5. Both illustrations indicate that the frequency distribution of assays (at least the central part) is approximately symmetrical. The distribution was arbitrarily censored successively at three points indicated by the small arrows (fig. 1A), and four abundance estimates were made using the complete and the partial data sets. The estimates, with intermediate computational values, are given in table 1.

In reference to figure 4:

- [1] The  $\text{MoS}_2$  assays are in arithmetic classes. Proceed to [2].
- [2] The frequency distribution of  $\text{MoS}_2$  assays,  $y$ , is approximately symmetrical (figs. 1A, 5A). Proceed to [3].
- [3] If the complete distribution is used, the abundance of  $\text{MoS}_2$  in the ore block is estimated, by the conventional method in equation 1, to be 0.359 percent.

If only assays equal to or greater than the arbitrary (or hypothetical) analytical cutoffs—0.25, 0.35,

and 0.40 percent—are used (table 1, col. 2), the quantities indicated in columns 6–14 (table 1) are derived from the indicated equations, and the successive abundance estimates are derived as shown in column 17. The abundance estimates are in agreement to two figures, even though as much as 63 percent of the data was arbitrarily censored (table 1, col. 8).

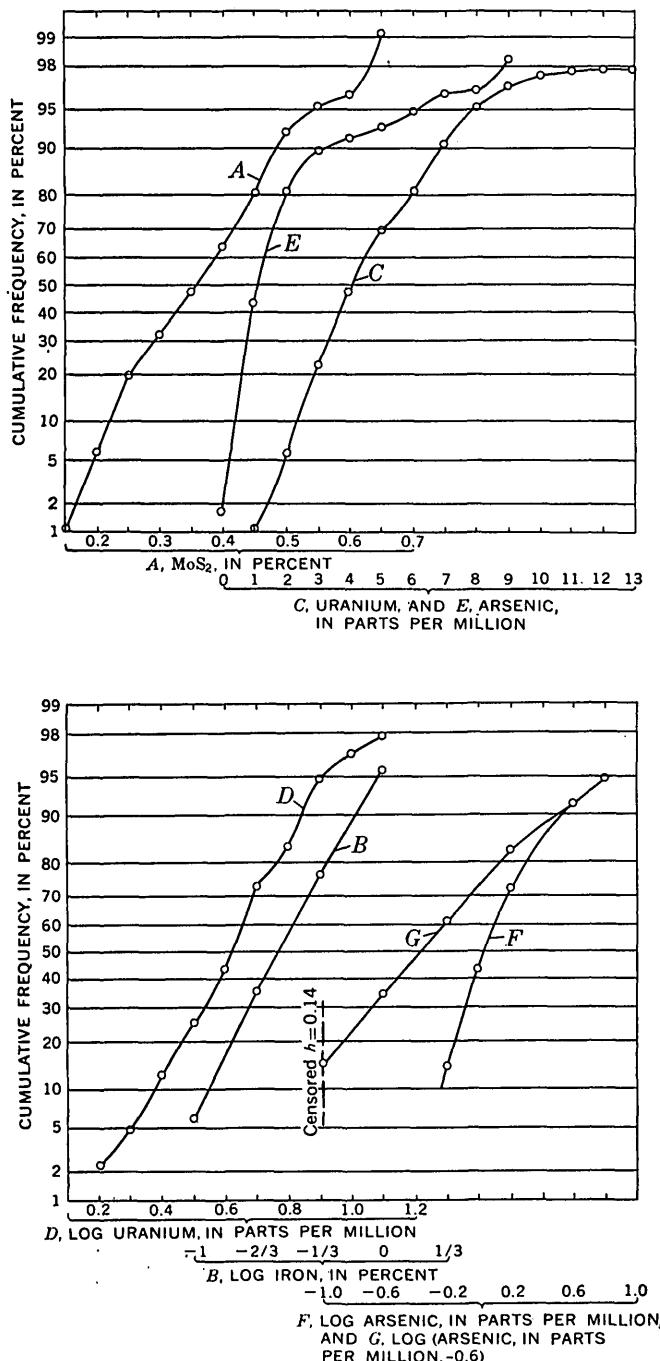


FIGURE 5.—Probability graphs of frequency distributions A–G in figure 1.

Had the data, in fact, been censored at the arbitrary cutoff values, it would have been necessary to judge the nature of the total frequency distribution from as little as 37 percent of the data occurring above the cutoffs. In this extreme example of data censoring, the judgment regarding the form of the frequency distribution could be made only on the basis of prior experience with similar data. Where the censored part of the distribution is minor (less than about one-third of the total distribution), the probability graphs may still be useful for this purpose.

#### IRON IN SANDSTONE

The iron concentration in 85 drill-core samples of quartzose sandstone from the Salt Wash Member of the Morrison Formation (Jurassic) in San Miguel County, Colo., are represented by histogram B in figure 1. The data are from Miesch (1963, pl. 3) and were obtained by means of a spectrographic method (Myers and others, 1961) wherein the results are reported in classes having the boundaries 0.046, 0.10, 0.22, 0.46, . . . percent. The corresponding logarithms of the boundaries are  $-1\frac{1}{3}$ ,  $-1$ ,  $-\frac{2}{3}$ ,  $-\frac{1}{3}$ , . . . . The boundaries, in percentage concentration values, are at geometric intervals and increase by a factor of 2.15. The boundaries, in log values, increase by an increment of one-third. We shall assume that the analyzed samples are an unbiased representation of the body of rock for which the geochemical abundance is to be estimated.

The estimate of iron abundance in the sandstone, derived from the grouped data form of equation 1, is 0.38 percent. The grouped-data form of equation 1 is:

$$\bar{x} = \frac{1}{n} \sum f_i x_i, \quad (9)$$

where  $f_i$  is the number of values in the  $i$ th class and  $x_i$  is the class midpoint. The midpoints used were the geometric centers of the classes—the values 0.07, 0.15, 0.32, 0.68, and 1.46 percent. There is little justification for accepting the value of 0.38 percent as a valid estimate of abundance other than the fact that it agrees closely with the abundance of 0.37 percent derived with the theoretically justified  $t$  estimator. It is not expected that equation 9, used in the manner described here (with geometric midpoints), will consistently lead to unbiased and efficient abundance estimates. The use of arithmetic midpoints (0.07, 0.16, 0.34, 0.73, and 1.58) leads to an abundance estimate of 0.40 percent. Although the estimate of 0.40 percent is not greatly different from that of 0.37, it does demonstrate the positive bias in the technique by which it was obtained. The positive bias is commonly much larger.

TABLE 1.—Summary of computations for estimating geochemical abundance from data represented in histograms in figure 1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Distribution <sup>1</sup>	Hypo-thetical analytical cutoff	Transformation	$\alpha$	$x_o$	$n'$	$n$	$h = \frac{n'}{n}$	$\bar{x}$ (eq 1) or $\bar{x}'$ (eq 3)	$s$ (eq 2) or $s'$ (eq 4)	$(s')^2$ $(x' - x_o)^2$	$\lambda$ (fig. 2)	$\hat{\mu}$ (eq 5)	$\hat{\sigma}$ (eq 6)	$t$ (eq 7)	$t\alpha$ (eq 8)	Estimated abundance	
Molybdenum sulfide (percent)																	
A	0.25	None			0	101	0	0.359	0.110	0.39	0.27	0.356	0.118			0.359	
A	do			0.25	20	101	.20	.396	.091	.58						.356	
A	.35	do		.35	48	101	.48	.445	.072	.94	.356	.117				.356	
A	.40	do		.40	64	101	.63	.475	.066	.77	1.55	.359	.114			.359	
Iron (percent)																	
B	10	$x = \log y$			-1.000	0	85	0	-.543	.314	.34	.08	-.544	.313	.37	.37	
B	do				-.667	5	85	.06	-.504	.281	.45	.59	.313	.37		.37	
B	.22	do			-.333	30	85	.35	-.355	.209	.45	.59	.318	.38		.38	
B	.46	do				65	85	.76	-.100	.133	.33	≈ 2.04	-.576	.359	.37	.37	
Uranium (ppm)																	
C	2.6	None			2.6	0	185	0	4.53	2.15	.80	.18	4.48	2.26		4.53	
C	do				4.0	23	185	.12	4.89	2.04	1.30	.94	4.11	2.66		4.48	
C	4.0				80	185	.43	5.78	2.03	.199						4.11	
C	$x = \log y$				0	185	0	.612	.199							4.54	
C	2.6	do			.415	23	185	.12	.660	.152	.39	.16	.621	.181	4.55	4.55	
C	4.0	do			.602	80	185	.43	.742	.125	.77	.86	.620	.181	4.54	4.54	
C	5.1	do			.708	135	185	.73	.845	.105	.58	2.00	.571	.220	4.24	4.24	
Arsenic (ppm)																	
E		None				0	58	0	1.69	1.78						1.69	
F		$x = \log y$				0	58	0	.096	.299						1.58	
F	.7	do				-.155	8	.14	.152	.283	.85	.22	.085	.318	1.58	1.58	
F	1.0	do				.000	22	.58	.257	.267	1.07	.75	.064	.348	1.59	1.59	
F	1.6	do				.204	42	.58	.72	.488	.244	.74	1.96	-.069	.467	1.51	
G	.7	$x = \log(y + \alpha)$			-0.6	-1.000	9	.58	.14	-.208	.510	.41	.20	-.367	.621		1.77
G	1.0	do			-.6	-.398	22	.58	.38	.032	.377	.76	.72	-.279	.526		1.77
G	1.6	do			-.6	.000	42	.58	.72	.375	.296	.62	1.92	-.345	.598	1.68	1.68
																1.75	

<sup>1</sup> The letters in this column refer to frequency distributions shown in figs. 1 and 5.

In reference to figure 4:

- [1] The data are grouped in geometric classes. Proceed to [7].
- [7] The frequency distribution, on a log scale, is shown in figure 1B. The probability graph for this distribution is given in figure 5. No important departure from the normal form is indicated, if the log scale is used. Proceed to [5].
- [5] The abundance of iron in the sandstone, using Sichel's  $t$  estimator, is estimated to be 0.37 percent. If the data are censored at the arbitrary points, 0.10, 0.22, and 0.46 percent, and 6, 35, and 76 percent of the data is effectively lost, the abundance estimates are 0.37, 0.38, and 0.37 percent, respectively (table 1, col. 17).

If only data values above the 0.46-percent class boundary (20 of the 85 values) are used, a large positive skewness is apparent from the fact that the median is between 0 and 0.46 percent—far below the central part of the total range of concentrations (0–2.2 percent). Therefore, a log transformation could be judged to be an appropriate step toward achieving a distribution closer to the normal form.

If the data had shown a significant departure from the lognormal form, a "STOP" sign would have been

encountered in figure 4. Because the data are in broad geometric classes, the  $\log(y + \alpha)$  transformation would have been awkward, and no method is available that could have been used to obtain a precise estimate of abundance.

#### URANIUM IN GRANITE

The data represented in histogram C (fig. 1) were originally from Coulomb (1959), but were reproduced by Hubaux and Smiriga-Snoeck (1964, p. 1207) and used in testing a digital-computer method for deriving means and standard deviations of censored distributions. Hubaux and Smiriga-Snoeck derived mean logarithms rather than arithmetic means or abundance estimates. The data represent uranium concentrations in samples from a homogeneous granitic massif; we shall assume that the sampling was unbiased.

The frequency distribution of the original data (fig. 1C) exhibits a clear positive skewness, and the data, therefore, were transformed logarithmically (fig. 1D). The frequency distribution, as noted by Hubaux and Smiriga-Snoeck (1964, p. 1206), is only imperfectly lognormal. The effectiveness of the log transformation in bringing the data closer to the normal form can be

seen by comparing the probability graph for distribution  $C$  with that of distribution  $D$  (fig. 5).

If all the data ( $n=185$ ) are used, the geochemical abundance of uranium in the granitic massif is estimated, as the arithmetic mean (eq. 1), to be 4.53 ppm (table 1, col. 17).

In reference to figure 4:

- [1] The uranium data are in arithmetic classes. Proceed to [2].
- [2] The frequency distribution of uranium analyses is unimodal and positively skewed; analytical discrimination is satisfactory (figs. 1C and 5C). Proceed to [4].
- [4] The frequency distribution of the logarithms of the uranium analyses is approximately normal (distribution  $D$ ; in figs. 1 and 5). Proceed to [5].
- [5] If the complete distribution is used ( $n=185$ ), the abundance, derived with the  $t$  estimator of Sichel (1952), is 4.54 ppm, virtually the same as the value, 4.53, derived by the conventional procedure.

After censoring the data at the arbitrary analytical cutoffs 2.6, 4.0, and 5.1 ppm and effectively discarding 12, 43, and 73 percent of the data, respectively, the abundance estimates, as derived with the  $t$  estimator, are 4.55, 4.54, and 4.24 ppm (table 1, col. 17).

Had the decision been made to use the original data rather than to transform them logarithmically (that is, had the distribution represented by histogram  $C$  in figure 1 been judged approximately symmetrical), the computation method would have proceeded as in the previous example for  $\text{MoS}_2$  assays. Censoring the data, then, at 2.6 and 4.0 ppm would result in abundance estimates of 4.48 and 4.11 ppm, respectively (table 1, col. 17). These are somewhat poorer than the corresponding values of 4.55 and 4.54 obtained by way of the  $t$  estimator, and the importance of the log transformation is apparent.

The problem of judging, from censored data, whether or not the total concentration values (fig. 1C) exhibit a symmetrical frequency distribution is not as difficult as for  $\text{MoS}_2$  assays in the previous example. Because only 27 percent of the uranium values are above 5.1 ppm, for example, the median of all the data is known to occur between 0 and 5.1 ppm. As the data extend to more than 14 ppm some positive skewness is evident. However, judging the degree to which the log transformation corrects the skewness is more difficult. Where a large proportion of the data is censored but the remainder is sufficient to indicate definite positive skewness in the original data, the log transformation might

be used as an approximation. Where a lesser proportion of the data is censored, the transformation using  $\log(y+\alpha)$  might be used if needed, and thereby offer a means for improving the accuracy of abundance estimates.

#### ARSENIC IN BASALTS AND DIABASES

The data with extreme positive skewness, represented by histogram  $E$  in figure 1, are from Onishi and Sandell (1955, tables 5, 10) and were used by Ahrens (1957, p. 207-209) in a discussion of frequency distributions of minor elements in igneous rocks. The data—arsenic determinations, in parts per million—were obtained on 58 samples of basalt and diabase from widely separated localities in the United States and from localities in Japan and Sicily. For purposes of illustrating the computational techniques we shall assume that the samples adequately represent the rock bodies from which they were taken.

The geochemical abundance of arsenic in the basalts and diabases, estimated by the conventional method (eq 1), is 1.69 ppm (table 1, col. 17).

In reference to figure 4:

- [1] The arsenic data are in arithmetic classes. Proceed to [2].
- [2] The frequency distribution is unimodal and positively skewed; analytical discrimination is satisfactory. Proceed to [4].
- [4] The frequency distribution of the logarithms of the arsenic analyses (distribution  $F$  in figs. 1, 5) retains a definite positive skewness. Proceed to [6].
- [6] If the methods described by Krige (1960, p. 236) are used, the appropriate constant,  $\alpha$ , is estimated to be  $-0.6$ . (Had the skewness of distribution  $F$  in fig. 1 been negative, the estimated constant would have been a positive value.) Eight of the analytical values are equal to or less than 0.6, and as  $\alpha$  is negative, the quantity  $\log(y+\alpha)$  for these values is undefined; the transformed data (distribution  $G$  in fig. 1) are censored with  $h=\%_8=0.14$ . The probability graph for the newly transformed data, drawn using  $\log(y-0.6)$ , is shown in figure 5 (distribution  $G$ ). Neither the histogram nor the probability graph indicates a large departure of the transformed data from the censored normal form.

If Cohen's technique for censored distributions (eq 5, 6) and Krige's  $t\alpha$  estimator are used, the abundance estimate for arsenic in the basalts and diabases is 1.77 ppm (table 1, col. 17). Because the number of analyses in this example is small ( $n=58$ ) and the data are highly skewed, it may

be argued that the estimate of 1.77 ppm is better than the estimate of 1.69 derived from the conventional procedure. However, the two estimates are in at least fair agreement.

When censoring the data at the arbitrary cutoffs, 1.0 and 1.6 ppm, and effectively losing 38 and 72 percent of the data, respectively (table 1, col. 8), the abundance estimates derived with the  $t_\alpha$  estimator are 1.68 and 1.75 ppm (table 1, col. 17).

Had the data actually been censored below 1.6 ppm, estimation of the constant  $\alpha$  may have been difficult, but some useful estimate might have been made unless the point of censoring was lower than about 1 ppm. Whether the point of censoring had been at either 0.7, 1.0, or 1.6 ppm, a high positive skewness would have been apparent from the fact that the median lies well below the central part of the known range of the data—0–10 ppm. If the log transformation is used (without adjusting the data by the constant  $\alpha$ ) and all the data are used, the abundance estimate is 1.58 ppm. If only data equal to or greater than 0.7, 1.0, and 1.6 ppm are used, the abundance estimates are 1.58, 1.59, and 1.51 ppm, respectively (table 1, col. 17). These estimates are notably poorer than those derived using the  $t_\alpha$  (rather than  $t$ ), estimator, but they are, nevertheless, sufficiently good for many types of geochemical studies.

### CONCLUSIONS

Computational methods given by Cohen (1959, 1961), Sichel (1952, 1966), and Krige (1960) are useful in providing accurate and efficient estimates of geochemical abundance in many problems where the conventional method for estimating population arithmetic means is not applicable (owing to censored data) or is inefficient (because of small data sets from nonnormal distributions). A combination of the methods may be useful where censored data are from an underlying frequency distribution that is nonnormal. Each of the methods requires that the data, or transformations of the data, be normally distributed. Two transformations that have proven satisfactory in much geochemical work are  $\log y$  and  $\log(y + \alpha)$ . Neither transformation is effective, however, where analytical discrimination has been poor. Moreover, selection of the appropriate transformation may be difficult where a large proportion of the distribution has been censored.

The methods discussed are recommended for estimating geochemical abundance primarily because they are the most efficient methods available. Indeed, the methods developed by Sichel and Cohen are the most efficient possible where the frequency distribution requirements are satisfied. Those developed by Sichel

and Krige, moreover, have been corrected for a small bias that exists where  $n$  is small. The method of Cohen does contain some bias where  $n$  is small, but the bias is probably much less than that introduced by most arbitrary methods used in handling the censored-data problem.

### LITERATURE CITED

- Ahrens, L. H., 1957, Lognormal-type distributions, [Pt.] 3 of The lognormal distribution of the elements—a fundamental law of geochemistry and its subsidiary: *Geochim. et Cosmochim. Acta*, v. 11, no. 4, p. 205–212.
- Aitchison, John, and Brown, J. A. C., 1957, The lognormal distribution, with special reference to its uses in economics: Cambridge Univ. Press, 176 p.
- Barnett, P. R., 1961, An evaluation of whole-order,  $\frac{1}{2}$ -order, and  $\frac{1}{4}$ -order reporting in semiquantitative spectrochemical analysis: U.S. Geol. Survey Bull. 1084-H, p. 183–206.
- Cohen, A. C., Jr., 1959, Simplified estimators for the normal distribution when samples are singly censored or truncated: *Technometrics*, v. 1, no. 3, p. 217–237.
- , 1961, Tables for maximum likelihood estimates; singly truncated and singly censored samples: *Technometrics*, v. 3, no. 4, p. 535–541.
- Coulomb, René, 1959, Contributions à la géochimie de l'uranium dans les granites intrusifs: France, Centre de Études Nucléaires de Saclay, rap. C.E.A. 1173.
- Dixon, W. J., and Massey, F. J., Jr., 1957, Introduction to statistical analysis: 2d ed., New York, McGraw-Hill Book Co., 488 p.
- Finney, D. J., 1941, On the distribution of a variate whose logarithm is normally distributed: *Jour. Royal Statistical Soc., Suppl.* 7, no. 2, p. 155–161.
- Fisher, R. A., 1950, Statistical methods for research workers: 11th ed., New York, Hafner Publishing Co., 354 p.
- Fleischer, Michael, and Chao, E. C. T., 1960, Some problems in the estimation of abundances of elements in the earth's crust: Internat. Geol. Cong., 21st, Copenhagen 1960, Rept., pt. 1, p. 141–148.
- Hald, A., 1952, Statistical tables and formulas: New York, John Wiley & Sons, Inc., 97 p.
- Hazen, S. W., Jr., and Berkenhotter, R. D., 1962, An experimental mine-sampling project designed for statistical analysis: U.S. Bur. Mines Rept. Inv. 6019, 111 p.
- Hubaux, A., and Smiriga-Snoeck, N., 1964, On the limit of sensitivity and the analytical error: *Geochim. et Cosmochim. Acta*, v. 28, no. 7, p. 1199–1216.
- Huff, L. C., 1955, A Paleozoic geochemical anomaly near Jerome, Arizona: U.S. Geol. Survey Bull. 1000-C, p. 105–118.
- Krige, D. G., 1960, On the departure of ore value distributions from the lognormal model in South African gold mines: South African Inst. Mining Metallurgy Jour., v. 61, no. 4, p. 231–244.
- Miesch, A. T., 1963, Distribution of elements in Colorado Plateau uranium deposits—a preliminary report: U.S. Geol. Survey Bull. 1147-E, p. E1–E57.
- Myers, A. T., Havens, R. G., and Dunton, P. J., 1961, A spectrochemical method for the semiquantitative analysis of rocks, minerals, and ores: U.S. Geol. Survey Bull. 1084-I, p. 207–229.

- Onishi, Hiroshi, and Sandell, E. B., 1955, Geochemistry of arsenic: *Geochim. et Cosmochim. Acta*, v. 7, nos. 1 and 2, p. 1-33.
- Sichel, H. S., 1947, An experimental and theoretical investigation of bias error in mine sampling, with special reference to narrow gold reefs: London, Inst. Mining Metallurgy Trans., v. 56, p. 403-474.
- Sichel, H. S., 1952, New methods in the statistical evaluation of mine sampling data: London, Inst. Mining Metallurgy Trans., v. 61, p. 261-288.
- , 1966, The estimation of means and associated confidence limits for small samples from lognormal populations: South African Inst. Mining and Metall. Jour., preprint no. 4, 17 p.

