Applications of

PROBABILITY GRAPHS

in

Mineral Exploration

by

Alastair J. Sinclair

Special Volume No. 4

THE ASSOCIATION OF EXPLORATION GEOCHEMISTS

Applications of

PROBABILITY GRAPHS

in

Mineral Exploration

by

Alastair J. Sinclair

Dept. of Geological Sciences

University of B.C.

Vancouver, B.C.

Special Volume No. 4

THE ASSOCIATION OF EXPLORATION GEOCHEMISTS

RICHMOND PRINTERS LTD.
583 Minoru Boulevard, Richmond, B. C., Canada

PREFACE

Cumulative probability plots are applicable to a wide variety of problems in many different fields. The present manual has been prepared principally for use by persons concerned with mineral exploration but the techniques covered should find much wider application. Text has been organized with the thought that the best introduction to a subject is a clearcut idealized approach followed by practical examples. Throughout, the author has attempted to use practical examples with which he is acquainted personally. Hopefully such provincialism will not obscure the more general usage of the methods proposed.

This manual evolved over the past few years as the writer became more and more involved in statistical analysis of many types of mineral exploration data. Three references (Tennant and White, 1959; Lepeltier, 1969; Bolviken, 1971) have been particularly instrumental in the development of the authors views and due acknowledgement is hereby made. Some sections of the manual have, at one time or another, served as a basis for lectures within the University and to Industry. The constructive response of many attendees of these lectures is greatly appreciated. Various research projects supported by the National Research Council of Canada and the Department of Energy, Mines and Resources have contributed to many of the ideas and examples cited in the text. Financial assistance from the Department of Energy, Mines and Resources in the form of a research agreement made preparation of the manuscript possible.

The manuscript has benefitted substantially from the constructive comments of R. G. Garrett, A. T. Miesch, A. W. Rose and R. F. Horsnail, members of the Computer Applications Committee of the Association of Exploration Geochemists, as well as my colleagues W. K. Fletcher and J. H. Montgomery. All errors and omissions, however, remain the responsibility of the author.

Technical assistance has been provided at various time by A. C. L. Fox, Asger Bentzen and J. F. W. Orr. Some of the geophysical examples used were kindly supplied by Mr. D. R. Cochrane. Illustrations were draughted by M. WaskettMyers and typing of various stages of the manuscript was done by Mrs. Joan Mullen and Mrs. Charlotte Heywood. I am grateful to all these people for their contributions.

Alastair J. Sinclair

Vancouver, B. C. Canada *February 19, 1976*

TABLE OF CONTENTS

CHAPTER I
INTRODUCTION

I-0: GENERAL STATEMENT

The use of probability plots requires only a general understanding of simple statistical concepts, with which the reader is assumed to be familiar. Such basic terms as arithmetic mean, variance, standard deviation, normal density distribution, etc. will not be dealt with explicitly and, if necessary, reference should be directed to any of a large number of introductory statistical texts.

An introduction to probability paper is deferred until Chapter II. In Chapter I several topics are considered that relate either to probability plots themselves or to the specific applications to which probability plots are put in this manual. Organization of the material is aimed at showing a simple progression from the wellknown histogram to probability plots.

I-1: MINERAL EXPLORATION DATA

Many variables measured in the course of mineral exploration programs are continuous or approximately so. For example, 700 soil samples analyzed for copper might provide values in the range 10 to 1000 ppm Cu. The variable (Cu in soil) is continuous between these limits because, theoretically at least, any intermediate value could be assumed by a sample. In practice, of course, a value would never be reported as 927.341 for example. In contrast to continuous variables are those referred to as discrete, that is, those variables that take only specific values. For example, the number of minerals in a polished section of an ore is a discrete variable. There might be 1, 2, 3, etc. minerals in a given polished section but not 1.372 mineral types. Here we confine our attention to those variables that are continuous or nearly so.

There are many approaches to the analysis of continuous, quantitative variables obtained during an exploration program and only those pertinent to a discussion of probability graphs will be considered. The most common method is subjective geological analysis of tabulated data or data that has been contoured empirically. Such an approach to data interpretation is being accompanied with increasing regularity by statistical analysis of varying degrees of complexity. Even the construction of a simple histogram can be considered an initial "statistical" step. Whatever the complexity of a statistical study of data, one can be assured that it will involve calculation of means and standard deviations, and probably some method of graphical representation of data. Calculations of means and more particularly standard deviations, become more arduous as the amount of data increases if manual methods are used. The practical difficulty of obtaining estimates of these parameters in the field commonly seems insurmountable. Fortunately, as we shall see, the use of cumulative probability paper overcomes these difficulties. to some extent.

Generally speaking a set of values represents a statistical sample of a population. Five hundred values obtained from a magnetometer survey of regular grid intersections represents one sampling of values. The grid might have started at some other point, for example, and a somewhat different set of 500 values would result. There are an infinite number of such statistical samples that might have been obtained, of which our one sample. (of <500 values) is but one realization. It is important to note the difference heree between statistical and geological usages of the term sample. A single soil sample is a. sample in a geological context, but it is simply one element in the statistical sample.

I-2: HISTOGRAMS

Histograms are a familiar method of displaying numerical information. Figure I-1 shows three histograms illustrating common variations in form that occur in the, case of populations encountered with mineral exploration data, ore grades in this case. Negatively skewed [figure I-1(a)], symmetric [figure I-1(b)], and positively skewed [figure, I-1(c)] density distributions are illustrated. Some obvious advantages of histograms as a means of visual representation of data are: (1) total range of data in a sample is apparent, (2) modes can be recognized easily, (3) the range of greatest abundance of values can be estimated rapidly, and (4) the general form of the density distribution of data is apparent. In some cases histograms are useful in distinguishing a threshold, between background and anomalous values, a purpose for which they have, found fairly, routine use. An additional advantage of an histogram is, that the preparatory grouping of data provides a relatively convenient form for calculating the mean and variance by the method of grouped data.

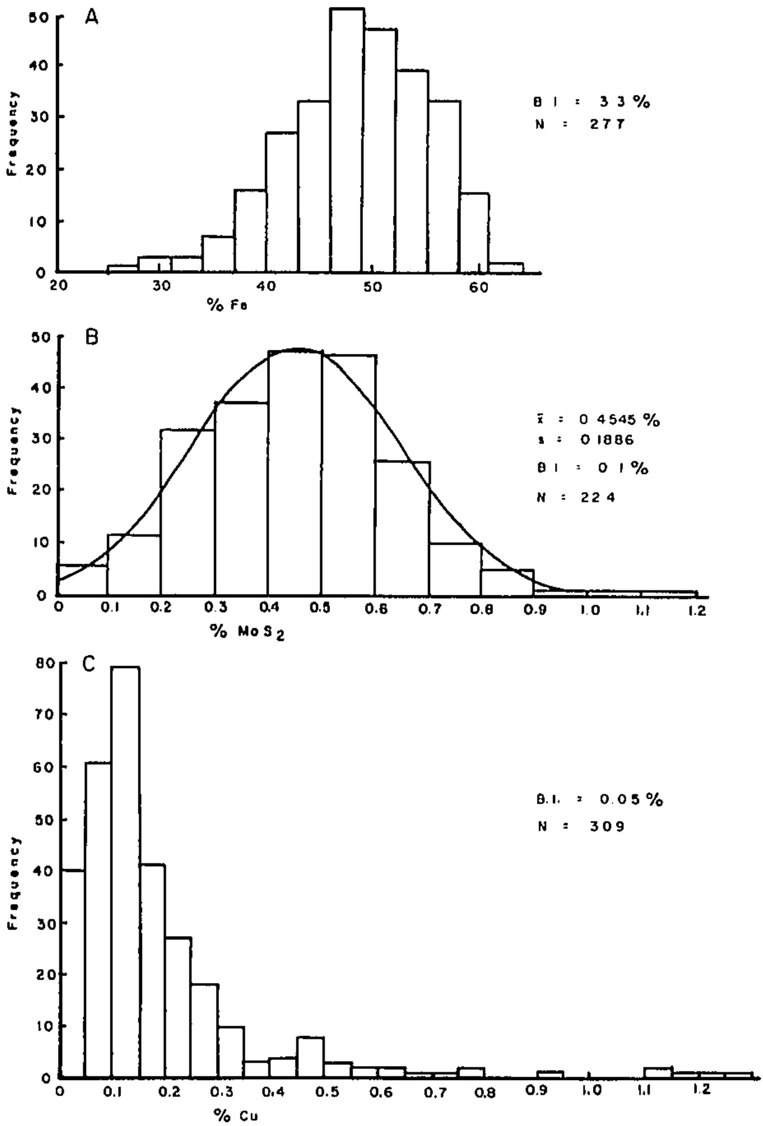
Our data must be of appropriate quality for the purpose on hand they must be representative (unbiased) and the measuring technique used to obtain the "numbers" must have adequate precision. Optimum data-collecting methods can be based on an orientation survey, a common procedure in many extensive geochemical and geophysical surveys.

There are several points concerning the construction of an histogram that are worthy of mention. First is the choice of class interval. According to Shaw (1964) a class interval is best chosen between one quarter and one-half the standard deviation of the data. If the class interval is too great the true form of the distribution is masked – if too small then too many gaps appear in the resulting histogram.

A second point in the construction of an histogram is the question of where to start the first class. The choice is not a serious matter as a rule but it seems reasonable to standardize the procedure, and have either one or two classes disposed symmetrically with respect to the mean value.

FIGURE I-1

Three examples of real density distributions of ore grades in histogram form showing (a) negative skewness, (b) symmetry, and (c) positive skewness. B.I. is class interval, N is sample size. A normal curve with the same mean and standard deviation as the real data has been fitted to the symmetric distribution (after Sinclair, 1972).



An histogram should be constructed with the ordinate (frequency) as a percentage if comparison is to be made with other histograms. Otherwise, it is awkward or impossible to make a meaningful visual comparison of two or more histograms, each based on substantially different numbers of values. In general, in construction of histograms it is good practice to include a listing of (1) title, (2) N, the sample size, (3) the class interval, and (4) the mean and standard deviation of the data.

I-3: CONTINUOUS DENSITY DISTRIBUTIONS

As the class interval of an histogram decreases for large samples, it is apparent that it becomes easier and easier to pass a smooth continuous curve through the tops of the classes. Consequently, it is possible to approximate many frequency distributions of continuous variables by a smooth mathematical curve known as a density distribution [see figure 1-1(b)]. One might imagine that many such mathematical models would be required to take into account all potential distributions of real data and while this might be true in theory it is fortunately not so in practice. A majority of variables in nature exhibit shapes of frequency distributions that can be approximated by a relatively small number of mathematical models. In fact, we will confine our attention in this manual to two specific forms, the normal and lognormal distributions, and will attempt to justify this position later in the present chapter. One must bear in mind that numerous discrete and continuous density distribution models are in use, including binomial, poisson and gamma, among others. For reference to these the reader should consult standard statistical texts.

The normal or Gaussian distribution was first put forward as a theory of error measurement. For example, we might wish to test the reproducibility of a chemical method of analyzing soils. A single well mixed sample might be divided into 10 sub samples, each of which is analyzed using the same method. The 10 values obtained will not necessarily be exactly the same due to random variations in analytical procedure. The distribution of measured values about the mean follows what is known as the normal or Gaussian density distribution given by the following formula:

$$y = 1\frac{1}{\sigma\sqrt{2\pi}}\ e^{-1/2\,(x-\mu)\,^2/\sigma^2}$$

where μ is the arithmetic mean, x is any measurement, and σ^2 is the variance of the population. The graphical expression is the familiar bell-shaped curve shown in Figure I-1(b).

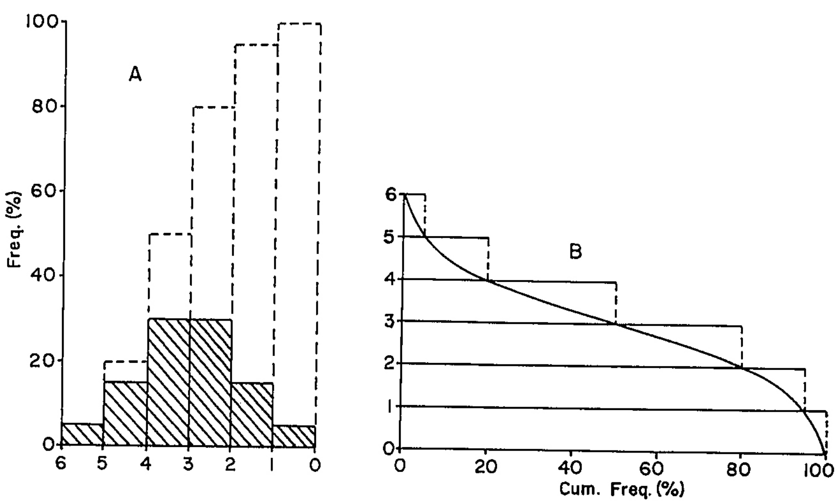
I-4: CUMULATIVE DISTRIBUTIONS

Data prepared for a standard histogram can also be presented as a cumulative histogram in which, to the frequency within any class is added the total frequencies of all preceding classes. Frequencies can be cumulated from either the high or low end of the range of values. This method of representation is common in the field of sedimentology for sediment size fractions, where frequencies are cumulated from coarse-grained to fine-grained fractions. An example is shown in figure I-2(a) where the relationship between a standard histogram and a cumulative histogram is apparent. The cumulative histogram and a smooth cumulative curve are repeated in figure I-2(b) but with coordinates reversed. In this hypothetical example data were cumulated from high values to low.

This method of representation of data is in the required form for plotting on probability graph paper (refer to section II-1). The concept of a cumulative histogram is straightforward and is fundamental to an understanding of probability plots.

FIGURE I-2

(a) Histogram and cumulative histogram of an hypothetical data set. Note that data are cumulated from high to low values. (b) The cumulative histogram of (a) is repeated but with coordinates reversed. A continuous curve (cumulative curve) has been fitted to the histogram. This is the graphic form of histogram that correlates most obviously with probability plots as used in this manual. Note that by an appropriate change in scale the cumulative curve can be transformed to a straight line!



I-5: LOGNORMAL DISTRIBUTIONS

In its simplest conceptual form the lognormal distribution is viewed as a normal distribution of the logarithms (to any base) of a set of data. It has been described in detail by Aitchison and Brown (1957) and can form a basis for theory of multiplicative errors just as the normal distribution is the basis of a theory of additive errors.

An extensive literature exists concerning the shape or form of density distributions of natural data (e.g. Ahrens, 1954; Shaw, 1961; Becker and Hazen, 1961; Rodionov, 1961). Numerous authors have concluded that many earth science variables of the type we are considering here have density distributions that are closely approximated by the lognormal law. For example, a close approach to lognormality is shown commonly by such variables as (1) minor elements in geochemistry (e.g. Shaw, 1961) (2) many geophysical variables (e.g. Slichter, 1955), (3) grades and tonnages of mineral deposits (e.g. Sinclair, 1974b), (4) sediment size data (e.g. Harris, 1958), (5) capacity of water reservoirs (e.g. Hazen, 1914), (6) sizes of oil pools (McCrossan, 1969), and so on. In addition, some variables dealt with routinely by earth scientists show normal distributions, but the very nature of the variables incorporates a log transformation. pH measurements and sediment size data (in phi units) are everyday examples.

Having emphasized the general support in the literature of lognormal distributions of an important proportion of earth science variables and the close approximation to lognormal density distributions shown by many variables measured routinely in mineral exploration, it is important to be aware of some of the vaguaries and complications to such a simple interpretation. For example, no fundamental "lognormal law" has been stated or should be implied. The lognormal model is merely an adequate approximation of reality! It is important to realize that real data depart most from fitted empirical models at the tails of the fitted density distributions. Exactly where a given model no longer applies to real data is difficult to determine but one can be fairly certain that a lognormal model, for example, cannot be applied with assurance beyond the range of data on which the model is based. Furthermore, distributions other than lognormal are encountered (e.g. Govett, 1975) and one must be constantly aware of such a possibility.

Perhaps the most serious problem in reconciling a lognormal model with much real data is encountered with polymodal distributions. Where component populations do not overlap appreciably each can be examined individually for lognormality. The frequency of occurrence of such polymodal lognormal distributions indicate that it is logical to expect overlapping populations to also approximate lognormal models. Certainly, this latter interpretation has worked out well in practice (e.g. Montgomery et al, 1915; Saager and Sinclair, 1974). Throughout this manual examples will be given that provide a test for the presence of combinations of lognormal populations.

Without entering a lengthy philosophic debate the author concludes that many variables commonly met in the earth sciences, particularly those measured routinely during mineral exploration programs, show a close approach to lognormality, or to a combination of two or more lognormal populations. Many others approximate normal density distributions. From a practical point of view this means that much data can be presented and interpreted usefully on standard logarithmic or arithmetic probability graphs.

CHAPTER II
CUMULATIVE PROBABILITY PAPER AND SINGLE POPULATIONS

II-0: GENERAL STATEMENT

Cumulative probability paper was first developed by Hazen (1914) as a means of simplifying the interpretation of reservoir storage data by eliminating the sharp curvature present in graphs with an arithmetic cumulative percentage scale. The method has since found fairly widespread applications in graphical study of numeric data in many fields. Numerous publications exist describing the applications and limitations of such plots (e.g. Goodrich, 1927; Rissik, 1941; Levi, 1946; Harding, 1949; and Cassie, 1954) but the technique has been widely publicized and used extensively in the earth sciences only recently.

Sedimentologists were the forerunners in the application of probability plots to geological problems – (e.g. Krumbein and Pettijohn, 1938; Otto, 1939; Tanner, 1958). Tennant and White (1959) appear to have been the first to recognize the potential of cumulative probability graphs in treating geochemical data, although Williams (1967) provides the first comprehensive treatment of the subject. More recently, Lepeltier (1969) has presented a succinct outline of the use of probability paper in interpretation of geochemical data with special reference to rapid, field-oriented procedures. Despite the abundant publicity relating to probability graphs and their applications, the method has been used in mineral exploration only spasmodically, but with somewhat increasing regularity, over the past few years.

According to Bolviken (1972), analysis by probability graph paper is a routine procedure used by the Geological Survey of Norway as an aid to interpreting various kinds of geochemical information. Published applications to other types of geological and/or mineral exploration data have not been widespread (with the exception of sedimentological studies), but certainly various kinds of geological and geophysical data are amenable to such analysis. Other potential applications include studies of size (tonnages) or average grades of mineral deposits (Sinclair, 1974b), capacity of gas reservoirs (See McCrossan, 1969), and innumerable others.

II-1: PROBABILITY GRAPH PAPER

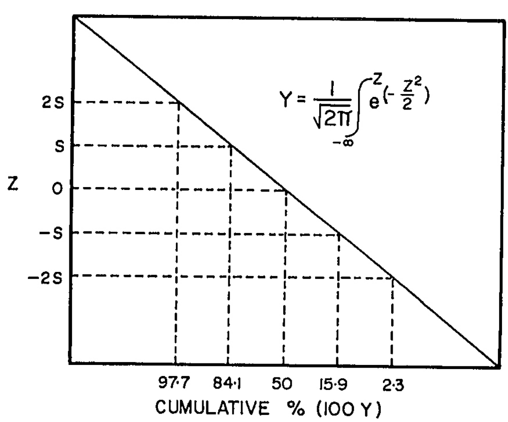
Standard cumulative probability paper, commonly referred to as arithmetic probability paper, has an arithmetic scale (the ordinate scale on most paper available in North America) and a somewhat unusual percentage (or probability) abscissa scale. Ordinate and abscissa commonly are reversed on probability paper available outside North America and for that used by sedimentologists. The probability (cumulative percentage) scale is arranged such that a cumulative normal density distribution will plot as a straight line. Derivation of the probability scale can be seen in figure II-1. It is apparent that the ordinate is arithmetic in terms of "numbers of standard deviations" from a central reference value, zero, that corresponds to the mean value of the distribution. A sloping straight line (any line) is then drawn in as a means of defining the probability scale by projecting known cumulative percentages to the probability scale. The cumulative percentage to be assigned to any point projected onto the probability scale is the value 100 y from the equation

$$y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{(-\frac{z^2}{2})}$$

where y represents the proportion of area (or values) in a normal standard density distribution that is below the specified standardized value "z". Hence, any percentage value on the probability scale can be determined by recourse to the foregoing integral equation. It is not necessary to work out such integrals because y and z values are tabulated in most elementary statistical texts. Note that the probability scale can be expanded or contracted by decreasing or increasing the slope of the line used in figure II-1 to project points to the probability scale.

FIGURE II-1

Diagram illustrating the direct relationship between Z values of a standardized normal distribution and cumulative percentage of the `probability" scale.



Although a graphical approach has been used to give an insight into the basis of probability paper, it is apparent from figure II-1, that cumulative percentage values on the probability scale are directly proportional to z values. Consequently, probability scales can be determined simply by multiplying z values by any appropriate constant. In practice, of course, probability graph paper with a variety of scales can be purchased. It is useful, however, to have some appreciation of the derivation of the probability scale if probability graph paper is to be used extensively.

A second type of probability paper is in common use, *logarithmic* probability paper. Log and arithmetic probability papers have comparable probability scales; the former, however, has a logarithmic scale for the second axis whereas the latter has an arithmetic scale. The logarithmic scale is to base 10.

Throughout the literature, probability graphs are presented using the probability scale as either ordinate or abscissa, and cumulating percentages beginning either with high or low values of a data set. Hence, there are four possible manners of presentation, all of which have been used by various authors. Throughout this manual the writer has chosen to standardize his presentation of probability plots. The probability scale is chosen as abscissa because most commercial probability graph paper readily available in North America is constructed in this manner. In addition, frequencies (probabilities) are cumulated from high towards low values. As Lepeltier (1969) has pointed out, this procedure avoids having the last, unplottable point (100 cumulative percent) at the high value end, which is commonly in the range of most interest in mineral exploration data. Furthermore, there are instances when such a procedure provides an additional point for plotting compared with the reverse procedure of cumulating from low towards high values (e.g. see section VIII-3).

II-2: ANALYSIS OF SINGLE POPULATIONS

The foregoing section shows that a single, cumulative, normal population plotted over the full probability range, defines a straight line on arithmetic paper because of the nature of the probability scale. Similarly, cumulative percentages of arithmetic values of a lognormal population define a straight line on log probability paper. Conversely, cumulative percentages of *logarithms* of a lognormal population define a straight line when plotted on arithmetic probability paper.

In practice, one plots precisely the same data that would be used to construct a cumulative histogram. The points thus obtained would only coincidently plot exactly on a straight line. In general, there would be a certain amount of scatter produced by expected sampling error. For the present, our discussion will be confined to *ideal* normal populations.

FIGURE II-2

Normal distributions with different mean values but identical standard deviations plot as a set of parallel straight lines on probability paper. Means can be read where the straight lines intersect the 50 percentile. Two standard deviations are estimated by the difference between readings at the 16 and 84 percentiles.

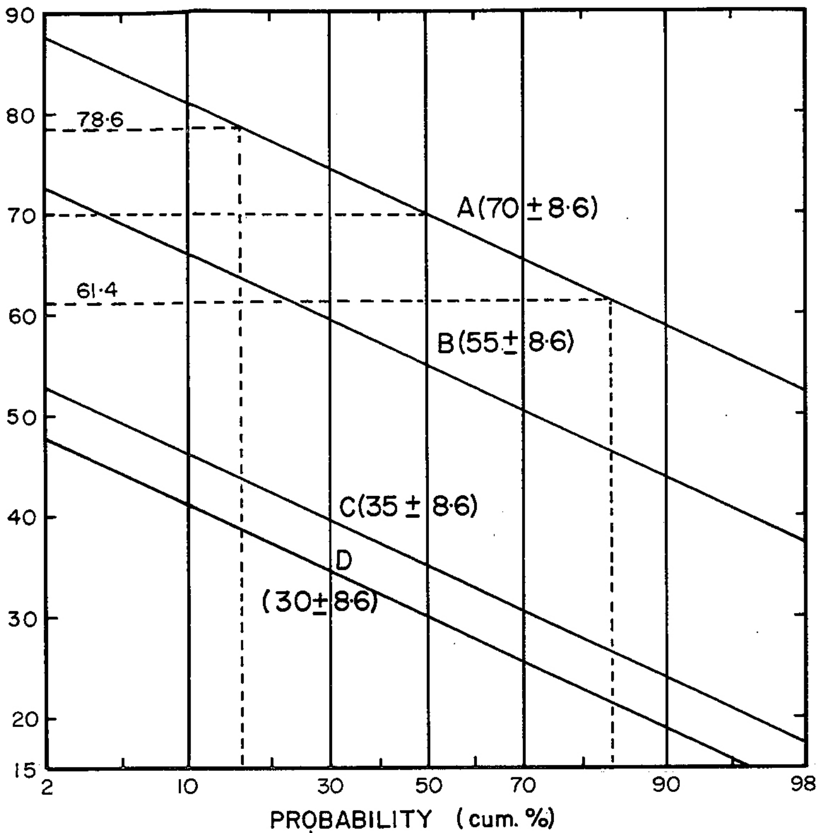


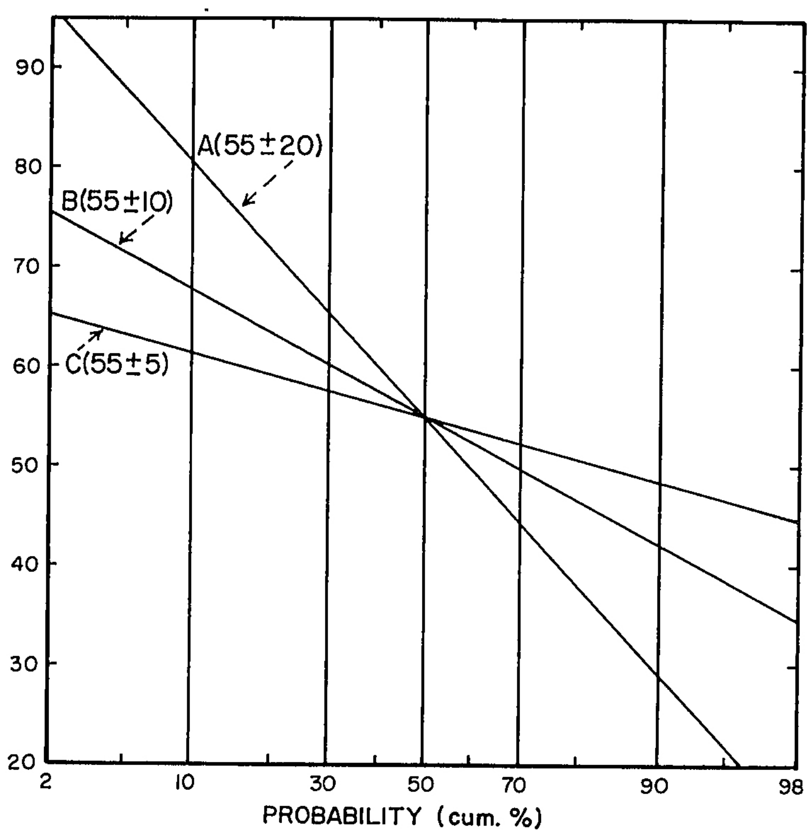
Figure II-2 shows several hypothetical populations, each represented by a straight line. All lines have the same slope but different mean values. The mean value estimate of each population can be read as the ordinate value corresponding to the 50 percentile. Because each of the populations 'is normal (a straight line plot on probability paper) the values of the mean, plus and minus one standard deviation, can be estimated fairly precisely by ordinate values that correspond to the 16 and 84,percentiles respectively. For example, in population A, the mean (x) is read at the 50 percentile as 70, and ($\bar{x}$ + s) and ($\bar{x}$ - s) can be read at the 84 and 16 cumulative percentiles as 78.6 and 61.4 respectively.

Consequently, estimates of parameters of population A are 70 ± 8.6. In practice, an estimate of 2s is obtained as the positive difference between values of the 84 and 16 cumulative percentiles. This difference is halved to estimates. Note that mean and standard deviation of normal distributions are quoted in this manual as $\bar{x}$ ±s, a form that should not be confused with representation of the standard error of the mean!

Three populations are shown as straight lines in figure II-3 that all intersect the 50 percentile line at an ordinate value of 55, indicating that each of the populations has the same mean value. Standard deviations of the three populations differ as do the slopes of lines that define the populations. In general, for data plotted to the same scale, a steeply sloping linear pattern indicates a relatively large standard deviation compared to a more gently sloping linear pattern.

FIGURE II-3

Normal distributions with the same mean but different standard deviations plot on probability paper as a set of straight lines that intersect at the 50 percentile.



In dealing with lognormal distributions, two approaches are possible:

1) Plot logarithmic values on arithmetic probability paper
2) Plot untransformed values on log probability paper.

The first method is used routinely in sedimentology where sediment size fractions are quoted in phi values (logarithms to base 2 of sieve diameters). Estimation of parameters ($\bar{x}$ and s) is done in the manner described previously. These estimates, however, refer to log values.

Ordinarily, with large numbers of values it is time-consuming and inconvenient to transform arithmetic values to logarithms, in which case the cumulative arithmetic data are plotted directly on log probability paper. This procedure avoids reference to tables of logarithms because the transformation is supplied graphically and automatically on plotting, as a result of the logarithmic ordinate scale. The estimated parameters read from a linear graph on log probability paper are *antilogs* of:

1) arithmetic mean of the logarithms of values
2) arithmetic mean of logarithms plus one standard deviation, and
3) arithmetic mean of logarithms minus one standard deviation.

Consequently, the mean value determined is the *geometric mean* of the original data and the two surrounding values that encompass approximately 68 percent of values in the distribution are located *asymmetrically* about this geometric mean. Throughout this manual the form adopted to designate lognormal distributions will be the geometric mean followed in brackets by antilogs of:

1) mean of logarithms plus one standard deviation, and
2) mean of logarithms minus one standard deviation.

As an example, consider the population 100 (250,40). Logarithms of these three values are 2, 2.3979 and 1.6021. These are equivalent to 2 ± 0.3979 and the symmetry of the distribution in log units becomes apparent.

Symbols used throughout this text are summarized in table II-1.

TABLE II-1
STATISTICAL SYMBOLS

Population	Example *	Symbol	Example
Arithmetic	100 ± 50	$\bar{x}$	100
Normal		s	50
Lognormal	2.00 ± 0.153	$\bar{x}$	2
(logarithmic values)		s	0.153
Lognormal	100 (142,70)	b	100
(antilogs)		$b + s_L$	142
		$b - s_L$	70

* Examples are those plotted in figures II-4 and II-5.
N = no. of data values that defines a population.
n_i = no. of values in subpopulation i where a density distribution is polymodal.
$\sum n_i = N$

II-3: NORMAL DISTRIBUTIONS PLOTTED ON LOG PROBABILITY PAPER (AND VICE VERSA)

It is useful to know the forms of cumulative curves that result if a normal population is plotted inadvertently on log probability paper (or vice versa). This is an obvious aid to interpretation in the case where data do not plot as straight lines.

Consider an arithmetic normal distribution (AN) with parameters 100 ± 50, shown as a straight line in figure II-4. The same distribution is shown on lognormal probability paper (figure II-5) where a pronounced curved pattern, concave downward, is apparent. Note that the greatest curvature is towards low values. This curve (AN in figure II-5) bears general resemblance to a top-truncated lognormal distribution discussed in the following section; but, as a rule, the two can be distinguished by noting whether greatest curvature is towards high or low values.

FIGURE II-4

The shapes of normal (AN) and lognormal (LN) distributions with comparable parameters are contrasted on arithmetic probability paper. The same two populations are shown in figure II-5.

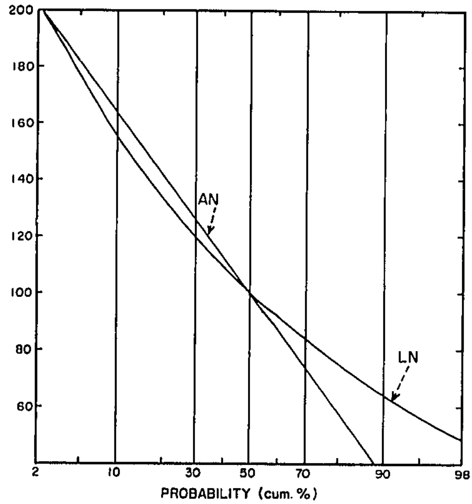
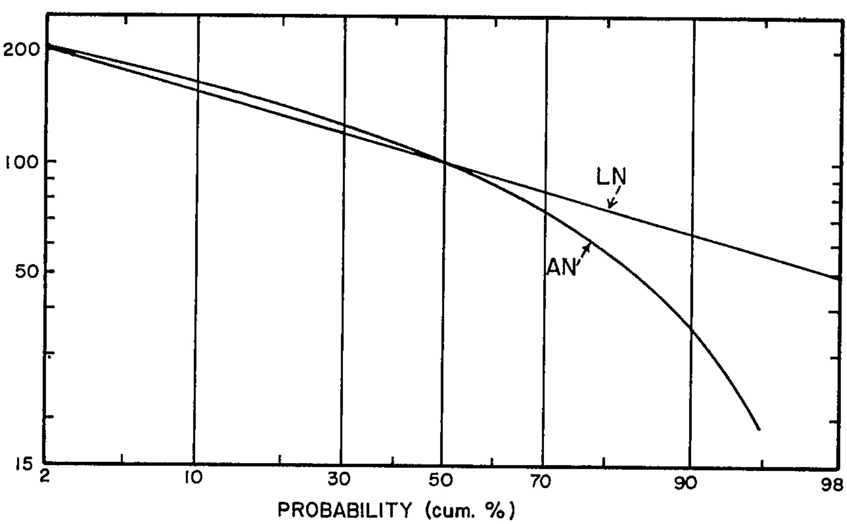


FIGURE II-5

The shapes of normal (AN) and lognormal (LN) distributions with comparable parameters are contrasted on log probability paper. The same two populations are shown in figure II-4.



Consider also a lognormal distribution (LN) with roughly comparable parameters, 100 (142, 70), shown as a straight line in figure II-5. The same lognormal distribution is shown in figure II-4 on arithmetic probability paper where it defines a pronounced curve, concave upward, with greatest steepening towards high values.

To summarize briefly, lognormal distributions plotted on arithmetic probability paper produce curved patterns concave upward; normal distributions plotted on log probability paper produce curved patterns concave downward. In general, concave-up graphs are skewed towards high values, and vice versa.

There are, of course, many possible types of density distributions. In section I-5 it was concluded that normal and lognormal distributions are approximated so commonly in nature that reference here will be confined to patterns arising from these two types of density distributions.

II-4: TRUNCATED DISTRIBUTIONS

A truncated distribution is defined as a density distribution, either normal or lognormal for our purposes here, for which all values above or below some particular value are not represented in available data. A singly truncated distribution thus can be described, as top-truncated or bottom-truncated depending on which end of the distribution is missing. Truncated distributions are not to be confused with censored distributions. A censored distribution is one for which measurements are known for items on one side of a specific value but only *number* of items is known on the other side. Common examples of censored distributions are sets of analytical data for which some values are reported as (a) greater than a particular value, or (b) zero or not detected or less than detection limit. If censored data are ignored in calculating cumulative percentages the resulting plot on probability paper is identical with a truncated distribution. Further discussion will be confined to truncated distributions.

Truncated distributions arise for a variety of reasons, both natural and artificial. Consider, for example, a sediment with a lognormal distribution of. particle sizes. Winnowing during sample collection might remove all particles below a specific size leaving a bottom truncated distribution of the remaining grains. Two examples dealing with ore tonnages are described in sections III-1 and V-3. In some cases, high analytical values have been purposely ignored in reporting data leading to top-truncated distributions from the observer's point of view (e.g. Brabec and White, 1971).

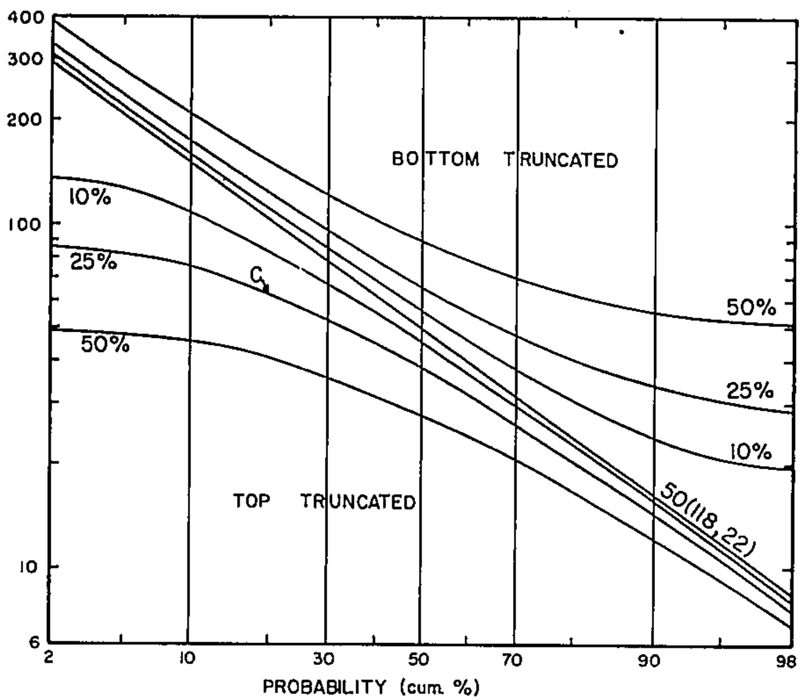
The effect of truncation on a lognormal distribution can be seen best by examining hypothetical truncated distributions as shown in figure II-6. Here, a single population with parameters 50 (118, 22) has been recalculated assuming that various proportions have been removed from its upper (top-truncated) and lower (bottom-truncated) ends. In both cases, curves have been drawn assuming that 10, 25 and 50 percent of the original symmetric population is missing.

Note the general features of these curves:

1) Pronounced curvature or departure from a linear pattern is most evident over the cumulative percentage range that has been truncated.
2) The curvature produces a "flattening" of the cumulative distribution at the end that has been truncated.

FIGURE II-6

Examples of a normal distribution truncated by removing various portions of the population and plotting the remainders of the population over the total probability scale. Percentage figures show the proportion of the upper (top) or lower (bottom) parts of the original population that were removed to reproduce the labelled curves. Point C on one curve indicates how a rough estimate of amount of truncation can be made at a percentile near the point of most rapid change in slope.



In a real situation representing a truncated distribution, recognition of top truncation or bottom truncation is evident from the end of the cumulative curve at which flattening occurs. It is possible, using a trial and error procedure, to estimate fairly precisely the percentage of the population that is missing. This is done by replotting the curved "real data" line assuming various proportions are missing, by exactly the reverse procedure to that used to obtain the curves of figure II-6 from the original straight line plot. Providing that less than 50 percent of a population is missing, a rough estimate of the missing percentage is obtained at the point where a curve begins to flatten pronouncedly. For example, in the curve for a 25 percent top-truncation a point near C (about 20 cumulative percent) divides the curve into three segments, a long relatively steep segment of very gentle curvature, a short segment of rapid curvature, followed by a flattened segment. A first trial assuming 20 percent of the population as missing would lead quickly to the true value of 25 percent missing, in one or two more trials.

The discussion thus far assumes that a truncated population can be recognized by the form of curvature of the probability plot. In practice, some ambiguity can exist. For example, in section II-3 and figure II-5, we have seen that a normal population plotted on log probability paper has a form not unlike that of a top-truncated lognormal distribution. Caution is therefore urged against arbitrarily interpreting curved .graphs such as those of figure II-6 as representing truncated distributions, particularly for top-truncated distributions, unless some reasonable possibility exists that truncation has occurred.

In mineral exploration data, some truncated distributions arise artificially as the result of some data being ignored. On the other hand it is possible that such patterns could arise naturally as the result of a biased sample of a population. One could imagine, for example, a large anomalous geochemical zone in which abundant peak values do not occur in the area sampled.

In a later section it will become apparent that truncated patterns bear vague resemblance to certain kinds of bimodal plots. The two generally can be distinguished by the presence of inflection points in bimodal cumulative curves and the absence of such inflection points in curves of truncated distributions.

CHAPTER III
PRACTICAL EXAMPLES OF SINGLE POPULATIONS

III-0: GENERAL STATEMENT

Thus far the discussion has been confined to ideal, hypothetical density distributions, their representation on probability graph paper, and techniques for extracting information from such plots. In this chapter, real examples will be presented that approximate comparable, ideal distributions. In some cases, information will be extracted that is not particularly pertinent to the problem represented by the specific data set, for the purpose of illustrating techniques developed in chapter II. The reader should bear in mind that some of the more useful applications of probability plots to exploration and other data are in the analysis of combinations of 2 or more density distributions, discussions of which are deferred to later chapters.

III-1: PRODUCTION TONNAGES FROM VEIN DEPOSITS AINSWORTH MINING CAMP, B.C.

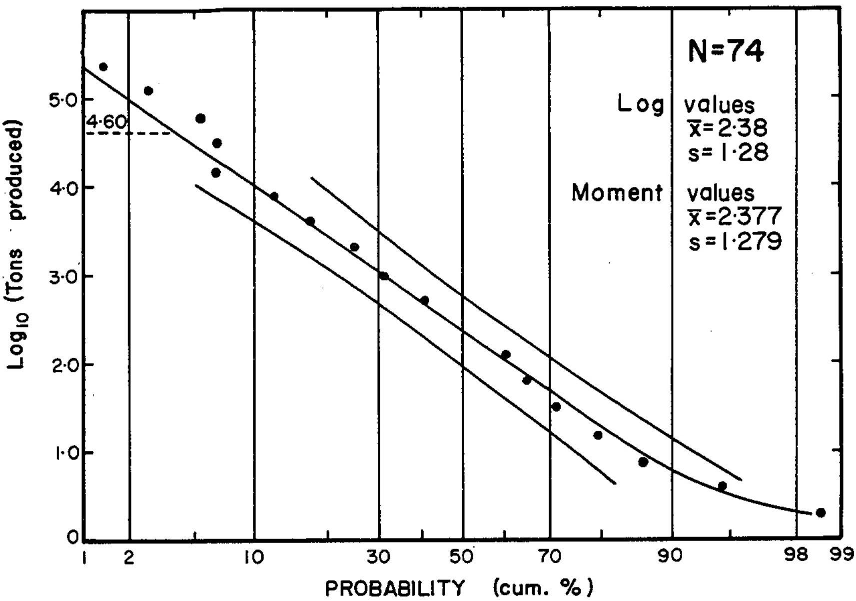
As an example of a single lognormal population consider production data for 74 Pb-Zn-Ag vein deposits in Ainsworth mining camp, southern British Columbia (Davidson, 1972). These data were compiled by Orr (1971) for all known deposits in the camp that had produced one ton or more of ore, the principal source being published reports of the B. C. Department of Mines and Petroleum Resources. A log probability plot (figure III-1) suggests that the data approximate a lognormal distribution and therefore can be represented by a straight line drawn through the plotted points. The amount of scatter of plotted points about the line is small considering the relatively small amount of data on which the graph is based and, as is generally the case, is more pronounced at the ends of the graph than towards the centre.

Note that curve-fitting does not follow standard rules because more weight is attached to values near the centre of a probability plot relative to those points near the extremities. Subjective weighting during the curve-fitting procedure can become a problem in interpreting the pattern of real data plotted on probability paper. Does real curvature exist or is the pattern a straight line? One useful rule is to accept real curvature if several consecutive points are progressively further and further from a linear trend, as opposed to random fluctuations either side of a linear trend. A second guide, suggested by

Woodsworth (1972) is to plot 95 percent confidence limits for a linear fit to plotted points, and assume curvature in the vicinity of those points plotting outside those limits.

FIGURE III-1

Probability plot of logarithms (base 10) of production for 74 vein deposits, Ainsworth Camp, B. C. Note flattening at bottom of curve, suggestive of bottom truncation. Bounding curves are 95 percent confidence limits of the normal distribution that best describes the data.



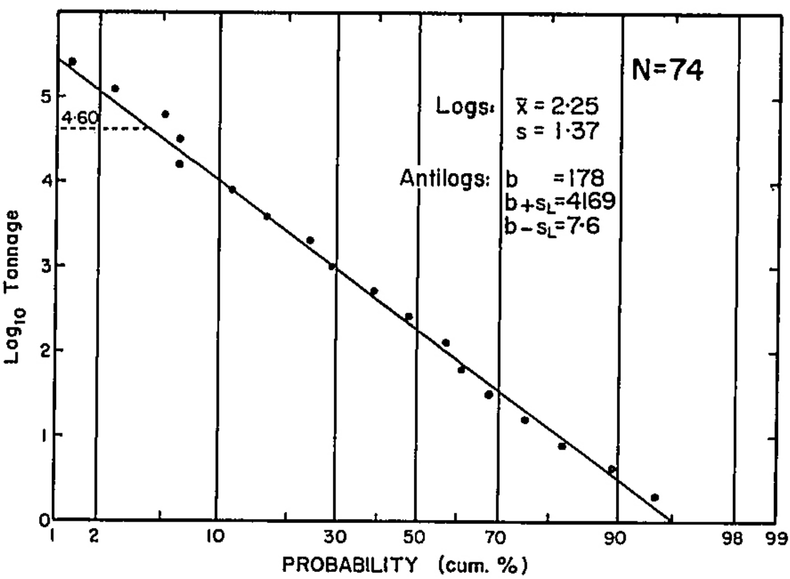
In this example, it is convenient to plot logarithms of production data on arithmetic probability paper rather than arithmetic data on log probability paper. The reason is that the data span more than five orders of magnitude and would be unwieldy using generally available log probability paper.

Estimates of parameters of the distribution can be read directly from the graph using as a basis the line fitted to the points and reading ordinates that correspond to the 16, 50 and 84 percentiles. They are estimated to be 2.38 ± 1.28 and compare with parameters obtained by the method of moments of 2.377 ± 1.279. The 95 percent confidence limits have been drawn in by a graphical estimation method described in section VIII-7 (cf. Lepeltier, 1969). All the plotted points are within these limits and we might conclude therefore that the assumption of lognormality and estimated parameters represent an adequate description of the population. Values of the parameters are given by their antilogs as 240 (4571, 12.5).

Despite the statistical soundness of the interpretation it might be questioned on the basis of the arbitrary choice of a minimum of one ton of ore produced. Why should the population coincide with this arbitrary lower cut off value? In fact, the lower end of the graph flattens appreciably and has the general form of a bottom-truncated distribution, as might be expected from the very nature of the data. Figure III-2 shows a plot of the original data recalculated on the *assumption* that the lower 5% of the population is not present. Points on the "recalculated" graph were obtained by multiplying the original cumulative percentage at each ordinate level by 0.95. The lower flattened part of the curve has disappeared and the data define a lognormal population with estimated logarithmic parameters 2.25 ±1.37 or in arithmetic terms, 178 (4169, 7.6).

FIGURE III-2

Probability plot of data of figure III-1 "corrected"for an assumed bottom truncation of 5 percent (Sinclair, 1974b).



In addition to defining the nature of the density distribution and allowing easy, rapid estimation of statistical parameters, plots such as these can be used to examine the probability that new finds in the camp will be greater than some specified size (Sinclair, 1974b). Let us assume that on the basis of our knowledge of the district, average ore grades and economic factors, a minimum deposit size for profitable production is estimated to be 40,000 tons ($\log_{10}$ = 4.6021). Assuming our data are representative of deposits in the area, the probability that a newly found deposit would equal or exceed this minimum is read as approximately 4% from either of the graphs. In this particular case the correction for bottom truncation has a negligible effect on the estimated probability of success. For much smaller minimum targets,, however, the probability would be appreciably less on the graph of figure III-2 that takes into account the bottom truncation. The simplicity of presentation and the ease with which probabilities of success can be estimated make such plots a useful adjunct to guides to exploration potential in districts for which adequate production data are available.

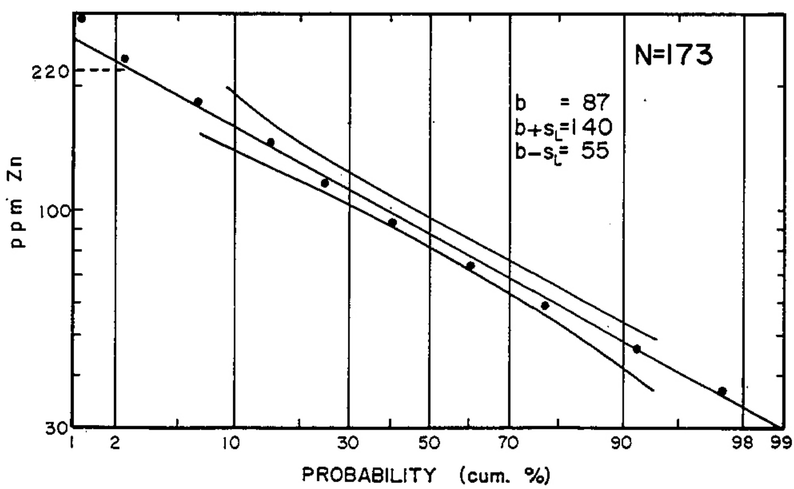
III-2: Zn IN SOILS, TCHENTLO LAKE AREA, CENTRAL BRITISH COLUMBIA.

One hundred and seventy three soil samples taken from an area underlain by more-or-less homogeneous dioritic bedrock, and largely covered by a thin layer of glacial till, are shown in figure III-3 as a probability plot. Scattered knolls of diorite outcrops occur here and there throughout the sampling grid, some of which are mineralized along joints with quartz, pyrite, and small amounts of molybdenite and chalcopyrite. Soil samples (B-horizon) were taken at intervals of 400 feet along E-W lines spaced 400 feet apart.

An examination of the probability plot indicates the presence of a single lognormal population. This is further supported by the fact•that the plotted points occur well within the 95 percent confidence limits estimated graphically, as outlined in section VIII-7. Parameters of the population, as determined from the graph, are 87 (140, 55).

FIGURE III-3

Log probability plot of 173 soil Zn (B-zone) values over a stockwork copper-molybdenum zone, Tchentlo Lake, B. C.



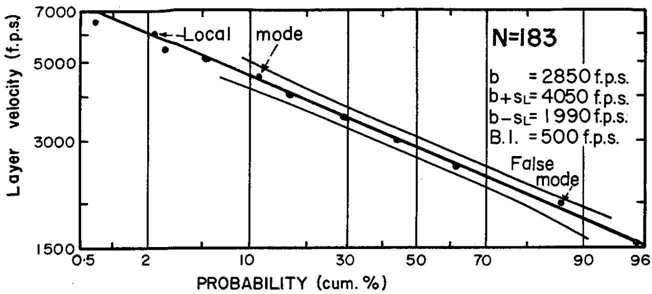
The population probably represents a background Zn population in the B-horizon of the soils. In a case such as this, it is wise *to assume* that some of the highest values are anomalous until proven otherwise. The traditional procedure recommended by Hawkes and Webb (1962) is to make the assumption that a threshold exists at (b + 2s) i.e. assume that the upper 2½ percent of values are anomalous. Applying this procedure to Tchentlo Lake data a threshold is established at 220 ppm Zn. The five highest values thus delimited were found to occur sporadically, away from known mineralized areas.

III-3: HAMMER SEISMIC DATA, SOUTHERN BRITISH COLUMBIA.

Figure III-4 is a cumulative probability plot of 183 seismic velocity values from a single subsurface layer in a survey area in southern British Columbia. Confidence limits are shown as estimated graphically using the procedure of Lepeltier (1969) and described in section VIII-7. It is apparent that the data closely follow a lognormal distribution law. None of the real data points are observed to plot outside the 95 percent confidence belt. The geometric mean (corresponding to the mode of arithmetic values) can be estimated directly from the straight line fitted to the data, as 2850 f.p.s. About 68 percent of samples lie between 4050 and 1990 f.p.s. Local modes are often of interest in such data. These can be recognized by a glance at the probability plots as "offset" points that lie "too much" towards the upper side of the line, relative to adjacent points on either side. To make use of such information, however, it should be recalled that the actual points plotted are at the lower extremity of the bar intervals they represent. Consequently, the mode itself is estimated at midway between an "offset" point and the next highest point. In some cases false modes are recognized in this manner because any bar interval for which the frequency is greater than that forecasted by the log-normal law will plot in the manner described. The significance of such modes, however, is questionable *if* they lie within the 95 percent confidence belt, in which case they could merely represent expected fluctuations due to ever-present sampling error.

FIGURE III-4

Log probability plot of 183 layer velocity measurements from a hammer seismic survey in southern British Columbia.



III-4: REPRESENTATION OF SEVERAL VARIABLES ON A SINGLE PROBABILITY GRAPH.

One hundred and twelve stream sediment samples from a semi-arid environment in southern British Columbia were analysed for Pb, Co, Cu, Zn and Ni. Results for each of these variables are shown as cumulative probability plots in figure III-5 and indicate that four of the five (Co excluded) can be considered as lognormal distributions. Parameters of these distributions have been determined graphically and are recorded in table III-1 where they are compared with corresponding values determined by the method of moments.

Parameters are comparable by the two methods, except for Pb. In this latter case the presence of a high proportion of "zero" values and a few high values that plot outside the limit of the graph, are the causes of the discrepancy.

Figure III-5 emphasizes the clarity with which several populations can be represented on a single probability graph and the ease with which significantly different populations (e.g. Co) can be recognized.

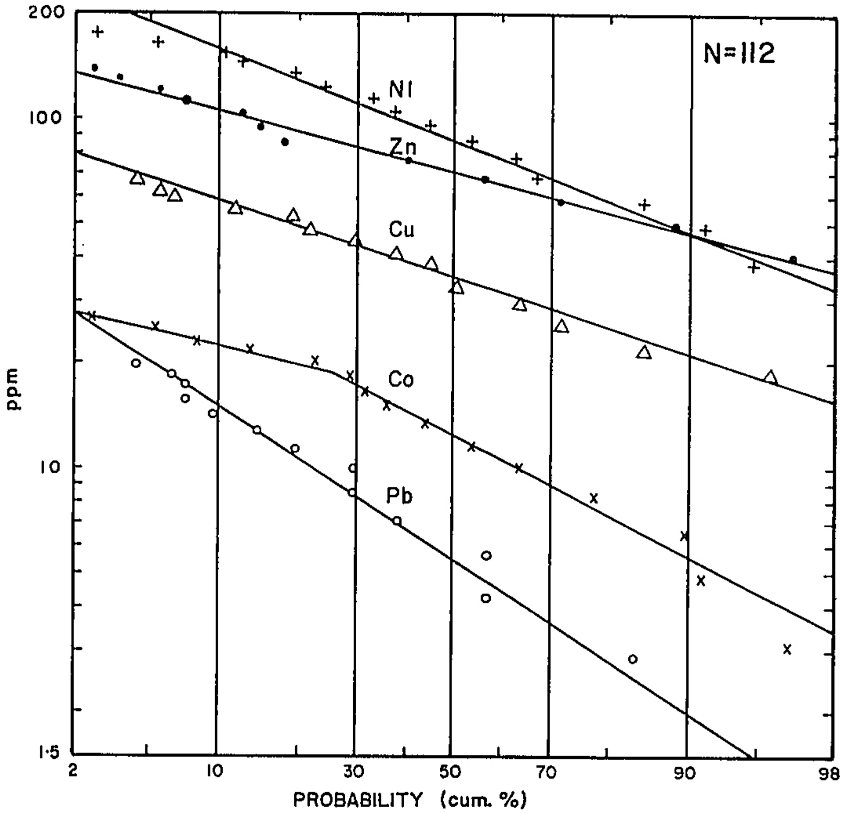
TABLE III-1

PARAMETERS OF POPULATIONS SHOWN IN FIGURE III-5 DETERMINED GRAPHICALLY AND BY MOMENTS

Variable	Graphical				Moments		
	N	b	$b+s_L$	$b-s_L$	b	$b+s_L$	$b-s_L$
Pb	112	5.4	12	2.4	3.6	19.1	0.7
Cu	112	35.1	52	23.2	34.2	51.6	22.7
Zn	112	70	96	51	71.1	106	47.5
Ni	112	86	148	53	88.8	149.3	56.6

FIGURE III-5

An example of several variables from a regional stream sediment survey, plotted on a single probability diagram. Note the clarity with which "different" elements such as Co stand out from the others.



CHAPTER IV
PROBABILITY PLOTS OF TWO HYPOTHETICAL POPULATIONS

IV-0: GENERAL STATEMENT

In preceding sections, consideration has been given to patterns of single populations, real and hypothetical, plotted as cumulative distributions on probability paper. The ease with which the forms of distributions and their parameters can be estimated, and the clarity of presentation of several variables on a single graph, give some indication of the usefulness of probability plots in presenting and analyzing many types of numeric data such as are the product of numerous mineral exploration programs.

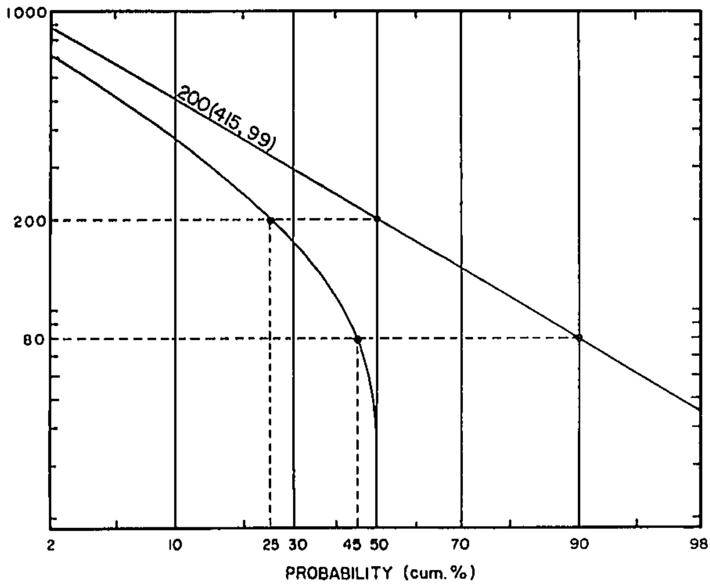
Commonly, however, data do not plot as a straight line on probability paper but have a definite curvature with a pronounced *inflection* point, or change in direction of curvature. As will be shown in this chapter, such patterns commonly result from the presence of two (or more) populations in a data set. Attention will be directed first to ideal patterns based on known hypothetical populations, from which an attempt will be made to make generalizations that aid in the reverse process of extracting constituent populations from real mixed populations. Procedures that are concerned with the estimation of constituent populations from a combination of two or more density distribution are known as *partitioning*.

IV-1: A SINGLE POPULATION PLOTTED OVER PART OF PROBABILITY RANGE

As an introduction to patterns produced by mixtures of data from two populations, it is instructive to first consider the graphical form of a single distribution plotted over only a part of the probability range. Consider the population 200 (415, 99) shown as a straight line in figure IV-1. Suppose this population were replotted on the assumption that it represents only the upper 50 percent of a data set, and ignore for the moment the lower 50 percent of the data. A series of points on the straight line can be replotted on their respective ordinate levels, each with half the cumulative probability indicated by the linear distribution. For example, at the 200 ordinate level we read 50 cumulative percent on the straight line and a new point on the 200 ordinate level is plotted at 25 cumulative percent. The same procedure is repeated at various ordinate levels.until sufficient points are obtained to draw in the smooth curve on 'figure IV-1. A practical example involving the reverse procedure is described in section VIII-2.

FIGURE IV-1

A single lognormal population with parameters 200 (415, 99) is replotted over 50 percent of the probability range to illustrate the pronounced departure from a straight line.



Curves drawn in the same manner and assuming that population 200 (415, 99) represents the upper 10, 25, 50, and 75 percentages of a data set, are reproduced in figure IV-2. Comparable curves for the same population representing comparable *lower* percentages of a data set are also shown. These graphs have several obvious characteristics:

1) The distributions are no longer linear on probability plots.
2) Curvature increases towards the centre of the plot and becomes asymptotic to the vertical percentile line that is the proportion of total data represented by the population.
3) The curves are convex upward or downward depending on whether the population represents the lower or upper fraction respectively, of a data set.

It is easy to imagine the pattern that would result if two curves of the sort shown in figure IV-2 were combined in a single data set; for example, one curve for 50 percent of a high population and a second curve for 50 percent of a much lower population such that the single resulting curve covering the entire probability range of interest, represents two populations (see figure IV-3). In fact, what has been done here is the development of a cumulative probability curve for a data set consisting of equal proportions of two populations with no effective overlap of values.

FIGURE IV-2

A single lognormal population (see figure IV-1) is replotted on the assumption that it represents various percentages at the upper or lower ends of a larger group of values.

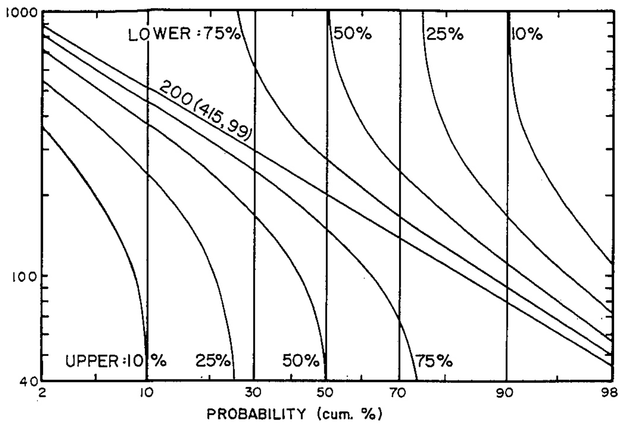
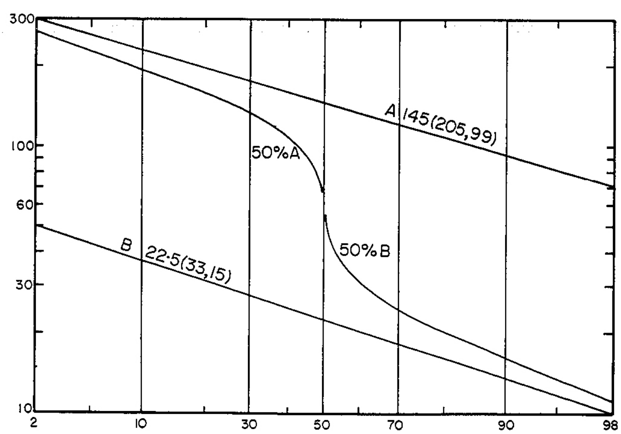


FIGURE IV-3

Two lognormal populations A and B are shown plotted over the upper and lower 50 percent probability ranges respectively, to illustrate the resulting curved distributions. Note that the two curved segments are both asymptotic to the 50 percentile, a feature that is never apparent in real data. Instead in practical cases the two curves are, joined as a single curve which in the example shown would have an inflection point at the 50 percentile.



IV-2: CONSTRUCTION OF PROBABILITY CURVES OF TWO HYPOTHETICAL POPULATIONS.

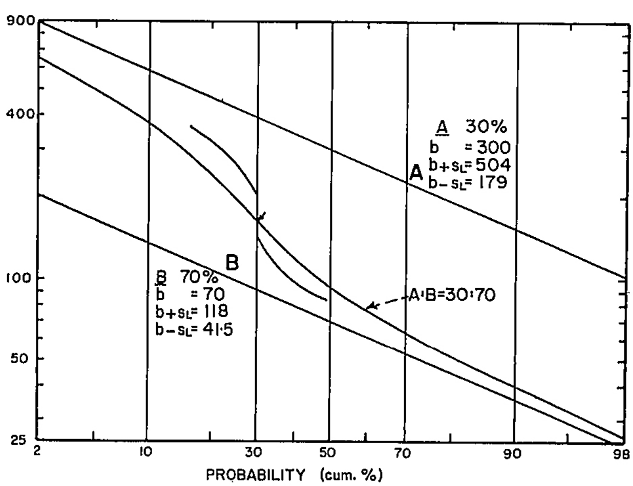
Bolviken (1971) has recently described a simple graphical method of determining the probability curve for two populations whose parameters are specified, combined in any desired proportions. Consider the two populations A, 300 (504, 174) and B, 70 (118, 43) that are to be combined in a single probability curve in the proportions 0.3A and 0.7B. At any particular ordinate level the cumulative percentage of the combined populations is equal to 0.3 times the cumulative percent of A plus 0.7 times the cumulative percent of B. This relationship can be generalized and expressed in equation form as follows:

$$P_{(A+B)} = f_A P_A + f_B P_B$$

where $P_{(A+B)}$ is the cumulative probability of the combined populations, P_A the cumulative probability of population A, P_B the cumulative probability of population B, fA the fraction of total data represented by population A, and f_B the fraction of total data represented by population B. Because the two populations comprise 100 percent of the data, $f_A + f_B = 1$.

FIGURE IV-4

A general example of two lognormal populations A and B with considerable overlap and similar standard deviations combined in the proportions 30% A and 70% B. The point of inflection is indicated by an arrow at the 30 percentile.



Populations A, 300 (504, 179) and B, 70 (118, 41.5) shown as straight lines in figure IV-4 are to be mixed in proportions 0.3 A and 0.7 B. By applying the general equation to successive ordinate levels, a series of points on the probability curve for the combined populations is obtained. Sufficient points must be calculated to permit a smooth curve to be drawn. As an example, consider the 200 ordinate level in figure IV-4 which intersects the A and B populations at 78% and 2.1% respectively. On applying the general equation the following result is obtained:

$$P_{(A + B)} = 0.3(78) + 0.7(2.1)$$
$$= 24.9\%$$

Hence, a point for the combined populations at the 200 ordinate level is calculated to be at 24.9 cumulative percent.

Similarly, at the 400 ordinate level $P_{(A + B)} = 0.3\ (29) + 0.7\ (0.0) = 8.7$. In this latter case the contribution of the lower population is negligible and approximates zero. For very low ordinate levels (below 90 ppm) the contribution of the higher population is essentially 100%. Note that the resulting smooth curve departs drastically from a linear pattern and, furthermore, that an inflection point or change in direction of curvature occurs at the 30 cumulative percentile. This example illustrates a fairly general situation of two populations, with a small but significant range of overlap, being combined in unequal proportions. It is not completely general in that the two hypothetical populations have the same standard deviation as indicated by equivalent slopes in figure IV-4.

A variety of probability curves of combinations of two populations have been constructed in the foregoing manner and are shown in figures IV-5 to IV-8 inclusive. These four diagrams can be looked upon as a sequence showing the effect on a combined probability graph as a population of low standard deviation is moved progressively upward in steps through the range of a population with a much larger standard deviation. The populations are labelled S and W in each of the diagrams to emphasize their *short* and *wide* dispersions (standard deviations) respectively. In addition, each of the four diagrams shows curves for combinations of S and W in various proportions. Consequently, diagrams presented thus far in this chapter define the range of patterns that result from ideal combinations of two lognormal (and normal) populations.

Examination of figures IV-3 to IV-8 inclusive reveals that two fundamentally different types of patterns occur. These are defined as *non-intersecting* bimodal curves and *intersecting* bimodal curves. The most common type, in the writer's experience, are non-intersecting curves typified by figures IV-5 and IV-8 where the pattern consists of a steeply sloping central segment flanked at either end by more gently sloping segments. Intersecting bimodal probability curves are illustrated by figures IV-6 and IV-7 where their pattern is seen to consist of a gently sloping central segment flanked by two more steeply sloping end segments.

The terms intersecting and non-intersecting, as applied to these curves, refer to the individual populations represented on probability paper. In the case of an intersecting bimodal probability curve, the two ideal populations that comprise the mixture are represented on probability paper by lines that *intersect* in the probability range of interest. Conversely, for a non-intersecting bimodal probability plot, the linear trends of the individual populations do *not* intersect in the probability range of interest. The use of "intersection" in this connection *should not* be confused with overlap. By their very nature intersecting populations have complete overlap in that one population is encompassed entirely within the range of another. Non-intersecting populations might overlap appreciably, but need not.

IV-3: PARTITIONING OF NON-INTERSECTING BIMODAL PROBABILITY CURVES.

In the preceding section, methods were presented by which bimodal probability curves could be constructed from specified ideal populations. Normally the reverse situation prevails – a set of data plots on probability paper with a form suggestive of a bimodal distribution, and the problem is to ascertain if such is the case and if so, to extract from the bimodal curve information about the two populations that comprise it. Procedures used to define parameters of individual populations within a polymodal population are known as partitioning.

Partitioning of non-intersecting bimodal probability curves is generally a straight-forward matter (c.f. Harding, 1949). The most important problem is to ascertain the proportions in which two populations are present. An examination of figures IV-3, IV-4, IV-5, and IV-8, shows that in every instance an inflection point occurs in the bimodal curve precisely at the cumulative percentile that represents the proportions of the two constituent populations. Consequently, in dealing with a bimodal probability curve, the first information to be extracted is the cumulative percentile at which the inflection point occurs. Once this figure has been obtained it is a routine matter to partition the bimodal distribution.

FIGURE IV-5

Examples of various combinations of two normal populations with different means and standard deviations: W has a large standard deviation (i.e. wide dispersion), S has low standard deviation (i.e. short dispersion). Figures IV-5 to IV-8 inclusive are a series that shows the range of patterns that result if two normal populations are combined. The S distribution can be viewed as being raised progressively through the effective range of the W distribution from the beginning to end of this series.

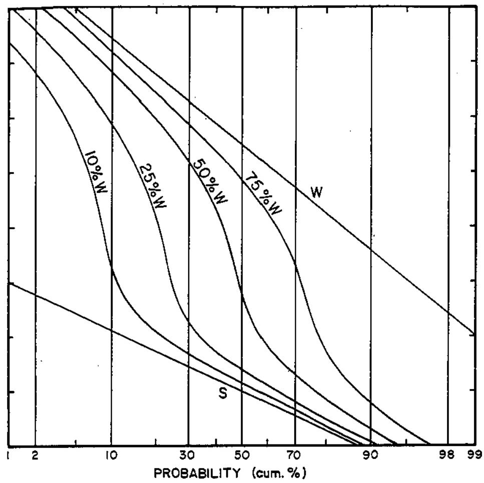


FIGURE IV-6

Mixtures of two intersecting populations W and S. See caption to figure IV-5.

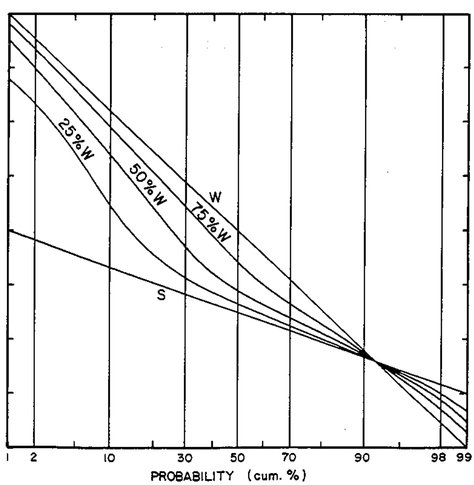


FIGURE IV-7

Mixtures of two intersecting populations, Wand S. See caption to figure IV-5.

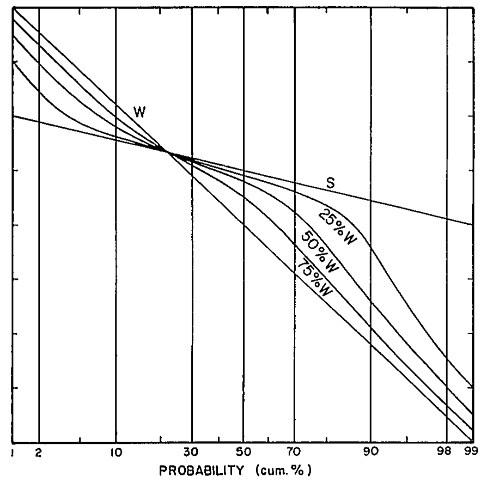


FIGURE IV-8

Mixtures of two non-intersecting populations, Wand S. See caption to figure IV-5.

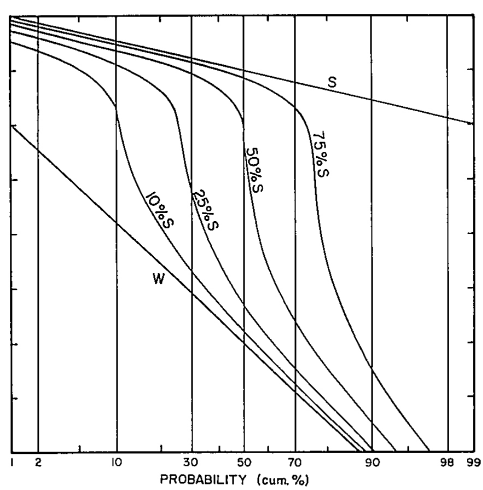
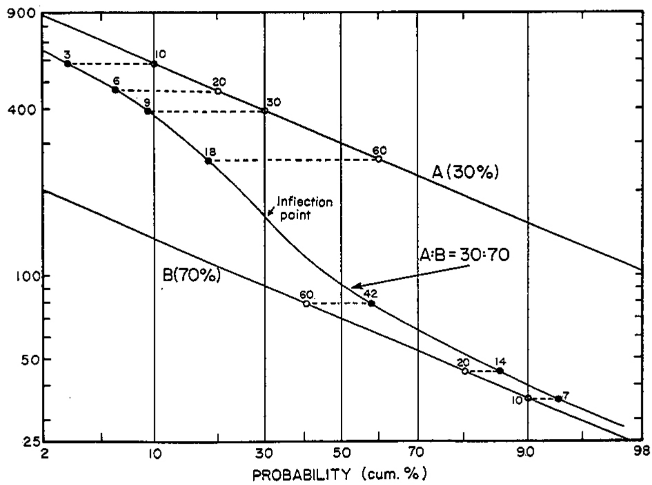


FIGURE IV-9

Illustration of partitioning procedure utilizing the two nonintersecting populations of figure IV-4. Black dots are convenient but arbitrarily chosen points on the mixed population that effectively represent either A or B. Thus no points are chosen in the central part, the range of effective overlap of A and B. These points can be recalculated readily as the open circles to define the A and B populations. Refer to text for details of partitioning procedure.



As an example, consider the curved distribution of figure IV-9. An, inflection point is apparent at or near the 30 cumulative percentile, indicating 30 percent of an upper population and 70 percent of a lower population. The sloping central segment indicates that there is appreciable effective overlap of ranges of the two populations. (If no overlap existed the central segment would be near vertical). The upper extremity of the curve, however, reflects only the upper population and the lower extremity only the lower population. Consequently, these two parts of the curve can be used to estimate the constituent populations. i.e. to partition the mixture.

Now, consider the cumulative percentile at any ordinate level near the upper extremity of the curve, say the 3 percentile for ease of calculation. This point represents only 3 percent of the total data but is (3/30) x 100 = 10 percent of the upper population. Hence, at this or dinate level a point has been defined at the 10 percentile that plots on the upper population. This procedure can be repeated for various or dinate levels that correspond to 6, 9, etc., cumulative percent on the curve, to define points on the line describing the upper population at 20, 30 etc., cumulative percent. The procedure can only be repeated until the effect of the lower population becomes significant at which point the calculated values depart from a linear pattern.

However, commonly in practice sufficient points with a linear trend can be determined to extrapolate the trend throughout the entire probability range and thus define the upper population quite precisely. The lower population can be estimated by the same procedure providing the probability scale is read in a complementary fashion (e.g. 80 cumulative percent is read as 100-80 = 20 cumulative percent).

The method is rapid with even a minimum of experience but results should always be checked, particularly in the intermediate range of overlap of the two populations where trends for both partitioned populations were extrapolated. Checking simply involves combining the partitioned populations in their indicated proportions to observe how closely ideal combinations fit the real, original curve. The procedure is that used in section IV-2.

In practical cases some difficulty is common in defining an inflection point within several cumulative percentage points. If an incorrect value has been chosen, ideal mixtures of the partitioned populations will not agree with the original curve and a second trial with a new inflection point must be attempted. Such trials should be repeated until acceptable agreement is obtained, or until the 2-population model (normal or lognormal as the case may be) is rejected.

The partitioning procedure outlined can be applied to real data that define a smooth curve reasonably comparable to the ideal curves of figures IV-5 and IV-8 and with relatively little scatter of plotted points about the curve. If considerable scatter occurs, a similar procedure is followed that differs only in the choice of ordinate levels to be used in partitioning. In such a case the original plotted points should be recalculated and each partitioned population estimated by lines through the recalculated points.

Many published interpretations of non-intersecting bimodal probability curves are based on fitting straight lines to each of the three sections of the curve and considering the two resulting intersections as important threshold values. In some cases this might be a useful procedure but in general such intersections have no particular significance. In other cases, the percentile corresponding to the mid point of the central segment is used as a basis for partitioning. Perusal of figures IV-5 and IV-8 shows, however, that in general the mid point of the central segment and the inflection point do not coincide and can, in some instances, be far removed from each other. Ordinarily there is no need to rely on approximate procedures such as these unless they are suited to a particular problem. The rapid but precise method recommended here simply involves drawing a smooth curve through plotted data points, recognition of an inflection point, partitioning of consituent populations, and checking the partition results, all based on the fundamental *curved* nature of bimodal probability plots.

Because of the frequency with which non-intersecting bimodal probability plots are encountered in nature, it is worthwhile to conclude this section with a brief summary of their characteristics:

1) Such plots describe a continuous curve with characteristic shape, including a central steeply sloping segment flanked by more gently sloping end segments.
2) An inflection point occurs within the central segment, at a percentile that defines relative proportions of the two constituent populations.
3) Some indication of the relative standard deviations of constituent populations is given by relative slopes of the two end segments providing the two populations are not grossly dissimilar in their proportions.
4) If the two populations have ranges without significant overlaps, the central segment of the curve is vertical. As the amount of overlap increases the slope of the central segment becomes progressively less.
5) The partitioning procedure includes the following stages in sequence:
   (a) draw a smooth curve through plotted data points
   (b) pick the inflection point
   (c) partition both upper and lower populations
   (d) check ideal mixtures of partitioned populations with the original curve describing the real data.

IV-4: PARTITIONING AN INTERSECTING BIMODAL PROBABILITY CURVE

Partitioning of intersecting bimodal probability curves is generally a more difficult procedure than is the case for non-intersecting curves and, as a rule, involves considerably more trial-and-error. There are several characteristics of ideal intersecting curves, however, that aid in determining the most efficient partitioning procedure.

1) The two curves for constituent populations intersect each other and the bimodal curve in the relatively flat central segment.
2) This triple intersection occurs at an inflection point in the bimodaL curve. Unfortunately this inflection point is generally difficult, if not impossible, to pick out precisely, particularly in real data for which points are scattered somewhat about a generalized curve.
3) The ordinate range spanned by the central, relatively flat segment defines, reasonably accurately, the effective range of the short dispersion population.
4) The percentage range spanned by the central segment provides a very rough estimate of the proportion of the short dispersion population represented in the bimodal curve. This estimate is generally high, but at least provides a starting point for trial-and-error partitioning. By the same token, the complementary estimate for the proportion of the wide dispersion population is generally low.

Providing all three segments of an intersecting bimodal curve are well defined, partitioning is not too difficult. A general procedure is as follows:

1) An estimate of the proportion of the wide dispersion population is made as in 4 above.
2) Using this estimate, points near the end of the two end segments are recalculated and plotted as a single population. If the recalculated points plot on a straight line, the estimated proportion is correct and the line defines the wide range population. If the points do not plot on a straight line, a new proportion (generally higher) must be used as a basis for new calculations. This procedure is repeated until a linear pattern is obtained for calculated points representing the partitioned population.
3) Once the wide range population has been defined, a series of points on the short range population can be calculated using the relationship

$$P_C = P_W\, f_W + P_S\, f_S$$

where, for any ordinate level, P_C is the cumulative probability of the combined populations (i.e. the real data curve), P_W and P_S are cumulative percentages of the wide dispersion and short dispersion populations respectively, and f_W and f_S are fractions (proportions) of the wide dispersion and short dispersion populations respec-tively. At this point in the partitioning procedure P_S is the only unknown in the equation.

4) The short dispersion population is then estimated as a straight line drawn through the points determined in the preceding steps.
5) The partitioning procedure should then be checked by recalculating ideal combinations of the partitioned populations at various ordinate levels for comparison with the bimodal curve for real data.

The foregoing partitioning method can be used most effectively in cases where the range of the wide dispersion population extends appreciably on either side of the short dispersion distribution. Considerably more guesswork is involved if the short dispersion population is shifted to one end of the effective range of the wide dispersion population.

An hypothetical example of the partitioning procedure is shown in figure IV-10. The curve was actually constructed for a 50:50 mixture of the two populations, W and S, shown. However, assuming we do not know the individual populations, partitioning would proceed as follows:

1) The percentage range spanned by the central "flat" segment of the curve is about (80-20) = 60, i.e. our first trial might assume that W represented (100-60) = 40 percent of the combined population.
2) Several points on each end segment are recalculated assuming W makes up 40 percent of the data. For example, on the upper segment 4 cumulative percent on the curve becomes (4/40 x 100) = 10 cumulative percent on a newly estimated W. On the lower end segment calculations

are done in exactly the same manner except that the complements of cumulative percentages are used, i.e. 96 cumulative percent is read as (100 - 96) = 4 cumulative percent.

3) It is apparent that the several points calculated for the two end segments do not define a single straight line. Consequently, calculations must be redone with a new estimate for the proportion of W present. Try 50 percent. Recalculated points are shown on either end of the line W and it is apparent that W and its proportion have been defined.

4) Populations S can now be estimated using the relationship $P_C = P_W f_W + P_S f_S$. For example, at the 150 ordinate level:

$$20.6 = .5 \times 33.5 + .5\ P_S \text{ from which } P_S = 7.8$$

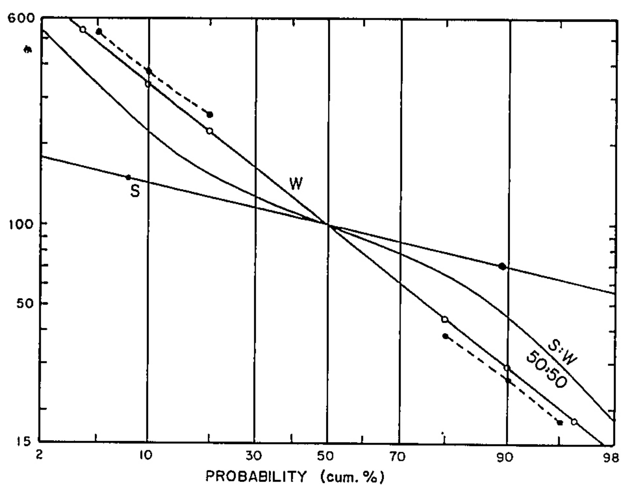
A second point on S is already known, the intersection of W and the curve. A third point calculated at the 70 ordinate level gives:

$$6.8 = 0.5 \times 64.4 + 0.5\ P_S \text{ or } P_S = 89.2$$

It is apparent that these three points describe a linear trend that defines population S.

FIGURE IV-10

Illustration of partitioning procedure for two intersecting populations W and S. The black dots joined by dashed lines indicate a first attempt based on an incorrect choice of relative proportions of S and W. With the proper choice of proportions the open circles at both extremities of the graph define a single (line) population, W. The three curves – S, Wand the mixture – intersect in a common point which in this case is coincidentally the 50 percentile. Other points on S can be determined as described in the text.



In this example, the results will not be checked because the bimodal population was originally constructed from S and W. Normally more points would be calculated for S and W than has been done here, where only a few points are shown in figure IV-10 to retain simplicity in the diagram and clarity of the partitioning procedure.

In the writer's experience, plots of real geochemical and geophysical data having an intersecting pattern, are not nearly as common as are those of the non-intersecting type. Furthermore, the scatter of points about a generalized pattern in real data can present ambiguities in interpretation. For example, certain combinations of 3 non-intersecting populations can produce patterns that closely approximate a bimodal, intersecting pattern. An example is described in section VI-4.

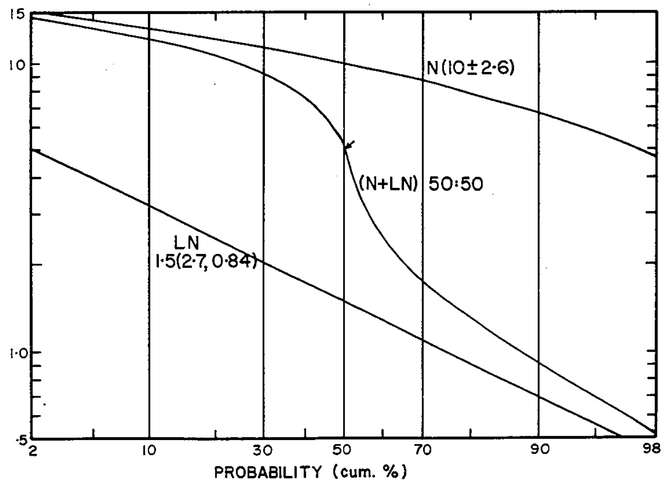
IV-5: COMBINATIONS OF NORMAL AND LOGNORMAL POPULATIONS ON LOG PROBABILITY PAPER.

Log probability graphs for data that consist of a combination of two populations, one normal and the other lognormal, can be analyzed best with an appreciation of idealized graphs generated from hypothetical populations. Certain types of data, such as ore grades and percentage minerals or elements in rock samples, offer.the possibility of a higher population that is normal and a lower population that is lognormal. Attention will be confined to this type of combination. Consider two such populations that overlap only slightly, as shown in figure IV-11, N a normal population (10 ± 2.6) and LN a lognormal population 1.5 (2.7, 0.84), that might simulate grades in some ore deposits.

The resulting cumulative probability plot of the two hypothetical populations combined in the proportions 50:50 is shown as a curve N + LN in figure IV-11. The curve is similar in general form to a mixture of two lognormal populations, and it is difficult to recognize criteria by which the two forms can be distinguished readily and with certainty.

FIGURE IV-11

Idealized pattern resulting if an upper normal population (N) and a lower lognormal population (LN) are combined on lognormal probability paper. In a real situation scatter of points defining the upper population could obscure its normal character and lead to it being interpreted as a lognormal population.



Note that an inflection point at the 50 cumulative percentile is readily discernible. However, if this were used as a basis for partitioning two lognormal populations, it would not be possible to get a close check with the real curve and points recalculated from the partitioned populations. In particular, divergences would occur on that part of the curve representing the normal population. The presence of such divergences associated with only one part of a bimodal curve should lead to consideration of the possibility that a normal population is present. This can be checked by plotting the end of the curve having divergences from "check" points on arithmetic probability paper and attempting partitioning.

Many arithmetic normal populations plotted over the full probability range on log probability paper can be crudely approximated by a straight line because their standard deviations are "short" ranges on a log scale. Hence one should keep in mind the possibility of normal and lognormal combinations in partitioning any curve that has the general form of two populations with *widely different standard deviations*.

Furthermore, there are particular types of data that might be *expected* to contain both normally and lognormally distributed populations, some of which were mentioned earlier in this section. An additional point to consider is the form of bimodal histograms if these are available.

CHAPTER V
EXAMPLES OF BIMODAL* PROBABILITY CURVES

V-0: GENERAL STATEMENT

Practical examples that could be cited in this chapter are almost infinite in terms of varieties of patterns and subject matter. The writer has chosen to restrict examples to topics of interest to the mineral exploration fraternity but, at the same time, to present real case histories sufficiently different in character and purpose to illustrate the wide range of applications of the techniques discussed in the preceding chapter.

Examples are drawn almost exclusively from the writer's personal experience and are limited in scope to that extent. It should be pointed out that a very important application, choice of threshold values, with particular reference to geochemical data, is reserved for a separate chapter (Chapter VII).

V-1: NON-INTERSECTING, BIMODAL DISTRIBUTION, 7325 LEVEL, EAGLE VEIN, NORTHERN BRITISH COLUMBIA

Eagle vein is a fairly regular, near vertical, copper sulphide-bearing deposit in the Liard Mining Division, northern British Columbia. Ninety-one Cu assays available for the 7325 level (Trimble 1972) are shown as a cumulative probability plot in figure V-1. A smooth curve drawn through the plotted points has the form of a non-intersecting bimodal distribution with an inflection point at the 62 cumulative percentile. This information was used to partition the curve into A and B populations, as shown in the figure using the method described in section IV-3. Ideal mixtures of the two curves, in the proportions 62 percent A and 38 percent B, were calculated as a check on the partitioning procedure and are shown as open triangles that coincide almost exactly with the smooth curve that describes original data points.

Consequently, the 91 data values consist of (0.62 x 91) = 56 A values and (0.38 x 91) = 35 B values. Parameters of the two estimated lognormal populations are given in table V-1.

Using the probability graph it is possible to group individual assay values as to population and then examine the geographic distribution of the different populations. In this case, assume thresholds at the lower 2 percentile of the A population (3.4% Cu) and the upper 2 percentile of the B population (5.8% Cu). Because only 2 percent of the A values (perhaps 1 value in this case) occur below 3.4 percent Cu, virtually all values below this limit are B population.

* Bimodal is used here in the ideal sense, indicating the presence of two populations.

By comparable reasoning, virtually all values above 5.8 percent Cu belong to A population. The intermediate range between 3.4 and 5.8 percent Cu contains representatives of both populations and any given value in this range cannot be assigned to a specific population. The number of values in the intermediate range is small, however, about 12 percent of the total, or about 10 or 11 values.

FIGURE V-1

Log probability plot of 91 Cu assay valves, 7325 level, Eagle vein, northern B. C. Original data are plotted as black dots, open circles are estimated partitioning points, open triangles are check points obtained by ideal combination of partitioned populations A and B. Inflection point is shown by a small arrow.

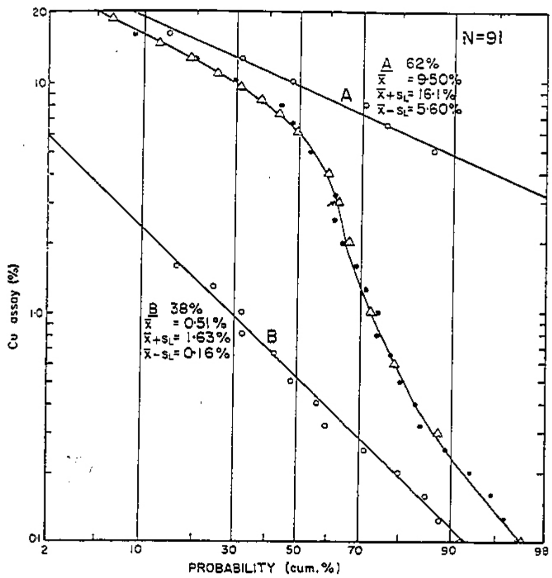


TABLE V-1

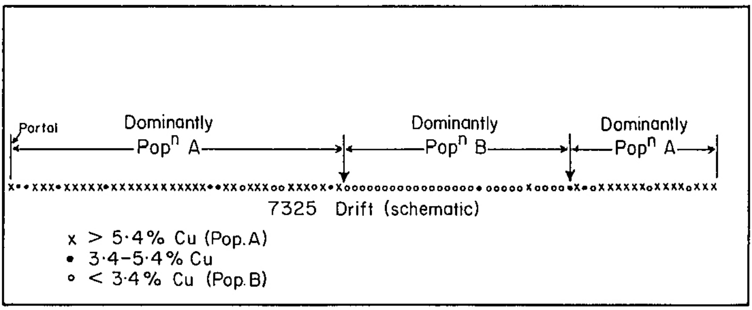
PARAMETERS OF PARTITIONED LOGNORMAL POPULATIONS, CU ASSAYS (%) 7325 LEVEL, EAGLE VEIN

Population	Proportion	N	b	b + s_L	b - s_L
A: High Grade	62	56	9.5	16.1	5.6
B: Low Grade	38	35	0.51	1.63	0.16
A + B	100	91			

Individual values can now be coded by colour or symbol on a plan of the level, according to the group they represent and the geographic distribution of populations can be examined. A schematic plan of this procedure is shown in figure V-2 which emphasizes the geographic grouping of the two populations. Values in the intermediate range plot in zones of the two distinctive populations in about the same proportions as expected from the interpretations. There are, in fact, 9 intermediate values compared with the 10 or 11 forecasted by the model.

FIGURE V-2

Relative distribution along 7325 level, Eagle vein, of samples of the two populations obtained by partitioning as shown in figure V-1 and explained in text.



In this case it has been possible to define a central segment of the vein (figure V-2) that consists almost exclusively of population B, with the remainder of the vein being characterized by population A. The geological character of these vein segments now can be investigated. On 7325 level, the central segment is distinguished from the rest of the vein by a much more pronounced shearing, fracturing and oxidation (Trimble 1972). This was recognized without recourse to probability plots but it is interesting to note that the probability analysis produced results identical with those based on field observation. In other cases such distinctions might not be so obvious without the help of "grouping" supplied by an analysis of a probability graph.

V-2: NON-INTERSECTING BIMODAL DISTRIBUTION – CHOICE OF CONTOUR VALUES.

Many types of mineral exploration data commonly are presented as contoured plans or sections. As a rule, the choice of contour values is a subjective decision based on a general, though perhaps detailed, perusal of the data. A more rigorous and standardized procedure is to take advantage of information gained from probability graphs to contour data in a meaningful and useful manner.

In the simplest case, that of a single population, it is convenient to use the mean value as a reference contour, and to use appropriate multiples of the standard deviation as other contour values. In particular, one of the contours could coincide with the value (b + 2s) to outline samples or areas that might be anomalous in the case of certain geochemical data (See Hawkes and Webb, 1962). Such contours have an additional value because, assuming the data are representative of an area or zone, one can determine rapidly the proportion of a sampled area that contains values above or below any specified contour. For this purpose it is convenient to choose contour values that are removed from the mean (both above and below) by simple multiples of the standard deviation. A contoured map of zinc in soils, described in section III-2, showed the upper 2.5 percent of values to occur as sporadic highs outside known mineralized areas.

In the case where two populations are represented in data, choice of contour values is somewhat more complex. A suggested procedure to follow involves the partitioning of the two populations as described in chapter IV. If no significant overlap of the two populations exists, contour values can be chosen from each of the individual partitioned populations as described previously. In addition, it is useful to choose another contour value that effectively separates the two populations.

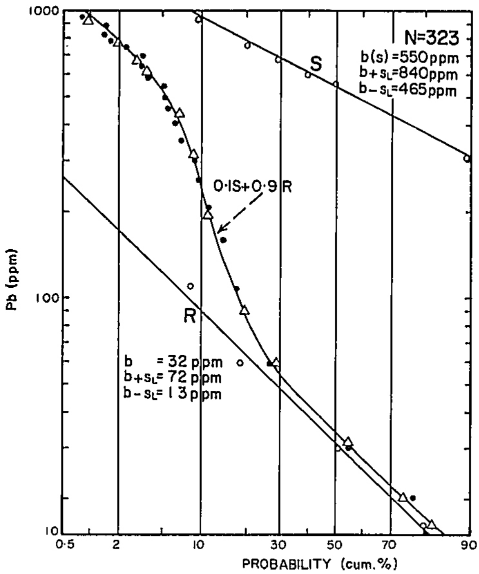
Where significant overlap of the two populations occurs, it is useful to define two thresholds as described in chapter VII, and use these as contours that divide the data into three groups, an upper group consisting essentially of a high population, a lower group consisting principally of a low population, and an intermediate group representing the effective range of overlap of the two populations and containing values belonging to both populations. Additional contour values above and below the threshold values could be chosen in the same manner as described for a single population.

For polymodal distributions containing more thann two populations the procedure for choosing contours becomes complicated, in part because of difficulties in interpreting the cumulative curve. In straightforward cases such as those described in chapter VII, however, a series of thresholds can be established in the same manner as ' for bimodal populations, and these thresholds provide useful contour values (eg. Montgomery et al, 1975).

As an example of the application of these procedures to a real situation, consider Pb assays from production blast hole samples, Brenda Mine, B. C. Brenda Mine is a large, low-grade Cu-Mo deposit in south-central British Columbia. Ore grade material is confined to a. single quartz diorite phase of the Lower Jurassic Brenda stock, although low grade material extends into other adjacent phases. As part of a study of the minor element distribution in and around the main pit area, 323 samples taken from blast hole cuttings from the 5060 level were analyzed for Pb (Oriel, 1972). Data are shown as a probability plot in figure V-3. The pattern is that of two nonintersecting lognormal populations. A smooth curve drawn through the plotted data points was partitioned into an upper S population and a lower R population based on an inflection point at the 10 cumulative percentile and using the method described in section IV-3. Overlap of the partitioned populations is negligible.

FIGURE V-3

Probability plot of 323 production blasthole Pb assays, 5060 level, Brenda Mine, southern B.C. Partitioned populations S and R have been interpreted as representing Pb in sulphides and silicates (rock) respectively. See figure V-1 for explanation of symbols.



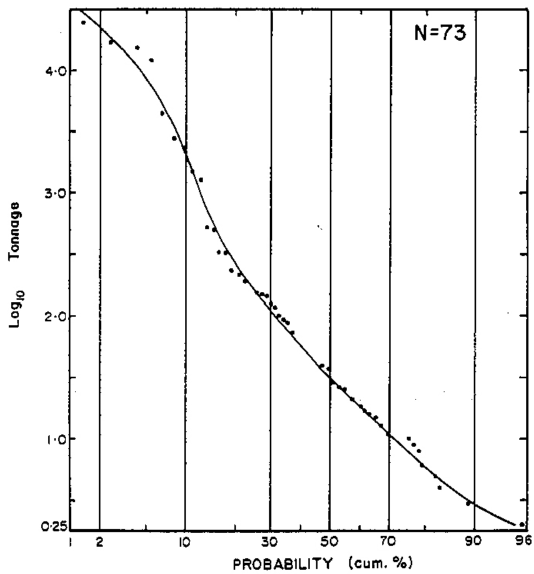
Oriel interprets the upper population, S, as indicative of sulphide lead (galena) and the lower population, R, as background lead held in the rock in silicate lattices. Because the study was a pilot project to investigate the possibility of zonal distribution of minor elements in and near the orebody, considerable importance was attached to selecting one or more contour values that would enhance the distinction between the two populations. In this case, a value of 280 ppm lead effectively separate samples and zones characterized by the two populations.

V-3: BOTTOM TRUNCATED, NON-INTERSECTING BIMODAL DISTRIBUTION — PRODUCTION TONNAGES, SLOCAN CITY MINING CAMP, B.C.

This example concerns total known production and reserves (tonnages) of 73 Pb-Zn-precious metal, vein deposits in Slocan City Mining Camp, south-central British Columbia (Sinclair, 1974b). The veins are small and although rich would probably not be of much interest to exploration companies at the present time because of their size and low gross value. Data, however, illustrate a method of analysis that is generally applicable to other mining camps, providing adequate information is available. Source of the data is given by Orr and Sinclair (1971).

FIGURE V-4

Probability plot of production tonnages of 73 vein deposits, Slocan City camp (Orr and Sinclair, 1971). Individual values have been cumulated. Note flattening at lower end, suggestive of bottom truncation.



A probability plot of the data, including all deposits that have produced at least one ton of ore, is given in figure V-4. Note that production tonnages span five orders of magnitude. Consequently, it was convenient to plot logarithmic values on arithmetic probability paper to avoid the awkward, elongated graph that would have resulted had 6-cycle log probability paper been used. Furthermore, the curve has been constructed by cumulating individual values (see section VIII-1) rather than grouping data in intervals. A smooth curve drawn through the data points of figure V-4, has the pattern of a non-intersecting bimodal distribution. The pronounced flattening of the curve at its lower end suggests the lower distribution has been truncated, a plausible likelihood in view of the arbitrary lower limit of 1 ton for available data. If the populations are truly lognormal, the percentage of missing data below the level of truncation can be estimated by a trial and error procedure. This was done using the method described in section 11-4, and indicated that the lower 10 percent of the data had been truncated. Original data were then replotted in corrected form, and is shown in figure V-5.

The curve in figure V-5 was partitioned by the method described in section IV-3, assuming an inflection point at the 10 cumulative percentile. Points used to estimate the partitioned populations H and L are shown as open circles. It is apparent that population L is well defined whereas some uncertainty exists in defining distribution H.

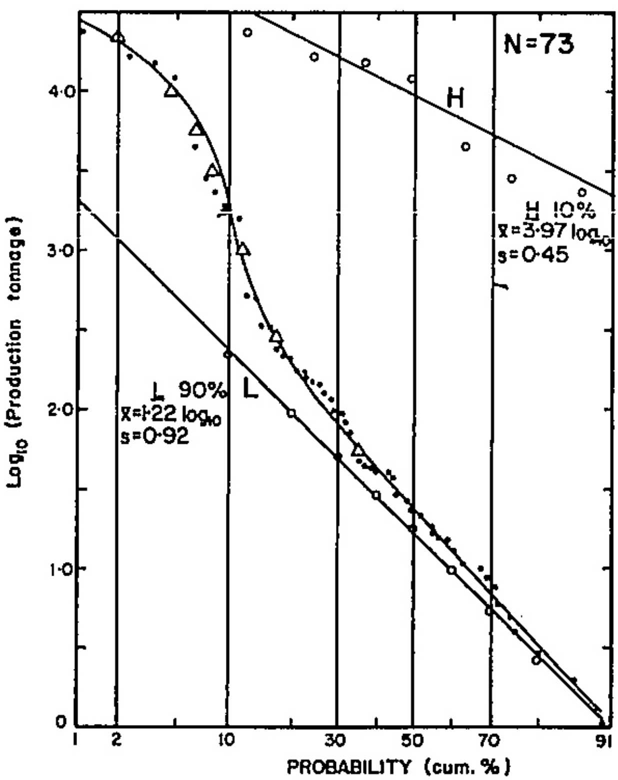
TABLE V-2

PARAMETERS OF PARTITIONED POPULATIONS, SLOCAN CITY TONNAGE DATA

			In Short Tons		
Population	**Proportion**	**N**	**b**	**b + s_L**	**b - s_L**
H: High tonnage	10	7	9120	26910	316
L: Low tonnage	90	66	16.2	132	2

FIGURE V-5

Probability plot of data of figure V-4 corrected for an assumed bottom truncation of 10 percent. See figure V-1 for explanation of symbols.



This uncertainty arises because of the small sample size and the relatively small proportion of population H in the total data. A number of checks were then made on the partitioning results by calculating ideal mixtures of H and L in the proportion 10:90. These are shown in figure V-5 as open triangles that almost coincide with the smooth curve describing the data.

Figure V-5 can now be used to extract several types of useful information. Parameters of the two distributions are shown in table V-2. We know that 10 percent of our 73 deposits (say 7) belong to the high tonnage population H and the remaining 66 deposits to a low tonnage population L. Because there is very little overlap of the effective ranges of the two populations, the 6 or 7 highest tonnage deposits can be assumed to belong to population H. Thus, the probability analysis has produced an important grouping of the data, that permits recognition of the specific population to which individual deposits belong. Geological characteristics of these groups can be examined for features that might permit rapid and certain distinction between the two tonnage groups. Such information could affect decisions involving additional exploration of a newly discovered deposit, or might point to a few "low tonnage" deposits with previously unrecognized high tonnage potential. In a detailed study of Slocan City Camp, Orr (1971) found that high tonnage deposits are characterized by the presence of carbonate in gangue, and a definite narrow range of vein orientations (strikes).

The probability plot of figure V-5 can also be used as a rapid method for estimating probabilities that a new find will meet some specified minimum exploration target. Assume for Slocan City Camp that a minimum size of 5000 [$\log_{10}$ (5000) = 3.699] tons has been estimated as necessary for profitable production. The probability that a newly found deposit will meet this minimum requirement is read from the curve as about 6% (i.e. 6 chances in 100). However, if the newly found deposit, can be classified into population H or L on the basis of observed geological features, the probability changes drastically. If recognized as a member of population H the probability that a deposit will exceed the minimum target is about 60 percent, whereas if it belongs to population L the probability of success is less than one percent.

Of course, an analysis of probability plots will not solve all problems. They simply provide an easy semi-quantitative approach to analysis of density distributions that allows the formulation of reasonable hypotheses concerning a given set of data and direct our attention by grouping data into specific populations. Interpretation of the significance of these populations is a matter for the investigator. The situation is not helped significantly if only a single population is found. On the other hand, the polymodal aspect of tonnage populations could be easily overlooked without the help of probability graphs.

V-4: INTERSECTING LOGNORMAL BIMODAL DISTRIBUTION - Sb IN STREAM SEDIMENTS, MT. NANSEN AREA, YUKON.

A probability plot of 158 Sb analyses of stream sediments from near Mt. Nansen, Yukon Territory (see Bianconi and Saager, 1971), is shown in figure V-6. Because of the low precision of analyses (the 95 percent confidence limits are ± 50 percent) the plotted points are somewhat scattered, but the general form of their probability curve drawn freehand through the original data is that of an intersecting bimodal lognormal distribution. Hence the curve was partitioned using the method described in section IV-4 to obtain estimated populations W and S as shown. Check points based on a combination of 40 percent W and 60 percent S, shown as open triangles in figure V-6, coincide almost exactly with the smooth curve. Estimated parameters of S and W are given in table V-3.

FIGURE V-6

Probability plot of 158 antimony analyses of stream sediments, Mt. Nansen area, Yukon. This is an example of an intersecting type. See figure V-1 for explanation of symbols.

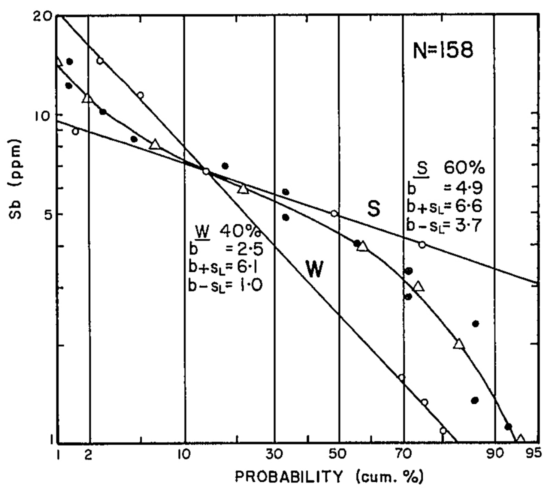


TABLE V-3

ESTIMATED PARAMETERS OF PARTITIONED POPULATIONS Sb IN STREAM SEDIMENTS - MT. NANSEN AREA, YUKON

Population	Proportion %	N	Values in ppm Sb b	$b + s_L$	$b - s_L$
S	60	95	4.9	6.55	3.35
W	40	63	2.44	6.1	0.98
S + W	100	158			

Information in table V-3 and figure V-6 can be used to estimate thresholds that group the data in the most useful manner for interpreting their geological significance (see chapter VII). Here, thresholds are arbitrarily chosen at ordinate levels that correspond to the 2.5 and 97.5 cumulative percentiles of population S. These percentiles correspond to 8.8 and 2.7 ppm Sb respectively. This range includes 95 percent of the S population and (43 - 8.5) = 34.5 percent of the W population, i.e. about 90 S values and 22 W values in this case. Therefore, in this intermediate range, S values are about 4 times as abundant as are W values, although individual values cannot be assigned specifically to either of the partitioned populations.

The remaining values, both above and below the intermediate range, belong predominantly to W population. In fact, only 4 or 5 values belong to S population whereas 41 or 42 represent W values. In other words, W values are 8 to 10 times more abundant than S values outside the central ppm range. The thresholds can be used to code values by colour or symbol on a plan of the area, in an attempt to define the fundamental geological significance of the two populations.

V-5: NON-INTERSECTING, NORMAL, BIMODAL DISTRIBUTION — AVERAGE Pb GRADES VEIN DEPOSITS, AINSWORTH CAMP.

Average Pb grades of 71 vein deposits, Ainsworth Mining Camp, southern British Columbia, are shown as a probability graph in figures V-7 and V-8. Data are shown on log probability paper (figure V-7) as an example of the form that *normal* distributions take when plotted on *lognormal* probability paper. The plot of figure V-8 has the form of a bimodal distribution that can be partitioned using the method described in. section IV-3, based on an inflection point at the 40 cumulative percentile. To check the partitioning procedure, a number of ideal mixtures of the two partitioned populations were calculated.

These are shown as a series of open triangles on figure V-8 that closely agree with a curve drawn freehand through original data points.

Parameters of the partitioned populations can now be estimated from the graph. These are 48 ± 12 percent Pb and 7.0 ± 3.9 percent Pb for the upper and lower populations respectively.

The interpretation is not unambiguous. An examination of the upper population suggests that it might in fact represent an averaging of two populations. Data are not adequate to provide convincing proof but the possibility of a third population should be kept in mind.

FIGURE V-7

Log probability plot of average Pb grades for 71 vein deposits, Ainsworth Camp, southern B.C. This pattern could arise by plotting a combination of two normal distributions on logarithmic probability paper. The same data are shown in figure V-8 plotted on arithmetic probability paper.

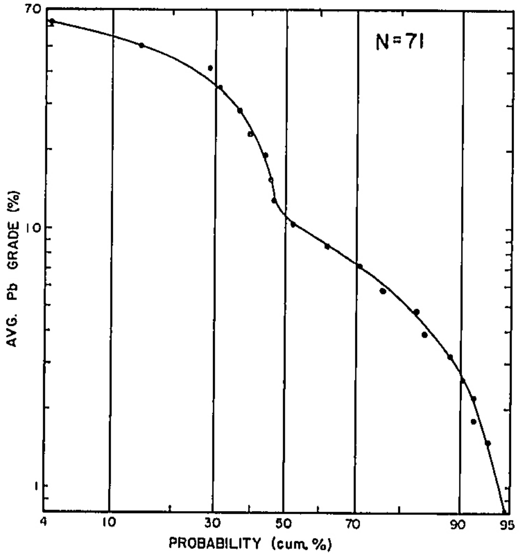
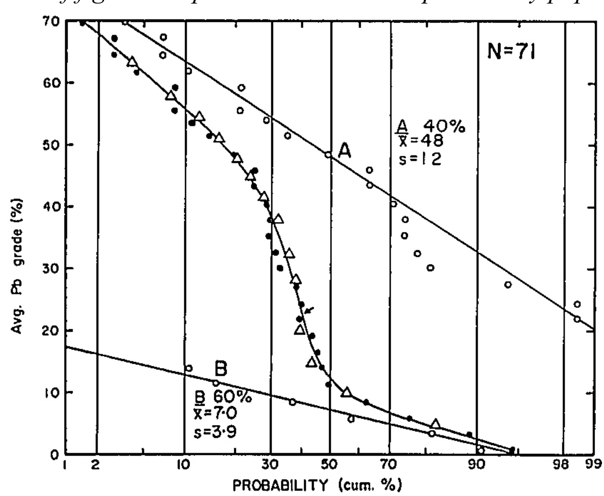


FIGURE V-8

Data of figure V-7 plotted on arithmetic probability paper.

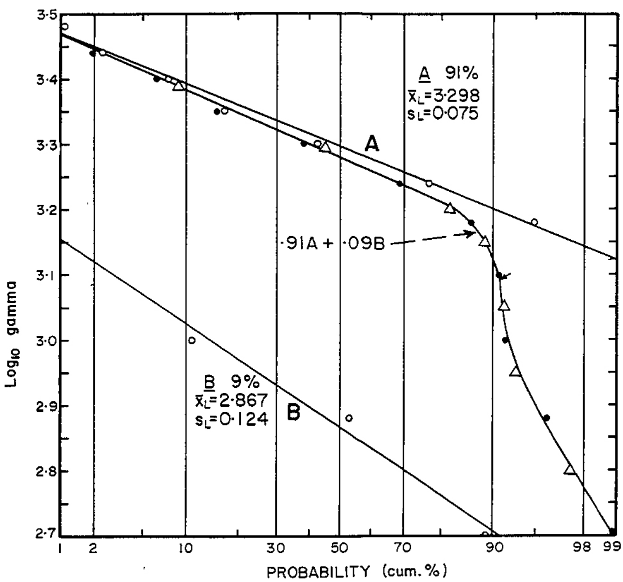


V-6: MAGNETOMETER DATA, CENTRAL BRITISH COLUMBIA.

Figure V-9 is a probability plot of values obtained from a ground magnetometer survey carried out on a grid over a claims group in central British Columbia. The curve has the form of a bimodal distribution of the non-intersecting type and is easily partitioned into populations A and B using the method described in Section IV-3, assuming an inflection point at the 91 cumulative percentile. Partitioned populations combine reasonably well to give the original curve, although it is apparent that agreement is not as good at the lower end as at the higher end.

FIGURE V-9

Probability plot of values obtained from a ground magnetometer survey central British Columbia. See figure V-1 for explanation of symbols.



The two populations have a very short range of effective overlap as indicated by the steepness of the central segment of the data curve. In fact, at least to a first approximation, it is possible to pick a single value (antilog 3.14 = 1380 gammas) that fairly effectively separates values within the two populations. The populations were found to correspond closely to the two dominant rock types in the survey area. Parameters of the two populations can be determined from the graph and serve as estimates that characterize values associated with individual rock types. These parameters are listed in table V-4.

TABLE V-4

Population	Proportion %	Parameters* (logarithms-base 10)	
		$\bar{x}$	s
A	91	3.298	0.075
B	9	2.867	0.123

* Antilogs of the parameters provide values in gammas

CHAPTER VI
PROBABILITY PLOTS OF COMBINATIONS OF THREE OR MORE POPULATIONS

VI-0: GENERAL STATEMENT

In general, as the number of populations represented in a single probability curve increases, more and more uncertainty is present in interpretation of the curve. This is particularly true in the cases of relatively few data values. To offset this difficulty, a useful procedure is to group data on the basis of some underlying physical or chemical property, such as rock type, pH, etc., to produce curves amenable to a simple and straight-forward, interpretation. In some cases, however, such grouping is not possible and a polymodal probability curve must form the basis of any interpretation attempted. As a rule, interpretation of distributions containing only three populations is straight-forward, although ambiguity can arise.

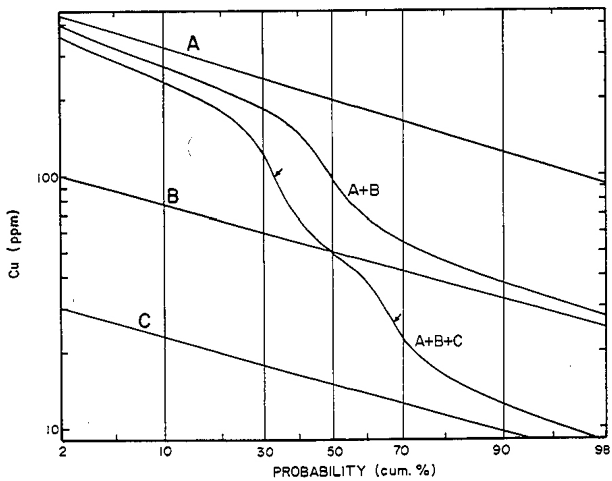
VI-1: CONSTRUCTION OF PROBABILITY CURVES FOR COMBINATIONS OF THREE POPULATIONS.

Consider the three lognormal populations, A, B, and C in figure VI-1, with approximate parameters 200 (290, 148), 50 (70, 36) and 15 (21, 10.7) respectively. Assume that these are to be combined in the proportions 1:1:1. It is convenient here to combine these populations in two successive stages using the method described in section IV-2. First, populations A and B are mixed in the proportions 1:1 to obtain the curve (A + B). Then (A + B) and C are mixed in proportion 2:1 to give the curve (A + B + C). An examination of curve (A + B + C) shows that two inflection points are present. As a general rule, the number of inflection points is one less than the number of populations present. Furthermore, as with the bimodal case, the inflection points provide accurate estimates of the proportions in which the three populations are present, i.e. at the 33 and 67 cumulative percentiles. These relationships form the basis for partitioning probability curves of real data.

Normally, if three populations are present and do not overlap too extensively, their partitioning is a straight-forward procedure not unlike that of a bimodal curve, but involving an additional step. Considerable ambiguity can be present, however, if relatively few data values are available, or if intersecting populations are present.

FIGURE VI-1

Diagram illustrating successive combination of 3 populations A, B and C. A and B are combined in the ratio 1:1 to produce (A + B). (A + B) is then combined with C in the ratio 2:1 to produce (A + B + C). Note the 2 inflection points on (A + B + C) at 33 and 67 percentiles.



In probability curves with the general form of figure VI-1, the partitioning procedure is fairly simple. The highest and lowest populations can be partitioned in exactly the same way as for bimodal curves, as described in section IV-3. With experience, it is possible in some cases to partition the intermediate population directly. More commonly, however, it is necessary to use a simpler but somewhat longer procedure that involves:

1) Combination of the high and low populations in the proportions in which they are present (derived from inflection points).
2) Application of the formula

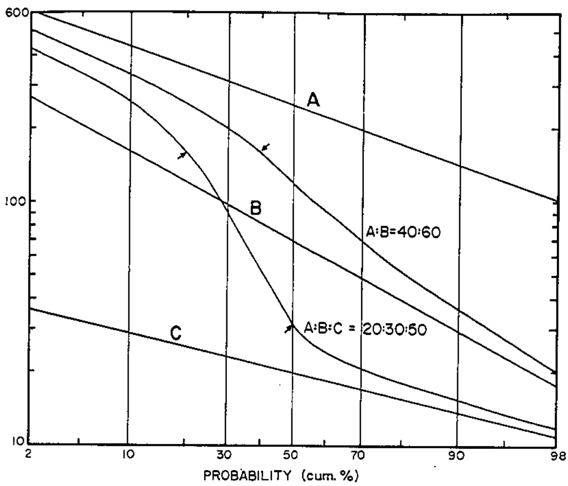
$$P_{(A+B+C)} = f_B P_B + f_{(A+C)} P_{(A+C)}$$

for various levels. In this equation $P_{(A+B+C)}$ and $P_{(A+C)}$ are read from the probability graph. The three populations thus partitioned should be recombined at various ordinate levels for comparison with the real data curve.

Figure VI-2 is presented to illustrate one of the ways in which practical difficulties can arise. The three hypothetical populations A, B, and C, with approximate parameters 250 (383, 162), 70 (133, 36) and 20 (26.5, 15.9) have been combined in the proportions 20:30:50 respectively, in stages, as shown in the figure. In this case, the upper inflection point at the 20 cumulative percentile is indiscernible, although the inflection point at the 50 cumulative percentile is apparent. This example differs from the preceding one in that the two higher populations overlap to a considerable extent. Furthermore, each of the upper two populations comprises a significantly lower proportion of total data than does the lower population. In comparable cases with real data, it is certain that the upper inflection point would be entirely masked and one would be inclined to attempt an interpretation involving only two partitioned populations. For this particular hypothetical case, partitioning of the curve on the basis of an assumed bimodal distribution would result in population C being well defined but the upper partitioned "population" would not show a linear pattern. Instead a bimodal curve comparable to (A + B) in figure VI-2 would emerge that could itself be partitioned to give estimates of the A and B populations.

FIGURE VI-2

An example of 3 combined populations that superficially resembles the combination of 2 non-intersecting populations. Note the ambiguity in defining inflection points.



VI-2: pH VALUES OF STREAMS, SOUTHERN BRITISH COLUMBIA.

pH measurements of streams are commonly an integral part of stream sediment and/or water geochemical surveys. A probability plot of pH values from one such survey in southern British Columbia, is shown in figure VI-3. The plot is on arithmetic probability paper, a logarithmic transform being built into the data because of the very nature of pH values. A smooth curve through the plotted points has the form of a trimodal distribution, with inflection points at the 16 and 85 cumulative percentiles. The curve has been partitioned using the method described in section VI-1 to obtain populations A, B, and C. Check points, based on ideal mixtures of the three populations in the proportions 16:69:15, are shown as open triangles that agree remarkably well with the real data curve. Estimated parameters of the three partitioned populations are summarized in table VI-1.

Thresholds arbitrarily chosen at the 99 cumulative percentiles of the A and B populations, and the cumulative percentiles of the B and C populations, as described in section VII-1, divide data into four groups (see table VI-2). This grouping could be of fundamental significance in interpretation of accompanying stream sediment analyses because of the important effect pH exerts on metal dispersion and concentration. Hence, the groups defined on the basis of the pH probability plot should be treated separately when various metal abundances are being examined as probability graphs.

FIGURE VI-3

Partitioned probability plot of pH values taken as part of a regional stream sediment survey, southern British Columbia. See figure V-1 for explanation of symbols.

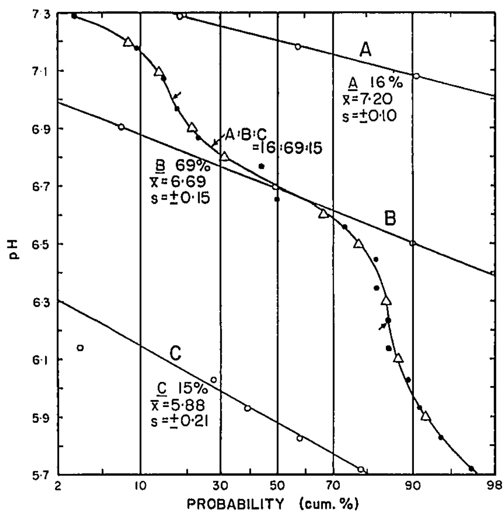


TABLE VI-1

ESTIMATED PARAMETERS OF PARTITIONED POPULATIONS, pH VALUES, SOUTHERN B.C.

Population	Proportion %	$\bar{x}$	s
A	16	7.2	0.1
B	69	6.69	0.15
C	15	5.88	0.21
A + B + C	100		

TABLE VI-2
ESTIMATED THRESHOLDS, pH VALUES, SOUTHERN B.C.

pH Thresholds	Principal Content of Range	Percent of Total Data
	POPULATION A	15
7.04		
	POPULATIONS A + B	4.5
6.93		
	POPULATION B	64.5
6.36		
	POPULATION C	16

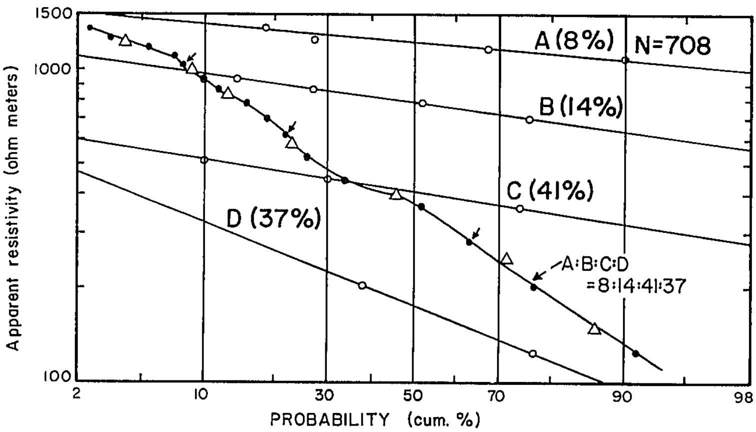
VI-3: APPARENT RESISTIVITY DATA, SOUTHERN BRITISH COLUMBIA.

Figure VI-4 is a cumulative probability plot of 708 apparent resistivity values obtained as part of an I.P. survey over a mineral prospect in southern British Columbia. The graph has a complex form with inflection points located approximately at the 8, 22, and 63 cumulative percentiles indicating the possible existence of four lognormal populations. Partitioning, proceeding on the assumption of lognormality, gives the four populations shown as straight lines. Control points used in obtaining the partitioned populations are shown as open circles.' The four partitioned populations were then recombined ideally in their respective proportions, at selected points shown as open triangles that virtually coincide with the original data curve. Consequently, the assumption of four lognormal populations in the original data set appears a plausible model and interpretation can proceed on this basis. In practice, interpretation of the significance of each population should be done in a manner comparable to that for geochemical data, i.e. selection of thresholds as described in chapter VII, and colour coding of various groups of values thus obtained for ease of comparison with geological maps or other data.

This example of partitioning assumes that because the original curve has a large data base (708 values) the inflection points are significant. Furthermore, the example illustrates the ease with which partitioned populations can be estimated when only relatively few points are available to define them. The actual partitioning procedure is of some interest. Populations A and D were partitioned in the normal way and the two were subtracted from the original data curve. The resulting "difference" curve had a simple bimodal distribution pattern that was easily partitioned into populations B and C.

FIGURE VI-4

Partitioned probability plot of 708 apparent resistivity readings over a mineral prospect in southern British Columbia. See figure V-1 for explanation of symbols.



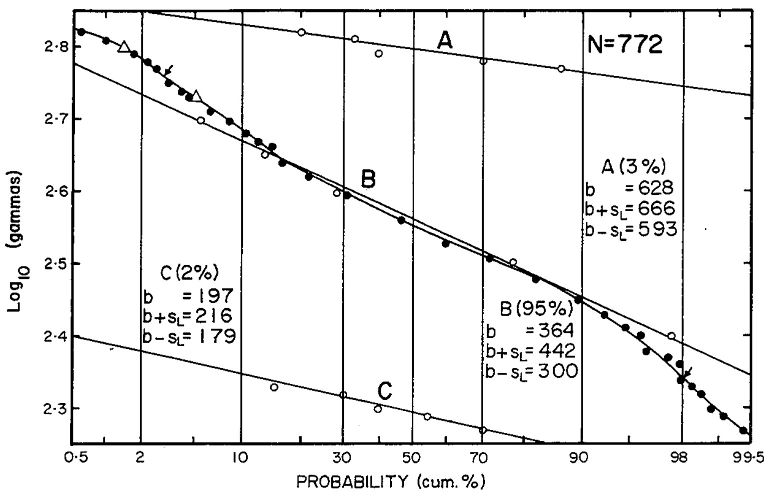
VI-4: GROUND MAGNETOMETER SURVEY, ASHNOLA PORPHYRY COPPER PROSPECT, SOUTHERN BRITISH COLUMBIA.

This example illustrates some of the ambiguities that arise in analysis of real polymodal data. Figure VI-5 is a log probability plot of 772 magnetometer readings obtained on a regular grid over the Ashnola porphyry copper prospect in southern British Columbia (Montgomery et al, 1975). Ignoring a slight variation at the upper end of the graph, one might easily interpret the general form as a combination of two intersecting populations. Alternatively, the pattern can be explained as the combination of *three non-intersecting* populations. Which of these interpretations is more likely? The probability graph itself does not necessarily provide an answer to this question. Where such ambiguities arise it seems wise to attempt *both* interpretations and evaluate them in terms of independent information, geological, for example. For the particular example quoted here, the model involving three non-intersecting populations seemed most reasonable because the partitioned populations could be related to rock types. In particular, the upper partitioned population, A of figure VI-5, which was ignored in the "two non-intersecting populations model" was found to correlate with several dyke-like intrusions of magnetite-bearing quartz monzonite.

This type of ambiguity is not uncommon in polymodal data in which one population dominates others by an order of magnitude or more. Alternate interpretations can easily escape detection if considerable scatter of data occurs in the probability plot, as might be expected with relatively few data.

FIGURE VI-5

Probability plot of 772 ground magnetometer readings over Ashnola por-phyry copper prospect, southern British Columbia.



CHAPTER VII
EFFECTIVE GROUPING OF POLYMODAL DATA — ESTIMATION OF THRESHOLDS IN GEOCHEMICAL DATA

VII-0: GENERAL STATEMENT

Threshold is a term commonly used in the mineral exploration industry to signify a specific value that effectively separates a set of data into high and low groups that result from different causes. Commonly, the term is applied to a value that distinguishes an upper anomalous set of data from a lower background set. In many types of data anomalous values are associated with mineralized areas. Hence, the estimation of a threshold value can be of considerable importance in directing exploration towards anomalous areas where the chances of discovery of an economic mineral deposit are greatly enhanced.

Thresholds are chosen in a variety of ways. The method recommended in several publications involves estimation of the statistical parameters of a data set and the arbitrary designation of values lying more than two standard deviations from the mean as being anomalous. In some cases this procedure might be adequate but the subjective approaches commonly used in practice, based on evaluation of an histogram or general perusal of tabulated data, suggests that the statistical approach is lacking in effectiveness. In fact, use of quantitative subjective approaches is a recognition of the fact pointed out by Bolviken (1971), that both anomalous and background values each represent their own density distributions. There is obviously no reason why only 2½ percent of any data set need be anomalous – why not 1 percent or 25 percent? In a case where no effective overlap of the two population occurs, it is a relatively easy matter to pick a threshold by rapid perusal of either a histogram or tabulated data (providing the data list is not too long). However, as the two populations overlap more and more, choice of a threshold becomes increasingly difficult. What is more, the effectiveness of a single threshold value decreases. If it is chosen in such a manner that all anomalous values are included, then a high proportion of background values are also included. If it is chosen at some point within the range of overlap then a number of anomalous values are missed due to inclusion with background values. Obviously a procedure is desirable for choosing threshold values that maximizes the recognition of anomalous values and minimizes the number of background values included with anomalous values. Cumulative probability plots provide an effective graphical technique to solve this problem (Sinclair, 1974a).

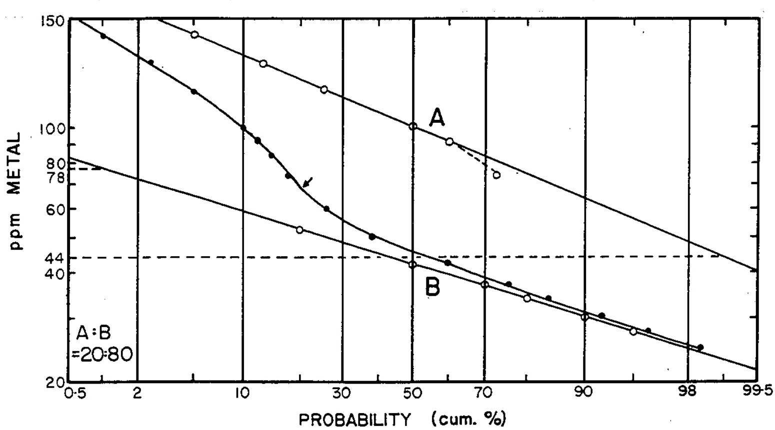
In a previous section (III-2), the writer suggested that the method of Hawkes and Webb (1962), of assuming the mean plus 2 standard deviations to be a threshold, values above which are anomalous, provides a useful safety factor when a single population is indicated on a probability graph of the data. In such a case, a very small proportion of the data might represent an anomalous population present in sufficiently small proportion that it does not show up on a probability graph. It is thus a wise procedure to assume the upper few values are anomalous until proven otherwise.

VII-1: CHOICE OF THRESHOLDS IN BIMODAL DISTRIBUTIONS

Many practical examples of geochemical data consist simply of combinations of a single background population and a single anomalous population. Such distributions commonly plot on log probability paper as curved distributions of the non-intersecting type discussed in chapters IV and V. An hypothetical example is shown in figure VII-1. Here the curved distribution is partitioned into two populations, A [100 (144, 71)] and B [42 (55, 33)] on the basis of an inflection point at the 20 cumulative percentile. Thresholds are chosen arbitrarily near the upper extremity of the B population and the lower extremity of the A population. For example, thresholds can be chosen at the 99 and 1 cumulative percentiles of the A and B populations respectively. There is nothing sacrosanct about these choices. In a given problem they could be different. In the writer's experience, the 98 or 99 and 1 or 2 cumulative percentiles have provided useful practical thresholds. These percentiles divide the data into three groups based on arbitrary thresholds at the 44 and 78 ppm levels. Sixteen percent of the total data is above the 78 ppm level. This consists of about 76 percent of the anomalous values and 1 percent of the background values. If the data set contained 100 values for example, the 16 greater than 78 ppm would consist of approximately 15 anomalous samples and 1 background sample.

FIGURE VII-1

An idealized bimodal distribution is partitioned into components A and B. Thresholds are chosen arbitrarily at I cumulative percent B and 99 cumulative percent A to give values of 78 and 44 ppm metal respectively.



The lower group, below 44 ppm, contains about 46 percent of the total data. It is made up of 57 percent of the background population and 1 percent of the anomalous population. In the case of a data set with 100 values, one anomalous value at most would be in this lower group and the remaining values would be background population.

The intermediate group between the arbitrary thresholds of 78 and 44 ppm contains about 38 percent of the values consisting of about 42 percent of the B population and 23 percent of the A population. In our hypothetical sample of 100 values, the intermediate group would contain four or five A values and thirty-three or thirty-four B values. Results are summarized in Table VII-1.

TABLE VII-1
DISTRIBUTION OF VALUES OF POPULATIONS A AND B AMONG GROUPS DEFINED BY ARBITRARY THRESHOLDS: HYPOTHETICAL EXAMPLE

	Total Data		**Population A**		**Population B**	
	%	**N**	**%**	**N**	**%**	**N**
GROUP I 78 ppm	16	16	76	15	1	0.8
GROUP II 44 ppm	38	38	23	4.8	42	33.6
GROUP III	46	46	1	0.2	57	45.6
TOTALS	100	100	100	20	100	80

This procedure has thus divided the anomalous values effectively into two groups, an upper group in which they outnumber background values by about 15 to 1, and a second group where they are outnumbered by background values by about 5 or 6 to 1. Priorities for further evaluation can now be assigned to each of the upper groups – top priority to the upper group because virtually all values in it are anomalous, and a lower priority to the central group because, although it contains essentially all remaining anomalous value, abundant background values are present.

VII-2: Cu IN STREAM SEDIMENTS, MT. NANSEN, AREA, YUKON TERRITORY

Cu analyses of 158 stream sediment samples from the Mount Nansen area, Yukon Territory, are shown as a probability plot in figure VII-2 (see Saager and Sinclair, 1974). A smooth curve through the plotted points has the form of a bimodal distribution consisting of two non-intersecting lognormal populations. An inflection point is evident at about 15 cumulative percent. The curve was partitioned, using the method described in section IV-3, to obtain populations A and B whose estimated parameters are given in Table VII-2. The partitioning procedure was checked at various ordinate levels by combining the two partitioned populations in the proportions 15 percent A and 85 percent B, with check points shown in figure VII-2 as open triangles that essentially coincide with the real data curve. In this example, some of the high values of population A are associated with known Cu-Mo sulphides related to porphyritic intrusions, and it seems reasonable to interpret the two populations as anomalous (A) and background (B).

Two arbitrary thresholds can be determined readily from the graph at the 1.0 and 99.0 cumulative percentiles of the B and A populations respectively. These percentiles coincide with values of 70 and 37 ppm Cu respectively. Hence, the data are divided into three groups, an upper group of predominantly anomalous values, a lower group of predominantly background values, and an intermediate group containing both anomalous and background values. Of the 158 values, about 23 are anomalous and 135 are background; 80 percent, or about 18 of the anomalous values are above the 70 ppm threshold, as are one or two background values. The remaining 5 values are, for all practical purposes, contained in the intermediate group with about 11 background values. The lower group consists of 91.5 percent of the background samples, i.e. about 124 values.

FIGURE VII-2

Partitioned log probability plot of 158 Cu determinations on stream sediments, Mt. Nansen area, Yukon. See figure V-1 for explanation of symbols.

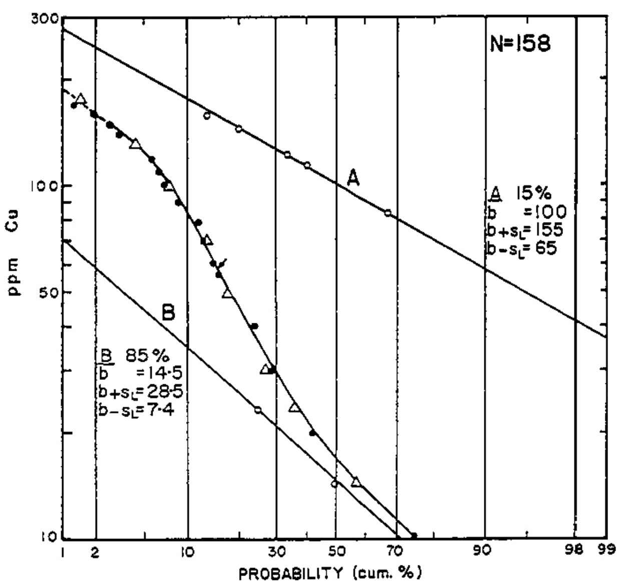


TABLE VII-2
ESTIMATED PARAMETERS OF PARTITIONED POPULATIONS, Cu IN STREAM SEDIMENTS, MT. NANSEN AREA — YUKON TERRITORY

Population	Proportion %	No.	Values in ppm Cu		
			b	$b + s_L$	$b - s_L$
A Anomalous	15	24	101	155	63
B Background	85	134	14.7	28.5	7.4
A + B	100	158			

This procedure has permitted definition of two ranges of ppm Cu that contain anomalous values, and to which priorities can be assigned for follow-up investigation. Values above 70 ppm Cu have top priority because they are almost entirely anomalous. Second priority is assigned to the 16 values between 37 and 70 ppm Cu because only 5 are anomalous.

Theoretically, individual values in the intermediate range cannot be assigned to either A or B populations. In practice, however, many, if not all, can commonly be recognized with a fair degree of certainty. Colour coding of values on a plan, according to their respective groups, permits recognition of those intermediate range samples that occur downstream from recognizable anomalous values. In many cases, virtually all anomalous samples in the intermediate range can be identified. A comparable procedure can be used in dealing with soil or whole rock analyses for which two thresholds are recognized. In this way, follow-up examination of second priority values can be cut to a minimum.

VII-3: Ni IN SOILS, NEAR HOPE, BRITISH COLUMBIA.

Figure VII-3 is a log probability graph of 166 Ni analyses of soils obtained from a grid superimposed on a known Cu-Ni showing related to ultramafic rocks enclosed in regionally metamorphosed fine-grained sedimentary strata near Hope, British Columbia. A smooth curve drawn through the plotted points indicates the presence of at least three populations, by inflection points at the 5.5 and 25 cumulative percentiles (see section VI-1). Two populations, A and C, were partitioned using techniques described in section IV-3. Estimation of the third population, B, was done as described in chapter VI, using the relationship

$$P_M = f_A P_A + f_B P_B + f_C P_C$$

where PM is the probability read from the real data curves, f_A, f_B, and f_C are proportions determined from inflection points, and P_A , P_B and P_C are probabilities on the three partitioned populations. Populations A, B, and C, were then combined ideally in the proportion 5.5 A, 19.5 B, and 75 C, at a number of Ni ppm levels. These ideal points are shown in figure VII-3 as open triangles that virtually coincide with the real data curve.

FIGURE VII-3

Partitioned log probability graph of 166 soil (B-horizon) Ni values from an area near Hope, B. C. that includes a small Cu-Ni showing. See figure V-1 for explanation of symbols.

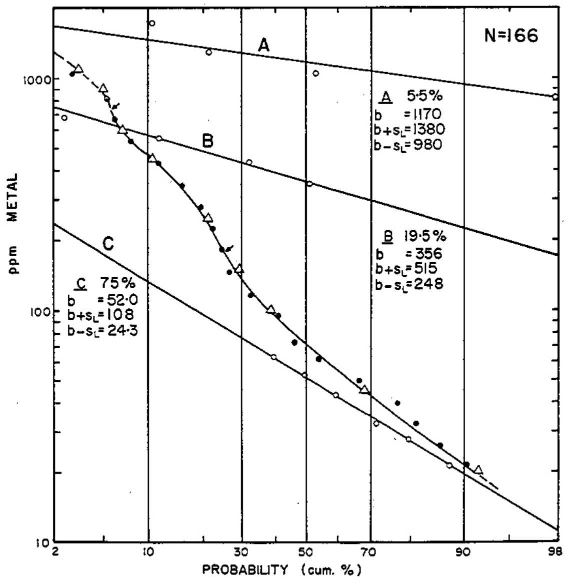


TABLE VII-3

ESTIMATED PARAMETERS OF PARTITIONED POPULATIONS, Ni IN SOILS, HOPE AREA, B.C.

Population	Proportion %	No. of samples	Values in ppm Ni b	b + s_L	b - s_L
A: Anomalous	5.5	9	1170	1380	980
B: Background (Ultramafic)	19.5	32	356	515	248
C: Background (Metaseds)	75	125	52	108	24.5
	100	166			

Population A is obviously not well defined, as indicated by the scattering of points used to define it. The main reason for this is that population A is only a small proportion of the total data and represents a very small sample on which to base the estimate of the entire population. Populations B and C, on the other hand, appear reasonably well defined. Estimated parameters are given in table VI-3. Populations B and C overlap somewhat and two thresholds are necessary. These thresholds are taken arbitrarily at the 2 cumulative percentile of population C (i.e. 236 ppm Ni) and the 98 cumulative percentile of population B (i.e. 170 ppm Ni). The three thresholds, reproduced in table VII-4, divide the data into four groups, three of which consist principally of single populations. (See table VI-3).

TABLE VII

ESTIMATED THRESHOLDS, Ni IN SOILS — HOPE AREA, BRITISH COLUMBIA

Threshold (ppm Ni)	Principal Content of Group
	Almost exclusively A Population
780	
	Almost exclusively B Population
236	
	Combination of B and C Populations
170	
	Almost exclusively C Population

Threshold values can be used as contour values on a plan of the grid, or as a basis to code sample locations by colours or symbols to aid in interpretation. In this case, population A is related to Ni-Cu mineralization and is therefore interpreted as anomalous. Population B corresponds to areas underlain by ultramafic rocks and population C to areas underlain by metasedimentary rocks.

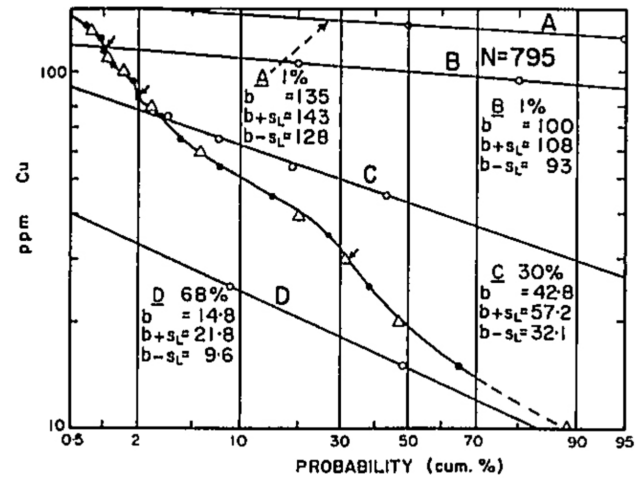
This example illustrates the arbitrary nature of threshold choice. Thresholds between, B and C population could equally well have been chosen, using 1 and 99 cumulative percentiles of the C and B populations respectively, as was done in the Mount Nansen example.

VII-4: Cu IN SOILS, SMITHERS AREA, BRITISH COLUMBIA

A probability plot of 795 soil copper analyses from an area near Smithers, British Columbia, is shown in figure VII-4. This. quantity of data is characteristic of that obtained from reconnaissance exploration surveys where large quantities of information are gathered during a relatively short time interval. The area sampled is underlain by acid to intermediate intrusive rocks, that cut a thick volcanic sequence.

FIGURE VII-4

Partitioned log probability plot of 795 soil Cu analyses from an area near Smithers, B. C. See figure V-1 for explanation of symbols.



Inflection points are evident on the curve at approximately the 1, 2, and 32 cumulative percentiles (see chapter VI). These populations can be estimated by partitioning the curve in stages. In this case, it is convenient to begin with population C for which most data points are available. Once C has been defined, D can be estimated, using, C and the original data curve. Both C and D populations can be defined reasonably well. The upper two populations, A and B, can be approximated roughly but cannot be delineated with much accuracy because of the small percentage of total data that each represents. Crude estimates of A and B are shown, based on available data.

A number of check points, shown as open triangles on the original data curve, were calculated for the partitioned populations A, B, C, and D, combined in the proportions 1:1:30:68. These points agree almost perfectly with the original curve, suggesting that the partitioning results are a plausible model for the data.

Comparison of data with a geological map suggests that populations C and D represent background Cu in soils over volcanic and plutonic rocks respectively. In the same way, A and B were interpreted as anomalous populations over volcanic and plutonic rocks respectively.

In choosing thresholds for distinction between anomalous and background values, there is no need, in this case, to consider either populations A or D. The critical part of the graph is the range of overlap of populations B and C. About

1.5 percent of the data, or 12 values, are above 100 ppm. Of these 12 values, 11 are anomalous, and 1 belongs to population C. Virtually all the anomalous values are above 85 ppm. The interval 85 to 100 ppm contains 5 anomalous and about 2 or 3 values from C population. Hence, two threshold values have been defined that contain all anomalous values and only a small number of background values.

There are several important points illustrated by this example, as follows:

1) In polymodal distributions it is not always necessary to begin partitioning at an end of the probability curve.
2) It is generally a wise procedure to carry through with a complete partitioning, even where this is not essential to the specific problem on hand.
3) Even in cases where specific populations cannot be defined with certainty, useful thresholds can be obtained, at least as useful as those resulting from other techniques in common use.
4) Inflection points on a probability curve based on abundant data are probably real, even if they indicate populations present in small proportions, and should be considered in the partitioning process.
5) An alternative approach would have been to group the data initially on the basis of the two predominant underlying rock types. This procedure was not followed because an interpretation deemed adequate for the problem was obtained by the simpler method used.

The foregoing examples show that the major advantage of probability plots is to provide a useful grouping of data. Commonly, this grouping is not simply for the purpose of obtaining thresholds between anomalous and background populations, but more generally is to derive thresholds between populations that aid in a general interpretation of the significance of the data.

TABLE VII-5
ESTIMATED PARAMETERS OF PARTITIONED POPULATIONS, Cu IN SOILS, SMITHERS AREA, BRITISH COLUMBIA.

Population	Proportion %	N	Values in ppm Cu		
			b	$b + s_L$	$b - s_L$
A	1	8	135	143	128
B	1	8	100	108	93
C	30	239	42.8	57.2	32.1
D	68	540	14.8	21.8	9.6
A+B+C+D	100	795			

CHAPTER VIII
ADDITIONAL TOPICS

VIII-0: GENERAL STATEMENT

This chapter contains a number of special procedures, examples and summaries that do not fit readily into the framework of previous sections. It is a "catch-all" chapter, devoted to specific topics that elucidate and complement material presented thus far.

VIII-1: CUMULATIVE PLOTS OF DATA CONSISTING OF A SMALL NUMBER OF VALUES.

Data cannot be used to construct a cumulative plot in the manner thus far outlined if they include only a small number of values. An appropriate method for the plotting of such data on probability paper described in some statistics texts (e.g. Schmitt, 1969), involves the cumulative frequencies of individual values rather than cumulative frequencies of class intervals as done heretofore.

As an extreme example of this technique consider the average gold assays of total production from each of 19 vein deposits in Ainsworth Mining Camp, British Columbia. Data are listed in the table VIII-1 and plotted in figure VIII-1. Note that in this case logarithms of assay values are plotted on arithmetic probability paper, because the arithmetic values cover four orders of magnitude and a log probability plot would have been unwieldy using commercially available graph paper.

FIGURE VIII-1

Probability graph of logarithms (base 10) of average gold grades of 19 vein deposits, Ainsworth Camp, B. C. Values are cumulated individually.

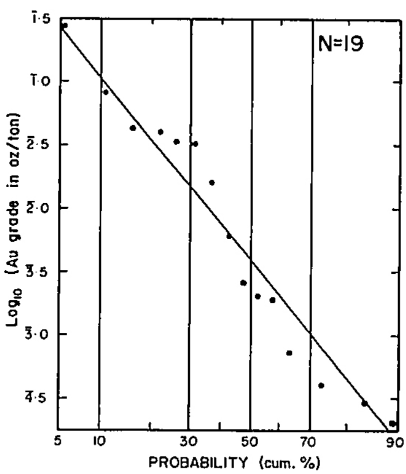


TABLE VIII-1
CUMULATIVE PERCENTAGE DATA FOR KNOWN Au GRADES OF MINERAL DEPOSITS IN AINSWORTH MINING CAMP, BRITISH COLUMBIA

Average Au Grade		Number of	Cumulative
Oz/Ton	**Log $_{10}$**	**Mines**	**Percent**
0.0001	$\bar{4}.000$	2	100
0.0002	$\bar{4}.301$	1	89.5
0.0003	$\bar{4}.477$	2	84.2
0.0004	$\bar{4}.602$	2	73.7
0.0007	$\bar{4}.845$	1	63.2
0.0019	$\bar{3}.279$	1	57.9
0.002	$\bar{3}.301$	1	52.6
0.0035	$\bar{3}.544$	1	47.4
0.0059	$\bar{3}.771$	1	42.1
0.016	$\bar{2}.204$	1	36.8
0.0325	$\bar{2}.512$	1	31.6
0.0342	$\bar{7}.534$	1	26.3
0.0416	$\bar{2}.619$	1	21.1
0.043	$\bar{2}.634$	1	15.8
0.0814	$\bar{2}.911$	1	10.5
0.2758	$\bar{1}.441$	1	5.3

Scatter of the 15 plotted values is considerable but the pattern can be approximated by a straight line, as shown in figure VIII-1. A qualitative interpretation is thus, that no evidence of more than a single population exists although we. cannot entirely exclude such a possibility. Furthermore, as might be expected, the data can be approximated adequately by a lognormal distribution model with parameters $\bar{3}.60 \pm 1.19$, or, in terms of ounces per ton, 0.0040 (0.0617, 0.0003). The graphically determined mean of logarithms, $\bar{3}.60$, compares with $\bar{3}.48$ determined by moment calculations. The agreement is adequate considering the scatter of plotted points about the "estimated" straight line.

The implications of such a plot to explorations considerations are worthy of some mention. Of 74 productive deposits in the area, only 19 have Au content recorded in publically available sources. These 19 values thus represent the total sample on which the entire population must be estimated, and this estimate is an important constituent of any potential ore that might be found in the camp.

VIII-2: REAL DATA CONTAINING A HIGH PROPORTION OF ZERO OR LESS THAN DETECTION LIMIT VALUES — CENSORED DATA

Sixty-seven pyrite samples from Endako molybdenum mine and vicinity, were analyzed spectrographically for twelve minor elements, including Bi (Dawson and Sinclair, 1974). Of these sixty-seven samples, Bi was not detected in twenty-six. A probability plot for Bi values is shown in figure VIII-2 as a curved line. The curve is relatively well defined and has the appearance of a single population plotted over part of the probability range (see section IV-1), in this case the upper 60 percent of the probability range. To check whether or not this curve might represent a single population, a number of points on the smooth curve were recalculated to 100 percent. These are shown in figure VIII-2, where it is apparent that a straight line can be drawn through them without difficulty.

One conclusion from this analysis is that two populations are represented in the data – an upper lognormal population, accounting for 60 percent of the values, lies above the analytical detection limit and is almost perfectly defined by the probability plot. A second population consisting of 40 percent of the data, lies below the detection limit and nothing can be said of its parameters or density distribution. The detection limit is coincidently an efficient threshold separating the two populations.

FIGURE VIII-2

Log probability graph of 67 Bi analyses of pyrites, Endako Molybdenum Mines, B.C. (after Dawson and Sinclair, 1974)

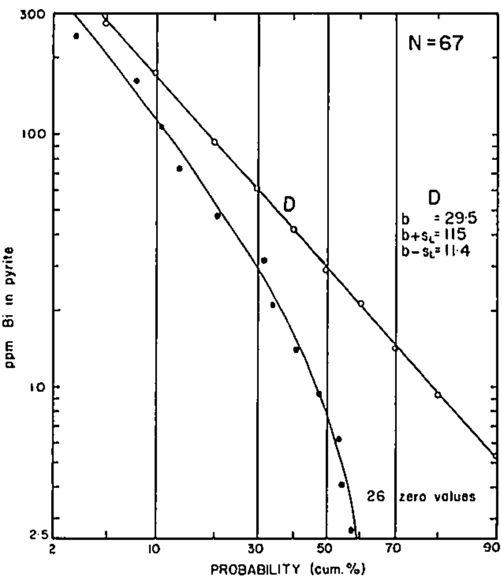


TABLE VIII-2

DATA FOR CALCULATION OF POINTS ON ASSUMED LOWER POPULATION (B), Pb IN PYRITE, MORRISON LAKE PORPHYRY COPPER DEPOSIT

Ordinate Level (ppm Pb)	Read from Graph $P_{(A+B)}$	P_A	Calculated P_B
200	28.5	89	2.57
100	34.3	97.8	7.14
60	37	100	10
40	40	100	14.3
30	42.5	100	17.9
20	45	100	21.4
10	51	100	30

$f_A = 0.30$, $f_B = 0.70$

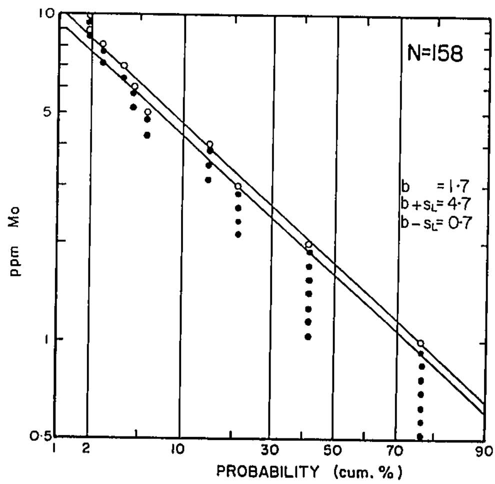
It is important to emphasize that the curvature of the probability graph in this example, based on the total sixty-seven samples, is the clue that two populations exist. If the plot had been straight we would have concluded that a single population was represented. In general, a useful procedure in considering data with a high proportion of zero (or analytically non-detectable) values, is to construct a probability plot based on *all* data, including zeros, and to examine it for presence or absence of curvature. If curvature is present, attempt to estimate the proportion of the population for which data are available, either by recognizing an inflection point, or by trial and error (see also section VIII-4).

VIII-3:. PLOTTING DATA WITH GAPS OR SHORT RANGES

Figure VIII-3 is a plot of Mo analyses of 158 stream sediment samples from the Mount Nansen area, Yukon. Values range from 0 to 10 ppm Mo. In such a case, the use of a bar interval of 1/4s or narrower, provides the type of plot shown by the solid dots of figure VIII-3. Note that the bar interval is significantly less than successive reported values of the variable (integer values in this case). In a normal histogram the resulting unit of pattern is several empty bars followed by a bar containing values, giving rise to the pronounced steplike pattern in figure VIII-3. The natural tendency is to pass a straight line through the centre of gravity of each vertical set of solid points. A little thought, however, shows that this would bias the distribution

FIGURE VIII-3

Log probability plot of 158 Mo values for stream sediment samples, Mt. Nansen area, Yukon. Black dots provide an ambiguous plot obtained by grouping data in classes with an interval much less than the interval between successive recorded values. Open circles are obtained by cumulating individual values.



towards too low a mean value. There are two ways of treating the data that in many cases give identical results. The first, recommended by Lepeltier (1969), is to choose a sufficiently wide bar interval such that no bars of a normal histogram are empty. The second method involves cumulating individual values (see also section VIII-1), and is illustrated by the open circles in figure VIII-3. Note the two lines drawn in figure VIII-3, the lower of which underestimates the mean due to being based on too narrow a bar interval relative to the precision of the analytical method used. This problem exists with much analytical data that has a short range relative to precision, and is particularly common with semi-quantitative chemical data where values are not interpolated between standards. Common geochemical examples include Mo, Hg, and Ag.

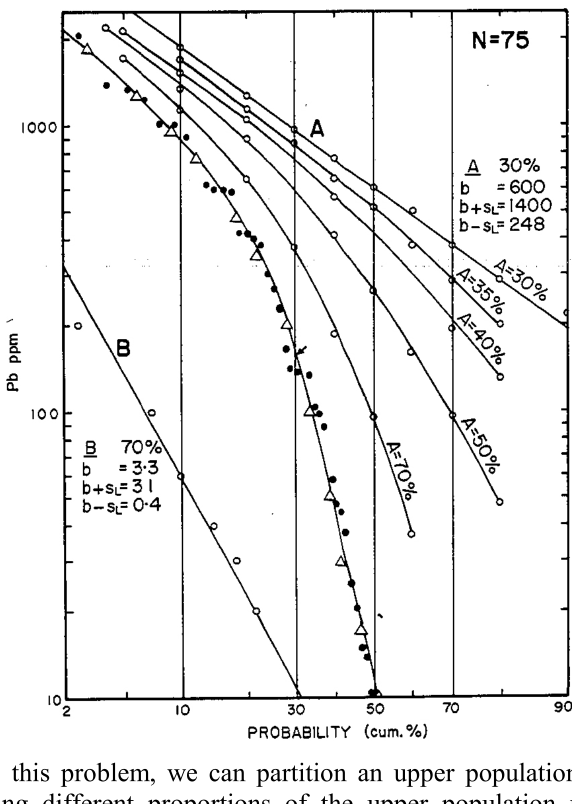
Note in figure VIII-3 that by cumulating data from high towards low values, we can plot ten data points. If data had been accummulated from low values towards high values, we would have precisely the same points except that only nine could be plotted. That is, the 100 cumulative percentile could not be plotted, nor could the zero molybdenum values!

VIII-4: PARTITIONING OF PARTIAL BIMODAL DISTRIBUTION—Pb IN PYRITE, MORRISON LAKE DEPOSIT.

Seventy-five pyrite samples from the Morrison Lake porphyry copper deposit in central British Columbia were analyzed by Wong (1972) for six elements, including Pb. Of the 75 samples, 35 contained Pb below the detection limit and were recorded as zero values. A probability plot of the data (figure VIII-4) is based on individual values, as described in section VIII-1, because of the number of values available. A relatively smooth curve is obtained that indicates the presence of at least two populations if lognormal distributions are assumed. However, an inflection point cannot be picked out easily.

FIGURE VIII-4

Log probability plot of 75 Pb analyses of pyrites, Morrison Lake porphyry copper deposit, B.C. See figure V-1 for explanation of symbols. Partitioning had to be done by a trial and error technique assuming various proportions of the component populations because the inflection point could not be specified with assurance.



To overcome this problem, we can partition an upper population by trial and error, assuming different proportions of the upper population until a linear pattern is obtained. This has been done for a variety of percentages from 70 to 30, and results are shown as curves whose curvature decreased progressively until at 30 percent the distribution is essentially linear. The linear pattern represents our estimate of an upper population, A. Now, assuming that only a single lower population exists, we can estimate it at various ordinate levels using the relationship $P_{(A + B)} = P_A f_A + P_B f_B$. The plotted points define a linear trend reasonably well. Parameters of the two populations can now be estimated. In addition, thresholds that group the data and provide useful contour values can be estimated, as outlined in sections VII-1.

VIII-5: RAPID METHODS FOR FIELD APPLICATION OF CUMULATIVE PROBABILITY PLOTS.

One of the major advantages of cumulative probability plots is their simplicity which leads to relative ease of application in the field. This is particularly true for geochemical and certain types of geophysical data. Lepeltier (1969) has treated the topic at some length.

Equipment necessary for field applications includes, in addition to logarithmic and arithmetic probability paper, a slide rule and, on occasion, a set of log tables. A straight edge and 'fish tail' curve are useful for interpolation.

In order to plot a probability graph, it is necessary to group data into classes of equal interval. In Chapter I, a class interval in the range one-quarter to one-half the standard deviation was recommended. In most cases, the vast majority of data is contained by a range of 2½ standard deviations about the mean. A class interval in the range ¼s to ½s, is equivalent to about ten or twenty groups. Suppose we choose fifteen groups of values as being a reasonable compromise, the appropriate interval for a set of data can then be estimated using the following formula:

$$\text{Log (Interval)} = \frac{Log\ R}{n}$$

where R is the ratio of highest to lowest value in the data set and n is the number of classes. R can be determined by inspection of the data and n should be in the range ten to twenty (say fifteen in most cases).

Log intervals calculated in this manner are not simple decimals but can be rounded to some convenient figure. The writer has found that for geochemical data an arbitrary class level (expressed logarithmically to base 10) of 0.05 or 0.1, depending on the range and magnitude of the data, is appropriate. Use of the above formula is simple and rapid.

Classes must be set up starting at some odd number to avoid the problem of values coinciding with class boundaries. A table can then be arranged as shown in figure VIII-5. Once this table has been organized the number of values in each class must be determined, a somewhat tedious exercise if the number of data values is large. A convenient procedure is to involve two people, one calling out values consecutively, and the other adding counters to appropriate class intervals

FIGURE VIII-5

Example of a tally sheet for field use in preparing probability graphs.

FREQUENCY DATA FOR____________________

Log_{10} interval = 0·1 N = ______

Lower Class Limit		Tally — Histogram	N	Class %	Cum. %
Log_{10}	Arith.	5 10 15 20 25 30 35 40 45 50			
3·33	2138				
3·23	1698				
3·13	1349				
3·03	1076				
2·93	851				
2·83	676				
2·73	537				
2·63	427				
2·53	339				
2·43	269				
2·33	214				
2·23	170				
2·13	135				
2·03	107				
1·93	85·1				
1·83	67·6				
1·73	53·1				
1·63	42·7				
1·53	33·9				
1·43	26·9				
1·33	21·4				
1·23	17·0				
1·13	13·5				
1·03	10·7				
0·93	8·51				
0·83	6·76	5 10 15 20 25 30 35 40 45 50			
		TOTALS			

Class frequencies as percentages are then easy to calculate, as are cumulative frequencies. The data are then in an appropriate form for construction of histogram, cumulative histograms, probability graphs, etc. Note that the table is arranged in descending order of magnitude and the classes are cumulated from high to low classes, as recommended by Lepeltier (1969) and used throughout this manual.

Cumulative probability plots prepared in the foregoing manner are, for the most part, equally as good as those prepared by more sophisticated approaches, such as computerized methods, but are obviously more tedious and thus more subject to human error. One advantage of such plots is that abnormalities in the curves can be identified early and checked for possible error, thus increasing efficiency of field work.

VIII-6: CONSTRUCTING AN HISTOGRAM FROM PROBABILITY PLOTS OF REAL AND HYPOTHETICAL DATA.

Generally speaking, cumulative plots are more explicit and meaningful than are histograms. In some cases, however, it is desirable to present data in histogram form, particularly as a means of presenting information to those unfamiliar with probability plots. Cumulative percentages are read and tabulated for intervals of ¼s, disposed symmetrically about the mean or mode. From this table the proportions of data within each class interval can be calculated and the results used to plot a histogram. In the case of a polymodal distribution, a minor problem arises in choosing an appropriate bar interval. An adequate choice, normally made by inspection, is to choose a bar interval that provides 15 to 20 bars. When a probability curve, based on real data, is used to construct a histogram, the resulting histogram can differ slightly in detail from that plotted directly from original data because of (1) a smoothing imposed on the probability graph, and (2) limitations of graphical interpolation.

VIII-7: PLOTTING CONFIDENCE LIMITS.

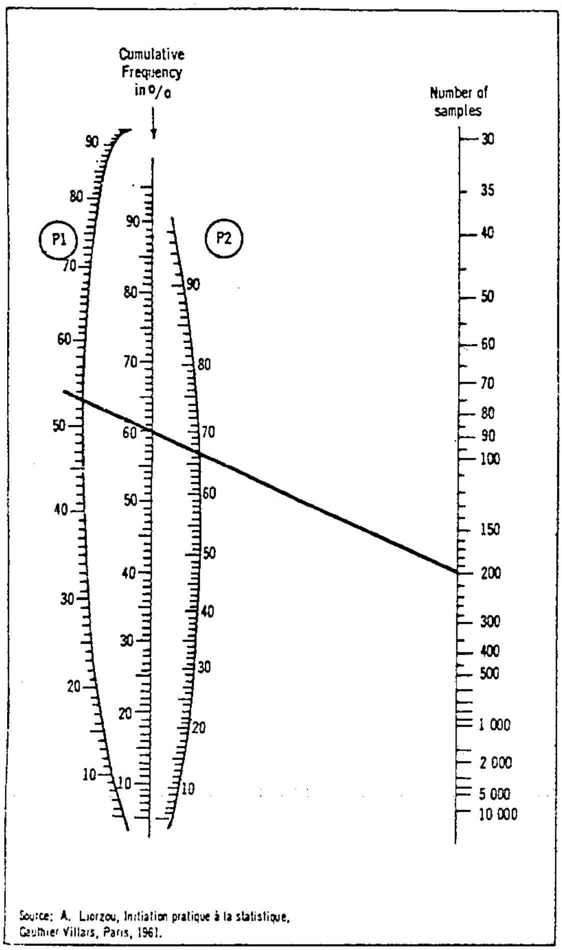
In some cases, it is desirable to plot confidence limits of a probability curve. This can be done graphically using a nomogram devised by Liorzou (1961), and reproduced from Lepeltier (1969) in figure VIII-6. Examples of its use are given in figures II-1, III-3, and III-4.

A single point on curves, defining the 95 percent confidence interval, is estimated by holding a straight edge on the number of samples and the cumulative percentage at a given ordinate level. Two values are then read from the graph and plotted. A number of such points are plotted and two smooth curves drawn to define the 95 percent confidence belt.

This technique is useful for a graphical comparison of two or more probability curves. It has been used by Woodsworth (1972) as a quantitative means of recognizing the presence of a "significant" change in slope of a single probability curve. If a point on a curve plots outside the 95 percent confidence belt, significant curvature is assumed.

FIGURE VIII-6

A nomogram for estimating the 95 percent confidence limits for normal distributions plotted on probability paper (after Lepeltier, 1969).



VIII-8: SUMMARY OF ADVANTAGES AND LIMITATIONS IN THE USE OF PROBABILITY PLOTS.

The main purpose of this manual has been to stress possible applications of probability paper in the analysis of various types of mineral exploration data. At the same time, the writer has attempted, in appropriate places, to indicate difficulties that arise. It seems useful, however, to summarize in one section, the general advantages and limitations of this simple graphical technique.

Advantages:

1) A simple form of graphical representation of data.
2) Rapid, qualitative analysis of density distributions.
3) Rapid estimation of parameters of normal and lognormal distributions.
4) Compact graphical representation of several sets of data on a single diagram.
5) Recognition of polymodal distributions.
6) Partitioning of polymodal distributions.
7) Estimation of thresholds and fundamental grouping of data.
8) Rapid recognition of certain abnormalities in data.
9) Specific applications (e.g. error representation in geochemical data, probability of success, etc.)

Disadvantages:

1) Data might not approximate normal or lognormal density distributions.
2) Data might be too sparse for meaningful analysis on probability paper. Some authors recommend a minimum of 100 values, but this is an arbitrary figure and many of the plots in this text based on as few as 70 data values appear valid and useful.
3) Tails of cumulative distributions commonly are not well-defined. This can cloud interpretation in the upper value range which is of particular interest in most mineral exploration data.
4) A small but significant proportion of plots appear to be uninterpretable, using procedures outlined in this text. This is particularly true if several populations are present with extensive overlap and/or different types of density distributions. Except in very special cases, four populations are about the maximum that can be treated successfully, and difficulties are not uncommon if three or four populations are present.
5) The lower the quality of data, the more ambiguous is the interpretation of the corresponding probability plot.

The most significant use of probability plots applied to mineral exploration data, is in therecognition of the number of populations in a data set, and the partial or complete partitioning of individual values into their respective groups or populations. Interpretation of the significance of the resulting groupings of data is then up to the user. Throughout the text the term anomalous has been used commonly to denote a high population, and to distinguish it from low background population(s). High populations *are not always* anomalous and their designation as such constitutes an interpretation.

VIII-9: SOME USEFUL HINTS ON PROCEDURE

1) Original data should be plotted on appropriate probability paper (arithmetic or lognormal), clearly and accurately.
2) Interpretations can be attempted on transparent overlays: In some cases, it is advantageous to plot original data on transparent probability paper.
3) Retain tabulated data used in the original plot and label it clearly to provide a . rapid check on details of the original plot and interpretations, if the need arises.
4) Probability plots should be labelled immediately as they are.constructed with all information pertaining to the partitioned populations' tabulated on them.
5) Never attempt to interpret and/or partition bimodal probability plots on the basis of assumed straight line segments. The curvature in such plots is an important element of interpretation.
6) Label assumed inflection points because they are not always obvious, particularly on a draughted copy.
7) A standardized procedure for labelling is useful (see below).
8) A complete partitioning and interpretation should always be done even where not apparently necessary, for a specific problem. The. time involved is slight as experience is gained and unexpected results are commonly forthcoming.

A standardized approach to plotting and labelling diagrams increased their clarity and avoids ambiguity. In this manual, filled black circles are used to represent original data, open circles indicate points used to estimate partitioned populations, and open triangles are check points. The use of different symbols for different data sets, plotted on the same graph paper, of course, avoids ambiguity.

Routine procedure in preparing and plotting data should include – (a) an examination of data to choose an appropriate bar interval, (b) grouping of the data into intervals and calculation of cumulative percentages for each interval, beginning with large values, and (c) plotting data on appropriate probability paper (normal, lognormal) with an appropriate choice of ordinate scale in the case of arithmetic probability plot.

If a single population is indicated, a straight line should be fitted to the plotted points and parameters estimated. If desired, confidence limits can be determined graphically and plotted. If plotted points have a curved pattern, a smooth curve should be drawn through them, with particular attention paid to possible inflection points. The form of the curve should be examined closely to ascertain the most probable interpretation of the number of populations represented, the possibility of truncation, etc.

Partitioning should be done following the appropriate procedures outlined in the text. A series of checks should be made on the par, titioning procedure by calculating ideal combinations of the partitioned populations for a number of ordinate levels. If these numerous check points essentially coincide with the real data curve, a plausible model for the data has been obtained by partitioning.

One can then proceed to the interpretation stage, perhaps using thresholds chosen on the basis of partitioned populations. For mineral exploration data, colour or symbol coding of groups of values defined by thresholds on a plan, commonly provides a useful means of analyzing the geological significance of each group.

BIBLIOGRAPHY

Ahrens, L. A., 1954. The lognormal distribution of the elements; Geochim. Cosmochim. Acta, Vol. 5, pp. 49-73.

Aitchison, J., and J. A. C. Brown, 1957. The lognormal distribution; Cambridge University Press.

Becker, R. M., and S. W. Hazen, Jr., 1961. Particle statistics of infinite populations as applied to mine sampling; U. S. Bur. Mines. Rept. Invest. 5669, 79 p.

Bianconi, F., and R. Saager, 1971. Reconnaissance mineral exploration in the Yukon Territory, Canada; Schweiz. Mineral. and Petrog. Mitt. Vol. 51, No. 1. pp. 139-154.

Bolviken, B., 1971. A statistical approach to the problem of interpretation in geochemical prospecting; Can. Inst. Min. Met., Spec. Vol. No. 11 pp. 564-567.

Bolviken, B., and R. Sinding-Larsen, 1972. Total error and other criteria in the interpretation of stream-sediment data; in Geochemical Exploration 1972, pp. 285-295.

Bolviken, B., 19.73. Statistical description of geochemical data; Norges Geol. Unders; No. 285, pp. 1-10 [in Norwegian with English summary].

Brabec, D., and W. H. White, 1971. Distribution of copper and zinc in rocks of the Guichon Creek batholith, British Columbia; Can. Inst. Min. Met., Spec. Vol. No. 11, pp. 291297.

Buchanan-Wollaston, H. J., and W. C. Hodgson, 1929. A new method of treating frequency curves in fishery statistics; Jour. Conseil. Int. Expl. Mer., Copenhagen; Vol. 4, pp. 207-225.

Cassie, R. M., 1954. Some uses of probability paper in the analysis of size frequency distributions; Aust. Jour. Marine and Freshwater Res., Vol. 5, pp. 513-522.

Clark, I., and R. H. T. Garnett, 1974. Identification of multiple mineralization phases by statistical methods; Trans. Inst. Min. Metall., Vol. 83, pp. A43-A52.

Davidson, R. A., 1972. A computer-oriented analysis of metal production from Ainsworth mining camp, southeastern B. C.; Thesis, Dept. of Geol., University of B. C., Vancouver, B. C. 60 p.

Dawson, K. M., and A. J. Sinclair, 1974. Factor analysis of minor element data for pyrites, Endako molybdenum mine, B. C., Canada; Econ. Geol. Vol. 69, pp. 404-411.

de Grys, A., 1964. Copper distribution patterns in soils and drainage in Central Chile; Econ. Geol., Vol. 59, p. 636-646.

de Grys, A., 1962. Seasonal variations in copper content in some Andean streams of Central Chile; Econ. Geol., Vol. 57, pp. 1031-1044.

Doust, J. F., and H. J. Josephs, 1941-42. A simple introduction to the use of statistics in telecommunication engineering; Post Office Eng. Jour. Vol. 34, pp. 36-41, pp. 78-84, pp. 139-144, and pp. 173178.

Finney, D. J., 1941. On the distribution of a variate whose logarithm is normally distributed; Jour. Roy. Stat. Soc., Suppl. 7, No. 2. pp. 155-161.

Goodrich, R. D. 1927. Straight line plotting of skew frequency data; Amer. Soc. Civil Eng. Trans. Vol. 91., pp. 1-43, Discussion pp. 44118.

Govett, G. J. S., W. D. Goodfellow, R. P. Chapman, and C. Y. Chork, 1975. Exploration geochemistry – distribution of elements and recognition of anomalies; Jour. Math. Geol., Vol. 7, pp. 415-446.

Green, E. J., 1963. Some comments on the bi-normal frequency distribution; Geochim. Cosmochim. Acta, Vol. 27, pp. 1071-1075.

Gumbel, E. J., 1958. Statistics of extremes; Columbia Univ. Press, New York, 375 pages (Esp. pp. 28-41).

Hald, A., 1952. Statistical theory with engineering applications; John Wiley and Sons, New York, 97 p.

Harding, J. P., 1949. The use of probability paper for the graphical analysis of polymodal frequency distributions; Jour. Marine Biol. Assoc. U. K., Vol. 28, pp. 141-163.

Harris, S. A., 1958. Probability curves and the recognition of adjustment to depositional environment; Jour. Sed. Pet., Vol. 28. pp. 151-163.

Hawkes, H. E., and J. S. Webb, 1962. Geochemistry in mineral exploration; Harper and Row, 415 p.

Hazen, A., 1914. Storage to be provided in impounding reservoirs for municipal water supply; Proc. Amer. Soc. Civil Eng., Vol. 39, pp. 1943-2044.

Hodgson, W. A., 1972. Optimum spacing for soil sample traverses; 10th APCOM Symp., pp. 75-78, Ed., South African Inst. Min. Met., Johannesburg.

Hubaux, A., 1961. Representation graphique des distributions d'olioelements; Ann. Soc. Geol. de Belg., T. 84, pp. 279-296. (in French). Koch, G. S., Jr. and R. F. Link, 1970. Statistical analysis of geological data, Vol. 1; John Wiley and Sons, New York, 375 p.

Kotller, F., 1950. The distribution of particle sizes (part II); Journ.

Franklin Inst., Vol. 250, pp. 419-441.

Krige, D. G., 1960. On the departure of ore value distributions from the lognormal model in South African gold mines; S. African Inst. Min. Metall. Jour., Vol. 61, pp. 231-244.

Krumbein, W. C., and F. A. Graybill, 1965. An introduction to statistical models in Geology; McGraw Hill Pub. Co., New York, 475 p.

Krumbein, W. C., and F. J. Pettijohn, 1938. Manual of sedimentary petrography; Appleton Century Co., New York, 549 p.

Lenz, C. J., 1972. Evaluation of geochemical data; Proc. of 10th AP

COM Symp., South African Inst. Min. Met. Johannesburg.

Lepeltier, C., 1969. A simplified statistical treatment of geochemical data by graphical representation; Econ. Geol. Vol. 64, pp. 538-550.

Levi, F., 1946. Graphical solutions of statistical problems; The Engineer, London, Vol. 182. pp. 338-340 and pp. 362-364.

Liorzou, A., 1961. Initiation pratique a la statistique; Gathier Villars, Paris.

McCrossan, R. G., 1969. An analysis of size frequency distributions of oil and gas reserves of Western Canada; Can. Jour. Earth Sci., Vol. 6, pp. 201-211.

Montgomery, J. H., D. R. Cochrane and A. J. Sinclair, 1975. Discovery and exploration of Ashnola porphyry copper deposit, near Keremeos, B. C.; a geochemical case history; in Geochemical Exploration 1974, pp. 85-100, Elsevier Pub. Co., Amsterdam.

Oriel, W. M., 1972. Detailed bedrock geology of the Brenda coppermolybdenum mine, Peachland, British Columbia; Thesis, Dept. of Geology., University of B. C., Vancouver, B. C. 94 p.

Orr, J. F. W., 1971. Mineralogy and computer-oriented study of mineral deposits in Slocan City camp, Nelson Mining Division, British Columbia; Thesis, Department of Geology, University of B. C., Vancouver, B. C., 143 p. 93

Orr, J. F. W., and A. J. Sinclair, 1971. A computer-processable file for mineral deposits in Slocan and Slocan City areas of British Columbia; Western Miner, Vol. 44, pp. 22-34.

Otto, G. H., 1939. A modified logarithmic probability graph for the interpretation of mechanical analysis of sediments; Jour. Sed. Pet., Vol. 9, pp. 62-76.

Parslow, G. R., 1974, Determination of background and threshold in exploration geochemistry; Jour. Geochem. Expl. Vol. 3, pp. 319-336.

Rissik, H., 1941. Probability graph paper and its engineering applications; The Engineer, London, Vol. 172, pp. 276-278 and pp. 296-298.

Rodionov. D. A., 1961, On the lognormal distribution of the elements in igneous rocks; Geokhimiya, No. 4, pp. 324-327.

Saager, R., and F. Bianconi, 1971. The Mount Nansen gold-silver deposit, Yukon Territory, Canada; Mineral. Deposita., Vol. 6, pp. 209-224.

Saager, R., and A. J. Sinclair, 1974. Factor analysis of stream sediment geochemical data from the Mount Nansen area, Yukon Territory, Canada; Mineral. Deposita, Vol. 9, pp. 243-252.

Schmitt, S. A., 1969. Measuring uncertainty; an elementary introduction to Bayesian statistics; Addison-Wesley Pub, Co., Reading, Mass., 400 p. (esp. p. 210-215).

Shaw, D. M., 1964. Interpretation geochimique des elements en trace dans les roches cristallines; Masson et Cie, Paris.

Shaw, D. M., 1961. Element distribution laws in geochemistry; Geochim. Cosmochim. Acta, Vol. 23, pp. 116-134.

Sinclair, A. J., 1975. Identification of multiple mineralization phases by statistical methods – a discussion; Trans. Inst. Min. Metall., Vol. 84, pp. A77-A78.

Sinclair, A. J., 1974(a). Selection of thresholds in geochemical data using probability graphs; Jour. Geochem. Expl., Vol. 3, pp. 129-149.

Sinclair, A. J., 1974(b). Probability graphs of ore tonnages in mining camps - guide to exploration; Bull. Can. Inst. Min. Met., Vol. 67, pp. 71-75.

Sinclair, A. J., 1972. Some statistical applications to problems in mineral exploration; B. C. Prof. Eng., Vol. 23, pp. 14-23.

Slichter, L. B., 1955. Geophysics applied to prospecting for ores; Econ. Geol. 50th Ann. Vol., Part II, pp. 885-969.

Spencer, D. W., 1963. The interpretation of grain size distribution curves of clastic sediments; Jour. Sed. Pet., Vol. 33, pp. 180-190.

Tanner, W. F., 1962. Detachment of gaussian component from zig-zag curves; Jour. App. Meterology, Vol. 2, pp. 119-121.

Tanner, W. F., 1958. Sample components obtained by the method of differences; Jour. Sed. Pet., Vol. 29, pp. 204-211.

Tennant, C. B., and M. L. White, 1959. Study of the distribution of some geochemical data; Econ. Geol., Vol. 54, pp. 1281-1290.

Trimble, R. J., 1972. Mineralogy and computer-oriented study of the Eagle vein complex, Liard Mining District, Northeastern B. C.; Thesis, Dept., of Geology, University of B. C., Vancouver, B. C., 78 p.

Wilding, I. G. P., D. Sampey and M. J. Erickson, 1970, Use of cumulative frequency plots to facilitate interpretation of geochemical data; Proc. 2nd Sem. in Geochem. Prospect. Methods and Tech., United Nations Mineral Resource Develop, Series, No. 38, pp. 361-368.

Williams, X. K., 1967. Statistics in the interpretation of geochemical data; New Zealand Jour. Geol. Geoph., Vol. 10, pp. 771-797.

Woodsworth, G. J., 1972. A geochemical drainage survey and its implications for metallogenesis, Central Coast Mountains, British Columbia; Econ. Geol., Vol. 66, pp. 1104-1120.

Wong, S. S., 1972. Minor Element study of pyrite from the Morrison Lake porphyry copper deposit, Central British Columbia; Thesis, Department of Geology, University of B. C., Vancouver, B. C. 75 p.