

# **Estimation of Probabilities of Three Kinds of Petrologic Hypotheses with Bayes Theorem**

J. Nicholls  
4204 Chippewa Rd NW  
Calgary, Alberta T2L 1A2  
Canada  
e-mail: nichollj@ucalgary.ca

A shorter version of this manuscript has been published by *Mathematical Geology*.

## Introduction

Petrologic hypotheses are physical-chemical explanations of the causes of variations in chemistry, mineralogy, and rock-type in rock bodies or rock suites. Predicted compositions derived from physical-chemical models (1995) can be compared to measured compositions of rocks and minerals. If the chemical, mineralogical, and rock-type data are consistent with the model, then the model is a viable hypothesis for explaining the causes of the variations, otherwise it is not. Predicted and measured values seldom are equal and one has the problem of defining consistency: How close should predicted and measured values be for consistency? Having established consistency, the next question is how confident or sure are we that the model or hypothesis is true? This latter question has not received much attention in the petrologic literature. Bayes theorem and Bayesian statistical methods can help answer both questions.

Numerical tests of petrologic hypotheses often fall into one of three categories: (1) Does a set of data points that are expected to have equal values have a variance greater than can be expected from analytical uncertainty? (2) Are the mean values of two sets of data points equal? (3) Does a data point fall on a trend predicted by a physical-chemical model? Examples discussed in this paper include estimates of the probability that a set of lava flows originated from a single magma batch and estimates of the probability that a set of rock analyses are samples of a magmatic crystallization sequence.

In symbols, Bayes theorem is:

$$Pr[H_i | D \& I] = \frac{Pr[D | H_i \& I]Pr[H_i | I]}{Pr[D | I]} \quad (1)$$

where  $H_i$  is an expression representing the hypothesis being tested,  $D$  is the data collected to test the hypothesis, and  $I$  is the prior information available about the hypothesis.  $Pr[H_i | D \& I]$  is the probability that  $H_i$  is correct or true, given the data ( $D$ ) and prior information ( $I$ ).  $Pr[D | H_i \& I]$  is the probability of observing or measuring the data, given that  $H_i$  is correct and given the prior information.  $Pr[H_i | I]$  is the probability of  $H_i$  being correct given the prior information and before taking into account the new data ( $D$ ).  $Pr[H_i | I]$  is called the prior probability.  $Pr[D | I]$  is the probability of observing or measuring the data, regardless of whether  $H_i$  or any other hypothesis is true.

Implicit in Bayes theorem is the idea that there are alternative hypotheses. If there is only one hypothesis and no alternatives, the probability for that hypothesis is one; it must be true if there are no other explanations. Consequently, applications of Bayes theorem become a comparison of two or more hypotheses. To evaluate the probabilities in Bayes theorem, the statements of the hypotheses and data are translated into statistical expressions.

Bayes factors are convenient devices for expressing Bayes theorem if there are two alternative hypotheses. A Bayes factor is:

$$B = \frac{Pr[D / H_0 \& I]}{Pr[D / H_1 \& I]} \quad (2)$$

where the subscripts distinguish the two hypotheses. Another way to express a Bayes factor is:

$$B = \frac{\frac{Pr[H_0 | D \& I]}{Pr[H_0 | I]}}{\frac{Pr[H_1 | D \& I]}{Pr[H_1 | I]}} \quad (3)$$

This last expression, which is a ratio of the ratios of the posterior probabilities to the prior probabilities for the two hypotheses, is the definition used by Jeffreys (1961) but which he labeled K instead of B. In the recent literature,  $B$  is the symbol used to label Bayes factors (e.g. Berger and Delampady, 1987; Jefferys and Berger, 1992).

With Bayes factors we can write Bayes theorem as:

$$Pr[H_0 | D \& I] = 1 - \frac{1 - p_0}{1 + p_0(B - 1)} \quad (4)$$

where  $p_0$ , the prior probability for  $H_0$ , is given by:

$$p_0 = Pr[H_0 | I] = 1 - Pr[H_1 | I] \quad (5)$$

For a fixed value of  $p_0$ , larger values of  $B$  correspond to higher values of  $Pr[H_0 | D \& I]$ . The variation of  $Pr[H_0 | D \& I]$  with  $B$  is shown on Figure 1. Bayes factors greater than one indicate an increased posterior probability for  $H_0$  over its prior probability whereas Bayes factors less than one indicate an increased posterior probability for  $H_1$  over its prior probability. If the Bayes factor is equal to one, the posterior probabilities equal the prior probabilities for the two hypotheses.

## Bayes Factors with Discrete Probabilities

The significance of Bayes factors is more apparent with discrete probabilities than with continuous distributions. The following example, adapted from (Schmitt, 1969), shows the properties of Bayes factors particularly well.

Suppose we have to identify an outcrop as belonging to either rock body X or rock body Y. If the outcrop contains aragonite, the outcrop is part of X. Although aragonite is difficult to identify in hand specimen, we can easily identify carbonate. In the map area, the total outcrop area of X is approximately four times the outcrop area of Y. From a study of thin sections, we know that the probability of finding carbonate in samples of X is 0.9 whereas we find carbonate in samples of Y only 2 out of 10 times. We arrange these data in the form:

<u>X</u>	<u>Y</u>	
0.8	0.2	Relative outcrop areas are the prior probabilities: $Pr(X)$ and $Pr(Y)$
0.9	0.2	Conditional probabilities of finding carbonate: $Pr(C X)$ and $Pr(C Y)$
0.1	0.8	Conditional probabilities of not finding carbonate: $Pr(N X)$ and $Pr(N Y)$

The conditional probabilities are derived from prior information obtained by looking at thin sections of rocks known to come from X or Y.

Multiplying the conditional probabilities for each rock type by the prior probabilities produces the joint probability distribution:

$Pr(C X)Pr(X)$	$Pr(C Y)Pr(Y)$	$Pr(C)$	$=$	$\underline{X}$ 0.72	$\underline{Y}$ 0.04	0.76
$Pr(N X)Pr(X)$	$Pr(N Y)Pr(Y)$	$Pr(N)$		0.08	0.16	0.24
$Pr(X)$	$Pr(Y)$			0.80	0.20	1.00

where  $Pr(C)$  is the probability of finding carbonate and  $Pr(N)$  is the probability of not finding carbonate at the outcrop. The elements of the joint distribution are, as the name implies, the probability of finding carbonate (or not finding carbonate) and, at the same time, of collecting from an outcrop of X (or from Y). For example, the first element of the joint distribution, 0.72, is the probability of finding carbonate while standing on an outcrop of X.

For this example, the formula in Bayes theorem can be written:

$$Pr(X|C) = \frac{Pr(X) Pr(C|X)}{Pr(C)} \quad (6)$$

where:

$$Pr(C) = Pr(X)Pr(C|X) + Pr(Y)Pr(C|Y) \quad (7)$$

The posterior probability of correctly identifying an outcrop as Y, given the presence of carbonate, is:

$$Pr(Y|C) = \frac{Pr(Y)Pr(C|Y)}{Pr(C)} \quad (8)$$

From Jeffreys' (1961) definition [equation (3)] we can obtain an expression for the Bayes factor:

$$B = \frac{Pr(X|C)}{Pr(Y|C)} \quad (9)$$

which is equation (2). The values for the two posterior probabilities,  $Pr(X|C)$  and  $Pr(Y|C)$ , calculated from equations (6) and (8), are 18/19 and 1/19. The Bayes factor [equation (9)] is 18. In other words, geologists finding carbonate in an outcrop have a probability equal to 18/19 ( $\sim 0.95$ ) of being correct if they call it X but have a probability of 1/19 ( $\sim 0.05$ ) of being correct if they call it Y. By finding carbonate, geologists have increased the posterior probability for X over its prior probability by a factor of 18 compared to the posterior probability for Y over its prior probability.

### **Bayes Factors and Continuous Distributions**

Statistical expressions of the hypotheses are usually written with probability density functions. For example:

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (10)$$

is a probability density function for a normally distributed variable,  $x$ .  $\mu$  and  $\sigma$  are parameters (= mean and standard deviation).

Bayes factors for continuous distributions can be calculated from (Berger and Delampady, 1987):

$$B = f(x | \mu)/m_g(x) \quad (11)$$

where:

$$m_g(x) = \int f(x | \theta) g(\theta) d\theta \quad (12)$$

$g(\theta)$  is the prior probability density function for  $\theta$ , with the condition that  $H_I$  is true:

$$g(\theta) = g(\theta | H_I) \quad (13)$$

and the range of integration is over all space available to  $\theta$ , commonly from  $-\infty$  to  $+\infty$ . A justification of equation (11) is given in the appendix.

In summary, the posterior probability of a hypothesis that can be represented by a statistical expression can be calculated if (1) there is at least one alternative and (2) the prior probability density functions for the competing hypotheses can be found.

### **Testing hypotheses represented by variance distributions**

The simplest tests of petrologic hypotheses compare the dispersion of the data with analytical uncertainty. The data can be ratios of isotopes or conserved elements, or even directly measured concentrations. If the data come from cogenetic samples, then we expect properly selected and transformed data sets to have variances the same as the analytical uncertainties. If the variance of the data sufficiently exceeds the analytical uncertainties, the conclusion is straight forward: the sources of the data are not cogenetic. On the other hand, there is no physical-chemical explanation if the analytical uncertainty sufficiently exceeds the variance in the data. Perhaps the reported analytical uncertainty is too large and the analyst, in an effort to be conservative, reported an analytical uncertainty as worse than it really is. Although unlikely, it is possible that we got lucky; the data may be better than we think they are. A third possibility is that the particular density

function used to describe the analytical uncertainties is incorrect.

Suppose we measure  $n$  values of a normally distributed random variable for which we expect the variance of the population to be  $\sigma_0^2$ . The variable,  $(n-1)s^2/\sigma_0^2$ , will also be a random variable but from a  $\chi^2_{(n-1)}$  distribution (Meyer, 1975).  $s^2$  is the sample variance. Consequently, the distribution function for calculating the Bayes factor is:

$$f\left[\frac{(n-1)s^2}{\sigma^2} \mid \sigma_0^2, n\right] = \chi^2_{(n-1)}\left[\frac{(n-1)s^2}{\sigma_0^2}\right] \quad (14)$$

The prior distribution for  $H_1$ ,  $g(\sigma)$ , is (Jeffreys, 1961; Schmitt, 1969):

$$g(\sigma) \propto 1/\sigma \quad (15)$$

Calculating the Bayes factor [equation (11)] gives:

$$B = \frac{(n-1)^{(n-1)/2} (n-3)s^{n-3}}{2^{(n-1)/2} \sigma_0^{(n-2)} [(n-3)/2]!} \exp\left[-\frac{(n-1)s^2}{2\sigma_0^2}\right] \quad (16)$$

Table 1 shows some examples of the use of this particular Bayes factor. The data in Table 1 were transformed from analyses of four picrites and one basalt erupted from Kilauea Volcano, Hawaii in 1968 (Wright, 1971; Wright et al., 1975). Fractionation or accumulation of olivine explains the variations in the data (Nicholls et al., 1986; Nicholls and Stout, 1988). If this hypothesis is correct, then each element ratio listed in the top part of Table 1 should be constant across all five samples because none of the elements in those ratios enters the olivine structure in measurable amounts. Analytical or measurement errors preclude the numbers being the same in all five analyses. However, we can test whether or not the dispersion of the data, as measured

by the standard deviation,  $s$ , of the five values of a particular ratio is different from the expected analytical uncertainty,  $\sigma_0$ .

The Bayes factors for the element ratios range from less than one (Ca/K and Na/K) to nearly 20 (P/K). Both Ti/K and P/K have associated Bayes factors that strongly support the hypothesis that the variance in the data is equal to the square of the analytical uncertainty and that, consequently, the variance in the data is due to analytical uncertainty. For Al/K, Ca/K, and Na/K, the associated Bayes factor is close enough to one that there is no reason to suspect the variance in the data arises from anything but analytical uncertainty, especially because the standard deviation of the data is less than the analytical uncertainty. Jeffreys (1961) suggests the following interpretations for ranges in  $B$ :

	$B > 1$	$H_0$ supported by the data.
$1 >$	$B > 10^{-1/2}$	Evidence against $H_0$ but not worth more than a bare mention.
$10^{-1/2} >$	$B > 10^{-1}$	Evidence against $H_0$ substantial.
$10^{-1} >$	$B > 10^{-3/2}$	Evidence against $H_0$ strong.
$10^{-3/2} >$	$B > 10^{-2}$	Evidence against $H_0$ very strong.
$10^{-2} >$	$B$	Evidence against $H_0$ decisive.

Nicholls and Stout (1988) hypothesized that the five lava flows, which provided the data in Table 1, were related by sorting (fractionation or accumulation of olivine). If the hypothesis is correct, the data should fall on a trend with a slope of two if plotted on a diagram with  $(\text{Fe} + \text{Mg})/\text{K}$  and  $\text{Si}/\text{K}$  as axial ratios (Figure 2). Another way of expressing this consequence is: If the hypothesis is true, the intercepts formed by drawing a line with a slope of two through each data point should be the same for all samples. The statistical equivalent,  $H_0$ , of this characteristic of the data, given that the

hypothesis is correct, is that the expected variance of the intercept values should be equal to the propagated analytical uncertainty. The mean values for the intercepts, X-ratios, and Y-ratios are listed in Table 1. The Bayes factors for the X- and Y-ratios are less than one and favor the alternative hypothesis,  $H_1$ , that the values in each set of ratios are not equal. Because the standard deviations of the data ( $s$ , Table 1) for these two ratios exceeds that expected from analytical uncertainty ( $\sigma_0$ ), it is likely that some process acted to cause the diversity, such as crystal sorting.

In contrast, the standard deviation of the intercept values ( $s = 0.7530$ ) is less than expected from analytical uncertainty ( $\sigma_0 = 5.4275$ ). Although the Bayes factor (0.0273) is very strong evidence against  $H_0$ , we cannot ascribe a physical-chemical cause to the difference. Physical-chemical causes would lead to values of  $s$  greater than  $\sigma_0$ . Rather the cause must be an over-estimation of the analytical uncertainties, the use of an inappropriate probability density function, or by luck, the data points plot closer to a single line with a slope of two than expected – somewhat like winning the lottery. A discussion of whether a normal or gaussian probability distribution function is an appropriate description for the dispersion of the intercept values is given in the appendix.

### **Testing hypotheses with differences in mean values**

This problem is one of testing whether an observation from a probability density function is equal to a given value or not. To see how the Bayes factor is calculated, suppose we measure a set of X's from a density distribution given by:

$$f(x|\mu, \sigma) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2 / n}\right] \quad (17)$$

Next, suppose the prior density function, assuming  $H_I$  is true, is also normal and is given by:

$$g(\theta|\mu, \tau) = \frac{1}{\tau\sqrt{2\pi}} \exp\left[-\frac{(\theta - \mu)^2}{2\tau^2}\right] \quad (18)$$

Integrating the product of equations (17) and (18) with respect to  $\theta$  from  $-\infty$  to  $+\infty$ , after substituting  $\theta$  for  $\mu$  in equation (17), and dividing the result into equation (17) gives (Berger and Delampady, 1987):

$$B = \sqrt{1 + \rho^{-2}} \exp\left\{-\frac{1}{2}\left[\frac{(t - \rho\eta)^2}{1 + \rho^2} - \eta^2\right]\right\} \quad (19)$$

where  $\rho = \sigma/(\tau\sqrt{n})$ ,  $\eta = (\mu_0 - \mu)/\tau$ , and  $t = \sqrt{n}(x - \mu_0)/\sigma$ .

Volcanic products have been used to define time-lines in the stratigraphic column. Correlations between outcrops are based on similarities in chemical parameters (e.g. Oviatt and Nash, 1989). Table 2 contains data from two outcrops of lava flows near Mount Edziza in northern British Columbia (Spooner, 1994). The problem is to decide whether the two outcrops are correlative (*i.e.* are the same age). If the lava flows are all products of the same magma batch, then isochronicity is more likely than if they are not from the same batch. If the lava flows are from different magma batches, then they can be the same age only by coincidence. Samples 141B,

142B, and 146B are from one locality whereas 194B and 197A are from another. Figure 3 shows two tests of the comagmatic hypothesis.

The problem has two parts. First is whether there is evidence that the rocks were formed from more than one batch. Second is whether the rocks from one locality are from one batch and the rocks from the other locality formed from a different batch.

The lava flows are approximately basaltic and if the chemical diversity is due to internal process (*e.g.* crystal fractionation), then sorting of olivine, plagioclase, clinopyroxene, Fe-Ti oxides, and possibly apatite should account for the diversity. The data should fall on a trend with a slope of one on a plot of:

$$[0.25 \text{ Al} + 0.5(\text{Fe} + \text{Mg}) + 1.5 \text{ Ca} + 2.75 \text{ Na}]/\text{K} \text{ versus } [\text{Si} + 1.5 \text{ Ti} + 2.5 \text{ P}]/\text{K}$$

if such a hypothesis is true (see Figure 3). We can look at the dispersion of the intercepts of lines with slope of one drawn through the data points to test the hypothesis. The results of the calculations are shown in Table 2. The Bayes factor for the variance of the intercepts is considerably less than one, suggesting the presence of more than one magma batch in the suite of samples. On the other hand, the dispersion of the P/K values leads to a Bayes factor greater than one ( $B = 6.0742$ , Table 2). Consequently, the P/K values are those expected from a single magma batch. This example demonstrates the occurrence of coincidence. Even though the P/K values are consistent with a single batch hypothesis, their small dispersion does not prove the truth of the hypothesis; petrologic hypotheses can seldom be proven true. The large dispersion in intercept values effectively disproves the single batch hypothesis. These results, however, do not directly answer the specific question of whether the samples from one locality belong to the same batch

as the samples from the second outcrop.

To test the specific question of whether the lava flows from the two localities belong to different magma batches, we need to examine measures of central tendency of the data, say the mean values of the data, for the two localities. We first calculate the weighted means and standard deviations for each location (see Meyer, 1975). These values are listed in Table 2. The variables in equation (19) are assigned as follows. Under  $H_0$ , the expected value of the difference in the weighted means,  $x$ , is zero because  $H_0$  is the hypothesis that the lava flows from the two localities are products of the same magma batch. Hence, we set  $\mu_0$  equal to zero. The observed value of  $x$  for P/K is 0.1101. Under  $H_1$ , the lava flows are from different batches and the most likely value for the difference in the weighted means is the observed difference,  $x = 0.1101$  for P/K. Thus, we set  $\mu$  equal to 0.1101 also.

We are assuming the distributions under both hypotheses are normal. The variance of the difference between two normally distributed variables is (Meyer, 1975):

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \quad (20)$$

Whether or not the samples are from the same magma batch, there is no reason to think the variances under the two hypotheses should be different. The variance arises from the analytical uncertainties, regardless of which hypothesis is true. Consequently, in equation (19), we have:

$$\sigma = \tau \quad (21)$$

We have only one estimate of the difference between the means, consequently  $n$  is equal to one. With these considerations, the Bayes factor becomes:

$$B = \sqrt{2} \exp\left(-\frac{t^2}{2}\right) \quad (22)$$

where  $t = (\mu_1 - \mu_2)/\sigma$ .

For both P/K and the intercept, the Bayes factors are less than one. Neither variable offers any support for  $H_0$  [see Jeffreys' (1961) criteria, p. 10]. An estimate of the probability that  $H_0$  is correct follows from Bayes theorem. If we are initially indifferent to whether the samples from the two localities are from the same magma batch, we can set the prior probabilities equal to 1/2 and apply equation (8) to the P/K value and then to the intercept value. The posterior probability from the first application becomes the prior probability for the second application. The results are:  $Pr[H_0 | D \ \& \ I]$  is 0.2015 after application to the P/K value and 0.0143 after the second application to the intercept value. Consequently, Bayes analysis provides an estimate of a probability near 0.99 that the lava flows from the two localities are from different magma batches.

### **Comparison of Observed and Theoretical Quantities**

Suppose a theory or hypothesis predicts a value for a quantity that can be measured or observed. Seldom will the observed and predicted values exactly agree. We would like to estimate the probability that the observed value,  $x$ , is really equal to the theoretical value,  $\mu_0$ . The hypothesis can be symbolized:

$$H_0: x = \mu_0 \quad (23)$$

The alternative hypothesis,  $H_1$ , is that  $x$  comes from some other probability distribution, say one with mean  $\mu$  and variance  $\tau^2$ . If a normal

distribution characterizes the distribution of  $x$ , regardless of hypothesis, and if the variance of the observed value is known,  $\sigma^2$ , then the Bayes factor is given by equation (19). In summary, if theory predicts a single number, we know how to calculate the Bayes factor.

Often, however, petrologic theory predicts a vector of quantities, say the composition of a solid or fluid solution. Commonly, predicted and observed quantities are shown on rectilinear diagrams such as Harker variation diagrams or Pearce element ratio diagrams. A point on such a graph can be treated as a number in the complex plane (Nicholls, 1990):

$$p = x + i y \quad (24)$$

with a variance given by:

$$\sigma_p^2 = \sigma_x^2 - \sigma_y^2 + 2 i \sigma_{xy} \quad (25)$$

$\sigma_x^2$  and  $\sigma_y^2$  are the variances of the variables plotted on the  $x$ - and  $y$ -axis, respectively and  $\sigma_{xy}$  is the covariance of the uncertainties between the two variables.  $i$  is the imaginary unit  $\sqrt{-1}$ . If  $x$  and  $y$  are normally distributed, then their joint distribution is bivariate normal:

$$f(x, y | \mu_{x0}, \mu_{y0}, \sigma_x, \sigma_y, r) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[ \left( \frac{x-\mu_{x0}}{\sigma_x} \right)^2 - 2r \frac{(x-\mu_{x0})(y-\mu_{y0})}{\sigma_x\sigma_y} + \left( \frac{y-\mu_{y0}}{\sigma_y} \right)^2 \right] \right\} \quad (26)$$

$\mu_{x0}$  and  $\mu_{y0}$  are the means of the two variables and  $r$  is the correlation coefficient. The covariance and variances are related to the correlation coefficient by:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (27)$$

If we assign the coordinates of the point predicted by theory or hypothesis to  $\mu_{x0}$  and  $\mu_{y0}$ , then equation (26) becomes the statistical expression representing our hypothesis:

$$H_0: x = \mu_{x0}, \quad y = \mu_{y0} \quad (28)$$

The alternate hypothesis,  $H_j$ , is for  $x$  and  $y$  to come from a density distribution with different means:

$$g(x, y | \mu_x, \mu_y, \sigma_x, \sigma_y, r) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[ \left( \frac{x-\mu_x}{\sigma_x} \right)^2 - 2r \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \left( \frac{y-\mu_y}{\sigma_y} \right)^2 \right] \right\} \quad (29)$$

but with the same variances and covariance because these quantities are derived from analytical uncertainties and are independent of which hypothesis is true.

Assuming the formula for calculating Bayes factors can be extended by analogy from the univariate to bivariate case, *i.e.* change the functions in equation (11) from univariate ones to bivariate ones, the formula for calculating bivariate Bayes factors is:

$$B = \frac{f(x, y | \mu_{x0}, \mu_{y0})}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y | \alpha, \beta) g(\alpha, \beta | \mu_x, \mu_y) d\alpha d\beta} \quad (30)$$

Performing the indicated mathematical operations gives with equations (26) and (29):

$$B = 2 \exp \left[ - \frac{2(u^2 - 2ruv + v^2) - (p^2 - 2rpq + q^2)}{4(1 - r^2)} \right] \quad (31)$$

where:

$$u = \frac{x - \mu_{x0}}{\sigma_x}, \quad v = \frac{x - \mu_{y0}}{\sigma_y}, \quad p = \frac{x - \mu_x}{\sigma_x}, \quad q = \frac{x - \mu_y}{\sigma_y}$$

Under the conditions of the alternate hypothesis,  $H_j$ , the most likely values for  $\mu_x$  and  $\mu_y$  are the observed or measured values themselves,  $x$  and  $y$ . Making these substitutions simplifies equation (31) considerably:

$$B = 2 \exp \left[ - \frac{(u^2 - 2ruv + v^2)}{2(1 - r^2)} \right] \quad (32)$$

An example is shown on Figure 4. Three samples of lava flows from Volcano Mountain, Yukon (Trupia, 1992; Trupia and Nicholls, 1996) have element ratios shown with filled circles: Xe11, Xe05, and VM01. The question is whether Xe05 and VM01 can be related to Xe11 by fractionation of olivine or clinopyroxene, either separately or together. If olivine is the only fractionating phase, then the data should fall on a line with a slope of one on Figure 4. If clinopyroxene is the only fractionating phase, the data should fall on a line with a slope of minus one. The two phases fractionating together should produce a trend with a slope everywhere between minus one and plus one. The specific path followed by melts formed by fractionating

Xe11 can be calculated through thermodynamic modeling. Fractionation paths are pressure dependent and three possible paths, calculated with the thermodynamic data base constructed by (Ghiorso et al., 1983), are shown as solid lines on Figure 4. The path at 3 GPa falls as close to the data point representing Xe05 as possible, given the thermodynamic data base. Paths at lower and higher pressure fall further from the data point.

Centered at each data point are elliptical contours of equal probability density (Meyer, 1975). All points  $(x,y)$  on the ellipse defined by:

$$k = \frac{1}{(1-r^2)} \left[ \left( \frac{x-x_0}{\sigma_x} \right)^2 - 2r \frac{(x-x_0)(y-y_0)}{\sigma_x \sigma_y} + \left( \frac{y-y_0}{\sigma_y} \right)^2 \right] \quad (33)$$

where  $k$  is a parameter that characterizes the contour, have the same probability density. Points inside the contour have a greater probability density than  $(x,y)$  whereas points outside the contour have a smaller probability density. The point  $(x_0,y_0)$  is the data point itself. Larger values of  $k$  characterize contours with smaller probability densities. Consequently, the most likely point on the fractionation path, the point predicted by theory, to equal the observed value is the tangent point of the path to a contour characterized by a particular value of  $k$ . In other words, given a fractionation path we calculate a value of  $k$  such that the elliptical contour is just tangent to the fractionation path.

Calculation of  $k$  and the coordinates of the point of tangency  $(X_p, Y_p)$  is relatively straight forward if the fractionation path is or can be approximated by a straight line. Given the equation of the fractionation curve, evaluated at  $(X_p, Y_p)$ , in the form:

$$Y_t = \lambda X_t + \beta \quad (34)$$

one can set the slope of the line,  $\lambda$ , equal to the appropriate derivative of the equation of the ellipse,  $dy/dx$ , also evaluated at  $(X_t, Y_t)$ . The resulting expression, with equation (35), and the equation for the ellipse, equation (34), are three equations in three unknowns,  $k$ ,  $X_t$ , and  $Y_t$ . The solutions are:

$$k = \frac{(\beta + \lambda x_0 - y_0)^2}{\lambda^2 \sigma_x^2 - 2 r \sigma_x \sigma_y + \sigma_y^2} \quad (35)$$

$$X_t = \frac{\lambda \sigma_x^2 y_0 \beta - r \sigma_x \sigma_y (\lambda x_0 + y_0 - \beta) + y_0 \sigma_y^2}{\lambda^2 \sigma_x^2 - 2 r \lambda \sigma_x \sigma_y + \sigma_y^2} \quad (36)$$

The y-coordinate,  $Y_t$ , can be calculated with equation (34) after calculating  $X_t$ . In order to calculate the Bayes factor, one simply assigns the coordinates of the point of tangency to  $\mu_{x0}$  and  $\mu_{y0}$  in equation (32).

Table 3 contains Bayes factors for hypotheses relating the Volcano Mountain samples by crystal fractionation. The factor for evaluating a fractionation hypothesis relating Xe05 to Xe11, the hypothetical initial magma, is  $2.83 \times 10^{-6}$ , a value so small that it practically guarantees the fractionation hypothesis with olivine and pyroxene is wrong. The path for fractionation of olivine alone is tangent to a smaller ellipse than is the closest olivine plus pyroxene path. The Bayes factor for the olivine-alone path is 0.081 (Table 3). Consequently, the data provide no support for either fractionation path between Xe05 and Xe11. If the two samples are related, they are related by some mechanism other than fractionation of olivine and clinopyroxene. The olivine fractionation hypothesis relating the second sample, VM01, to Xe11 has a Bayes factor of 1.956; a value that supports the

fractionation hypothesis.

## Discussion - Concerns and Problems

### *Comparison of Bayesian and Standard Probabilities*

Standard tests of hypotheses usually follow a recipe similar to the following: (1) Compute a statistic, such as the mean or standard deviation from the data. (2) Compute the probability or *P-value* of getting that value or a larger value given a probability distribution dictated by some null hypothesis. If the computed statistic has a low probability of occurring, then reject the null hypothesis for the data in question (Press et al., 1992 p. 603). The probability calculated with this method is the probability of measuring data that produce statistics **greater than or equal to the observed value**. Consequently, rejection or acceptance of a null hypothesis depends more on values that were not measured than on data that were measured. Because there are always more unmeasured than measured values, probabilities assigned to the null hypothesis with standard calculations are generally larger than those calculated with Bayesian methods (Berger and Berry, 1988).

Although statisticians point out that *P-values* are not the probabilities of the null hypothesis (Berger and Berry, 1988; Berger and Sellke, 1987), most of us nonstatisticians have certainly interpreted them to be. Just as mathematical models are used to represent physical processes, statistical statements are used to represent hypotheses. Standard methods never explicitly consider alternative hypotheses; the null hypothesis is the only one stated with any precision. Bayesian methods force attention to alternative hypotheses and cause us, the investigators, to consider inverting the statistical representation back into theories of physical-chemical processes.

### *Probability Density Functions instead of Cumulative Frequency Functions*

Statistics texts, (e.g. Ash, 1993), are careful to point out that a number calculated with a probability density function [e.g. equation (10)], is not itself a probability. Rather, probabilities are calculated by integrating probability density functions to form cumulative frequency distributions. Probabilities calculated with Bayes factors or Bayesian methods are often expressed with probability density. Because Bayes factors can be expressed as the ratio of two probabilities [see equation (2)], they can be written:

$$B = \frac{f_0(x)\Delta x}{f_1(x)\Delta x} \quad (37)$$

if  $\Delta x$  is small. By multiplying the two probability density functions,  $f_0(x)$  and  $f_1(x)$ , by  $\Delta x$ , both the numerator and the denominator acquire the proper probability 'units'.

More correctly, Bayes factors can be written:

$$B = \frac{\int_x^{x+\Delta x} f_0(x) dx}{\int_x^{x+\Delta x} f_1(x) dx} \quad (38)$$

Equation (38) is the ratio of the cumulative distribution functions and both the numerator and denominator correctly represent probabilities.

Equation (37) transforms into a ratio of probability distribution functions just by cancelling the  $\Delta x$  terms (see Schmitt, 1969). The same result comes from equation (38) by taking the limit as  $\Delta x$  approaches zero. Consequently, it is legitimate to calculate Bayes factors and ratios of

probabilities from the density functions.

### *Limitations of Equations (22) and (32)*

Bayes factors should vary from zero to infinity if posterior probabilities as calculated with equation (4) are to vary from zero to one at a fixed value of  $p_0$ . Equations (22) and (32) can only result in values of  $B$  that are less than or equal to some finite maximum. Figure 5 demonstrates this fact for equation (22).  $B$  can reach its maximum value of  $\sqrt{2}$  only if the observed value of the variable  $\mu$  matches the hypothetical value  $\mu_0$ . Suppose a maximum value for  $B$  is obtained. Equation (4) calculates the posterior probability for a given  $p_0$ . The increase in  $p$  over  $p_0$  depends on  $p_0$ . The maximum increase possible is 0.086 if the  $p_0$  is 0.414. Although more complicated, equation (32) behaves in a similar manner; the principal differences are: 1) the maximum value possible for  $B$  is 2 and 2) the dependence of  $B$  on the correlation coefficient,  $r$ , and the variables  $u$  and  $v$ . Some of these dependencies are shown on Figure 6.

Equations (22) and (32) are both particular cases of more general equations. Simplifications result from setting the variances and covariances of the statistical representations of the hypotheses equal. By setting these statistical parameters equal, we are effectively limiting the possible range of hypotheses and restricting the range of values the Bayes factors can have.

Probabilities favoring hypotheses that lead to equations (22) and (32) cannot become large with a single datum. Several data points must be entered in succession into a Bayes analysis to get a high probability. Perhaps this is fitting for petrologic hypotheses. We can never be sure they are correct, being consistent with the data is the best we can ask. Consequently, high probabilities that they are true should be rare.

## Acknowledgments

This paper benefited from comments by M.Z. Stout, J.K. Russell, Ben Edwards, T.M. Gordon, and E.D. Ghent. They, of course, are not in any way responsible for its defects. George Snyder of the U.S.G.S. once asked how sure I was of a petrologic interpretation. I had no answer. This paper is a belated attempt to answer his question. A computer program that calculates the statistics for the three kinds of probability estimations described in this paper can be obtained by writing. Support was provided by NSERC Research Grant A7372.

## Appendix

*Justification of equation (11):*

Suppose we have two hypotheses with initial probabilities  $p_0$  and  $1 - p_0$ . The statement of the first hypothesis,  $H_0$ , is that the data should come from a probability density function  $f(x|\theta_0)$  characterized by a parameter,  $\theta$ , equal to a particular value,  $\theta_0$ . We are less certain about what  $\theta$  should be if the second hypothesis,  $H_1$ , is true but we know that  $\theta$  belongs to a set of values with its own probability density function,  $g(\theta)$ . For each possible value of theta, given  $H_1$ , we could observe a set of data,  $x$ . The total number of ways we could get sets of data is the "sum" of the products of the two density functions. As usual, because the density functions are continuous, we integrate rather than sum over all possible values of  $\theta$ :

$$\int f(x|\theta) g(\theta) d\theta \quad (\text{A-1})$$

The density distribution of  $x$  follows from a procedure analogous to finding the distributions for discrete probabilities:

$$p(x) = p_0 f(x|\theta_0) + (1 - p_0) \int f(x|\theta) g(\theta) d\theta$$

From Bayes theorem:

$$p(\theta_0|x) = p_0 f(x|\theta_0)/p(x) \quad (\text{A-2})$$

The expression for the Bayes factor [equation (11)] follows.

*Adequacy of the normal distribution for uncertainties in ratios*

The analytical uncertainty associated with each element in the ratio,  $\mu_x = \mu_X/\mu_Z$ , will be denoted by  $\sigma_x$  and  $\sigma_Z$ . The uncertainty in the ratio itself is, to first order (Nicholls, 1990):

$$\sigma_x^2 = \sigma_X^2/\mu_Z^2 + x^2 \sigma^2 \quad (\text{A-3})$$

where:

$$\sigma^2 = \sigma_Z^2/\mu_Z^2 \quad (\text{A-4})$$

If we assume the original analytical values themselves come from normal probability distribution functions with means  $\mu_X$  and  $\mu_Z$  and variances  $\sigma_X^2$  and  $\sigma_Z^2$ , then the probability density function for their ratio can be calculated from (Meyer, 1975):

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \int_{-\infty}^{+\infty} |Z| e^{-\frac{(xZ - \mu_x)^2}{2\sigma_x^2}} e^{-\frac{(Z - \mu_Z)^2}{2\sigma_Z^2}} dZ \quad (\text{A-5})$$

The result is:

$$f(x) = \frac{\sqrt{2} t e^{\frac{x\mu_x\mu_z}{q^2} - \frac{p^2}{2q^2}} \left[ \operatorname{erf} \left( \frac{\sqrt{2} t}{2\sigma_x\sigma_z q} + 1 \right) \right]}{2\sqrt{\pi} q^3} + \frac{\sigma_x\sigma_z}{\pi q^2} e^{-\frac{\mu_x^2}{2\sigma_x^2}} e^{-\frac{\mu_z^2}{2\sigma_z^2}} \quad (\text{A-6})$$

where:

$$\begin{aligned} p^2 &= \mu_x^2 + x^2 \mu_z^2 \\ q^2 &= \sigma_x^2 + x^2 \sigma_z^2 \\ t &= \sigma_z^2 \mu_x x + \sigma_x^2 \mu_z \end{aligned}$$

In comparison, we calculated the Bayes factors in Table 1 on the assumption that the ratios themselves, rather than the original measurements, were from normal probability densities:

$$g(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \quad (\text{A-6})$$

The two distributions, calculated from equations (A-5) and (A-6), for the Ca/K ratio listed in Table 1 are shown on Figure A-1. The maximum difference between the two distributions is small and is approximately one-tenth of the value of either distribution. Consequently, treating the ratios as normally distributed variables should not lead to erroneous Bayes factors. The intercept is a linear combination of ratios and it also should be normally distributed to the same approximation as are the ratios. The explanation for the small Bayes factor for the intercepts is either inappropriate analytical uncertainties or we hit the lottery.

## References Cited

- Ash, C., 1993, *The Probability Tutoring Book*: Piscataway, NJ, USA, IEEE Press, 470 p.
- Berger, J. O., and Berry, D. A., 1988, Statistical analysis and the illusion of objectivity: *American Scientist*, v. 76, p. 159-165.
- Berger, J. O., and Delampady, M., 1987, Testing precise hypotheses: *Statistical Science*, v. 2, p. 317-352.
- Berger, J. O., and Sellke, T., 1987, Testing a point null hypothesis: The irreconcilability of P values and evidence: *Journal of the American Statistical Association*, v. 82, p. 112-122.
- Ghiorso, M. S., Carmichael, I. S. E., Rivers, M. L., and Sack, R. O., 1983, The Gibbs free energy of mixing of natural silicate liquids; an expanded regular solution approximation for the calculation of magmatic intensive variables: *Contributions to Mineralogy and Petrology*, v. 84, p. 107-145.
- Ghiorso, M. S., and Sack, R. O., 1995, Chemical mass transfer in magmatic processes IV. A revised and internally consistent thermodynamic model for the interpolation and extrapolation of liquid-solid equilibria in magmatic systems at elevated temperatures and pressures: *Contributions to Mineralogy and Petrology*, v. 119, p. 197-212.
- Jefferys, W. H., and Berger, J. O., 1992, Ockham's razor and Bayesian analysis: *American Scientist*, v. 80, p. 64-72.
- Jeffreys, H., 1961, *Theory of Probability*: Oxford, England, Oxford University Press, 459 p.
- Meyer, S. L., 1975, *Data Analysis for Scientists and Engineers*: Inc, John Wiley and Sons.
- Nicholls, J., Russell, K., J., Stout, and Z., M., 1986, Testing magmatic hypotheses with thermodynamic modelling, *Short Course in Silicate Melts*: Toronto, M. Scarfe. *Mineralogical Assoc. of Canada*, p. 210-235.

- Nicholls, J., 1990, Principles of thermodynamic modeling of igneous processes, *Modern Methods of Igneous Petrology - Understanding Magmatic Processes: Reviews in Mineralogy*, Washington, D.C., Nicholls and J. K. Russell. Mineralogical Society of America, p. 1-23.
- Nicholls, J., and Stout, M. Z., 1988, Picritic melts in Kilauea - Evidence from the 1967-1968 Halemaumau and Hiiaka eruptions: *Journal of Petrology*, v. 29, p. 1031-1057.
- Oviatt, C. G., and Nash, W. P., 1989, Late Pleistocene basaltic ash and volcanic eruptions in the Bonneville basin, Utah: *Geological Society of America Bulletin*, v. 101, p. 292-303.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 1992, *Numerical Recipes in FORTRAN, The Art of Scientific Computing*: Cambridge, England, Cambridge University Press, 963 p.
- Schmitt, S. A. p., 1969, *Measuring Uncertainty, An Elementary Introduction to Bayesian Statistics*: Reading, MA, USA, Addison-Wesley Publishing Company, 400 p.
- Spooner, I. S., 1994, *Quaternary Environmental Change in the Stikine Plateau Region, Northwestern British Columbia, Canada [Ph. D. thesis]*: University of Calgary.
- Trupia, S., 1992, *Petrology of nephelinites and associated ultramafic nodules of Volcano Mountain, Yukon Territory [M Sc thesis]*: University of Calgary, 123 p.
- Trupia, S., and Nicholls, J., 1996, Petrology of Recent lava flows, Volcano Mountain, Yukon Territory, Canada: *Lithos*, v. 37, p. 61-78.
- Wright, T. L., 1971, *Chemistry of Kilauea and Mauna Loa in space and time: United States Geological Survey Professional Paper*, v. 735, p. 1-40.
- Wright, T. L., Swanson, D. A., and Duffield, W. A., 1975, Chemical compositions of Kilauea east-rift lava, 1968-1971: *Journal of Petrology*, v. 16, p. 110-133.