Statistical uncertainty associated with histograms in the Earth sciences

Pieter Vermeesch

Department of Geological and Environmental Sciences, Stanford University, Stanford, California, USA

Received 13 October 2004; revised 17 November 2004; accepted 24 November 2004; published 24 February 2005.

[1] Two types of quantitative information can be distinguished in the Earth sciences: categorical data (e.g., mineral type, fossil name) and continuous data (e.g., apparent age, strike, dip). Many branches of the Earth sciences study populations of such data by collecting a random sample and binning it into a histogram. Histograms of categorical data follow multinomial distributions. All possible outcomes of a multinomial distribution with M categories must plot on a $(M - 1)$ simplex Δ_{M-1} because they are subject to a constant sum constraint. Confidence regions for such multinomial distributions can be computed using Bayesian statistics. The conjugate prior/posterior to the multinomial distribution is the Dirichlet distribution. A 100(1-α)% confidence interval for the unknown multinomial population given an observed sample histogram is a polygon on Δ_{M-1} containing 100(1-α)% of its Dirichlet posterior. The projection of this polygon onto the sides of the simplex yields M confidence intervals for the M bin counts. These confidence intervals are "simultaneous" in the sense that they form a band completely containing the 100(1-α)% most likely multinomial populations. As opposed to categorical variables, adjacent bins of histograms containing continuous variables are not mutually independent. If this "smoothness" of the unknown population is not taken into account, the Bayesian confidence bands described above will be overly conservative. This problem can be solved by introducing an ad hoc prior of "smoothing weights" $w = e^{-sr}$, where r is the integrated squared second derivative of the histogram and s is a "smoothing parameter."

Citation: Vermeesch, P. (2005), Statistical uncertainty associated with histograms in the Earth sciences, *J. Geophys. Res.*, *110*, B02211, doi:10.1029/2004JB003479.

1. Introduction

[2] Consider a jar filled with infinitely many balls of M different colors. Suppose that we want to estimate the proportions of the colors in the jar by drawing a sample of N balls from it and counting the number of times each of the colors occurs in this sample: $\mathbf{n} = \{n_1, n_2, \ldots n_M | \sum_{j=1}^{M} n_j = N\}$. Then our best guess (the so-called "maximum likelihood estimate") for the M proportions is $\mathbf{p} = \{p_1 = n_1/N, p_2 = n_2/N, \ldots, p_M = n_M/N | \sum_{j=1}^{M} p_j = 1\}$. Now we ask ourselves the question: how confident are we about $\mathbf{p}$? In other words: Are there any other sets of proportions $\mathbf{p}' = \{p'_1, p'_2, \ldots, p'_M | \sum_{j=1}^{M} p'_j = 1\}$ that could have yielded the observations $\mathbf{n}$ with reasonable probability?

[3] This simple statistical problem frequently occurs in geological applications. Of course, geologists are not counting "balls" but things like sediment grains or faults. Neither are they interested in "color" (although sometimes they do) but in mineral type, age, or angle. In such studies, the information that is interpreted is not represented by the measurements themselves but by estimates of their probability distribution, which are most often represented by some sort of histogram. When reporting analytical data, it is considered good scientific practice to provide an estimate of the associated statistical uncertainties. This paper presents a method to extend this practice to the kind of point-counting studies described above. In section 2, we will introduce a number of examples of histograms in the Earth sciences, as a further motivation of the present study. We will distinguish between two types of histograms. A first type is used to represent categorical variables, such as color or mineral type. Here, we will also discuss the ternary diagram, which is a different way of visualizing histograms with only three bins, that is quite popular in sedimentary petrography. A second type of histogram which we will discuss contains continuous, or time series data. The prime examples of this kind of histograms are detrital thermochronological grain age histograms, which tally the number of times a range of apparent grain ages occur in a detrital sample. However, continuous histograms need not necessarily contain age data, and we will see an alternative example for which they do not. The fundamental difference between the aforementioned two types of histograms is that the bins of categorical histograms are mutually independent, while adjacent bins of continuous histograms are correlated to some degree. As a consequence, the method for constructing their respective confidence bands will be somewhat different.

0148-0227/05/2004JB003479$09.00

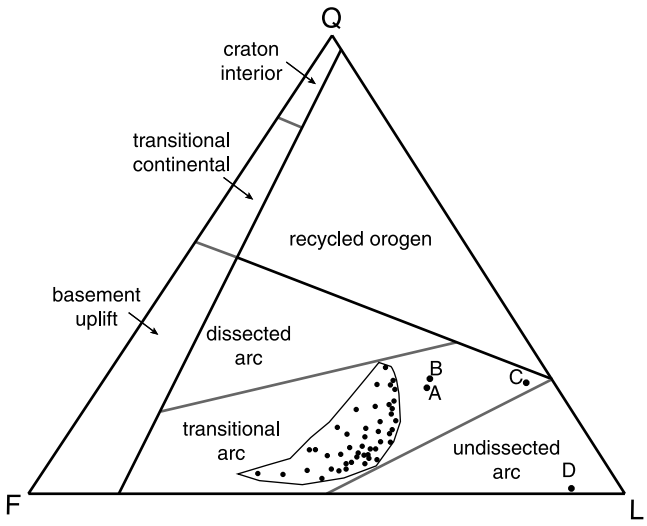


Figure 1. Petrographic QFL diagram with tectonic discrimination fields by *Dickinson et al.* [1983]. All samples except A represent 400 synthetic point counts. Sample A is based on only 200 counts.

[4] After section 2 has set the stage, we can begin developing the statistics of the actual method itself. The simultaneous confidence bands discussed in this paper will be derived according to the so-called ''Bayesian'' paradigm, as opposed to the more traditional ''frequentist'' paradigm. In section 3, these terms will be explained using a simple binomial example, which is a degenerate case of the problem this paper addresses. So by the end of section 3, we should be in a good shape to compute simultaneous confidence bands for multinomial proportions, which is the subject of section 4. Section 4.1 will explain why frequentist confidence intervals do not easily generalize to histograms with more than three bins. As an alternative, a Bayesian method to construct confidence bands for categorical histograms will be developed in section 4.2. Finally, section 4.3 gives an ad hoc way to modify the method of section 4.2 so that it takes into account the autocorrelation of continuous histograms. Section 5 revisits the examples of section 2 and answers the questions that were raised in it. Section 6 wraps up the paper with some summarizing conclusions.

2. Setting the Stage: Examples of Histograms in the Earth Sciences

2.1. Categorical Histograms

[5] The framework composition of sandstones contains useful information about their provenance, transport history and postdepositional evolution, and is used to reconstruct the plate tectonic setting of sedimentary basins [e.g., *Dickinson et al.*, 1983; *Dickinson*, 1985]. Framework compositions are measured by petrographic point counting of thin sections. The results are often plotted on ternary diagrams, the most popular of which is the QFL diagram, which depicts quartz, feldspar, and lithic fragments (Figure 1). As discussed in section 1, one of the questions this paper will answer is how to estimate the statistical uncertainties for such point-counting measurements. *Van der Plas and Tobi* [1965] discuss the construction of confidence intervals for individual point-counting proportions, for example the percentage of quartz in a thin section. However, we are rarely interested in just a single proportion. This paper develops a Bayesian method to compute simultaneous confidence bands for categorical histograms. This method will allow an estimation of the likelihood that a specific sample falls into one particular field of tectonic affinity on the QFL plot (Figure 1). To avoid confusion, we should remark that while this paper will discuss the statistical uncertainties of individual point-counting measurements (one sample), it will not talk about the uncertainties on populations of several measurements. Whereas the former follows a multinomial distribution, the latter can take many forms, such as the logistic normal distribution. Many interesting issues are associated with ternary populations, but the reader is referred to *Weltje* [2002] for a discussion of them. Figure 1 shows a petrographic QFL diagram with tectonic discrimination fields by *Dickinson et al.* [1983]. The ''cloud'' of points and the associated hand-drawn contour mark a detrital population. For the discussion of how to compute this contour in a statistically more rigorous way, we again refer to *Weltje* [2002]. The present paper will address the following questions: (1) How different are samples A and B? (2) Is it possible that samples A and B belong to the contoured population? (3) How certain are we that sample C falls into the ''transitional arc'' field? Could it be that it actually belongs to one of the neighboring fields? (4) How does the number of grains affect the precision of our point-counting results? Since the ternary diagram plots ratios, we lose information on the actual number of grains counted. For example, sample A represents 200 counts, while sample B represents 400 and there is no way to tell this from Figure 1.

[6] The ternary diagram is very popular in sedimentary petrography, but when more than three components need to be plotted, we must use another device: the histogram. This is the case in heavy mineral analysis [e.g., *Faupl et al.*, 2002], and in clast counting, which is a scaled-up version of petrographic point counting [e.g., *Yue et al.*, 2001]. Figure 2 shows two heavy mineral analyses by *Faupl et al.* [2002]. For each sample, 200 grains were counted. The basic questions that arise when doing this sort of analysis are

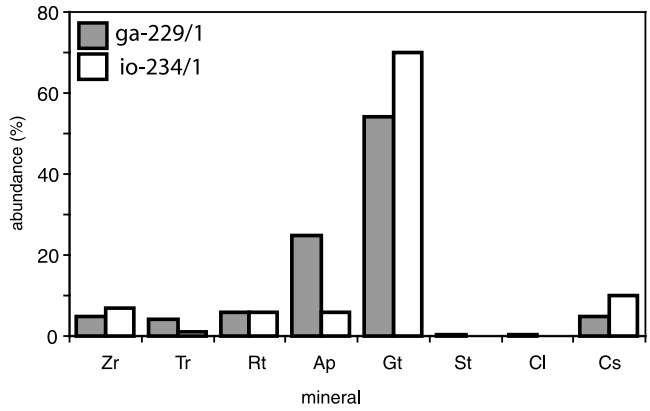


Figure 2. Heavy mineral analysis of two samples from the Peloponnese (Greece) by *Faupl et al.* [2002]. Zr, zircon; Tr, tourmaline; Rt, rutile; Ap, apatite; Gt, garnet; St, staurolite; Cl, chloritoid; Cs, chrome spinel.

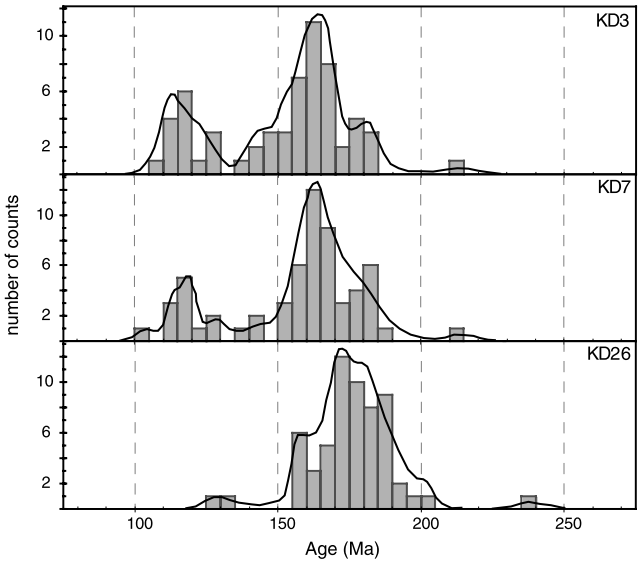


Figure 3. Three U-Pb grain age histograms and corresponding kernel density estimates for samples of detrital zircon from the Cretaceous Methow basin [*DeGraaff-Surpless et al.*, 2003].

the same as for the ternary example: (1) What is the precision of the estimated mineral fractions? (2) How would the precision be affected by increasing N, the total number of grains counted? (3) Are samples ga-229/1 and io-234/1 compatible with each other? Regarding the last question, it is useful to remark that when a very large number of grains are counted, it is almost certain that a statistically significant difference between two samples of the same rock will be found. No two samples collected on the field have an exactly identical composition. As the number of counts increases, our power to resolve even the smallest differences will increase. It is when this point is reached that the petrographic composition of the sample has been properly characterized and we can begin to study populations of samples. As a comforting note, we can already tell here that the guidelines of *Van der Plas and Tobi* [1965] fulfill this requirement most of the time.

2.2. Continuous Histograms (Time Series)

[7] Alternatively, histograms can also be used for continuous data. Detrital thermochronology tries to find the provenance area of sedimentary rocks and unravel its geologic history, by dating individual mineral grains in the sample [e.g., *Avigad et al.*, 2003; *DeGraaff-Surpless et al.*, 2003]. Figure 3 shows three detrital zircon U-Pb grain age distributions from the Methow Basin, in the southern Canadian cordillera [*DeGraaff-Surpless et al.*, 2003]. For each sample, Figure 3 not only shows the grain age histogram, but also the continuous "kernel density estimate." Unlike categorical point-counting data, the grain ages that are used in detrital thermochronology can have significant analytical uncertainties. This is a second source of error (the first one being counting statistics) that is not taken into account by the histogram. The kernel density estimator is an alternative estimator of probability density that does take into account measurement uncertainties. However, it is not easy to estimate the effect of counting statistics on kernel density estimates. In this paper, we will ignore measurement uncertainties, and just focus on the effect of counting statistics. We will later see that in order to get a better idea of the importance of both factors, it is good practice to use histograms in conjunction with kernel density estimates. The reader is referred to *Silverman* [1986] and *Sircombe and Hazelton* [2004] for a discussion of the kernel density estimator and some issues that are associated with it. This paper will answer the following questions concerning detrital grain age histograms: (1) What is the uncertainty on the bin counts? (2) How certain are we that empty bins actually correspond to missing age fractions? (3) Are grain age histograms such as the three shown in Figure 3 compatible with or significantly different from each other?

[8] It is easy to see that detrital grain age histograms represent time series. However, continuous histograms are not restricted to the time dimension. Figure 4 shows a histogram of dip estimates for 33 reverse faults reported by *Collettini and Sibson* [2001]. Although the units of this histogram are not time, but angle (in degrees), it still represents a continuous function, or "time series". One of the observations made by Collettini and Sibson about this histogram is that it is bimodal, with one peak at 30° and a second at 50°. The simultaneous Bayesian confidence intervals described in this paper will tell us if this bimodality is statistically significant on for example a 95% confidence level. Whereas categorical data follow multinomial distributions, where the bins are mutually independent apart from the fact that they must sum to a fixed number (the sample size), time series are autocorrelated to some degree, and this must be taken into account when computing confidence intervals. This paper will assess the importance of this problem and propose a Bayesian solution to it in the form of an ad hoc smoothing prior.

3. Definition of a Confidence Interval

[9] In this section, we will introduce some fundamental statistical principles and nomenclature which will be needed in section 4. Surprisingly enough, there is no general agreement in the statistics community on the definition of

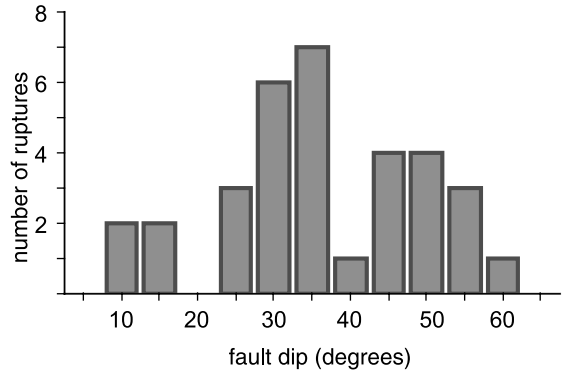


Figure 4. Histogram of 33 reverse fault dip estimates. Although the measurements are in degrees, the histogram can still be considered a "time series" because it is expected to fit a more or less smooth, autocorrelated function.

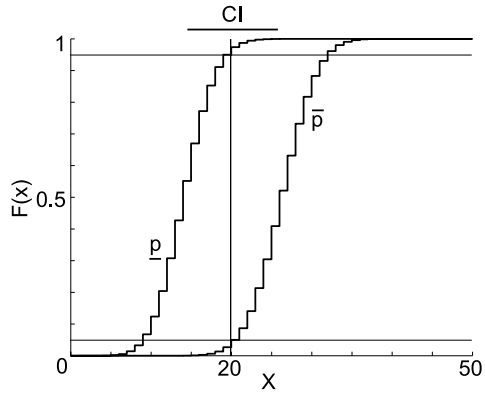


Figure 5. The 90% frequentist confidence bounds on p for $n = 20$, $N = 50$. The step functions represent the cumulative binomial distribution with parameters (N, $\underline{p}$) and (N, $\bar{p}$), respectively; $\underline{p}$ is the lowest value for p for which 20 out of 50 counts would occur more than 5% of the time. Likewise, $\bar{p}$ is the highest value for p that would yield the observed ratio of 20/50 with more than 5% probability.

a confidence interval. There are two points of view: the frequentist and the Bayesian point of view. To explain the difference between these two paradigms, we will consider a degenerate case of the problem at hand. Revisiting the metaphor from section 1, we now consider a jar with balls of only two colors, say black and white. Drawing N balls from this jar as before, we count the number n of black balls. For this binomial experiment, the maximum likelihood estimate for the proportion of black balls in the jar is p = n/N. How do we construct a 100(1-α)% confidence interval for this estimate? An approximate solution to this problem is given by *Van der Plas and Tobi* [1965], but both the frequentist and the Bayesian methods which will be discussed next are exact.

3.1. Frequentist Approach

[10] According to the "frequentist", a confidence interval for a parameter θ "consists precisely of all those values of θ_0 for which the null hypothesis H_0: $\theta = \theta_0$ is accepted" [*Rice*, 1995]. For example, we saw earlier that histograms represent the outcome of a multinomial experiment. The probability distribution of each of the bin counts of a histogram is the marginal of a multinomial distribution, which is the binomial distribution. Consider a bin containing n out of N measurements. The maximum likelihood estimate for the binomial parameter p then is $\hat{p}_{\mathrm{MLE}} = n/N$. Now consider the null hypothesis H_0: $p = p_o$ versus the alternative H_a: $p \neq p_o$. H_0 is accepted on a 100(1-α)% confidence level if:

$$\sum_{i=n}^{N} \binom{N}{i} p_o^i p_o^{N-i} < \frac{\alpha}{2} < \sum_{i=0}^{n} \binom{N}{i} p_o^i p_o^{N-i} \tag{1}$$

Now, according to the definition, a two-sided confidence interval contains all those values for p_o which pass the test given by equation (1). The solution can be found by numerical iteration and/or interpolation [*Clopper and Pearson*, 1934]. An example for $N = 50$, $n = 20$ and $\alpha = 0.1$ is given in Figure 5.

[11] It can be shown [e.g., *Blyth*, 1986] that equation (1) is mathematically equivalent to

$$\underline{p} = B\left(1 - \frac{\alpha}{2}, n+1, N-n\right) < p < B\left(\frac{\alpha}{2}, n, N-n+1\right) = \bar{p} \tag{2}$$

Where $B(\alpha, a, b)$ is the 100α percentile of the β distribution with parameters a and b:

$$\beta(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1-p)^{b-1} \tag{3}$$

where $\Gamma(x)$ is the gamma function, which can be considered the continuous version of the factorial operator. For example, if x is an integer, then $\Gamma(x + 1) = x!$. Likewise, the β distribution can be thought of as being a continuous version of the binomial distribution. Notice that for $n = 0$ and $n = N$, equation (2) breaks down. Instead, the following expressions should be used:

$$\underline{p} = 0 < p < 1 - \alpha^{1/N} = \bar{p} \quad \text{if n} = 0, \text{ or} \tag{4}$$

$$\underline{p} = (1-\alpha)^{1/N} < p < 1 = \bar{p} \quad \text{if n} = \text{N} \tag{5}$$

3.2. Bayesian Approach

[12] For a "Bayesian", a 100(1-α)% confidence (or "credibility") interval for a parameter θ given some data $\mathbf{x}$ is an interval for θ that covers 100(1-α)% of its *posterior distribution* $P(\theta|\mathbf{x})$, where the latter is given by

$$P(\theta|\mathbf{x}) \propto P(\mathbf{x}|\theta)P(\theta) \tag{6}$$

with $P(\theta)$ a "prior distribution" on θ and $P(\mathbf{x}|\theta)$ the "likelihood function" of the data given the parameter. The subjectivity of the Bayesian approach lies in the choice of the prior distribution. A uniform distribution ("flat prior") is often taken if no prior information exists as to what the value of θ should be. However, whether or not this is a good "noninformative" prior has been challenged. The uniform distribution does not yield posterior distributions that are invariant under reparameterization [*Jeffreys*, 1946]. We will soon see an example of an alternative prior distribution that does have this invariance.

[13] We now return to the problem of independent credibility intervals for multinomial proportions. Again, we consider a bin with n counts out of N and want to construct a 100(1-α)% credibility interval for p = n/N. The likelihood function is binomial: $P(n|p) = \binom{N}{n} p^n p^{N-n}$. If we take a flat prior for P(p), then the posterior is a $\beta(n+1, N-n+1)$ distribution [*Bayes*, 1763]:

$$P(q < p < r|n) = \frac{\Gamma(N+2)}{\Gamma(n+1)\Gamma(N-n+1)} \int_q^r p^n (1-p)^{N-n} dp \tag{7}$$

Therefore

$$\underline{p} = B\left(1 - \frac{\alpha}{2}, n+1, N-n+1\right) < p < B\left(\frac{\alpha}{2}, n+1, N-n+1\right) = \bar{p} \tag{8}$$

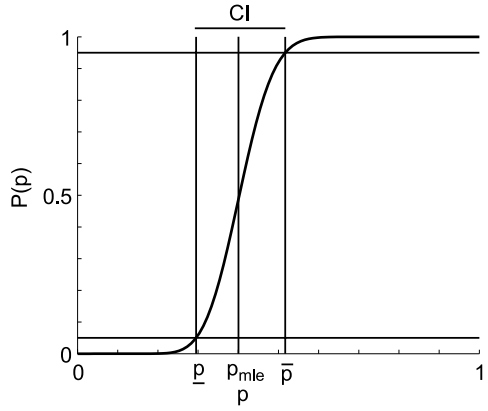


Figure 6. The 90% Bayesian credibility bounds on p for $n = 20$, $N = 50$. The curve represents the cumulative β distribution function with parameters $n + 1$ and $N - n + 1$, using a flat prior. The credibility interval $[\underline{p} < p < \bar{p}]$ is a (symmetric) interval for p that covers 90% of the area under this posterior distribution.

Notice the similarities between equations (2) and (8). However, as opposed to the frequentist equation (2), the Bayesian equation (7) does not require a special case for $n = 0$ and $n = N$. The β distribution is an example of a "conjugate prior." This means that if we take a β-distributed prior, and a binomial sampling distribution, then the posterior will also have a β distribution. The uniform distribution is a special case of the β distribution for $a = b = 1$ (i.e., $\beta(1, 1)$). $\beta(\frac{1}{2}, \frac{1}{2})$ is a noninformative prior ("Jeffreys' prior") for the binomial distribution that is invariant under reparameterization [e.g., *Gill*, 2002, p. 124]. The posterior distribution then becomes $\beta(n + \frac{1}{2}, N - n + \frac{1}{2})$. Taking the same example as for section 3.1 (i.e., $n = 20$, $N = 50$, $\alpha = 0.1$), Figure 6 shows a two-sided Bayesian credibility interval for p.

4. Simultaneous Confidence Intervals for Multinomial Proportions

[14] As shown in section 3, it is relatively easy to construct independent confidence intervals for each of the M bin counts n_m ($1 \leq m \leq M$) that make up a histogram, both under the frequentist and the Bayesian paradigm. However, we need to be more ambitious than that. In order to be able to compare two samples and test if they are significantly different, we would like to construct simultaneous confidence intervals for all of the M histogram bins. Like we did for the binomial case in section 3, we will again discuss first the frequentist and then the Bayesian solution to this problem. It will soon become clear why the Bayesian method is more appropriate for our purposes.

4.1. Frequentist Confidence Regions

[15] As discussed before, histograms are representations of multinomial distributions. Suppose we have N numbers ("balls"), distributed over M bins ("colors"), corresponding to M multinomial proportions. The bin counts $(n_1, .., n_M)$ must fulfill the condition $\sum_{m=1}^{M} n_m = N$. Therefore all possible multinomial distributions must fall on an "M simplex" Δ_{M-1}. An example of a three simplex (which just is another word for "ternary diagram") is shown on Figure 7. Consider a histogram with M bins, representing a sample of N numbers: $X_N = \{x_1, .., x_N\}$. This histogram corresponds to one point on Δ_{M-1}, the "maximum likelihood estimate" (MLE) of the bin counts. Under the frequentist paradigm, outlined in section 3.1, a 100(1-α)% confidence region on Δ_{M-1} consists of all those probability vectors $\mathbf{p} = (p_1, .., p_M | \sum_{m=1}^{M} p_m = 1)$ which are capable of yielding observations as extreme as $\mathbf{n} = (n_1, .., n_M | \sum_{m=1}^{M} n_m = N)$ with at least 100(1-α)% probability.

[16] In order to find this region, a grid of possible $\mathbf{p}^{kl} = (p_1^{kl}, .., p_M^{kl} | \sum_{m=1}^{M} p_m^{kl} = 1)$ is evaluated. For each of these "test populations" (e.g., the black dot on Figure 8) a large number of synthetic "samples" (the white dots on Figure 8) of N numbers were generated, following an algorithm given in Appendix A. Next, we construct the 100α% "convex hull" of these synthetic samples. This is a polygon containing 100α% (the so-called "hull percentile") of them. We test to see if $\mathbf{p}^{\mathrm{MLE}}$ (the black square on Figure 8) falls within the convex hull of $\mathbf{p}^{kl}$. If this is not the case, then $\mathbf{p}^{kl}$ falls outside the 100α% confidence region of $\mathbf{p}^{\mathrm{MLE}}$. This procedure is repeated for the entire grid ($k = 1..K$, $l = 1..L$). On Figure 9, the contour lines contain all those grid points for which the MLEs fall within their 95 percentile hull.

[17] Figures 8 and 9 just serve as an illustration of the frequentist paradigm on Δ_2. A more efficient way to

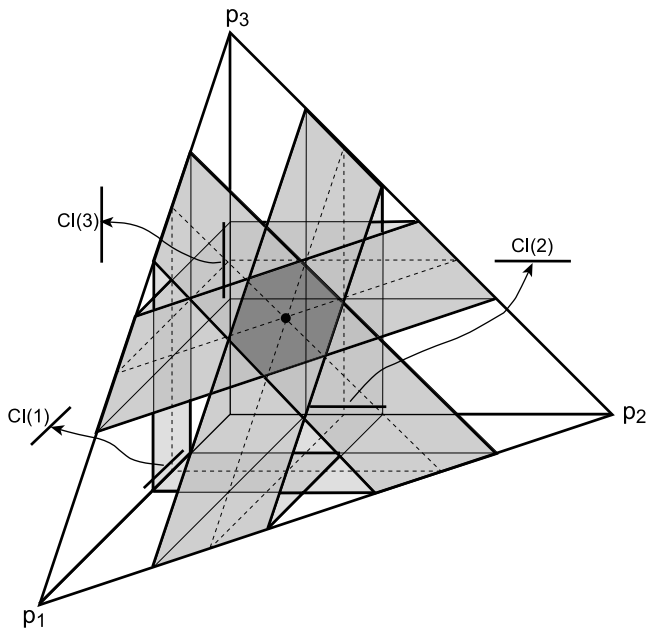


Figure 7. All possible outcomes of a trinomial experiment, for example, a three-bin histogram of N measurements x_i plot on a three simplex. The maximum likelihood estimate for the multinomial proportions is given by p_m^{MLE} = (number of x_i in mth bin)/N(m = 1, 2, or 3). The maximum likelihood estimate (MLE) is represented by a solid circle. The posterior distribution of the unknown parameters p_1, p_2, and p_3 is given by a Dirichlet distribution. To find simultaneous 100(1-α)% confidence bounds for these parameters, we need to find a polygon on the simplex that contains 100(1-α)% of the posterior distribution.

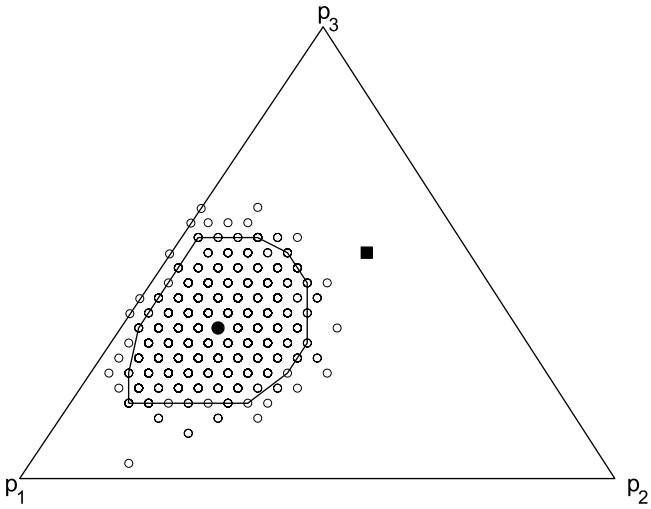


Figure 8. Test if the trinomial distribution marked by the black dot ($p_1 = 1/2$, $p_2 = 1/6$, $p_3 = 1/3$) belongs to the 95% confidence region of the trinomial experiment marked by the black square ($n_1 = 5$, $n_2 = 10$, $n_3 = 15$). A large number (1000) of trinomial samples of $N = 30$ numbers were generated from this distribution. They are represented by the open circles. The black contour line represents the 95% convex hull. Since the solid square does not fall within this hull, the solid circle falls outside the 95% confidence region of the trinomial experiment.

compute approximate frequentist confidence regions on the ternary diagram is described by *Weltje* [2002, p. 246]. Projecting the frequentist confidence region onto the axes of the simplex would not represent that region, but the smallest polygon circumscribing it. Therefore it is not possible to accurately "translate" a frequentist contour plot to error bars on a histogram, which makes it impossible to easily visualize the frequentist uncertainties of histograms with more than three bins. The Bayesian credibility regions discussed next solve this problem.

4.2. Bayesian Credibility Regions

[18] It is relatively easy to generalize the methodology outlined in section 3.2 from a binomial to a multinomial situation. Recall that the conjugate prior to a binomial distribution is the β distribution. The conjugate prior to a multinomial distribution is the "Dirichlet distribution":

$$D_{\mathbf{a}}(p_1, .., p_M) = \frac{\Gamma\left(\sum_{i=1}^{i=M} a_i\right)}{\prod_{i=1}^{i=M} \Gamma(a_i)} \prod_{i=1}^{i=M} p_i^{a_i - 1} \tag{9}$$

[19] The multinomial uniform distribution is a special case of the Dirichlet distribution with all $a_i = 1$. If $\mathbf{n}$ is a vector of M bin counts, then the posterior distribution under such a flat prior is $D_{\mathbf{n}+1}(p_1, .., p_M)$. The choice of a prior that is truly noninformative and invariant under reparameterization is more controversial for the Dirichlet than it was for the β distribution. Jeffreys suggested taking $a_i = 1/2$, while Perks recommended using $a_i = 1/M$ ($\forall i = 1..M$) [*Good*, 1965]. Similar to the binomial case (section 3.2), simultaneous Bayesian credibility bands for the multinomial distribution are intervals that cover 100(1-α)% of the area under the posterior distribution. A few examples of Dirichlet posteriors are shown on Figure 10. As opposed to the β distribution, there are no tables of the percentiles of the Dirichlet distribution. In order to integrate this multidimensional function ourselves, we have to numerically sample from it, as described by *Devroye* [1986] and in Appendix B.

[20] Thus a collection of B "sample histograms" can be constructed, representing B samples from the posterior Dirichlet distribution (Figure 11). All these histograms correspond to points on Δ_{M-1}. Asymptotically, independent 100(α/2) and 100(1 − α/2) percentiles for the replicates of each of the histogram bin counts will converge to the independent credibility intervals of equation (8). However, it is also possible to obtain simultaneous credibility bands. The Bayesian way of doing this is to find M credibility intervals that define a polygon on Δ_{M-1} containing 100(1-α)% of the posterior distribution (Figures 7 and 11). The algorithm for finding this polygon is given in Appendix B. Figures 12 and 13 show the effect of different priors on the posterior distribution and its corresponding credibility polygon.

[21] This Bayesian method yields nonzero credibility intervals, even for empty bins. It works for histograms but not for kernel density estimates, which are continuous functions that cannot be easily represented on a simplex. As histograms traditionally do not take into account measurement uncertainties, the Bayesian credibility bands only reflect the uncertainties induced by the counting statistics, and not those caused by analytical imprecision. A final remark to be made is that, strictly speaking, the way we have defined simultaneous Bayesian credibility regions is only exact for "categorical histograms", such as those

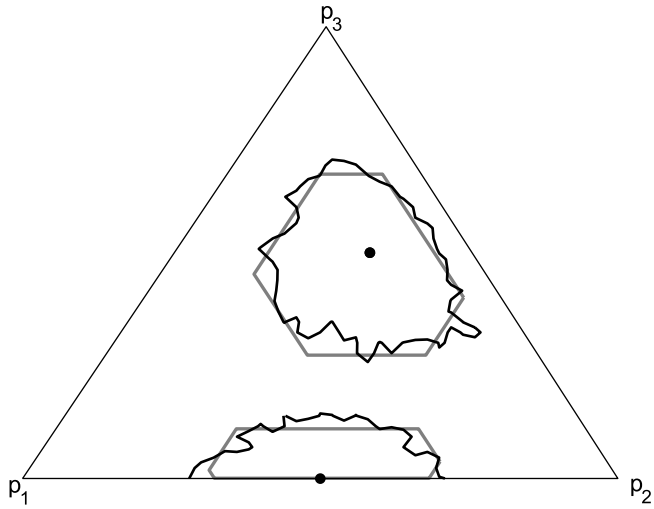


Figure 9. Maximum likelihood estimates (MLE) (solid circles) for two trinomial experiments ($\{n_1 = 5, n_2 = 10, n_3 = 15\}$ and $\{n_1 = 15, n_2 = 15, n_3 = 0\}$). The black contours represent the frequentist confidence regions, obtained by repeating the experiment shown in Figure 7 on a 1250 point grid. For each of the grid points, $B = 200$ trinomial samples were generated. The gray lines outline the Bayesian credibility regions (using a flat prior). The agreement between the two methods is surprisingly good.

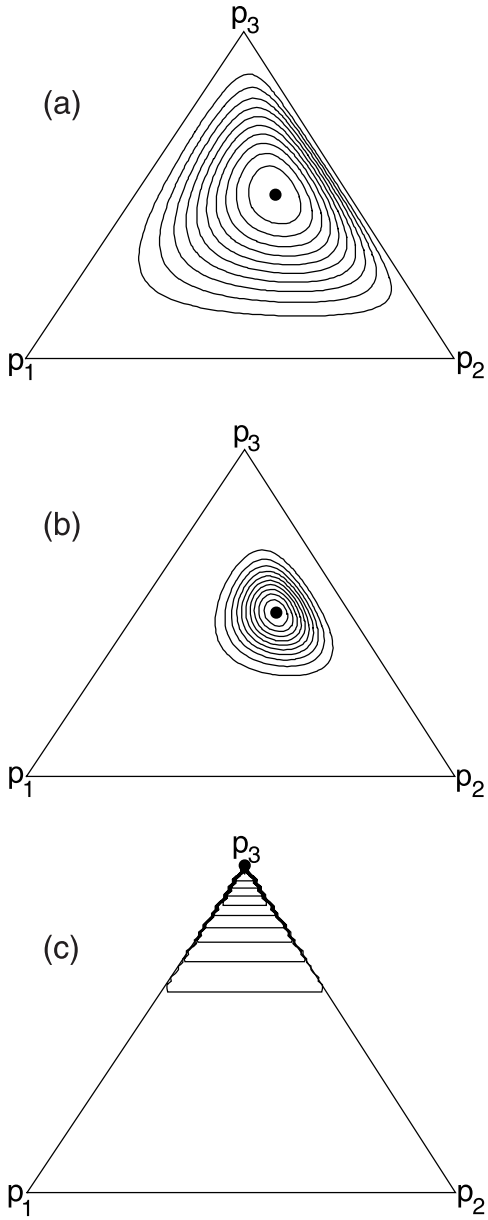


Figure 10. Analytical contours for different posterior Dirichlet distributions. The parameters are (a) (1, 2, 3), (b) (5, 10, 15), and (c) (0, 0, 5). Comparison of Figures 10a and 10b shows how the posterior Dirichlet distribution is more tightly constrained when more data are used. Figure 10c shows how, even when two bins are empty, meaningful confidence intervals can be computed.

obtained by point-counting mineral assemblages. However, if the histogram represents a time series, which is the case in detrital thermochronology, it will have some "smoothness" to it. This effect will not be captured by the Bayesian credibility regions discussed before. The categorical Bayesian credibility bands can be overly conservative if applied to such "autocorrelated" data. Section 4.3 discusses this issue.

4.3. Bayesian Credibility Bands for Smooth Histograms

[22] Strictly speaking, Bayesian credibility bands are only applicable to nonsmooth or categorical data (section 4.2). In this section, we will discuss the importance of this problem and a way to solve it. We can express the "roughness" r of a time series $g(t)$ as a function f of its second derivative:

$$r(g(t)) = f\left(\int \left(\frac{d^2 g(t)}{dt^2}\right)^2 dt\right) \tag{10}$$

For example, for the discrete case of a histogram with three bins $\mathbf{n} = (n_1, n_2, n_3)$, we could write

$$r(\mathbf{n}) = (n_1 - 2n_2 + n_3)^2 \tag{11}$$

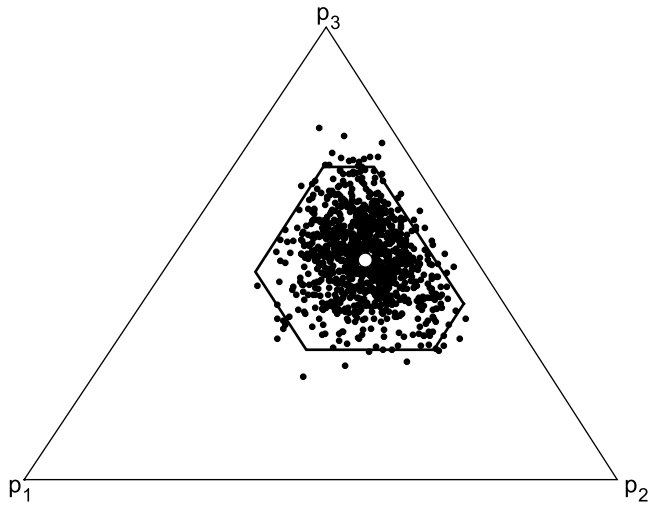


Figure 11. Numerical integration of the area under posterior distributions such as those in Figure 10. The open circle shows the maximum likelihood trinomial distribution for a particular sample ($n_1 = 5$, $n_2 = 10$, $n_3 = 15$). The solid circles represent 1000 random samples from the Dirichlet posterior distribution $D_{6,11,16}$ on the three simplex. The polygon contains 95% of these points. The projection of this polygon onto the three parameter axes yields the simultaneous Bayesian credibility intervals.

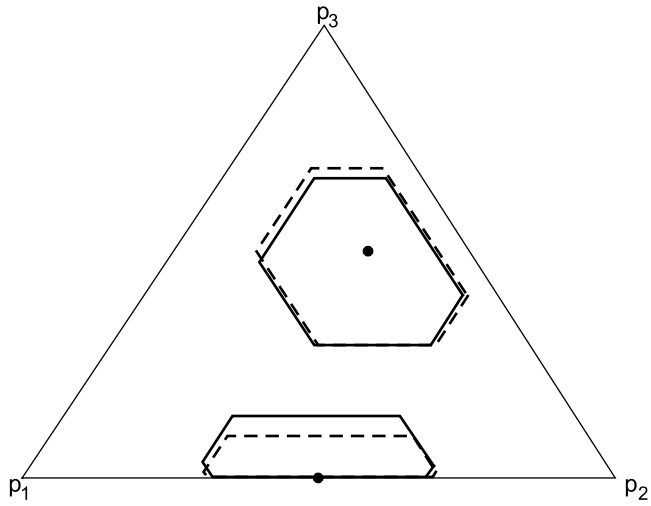


Figure 12. Three simplex with sample bin counts (5, 10, 15) and (15, 15, 0) (solid circles). The solid polygons represent their respective simultaneous 95% Bayesian credibility polygons with uniform prior $D_{1,1,1}$, while for the dashed polygons, Perks' prior $D_{\frac{1}{3},\frac{1}{3},\frac{1}{3}}$ was used.

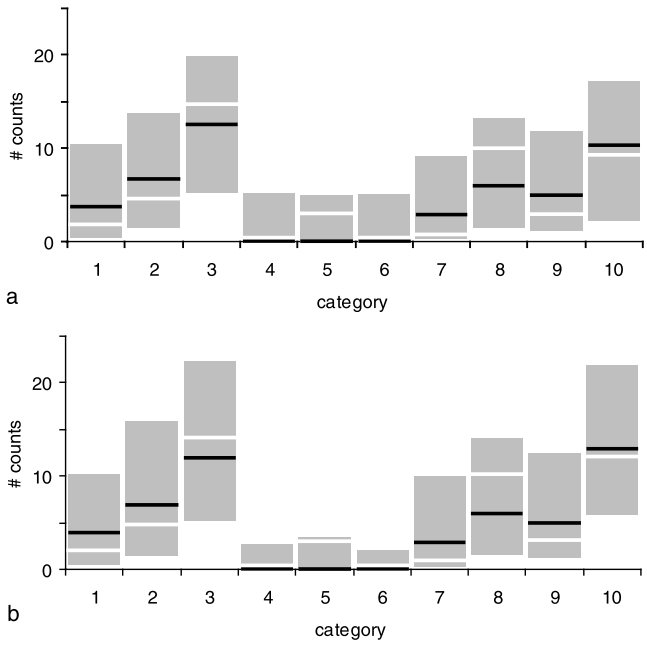


Figure 13. White histograms representing a synthetic population of categorical variables and black histogram representing a sample of 50 items from it. Note that the sample completely missed the fifth population bin. The gray band covers 95% of the posterior distribution for (a) a flat (uniform) prior and (b) Perks' prior. Both credibility bands correctly contain the original population.

We now define the "smoothness weight" w as

$$w = e^{-sr} \tag{12}$$

where s is the "smoothing parameter". Figure 14 shows the trinomial smoothing weights for different values of s (for fixed $N = n_1 + n_2 + n_3$). The distribution of the weights can be used to "filter" the posterior distribution, thereby in effect serving as a "prior distribution" (a "smoothing prior"). The algorithmic details for this procedure are given in Appendix C.

[23] Figure 15 shows the results of this kind of posterior filtering for a trinomial distribution on Δ_2. A logical extension of this method from three to M bins would be to replace equation (11) by

$$r(\mathbf{n}) = \sum_{m=2}^{M-1} (n_{m-1} - 2n_m + n_{m+1})^2 \tag{13}$$

By generating samples from the posterior distribution and accepting or rejecting them based on the smoothing weights given by equations (12) and (13), B samples from the smoothed posterior could be obtained. However, the amount of computation time that would be required for this process increases exponentially with M. The "sliding window" approach explained next does a similar job in linear time. The roughness penalty embodied by equation (12) depends on the second derivative only. This means that we only consider the influence of immediately adjacent bins on each other. For example, the mth bin is directly correlated with the $(m - 1)$th bin and the $(m + 1)$th bin, but not with the $(m - 2)$th bin and the $(m + 2)$th bin. This warrants the use of a three bin wide "sliding window," that recursively smooths the posterior distribution from the left to the right (or vice versa) one bin at a time. The details of this method are given in Appendix C.

[24] Figures 16 and 17 illustrate the results of the sliding window procedure on a synthetic data set. Figures 16 and 17 demonstrate how using a smoothing prior filters out the roughest "spikes" from the posterior sample set. It is such often small minority of outliers that makes the unsmoothed, categorical confidence bands of section 4.3 too wide, meaning too conservative, for continuous histograms. An example of the sliding window approach on real data is

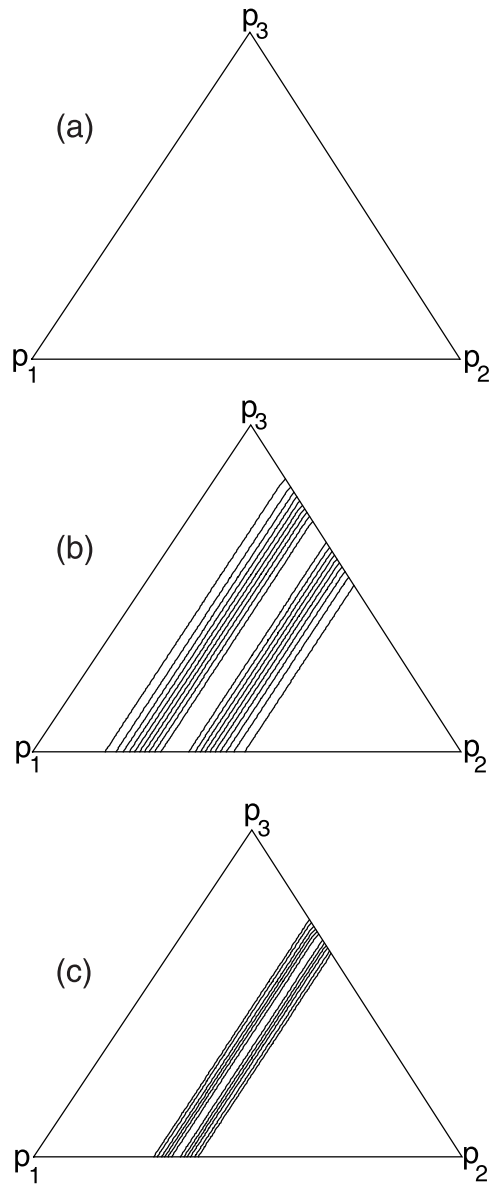


Figure 14. Contoured smoothing priors for different smoothing parameters, with $N = n_1 + n_2 + n_3 = 10$ for (a) $s = 0$ (no smoothing), (b) $s = 0.1$ (weak smoothing), and (c) $s = 1$ (strong smoothing). The strongest weights are located along the $p_2 = n_2/N = 1/3$ line, which connects all possible histograms with zero second derivative.

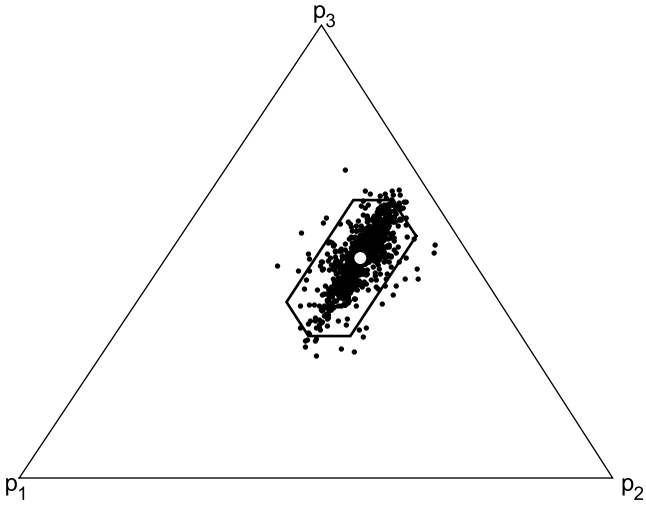


Figure 15. Smoothed version of Figure 11, using $s = 0.1$.

deferred to section 5.2, where we will also discuss which values for the smoothing parameter s to choose, and why the rather ad hoc nature of the smoothing method discussed in this section is probably not a big issue after all.

5. Case Studies

5.1. Categorical Histograms

[25] We first return to the example previously discussed in section 2.1 and illustrated by Figure 1. Figure 18 shows the simultaneous 95% Bayesian credibility regions for samples A–D on the ternary QFL diagram in their now familiar polygonal form. The difference between the credibility regions of samples A and B clearly stands out. Recall that 400 points were counted for sample B, as opposed to only 200 for sample A. As a result, the uncertainty polygon of A is substantially larger than that of B. It is quite possible that A and B were sampled from the same distribution. Whereas it very unlikely that sample B could have been derived from the population outlined by the contour, this conclusion cannot be made for sample A. A similar situation exists for sample C. This sample plots into the "transitional arc" field of the QFL diagram, but its 95% uncertainty region partly falls inside the "undissected arc" and "recycled orogen" fields. Although we have not done so, it would even be possible to compute the respective probabilities for the three fields, by counting the number of numerical Bayesian replicates that fall into them. Finally, sample D contains only one percent (4/400) of quartz. Its uncertainty hexagon is highly asymmetric but falls entirely inside the simplex, as it should.

[26] We now proceed to the multinomial example of heavy mineral analysis by *Faupl et al.* [2002]. Figure 19 shows the 95% credibility intervals for the eight heavy mineral proportions, using Jeffreys' prior. When 200 grains are counted, the percentage error is between 2 (staurolite in ga-229/1) and 20% (garnet in io-234/1) (Figure 19a). There is less than 5% probability that the heavy mineral distribution of sample io-234/1 is compatible with that of sample ga-229/1 because the observed apatite and garnet fractions of io-234/1 fall outside the simultaneous 95% credibility band of ga-229/1. Likewise, it is less than 5% likely that sample ga-229/1 is compatible with io-234/1 because the apatite and garnet fractions of the former fall outside the confidence bands of the latter. The statement that ga-229/1 and io-234/1 are mutually compatible is true with less than 2.5%, and not 5% probability, because it involves two simultaneous tests. This is a consequence of the so-called Bonferroni inequality [*Rice*, 1995]. The Bonferroni rule is on the conservative side, especially considering the fact that the two tests are not entirely independent from each other. Figure 19b shows that if the percentages reported by Faupl et al. had been the result of counting 1000 instead of 200 grains, the percentage errors would have been between 0.5 and 9%.

5.2. Continuous Histograms

[27] We first consider a data set of 157 concordant U-Pb SHRIMP ages on detrital zircon from the Cambrian Nubian Sandstone of southern Israel [*Avigad et al.*, 2003]. The vast majority of these grains are of Pan-African age (900–540 Ma), and likely derived from the Arabian-Nubian shield but there are some older grains as well, which could have come from as far as central Africa [*Avigad et al.*, 2003]. Figure 20 shows the kernel density plot of this data set and its grain age histogram with 95% credibility band. Figure 20 contains an optimal amount of information: the kernel density estimate shows the sample taking into account measurement uncertainties, while the histogram represents the estimated population and the uncertainties caused by counting statistics. The credibility band also allows a better assessment of the likelihood that empty bins actually correspond to missing age fractions,

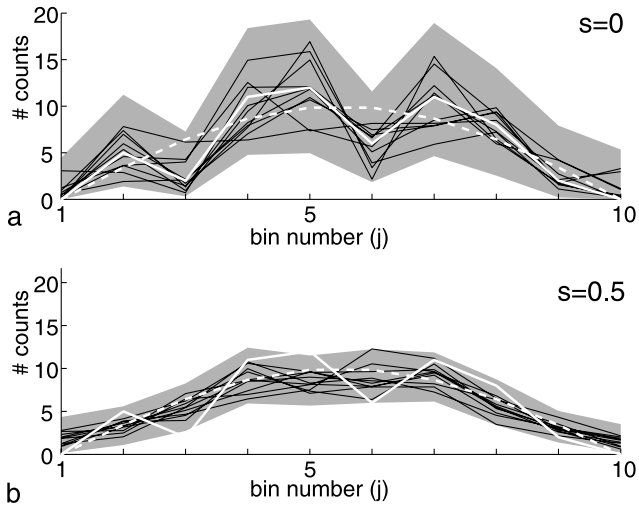


Figure 16. Smooth population (sine function). Dashed white lines show the histograms, or rather "frequency polygons" [*Scott*, 1992] of a smooth population (sine function). The solid white lines show the frequency polygons of a sample of 57 numbers that were randomly drawn from this population. The gray areas mark the simultaneous 95% credibility bands, obtained by the Bayesian method and based on 500 samples from the posterior distribution. Ten of these samples are shown in black to illustrate the effect of the smoothing prior. (a) The nonsmoothed Bayesian credibility band is about twice as wide as (b) the smoothed credibility band.

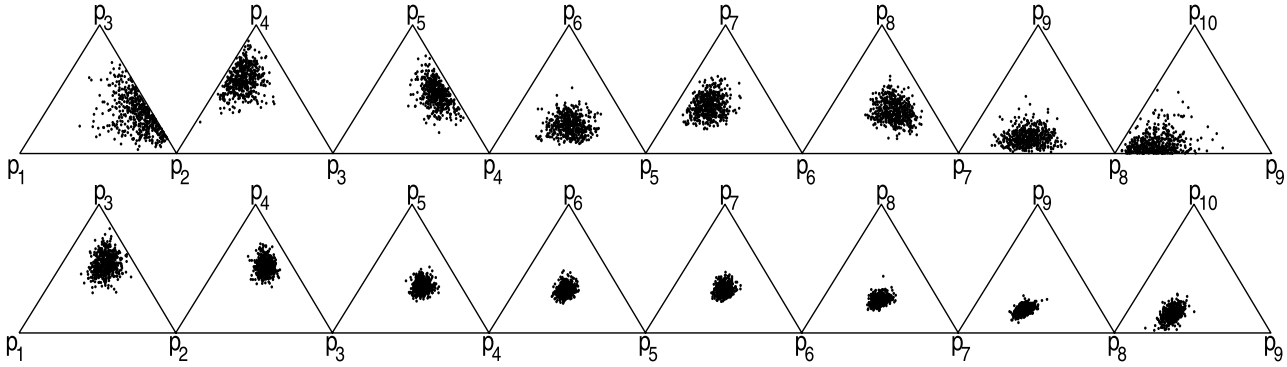


Figure 17. Trinomial "slices" through the (top) unsmoothed and (bottom) smoothed posteriors of Figure 16.

and of the statistical significance of some of the minor pre Pan-African peaks. The Nubian Sandstone example does not follow a very smooth distribution. Therefore it might not be necessary to apply any smoothing to it at all.

[28] This is certainly not the case for another data set, containing the ages of 155 lunar spherules, dated with the $^{40}Ar/^{39}Ar$ method, and published by *Culler et al.* [2000]. Figure 21a shows the simultaneous 95% credibility band of this age histogram without smoothing ($s = 0$), whereas for Figure 21b, a smoothing prior with s = 0.25 was used. The resulting credibility band is significantly narrower. To study the effect of the smoothing parameter s on the credibility band, the experiment of Figure 21 was repeated for a range of s values and the average width of the credibility band was calculated for each of them. Figure 22 shows how a moderate amount of smoothing can reduce the width of the credibility band by a factor of two, but that smoothing even more does not have much effect. This is because the exaggerated width of the unsmoothed credibility bands is mostly caused by just a few anomalous spikes (see Figure 16a). The sharpest of these excursions have the greatest effect on the width of the credibility bands, and will be filtered out the easiest. Smoothing out posterior samples that are less rough takes a lot more effort while having a much smaller effect. The fact that the magnitude of the smoothing parameter is not all that important reassures us that the rather arbitrary nature of the smoothing prior as defined by equations (10) and (12) is not a problem.

[29] *DeGraaff-Surpless et al.* [2003] presented a statistical analysis of the detrital U-Pb zircon age data sets shown in Figure 3. Using the Kolmogorov-Smirnov (K-S) test, they compared the different samples to see if these could have been derived from the same population. The conclusion was that samples KD3 and KD7 were compatible with each other on the 95% confidence level, but that sample KD26 was not. The same test can be done using Bayesian credibility bands, as shown on Figure 23. No smoothing prior was used for the construction of Figure 23 because it is not our goal to constrain the distribution of the underlying population, but only to see if the different observations are compatible with each other. As soon as any part of one histogram falls outside the 95% credibility band of the other, the former is not compatible with the latter. However, as discussed in section 5.1, in order to test if two samples are mutually compatible, we must construct two 97.5% credibility bands. If each of the two histograms completely falls inside the 97.5% credibility band of the other, there is at least 5% chance that they are compatible with each other. In addition to the K-S test and the Bayesian credibility bands, the χ^2 test is a third statistical method that was used to test the compatibility of the three samples. Its results are also shown on Figure 23. The three methods yield the same conclusions: samples KD3 and KD7 are compatible with each other, while KD26 is not. As a word of caution, we should repeat the remark made in section 2.1. Provided the number of measurements is large enough, eventually any test will fail, no matter how small the difference between the distributions. Instead of blindly looking whether or not a test has failed, it is better to consider the relative variation of the p values, ensuring that samples of roughly the same size are compared, or to use a different measure of "distance" between distributions [e.g., *Sircombe and Hazelton*, 2004].

[30] Finally, we return to the histogram of dip angles of 33 reverse faults from *Collettini and Sibson* [2001]. Figure 24 shows the simultaneous 95% credibility band

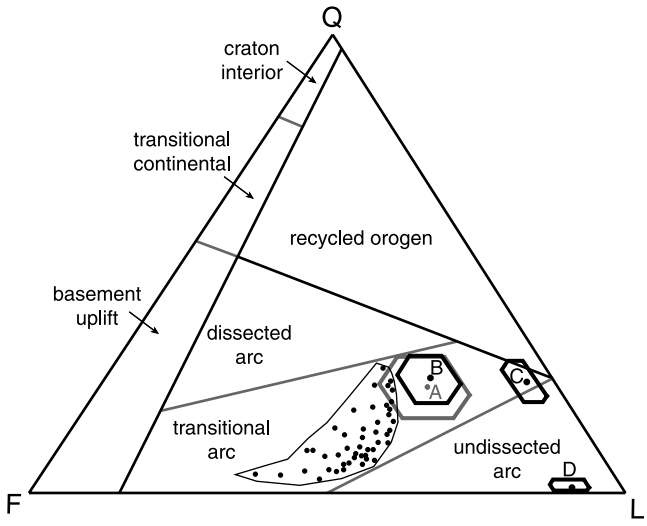


Figure 18. QFL diagram of Figure 1, with 95% credibility regions for samples A–D. The hexagon of sample A (200 point counts) is markedly larger than that of sample B (400 point counts). Sample C most likely falls inside the "transitional arc" field, but there is more than 5% likelihood that it belongs to either the "undissected arc" or "recycled orogen" field.

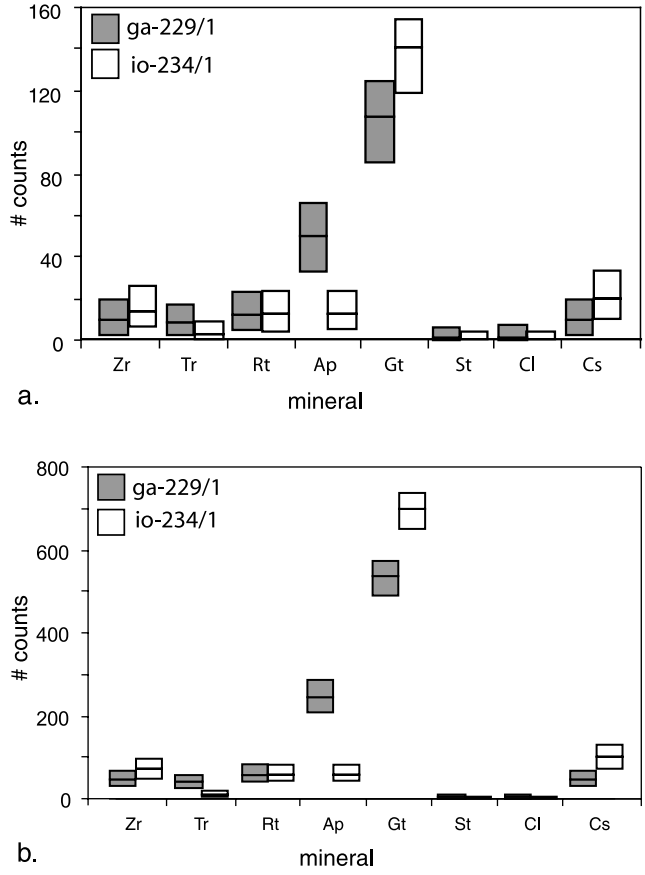


Figure 19. Heavy mineral analyses of Figure 2, with their 95% credibility intervals, using Jeffreys' prior (a) for the 200 counts of *Faupl et al.* [2002] and (b) if the same proportions had been obtained by counting 1000 grains. See Figure 2 for the key to the mineral abbreviations.

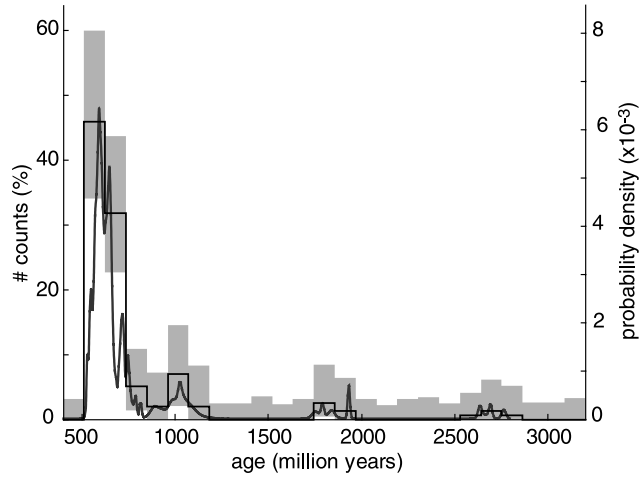


Figure 20. Detrital grain age histogram and kernel density function for 157 SHRIMP U-Pb zircon ages from the Cambrian Nubian Sandstone of southern Israel from *Avigad et al.* [2003]. The 95% credibility band for the histogram was calculated using Jeffreys' prior. We are more than 95% certain that each of the empty histogram bins contains less than roughly 5% of the population.

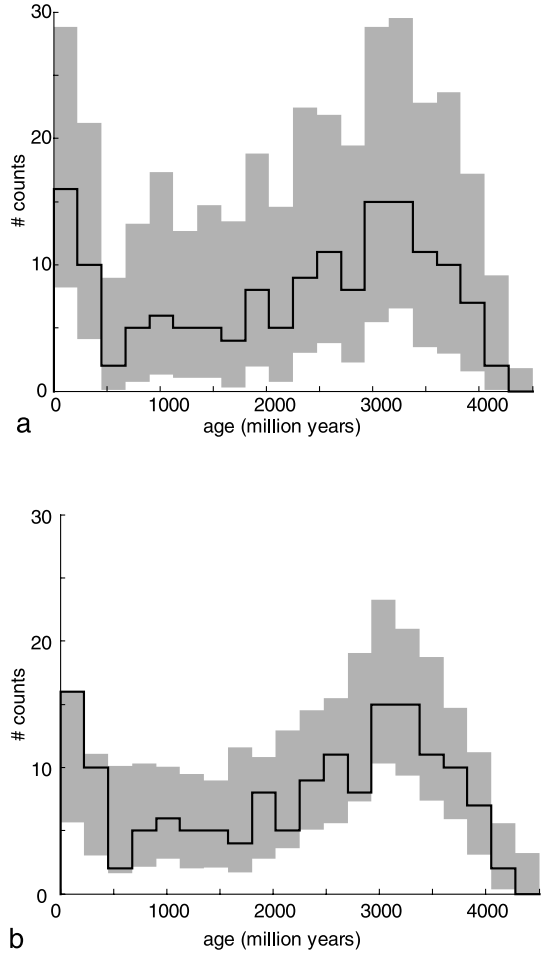


Figure 21. Histograms showing the ages of 155 lunar spherules, droplets of molten rock that result from meteorite impacts, measured with the $^{40}Ar/^{39}Ar$ method [*Culler et al.*, 2000]. These spherules record a time series of impact activity on the Moon's surface, which should be a more or less smooth function. (a) The 95% credibility band of histogram was calculated without a smoothing prior ($s = 0$). (b) Histogram calculated with a moderately strong smoothing prior ($s = 0.25$), resulting in a much narrower credibility band.

for this histogram. For Figure 24a, no smoothing prior was used (s = 0), while for Figure 24b, the smoothing parameter was s = 1. In either case, it is easy to find a monomodal histogram that integrates to 33, while completely fitting within the credibility band. This means that the apparent bimodality is not statistically significant on a 95% confidence level. Since the author is not a structural geologist, he cannot assess if such bimodality is an expected feature. If so, a bimodal prior distribution could be used instead of a uniform one. In that case, it is possible that the bimodality is statistically significant. However, without such prior information, it is not.

6. Conclusions

[31] In the Earth sciences, it is often not the data itself, but an estimate of its probability distribution (density) that is interpreted. This paper addressed the problem of how to

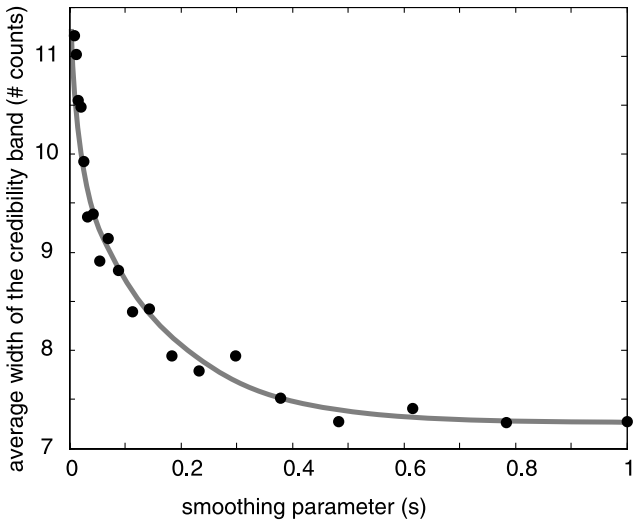


Figure 22. Exercise shown in Figure 21 repeated for a range of s values. This graph shows the evolution of the average credibility band width with s, suggesting that it is not necessary to use very strong smoothing.

quantify the statistical uncertainty on such interpretations. The histogram is one of the most convenient ways to represent data like petrographic point counts. We showed how to construct simultaneous Bayesian credibility bands for such categorical histograms. When only three categorical variables are studied, the ternary diagram is an alternative method of visualization, to which the method developed in this paper is equally applicable. Credibility bands allow an assessment of the precision of point-counting results and a way to intercompare multiple samples. Histograms can also be used to estimate probability densities of continuous variables, such as radiometric ages in detrital thermochronology. The main alternative to the histogram for this purpose is the kernel density estimator. The advantages of the latter to the former are that (1) kernel density estimates yield continuous, rather than stepwise functions and (2) they explicitly take into account measurement uncertainties, whereas histograms do not. On the other hand, histograms have the substantial advantage that it is possible to compute confidence bands for them, as described in this paper. This is far less obvious for kernel density estimates. When analytical uncertainties exist, it is good practice to use kernel density estimates in conjunction with histograms, including their credibility bands.

[32] Credibility bands provide a measure of the influence of counting statistics on density estimates. They also allow a better judgment of the possible similarities between different populations. If measurement uncertainties are small, the histogram is a good estimator of probability density, for which exact Bayesian credibility bands can be calculated. These have nonzero width even over intervals that were not

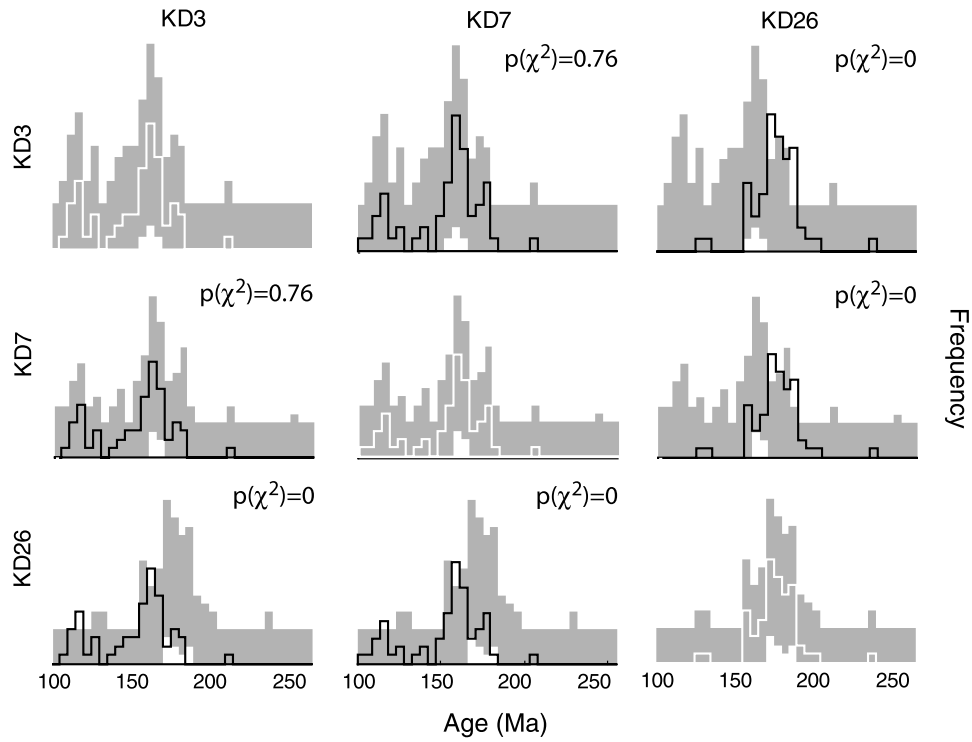


Figure 23. Intercomparison of three samples of *DeGraaff-Surpless et al.* [2003]. Since the histograms of samples KD3 and KD7 fall completely within each other's 97.5% credibility bands (using Jeffreys' prior), the two histograms are statistically "compatible" on the 95% confidence level. Parts of sample KD26 fall outside the 97.5% credibility bands of samples KD3 and KD7, and vice versa. Therefore a statistically significant difference exists between sample KD26 and the other two samples. The same conclusions were reached by doing a Kolmogorov-Smirnov test [*DeGraaff-Surpless et al.*, 2003] and a χ^2 test. The p values of the latter are also shown.

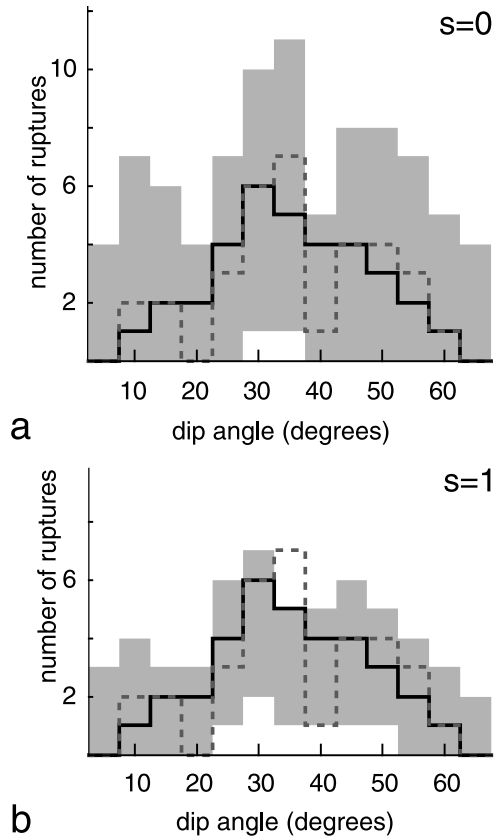


Figure 24. Data set of Figure 4 (dashed gray line) and simultaneous 95% credibility bands (gray shaded area) computed (a) without smoothing and (b) using a smoothing prior with smoothing parameter s = 1. In either case, the apparent bimodality observed in Figure 4 turns out not to be statistically significant because it is easy to fit a monomodal histogram (black line) inside the credibility band. Prior information or more data are needed to prove bimodality.

sampled. This is important in disciplines such as detrital thermochronology, where not just the presence, but also the absence of certain (age) components is important. The degree of confidence that certain age intervals are absent in a detrital population can be calculated analytically [*Vermeesch*, 2004]. For example, if 100 sediment grains are counted, there is up to 11% chance that at least one fraction $\geq$0.05 of the population was missed by that sample. (Bayesian) credibility bands such as those on Figure 20 are an alternative way to express this kind of uncertainty. Continuous histograms often represent time series, which are autocorrelated to some degree. In such cases, adjacent bins are not independent from each other, as was the case for categorical histograms. The Bayesian way to deal with this problem is to use an ad hoc smoothing prior, which can be completely determined by a single smoothing parameter. Smoothing helps to better constrain the probability distribution that underlies the data. Smoothing is not necessary if we merely want to see which other experimental outcomes are compatible with the observations.

[33] This paper describes a computer program named EPDU (an estimator of probability density and its uncertainties) that runs on both PC and Macintosh computers. This program is available online at http://pangea.stanford.edu/research/noble/epdu.

Appendix A

[34] To generate a large number B of synthetic samples from a multinomial distribution $\mathbf{p} = (p_1, .., p_M | \sum_{m=1}^{M} p_m = 1)$, we use a "bars and stars" procedure:

[35] 1. Generate the following vector of M + 1 numbers: $\mathcal{A} = (0, (p_1), (p_1 + p_2), .., (\sum_{m=1}^{M} p_m = 1))$. This represents the edges ("bars") of a histogram. The gaps between subsequent entries in this array represent the multinomial probabilities $(p_1, .., p_M)$.

[36] 2. Create a matrix $\mathcal{B}$ of size B $\times$ N with random numbers between 0 and 1, drawn from a uniform distribution. This represents B synthetic "samples" of N values ("stars").

[37] 3. For each row of $\mathcal{B}$, count the number of "stars" that fall in between the "bars" of $\mathcal{A}$. This procedure yields a matrix $\mathcal{H}$ of size B $\times$ M with multinomial replications of $\mathbf{p}$.

Appendix B

[38] The following procedure produces a random sample from a Dirichlet distribution $D_{\mathbf{a}}$: generate a vector $\mathbf{x} = (x_1, .., x_m, .., x_M)$ by drawing each of the x_ms from a gamma distribution with shape parameter a_m. Then $\Theta = (\theta_1, .., \theta_m, .., \theta_M)$ with $\theta_m = x_m / \sum_{m=1}^{M} x_m$ has the desired Dirichlet distribution [*Devroye*, 1986]. Alternatively, it is also possible to obtain a sample of the posterior distribution using a procedure named the "Bayesian bootstrap" [*Rubin*, 1981].

[39] To numerically integrate the Dirichlet distribution, we use either the "traditional method" of *Devroye* [1986], or the Bayesian bootstrap. Both methods give the same results. Thus we generate a $B \times M$ matrix $\mathcal{H}$ containing B random samples from the Dirichlet posterior of interest, each representing a histogram of M bins. The following procedure finds a polygon on Δ_{M-1}, containing 100(1-α)% of the posterior distribution:

[40] 1. Construct a two-sided 100(1-γ)% credibility interval for each of the columns ("bins") of this matrix. This can be done either analytically with equation (8), or numerically by computing the 100(α/2) and 100(1-α/2) percentiles. This yields M independent credibility intervals.

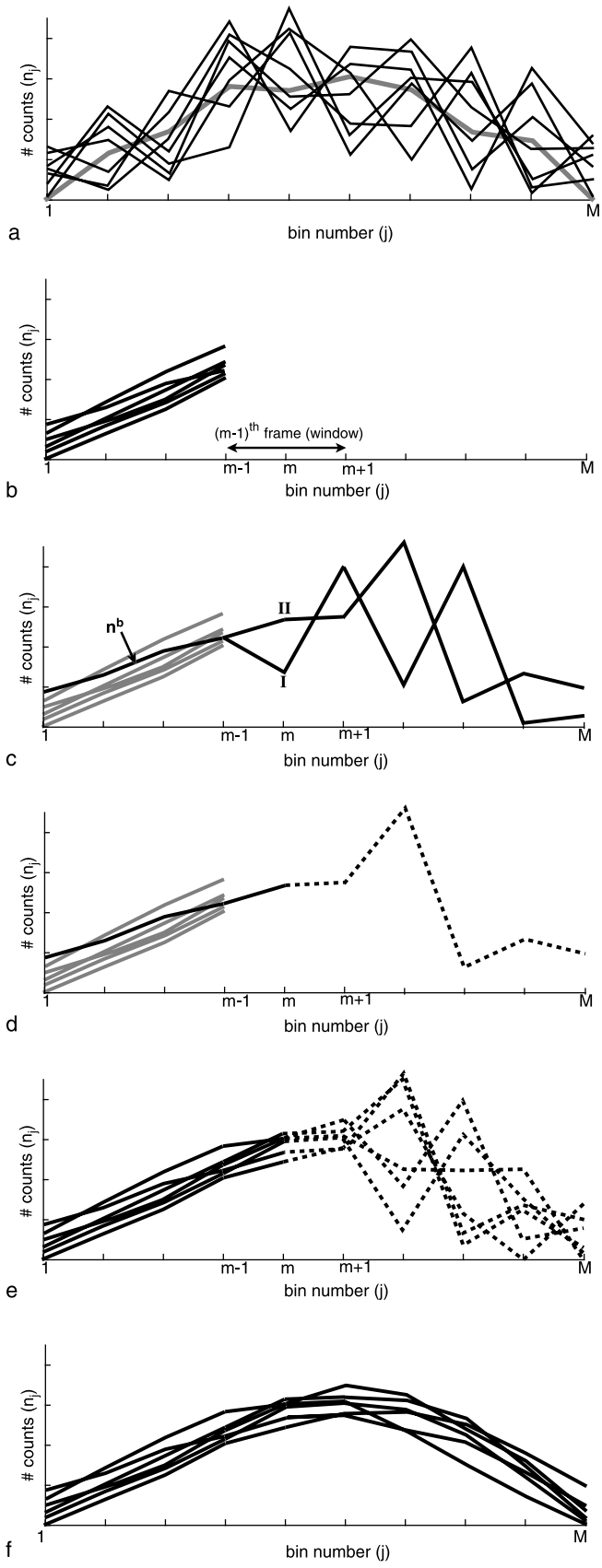
[41] 2. For each column, accept those values (rows) that fall within its respective credibility interval and reject those rows that fall outside of it. Divide the number of rejected rows by B (the total number of rows), and call this fraction ρ. If $\delta = \rho - \alpha > 0$, repeat steps 1 and 2 for a larger γ. If $\delta <$ 0, repeat them for a smaller γ.

[42] 3. Stop the iteration if δ is small enough (e.g., <0.001). The independent 100(1-γ)% credibility intervals for each of the bins then correspond to simultaneous 100(1-α)% credibility bands for the entire histogram.

Appendix C

[43] First, we will explain how to smooth a Dirichlet posterior on Δ_2:

[44] 1. Generate a random Bayesian replicate $\mathbf{n}^b = (n_1^b, n_2^b, n_3^b)$ from the unsmoothed Dirichlet posterior as in the work by *Devroye* [1986] or Appendix B.



[45] 2. Calculate the roughness r of this sample with equation (11) and the smoothness weight w with equation (12). The latter is a number between zero (infinite roughness) and one (zero roughness).

[46] 3. Generate a random number between zero and one. If this number is greater than w, reject $\mathbf{n}^b$. If it is less than w, accept $\mathbf{n}^b$.

[47] 4. Repeat steps 1–3 for $b = 1 \ldots B$ until a large number B (e.g., 500) of samples from the posterior distribution have been accepted.

[48] Next, we describe the method for extending this method to histograms of more than three bins, using a "sliding window" approach. Figure C1a shows B Bayesian replicates from an unsmoothed posterior distribution. On Figure C1a, the gray line shows the observed bin counts $\mathbf{n} = (n_1, \ldots, n_M)$ in the form of a "frequency polygon" [*Scott*, 1992]. The solid lines show B multinomial replicates from the Dirichlet posterior: $\mathbf{n}^b = (n_1^b, \ldots, n_M^b)$, with $1 \leq b \leq B$. In this example $B = 7$, but in real applications a more typical value would be $B > 500$. Assume that the first $m - 1$ bins have already been smoothed (Figure C1b). Figures C1c and C1d then show how to find the bth replicate for the next bin. We generate an array $\mathbf{n}^b = (n_1^b, .., n_M^b)$, where $(n_1^b, .., n_{m-1}^b)$ are "inherited" from the previous smoothing steps and $(n_m^b, .., n_M^b)$ are generated at random from $M - m + 1$ gamma distributions with parameters $(n_m, \ldots, n_M)$, respectively. Dividing $\mathbf{n}^b$ by $\sum_{j=1}^{M} n_j^b$ yields a sample $\mathbf{p}^b$ from the Dirichlet posterior ($\sum_{j=1}^{M} p_j^b = 1$) [*Devroye*, 1986]. Multiplying $\mathbf{p}^b$ by $N = \sum_{j=1}^{M} n_j$ gives a Bayesian replication of the original histogram $\mathbf{n}$. We now only consider one "trinomial frame" of this Bayesian replicate $(n_{m-1}^b, n_m^b, n_{m+1}^b)$ and calculate its roughness and the corresponding smoothness weight with equation (12). Using the same decision rule as before, we either accept or reject the mth bin. Two examples of such "multinomial extensions" to the bth Bayesian replicate are shown on Figure C1c, labeled I and II. The latter candidate will have a much greater chance of being accepted than the former.

[49] After repeating this procedure B times, we obtain a set of B multinomial extensions to the previously smoothed set of samples from the Dirichlet posterior (Figure C1e). We discard the replicates from the $(m + 1)$th bin onward and just

Figure C1. Illustration of the "sliding window" approach to smoothing the posterior distribution. (a) The solid black lines represent $B = 7$ Bayesian replicates generated without smoothing ($s = 0$). The dashed line marks the frequency polygon of the true population. (b) Suppose that the first $(m - 1)$ bins have already been smoothed, then B smoothed replicates for the mth step can be generated as follows: (c) to get the first replicate, first try the multinomial extension I and randomly accept or reject it based on its smoothness weight (equation (12)). Since replicate I is quite rough at the mth bin, it is very likely to be rejected. If this is the case, generate a new replicate. (d) Replicate II happens to be much smoother at the mth bin, giving it a greater chance of being accepted. If accepted, discard the replicates from $(m + 1)$ onward. (e) Repeat steps in Figures C1c and C1d until B values for the mth value have been accepted. (f) Repeat steps in Figures C1c–C1e until all M bins have been filled.

keep the values for the mth bin. The procedure then recursively smooths the remaining bins one by one, in exactly the same way as described before. However, at the $(M - 1)$th frame, we have to accept not only the $(m - 1)$th but also the mth bin in order to end the recursive process. Likewise, to start the recursive process, we must accept both the first and the second bin of the first trinomial "frame" during the first step of the smoothing procedure. To avoid any "edge effects", we "pad" the vector **n** of observed bin counts with zeros. This is more than just a trick, because it is an implicit assumption of the histogram that the number of observations outside its support is zero.

[50] At the end of the smoothing procedure, the set of posterior samples looks like Figure C1f. The smoothed replicates are "more parallel" than the unsmoothed ones shown in Figure C1a. Thus it becomes clear why the width of the smoothed credibility band is smaller than that of its unsmoothed counterpart, as illustrated by an example of the stepwise smoothing procedure on synthetic data shown in Figures 16 and 17.

[51] **Acknowledgments.** The author wishes to thank Susan Holmes and Albert Tarantola for introducing him to Bayesian statistics and advising him on the writing of this paper. However, these people are not responsible for any inaccuracies or errors that may exist in the paper.

References

Avigad, D., K. Kolodner, M. McWilliams, H. Persing, and T. Weissbrod (2003), Provenance of northern Gondwana Cambrian sandstone revealed by detrital zircon SHRIMP dating, *Geology*, *31*(3), 227–230.

Bayes, F. R. S. (1763), An essay towards solving a problem in the doctrine of chances, *Philos. Trans. R. Soc. London*, *53*, 370–418.

Blyth, C. R. (1986), Approximate binomial confidence limits, *J. Am. Stat. Assoc.*, *81*(395), 843–855.

Clopper, C. J., and E. S. Pearson (1934), The use of confidence or fiducial limits illustrated in the case of the binomial, *Biometrika*, *26*(4), 404–413.

Collettini, C., and R. H. Sibson (2001), Normal faults, normal friction?, *Geology*, *29*(10), 927–930.

Culler, T. S., T. A. Becker, R. A. Muller, and P. R. Renne (2000), Lunar impact history from $^{40}Ar/^{39}Ar$ dating of glass spherules, *Science*, *287*(5459), 1785–1788.

DeGraaff-Surpless, K., J. B. Mahoney, J. L. Wooden, and M. O. McWilliams (2003), Lithofacies control in detrital zircon provenance studies: Insights from the Cretaceous Methow Basin, southern Canadian Cordillera, *Geol. Soc. Am. Bull.*, *115*(8), 899–915.

Devroye, L. (1986), *Non-uniform Random Variate Generation*, 843 pp., Springer, New York.

Dickinson, W. R. (1985), Interpreting provenance relations from detrital modes of sandstones, in *Provenance of Arenites*, *NATO ASI Ser. C, Math. Phys. Sci.*, vol. 148, edited by G. G. Zuffa, pp. 333–361, Springer, New York.

Dickinson, W. R., L. S. Beard, G. R. Brakenridge, J. L. Erjavec, R. C. Ferguson, K. F. Inman, R. A. Knepp, F. A. Lindberg, and P. T. Ryberg (1983), Provenance of North American Phanerozoic sandstones in relation to tectonic setting, *Geol. Soc. Am. Bull.*, *94*(2), 222–235.

Faupl, P., A. Pavlopoulos, and G. Migiros (2002), Provenance of the Peloponnese (Greece) flysch based on heavy minerals, *Geol. Mag.*, *139*(5), 513–524.

Gill, J. (2002), *Bayesian Methods: A Social and Behavioral Sciences Approach*, 459 pp., CRC Press, Boca Raton, Fla.

Good, J. I. (1965), *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*, 109 pp., MIT Press, Cambridge, Mass.

Jeffreys, H. (1946), An invariant form for the prior probability in estimation problems, *Proc. R. Soc. London, Ser. A*, *186*, 453–461.

Rice, J. A. (1995), *Mathematical Statistics and Data Analysis*, 602 pp., Duxbury, Pacific Grove, Calif.

Rubin, D. B. (1981), The Bayesian bootstrap, *Ann. Stat.*, *9*(1), 130–134.

Scott, D. W. (1992), *Multivariate Density Estimation: Theory, Practice, and Visualization*, 317 pp., John Wiley, New York.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, 175 pp., CRC Press, Boca Raton, Fla.

Sircombe, K. N., and M. L. Hazelton (2004), Comparison of detrital age distributions by kernel functional estimation, *Sediment. Geol.*, *171*, 91–111.

Van der Plas, L., and A. C. Tobi (1965), A chart for judging the reliability of point counting results, *Am. J. Sci.*, *263*(1), 87–90.

Vermeesch, P. (2004), How many grains are needed for a provenance study?, *Earth Planet. Sci. Lett.*, *224*(3–4), 441–451.

Weltje, G. J. (2002), Quantitative analysis of detrital modes; statistically rigorous confidence regions in ternary diagrams and their use in sedimentary petrology, *Earth Sci. Rev.*, *57*(3–4), 211–253.

Yue, Y., B. D. Ritts, and S. A. Graham (2001), Initiation and long-term slip history of the Altyn Tagh Fault, *Int. Geol. Rev.*, *43*(12), 1087–1093.

P. Vermeesch, Department of Geological and Environmental Sciences, Stanford University, 450 Serra Mall, Room 320-305, Stanford, CA 94305, USA. (pvermees@pangea.stanford.edu)