15

A computational Bayes approach to some common archaeological problems

C. E. Buck

C. D. Litton

(Department of Mathematics, University of Nottingham, Nottingham, NG7 2RD)

15.1 Introduction

Over the past few years it has become abundantly clear that much more information can be obtained from archaeological data by postulating a mathematical model and using the data to estimate the parameters of the model than by merely using simple descriptive statistics to summarise the data. Indeed Fieller and Flenley (1988) neatly make this point in their paper on particle size analysis in which they observe that:

> 'Analysis of particle size data can proceed by either of two distinct routes. The first attempts only to obtain a simple numeric summary and description of the observed data; the second postulates a statistical model for the distribution of the sizes and proceeds to estimate the parameters of the model by statistical techniques. The first approach describes only the actual data obtained, the second attempts to investigate the process underlying the data.'

We believe that this is just as true for other archaeological problem areas. If it is possible to investigate the underlying processes then this is the route we should take, provided only that computational resources allow. Of course, the model must be realistic, that is its assumptions should be reasonable from both archaeological and mathematical view points. Furthermore, it is desirable that these assumptions be examined and tested in some way.

In addition to an increased awareness of the necessity for mathematical model building in archaeology, it has been argued by both Orton (1980, p.220) and Ruggles (1986) that, as opposed to a classical approach for making inferences, there are significant advantages to using Bayesian methods. The Bayesian approach provides a logical and coherent framework within which prior beliefs about a problem may be updated in the light of sample data. As such it is a systematic learning process and has we believe much to offer archaeology. Bayesian analysis has been used to solve a number particular problems, for example: Freeman (1976) investigated the megalithic yard; Naylor and Smith (1988a) examined how to make inferences using several radiocarbon dates; Kadane and Hastorf (1987) have applied Bayesian analysis to the study of burnt plant remains as part of paleoethnobotanical investigations at a prehistoric site in Peru and Buck *et al.* (1988) and Buck and Litton (1989) developed methods for analysing archaeological field survey data. All the same, Bayesian analysis has not been used widely for making inferences regarding archaeological models.

There are two major reasons for this. Firstly, the implementation of Bayesian methods has been restricted because efficient procedures necessary to obtain appropriate summary statistics from which to make inferences have not been available. This is no longer the case as recent advances in numerical techniques, based upon the Gibbs sampler (Gelfand & Smith 1990), make the calculations simple to implement, if still rather time consuming. Secondly, archaeologists have found some of the concepts associated with Bayesian analysis rather difficult to grasp. This may well be associated as much with the statisticians inability to convey the advantages of Bayesian methods as with the archaeologists inability to comprehend. This situation cannot be resolved without real collaboration between archaeologists and statisticians.

In this paper we provide two illustrations of the use of the Bayesian methodology applied to archaeological methods, namely cluster analysis and seriation. Classical versions of both techniques are widely used in archaeology and/or archaeometry and have a large associated literature both in archaeological and statistical journals. We hope that through these examples more archaeologists will appreciate the wide applicability of Bayesian analysis and so be prepared to take part in the collaboration that is so highly desirable.

15.2 The Bayesian paradigm

We assume that we have a statistical model which involves a vector, θ, of k unknown parameters $\theta_1, \theta_2, \ldots, \theta_k$ and the sample data $\boldsymbol{x}$. The model defines a relationship between θ and $\boldsymbol{x}$ which gives rise to a *likelihood function* $l(\boldsymbol{x};\theta)$. In the Bayesian paradigm we view the unknown parameter vector θ as a realisation of a random variable Θ having a *prior density* represented by $p(\theta)$. Inferences about θ, given the data $\boldsymbol{x}$, are based on the *posterior density*, $p(\theta|\boldsymbol{x})$, which is calculated using Bayes theorem as

$$p(\theta|\boldsymbol{x}) = \frac{l(\boldsymbol{x};\theta)p(\theta)}{\int l(\boldsymbol{x};\theta)p(\theta)\mathrm{d}\theta},$$

where the integration is carried out over the appropriate range of θ. We note that in this formulation all our prior information is encapsulated in the prior density $p(\theta)$ and the likelihood carries the information provided by the data. Hence the posterior density describes our knowledge about θ after both the knowledge provided by the prior information and that from the data have been combined in a logical fashion.

The essential feature of Bayesian analysis is captured in the above equation. In any practical situation considerable effort and thought must be put into the development of

a suitable statistical model which captures its important features and also into how to specify our prior knowledge about the vector $\boldsymbol{\theta}$.

However, the real stumbling block to the practical implementation of the Bayesian paradigm, has been the difficulty in carrying out any necessary integrations. For example suppose we wish to make inferences regarding θ_1, just one of the parameters, then the marginal posterior density of θ_1 is

$$p(\theta_1|\boldsymbol{x}) = \int p(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\theta_2, \ldots, \mathrm{d}\theta_n$$

and the expectation of θ_1 is

$$E[\theta_1|\boldsymbol{x}] = \int \theta_1 p(\boldsymbol{\theta}|\boldsymbol{x})\mathrm{d}\boldsymbol{\theta}.$$

Both these results involve integrations for which analytic expressions can be found only in a very limited number of special cases. As a result over the last few years a considerable amount of research effort has been channeled towards providing suitable numerical and analytic approximation techniques for such calculations (see Naylor & Smith 1982, 1988b, Tierney & Kadane 1986). Even so the use of these techniques requires considerable expertise together with special computer software and so the problem remains intractable except for a highly trained specialist. However in a recent paper Gelfand and Smith (1990) have suggested a method based on the Gibbs sampler that, although not without difficulties, is comparatively simple to implement. Extensive applications of the method are given in Gelfand *et al.* (forthcoming) and Carlin *et al.* (forthcoming).

15.3 The Gibbs sampler

In order to implement the Gibbs sampler, we must be dealing with collections of random variables $\theta_1, \theta_2, \ldots, \theta_k$ for which their joint density is uniquely determined by the conditional density of θ_s given $\theta_1, \theta_2, \ldots, \theta_{s-1}, \theta_{s+1}, \ldots, \theta_k$ for $s = 1, 2, \ldots, k$. As stated in the previous section, in order to make inferences about θ_s we need to calculate the marginal density of θ_s for $s = 1, 2, \ldots, k$. Geman and Geman (1984) develop an algorithm, known as the Gibbs sampler, for extracting the marginal densities from the conditional densities. The algorithm is based upon simulating in a systematic fashion from the conditional densities as follows.

1. Choose some arbitrary starting values $\theta_1^{(0)}, \theta_2^{(0)}, \ldots, \theta_k^{(0)}$.
2. Simulate $\theta_1^{(1)}$ from the conditional density of $\theta_1|\theta_2^{(0)}, \ldots, \theta_k^{(0)}$.
   Simulate $\theta_2^{(1)}$ from the conditional density of $\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \ldots, \theta_k^{(0)}$.
   .....
   Simulate $\theta_k^{(1)}$ from the conditional density of $\theta_k|\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_{k-1}^{(0)}$.
   This completes one iteration of the algorithm.
3. Repeat 2 r-times with $\theta_i^{(0)}$ and $\theta_i^{(1)}$ replaced by $\theta_i^{(t-1)}$ and $\theta_i^{(t)}$ respectively $t = 2, 3, \ldots, r$.

Under fairly general conditions Geman and Geman (1984) show that the distribution of the simulated values $\theta_i^{(r)}$ tends to the distribution of θ_i. Thus if the whole process is repeated, say l times, the resulting collection of $\theta_i^{(r)}$ may be used as an estimate of the posterior density of θ_i.

15.4 Application to cluster analysis

Cluster analysis is one of the statistical techniques most widely used in the analysis of archaeological data sets. It is commonly employed as an exploratory technique for the identification of structure in complex, high dimensional data sets and used in this manner, no underlying model is assumed. However, in certain situations, models may be postulated (although they are not always explicitly stated) and some of these will be examined in this section. Specifically, in many of these cases there is information in addition to the actual data being studied which points to possible cluster membership. Two examples may be cited of situations in which such information exists; they belong to the general field of provenancing. In the first Leese *et al.* (1986) describe the use of cluster analysis to group medieval floor tiles from the East Midlands on the basis of their chemical composition. Here prior information is available for some of the tiles from the kiln sites. Hammond *et al.* (1977) attempt to locate the source of jade used to make artefacts found on Mayan sites in America. Here the prior information takes the form of chemical analyses from modern sources of jade and jadeite that might have been available to the Maya.

Alvey *et al.* (1979) describe another application in which cluster analysis distinguished groups of clay pipes found in Nottingham that could have been made from the same mould. Only those pipes were used in the study that could be clearly identified, by means of a stamp, as having been made by a particular maker. Thirteen measurements of each pipe were taken in order to provide an adequate description of its shape and size (see Alvey *et al* 1979 for more details). A standard cluster analysis program was used in an attempt to identify groups of similar pipes based upon these measurements. However, prior information regarding the variability of the measurements of pipes made from the same mould, although available at the time, could not be incorporated into the analysis. This prior information comes from a collection of twenty-one pipes from the Bristol City Museum (see Jackson & Price 1974) for which there is very strong evidence that the pipes were made from the same mould. The technique we describe here permits the inclusion of such prior information.

The measurements from twenty-one pipes made from the same mould were examined for normality. It appears reasonable to suppose that measurements of pipes made from the same mould have a multivariate normal distribution (see Taylor 1989 for details of this analysis).

Returning to the collection of Nottingham pipes. Let $\boldsymbol{X}_i$ represent the measurements on the ith pipe and assume that

there are M possible moulds. If the ith pipe was made from the mth mould then

$$X_i \sim N(\mu_m, \Sigma_m),$$

where μ_m and Σ_m are the mean and covariance matrix for the mth mould. Assuming that the variability does not depend upon the mould, it is reasonable to take $\Sigma_m = \Sigma$, for $m = 1, 2, \ldots, M$. Thus the clustering problem may be viewed as the determination of the number of components in a mixture of multivariate normals with different but unknown means and a common (unknown) covariance matrix. Previous work on this problem has been carried out by Wolfe (1970), Day (1969), Binder (1978) and Symons (1981). Within the CLUSTAN package this is carried out using the NORMIX option with maximum likelihood being used for the parameter estimation. However the number of clusters has to be chosen by the user as this is not automatically done by the software. CLUSTAN does output a 'likelihood ratio' criterion for examining the relative likelihood of various numbers of clusters. This is very much an approximate test and almost inevitably the number of clusters is arrived at subjectively. If normality is not assumed then a suitable form of hierarchical clustering could be used. In this case the number of components or non-empty clusters is arrived at subjectively by examining what is known as the 'fusion plot'.

We now formulate the model in a way suitable for use with the Gibbs sampler. The density of X_i is given by

$$f(x_i|\lambda_m, \mu, \Sigma) = \sum_{m=1}^{M} \lambda_m f(x_i|\mu_m, \Sigma),$$

where the λ_m are the mixture parameters. Let $\phi_i = m$ if the ith pipe comes from the mth mould or equivalently X_i comes from the mth component with probability λ_m.

For prior densities we assume that λ has a Dirichlet distribution with parameter vector α, that is

$$\lambda' = (\lambda_1, \lambda_2, \ldots, \lambda_M) \sim D(\alpha)$$

We assume that the prior density for the mean of the mth cluster, μ_m, has a normal distribution with expectation η_m and covariance matrix $\rho_m^{-1}\Sigma$, where η_m is assumed to be normally distributed with expectation η_0 and covariance matrix Σ_0. Finally we assume that Σ^{-1}, the inverse of the covariance matrix, has a Wishart distribution with β degrees of freedom and precision matrix τ.

Then the conditional densities are

1. $\Sigma^{-1}|X, \lambda, \phi, \mu_1, \mu_2, \ldots, \mu_M \sim$ Wishart with degrees of freedom $\beta+n_1+n_2+, \ldots, +n_M$ and precision matrix

$$\tau^* = \tau + \sum_{m=1}^{M} s_m + \sum_{m=1}^{M} \frac{\rho_m n_m}{\rho_m + n_m}(\eta_m - \bar{x}_m)(\eta_m - \bar{x}_m)'$$

2. $\mu_m|X, \Sigma, \lambda, \phi, \mu_i (i \neq m) \sim N(\frac{\rho_m \eta_m + n_m \bar{x}_m}{\rho_m + n_m}, (\rho_m + n_m)^{-1}\Sigma)$
3. $\lambda|X, \Sigma, \phi, \mu \sim D(\alpha + n)$
4. $p(\phi_i = m|X, \mu_1, \mu_2, \ldots, \mu_M, \Sigma, \lambda) = p_m \propto \lambda_m \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp -\frac{1}{2}(x_i - \mu_m)'\Sigma^{-1}(x_i - \mu_m)$,

where n_m is the number of pipes in cluster m, $\bar{x}_m$ and $n_m^{-1}s_m$ are the sample mean and sample covariance matrix respectively for cluster m.

To illustrate our methodology we consider a collection of seventy-two pipes known to be made by the Brinsley family of Nottingham. We use the data from the Bristol collection as our prior information regarding the variability of pipes made from the same mould. We take $\beta = 20$ and $\tau = \beta s A s'$ where

$$s' = (0.29, 0.75, 0.69, 0.52, 0.55, 0.37, 0.04, 0.01, 0.33, 0.12, 0.88, 0.38, 0.47)$$

is the vector of standard deviations of the Bristol data; the corresponding correlation matrix is shown in Table 15.1.

To reflect our lack of prior knowledge of the mixture parameters we take

$$\alpha' = (0.5, \ldots, 0.5)$$

We take M, the maximum number of clusters, to be 50. We have to specify our beliefs about η_m for each cluster. This we do by setting

$$\eta_0' = (5.0, 25.0, 22.0, 6.5, 0.44, 8.9, 0.43, 1.0, 8.5, 2.1, 4.8, 7.1, 5.9)$$

which is the vector of sample means for the Bristol data,

$$\Sigma_0 = 25 s A s'$$

and restricting the η_m to be within an appropriate region. The region used was chosen to be slightly larger than the region bounded by the maximum and minimum values of each variable for the seventy-two pipes in the study. In this way, the η_m are spread over the region of interest. At each iteration, new values of η_m are simulated for each cluster. The effect of this is to discourage the formation of small groups of one or two pipes. The pipes are allocated to initial clusters using an equation similar to 4 above and each simulation is run for $r = 20$ iterations. A hundred replications (i.e. $l = 100$) of the experiment result in the pipes being divided into four, five and six clusters on seventy-five, twenty-four and one occasion respectively. From this we conclude that it is highly likely that four moulds were used to produce this collection of pipes.

These results should be compared with the 'likelihood ratio' criterion obtained using the NORMIX option of CLUSTAN.

H_0	vs.	H_1	Test Statistic	Approximate Probability of a More Extreme Value
3 clusters		4 clusters	31.9	0.003
4 clusters		5 clusters	23.3	0.087
5 clusters		6 clusters	19.0	0.299
6 clusters		7 clusters	12.8	0.812

This suggests that four or five clusters are likely although it must be remembered that this test is approximate and its properties are not fully understood. From the 'fusion plot' produced by hierarchical clustering it is also unclear whether four or five groups are present.

$$A = \begin{bmatrix} 1.00 & 0.07 & 0.47 & 0.25 & 0.11 & 0.30 & 0.23 & 0.38 & 0.16 & 0.01 & -0.26 & 0.74 & 0.21 \\ 0.07 & 1.00 & -0.13 & 0.09 & 0.02 & -0.29 & -0.70 & 0.16 & -0.46 & 0.22 & 0.40 & 0.15 & -0.01 \\ 0.47 & -0.13 & 1.00 & 0.62 & -0.69 & 0.38 & 0.26 & -0.09 & 0.47 & -0.28 & -0.49 & 0.30 & -0.06 \\ 0.25 & 0.09 & 0.62 & 1.00 & -0.51 & -0.03 & -0.13 & -0.01 & 0.33 & -0.02 & 0.02 & 0.15 & 0.20 \\ 0.11 & 0.02 & -0.69 & -0.51 & 1.00 & 0.15 & -0.05 & 0.27 & -0.30 & 0.54 & 0.16 & 0.13 & 0.01 \\ 0.30 & -0.29 & 0.38 & -0.03 & 0.15 & 1.00 & 0.24 & -0.04 & -0.05 & 0.25 & -0.37 & 0.21 & -0.17 \\ 0.23 & -0.70 & 0.26 & -0.13 & -0.05 & 0.24 & 1.00 & -0.39 & 0.41 & -0.24 & -0.48 & 0.25 & 0.00 \\ 0.38 & 0.16 & -0.09 & -0.01 & 0.27 & -0.04 & -0.39 & 1.00 & -0.04 & 0.14 & -0.16 & 0.22 & -0.17 \\ 0.16 & -0.46 & 0.47 & 0.33 & -0.30 & -0.05 & 0.41 & -0.04 & 1.00 & -0.29 & -0.49 & 0.12 & -0.08 \\ 0.01 & 0.22 & -0.28 & -0.02 & 0.54 & 0.25 & -0.24 & 0.14 & -0.29 & 1.00 & -0.08 & 0.04 & -0.03 \\ -0.26 & 0.40 & -0.49 & 0.02 & 0.16 & -0.37 & -0.48 & -0.16 & -0.49 & -0.08 & 1.00 & -0.12 & 0.42 \\ 0.74 & 0.15 & 0.30 & 0.15 & 0.13 & 0.21 & 0.25 & 0.22 & 0.12 & 0.04 & -0.12 & 1.00 & -0.01 \\ 0.21 & -0.01 & -0.06 & 0.20 & 0.01 & -0.17 & 0.00 & -0.17 & -0.08 & -0.03 & 0.42 & -0.01 & 1.00 \end{bmatrix}$$

Table 15.1:

Archaeologists will be rightly concerned as to how sensitive our results are to changes in the prior information. For example suppose that one hundred and one Bristol pipes instead of twenty-one had generated the same values of s and A. Setting $\beta = 100$ and $\tau = 100sAs'$ resulted in the pipes being grouped into four, five, six and seven clusters on twenty-nine, fifty-seven, eleven and three occasions respectively. That is the stronger our belief that variability in the Bristol pipes is applicable to the Nottingham pipes, the more groups we find.

Another experiment we have conducted is to multiply the standard deviations in s by factors of 0.9 and 1.1 while keeping $\beta = 20$. In both cases the distribution of the number of clusters remained as in our original experiment. This also occured for $\beta = 100$.

We have described here how prior information may be incorporated into cluster analysis using Bayesian techniques. Clearly such an approach is not limited to the analysis of pipe data, in fact it has wide applicability to a variety of archaeological situations. The nature of the prior information available may be varied, relating for example to the cluster means, cluster membership or even the number of clusters. Prior information abounds in archaeology, but is commonly not made explicit and invariably is not incorporated into the statistical analysis. It is usually used only in retrospect to assess the validity or otherwise of the cluster membership or the number of clusters.

We appreciate that archaeologists will not only have problems understanding the Bayesian approach and its emphasis on prior information, but will encounter real practical problems in specifying the latter's form. Close collaboration between archaeologists and statisticians is obviously required.

15.5 Application to seriation

Seriation has developed as a technique for ordering chronologically, archaeological contexts such as graves, by using the incidence or relative frequencies of particular artefact types found in them. Robinson (1951) and Kendall (1971a) describe an underlying model of seriation in which the relative popularity of an artefact never falls then rises again. Although not always explicitly stated this model is the only one regularly adopted for general chronological seriation. Both Kendall (1971b) and Laxton (1976) develop the theory of this model and show, by using a similarity measure, how to test whether a data set fits it. However, many archaeological data sets clearly cannot be seriated using these methods because no such ordering of the data exists. To overcome this Laxton and Restorick (1989) have proposed a method which takes account of the 'noisy' nature of the data. Other responses to this problem have been the application of correspondence analysis (see Djindjian 1989 and Madsen 1988) and multidimensional scaling (see Boneva 1971). As far as we are aware, one drawback of all the currently used methods of seriation is that they produce an ordering but give no indication as to other orderings that may fit the model almost as well. That is a method finds the 'best' ordering using the algorithm incorporated in it, but fails to inform the user of the second, third or fourth 'best' orderings. In this paper we provide a means of identifying other possible orderings together with some assessment of their relative chances of occurring.

We now develop a stochastic model which we will use for seriation. The parameters of the model will be estimated using the Bayesian methodology outlined earlier. Let $\theta_j(t)$ be the underlying proportion of artefact j present in a society at time t. Suppose that context i occurred at (unknown) time t_i and let for simplicity

$$\theta_{ij} = \theta_j(t_i).$$

That is θ_{ij} is the underlying proportion of artefact type j when context i was deposited. Suppose we observe n_{ij} artefacts of type j associated with the ith context and that we assume that the n_{ij} $j = 1, \ldots, J$ are a sample from the population proportions θ_{ij} $j = 1, \ldots, J$. If we view the problem in this manner we may assume that $n_i = (n_{i1}, \ldots, n_{iJ})$ has a multinomial distribution with parameters $\theta_i = (\theta_{i_1}, \ldots, \theta_{i_J})$ and $m_i = \sum_{j=1}^{J} n_{ij}$.

Of course, we acknowledge that there are considerable problems with this approach. How are the underlying populations defined? How can it be verified that the observed frequencies arise from some form of random sampling from the population? Madsen (1988) provides a powerful argument as to why this approach based upon the assumption of random sampling in seriation is very suspect.

Nevertheless we will proceed with our analysis in order to demonstrate our overall methodology. No doubt better, more realistic models could be formulated, but the overall strategy would be the same. Therefore, granting that the

model is reasonable and using Bayes Theorem the posterior density of $\theta_i|n_i$ is given by

$$p(\theta_i|n_i) \propto p(n_i|\theta_i)p(\theta_i),$$

where

$$p(n_i|\theta_i) \propto \prod_{j=1}^{J} \theta_{ij}^{n_{ij}}$$

and

$p(\theta_i)$ is the prior density of θ_i.

The natural choice as a prior for θ_i is the Dirichlet distribution with parameter vector α_i, that is

$$p(\theta_i) \propto \prod_{j=1}^{J} \theta_{ij}^{\alpha_{ij}-1}.$$

In which case, the posterior density of $\theta_i|n_i$ is

$$D(n_i + \alpha_i).$$

Vague prior information regarding θ_i may be expressed by taking $\alpha_{ij} = 0.5$.

Given this model and the Gibbs sampler methodology, the analysis is easily implemented as we now describe.

1. For each context i, simulate θ_i from a $D(n_i + \alpha_i)$ distribution.
2. Test whether in some order $\theta_1, \theta_2, \ldots, \theta_I$ seriate. If so, record that order. If not, no order is recorded.
3. Repeat 1 and 2.

As an illustration, we apply our methodology to a highly simplified data set given in Laxton and Restorick (1989) and reproduced in table 15.2.

Using Kendall's method, the seriation of the sites is 2, 5, 3, 6, 1, 4. However, as Laxton and Restorick (1989) point out, correspondence analysis produces the order 3, 6, 5, 2, 1, 4 which does not satisfy Kendall's model as, in this order, the relative artefact frequencies do fall and then rise again in popularity.

Using vague prior information, we applied the Bayesian methodology to this data set and ran the simulation until 1000 complete orderings of the six sites were obtained. The results are given in 15.3.

It is clear from 15.3 that the order produced using Kendall's method is the most likely seriation of the data given his model. We note that there are a number of other orders which represent possible seriations of the data. It is less likely for any of the orderings, other than that of Kendall, to be correct. Kendall's ordering has a posterior probability of 0.679, while the next most likely alternative has a posterior probability of only 0.128. This second alternative also does not satisfy Kendall's model.

It is interesting that the order 3, 6, 5, 2, 1, 4 (that obtained using correspondence analysis) is also represented in the results from our analysis. However it occurs with a posterior probability of only 0.016 and is not therefore seen as a likely correct ordering of the data.

We suggest that the method given above may be useful for assessing the robustness of the answer produced by a seriation algorithm. All we have to do is perturbate the data in some reasonable fashion. In our example we have done this by viewing the observed frequencies as realisations of a multinomial distribution and then sampled from the posterior distribution of the underlying proportions. Other more justifiable assumptions could no doubt be made here and we invite suggestions. However, once we have perturbated the original data, we process it through our seriation algorithm. We then repeat this many times and hence assess the robustness or otherwise of our original answer. In doing so we will give the archaeologist some idea of how good the original answer is and what other possible orders there are together with some indication of their relative importance. Furthermore, using this technique, other models of seriation can easily be assessed. In order to arrive at such alternative models, further collaboration between archaeologists and statisticians will be necessary.

15.6 Conclusions

We hope that it is clear from what goes before that in many situations much more can be obtained from archaeological data by using statistical modelling than by using simple descriptive statistics. We appreciate that the implementation of this modelling approach has up to now been hampered by the difficulty of developing models whose parameters can readily be estimated. By using Bayesian techniques this is not really a problem. In previous work (Buck & Litton 1989) we have pointed out that Bayesian analysis has applications in the processing of archaeological geochemical data. In this paper we have introduced applications in two other fields and hope that from this archaeologists will be able to appreciate the wide ranging applicability of the techniques. What is needed now is real collaboration between archaeologists and statisticians in order to develop explicit models for the underlying processes which give rise to other types of archaeological data. This will then allow statisticians to demonstrate the power that Bayesian inference can have in these fields as well.

Acknowledgements

We are grateful to our colleagues Dr. R. R. Laxton, Prof. A. F. M. Smith and Dr. D. A. Stephens for their encouragement in this work and for their comments on earlier drafts of the paper. The first named author acknowledges support from the U.K. Science and Engineering Research Council's Complex Stochastic Systems Initiative.

Bibliography

ALVEY, R. C., R. R. LAXTON, & G. F. PAECHTER 1979. "Statistical analysis of some Nottingham clay tobacco pipes", *in* Davey, P., (ed.), *The Archaeology of the Clay Tobacco Pipe*, pp. 229–253. British Archaeological Reports, Oxford. British Series

Site	Artefact Type						
	1	2	3	4	5	6	7
1	20	3	4	42	18	0	13
2	85	3	12	0	0	0	0
3	26	40	8	0	0	26	0
4	20	1	4	13	58	0	4
5	67	10	23	0	0	0	0
6	26	29	8	3	0	33	1

Table 15.2: Artefact counts for the six sites

Ordering	Frequency	Posterior Probability
2, 5, 3, 6, 1, 4	679	0.679
2, 5, 3, 6, 4, 1	128	0.128
2, 5, 6, 3, 1, 4	86	0.086
4, 1, 6, 3, 2, 5	46	0.046
2, 5, 6, 3, 4, 1	28	0.028
3, 6, 5, 2, 1, 4	16	0.016
4, 1, 3, 6, 2, 5	8	0.008
3, 6, 5, 2, 4, 1	4	0.004
5, 2, 3, 6, 4, 1	3	0.003
4, 1, 2, 5, 3, 6	1	0.001
5, 2, 6, 3, 4, 1	1	0.001

Table 15.3: Results from 1000 replications of ordering the six sites

63.

BINDER, D. A. 1978. "Bayesian cluster analysis", *Biometrika*, 65: 31–38.

BONEVA, L. I. 1971. "A new approach to a problem of chronological seriation associated with the with the works of Plato", *in* Hodson, F. R., Kendall, D. G., & Tautu, P., (eds.), *Mathematics in the Archaeological and Historical Sciences*, pp. 173–185. Edinburgh University Press, Edinburgh.

BUCK, C. E., W. G. CAVANAGH, & C. D. LITTON 1988. "The spatial analysis of site phosphate data", *in* Rahtz 1988, pp. 151–160.

BUCK, C. E. & C. D. LITTON 1989. "Image segmentation for archaeological geochemical data", *in* Rahtz & Richards 1989, pp. 133–144.

CARLIN, B. P., A. E. GELFAND, & A. F. M. SMITH forthcoming. "Hierarchical Bayesian analysis of change-point problems". forthcoming.

DAY, N. E. 1969. "Estimating the components of a mixture of normal distributions", *Biometrika*, 56 (3): 463–474.

DJINDJIAN, F. 1989. "Fifteen years of the data analysis French school to quantitative archaeology", *in* Rahtz & Richards 1989, pp. 207–218.

FIELLER, N. & E. FLENLEY 1988. "Statistical analysis of particle size and sediments", *in* Rahtz 1988, pp. 79–94.

FREEMAN, P. R. 1976. "A Bayesian approach to the megalithic yard", *Journal of the Royal Statistical Society A*, 139: 20–55.

GELFAND, A. E., S. E. HILLS, A. RACINE-POON, & A. F. M. SMITH 1988. "Illustration of Bayesian inference in normal data models using Gibbs sampling", *Journal of the American Statistical Association*.

GELFAND, A. E. & A. F. M. SMITH 1990. "Sampling based approaches to calculating marginal densities", *Journal of the American Statistical Association*, 85: 398–409.

GEMAN, S. & D. GEMAN 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6 (6): 721–741.

HAMMOND, N., A. ASPINALL, S. FEATHER, J. HAZELDEN, T. GAZARD, & S. AGRELL 1977. "Maya jade: source location and analysis", *in* Earle, T. K. & Ericson, J. E., (eds.), *Exchange Systems in Prehistory*, pp. 35–67. New York Academic Press, New York.

JACKSON, R. G. & R. H. PRICE 1974. *Bristol clay pipes, a study of makers and their marks*. Bristol City Museum: Research Monograph, No. 1.

KADANE, J. B. & C. A. HASTORF 1987. "Bayesian paleoethnobotany", *in Third Valencia International Meeting on Bayesian Statistics (Invited Papers)*.

KENDALL, D. G. 1971a. "Abundance matrices and seriation in archaeology", *Zeit Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 17: 104–112.

KENDALL, D. G. 1971b. "Seriation from abundance matrices", *in* Hodson, F. R., Kendall, D. G., & Tautu, P., (eds.), *Mathematics in the Archaeological and Historical Sciences*, pp. 215–252. Edinburgh University Press, Edinburgh.

LAXTON, R. R. 1976. "A measure of pre-Q-ness with applications to archaeology", *Journal of Archaeological Science*, 3: 43–54.

LAXTON, R. R. 1987. "Some mathematical problems in seriation", *Acta Applicandae Mathematicae*, 10: 213–235.

LAXTON, R. R. & J. RESTORICK 1989. "Seriation by similarity and consistency", *in* Rahtz & Richards 1989, pp. 229–240.

LEESE, M. N., M. J. HUGHES, & J. CHERRY 1986. "A scientific study of N. Midlands medieval tile production", *Oxford Journal of Archaeology*, 5 (3): 355–370.

MADSEN, T. 1988. "Multivariate statistics and archaeology", *in* Madsen, T., (ed.), *Multivariate Archaeology: Numerical Approaches in Scandinavian Archaeology*, pp. 7–27. Jutland Archaeological Society Publications XXI, Moesgard.

NAYLOR, J. C. & A. F. M. SMITH 1982. "Applications of a method for the efficient computation of posterior distributions", *Applied Statistics*, 31: 214–225.

NAYLOR, J. C. & A. F. M. SMITH 1988a. "An archaeological inference problem", *Journal of the American Statistical Association*, 83 (403): 588–595.

NAYLOR, J. C. & A. F. M. SMITH 1988b. "Econometric illustrations of novel numerical integration strategies for Bayesian inference", *Journal of Econometrics*, 38: 198–126.

ORTON, C. 1980. *Mathematics in Archaeology*. Collins, London.

RAHTZ, S. P. Q., (ed.) 1988. *Computer and Quantitative Methods in Archaeology 1988*, International Series 446, Oxford. British Archaeological Reports.

RAHTZ, S. P. Q. & J. D. RICHARDS, (eds.) 1989. *Computer Applications and Quantitative Methods in Archaeology 1989*, International Series 548, Oxford. British Archaeological Reports.

ROBINSON, W. S. 1951. "A method for chronologically ordering archaeological deposits", *American Antiquity*, 16: 293–301.

RUGGLES, C. 1986. "You can't have one without the other? I. T. and Bayesian statistics and their possible impact within archaeology.", *Science and Archaeology*, 28: 8–15.

SYMONS, M. J. 1981. "Clustering criteria and multivariate normal mixtures", *Biometrics*, 37: 35–43.

TAYLOR, K. 1989. "On measuring clay tobacco pipes". Undergraduate Project, Department of Mathematics, University of Nottingham.

TIERNEY, L. & J. B. KADANE 1986. "Accurate approximations for posterior moments and marginal densities", *Journal of the American Statistical Association*, 81: 82–8.

WOLFE, J. H. 1970. "Pattern clustering by multivariate analysis", *Multivariate Behavioural Research*, 5: 329–350.