Dingwall, L., S. Exon, V. Gaffney, S. Laflin and M. van Leusen (eds.) 1999. Archaeology in the Age of the Internet. CAA97. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 25th Anniversary Conference, University of Birmingham, April 1997 (BAR International Series 750, CD-ROM). Archaeopress, Oxford.

Bayesian Seriation as a Tool in Archaeology

Ulrich Halekoh and Werner Vach

Abstract

The seriation of objects based on findings or traits accompanying them is a major task in archaeological research. An example is the chronological ordering of graves based on the gravegoods found therein. Correspondence analysis is a well established tool to solve this task but has only limited capability to show the accuracy of the provided seriation. We introduce a new approach based on a stochastic model and a Bayesian analysis which allows for a given dataset not only to determine an estimate for the underlying order, but also to asses its variability. This hints to subgroups of graves that are nearly exchangeable with respect to their relative chronological position and hence contributes to prevent over-interpretation. We also provide means to characterize traits that carry substantial chronological information and may be useful for the classification of new graves. Additionally, our approach may enable the archaeologist to incorporate prior information as sex-specific classifications of graves or their stratigraphic relationships.

1 Introduction

Seriation or the reconstruction of the chronological order of finds is a major task in archaeological research. Lacking direct information (e.g. C^{14} analyses) one has to derive a chronological order by the combination of traits in the finds. Such an analysis is based on the assumption that each trait occurs only in a subperiod of the time covered by all findings and that within this subperiod its frequency first increases and then decreases (Robinson 1951, Mayer-Oaks 1955).

One well established exploratory tool for this task is correspondence analysis (Benzecri 1973, Hill 1974, Greenacre 1984, Baxter 1994). Although it is capable of detecting an underlying dominant chronological order it does not allow to draw conclusions on the precision of an estimated order and due to the lack of an explicit model it remains often unclear how to incorporate available additional information. To overcome these difficulties we present in this paper a new approach based on a stochastic model and a Bayesian analysis.

We introduce first the stochastic model and our choice for the prior distributions of the parameter necessary for the application of the Bayesian method. Then we focus on the analysis of the posterior distribution, first with respect to the chronological order and secondly with respect to the identification of traits most useful for chronological classification. Finally, we illustrate the analysis with data from two excavation fields and demonstrate by an artificial example an important difference between our approach and correspondence analysis if we analyse data carrying more information than only a chronological order.

2 The method

In general, a Bayesian analysis consists in specifying a stochastic model $P(Y|\theta)$, , describing the probability to observe the data Y in dependence of a parameter vector θ and the specification of a prior distribution $P(\theta)$. The latter reflects our knowledge on the parameter θ prior to observing the data. Our knowledge on the parameters after observing the data is then given by the so called posterior distribution, i.e. the distribution of the parameters given the data.

This can be derived by application of the Bayes rule, i.e.

$$P(\theta|Y) = \frac{P(Y|\theta)P(\theta)}{\int P(Y|\theta)P(\theta)d\theta}.$$

Any inference is based on this posterior distribution. In our setting the data are given as a binary

incidence matrix

$$Y: = \left(Y_{i,j}\right)_{i=1,\dots,I;\, j=1,\dots,J} \; with \; Y_{i,j} \in \{0,1\},$$

where $Y_{i,j}$ = 1indicates the presence of trait j at object i and we have overall I objects and J traits.

The stochastic model assumes, that all incidences $Y_{i,j}$ are independent given θ and that

$P(Y_{i,j} = 1 | \theta) = g_i p_{Ri,j}$,

Where R_i denotes the chronological position of the ith object, $p_{r,j}$ is the probability to observe trait j at chronological position r, and g_i is a quality indicator for each object i, representing for example the richness of a grave. Hence the parameter vector $\theta = (g, R, \theta_p)$ consists in our setting of three subvectors. $g = (g_1, ..., g_I)$ is the vector of the quality indicators and $R = (R_1, ..., R_I)$ is the rank vector representing the chronological order. The third vector θ_p contains for each trait j the parameters $p_{r,j}$ describing the course of the incidence in time, which are defined for $r \in O: \{1 - I, ..., 2I\}$. We assume an unimodal course in time which is described by additional auxiliary parameters $a_j, b_j, c_j \in O$ with $a_j \leq b_j \leq c_j$ and $p_j^{<a} > 0$, $p_j^{>c} > 0$ such that

$p_{r,j} < p_{(r+1),j}$ for $a_j - 1 \leq r < b_j$,

$p_{(r-1),j} > p_{r,j}$ for $b_j < r \leq c_j + 1$,

$p_{r,j} < p_j^{<a}$ for $r = 1 - I, ..., a_j - 1$,

$p_{r,j} < p_j^{>c}$ for $r = c_j + I, ..., 2I$.

i.e. we require that within the interval $[a_j, c_j]$ the parameters $p_{r,j}$ are first increasing, culminating at $b_j$and then decreasing. This is in accordance with the expectation in most archaeological settings that a specific trait comes into use at some point in time, becomes more and more widespread but is superseded by another trait and finally dies out. Outside the interval $[a_j, c_j]$ we allow small incidences $p_j^{<a}$and $p_j^{>c}$, respectively, to account for some violation of this strong expectation, for example due to a misclassification of a trait or the delayed occurrence of a trait due to passing it over several generations in a family. Note that we avoid any assumption on a uniform development in time; the traits are allowed to have different lifespans, a rapid or slow appearance or dying out, and several traits can occur or die out at the same time. Additionally, by defining the $p_{r,j}$ on an interval larger than the span of the data we achieve a natural solution to the boundary problem.

We turn now to the specification of the prior distributions for the parameters. Having no substantial prior knowledge we try to use uninformative priors in order to reflect our state of vague prior information. The components g_i of g are assumed to be independently and identically distributed uniformly on the interval [0,1]. R is assumed to be uniform on all permutations of $\{1, ..., I\}$ g,R and the components of θ_p for different traits are assumed to be independent. The prior distribution of $(p_{r,j})_{r \in O}$ is roughly spoken uniform on all vectors with unimodal structure. The actual implementation of this notion is described with the auxiliary parameters via successive conditioning. We first give the unconditional prior distribution of a_j and c_j, then conditionally on these parameters that of b_j, then the prior distribution of $p_{bj,j}$ given a_j, b_j, c_j, then the prior distribution of $p_{a_j,j}, \dots, p_{b_j-1,j}, p_{b_j+1,j}, \dots, p_{c_j,j}$ given $p_{b_j,j}$ and a_j, b_j, c_j, and finally the distribution of $p_j^{<a}$and $p_j^{>c}$ given $p_{a_j,j}$ and $p_{c_j,j}$ respectively. The prior of $(a_j, c_j)$is uniform on all pairs of integers in O. b_j is uniformly distributed between a_j and c_j. The maximum $p_{b_j,j}$ is assumed to be uniformly distributed on [0,1]. The vector $(p_{a_j,j}, \dots, p_{b_j-1,j}, p_{b_j+1,j}, \dots, p_{c_j,j})$is uniformly distributed on all vectors obeying the monotonicity and unimodality restriction. Finally, the parameters $p_j^{<a}$and $p_j^{>c}$ are uniform on the intervals $[0, p_{a_j,j}]$, and $[0, p_{c_j,j}]$ respectively.

3 Analysis of the posterior distribution of θ

In the analysis of the posterior $P(Y|\theta)$ we will first focus on the analysis of the most interesting aspect, the chronological order, i.e. we analyse the marginal posterior distribution of the ranks R. It is not possible to compute this distribution analytically. However, we are able to produce a sample of rank vectors from this distribution via application of Markov chain Monte Carlo methods (Gelfand and Smith 1990, Geyer 1991, Gilks and Wild 1992, Gilks et al. 1996). Details are given in Halekoh and Vach (in prep). Description of the posterior is then based on this sample.

3.1 Checking for multimodality

It may happen that there are two distinctly different orders underlying a data set. In this case the posterior distribution has two modes, i.e. there a two subregions in the space of all orders with high probability mass, but the space between the two regions has low probability mass. It is important to detect this situation because otherwise an analysis may yield misleading results. To detect multimodality, we need to introduce a distance measure on the rank vectors.

For two rank vectors R and $\tilde{R}$ we use the L_1-norm distance divided by I, i.e.

$$d*\left(R,\tilde{R}\right) := \frac{1}{I}\sum_{i=1}^{1}\left|R_i - \tilde{R}_i\right|.$$

It can be interpreted as the numbers of positions which each object has to move on average to change one directed order into the other. However, the stochastic model and the seriation problem itself is symmetric with respect to the direction of the object order. Hence each rank vector R represents simultaneously a second rank vector $\tilde{R}$, which represents the same order with opposite direction, i.e.

$$\overleftarrow{R}_i = I + 1 - R_i$$

Hence we have to define the distance of two rank vectors as

$$d\left(R,\tilde{R}\right) := \min\left(d*\left(R,\tilde{R}\right), d*\left(\overleftarrow{R},\tilde{R}\right)\right).$$

Now in order to check for multimodality we determine for each sampled vector the proportion of sampled rank vectors within its neighbourhood of diameter λ as a nonparametric estimator for its probability. A mode is then defined as a rank vector for which all rank vectors in its neighbourhood show a smaller probability. Then we can consider the number of modes as a function of λ. Checking for multimodality is then possible as in the setting of univariate densities (Silverman 1981): unimodality can be assumed if several modes occur only for very small values of λ.

3.2 Analysis of the object order

The most simple way to describe the posterior distribution of the object order is to consider the distribution of the rank of each object. However, this needs some care, as due to the symmetry of the problem the same order may be represented by two different rank vectors R and $\overleftarrow{R}$ in our sample. In the case of a single mode one can solve this problem by selecting one of the two possible directions of the mode order - which may be either artificially chosen or based on external archaeological information - and choosing for each order in the sample that rank vector with minimal distance to the directed mode order.

After directing each sampled rank vector we can speak of the distribution of the chronological position of each object. This distribution is easily graphically summarized by depicting for each object the boxplot of this distribution and one gets not only an impression of the mean chronological position of each object but also of its positional variability. Additionally, we can look for each pair of objects at the posterior probability that the first object precedes the other. The larger this probability the more confident we can be about the relative chronological order of the two objects.

Besides the mode order we can use that estimate for the chronological order which is induced by the order of the posterior means of the ranks of the objects. Alternatively, we can choose the order with minimal average distance to all sampled rank vectors, i.e. a Bayes estimate minimizing the expected loss. Global estimates of the variability can also be obtained, for example using the expected distance to the Bayes estimate.

3.3 Classification of traits

Traits which occur with sufficient frequency within a narrow time span provide the archaeologist with the ability to separate periods of time and enables him or her to develop classification rules for new objects. Since the parameters $p_{r,j}$ describe the probability of the occurrence of a trait the depiction of their posterior means can give an impression of the lifespan of the trait and its characteristic period. Traits where the posterior mean values for $p_{r,j}$ are high for a relatively small range of positions and low for the others will be the best candidates for identifying an underlying period. Traits where these parameters only gradually decrease or have high values for a large range of r will provide only weak information with respect to separating periods.

4 Examples

Example 1: Seriation of Merovingian graves based on bead data.

The data of our first example stem from an excavation of Merovingian graves from the burial site of Eichstetten, South Germany. Besides other goods, beads differing by size, colour, material, production technique and shape were found in the graves. We perform a seriation of the graves (i.e. objects) using as traits different predefined types of beads (Sasse and Theune 1996).

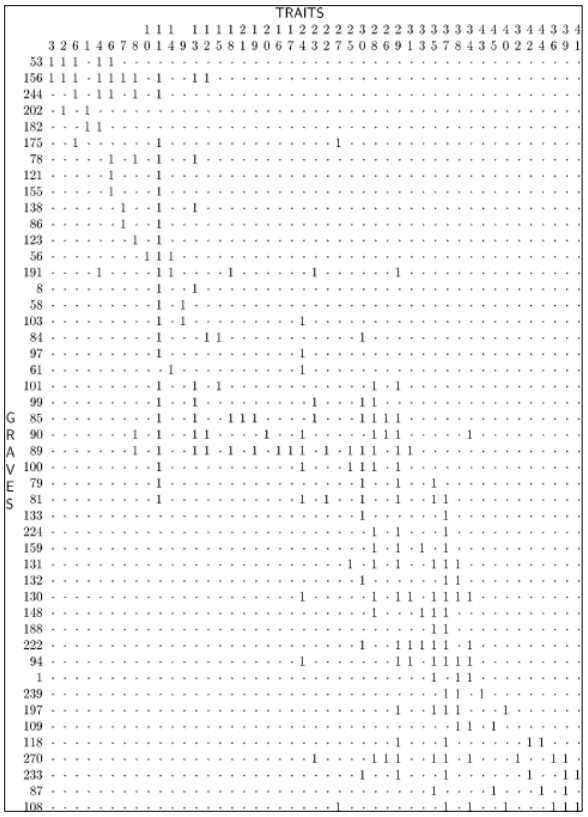

Figure 1. Incidence matrix of the Eichstetten data (columns=traits, rows=graves). The traits are numbered according to the original publication (Sasse and Theune 1996). The graves are ordered according to their mean rank position. The traits are ordered with respect to the mean rank of the objects associated with the trait.

The data are shown in Figure 1 where we have already used the order according to the mean posterior rank in our Bayesian analysis. For these data the analysis of multimodality gives no hints for several modes: all values of $\lambda \geq 3.9$ define the same mode estimate and the Bayes estimate as well as the order in Figure 1 differ only slightly from it. All estimates are very similar to the solution found by a correspondence analysis. The overall variability can be described by the expected posterior distance to the Bayes estimate which is 4.4. Furthermore, 95% of the posterior distribution is within a distance of 6.2 from the posterior mode, i.e. all these orders can be reached from the mode by moving each object no more than 7 positions on average.

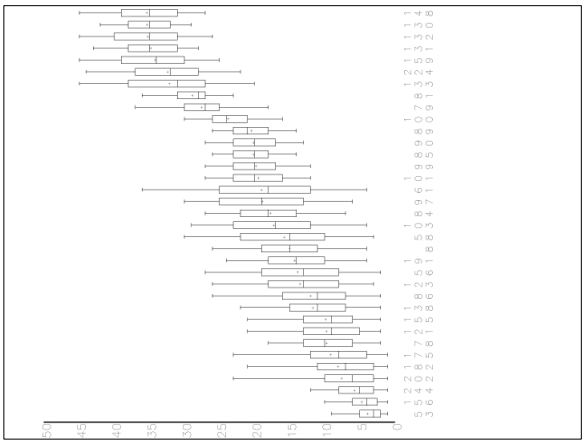

Figure 2. Boxplots of the posterior distribution of the rank of each object. The whiskers refer to the lower and upper 5% quantile, the '+' sign to the posterior mean.

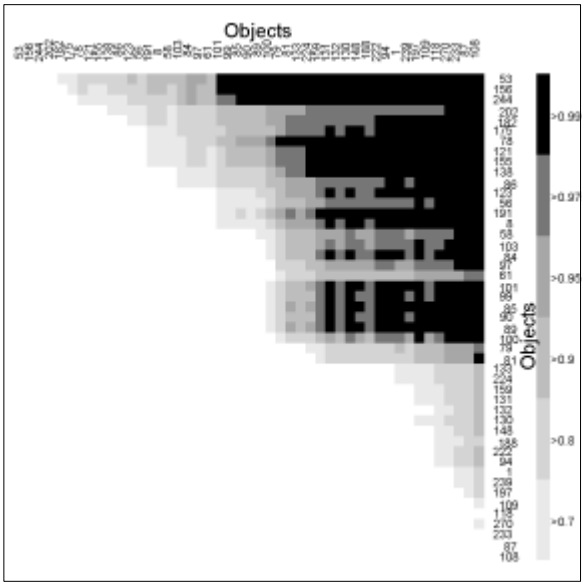

Figure 3. Posterior probabilities that an object shown in the row precedes an object shown in the column.

Figures 2 and 3 depict further aspects of the posterior distribution of the object order. Figure 2 summarizes the posterior distribution of the position of each object in boxplots. In this figure we can detect several groups of graves with a very similar posterior distribution of their ranks, for example the graves 202 to 155, 101 to 89 or 159 to 109. Within these groups any ordering seems to be more or less equally likely. Figure 3 shows for each pair of objects the posterior probability to find one object before the other. In this graph one can easily detect two groups of objects (53,....,89) and (133,....,108) which are well separated but show a highly variable internal chronological

structure. Additionally, within the first group the objects (101,....,89) can be nearly definitely placed after the first three objects of this group. Also Figure 2 conveys the impression that we cannot do much more but to distinguish early and late graves. Hence the proposal of Sasse and Theune (1996) to divide the chronological course of the objects in five small phases may be regarded as somewhat too optimistic.

One may argue that the variability of the position of an object is merely a function of the number of traits of the object and hence the efforts to produce Figures 2 and 3 would not be not justified. However, this is not true. For example the two objects 61 and 123 have both only two traits, but comparing the boxplots of Figure 2 object 123 has a much smaller variability. Indeed, object 123 has the two traits 8 and 10 which are both only present in the first half of the ordered objects and trait 10 is a 'strong' trait occurring at nearly all objects in this period. This results in a high probability to put this object before object 159 or later ones (cf. Figure 3). In contrast, object 61 has the two traits 13 and 25, which are both sparse and not very typical for a subperiod, such that positioning of this object is difficult.

Looking at the posterior distribution of $p_{r,j}$ in this small example is not very informative hence we switch now to the larger burial site of Weingarten where the same traits has been analysed by Sasse and Theune(1996).

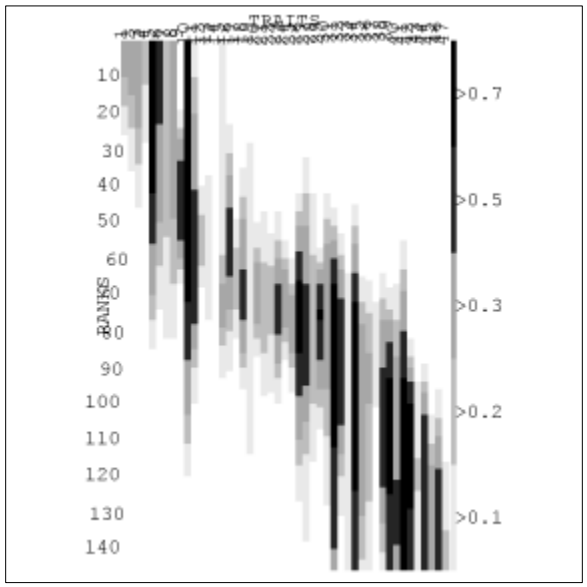

Figure 4. Posterior means of the occurrence probabilities $p_{r,j}$ for all traits j of the Weingarten data (Sasse and Theune 1996).

In Figure 4 we show the posterior means of $p_{r,j}$ in a grayscale figure. First, about half of the traits seem to be not very useful for chronological classification, as for none of their chronological positions the posterior mean of $p_{r,j}$ is larger than 0.5. Good candidates are the traits 16, 18, 23, and 29 allowing a rather sharp chronological classification of objects since the high values of the posterior means of their occurrence probabilities are bounded to a narrow chronological interval and outside this interval they decrease rapidly. In contrast, traits 6, 11, and 27 are less useful although they have a similar range of posterior means larger than 0.5, but simultaneously they have non negligible probabilities over a large chronological span. If we are only interested in a very coarse chronological classification traits 5, 10, 31 and 34 seem to be useful due to their long span with high posterior means and their rapid increase and decrease.

Example 2: An artificial data set carrying a chronological and a cluster structure

The results of a correspondence analysis can be erroneous if some of the selected traits do not represent the chronological order but other structures, e.g. sex groups, social groups or regional differences. With an artificial data set we want to demonstrate that our approach is also able to detect different structures if they are simultaneously present in the data. The example consists of 24 graves and 20 traits.

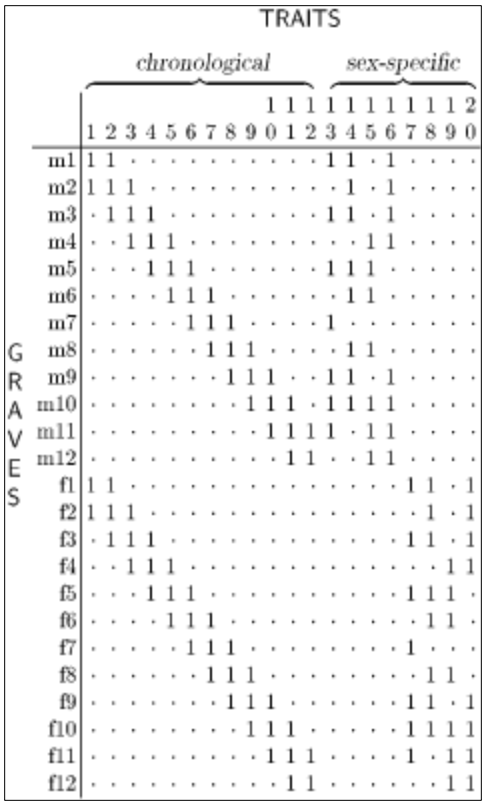

Figure 5. Example of an incidence matrix with chronological and group specific structure.

The first 12 traits carry chronological information and the last 8 traits group specific characteristics. The incidence matrix (see Fig. 5) may be regarded as an example for an excavation field with chronologically well ordered graves containing also sex specific gravegoods.

1.	2.	3.	1. mode	2. mode
f7	f6	f1	m2	f1
m7	f6	f10	m1	f12
f8	f7	f3	f2	f5
m8	m7	f9	f1	f10
f6	f5	f11	m3	f11
m6	m5	f12	f3	f3
f9	f2	f5	m4	f9
m9	m2	f2	f4	f2
f5	f3	f4	f5	f6
m5	m3	f6	m5	f8
f10	f4	f8	f6	f4
m10	m4	f7	m6	f7
f11	f1	m7	f7	m7
m11	m1	m6	m7	m6
f12	f8	m8	f8	m8
m12	m8	m4	m8	m3
f4	f9	m2	f9	m5
m4	m9	m5	m9	m4
f3	f10	m12	f10	m3
m3	m10	m11	m10	m10
f1	f11	m3	m11	m1
m1	m11	m9	m12	m9
f2	f12	m10	f12	m12
m2	m12	m1	f11	m11

Figure 6. Results of two analyses of the data shown in Figure 5. Columns 1-3: Object orders according to the eigenvectors of the first three nontrivial eigenvalues of a correspondence analysis; Columns 4 and 5: Modes from a Bayesian analysis.

Correspondence analysis is well known to be able to detect different structures in a data set by looking not only at the first eigenvector (e.g. Zimmermann 1995, Baxter 1994). However, for the given example, correspondence analysis fails to detect the chronological order (see Fig. 6); the first eigenvector puts the early male and female graves (m1-m4,f1-f4) just beside the late ones (m10-m12,f10-f12), the second eigenvector puts a middle group (f5-f7, m5-m7) at the top and the third eigenvector represents only the sex difference. The Bayesian analysis reveals two distinct modes; the first represents nearly exactly the joint chronological order of male and female graves, the second the two sex groups.

5 Outlook

We have so far concentrated on exploiting our Bayesian approach to describe the variability of a seriation of objects and did not use the ability of the approach to incorporate substantial prior information into the analysis. In the analysis of the artificial example 2 we considered both structures, the chronological and the sex-specific one, as unknown. It is quite conceivable that classification of the graves

with respect to sex is possible a priori from the group-specific gravegoods. This information can be incorporated into our analysis by extending our model to $P(Y_{i,j} = 1| \theta) = g_i p_{Ri,j} k_{i,j}$ where $k_{i,j} = 0$ if trait j is not compatible with the sex of the individual of grave i and $k_{i,j} = 1$ otherwise. With this extension of our parameterization only the (chronological) order remains as the unknown structure to be extracted in the analysis. By more complicated extensions we can also handle other factors, for example differences in the trait frequencies between different spatial regions.

Another straightforward extension is the use of prior knowledge on the order of objects. In many archaeological excavations one can often see several strata in the earth ground, such that objects found in one stratum must be earlier than objects found in a stratum above. The implied partial order of the objects is known as the 'Harris Matrix' in archaeology. To use this prior information we have to restrict the set of object orders to those compatible with the partial order. Another type of prior information exists in analysing graves of larger burial sites using the spatial distribution of the graves. Here it can be assumed that spatial closeness reflects partially a chronological order. Hence we can put higher prior probability on orders showing some degree of correlation with spatial closeness.

6 Discussion

Using an explicit stochastic model and a Bayesian analysis we are able to develop a new approach for seriation problems in archaeology. The main advantage of the approach is the possibility to describe the variability of the object order implied by the available data. This description may prevent overinterpreting of single point estimates and hence may lead to a more appropriate interpretation of archaeological results. Besides describing the variability of the order the Bayesian analysis allows also other useful investigations. We have shown the additional use in analysing traits with respect to their value for chronological classification. Other investigations try to detect outliers or to analyse the shape of the course in time. In the long run the Bayesian analysis will be probably most fruitful in incorporating external information either in the stochastic model or in the prior information.

Note that our Bayesian approach is different from the one proposed by Buck and Litton (1991) and described in Buck, Cavanagh and Litton (1996) and Laxton (1997). They do not incorporate the unimodality restriction into the prior distribution of the parameters but analyse the posterior by retaining only samples obeying this restriction. For large data sets this method may be computationally not feasible since these samples have only small posterior probability and hence a large amount of the samples must be disregarded.

Bayesian analysis has also some drawbacks. First, choice of the prior distribution is subjective and there exists no unique way to represent the state of 'no prior knowledge'. A partial solution to this problem are sensitivity analyses trying different prior distributions and comparing the results. In Halekoh and Vach (in prep.) we present a framework for such a sensitivity analysis, and in the examples presented in this article no great impact of the way of specifying uninformative priors was found. Secondly, generation of a sequence of samples from the posterior distribution by Markov Chain Monte Carlo techniques is not a trivial task and requires careful programming and diagnostics; a point further discussed in Halekoh and Vach (in prep.), too.

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft.

Bibliography

Baxter, M J, 1994 *Exploratory Multivariate Analysis in Archaeology*, Edinburgh University Press, Edinburgh

Benzecri, J P, 1973 *L'Analyse des donnees II: l'analyse de correspondances* (in French), Paris Dunod

Buck, C E, and Litton, C D, 1991 A computational Bayes approach to some common archaeological problems, in *Computer Applications and Quantitative Methods in Archaeology* (eds K Lockyear and S Rahtz), BAR International Series, 565, 93-99

Gelfand, A E, and Smith, A F M, 1990 Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, 85, 398-409

Geyer, C. J., 1991 Markov chain Monte Carlo maximum likelihood, in *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface* (ed. E.M. Keramides),

Gilks, W, Richardson, S, and Spiegelhalter, D J (eds), 1996 *Markov Chain Monte Carlo in Practice*, London: Chapman and Hall
Gilks, W R, and Wild, P, 1992 Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, 41, 337-348
Greenacre, M, 1984 *Theory and applications of correspondence analysis*, New York Academic Press
Halekoh, U, and Vach, W, in preparation *A Bayesian Approach to Seriation Problems in Archaeology*
Hill, M O, 1974 Correspondence analysis: a neglected multivariate method, *Applied Statistics*, 23, 340-354
Laxton, R, R, 1997 Seriation in Archaeology. Modelling, Methods and Prior Information, in *Classsification and Knowledge Organization Proceedings of the 20th Annual Conference of the Gesellschaft für Klassifikation* (R Klar and S Opitz), Heidelberg: Springer 617-630
Mayer-Oaks, W J, 1955 Prehistory of the Upper Ohio Valley. An introductory archaeological study, *Annals of the Carnegie-Museum*, 34
Robinson, W F, 1951 A method for chronologically ordering archaeological deposits, *American Antiquity*, 16, 293-301
Sasse, B, and Theune, C, 1996 Perlen als Leittypen der Merowingerzeit (in German), *Germania*, 74, 187-231
Silverman, B W, 1981 Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society B*, 43, 97-99
Zimmermann, A, 1995 Some aspects of the application of correspondence analysis in archaeology, in *Whither Archaeology? - Papers in honour of Evzen Neustupny* (M Kuna and N Venclova), Institute of Archaeology, Praha 255-263

Contact details

Ulrich Halekoh, Werner Vach
Centre for Data Analysis and Model Building
Eckerstrasse 1
D-79104 Freiburg
GERMANY
e-mail: *uh@fdm.uni-freiburg.de*