Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 45 (2004) 651 – 673
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A Bayesian approach to seriation problems in archaeology

U. Halekoh[a,*], W. Vach[b]

[a] Centre for Data Analysis and Modelling, University of Freiburg, Eckerstraße 1, D-79110 Freiburg i. Br., Germany

[b] Department of Statistics and Demography, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

Received 6 February 2001; received in revised form 6 December 2002

Abstract

In archaeology, the reconstruction of the relative chronology of objects (e.g. graves) is often based on absence/presence of information about finds (e.g. grave goods). Traditionally, this task is known as seriation. In this article the task is tackled by formulating a stochastic model for the relationship between the underlying grave order and the observed incidences and by analysing the data using the Bayesian method. In selecting a prior distribution the attempt has been to reflect the archaeological context, especially a potential preselection of specific types of finds suitable for the seriation task. In contrast to established methods for seriation, such as correspondence analysis, it is possible directly to describe the variability of the estimated order by analysing the posterior distribution of the order. Because the order of the graves is a non-numerical and high-dimensional parameter, special techniques for the analysis of the posterior distribution are required. Construction of a Markov chain Monte Carlo method to approximate the posterior distribution is also partially non-standard, since the distribution can be multi-modal and because a huge number of nuisance parameters are introduced to avoid parametric assumptions on the shape of the distribution of the types through time. An example illustrates the techniques and demonstrates the need for a sensitivity analysis in this setting. The framework of our approach can easily be extended either to adjust for known factors which influence the absence/presence or in order to incorporate prior information on the grave order.

Keywords: Archaeology; Bayesian analysis; Markov chain Monte Carlo; Seriation

* Corresponding author. Biometry Research Unit, Department of Animal Breeding and Genetics, Danish Institute of Agricultural Sciences, PO Box 50, DK-8830 Tjele, Denmark. Tel.: +4589991825; fax: +4589991200.

E-mail addresses: ulrich.halekoh@agrsci.dk (U. Halekoh), werner.v@statdem.sdu.dk (W. Vach).

doi:10.1016/S0167-9473(03)00034-3

1. Introduction

The reconstruction of the chronological order of finds is a major task in archaeological research. If neither direct information, such as stratigraphic sequences, written records or coins, is available nor sources like radiocarbon or dendrochronological dating can be used, it is necessary to gain as much relative chronological information as possible from the finds obtained during an excavation. In practice, particular features of such finds (e.g. decorative features or techniques used in manufacture) are used to define types. The seriation is then based on the basic assumption that each type occurs only in a sub-period of the time covered by the finds and that within this sub-period its frequency first increases and then decreases (Robinson, 1951; Mayer-Oakes, 1955).

For the purpose of seriation the relevant data from an excavation is often summarized as an incidence matrix. In the case of an excavation of a burial site the matrix indicates the absence/presence of grave good types in the graves. We will consider as an example results from an excavation of the cemetery at Münsingen–Rain. For 59 tombs (i.e. graves) the presence/absence of 70 grave good types like fibulae, torcs and bracelets was recorded. For the purpose of illustration Table 1 gives an extract of the full incidence matrix presented by Hodson (1968, plate 123). The basic data analytical task is now to reorder the rows and columns of the incidence matrix so that it shows approximately a diagonal structure. This diagonal structure reflects the basic assumption mentioned above. Heuristic procedures for this task (e.g. Ford, 1962; Goldmann, 1972) date back to Petrie (1899). Approaches of Kendall (1971a, b) and Laxton (1976) are based on specific properties of such matrices. Correspondence analysis (Benzecri, 1973; Hill, 1974; Greenacre, 1984) uses the similarity of the reordered matrix with a scatterplot of highly correlated variables and today it is a standard tool for seriation in archaeology (Madsen, 1988; Jørgensen, 1992; Baxter, 1994). All these methods lack an explicit stochastic model although there is a strong connection between correspondence analysis and Gaussian ordination (Ihm and van Groenewoud, 1984).

Table 1
Extract of the incidence matrix for the findings from the excavation of the cemetery at Münsingen–Rain

Graves	Traits									
	Open brace	Torc	Fibula bronze	Fibula iron	Wire ring	Globular-sleeve	Swollen brace	Antler disc	Band ring	Eye-ended brace	…
13	1	0	1	0	0	0	0	0	0	0	
32	1	1	1	0	0	0	0	0	0	0	
7	1	1	0	1	0	0	0	0	0	0	
9	0	1	1	0	1	0	0	0	0	0	
16	0	0	0	1	1	0	0	0	0	0	
23	0	1	1	1	0	1	0	0	0	0	
44	0	1	1	1	0	0	1	0	0	0	
12	0	1	1	1	1	1	0	1	0	0	
8a	0	1	0	1	0	1	0	0	0	0	
8b	0	0	1	0	0	0	0	0	1	0	
…											

Although correspondence analysis is a useful exploratory tool it has some drawbacks if it is to be used for statistical inference. It does not allow to draw conclusions directly about the precision of an estimated order and, due to the lack of an explicit model, it often remains unclear how we should incorporate available additional information. To overcome these difficulties, this paper presents an approach based on an explicit stochastic model and a Bayesian analysis. The stochastic model and the prior distribution chosen are described in Section 2. Methods to analyse the posterior distribution are described in Section 3, and Section 4 is devoted to an example. To approximate the posterior distribution we have made use of the Metropolis-coupled Markov chain Monte Carlo (MCMCMC) technique proposed by Geyer (Geyer, 1991; Geyer and Thompson, 1995). For the generation of the single chains, we use a sequence of Metropolis-Hastings steps; details are given in Section 5. The final discussion, in Section 6, summarizes the basic results and gives hints to some useful extensions.

There have already been some attempts at applying Bayesian methodology to seriation problems. Buck and Litton (1991) (see also Buck et al., 1996) and Buck and Sahu (2000) fit the traditional Q-matrix approach as well as correspondence analysis into a Bayesian framework and compare the fit of different parametric models. Our approach is different in the sense that we start from scratch and only incorporate the key assumption of seriation, namely the uni-modal course of types over time, into the model.

2. Stochastic model and prior distribution

2.1. Stochastic model

In this paper we will consider the analysis of the binary incidence matrix

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

where $Y_{i,j} = 1$ indicates the presence of grave good type j in grave i and there are I graves and J types.

The main parameter of interest is the relative chronological position or rank R_i of grave i among the I graves. To describe the course of the incidences of type j along the chronological positions r additional parameters $p_{r,j} \in [0,1]$ will be introduced. These parameters reflect the basic assumption mentioned in the introduction. Hence a position a_j exists, where a type comes into use, and a position c_j, after which it is no longer used. The $(p_{r,j})$ are uni-modal on the interval $[a_j, c_j]$, and the position of the mode is denoted by b_j. However, there is no necessity for a_j and c_j to lie within the interval $[1, I]$, hence allowance has been made for the possibility that the a_j, c_j and the mode b_j lie outside this interval. Moreover, allowance has been made that outside the interval $[a_j, c_j]$ small incidences $p_j^{<a}$ and $p_j^{>c}$, respectively, account for some violation. This could be, for example, due to a misclassification of a type or due to the delayed occurrence of a type which is passed on several generations in a family. Summarizing, we assume for each type j the existence of $a_j \leqslant b_j \leqslant c_j$ and $p_j^{<a} > 0$ and $p_j^{>c} > 0$, such that $p_{r,j} = p_j^{<a}$ for $r < a_j$, $p_{r,j} < p_{r+1,j}$ for $a_j - 1 \leqslant r < b_j$, $p_{r-1,j} > p_{r,j}$ for

$b_j < r \leqslant c_j + 1$, and $p_{r,j} = p_j^{>c}$ for $r > c_j$. Finally, for each grave i a parameter $g_i \in [0, 1]$ is introduced, reflecting the richness of the grave. At this point is possible to summarize all parameters into a parameter vector θ consisting of three components $\theta = (g, R, \psi)$ with $g := (g_1, \ldots, g_I)$, $R := (R_1, \ldots, R_I)$, a permutation of $1, \ldots, I$, and $\psi := (\psi_j)_{j=1,\ldots,J}$, with $\psi_j := (p_j, \eta_j, p_j^{<a}, p_j^{>c})$ where $p_j := (p_{r,j})_{r\in\mathbb{Z}}$, $\eta_j := (a_j, b_j, c_j)$.

In the stochastic model the probability of the occurrence of type j in grave i has to be linked to the chronological position R_i, the course over time p_j and the richness g_i. Obviously, this probability should increase with $p_{R_i,j}$ and g_i. Hence the choice of a simple multiplicative model

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

Adding the assumption that $(Y_{i,j})_{i=1,\ldots,I;j=1,\ldots,J}$ are independent given θ the likelihood of θ is equal to

$$\prod_{i=1,\ldots,I,j=1,\ldots,J} (g_i\, p_{R_i,j})^{Y_{i,j}} (1 - g_i\, p_{R_i,j})^{1-Y_{i,j}}.$$

2.2. Specification of the prior distribution

With respect to the choice of the prior distribution of θ, this paper will only consider the situation where there is no prior knowledge about the chronological order of the graves, i.e. where the data alone should be sufficient in answering the question of interest. This situation corresponds to the typical use of correspondence analysis as a tool for seriation. Hence, we assume that R has a uniform prior distribution on the permutations of $\{1, \ldots, I\}$. Similarly, we assume that we have no prior knowledge about g_i and choose uniform priors on $[0, 1]$ for each g_i. The prior density of ψ_j is a product of conditional densities specified below. Especially, the prior density of ψ_j is the product of the prior density of the pair (a_j, c_j) (1), the prior density of b_j given a_j and c_j (2), the prior density of $p_{b_j,j}$ given η_j (3), the prior density of $p_{a_j}, p_{a_j+1}, p_{b_j-1,j}, p_{b_j+1,j}\, p_{c_j}$ given $p_{b_j,j}$ and η_j (4) and finally the prior densities of $p_j^{<a}$ (5) and $p_j^{>c}$ (6) given the other parameters. We further assume that g, R and $\psi_1, \ldots, \psi_J$ are independent. Hence the prior density of θ is the product of the prior densities for g, R and the ψ_j.

With respect to the choice of the prior distribution of ψ_j, the fact must be taken into account that an archaeological data set is often the result of a long process, where the types have been selected for the task of seriation. Hence we may argue that, for example, short ranges of $c_j - a_j$ are a priori more probable than longer ranges. However, it is difficult to describe this prior knowledge exactly or to elicit it, especially if a published data set is analysed as in our example. On the other hand, some sensitivity of the posterior distribution for R to the specification of the prior distribution for $(\eta_j)_{j=1,\ldots,J}$, must be expected, especially since the ranges $(c_j - a_j)_{j=1,\ldots,J}$ influence the possible variation of R. Therefore, it seems sensible to consider a flexible class of prior distributions for ψ, and later on to investigate the impact of the choice of the prior distribution for the nuisance parameters, ψ, on the posterior distribution of R. In order to achieve such a framework, an indexing parameter $\alpha \in (0, 1]$ is introduced and the prior probability of a pair (a_j, c_j) is set proportional to $\alpha^{c_j-a_j}$

$$P(a_j, c_j) \propto \alpha^{c_j-a_j}. \tag{1}$$

Therefore, smaller values of $c_j - a_j$ will obtain relatively higher mass with decreasing α. Furthermore, the selection process of the types may have favoured types with a shape of $(p_{r,j})_{r\in\mathbb{Z}}$ roughly symmetric around b_j and maximal value $p_{b_j,j}$ close to 1, as such types are regarded as most valuable by the archaeologists. This must be taken into account in choosing the conditional distribution $v(b_j|a_j, c_j)$ for b_j given a_j and c_j and in choosing the conditional prior distribution for the maximal value $p_{b_j,j}$ given η_j. Hence, for $v(b_j|a_j, c_j)$ first the uniform distribution on $\{a_j, \ldots, c_j\}$ will be considered, i.e.

$$v(b_j|a_j, c_j) \propto I_{\{a_j,\ldots,c_j\}}(b_j). \tag{2}$$

Secondly, for the purpose of a sensitivity analysis, a triangular prior distribution on $\{a_j, \ldots, c_j\}$ will be used, in which $v(b_j|a_j, c_j)$ is modelled as proportional to $b_j - a_j + 1$, if $b_j \leqslant (a_j + c_j)/2$, and proportional to $c_j - b_j + 1$ otherwise. For the distribution of $p_{b_j,j}$ first a uniform distribution on the interval $[0, 1]$ will be considered, and subsequently, for the purpose of a sensitivity analysis, choices will be made from a class with densities

$$u(p_{b_j,j}) = (t + 1)p_{b_j,j}^t I_{(0,1)}(p_{b_j,j}) \quad \text{with } t \in [0, \infty). \tag{3}$$

For the distribution of the components $p_j^{(-b_j)} := (p_{a_j,j}, \ldots, p_{b_j-1,j}, p_{b_j+1,j}, \ldots, p_{c_j,j})$, given η_j and $p_{b_j,j}$, we will assume a uniform distribution on the possible values satisfying non-negativity and the given restrictions, i.e.

$$f(p_j^{(-b_j)}|p_{b_j,j}, \eta_j) = \frac{(b_j - a_j)!(c_j - b_j)!}{p_{b_j,j}^{(c_j-a_j)}} \times \prod_{r=a_j}^{b_j-1} I_{(0,\, p_{r+1,j})}(p_{r,j}) \prod_{r=b_j+1}^{c_j} I_{(0,\, p_{r-1,j})}(p_{r,j}). \tag{4}$$

Furthermore, we assume that, given η_j and $p_{a_j,j}, \ldots, p_{c_j,j}$, the two parameters $p_j^{<a}$ and $p_j^{>c}$ are independently, uniformly distributed on $(0, p_{a_j,j})$ and $(0, p_{c_j,j})$, respectively, such that

$$f(p_j^{<a}|p_j^{(-b_j)}, p_{b_j,j}, \eta_j) := \frac{1}{p_{a_j,j}} I_{(0,\, p_{a_j,j})}(p_j^{<a}), \tag{5}$$

$$f(p_j^{>c}|p_j^{(-b_j)}, p_{b_j,j}, \eta_j) := \frac{1}{p_{c_j,j}} I_{(0,\, p_{c_j,j})}(p_j^{>c}). \tag{6}$$

Hence, the prior distribution of p_j is a mixture of distributions $f(p_j|\eta_j)$ with densities composed of the factors (3)–(6), and with mixing probabilities proportional to $v(b_j|a_j, c_j)\alpha^{c_j-a_j}$.

So far we have assumed that a_j and c_j can take any values in $\mathbb{Z}$. For $\alpha = 1$ the prior given above is improper and this, unfortunately, implies an improper posterior. To circumvent this problem, it is necessary to restrict the possible set of values. To avoid a special treatment of the boundaries, it is convenient to allow a larger interval than simply the values $1,\dots,I$. Hence the original range is tripled and a_j and c_j are restricted to $\{1-I,\dots,2I\}$.

Finally, we note that the above model only allows identification and estimation of undirected grave orders, since it is reversible with respect to the grave order and the course of the incidences. This is specific to the problem, because the analysis of the combinations of types gives no indication of the beginning or ending of the order unless information about asymmetries in the course of $p_{r,j}$ is available a priori. Consequently, one should regard any value of (R, p) as representative of two equivalent values where the second value (R', p') satisfies $R'_i = I + 1 - R_i$ for all i and $p'_{r,j} = p_{I+1-r,j}$ for all r and j.

3. Analysis of the posterior distribution

In Section 5 we will describe the MCMC technique used to obtain a finite sequence $\theta^1,\dots,\theta^N$, which can be regarded as an approximation to the posterior distribution. Our major interest concerns the grave order, wherefore our focus will be on the analysis of the marginal posterior distribution of the grave order. A major prerequisite is an appropriate distance measure on the grave orders. This choice is discussed in the first subsection. Based on this measure we will consider a strategy to investigate a possible multi-modality of the posterior distribution and suggest point estimators for a grave order. Then we will discuss several tools to describe the variability of the posterior distribution of the grave orders. In this section, the N rank vectors in our sample are denoted by $R^1,\dots,R^N$.

3.1. Choice of the distance measure

Let $\vec{R}$ and $\vec{\tilde{R}}$ be two rank vectors representing *directed* orders as indicated by the "→" above the symbol. To compare the directed orders we suggest using the L_1-norm distance of the rank vectors divided by the number of graves, i.e.

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

The value of this distance measure can be interpreted as the number of positions which each grave has to move on average to change one directed order into the other. However, in our stochastic model each rank vector is only a representative for an undirected order or—to put it another way—for a pair of directed orders. Hence for two such rank vectors R and $\tilde{R}$, representing the pairs $(\vec{R},\vec{R}')$ and $(\vec{\tilde{R}},\vec{\tilde{R}}')$ of rank vectors, we consider the minimal possible L_1-norm, i.e.

$$d(R,\tilde{R}) := \min(d^*(\vec{R},\vec{\tilde{R}}), d^*(\vec{R},\vec{\tilde{R}}')).$$

Alternative distance measures could be based on antitone functions of the absolute value for coefficients measuring the degree of association between two ordinal variables; like

Spearman's ϱ or Kendall's τ. Our choice is related to Spearman's footrule (Spearman, 1906).

3.2. Checking the existence of multiple modes

Because the posterior probability of a single order is typically too small to be approximated in a reasonable manner, we have to base the analysis of multi-modality on smoothed versions of the posterior density. This allows us to analyse multi-modality in a similar manner as kernel density estimates can be used for this task for continuous variables (Silverman, 1981). Therefore, we consider (for a fixed bandwidth b and rank vector R^i of the sample) the posterior probability of the neighbourhood of R^i with radius b approximated by $\hat{P}(b,R^i) := \#\{j \mid d(R^i,R^j) \leqslant b\}/N$. This is—up to a normalizing factor only depending on b—a kernel density estimate. If for all R^j in this neighbourhood $\hat{P}(b,R^j)$ is less than or equal to $\hat{P}(b,R^i)$, we call R^i a b-mode. Finally, we can assign to each mode, R^i, a rough estimate for the posterior mass of its neighbourhood by determining the fraction of the rank vectors within the sample which are closest to R^i in comparison with the other modes. To find all modes, we have to perform this step for several bandwidths and to select the smallest b, for which we find several b-modes with substantial posterior probability.

3.3. Artificial directing

In further analysis of the posterior distribution, we are always faced with the problem that terms like the 'position' of a grave cannot be used due to the ambiguity of undirected orders. In general, an artificial definition of a direction is possible by choosing a reference order $\vec{R}^0$ and selecting from each undirected order represented by the pair $(\vec{R},\vec{R}')$ either $\vec{R}$, if $d^*(\vec{R},\vec{R}^0) < d^*(\vec{R}',\vec{R}^0)$, or $\vec{R}'$ if $d^*(\vec{R},\vec{R}^0) > d^*(\vec{R}',\vec{R}^0)$. Such an artificial directing can be misleading. However, it seems to be allowed if there is only negligible posterior probability on orders with an ambiguity in the directing. The latter can easily be checked by approximating the posterior probabilities $A(\gamma) := P_{\text{post}}(|d^*(\vec{R},\vec{R}^0) - d^*(\vec{R}',\vec{R}^0)| \leqslant \gamma)$, where $(\vec{R},\vec{R}')$ denote the two possible rank vectors representing the grave order. If $\hat{A}(\gamma)$ is away from 0 for small values of γ, an artificial directing is prohibited.

Candidates for R^0 are the point estimates discussed in the next section. The selection of one direction $\vec{R}^0$ is usually no problem, as in most archaeological applications there is some external knowledge about what is early and what is late. In the absence of such knowledge the decision can be made arbitrarily without hampering interpretation.

3.4. Point estimation of a grave order

The b-modes (defined in Section 3.2) provide first point estimates in the sense of posterior modes. Alternatively we can define a Bayes estimate, minimizing the expectation of a given loss function. One proposal for such a loss function is the distance d, such that the Bayes estimate R^{B} is defined by minimizing $E_{\text{post}}[d(R^{\mathrm{B}},R)]$. To compute this Bayes estimate approximately, we try to minimize the function $f(R) := \sum_{s=1}^{N} d(R,R^s)$

by performing permutations within quadruples of neighbouring graves, until no such permutation allows further improvement. As a start value we use the R^s with minimal $f(R^s)$ among $R^1, \dots, R^N$.

Once we have performed an artificial directing, we can also compute estimates for the position of each grave i, for example the posterior means $1/N \sum_{s=1}^{N} R_i^s$, and define a point estimate of the grave order by the order of the posterior means.

3.5. Summarizing the variability of the grave order

The analysis of multi-modality provides first measures for the variability by reporting those values of b for which the b-neighbourhood(s) of the mode(s) cover a given percentage of the posterior distribution. A further global measure is the expected loss of the Bayes estimate R^{B} approximated by $1/N \sum_{s=1}^{N} d(R^{\mathrm{B}}, R^s)$.

The major archaeological interest is of course in the position of a single grave or in the relative positioning of several graves. Without artificial directing this task is difficult to fulfill, the most simple analysis being to report for each triple of graves the posterior probability to find the first grave between the two others. But it is hard to report these numbers in a readable tabular or graphical summary. After directing, we can use a variety of methods to analyse the posterior distribution of the position, i.e. the rank of a grave. From our experience the most useful tool in the case of a uni-modal posterior distribution is a simultaneous graph of boxplots summarizing the posterior distribution for each grave. An example can be found in Halekoh and Vach (1999). Additionally, highest probability density intervals should be computed, i.e. the smallest set of possible positions of a grave with a prespecified posterior probability, as this may allow us to detect bi-modality. Besides the variability of the position of a grave, it is also helpful to know which graves can be regarded as earlier than others. This can be analysed by the posterior probabilities $P_{\mathrm{post}}(\vec{R}_i < \vec{R}_j)$, which we approximate by $1/N\#\{s|\vec{R}_i^s < \vec{R}_j^s\}$. Representing these probabilities on a grey scale in a triangular matrix results in a further simple tool to aid in interpretation.

3.6. Point estimation of the order of types

In archaeology there is not only an interest in the order of the graves but also in an order of the types, as one wants to use types for classifying further graves with respect to their chronological position. There are different possibilities for defining the order of the types as a function of the parameters in our model. In particular, after artificial directing, we can use the order defined by the values $(a_j)_{j=1,\dots,J}, (b_j)_{j=1,\dots,J}$ or $(c_j)_{j=1,\dots,J}$. However, all these parameters have the drawback of being highly variable which makes analysis of the posterior distribution cumbersome. More feasible is an analysis based on the 'centres of gravity' defined by $z_j := (\sum_{r=a_j}^{c_j} r p_{r,j}) / \sum_{r=a_j}^{c_j} p_{r,j}$. Once we have selected one definition for the order of types, point estimates can be derived in an analogous manner to Section 3.4.

A specific situation in which we need a point estimate of the order of the types is when we wish to represent the data matrix $(Y_{i,j})_{i=1,\dots,I, j=1,\dots,J}$ with the graves ordered according to some point estimate. Here we propose to consider the posterior distribution of the order of the types given the estimated order of the graves. This can easily be approximated by rerunning the MCMC algorithm without updating R.

4. Example

In this section, we will analyse the data of the example given in the introduction, where for 59 graves the presence/absence of 70 grave good types was recorded. Kendall (1971b, 1988) reproduces the incidence matrix given in Hodson (1968, plate 123) and discusses several approaches for seriating the data.

Analysis of multi-modality (cf. Table 2, first row) gives a distinct hint of several modes. For example for $b = 4.7$ we observe six modes with $\hat{P}(b, R^0)$ of 0.157, 0.090, 0.067, 0.014, 0.008 and 0.005 and rough measures for their mass of 0.432, 0.227,

Table 2
Results of a standard analysis ($\alpha = 1$, b_j uniform, $t = 0$) and a sensitivity analysis for the example

			$ABCD$	$\bar{A}BCD$	$AB\bar{D}\bar{C}$	$\bar{A}B\bar{D}\bar{C}$	$ACBD$	$ACDB$	Other
α	b	EL	%	%	%	%	%	%	%
1.0	4.7	5.97	43.2	29.3	2.3	1.15	22.7	1.4	
0.99	4.7	6.19	43.8	17.6	14.9	8.4	13.6		1.8 $\bar{A}CBD$
0.975	4.5	6.57	29.8	21.8	29.4	10.5	4.1	4.4	
0.95	4.5	6.38	12.9	3.2	32.4	39.0	0.5	11.7	0.3 $\bar{B}\bar{A}CD$
0.925	4.5	5.19	4.4	1.0	42.4	50.7	1.3	0.2	
0.9	4.7	6.14	4.3	6.2	72.6		0.3	16.3	0.3 $BACD$
0.875	4.9	5.62	9.6	3.3	78.7		1.5	6.9	
0.85	5.3	5.20	10.7	1.3	83.5		4.1		0.4 $\bar{A}CBD$
0.825	5.4	4.90	7.5	1.0	88.1		3.5		
0.8125	5.4	5.88	31.5	1.5	66.4		0.6		
0.800	6.1	6.65	39.4		56.5		4.2		
0.750	6.5	7.07	39.9		58.3		1.9		
0.700	6.1	7.80	53.6		43.0		3.4		
0.6	7.8	8.61	36.4		63.7				
0.5	8.2	9.28	57.0		43.0				
0.4	8.8	9.60	49.3		50.7				
$t = 2$	4.8	5.51	36.0	45.4	0.8		17.8		
$t = 10$	4.4	4.94	20.8	69.4			9.8		

EL denotes the expected loss of the Bayes estimate. b is the value of the smoothing parameter chosen for the analysis of multi-modality. Each mode is characterized by the ordering of the four groups A, B, C and D and a potential inversion (denoted with a bar) within each group in comparison with the first mode of the analysis with $\alpha = 1.0$. '%' denotes the rough estimate of the posterior mass of each mode. In the sensitivity analysis with varying values of α, we fix b_j to be uniform on $[a_j, c_j]$ and $t = 0$. The two lines at the bottom of the table refer to the analysis with b_j triangular on $[a_j, b_j]$ and $\alpha = 1.0$.

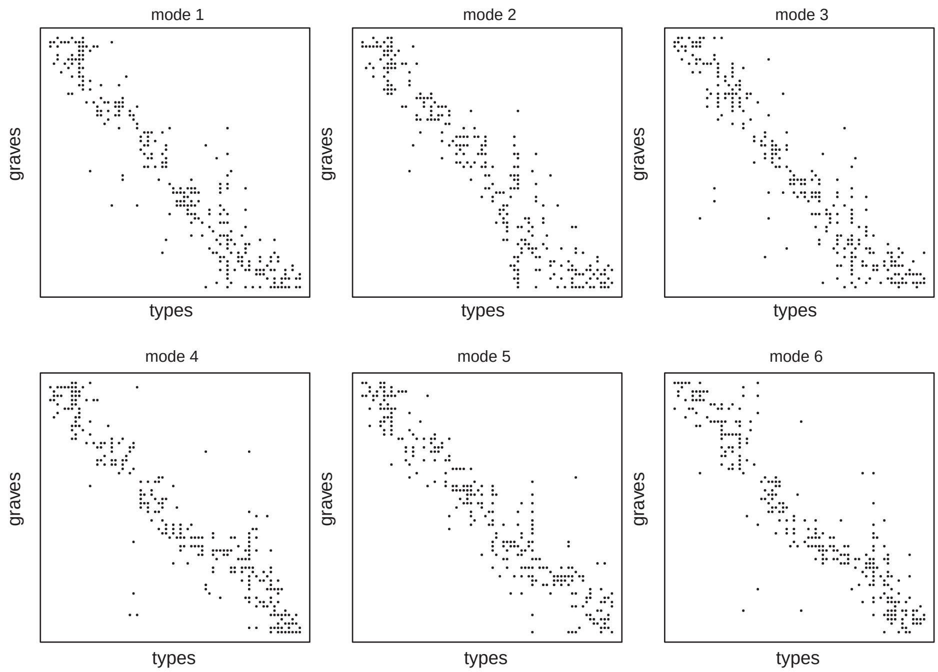

Fig. 1. The incidence matrix of the Münsingen–Rain data given for different grave orders according to the first six modes. In each subfigure the types are ordered with respect to the posterior mean of the 'centre of gravity' given the order of the graves.

0.293, 0.023, 0.014 and 0.012, respectively. Fig. 1 shows, that indeed all six solutions correspond to an incidence matrix with a similar degree of diagonalization of the data matrix. The conceptual connection between these modes can be described as a reordering of subgroups of the graves. Note that an artificial directing using the first mode as a reference order is allowed here, because $\hat{A}(\gamma)$ is greater than 0 only for $\gamma \geqslant 15.4$. We can compare the different modes in Fig. 2, where four groups A, B, C and D are distinguished which are nearly exactly separated with respect to the chronological order given by the first mode.

Comparing the first and the second mode, the position of the groups B and C is exchanged, whereas their internal structure is kept roughly the way it was. The third mode differs from the first mode by a reversal of group A. The comparison of the fourth mode with the first one reveals no change in the chronological order within the groups A and B, but C and D are exchanged, accompanied by an internal reversal. In the fifth mode, group B has moved from the second to the fourth position, and the internal ordering of C and D is changed. Finally, the sixth mode may be derived from the fourth one by reversing group A. For $b = 4.8$ the fifth mode disappears and for $b = 4.9$ to 5.3 we have only three modes where the third mode is now the previous fourth mode. For $b = 5.4$ we have only two modes and for $b \geqslant 5.5$ just one mode remains.

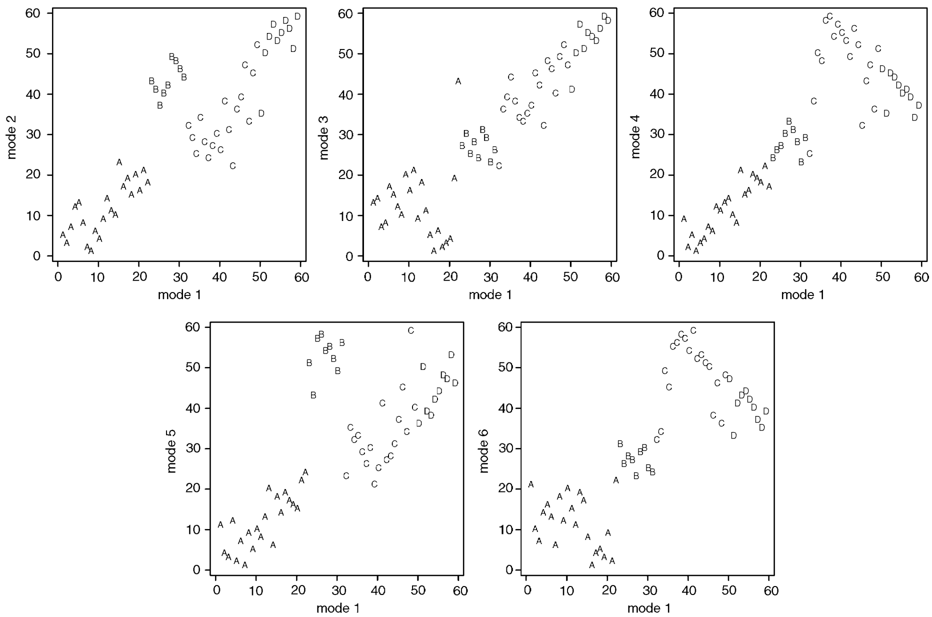

Fig. 2. Scatterplots of the rank vectors of the 2–6th mode versus the 1st mode. The choice of *A–D* is explained in the text.

It may be interesting to compare this result with that of a correspondence analysis. In the latter, the order of graves is that imposed by the scores of the eigenvector corresponding to the largest non-trivial eigenvalue. It is common practice in archaeology to check a possible ambiguity of this solution by looking at scatterplots of the eigenvectors of the largest eigenvalues, as in Fig. 3.

In the case of an unambiguous underlying gradient, these scatterplots should show parabolic or cubic curves (Greenacre, 1984, p. 227, 228), and we find these types of curve in Fig. 3. A closer look reveals a possible special role of the grave group *B* because this is separated from the others by the scores both of the first and the second eigenvector. According to the order with respect to the first eigenvector, group *B* is placed chronologically after the other three groups which have an ordering similar to the first mode of the Bayesian analysis. Consequently, the fifth mode strongly resembles the order suggested by a correspondence analysis. However, the second eigenvector suggests to mix up the graves of the groups *A*, *C* and *D*. The modes of the Bayesian analysis yield a similar degree of diagonalization of the incidence matrix as a correspondence analysis does. This indicates, that the latter fails to represent all the essential structures in the data.

The multi-modality and the striking difference between our results and those from a correspondence analysis may raise questions about the influence of our choice of the prior distribution. Hence, in a sensitivity analysis we first varied the value of α. We consistently obtained a multi-modal posterior distribution with varying number and

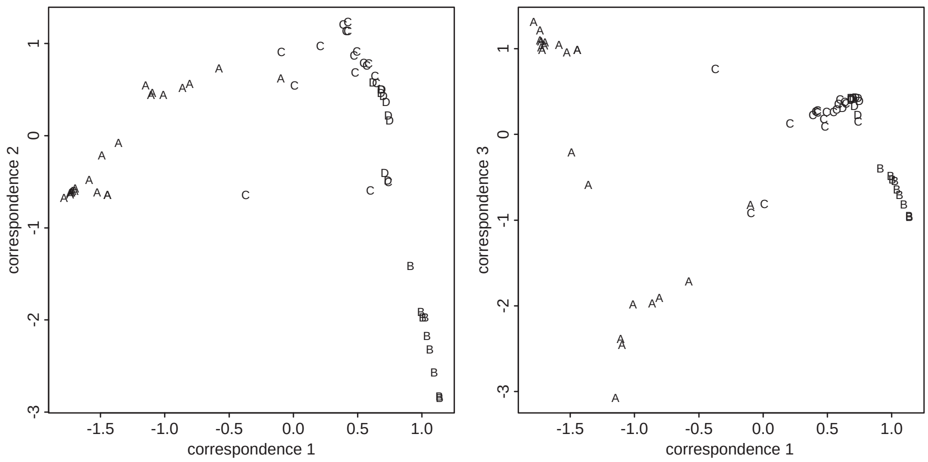

Fig. 3. Scatterplots of the scores of the eigenvectors corresponding to the three largest non-trivial eigenvalues of the correspondence analysis. The labelling *A–D* of the graves is the same as in Fig. 2.

composition of modes (Table 2), but however, the modes similarly consistent show a reordering of the four groups A, B, C and D with partial inversion within the groups. Overall, the composition of the modes is roughly a smooth function of α: from $\alpha = 1.0$ down to $\alpha = 0.925$ we observe a change from a dominance of the three modes $ABCD$, $\bar{A}BCD$ and $ACBD$ to a dominance of $AB\bar{D}\bar{C}$ and $\bar{A}B\bar{D}\bar{C}$. For $\alpha \leqslant 0.9$ we cannot discriminate these two modes further, they become one mode. Down to $\alpha = 0.825$ we have a rather stable posterior distribution with dominance of the single mode $AB\bar{D}\bar{C}$, and hence we observe here the smallest values for the expected loss. Around $\alpha = 0.8125$ we have a rapid change to a posterior distribution with high mass for both $ABCD$ and $AB\bar{D}\bar{C}$, which is stable down to $\alpha = 0.4$. Keeping $\alpha = 1.0$ fixed, but considering a triangular prior distribution of b_j and values $t > 0$, we can again observe a decrease of the overall variability, but this is now due to the dominance of a mode corresponding to the ordering $\bar{A}BCD$.

In summary, application of our approach to this data set reveals that extracting a single order of the graves using the incidence information alone is not justifiable. We can conclude, that the graves within the groups A, B, C and D are probably temporally proximal, but also that several orderings of the groups are plausible. Hodson (1968) hints to the weak link between the graves of group B (considered by him as late) and the other graves. He also criticizes a former analysis as having too rigidly aligned the seriation to the spatial neighbourhood of the graves. A final decision on a correct grouping and the correct direction within each group needs additional archaeological information.

Convergence of the approximating sequence can be a crucial problem in the case of multi-modality. It is necessary to ensure that the sequence is not stuck in one or a few modes for a long period of time. Therefore, the space of R has been divided in the same way as it had been done by determining the mass of each mode. Dividing the sequence in 10 blocks of size 1000, we can observe (for the case $\alpha = 1$), that the three modes with substantial posterior probability appear within each block, and that each of the 3 modes with small posterior probability appear in at least 5 of the 10 blocks. Moreover, the sequence shows 3698 jumps between modes and the distribution of the waiting time between jumps shows a 95% quantile of 9 with a maximum of 163. This indicates a sufficient mixing of the sequence. This was corroborated by inspecting sample paths of the rank of single graves as well as by the fact, that in our sensitivity analysis we were able to observe a rather smooth dependence of the results on α, which would be unlikely in the case of convergence problems.

5. Approximation of the posterior distribution

Within the last decade, Markov chain Monte Carlo (MCMC) techniques have become a widely used tool for approximating posterior distributions (Geman and Geman, 1984; Gelfand and Smith, 1990; Smith and Roberts, 1993; Tierney, 1994; Gilks et al. 1996). Many problems associated with the use of MCMC techniques have been addressed in the literature, and one of the most crucial ones is to achieve rapid convergence in the case of multi-modal distributions. It is not unlikely for the seriation problem—and our example illustrates this—that the posterior distribution may be multi-modal. Hence it is important to consider a technique which prevents the generated chain from getting stuck in a single mode. We have made use of the MCMCMC technique (Geyer, 1991; Geyer and Thompson, 1995). The idea underlying this approach is to generate not just one chain but K chains with different stationary distributions $(\pi_k)_{k=1,\dots,K}$, where π_1 is the posterior distribution of interest, and $\pi_2, \dots, \pi_K$ are distributions with increasing degree of flatness, such that π_K is close to a uniform distribution. In each step of this technique, each chain is updated by an appropriate MCMC technique, and additionally the values of the chains are allowed to be exchanged by a suitable Metropolis-Hastings step such that all chains together converge to the product of the distributions $\pi_1, \dots, \pi_K$. Because the distributions of $\pi_2, \dots, \pi_K$ are flatter than that of π_1, the corresponding chains are more likely to jump between their modes. Exchanging the values between the chains results also in jumps between the modes of the posterior distribution π_1 in the first chain—the only one that is analysed.

One general strategy for generating flat distributions which are still similar to the distribution of interest π_1 is to take powers of π_1 such that $\pi_k \propto \pi_1^{\lambda_k}$ with $1 = \lambda_1 > \lambda_2 > \cdots > \lambda_K \geqslant 0$. We use a slightly different idea by taking only the power of the likelihood and not of the complete posterior distribution such that $\pi_k \propto l(Y|\theta)^{\lambda_k} f(\theta)$. This facilitates the computations but still serves the main task: in the last chain K the main parameter of interest, i.e. R, is generated from a uniform distribution. With this variant, the acceptance probability to exchange the values θ_1 and θ_2 of the chains k_1 and k_2 is equal to

$$\min\left[1, \left(\frac{l(Y|\theta_2)}{l(Y|\theta_1)}\right)^{\lambda_{k_1} - \lambda_{k_2}}\right].$$

In our implementation we only allow exchange between neighbouring chains. To be more precise, after updating each single chain we sequentially propose to exchange the values between chain k and $k+1$ with k running from 1 to $K-1$.

It is necessary to decide on the total number of chains, K, and the values $\lambda_1,\dots,\lambda_K$. Our aim has been to choose these values in a manner such that for any pair of neighbouring chains the acceptance frequency of the proposal for an exchange is about 50%. In order to achieve such values an adaptive procedure has been applied in a similar manner to that of Geyer and Thompson (1995). Details are given in Appendix A. We have obtained $K=31$ with the values (in %): 100, 95.1, 90.3, 85.7, 81.1, 76.8, 72.4, 68.1, 63.8, 59.6, 55.7, 51.7, 48.1, 44.7, 41.3, 37.8, 34.2, 30.6, 27.2, 23.8, 20.5, 17.5, 14.6, 12.0, 9.6, 7.4, 5.5, 3.9, 2.4, 1.1, 0. Each analysis of Section 4 is based on a sequence of 10,000 cycles after a burn-in phase of 2500 cycles with the values given above. For all of the pairs of neighbouring chains we observed acceptance frequencies larger than 0.49, such that our aim has been achieved.

In addition to the exchange of the values between two neighbouring chains additional steps are necessary in order to change the values of the single chains. These steps and their arrangement are described in what follows. For Metropolis-Hastings steps the acceptance frequencies observed in the first chain (with $\lambda=1$) in the analysis of our example with $\alpha=1.0$ are reported. For the other chains they are usually larger due to the increasing flatness of the stationary distributions and for $\lambda=0$ they approach 1.0. Any exception is noted in the text.

5.1. Sampling of R

Here we use Metropolis-Hastings steps proposing to reverse the order of graves within intervals of the present order. For two given positions $r_1<r_2$ the proposal distribution is a Dirac distribution on the rank vector R^* with

$$R_r^{*-1}=\begin{cases} R^{-1}_{r_2+r_1-r} & \text{if } r_1\leqslant r\leqslant r_2,\\ R_r^{-1} & \text{otherwise.}\end{cases}$$

In computing the acceptance probabilities both the proposal density and the prior cancel and it remains from the likelihood

$$\tilde{l}(R_{r_1}^{-1},\dots,R_{r_2}^{-1}):=\prod_{j=1}^{J}\prod_{r=r_1}^{r_2}(g_{R_r^{-1}}p_{r,j})^{Y_{R_r^{-1},j}}(1-g_{R_r^{-1}}p_{r,j})^{1-Y_{R_r^{-1},j}}$$

and the acceptance probability is

$$\min\left(1,\left[\frac{\tilde{l}(R_{r_1}^{*-1},\dots,R_{r_2}^{*-1})}{\tilde{l}(R_{r_1}^{-1},\dots,R_{r_2}^{-1})}\right]^{\lambda}\right).$$

Naturally, the acceptance probabilities of these steps decrease with increasing length r_2-r_1+1 of the interval. In the example, we observe in preliminary investigations that acceptance frequencies are smaller than 1% for intervals larger than 10 positions. Hence we restrict application of this step to intervals with a maximal length of 10.

Additionally, we do not apply this step to all intervals in each cycle but only to a random selection of the intervals (see Section 5.4).

5.2. Sampling of g_i

To change the value of a single component g_i we draw from the full conditional distribution of g_i. As the prior of g_i is uniform on (0,1) the density of the full conditional distribution is proportional to $\tilde{l}(g_i)^{\lambda} I_{(0,1)}(g_i)$ with

$$\tilde{l}(g_i) := \prod_{j=1}^{J} g_i^{Y_{i,j}} (1 - g_i\, p_{R_i,j})^{1-Y_{i,j}}.$$

The mapping $g_i \mapsto \tilde{l}(g_i)^{\lambda}$ is a log-concave function on the unit interval and therefore we can apply adaptive rejection sampling to draw from the conditional posterior distribution (Gilks and Wild, 1992).

5.3. Sampling of ψ_j

A first straightforward idea for changing the value of ψ_j is to propose a change to the value of a single $p_{r,j}$ accompanied by a possible change of a_j, b_j and c_j. For example one can try to draw from the full conditional distribution of $(p_{r,j}, a_j, b_j, c_j)$. However, due to the order restrictions such a step changes $p_{r,j}$ only to a very small amount which implies slow mixing. We can overcome this difficulty by defining a Metropolis-Hastings step that allows a simultaneous change of many components of $(p_{r,j})_{r=1-I,\dots,2I}$. However, this step does not allow us to change a_j, b_j, c_j and $p_{b_j,j}$, which has to be done by additional Metropolis-Hastings steps. In the end, we have five steps in order to change the values of the components of ψ_j. These are described in Appendix B.

5.4. Arrangement of steps

So far we have described several steps to update components of the parameter vector θ. The final decision concerns the arrangement of these steps. As we have to expect some degree of correlation between ψ and R, and since we use a lot of single steps to change both parameter vectors, it is advisable to alternate between changing ψ and R. Following this idea, we implement the following sequence comprising one cycle: 1. Update R, 2. Update $g_1, \dots, g_I$, 3. Update $(p_{r,j})_{r<b_j}$ and $(p_{r,j})_{r>b_j}$ for all j, 4. Update a_j and c_j five times for each j, 5. Update R, 6. Update $p_{b_j,j}$ for all j, 7. Propose exchanging ψ_j between neighbouring chains for all j, 8. Update a_j and c_j five times for each j, 9. Update R, 10. Update ψ_j by shift proposals for each j, 11. Update a_j and c_j five times for each j, 12. Propose exchanging θ between neighbouring chains. The term 'Update R' summarizes two steps. In the first step we propose to exchange the values R_r^{-1} and R_{r+1}^{-1} with r running from 1 to $I-1$ and then from $I-1$ down to 1. In the second step we consider proposals to invert intervals for a random selection of all intervals of size 3 to 10 with selection probability 0.1.

6. Discussion

Using an explicit stochastic model and tools of Bayesian inference it is possible to develop a new approach for seriation problems in archaeology as presently demonstrated. The main advantage of the approach is that it becomes possible to describe the variability of the grave order implied by the available data. This adequate description may prevent over-interpretation of single point estimates and hence may lead to a more adequate interpretation of archaeological results. In our example, the description of variability leads to the detection of several clearly different orders compatible with the data. These were not visible when using correspondence analysis, the standard tool for seriation in archaeology.

Our approach relies on a choice of an underlying model and a prior distribution. We have chosen our model assuming as little as possible about the shape of the course of incidences over time, making it a flexible tool for all situations appearing in archaeological research. The main difficulty in choosing a prior distribution arises from the fact that the grave good types used in an analysis are usually the result of a long selection process. This may imply that a priori types with a long life-span are unlikely. However, it seems to be difficult to formalize or elicit this prior information. Hence, we propose to perform a sensitivity analysis to study the impact of assumptions on the prior distribution for life times and for the shape of the course over time.

Our approach can be extended in several directions relevant for archaeological research. First, we can incorporate additional, known factors with impact on the incidence. As an example, let us consider an analysis where some of the grave goods are sex specific, as for example pieces of jewellery for women and weapons for men. Then we have $P(Y_{i,j}=1)=g_i\, p_{R_i,j} k_{i,j}$ where $k_{i,j}=0$ if type j is not compatible with the sex of the individual of grave i and otherwise $k_{i,j}=1$. By more complicated extensions it is also possible to handle other factors, for example differences in the type frequencies between different spatial regions. In analysing sex-specific grave goods, a simple correspondence analysis will ideally show two factors, one separating the two sexes and one representing time; less ideally it will mix up both in one factor. The extension of our approach guarantees that only the (chronological) order remains as the unknown factor to be extracted in the analysis.

Second, there might exist additional information on the relative chronology from other sources. It is a good tradition in archaeology to make use of this additional information to validate results of a seriation, as a seriation may indicate an underlying gradient different from time. However, it may happen that all single sources are too weak to give satisfying results, and hence need to be combined. A typical example occurs in excavations, where it is often possible to see several strata in the excavated ground, such that graves found in one stratum must be earlier than graves found in a stratum above. The implied partial order of the graves identified during excavation is often illustrated by archaeologists using a graphical tool known as a ‘Harris Matrix’ (Harris, 1989). To use this prior information it is necessary to restrict the set of grave orders to those compatible with the partial order, and our MCMC algorithm can be used with minor modifications.

Another type of prior information occurs in analysing graves of larger burial sites using the spatial distribution of the graves. Here it might be known that spatial proximity partially reflects a chronological order. Hence, we can put higher prior probability on orders showing some degree of correlation with spatial proximity.

Third, in archaeology there are often counts of types, for example in the case of a seriation of a collection of excavated settlements. In such a case it would be necessary to extend our model appropriately, e.g. by choosing Poisson or zero-inflated Poisson models.

Our research seems also to be of some interest from a statistical point of view. First, a non-Bayesian analysis based on the proposed stochastic model seems to be out of reach, especially with respect to the precision of estimates. Classical likelihood theory cannot be used as the number of parameters increase with the sample size and the main parameter of interest is not numerical. Also alternatives, such as the bootstrap, are not directly applicable as we have no simple iid units. By contrast, a Bayesian approach yields a simple description of variability although a frequentistic interpretation of highest posterior density regions as confidence intervals has yet to be investigated. Secondly, we have shown, that it is possible to use MCMC techniques for problems with a lot of order restrictions on the parameters, where a simple, component-wise Gibbs sampler would suffer from bad mixing. In particular, we were able to use the order restrictions in a manner which allows us to define Metropolis-Hastings steps, changing many components simultaneously but nevertheless having high acceptance probabilities, even when the proposals are just drawn from the prior distribution. Thirdly, the use of the MCMC techniques was necessary here for other reasons than avoiding numerical integration. In our setting, computation of the posterior distribution requires—at least for $\alpha = 1.0$—only integrals which can be solved symbolically. The computation however, requires summing up over all possible orders and over all values of $(a_j, c_j)_{j=1,\dots,J}$ and this combinatorial difficulty is circumvented by using MCMC techniques. Fourthly, we have shown that the MCMCMC technique of Geyer (1991) can be successfully used in situations with multi-modal posterior distributions. We made a slight addition in that we used multiple chains not only for exchanging the complete parameter vector, but also for generating proposals for single parameters. This idea can be extended in many ways. For example we can use the value of $c_j - a_j$ in a neighbouring chain to determine how often we have to update a_j and c_j. What remains is to find methods to make use of the information contained in the additional chains, as the current way seems to be rather wasteful of information.

As a referee pointed out that there exists a slight non-identifiability in our model. Since the probabilities are modelled as a product, there exists the possibility to multiply all g's with a constant and to divide all p's by the same constant without changing the probabilities. Such a lack of identifiability poses no problem in a Bayesian analysis, as long as one considers the transformation of a prior distribution and a likelihood into a posterior distribution (e.g. Lindley, 1971; Lindley and Smith, 1972; Nobile and Green, 2000, p. 19). Within a set C of equivalent parameter values the likelihood is just constant and hence the posterior reflects within such a class only the prior. However, non-identifiability may have some impact on the analysis of the posterior distribution and on the convergence of MCMC methods. Analysing the posterior distribution it is wise to focus on functions of the parameters, which are constant within classes of equivalent parameters. This problem does not occur in our case, because we do not analyse the posterior distribution of g or p. However, we believe that in our context one can try to study for example the shape of the posterior distribution of the temporal course of a single trait. With respect to convergence of MCMC methods non-identifiability may lead to a flat posterior and hence it may be a reason for slow mixing. However, in our case slow mixing is due to multi-modality and the impact of non-identifiability is negligible.

The main drawback of our approach seems to be the high computational burden. In our example, for the 12,500 cycles we needed more than 60 h on a SPARC station 20. The choice of K and λ_k and each sensitivity analysis require additional runs. However, the additional insights possible with the new approach are worth these efforts. Implementation of the described algorithm requires some care to avoid overflow or other numerical instabilities, and the authors are glad to provide their PASCAL program on request.

Acknowledgements

The work of both authors has been supported by the Deutsche Forschungsgemeinschaft. The authors would like to thank Caitlin Buck and three referees for helpful comments.

Appendix A

Suppose we have chosen K and $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$. Running chains with these values we observe the frequencies $\hat{h}_k$ for accepting the proposal to exchange the values of chains k and $k+1$. We assume that the acceptance probability $h(\lambda_k, \lambda_{k+1})$ of the proposal to exchange the values of two neighbouring chains with arbitrary values λ_k and λ_{k+1} is approximately

$$h(\lambda_k, \lambda_{k+1}) = \exp\left(- \int_{\lambda_k}^{\lambda_{k+1}} b(s)\,\mathrm{d}s \right) \tag{A.1}$$

with an unknown function $b(s)$. Assuming further, that $b(s)$ is a step function with jumps only at $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_K$, we can achieve an estimate of $b(s)$ using (A.1) by requiring $h(\tilde{\lambda}_k, \tilde{\lambda}_{k+1}) = \hat{h}_k$, and hence we can determine new values of K and $\lambda_1, \ldots, \lambda_K$ using (A.1) again by requiring $h(\lambda_k, \lambda_{k+1}) = h_0$. We start this adaptive procedure with $K = 21$ and equally spaced values of λ_k on $[0, 1]$. We cycle twice through this procedure choosing h_0 always as the smallest value larger than 0.5 such that λ_K coincides with 0, and in each step we estimate $\hat{h}_k$ based on 2000 repetitions after a burn-in phase of 2500.

Appendix B

The first step allows us to change simultaneously all values $(p_{r,j})_{r<b_j}$ (or $(p_{r,j})_{r>b_j}$, respectively) by drawing proposals from the conditional prior distribution of these parameters, given a_j, b_j, c_j and $p_{b_j,j}$. Hence the proposal density is

$$h((p_{r,j})_{a_j \leqslant r<b_j}, p_j^{<a}) = f((p_{r,j})_{a_j \leqslant r<b_j}, p_j^{<a} | \theta \setminus (p_{r,j})_{a_j \leqslant r<b_j}, p_j^{<a})$$

$$\propto \frac{1}{p_{a_j,j}} I_{(0,\, p_{a_j,j})}(p_j^{<a}) I_{(0,\, p_{a_j+1,j})}(p_{a_j,j}) \ldots I_{(0,\, p_{b_j,j})}(p_{b_j-1,j}).$$

From this density we sample via successive conditioning, starting with drawing $p_{b_j-1,j}$ from its marginal distribution and ending with $p_{a_j,j}$ and $p_j^{<a}$. Note that the conditional densities of $p_{r,j}$ given $p_{r+1,j}, \ldots, p_{b_j-1,j}$ are proportional to $p^{r-a_j} I_{(0,\, p_{r+1,j})}(p)$ such that sampling can be based on the inverse distribution function. Because the proposal density is the conditional prior, computation of the acceptance probability requires only to evaluate

$$\tilde{l}(p_{b_j-1,j}, \ldots, p_{a_j,j}, p_j^{<a}) := \prod_{r=1}^{a_j-1} (g_{R_r^{-1}} p_j^{<a})^{Y_{R_r^{-1},j}} (1 - g_{R_r^{-1}} p_j^{<a})^{1-Y_{R_r^{-1},j}}$$

$$\times \prod_{r=a_j+1}^{b_j-1} (g_{R_r^{-1}} p_{r,j})^{Y_{R_r^{-1},j}} (1 - g_{R_r^{-1}} p_{r,j})^{1-Y_{R_r^{-1},j}}$$

at the proposed and actual parameter values, respectively, to raise the ratio to the power of λ and to set it to 1 if it exceeds 1. One may conjecture, that this type of Metropolis-Hastings step has a low acceptance probability as one would usually expect a great difference between the prior and the posterior distribution. However, the average acceptance frequency of this step is about 80% in our examples. This is due to the fact, that the information about the shape of $(p_{r,j})_{r\in\mathbb{Z}}$ in the posterior is mainly reflected by the posterior distribution of a_j, b_j, c_j and $p_{b_j,j}$, whereas, if the values of these parameters are fixed, the conditional posterior is mainly determined by the conditional prior, forcing the uni-modal shape of $(p_{r,j})_{r\in\mathbb{Z}}$. A more subtle analysis of the acceptance frequencies shows, that they depend on the size of $b_j - a_j$, increasing from 69% for $a_j = b_j$ to values of about 80–90% for $b_j - a_j \geqslant 15$. In accordance with the above argument, again this can be explained by the relative role of likelihood and prior in determining the conditional posterior distribution. If $b_j - a_j$ is small, the prior allows some variation in $(p_j^{<a})$, such that the likelihood of the observations $Y_{i,j}$ with $R_i \leqslant a_j$ can play a role, whereas for large $b_j - a_j$ the prior forces $p_j^{<a}$ to be close to zero.

The second step is a Metropolis-Hastings step in order to change $p_{b_j,j}$. The proposal is again drawn from the conditional prior distribution, therefore the proposal density is proportional to

$$h(p_{b_j,j}) \propto p_{b_j,j}^{z} I_{(q,1)}(p_{b_j,j}),$$

where $q = \max(p_{b_j-1,j}, p_{b_j+1,j})$ and

$$z = \begin{cases} t + a_j - c_j & \text{if } a_j \neq b_j \neq c_j, \\ t + a_j - c_j - 1 & \text{if } a_j = b_j \neq c_j \vee a_j \neq b_j = c_j, \\ t + a_j - c_j - 2 & \text{if } a_j = b_j = c_j. \end{cases}$$

Again, computation of the acceptance probability reduces mainly to evaluation of $(g_{R_{b_j}^{-1}} p_{b_j,j})^{Y_{R_{b_j}^{-1},j}}(1 - g_{R_{b_j}^{-1}} p_{b_j,j})^{1-Y_{R_{b_j}^{-1},j}}$ for the actual and proposed values of $p_{b_j,j}$, and the acceptance frequency in our example is greater than 98%.

The third step realizes the change of a_j (or c_j respectively). This task is rather crucial as changing a_j implies at least a change of $p_{a_j,j}$. Hence the parameters to be changed depend on the value of a parameter to be changed simultaneously. This is a situation closely related to the setting considered by Green (1995). Green pointed out, that the Metropolis-Hastings construction principle can also be used in this setting, but that there is no general technique like the Gibbs sampler to ensure an acceptance probability of 1. Our final choice to generate proposals is the following: If $1 - I < a_j < b_j$ we propose $a_j^* = a_j - 1$ with probability $\alpha/(1+\alpha)$, and $a_j^* = a_j + 1$ with probability $1/(1+\alpha)$. In the first case we then draw a value p^* from the uniform distribution on $[p_j^{<a}, p_{b_j,j}]$, determine $r' = \max\{r | a_j - 1 \leqslant r < b_j, p_{r,j} < p^*\}$, set $p_{r,j}^* := p_{r+1,j} \forall r \in \{a_j - 1, \ldots, r' - 1\}$, set $p_{r',j}^* = p^*$ and keep all other parameters fixed. In the second case we choose randomly one $r' \in \{a_j, \ldots, b_j - 1\}$, set $p_{r,j}^* = p_{r-1,j} \forall r \in \{a_j, \ldots, r'\}$ and keep all other parameters fixed. If $1 - I < a_j = b_j$, we set $a_j^* = a_j - 1$ and proceed as in the first case. If $1 - I = a_j < b_j$ we set $a_j^* = a_j + 1$ and proceed as in the second case. We propose no change if $1 - I = a_j = b_j$. The corresponding proposal density—taking the appropriate parameter restrictions into account—consists of only two factors. The first factor is $\alpha/(1 + \alpha)$ or $1/(1 + \alpha)$, respectively, determining whether we go up or down and 1 otherwise. The second factor is $1/(p_{b_j,j} - p_j^{<a}) I_{(p_j^{<a}, p_{b_j,j})}(p_{r',j}^*)$ in the first case and $1/(b_j - a_j)$ in the second case. Now let

$$\tilde{l}_{a_j}(p_j) := \prod_{r=\max(1,a_j)}^{\min(r',I)} p_{r,j}^{Y_{R_r^{-1},j}} (1 - g_{R_r^{-1}} p_{r,j})^{1-Y_{R_r^{-1},j}}$$

denote the essential part of the likelihood. If $a_j^* = a_j - 1$ then the acceptance probability is equal to

$$\alpha = \min\left(1, \left(\frac{\tilde{l}_{a_j-1}(p_j^*)}{\tilde{l}_{a_j-1}(p_j)}\right)^{\lambda} \frac{v(b_j|a_j - 1, c_j)}{v(b_j|a_j, c_j)} \frac{p_{a_j,j}}{p_{a_j-1,j}^*} \frac{p_{b_j,j} - p_j^{<a}}{p_{b_j,j}} \tau(a_j, b_j)\right),$$

where

$$\tau(a_j, b_j) = \begin{cases} 1 & \text{if } 2 - I < a_j < b_j, \\ 1 + \alpha & \text{if } 2 - I = a_j < b_j, \\ \alpha/(1 + \alpha) & \text{if } 2 - I < a_j = b_j, \\ \alpha & \text{if } 2 - I = a_j = b_j. \end{cases}$$

If $a_j^* = a_j + 1$ the acceptance probability is equal to

$$\alpha = \min\left(1, \left(\frac{\tilde{l}_{a_j}(p_j^*)}{\tilde{l}_{a_j}(p_j)}\right)^{\lambda} \frac{v(b_j|a_j+1, c_j)}{v(b_j|a_j, c_j)} \frac{p_{a_j,j}}{p^*_{a_j+1,j}} \frac{p_{b_j,j}}{p_{b_j,j} - p_j^{<a}} \frac{1}{\tau(a_j+1, b_j)}\right).$$

Note that, in the formulae for the acceptance probabilities, all ratios are close to 1, since numerator and denominator are of the same magnitude. This justifies also the asymmetry in the construction of the two cases with respect to the choice of r'. The acceptance probability of this step is about 86% in the first chain, and it is at a similar level for the other chains. It does not approach 100% for $\lambda = 0$, because we are not drawing the proposals from the prior distribution. Obviously, in order to obtain rapid mixing it is not sufficient to execute this step once in each cycle as it allows us to change a_j maximally by one position. In our example we used it 15 times within each cycle, cf. Section 5.4.

As the fourth step one might expect a similar suggestion for changing the value of b_j. However, we consider instead a proposal to shift the complete vector ψ_j. Denote with $T_r^j(\theta)$ the transformation such that $\tilde{\theta} := T_r^j(\theta)$ leaves θ unchanged except for $\tilde{\psi}_j := ((p^{(r)}_{r',j})_{r' \in \mathbb{Z}}, (a_j - r, b_j - r, c_j - r), p_j^{<a}, p_j^{>c})$ with

$$p^{(r)}_{r',j} := \begin{cases} p_{r'+r,j} & r' := a_j - r, \ldots, c_j - r, \\ p_j^{<a} & r' < a_j - r, \\ p_j^{>c} & r' > c_j - r. \end{cases} \tag{B.1}$$

Then we consider, for each type j separately, the proposals $T_r^j(\theta)$ with $r \in U_j(\theta) = \{c_j - 2I, \ldots, a_j - (1 - I)\}$ and proposal density proportional to $\pi_k(T_r^j(\theta))$. With

$$\tilde{l}_j(r) := \prod_{r'=1}^{I} [g_{R^{-1}_{r'}} p^{(r)}_{r',j}]^{Y_{R^{-1}_{r'},j}} [1 - g_{R^{-1}_{r'}} p^{(r)}_{r',j}]^{(1 - Y_{R^{-1}_{r'},j})},$$

the proposal density reduces to $\tilde{l}_j(r)^{\lambda} / \sum_{r' \in U_j(\theta)} \tilde{l}_j(r')^{\lambda}$. The acceptance probability is equal to 1. One rational for this step is based on the observation that for a type being expressed only at few graves the posterior distribution puts non-negligible mass on intervals where simultaneously $Y_{R_r^{-1},j}{=}0$ for all $r{=}a_j, \ldots, c_j$ holds. Now, if the sampling process reaches such a point, the sequence of the subsequently sampled intervals $[a_j, c_j]$ behaves like a random walk and hence we have rare but long subsequences where the above condition occurs and therefore we have no good mixing.

One may argue that all these steps are not sufficient to achieve rapid mixing because a distinct change of $b_j - a_j$, $c_j - b_j$ and $p_{b_j,j}$ determining mainly the shape of the course of $(p_{r,j})_{r \in \mathbb{Z}}$ is unlikely in one cycle. To overcome this difficulty we consider as a fifth step another Metropolis-Hastings step proposing to exchange the values of ψ_j completely between neighbouring chains. The acceptance probability of exchanging the values ψ_j^1 and ψ_j^2 of ψ_j of two chains with powers λ_1 and λ_2 is

$$\alpha = \min\left[1, \left(\frac{\tilde{l}(Y|\psi_j^2, g^1, R^1)}{\tilde{l}(Y|\psi_j^1, g^1, R^1)}\right)^{\lambda_1} \left(\frac{\tilde{l}(Y|\psi_j^1, g^2, R^2)}{\tilde{l}(Y|\psi_j^2, g^2, R^2)}\right)^{\lambda_2}\right]$$

with $\tilde{l}(\psi_j, g, R) = \prod_{i=1}^{I} (g_i p_{R_i,j})^{Y_{i,j}} (1 - g_i p_{R_i,j})^{1-Y_{i,j}}$ and g^1, R^1 and g^2, R^2 denoting the values of g and R for the two chains. In implementing this step, we have to invert the order in one chain, if this reduces the distance of the rank vectors. The average acceptance probability is 64%.

References

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

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

Buck, C.E., Litton, C.D., 1991. A computational Bayes approach to some common archaeological problems. In: Lockyear, K., Rahtz, S. (Eds.), Computer Applications and Quantitative Methods. No. 565 in BAR International Series. Archaeopress, Oxford, pp. 93–99.

Buck, C.E., Sahu, S.K., 2000. Bayesian models for relative archaeological chronology building. Appl. Statist. 49, 423–440.

Buck, C.E., Cavanagh, W.G., Litton, C.D., 1996. Bayesian Approach to Interpreting Archaeological Data. Wiley, Chichester.

Ford, J.A., 1962. A quantitative method for deriving cultural chronology. Technical Report, Pan American Union, I. Washington, DC.

Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. J. Amer. Statist. Assoc. 85, 398–409.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Trans. Pattern. Anal. Mach. Intell. 6, 721–741.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood. In: Keramidas, E.M. (Ed.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Interface, Foundation, Fairfax Station, pp. 156–163.

Geyer, C.J., Thompson, E.A., 1995. Annealing Markov chain Monte Carlo with applications to ancestral inference. J. Amer. Statist. Assoc. 90, 909–920.

Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. Appl. Statist. 41, 337–348.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.

Goldmann, K., 1972. Zwei Methoden chronologischer Gruppierung. Acta Praehist. Archaeol. 3, 1–34 (in German).

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Greenacre, M., 1984. Theory and Applications of Correspondence Analysis. Academic Press, New York.

Halekoh, U., Vach, W., 1999. Bayesian seriation as a tool in archaeology. In: Dingwall, L., Exon, S., Gaffney, V., Laflin, S., van Leusen, M. (Eds.), Archaeology in the Age of the Internet—CAA'97—Computer Applications and Quantitative Methods in Archaeology. BAR S750. Archaeopress, Oxford, 25th Anniversary Conference, University of Birmingham, April 1997.

Harris, E.C., 1989. Principles of Archaeological Stratigraphy, 2nd Edition. Academic Press, London.

Hill, M.O., 1974. Correspondence analysis: a neglected multivariate method. Appl. Statist. 23, 340–354.

Hodson, F.R., 1968. The La Téne Cemetery at Münsingen-Rain. Stämpfli, Bern.

Ihm, P., van Groenewoud, H., 1984. Correspondence analysis and Gaussian ordination. No. 3 in Compstat lectures. Physica, Wien, pp. 5–59.

Jørgensen, L., 1992. Chronological studies of Anglo-Saxon England, Lombard Italy and Vendel Period Sweden. Arkaeol. Skrifter 5.

Kendall, D.G., 1971a. Abundance matrices and seriation in archaeology. Z. Wahrscheinlichkeitstheorie 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. Edinburgh University Press, Edinburgh, pp. 215–232.

Kendall, D.G., 1988. Seriation. In: Kotz, S., Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, Vol. 8. Wiley, Chichester, pp. 417–424.

Laxton, R.R., 1976. A measure of the pre-q-ness with application to archaeology. J. Archaeol. Sci. 3, 43–54.

Lindley, D.V., 1971. Bayesian Statistics: A Review. SIAM, Philadelphia, PA.

Lindley, D.V., Smith, A.F.M., 1972. Bayes estimates for the linear model (with discussion). J. Roy. Statist. Soc. B 34, 1–42.

Madsen, T. (Ed.), 1988. Multivariate Archaeology, Vol. XXI. Jutland Archaeological Society, Aarhus, Denmark.

Mayer-Oakes, W.J., 1955. Prehistory of the Upper Ohio Valley. An introductory archaeological study. Ann. Carnegie-Mus. 34.

Nobile, A., Green, P.J., 2000. Bayesian analysis of factorial experiments by mixture modelling. Biometrika 87, 15–35.

Petrie, W.M.F., 1899. Sequences in prehistoric remains. J. Anthropol. Inst. 29, 295–301.

Robinson, W.F., 1951. A method for chronologically ordering archaeological deposits. Amer. Antiquity 16, 293–301.

Silverman, B.W., 1981. Using kernel density estimates to investigate multimodality. J. Roy. Statist. Soc. B 43, 97–99.

Smith, A.F.M., Roberts, G.O., 1993. Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J. Roy. Statist. Soc. B 55, 3–23 (with discussion p. 53–102).

Spearman, C., 1906. 'Footrule' for measuring correlation. Br. J. Physchol. 2, 89–108.

Tierney, L., 1994. Markov chains for exploring posterior distributions. Ann. Statist. 22, 1701–1762 (with discussion).