

Bayesian Hierarchical Classes Analysis

Iwin Leenen

Universidad Complutense de Madrid

University of Leuven

Iven Van Mechelen

University of Leuven

Andrew Gelman

Columbia University

Stijn De Knop

University of Leuven

Author notes:

Iwin Leenen is now at the Instituto Mexicano de Investigación de Familia y Población (IMIFAP), Mexico. The research reported in this paper was partially supported by the Spanish Ministerio de Educación y Ciencia (programa Ramón y Cajal) and by the Research Council of K.U.Leuven (PDM/99/037, GOA/2000/02, and GOA/2005/04).

The authors are grateful to Johannes Berkhof for fruitful discussions.

Authors' addresses:

Iwin Leenen, IMIFAP, Málaga Norte 25, Col. Insurgentes Mixcoac, C.P. 03920, Mexico D.F.; email: iwin@imifap.org.mx.

Iven Van Mechelen and Stijn De Knop, Department of Psychology, K.U.Leuven, Tiensestraat 102, B-3000 Leuven, Belgium; email: Iven.VanMechelen@psy.kuleuven.be.

Andrew Gelman, Department of Statistics, Columbia University, NY 10027, New York; email: gelman@stat.columbia.edu.

Running head: Bayesian HICLAS

Bayesian Hierarchical Classes Analysis

Abstract

Hierarchical classes models are models for N -way N -mode data that represent the association among the N modes and simultaneously yield, for each mode, a hierarchical classification of its elements. In this paper, we present a stochastic extension of the hierarchical classes model for two-way two-mode binary data. In line with the original model, the new probabilistic extension still represents both the association among the two modes and the hierarchical classifications. A fully Bayesian method for fitting the new model is presented and evaluated in a simulation study. Furthermore, we propose tools for model selection and model checking based on Bayes factors and posterior predictive checks. We illustrate the advantages of the new approach with applications in the domain of the psychology of choice and psychiatric diagnosis.

Hierarchical classes models (Ceulemans, Van Mechelen, & Leenen, 2003; De Boeck & Rosenberg, 1988; Leenen, Van Mechelen, De Boeck, & Rosenberg, 1999; Van Mechelen, De Boeck, & Rosenberg, 1995), dubbed HICLAS, are a family of deterministic models for N -way N -mode data. In this paper, we will focus on hierarchical classes models for two-way two-mode binary data $(y_{ij})_{m \times n}$ (i.e., data that can be represented in a 0/1 matrix), although extensions to the more general case can be considered. As examples of this type of data, one may think of person by item success/failure data, object by attribute presence/absence data, and so forth. HICLAS models for two-way two-mode data imply the representation of three types of relations in the data: (a) an association relation that links the two modes together; (b) an equivalence relation defined on each mode, yielding a two-sided classification; and (c) an if-then type hierarchical relation on the classes (and the elements) in each mode (where the term *hierarchical* is used in the general sense of a partial order, see Sneath & Sokal, 1973; Van Mechelen, Rosenberg, & De Boeck, 1997). The various members of the family of hierarchical classes models differ in the way they represent these three types of relations. The HICLAS approach has been successfully applied in various substantive domains including person perception (e.g., Cheng, 1999), personality psychology (e.g., ten Berge & de Raad, 2001), and the psychology of learning (e.g., Luyten, Lowyck, & Tuerlinckx, 2001). For a review of two-way two-mode clustering methods, and the unique position of hierarchical classes models within this domain, see Van Mechelen, Bock, and De Boeck (2004).

The HICLAS model, as a deterministic model for a binary matrix, predicts the value y_{ij} for each cell (i, j) as a single number, either 0 or 1. Strictly speaking, one should reject the model as soon as a single discrepancy (i.e., a cell value that is mispredicted) is found. In analyses of data, however, discrepancies are allowed for: Algorithms were developed (De Boeck & Rosenberg, 1988; Leenen & Van Mechelen, 2001) to find a hierarchical classes model that has a minimal number of discrepancies with a given data set. Although this combinatorial approach has yielded satisfactory results in many applications, some drawbacks are to be considered. First, the deterministic model is incomplete because the relation between the model and the data is not specified. Second, the proposed minimization algorithms yield a single solution (which in the best case is a global optimum), whereas several solutions may exist that have an optimal or near-optimal fit to the

data and that may be interesting from a substantive point of view. Third, as a consequence of the model's incompleteness, no statistical testing tools for model checking can be developed.

Maris, De Boeck, and Van Mechelen (1996) proposed a probabilistic variant of the HICLAS family, called probability matrix decomposition (PMD) models, which to some extent meet the objections above. However, in these models, the representation of the classifications and hierarchical relations is lost. The classifications and hierarchy relations in binary data have often been of important substantive interest, though; examples include concept analysis (Ganter & Wille, 1996, pp. 1–15) and knowledge space theory (Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990).

In this paper, we introduce an alternative stochastic extension of the HICLAS model that does retain the hierarchical classifications. The extension is based on a generic Bayesian methodology proposed by Gelman, Leenen, Van Mechelen, De Boeck, and Poblome (2003). As a result, many tools for model estimation and checking that are common for stochastic models become available for deterministic models as well.

In the remainder of this paper, we will first recapitulate the deterministic HICLAS model (Section 1); next, we introduce the new stochastic extension and the Bayesian estimation procedure (Section 2). Furthermore, we present the results of a simulation study to show how well the proposed procedure succeeds in recovering an underlying true HICLAS model (Section 3). For brevity's sake, we have opted to limit the exposition in the latter three sections to Van Mechelen et al.'s (1995) conjunctive HICLAS model, as the proposed extension is straightforward and fully analogous for other members of the HICLAS family. In Section 4, we will illustrate the advantages of the new approach with applications from two different domains. Finally, we present some concluding remarks and a discussion on possible further extensions of the proposed model (Section 5).

1 The deterministic hierarchical classes model

1.1 Model

Consider a binary observed data matrix $Y = (y_{ij})_{m \times n}$. In a typical deterministic hierarchical classes analysis, Y is approximated by a binary reconstructed data matrix $\hat{Y} = (\hat{y}_{ij})_{m \times n}$ for which

a hierarchical classes model (of a prespecified complexity) exists. Throughout this section, we will use the hypothetical child by item matrix in the top panel of Table 1 as an example of a reconstructed data matrix. The items are sums of fractions presented to the children, and we denote $\hat{y}_{ij} = 1$ if child i succeeds in item j and $\hat{y}_{ij} = 0$ otherwise. A child is considered to succeed in an item if (s)he returns the correct solution in its simplest form (such as in, e.g., $\frac{5}{9} + \frac{7}{9} = 1\frac{1}{3}$). We now formally define the three types of relations in the reconstructed data matrix \hat{Y} that are represented by a hierarchical classes model.

The *association relation* is the binary relation between the row elements and column elements as defined by the 1-entries in \hat{Y} . From the data matrix in Table 1, for example, we can read that (John, $\frac{5}{9} + \frac{7}{9}$) is an element of the relation between children and items, while (John, $\frac{2}{3} + \frac{1}{2}$) is not.

Two *equivalence relations* are defined, one on the row elements and one on the column elements of \hat{Y} . Two row [resp. column] elements i and i' [resp. j and j'] are equivalent (denoted as $i \sim_{\text{Row}} i'$ [resp. $j \sim_{\text{Col}} j'$]) iff they are associated with the same column [resp. row] elements. For example, John and Dave are equivalent because the sets of items they succeed in are identical. Likewise, ($\frac{6}{7} + \frac{4}{7}$) and ($\frac{3}{5} + \frac{4}{5}$) are equivalent: They are solved by the same persons. The equivalence relations induce a partitioning of the row and column elements into a number of partition classes. For example, $\{\frac{2}{3} + \frac{1}{2}, \frac{3}{4} + \frac{5}{6}, \frac{2}{5} + \frac{3}{4}\}$ constitute an item class for the data matrix in Table 1. As we will explain below, the maximal number of row [resp. column] classes is constrained (i.e., this number is substantially less than the 2^n [resp. 2^m] possible row [resp. column] patterns in an $m \times n$ binary matrix) due to \hat{Y} being a reconstructed model matrix.

The *hierarchical relations* are if-then type relations defined on the rows and on the columns. A row i [resp. column j] is considered hierarchically below row i' [resp. column j'] (denoted as $i \prec_{\text{Row}} i'$ [resp. $j \prec_{\text{Col}} j'$]) iff the set of column [resp. row] elements that i [resp. j] is associated with constitutes a proper subset of the set of column [resp. row] elements that i' [resp. j'] is associated with. For example, Lindsay is hierarchically below Dave as Lindsay succeeds in only a subset of the items that Dave succeeds in; or, alternatively, *if* Lindsay succeeds in an item, *then* Dave succeeds in this item as well, with as a possible substantive interpretation that Dave has reached a higher stage of mastery (or knowledge state) than Lindsay. With respect to the

columns, $(\frac{5}{9} + \frac{7}{9})$ is hierarchically below $(\frac{3}{5} + \frac{4}{5})$ because *if* a child solves $(\frac{5}{9} + \frac{7}{9})$ *then* (s)he solves $(\frac{3}{5} + \frac{4}{5})$ as well. The latter implication relation may be interpreted as a prerequisite relation on the items: Solving $(\frac{3}{5} + \frac{4}{5})$ is prerequisite for solving $(\frac{5}{9} + \frac{7}{9})$. Obviously, the hierarchical relation on the row [resp. column] elements directly implies a hierarchical relation on the row [resp. column] classes. The equivalence and hierarchy relations are called *set-theoretical relations* and we denote $i \preceq_{\text{Row}} i'$ [resp. $j \preceq_{\text{Col}} j'$] if either $i \sim_{\text{Row}} i'$ [resp. $j \sim_{\text{Col}} j'$] or $i \prec_{\text{Row}} i'$ [resp. $j \prec_{\text{Col}} j'$].

A hierarchical classes model implies a decomposition of \widehat{Y} into two binary matrices, one for each mode, which represent the three above-mentioned types of relations. Those matrices are called bundle matrices and are denoted by $S = (s_{ik})_{m \times r}$ and $P = (p_{jk})_{n \times r}$ for the rows and the columns, respectively, with entries $s_{ik}, p_{jk} \in \{0, 1\}$; the integer r is called the rank of the model and the r columns of S and P are called bundles.¹ For a given row i [resp. column j], the set of bundles with $s_{ik} = 1$ [resp. $p_{jk} = 1$] will be referred to as the bundle pattern of i [resp. j].

We now explain how the bundle matrices represent the association, equivalence, and hierarchy relations in Van Mechelen et al.'s (1995) conjunctive HICLAS model. The model correctly represents the *association relation* defined on \widehat{Y} iff the bundle matrices S and P satisfy, for any i and j ,

$$\hat{y}_{ij} = \begin{cases} 1 & \text{if } \forall k(1 \leq k \leq r) : s_{ik} \geq p_{jk} \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Equation (1) means that a row i is associated with a column j iff the bundle pattern of column j is a (proper or improper) subset of the bundle pattern of row i . Such a conjunctive rule (formally expressed by the quantifier \forall) may be a plausible formalization of a psychological theory where each column element imposes a set of requirements that may or may not be met by a row element. In that case, the bundles formally correspond with requirements and (1) means that a row is associated with a column iff the row meets all the requirements posed by the column. Applied to our hypothetical example, items may require one or more abilities for being successfully solved, while children either have or do not have these abilities. In particular, the example was constructed with three underlying abilities in mind: (I) finding the lowest common multiple of two integers, (II) finding the greatest common divisor of two integers, and (III) dealing with fractions where the numerator exceeds the denominator. The bottom panel of Table 1 shows the bundle matrices

S and P of a conjunctive model of rank 3 for the hypothetical data. We see, for example, that John masters abilities II and III; therefore, he is able to solve item $(\frac{6}{7} + \frac{4}{7})$ (because this item only requires ability III), whereas he does not succeed in item $(\frac{1}{4} + \frac{2}{3})$ (because it requires ability I).

The conjunctive HICLAS model correctly represents the *relations of equivalence and hierarchy* iff the following restrictions hold on S and P :

$$\forall i, i' : i \preceq_{\text{Row}} i' \text{ iff } \forall k(1 \leq k \leq r) : s_{ik} \leq s_{i'k} \quad (2a)$$

$$\forall j, j' : j \preceq_{\text{Col}} j' \text{ iff } \forall k(1 \leq k \leq r) : p_{jk} \geq p_{j'k}. \quad (2b)$$

A pair of bundle matrices (S, P) for which (2a) and (2b) hold are called *set-theoretically consistent*. Equation (2a) [resp. (2b)] implies that equivalent rows [resp. columns] have identical bundle patterns. Equation (2a) further implies that, if row i is hierarchically below row i' , then the bundle pattern of i is a proper subset of the bundle pattern of i' . Mary's abilities $(\{\text{I, II, III}\})$, for example, include as a proper subset those of John and Dave $(\{\text{II, III}\})$, who are hierarchically below Mary. On the other hand, (2b) implies that if column j is hierarchically below column j' then the bundle pattern of j is a proper *superset* of the bundle pattern of j' . The items in class $\{\frac{2}{3} + \frac{1}{2}, \frac{3}{4} + \frac{5}{6}, \frac{2}{5} + \frac{3}{4}\}$, for example, which is hierarchically below $\{\frac{6}{7} + \frac{4}{7}, \frac{3}{5} + \frac{4}{5}\}$, require a proper superset of the abilities required by the latter items $(\{\text{I, III}\} \supset \{\text{III}\})$. Within an ability context, the inverse representation of the item hierarchy naturally follows from the fact that the more abilities are required by an item, the fewer children are able to solve it.

1.2 Graphical representation

Van Mechelen et al. (1995) proposed a graphical representation that gives a full account of the relations represented by the conjunctive HICLAS model. A graphical representation of the model in Table 1 is found in Figure 1. Child and item *classes* appear as paired boxes, the upper box of each pair being a child class and the lower box an item class. The lines connecting the classes represent the *hierarchical relations*, with the item hierarchy to be read upside down. The *association relation* can be read from the graph as follows: A child succeeds in a item iff the child and item class form a paired box or a downward path exists from the child to the item class.

1.3 Data analysis

Although a HICLAS model can be found for any binary matrix Y , an exact decomposition of an observed data matrix almost always requires a model of high rank. Because models of high rank are very complex to interpret, one typically searches for an approximate reconstructed data matrix \hat{Y} that can be represented by a model of a low, prespecified rank r and for which the loss function

$$D(\hat{Y}, Y) = \sum_{i=1}^m \sum_{j=1}^n (\hat{y}_{ij} - y_{ij})^2 \quad (3)$$

has minimal value (De Boeck & Rosenberg, 1988; Leenen & Van Mechelen, 2001). Note that, since Y and \hat{Y} only contain 0/1 values, the least squares loss function in (3) comes down to a least absolute deviations loss function and equals the number of discrepancies between Y and \hat{Y} .

1.4 Model checking and drawing inferences

Although yielding quite satisfactory results in general, the widespread use of finding a best-fitting HICLAS model \hat{Y} for given data Y is conceptually unclear. For, in an approximate HICLAS model the relation to the data is not specified (i.e., it is not specified how and why the model's predictions may differ from the observed data), and hence, the slightest deviation between model and data strictly speaking implies a rejection of the model. Furthermore, because the model either fully holds or is rejected as a whole, it does not make sense to develop checks for specific model assumptions.

Moreover, in practical applications, the basis for pre-specifying the rank r of the approximate HICLAS model is often unclear. A number of heuristics have been developed to infer a rank from the data, most often based on comparing the goodness of fit of models of successive ranks (such as, e.g., the scree test and its variants). Although these heuristics may yield satisfactory results, in some cases the selection of a particular rank seems arbitrary and susceptible to chance.

2 Stochastic extension within a Bayesian framework

2.1 Model

Traditional HICLAS analyses, with their search for an approximate solution that is minimally discrepant to a given data set, are often implicitly justified by assuming that an underlying stochastic

process may cause a cell's observed value to be different from the value predicted by the model. This implicit assumption can be formalized by introducing a parameter π which is defined as:

$$\forall i, j : \Pr [Y_{ij} = \hat{y}_{ij}(S, P) \mid S, P, \pi] = 1 - \pi, \quad (4)$$

where $\hat{y}_{ij}(S, P)$ is the general notation for the predicted value in cell (i, j) obtained by combining S and P by (1). Clearly, π reflects the expected proportion of discrepancies in the model.

As an alternative, a slightly more general model may be considered, which allows predicted values 0 and 1 to have different error probabilities:

$$\forall i, j : \Pr [Y_{ij} = \hat{y}_{ij}(S, P) \mid S, P, \pi_0, \pi_1] = \begin{cases} 1 - \pi_0 & \text{if } \hat{y}_{ij}(S, P) = 0 \\ 1 - \pi_1 & \text{if } \hat{y}_{ij}(S, P) = 1 \end{cases} \quad (5)$$

In both (4) and (5), it is additionally assumed that the Y_{ij} are locally independent. The two models will be referred to as the one-error and the two-error probability model, respectively.

As in the original HICLAS model, we will add the requirement that the pair of bundle matrices (S, P) correctly represents the relations of equivalence and hierarchy in $\hat{Y}(S, P)$ (i.e., these matrices must satisfy (2a) and (2b)). As such, the newly presented model preserves the hierarchical classification of the deterministic model. In this respect, it differs from Maris et al.'s (1996) PMD model, which is another probabilistic variant of the original HICLAS model. In a PMD model, bundle membership is probabilistic and the value predicted by the model (i.e., the analogue of $\hat{y}_{ij}(S, P)$) is a probability as well; the definitions of the set-theoretical relations \preceq_{Row} and \preceq_{Col} no longer apply and the representations of classification and hierarchy are lost.

A second difference between the new model and PMD is the place where the random process enters the model. On the one hand, in a PMD model, each entry in the bundle matrices pertains to a Bernoulli random variable, the realizations of which are (deterministically) combined by an association rule like, for example, (1), to give rise to a binary observed value y_{ij} . On the other hand, the stochastic extension in (4) [resp. (5)] implies the entries in the bundle matrices to be fixed; they are combined to a binary predicted value \hat{y}_{ij} , which is randomly altered (from 0 to 1 or vice versa) with a probability π [resp. π_0 or π_1] to give rise to the observed y_{ij} . This is somewhat similar to ordinary linear regression, where it is assumed that a random perturbation acts on the value predicted by the model.

2.2 Parameter estimation

The likelihood of the data Y under the one-error probability model can readily be shown to be:

$$p(Y|S, P, \pi) = \pi^{D(\hat{Y}(S, P), Y)} (1 - \pi)^{C(\hat{Y}(S, P), Y)} \quad (6)$$

with $D(\hat{Y}(S, P), Y)$ denoting the number of discrepancies (as defined in (3)) and $C(\hat{Y}(S, P), Y)$ the number of concordances [i.e., $mn - D(\hat{Y}(S, P), Y)$]. Equation (6) may be used to obtain maximum likelihood estimates for (S, P, π) . Incidentally, $(\hat{S}, \hat{P}, \hat{\pi})$ is a maximum likelihood estimate iff $D(\hat{Y}(\hat{S}, \hat{P}), Y)$ has minimal value and $\hat{\pi} = D(\hat{Y}(\hat{S}, \hat{P}), Y)/mn$.² However, for at least two reasons, one may wish to go beyond finding maximum likelihood estimates for (S, P, π) . Firstly, apart from a trivial kind of nonidentifiability due to a joint permutability of the columns of S and P , in some cases different pairs of bundle matrices may combine to the same \hat{Y} . If the latter is the case, the model is called *nonunique*. Secondly, different matrices $\hat{Y}(S, P)$ may exist with (almost) identical minimal values on the loss-function (3) and, hence, with (almost) maximal likelihood values. Therefore, we propose to adopt a Bayesian perspective (for an introduction, see, e.g., Gelman, Carlin, Stern, & Rubin, 2004) where the parameters are considered random variables and where we inspect the posterior distribution of the parameters given the data:

$$p(S, P, \pi|Y) \propto p(Y|S, P, \pi) p(S, P, \pi).$$

Throughout this paper, we will assume a uniform prior on (S, P, π) , which represents a minimal expansion of the existing deterministic model. Other priors can be considered as well (based on, e.g., additional information or in the case of restricted HICLAS models, see also Section 5).

For the two-error probability model, the likelihood of the data Y is given by:

$$p(Y|S, P, \pi_0, \pi_1) = \pi_0^{n_{10}(Y, \hat{Y}(S, P))} (1 - \pi_0)^{n_{00}(Y, \hat{Y}(S, P))} \pi_1^{n_{01}(Y, \hat{Y}(S, P))} (1 - \pi_1)^{n_{11}(Y, \hat{Y}(S, P))}, \quad (7)$$

where $n_{hh'}(Y, \hat{Y}(S, P)) = \#\{(i, j) \mid y_{ij} = h \text{ and } \hat{y}_{ij}(S, P) = h'\}$ ($h, h' \in \{0, 1\}$). Again, the posterior distribution $p(S, P, \pi_0, \pi_1|Y)$ will be considered and a uniform prior will be assumed.

2.3 Computation of the posterior distribution

In this section, we describe a Metropolis algorithm that was developed along the lines set out by Gelman et al. (2003) and that can be used to simulate the posterior distribution for either

the one-error or the two-error probability model. We will immediately turn to the algorithm for obtaining a simulated posterior distribution for the parameter vector (S, P, π_0, π_1) in the two-error probability case, the algorithm for the one-error case involving some obvious simplifications only.

We run independently z (≥ 2) parallel sequences (or “chains”), in each of them proceeding with the following steps:

Step 0: Initial estimates $S^{(0)}$ and $P^{(0)}$ are obtained, $s_{ik}^{(0)}$ and $p_{jk}^{(0)}$ being realizations of *iid* Bernoulli variables with probability parameters such that $\Pr[\hat{y}_{ij}(S^{(0)}, P^{(0)}) = 1]$ equals the proportion of 1-entries in the data matrix Y .³ Subsequently, $\pi_0^{(0)}$ and $\pi_1^{(0)}$ are initialized:

$$\begin{aligned}\pi_0^{(0)} &\leftarrow \frac{n_{10}(Y, \hat{Y}(S^{(0)}, P^{(0)})) + 1}{n_{\cdot 0}(Y, \hat{Y}(S^{(0)}, P^{(0)})) + 2} \\ \pi_1^{(0)} &\leftarrow \frac{n_{01}(Y, \hat{Y}(S^{(0)}, P^{(0)})) + 1}{n_{\cdot 1}(Y, \hat{Y}(S^{(0)}, P^{(0)})) + 2},\end{aligned}$$

where $n_{\cdot h}(Y, \hat{Y}(S^{(0)}, P^{(0)})) = n_{0h}(Y, \hat{Y}(S^{(0)}, P^{(0)})) + n_{1h}(Y, \hat{Y}(S^{(0)}, P^{(0)}))$ ($h \in \{0, 1\}$). As a final step in the initialization, we set $t \leftarrow 0$.

Step 1: A candidate pair of bundle matrices (S^*, P^*) is constructed in four steps as follows: (a) We initialize $(S^*, P^*) \leftarrow (S^{(t)}, P^{(t)})$; (b) a strictly positive integer w is drawn from a $\text{Poisson}(\lambda)$ distribution (computational tests suggest that values of λ between 1 and 5 yield a reasonable performance); (c) a set of w cells is randomly selected from matrices (S^*, P^*) with all $(m+n)r$ cells having equal selection probability and (d) the value in the selected cells is switched (either from 0 to 1 or from 1 to 0). This procedure implicitly defines a jumping distribution $J(S^*, P^* | S, P)$, which gives the probability that the candidate (S^*, P^*) is obtained from (S, P) (Hastings, 1970; Gilks, Richardson, & Spiegelhalter, 1996, p. 5–12). It follows that

$$J(S^*, P^* | S, P) = \frac{P(W = w)}{\binom{(m+n)r}{w}}, \quad (8)$$

with w being the number of cells where (S^*, P^*) differs from (S, P) , and W having a truncated Poisson distribution with parameter λ (truncated so as to exclude 0).

Step 2: Next, we compute the probability of acceptance

$$\alpha \left(S^*, P^* | Y, S^{(t)}, P^{(t)}, \pi_0^{(t)}, \pi_1^{(t)} \right) = \min \left(1, \frac{p(S^*, P^*, \pi_0^{(t)}, \pi_1^{(t)} | Y)}{p(S^{(t)}, P^{(t)}, \pi_0^{(t)}, \pi_1^{(t)} | Y)} \right) \quad (9)$$

and replace $(S^{(t)}, P^{(t)})$ with (S^*, P^*) with probability $\alpha \left(S^*, P^* | Y, S^{(t)}, P^{(t)}, \pi_0^{(t)}, \pi_1^{(t)} \right)$.

Step 3: $\pi_0^{(t)}$ and $\pi_1^{(t)}$ are updated by drawing from a Beta $\left[n_{10}(Y, \widehat{Y}(S^{(t)}, P^{(t)})) + 1, n_{00}(Y, \widehat{Y}(S^{(t)}, P^{(t)})) + 1 \right]$ and a Beta $\left[n_{01}(Y, \widehat{Y}(S^{(t)}, P^{(t)})) + 1, n_{11}(Y, \widehat{Y}(S^{(t)}, P^{(t)})) + 1 \right]$ distribution, respectively.

Step 4: If the new pair of bundle matrices $(S^{(t)}, P^{(t)})$ is set-theoretically consistent, then we set

$$(a) \left(S^{(t+1)}, P^{(t+1)}, \pi_0^{(t+1)}, \pi_1^{(t+1)} \right) \leftarrow \left(S^{(t)}, P^{(t)}, \pi_0^{(t)}, \pi_1^{(t)} \right) \text{ and, subsequently, (b) } t \leftarrow t + 1.$$

Steps 1 through 4 are repeated independently in each of the z sequences. Each iteration in each sequence yields a draw (S, P, π_0, π_1) from the posterior distribution. The process is stopped when for each individual model parameter the sequences appear to have converged. To monitor convergence, we use Gelman and Rubin's (1992) \hat{R} -statistic, which for a given (scalar) parameter compares the between- and within-sequence variances of that parameter. \hat{R} -values near 1 for all parameters are considered an indication that the z sequences have converged. To account for possible absence of within-sequence variance in a binary bundle parameter, we additionally define $\hat{R} = 1$ for a parameter that assumes the same value throughout all sequences and $\hat{R} = +\infty$ for a parameter that in one sequence has a constant value being different from those in other sequences.

After convergence, a subset of L draws of the parameter vector is selected from the converged part⁴ of the z simulated sequences (e.g., by selecting each hundredth or thousandth draw) and forms the simulated posterior distribution of (S, P, π_0, π_1) . The choice of L basically depends on the desired precision of posterior inferences. With respect to z , we recommend to run the Metropolis algorithm with at least four sequences (rather than the minimal number of two) to enhance the chance that all important areas of the target distribution are covered.

To account for permutational freedom —when models from different chains are identical upon a permutation of the bundles, the \hat{R} -criterion will indicate nonconvergence—, each model is permuted such that the difference with some prespecified reference model (\tilde{S}, \tilde{P}) (e.g., the model obtained from the original deterministic algorithm) is minimal. That is, after each iteration t a permutation of the r bundles in $(S^{(t)}, P^{(t)})$ is looked for such that

$$\sum_{k=1}^r \left[\sum_{i=1}^m (s_{ik}^{(t)} - \tilde{s}_{ik})^2 + \sum_{j=1}^n (p_{jk}^{(t)} - \tilde{p}_{jk})^2 \right] \quad (10)$$

has minimal value. Note that posterior inferences with respect to the association, equivalence, and hierarchy relations remain unaffected by the choice of (\tilde{S}, \tilde{P}) .

2.4 Representation of the posterior distribution

In contrast with the deterministic model, where for a particular pair of elements (or classes) the association, equivalence, and hierarchy relations either hold or do not hold, the new stochastic extension implies a (marginal) posterior probability instead. In this section, some methods are proposed to graphically represent the information in the simulated posterior distribution.

To chart out the uncertainty in the association relation, one may estimate, for each combination of a row i and a column j , the marginal posterior probability of row i being associated with column j ; that is, one may calculate the proportion of draws —within the set of L selected simulation draws— for which $\hat{y}_{ij}(S, P) = 1$. The obtained probability estimate can then be converted to a grey value and cell (i, j) of an $m \times n$ grid can be colored accordingly. Repeating this process for all cells (i, j) , a graphical representation is obtained that highlights regions of low, medium, and high posterior probability in the reconstructed data matrix.

With respect to the posterior information about the classification and hierarchical relations, we propose an extension of Van Mechelen et al.'s (1995) original representation of the deterministic model. As before, each class is represented as a box. Considering that now all elements have some posterior probability of belonging to a particular class, only elements with a posterior probability exceeding a prespecified cut-off are displayed in the box representing that class. The uncertainty in the hierarchical relation among classes can further be represented in the graph by varying the line thickness or by labeling the edges. To quantify the degree of certainty of the hierarchical relation $C_1 \prec C_2$ between any pair of row classes (C_1, C_2) , we propose the following measure:

$$p(C_1 \prec_{\text{Row}} C_2) = \frac{\sum_{i=1}^m \sum_{i'=1}^m p(i \in C_1) p(i' \in C_2) p(i \prec_{\text{Row}} i')}{\sum_{i=1}^m \sum_{i'=1}^m p(i \in C_1) p(i' \in C_2)},$$

where $p(i \in C_k)$ is the posterior probability of row i belonging to row class C_k and $p(i \prec_{\text{Row}} i')$ is the posterior probability of row i being hierarchically below row i' . The uncertainty in the hierarchical relation between two column classes can be calculated through a similar formula. We will illustrate the proposed graphical representations in the applications.

2.5 Model checking and drawing inferences

Within a Bayesian framework, a broad range of tools for model checking and model selection is available, including the Bayes factor (see Kass & Raftery, 1995) and posterior predictive checks (PPC's) (see Gelman, Meng, & Stern, 1996; Rubin, 1984). Bayes factors are excellent tools for choosing between two (or more) competing models, while PPC's are particularly useful for the examination of (aspects of) a model's fit. Following Berkhof, Van Mechelen, and Gelman (2003), we argue in favour of a combined use of Bayes factors and PPC's. In this section, we briefly discuss the basic principles of both tools and illustrate their application in the context of Bayesian HICLAS.

2.5.1 Bayes factors

Given two hypotheses H_1 and H_2 , and data Y that are assumed to have arisen under either H_1 or H_2 (e.g., H_1 being the hypothesis of a one-error probability model and H_2 of a two-error probability model), the Bayes factor is formally defined as the ratio of the posterior to the prior odds:

$$B_{12} = \frac{p(H_1|Y)/p(H_2|Y)}{p(H_1)/p(H_2)}. \quad (11)$$

In the case that both hypotheses are considered equally likely a priori, the Bayes factor comes down to the posterior odds in favour of H_1 . Applying Bayes's theorem to (11) yields

$$B_{12} = \frac{p(Y|H_1)}{p(Y|H_2)},$$

where $p(Y|H_k)$ ($k = 1, 2$) is called the *marginal likelihood* under hypothesis H_k .

In the remainder of this section, we discuss calculation of the marginal likelihood under the hypothesis H that the data have arisen under a two-error probability HICLAS model of rank r (the proposed strategy being similar for other hypotheses). In that case, by the law of total probability, the marginal likelihood can be expressed as follows:

$$p(Y|H) = \sum_{(S,P) \in \mathcal{M}_r} \int_0^1 \int_0^1 p(Y|S, P, \pi_0, \pi_1) p(S, P, \pi_0, \pi_1) d\pi_1 d\pi_0, \quad (12)$$

where \mathcal{M}_r denotes the set of all deterministic HICLAS models (S, P) of rank r (i.e., with the restriction that (S, P) is set-theoretically consistent). Due to the usually huge number of elements in \mathcal{M}_r , evaluating the sum in (12) is almost never feasible in practical applications, though.

Following Chib (1995; Chib & Jeliazkov, 2001), an estimate of the marginal likelihood can be alternatively obtained by elaborating the *basic marginal likelihood identity*, which is given by:

$$p(Y) = \frac{p(Y|\tilde{\theta}) p(\tilde{\theta})}{p(\tilde{\theta}|Y)}, \quad (13)$$

where we have suppressed the dependence on H for notational convenience. Although this identity holds for any $\tilde{\theta} = (\tilde{S}, \tilde{P}, \tilde{\pi}_0, \tilde{\pi}_1)$, for estimation efficiency, $\tilde{\theta}$ will be taken to be a point with an (anticipated) high posterior density (obtained, e.g., on the basis of a deterministic analysis). In evaluating the right-hand side of (13), the likelihood $p(Y|\tilde{\theta})$ can be directly computed by (7), while the probability $p(\tilde{\theta})$ under a uniform prior distribution is inversely related with the number of elements in \mathcal{M}_r and can be easily estimated by

$$\hat{p}(\tilde{\theta}) = \left(2^{r(m+n)} \cdot \hat{p}_{\text{STC}}\right)^{-1},$$

where \hat{p}_{STC} is the proportion of set-theoretically consistent pairs (S, P) in a random sample from the population of all $2^{r(m+n)}$ pairs of a binary $m \times r$ matrix S and a binary $n \times r$ matrix P .

An estimate of the denominator in (13) is obtained after decomposing the posterior as:

$$p(\tilde{S}, \tilde{P}, \tilde{\pi}_0, \tilde{\pi}_1|Y) = p(\tilde{S}, \tilde{P}|Y) p(\tilde{\pi}_0, \tilde{\pi}_1|Y, \tilde{S}, \tilde{P}).$$

While $p(\tilde{\pi}_0, \tilde{\pi}_1|Y, \tilde{S}, \tilde{P})$ can be computed directly (since by (7) it is the product of two independent Beta densities), $p(\tilde{S}, \tilde{P}|Y)$ needs to be estimated. Such an estimate can be obtained relying on a generic strategy proposed by Chib and Jeliazkov (2001) and which applies in the case of a posterior distribution obtained by a Metropolis algorithm.

In the applications of Section 4, Bayes factors will be applied to compare models for different ranks and to compare one-error and two-error probability models. We will adhere to reporting twice the natural logarithm of the Bayes factor, so as to put it on a comparable scale as the familiar deviance and likelihood ratio test statistics (Kass & Raftery, 1995).

2.5.2 Posterior predictive checks

The rationale of PPC's is the comparison, via some test quantity, of the observed data with data that could have been observed if the actual experiment were replicated under the model with the

same parameters (Gelman et al., 2004). The test quantity may be a statistic $T(Y)$, summarizing some aspect of the data, or, more generally, a discrepancy measure $T(Y, S, P, \pi_0, \pi_1)$ quantifying the model's deflection from the data in some respect (Meng, 1994).

A PPC procedure consists of three steps. First, for each of the draws $\theta^{(l)} = (S^{(l)}, P^{(l)}, \pi_0^{(l)}, \pi_1^{(l)})$ ($l = 1, \dots, L$) from the simulated posterior distribution, a replicated data set $Y^{\text{rep } l}$ is simulated, each cell entry $y_{ij}^{\text{rep } l}$ being an independent realization of a Bernoulli variable with probability parameter $|h - \pi_h^{(l)}|$, where $h = \hat{y}_{ij}(S^{(l)}, P^{(l)})$. Second, for each simulated draw, the value of the test quantity is calculated for both the observed data Y^{obs} (i.e., $T(Y^{\text{obs}}, \theta^{(l)})$) and the associated replicated data $Y^{\text{rep } l}$ (i.e., $T(Y^{\text{rep } l}, \theta^{(l)})$). As a final step, the values for the observed and the replicated data are compared and the proportion of simulation draws for which $T(Y^{\text{rep } l}, \theta^{(l)}) > T(Y^{\text{obs}}, \theta^{(l)})$ is considered as an estimate of the posterior predictive p -value.

Different aspects of the model can be checked using PPC's with appropriate choices of the test quantity. By way of illustration, we now propose two novel test quantities that are tailor-made for the Bayesian HICLAS model: one to check the rank of the model and an other to check the conformity of the set-theoretical relations in the model with those in the data.

Rank selection For rank selection, we define the test quantity:

$$T(Y) = D(\tilde{Y}^r, Y) - D(\tilde{Y}^{r+1}, Y), \quad (14)$$

where \tilde{Y}^k ($k = r, r+1$) is the deterministic model matrix obtained by applying the original HICLAS algorithm in rank k to the data set Y . The statistic $T(Y)$ quantifies the decrease in number of discrepancies that is obtained if a model with an extra bundle is fitted to the data. For the replicated data, which are constructed under the rank r model, this extra bundle only fits noise. Therefore, a PPC with $T(Y)$ checks whether the increase of the model's fit to the observed data due to adding an extra bundle is significantly larger than expected by chance.

Set-theoretical relations As a check of the set-theoretical relations represented by the model, we will compare, for each pair of rows i and i' [resp. columns j and j'], the presence/absence of a set-theoretical relation $i \preceq_{\text{Row}} i'$ [resp. $j \preceq_{\text{Col}} j'$] in the model with their implicational strength in

the data. As a measure for the latter, we will use the conditional proportions

$$\tilde{p}(\mathbf{y}_{i' \cdot} | \mathbf{y}_{i \cdot}) = \frac{\sum_j y_{ij} y_{i'j}}{\sum_j y_{ij}} \quad \text{and} \quad \tilde{p}(\mathbf{y}_{\cdot j'} | \mathbf{y}_{\cdot j}) = \frac{\sum_i y_{ij} y_{ij'}}{\sum_i y_{ij}}$$

($\tilde{p}(\mathbf{y}_{i' \cdot} | \mathbf{y}_{i \cdot})$ [resp. $\tilde{p}(\mathbf{y}_{\cdot j'} | \mathbf{y}_{\cdot j})$] being undefined whenever $\sum_j y_{ij} = 0$ [resp. $\sum_i y_{ij} = 0$]). Then, the test quantity used in the PPC for the row set-theoretical relations is defined as:

$$T(Y, S) = \sum_{i=1}^m \sum_{i'=1}^m \left| \tilde{p}(\mathbf{y}_{i' \cdot} | \mathbf{y}_{i \cdot}) - I_S(i \preceq_{\text{Row}} i') \right|, \quad (15a)$$

where the first sum is across rows i with $\sum_j y_{ij} \neq 0$; $I_S(i \preceq_{\text{Row}} i')$ takes a value of 1 if the bundle matrix S implies $i \preceq_{\text{Row}} i'$, and 0 otherwise. A similar test quantity for the column set-theoretical relations may be defined:

$$T(Y, P) = \sum_{j=1}^n \sum_{j'=1}^n \left| \tilde{p}(\mathbf{y}_{\cdot j'} | \mathbf{y}_{\cdot j}) - I_P(j \preceq_{\text{Col}} j') \right|, \quad (15b)$$

where the first sum is across columns j with $\sum_i y_{ij} \neq 0$.

3 Simulation study

Here we present an evaluation of the proposed methodology, applying the Metropolis algorithm described in Section 2.3 to artificial data generated under the stochastic HICLAS model. Two questions will be addressed: (a) What is the convergence rate, that is, does the algorithm converge for all parameters and, if so, how fast? (b) How does the obtained posterior distribution relate to the underlying true model? As pointed out by Gelman et al. (2004), even if the Markov chains appear to have converged, it is still possible that some important areas of the target distribution have not been captured, for example, because they could not be reached by the simulation algorithm.

3.1 Design and procedure

As a first step, true two-error probability models (S, P, π_0, π_1) were created while systematically varying the following four factors:

1. The *Rank* r of the model (i.e., the number of columns in S and P) was set to either 2 or 3.
2. The *Size* $m \times n$ of the predicted matrix $\widehat{Y}(S, P)$ was set to either 25×25 or 60×40 .

3. The value of π_0 varied at three levels: .05, .10, and .20.
4. The value of π_1 varied at three levels: .05, .10, and .20.

These four factors were orthogonally crossed and for each cell in the design, 10 true models (S, P, π_0, π_1) were generated, where for each model the entries in the pair (S, P) were independent realizations of a Bernoulli variable:

$$\begin{aligned} S_{ik} &\stackrel{\text{iid}}{\sim} \text{Bernoulli}(1 - \mu) && (i = 1 \dots m, k = 1 \dots r) \\ P_{jk} &\stackrel{\text{iid}}{\sim} \text{Bernoulli}(\mu) && (j = 1 \dots n, k = 1 \dots r). \end{aligned} \tag{16}$$

For each model, a value of μ was randomly selected in such a way that $\Pr(Y_{ij} = 1 \mid r, m, n, \pi_0, \pi_1, \mu)$, with Y_{ij} defined as in (1), can be considered a draw from a uniform distribution on $[0.30, 0.70]$.⁵ In the next step, a data matrix Y was created from each model in accordance with (5). As a result of this procedure, we obtained in total $10 \times 2(\text{Rank}) \times 2(\text{Size}) \times 3(\pi_0) \times 3(\pi_1) = 360$ data matrices, which subsequently were analyzed using the proposed Metropolis algorithm. As to the algorithmic options, we opted for four sequences, which were considered to have converged as soon as $\hat{R} < 1.05$ for all parameters. Furthermore, we chose $\lambda = 3$ for the parameter of the Poisson distribution used in generating new candidates. To save computer resources, only each hundredth draw was saved and included in the calculations for \hat{R} .

For the analyses we used a Pentium IV processor with a clock frequency of 3.06 GHz. The analysis of a data set was stopped and the chains were considered to have not converged whenever the number of iterations exceeded 10,000,000.

3.2 Analysis and results

For 359 out of the 360 data sets, \hat{R} showed convergence of the chains. Overall, for about half of the data sets, convergence was reached at less than 210,000 iterations (and in less than one minute and a half); in 5% of the cases, more than 1,000,000 iterations (or more than 15 minutes were needed). For the 359 data sets that showed convergence, we now examine how well the true model parameters (viz., π_0 , π_1 , s_{ik} , and p_{jk}), can be recovered from the simulated posterior distribution. In particular, we calculated for each parameter in each data set the 95% Bayesian

confidence interval defined by the 2.5 and 97.5 percentiles of the marginal posterior distribution for the parameter, which allowed us to assess goodness of recovery by examining (a) whether or not the true value of the parameter falls within the interval and (b) the width of the interval.

In Figure 2 we present summary results for both statistics, separately for different sizes and ranks of the model. With respect to the error parameters, the 95% confidence intervals have a mean width of 0.06, but are too narrow on average, as for only 92% of the data sets, the confidence interval effectively contained the true value. With respect to the goodness of recovery of a bundle value parameter s_{ik} or p_{jk} , remind that the true value is either 0 or 1 and that also the Bayesian confidence interval has a width of either 0 (i.e., the 2.5 and 97.5 percentiles have the same value) or 1 (i.e., the 2.5 and 97.5 percentiles equal 0 and 1, respectively). From the left-hand panels in Figure 2, we read then that the latter occurs in somewhat more than 10% of the cases on average (with some variation). It follows that in about 90% of the cases the confidence interval for a bundle parameter reduces to a single value. Given that the Bayesian confidence interval almost always (i.e., in 99.9% of the cases) contains the true value, we can derive that a Bayesian analysis succeeds in recovering the true value of about 90% of the cells in S and P with virtual certainty. As an aside, we note that for the goodness of recovery of the association, equivalence, and hierarchy relations, we found similar results as compared to those of the parameters s_{ik} and p_{jk} on which they depend.

A point estimate for a parameter is usually obtained by its posterior mean. For an alternative, more graphical way to look at the recovery of the binary bundle parameters, we calculated the posterior mean for each bundle parameter (s_{ik} as well as p_{jk}) in each data set. Subsequently, the set of all (67,500) bundle parameters was partitioned in groups of bundle parameters with identical posterior means (after rounding to the nearest 0.01 unit) and, for each group, the mean of the true bundle parameter values was calculated and plotted in Figure 3 (along with a 95%-confidence interval). We note a strong 1–1 relation: The posterior mean obtained for a bundle value closely resembles the probability that the corresponding true parameter equals 1. This means that the posterior distribution correctly represents the uncertainty about the true parameter's value.

In this simulation study, the true HICLAS models were randomly constructed, which makes it rather unlikely that the true model is nonunique. Considering that an important motivation

for developing the Bayesian HICLAS model was to trace possible identifiability problems, it is of particular interest to see how the proposed algorithm behaves if the posterior density is multimodal (with the multimodality being due to nontrivial identifiability problems). To this end, a second simulation study was set up, with 20 data sets that were constructed in such a way that the posterior density was (at least) trimodal with the three modes corresponding to different variants of the same nonunique model. Importantly, we found that in all cases the simulated posterior distribution generated by the proposed Metropolis algorithm included these three modes, as such revealing, indeed, lack of identifiability of these models at the mode.⁶

4 Applications

We now illustrate the Bayesian approach in hierarchical classes modeling by a reanalysis of two data sets that were previously analyzed with the deterministic model. For the first application, in the area of the psychology of choice, we use person by object select/nonselect data (i.e., so-called pick any/ n data, Coombs, 1964), presented in Van Mechelen and Van Damme (1994). In the second application, an analysis of patient by symptom presence/absence data, originally presented by Van Mechelen and De Boeck (1989), illustrates the approach in the domain of clinical psychology.

4.1 Revealing latent choice criteria

Studies on decision-making processes have shown that subjects who look for an optimal choice in a large set of complex choice alternatives often apply a two-stage strategy: First, a subset (the so-called “consideration set”) of alternatives is selected applying predominantly noncompensatory (disjunctive/conjunctive) combination rules and, subsequently, the final choice is made from that reduced set based on compensatory rules (see, e.g., Beach, 1990; Ogilvie & Schmitt, 1979; Westenberg & Koele, 1992). Van Mechelen and Van Damme (1994) simulated the first stage of this decision-making process by presenting each of 26 second-year psychology students with 25 index cards with room descriptions from the Housing Service of the University of Leuven and asking him/her to select those rooms (s)he would decide to visit when looking for a student room. (Most

students in Leuven rent a room for the academic year and pass by the Housing Service to get an overview of the available rooms.) This resulted in a 26×25 binary data matrix Y with $y_{ij} = 1$ if student i selected room j , and $y_{ij} = 0$ otherwise.

In line with the earlier studies, Van Mechelen and Van Damme (1994) assumed that each student evaluated each room on a number of latent choice criteria and that (s)he selected those rooms that meet all choice criteria that (s)he considered important. This decision process is formalized by a variant of the conjunctive HICLAS model with the bundles corresponding to the latent choice criteria, that is, with s_{ik} indicating whether or not student i considers criterion k important and p_{jk} indicating whether room j meets criterion k . As such, the following association rule predicts whether student i includes room j in his (her) consideration set:⁷

$$\hat{y}_{ij}(S, P) = \begin{cases} 1 & \text{if } \forall k(1 \leq k \leq r) : p_{jk} \geq s_{ik} \\ 0 & \text{otherwise.} \end{cases}$$

That is, a student selects a room if, for all latent criteria, the room either meets the criterion or the student does not consider it important. A substantive challenge of the study was to induce from the data the criteria that guided the decision process. Similar to the child by item example in Section 1.1, hierarchical classifications of students and rooms (based on the relations among sets of latent criteria) naturally follow from the conjunctive theory underlying the model.

Here we apply a Bayesian HICLAS model to the same data. The error process implied by the stochastic extension means that rooms may turn up in a student's consideration set, while they are not predicted to be there (or vice versa), which is likely to happen given the complex descriptions of the rooms. We fitted models with one and two error probability parameters and with ranks varying from 1 to 4, applying the described Metropolis algorithm with 5 sequences and the convergence criterion $\hat{R} < 1.05$ for all parameters. For all the fitted models, the Bayesian analysis made clear that a deterministic approach would suffer from multimodality as several different HICLAS models (S, P) were found with a minimal number of discrepancies $D(\hat{Y}(S, P), Y)$. For example, for the rank 1 one-error probability model, nine pairs (S, P) were found with a minimum of 197 discrepancies. Moreover, many more models come close to this lower bound. Yet, applying the original deterministic algorithm would return a single model, keeping the user unaware of possible

alternative models with an (almost) equally low number of discrepancies.

As to rank selection, we both calculated Bayes factors and applied the PPC procedure based on the statistic in (14). Whereas the latter chooses a rank 2 model (as testing rank 1 versus rank 2 yields $T = 54$, PPC- $p = .02$, while testing rank 2 against rank 3 yields $T = 27$, PPC- $p = .12$, indicating that the more complex rank 3 model does not significantly improve the fit), the Bayes factor shows a preference for a higher rank ($2\log_e B_{32} = 18.56$, $2\log_e B_{42} = 18.34$, $2\log_e B_{34} = 0.22$). In case of such conflicting results, the final decision will generally be based on substantive considerations. Here, for simplicity's sake, we will retain the more parsimonious model in rank 2 for further discussion (which was also the rank of the model in the original paper).

Some further checking of the retained model shows a reasonable fit to the data. The Bayes factor comparing the hypothesis of two error probability parameters with that of one error parameter shows no need to differentiate the error probabilities for predicted values of 0 and 1 ($2\log_e B_{12} = 2.87$, showing a slight preference for the one-error probability model). Furthermore, no inconsistency between the set-theoretical relations as represented in the model and the corresponding if-then relations in the data is found when a PPC is performed using test quantities (15a) and (15b) (PPC- $p = .41$ and PPC- $p = .40$, respectively).

Figure 4 shows a graphical representation of the rank 2 Bayesian HICLAS model, along the lines described in Section 2.4. The three types of relations can be read from the figure as follows: (a) Student and room *classes* are represented by solid-framed and dotted-framed boxes, respectively. As to the uncertainty with respect to the classifications, for each element, the marginal posterior probability of belonging to the class is added. So, we can read, for instance, that Room 7 is found to be in cluster C in all the simulation draws whereas Room 12 is found about evenly in the clusters A and C. In this respect, we emphasize that the appearance of the same element in different classes does not imply the clustering to be overlapping but rather reflects uncertainty of the element's location. (b) The room and student *hierarchies* are represented by solid and dotted arrows between classes, respectively, which are labeled with the posterior probability that the corresponding classes are hierarchically related. We read, for example, that it is rather uncertain (.42) that student class a is hierarchically below class d , while room class B is almost surely (.90)

hierarchically below room class C. (c) The *association* relation can be read deterministically as in the original model: Rooms are selected by any student who is below it. The uncertainty in the association relation is not represented in the graph, though. We use Figure 5 to represent the marginal posterior probability for each room-student association by using grey values, with darker values indicating a higher probability of the student selecting the room. The dot in the center of each cell, either black or white, represents the observed data value and visualizes the degree of discrepancy between the model and the data with respect to the association relation.

Figures 4 and 5 illustrate that, although many different pairs of bundle matrices (S, P) with a minimal or near-minimal value on the loss-function exist, the hierarchical classes structure does have a stable part. For example, in almost all simulation draws, Rooms 7, 22, and 13, were found in the same class (at the top of the hierarchy), and with respect to the association relation, Figure 5 shows that for most cells, there is little uncertainty about the student-room associations (with most squares being either very dark or very light). With regard to the substantive interpretation of the deterministic core of the model, we largely follow Van Mechelen and Van Damme (1994), who interpreted the two latent choice criteria in terms of quality and price of the rooms.

4.2 Revealing implicit taxonomy in psychiatric diagnosis

Cognitive theories on the diagnostic process in psychiatry often assume that underlying a clinician's judgements and decisions, there is an implicit and personal diagnostic classification system (see, e.g., Chan & Jackson, 1979, 1982; Garb, 1998, 2005). The implicit taxonomy of a diagnostician indicates for each syndrome which symptoms are considered to be characteristic of it and acts as a mediator inbetween the available patient information and his/her final judgements and decisions.

Van Mechelen and De Boeck (1989) obtained data asking a single psychiatrist to judge the absence/presence of 22 symptoms in each of 30 of his patients. The data were organized in a 30×22 matrix Y with $y_{ij} = 1$ if symptom j is judged to be present in patient i , and 0 otherwise. In line with the cognitive theories presented above, Van Mechelen and De Boeck assumed that the psychiatrist judges a symptom to be present in a patient iff the symptom is characteristic of one of the syndromes in his implicit taxonomy from which the patient suffers (according to the

information he has available). This association rule is formalized in a disjunctive HICLAS model (De Boeck & Rosenberg, 1988) with the bundles corresponding to latent syndromes, that is, s_{ik} indicating whether or not patient i suffers from syndrome k and p_{jk} whether or not symptom j is characteristic of syndrome k . The association rule of the disjunctive model is then given by:

$$\hat{y}_{ij}(S, P) = \begin{cases} 1 & \text{if } \exists k(1 \leq k \leq r) : s_{ik} = p_{jk} = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

In words, symptom j is judged to be present in patient i iff *at least one* syndrome k exists of which symptom j is characteristic (i.e., $p_{jk} = 1$) and from which patient i suffers (i.e., $s_{ik} = 1$).

For a Bayesian analysis of these data, we implemented a stochastic extension of the disjunctive HICLAS model along the same lines as described for the conjunctive model in Section 2. With both the psychiatric symptoms and the information from the patients being highly complex, misjudging the presence/absence of a symptom in a patient is likely to occur. We fitted models with ranks ranging from 1 to 4 and found that a PPC with test statistic (14) pointed to the rank 4 model. Because the Bayes factor was practically indifferent between the models of rank 2, 3, and 4, we retained the model of rank 4 for further discussion. A Bayes factor comparing the one-error and two-error probability models of rank 4 showed that there is no need to differentiate the probability of misses and false alarms ($2 \log_e B_{12} = 14.91$).

Figure 6 shows a graphical representation of the retained rank 4 model (which is structurally different from a graph for a conjunctive model). Patient and symptom classes are represented by boxes with the patient hierarchy at the top half of the graph and the symptom hierarchy, upside down, in the bottom half; furthermore, as to the association relation, a patient is predicted to have a symptom iff a downward path exists from the patient's class to the symptom's class. Similar to the conjunctive case, the uncertainty with respect to the classifications is indicated by adding the posterior probability behind each patient/symptom and the uncertainty of the hierarchy relations among classes by labeling each connecting line as explained in Section 2.4. Note that many of these posterior probabilities are rather low in general: Inferences about the set-theoretical relations among patients or symptoms must be made with care.

We will now compare the Bayesian model with (a) the deterministic model of the original paper

and (b) a PMD model fitted to the same data (using Maris et al.'s, 1996, algorithm). Table 2 shows the estimates for the bundle parameters obtained by each type of analysis.

The deterministic rank 4 model presented by Van Mechelen and De Boeck (1989) showed 70 (10.6%) discrepancies to the data. The Bayesian analysis now reveals that this result was a local optimum, as in the simulated posterior distribution 19 (distinct) models (S, P) were found with less discrepancies. More importantly, though, than a gain of a few discrepancies less, is the finding that many alternative HICLAS models may be considered for these data (e.g., more than 300 models exist with 71 or less discrepancies). The Bayesian approach elegantly maps the aspects that are robust across the large set of good models for these data.

Comparing the values in the symptom bundle matrix P of the deterministic analysis with the corresponding posterior means of the Bayesian analysis (Table 2) confirms the gist of Van Mechelen and De Boeck's (1989) interpretation of the syndromes, although some remarks should be added: The implicit syndromes corresponding to Bundle I and II seem to represent Affective Disorders, but where the deterministic output does not allow to differentiate them, based on the Bayesian output, we can identify the first syndrome as pure Depression (with little uncertainty that the symptoms depression, suicide, and social isolation are characteristic of it) while Anxiety seems more specific to the second syndrome. About the interpretation of the fourth implicit syndrome (Bundle IV) as Psychotic Disorder there is virtually no doubt (with posterior probabilities for hallucinations/delusions, suspicion, and agitation close to 1). The interpretation of the third syndrome (Bundle III), however, is less clear: Whereas in the original paper it was labeled Substance Abuse Disorder, the posterior probability of the symptom substance abuse belonging to the third syndrome turns out to be only 0.27. Because no symptoms are found that are exclusively related to Bundle III, finding an adequate interpretation for this syndrome is difficult, anyhow.

The output of the PMD and Bayesian HICLAS analysis in Table 2 looks very similar, with for each cell in the bundle matrices a probability value. However, the interpretation of the probabilities is different. In the PMD model, they are the parameter of bundle-specific Bernoulli random variables: The observed value in cell (i, j) of the data matrix is explained by a two-step process where, first, the r Bernoulli variables of both patient i and symptom j are realized (to indicate, respectively,

which syndromes the patient suffers from and which syndromes the symptom is characteristic of) and, second, the thus obtained realizations are combined by (17) to give rise to the observed association. For a Bayesian analysis, we assumed that the relations between patients and syndromes and symptoms and syndromes are fixed but unknown and that the posterior probabilities quantify the uncertainty about these relations after taking into account the information in the data. In the current context, the PMD assumption that the bundle patterns of patients and symptoms are realized again for each cell⁸ seems unrealistic (as it means that the syndromes that a symptom is characteristic of [resp. a patient suffers from] may change with each new patient [resp. syndrome]).

As an aside, we mention that comparing the fit of the PMD and Bayesian HICLAS model to these data turned out to be slightly in favour of the latter. Considering that both models yield a probability of a patient and a symptom being associated in the data, one may compare the correspondence —as quantified by, for example, a correlation— between these probabilities and the observed data. For the current data, these correlations were 0.71 and 0.80 for the PMD and the Bayesian HICLAS model, respectively (which is a significant difference, $z = 8.45$, $p < .01$, Meng, Rosenthal, & Rubin, 1992). An alternative strategy for selecting among competing models could be based on cross-validation techniques (see, e.g., Eastment & Krzanowski, 1982, who proposed such a strategy in the context of principal component analysis).

Apart from a quantification of the uncertainty about bundle membership, the Bayesian HICLAS model also includes a classification and hierarchy for patients and symptoms and a quantification of the uncertainty about the latter. The discussion in Van Mechelen and De Boeck (1989) shows the importance of the hierarchical classification in this study. In a PMD model, this information is lost, and, although one may try to recycle it (e.g., by dichotomizing the output, or by combining in some way the probability parameters of patients or symptoms), it stays outside the model's framework (i.e., it is arbitrary or additional assumptions are needed).

5 General discussion

This paper presented a new stochastic extension of the original deterministic hierarchical classes model (De Boeck & Rosenberg, 1988; Van Mechelen et al., 1995) in response to a number of con-

ceptual and practical problems. Conceptually, the benefit of the new extension is that the relation between the predicted values and the observed values is made explicit thanks to the inclusion of one or two error-probability parameters. This enabled us to consider the model within a Bayesian framework and to make use of a variety of tools for model estimation and model checking. As shown in the illustrative applications, this approach leads to a more differentiated view, resulting in a fine-tuned interpretation where core and peripheral aspects of the resulting structure are distinguished. Unlike the probabilistic variant developed by Maris et al. (1996), this differentiation is obtained without leaving or “fuzzifying” the set-theoretical or logical framework of the original HICLAS model. As such, the new model on the one hand fully retains the deterministic core of the original model, including the representation of the set-theoretical relations; on the other hand, by superimposing a stochastic framework, discrepancies between the underlying deterministic structure and the actual data are fully accounted for, and by means of a Bayesian approach uncertainty in our knowledge about the deterministic structure (including the set-theoretical relations) is fully captured.

As an illustration of the psychological relevance of the new model and of its potential range of application, we notice here the particularly close link between the stochastic extension of the conjunctive HICLAS model when applied to person by item success/failure data and knowledge space theory (Doignon & Falmagne, 1999; Falmagne et al., 1990). In particular, the bundels in such a HICLAS model can be considered formally equivalent with the skills (or competences) that are assumed to underly a set of items in knowledge space theory, while the person classes correspond with knowledge states. The hierarchical relations among the knowledge states as represented by the HICLAS model are of utmost interest as they reveal possible learning paths people may pass through in going from the empty knowledge state to complete mastery of the domain. The error parameters π_0 and π_1 further correspond to the concepts within knowledge space theory of probabilities of a lucky guess and a careless error, respectively.

The simulation study and applications show how the Bayesian approach reveals the full complexity of the solution space when searching for a deterministic HICLAS model that is minimally discrepant with a (medium-sized) data matrix. Three findings in particular shed a new light on

the optimization problem. First, in many cases, there is not a single optimum, that is, the solution space is multimodal in a strict sense. Second, for all data matrices that we studied (both artificial and real), a large set of near-optimal HICLAS models was found (i.e., models with only one or a few discrepancies more than the optimal), with the differences among these models often corresponding to important substantive differences. Third, the simulation study showed that the underlying “true” model often has considerably more discrepancies than the best fitting model(s), particularly when the underlying truth is complex and highly sensitive to error. These three findings show that interpreting a single, arbitrary selected, (near-)optimal solution (such as the model returned by a deterministic HICLAS algorithm) may be a perilous undertaking. Arguably, the most important practical contribution of the Bayesian approach is that it provides an overall view, highlighting model aspects that are robust across the large set of plausible models for the given data set.

Finally, the new model extension as presented here may be extended further in various directions. First, while the presented model includes only one or two probability parameters, additional error parameters may be introduced based on hypotheses about the error process at hand. Such hypotheses may be either a priori or may result from model checks. A second direction in which the model can be further extended relates to the choice of the prior distribution. In this paper, we have chosen a uniform distribution for the set of parameters, implying that any hierarchical classes model is considered equally likely a priori and that the posterior distribution is proportional to the likelihood. Alternative prior distributions may be considered, motivated either by former results or by restrictions one may wish to add to the HICLAS model (for example, the restriction that the hierarchies constitute total orders rather than partial orders, or that the hierarchical relationship among some elements is fixed, see Ceulemans, Van Mechelen, & Kuppens, 2004). Such restrictions can easily be incorporated by assigning a zero probability in the prior distribution to models that do not satisfy them. The Metropolis algorithm for simulating the posterior distribution can be adapted accordingly either by changing the subroutine that generates the new model candidates or simply by dropping models that do not satisfy the restrictions. Finally, an interesting direction for further research is charting out by means of a sensitivity analysis how and to which extent the obtained simulated posterior distribution may depend on the selection of the prior.

References

- Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts*. Chichester, England: Wiley.
- Berkhof, J., Van Mechelen, I., & Gelman, A. (2003). A Bayesian approach to the selection and testing of mixture models. *Statistica Sinica*, *13*, 423–442.
- Ceulemans, E., Van Mechelen, I., & Kuppens, P. (2004). Adapting the formal to the substantive: Constrained Tucker3-HICLAS. *Journal of Classification*, *21*, 19–50.
- Ceulemans, E., Van Mechelen, I., & Leenen, I. (2003). Tucker3 hierarchical classes analysis. *Psychometrika*, *68*, 413–433.
- Chan, D. W., & Jackson, D. N. (1979). Implicit theory of psychopathology. *Multivariate Behavioral Research*, *14*, 3–19.
- Chan, D. W., & Jackson, D. N. (1982). Individual differences in the perception and judgment of psychopathology. *Multivariate Behavioral Research*, *17*, 3–32.
- Cheng, C. F. G. (1999). A new approach to the study of person perception: The hierarchical classes analysis. *Chinese Journal of Psychology*, *41*, 53–64.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, *90*, 1313–1321.
- Chib, S., & Jeliazkov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association*, *96*, 270–281.
- Coombs, C. H. (1964). *A theory of data*. New York: Wiley.
- De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika*, *53*, 361–381.
- Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces*. Berlin, Germany: Springer.
- Eastment, H. T., & Krzanowski, W. J. (1982). Cross-validatory choice of the number of components from a principal component analysis. *Technometrics*, *24*, 73–77.
- Falmagne, J.-C., Koppen, M., Villano, M., Doignon, J.-P., & Johannesen, L. (1990). Introduction to knowledge spaces: How to build, test and search them. *Psychological Review*, *97*, 201–224.
- Ganter, B., & Wille, R. (1996). *Formale Begriffsanalyse: Mathematische Grundlagen* [Formal

- concept analysis: Mathematical foundations]. Berlin, Germany: Springer.
- Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.
- Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis* (2nd ed.). London: Chapman & Hall.
- Gelman, A., Leenen, I., Van Mechelen, I., De Boeck, P., & Poblome, J. (2003). *Bridges between deterministic and probabilistic classification models*. Manuscript submitted for publication.
- Gelman, A., Meng, X.-L., & Stern, H. S. (1996). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, *6*, 733–807.
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*, 457–511.
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (Eds.). (1996). *Markov chain Monte Carlo in practice*. London: Chapman & Hall.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*, 97–109.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*, 773–795.
- Leenen, I., & Van Mechelen, I. (2001). An evaluation of two algorithms for hierarchical classes analysis. *Journal of Classification*, *18*, 57–80.
- Leenen, I., Van Mechelen, I., & De Boeck, P. (2001). Models for ordinal hierarchical classes analysis. *Psychometrika*, *66*, 389–404.
- Leenen, I., Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1999). INDCLAS: A three-way hierarchical classes model. *Psychometrika*, *64*, 9–24.
- Luyten, L., Lowyck, J., & Tuerlinckx, F. (2001). Task perception as a mediating variable: A contribution to the validation of instructional knowledge. *British Journal of Educational Psychology*, *71*, 203–223.

- Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika*, *61*, 7–29.
- Meng, X.-L. (1994). Posterior predictive p -values. *Annals of Statistics*, *22*, 1142–1160.
- Meng, X.-L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172–175.
- Meulders, M., De Boeck, P., & Van Mechelen, I. (2003). A taxonomy of latent structure assumptions for probability matrix decomposition models. *Psychometrika*, *68*, 61–77.
- Ogilvie, J. R., & Schmitt, N. (1979). Situational influences on linear and nonlinear use of information. *Organizational Behavior and Human Performance*, *4*, 337–352.
- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, *12*, 1151–1172.
- Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy: The principles and practice of numerical classification*. San Francisco, CA: Freeman.
- ten Berge, M., & de Raad, B. (2001). The construction of a joint taxonomy of traits and situations. *European Journal of Personality*, *15*, 253–276.
- Van Mechelen, I., Bock, H.-H., & De Boeck, P. (2004). Two-mode clustering methods: A structured overview. *Statistical Methods in Medical Research*, *13*, 363–394.
- Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. *Journal of Social and Clinical Psychology*, *8*, 276–287.
- Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika*, *60*, 505–521.
- Van Mechelen, I., Rosenberg, S., & De Boeck, P. (1997). On hierarchies and hierarchical classes models. In B. Mirkin, F. R. McMorris, F. S. Roberts, & A. Rzhetsky (Eds.), *Mathematical hierarchies and biology* (pp. 291–298). Providence, RI: American Mathematical Society.
- Van Mechelen, I., & Van Damme, G. (1994). A latent criteria model for choice data. *Acta Psychologica*, *87*, 85–94.
- Westenberg, M. R. M., & Koele, P. (1992). Response modes, decision processes and decision outcomes. *Acta Psychologica*, *80*, 169–184.

Footnotes

¹Mathematically, the decomposition of \widehat{Y} into S and P implies that \widehat{Y} equals the Boolean matrix product of the two bundle matrices (possibly after taking complements of \widehat{Y} , S , and P). We limit the discussion here to binary bundle matrices, although also generalizations have been proposed where the bundles are ordinal variables (Leenen, Van Mechelen, & De Boeck, 2001). For more details, we refer to the original papers.

²Strictly speaking, this is only true in general under the condition that $\hat{\pi} \leq 0.50$.

³Because of the asymmetrical relation between the s_{ik} and p_{jk} (see (1)), the parameters of the Bernoulli random variables are chosen to be complementary for $S_{ik}^{(0)}$ and $P_{jk}^{(0)}$. Then, it follows that:

$$S_{ik}^{(0)} \stackrel{\text{iid}}{\sim} \text{Bernoulli} \left(1 - \sqrt{1 - p_1^{1/r}} \right) \quad \text{and} \quad P_{jk}^{(0)} \stackrel{\text{iid}}{\sim} \text{Bernoulli} \left(\sqrt{1 - p_1^{1/r}} \right),$$

where p_1 denotes the proportion of one-entries in Y .

⁴The converged part of a sequence includes only the draws used in the calculations of \hat{R} . Gelman and Rubin (1992) recommend to discard the first half of each sequence (to diminish the effect of the starting values).

⁵From Equations (1), (5), and (16), it follows that:

$$\Pr(Y_{ij} = 1 \mid r, m, n, \pi_0, \pi_1, \mu) = (1 - \mu^2)^r (1 - \pi_0 - \pi_1) + \pi_0.$$

⁶Full details of this second simulation study are available from the first author on simple request.

⁷Mathematically, the conjunctive HICLAS model variant used in this section comes down to the conjunctive HICLAS model as introduced in Section 1 for the transposed data.

⁸Recently developed variants of the PMD model would allow for the possibility that the bundle patterns of either patients or symptoms are fixed (i.e., they are not realized again for each cell entry, see Meulders, De Boeck, & Van Mechelen, 2003). However, a PMD model with fixed bundle patterns for both modes of a binary data matrix coincides with the deterministic HICLAS model.

Figure captions

Figure 1: Graphical representation of the hierarchical classes model in Table 1.

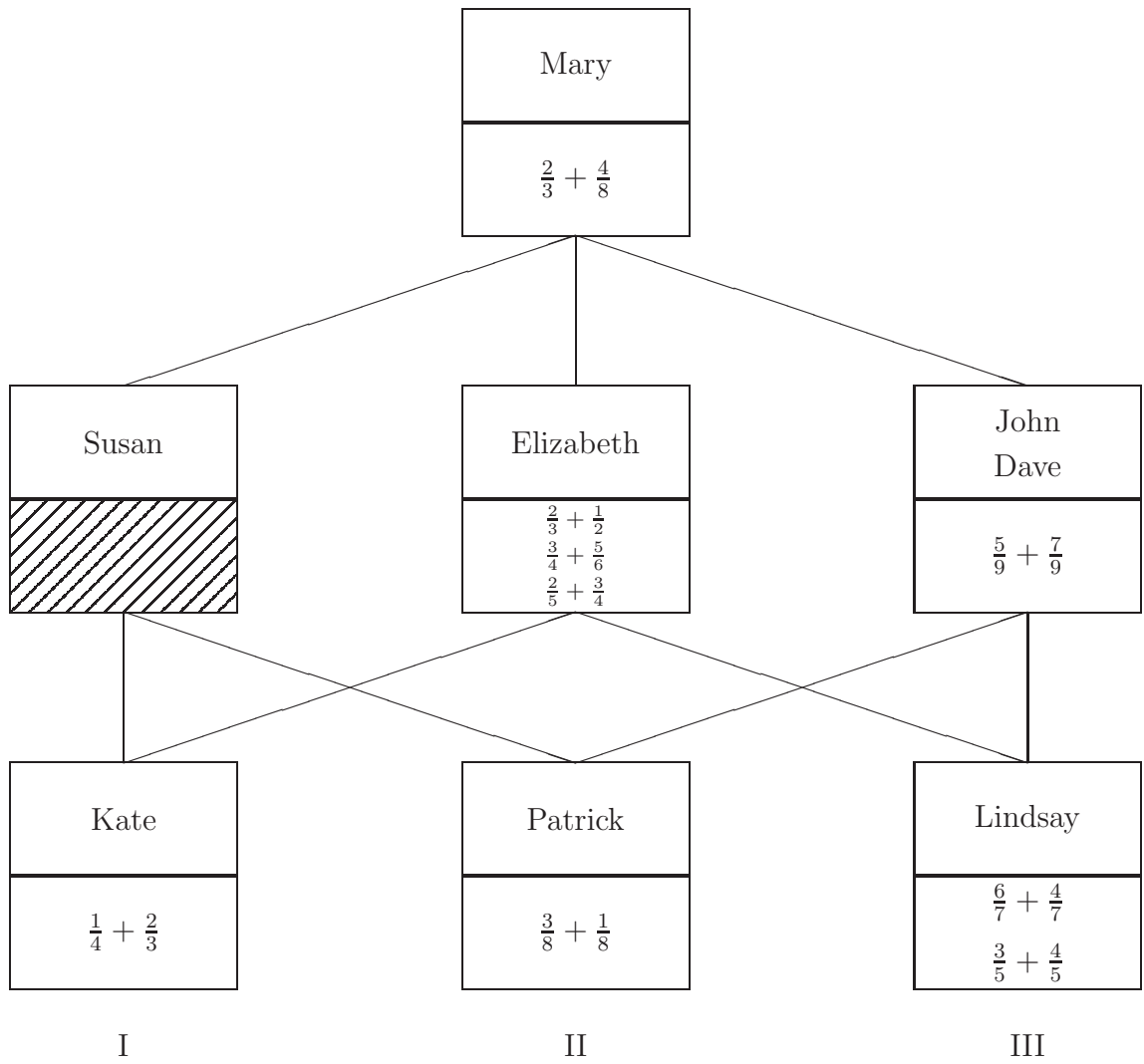
Figure 2: Recovery statistics for π_0 , π_1 , s_{ik} , and p_{jk} for different sizes of the data matrix and ranks of the underlying true model. CI denotes confidence interval.

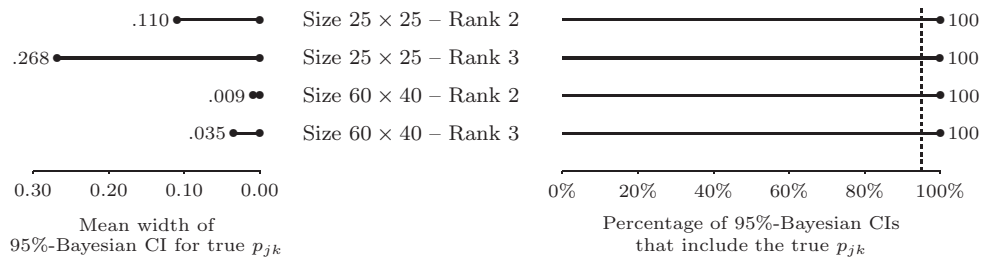
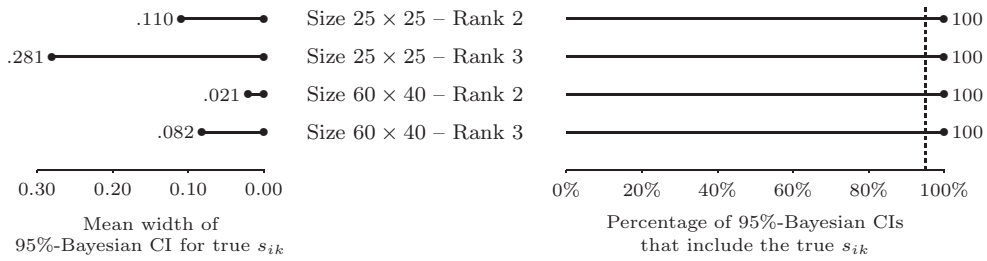
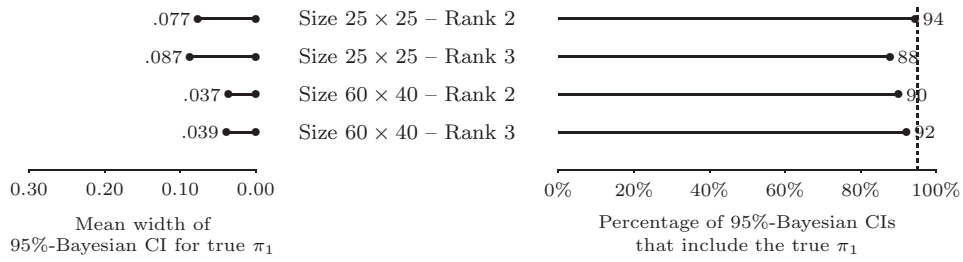
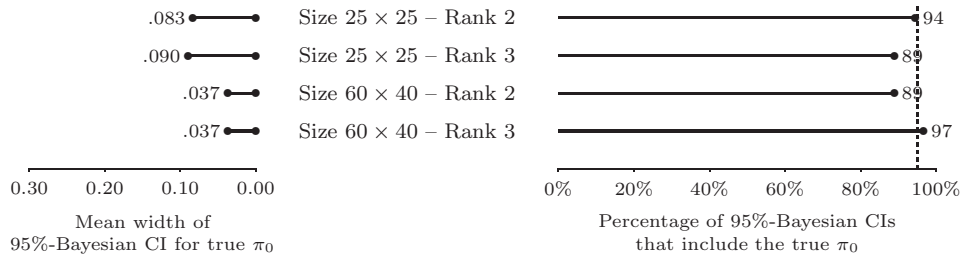
Figure 3: Mean true values of parameters in S and P (with 95% confidence intervals) given the estimated posterior probability.

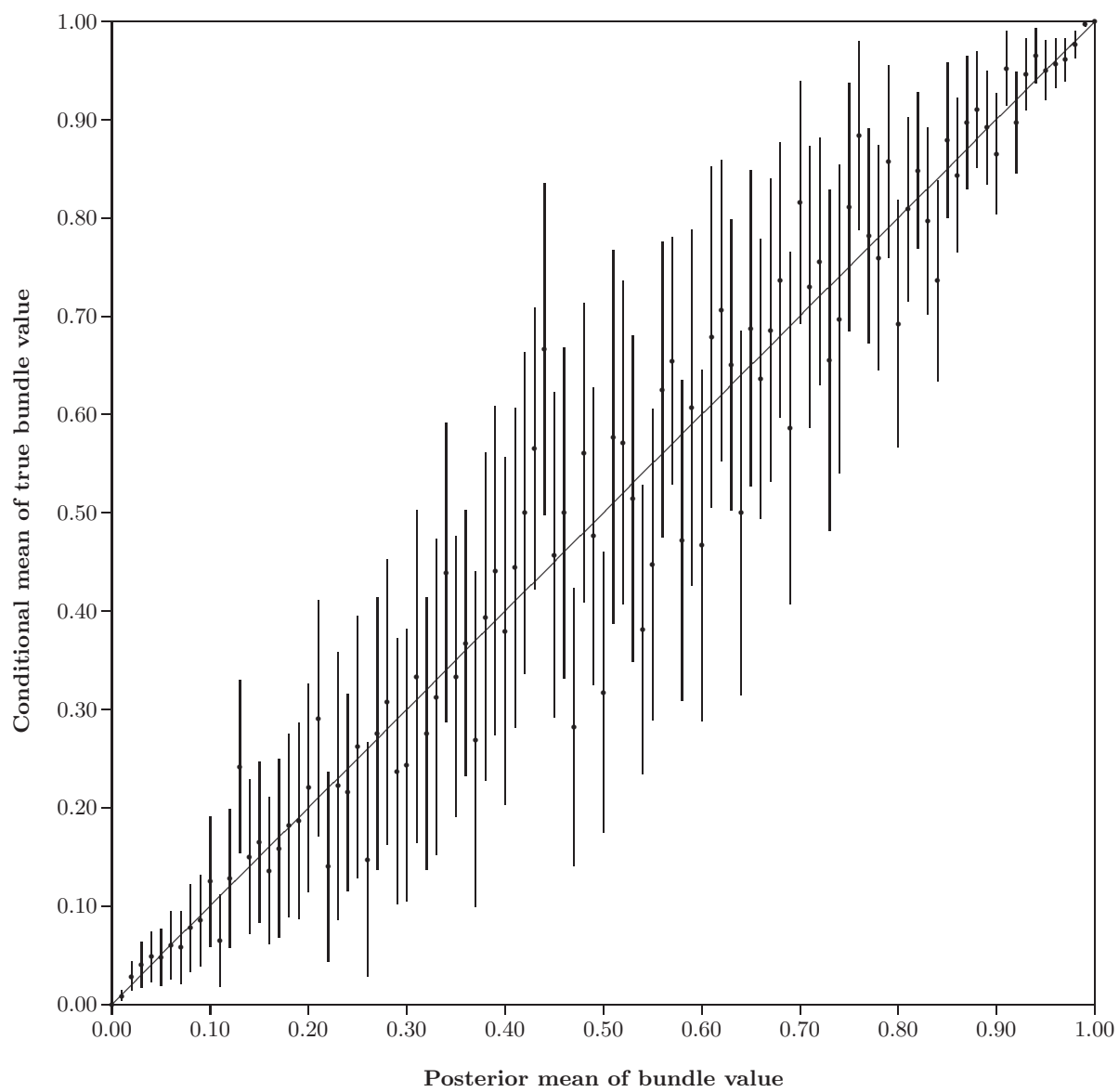
Figure 4: Graphical representation of the conjunctive Bayesian HICLAS model in rank 2 for the room data. Numbers in italic behind each student/room indicate the posterior probability of class membership; the numbers attached to the arrows indicate posterior probabilities of hierarchical relations between classes. Elements in classes and hierarchical relations among classes are only represented if the corresponding posterior probability exceeds .33.

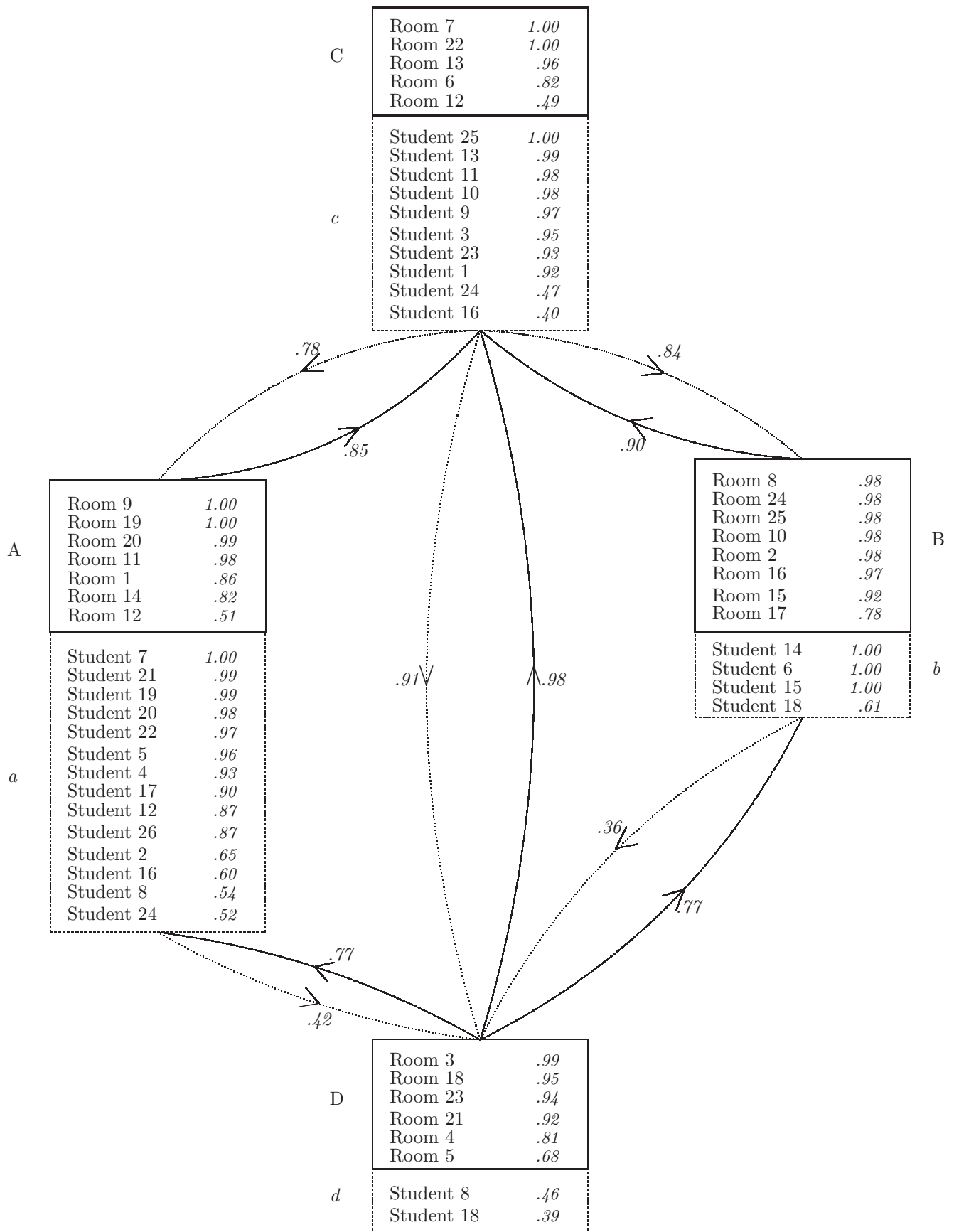
Figure 5: Posterior probabilities for the (room, student)-association relation using grey values, with darker values denoting higher probabilities. The rows [resp. columns] have been re-ordered in order to represent student [resp. room] classes (based on each element's highest posterior class membership probability). The circle in the middle of each cell represents the observed data value (with black and white for observed values of 1 and 0, respectively).

Figure 6: Graphical representation of the Bayesian disjunctive HICLAS model in rank 4 for the psychiatrist data. Numbers in italic behind each patient/symptom indicate the posterior probability of class membership; the numbers attached to the arrows indicate posterior probabilities of hierarchical relations between classes. A patient's final diagnosis is added between parenthesis (A = 'Affective or Anxiety Disorder'; P = 'Psychotic Disorder'; S = 'Substance Use or Substance-Induced Disorder'; X = 'Other').









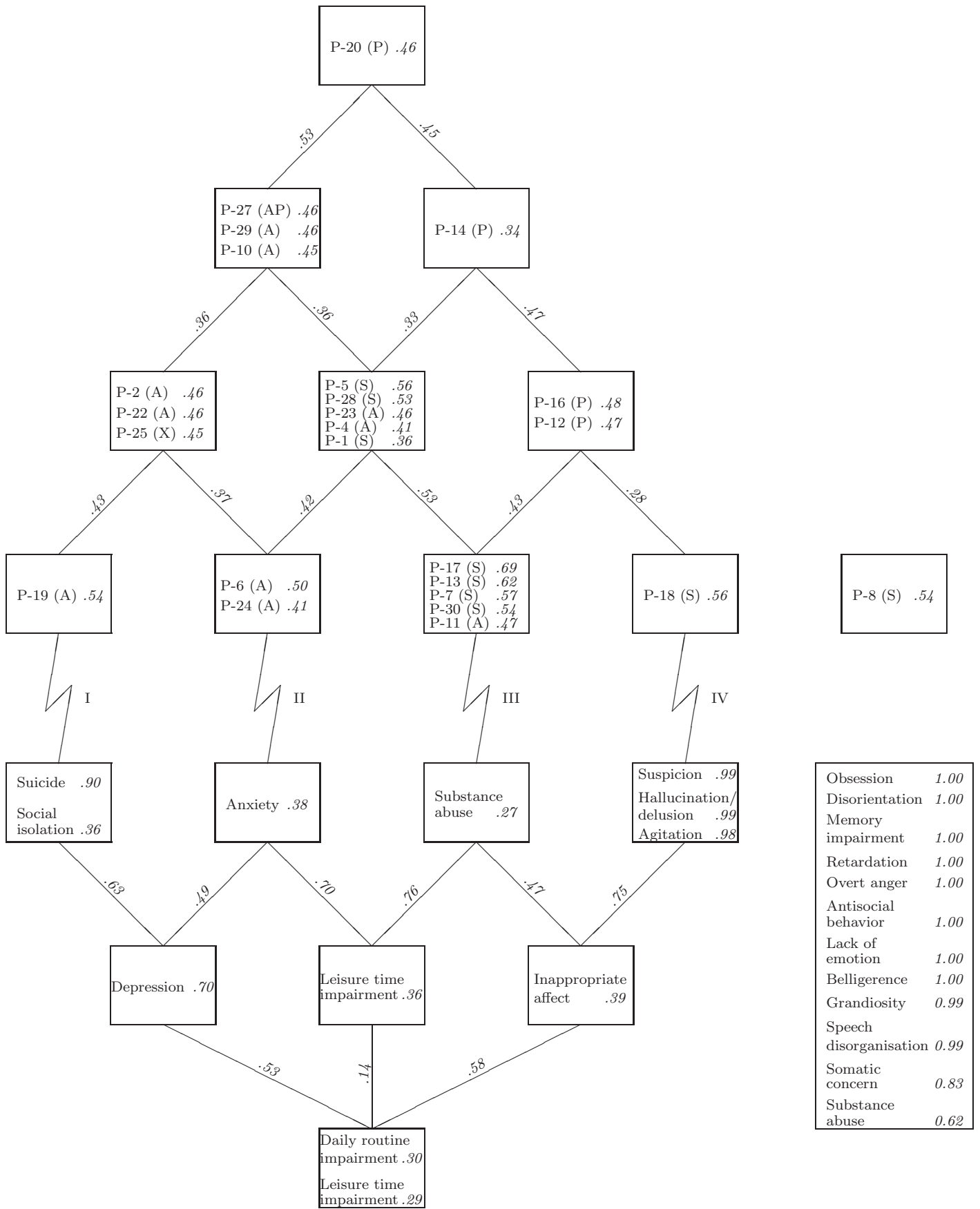


Table 1: *Hypothetical Two-Way Two-Mode Binary Data and Hierarchical Classes Model*

| Data | Items | | | | | | | | |
|-----------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| | $\frac{5}{9} + \frac{7}{9}$ | $\frac{2}{3} + \frac{1}{2}$ | $\frac{3}{4} + \frac{5}{6}$ | $\frac{6}{7} + \frac{4}{7}$ | $\frac{3}{8} + \frac{1}{8}$ | $\frac{1}{4} + \frac{2}{3}$ | $\frac{2}{5} + \frac{3}{4}$ | $\frac{3}{5} + \frac{4}{5}$ | $\frac{2}{3} + \frac{4}{8}$ |
| John | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| Mary | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Lindsay | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| Elizabeth | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 |
| Susan | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| Dave | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| Kate | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Patrick | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

Conjunctive Hierarchical Classes Model

| S | Row Bundles | | | P | Column Bundles | | |
|-----------|-------------|----|-----|-----------------------------|----------------|---|----|
| | I | II | III | | Columns | I | II |
| John | 0 | 1 | 1 | $\frac{5}{9} + \frac{7}{9}$ | 0 | 1 | 1 |
| Mary | 1 | 1 | 1 | $\frac{2}{3} + \frac{1}{2}$ | 1 | 0 | 1 |
| Lindsay | 0 | 0 | 1 | $\frac{3}{4} + \frac{5}{6}$ | 1 | 0 | 1 |
| Elizabeth | 1 | 0 | 1 | $\frac{6}{7} + \frac{4}{7}$ | 0 | 0 | 1 |
| Susan | 1 | 1 | 0 | $\frac{3}{8} + \frac{1}{8}$ | 0 | 1 | 0 |
| Dave | 0 | 1 | 1 | $\frac{1}{4} + \frac{2}{3}$ | 1 | 0 | 0 |
| Kate | 1 | 0 | 0 | $\frac{2}{5} + \frac{3}{4}$ | 1 | 0 | 1 |
| Patrick | 0 | 1 | 0 | $\frac{3}{5} + \frac{4}{5}$ | 0 | 0 | 1 |
| | | | | $\frac{2}{3} + \frac{4}{8}$ | 1 | 1 | 1 |

Table 2: *Bundle Parameter Estimates from Deterministic, PMD, and Bayesian HICLAS analyses for the Psychiatrist Data of Section 4.2.*

| | Deterministic | | | | PMD | | | | Bayesian HICLAS | | | |
|---|---------------|----|-----|----|-----|-----|-----|-----|-----------------|-----|-----|-----|
| | I | II | III | IV | I | II | III | IV | I | II | III | IV |
| <i>Patients (and diagnosed disorders)</i> | | | | | | | | | | | | |
| P-1 (Substance) | 0 | 1 | 1 | 0 | .41 | .45 | .69 | .26 | .45 | .65 | .79 | .02 |
| P-2 (Affective) | 1 | 1 | 0 | 0 | .74 | .41 | .27 | .23 | .91 | .62 | .27 | .00 |
| P-3 (Affective) | 1 | 1 | 0 | 0 | .63 | .35 | .34 | .31 | .87 | .45 | .26 | .07 |
| P-4 (Affective) | 0 | 1 | 0 | 0 | .55 | .61 | .41 | .30 | .39 | .82 | .70 | .01 |
| P-5 (Substance) | 0 | 1 | 1 | 0 | .53 | .65 | .61 | .27 | .37 | .92 | .91 | .02 |
| P-6 (Affective) | 0 | 1 | 0 | 0 | .37 | .51 | .25 | .20 | .20 | .66 | .37 | .00 |
| P-7 (Substance) | 0 | 0 | 1 | 0 | .33 | .40 | .73 | .34 | .12 | .33 | .74 | .02 |
| P-8 (Substance) | 0 | 0 | 0 | 0 | .26 | .22 | .29 | .15 | .33 | .11 | .07 | .00 |
| P-9 (Psychotic) | 1 | 1 | 0 | 1 | .51 | .39 | .27 | .68 | .87 | .50 | .48 | .99 |
| P-10 (Affective) | 1 | 1 | 0 | 0 | .74 | .54 | .35 | .25 | .91 | .78 | .66 | .01 |
| P-11 (Affective) | 0 | 1 | 0 | 0 | .28 | .45 | .32 | .26 | .09 | .45 | .69 | .02 |
| P-12 (Psychotic) | 0 | 0 | 0 | 1 | .22 | .29 | .31 | .78 | .02 | .18 | .50 | .99 |
| P-13 (Substance) | 0 | 0 | 1 | 0 | .22 | .28 | .72 | .22 | .04 | .24 | .63 | .02 |
| P-14 (Psychotic) | 0 | 1 | 0 | 1 | .36 | .40 | .30 | .72 | .22 | .68 | .64 | .99 |
| P-15 (Psychotic) | 0 | 1 | 0 | 1 | .42 | .36 | .29 | .52 | .44 | .55 | .36 | .71 |
| P-16 (Psychotic) | 0 | 0 | 0 | 1 | .21 | .28 | .28 | .75 | .02 | .18 | .49 | .99 |
| P-17 (Substance) | 0 | 0 | 1 | 0 | .21 | .28 | .58 | .23 | .04 | .26 | .71 | .02 |
| P-18 (Substance) | 0 | 0 | 1 | 1 | .19 | .26 | .33 | .64 | .06 | .18 | .32 | .99 |
| P-19 (Affective) | 1 | 0 | 0 | 0 | .37 | .24 | .16 | .15 | .65 | .15 | .08 | .00 |
| P-20 (Psychotic) | 1 | 1 | 0 | 1 | .75 | .67 | .45 | .69 | .90 | .86 | .71 | .88 |
| P-21 (Affective) | 1 | 1 | 0 | 0 | .55 | .49 | .28 | .23 | .80 | .66 | .66 | .01 |
| P-22 (Affective) | 1 | 1 | 0 | 0 | .72 | .36 | .24 | .19 | .89 | .63 | .26 | .00 |
| P-23 (Affective) | 0 | 1 | 0 | 0 | .36 | .63 | .29 | .24 | .23 | .75 | .72 | .01 |
| P-24 (Affective) | 0 | 1 | 0 | 0 | .54 | .52 | .27 | .20 | .36 | .73 | .34 | .00 |
| P-25 (Other) | 1 | 1 | 0 | 0 | .72 | .36 | .24 | .19 | .89 | .62 | .25 | .00 |
| P-26 (Affective) | 1 | 1 | 0 | 0 | .63 | .49 | .31 | .24 | .88 | .60 | .64 | .01 |
| P-27 (Affective+Psychotic) | 1 | 1 | 0 | 0 | .72 | .49 | .29 | .23 | .92 | .80 | .68 | .01 |
| P-28 (Substance) | 0 | 1 | 1 | 0 | .36 | .54 | .54 | .42 | .25 | .77 | .87 | .13 |
| P-29 (Affective) | 1 | 1 | 0 | 0 | .74 | .54 | .35 | .25 | .91 | .80 | .68 | .01 |
| P-30 (Substance) | 0 | 0 | 1 | 0 | .20 | .25 | .47 | .34 | .04 | .24 | .61 | .14 |
| <i>Symptoms</i> | | | | | | | | | | | | |
| inappropriate affect | 0 | 1 | 0 | 1 | .34 | .63 | .55 | .72 | .10 | .38 | .75 | .76 |
| belligerence | 0 | 0 | 0 | 0 | .13 | .15 | .30 | .13 | .00 | .00 | .00 | .00 |
| agitation | 0 | 0 | 0 | 1 | .11 | .12 | .17 | .64 | .00 | .00 | .01 | .98 |
| retardation | 0 | 0 | 0 | 0 | .16 | .16 | .26 | .15 | .00 | .00 | .00 | .00 |
| lack of emotions | 0 | 0 | 0 | 0 | .10 | .10 | .11 | .16 | .00 | .00 | .00 | .00 |
| speech disorganization | 0 | 0 | 0 | 0 | .09 | .11 | .15 | .21 | .00 | .00 | .00 | .01 |
| grandiosity | 0 | 0 | 0 | 0 | .12 | .13 | .13 | .26 | .00 | .00 | .00 | .01 |
| suspicion | 0 | 0 | 0 | 1 | .10 | .11 | .12 | .66 | .00 | .00 | .01 | .99 |
| hallucination/delusion | 0 | 0 | 0 | 1 | .10 | .11 | .12 | .66 | .00 | .00 | .01 | .99 |
| overt anger | 0 | 0 | 0 | 0 | .08 | .11 | .16 | .19 | .00 | .00 | .00 | .00 |
| depression | 1 | 1 | 0 | 0 | .86 | .71 | .34 | .27 | .96 | .74 | .27 | .01 |
| anxiety | 1 | 1 | 0 | 0 | .71 | .67 | .26 | .30 | .45 | .67 | .27 | .09 |
| obsession | 0 | 0 | 0 | 0 | .08 | .10 | .10 | .10 | .00 | .00 | .00 | .00 |
| suicide | 1 | 0 | 0 | 0 | .75 | .20 | .15 | .17 | .91 | .09 | .01 | .00 |
| somatic concern | 0 | 0 | 0 | 0 | .19 | .33 | .17 | .19 | .02 | .09 | .06 | .01 |
| social isolation | 0 | 1 | 0 | 0 | .83 | .43 | .36 | .32 | .70 | .49 | .15 | .02 |
| daily routine impairment | 0 | 1 | 1 | 1 | .82 | .80 | .77 | .71 | .55 | .95 | .97 | .55 |
| leisure time impairment | 0 | 1 | 1 | 1 | .80 | .79 | .77 | .61 | .50 | .96 | .97 | .45 |
| antisocial behavior | 0 | 0 | 0 | 0 | .08 | .10 | .26 | .11 | .00 | .00 | .00 | .00 |
| substance abuse | 0 | 0 | 1 | 0 | .11 | .14 | .66 | .18 | .03 | .08 | .27 | .01 |
| disorientation | 0 | 0 | 0 | 0 | .07 | .08 | .12 | .10 | .00 | .00 | .00 | .00 |
| memory impairment | 0 | 0 | 0 | 0 | .08 | .09 | .19 | .11 | .00 | .00 | .00 | .00 |

Note. The psychiatrist's final diagnosis for each patient is added between parenthesis.