Bayesian Inference with Probability Matrix Decomposition Models

Michel Meulders
Paul De Boeck
Iven Van Mechelen
Katholieke Universiteit Leuven

Andrew Gelman
Columbia University

Eric Maris
Katholieke Universiteit Nijmegen

Probability Matrix Decomposition models may be used to model observed binary associations between two sets of elements. More specifically, to explain observed associations between two elements, it is assumed that B latent Bernoulli variables are realized for each element and that these variables are subsequently mapped into an observed data point according to a prespecified deterministic rule. In this paper, we present a fully Bayesian analysis for the PMD model making use of the Gibbs sampler. This approach is shown to yield three distinct advantages: (a) in addition to posterior mean estimates it yields $(1 - \alpha)\%$ posterior intervals for the parameters, (b) it allows for an investigation of hypothesized indeterminacies in the model's parameters and for the visualization of the best possible reduction of the posterior distribution in a low-dimensional space, and (c) it allows for a broad range of goodness-of-fit tests, making use of the technique of posterior predictive checks. To illustrate the approach, we applied the PMD model to opinions of respondents of different countries concerning the possibility of contracting AIDS in a specific situation.

Probability Matrix Decomposition (PMD) models have been introduced by Maris, De Boeck and Van Mechelen (1996) as a method of data analysis for two-way two-mode frequency data (using Tucker's terminology of types of data, see Carroll & Arabie, 1980). In general, the rows and the columns of the data matrix refer to different types of elements, denoted as *objects* and *attributes*. In most applications of PMD models, the observed frequencies reflect the number of raters who have the opinion that objects and attributes are related. However, one may also consider other types of repeated measurements, such as the judgments of one rater at different occasions. In these cases we use the term *replications* for repeated measurements. By making a specific choice of objects, attributes, and replications a wide variety of phenomena in educational and behavioral research may be investigated.

For instance, Maris, De Boeck, and Van Mechelen (1996) used PMD models to study the process of psychiatric diagnosis. In this study, different clinicians had to judge whether patients have a certain symptom. Candel and Maris (1997) illustrated the use of PMD models in marketing research and made a comparison with other methods of data analysis for two-way two-mode frequency data, namely, latent class analysis and correspondence analysis. To illustrate the use of PMD models in marketing research they asked respondents to judge whether products have a certain attribute. De Bonis, De Boeck, Pérez-Diaz, and Nahas (1999) used PMD models to study emotion perception in facial expressions. These authors asked raters to judge whether they perceive a certain emotion in a photographed facial expression. In addition, PMD models may also be useful in personality research. For instance, in order to analyze the hostile behavior of people in different frustrating situations persons were asked to judge whether they would display a certain hostile reaction in a certain frustrating situation (Meulders, De Boeck, & Van Mechelen, 2000).

Furthermore, PMD models could also be used to analyze some broad categories of judgments that are often studied in educational and behavioral research. For instance, they could be used in the context of social network analysis (Wasserman & Faust, 1994) to explain the ties between members of a network on some social relation (e.g. friendship, communicating, etc.). The two-way two-mode data matrix to be used in such analysis may be obtained by asking each actor in the network to report all those to whom he or she is tied, on the social relationship in question, and by afterwards counting the number of ties between each pair of actors. Another research area where PMD models could be meaningfully applied is that of consumer behavior. More specifically, one could record the number of times consumers buy a certain product and analyze the resulting consumers by products matrix. As a final example, PMD models could be used in the context of educational measurement to study the extent to which pupils master certain attainment targets. In this case the data may reflect the number of judges who indicate that a particular pupil masters a certain attainment target.

PMD models assume a two-fold process to explain associations between objects and attributes: First, when judging the association between elements (objects and attributes), it is assumed that raters may (or may not) perceive each of a set of *latent features* in each element. Second, it is assumed that judgments regarding the association of two elements are the result of combining the feature patterns of these elements according to a specific rule. In general, all kinds of rules may be considered but in the literature especially deterministic rules that are interesting from a substantive point of view have been proposed (see Maris, 1995; Van Mechelen, De Boeck, & Rosenberg, 1995; Maris, De Boeck, & Van Mechelen, 1996). For instance, according to a *disjunctive communality* rule, the perception of a certain feature in both elements (object and attribute) is a sufficient condition for the elements to be associated, whereas, with a *conjunctive dominance* rule it is necessary that all the features that are perceived in one element (object) are also perceived in the other element (attribute).

To make this abstract description of the PMD model more specific it suffices to elaborate on the nature of the latent features and the mapping rule in some of the previously mentioned applications. In the application on psychiatric diagnosis, the latent features were conceived as psychiatric syndromes. This means that clinicians, when judging whether a patient has a symptom, covertly classify the patient according to which syndromes he or she has and also covertly classify the symptom as to which syndromes it belongs to. Furthermore, a disjunctive mapping rule was proposed which means that a patient has a particular symptom according to a clinician if the patient has at least one syndrome which also contains that symptom.

In the application on the facial perception of emotions, the latent features represent properties of the face (e.g., the eyes of a fearful face, the mouth of a happy face). Applying a PMD model to these data, it is assumed that, when judging whether they perceive a certain emotion in a certain facial expression, raters may (or may not) perceive each of a set of relevant latent features in the face. In addition, it is assumed that in order to be perceived in a facial expression, emotions may require the perception of certain features. Furthermore, a conjunctive rule was used which means that an emotion is only perceived in a face if all the features that are required by the emotion are also perceived in the face.

When using the PMD model for statistical inference, the following three problems need to be solved: (a) parameter estimation, (b) model identifiability and (c) checking the model's goodness of fit.

For the first problem, Maris et al. (1996) describe an EM algorithm to obtain maximum likelihood estimates and posterior mode estimates of the parameters of the PMD model, making use of the fact that the distribution of the augmented data has a simple structure. A method to estimate the standard errors of the parameters, however, is not presented by these authors.

The second problem, model identifiability, is still to be investigated. It concerns the existence of trade-off relations in the model's parameters as well as the

uniqueness of the global maximum: First, the problem of possible indetermina-
cies or trade-off relations in the parameters that result from the application of the
model to the data at hand is still unsolved. More specifically, such indetermina-
cies imply the existence of different solutions with the same posterior density
(e.g. a ridge in the surface of the posterior density function). Secondly, no
analytical results are available from which it can be derived that the posterior
density function has a single global maximum, which means that, in general, it
may be multimodal. That is, local maxima having a different posterior density
may exist.

The third problem, checking the fit of the model, requires the development of
one or more goodness-of-fit statistics with a known distribution under the
model. Goodness-of-fit statistics that measure the absolute fit of the model have
not yet been developed for the PMD model. To test the relative fit of models
with different numbers of features, a standard likelihood ratio test statistic is not
readily applicable because its asymptotic distribution is unknown.

In this paper we will show that the three previously mentioned problems can
be solved within a Bayesian framework:

1. Regarding the estimation problem, a Gibbs sampling algorithm (Gelfand
    & Smith, 1990; Geman & Geman, 1984; Smith & Roberts, 1993; Tanner
    & Wong, 1987), also labeled "chained data augmentation" by Tanner
    (1996, p. 137), will be proposed to compute a sample of the entire
    posterior distribution of the parameters. In this way the posterior mean
    estimates of the parameters as well as $100*(1 - \alpha)\%$ posterior intervals
    can be computed. Like the EM algorithm, a Gibbs sampling algorithm
    that involves the use of latent variables will appear to gain computational
    advantage from the fact that the augmented posterior has a simple struc-
    ture.
2. Concerning the identifiability of the model, we will first show how a
    multivariate analysis technique like principal components analysis can be
    helpful in identifying in a confirmatory way hypothesized indetermina-
    cies which influence the pattern of posterior covariances in a systematic
    manner. Second, we show that from a purely exploratory point of view,
    the results of principal components analysis may be used to visualize
    projections of the posterior distribution in a low-dimensional space.
3. With respect to model checking, we will indicate how the posterior
    sample may be used to assess the fit of the model with the technique of
    posterior predictive checks (Gelman, Carlin, Stern, & Rubin, 1995; Gel-
    man, Meng, & Stern, 1996). More specifically, we will discuss how this
    approach can be used to assess the general goodness of fit of the model
    and to determine the dimensionality of the PMD model, that is, the
    number of features.

In the following paragraphs we will first briefly recapitulate the model and the
estimation of its parameters with the EM algorithm. Second, we will deal with

the Bayesian estimation of the parameters via a Gibbs sampling algorithm and we will show how the obtained posterior sample can be used to solve the three problems mentioned above. Third, we will illustrate the approach with data on opinions of respondents in different countries concerning the possibility of contracting AIDS in a specific situation.

Probability Matrix Decomposition Models

Model

In general, the observed data for a PMD model are frequencies of 0/1 responses regarding the associations between objects and attributes. Stated differently, the data can be conceived as a two-way two-mode array with (sums of) multiple replications of binary associations in each cell. In a common application, the objects could be persons, groups, countries, and so on; the attributes could be items, characteristics, opinions, and so on. Finally, the replications could be based on several raters, time points, situations and so on. The observed variable Y_i^{oa} with realizations y_i^{oa}, which denotes the binary association between object o and attribute a ($o = 1, \ldots, O$; $a = 1, \ldots, A$) at the ith replication ($i = 1, \ldots, I_{oa}$), equals 1 if object o has attribute a at the ith replication and 0 otherwise. The PMD model has two parts, which will now be successively discussed, assuming that raters constitute the replication mode.

1. PMD models explain the associations between objects and attributes through binary *latent responses* at both the object and the attribute side. That is to say, an observation y_i^{oa} is assumed to be based on realizations s_{ai}^{ob} and p_{oi}^{ab} of latent binary variables S_{ai}^{ob} and P_{oi}^{ab} ($b = 1, \ldots, B$). In general, a latent response variable indicates for objects and attributes whether they have the corresponding feature at replication i. In particular, s_{ai}^{ob} equals 1 if object o has feature b according to rater i when he or she is judging the association between object o and attribute a, and 0 otherwise. Likewise p_{oi}^{ab} equals 1 if attribute a has feature b according to rater i when he or she is judging the association between o and a, and 0 otherwise. In contrast to the corresponding deterministic models (De Boeck & Rosenberg, 1988; Van Mechelen, De Boeck, & Rosenberg, 1995), the PMD model assumes that having a feature is essentially a probabilistic process. More specifically, it is assumed that the latent response variables $S_{ai}^{ob} \sim$ Bern (ρ_{ob}) and similarly $P_{oi}^{ab} \sim$ Bern (τ_{ab}).

2. The latent responses s_{ai}^{ob} and p_{oi}^{ab} ($b = 1, \ldots, B$) are mapped into the observed responses y_i^{oa} according to some prespecified rule. Maris et al. (1996) consider several deterministic rules that map the realized latent responses into the observed response. In this paper we will use one of the rules, namely a disjunctive communality rule, which is defined as follows:

$$Y_i^{oa} = 1 \Leftrightarrow \exists b : S_{ai}^{ob} = P_{oi}^{ab} = 1 \ (b = 1, \ldots, B).$$

157

Maris et al. (1996) show that this rule leads to the following probability that Y_i^{oa} equals 1:

$$P(Y_i^{oa} = 1 \mid \boldsymbol{\rho}_o, \tau_a) = 1 - \prod_b (1 - \rho_{ob}\tau_{ab}). \tag{1}$$

Estimation

Before summarizing the estimation of the parameters with the EM algorithm, we introduce some notation. The number of observed 1— and 0-responses with respect to the pair (o,a) is denoted by f_1^{oa} and f_0^{oa}, respectively. Let **Y** be the vector of observed response variables Y_i^{oa} ($o = 1, \ldots, O; a = 1, \ldots, A; i = 1, \ldots, I_{oa}$) and let **S** and **P** contain the latent response variables S_{ai}^{ob} and P_{oi}^{ab} respectively ($o = 1, \ldots, O; a = 1, \ldots, A; b = 1, \ldots, B; i = 1, \ldots, I_{oa}$). Furthermore, let **Z** = (**S**, **P**) comprise the total set of latent response variables. Finally we define $\boldsymbol{\theta} = (\boldsymbol{\rho}, \tau)$ as the vector of all the parameters in the model, with the vector $\boldsymbol{\rho}$ containing the probabilities ρ_{ob} that object o has feature b ($o = 1, \ldots, O; b = 1, \ldots, B$) and the vector τ containing the probabilities τ_{ab} that attribute a has feature b ($a = 1, \ldots, A; b = 1, \ldots, B$).

The observed posterior can now be expressed as follows:

$$p(\boldsymbol{\theta} \mid \mathbf{Y}) \propto p(\boldsymbol{\theta})p(\mathbf{Y} \mid \boldsymbol{\theta}) = p(\boldsymbol{\theta}) \prod_o \prod_a P(Y_i^{oa} = 1 \mid \boldsymbol{\theta})^{f_1^{oa}} P(Y_i^{oa} = 0 \mid \boldsymbol{\theta})^{f_1^{oa}}.$$

In the above formula we still have to make a choice with respect to the joint prior distribution $p(\boldsymbol{\theta})$. One possibility is to assume that the individual parameters θ_j are iid with $p(\theta_j) \propto 1$ so that the observed posterior is proportional to the likelihood of the data. Maris et al. (1996) indicate, however that this choice for $p(\boldsymbol{\theta})$ may be problematic in that, depending on the particular set of observations, posterior mode estimates may not exist in the interior of the parameter space. Maris et al. (1996) also propose an alternative prior distribution that guarantees the existence of posterior mode estimates in the interior of the parameter space, namely $p(\theta_j) \overset{iid}{\sim}$ Beta $(\theta_j \mid 2,2)$. The joint prior distribution then is proportional to:

$$p(\boldsymbol{\theta}) \propto \prod_o \prod_b \rho_{ob}(1 - \rho_{ob}) \prod_a \prod_b \tau_{ab}(1 - \tau_{ab}). \tag{2}$$

We notice that the Beta$(\theta_j \mid 2, 2)$ prior, which is specified for purely technical reasons here, a priori assumes that parameters are moderate rather than extreme. In particular, this prior is a concave function on [0, 1] with a mode at 0.5. Furthermore, the Beta prior is also conjugate, which offers the advantage of being interpretable as additional data. In particular, the Beta $(\theta_j \mid 2, 2)$ prior adds one latent observation of each type (0 and 1) to the entire set of latent observations regarding θ_j.

The maximization of the observed posterior can be accomplished via an EM algorithm (Dempster, Laird, & Rubin, 1977). In general, such an algorithm consists of two steps: an E step (Expectation step) and an M step (Maximization step). In the E step of the $(m + 1)$th iteration, the expectation of the log augmented posterior is computed with respect to the distribution of the latent data, **Z**, conditional on the observed data, **Y**, and the current guess of the posterior mode ($\boldsymbol{\theta}^{(m)}$). Formally:

$$Q(\theta, \theta^{(m)}) = \int_Z \ln[\ p(\theta|Y,Z)]p(Z|\theta^{(m)},Y)dZ.$$

In the M step this function is maximized with respect to θ. The algorithm has converged to a stationary point (local maximum or saddle point) if

$$\frac{\partial p(\theta|Y)}{\partial \theta}\bigg|_{\hat{\theta}} \approx 0.$$

Bayesian Analysis with the PMD Model: Estimation

A Gibbs sampling algorithm, also called "chained data augmentation" (Tanner, 1996, p. 137), can be used to obtain a sample from an observed posterior distribution. Similar to the EM algorithm, this approach is based on the mathematical tractability of the augmented posterior relative to the complexity of the observed posterior.

Given the mth simulation draw $\theta^{(m)}$, the $(m + 1)$th iteration of the chained data augmentation algorithm consists of the following two steps:

1. Imputation step: generate latent data $Z^{(m+1)}$ from the conditional predictive distribution, $p(Z|\theta^{(m)}, Y)$.
2. Posterior step: draw a simulation $\theta^{(m+1)}$ of the parameter vector from the augmented posterior distribution, $p(\theta|Y, Z^{(m+1)})$.

The implementation of both steps is straightforward for the PMD model.

1. To summarize the imputation step, we first introduce some notation. Let $Z_i^{oa} = (S_{ai}^{ol}, \ldots, S_{ai}^{oB}, P_{oi}^{al}, \ldots, P_{oi}^{aB})$ and similarly $z_i^{oa} = (s_{oi}^{ol}, \ldots s_{ai}^{oB}, p_{oi}^{al}, \ldots, p_{oi}^{aB})$. Moreover, $C(Z_i^{oa}) = Y_i^{oa}$ is used to denote the mapping of latent response variables S_{ai}^{ob} and P_{oi}^{ab} $(b = 1, \ldots, B)$ into the observed response variable Y_i^{oa}. The conditional predictive distribution $P(Z_i^{oa} = z_i^{oa}|\theta^{(m)}, Y_i^{oa} = y_i^{oa})$ is defined as follows:

$$
\begin{cases}
\dfrac{P(Z_i^{oa} = z_i^{oa}|\theta^{(m)})}{P(Y_i^{oa} = y_i^{oa}|\theta^{(m)})} & \text{if } C(z_i^{oa}) = y_i^{oa} \\
0 & \text{if } C(z_i^{oa}) \neq y_i^{oa}.
\end{cases}
$$

The denominator of the first term depends on the specific type of mapping rule. For disjunctive communality it is given by (1). The numerator is equal to:

$$\prod_b (\rho_{ob}^{(m)})^{s_{ai}^{ob}}(1 - \rho_{ob}^{(m)})^{1-s_{ai}^{ob}}(\tau_{ab}^{(m)})^{p_{oi}^{ab}}(1 - \tau_{ab}^{(m)})^{1-p_{oi}^{ab}}.$$

In order to draw a vector from this discrete conditional predictive distribution we make use of the inverse cumulative distribution function. Suppose that for a data point y_i^{oa} the candidate vectors z_i^{oa} are numbered as $z_1, z_2, \ldots, z_K$. The cumulative distribution function is then easily tabulated as:

$$F(z_x) = \sum_{k=1}^x P(Z_i^{oa} = z_k|\theta^{(m)}, y_i^{oa}) \quad x = 1, \ldots, K.$$

To draw at random one of the candidates $z_x(x = 1, \ldots, K)$, we first draw a random number u from a uniform distribution $U(0,1)$ and afterwards we use $F^{-1}(u)$ to determine which of the candidates has been chosen.

2. To describe the posterior step, we first define the following statistics involving the latent response variables:

$$s_1^{ob} = \sum_{a=1}^A \sum_{i=1}^{I_{oa}} s_{ai}^{ob} \quad \text{and} \quad s_0^{ob} = \sum_{a=1}^A \sum_{i=1}^{I_{oa}} (1 - s_{ai}^{ob})$$

$$p_1^{ab} = \sum_{o=1}^O \sum_{i=1}^{I_{oa}} p_{oi}^{ab} \quad \text{and} \quad p_0^{ab} = \sum_{o=1}^O \sum_{i=1}^{I_{oa}} (1 - p_{oi}^{ab}).$$

Using this notation and taking (2) as the joint prior distribution, it is easy to see that the augmented posterior distribution is proportional to the following product:

$$p(\theta|\mathbf{Y},\mathbf{Z}) \propto \prod_o \prod_b (\rho_{ob})^{(1+s_1^{ob})}(1 - \rho_{ob})^{(1+s_0^{ob})} \prod_a \prod_b (\tau_{ab})^{(1+p_1^{ab})}(1 - \tau_{ab})^{(1-p_0^{ab})}.$$

This implies that the augmented posterior distribution is equal to a product of Beta distributions, namely:

$$p(\theta|\mathbf{Y},\mathbf{Z}) = \prod_o \prod_b \text{Beta}(\rho_{ob}|2 + s_1^{ob}, 2 + s_0^{ob}) \prod_a \prod_b \text{Beta}(\tau_{ab}|2 + p_1^{ab}, 2 + p_0^{ab}).$$

In the $(m + 1)$-th iteration of the posterior step we draw the individual parameters from their corresponding Beta distribution:

$$\rho_{ob}^{(m+1)} \sim \text{Beta}(2 + (s_1^{ob})^{(m+1)}, 2 + (s_0^{ob})^{(m+1)})$$

$$\tau_{ab}^{(m+1)} \sim \text{Beta}(2 + (p_1^{ab})^{(m+1)}, 2 + (p_0^{ab})^{(m+1)}).$$

Tanner and Wong (1987) show that, under some regularity conditions, the subsequent values $\theta^{(1)}, \theta^{(2)}, \ldots$ form a Markov chain which converges to the true posterior distribution. An important aspect in the implementation of the algorithm is the required number of iterations to approximate convergence (Gilks, Richardson, & Spiegelhalter, 1995; Cowles & Carlin, 1996). We follow the approach of Gelman and Rubin (1992) by simulating multiple chains from different starting points and judging approximate convergence based on the statistic $\hat{R}^{1/2}$, which measures the ratio of a weighted sum of between-chain variation and within-chain variation to within-chain variation, for each scalar estimand of interest. The statistic $\hat{R}^{1/2}$ may be interpreted as an estimate of the factor by which the scale of the current distribution for a parameter might be reduced if the simulations were continued in the limit $m \to \infty$. Values of $\hat{R}^{1/2}$ near 1 indicate convergence; in practice $\hat{R}^{1/2}$ smaller than 1.1 for each scalar estimand of interest is a reasonable criterion to stop the simulations (Gelman et al., 1995). In this computation it is necessary to discard the initial part of each chain to fit the target distribution more accurately. We notice that the convergence diagnostic $\hat{R}^{1/2}$ is based on normal theory approximations so that it is appropriate to

transform the scalar estimands to be approximately normally distributed. As the parameters of the PMD model are in the [0,1] interval, a logit transformation is used.

Bayesian Analysis with the PMD Model: Indeterminacies

In a classical maximum likelihood framework, an important question is whether some obtained solution corresponds to the global maximum of the likelihood. In this respect, it often cannot be shown analytically that the likelihood is unimodal. Usually, it is only possible to check whether a particular solution is a local maximum (Goodman, 1974; Formann, 1992); to gain further evidence about the optimality of a solution it is common to run the algorithm several times starting from different points in the parameter space.

A fully Bayesian analysis offers two additional possible ways to investigate indeterminacies in a model's parameters, which reflect regions of high posterior density.

First, it is possible to investigate in a *confirmatory* way to what extent certain hypothesized indeterminacies explain the posterior uncertainty in the parameters. The indeterminacies we have in mind have two characteristics: (a) they imply a specific pattern of covariation between the parameters and (b) their impact on the posterior uncertainty of the parameters depends on the data at hand. Because of the latter characteristic, they may be contrasted with indeterminacies which occur independently of the data at hand, such as, for instance, indeterminacies related to rotational freedom for the components in a principal components analysis. The latter type of indeterminacy is often trivial and should be taken into account during the parameter estimation.

To depict an overview of the joint influence of the different hypothesized indeterminacies, a principal components analysis of the posterior covariance matrix may be helpful. Furthermore, rotating the loadings of parameters towards the hypothetical pattern of covariances that is expected from the different types of indeterminacy, can help to identify the components. To rotate the loadings towards a hypothetical pattern, an orthogonal Procrustes rotation can be used (Schönemann, 1966).

Second, from a purely *exploratory* viewpoint, a fully Bayesian analysis allows us to visualize projections of the posterior distribution in low-dimensional space, by means of multivariate analysis techniques—in particular, principal components analysis. Such projections may indicate areas of high posterior density and they may also reveal whether the posterior has one or multiple modes.

We will now describe three types of indeterminacy that may exist in the parameters of the PMD model. A first type, labeled *permutation* indeterminacy, is a trivial type of indeterminacy that also occurs in mixture models. In the case of a disjunctive communality mapping rule, permutation indeterminacy between the parameters results from the fact that the left-hand side of (1) is invariant to permutation of features. Permutation indeterminacy implies that, for a PMD model with B features, the parameter space contains $B!$ identical regions of

posterior mass. A possible solution to this problem is to identify for each simulated Markov chain the region that is sampled from, based for instance, on the posterior mean estimates of the parameters. In this approach one must make sure that the simulated parameter values do not switch between different posterior regions during the simulation of one Markov chain. If the feature probabilities of a specific object/attribute differ substantially, then a permutation of the features can be identified through a visual inspection of the iteration history of the simulated parameter values for this object/attribute.

A second and less trivial type of indeterminacy, labeled *within-feature* indeterminacy, may exist between parameters of objects and attributes regarding one feature. That is to say, multiplying all the feature probabilities of objects by a constant and dividing all the feature probabilities of attributes by the same constant does not affect the likelihood of the model. Notice that the value of the constant is restricted by the condition that all parameters have to take values within the [0, 1] interval, so that in special cases this type of indeterminacy can be negligible, which is why the indeterminacy depends on the data. Within-feature indeterminacy implies a specific pattern of covariation between parameters. In particular, one may predict positive posterior covariances among the ρ_{ob}'s and among the τ_{ab}'s for each feature b, and negative covariances between ρ_{ob}'s and τ_{ab}'s, also for each feature b.

A third type of indeterminacy, labeled *between-feature* indeterminacy, may exist between feature probabilities concerning different features. In general, it is similar to the second type, although its implications for single parameters are much less straightforward. It can be described as follows. The probability of observing a one at the i-th replication in cell (o, a) is given in (1). Now it is easy to see that multiplying one term of the product, for example $(1 - \rho_{o1}\tau_{a1})$, by a constant and dividing another, for example $(1 - \rho_{o2}\tau_{a2}$, by the same constant, does not affect the likelihood of the model. Also for this type of indeterminacy, the value of the constant is restricted by the condition that all the parameters should take values in the $[0,\tau1]$ interval, so that in special cases this indeterminacy can be negligible, meaning that it depends on the data. Besides, between-feature indeterminacy also implies a specific pattern of covariation between parameters. More specifically, negative covariances between ρ_{ob}'s and τ_{ab}'s of different features may be expected.

It should be stressed that, in a specific application, the pattern of posterior covariances may be the result of both within- and between-feature indeterminacy. For example, a positive covariance between ρ_{ob}'s of one feature and τ_{ab}'s of another may result from a negative covariance between ρ_{ob}'s and τ_{ab}'s within a feature, and a negative covariance between features. As a consequence, just looking at pairwise covariances may be misleading. A principal components analysis of the posterior covariance matrix and Procrustes rotation of the loadings towards the hypothetical pattern of covariances that is expected from both types of indeterminacy, can help to identify the components. For example, in the case of two features, three components are expected: one for each feature to

reflect within-feature indeterminacy, and one contrasting both features to reflect between-feature indeterminacy. For more than two features, for each pair of features, a bipolar between-feature component can be hypothesized. Finally, it is not guaranteed that these are the only types of indeterminacy that occur in a specific application. In other words, it is still possible that other, more complex types of indeterminacy, which depend on the data at hand, may occur in the model's parameters.

Bayesian Analysis with the PMD Model: Model Checking

Besides the estimation of the parameters, an appropriate analysis also implies checking the goodness of fit of the model. In this respect, two questions are of particular importance:

1. What is the relative goodness of fit of models with different numbers of features? In other words, how much does the goodness of fit improve by adding one feature and is the improvement statistically significant? This is usually an important question in models that are used to represent the data as well as possible with a low number of dimensions.
2. What is the absolute goodness of fit of the model? More specifically, to what extent do the expected frequencies under the model approximate the observed frequencies?

Within a classical maximum likelihood framework, such model checking requires the construction of test quantities with a known distribution under the model. For the PMD model, a standard likelihood ratio test statistic cannot be used to test the relative fit of models with different numbers of features because this statistic is not asymptotically chi-square distributed. The latter is due to the fact that the null hypothesis of a PMD model with B features corresponds to a boundary of the parameter space of the alternative model with $B + r$ features (McLachlan & Basford, 1988). On the other hand, goodness-of-fit statistics that measure the absolute fit of the PMD model have not been developed yet. A valuable alternative for the construction of test quantities with a known distribution under the model is to simulate the distribution of a test quantity under the model.

Within a Bayesian framework, model checking is basically a matter of comparing observed data with data that could have been observed under the model if the actual experiment were replicated with values of the posterior distribution of θ (Gelman, Carlin, Stern, & Rubin, 1995). The reference distribution for a replicated observation Y^{rep}, also called the posterior predictive distribution, is defined as follows:

$$p(Y^{rep}|Y) = \int p(Y^{rep}|\theta)p(\theta|Y)d\theta.$$

In order to compare Y^{rep} and Y, one may define a test quantity $T(Y)$ that is a function of the data only, or a test quantity $T(Y, \theta)$ that is a function of both the data and the parameters.

Rubin (1984) defines the posterior predictive p value as the probability that $T(\mathbf{Y}^{\text{rep}})$ exceeds or equals $T(\mathbf{Y})$; that is:

ppc p — value $= P[T(\mathbf{Y}^{\text{rep}}) \geq T(\mathbf{Y})|\mathbf{Y}] = \int I[T(\mathbf{Y}^{\text{rep}}) \geq T(\mathbf{Y})]p(\mathbf{Y}^{\text{rep}}|\mathbf{Y})d\mathbf{Y}^{\text{rep}}$,

with $I[\cdot]$ being the indicator function. An extreme p value indicates that $T(\mathbf{Y})$ is unlikely to occur under the model and places doubt on the model aspect measured by $T(\cdot)$. The computation of the p value is straightforward once a sample of the posterior distribution is available. In particular, the following steps need to be performed for each draw $\boldsymbol{\theta}^{(m)}$ ($m = 1, \ldots, M$) of the observed posterior distribution:

1. Generate a replicated data set $\mathbf{Y}^{\text{rep},m}$ from $p(\mathbf{Y}|\boldsymbol{\theta}^{(m)})$.
2. Compute $T(\mathbf{Y}^{\text{rep},m})$.

Afterwards the p value is calculated as the proportion of simulated values $T(\mathbf{Y}^{\text{rep},m})$ that exceed or equal $T(\mathbf{Y})$. Beyond the actual p value, a useful graphical device consists in the location of $T(\mathbf{Y})$ in the histogram of simulated values $T(\mathbf{Y}^{\text{rep},m})$.

Meng (1994) describes the posterior predictive check approach with test quantities $T(\mathbf{Y}, \boldsymbol{\theta})$ that are a function of both data and parameters. Such quantities are labeled realized *discrepancy measures*. Gelman, Meng, and Stern (1996) elaborated this approach further and discussed several discrepancy measures to evaluate the goodness of fit of a model. In this case, the p value is defined as the probability that the replicated discrepancy measure exceeds or equals the realized discrepancy measure:

ppc p — value $= P(T(\mathbf{Y}^{\text{rep}},\boldsymbol{\theta}) \geq T(\mathbf{Y},\boldsymbol{\theta})|\mathbf{Y})$
$$= \int\int I[T(\mathbf{Y}^{\text{rep}},\boldsymbol{\theta}) \geq T(\mathbf{Y},\boldsymbol{\theta})]p(\mathbf{Y}^{\text{rep}}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}d\mathbf{Y}^{\text{rep}}.$$

To compute the p value, the following steps are required for each draw $\boldsymbol{\theta}^{(m)}$ ($m = 1, \ldots, M$) of the posterior:

1. Generate a replicated data set $\mathbf{Y}^{\text{rep},m}$ from $p(\mathbf{Y}|\boldsymbol{\theta}^{(m)})$.
2. Compute $T(\mathbf{Y}, \boldsymbol{\theta}^{(m)})$.
3. Compute $T(\mathbf{Y}^{\text{rep},m}, \boldsymbol{\theta}^{(m)})$.

The p value is subsequently calculated as the proportion of simulated values $T(\mathbf{Y}^{\text{rep},m}, \boldsymbol{\theta}^{(m)})$ that exceed or equal their counterpart $T(\mathbf{Y}, \boldsymbol{\theta}^{(m)})$. In addition, one may plot the values $T(\mathbf{Y}^{\text{rep},m}, \boldsymbol{\theta}^{(m)})$ against their counterparts $T(\mathbf{Y},\boldsymbol{\theta}^{(m)})$ ($m = 1, \ldots, M$).

For PMD models we will focus on two test quantities that measure the general goodness of fit of the model and one test quantity that can be used to decide between models with different numbers of features. A first test quantity that we propose to assess the general goodness of fit of the model is the Pearson chi-square discrepancy measure, defined as the sum of standardized squared deviations between observed frequencies and expected frequencies under the

model:

$$\chi^2(\mathbf{Y},\boldsymbol{\theta}) = \sum_o \sum_a \left\{ \frac{f_1^{oa} - \mathrm{E}(f_1^{oa}|\boldsymbol{\theta})}{[\mathrm{Var}(f_1^{oa}|\boldsymbol{\theta})]^{\frac{1}{2}}} \right\}^2,$$

with $\mathrm{Var}(f_1^{oa}|\boldsymbol{\theta}) = [\mathrm{E}(f_1^{oa}|\boldsymbol{\theta})\mathrm{E}(f_0^{oa}|\boldsymbol{\theta})]/I_{oa}$, and in which I_{oa} is the number of replications for pair (o, a), as defined earlier. Another test quantity that serves the same purpose is the likelihood ratio chi-square discrepancy measure:

$$L^2(\mathbf{Y},\boldsymbol{\theta}) = 2 \sum_o \sum_a \left\{ f_1^{oa} \log \left[\frac{f_1^{oa}}{\mathrm{E}(f_1^{oa}|\boldsymbol{\theta})} \right] + f_0^{oa} \log \left[\frac{f_0^{oa}}{\mathrm{E}(f_0^{oa}|\boldsymbol{\theta})} \right] \right\}.$$

The Pearson chi-square and the likelihood ratio chi-square discrepancy measure are of the discrepancy measure type as the expected frequencies $\mathrm{E}(f_1^{oa}|\boldsymbol{\theta})$ and $\mathrm{E}(f_0^{oa}|\boldsymbol{\theta})$ are a function of the parameters. In particular, the expected frequency of one and zero responses in cell (o, a) may be computed as $I_{oa}P(Y_i^{oa} = 1|\boldsymbol{\theta})$ and $I_{oa}P(Y_i^{oa} = 0|\boldsymbol{\theta})$, respectively, with the probability of observing a one or a zero depending on the specific type of mapping rule that is used (see (1)).

The likelihood ratio chi-square discrepancy measure is also a building block for a measure that can be used to choose between models with different numbers of features. That is to say, to test a model with B_r features against a model with B_u features, we propose the test quantity:

$$L_{\mathrm{diff}}^2(\mathbf{Y}) = L_4^2(\mathbf{Y},\hat{\boldsymbol{\theta}}) - L_u^2(\mathbf{Y},\hat{\boldsymbol{\theta}}), \tag{3}$$

which is a function of the data only as dependence on the parameters is eliminated by substituting the posterior mode estimate $\hat{\boldsymbol{\theta}}$ at the right hand side.

In order to simulate the reference distribution of (3) we generate M-replicated data sets under the restricted model and compute for each data set the posterior mode by running the EM algorithm 10 times and choosing the solution with the highest posterior density. Afterwards we compute the quantity $L_{\mathrm{diff}}^2(\mathbf{Y}^{\mathrm{rep},m}) = L_r^2(\mathbf{Y}^{\mathrm{rep},m}, \hat{\boldsymbol{\theta}}^{(m)}) - L_u^2(\mathbf{Y}^{\mathrm{rep},m}, \hat{\boldsymbol{\theta}}^{(m)})$ $(m = 1, \ldots, M)$ for each replicated data set. We typically simulate 500 values of $L_{\mathrm{diff}}^2(\mathbf{Y}^{\mathrm{rep}})$, since finding the posterior mode for each data set is computationally intensive. Finally, we note that a quantity similar to that in (3) was used by Rubin and Stern (1994) to determine the number of latent classes in latent class analysis.

Example

As an illustration of the approach, the PMD model is now applied to real data on opinions of respondents of different countries concerning the possibility of contracting AIDS in a specific situation. The data were supplied by the Zentralarchiv für Empirische Sozialforschung at Cologne (Reif & Melich, 1992). The raw data are the answers from 23,397 respondents in 13 countries with respect to 10 items in the format: "In your opinion, in each of the following situations, can AIDS be contracted . . . yes, possibly or not?". Table 1 contains a description of the situations. Since the PMD model can only be used to explain

TABLE 1
Ten Situations in which AIDS Might Possibly be Contracted

Situation
1	Eating a meal prepared by someone who has AIDS or is HIV positive
2	Handling objects touched by someone who has AIDS or is HIV positive
3	Drinking from a glass which is used by someone who has AIDS or is HIV positive
4	Sitting on the same toilet seat as someone who has AIDS or is HIV positive
5	Being injected with a needle which has been used by someone who has AIDS or is HIV positive
6	Receiving blood from someone who has AIDS or is HIV positive
7	Shaking the hand of someone who has AIDS or is HIV positive
8	Kissing the mouth of someone who has AIDS or is HIV positive
9	Having sex with someone who has AIDS or is HIV positive
10	Taking care of someone who has AIDS or is HIV positive

TABLE 2
Frequency of Respondents in 13 Countries with the Opinion that AIDS can be Contracted in 10 Situations (Random Sample with N = 50)

Country	Situation									
	Eating	Object	Glass	Toilet	Needle	Blood	Hand	Kissing	Sex	Care
France	12	10	24	25	49	50	7	24	50	24
Belgium	11	6	21	21	50	50	4	26	50	23
Netherlands	9	4	15	14	50	50	4	26	48	14
Germany	18	10	23	26	50	50	10	38	49	30
Italy	11	6	17	24	50	50	4	28	50	24
Luxemburg	5	5	21	20	50	50	6	25	50	15
Denmark	1	5	14	13	50	50	0	24	49	13
Ireland	15	11	22	27	50	50	8	36	50	16
Great Britain	11	8	26	14	49	49	3	34	49	19
Northern Ireland	7	6	12	16	48	48	5	32	48	17
Greece	13	9	25	27	50	50	10	38	50	25
Spain	20	17	29	29	48	48	14	40	48	21
Portugal	29	24	32	38	48	49	24	45	49	34

binary associations between countries and situations, the raw data are dichoto-mized. More specifically, for each item, the categories "possibly" and "yes" are combined, since both imply that there is risk involved. In order to illustrate the modeling in a small-sample scenario, we analyze here a random subsample of 50 respondents per country. Table 2 contains the resulting data set. The frequencies

TABLE 3

Posterior Mean Estimate and 95% Posterior Interval (PI) of the Feature Probabilities for a One-feature Model

Country	Mean	95% PI	Situation	Mean	95% PI
France	.980	[.951, .997]	Eating	.258	[.224, .293]
Belgium	.984	[.956, .998]	Object	.193	[.163, .225]
Netherlands	.958	[.913, .988]	Glass	.445	[.407, 484]
Germany	.986	[.967, .998]	Toilet	.466	[.426, .505]
Italy	.985	[.960, .998]	Needle	.995	[.986, .999]
Luxemburg	.980	[.946, .997]	Blood	.996	[.988, .999]
Denmark	.955	[.902, .992]	Hand	.159	[.132, .189]
Ireland	.988	[.968, .998]	Kissing	.657	[.618, .694]
Great Britain	.967	[.931, .991]	Sex	.994	[.986, .999]
Northern Ireland	.934	[.884, .971]	Care	.436	[.398, .475]
Greece	.989	[.971, .999]			
Spain	.969	[.942, .989]			
Portugal	.984	[.967, .996]			

in this table indicate the number of respondents in a country with the opinion that AIDS can be contracted in a specific situation.

Estimation

The one-feature and the two-feature model are estimated using the Gibbs sampling algorithm. For both models five independent chains are simulated using random starting points generated from a uniform distribution $U(0, 1)$. The algorithm is stopped if the statistic $\hat{R}^{1/2}$, computed on the second halves of the chains, is smaller than 1.1 for each logit transformed parameter. For the one- and two-feature models this occurs after 2,000 and 3,000 iterations, respectively. $\hat{R}^{1/2}$ values for the parameters of these models are in the range [1.00, 1.01] and [1.00, 1.09], respectively. Finally, for each model a sample of 5,000 draws is constructed by taking 1000 evenly spaced draws from the second halves of the five simulated chains.

Table 3 and Table 4 show the posterior mean estimates and the 95% posterior intervals for the parameters of the one-feature model and the two-feature model, respectively. The estimates are probabilities, to be interpreted as follows: The feature probability for a country equals the probability that respondents in that country have the opinion associated with the feature. On the other hand, the feature probability for a situation equals the *conditional probability* that respondents agree that the situation is a risk, given they have the opinion associated with the feature.

In the one-feature model (see Table 3), the feature probabilities for the situations can be interpreted as reflecting the objectively true view on which situations could lead to AIDS. "Being injected with a needle which has been used by someone with AIDS", "receiving blood from someone with AIDS" or

TABLE 4
Posterior Mean Estimate and 95% Posterior Interval (PI) of the Feature Probabilities for a Two-feature Model

Country	Feature 1		Feature 2		Situation	Feature 1		Feature 2	
	Mean	95% PI	Mean	95% PI		Mean	95% PI	Mean	95% PI
France	.973	[.934, .996]	.398	[.277, .530]	Eating	.050	[.010, .109]	.531	[.409, .667]
Belgium	.982	[.952, .998]	.305	[.191, .437]	Object	.057	[.018, .106]	.358	[.243, .475]
Netherlands	.972	[.933, .995]	.170	[.075, .277]	Glass	.283	[.211, .356]	.542	[.376, .707]
Germany	.967	[.915, .994]	.591	[.452, .724]	Toilet	.251	[.176, .326]	.687	[.540, .829]
Italy	.983	[.950, .998]	.322	[.204, .451]	Needle	.994	[.983, .999]	.783	[.550, .949]
Luxemburg	.984	[.954, .998]	.213	[.112, .327]	Blood	.994	[.984, .999]	.826	[.528, .969]
Denmark	.982	[.953, .997]	.058	[.009, .132]	Hand	.021	[.003, .055]	.351	[.261, .449]
Ireland	.978	[.937, .997]	.479	[.344, .617]	Kissing	.464	[.384, .534]	.846	[.719, .962]
Great Britain	.962	[.916, .991]	.319	[.201, .445]	Sex	.991	[.976, .999]	.857	[.625, .982]
Northern Ireland	.942	[.893, .978]	.222	[.109, .350]	Care	.272	[.195, .346]	.545	[.380, .710]
Greece	.976	[.929, .997]	.544	[.407, .692]					
Spain	.871	[.740, .952]	.715	[.555, .846]					
Portugal	.842	[.671, .952]	.945	[.846, .993]					

"having sex with someone with AIDS" have a high probability for this feature: .995, .996, and .994, respectively; they are in fact also the true risks. Other situations, like "handling objects touched by someone with AIDS", "eating a meal prepared by someone with AIDS", "shaking the hand of someone with AIDS" have a low probability for this feature: .19, .26, and .16, respectively, and they refer to beliefs not corroborated by scientific knowledge. Finally "kissing the mouth of someone with AIDS" has a moderate probability, namely .66, and also from an objective scientific point of view, implies a slight risk.

Contrary to the large differences between the feature probabilities for the situations, in the one-feature model there are only small differences between the feature probabilities of the countries, meaning that respondents in different countries have approximately the same opinion about the possibility of contracting AIDS in a situation. As will be explained in the section of model checking, this prediction does not fit the observed data. As a matter of fact, the one-feature model is to be rejected against the two-feature model, whereas the latter model does sufficiently account for the data.

In the two-feature model (see Table 4), the first feature can be interpreted as reflecting the "correct opinion". Situations with high probabilities for this first feature correspond to the true risks. The second feature reflects an "alternative opinion" about the possibilitiy of contracting AIDS in a situation. More specifically, the true risks are underestimated and some situations are considered risky whereas objectively this is not the case. The situations "being injected with a needle which has been used by someone with AIDS", "receiving blood from someone with AIDS" and "having sex with someone with AIDS" have a lower probability for this feature to carry AIDS (.78, .83 and .86), whereas "sitting on the same toilet as someone with AIDS", "kissing the mouth of someone with AIDS" and "eating a meal prepared by someone with AIDS" have a higher probability for this feature (.69, .85, .53, respectively).

The probability of a country for the first feature can now be interpreted as the probability that respondents in that country have the opinion that is in accordance with the facts and the probabililty of a country for the second feature can be interpreted as the probability that respondents in that country have the alternative opinion. The estimates for the two-feature model show that respondents in most countries have a high probability of having the correct opinion. Exceptions are respondents in Portugal and Spain, who have a slightly lower probability to have this opinion (.84 and .87, respectively). In contrast, country probabilities for the alternative opinion vary from rather low for Denmark (.06) to rather high for Portugal (.95).

It should be clarified now that having an opinion means responding to items on the basis of the opinion in question, and that respondents can combine opinions in responding, as the probabilities do not add up to 1.0 over features. As explained earlier, the combination is of a disjunctive type. Therefore, given the same probability for the correct opinion, countries with a higher probability for the alternative opinion will tend to consider situations with a high probability

for the alternative opinion as being more risky. In other words, countries having a high probability for both opinions differ from countries having only a high probability for the correct opinion in that they tend to consider safe situations as being more risky. In this respect, it must be noticed that the rather high probabilities of the true risks "needle", "blood" and "sex" for the feature reflecting the alternative opinion are not problematic, but only imply that respondents having the alternative opinion also consider true risks as being risky. As an example, consider the probabilities that respondents in Denmark (having a high probability for the correct opinion (.98), and a low probability for the alternative opinion (.06)) and Greece (having a high probability for the correct opinion (.98) and a moderate probability for the alternative opinion (.54)) consider "receiving blood from someone with AIDS" or "sitting on the same toilet as someone with AIDS" as being a risk: Respondents from both Denmark and Greece have a probability of .98 to consider the former situation risky, whereas respondents in Greece have a somewhat higher probability (.53) than respondents in Denmark (.28) to consider the latter situation a risk.

The 95% posterior intervals of most parameters (see Table 4) are rather small: For the first and the second feature, the median range of the 95% posterior intervals equals .07 and .25, respectively. Some exceptions having a somewhat larger posterior interval are the parameter of Portugal for the first feature ([.67, .95]) and the parameters of "needle", "blood", and "sex" for the second feature ([.55, .95], [.53, .97] and [.63, .98], respectively). In the second on the identifiability of the model we will see that these larger posterior intervals may be partly explained by hypothesized indeterminacies in the model's parameters. However, as will be explained, a substantial part of the uncertainty in these parameters must be due to other more specific sources.

Indeterminacies

First, the matrix of posterior covariances between parameters is analyzed with a principal components analysis. The first seven components of this analysis account for 17.3%, 12.6%, 7.9%, 6.4%, 5.4%, 5.0% and 4.6% of the variance, respectively. A scree test (Cattell, 1966) indicates that a model with four components is appropriate. Secondly, an orthogonal Procrustes rotation is used to rotate the loadings of parameters on the four components towards the hypothetical structure that is expected under between- and within-feature indeterminacy. In specifying the hypothetical loadings, we make sure that the proportion of explained variance V_{θ_j} of a parameter θ_j is the same as in the unrotated solution.

Table 5 contains the hypothetical loadings of object and attribute parameters for the case of two features; three components are distinguished, one for between-feature indeterminacy and two for within-feature indeterminacy. For the component reflecting between-feature indeterminacy, the hypothetical loadings of parameters concerning the same feature have the same sign, unlike parameters concerning different features. For a component reflecting within-feature indeterminacy for a given feature, the hypothetical loadings of object and

TABLE 5

*Hypothetical Loadings of Parameters on Components that Reflect
Between-and Within-feature Indeterminacy*

Parameter	Between	Within	
		Feature 1	Feature 2
ρ_{o1}	$\sqrt{\dfrac{V_{\rho o1}}{2}}$	$-\sqrt{\dfrac{V_{\rho o1}}{2}}$	0
ρ_{o2}	$-\sqrt{\dfrac{V_{\rho o2}}{2}}$	0	$-\sqrt{\dfrac{V_{\rho o2}}{2}}$
τ_{a1}	$\sqrt{\dfrac{V_{\tau a1}}{2}}$	$\sqrt{\dfrac{V_{\tau a1}}{2}}$	0
τ_{a2}	$-\sqrt{\dfrac{V_{\tau a2}}{2}}$	0	$\sqrt{\dfrac{V_{\tau a2}}{2}}$

attribute parameters concerning that feature have opposite signs; the parameters
regarding the other feature having zero loadings on this component. It is as-
sumed that within-feature indeterminacy independently occurs for each of the
two features.

Table 6 shows the congruence between the hypothetical loadings and the
loadings that are obtained after Procrustes rotation. For the first and the third
component the congruence coefficients are high (.83 and .93, respectively).
Therefore we may conclude that these components reflect between-feature inde-
terminacy and within-feature indeterminacy for the second feature, respectively.
On the other hand, the congruence for the second component is only moderate
(.30). Therefore, the interpretation of this component as within-feature indeter-
minacy for the second feature is problematic. The four rotated components
account for 12.5%, 6.7%, 16.8% and 8.0% of the variance, respectively. Hence,
especially the first and the third component seem important, and precisely these
two could be easily identified as expected types of indeterminacy.

Two comments may be added to clarify the only moderate congruence coeffi-
cient concerning within-feature indeterminacy for the first feature. First, from
the country and the situation parameter values of the first feature, it may be
concluded that there is not much room for within-feature indeterminacy since
parameters, at both sides, that is, of most countries (except for Portugal and
Spain) and of some situations (i.e., "needle", "blood", and "sex") are already
near the boundary of the parameter space. In case there are at both sides (for
objects and attributes) parameter values near 1, then multiplying at one side and
dividing at the other is not possible without transgressing the boundary of 1.00.
Secondly, in comparison with the other two components, the second component
only accounts for a small part (6.7%) of the variance in the parameters. The
fourth component of the rotated solution is more specific in nature, as it

TABLE 6

Congruence Between Hypothetical Loadings and Loadings Obtained after Procrustes Rotation

	Rotated		
	Between	Within	
		Feature 1	Feature 2
	.826	−.052	−.144
Hypothetical	−.146	.303	.095
	−.173	.041	.928

concerns only a few parameters. It is also less important since it only accounts for a smaller part of the variance (8.0%).

The loadings of the individual parameters on the rotated components may be used to estimate the proportion of the variance per parameter that is due to the corresponding type of trade-off. We will only summarize the results for the two well-identified components (1 and 3). For the first component reflecting between-feature indeterminacy, the smallest, median and largest percentage of posterior variance accounted for in an individual parameter equals .02%, 4.9% and 36.2% respectively; for the third component, reflecting within-feature indeterminacy for the second feature, the smallest, median and largest percentage of posterior variance accounted for in parameters regarding the second feature equals .0001%, 4.8% and 39.9% respectively.

In the estimation section we noted that the parameter of Portugal for the first feature and the parameters of "needle", "blood" and "sex" for the second feature have a somewhat larger 95% posterior interval. It can now be examined to what extent this is due to the two identified types of indeterminacy. From the matrix of component loadings it can be derived that 26.9% of the variance in the parameter of Portugal regarding the first feature, is accounted for by between-feature indeterminacy. Furthermore, only 3.6% of the variance in this parameter is accounted for by the component reflecting within-feature indeterminacy for the second feature. For the parameters of "needle", "blood" and "sex" concerning the second feature, the component reflecting between-feature indeterminacy accounts for 36.2%, 25.9%, and 13.9%, respectively, of the variance. Besides, the component reflecting within-feature indeterminacy for the second feature accounts for 9.6%, 23.9%, and 5.4%, respectively, of the variance in these parameters. Nevertheless, the posterior uncertainty of most parameters seems to be relatively small in general, and indeterminacies in the model's parameters only account for a small part of this uncertainty.

In this respect it is interesting to note that the present analysis yields even more reliable parameter estimates (smaller 95% posterior intervals) and fewer indeterminacies in the model's parameters than an alternative analysis in which

the data were dichotomized in a slightly different way, namely by adding the "possibly" category to the category with the lowest frequency. The latter procedure yielded different observed data only for the items "needle", "blood" and "sex": Fewer respondents (especially for Portugal and to a lesser extent for Spain) had the opinion that AIDS can be contracted in these situations.

In this alternative analysis two hypothesized components could also be identified, as in the analysis reported on earlier, namely one reflecting between-feature indeterminacy and another reflecting within-feature indeterminacy for the second feature. These components accounted for 24.4% and 16.8% of the variance in the sample of the posterior, respectively. The 95% posterior intervals of the parameter of Portugal for the first feature ([.25, .76]) and of the parameters of "needle", "blood" and "sex" for the second feature ([.27, .87], [.35, .88], [.30, .88], respectively) were larger than in the present analysis and more of this posterior uncertainty could be explained by the identified components. In particular, the component reflecting between-feature indeterminacy accounted for 50.8% of the variance in the parameter of Portugal regarding the first feature and it accounted for 38.2%, 35.5% and 36.7% of the variance in the parameters of "needle", "blood" and "sex" concerning the second feature. Furthermore, the component reflecting within-feature indeterminacy for the second feature. accounted for 20.1% of the variance in the parameter of Portugal regarding the first feature and it accounted for 18.7%, 16.8% and 20.1% of the variance in the parameters of "needle", "blood" and "sex" regarding the second feature.

Figure 1 shows a plot of the two-dimensional histogram of the scores of the posterior draws on the first two unrotated principal components from the original analysis (with "possibly" and "yes" coded as 1). The plot approximates the joint posterior density of the principal components and so it may be interpreted as the best possible two-dimensional reduction of the observed posterior distribution. Figure 1 does not show any clearly separated regions of high posterior density. Another way to investigate whether separate regions of posterior density exist, is through visual inspection of (a) the marginal posterior distributions of all the parameters and (b) the joint posterior distributions of all pairs of parameters. In these plots as well, no separate regions of high posterior density could be identified.

One may wonder whether some of the indeterminacies discussed above could be eliminated during the estimation procedure by restricting the parameter space. Regarding within-feature indeterminacy, this could for instance be done by fixing for each feature one parameter at the posterior mode estimate. Applying the latter procedure, however, appeared to have little effect on the pattern of posterior covariances and on the results of the principal components analysis. In other words, within-feature indeterminacy can still be an issue if one parameter for each feature is fixed.
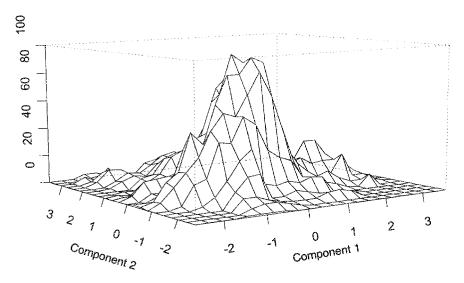
FIGURE 1. *Two-dimensional histogram of the posterior density for the two-feature model, based on the scores of the posterior draws on the first two principal components*
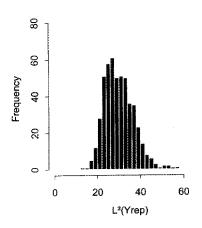
Model Checking

The test quantity $L^2_{\text{diff}}(Y)$ is used to test models with different numbers of features against each other. In particular models with one, two, and three features are considered.

Figure 2 shows the results of this procedure for a one-feature model versus a two-feature model and a two-feature model versus a three-feature model. The left panel displays the simulated reference distribution $L^2_{\text{diff}}(Y^{\text{rep}})$ for one versus two features. The quantity $L^2_{\text{diff}}(Y)$, which is based on the observed data, is not displayed in the figure because it has a very large value, namely 297.2. The corresponding p value equals .00. Hence, a model with two features is clearly superior to a model with one feature. On the other hand the right panel shows the simulated reference distribution $L^2_{\text{diff}}(Y^{\text{rep}})$ for two versus three features. This figure also displays the quantity $L^2_{\text{diff}}(Y)$ which equals 16.0. The corresponding p value equals .88, which means that there is no reason to prefer a three-feature over a two-feature model.

After having determined the appropriate number of features, we also assess the absolute fit of the selected model using a Pearson chi-square discrepancy measure. Figure 3 shows a plot of a sample of $\chi^2(Y^{\text{rep},m},\theta^{(m)})$ values against the correpsonding $\chi^2(Y,\theta^{(m)})$ values. The replicated discrepancies $\chi^2(Y^{\text{rep},m},\theta^{(m)})$ exceed or equal the observed discrepancies $\chi^2(Y,\theta^{(m)})$ for about 66% of the replicated data sets. The corresponding p value equals .66. In other words, the observed frequencies do not deviate systematically from frequencies which were
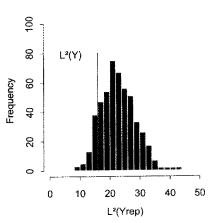
1 Versus 2 Features

2 Versus 3 Features

FIGURE 2. *The simulated reference distribution of the quantity* $L^2_{diff}(Y^{rep})$ *for the comparison of models with (a) one versus two features and (b) two versus three features*

generated from the posterior density of θ given the real data. Thus it can be concluded that the two-feature model fits the data fairly well.

Conclusion

In general, a fully Bayesian analysis, which implies simulating a sample of the entire posterior distribution, has three important advantages. First, the posterior sample provides the entire marginal posterior distribution of any estimand of interest, and thus goes beyond only locating the posterior mode. Second, the posterior sample provides information about indeterminacies in a model's parameters. More specifically, principal components analysis may be used in a confirmatory way to identify hypothesized indeterminacies which imply a specific pattern of covariances between parameters and to compute their impact on the posterior uncertainty of the parameters. In addition, the results of the principal components analysis may be used from an exploratory viewpoint to visualize the posterior distribution in a low-dimensional space, which may reveal interesting features of the posterior distribution. Third, the posterior sample may be used to assess the fit of the model with the technique of posterior predictive checks.

In the present paper we illustrated each of these advantages in a specific application with the PMD model. First, the use of a Gibbs sampling algorithm to compute a sample of the entire posterior distribution was shown to be straightforward since the augmented posterior distribution has a mathematically tractable form. As a result, a $(1 - \alpha)\%$ posterior interval of any estimand of interest was easily obtained without any extra computational effort and without relying
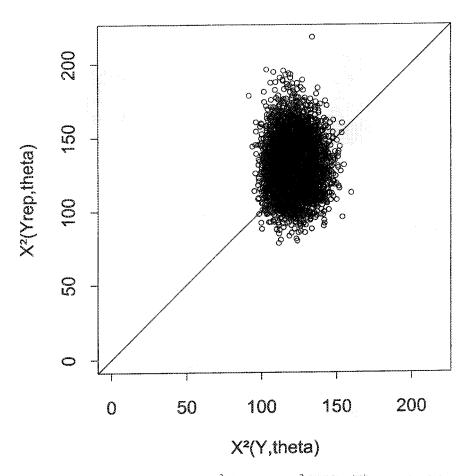
175

Two-feature Model

FIGURE 3.   *Scatterplot of replicated χ^2 discrepancy, $\chi^2(Y^{rep,m}, \theta^{(m)})$ vs. realized discrepancy, $\chi^2(Y,\theta^{(m)})$, for 5000 random draws of $(\theta^{(m)}, Y^{rep,m})$ from the posterior distribution of the two-feature model*

on normal approximations, in contrast to the computation of standard errors in the context of the EM algorithm. Second, a principal components analysis was shown to be an informative tool for tracing regions with a relatively high posterior density. In particular, Procrustes rotation of the obtained component loadings towards a hypothetical structure, which was expected from between- and within-feature indeterminacies, was shown to help to identify two compo-

nents. In addition, a histogram of the scores of the posterior draws on the first two principal components was shown to be a useful way to approximate the observed posterior distribution in a two-dimensional space. Third, the posterior sample was used to assess the fit of the PMD model both in a relative sense and in an absolute sense.

Finally, it may be noted that the approach presented in this paper can be extended to several other models inside as well as outside the PMD family. First, it is straightforward to implement the Gibbs sampler for PMD models with other deterministic mapping rules (Maris, De Boeck, & Van Mechelen, 1996). As a matter of fact, every possible mapping of latent variables into an observed data point can be used. Furthermore, it is possible to extend the approach to PMD models with a stochastic mapping rule; for this the vectors of realized latent response variables for an observed data point are sampled according to a prespecified probability distribution. Finally, given the promising results for the PMD model it may be worthwhile to extend the novel multivariate posterior analysis method we proposed here, to models outside the PMD family for which a full Bayesian approach is currently a topic of interest (Gelman, Carlin, Stern, & Rubin, 1995; Tanner, 1996; Hoijtink & Molenaar, 1997).

References

Candel, M.J.J.M., & Maris, E. (1997). Perceptual analysis of two-way two-mode frequency data: Probability matrix decomposition and two alternatives. *International Journal of Research in Marketing, 14*, 321–339.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig and L. W. Porter (Eds.), *Annual Review of Psychology, 31*, 607–649.

Cattell, R. (1966). *Handbook of multivariate experimental psychology.* Chicago: Rand McNally.

Cowles, K., & Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association, 91*, 883–904.

De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53*, 361–381.

de Bonis, M., De Boeck, P., Pérez-Diaz, F., & Nahas, M. (1999). A two-process theory of facial perception of emotions. *Life Sciences, 322*, 669–675.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*, 1–38.

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *Journal of the American Statistical Association, 87*, 476–486.

Gelfand, A. E., & Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*, 398–409.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* London: Chapman & Hall.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*, 457–472.

Gelman, A., Meng, X. M., & Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica, 4,* 733–807.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6,* 721–741.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1996). *Practical Markov chain Monte Carlo.* New York: Chapman & Hall.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika, 61,* 215–231.

Hoijtink, H., & Molenaar, I. W. (1997). A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks. *Psychometrika, 62,* 171–189.

Maris, E. (1995). Psychometric latent response models. *Psychometrika, 60,* 523–547.

Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika, 61,* 7–29.

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models.* New York: Marcel Dekker.

Meng, X. L. (1994). Posterior predictive p values. *The Annals of Statistics, 22,* 1142–1160.

Meulders, M., De Boeck, P., & Van Mechelen, I. (2000). *A taxonomy of latent structure assumptions for probability matrix decomposition models.* Manuscript submitted for publication.

Reif, K., & Melich, A. (1992). *Eurobarometer 32: The single European market, drugs, alcohol and cancer, November 1989* [computer file]. Conducted by INRA (Europe), Brussels. 2nd ICPSR ed. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and distributor].

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics, 12,* 1151–1172.

Rubin, D. B., & Stern, H. S. (1994). Testing in latent class models using a posterior predictive check distribution. In A. Von Eye & C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 420–438). Thousand Oaks, CA: Sage.

Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika, 31,* 1–10.

Smith, A.F.M., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B, 55,* 3–23.

Tanner, M. A. (1996). *Tools for statistical inference: Methods for exploration of posterior distributions and likelihood functions,* third edition. New York: Springer-Verlag.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association, 82,* 528–540.

Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive hierarchical classes model. *Psychometrika, 60,* 505–521.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* New York: Cambridge University Press.

Authors

MICHEL MEULDERS, Professor, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, B-3000 leuven, Belgium; Michel.Meulders@ psy.kuleuven.ac.be. He specializes in psychometric models.

PAUL DE BOECK, Professor, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, B-3000 leuven, Belgium; Paul.DeBoeck@psy.kuleuven.ac.be. He specializes in psychometric models.

IVEN VAN MECHELEN, Professor, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, B-3000 leuven, Belgium; Iven.VanMechelen@psy. kuleuven.ac.be. He specializes in psychometric models.

ANDREW GELMAN, Professor, Department of Statistics, Columbia University, New York, NY 10027; Gelman@neyman.stat.columbia.edu. He specializes in Bayesian data analysis.

ERIC MARIS, Professor, Nijmegen Institute of Cognition and Information, Katholieke Universiteit Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands; Maris. nici.kun.nl. He specializes in mathematical models.

TITLE: Bayesian inference with probability matrix decomposition
models
SOURCE: Journal of Educational and Behavioral Statistics 26 no2
Summ 2001
WN: 0119604897002