PSYCHOMETRIKA—VOL. 86, NO. 1, 30–64
MARCH 2021
https://doi.org/10.1007/s11336-021-09750-9

INFERRING THE NUMBER OF ATTRIBUTES FOR THE EXPLORATORY DINA MODEL

YINGHAN CHEN

UNIVERSITY OF NEVADA, RENO

YING LIU, STEVEN ANDREW CULPEPPER AND YUGUO CHEN

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Diagnostic classification models (DCMs) are widely used for providing fine-grained classification of a multidimensional collection of discrete attributes. The application of DCMs requires the specification of the latent structure in what is known as the $\boldsymbol{Q}$ matrix. Expert-specified $\boldsymbol{Q}$ matrices might be biased and result in incorrect diagnostic classifications, so a critical issue is developing methods to estimate $\boldsymbol{Q}$ in order to infer the relationship between latent attributes and items. Existing exploratory methods for estimating $\boldsymbol{Q}$ must pre-specify the number of attributes, K. We present a Bayesian framework to jointly infer the number of attributes K and the elements of $\boldsymbol{Q}$. We propose the crimp sampling algorithm to transit between different dimensions of K and estimate the underlying $\boldsymbol{Q}$ and model parameters while enforcing model identifiability constraints. We also adapt the Indian buffet process and reversible-jump Markov chain Monte Carlo methods to estimate $\boldsymbol{Q}$. We report evidence that the crimp sampler performs the best among the three methods. We apply the developed methodology to two data sets and discuss the implications of the findings for future research.

Key words: diagnostic models, dimensionality, Dirichlet processes, Indian buffet process, reversible jump MCMC, Bayesian.

1. Introduction

Diagnostic latent class models continue to receive attention by educational and psychometric researchers. Researchers applied diagnostic classification models (DCMs) to language testing (von Davier 2008), fraction-subtraction (Chen et al. 2015; Chen et al. 2018; de la Torre and Douglas 2004, 2008; Tatsuoka 1984; Tatsuoka 2002), situational judgment (Sorrel et al. 2016), and social anxiety (Chen et al. 2015). Furthermore, recent research developed longitudinal diagnostic models to track student learning trajectories (Chen et al. 2018; Kaya and Leite 2017; Li et al. 2016; Madison and Bradshaw 2018; Wang et al. 2017; Wang et al. 2018; Zhang et al. in press) and detect skill changes (Ye et al. 2016).

Confirmatory applications of diagnostic models require detailed, a priori knowledge about the underlying structure as to how latent skills relate to items. One challenge with widespread application of confirmatory DCMs is that expert knowledge may be unavailable to pre-specify the latent structure. Consequently, several studies proposed exploratory methods to infer the latent structure for DCMs (Chen et al. 2015; Chen et al. 2018; Culpepper and Chen 2018; Culpepper 2019a; Liu et al. 2013; Xu 2017; Xu and Shang 2018) using parsimonious DCMs and more general diagnostic modeling frameworks (for general models see de la Torre 2011; Henson et al. 2009; von Davier 2008).

This paper contributes to the literature on exploratory DCMs by proposing new methods for joint inference concerning the number of latent skills, K, and the underlying structure that

Correspondence should be made to Steven Andrew Culpepper, Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820, USA. Email: sculpepp@illinois.edu

describes how latent skills relate to items. That is, existing exploratory DCMs assume that the number of attributes, K, underlying item responses is known. In cases where the latent structure is unknown, it is also unexpected for researchers to know K a priori. Correct specification of K is fundamental for accurately recovering the latent structure. If K is too small, some attributes are absent and the missing attributes cannot be correctly classified. Likewise, a value of K that is too large possibly introduces spurious attributes. In short, incorrectly specifying K results in a misspecified model. Current applications of exploratory DCMs consider the problem of inferring K by comparing models of varying dimensions with fit indices such as the BIC (Xu and Shang 2018) and deviance information criterion (DIC; Culpepper and Chen 2018). Whereas existing approaches that use fit indices to infer K are easy to implement, we report Monte Carlo simulation results to support that our new methods more accurately infer the number of attributes with a smaller computational burden. In fact, our findings support the conclusion of Sen and Bradshaw (2017) that traditional fit indices are inadequate for comparing the fit of diagnostic models.

There are several benefits of our strategy for inferring K. First, we propose a "crimp sampling" algorithm that is designed to infer the latent structure and number of attributes while also enforcing model identifiability conditions (e.g., see Chen et al. 2020a; Culpepper 2019b; Fang et al. 2019; Gu and Xu 2019; Xu 2017; Xu and Zhang 2016). The term "crimp sampling" is based upon a technique in rock-climbing where climbers use just their fingertips to proceed up narrow edges. That is, as in rock-climbing, our crimp sampler involves as few items as possible to transition in an irreducible manner between identified model spaces of different dimensions. Second, we provide Monte Carlo evidence that the "crimp sampler" outperforms existing methods such as using the DIC and Bayesian procedures for inferring dimensionality such as the Dirichlet process/Indian buffet process (IBP; e.g., Gershman and Blei 2012; Griffiths and Ghahramani 2005, 2011; Sethuraman 1994; Teh et al. 2007; Thibaux and Jordan 2007) and reversible jump Markov chain Monte Carlo (RJ-MCMC; Green 1995). Third, we consider methods for simultaneous inference of $\boldsymbol{Q}$, K, and the item and structural parameters, which accounts for all sources of modeling error and provides an approach for automatic selection of the number of attributes. Fourth, we improve existing algorithms for estimating the DINA $\boldsymbol{Q}$ matrix (e.g., see Chen et al. 2018). Specifically, we use a collapsed Gibbs sampler (Liu, 1994) for updating attributes to improve mixing of the Markov chains. Furthermore, our proposed methodology imposes weaker necessary and sufficient identifiability conditions on the posterior distribution for $\boldsymbol{Q}$ and, as discussed below, it admits zero rows in $\boldsymbol{Q}$, which allows for the possibility that some items do not load on the attributes shared by the majority of items. It is important to note that imposing identifiability on the posterior distribution does not guarantee the true $\boldsymbol{Q}$ matrix is identified. In fact, there is no consistent estimator if the true $\boldsymbol{Q}$ is unidentified. We enforce identifiability on the posterior distribution for at least two reasons. First, we assume the true $\boldsymbol{Q}$ is identified and it is reasonable to restrict the search to the identified space. Second, unrestricted sampling will yield a posterior with support on unidentified $\boldsymbol{Q}$ matrices and draws from the unidentified space cannot be coherently interpreted. That is, if we freely sample $\boldsymbol{Q}$, we may have some posterior samples that are identified and some that are not identified. One option could be to deploy post-processing to filter out the unidentified cases, but the drawback may be that we would require longer chains to sample from the identified space.

The remainder of this paper is organized as follows. Section 2 reviews the problem of dimensionality for DCMs, provides an overview of approaches in statistics literature such as IBP and RJ-MCMC, and adapts them to estimate K in DCMs. Section 3 introduces the crimp sampling algorithm and shows that it is irreducible and satisfies identifiability constraints. Section 4 includes a summary of algorithms of the crimp sampler, IBP, and RJ-MCMC for inferring K in the deterministic inputs, noisy "and" gate model (DINA). Section 5 reports results from Monte Carlo simulation studies regarding the performance of the aforementioned algorithms in recovering the

DINA model K. Section 6 presents two applications of the crimp sampling method to infer K; the first application is to Tatsuoka's fraction-subtraction data (Tatsuoka 1984), and the second is to the data set on Problems in Elementary Probability Theory, which is available in the "Probabilistic Knowledge Structures" pks R package (Heller and Wickelmaier 2013). Finally, Sect. 7 concludes the paper with a discussion of the implications of the study and recommendations for future research.

2. Adapting RJ-MCMC and IBP to Estimate K

In this section, we review the problem of estimating K for DCMs, provide a Bayesian framework for this problem, and adapt two approaches from the statistics literature to infer K and other DCM parameters.

2.1. Overview

Applications of diagnostic latent class models are based on the specification of a binary $\boldsymbol{Q}$-matrix, which defines which latent skills relate to each item. Let J denote the number of items and K the number of attributes. The $J \times K$ structure matrix is defined as $\boldsymbol{Q} = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_J)^\top$, where $\boldsymbol{q}_j^\top = (q_{j1}, \ldots, q_{jK})$ is a binary K-vector and $q_{jk} = 1$ indicates item j requires the mastery of attribute k and 0 otherwise. Let $\boldsymbol{\alpha}_i^\top = (\alpha_{i1}, \ldots, \alpha_{iK})$ be the latent attribute profile of subject i, where $\alpha_{ik} = 1$ indicates subject i possesses attribute k and 0 otherwise. Let Y_{ij} be the binary response of subject i $(1 \leq i \leq N)$ to item j $(1 \leq j \leq J)$. In diagnostic models Y_{ij} is modeled by an item response function $P(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}_j, \boldsymbol{q}_j)$, where y_{ij} is a realized value and $\boldsymbol{\Omega}_j$ indicates parameters specific to item j. We assume individual i's responses to the J items are mutually independent when conditioned upon $\boldsymbol{\alpha}_i$. Let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})^\top \in \{0, 1\}^J$ be a random vector of responses for individual i with realization $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iJ})^\top$. The assumption of local independence implies that the likelihood of $\boldsymbol{Y}_i = \boldsymbol{y}_i$ is

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}, \boldsymbol{Q}) = \prod_{j=1}^{J} P(Y_{ij} = y_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}_j, \boldsymbol{q}_j), \tag{1}$$

where $\boldsymbol{\Omega} = (\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_J)^\top$. Let $\boldsymbol{\pi}$ be a 2^K vector of structural probabilities where element c is defined as $P(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) = \pi_c$ and $\boldsymbol{v} = (2^{K-1}, \ldots, 1)^\top$ is a vector used to create a bijection between the binary attributes and integers $c \in \{0, \ldots, 2^K - 1\}$. The likelihood of observing $\boldsymbol{y}_i$ given $\boldsymbol{\pi}$, $\boldsymbol{\Omega}$, and $\boldsymbol{Q}$ is

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{\pi}, \boldsymbol{\Omega}, \boldsymbol{Q}) = \sum_{c=0}^{2^K-1} \pi_c P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}, \boldsymbol{Q}), \tag{2}$$

The likelihood of observing an independent sample of $i = 1, \ldots, N$ respondents is therefore

$$P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\Omega}, \boldsymbol{Q}) = \prod_{i=1}^{N} P(\boldsymbol{Y}_i = \boldsymbol{y}_i|\boldsymbol{\pi}, \boldsymbol{\Omega}, \boldsymbol{Q}), \tag{3}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)$ denotes the observed binary responses for individual $i = 1, \ldots, N$ for the random responses $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$.

This paper focuses on inferring $\boldsymbol{Q}$ and K. In general, the conditional posterior distribution of $\boldsymbol{Q}$ and K given $\boldsymbol{Y}$, $\boldsymbol{\pi}$, and $\boldsymbol{\Omega}$ is

$$p(\boldsymbol{Q}, K|\boldsymbol{Y} = \boldsymbol{y}, \boldsymbol{\pi}, \boldsymbol{\Omega}) \propto P(\boldsymbol{Y} = \boldsymbol{y}|\boldsymbol{\pi}, \boldsymbol{\Omega}, \boldsymbol{Q}, K)p(\boldsymbol{\Omega}, \boldsymbol{\pi}|K)p(\boldsymbol{Q}|K)p(K), \tag{4}$$

where $p(\boldsymbol{\Omega}, \boldsymbol{\pi}|K)$ is the prior of $\boldsymbol{\Omega}$ and $\boldsymbol{\pi}$ given K, $p(\boldsymbol{Q}|K)$ is the conditional prior of $\boldsymbol{Q}$ given K, and $p(K)$ is the prior distribution for K. Equation 4 shows that a prior for $\boldsymbol{Q}$ is needed to infer $\boldsymbol{Q}$ in the Bayesian framework. We next discuss two priors for $\boldsymbol{Q}$.

2.2. *Priors for* $\boldsymbol{Q}|K$

Existing Bayesian methods consider two types of priors for binary matrices such as $\boldsymbol{Q}$. One approach is to adapt RJ-MCMC and consider that K is finite on a set of integers. A second strategy treats K as infinite and applies the IBP algorithm to infer $\boldsymbol{Q}$. We next review both approaches and discuss their implementation for inferring $\boldsymbol{Q}$ in diagnostic models.

2.2.1. RJ-MCMC for Finite K According to the finite feature model (Griffiths and Ghahramani 2005), we consider K finite and specify a model for elements of $\boldsymbol{Q}$. A common approach is to assume that the elements in column k of $\boldsymbol{Q}$ (i.e., $\boldsymbol{Q}_k = (q_{1k}, \ldots, q_{Jk})^\top$) are independent and identically distributed conditional on a column-specific probability, ω_k. The hierarchical model formulation is

$$q_{jk}|\omega_k \overset{iid}{\sim} Bernoulli(\omega_k), \ j = 1, \ldots, J, \tag{5}$$

$$\omega_k|\lambda \overset{iid}{\sim} Beta(\lambda/K, 1), \ k = 1, \ldots, K, \tag{6}$$

where λ is a hyper-prior parameter. We also assume the columns are independent with each other, so the probability of $\boldsymbol{Q}$ given $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_K)$ and K is

$$p(\boldsymbol{Q}|K, \boldsymbol{\omega}) = \prod_{k=1}^{K}\prod_{j=1}^{J} \omega_k^{q_{jk}}(1-\omega_k)^{1-q_{jk}} = \prod_{k=1}^{K} \omega_k^{m_k}(1-\omega_k)^{J-m_k}, \tag{7}$$

where $m_k = \sum_{i=1}^{J} q_{ik}$ is the number of items requiring attribute k. Integrating $\boldsymbol{\omega}$ yields the prior for $\boldsymbol{Q}$ given K,

$$\begin{aligned} p(\boldsymbol{Q}|K) &= \int p(\boldsymbol{Q}|K, \boldsymbol{\omega})p(\boldsymbol{\omega}|K)d\boldsymbol{\omega} \\ &= \prod_{k=1}^{K} \frac{\frac{\lambda}{K}\Gamma\left(m_k + \frac{\lambda}{K}\right)\Gamma\left(J - m_k + 1\right)}{\Gamma\left(J + 1 + \frac{\lambda}{K}\right)}. \end{aligned} \tag{8}$$

We assume a discrete uniform distribution for finite K,

$$K \sim \text{uniform}\left(K_{\min}, K_{\max}\right). \tag{9}$$

Note $K_{\min}$ and $K_{\max}$ can be specified with prior knowledge or based upon limits defined by model identifiability conditions (e.g., see a discussion pertaining to the DINA model implementation in Sect. 4).

The RJ-MCMC is available for approximating the posterior distribution when K is finite. That is, RJ-MCMC applies the Metropolis–Hastings (MH) algorithm to transit between models of varying dimensions. Specifically, the RJ-MCMC consists of either a "birth" or "death" move. Let the model parameters for a given K be denoted by $\boldsymbol{\xi}^K = \{\boldsymbol{Q}, \boldsymbol{\pi}, \boldsymbol{\Omega}, K\}$. A "birth" move adds a column to $\boldsymbol{Q}$, adds elements to $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$, and correspondingly transits from a lower dimension to a higher dimension, i.e., $\boldsymbol{\xi}^K \to \boldsymbol{\xi}^{K+1}$. In contrast, a "death" move transits from $\boldsymbol{\xi}^K \to \boldsymbol{\xi}^{K-1}$ and therefore deletes a column of $\boldsymbol{Q}$ and removes elements of $\boldsymbol{\pi}$ and $\boldsymbol{\Omega}$.

The RJ-MCMC implementation for DCMs includes four essential steps. First, a candidate $\boldsymbol{\xi}^* = \{\boldsymbol{Q}^*, \boldsymbol{\pi}^*, \boldsymbol{\Omega}^*, K^*\}$ is sampled (i.e., $K^* = K + 1$ for a birth move and $K^* = K - 1$ for a death move). Second, the Jacobian of transformation, $|J|$, is computed for parameters that change dimensions. Third, the acceptance ratio is computed as

$$A = \frac{P(\boldsymbol{Y}|\boldsymbol{\xi}^*)}{P(\boldsymbol{Y}|\boldsymbol{\xi}^K)} \cdot \frac{p(\boldsymbol{\xi}^*)}{p(\boldsymbol{\xi}^K)} \cdot \frac{p(\boldsymbol{\xi}^K|\boldsymbol{\xi}^*)}{p(\boldsymbol{\xi}^*|\boldsymbol{\xi}^K)} \cdot |J|, \tag{10}$$

where $p(\boldsymbol{\xi}^*)$ and $p(\boldsymbol{\xi})$ are priors and $p(\boldsymbol{\xi}^K|\boldsymbol{\xi}^*)$ and $p(\boldsymbol{\xi}^*|\boldsymbol{\xi}^K)$ denote the distributions for sampling candidates. Finally, the candidate is accepted with probability $\min(1, A)$ (see Algorithm 4 and Appendix A for details).

2.3. IBP for Infinite K

The alternative to considering K as finite is to instead suppose it is infinite. In fact, the IBP considers K to be infinite rather than finite. Therefore, we will implement it in diagnostic models. The IBP is a stochastic process that models the probability distribution of binary matrices with a fixed number of rows and potentially infinite number of columns. Following Griffiths and Ghahramani (2005), the IBP prior for $\boldsymbol{Q}$ is:

$$p(\boldsymbol{Q}) = \frac{\lambda^{K_+}}{\prod_{h=1}^{2^J-1} K_h!} \exp\{-\lambda H_J\} \prod_{k=1}^{K_+} \frac{(J - m_k)!\,(m_k - 1)!}{J!}, \tag{11}$$

where K_+ is the number of nonzero columns in $\boldsymbol{Q}$, i.e., the number of attributes $\boldsymbol{Q}$ possesses, K_h is the number of attributes of pattern h (i.e., of one of the $2^J - 1$ forms of column configuration), $H_J = \sum_{j=1}^{J} 1/j$ is the Jth harmonic number, m_k is the number of rows possessing feature k, and λ is the parameter of IBP that influences the expected number of columns. The binary matrix can be updated through Gibbs sampling using the exchangeability of IBP (Griffiths and Ghahramani 2011).

We apply Gibbs sampling to generate samples from the posterior distribution $p(\boldsymbol{Q}|\boldsymbol{Y}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N, \boldsymbol{\Omega})$ using Eq. 11 as the prior for $\boldsymbol{Q}$. We start with an arbitrary binary matrix $\boldsymbol{Q}$ and then sequentially update each row of $\boldsymbol{Q}$. Let $m_{(j)k}$ be the number of items possessing feature k excluding item j. If $m_{(j)k}$ is greater than 0, we set $q_{jk} = 1$ with probability

$$P(q_{jk} = 1|\boldsymbol{Q}_{(jk)}, \boldsymbol{Y}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N, \boldsymbol{\Omega}) \propto p(\boldsymbol{Y}|\boldsymbol{Q}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N, \boldsymbol{\Omega}) P(q_{jk} = 1|\boldsymbol{Q}_{(jk)}), \tag{12}$$

where $\boldsymbol{Q}_{(jk)}$ denotes the entries of $\boldsymbol{Q}$ other than q_{jk}; otherwise, we delete column k. Upon finishing all elements in row j, we add additional n_j elements of 1s, $n_j \sim \text{Poisson}(\lambda/J)$, that is, we add n_j new columns which have only ones at row j and zeros elsewhere.

An important difference between the RJ-MCMC and IBP is that the birth move and death move in RJ-MCMC add or delete one column based on the acceptance probability, whereas the IBP method changes the number of columns automatically in each iteration when updating $\boldsymbol{Q}$.

3. Crimp Sampler

This section introduces the crimp sampler for a finite K. Note that neither the IBP nor RJ-MCMC are explicitly designed to address the issue of identifiability constraints in DCMs. For instance, the exchangeability of the rows in IBP is no longer valid if constraints are applied (Doshi-Velez and Williamson 2017) and the simple updating strategy for $\boldsymbol{Q}$ of the IBP is no longer feasible. As for the RJ-MCMC, if we add the identifiable constraints to the prior of $\boldsymbol{Q}$, the determinant of Jacobian matrix J would become hard to calculate. Therefore, we aim to propose a novel algorithm that can move between different K and address the issue of identifiability constraints.

3.1. Crimp Sampling

Let $\mathcal{Q}_K$ be the set of all $\boldsymbol{Q}$ matrices with K columns that satisfy model identifiability constraints. For the crimp sampler, we consider a range of values for $K \in \{K_{\min}, \ldots, K_{\max}\}$ and develop a carefully designed Metropolis–Hastings transit proposal to sample $\boldsymbol{Q}$ matrices from the identifiable space $\cup_{K=K_{\min}}^{K_{\max}} \mathcal{Q}_K$. The prior specification for $\boldsymbol{Q}$ and K is

$$p(\boldsymbol{Q}|K) \propto \mathcal{I}(\boldsymbol{Q} \in \mathcal{Q}_K) \tag{13}$$

$$p(K) = \frac{|\mathcal{Q}_K|}{\sum_{K=K_{\min}}^{K_{\max}} |\mathcal{Q}_K|}, \tag{14}$$

where $\mathcal{I}(\cdot)$ denotes the indicator function and $|\mathcal{Q}_K|$ is the cardinality of the identified space for $\mathcal{Q}_K$. Accordingly, the joint prior for $\boldsymbol{Q}$ and K is uniform on the identified sample space, i.e.,

$$p(\boldsymbol{Q}, K) \propto \mathcal{I}(\boldsymbol{Q} \in \cup_{K=K_{\min}}^{K_{\max}} \mathcal{Q}_K). \tag{15}$$

We propose a Metropolis–Hastings (MH) step to update the current $\boldsymbol{Q}$ to a new $\boldsymbol{Q}^*$ within the identifiable space. As outlined in Algorithm 1, the MH update for the crimp sampler involves four steps. First, we sample a candidate $(\boldsymbol{Q}^*, K^*)$ from the identified space. We propose to add a new column with probability p_1 if there is at least one row of all zeros in the current $\boldsymbol{Q}$. It is important to note that we only allow all zero rows if the remaining elements of $\boldsymbol{Q}$ satisfy the identifiability constraints. Note that items with zero rows do not relate to the current set of underlying attributes. In cases where there are zero rows for a set of items, our inference would be that there is no evidence the set of items with zero rows load onto the current set of attributes. Furthermore, when there are zero rows, there is an opportunity to add attributes to describe items with zero rows. In fact, our algorithm proposes to add a column to match the items with zero rows, so long as the number of columns does not exceed the upper bound for K. Furthermore, we propose to remove an existing column with probability p_2; otherwise, the number of columns in $\boldsymbol{Q}^*$ remains the same. Second, we compute the transition probabilities $T(\boldsymbol{Q}, K; \boldsymbol{Q}^*, K^*)$ and $T(\boldsymbol{Q}^*, K^*; \boldsymbol{Q}, K)$ for moves from $(\boldsymbol{Q}, K) \to (\boldsymbol{Q}^*, K^*)$ and $(\boldsymbol{Q}^*, K^*) \to (\boldsymbol{Q}, K)$, respectively. Third, we compute

the ratio of posteriors conditioned upon the item parameters $\boldsymbol{\Omega}$ and the current $N \times K$ matrix of attributes $\boldsymbol{A}_K = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)^\top$. Finally, we compute the acceptance probability

$$r = \min\left(1, \frac{p(\boldsymbol{Q}^*, K^*|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}, K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} \cdot \frac{T(\boldsymbol{Q}^*, K^*; \boldsymbol{Q}, K)}{T(\boldsymbol{Q}, K; \boldsymbol{Q}^*, K^*)}\right), \tag{16}$$

where

$$\frac{p(\boldsymbol{Q}^*, K^*|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}, K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}^*, K^*)p(\boldsymbol{A}_K, \boldsymbol{\pi}_K|K^*)p(\boldsymbol{Q}^*, K^*)}{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}, K)p(\boldsymbol{A}_K, \boldsymbol{\pi}_K|K)p(\boldsymbol{Q}, K)}. \tag{17}$$

Note that we assume the dimension of item parameters does not depend on K.

There are two types of transitions to consider for Eq. 17. First, transitions from higher to lower dimensions are denoted by $(\boldsymbol{Q}, K) \to (\boldsymbol{Q}^*, K-1)$, which for simplicity we write as $\boldsymbol{Q}_K \to \boldsymbol{Q}^*_{K-1}$. Without loss of generality suppose the Kth attribute is proposed to be deleted, then $\boldsymbol{A}_{K-1} = (\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N)^\top$ where $\tilde{\boldsymbol{\alpha}}_i$ represents the first $K-1$ attributes, i.e., $\boldsymbol{\alpha}_i = (\tilde{\boldsymbol{\alpha}}_i^\top, \alpha_{iK})^\top$. Equation 17 for K to $K-1$ can be simplified as:

$$\frac{p(\boldsymbol{Q}^*_{K-1}|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}_K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{A}_{K-1}, \boldsymbol{\Omega}, \boldsymbol{Q}^*_{K-1})\left(\prod_{c=0}^{2^{K-1}-1} {\pi_c}^{n_c+1}\right)}{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_K)\left(\prod_{c=0}^{2^{K-1}-1} \pi_{c1}^{n_{c1}}\pi_{c0}^{n_{c0}}\right)}, \tag{18}$$

where n_c is the number of individuals with $K-1$ attributes equal to $\boldsymbol{\alpha}_c$, π_c is the corresponding probability of having attribute profile $\boldsymbol{\alpha}_c$, n_{c0} and n_{c1} are the numbers of individuals with the first $K-1$ attributes equal to $\boldsymbol{\alpha}_c$ and the Kth attribute equal to 0 and 1, respectively, and π_{c0} and π_{c1} are the probabilities of having attribute profiles corresponding to those n_{c0} and n_{c1} individuals, respectively. As for $(\boldsymbol{Q}, K) \to (\boldsymbol{Q}^*, K+1)$ transitions, we rewrite it as $\boldsymbol{Q}_K \to \boldsymbol{Q}^*_{K+1}$. The acceptance ratio is obtained by integrating out $\left(\alpha_{1,K+1}, \ldots, \alpha_{N,K+1}\right)$ and the additional elements of $\boldsymbol{\pi}$. Equation 17 for transitions from K to $K+1$ can be simplified as:

$$\frac{p(\boldsymbol{Q}^*_{K+1}|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}_K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}^*_{K+1})}{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_K)}. \tag{19}$$

The full derivation of Eq. 17 for both directions is included in Appendix B.

3.2. Irreducibility and Proposed Transition Probability

Chen et al. (2018) showed that the transition between any candidate $\boldsymbol{Q}^* \in \mathcal{Q}_K$ with K fixed is irreducible. It is left to show that the proposed Algorithm 1 is able to transit between any candidates with all possible number of columns through adding or deleting one selected column. We relax the constraints to allow rows with all zeros, so that the proposed method can reach the state where $\boldsymbol{Q}^*$ has more than one all-zero rows, and thus, the transition to $\boldsymbol{Q}^*_{K+1}$ is feasible. Similarly, the proposed method can reach the state where a column is chosen to be deleted, and the updated $\boldsymbol{Q}^*$ still satisfies identifiable constraints; thus, the transition to $\boldsymbol{Q}^*_{K-1}$ is feasible.

The proposed transition between candidates with the same number of columns is symmetric. The transition probability from $(\boldsymbol{Q}, K)$ to $(\boldsymbol{Q}^*, K+1)$ is

$$T(\boldsymbol{Q}, K; \boldsymbol{Q}^*, K+1) = p_1 \times \frac{1}{2^m - 1} \times \frac{1}{\prod_{k=1}^{K}(2^{n_{e_k}} - 1)} \times \frac{1}{2^{J-m-\sum_{k=1}^{K} n_{e_k}} - a} \tag{20}$$

Algorithm 1
Crimp sampler proposal of identifiable candidate $\boldsymbol{Q}^*$

For current $(\boldsymbol{Q}, K)$, sample $u \sim \text{Uniform}(0, 1)$. Let m be the number of rows of all zeros in $\boldsymbol{Q}$. We propose a candidate $\boldsymbol{Q}^*$ with specified probabilities p_1 and p_2 where $p_1 + p_2 < 1$:
if $m > 0$, $K < K_{\max}$ and $u < p_1$ **then**
  We add the $(K + 1)$th column to $\boldsymbol{Q}$ as follows:
  For the m rows of all zeros in $\boldsymbol{Q}$, let c of them be ones and $m - c$ of them be zeros in the $(K + 1)$th column, $1 \le c \le m$.
  Let $\boldsymbol{e}_k$ be the binary vector with only the kth element equal to one. For the n_{e_k} rows which are $\boldsymbol{e}_k$s from the identity matrix in $\boldsymbol{Q}$, let c_k of them be zeros and $n_{e_k} - c_k$ of them be ones in the $(K + 1)$th column, $1 \le c_k \le n_{e_k}, k = 1, \ldots, K$. Denote n_c the number of ones currently in the $(K + 1)$th column.
  Let the remaining $J - m - \sum n_{e_k}$ elements of the $(K + 1)$th column be uniform from all possible configurations that satisfy the identifiability constraints. There are $2^{J-m-\sum n_{e_k}} - a$ possible configurations, where a is a correction to make sure that every column sum is greater than or equal to 3. Here $a = 1 + J - m - \sum n_{e_k}$ for $n_c = 1$; $a = 1$ for $n_c = 2$, and $a = 0$ for $n_c \ge 3$.
else if $K > K_{\min}$ and $u > 1 - p_2$ **then**
  Randomly remove one of the K columns.
else
  Propose $\boldsymbol{Q}^*$ with the same number of columns K through updating elements in $\boldsymbol{Q}$ by constrained Gibbs sampler or Metropolis-Hastings sampler in Chen, Culpepper, Chen, and Douglas (2018).
end if

and the transition probability from $(\boldsymbol{Q}^*, K + 1)$ to $(\boldsymbol{Q}, K)$ is

$$T(\boldsymbol{Q}^*, K + 1; \boldsymbol{Q}, K) = p_2 \times \frac{1}{K + 1}, \tag{21}$$

where p_1 and p_2 are prespecified probabilities of adding and deleting a column, n_{e_k} is the number of $\boldsymbol{e}_k$ rows in $(\boldsymbol{Q}, K)$, and a is the correction defined in Algorithm 1.

3.3. *Estimating Attribute Profiles*

With the proposed $\boldsymbol{Q}^*$, we can update item parameters and attribute profiles through Gibbs sampling (Culpepper 2015; Chen et al. 2018). In particular, when updating $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\pi}$ of varying dimensions, we apply the collapsed Gibbs step to integrate $\boldsymbol{\pi}$ out. We assume $\boldsymbol{\pi}|K \sim \text{Dirichlet}(\boldsymbol{\delta}_0)$. When $\boldsymbol{\delta}_0 = \mathbf{1}_{2^K}$, the conditional prior distribution for $\boldsymbol{\alpha}_i$ after integrating over $\boldsymbol{\pi}$ is

$$p(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c | \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_N, K) = \frac{n'_c + 1}{N + 2^K - 1}, \tag{22}$$

where n'_c is the number of attribute profile $\boldsymbol{\alpha}_c$ of all but the i-th subject. Accordingly, when the same dimension K or a lower dimension $K-1$ is proposed, we can update the current $\boldsymbol{\alpha}_i$ or collapsed $\tilde{\boldsymbol{\alpha}}_i$ sequentially with weight proportional to $p(\boldsymbol{Y}_i|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{Q}, \boldsymbol{\alpha}_c)(n'_c+1), c = 0, \ldots, 2^K-1$ or $2^{K-1}-1$ respectively.

When a higher dimension is proposed, the conditional prior for the $K + 1$ attribute profile $\boldsymbol{\alpha}^*_i = (\boldsymbol{\alpha}_i, \alpha^*_i)$ is

$$p(\boldsymbol{\alpha}^*_i = (\boldsymbol{\alpha}_c, \alpha^*) | \boldsymbol{\alpha}^*_1, \ldots, \boldsymbol{\alpha}^*_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_N) = \frac{n_{c\alpha^*,i-1} + 1}{n_{c,i-1} + 2} \cdot \frac{n'_c + 1}{N + 2^K - 1}, \tag{23}$$

where $\alpha^* = 0$ or 1, $n_{c\alpha^*,i-1}$ is the number of the first $i-1$ subjects who possess attribute pattern $(\boldsymbol{\alpha}_c, \alpha^*)$, and $n_{c,i-1}$ is the number of the first $i-1$ subjects with the first K attributes of pattern $\boldsymbol{\alpha}_c$. Then, we can update $\boldsymbol{\alpha}_i^*$ to $(\boldsymbol{\alpha}_c, \alpha^*)$ with weight proportional to $p(\boldsymbol{Y}_i|\boldsymbol{\alpha}_i^* = (\boldsymbol{\alpha}_c, \alpha^*), \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \times \frac{n_{c\alpha^*,i-1}+1}{n_{c,i-1}+2} \times \frac{n_c'+1}{N+2^K-1}$, $c = 0, \ldots, 2^K-1, \alpha^* = 0, 1$. The full derivation of Eqs. 22 and 23 is in Appendix C.

4. Algorithms for the DINA Model

In this section, we outline the algorithms for the crimp sampler and the two other methods, IBP and RJ-MCMC, for estimating $\boldsymbol{Q}$ and other parameters for the DINA model. Recall the DINA item response function is

$$P(Y_{ij} = 1|\boldsymbol{\alpha}_i, \boldsymbol{q}_j, \boldsymbol{\Omega}_j) = (1-s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}, \quad \eta_{ij} = \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{q}_j \geq \boldsymbol{q}_j^\top \boldsymbol{q}_j), \tag{24}$$

where $\boldsymbol{\Omega}_j = (s_j, g_j)$ denotes the slipping and guessing parameters for item j. Also, we let $\boldsymbol{s} = (s_1, \ldots, s_J)^\top$ and $\boldsymbol{g} = (g_1, \ldots, g_J)^\top$ denote all the DINA slipping and guessing parameters. We sample $\boldsymbol{Q}$ and K subject to identifiability constraints. Note that Gu and Xu (2019) and Gu and Xu (in press) showed that the following conditions are necessary and sufficient to identify the DINA model parameters:

1. With proper row permutations, $\boldsymbol{Q} = (\boldsymbol{I}_K, (\boldsymbol{Q}')^\top)^\top$ where columns in $\boldsymbol{Q}'$ are distinct;
2. Each attribute loads onto at least three items.

Recall that we consider $K \in [K_{\min}, K_{\max}]$. The identifiability conditions imply $K_{\max}$ has an upper bound. For instance, the upper bound for the DINA model is that $K_{\max} \leq J/2$. Note that this inequality is derived from the fact that if $K = J/2$, then there must be $J/2$ items with simple structure and the remaining $J/2$ rows of Q must be linearly independent with at least two ones per column.

Despite the different proposal of sampling $\boldsymbol{Q}$ and K in the three methods, the updating of item parameters and latent attribute profiles can all be implemented through Gibbs sampling (Culpepper 2015; Chen et al. 2018). The prior specifications for the DINA model parameters and latent attributes are

$$p(\boldsymbol{\alpha}_i|\boldsymbol{\pi}) \propto \prod_{c=1}^{C} \pi_c^{\mathcal{I}(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c)}, \quad 0 < \pi_c < 1, \ \sum_{c=1}^{C} \pi_c = 1, \ C = 2^K, \tag{25}$$

$$\boldsymbol{\pi}|K \sim \text{Dirichlet}(\boldsymbol{\delta}_0), \quad \boldsymbol{\delta}_0 = (\delta_{01}, \ldots, \delta_{0C}), \tag{26}$$

$$p(s_j, g_j) \propto s_j^{\alpha_s - 1}(1-s_j)^{\beta_s - 1} g_j^{\alpha_g - 1}(1-g_j)^{\beta_g - 1} \mathcal{I}(0 \leq g_j < 1 - s_j \leq 1). \tag{27}$$

Algorithm 2 shows the proposed crimp sampler, Algorithm 3 shows the application of Indian buffet process in estimating $\boldsymbol{Q}$, and Algorithm 4 shows the reversible jump MCMC of estimating $\boldsymbol{Q}$.

5. Monte Carlo Simulation Studies

This section reports the performance of the three algorithms on simulated data sets. We use the recovery rate of the number of columns and the estimation accuracy rate of the $\boldsymbol{Q}$ matrix to assess

Algorithm 2
Crimp sampler for $\boldsymbol{Q}$

Initialize with an identifiable $\boldsymbol{Q}_K^{(0)}$ matrix, attribute profiles $\boldsymbol{\alpha}^{(0)}$, attribute categorical probabilities $\boldsymbol{\pi}^{(0)}$ and other item parameters $\boldsymbol{s}^{(0)}$ and $\boldsymbol{g}^{(0)}$.
for t in $1:T$ **do**
  Update $\boldsymbol{Q}^{(t)}$ using Metropolis-Hastings steps as Algorithm 1.
  for i in $1:N$ **do**
    if $K^{(t)} = K^{(t-1)}$ or $K^{(t)} = K^{(t-1)} - 1$ **then**
      Update $\boldsymbol{\alpha}_i^{(t)}$ to $\boldsymbol{\alpha}_c$ with weight proportional to $p(\boldsymbol{Y}_i|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{Q}, \boldsymbol{\alpha}_c)(n'_c + 1),\ c = 0, \ldots, 2^{K^{(t)}} - 1$.
    else if $K^{(t)} = K^{(t-1)} + 1$ **then**
      Update $\boldsymbol{\alpha}_i^{(t)}$ to $(\boldsymbol{\alpha}_c, \alpha^*)$ with weight proportional to $p(\boldsymbol{Y}_i|(\boldsymbol{\alpha}_c, \alpha^*), \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \cdot \frac{n_{c\alpha^*,i-1}+1}{n_{c,i-1}+2} \cdot \frac{n'_c+1}{N+2^K-1}, c = 0, \ldots, 2^{K^{(t-1)}} - 1, \alpha^* = 0, 1$.
    end if
  end for
  Update $\boldsymbol{\pi}^{(t)}|\boldsymbol{\alpha}^{(t)} \sim Dirichlet(\tilde{\boldsymbol{N}} + \boldsymbol{\delta}_0)$, where $\tilde{\boldsymbol{N}} = (n_1, \ldots, n_c)^\top$ represents the frequencies of each attribute pattern $\boldsymbol{\alpha}_c$, $c = 0, \ldots, 2^{K^{(t)}} - 1$.
  Update $\boldsymbol{s}^{(t)}, \boldsymbol{g}^{(t)}|\boldsymbol{Y}, \boldsymbol{\alpha}^{(t)}, \boldsymbol{Q}^{(t)} \sim Beta(a_s, b_s)Beta(a_g, b_g)\mathcal{I}(0 < g < 1 - s < 1)$, i.e., sample $\boldsymbol{s}^{(t)}$ and $\boldsymbol{g}^{(t)}$ independently from $Beta(a_s, b_s)$ and $Beta(a_g, b_g)$ truncated in the region $0 < g < 1-s < 1$.
end for

Algorithm 3
IBP Gibbs sampling for $\boldsymbol{Q}$

Initialize with an identifiable $\boldsymbol{Q}_K^{(0)}$ matrix, attribute profiles $\boldsymbol{\alpha}^{(0)}$, attribute categorical probabilities $\boldsymbol{\pi}^{(0)}$, item parameters $\boldsymbol{s}^{(0)}$ and $\boldsymbol{g}^{(0)}$, and hyperparameter $\lambda^{(0)}$.
for t in $1:T$ **do**
  for all j in $1:J$ and k in $1:K^{(t)}$ **do**
    Update $\boldsymbol{q}_{jk}^{(t)}$ to 1 with probability proportional to $p(\boldsymbol{Y}|\boldsymbol{Q}_1, \boldsymbol{s}^{(t-1)}, \boldsymbol{g}^{(t-1)}, \boldsymbol{A}^{(t-1)})\frac{m_{(j)k}}{J}$ and 0 otherwise.
    if $m_{(j)k} = 0$ **then**
      Delete column k and add $r_j \sim$ Poisson(λ/J) new attributes to item j.
    end if
  end for
  Update $\lambda^{(t)} \sim Gamma(K^{(t)} + 1, H_N + 1)$ (see Appendix D for details.)
  for i in $1:N$ **do**
    Update $\boldsymbol{\alpha}_i^{(t)}$ to $\boldsymbol{\alpha}_c$ with weight proportional to $p(\boldsymbol{Y}_i|\boldsymbol{s}, \boldsymbol{g}, \boldsymbol{Q}, \boldsymbol{\alpha}_c)(n_{c,i} + 1),\ c = 0, \ldots, 2^K - 1$.
  end for
  Update $\boldsymbol{\pi}^{(t)}, \boldsymbol{s}^{(t)}, \boldsymbol{g}^{(t)}$ as in Algorithm 2.
end for

performance. We examined the performance of the crimp sampler, IBP, and RJ-MCMC under different sample sizes (i.e., $N = 500$, 1000, and 2000), numbers of attributes (i.e., $K = 3, 4$ and 5), and correlations among the latent attributes (i.e., $\rho = 0, 0.25$, and 0.5). Note we also compare the MCMC procedures to a simpler strategy that uses the DIC to infer K. That is, we estimated models with $K = 2$ to 6 and inferred K by selecting the model with the smallest DIC. For each combination of the settings we replicate 100 data sets.

Algorithm 4
RJ-MCMC for $\boldsymbol{Q}$

Initialize with an identifiable $\boldsymbol{Q}_K^{(0)}$ matrix, attribute profiles $\boldsymbol{\alpha}^{(0)}$, attribute categorical probabilities $\boldsymbol{\pi}^{(0)}$, item parameters $\boldsymbol{s}^{(0)}$ and $\boldsymbol{g}^{(0)}$, and hyperparameter $\lambda^{(0)}$. Let b_K be the probability of birth move and $1 - b_K$ be the probability of death move.
for t in $1 : T$ **do**
  for all j in $1 : J$ and k in $1 : K^{(t)}$ **do**
    Update $\boldsymbol{q}_{jk}^{(t)}$ to 1 with probability proportional to $p(\boldsymbol{Y}|\boldsymbol{Q}_1, \boldsymbol{\alpha}, \boldsymbol{\Omega}) \cdot (m_{(j)k} + \frac{\lambda}{K^{(t)}})/(J + \frac{\lambda}{K^{(t)}})$ and 0 otherwise.
  end for
  For current $\boldsymbol{\xi}^{(t)} = \{\boldsymbol{Q}_K, \boldsymbol{\pi}, K\}$, sample $p \sim$ Uniform(0, 1). Let $C = 2^K$. We propose a candidate $\boldsymbol{\xi}^* = \{\boldsymbol{Q}^*, \boldsymbol{\pi}^*, K^*\}$ as follows:
  if $p \leq b_K$ **then**
    for i in $1 : C$ **do**
      draw $u_i \sim Beta(2, 2)$; let $\pi_i^* = \pi_i u_i$ and $\pi_{i+C}^* = \pi_i(1 - u_i)$.
    end for
    for j in $1 : J$ **do**
      Sequentially draw $q_{j,K+1} \sim Bernoulli(p^*)$, where
      $p^* = p(q_{j,K+1} = 1|q_{1,K+1}, ..., q_{j-1,K+1}) = (m_{(j)K+1} + \frac{\alpha}{K+1})/(j + \frac{\alpha}{K+1})$.
    end for
    Accept candidate $\boldsymbol{\xi}^{(t+1)} = \boldsymbol{\xi}^* = \{\boldsymbol{Q}^{K+1}, \boldsymbol{\pi}^*, K + 1\}$ with probability $\min(1, A)$ (see Appendix A for specific expression of A in Eq. 10); let $\boldsymbol{\xi}^{(t+1)} = \boldsymbol{\xi}^{(t)}$ otherwise.
  else
    Randomly select a column in $\boldsymbol{Q}_K$ to delete.
    For $k = 1, \ldots, (C/2)$, let $\pi_k^* = \pi_k + \pi_{k+C/2}$.
    Accept candidate $\boldsymbol{\xi}^{(t+1)} = \boldsymbol{\xi}^* = \{\boldsymbol{Q}^{K-1}, \boldsymbol{\pi}^*, K - 1\}$ with probability $\min(1, A)$; let $\boldsymbol{\xi}^{(t+1)} = \boldsymbol{\xi}^{(t)}$ otherwise.
  end if
  Update $\boldsymbol{\alpha}_i^{(t)}$ to $\boldsymbol{\alpha}_c$ with weight proportional to $p(\boldsymbol{\alpha}_i = \boldsymbol{\alpha}_c|\boldsymbol{Y}_i, \boldsymbol{\pi}, \boldsymbol{s}^{(t-1)}, \boldsymbol{g}^{(t-1)}), c = 0, \ldots, 2^{K^*} - 1$; then update $\boldsymbol{\pi}^{(t)}$, $\lambda^{(t)}$, $\boldsymbol{s}^{(t)}$, $\boldsymbol{g}^{(t)}$ as in Algorithm 3.
end for

For the crimp sampler, we use a chain length of 20,000 with a 10,000 burn-in period for $K = 3$ and 4, and a chain length of 35,000 with 20,000 burn-in period for $K = 5$. The range of columns is 2 to 6 for $K = 3$ and 4, and 3 to 8 for $K = 5$. We set the probability of adding a column p_1 as 0.25 and the probability of deleting a column p_2 as 0.1 based on preliminary simulation results. The average run-time for the crimp sampler for a sample size of 500 using a 2.4 GHz processor was 1.7, 3.2, and 5.1 minutes for $K = 3$, 4, and 5, respectively. For IBP, we use the same simulation settings as the crimp sampler. The range of columns is 1 to 8 for all possible values of K. The run-time for the IBP for a sample size of 500 using a 2.50 GHz processor was 4.6, 5.5, and 7.5 min for $K = 3$, 4, and 5. For the DIC, the average run-time for the DIC for a sample of size 500 using a 2.590 GHz processor was 67, 83, and 196 minutes for $K = 3$, 4, and 5.

For RJ-MCMC, we use a chain length of 30,000 with a 20,000 burn-in period for $K = 3$, a chain length of 40,000 with a 30,000 burn-in period for $K = 4$, and a chain length of 50,000 with a 40,000 burn-in period for $K = 5$. Note we use longer chains for the RJ-MCMC, to ensure convergence. The range of columns is 2 to 8 for $K = 3$ and 4, and 2 to 9 for $K = 5$. We set the probabilities of birth move b_k and death move d_k to 0.5. The average run-time for the RJ-MCMC

for a sample size of 500 using a 2.50 GHz processor was 11.9, 22.5, and 37 minutes for $K = 3, 4,$ and 5.

The true $\boldsymbol{Q}$ matrices we use are shown in Table 1. The true slipping and guessing parameters were set to be 0.2. When $\rho = 0$, we define the true $\boldsymbol{\pi} = (1/2^K, \ldots, 1/2^K)$ and sample the latent attributes uniformly from all attribute configurations. For the cases when $\rho = 0.25$ and 0.5, we sample the true latent attributes from a multivariate probit model and assume the true $\boldsymbol{\pi}$ equals the frequency of each attribute pattern as generated by

$$\alpha_{ik} = \begin{cases} 1 & \text{if } \theta_{ik} \geq \Phi^{-1}(\frac{k}{K+1}) \\ 0 & \text{otherwise} \end{cases}, \ k = 1, \ldots, K. \tag{28}$$

The prior parameters we use for $\boldsymbol{\pi}$, $\boldsymbol{s}$, and $\boldsymbol{g}$ in simulation are $\boldsymbol{\delta}_0 = \mathbf{1}$ for Eq. 26 and $\alpha_s = \beta_s = \alpha_g = \beta_g = 1$ for Eq. 27.

We report two measures regarding estimation accuracy of $\boldsymbol{Q}$. First, we report the elementwise accuracy rate (EAR) for all replications of the Monte Carlo simulation study. That is, for each replication we computed the proportion of elements of the true $\boldsymbol{Q}$ that are correctly estimated. It is important to note that from replications the estimated K ($\hat{K}$) may not match the true K, so we use the true number of elements of $\boldsymbol{Q}$ (i.e., $J \times K$) for the denominator in the computation of the proportion correct. In cases where $\hat{K} < K$, entire columns of the $\boldsymbol{Q}$ are missing. For instance, if $K = 4$ and $\hat{K} = 3$, the denominator is $4J$ and the numerator is calculated based upon the agreement in the elements of $\hat{\boldsymbol{Q}}$. We compute the EAR for each replication as:

$$\text{EAR}^{(r)} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \mathcal{I}(q_{jk} = \hat{q}_{jk}^{(r)}) \tag{29}$$

and the average EAR we report by simulation conditions is $\text{EAR} = \frac{1}{R} \sum_{r=1}^{R} \text{EAR}^{(r)}$, where R is the number of replications. Second, we report the number of over-specified elements (NOSE) to quantify accuracy for cases when $\hat{K} > K$ and there are additional columns of $\boldsymbol{Q}$. Specifically, for each replication we compute the number of additional elements as:

$$\text{NOSE}^{(r)} = \sum_{k=K+1}^{\hat{K}} \hat{q}_{jk}. \tag{30}$$

We report the average $\text{NOSE} = \frac{1}{R} \sum_{r=1}^{R} \text{NOSE}^{(r)}$ across replications by simulation condition.

5.1. Simulation Results

Table 2 reports the occurrence of correctly estimated K out of 100 replications and the average elementwise accuracy rate (EAR) of the estimated $\boldsymbol{Q}$ matrix when K is correctly estimated for the conditions with $\rho = 0$ and 0.25. The crimp sampler outperformed IBP and RJ-MCMC on estimating the number of attributes for all Ks and ρs in Table 2 except for one case where the RJ-MCMC is slightly better (i.e., $K = 3$ and $\rho = 0.25$). Furthermore, the crimp sampler improved upon the DIC in all but the $N = 2000$, $K = 3$, and $\rho = 0.25$ case in Table 2.

The crimp sampler also outperformed IBP and RJ-MCMC in recovering the correct $\boldsymbol{Q}$ matrix; its EAR is above 98% for $\rho = 0$ and is above 90% for $\rho = 0.25$. Table 2 also reports information about the number of over-specified elements for each algorithm. The results suggest the

TABLE 1.
True $\boldsymbol{Q}$ matrices for $K = 3, 4$ and 5.

J	$K=3$			$K=4$				$K=5$				
1	1	0	0	1	0	0	0	1	0	0	0	0
2	0	1	0	0	1	0	0	0	1	0	0	0
3	0	0	1	0	0	1	0	0	0	1	0	0
4	1	0	0	0	0	0	1	0	0	0	1	0
5	0	1	0	1	0	0	0	0	0	0	0	1
6	0	0	1	0	1	0	0	1	0	0	0	0
7	1	0	0	0	0	1	0	0	1	0	0	0
8	0	1	0	0	0	0	1	0	0	1	0	0
9	0	0	1	1	1	0	0	0	0	0	1	0
10	1	1	0	1	0	1	0	0	0	0	0	1
11	1	0	1	1	0	0	1	0	0	0	1	1
12	0	1	1	0	1	1	0	0	1	0	0	1
13	1	1	0	0	1	0	1	1	0	0	0	1
14	1	0	1	0	0	1	1	1	0	1	0	0
15	0	1	1	1	1	1	0	1	1	0	0	0
16	1	1	1	1	1	0	1	0	0	1	1	1
17	1	1	1	1	0	1	1	0	1	0	1	1
18	1	1	1	0	1	1	1	0	1	1	0	1
19	–	–	–	–	–	–	–	1	0	0	1	1
20	–	–	–	–	–	–	–	1	1	1	0	0

crimp sampler and RJ-MCMC outperformed IBP. Note that RJ-MCMC had fewer over-specified elements for larger K because it tends to underestimate K.

Table 3 reports the Monte Carlo results for the conditions with $\rho = 0.5$. It is important to note that a larger correlation implies a strong signal in the data from a higher-order model for attributes, which may be more closely aligned with a continuous item response model. In this setting, the attributes are highly correlated and the latent attribute structure becomes more sparse. Consequently, we can expect all algorithms will be less accurate at inferring K when there is more overlap among the latent attributes. The results in Table 3 confirm the difficulty in estimating K for all methods when attributes are highly correlated. For $K = 3$ and 4, there are cases where the crimp sampler, DIC, and RJ-MCMC demonstrate the best performance. For $K = 5$, the crimp sampler outperforms all methods for inferring K.

Tables 4 and 5 report the frequency of estimated K in the 100 replications. The tables provide insight regarding the sampling variability of estimates for K for the crimp sampler, IBP, and RJ-MCMC. The results in Table 4 suggest the occurrences of estimated $\hat{K}$ are centered around the true value for the crimp sampler and RJ-MCMC, while the estimated values are more dispersed for IBP. Table 5 shows that there are instances where the estimated K is incorrectly estimated in the majority of replications for the RJ-MCMC and IBP. In contrast, the crimp sampler most frequently recovers the true K when $K = 3$ and 4 and it is relatively less accurate for the $K = 5$ and $N = 2000$ case. Note we report the frequency of estimated K using DIC selection over 100 replications as well as RMSEs of estimated item parameters in Appendix F.

TABLE 2.
Summary of recovery rate of $\hat{K}$ and $\hat{\boldsymbol{Q}}$ by sample size N, number of true attributes K and attribute dependence $\rho = 0$ and 0.25.

			Crimp sampler			DIC	IBP				RJ-MCMC			
N	K	ρ	$\hat{K} = K$	EAR	NOSE	$\hat{K} = K$	$\hat{K} = K$	EAR	NOSE	Ident.	$\hat{K} = K$	EAR	NOSE	Ident.
500	3	0.00	97	99.73	0.00	96	91	98.34	0.90	97.97	95	99.73	0.17	95.60
1000	3	0.00	95	99.16	0.00	92	19	84.20	11.91	96.46	87	99.21	0.29	96.60
2000	3	0.00	95	99.20	0.00	90	55	73.45	1.60	85.90	76	98.29	0.55	98.71
500	4	0.00	100	100.00	0.00	99	52	94.35	1.79	98.10	92	99.79	0.18	96.24
1000	4	0.00	94	99.20	0.00	90	14	86.48	8.00	100.00	78	98.80	0.80	99.62
2000	4	0.00	91	98.67	0.13	76	38	59.35	0.40	24.00	66	94.06	1.20	96.73
500	5	0.00	97	99.58	0.00	90	35	93.99	2.13	96.12	90	99.54	0.33	96.41
1000	5	0.00	90	98.61	0.13	65	28	87.02	5.53	80.27	79	97.12	0.53	98.92
2000	5	0.00	84	98.27	0.26	57	24	62.51	0.00	9.02	59	90.23	0.63	90.83
500	3	0.25	92	98.67	0.00	82	67	92.44	2.87	90.69	99	96.98	0.00	95.14
1000	3	0.25	90	98.57	0.00	85	19	83.57	10.02	99.26	92	97.91	0.21	98.29
2000	3	0.25	83	97.18	0.19	87	23	81.42	10.60	84.94	82	96.91	0.38	97.89
500	4	0.25	85	97.82	0.14	80	65	88.67	0.93	54.36	63	90.86	0.21	77.41
1000	4	0.25	80	97.02	0.34	68	32	83.79	5.85	84.87	54	91.81	0.60	92.84
2000	4	0.25	79	96.73	0.34	65	26	62.02	0.74	18.65	47	90.50	0.20	95.55
500	5	0.25	67	93.01	1.14	41	44	83.07	1.31	38.99	29	86.55	0.55	61.41
1000	5	0.25	54	91.68	1.10	32	33	81.18	4.25	0.00	23	88.28	0.15	88.53
2000	5	0.25	48	90.65	1.35	21	6	63.18	0.00	6.46	23	85.96	0.04	87.30

$\hat{K} = K$ reports the occurrence of correctly estimated $\hat{K}$ out of 100 replications in each case. EAR reports the elementwise accuracy of estimated $\hat{\boldsymbol{Q}}$. NOSE = the average number of over-specified elements of $\boldsymbol{Q}$. "Ident." reports the averaged percentage of identifiable $\boldsymbol{Q}$ matrices after burn-in period among 100 replications in IBP and RJ-MCMC.

TABLE 3.
Summary of recovery rate of $\hat{K}$ and $\hat{\boldsymbol{Q}}$ by sample size N, number of true attributes K and attribute dependence $\rho = 0.5$.

			Crimp sampler			DIC	IBP				RJ-MCMC			
N	K	ρ	$\hat{K} = K$	EAR	NOSE	$\hat{K} = K$	$\hat{K} = K$	EAR	NOSE	Ident.	$\hat{K} = K$	EAR	NOSE	Ident.
500	3	0.5	83	93.03	0.40	79	82	90.65	1.62	34.23	92	93.24	0.00	58.64
1000	3	0.5	76	93.24	0.91	74	25	82.89	9.18	54.37	74	92.17	0.01	86.67
2000	3	0.5	71	93.85	1.14	74	28	82.08	9.95	70.15	66	91.45	0.19	95.22
500	4	0.5	77	93.74	0.58	76	47	82.15	0.63	0.68	37	84.09	0.02	5.30
1000	4	0.5	53	92.42	1.54	65	36	82.76	4.09	0.84	27	85.26	0.14	13.74
2000	4	0.5	32	92.18	3.27	37	22	63.25	1.55	8.37	23	84.68	0.18	15.54
500	5	0.5	62	89.31	2.63	37	27	78.73	0.48	0.02	8	81.75	0.00	0.00
1000	5	0.5	31	84.20	4.43	20	30	79.21	3.49	0	8	83.30	0.04	0.00
2000	5	0.5	18	84.53	4.93	15	2	63.65	0.00	3.33	10	82.12	0.03	0.00

$\hat{K} = K$ reports the occurrence of correctly estimated $\hat{K}$ out of 100 replications in each case. EAR reports the elementwise accuracy of estimated $\hat{\boldsymbol{Q}}$. NOSE = the average number of over-specified elements of $\boldsymbol{Q}$. "Ident." reports the averaged percentage of identifiable $\boldsymbol{Q}$ matrices after burn-in period among 100 replications in IBP and RJ-MCMC.

TABLE 4.
Summary of the frequency of estimated $\hat{K}$ by sample size N, number of true attributes K and attribute dependence $\rho = 0$ and 0.25.

			Crimp sampler $\hat{K}$							IBP $\hat{K}$								RJ-MCMC $\hat{K}$						
N	K	ρ	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	2	3	4	5	6	7	8	9
500	3	0.00	3	97	0	0	0	–	–	0	2	91	2	5	0	0		0	0	95	2	3	0	0	0 –
1000	3	0.00	5	95	0	0	0	–	–	0	4	19	19	45	11	2		0	1	87	9	1	0	2	0 -
2000	3	0.00	5	95	0	0	0	–	–	7	10	55	24	3	0	1		0	0	76	18	5	1	0	0 -
500	4	0.00	0	0	100	0	0	–	–	2	4	10	52	19	11	2		0	0	0	92	7	1	0	0 -
1000	4	0.00	0	6	94	0	0	–	–	0	5	14	14	33	21	11		2	0	0	78	17	4	0	1 -
2000	4	0.00	1	5	91	3	0	–	–	15	15	18	38	12	2	0		0	0	0	66	33	1	0	0 -
500	5	0.00	–	0	3	97	0	0	0	1	8	17	35	31	7	1		0	0	0	0	90	9	1	0 0
1000	5	0.00	–	2	6	90	1	1	0	0	1	6	16	28	26	20		3	0	0	4	79	14	3	0 0
2000	5	0.00	–	0	11	84	5	0	0	3	10	18	45	24	0	0		0	0	0	17	59	23	1	0 0
500	3	0.25	8	92	0	0	0	–	–	0	5	67	25	3	0	0		0	1	99	0	0	0	0	0 -
1000	3	0.25	10	90	0	0	0	–	–	1	7	19	41	25	5	2		0	4	92	3	1	0	0	0 -
2000	3	0.25	12	83	5	0	0	–	–	1	4	23	34	21	12	3		2	7	82	8	2	1	0	0 -
500	4	0.25	3	9	85	3	0	–	–	1	2	20	65	10	2	0		0	0	33	63	1	3	0	0 -
1000	4	0.25	3	10	80	7	0	–	–	1	7	7	32	33	16	4		0	0	34	54	8	3	0	0 -
2000	4	0.25	2	12	79	7	0	–	–	7	12	44	26	6	2	1		2	0	45	47	7	1	0	0 -
500	5	0.25	–	5	10	67	5	13	0	0	5	11	23	44	14	3		0	0	5	56	29	9	1	0 0
1000	5	0.25	–	11	16	54	9	10	0	0	2	6	16	33	25	17		1	0	8	67	23	1	1	0 0
2000	5	0.25	–	12	20	48	11	9	0	1	12	34	47	6	0	0		0	0	9	64	23	4	0	0 0

The specified range of K for the crimp sampler is 2 to 6 when $K = 3$ and 4, 3 to 8 when $K = 5$; the specified range of K for IBP is 1 to 8 for all cases; and the specified range of K for RJ-MCMC is 2 to 8 when $K = 3$ and 4, and 2 to 9 when $K = 5$.

TABLE 5.
Summary of the frequency of estimated $\hat{K}$ using crimp sampler, IBP and RJMCMC by sample size N and number of true attributes K when attribute dependence $\rho = 0.5$.

			Crimp sampler $\hat{K}$							IBP $\hat{K}$								RJ-MCMC $\hat{K}$						
N	K	ρ	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	2	3	4	5	6	7	8
500	3	0.5	4	83	13	0	0	–	–	0	3	82	13	2	0	0	0	8	92	0	0	0	0	0
1000	3	0.5	0	76	24	0	0	–	–	0	3	25	48	20	4	0	0	25	74	1	0	0	0	0
2000	3	0.5	0	71	29	0	0	–	–	0	3	28	37	14	7	10	1	29	66	4	1	0	0	0
500	4	0.5	0	13	77	9	1	–	–	0	4	41	47	8	0	0	0	0	62	37	1	0	0	0
1000	4	0.5	0	23	53	20	4	–	–	0	5	12	36	37	8	2	0	1	68	27	4	0	0	0
2000	4	0.5	0	24	32	29	15	–	–	5	7	49	22	6	8	2	1	4	68	23	4	1	0	0
500	5	0.5	–	0	3	62	5	30	0	2	2	16	47	27	6	0	0	0	14	78	8	0	0	0
1000	5	0.5	–	0	4	31	35	30	0	0	2	9	18	30	29	11	1	0	20	70	8	2	0	0
2000	5	0.5	–	0	13	18	32	37	0	1	7	42	48	2	0	0	0	0	30	58	10	2	0	0

The specified range of K for the crimp sampler is 2 to 6 when $K = 3$ and 4, 3 to 8 when $K = 5$; the specified range of K for IBP is 1 to 8 for all cases; and the specified range of K for RJ-MCMC is 2 to 8 when $K = 3$ and 4, and 2 to 9 when $K = 5$.

6. Applications

In this section, we apply the crimp sampler to Tatsuoka's Fraction-Subtraction data set (Tatsuoka 1984; Tatsuoka 2002) and the Problems in Elementary Probability Theory data set (Heller and Wickelmaier 2013). For each data set, we ran 50 chains to demonstrate the posterior distribution of the random variable K and used the posterior mode across chains as the final estimation and to ensure convergence. Note that running multiple chains also addresses the issue of a multimodal posterior distribution as noted by Liu et al. (2020).

6.1. Tatsuoka's Fraction-Subtraction Data

The data set contains responses from $N = 536$ students to $J = 20$ fraction-subtraction questions. The expert $\boldsymbol{Q}$ in Table 6 presents the pre-specified $\boldsymbol{Q}$ matrix based on the following eight skills:

1. Convert a whole number to fraction,
2. Separate a whole number from fraction,
3. Simplify before subtraction,
4. Find a common denominator,
5. Borrow from the whole number part,
6. Column borrow to subtract the second numerator from the first,
7. Subtract numerators,
8. Reduce answers to simplest form.

We set the range for K from 2 to 8 and repeated estimation 50 times. Recall the expert $\boldsymbol{Q}$ for this data set included eight attributes. Note the DINA identifiability conditions imply that the largest value is $K = 10$ for the fraction-subtraction data given there are 20 items. In practice, researchers can specify the entire range of identifiable values for K. Researchers can easily assess whether the specified range for K is too narrow. Specifically, issues with using too narrow a range for K would manifest as the algorithm sampling values near or at the upper bound. For the application reported in this subsection, we report that the estimated K was significantly less than the expert K, so there was no apparent issue with the pre-specified range.

We ran 50 replications of the algorithm on the fraction-subtraction data. Of the 50 replications, 39 reported an estimated $\hat{K} = 3$, 5 reported $\hat{K} = 4$, 4 reported $\hat{K} = 5$, and 2 reported $\hat{K} = 6$ and 7, respectively. Among the 39 replications where $\hat{K} = 3$, the most frequently occurring estimated $\hat{\boldsymbol{Q}}$ is as shown in Table 6.

Figure 1 shows the plot of maximum proportional scale reduction factor (max $\hat{R}$) (Brooks and Gelman 1998) for diagnosing convergence of Markov chains with multivariate parameters. The max $\hat{R}$ is below 1.1 after 8000 iterations, so it is reasonable to use a chain length of 20,000 with 10,000 burn-in.

One concern researchers may have with using the DINA model is that the model is misspecified and a more general model may provide better fit. Accordingly, we deploy two strategies to evaluate relative model fit. First, we computed the DIC to compare the fit of our exploratory DINA with crimp sampling to two confirmatory general diagnostic models (GDM) that use the expert $\boldsymbol{Q}$: (1) a main-effect-only model; and (2) a saturated model with all main-effects and interaction-effects. The $\hat{\boldsymbol{Q}}$ estimated using our new method had an average DIC over replications of 9133.34, whereas the DIC for the confirmatory GDMs was 13893.19 for a main-effects only model and 13,874.14 for the saturated model. Second, we compared the aforementioned models by computing the tenfold cross-validated deviance (i.e., minus two times the log-likelihood). The DIC is intended to estimate the cross-validated deviance, and we directly computed it to ensure the DIC was providing an accurate assessment of relative model fit. The cross-validated deviance

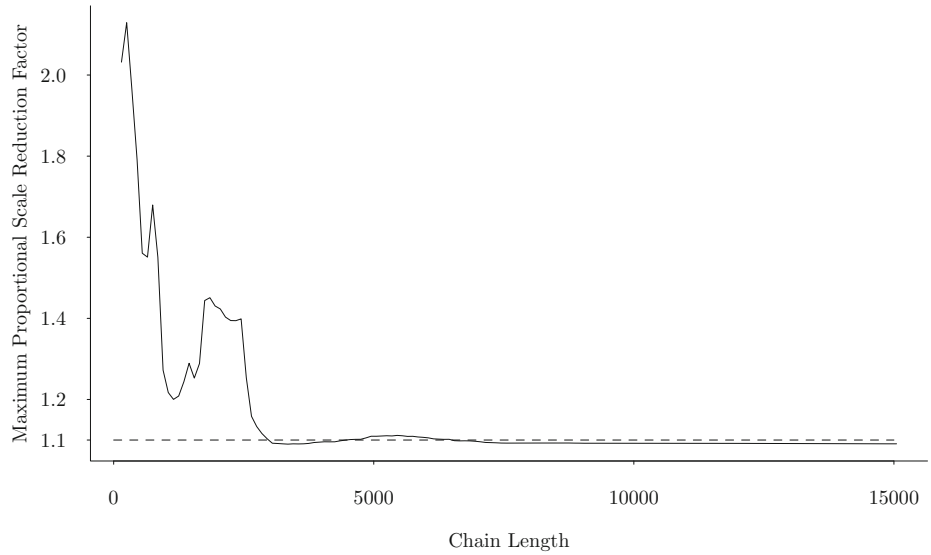

FIGURE 1.
The maximum $\hat{R}$ for Fraction-Subtraction data.

also agreed with the DIC (i.e., it equaled 9155.87 for our exploratory method, 9766.30 for a confirmatory model with main-effects, and 9602.75 for a saturated confirmatory model), and we next report empirical results for our crimp sampling algorithm.

The estimated $\boldsymbol{Q}$ describes a different underlying structure for the fraction-subtraction items than the expert $\boldsymbol{Q}$. For instance, the expert $\boldsymbol{Q}$ includes eight attributes, whereas the estimated $\boldsymbol{Q}$ includes three. The discrepancy between the two $\boldsymbol{Q}$ matrices provides an example as to how structure formulated by experts could differ from patterns uncovered in multivariate response patterns by machine learning tools. For instance, the expert $\boldsymbol{Q}$ specifies eight fine-grained operations students must perform, whereas the estimated $\boldsymbol{Q}$ identifies underlying features specified by the exploratory DINA model that best describe statistical dependency among the items. That is, the first attribute of the estimated $\boldsymbol{Q}$ distinguishes between items that require students to find a common denominator and the second attribute characterizes items that require borrowing from the whole number part. Additionally, the third attribute is related to subtracting both integer and fraction.

We also estimated the structural class probabilities. Specifically, the proportions of each attribute pattern in increasing order of binary-to-integer bijection $\boldsymbol{\alpha}_c^\top \boldsymbol{v}$ (from all 0's to all 1's) are: 20.7%, 14.6%, 2.0%, 3.2%, 1.5%, 13.2%, 1.2%, 43.6%.

6.2. Problems in Elementary Probability Theory

The data set contains responses from $N = 504$ students to two sets of $J = 12$ elementary probability questions before and after instructions. We used the first set of the questions. We set the range of K from 2 to 6 and repeated estimation 50 times. Figure 2 shows the plot of max $\hat{R}$ for the probability data. The max $\hat{R}$ is below 1.1 after 2000 iterations, so it is reasonable to use a chain length of 20,000 with 10,000 burn-in.

Of the 50 replications, 42 reported an estimated $\hat{K} = 3$, 7 reported $\hat{K} = 4$ and 1 reported $\hat{K} = 1$. Among the 42 replications where $\hat{K} = 3$, the most occurrence of estimated $\hat{\boldsymbol{Q}}$ is as shown in Table 7; the estimated $\hat{\boldsymbol{Q}}$ appeared most often for 31 out 42. We also compared the DIC between models fitted by the estimated $\hat{\boldsymbol{Q}}$ and the expert $\boldsymbol{Q}$ using the main-effects and a saturated GDM

TABLE 6.
Estimated $\hat{\boldsymbol{Q}}$, slipping $\hat{s}_j$ and guessing $\hat{g}_j$ parameters from the crimp sampler for Fraction-Subtraction Data.

Item	Expert $\boldsymbol{Q}$									$\hat{\boldsymbol{Q}}$		$\hat{s}_j$	$\hat{g}_j$
$\frac{5}{3} - \frac{3}{4}$	0	0	0	1	0	1	1	0	1	0	0	0.133	0.043
$\frac{3}{4} - \frac{3}{8}$	0	0	0	1	0	0	1	0	1	0	0	0.069	0.054
$\frac{5}{6} - \frac{1}{9}$	0	0	0	1	0	0	1	0	1	0	0	0.140	0.010
$3\frac{1}{2} - 2\frac{3}{2}$	0	1	1	0	1	0	1	0	0	1	0	0.125	0.180
$4\frac{3}{5} - 3\frac{4}{10}$	0	1	0	1	0	0	1	1	1	0	1	0.216	0.309
$\frac{6}{7} - \frac{4}{7}$	0	0	0	0	0	0	1	0	0	0	1	0.043	0.301
$3 - 2\frac{1}{5}$	1	1	0	0	0	0	1	0	1	0	1	0.342	0.032
$\frac{2}{3} - \frac{2}{3}$	0	0	0	0	0	0	1	0	0	0	1	0.053	0.580
$3\frac{7}{8} - 2$	0	1	0	0	0	0	0	0	0	0	1	0.250	0.348
$4\frac{4}{12} - 2\frac{7}{12}$	0	1	0	0	1	0	1	1	0	1	1	0.227	0.032
$4\frac{1}{3} - 2\frac{4}{3}$	0	1	0	0	1	0	1	0	0	1	1	0.072	0.072
$\frac{11}{8} - \frac{1}{8}$	0	0	0	0	0	0	1	1	0	0	1	0.093	0.195
$3\frac{3}{8} - 2\frac{5}{6}$	0	1	0	1	1	0	1	0	1	1	1	0.351	0.019
$3\frac{4}{5} - 3\frac{2}{5}$	0	1	0	0	0	0	1	0	0	0	1	0.068	0.138
$2 - \frac{1}{3}$	1	0	0	0	0	0	1	0	1	0	1	0.253	0.068
$4\frac{5}{7} - 1\frac{4}{7}$	0	1	0	0	0	0	1	0	0	0	1	0.112	0.117
$7\frac{3}{5} - \frac{4}{5}$	0	1	0	0	1	0	1	0	0	1	1	0.139	0.050
$4\frac{1}{10} - 2\frac{8}{10}$	0	1	0	0	1	1	1	0	0	1	1	0.155	0.132
$4 - 1\frac{4}{3}$	1	1	1	0	1	0	1	0	1	1	1	0.327	0.028
$4\frac{1}{3} - 1\frac{5}{3}$	0	1	1	0	1	0	1	0	0	1	1	0.186	0.016

The expert $\boldsymbol{Q}$ can be found in de la Torre and Douglas (2004).

with all main-effects and interaction-effects. The estimated $\hat{\boldsymbol{Q}}$ gives an average DIC of 4763.75 (based on 31 replications), whereas the expert $\boldsymbol{Q}$ using the main-effects GDM gives a DIC of 6369.73 and the DIC was 6346.04 for the confirmatory model with all effects (saturated model). Furthermore, the cross-validated deviance equaled 4970.36 for our exploratory method, 4994.07 for a confirmatory model with only main-effects, and 4976.84 for a saturated confirmatory model. Consequently, our exploratory model provided the best fit to the data and we next report model parameter estimates.

Table 7 presents the expert $\boldsymbol{Q}$, our estimate $\hat{\boldsymbol{Q}}$ and item parameters. By comparing $\hat{\boldsymbol{Q}}$ with the problem set (see Appendix E), Attribute 1 (represented by Questions 1, 5, and 9) is related to calculation of classical probability, Attribute 2 (represented by Questions 2, 3 and 8) is related to applying probability models, and Attribute 3 (appeared in Questions 4, 10, 11 and 12) might be related to the understanding of independence. The estimated proportions of each attribute pattern in increasing order of binary-to-integer bijection $\boldsymbol{\alpha}_c^\top \boldsymbol{v}$ (from all 0's to all 1's) are: 10.5%, 1.7%, 1.8%, 1.9%, 3.5%, 13.2%, 1.8%, 65.6%.

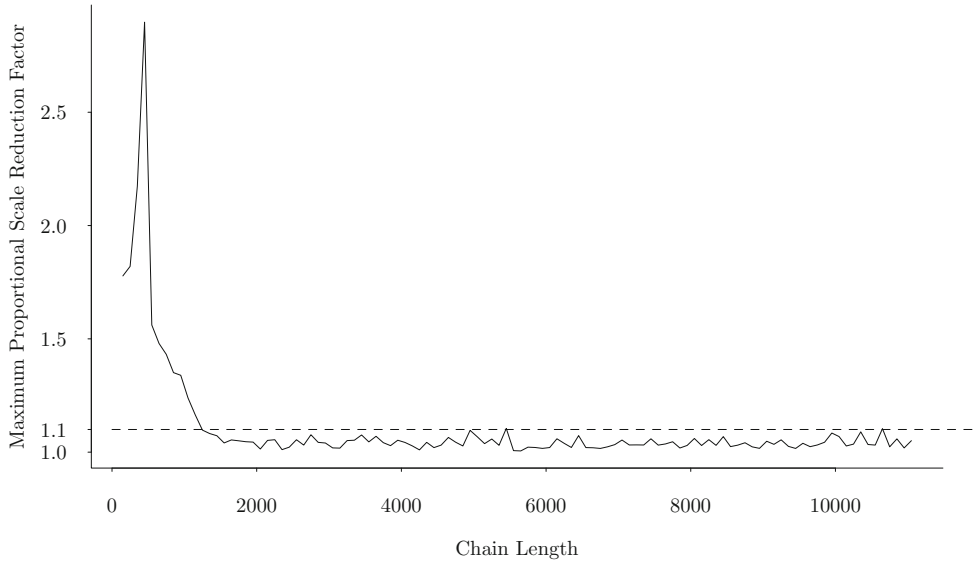

FIGURE 2.
The maximum $\hat{R}$ for probability data.

TABLE 7.
Estimated $\hat{\boldsymbol{Q}}$, slipping $\hat{s}_j$ and guessing $\hat{g}_j$ parameters from the crimp sampler for Elementary Probability Theory Data.

	Expert $\boldsymbol{Q}$				$\hat{\boldsymbol{Q}}$			$\hat{s}_j$	$\hat{g}_j$
1	0	0	1	0	1	0	0	0.082	0.213
2	1	0	0	0	0	1	0	0.038	0.371
3	0	0	0	1	0	1	0	0.043	0.260
4	0	1	0	0	0	0	1	0.040	0.131
5	1	0	1	0	1	0	0	0.152	0.249
6	1	0	1	0	1	0	0	0.048	0.265
7	0	0	1	1	1	1	0	0.067	0.366
8	0	0	1	1	0	1	0	0.050	0.451
9	0	1	1	0	1	0	0	0.253	0.089
10	1	1	0	0	0	0	1	0.192	0.065
11	1	1	1	0	0	0	1	0.304	0.037
12	0	1	1	1	1	1	1	0.188	0.038

The expert $\boldsymbol{Q}$ can be found in Heller and Wickelmaier (2013).

7. Discussion

The research problem of validation and/or estimation of $\boldsymbol{Q}$ matrix in latent class models has been advanced in recent years. Existing methods of validation or estimation of $\boldsymbol{Q}$ usually assume a given number of attributes K. However, when $\boldsymbol{Q}$ remains partially or completely unknown, it seems unrealistic to assume a known K. We proposed a Bayesian framework to incorporate the unknown K into diagnostic models, and developed a novel 'crimp sampler' algorithm to estimate K and $\boldsymbol{Q}$ simultaneously. We also adapted RJ-MCMC and IBP from the statistics literature to be implemented in $\boldsymbol{Q}$ estimation in diagnostic models. Our study is the first to examine the problem of inferring the dimensionality of $\boldsymbol{Q}$ in exploratory settings and offers new results about inferring

the number of attributes. In short, we found evidence that our method improved upon all other methods, such as using the DIC fit index, reversible jump MCMC, and Indian buffet process. We also demonstrate that selecting the true K is a difficult statistical problem and correct inference is most challenging in cases where K is larger (e.g., $K = 5$) and the attributes are correlated.

We showed that the crimp sampler proposes irreducible transits between states in different dimensions, and the simulation study indicates the crimp sampler outperforms the other two algorithms. One advantage of the crimp sampler is that the proposed $\boldsymbol{Q}$ always satisfies the identifiability constraints, which ensures that, if the true $\boldsymbol{Q}$ is identified, the posterior can be used to infer the latent structure. Another advantage is that the crimp sampler uses a uniform prior for $\boldsymbol{Q}$ and K from its candidate space and proposes a unique transit between different candidates. It is feasible to adjust the proposed Metropolis–Hastings step for other constraints or restricted candidate space, and the uniform prior for $\boldsymbol{Q}$ and K is still the same. For instance, we can incorporate some expert knowledge as partially fixed $\boldsymbol{Q}$ and propose moves within the restricted identifiable candidate space. Both the crimp sampler and RJ-MCMC use Metropolis–Hastings steps. The main difference between those two is that the crimp sampler applies the properties of Dirichlet processes to integrate the augmented variables of higher dimensions into explicit formulas when K increases and collapses existing variables into lower dimensions when K decreases, whereas the RJ-MCMC calculates the Jacobian matrix of the mapping between different dimensions. Therefore, the crimp sampler is more computationally efficient than RJ-MCMC algorithm. Furthermore, we reported Monte Carlo simulation evidence that the crimp sampler outperformed both the IBP and RJ-MCMC in terms of recovery of the true K. In particular, the crimp sampler demonstrated improved recovery relative to the IBP and RJ-MCMC in more difficult cases where attributes are more correlated and $K = 5$. Additional research is needed to accurately recover $\boldsymbol{Q}$ and K as the sample size and number of attributes increase.

Our method for inferring the DINA $\boldsymbol{Q}$ matrix is applicable in many settings. For instance, the method can be used to cluster multivariate binary data with the DINA model without requiring prespecification of K. Furthermore, the DINA model is a popular CDM in education, and researchers can use our algorithm to estimate the DINA model parameters. In fact, we provided two examples where the exploratory DINA model improved the fit to real data in comparison with more general confirmatory models.

We implemented the algorithms with the DINA model, where the dimensions of the item parameters (slipping and guessing) remain the same when the number of attributes varies. One direction for future research is to extend the crimp sampler to general diagnostic models where the dimensions of item parameters might change along with the varying number of attributes. It is important to note that the algorithms we developed for the IBP and RJ-MCMC can be directly applied to a general diagnostic model. That is, the IBP and RJ-MCMC steps for updating $\boldsymbol{Q}$ and K can be included along with MCMC steps for updating item parameters of general exploratory DCMs (e.g., see Chen et al. 2020b; Culpepper 2019a). The crimp sampler we propose here is also applicable with general restricted latent class models. The main concern with extending the crimp sampler to a general model is the need for a careful choice of priors and corresponding calculations of posteriors in varying dimensions. An additional area of future research is to consider the problem when the true $\boldsymbol{Q}$ is partially identifiable, i.e., the case when only a subset of the columns of $\boldsymbol{Q}$ satisfy the model identifiability conditions. The algorithms developed in this paper can be applied with or without identifiability restrictions, and future research is needed to understand the implications of partial identifiability. Moreover, another topic for future research relates to the fact that the crimp sampler can also be adapted into other frameworks of research problems concerning membership classification with an unknown number of classes.

Additionally, the goal and scope of our paper were to explore options for inferring K in the Bayesian framework. We focused our inquiry on the case where the true $\boldsymbol{Q}$ satisfied model identifiability conditions. In practice, the true $\boldsymbol{Q}$ matrix may be incomplete and additional research is

needed to investigate the impact a non-identifiable $\boldsymbol{Q}$ matrix has on the performance of the proposed algorithms for jointly inferring $\boldsymbol{Q}$ and K. Furthermore, we addressed the fundamental goal of inferring the latent structure, which is necessary for applications of diagnostic models. Additional research is needed to examine the implications of model identifiability on the classification accuracy of exploratory methods.

A final area of future research relates to evaluating the utility of using the latent structure derived from data-driven techniques to inform instructional decisions. Results from an exploratory method can be used to guide theory development for researchers and practitioners. That is, experts may have a belief about $\boldsymbol{Q}$ and the methods described in this paper can be used to validate and possibly uncover new structure. Exploratory diagnostic models have the potential to advance the scientific method in educational research where researchers propose a theory via a $\boldsymbol{Q}$ matrix and test its feasibility.

Acknowledgments

This research was partially supported by National Science Foundation Methodology, Measurement, and Statistics program Grant #1758631 and #1758688.

Appendix A

Reversible Jump MCMC Algorithm

We use reversible jump MCMC to move between $\boldsymbol{Q}$ matrices of different numbers of columns. With the probabilities for "birth" and "death" steps $b_K + d_K = 1$, we have two possible moves:

1. **Birth move**: We propose $K \rightarrow K + 1$ with probability b_K.

To make our algorithm more efficient, we apply the collapsed Gibbs sampling (Liu 1994) by integrating $\boldsymbol{\alpha}_i$ out:

$$P(\boldsymbol{Y}|\boldsymbol{g}, \boldsymbol{s}, \mathbf{Q}, \boldsymbol{\pi}, K) = \prod_{i=1}^{N} \sum_{\boldsymbol{\alpha}_c \in \{0,1\}^K} \pi_c \prod_{j=1}^{J} P(Y_{ij}|s_j, g_j, \boldsymbol{\alpha}_c, \boldsymbol{q}_j),$$

so that we can skip the step of sampling $\boldsymbol{\alpha}_i$.

Following the steps described in Algorithm 4, let $\boldsymbol{Q}_K = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_K)$, we propose the move from $\{\boldsymbol{Q}_K, \boldsymbol{\pi}, K\} \rightarrow \{\boldsymbol{Q}_{K+1}, \boldsymbol{\pi}^*, K + 1\}$, which is

$$\begin{bmatrix} \boldsymbol{q}_1 \\ \vdots \\ \boldsymbol{q}_K \\ \boldsymbol{q}_{K+1} \\ \pi_1 \\ \vdots \\ \pi_C \\ u_1 \\ \vdots \\ u_C \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{q}_1 \\ \vdots \\ \boldsymbol{q}_K \\ \boldsymbol{q}_{K+1} \\ \pi_1^* \\ \vdots \\ \pi_C^* \\ \pi_{C+1}^* \\ \vdots \\ \pi_{2C}^* \end{bmatrix}. \tag{31}$$

The RJ-MCMC acceptance ratio is

$$A = \frac{P(\boldsymbol{Y}|\boldsymbol{g}, \boldsymbol{s}, \boldsymbol{Q}_{K+1}, \boldsymbol{\pi}^*, K+1)}{P(\boldsymbol{Y}|\boldsymbol{g}, \boldsymbol{s}, \boldsymbol{Q}_K, \boldsymbol{\pi}, K)} \times \frac{P(K+1)P(\boldsymbol{\pi}^*|\boldsymbol{\delta}_0^*, K+1)P(\boldsymbol{Q}_{K+1}|K+1)}{P(K)P(\boldsymbol{\pi}|\boldsymbol{\delta}_0, K)P(\boldsymbol{Q}_K|K)}$$
$$\times \frac{d_{K+1}\dfrac{1}{K+1}}{b_K \prod_{j=1}^{J} p(q_{j,K+1}|q_{1,K+1}, \ldots, q_{j-1,K+1}) \prod_{i=1}^{C} p(u_i^*)} \times |J|, \tag{32}$$

where

$$\frac{P(\boldsymbol{\pi}^*|\boldsymbol{\delta}_0^*, K+1)}{P(\boldsymbol{\pi}|\boldsymbol{\delta}_0, K)} = \frac{\Gamma(2^{K+1})}{\Gamma(2^K)}, \qquad (\boldsymbol{\delta}_0 = \mathbf{1}_{2^K}, \boldsymbol{\delta}_0^* = \mathbf{1}_{2^{K+1}}),$$

and the Jacobian matrix J is

$$|J| = \left| \frac{\partial \boldsymbol{\pi}^*}{\partial(\boldsymbol{\pi}, \boldsymbol{u}^*)} \right|$$

$$= \left| \begin{pmatrix} u_1 & & \\ & u_2 & \\ & & \ddots \\ & & u_C \end{pmatrix} \begin{pmatrix} 1-u_1 & & \\ & 1-u_2 & \\ & & \ddots \\ & & 1-u_C \end{pmatrix} \begin{pmatrix} \pi_1 & & \\ & \pi_2 & \\ & & \ddots \\ & & \pi_C \end{pmatrix} \begin{pmatrix} -\pi_1 & & \\ & -\pi_2 & \\ & & \ddots \\ & & -\pi_C \end{pmatrix} \right| \tag{33}$$

$$= \prod_{i=1}^{C} \pi_i.$$

The acceptance probability for the proposed birth move is $\min(1, A)$.

2. **Death move**: We propose $K \rightarrow K-1$ with probability d_K.
We randomly select a column in $\boldsymbol{Q}_K$ to delete. The acceptance probability is $\min(1, A^{-1})$, where A^{-1} can be computed in a similar way as Eq. 32.

Appendix B

Acceptance Ratio of Crimp Sampler

The crimp sampler moves between identifiable spaces with deterministic transformations of the model parameters. The posterior distribution of $\boldsymbol{\xi}^K = (\boldsymbol{Q}_K, \boldsymbol{\pi}_K, \boldsymbol{A}_K, K)$ and $\boldsymbol{\Omega}$ is

$$p(\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{Q}_K, K, \boldsymbol{\Omega}|\boldsymbol{Y}) \propto p(\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{Q}_K, K|\boldsymbol{Y}, \boldsymbol{\Omega})p(\boldsymbol{\Omega}), \tag{34}$$

where

$$\begin{aligned} p(\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{Q}_K, K|\boldsymbol{Y}, \boldsymbol{\Omega}) &\propto p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_K)p(\boldsymbol{A}_K|\boldsymbol{\pi}_K)p(\boldsymbol{\pi}_K)p(\boldsymbol{Q}_K, K) \\ &\propto p(\boldsymbol{A}_K, \boldsymbol{\pi}|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{Q}_K)p(\boldsymbol{Q}_K, K). \end{aligned} \tag{35}$$

Let $\boldsymbol{u} = (u_0, \ldots, u_{2^K-1})$ be a vector of independent random variables. The conditional posterior of $\boldsymbol{\xi}^K$ and $\boldsymbol{u}$ is

$$p(\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{u}, \boldsymbol{Q}_K, K|\boldsymbol{Y}, \boldsymbol{\Omega}) \propto p(\boldsymbol{A}_K, \boldsymbol{\pi}_K|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{Q}_K)p(\boldsymbol{Q}_K, K)p(\boldsymbol{u}). \tag{36}$$

Note that for $\boldsymbol{\pi}_K \sim \text{Dirichlet}(\boldsymbol{d}_K)$ where $\boldsymbol{d}_K = (d_0, \ldots, d_{2^K-1})$,

$$p(\boldsymbol{A}_K|\boldsymbol{\pi}_K)p(\boldsymbol{\pi}_K) = C_K \prod_{c=0}^{2^K-1} \pi_c^{n_c+d_c-1}, \tag{37}$$

where n_c is the number of respondents in class c and the normalizing constant is

$$C_K = \frac{\Gamma(\sum_{c=0}^{2^K-1} d_c)}{\prod_{c=0}^{2^K-1} \Gamma(d_c)}. \tag{38}$$

Moves from $K \rightarrow K+1$ involve transiting from $\boldsymbol{\xi}^K = (\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{Q}_K, K, \boldsymbol{u}) \rightarrow \boldsymbol{\xi}^{K+1} = (\boldsymbol{A}_{K+1}, \boldsymbol{\pi}_{K+1}, \boldsymbol{Q}_{K+1}, K+1)$. That is, the move to a higher dimension involves a transformation of $\boldsymbol{\pi}_K$ and $\boldsymbol{u}$ into $\boldsymbol{\pi}_{K+1}$. Specifically, let

$$\begin{aligned} \pi_{c0} &= \pi_c(1-u_c), \\ \pi_{c1} &= \pi_c u_c, \end{aligned} \tag{39}$$

where π_{c0} and π_{c1} are elements of $\boldsymbol{\pi}_{K+1}$ for $c = 0, \ldots, 2^K - 1$. Note the Jacobian of transformation equals $|J| = \prod_{c=0}^{2^K-1}(\pi_{c0} + \pi_{c1})^{-1}$.
Let $\boldsymbol{a}_{K+1} = (\alpha_{1,K+1}, \ldots, \alpha_{N,K+1})^\top$ be an N vector of the proposed $K+1$ binary attributes. The full conditional distribution for the proposed $K+1$ state prior to transformation is

$$\begin{aligned} &p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}, \boldsymbol{Q}_{K+1}, K+1|\boldsymbol{Y}, \boldsymbol{\Omega}) \\ &\quad \propto p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1})p(\boldsymbol{a}_{K+1}|\boldsymbol{A}_K, \boldsymbol{u})p(\boldsymbol{A}_K|\boldsymbol{\pi}_K)p(\boldsymbol{\pi}_K)p(\boldsymbol{u})p(\boldsymbol{Q}_{K+1}, K+1), \end{aligned} \tag{40}$$

where $p(\boldsymbol{a}_{K+1}|\boldsymbol{A}_K, \boldsymbol{u}) = \prod_{i=1}^{n} \sum_{c=0}^{2^K-1} \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) \left(u_c^{\alpha_{i,K+1}}(1-u_c)^{1-\alpha_{i,K+1}}\right)$.
Let $u_c \sim \text{Beta}(1,1)$ for $c = 0, \ldots, 2^K - 1$ and consider $\boldsymbol{d} = \mathbf{1}_{2^K}$. The conditional prior distribution of $\boldsymbol{a}_{K+1}$, $\boldsymbol{A}_K$, $\boldsymbol{\pi}_K$, and $\boldsymbol{u}$ given $\boldsymbol{Q}_{K+1}$ is

$$p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}) \propto \left(\prod_{c=0}^{2^K-1} u_c^{n_{c1}}(1-u_c)^{n_{c0}}\right) C_K \prod_{c=0}^{2^K-1} \pi_c^{n_c+d_c-1}, \tag{41}$$

where $n_{c0} + n_{c1} = n_c$, n_{c0} is the number of individuals in class c (for attributes 1 to K) with attribute $K+1$ equal to 0, and n_{c1} corresponds to the number of individuals in class c with attribute $K+1$ equal to 1. Applying the transformation implies

$$\begin{aligned} p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_{K+1}) &\propto C_K \prod_{c=0}^{2^K-1} \left(\frac{\pi_{c1}}{\pi_{c0}+\pi_{c1}}\right)^{n_{c1}} \left(\frac{\pi_{c0}}{\pi_{c0}+\pi_{c1}}\right)^{n_{c0}} (\pi_{c0}+\pi_{c1})^{n_c} \times |J_c| \\ &= C_K \prod_{c=0}^{2^K-1} \pi_{c1}^{n_{c1}} \pi_{c0}^{n_{c0}} (\pi_{c0}+\pi_{c1})^{-1}. \end{aligned} \tag{42}$$

The transformed conditional posterior distribution is therefore

$$\begin{aligned} &p(\boldsymbol{A}_{K+1}, \boldsymbol{\pi}_{K+1}, \boldsymbol{Q}_{K+1}, K+1|\boldsymbol{Y}, \boldsymbol{\Omega}) \\ &\quad \propto p(\boldsymbol{Y}|\boldsymbol{A}_{K+1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) p(\boldsymbol{A}_{K+1}, \boldsymbol{\pi}_{K+1}) p(\boldsymbol{Q}_{K+1}, K+1), \end{aligned} \tag{43}$$

where $\boldsymbol{A}_{K+1} = (\boldsymbol{A}_K, \boldsymbol{a}_{K+1})$.
The transformed posterior for $\boldsymbol{\xi}^{K+1}$ requires we propose values for $\boldsymbol{a}_{K+1}$ and the 2^K $\boldsymbol{u}$, which form the additional elements of $\boldsymbol{\pi}_{K+1}$ (in fact, RJ-MCMC transitions between spaces this way). Proposals for these parameters must be carefully constructed to ensure proper mixing and acceptance rates. Rather than proposing values we integrate out $\boldsymbol{a}_{K+1}$ and $\boldsymbol{u}$. That is, we implement the inverse transformations

$$\begin{aligned} \pi_c &= \pi_{c0} + \pi_{c1}, \\ u_c &= \frac{\pi_{c1}}{\pi_{c0}+\pi_{c1}}, \end{aligned} \tag{44}$$

sum over $\boldsymbol{a}_{K+1}$, and integrate out $\boldsymbol{u}$ to find,

$$\begin{aligned} &p(\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{Q}_{K+1}, K+1|\boldsymbol{Y}, \boldsymbol{\Omega}) \\ &\propto \int \sum_{\alpha_{1,K+1}=0}^{1} \cdots \sum_{\alpha_{N,K+1}=0}^{1} p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}, \boldsymbol{Q}_{K+1}, K+1|\boldsymbol{Y}, \boldsymbol{\Omega}) d\boldsymbol{u} \\ &= \left\{ \int \left[\prod_{i=1}^{N} \sum_{\alpha_{i,K+1}=0}^{1} p(\boldsymbol{Y}_i|\boldsymbol{\alpha}_i, \alpha_{i,K+1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) p(\alpha_{i,K+1}|\boldsymbol{u}, \boldsymbol{\alpha}_i) \right] p(\boldsymbol{u}) d\boldsymbol{u} \right\} \\ &\quad \times p(\boldsymbol{A}_K|\boldsymbol{\pi}_K) p(\boldsymbol{\pi}_K) p(\boldsymbol{Q}_{K+1}, K+1) \\ &= p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) p(\boldsymbol{A}_K|\boldsymbol{\pi}_K) p(\boldsymbol{\pi}_K) p(\boldsymbol{Q}_{K+1}, K+1). \end{aligned} \tag{45}$$

Therefore, for $K \rightarrow K+1$, with the uniform prior for $\boldsymbol{Q}_K$ and K, all terms but the likelihood functions will cancel out in the MH acceptance ratio, i.e.,

$$\frac{p(\boldsymbol{Q}^*_{K+1}|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}_K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}^*_{K+1})}{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_K)}. \tag{46}$$

Transitions to lower dimensions involve a move from $\boldsymbol{\xi}^K \rightarrow \boldsymbol{\xi}^{K-1} = (\boldsymbol{A}_{K-1}, \boldsymbol{\pi}_{K-1}, \boldsymbol{Q}_{K-1}, K-1)$. Without loss of generality, suppose the Kth attribute is proposed to be deleted. Therefore, $\boldsymbol{\alpha}_i = (\tilde{\boldsymbol{\alpha}}_i^\top, \alpha_{iK})^\top$ and $\boldsymbol{A}_{K-1} = (\tilde{\boldsymbol{\alpha}}_1, \ldots, \tilde{\boldsymbol{\alpha}}_N)^\top$.
An important observation is that deleting a column of $\boldsymbol{Q}$ implies an attribute is no longer related to observed responses, which implies

$$p(\boldsymbol{Y}_i|\tilde{\boldsymbol{\alpha}}_i, \alpha_{iK}=0, \boldsymbol{\Omega}, Q_{K-1}) = p(\boldsymbol{Y}_i|\tilde{\boldsymbol{\alpha}}_i, \alpha_{iK}=1, \boldsymbol{\Omega}, Q_{K-1}). \tag{47}$$

The conditional posterior after summing over $\boldsymbol{a}_K$ is

$$\begin{aligned}
&p(\boldsymbol{A}_{K-1}, \boldsymbol{\pi}_K, \boldsymbol{Q}_{K-1}, K-1|\boldsymbol{Y}, \boldsymbol{\Omega}) \\
&\quad\propto \left[\prod_{i=1}^{N} \sum_{\alpha_{iK}=0}^{1} p(\boldsymbol{Y}_i|\tilde{\boldsymbol{\alpha}}_i, \alpha_{iK}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K-1}) p(\boldsymbol{\alpha}_i|\boldsymbol{\pi}_K)\right] p(\boldsymbol{\pi}_K) p(\boldsymbol{Q}_{K-1}, K-1) \\
&\quad= \left[\prod_{i=1}^{N} p(\boldsymbol{Y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Omega}, \boldsymbol{Q}_{K-1}) \sum_{\alpha_{iK}=0}^{1} p(\boldsymbol{\alpha}_i|\boldsymbol{\pi}_K)\right] p(\boldsymbol{\pi}_K) p(\boldsymbol{Q}_{K-1}, K-1) \\
&\quad= p(\boldsymbol{Y}|\boldsymbol{A}_{K-1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K-1}) p(\boldsymbol{A}_{K-1}|\boldsymbol{\pi}_K) p(\boldsymbol{\pi}_K) p(\boldsymbol{Q}_{K-1}, K-1),
\end{aligned} \tag{48}$$

where

$$p(\boldsymbol{A}_{K-1}|\boldsymbol{\pi}_K) = \prod_{c=0}^{2^{K-1}-1} (\pi_{c0} + \pi_{c1})^{n_c} \tag{49}$$

and n_c is the number of individuals in the collapsed class c. The joint conditional prior for $\boldsymbol{A}_{K-1}$ and $\boldsymbol{\pi}_K$ is

$$p(\boldsymbol{A}_{K-1}, \boldsymbol{\pi}_K) = p(\boldsymbol{A}_{K-1}|\boldsymbol{\pi}_K) p(\boldsymbol{\pi}_K) = \left(\prod_{c=0}^{2^{K-1}-1} (\pi_{c0} + \pi_{c1})^{n_c}\right) \left(C_K \prod_{c=0}^{2^{K-1}-1} \pi_{c0}^{d_{c0}-1} \pi_{c1}^{d_{c1}-1}\right). \tag{50}$$

We then transform $\boldsymbol{\pi}_K$ by defining

$$\begin{aligned}
\pi_c &= \pi_{c0} + \pi_{c1}, \\
u_c &= \frac{\pi_{c1}}{\pi_{c0} + \pi_{c1}},
\end{aligned} \tag{51}$$

for $c = 0, \ldots, 2^{K-1} - 1$ where π_c is an element of $\boldsymbol{\pi}_{K-1}$. The transformation implies

$$\begin{aligned} p(\boldsymbol{A}_{K-1}, \boldsymbol{u}, \boldsymbol{\pi}_{K-1}) &= \left(\prod_{c=0}^{2^{K-1}-1} \pi_c^{n_c} \right) \left(C_K \prod_{c=0}^{2^{K-1}-1} \pi_c^{d_{c1}+d_{c0}-2} u_c^{d_{c1}-1} (1-u_c)^{d_{c0}-1} \right) |J|^{-1} \\ &= C_K \left(\prod_{c=0}^{2^{K-1}-1} \pi_c^{n_c+d_{c0}+d_{c1}-1} \right) \left(\prod_{c=0}^{2^{K-1}-1} u_c^{d_{c1}-1} (1-u_c)^{d_{c0}-1} \right) \\ &= p(\boldsymbol{A}_{K-1}|\boldsymbol{\pi}_{K-1}) p(\boldsymbol{\pi}_{K-1}) p(\boldsymbol{u}). \end{aligned} \tag{52}$$

Therefore, integrating over $\boldsymbol{u}$ yields

$$\begin{aligned} & p(\boldsymbol{A}_{K-1}, \boldsymbol{\pi}_{K-1}, \boldsymbol{Q}_{K-1}, K-1|\boldsymbol{Y}, \boldsymbol{\Omega}) \\ \propto & \int p(\boldsymbol{A}_{K-1}, \boldsymbol{\pi}_{K-1}, \boldsymbol{u}, \boldsymbol{Q}_{K-1}, K-1|\boldsymbol{Y}, \boldsymbol{\Omega}) d\boldsymbol{u} \\ = & p(\boldsymbol{Y}|\boldsymbol{A}_{K-1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K-1}) p(\boldsymbol{A}_{K-1}|\boldsymbol{\pi}_{K-1}) p(\boldsymbol{\pi}_{K-1}) p(\boldsymbol{Q}_{K-1}, K-1) \int p(\boldsymbol{u}) d\boldsymbol{u}, \end{aligned} \tag{53}$$

where

$$\int p(\boldsymbol{u}) d\boldsymbol{u} = \prod_{c=0}^{2^{K-1}-1} \int u_c^{d_{c1}-1} (1-u_c)^{d_{c0}-1} du_c = \prod_{c=0}^{2^{K-1}-1} \text{Beta}(d_{c0}, d_{c1}). \tag{54}$$

If the prior for $\boldsymbol{\pi}_K \sim \text{Dirichlet}(\mathbf{1}_{2^K})$, then the integral is one. Since we use the uniform prior of $p(\boldsymbol{Q}, K)$, the acceptance ratio becomes

$$\frac{p(\boldsymbol{Q}^*_{K-1}|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})}{p(\boldsymbol{Q}_K|\boldsymbol{A}_K, \boldsymbol{\pi}_K, \boldsymbol{\Omega}, \boldsymbol{Y})} = \frac{p(\boldsymbol{Y}|\boldsymbol{A}_{K-1}, \boldsymbol{\Omega}, \boldsymbol{Q}^*_{K-1}) \left(\prod_{c=0}^{2^{K-1}-1} \pi_c{}^{n_c+1} \right)}{p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{\Omega}, \boldsymbol{Q}_K) \left(\prod_{c=0}^{2^{K-1}-1} \pi_{c1}^{n_{c1}} \pi_{c0}^{n_{c0}} \right)}. \tag{55}$$

Appendix C

Gibbs Sampling Step for Updating $\boldsymbol{\alpha}$ in Crimp Sampler

The conditional posterior distribution of $\boldsymbol{a}_{K+1}$, $\boldsymbol{A}_K$, $\boldsymbol{\pi}_K$, and $\boldsymbol{u}$ is

$$p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \propto p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \cdot p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}); \tag{56}$$

then, the marginal posterior of $\boldsymbol{a}_{K+1}$ and $\boldsymbol{A}_K$ is

$$p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}|\boldsymbol{Y}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \propto p(\boldsymbol{Y}|\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \cdot \int \int p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}) d\boldsymbol{\pi}_K d\boldsymbol{u},$$

where

$$p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}, \boldsymbol{\pi}_K, \boldsymbol{u}) = p(\boldsymbol{A}_K|\boldsymbol{\pi}_K) \cdot p(\boldsymbol{\pi}_K) \cdot p(\boldsymbol{u}) \cdot p(\boldsymbol{a}_{K+1}|\boldsymbol{A}_K, \boldsymbol{u}) \propto \left(\prod_{c=0}^{2^K-1} u_c^{n_{c1}} (1-u_c)^{n_{c0}} \right) \prod_{c=0}^{2^K-1} \pi_c^{n_c+d_c-1}.$$

Therefore,

$$\begin{aligned} p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}) &\propto \int \left(\prod_{c=0}^{2^K-1} u_c^{n_{c1}} (1-u_c)^{n_{c0}} \right) d\boldsymbol{u} \int \left(\prod_{c=0}^{2^K-1} \pi_c^{n_c+d_c-1} \right) d\boldsymbol{\pi}_K \\ &\propto \left(\prod_{c=0}^{2^K-1} B(n_{c1}+1, n_{c0}+1) \right) \cdot \frac{\prod_{c=0}^{2^K-1} \Gamma(n_c+1)}{\Gamma(N+2^K)}. \end{aligned} \tag{57}$$

Let n'_c be the number of individuals other than i with K attributes equal to $\boldsymbol{\alpha}_c$, and let n'_{c0} and n'_{c1} be the number of such individuals with $\alpha_{c,K+1} = 0$ and $\alpha_{c,K+1} = 1$, respectively. So $n'_c = n'_{c0} + n'_{c1}$. That is,

$$\begin{aligned} n_c &= n'_c + \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c), \\ n_{c0} &= n'_{c0} + (1-\alpha_{i,K+1}) \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c), \\ n_{c1} &= n'_{c1} + \alpha_{i,K+1} \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c). \end{aligned}$$

Recall the following properties for the beta and gamma functions:

$$\begin{aligned} B(x+1, y) &= B(x, y) \cdot \frac{x}{x+y}, \\ B(x, y+1) &= B(x, y) \cdot \frac{y}{x+y}, \\ \Gamma(x+1) &= x \cdot \Gamma(x). \end{aligned}$$

Then, Eq. 57 can be written as:

$$\begin{aligned} &p(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N, \alpha_{1,K+1}, \ldots, \alpha_{N,K+1}) \propto \\ &\left(\prod_{c=0}^{2^K-1} B(n'_{c1} + \alpha_{i,K+1} \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) + 1, n'_{c0} + (1-\alpha_{i,K+1}) \mathcal{I}(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c) + 1) \right) \cdot \frac{\prod_{c=0}^{2^K-1} \Gamma(n_c+1)}{\Gamma(N+2^K)}. \end{aligned} \tag{58}$$

Let $\boldsymbol{A}_{(i)} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_N)$ and $\boldsymbol{\alpha}_{(i),K+1} = (a_{1,K+1}, \ldots, \alpha_{i-1,K+1}, \alpha_{i+1,K+1}, \cdots, a_{N,K+1})$. By integrating out $\boldsymbol{\alpha}_i$ and $\alpha_{i,K+1}$, we have

$$\begin{aligned} &p(\boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) = \\ &\sum_{c=0}^{2^K-1} \left[p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \alpha_{i,K+1} = 0, \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) + p(\alpha_i^\top \boldsymbol{v} = c, \alpha_{i,K+1} = 1, \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) \right]. \end{aligned} \tag{59}$$

Note that

$$B(n'_{c1}+1, n'_{c0}+2) = \frac{n'_{c0}+1}{n'_c+2} B(n'_{c1}+1, n'_{c0}+1),$$
$$B(n'_{c1}+2, n'_{c0}+1) = \frac{n'_{c1}+1}{n'_c+2} B(n'_{c1}+1, n'_{c0}+1),$$
$$\Gamma(n'_c+2) = (n'_c+1)\Gamma(n'_c+1),$$

which implies

$$\begin{aligned} &p(\alpha_i^\top \boldsymbol{v} = c, \alpha_{i,K+1} = 0, \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) \\ =&\frac{n'_{c0}+1}{n'_c+2} \left(\prod_{c=0}^{2^K-1} B(n'_{c1}+1, n'_{c0}+1) \right) \cdot \frac{(n'_c+1)\prod_{c=0}^{2^K-1}\Gamma(n'_c+1)}{\Gamma(N+2^K)}, \end{aligned}$$

and

$$\begin{aligned} &p(\alpha_i^\top \boldsymbol{v} = c, \alpha_{i,K+1} = 1, \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) \\ =&\frac{n'_{c1}+1}{n'_c+2} \left(\prod_{c=0}^{2^K-1} B(n'_{c1}+1, n'_{c0}+1) \right) \cdot \frac{(n'_c+1)\prod_{c=0}^{2^K-1}\Gamma(n'_c+1)}{\Gamma(N+2^K)}. \end{aligned}$$

Then, Eq. 59 should be

$$\begin{aligned} p(\boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) &= \left(\prod_{c=0}^{2^K-1} B(n'_{c1}+1, n'_{c0}+1) \right) \cdot \frac{\prod_{c=0}^{2^K-1}\Gamma(n'_c+1)}{\Gamma(N+2^K)} \left(\sum_{c=0}^{2^K-1} (n'_c+1) \right) \\ &= \left(\prod_{c=0}^{2^K-1} B(n'_{c1}+1, n'_{c0}+1) \right) \cdot \frac{\prod_{c=0}^{2^K-1}\Gamma(n'_c+1)}{\Gamma(N+2^K-1)}. \end{aligned} \tag{60}$$

Therefore, the full conditional distribution for $\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c$ and $\alpha_{i,K+1} = 1$ is

$$\begin{aligned} p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \alpha_{i,K+1}+1 | \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1}) &= \frac{p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \alpha_{i,K+1}+1, \boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1})}{p(\boldsymbol{A}_{(i)}, \boldsymbol{\alpha}_{(i),K+1})} \\ &= \frac{n'_{c1}+1}{n'_c+2} \cdot \frac{n'_c+1}{N+2^K-1}. \end{aligned} \tag{61}$$

In order to update $\alpha_{i,K+1}$ sequentially, we need to find

$$p(\boldsymbol{\alpha}_i, \alpha_{i,K+1} | \boldsymbol{A}_{(i)}, \alpha_{1,K+1}, \ldots, \alpha_{i-1,K+1}) = \frac{p(\boldsymbol{A}_K, \alpha_{1,K+1}, \ldots, \alpha_{i,K+1})}{p(\boldsymbol{A}_{(i)}, \alpha_{1,K+1}, \ldots, \alpha_{i-1,K+1})}. \tag{62}$$

The probability in the numerator is calculated by summing Eq. 57 over $\alpha_{i+1,K+1}, \ldots, \alpha_{N,K+1}$:

$$p(\boldsymbol{A}_K, a_{1,K+1}, \ldots, a_{i,K+1})$$

$$
\begin{aligned}
&= \sum_{\alpha_{i+1,K+1}=0}^{1} \cdots \sum_{\alpha_{N,K+1}=0}^{1} p(\boldsymbol{A}_K, \boldsymbol{a}_{K+1}) \\
&= \frac{\prod_{c=0}^{2^K-1} \Gamma(n_{c,N}+1)}{\Gamma(N+2^K)} \cdot \sum_{\alpha_{i+1,K+1}=0}^{1} \cdots \sum_{\alpha_{N,K+1}=0}^{1} \left(\prod_{c=0}^{2^K-1} B(n_{c1,N}+1, n_{c0,N}+1) \right). \qquad (63)
\end{aligned}
$$

Given that

$$
\begin{aligned}
n_{c,N} &= n_{c,N-1} + \mathcal{I}(\boldsymbol{\alpha}_N^\top \boldsymbol{v} = c), \\
n_{c1,N} &= n_{c1,N-1} + \alpha_{N,K+1} \cdot \mathcal{I}(\boldsymbol{\alpha}_N^\top \boldsymbol{v} = c), \\
n_{c0,N} &= n_{c0,N-1} + (1-\alpha_{N,K+1}) \cdot \mathcal{I}(\boldsymbol{\alpha}_N^\top \boldsymbol{v} = c),
\end{aligned}
$$

we have

$$
\begin{aligned}
&\sum_{\alpha_{N,K+1}=0}^{1} \left(\prod_{c=0}^{2^K-1} B(n_{c1,N}+1, n_{c0,N}+1) \right) \\
&= \prod_{c=0}^{2^K-1} B(n_{c1,N-1}+1, n_{c0,N-1}+\mathcal{I}(\boldsymbol{\alpha}_N^\top \boldsymbol{v} = c)+1) + \prod_{c=0}^{2^K-1} B(n_{c1,N-1}+\mathcal{I}(\boldsymbol{\alpha}_N^\top \boldsymbol{v} = c)+1, n_{c0,N-1}+1) \\
&= \left[\prod_{c=0}^{2^K-1} B(n_{c1,N-1}+1, n_{c0,N-1}+1) \right] \left(\frac{n_{c0,N-1}+1}{n_{c1,N-1}+n_{c0,N-1}+2} + \frac{n_{c1,N-1}+1}{n_{c1,N-1}+n_{c0,N-1}+2} \right) \\
&= \prod_{c=0}^{2^K-1} B(n_{c1,N-1}+1, n_{c0,N-1}+1).
\end{aligned}
$$

Therefore, Eq. 63 can be written as:

$$
p(\boldsymbol{A}_K, \alpha_{1,K+1}, \ldots, \alpha_{i,K+1}) = \frac{\prod_{c=0}^{2^K-1} \Gamma(n_{c,N}+1)}{\Gamma(N+2^K)} \cdot \prod_{c=0}^{2^K-1} B(n_{c1,i}+1, n_{c0,i}+1). \qquad (64)
$$

For the probability in the denominator,

$$
\begin{aligned}
&p(\boldsymbol{A}_{(i)}, \alpha_{1,K+1}, \ldots, \alpha_{i-1,K+1}) \\
&= \sum_{\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c^* = 0}^{2^K-1} \sum_{\alpha_{i-1,K+1}=0}^{1} p(\boldsymbol{A}_K, \alpha_{1,K+1}, \ldots, \alpha_{i,K+1}) \\
&= \sum_{c^*=0}^{2^K-1} \left[\frac{\prod_{c=0}^{2^K-1} \Gamma(n_{c,N}+1)}{\Gamma(N+2^K)} \cdot \prod_{c=0}^{2^K-1} B(n_{c1,i-1}+1, n_{c0,i-1}+1) \right] \\
&= \frac{\prod_{c=0}^{2^K-1} B(n_{c1,i-1}+1, n_{c0,i-1}+1)}{\Gamma(N+2^K)} \cdot \sum_{c^*=0}^{2^K-1} (n'_{c^*,N}+1) \cdot \prod_{c=0}^{2^K-1} \Gamma(n'_{c,N}+1)
\end{aligned}
$$

$$= \left(\prod_{c=0}^{2^K-1} B(n_{c1,i-1}+1, n_{c0,i-1}+1) \right) \cdot \frac{\prod_{c=0}^{2^K-1} \Gamma(n'_{c,N}+1)}{\Gamma(N+2^K-1)}. \tag{65}$$

Combining Eqs. 64 and 65, Eq. 62 is

$$\begin{aligned} & p(\boldsymbol{\alpha}_i^\top \boldsymbol{v}, \alpha_{i,K+1} | \boldsymbol{A}_{(i)}, \alpha_{1,K+1}, \cdots, \alpha_{i-1,K+1}) \\ =& \frac{\frac{\prod_{c=0}^{2^K-1} \Gamma(n_{c,N}+1)}{\Gamma(N+2^K)} \cdot \prod_{c=0}^{2^K-1} B(n_{c1,i}+1, n_{c0,i}+1)}{\left(\prod_{c=0}^{2^K-1} B(n_{c1,i-1}+1, n_{c0,i-1}+1) \right) \cdot \frac{\prod_{c=0}^{2^K-1} \Gamma(n'_{c,N}+1)}{\Gamma(N+2^K-1)}} \\ =& \frac{1}{N+2^K-1} \cdot \left(\prod_{c=0}^{2^K-1} \frac{B(n_{c1,i}+1, n_{c0,i}+1)\Gamma(n_{c,N}+1)}{B(n_{c1,i-1}+1, n_{c0,i-1}+1)\Gamma(n'_{c,N}+1)} \right). \end{aligned} \tag{66}$$

If $\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c$ and $\alpha_{i,K+1} = 1$, then Eq. 66 can be simplified as

$$\frac{n_{c1,i-1}+1}{n_{c,i-1}+2} \cdot \frac{n'_c+1}{N+2^K-1},$$

and if $\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c$ and $\alpha_{i,K+1} = 0$, Eq. 66 is

$$\frac{n_{c0,i-1}+1}{n_{c,i-1}+2} \cdot \frac{n'_c+1}{N+2^K-1}.$$

Therefore, we can update $(\boldsymbol{\alpha}_i, \alpha_{i,K+1})$ sequentially to $(\boldsymbol{\alpha}_c, \alpha^*)$ with probability proportional to $p(\boldsymbol{Y}_i | \boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \alpha_{i,K+1} = 1, \boldsymbol{\Omega}, \boldsymbol{Q}_{K+1}) \cdot \left(\frac{n_{c\alpha^*,i-1}+1}{n_{c,i-1}+2} \cdot \frac{n'_c+1}{N+2^K-1} \right)$, $c = 0, \ldots, 2^K$, $\alpha^* = 0, 1$.

Appendix D

Posterior Distribution of λ in Indian Buffet Process Algorithm

Assume $\lambda \sim Gamma(a, b)$, and λ is only related to $\boldsymbol{Q}$, so its posterior can be updated by

$$\begin{aligned} p(\lambda^{(t)} | \boldsymbol{Q}^{(t)}) &\propto p(\boldsymbol{Q}^{(t)} | \lambda) p(\lambda) \\ &\propto \lambda^K \exp\{-\lambda H_N\} \lambda^{a-1} \exp\{-\lambda b\} \\ &\propto \lambda^{K+a-1} \exp\{-\lambda(H_N + b)\}, \end{aligned} \tag{67}$$

then $\lambda^{(t)}$ can be sampled from Gamma$(K+a, H_N+b)$, where K is the number of features in current $\boldsymbol{Q}^{(t)}$ and H_N the Nth harmonic number.

TABLE 8.
Summary of RMSE of estimated item parameters for $K = 3$, $\rho = 0$ by sample size N.

Crimp sampler						IBP						RJMCMC					
$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$	
$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$
0.027	0.018	0.024	0.011	0.019	0.009	0.035	0.066	0.028	0.042	0.134	0.140	0.032	0.031	0.021	0.024	0.013	0.014
0.030	0.042	0.020	0.026	0.011	0.012	0.029	0.055	0.030	0.048	0.135	0.134	0.027	0.034	0.023	0.021	0.017	0.016
0.034	0.020	0.028	0.019	0.017	0.016	0.031	0.052	0.030	0.046	0.128	0.139	0.029	0.030	0.022	0.021	0.014	0.014
0.054	0.025	0.029	0.024	0.018	0.011	0.034	0.067	0.032	0.047	0.134	0.137	0.028	0.028	0.020	0.023	0.014	0.015
0.028	0.025	0.024	0.019	0.020	0.012	0.032	0.055	0.027	0.047	0.138	0.134	0.031	0.030	0.023	0.024	0.014	0.014
0.039	0.024	0.025	0.017	0.022	0.016	0.030	0.048	0.032	0.044	0.130	0.138	0.032	0.029	0.023	0.021	0.013	0.014
0.027	0.030	0.015	0.019	0.012	0.016	0.032	0.066	0.031	0.040	0.132	0.138	0.029	0.032	0.020	0.021	0.015	0.013
0.039	0.024	0.028	0.019	0.017	0.013	0.034	0.051	0.029	0.049	0.135	0.131	0.035	0.030	0.024	0.020	0.015	0.015
0.031	0.058	0.026	0.032	0.012	0.014	0.030	0.050	0.028	0.042	0.129	0.136	0.032	0.033	0.023	0.023	0.015	0.014
0.029	0.047	0.015	0.015	0.011	0.009	0.047	0.025	0.030	0.030	0.221	0.020	0.045	0.023	0.028	0.016	0.020	0.010
0.044	0.018	0.028	0.023	0.018	0.016	0.041	0.031	0.032	0.028	0.220	0.022	0.037	0.020	0.029	0.015	0.021	0.012
0.035	0.019	0.012	0.009	0.011	0.010	0.042	0.031	0.030	0.021	0.218	0.022	0.042	0.022	0.032	0.017	0.022	0.011
0.039	0.032	0.030	0.036	0.029	0.023	0.051	0.030	0.030	0.022	0.224	0.021	0.040	0.023	0.031	0.015	0.021	0.011
0.054	0.043	0.036	0.020	0.032	0.017	0.036	0.032	0.030	0.028	0.216	0.025	0.043	0.023	0.028	0.015	0.021	0.010
0.037	0.029	0.015	0.025	0.011	0.018	0.048	0.030	0.041	0.020	0.219	0.021	0.040	0.021	0.027	0.016	0.022	0.012
0.029	0.025	0.018	0.016	0.017	0.015	0.051	0.020	0.037	0.017	0.250	0.033	0.047	0.020	0.041	0.014	0.028	0.011
0.049	0.020	0.026	0.019	0.018	0.013	0.054	0.021	0.042	0.015	0.249	0.031	0.052	0.019	0.036	0.014	0.027	0.010
0.044	0.027	0.020	0.019	0.014	0.019	0.056	0.021	0.039	0.014	0.249	0.032	0.056	0.021	0.034	0.012	0.027	0.011

Results are based on replications for which $\hat{K}$ is correctly estimated.

Appendix E

Problem set of Elementary Probability Theory

The twelve questions from R package pks (Heller and Wickelmaier 2013) are

1. A box contains 30 marbles in the following colors: 8 red, 10 black, 12 yellow. What is the probability that a randomly drawn marble is yellow?
2. A bag contains 5-cent, 10-cent, and 20-cent coins. The probability of drawing a 5-cent coin is 0.35, that of drawing a 10-cent coin is 0.25, and that of drawing a 20-cent coin is 0.40. What is the probability that the coin randomly drawn is not a 5-cent coin?
3. A bag contains 5-cent, 10-cent, and 20-cent coins. The probability of drawing a 5-cent coin is 0.20, that of drawing a 10-cent coin is 0.45, and that of drawing a 20-cent coin is 0.35. What is the probability that the coin randomly drawn is a 5-cent coin or a 20-cent coin?
4. In a school, 40% of the pupils are boys and 80% of the pupils are right-handed. Suppose that gender and handedness are independent. What is the probability of randomly selecting a right-handed boy?
5. Given a standard deck containing 32 different cards, what is the probability of not drawing a heart?
6. A box contains 20 marbles in the following colors: 4 white, 14 green, 2 red. What is the probability that a randomly drawn marble is not white?
7. A box contains 10 marbles in the following colors: 2 yellow, 5 blue, 3 red. What is the probability that a randomly drawn marble is yellow or blue?
8. What is the probability of obtaining an even number by throwing a dice?
9. Given a standard deck containing 32 different cards, what is the probability of drawing a 4 in a black suit?

TABLE 9.
Summary of RMSE of estimated item parameters for $K = 4$, $\rho = 0$ by sample size N.

Crimp sampler						IBP						RJMCMC					
$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$	
$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$
0.019	0.058	0.017	0.032	0.013	0.023	0.047	0.058	0.071	0.085	0.197	0.149	0.034	0.035	0.020	0.025	0.021	0.019
0.039	0.018	0.028	0.021	0.021	0.013	0.045	0.056	0.066	0.079	0.194	0.151	0.038	0.033	0.023	0.024	0.018	0.019
0.031	0.017	0.019	0.014	0.010	0.012	0.058	0.059	0.057	0.082	0.192	0.152	0.031	0.037	0.023	0.022	0.014	0.019
0.031	0.043	0.040	0.037	0.029	0.028	0.052	0.062	0.059	0.068	0.193	0.153	0.032	0.035	0.025	0.023	0.014	0.019
0.018	0.045	0.019	0.021	0.019	0.020	0.049	0.057	0.074	0.081	0.198	0.149	0.034	0.034	0.022	0.028	0.018	0.018
0.019	0.020	0.018	0.013	0.010	0.015	0.051	0.059	0.068	0.085	0.195	0.151	0.037	0.031	0.025	0.022	0.016	0.017
0.036	0.027	0.038	0.028	0.045	0.029	0.060	0.065	0.049	0.079	0.193	0.153	0.031	0.031	0.024	0.026	0.013	0.021
0.037	0.034	0.034	0.028	0.24	0.016	0.046	0.068	0.053	0.071	0.194	0.151	0.034	0.037	0.022	0.026	0.018	0.017
0.044	0.26	0.035	0.016	0.022	0.010	0.072	0.029	0.101	0.027	0.330	0.029	0.044	0.024	0.033	0.018	0.025	0.011
0.102	0.045	0.070	0.042	0.037	0.021	0.080	0.030	0.090	0.028	0.329	0.027	0.051	0.022	0.039	0.016	0.022	0.011
0.043	0.015	0.023	0.020	0.012	0.019	0.072	0.030	0.093	0.025	0.326	0.030	0.048	0.026	0.038	0.015	0.025	0.013
0.033	0.026	0.025	0.021	0.015	0.016	0.083	0.027	0.077	0.025	0.325	0.028	0.045	0.025	0.031	0.015	0.018	0.013
0.063	0.025	0.036	0.027	0.013	0.012	0.075	0.031	0.090	0.029	0.326	0.028	0.048	0.024	0.034	0.018	0.028	0.012
0.050	0.027	0.042	0.027	0.031	0.031	0.086	0.028	0.080	0.027	0.326	0.026	0.044	0.025	0.034	0.016	0.026	0.012
0.087	0.037	0.075	0.021	0.034	0.018	0.095	0.021	0.116	0.017	0.404	0.041	0.061	0.022	0.035	0.15	0.034	0.010
0.032	0.024	0.070	0.016	0.042	0.015	0.087	0.020	0.111	0.015	0.408	0.043	0.067	0.021	0.046	0.015	0.030	0.009
0.047	0.017	0.053	0.013	0.028	0.014	0.095	0.020	0.101	0.015	0.402	0.041	0.062	0.021	0.040	0.016	0.030	0.009
0.066	0.017	0.022	0.020	0.015	0.018	0.092	0.022	0.104	0.019	0.406	0.042	0.066	0.021	0.043	0.015	0.037	0.009

Results are based on replications for which $\hat{K}$ is correctly estimated.

TABLE 10.
Summary of RMSE of estimated item parameters for $K = 5$, $\rho = 0$ by sample size N.

Crimp-Sampler						IBP						RJMCMC					
$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$		$N = 500$		$N = 1000$		$N = 2000$	
$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$	$\hat{s}_j$	$\hat{g}_j$
0.035	0.026	0.038	0.019	0.034	0.013	0.059	0.062	0.062	0.089	0.199	0.177	0.028	0.042	0.025	0.033	0.024	0.013
0.043	0.012	0.036	0.025	0.030	0.019	0.086	0.083	0.061	0.091	0.195	0.185	0.036	0.037	0.026	0.022	0.015	0.014
0.027	0.027	0.025	0.033	0.023	0.020	0.105	0.098	0.126	0.119	0.250	0.220	0.034	0.032	0.026	0.030	0.020	0.040
0.045	0.035	0.048	0.038	0.027	0.018	0.112	0.109	0.095	0.138	0.214	0.200	0.035	0.034	0.028	0.030	0.020	0.095
0.042	0.013	0.025	0.013	0.015	0.017	0.047	0.058	0.046	0.088	0.135	0.091	0.033	0.035	0.019	0.020	0.035	0.100
0.023	0.026	0.022	0.046	0.018	0.019	0.058	0.063	0.062	0.088	0.200	0.178	0.033	0.034	0.022	0.027	0.013	0.037
0.026	0.019	0.022	0.051	0.027	0.011	0.083	0.083	0.057	0.087	0.198	0.186	0.037	0.033	0.025	0.028	0.023	0.008
0.046	0.032	0.052	0.027	0.027	0.022	0.107	0.097	0.120	0.119	0.249	0.221	0.042	0.031	0.022	0.024	0.025	0.038
0.044	0.055	0.042	0.055	0.028	0.028	0.117	0.105	0.092	0.142	0.215	0.198	0.038	0.034	0.029	0.036	0.013	0.101
0.059	0.051	0.037	0.020	0.029	0.020	0.042	0.057	0.047	0.088	0.134	0.094	0.033	0.032	0.019	0.025	0.045	0.105
0.021	0.023	0.041	0.037	0.013	0.018	0.129	0.027	0.095	0.029	0.288	0.024	0.045	0.025	0.033	0.016	0.014	0.012
0.094	0.043	0.016	0.014	0.015	0.014	0.103	0.033	0.066	0.029	0.279	0.025	0.050	0.022	0.033	0.016	0.024	0.033
0.044	0.012	0.054	0.023	0.029	0.018	0.066	0.029	0.071	0.035	0.281	0.030	0.047	0.023	0.039	0.016	0.043	0.041
0.012	0.017	0.027	0.030	0.011	0.021	0.125	0.028	0.130	0.033	0.373	0.066	0.051	0.022	0.032	0.019	0.028	0.015
0.045	0.034	0.024	0.031	0.013	0.023	0.107	0.028	0.077	0.031	0.334	0.049	0.055	0.025	0.038	0.018	0.028	0.007
0.050	0.038	0.057	0.049	0.026	0.015	0.184	0.024	0.167	0.020	0.414	0.028	0.067	0.018	0.047	0.016	0.063	0.013
0.035	0.019	0.029	0.019	0.036	0.011	0.162	0.025	0.106	0.016	0.377	0.038	0.062	0.023	0.052	0.015	0.034	0.015
0.048	0.018	0.053	0.033	0.023	0.018	0.157	0.022	0.146	0.020	0.404	0.031	0.066	0.022	0.050	0.013	0.028	0.012
0.088	0.029	0.041	0.023	0.017	0.013	0.140	0.022	0.113	0.018	0.381	0.038	0.070	0.022	0.048	0.014	0.053	0.009
0.038	0.045	0.041	0.027	0.033	0.013	0.165	0.023	0.142	0.019	0.446	0.030	0.078	0.022	0.047	0.017	0.018	0.013

Results are based on replications for which $\hat{K}$ is correctly estimated.

TABLE 11.
Summary of the frequency of estimated $\hat{K}$ by sample size N, number of true attributes K, and attribute dependence ρ using DIC selection.

			$\hat{K}$ selected by lowest DIC						
N	K	ρ	2	3	4	5	6	7	8
500	3	0.00	0	96	4	0	0	0	0
1000	3	0.00	0	92	8	0	0	0	0
2000	3	0.00	0	90	10	0	0	0	0
500	4	0.00	0	1	99	0	0	0	0
1000	4	0.00	0	4	90	6	0	0	0
2000	4	0.00	0	1	76	19	4	0	0
500	5	0.00	0	0	2	90	8	0	0
1000	5	0.00	0	0	17	65	18	0	0
2000	5	0.00	0	0	39	57	4	0	0
500	3	0.25	0	82	18	0	0	0	0
1000	3	0.25	0	85	15	0	0	0	0
2000	3	0.25	0	87	13	0	0	0	0
500	4	0.25	0	11	80	9	0	0	0
1000	4	0.25	0	10	68	22	0	0	0
2000	4	0.25	0	13	65	22	0	0	0
500	5	0.25	0	0	59	41	0	0	0
1000	5	0.25	0	0	67	32	1	0	0
2000	5	0.25	0	1	66	21	12	0	0
500	3	0.5	0	79	19	2	0	0	0
1000	3	0.5	0	74	25	1	0	0	0
2000	3	0.5	0	74	23	3	0	0	0
500	4	0.5	0	15	76	9	0	0	0
1000	4	0.5	0	13	65	21	1	0	0
2000	4	0.5	0	28	37	32	3	0	0
500	5	0.5	0	0	52	37	11	0	0
1000	5	0.5	0	0	69	20	11	0	0
2000	5	0.5	0	1	67	15	13	4	0

The specified range of K for DIC selection is 2 to 7 for $K = 3$ and 4, and 2 to 8 for $K = 5$.

10. A box contains marbles that are red or yellow, small or large. The probability of drawing a red marble is 0.70, the probability of drawing a small marble is 0.40. Suppose that the color of the marbles is independent of their size. What is the probability of randomly drawing a small marble that is not red?
11. In a garage there are 50 cars. 20 are black and 10 are diesel powered. Suppose that the color of the cars is independent of the kind of fuel. What is the probability that a randomly selected car is not black and it is diesel powered?
12. A box contains 20 marbles. 10 marbles are red, 6 are yellow and 4 are black. 12 marbles are small and 8 are large. Suppose that the color of the marbles is independent of their size. What is the probability of randomly drawing a small marble that is yellow or red?

Appendix F

Supplementary Simulation Results

Tables 8, 9 and 10 report comparison on the root-mean-squared error (RMSE) of the estimated item parameters for $\rho = 0$ for the cases where K was correctly estimated. The tables show that both the crimp sampler and RJ-MCMC give a smaller RMSEs on item parameter estimation than IBP across different sample sizes and K, and the crimp sampler is slightly better than RJ-MCMC

for the $K = 5$ case. The results for $\rho = 0.25$ are omitted here because the performance was similar. Table 11 reports the performance of inferring K using DIC selection.

References

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*(4), 434–455.

Chen, Y., Culpepper, S., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika*, *85*, 121–153.

Chen, Y., Culpepper, S., & Liang, F. (2020b). A sparse latent class model for cognitive diagnosis. *Psychometrika*, 1–33.

Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. (2018). Bayesian estimation of the DINA Q. *Psychometrika*, *83*(1), 89–108.

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. A. (2018). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills. *Applied Psychological Measurement*, *42*, 5–23.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866.

Culpepper, S. A. (2015). Bayesian estimation of the DINA model with Gibbs sampling. *Journal of Educational and Behavioral Statistics*, *40*(5), 454–476.

Culpepper, S. A. (2019a). Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. *Psychometrika*, *84*(2), 333–357.

Culpepper, S. A. (2019b). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika*, *84*(4), 921–940.

Culpepper, S. A., & Chen, Y. (2018). Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics*, *44*, 3–24.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, *69*(3), 333–353.

de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, *73*(4), 595–624.

Doshi-Velez, F., & Williamson, S. A. (2017). Restricted Indian buffet processes. *Statistics and Computing*, *27*(5), 1205–1223.

Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19–40.

Gershman, S. J., & Blei, D. M. (2012). A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, *56*(1), 1–12.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732.

Griffiths, T. L., & Ghahramani, Z. (2005). *Infinite latent feature models and the Indian buffet process (Technical Report 2005-001)*. London: Gatsby Computational Neuroscience Unit.

Griffiths, T. L., & Ghahramani, Z. (2011). The Indian buffet process: An introduction and review. *Journal of Machine Learning Research*, *12*(Apr), 1185–1224.

Gu, Y., & Xu, G. (2019). The sufficient and necessary condition for the identifiability and estimability of the dina model. *Psychometrika*, *84*(2), 468–483.

Gu, Y., & Xu, G. (in press). Sufficient and necessary conditions for the identifiability of the Q-matrix. Statistica Sinica. https://doi.org/10.5705/ss.202018.0410

Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics*, *42*, 49–56.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210.

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement*, *77*(3), 369–388.

Li, F., Cohen, A., Bottge, B., & Templin, J. (2016). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, *76*(2), 181–204.

Liu, C.-W., Andersson, B., & Skrondal, A. (2020). A constrained metropolis-hastings robbins-monro algorithm for q matrix estimation in dina models. *Psychometrika*, *85*(2), 322–357.

Liu, J., Xu, G., & Ying, Z. (2013). Theory of the self-learning Q-matrix. *Bernoulli*, *19*(5A), 1790–1817.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, *89*(427), 958–966.

Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*, 963–990.

Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied Psychological Measurement*, *41*(6), 422–438.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, *4*, 639–650.

Sorrel, M. A., Olea, J., Abad, F. J., de la Torre, J., Aguado, D., & Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods*, *19*(3), 506–532.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*(3), 337–350.
Tatsuoka, K. K. (1984). *Analysis of errors in fraction addition and subtraction problems*. Computer-Based Education Research Laboratory: University of Illinois at Urbana-Champaign.
Teh, Y. W., Grür, D., & Ghahramani, Z. (2007). Stick-breaking construction for the Indian buffet process. In *Artificial intelligence and statistics* (pp. 556–563). San Juan, Puerto Rico.
Thibaux, R., & Jordan, M. I. (2007). Hierarchical beta processes and the Indian buffet process. In *Artificial intelligence and statistics* (pp. 564–571). San Juan, Puerto Rico.
von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307.
Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. (2017). Tracking skill acquisition with cognitive diagnosis models: A higher-order hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics*, *43*(1), 57–87.
Wang, S., Zhang, S., Douglas, J., & Culpepper, S. A. (2018). Using response times to assess learning progress: A joint model for responses and response times. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 45–58.
Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *Annals of Statistics*, *45*(2), 675–707.
Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*(523), 1284–1295.
Xu, G., & Zhang, S. (2016). Identifiability of diagnostic classification models. *Psychometrika*, *81*(3), 625–649.
Ye, S., Fellouris, G., Culpepper, S. A., & Douglas, J. (2016). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*(2), 139–158.
Zhang, S., Douglas, J., Wang, S., & Culpepper, S. A. (in press). Reduced reparameterized unified model applied to learning spatial reasoning skills. In M. von Davier & Y. Lee (Eds.), *Handbook of diagnostic classification models*. New York: Springer.

Manuscript Received: 30 SEP 2019
Final Version Received: 15 FEB 2021
Accepted: 19 FEB 2021
Published Online Date: 22 MAR 2021