应用概率统计 第 32 卷
第 2 期 2016 年 4 月

Chinese Journal of Applied Probability and Statistics
Apr., 2016, Vol. 32, No. 2, pp. 157-183
doi: 10.3969/j.issn.1001-4268.2016.02.005

Bayesian Semiparametric Analysis and Model Comparison for Confirmatory Factor Model *

XIA Yemao LIU Yingan

(*Department of Applied Mathematics, Nanjing Forestry University, Nanjing, 210037, China*)

Abstract: A Bayesian semiparametric procedure for confirmatory factor analysis model is proposed to address the heterogeneity of the multivariate responses. The approach relies on the use of a prior over the space of mixing distributions with finite components. Blocked Gibbs sampler is implemented to cope with the posterior analysis. For model comparison, the L_ν measure and Bayes factor are developed. A generalized weighted Chinese restaurant algorithm is suggested to compute the likelihood of data. Empirical results are presented to illustrate the effectiveness of the methodologies.

Keywords: confirmatory factor model; finite dimensional Dirichlet prcocess prior; blocked Gibbs sampler; model comparison

2010 Mathematics Subject Classification: 62H25; 62G07

§1. Introduction

Confirmatory factor analysis model (CFA), a widely appreciated multivariate statistical method to explore latent variables (factors) that are related to the manifest variables and explain correlations among the observed data, has been received substantial concerns in behavioral, educational, medical and social-psychological sciences. Commonly used softwares such as LISREL VIII [1], EQS 6 [2], and LISCOMP [3] have been developed to popularize CFA.

Often statistical results based on a single CFA are seriously distorted when data take on heterogeneity [4]. Heterogenous data arise when the underlying population consists of a number of distinct components. In such setting, one popular choice is to resort to finite mixture model [5–7] to identify the underlying structure of multivariate data and establish multivariate mixture model. There is rich literature on this important issue. Within the frequency statistical framework, Blåfield [8] developed a quasi-Newton routine to analyze unconditional multivariate normal mixture model incorporating first and second CFA excluded means; Yung [9] presented a multivariate normal mixture model incorporating

*The project was supported by the National Natural Science Foundation of China (11471161), the Nanjing Forestry University Grant (163101004), and the Nanjing Science and Technology Innovation Item (013101001).
Received April 15, 2014. Revised July 12, 2014.

the CFA with structured means and developed complex sampling schemes to solve the estimation problem; Arminger and Stein [10] adopted two-stage procedure to analyze the mixture model that includes CFAs and structured regression models. Relying on the EM algorithm [11], Jedidi et al. [12] discussed the maximum likelihood analysis for an unconditional normal mixture model that includes CFAs and a full structure equation model (the full LISREL model); McLachlan et al. [13] applied AECM algorithm [14] to analyze the mixture of orthogonal factor analysis model and establish density estimation procedure. In the framework of Bayesian analysis, Hoshino [15] considered the multivariate normal mixtures of which means and covariance matrices are structured as CFA; Zhu and Lee [16] proposed a Bayesian analysis of finite mixture in the linear structure equation model using the idea of augmenting the observed data with latent variables and allocation variables.

The aforementioned approaches are along with the lines of a parametric method, which need to handle the very large number of parameters. More flexible and feasible approach is to work with Dirichlet process mixture modeling [17, 18] which is referred to as Bayesian semiparametric modeling (see [19–22] and references therein). An attractive summary and review of this literature was given by Walker et al. [23] and recently synthesized by Müller and Quintana [24]. An undesirable side effect in Dirichlet process mixture modeling is to touch on the infinite dimensional distribution. This leads to the direct posterior inferences of functionals of the population distribution intractable. Recently, Ishwaran and Zarepour [25] proposed a finite dimensional Dirichlet prior to approximate Dirichlet process and developed a semiparametric Bayesian procedure for analyzing finite mixture model. The key point in their development is to break the mixture model into the hierarchical model and place the finite dimensional Dirichlet prior over the space of mixing distribution with at most N components. This method facilitates the direct sampling scheme to explore the nonparametric posterior analysis, consequently, resulting in the computational and inferential advantages in the finite mixture analysis.

More recently, Bayesian semiparametric analysis for the latent variable models has been becoming increasingly popular (see for example, [26–29] and among others). The key to these methods is to relax the parametric model of the factor by the nonparametric alternative and develop robust procedures to defend against the distribution deviations. Although working successfully in these contexts, the developed approach may not be more effective in solving the heterogeneity of data. It does not allow for the factor loadings and variances of the unique errors to vary across the subjects.

In this paper, we focus on the Bayesian semiparametric analysis for CFA. Comparing to the methods mentioned above, the methodology developed here lies in two different directions. First, our modeling is to introduce the finite dimensional Dirichlet prior as a prior of the distributions of the mean and variance-covariance parameters. Utilizing the

discreteness of realizations of the finite dimensional DP prior, we select the underlying distributions of mean and variance-covariance parameters from the space of distributions with at most finite atoms. This approach is therefore in line with the nonparametric Bayes analysis. Second, although the resulting model for the observed data is the multivariate normal mixture model, our estimation procedure is different from these existing methods for analyzing the mixture CFA model. By introducing latent variables, the proposed method avoids estimating unknown parameters involved in the mixture components, hence allowing for significant simplification of computational complexity.

Another contribution in this paper is to develop a procedure for model selection. For the semiparametric CFA analysis, it is well known that the computation of Bayes factor [30] is rather difficult since it involves the integration of unknown random distributions over the space of mixing distributions. Basu and Chib [31] devised an effective method in computing the marginal likelihood of the semiparametric model based on the framework of [32]. Though, this method can be effective in computing marginal likelihood of the Dirichlet process mixture model, it is not clear how to work under the finite dimensional Dirichlet prior measure, that is, when no pólya urn scheme is available. Our approach here concerns, in essence, an important extension from those in [33] for normal mean mixture model in one dimension to the CFA in multiple dimensions. Based on the Chinese restaurant algorithm, the observed likelihood is computed. Consequently, the computation of marginal likelihood of the observed data is addressed via Chib's method. To compare with this, we introduced the multivariate version of L_ν measure [34–36] as a reference for model selection. To explore the posterior inferences, we adopt the blocked Gibbs sampler to implement the Markov chains Monte Carlo (MCMC) sampling. Posterior predictive distributions of observations are obtained via simulated observations from the posterior distribution.

§2. Model and Method

2.1 Review

In this section, we briefly review the finite dimensional Dirichlet prior. Let $\mathbb{R}^d$ be a d-dimesional Euclidian space equipped with a Borel field $\mathscr{B}^d$, and let $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_N$ be the independent and identically distributed random vectors with the common distribution G_0 on $(\mathbb{R}^d, \mathscr{B}^d)$. A random probability measure P on $(\mathbb{R}^d, \mathscr{B}^d)$ is said to have a finite dimensional Dirichlet prior with level N and parameters c and G_0 (denoted by $\mathscr{D}_N(c\,G_0)$), if each realization of P is

$$\mathsf{P}(\cdot) = \sum_{k=1}^{N} \pi_k \delta_{\boldsymbol{V}_k}(\cdot), \tag{1}$$

where $\delta_{\boldsymbol{V}_k}(\cdot)$ is a point mass measure concentrated at $\boldsymbol{V}_k$, and π_k is the random weight satisfying $\pi_k > 0$ and $\sum\limits_{k=1}^{N} \pi_k = 1$, a.s. Moreover, $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$ is independent of $\{\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_N\}$ and has the symmetrical Dirichlet distribution Dirichlet$(c/N, \ldots, c/N)$ for some $c > 0$. If a probability distribution measure P is distributed with $\mathscr{D}_N(c\,G_0)$, we will denote $\mathsf{P} \sim \mathscr{D}_N(c\,G_0)$ or $\mathsf{P} \sim \sum\limits_{k=1}^{N} \pi_k \delta_{\boldsymbol{V}_k}(\cdot)$ in the different contexts.

Finite dimensional Dirichlet prior was formally proposed by Ishwaran and Zarepour [25] to approximate the well-known Dirichlet process (i.e. $N = \infty$) in the nonparametric Bayesian analysis. As a prior for the probability distribution measure, Ishwaran and James [37] exploited some theoretical properties of the posterior and applications of P in the univariate normal mixture models. The following theorem reveals the properties of functionals related to the distribution measure P, which can be obtained by the similar routine in [17].

Theorem 1 Let $\mathsf{P} \sim \mathscr{D}_N(c\,G_0)$; $\psi(x)$, $\psi_1(x)$ and $\psi_2(x)$ are measurable real valued functions defined on $(\mathbb{R}^d, \mathscr{B}^d)$.

(i) If $\int_{\mathbb{R}^d} |\psi(x)| \mathrm{d}G_0(x) < \infty$, then $\int_{\mathbb{R}^d} \psi(x) \mathsf{P}(\mathrm{d}x) < \infty$, a.s., and

$$\int \int_{\mathbb{R}^d} \psi(x) \mathsf{P}(\mathrm{d}x) \mathscr{D}_N(\mathrm{d}\mathsf{P} | c\,G_0) = \mathsf{E}_{G_0} \psi(X),$$

where X is a $d \times 1$ dimensional random vector with distribution G_0.

(ii) If $\int_{\mathbb{R}^d} |\psi_i(x)| \mathrm{d}G_0(x) < \infty$, and $\int_{\mathbb{R}^d} |\psi_1(x)\psi_2(x)| \mathrm{d}G_0(x) < \infty$ then

$$\mathsf{Cov}\Big(\int_{\mathbb{R}^d} \psi_1(x) \mathsf{P}(\mathrm{d}x), \int_{\mathbb{R}^d} \psi_2(x) \mathsf{P}(\mathrm{d}x) \Big) = \frac{\sigma_{12}}{c+1} \Big(1 + \frac{c}{N}\Big),$$

where $\sigma_{12} = \mathsf{Cov}_{G_0}(\psi_1(X), \psi_2(X))$.

The first two moments of P can be obtained by setting $\psi_i(x) = \psi(x) = I_A(x)$ in Theorem 1. In particular, for any $A \in \mathscr{B}^d$, $\mathsf{E}(\mathsf{P}(A)) = G_0(A)$ and

$$\mathsf{Var}\,(\mathsf{P}(A)) = \frac{G_0(A)(1 - G_0(A))}{c+1} \Big(1 + \frac{c}{N}\Big), \tag{2}$$

which shows that P is centered at the prior guess G_0, and c and N determine the concentration of the prior around G_0. Clearly, when N tends to infinity, (2) is consistent with that in Dirichlet process (see [17]).

Let $X_1, X_2, \ldots, X_n$ be a random sample from P. Since $\mathscr{D}_N(c\,G_0)$ put probability one on the discrete probability measure, the probability distribution selected by it is naturally discrete with at most N atoms. Therefore, the joint distribution of $X_1, X_2, \ldots, X_n$

is degenerate with respect to the L-measure on $(\mathbb{R}^{nd}, \mathscr{B}^{nd})$. This problem can be addressed by decomposing sample space into some exclusive nonempty subspaces. Restricted on the hyperplane of sample space, the joint density of $X_1, X_2, \ldots, X_n$ can be obtained easily by multiplying the density of distinct random variables. To be specific, let $\boldsymbol{P} = \{C_1, C_2, \ldots, C_{n(\boldsymbol{P})}\}$ be a partition of $\{1, 2, \ldots, n\}$, where $n(\boldsymbol{P})$ is the number of cells in the partition and C_i is the i-th cell of the partition $\boldsymbol{P}$ with size $e_i = \#C_i$. Define

$$\boldsymbol{S}_{C_1,C_2,\ldots,C_{n(\boldsymbol{P})}} = \{(x_1, x_2, \ldots, x_n) : x_i = x_j, \text{ if } i, j \in C_m \text{ for some } m = 1, 2, \ldots, n(\boldsymbol{P})\}.$$

Obviously, $\mathbb{R}^{nd}$ can be expressed as the union of $\boldsymbol{S}_{C_1,C_2,\ldots,C_{n(\boldsymbol{P})}}$ for all possible partitions $\boldsymbol{P}$. The joint density function of $X_1, X_2, \ldots, X_n$ is given by the following theorem which can be considered as an extension of Theorem 1 in [19].

Theorem 2 Let $\mathsf{P} \sim \mathscr{D}_N(c\,G_0)$; $(X_1, X_2, \ldots, X_n)$ is a sample of size n from P. If G_0 is nonatomic with density function $g_0(x)$, then, the joint density function of $X_1, X_2, \ldots, X_n$ restricted on $\boldsymbol{S}_{C_1,C_2,\ldots,C_{n(\boldsymbol{P})}}$ is

$$p(x_1, x_2, \ldots, x_n) = \frac{c^{n(\boldsymbol{P})}}{c^{[n]}} \frac{N!}{N^{n(\boldsymbol{P})}(N - n(\boldsymbol{P}))!} \prod_{j=1}^{n(\boldsymbol{P})} \left(1 + \frac{c}{N}\right)^{[e_j - 1]} \prod_{j=1}^{n(\boldsymbol{P})} g_0(x_j^*), \tag{3}$$

where the x_j^*'s are the distinct values of x_i's, and $c^{[k]} = c(c+1)\cdots(c+k-1)$ for any $k > 0$ and 1 for $k = 0$.

Theorem 2 provides a critical result in computing likelihood of the observed data. See Section 4.1 for more details.

2.2 Bayesian Semiparametric Framework

We now turn to the semiparametric CFA model. Recall that a finite mixture model with N components for a $p \times 1$ random observed vector $\boldsymbol{y}$ is typically defined as

$$f(\boldsymbol{y}|\boldsymbol{\zeta}) = \sum_{m=1}^{N} \pi_m p_m(\boldsymbol{y}|\boldsymbol{\mu}_m, \boldsymbol{\eta}_m), \tag{4}$$

where π_m is the mixing proportion such that $\pi_m > 0$ and $\sum_{m=1}^{N} \pi_m = 1.0$; $p_m(\cdot|\boldsymbol{\mu}_m, \boldsymbol{\eta}_m)$ is the multivariate normal density function with an unknown mean vector $\boldsymbol{\mu}_m$ and a covariance structure $\boldsymbol{\Sigma}_m = \boldsymbol{\Sigma}_m(\boldsymbol{\eta}_m)$ that depends on an unknown parameter vector $\boldsymbol{\eta}_m$; $\boldsymbol{\zeta}$ is the parameter vector that contains all unknown parameters in π_m, $\boldsymbol{\mu}_m$ and $\boldsymbol{\eta}_m$. A finite CFA mixture model is defined by specifying the m-th mixture component in (4) through the well-known CFA:

$$\boldsymbol{y} = \boldsymbol{\mu}_m + \boldsymbol{\Lambda}_m \boldsymbol{\omega}_m + \boldsymbol{\epsilon}_m, \tag{5}$$

where $\boldsymbol{\mu}_m$ is a $p \times 1$ intercept vector, $\boldsymbol{\Lambda}_m$ is a $p \times q$ factor loading matrix, $\boldsymbol{\omega}_m$ is a $q \times 1$ random vector of latent variables and $\boldsymbol{\epsilon}_m$ is the vector of measurement errors. Following the standard convention for CFA, we assume that $\boldsymbol{\omega}_m$ is distributed according to $N_q(\mathbf{0}, \boldsymbol{\Phi}_m)$ with positive finite matrix $\boldsymbol{\Phi}_m$, and $\boldsymbol{\epsilon}_m$, independent of $\boldsymbol{\omega}_m$, is distributed according to $N_p(\mathbf{0}, \boldsymbol{\Psi}_m)$ with a diagonal matrix $\boldsymbol{\Psi}_m$. Based on these settings, the covariance structure of the m-th component is $\boldsymbol{\Sigma}_m(\boldsymbol{\eta}_m) = \boldsymbol{\Lambda}_m \boldsymbol{\Phi}_m \boldsymbol{\Lambda}_m^{\mathsf{T}} + \boldsymbol{\Psi}_m$, where $\boldsymbol{\eta}_m$ is the unknown parameter vector formed by the free unknown parameters in $\boldsymbol{\Lambda}_m$, $\boldsymbol{\Psi}_m$ and $\boldsymbol{\Phi}_m$.

Let $\boldsymbol{\theta}_m = \{\boldsymbol{\mu}_m, \boldsymbol{\Lambda}_m, \boldsymbol{\Psi}_m, \boldsymbol{\Phi}_m\}$ be the vector formed by the location and variance-covariance parameters in $p(\cdot|\boldsymbol{\mu}_m, \boldsymbol{\eta}_m)$. To motivate our modeling, we rewrite the mixture model (4) as

$$f(\boldsymbol{y}|\boldsymbol{\zeta}) = \int p(\boldsymbol{y}|\boldsymbol{\theta}) \mathsf{Q}_N(\mathrm{d}\boldsymbol{\theta}),$$

where $\mathsf{Q}_N(\mathrm{d}\boldsymbol{\theta}) = \sum_{m=1}^{N} \pi_m \delta_{\boldsymbol{\theta}_m}(\mathrm{d}\boldsymbol{\theta})$ is the finite mixing distribution. The approach commonly used in finite mixture analysis is to treat N as fixed and known. Bayesian analysis is taken within the following parameter space

$$\boldsymbol{\Xi}^N \times \Big\{(\pi_1, \pi_2, \ldots, \pi_N) : \pi_m > 0, \sum_{m=1}^{N} \pi_m = 1.0\Big\} \tag{6}$$

corresponding to the atoms $\boldsymbol{\theta}_m$'s and weights π_m for the mixing distribution. Obviously, when N is larger one need to handle a large number of parameters. Rather than treating N to be known, we assume the underlying mixing distribution is random and assign a finite dimensional Dirichlet prior (1) on it. Specifically, we assume that

$$\begin{cases} p(\boldsymbol{y}|\mathsf{P}) = \displaystyle\int p(\boldsymbol{y}|\boldsymbol{\theta}) \mathsf{P}(\mathrm{d}\boldsymbol{\theta}), \\ \mathsf{P} \sim \mathscr{D}_N(c\, G_0) \end{cases} \tag{7}$$

for some baseline distribution G_0. As a result, the distribution of $\boldsymbol{y}$ is the normal mixture model with the mixing distribution P selected from the parameter space given by

$$\mathscr{Q}_N = \bigcup_{k=1}^{N} \mathscr{Q}(k),$$

where $\mathscr{Q}(k)$ is the space of finite mixtures over $\boldsymbol{\Xi}^k$ with exactly k atoms. As pointed out by Ishwaran and Zarepour [25], there are both conceptual and computational advantages to work with $\mathscr{Q}_N$ as parameter space. It puts us more on par with non-Bayesian methods from the theoretical perspective and provides computational convenience for drawing values directly from the posterior of the mixing distribution.

Let $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ be the random observations of size n from model (7) associated with $p(\boldsymbol{y}|\boldsymbol{\theta})$ given by (5). By introducing latent variables $\boldsymbol{\theta}_i = \{\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i, \boldsymbol{\Psi}_i, \boldsymbol{\Phi}_i\}$, we can rewrite (7) and (5) as the following hierarchical model

$$\begin{cases} \boldsymbol{y}_i|\boldsymbol{\omega}_i, \boldsymbol{\theta}_i \overset{\text{ind}}{\sim} N_p(\boldsymbol{\mu}_i + \boldsymbol{\Lambda}_i\boldsymbol{\omega}_i, \boldsymbol{\Psi}_i), \\ \boldsymbol{\omega}_i|\boldsymbol{\Phi}_i \overset{\text{ind}}{\sim} N_q(\boldsymbol{0}, \boldsymbol{\Phi}_i), \\ \boldsymbol{\theta}_i|\mathsf{P} \overset{\text{i.i.d.}}{\sim} \mathsf{P}, \\ \mathsf{P} \sim \sum\limits_{m=1}^{N} \pi_m \delta_{\boldsymbol{V}_m}, \end{cases} \tag{8}$$

where $\boldsymbol{\omega}_i$ is the factor vector and $\boldsymbol{V}_m = \{\boldsymbol{V}_m^{\mu}, \boldsymbol{V}_m^{\lambda}, \boldsymbol{V}_m^{\psi}, \boldsymbol{V}_m^{\phi}\}$ is the atom, in which $\boldsymbol{V}_m^{\mu}$ is the $p \times 1$ random vector, $\boldsymbol{V}_m^{\lambda}$ is the $p \times q$ matrix, $\boldsymbol{V}_m^{\psi}$ is the $p \times p$ diagonal matrix with the j-th diagonal element V_{mj}^{ψ}, and $\boldsymbol{V}_m^{\phi}$ is the $q \times q$ positive finite matrix.

For the baseline distribution G_0, we follow the common choice for the CFA model ([27]) and assume

$$\mathrm{d}G_0(\boldsymbol{v}_m^{\mu}, \boldsymbol{v}_m^{\lambda}, \boldsymbol{v}_m^{\psi}, \boldsymbol{v}_m^{\phi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}) = p(\boldsymbol{v}_m^{\mu}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\boldsymbol{v}_m^{\lambda}, \boldsymbol{v}_m^{\psi})\, p(\boldsymbol{v}_m^{\phi}|\boldsymbol{\Phi})\mathrm{d}\boldsymbol{v}_m^{\mu}\mathrm{d}\boldsymbol{v}_m^{\lambda}\mathrm{d}\boldsymbol{v}_m^{\psi}\mathrm{d}\boldsymbol{v}_m^{\phi},$$

in which

$$\begin{aligned} p(\boldsymbol{v}_m^{\mu}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) &\overset{\mathrm{d}}{=} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad p(\boldsymbol{v}_m^{\phi}|\boldsymbol{\Phi}) \overset{\mathrm{d}}{=} \text{Wishart}_q^{-1}(\rho_{0\phi}, \boldsymbol{\Phi}^{-1}), \\ p(\boldsymbol{v}_m^{\lambda}, \boldsymbol{v}_m^{\psi}) = p(\boldsymbol{v}_m^{\psi})\, p(\boldsymbol{v}_m^{\lambda}|\boldsymbol{v}_m^{\psi}) &\overset{\mathrm{d}}{=} \prod_{j=1}^{p} \text{Gamma}^{-1}(\alpha_{j0}, \beta_{j0}) \cdot N_q(\boldsymbol{\Lambda}_{j0}, v_{mj}^{\psi}\boldsymbol{H}_{j0}). \end{aligned} \tag{9}$$

Here, $\text{Gamma}^{-1}(\alpha_{j0}, \beta_{j0})$ is the inverse gamma distribution with shape $\alpha_{j0} > 0$ and scale $\beta_{j0} > 0$, $\boldsymbol{\Lambda}_{j0}$ is a $q \times 1$ row vector and $\boldsymbol{H}_{j0}$ is the $q \times q$ positive definite matrix; $\text{Wishart}_q^{-1}(\rho, \boldsymbol{\Phi}^{-1})$ is the $q \times q$ inverse wishart distribution with ρ degrees of freedom and scale matrix $\boldsymbol{\Phi}^{-1}$.

The following conjugate type prior distributions are used for hyper-parameters $\{\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Phi}\}$:

$$\boldsymbol{\mu} \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \qquad \boldsymbol{\Sigma} \sim \text{Wishart}_p^{-1}(\rho_{0\Sigma}, \boldsymbol{R}_{0\Sigma}^{-1}), \qquad \boldsymbol{\Phi} \sim \text{Wishart}_q^{-1}(s, \boldsymbol{R}_{0\phi}^{-1}). \tag{10}$$

The hyperparameters α_{0j}, β_{0j}, $\rho_{\phi 0}$, $\rho_{0\Sigma}$, s, the vector $\boldsymbol{\mu}_0$ and $\boldsymbol{\Lambda}_{0j}$, and the matrices $\boldsymbol{\Sigma}_0$, $\boldsymbol{H}_{0k}$, $\boldsymbol{R}_{0\Sigma}$ and $\boldsymbol{R}_{0\phi}$ involved in (9) and (10) are assumed to be known. As pointed out by Zhu and Lee [16] that the conjugate type prior distributions are sufficiently flexible in most applications, and for situations with a reasonable amounts of data available, the hyper-parameters scarcely affect the analysis.

§3. Posterior Analysis

3.1 Sampling Scheme and Full Conditional Distributions

Let $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_n\}$ be the collection of the observed variables, $\boldsymbol{\Omega}=\{\boldsymbol{\omega}_1,\boldsymbol{\omega}_2,\ldots,\boldsymbol{\omega}_n\}$ be the collection of latent factor variables, and $\boldsymbol{V}=\{\boldsymbol{V}_1,\boldsymbol{V}_2,\ldots,\boldsymbol{V}_N\}$ be the collection of atoms; $\boldsymbol{\vartheta}=\{\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{R}\}$. In a standard Bayesian analysis, we require evaluating the complicated posterior distribution. The trick for achieving efficient Markov chain Monte Carlo sampling of the semiparametric hierarchical model (8) is to introduce the classification variable L_i which takes values in $\{1,2,\ldots,N\}$ and recast the model completely as follows:

$$\begin{cases}\boldsymbol{y}_i|\boldsymbol{\omega}_i,\boldsymbol{V},\boldsymbol{L}\overset{\text{ind}}{\sim}N_p(\boldsymbol{\mu}_i+\boldsymbol{\Lambda}_i\boldsymbol{\omega}_i,\boldsymbol{\Psi}_i),\\ \boldsymbol{\omega}_i|\boldsymbol{V},\boldsymbol{L}\overset{\text{ind}}{\sim}N_q(\boldsymbol{0},\boldsymbol{\Phi}_i),\\ L_i=\cdot|\boldsymbol{\pi}\sim\sum\limits_{k=1}^{N}\pi_k\delta_k(\cdot),\\ \boldsymbol{\pi}\sim D(c/N,\ldots,c/N),\end{cases}\tag{11}$$

in which $\boldsymbol{L}=\{L_1,L_2,\ldots,L_n\}$, $\boldsymbol{\mu}_i=\boldsymbol{V}_{L_i}^{\mu}$, $\boldsymbol{\Psi}_i=\boldsymbol{V}_{L_i}^{\psi}$, $\boldsymbol{\Lambda}_i=\boldsymbol{V}_{L_i}^{\lambda}$ and $\boldsymbol{\Phi}_i=\boldsymbol{V}_{L_i}^{\phi}$. By rewriting the model as (11), we can devise a blocked Gibbs sampler for exploring the posterior $p(\boldsymbol{\Omega},\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{\vartheta}|\boldsymbol{Y})$. The blocked Gibbs sampler is implemented iteratively by drawing (i) $\boldsymbol{\Omega}$ from $p(\boldsymbol{\Omega}|\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{\vartheta},\boldsymbol{Y})$, (ii) $\boldsymbol{\vartheta}$ from $p(\boldsymbol{\vartheta}|\boldsymbol{\Omega},\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{Y})$, (iii) $(\boldsymbol{\pi},\boldsymbol{V})$ from $p(\boldsymbol{\pi},\boldsymbol{V}|\boldsymbol{\Omega},\boldsymbol{L},\boldsymbol{\vartheta},\boldsymbol{Y})$, and (iv) $\boldsymbol{L}$ from $p(\boldsymbol{L}|\boldsymbol{\Omega},\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{\vartheta},\boldsymbol{Y})$.

Firstly, noting that

$$p(\boldsymbol{\Omega}|\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{\vartheta},\boldsymbol{Y})\propto\prod_{i=1}^{n}p(\boldsymbol{y}_i|\boldsymbol{\omega}_i,\boldsymbol{\theta}_i)\,p(\boldsymbol{\omega}_i|\boldsymbol{\Phi}_i),$$

it can be shown that

$$p(\boldsymbol{\Omega}|\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{\vartheta},\boldsymbol{Y})=\prod_{i=1}^{n}p(\boldsymbol{\omega}_i|\boldsymbol{\theta}_i,\boldsymbol{y}_i)\overset{\text{D}}{=}\prod_{i=1}^{n}N_q(\boldsymbol{\mu}_{\omega i},\boldsymbol{\Xi}_i)$$

with $\boldsymbol{\mu}_{\omega i}=\boldsymbol{\Xi}_i\boldsymbol{\Lambda}_i^{\intercal}\boldsymbol{\Psi}_i^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}_i)$, and $\boldsymbol{\Xi}_i=(\boldsymbol{\Lambda}_i^{\intercal}\boldsymbol{\Psi}_i^{-1}\boldsymbol{\Lambda}_i+\boldsymbol{\Phi}_i^{-1})^{-1}$.

Secondly, it is obvious that given $\boldsymbol{V}$, the conditional distribution $p(\boldsymbol{\vartheta}|\boldsymbol{\pi},\boldsymbol{V},\boldsymbol{L},\boldsymbol{\Omega},\boldsymbol{Y})$ is independent of $\boldsymbol{\pi}$, $\boldsymbol{L}$, $\boldsymbol{\Omega}$ and $\boldsymbol{Y}$. Under the informative prior (10), it can be shown that the posterior distribution of $\boldsymbol{\vartheta}$ is given as follows

$$\begin{cases}\boldsymbol{\mu}|\boldsymbol{\Sigma},\boldsymbol{V}\sim N_p(\boldsymbol{m}_{\mu},\boldsymbol{A}_{\mu}),\\ \boldsymbol{\Sigma}|\boldsymbol{\mu},\boldsymbol{V}\sim\text{Wishart}_p^{-1}(N+\rho_{0\Sigma},\boldsymbol{R}_{\Sigma}^{-1}),\\ \boldsymbol{\Phi}|\boldsymbol{V}\sim\text{Wishart}_q^{-1}(N\rho_{0\phi}+s,\boldsymbol{R}_{\phi}^{-1}),\end{cases}$$

in which

$$\boldsymbol{m}_\mu = \boldsymbol{A}_\mu\Big(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 + \boldsymbol{\Sigma}^{-1}\sum_{k=1}^{N}\boldsymbol{V}_k^\mu\Big), \qquad \boldsymbol{A}_\mu = (\boldsymbol{\Sigma}_0^{-1} + N\boldsymbol{\Sigma}^{-1})^{-1},$$

$$\boldsymbol{R}_\Sigma = \Big(\sum_{k=1}^{N}(\boldsymbol{V}_k^\mu - \boldsymbol{\mu})(\boldsymbol{V}_k^\mu - \boldsymbol{\mu})^\tau + \boldsymbol{R}_{0\Sigma}^{-1}\Big)^{-1}, \qquad \boldsymbol{R}_\phi = \Big(\sum_{k=1}^{N}\boldsymbol{V}_k^{\phi-1} + \boldsymbol{R}_{0\phi}^{-1}\Big)^{-1}.$$

Thirdly, since $p(\boldsymbol{\pi}, \boldsymbol{V}|\boldsymbol{\vartheta}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y}) = p(\boldsymbol{\pi}|\boldsymbol{\vartheta}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y})\, p(\boldsymbol{V}|\boldsymbol{\vartheta}, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y})$, drawing $(\boldsymbol{\pi}, \boldsymbol{V})$ is accomplished by first drawing $\boldsymbol{\pi}$ from $p(\boldsymbol{\pi}|\boldsymbol{\vartheta}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y})$ and then drawing $\boldsymbol{V}$ from $p(\boldsymbol{V}|\boldsymbol{\vartheta}, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y})$. Let n_k denote the total number of i such that $L_i = k$ $(k = 1, 2, \ldots, N)$, then, the full conditional distribution of $\boldsymbol{\pi}$ is given by

$$p(\boldsymbol{\pi}|\boldsymbol{\vartheta}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{Y}) \sim \text{Dirichlet}(c/N + n_1, c/N + n_2, \ldots, c/N + n_N),$$

which only depends on $\boldsymbol{L}$. Drawing $\boldsymbol{\pi}$ from $p(\boldsymbol{\pi}|\boldsymbol{L})$ can be accomplished by drawing N independent gamma$(c/N + n_k, 1)$ random variables divided by their sum, see [17].

Let $\{L_1^*, L_2^*, \ldots, L_m^*\}$ be the unique set of L_i values, $\boldsymbol{V}_L = \{\boldsymbol{V}_{L_1^*}, \boldsymbol{V}_{L_2^*}, \ldots, \boldsymbol{V}_{L_m^*}\}$ and $\boldsymbol{V}_{(-L)}$ corresponds those values in $\boldsymbol{V} = \{\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_N\}$ with $\boldsymbol{V}_L$ removed. Then,

$$p(\boldsymbol{V}|\boldsymbol{\vartheta}, \boldsymbol{L}, \boldsymbol{Y}) = p(\boldsymbol{V}_{(-L)}|\boldsymbol{\vartheta})\, p(\boldsymbol{V}_L|\boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}).$$

The components of $(\boldsymbol{V}_{(-L)}|\boldsymbol{\vartheta})$ are i.i.d. with distribution given in (9) and easy to sample. Clearly,

$$p(\boldsymbol{V}_L|\boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}) = \prod_{j=1}^{m} p(\boldsymbol{V}_{L_j^*}|\boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}).$$

The emphasis is placed on the full conditional distribution $p(\boldsymbol{V}_L^\mu, \boldsymbol{V}_L^\lambda, \boldsymbol{V}_L^\psi, \boldsymbol{V}_L^\phi|\boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y})$. It can be shown that

$$\begin{aligned}
&[\boldsymbol{V}_{L_j^*}^\mu|\boldsymbol{V}_{L_j^*}^\lambda, \boldsymbol{V}_{L_j^*}^\psi, \boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}] \overset{\text{ind}}{\sim} N_p(\boldsymbol{m}_j^*, \boldsymbol{A}_j^*),\\
&[V_{L_j^*k}^\psi|\boldsymbol{V}_{L_j^*}^\mu, \boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}] \overset{\text{ind}}{\sim} \text{Gamma}^{-1}(n_j^*/2 + \alpha_{0k}, \beta_{jk}^*),\\
&[\boldsymbol{V}_{L_j^*k}^\lambda|\boldsymbol{V}_{L_j^*k}^\psi, \boldsymbol{V}_{L_j^*}^\mu, \boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}] \overset{\text{ind}}{\sim} N(\boldsymbol{\Lambda}_{jk}^*, V_{L_j^*k}^\psi \boldsymbol{H}_{jk}^*),\\
&[\boldsymbol{V}_{L_j^*}^\phi|\boldsymbol{L}, \boldsymbol{\vartheta}, \boldsymbol{Y}] \overset{\text{ind}}{\sim} \text{Wishart}_q^{-1}\Big(n_j^* + \rho_{\phi 0}, \sum_{\{i:L_i = L_j^*\}}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^\tau + \boldsymbol{\Phi}^{-1}\Big),
\end{aligned}$$

where n_j^* is the number of L_i equal to L_j^*, and

$$\boldsymbol{m}_j^* = \boldsymbol{A}_j^*\Big\{\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{V}_{L_j^*}^{\psi-1}\sum_{\{i:L_i = L_j^*\}}(\boldsymbol{y}_i - \boldsymbol{V}_{L_j^*}^\lambda\boldsymbol{\omega}_i)\Big\}, \qquad \boldsymbol{A}_j^* = (\boldsymbol{\Sigma}^{-1} + m_j^*\boldsymbol{V}_{L_j^*}^{\psi-1})^{-1},$$

$$\beta_{jk}^* = \beta_{0k} + \frac{1}{2}\Big\{\boldsymbol{\Lambda}_{0k}\boldsymbol{H}_{0k}^{-1}\boldsymbol{\Lambda}_{0k}^\tau - \boldsymbol{\Lambda}_{jk}^*\boldsymbol{H}_{0k}^*\boldsymbol{\Lambda}_{jk}^{*\tau} + \sum_{\{i:L_i = L_j^*\}}(y_{ik} - V_{L_j^*k}^\mu)^2\Big\},$$

$$\boldsymbol{\Lambda}^*_{jk} = \boldsymbol{H}^*_k\Big\{\boldsymbol{H}_{0k}\boldsymbol{\Lambda}_{0k} + \sum_{\{i:L_i=L^*_j\}}(y_{ik} - V^{\mu}_{L^*_j,k})\boldsymbol{\omega}_i\Big\}, \qquad \boldsymbol{H}^*_{jk} = \Big(\boldsymbol{H}^{-1}_{0k} + \sum_{\{i:L_i=L^*_j\}}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^{\intercal}\Big)^{-1}.$$

Lastly, the full conditional distribution of $\boldsymbol{L}$ is given by

$$L_i|\boldsymbol{V},\boldsymbol{\pi},\boldsymbol{Y} \overset{\text{ind}}{\sim} \sum_{k=1}^{N}\pi^*_{ik}\delta_k(\cdot),$$

in which π_{ik} is the random weights proportional to $\pi_k p(\boldsymbol{y}_i|\boldsymbol{\vartheta},\boldsymbol{V}_k)\,p(\boldsymbol{\omega}_i|\boldsymbol{V}^{\phi}_k)$ such that $\sum_{k=1}^{N}\pi_{ik} = 1.0$ $(i = 1,2,\ldots,n)$.

3.2 Posterior Predictive Density Estimation

Many authors including Lo [19], Escobar and West [21], Müller et al. [38], Roeder and Wasserman [39], and among others, used normal mixture models to take Bayesian density estimate. In the context of univariate normal mean mixture model, Ishwaran and Zarepour [25] approximate the unknown predictive density of the observations based on the truncated Dirichlet process. The work developed by Ishwaran and Zarepour can be extend to the semiparametric CFA. More specifically, let $p(\boldsymbol{y}_{n+1}|\boldsymbol{Y})$ represent the predictive density for $\boldsymbol{y}_{n+1}$ conditional on the data $\boldsymbol{Y}$, and let $\boldsymbol{\theta}_{n+1}$ be the corresponding unobserved $\boldsymbol{\theta}$ value; $\boldsymbol{\Theta}_n = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\ldots,\boldsymbol{\theta}_n)$, then

$$p(\boldsymbol{y}_{n+1}|\boldsymbol{Y}) = \mathsf{E}[\mathsf{E}[p(\boldsymbol{y}_{n+1}|\boldsymbol{\theta}_{n+1})|\boldsymbol{\Theta}_n]|\boldsymbol{Y}] = \sum_{k=1}^{N}\mathsf{E}[\pi_k p(\boldsymbol{y}_{n+1}|\boldsymbol{V}_k)|\boldsymbol{Y}]. \tag{12}$$

Consequently, $p(\boldsymbol{y}_{n+1}|\boldsymbol{Y})$ can be approximated by averaging the mixture of normal density over the simulated observations. A predictive density estimate can be derived by evaluating the averaged density over a refined partition. More specifically, suppose that $\{(\boldsymbol{\pi}^{(m)},\boldsymbol{V}^{(m)}) : m = 1,2,\ldots,M\}$ are the simulated observations obtained from the blocked Gibbs sampler, then, a consistent estimate for $p(\boldsymbol{y}_{n+1}|\boldsymbol{Y})$ is given by

$$\begin{aligned}\widehat{p}(\boldsymbol{y}_{n+1}|\boldsymbol{Y}) = \frac{1}{M}\sum_{m=1}^{M}\sum_{g=1}^{N}\pi^{(m)}_g\frac{1}{(\sqrt{2\pi}\,)^p|\boldsymbol{\Sigma}^{(m)}_g|^{1/2}}\\ \times\exp\Big\{-\frac{1}{2}(\boldsymbol{y}_{n+1}-\boldsymbol{V}^{\mu(m)}_g)^{\intercal}\boldsymbol{\Sigma}^{(m)-1}_g(\boldsymbol{y}_{n+1}-\boldsymbol{V}^{\mu(m)}_g)\Big\}\end{aligned}$$

with $\boldsymbol{\Sigma}^{(m)}_g = \boldsymbol{V}^{\lambda(m)}_g\boldsymbol{V}^{\phi(m)}_g\boldsymbol{V}^{\lambda(m)\intercal}_g + \boldsymbol{V}^{\psi(m)}_g$.

§4. Model Comparison

4.1 Marginal Likelihood and Bayes Factor

Based on the well-known identity ([32]), the marginal log-likelihood of $\boldsymbol{Y}$ is given by

$$\ln m_N(\boldsymbol{Y}) = \ln p_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*) + \ln p(\boldsymbol{\vartheta}^*) - \ln p_N(\boldsymbol{\vartheta}^*|\boldsymbol{Y}),$$

where $\boldsymbol{\vartheta}^*$ is some point in the parameter space, $p(\boldsymbol{\vartheta}^*)$ and $p_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*)$ are respectively the prior density and the posterior density of parameters evaluated at that same point. $p_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*)$ is the likelihood of observed data $\boldsymbol{Y}$. Hence, if we find estimates $\widehat{p}_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*)$ and $\widehat{p}_N(\boldsymbol{\vartheta}^*|\boldsymbol{Y})$ of the likelihood and posterior ordinates, it follows that we can conveniently estimate the marginal likelihood as

$$\ln \widehat{m}_N(\boldsymbol{Y}) = \ln \widehat{p}_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*) + \ln p(\boldsymbol{\vartheta}^*) - \ln \widehat{p}_N(\boldsymbol{\vartheta}^*|\boldsymbol{Y}). \tag{13}$$

It is obvious that $p(\boldsymbol{\vartheta}^*|\boldsymbol{Y})$ can be expressed as

$$p(\boldsymbol{\vartheta}^*|\boldsymbol{Y}) = p(\boldsymbol{\Phi}^*|\boldsymbol{Y})\, p(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{Y})\, p(\boldsymbol{\mu}^*|\boldsymbol{\Phi}^*, \boldsymbol{\Sigma}^*, \boldsymbol{Y}),$$

in which

$$\begin{aligned}
p(\boldsymbol{\Phi}^*\boldsymbol{Y}) &= \int p(\boldsymbol{\Phi}^*|\boldsymbol{V}, \boldsymbol{Y})\, p(\boldsymbol{V}|\boldsymbol{Y})\mathrm{d}\boldsymbol{V} = \int p(\boldsymbol{\Phi}^*|\boldsymbol{V})\, p(\boldsymbol{V}|\boldsymbol{Y})\mathrm{d}\boldsymbol{V},\\
p(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{Y}) &= \int p(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{\mu}, \boldsymbol{V}, \boldsymbol{Y})\, p(\boldsymbol{\mu}, \boldsymbol{V}|\boldsymbol{\Phi}^*, \boldsymbol{Y})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{V}\\
&= \int p(\boldsymbol{\Sigma}^*|\boldsymbol{\mu}, \boldsymbol{V})\, p(\boldsymbol{\mu}, \boldsymbol{V}|\boldsymbol{\Phi}^*, \boldsymbol{Y})\mathrm{d}\boldsymbol{\mu}\mathrm{d}\boldsymbol{V},\\
p(\boldsymbol{\mu}^*|\boldsymbol{\Sigma}^*, \boldsymbol{\Phi}^*, \boldsymbol{Y}) &= \int p(\boldsymbol{\mu}^*|\boldsymbol{\Phi}^*, \boldsymbol{\Sigma}^*, \boldsymbol{V}, \boldsymbol{Y})\, p(\boldsymbol{V}|\boldsymbol{\Sigma}^*, \boldsymbol{\Phi}^*, \boldsymbol{Y})\mathrm{d}\boldsymbol{V}\\
&= \int p(\boldsymbol{\mu}^*|\boldsymbol{\Sigma}^*, \boldsymbol{V})\, p(\boldsymbol{V}|\boldsymbol{\Sigma}^*, \boldsymbol{\Phi}^*, \boldsymbol{Y})\mathrm{d}\boldsymbol{V}
\end{aligned}$$

are the reduced conditional density ordinates. The first ordinate, $p(\boldsymbol{\Phi}^*|\boldsymbol{Y})$ can be estimated in an obvious way, by taking the ergodic average of the full conditional density with the posterior draws of $\boldsymbol{V}$ leading to estimate

$$\widehat{p}(\boldsymbol{\Phi}^*|\boldsymbol{Y}) = M^{-1} \sum_{m=1}^{M} p(\boldsymbol{\Phi}^*|\boldsymbol{V}^{(m)}).$$

The second ordinate $p(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{Y})$ is more intractable since the integral is taken with respect to $p(\boldsymbol{\mu}, \boldsymbol{V}|\boldsymbol{\Phi}^*, \boldsymbol{Y})$ which can not take advantage of the output directly from the Gibbs outputs. A simple solution is to run an additional M iterations with the complete conditional densities

$$p(\boldsymbol{V}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{\Phi}^*, \boldsymbol{Y}), \quad p(\boldsymbol{\mu}|\boldsymbol{V}, \boldsymbol{\Sigma}, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\Omega}, \boldsymbol{\Phi}^*, \boldsymbol{Y}) \quad \text{and} \quad p(\boldsymbol{\Sigma}, \boldsymbol{\pi}, \boldsymbol{L}, \boldsymbol{\Omega}|\boldsymbol{\mu}, \boldsymbol{V}, \boldsymbol{\Phi}^*, \boldsymbol{Y}).$$

From MCMC theory, with $\{\boldsymbol{\mu}^{(m)}, \boldsymbol{V}^{(m)}\}$ simulated from this run, a consistent estimate of $p(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{Y})$ is obtained by

$$\widehat{p}(\boldsymbol{\Sigma}^*|\boldsymbol{\Phi}^*, \boldsymbol{Y}) = M^{-1} \sum_{m=1}^{M} p(\boldsymbol{\Sigma}^*|\boldsymbol{\mu}^{(m)}, \boldsymbol{V}^{(m)}).$$

A similar technique, with an important twist, can be invoked to obtain the reduced conditional ordinate $p(\boldsymbol{\mu}^*|\boldsymbol{\Phi}^*, \boldsymbol{\Sigma}^*, \boldsymbol{Y})$. Note that the full conditional distributions $p(\boldsymbol{\Phi}^*|\boldsymbol{V})$, $p(\boldsymbol{\Sigma}^*|\boldsymbol{\mu}, \boldsymbol{V})$ and $p(\boldsymbol{\Sigma}^*|\boldsymbol{\mu}^*, \boldsymbol{V})$ are standard, hence, the computation is straightforward.

The key quantity in (13) is to compute the observed likelihood $\widehat{p}_N(\boldsymbol{Y}|\boldsymbol{\vartheta}^*)$. Basu and Chib [31] adopted the sequential importance sampling (SIS) algorithm to compute the likelihood ordinates for the Dirichlet process mixture models. Since no predictive rule of sampling can be available for the finite dimensional Dirichlet prior mixture model, here we extend the generalized weighted Chinese restaurant algorithm [33] to compute observed likelihood (see Appendix).

4.2 L_ν-Measure

Another attractive measure for model comparisons is the L_ν-measure which is based on the posterior predictive density $p(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y})$, where $\boldsymbol{y}_i^{\rm rep}$ denotes the future values of a replicate experiment. Let $\boldsymbol{Y}^{\rm rep} = \big(\boldsymbol{y}_1^{\rm rep\,\intercal}, \boldsymbol{y}_2^{\rm rep\,\intercal}, \ldots, \boldsymbol{y}_n^{\rm rep\,\intercal}\big)^\intercal$ be a collection set of future responses of $\boldsymbol{y}_i$ in a replicate experiment. For some $0 \leqslant \nu < 1$, we consider the following multivariate version of L_ν-measure:

$$L_\nu(\boldsymbol{Y}) = \sum_{i=1}^n \mathrm{tr}\big[\mathsf{Cov}\,(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y})\big] + \nu \sum_{i=1}^n \mathrm{tr}\big[\big\{\mathsf{E}(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y}) - \boldsymbol{y}_i\big\}\big\{\mathsf{E}(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y}) - \boldsymbol{y}_i\big\}^\intercal\big], \tag{14}$$

where the expectation is taken with respect to the posterior predictive distribution of $p(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y})$ given by

$$p(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y}) = \int p(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{V}, \boldsymbol{\pi})\, p(\boldsymbol{V}, \boldsymbol{\pi}|\boldsymbol{Y}) \mathrm{d}\boldsymbol{V}\, \mathrm{d}\boldsymbol{\pi}. \tag{15}$$

Clearly, small values of the L_ν-measure indicate that the model gives predictions closed to the observed values, and the variability in the predictions is low as well. Hence, the model with the smallest L_ν-measure is selected from a collection of competing models. It has been shown that L_ν-measure with $\nu = 0.5$ has nice theoretical properties [35]. Thus, this value of ν will be used in our empirical illustrations.

Two quantities $\mathsf{E}(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y})$ and $\mathsf{E}(\boldsymbol{y}_i^{\rm rep}\boldsymbol{y}_i^{\rm rep\,\intercal}|\boldsymbol{Y})$ are involved in calculating $L_\nu(\boldsymbol{Y})$:

$$\mathsf{E}(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{Y}) = \mathsf{E}(\mathsf{E}(\boldsymbol{y}_i^{\rm rep}|\boldsymbol{V}, \boldsymbol{\pi})|\boldsymbol{Y}) = \sum_{k=1}^N \mathsf{E}(\pi_k \boldsymbol{V}_k^\mu|\boldsymbol{Y})$$

and

$$\mathsf{E}(\boldsymbol{y}_i^{\rm rep}\boldsymbol{y}_i^{\rm rep\,\intercal}|\boldsymbol{Y}) = \mathsf{E}[\mathsf{E}(\boldsymbol{y}_i^{\rm rep}\boldsymbol{y}_i^{\rm rep\,\intercal}|\boldsymbol{V}, \boldsymbol{\pi}]|\boldsymbol{Y}] = \sum_{k=1}^N \mathsf{E}[\pi_k(\boldsymbol{V}_k^\lambda \boldsymbol{V}_k^\phi \boldsymbol{V}_k^{\lambda\intercal} + \boldsymbol{V}_k^\psi + \boldsymbol{V}_k^\mu \boldsymbol{V}_k^{\mu\intercal})|\boldsymbol{Y}].$$

Based on MCMC sample $\{(\boldsymbol{V}^{(t)}, \boldsymbol{\pi}^{(t)}, \boldsymbol{L}^{(t)}, \boldsymbol{\vartheta}^{(t)}) : t = 1, 2, \ldots, T\}$ already available in the estimation, the consistent estimate for L_ν statistic can be approximated via ergodic average.

§5. Empirical Illustrations

5.1 Simulation Study

A simulation study is presented to give some ideas on the performance of the proposed approach. The data set is generated from a mixture of factor analysis models with two components, of which each component involves six manifest variables which are related to two latent factors. The population values of elements in $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are taken as:

$$\boldsymbol{\Lambda}_1^{\top} = \begin{bmatrix} 0.4 & 0.4 & 0.4 & 0.0^* & 1.0^* & 0.0^* \\ 0.0^* & 0.4 & 0.4 & 0.4 & 0.0^* & 1.0^* \end{bmatrix},$$

$$\boldsymbol{\Lambda}_2^{\top} = \begin{bmatrix} 0.8 & 0.8 & 0.8 & 0.0^* & 1.0^* & 0.0^* \\ 0.0^* & 0.8 & 0.8 & 0.8 & 0.0^* & 1.0^* \end{bmatrix},$$

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}, \qquad \boldsymbol{\Phi}_2 = \begin{bmatrix} 1.0 & -0.3 \\ -0.3 & 1.0 \end{bmatrix}, \tag{16}$$

in which the parameters with asterisks are treated as known for model identification. The true population values of other unknown parameters are given by: $\boldsymbol{\mu}_1 = -1.5 \times \mathbf{1}_6$, $\boldsymbol{\mu}_2 = 2.5 \times \mathbf{1}_6$, where $\mathbf{1}_6$ is a 6×1 vector with all elements equal to 1; $\psi_{1j} = 0.36$, $\psi_{2j} = 0.64$ for all $j = 1, 2, \ldots, 6$. The mixing proportions are $\pi_1 = 0.3$, $\pi_2 = 0.7$. With these settings, the data set with sample size 200 are generated from the mixture of two CFA models.

For the Bayesian analysis, the following two inputs of the hyperparameters in the prior distribution (10) are considered: (I) $\boldsymbol{\mu}_0 = \overline{\boldsymbol{y}}$, $\boldsymbol{\Sigma}_0 = \boldsymbol{S}_y$ where $\overline{\boldsymbol{y}}$ and $\boldsymbol{S}_y$ are the sample mean and sample covariance matrix of the simulated sample; $\rho_{0\Sigma} = \rho_{0\phi} = s = 10.0$, $\lambda_{0jk} = 0.68$ in $\boldsymbol{\Lambda}_0$, $\boldsymbol{H}_{0j} = \boldsymbol{I}_2$, $\alpha_{j0} = 9.0$, $\beta_{j0} = 8.0$ for all $j = 1, 2, \ldots, 6$, $\boldsymbol{R}_{0\Sigma}^{-1} = \boldsymbol{I}_6$ and $\boldsymbol{R}_{0\phi}^{-1} = \boldsymbol{I}_2$; (II) $\boldsymbol{\mu}_0 = \mathbf{0}$, $\boldsymbol{\Sigma}_0 = 100.0\boldsymbol{I}_6$, $\rho_{0\Sigma} = \rho_{0\phi} = s = 10.0$, $\lambda_{0jk} = 0$, $\boldsymbol{H}_{0j} = 100\boldsymbol{I}_2$, $\alpha_{j0} = \beta_{j0} = 2.0$ for all $j = 1, 2, \ldots, 6$, $\boldsymbol{R}_{0\Sigma}^{-1} = 100.0\boldsymbol{I}_6$ and $\boldsymbol{R}_{0\phi}^{-1} = 100.0\boldsymbol{I}_2$. Since we choose rather dispersed covariance and scale matrices in the priors of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$, these priors are approximatively noninformative. Moreover, we fix the smaller values for both shape α_{0j} and scale β_{0j} in (9) to encourage larger or smaller values for V_{mj}^{ψ}. The value c should be selected with care because it controls the amount of smoothing and directly affects the number of estimated clusters. A natural choice is to assign a gamma prior for it. Unfortunately, under this setting, we will need to resort to a Metropolis-Hastings step which involves the complex psi function. This will drastically cause the computational burden. We address this problem by tuning value of c and take $c = n^{-1}$, $c = n^0$, $c = n^1$ and $c = n^2$ to reflect smaller, moderate, and larger departures from the single parametric CFA.

We first conduct a few test runs as a pilot study to obtain some ideas about the

number of the Gibbs sampler iterations in getting convergence. We found that in all these runs, the Gibbs sampler converges in about 2 000 iterations. To be conservative, we collect 1 000 observations after 2 000 burn-ins for posterior analysis.

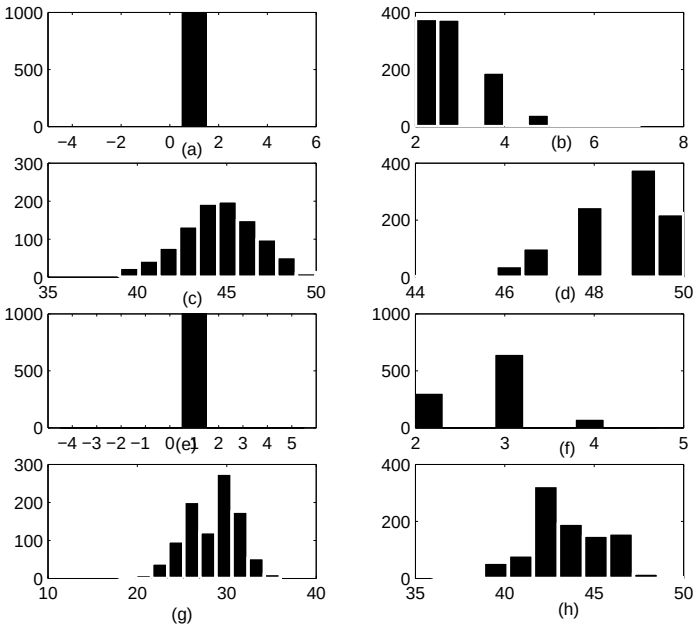

Figure 1 **Histograms of the number of the distinct values of $\boldsymbol{\theta}_i$ under prior (I) and (II) in the simulated data with size $n = 200$ and level $N = 50$. (a) and (e): $c = n^{-1}$; (b) and (d): $c = n^0$; (c) and (g): $c = n^1$; (d) and (h): $c = n^2$; Panels (a)-(d) under prior (I) and panels (e)-(h) for prior (II).**

Figure 1 illustrates the number of the distinct values of $\boldsymbol{\theta}_i$s for different values of c under priors (I) and (II) with $N = 50$. As is expected, there is only one cluster among the $\boldsymbol{\theta}_i$'s when c approaches to zero and the finite dimensional mixture model is equivalent to the single CFA. As c increases, the number of clusters also increases and amounts to N when $c = n^2$. This is similar to those given by Escobar who deals with the univariate normal mean mixture model using the infinite dimensional Dirichlet process. An interesting result is that prior (I) tends to give more numbers of estimating clusters than prior (II). We compute the logarithms of the marginal likelihood and $L_{0.5}$ measures of the data under prior (I) for different values of c and N through the proposed approach. For the generalized Chinese restaurant algorithm, we collect 3 000 observations after convergence to compute $\ln p(\boldsymbol{\vartheta}^*|\boldsymbol{Y})$ and take 100 replications in the sequential sampling. For each replication, We sample 2 000 independent observations in approximating integrals involved in the GWCR algorithm. Table 1 gives the summary of the logarithms of marginal likelihoods and values of $L_{0.5}$-measure. Based on Table 1, the log-likelihoods of the observed data

and $L_{0.5}$-measure both seem to support moderate value of c.

Table 1 Logarithm of likelihood and $L_{0.5}$-measure for different values and truncated levels in the simulated data: prior (I)

		$\ln m(\boldsymbol{Y})$		$L_{0.5}(\boldsymbol{Y})$	
c		$N=50$	$N=100$	$N=50$	$N=100$
$c=n^{-1}$:					
	Mean	-2 149.12	-2 178.03	7 068.78	6 998.02
	Med.	-2 108.61	-2 136.26	7 176.44	7 006.78
	Sd.	122.86	179.08	594.10	596.52
$c=1.0$:					
	Mean	-2 118.69	-2 079.87	7 115.19	7 061.47
	Med.	-2 071.69	-2 054.38	7 256.65	7 136.67
	Sd.	158.87	90.40	597.49	573.29
$c=n$:					
	Mean	-1 923.47	-1 900.39	6 902.17	6 732.32
	Med.	-1 922.09	-1 896.83	6 999.53	6 806.66
	Sd.	26.84	30.01	591.42	553.17
$c=n^2$:					
	Mean	-1 966.70	-1 992.86	7 185.39	7 218.48
	Med.	-1 942.83	-1 989.78	7 267.93	7 285.67
	Sd.	55.33	79.65	599.74	581.62

Synthetic data set generated from the proposal model (11) is analyzed to show the effectiveness of the L_v measure and BF. For $n=500$, we first generate L_i from discrete distribution $\sum_{m=1}^{5}\pi_m\delta_m(\cdot)$ with $\boldsymbol{\pi}$ drawn from Dirichlet$(0.1,\ldots,0.1)$, then draw $\boldsymbol{\omega}_i$ from $N_2(\mathbf{0},\boldsymbol{\Phi}_{L_i})$ and sample $\boldsymbol{y}_i$ from $N(\boldsymbol{\mu}_{L_i},\boldsymbol{\Lambda}_{L_i}\boldsymbol{\omega}_i,\boldsymbol{\Psi}_{L_i})$, in which $\boldsymbol{\mu}_g$, $\boldsymbol{\Lambda}_g$, $\boldsymbol{\Psi}_g$, and $\boldsymbol{\Phi}_g$ are given by $\boldsymbol{\mu}_g=(-2.5+0.5*(g-1))\mathbf{1}_6$, $\boldsymbol{\Lambda}_g=\boldsymbol{\Lambda}_2$, $\boldsymbol{\Psi}_g=\boldsymbol{\Psi}_2$ and $\boldsymbol{\Phi}_g=\boldsymbol{\Phi}_2$. The elements in $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Phi}_2$ are given in (16). The $L_{0.5}$ measures and BFs under different N and c are calculated. We choose six pairs for (N,c): $(5,5)$, $(50,5)$, $(100,5)$, $(5,10)$, $(50,10)$ and $(100,10)$. Note that the first corresponds to the true model. We do 100 replications and find that the correct rates via $L_{0.5}$ measure and BF are about 0.88 and 0.93, respectively. Since our modeling focuses on exploring the underlying clusters among the data, selections of particular parametric structures involved in CFA are not presented here and will constitute the further study.

Figure 2 presents the contours of pair (y_1,y_2) overlaid on the simulated data for $N=50$ based on the 60×60 grids. It is clear that there is only one mode in the posterior predictive density when $c=n^{-1}$ and more than two modes in the posterior density when $c=1.0$.

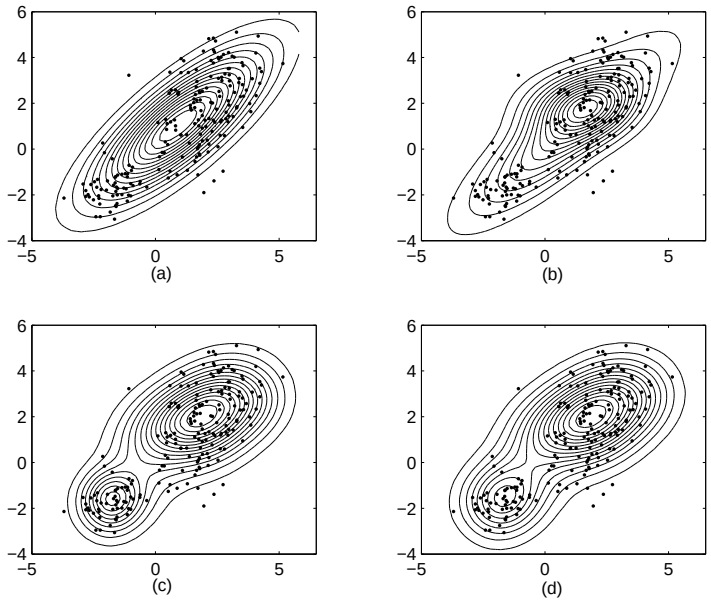

Figure 2 **Bivariate posterior predictive density estimates for pair (y_1, y_2) overlaid on corresponding plots of simulated data with size $n = 200$ and $N = 50$: (a) $c = n^{-1}$; (b) $c = n^0$; (c) $c = n$; (d) $c = n^2$.**

To investigate the differences of performance between our proposal methodology (denoted by MPH) and Lee, Lu and Song's modeling [26] (denoted by MFA), we consider the situation in which the factor model is the mixture model with two components:

$$\boldsymbol{\omega}_i \sim 0.3N_2(0, \boldsymbol{\Phi}_1) + 0.7N_2(0, \boldsymbol{\Phi}_2).$$

The true values of unknown parameters are taken as: $\boldsymbol{\mu} = \mathbf{0}_6$, $\psi_{\epsilon j} = 1.0$ $(j = 1, 2, \ldots, 6)$,

$$\boldsymbol{\Lambda}^{\tau} = \begin{bmatrix} 1.0^* & 0.8 & 0.8 & 0.0^* & 0.0^* & 0.0^* \\ 0.0^* & 0.0^* & 0.0^* & 1.0^* & 0.8 & 0.8 \end{bmatrix},$$

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 1.0 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\Phi}_2 = \begin{bmatrix} 1.0 & -0.3 \\ -0.3 & 1.0 \end{bmatrix}.$$

Based on the above settings, a data set with sample size 500 is generated. We implement our proposal and Lee, Lu and Song's procedure to analyze such data simultaneously. For comparison, we assign finite dimensional DP to the distributions of $\boldsymbol{\Phi}$ and $\boldsymbol{\omega}$, respectively. Note that the latter corresponds to Lee, Lu and Song's model. The following inputs for hyper-parameters involved in the parametric components and semiparametric components are used: $\boldsymbol{\mu}_0 = \overline{\boldsymbol{y}}$, $\boldsymbol{\Sigma}_0 = \boldsymbol{S}_y$, $\lambda_{0jk} = 0.8$ in $\boldsymbol{\Lambda}_0$, $\boldsymbol{H}_{0j} = \boldsymbol{I}_2$, $\alpha_{j0} = 9.0$, $\beta_{j0} = 8.0$ for all $j = 1, 2, \ldots, 6$, $\rho_0 = 10.0$ and $\boldsymbol{R}_{0\phi}^{-1} = \boldsymbol{I}_2$. Moreover, we take $G = 50$ and assume

$\alpha \sim \text{Gamma}(2.0, 2.0)$ in the finite dimensional DP. Table 2 gives the biases (BIAS), root mean squares (RMS) and standard deviations (SD) of estimates of unknown parameters under two models based on 100 replications. It can be seen clearly that under MFA, there exist serious biases for estimates of factor loadings and variance parameters of unique errors. The underlying reason is that under our setting, the variance and covariance parameters of observed variables are flatted, which may not be identified by MFA model. It also indicates that MFA can not capture the clustering effect among $\mathbf{\Phi}$s. We also considered the case when $\boldsymbol{y}$ is from a mixture model with $\boldsymbol{\mu}_1 = -1.5\mathbf{1}_6$ and $\boldsymbol{\mu}_2 = 0.5\mathbf{1}_6$ but to keep $\mathbf{\Lambda}$, $\mathbf{\Psi}_\epsilon$ invariant across two components. The results (not presented here) are similar to those in Table 2, hence omitted for saving space.

Table 2 Summary of estimates of unknown parameters under MFA and MPH: simulation study

	MFA			MPH		
Par.	BIAS	RMS	SD	BIAS	RMS	SD
μ_1	-0.223	0.603	0.354	0.025	0.063	0.063
μ_2	-0.018	0.249	0.338	-0.004	0.046	0.057
μ_3	-0.225	0.488	0.467	0.022	0.044	0.057
μ_4	-0.048	0.244	0.281	0.01	0.072	0.062
μ_5	-0.042	0.254	0.363	-0.023	0.051	0.058
μ_6	-0.127	0.366	0.384	0.005	0.076	0.057
λ_{21}	-0.241	0.405	1.101	0.073	0.113	0.093
λ_{31}	0.139	0.296	1.238	0.02	0.057	0.089
λ_{52}	0.025	0.183	1.196	0.057	0.133	0.095
λ_{62}	-0.103	0.345	1.194	0.017	0.144	0.091
$\psi_{\epsilon 1}$	0.966	0.972	0.123	0.052	0.126	0.112
$\psi_{\epsilon 2}$	0.664	0.668	0.103	-0.07	0.105	0.09
$\psi_{\epsilon 3}$	0.628	0.633	0.103	-0.026	0.058	0.087
$\psi_{\epsilon 4}$	0.996	1.004	0.125	-0.04	0.107	0.11
$\psi_{\epsilon 5}$	0.643	0.648	0.103	0.008	0.057	0.092
$\psi_{\epsilon 6}$	0.621	0.632	0.102	-0.035	0.094	0.085

5.2 Diabetic Nephropathy Dat

We illustrate the proposed method through analyzing a data set about diabetic nephropathy (kidney disease) of type 2 diabetes patients. Data are obtained from high risk diabetes patients who were participated in an applied genomic program conducted by the Institute of Diabetes, and underwent a comprehensive assessment of complications based on the European Diabetes protocol. The endogenous (outcome) variable of dia-

betic nephropathy is assessed as a latent variable reflected by two observed continuous phenotype variables: logarithm urinary albumin creatinine ratio (lnACR) and logarithm plasma creatinine (PCr). Based on some preliminary data analysis, and motivated by some medical findings [40], we select an exogenous latent variable related to lipid control via continuous phenotype observed variables: non-high density lipoprotein cholesterol (non-HDL), lower density lipoprotein cholesterol (LDL), and logarithm plasma triglyceride (TG) that may have significant effects on diabetic nephropathy. Furthermore, continuous phenotype observed variables: fasting plasma glucose (FPG) and glycated hemoglobin (HbA1c) are selected to form a latent variable related to glycemic control. With missing responses removed, the sample size of the data set is 648.

Let $\boldsymbol{y} = (\text{lnACR, PCr, HbA1c, FPG, non-HDL, LDL, TG})^{\mathsf{T}}$ be the vector of the observed variables. Based on the objective of this example and the medical motivation given in the last paragraph, it is natural to group (i) {lnACR, PCr} to an endogenous latent variable that can be interpreted as "diabetic nephropathy, η"; (ii) {HbA1c, FPG} and {non-HDL, LDL, TG} to two exogenous phenotype latent variables that can be respectively interpreted as "glycemic control, ξ_1", and "lipid control ξ_2". Hence, the following loading matrix $\boldsymbol{\Lambda}_i$ in the measurement equation with $\boldsymbol{\omega}_i = (\eta_i, \xi_{i1}, \xi_{i2})^{\mathsf{T}}$ is considered:

$$\boldsymbol{\Lambda}_i^{\mathsf{T}} = \begin{bmatrix} 1^* & \lambda_{i21} & 0^* & 0^* & 0^* & 0^* & 0^* \\ 0^* & 0^* & 1^* & \lambda_{i42} & 0^* & 0^* & 0^* \\ 0^* & 0^* & 0^* & 0^* & 1^* & \lambda_{i63} & \lambda_{i73} \end{bmatrix}.$$

Although other structures of $\boldsymbol{\Lambda}_i$ could be used, here we consider an non-overlapped structure for clear interpretation of the latent variables.

The proposed approach with $N = 50$ is applied to take the posterior analysis with the following prior inputs for the hyper-parameters in the conjugate prior distributions: $\boldsymbol{\mu}_0 = \widetilde{\boldsymbol{\mu}}_0$, $\boldsymbol{\Sigma}_0 = 1.0$, $\alpha_{0k} = \beta_{0k} = 4.0$, for $k = 1, 2, \ldots, 7$, $\lambda_{0jk} = \widetilde{\lambda}_{0jk}$, $\boldsymbol{H}_{j0} = \boldsymbol{I}_3$, $\rho_{0\Sigma} = 20$, $\rho_{0\phi} = s = 10.0$, $\boldsymbol{R}_{0\phi} = (\rho_{0\phi} - 4)\widetilde{\boldsymbol{\Phi}}_0$ where $\widetilde{\boldsymbol{\mu}}_0$, $\widetilde{\boldsymbol{\Lambda}}_0$, and $\widetilde{\boldsymbol{\Phi}}_0$ denote the maximum likelihood estimates obtained from analyzing a 'control-group' sample of diabetes patients under single CFA.

We take $c = 0.5$, 5.0, 50.0 and 500.0 in our posterior analysis. The log-likelihoods of observed data and the values of $L_{0.5}$-measure under $c = 0.5$, 5.0, 50.0 and 500.0 are $\{-6\,766.74, \mathbf{-6\,227.41}, -6\,257.57, -6\,430.85\}$ and $\{6\,285.84, \mathbf{6\,171.21}, 6\,227.49, 6\,347.50\}$, respectively. It appears that the moderate value of c favors the data fit. Furthermore, we also compute the $L_{0.5}$-measure for the single CFA. The resulting summary is $\ln p(\boldsymbol{Y}) = -6\,859.56$ and $L_{0.5}(\boldsymbol{Y}) = 6\,286.87$, which shows that the finite dimensional DP mixture model in analyzing such data is appropriate.

The histograms, Bayes density estimates using the proposed model and the Bayes

density estimates under parametric CFA for three manifest variables "logAcr", "Pcr" and "TG" are given in Figure 3. The histograms illustrate that the distributions of selected variables are deviated from normality in terms of bimodality and skewness. The superior predictive performance of the finite dimensional DP mixture over the common CFA is evident from Figure 3. The proposed model is successful in capturing these features while the common CFA fails. For the computation of Bayesian nonparametric density, we choose 40 refined grids in interval $[-4, 6]$ and collect 1 000 simulated observations from the blocked Gibbs sampler at each point after an initial 5 000 iteration burn-ins.

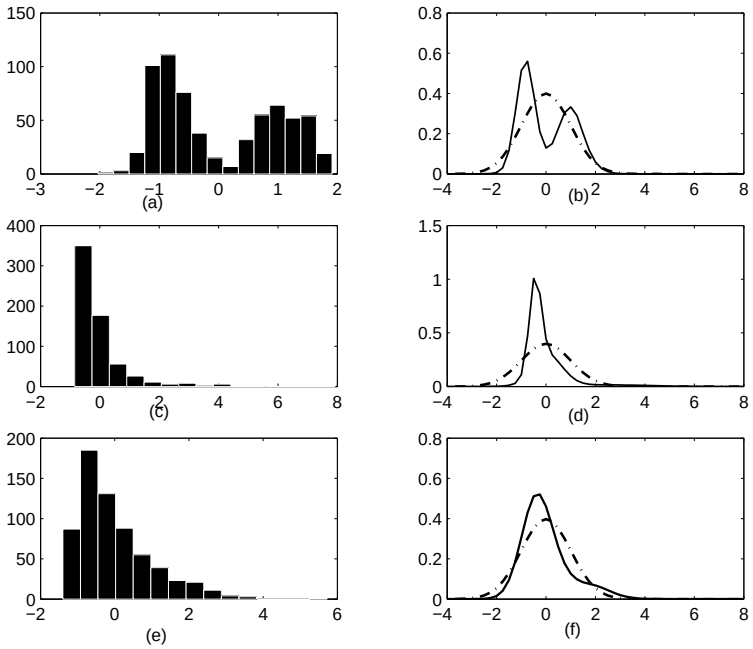

Figure 3 **Diabetic nephropathy data. (a), (c) and (e): histograms for logAcr, Pcr and TG; (b), (d) and (f): Bayesian posterior density estimates for logAcr, Pcr and TG: solid lines correspond to finite dimensional DP with $c = 5.0$ and $N = 50$ and dashed lines represent the parametric CFA.**

Figure 4 illustrates the bivariate posterior predictive densities for pairs of first four observed variables, which are based on 50×50 grids and 5 000 output posterior samples after 3 000 burn-in iterations. Our model-based predictive inference is more accurate and has captured well the higher frequency regions. The posterior number of the distinct $\boldsymbol{\theta}_i$ values is given in Figure 5, although this method generally tends to overestimate the number of the components. Based on Figure 5, we find that most of the posterior distribution are concentrated on anywhere from 6 to 10 clusters, thus presenting evidence for the presence of at least 5 components.

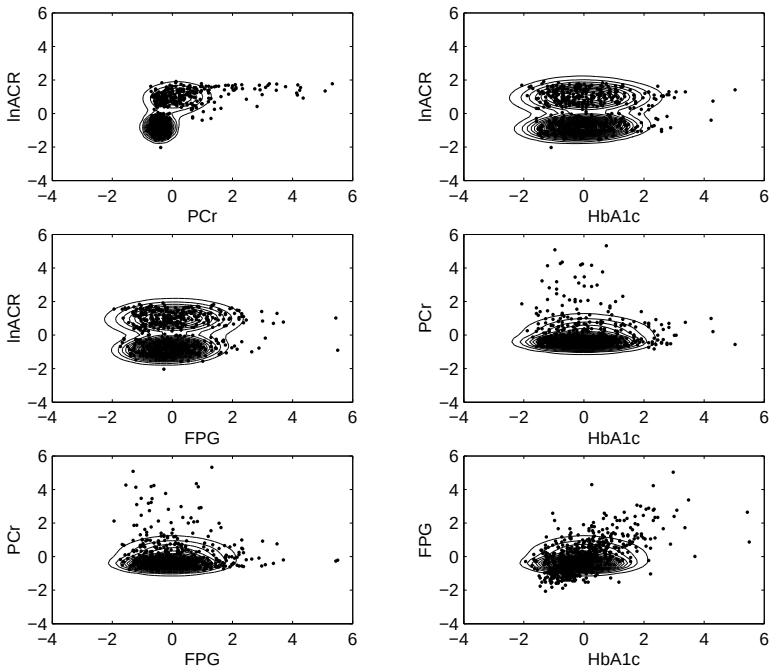

Figure 4 **Bivariate posterior predictive density estimates for pairs (logAcr, Pcr), (logAcr, HbA1c), (logAcr, FPG), (Pcr, HbA1c), (Pcr, FPG) and (HbA1c, FPG) under the finite dimensional DP with $c = 5.0$ and $N = 50$. All plots are overlaid on corresponding plots of Diabetic nephropathy data.**

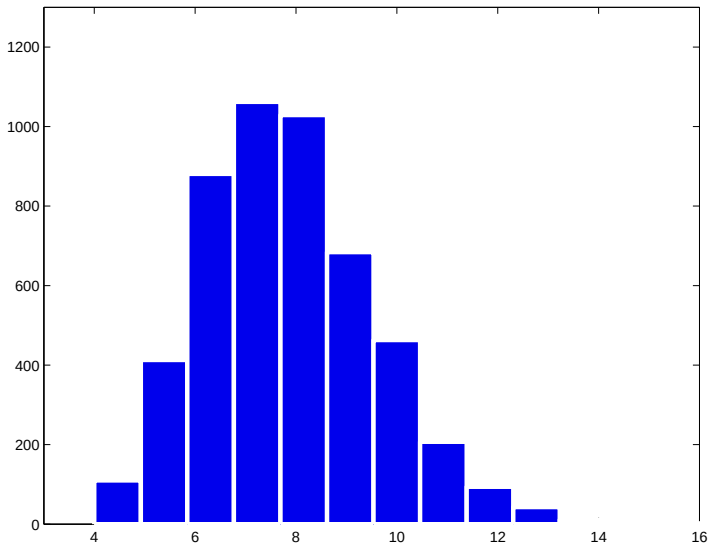

Figure 5 **Posterior distribution of the number of distinct θ_i values in the analysis of the diabetic nephropathy data with $c = 5.0$ and $N = 50$.**

§6. Concluding Remarks

Parametric modeling for CFA has long dominated Bayesian inference work, typically developed within standard exponential family. Such modeling is often confused with handling the multimodal and unknown heterogeneous problems. In dealing with multimodality or increased heterogeneity in data one is naturally to resort to the finite mixture model which is more flexible and feasible to implement due to advances in simulated-based model fitting.

Rather than handling the very large number of parameters resulting from the finite mixture models, in this article, we consider the finite-dimensional Dirichlet process mixture model for CFA with continuous responses. The core of our proposal is to model the mean vector and variance-covariance parameters of unique errors and latent variables into the finite-dimensional Dirichlet prior, which allows local dependence structure such as classification groups and clustering among the data. For posterior analysis, the blocked Gibbs sampler developed by Ishwaran and Zarepour [25] is extended to cope with the posterior inferences. The existing applications of the proposed methodologies can be applied to more general latent variable models that include the structure equation modeling, the multilevel SEMs [41], and longitudinal latent trait models [42] with discrete variables.

Appendix

A.1 Proof of Theorem 1 and 2

Proof of Theorem 1 The first part of conclusion (i) follows from that

$$\mathsf{E}\Big[\int_{\mathbb{R}^d}|\psi(x)|\mathsf{P}(\mathrm{d}x)\Big]=\mathsf{E}\sum_{k=1}^{N}p_k|\psi(V_k)|=\mathsf{E}|\psi(V_1)|\mathsf{E}\Big(\sum_{k=1}^{N}p_k\Big)=\mathsf{E}_{G_0}|\psi(V_1)|<\infty,$$

where the last equality holds since $\sum_{k=1}^{N}p_k=1$, a.s. Hence,

$$\int\int_{\mathbb{R}^d}\psi(x)\mathsf{P}(\mathrm{d}x)\mathscr{D}_N(\mathrm{d}\mathsf{P}|c\,G_0)=\mathsf{E}\sum_{k=1}^{N}p_k\psi(V_k)=\mathsf{E}_{G_0}\psi(V_1)=\mathsf{E}_{G_0}\psi(X).$$

For (ii),

$$\begin{aligned}\mathsf{E}\Big[\int_{\mathbb{R}^d}\psi_1(x)\mathsf{P}(\mathrm{d}x)\int_{\mathbb{R}^d}\psi_2(x)\mathsf{P}(\mathrm{d}x)\Big]&=\mathsf{E}\sum_{k=1}^{N}\sum_{l=1}^{N}p_kp_l\psi_1(V_k)\psi_2(V_l)\\&=\mathsf{E}\sum_{k=1}^{N}\psi_1(V_k)\psi_2(V_k)p_k^2+\mathsf{E}\sum_{k\neq l}^{N}\psi_1(V_k)\psi_2(V_l)p_kp_l\\&=\mathsf{E}[\psi_1(V_1)\psi_2(V_1)]\sum_{k=1}^{N}\mathsf{E}p_k^2+\sum_{k\neq l}^{N}\mathsf{E}[\psi_1(V_k)\psi_2(V_l)p_kp_l]\end{aligned}$$

$$= (\sigma_{12} + \mu_1\mu_2)\sum_{k=1}^{N} \mathsf{E}p_k^2 + \mu_1\mu_2 \sum_{k\neq l}^{N} \mathsf{E}[p_k p_l]$$
$$= \mu_1\mu_2 + \sigma_{12}\sum_{k=1}^{N} \mathsf{E}p_k^2 = \mu_1\mu_2 + \frac{\sigma_{12}}{c+1}\Big(1+\frac{c}{N}\Big),$$

where $\mu_i = \mathsf{E}_{G_0}\psi_i(X)$. The last equity holds since $p_k \sim \text{Beta}(c/N, c - c/N)$. □

Proof of Theorem 2 Note that given $\boldsymbol{P}$, $X_1^*, X_2^*, \ldots, X_{n(\boldsymbol{P})}^*$ are i.i.d. with the common density function $g(x)$, hence,

$$(X_1, X_2, \ldots, X_n)|\boldsymbol{P} = \prod_{j=1}^{n(\boldsymbol{P})} g(x_j^*).$$

Following from [37], the probability mass function of any partition $\boldsymbol{P}$ is given by

$$\varpi(\boldsymbol{P}) = \sum_{(i_1,i_2,\ldots,i_{n(\boldsymbol{P})})} \mathsf{E}\Big[\prod_{j=1}^{n(\boldsymbol{P})} \pi_{i_j}^{e_j}\Big] = \frac{(c/N)^{n(\boldsymbol{P})}N!}{c^{[n]}(N-n(\boldsymbol{P}))!}\prod_{j=1}^{n(\boldsymbol{P})}\Big(1+\frac{c}{N}\Big)^{[e_j-1]}, \tag{17}$$

which only depends on the magnitude of c and N. Multiplying them together gives the joint distribution of $X_1, X_2, \ldots, X_n$ and $\boldsymbol{P}$. The conclusion follows by restricting joint density on $\boldsymbol{S}_{C_1,C_2,\ldots,C_{n(\boldsymbol{P})}}$. □

A.2 GWCR Algorithm

Following the notation in Theorem 2, the likelihood of the observed data $\boldsymbol{Y}$ is given by

$$\begin{aligned} m(\boldsymbol{Y}|\boldsymbol{\vartheta}^*) &= \int\Big(\int\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}_i)\mathsf{P}(\mathrm{d}\boldsymbol{\theta}_i)\Big)\mathscr{D}_N(\mathrm{d}\mathsf{P}|c\,G_{0\vartheta^*}) \\ &= \int\int\int\prod_{i=1}^{n} p(\boldsymbol{y}_i|\boldsymbol{\theta}_i)\prod_{i=1}^{n}\Big(\sum_{k=1}^{N}\pi_k\delta_{\boldsymbol{V}_k}(\mathrm{d}\boldsymbol{\theta}_i)\Big)p(\mathrm{d}\boldsymbol{\pi})G_0(\mathrm{d}\boldsymbol{V}|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*,\boldsymbol{\Phi}^*) \\ &= \sum_{\boldsymbol{P}\in S}\varpi(\boldsymbol{P})(\boldsymbol{P})p(\boldsymbol{Y}|\boldsymbol{P},\boldsymbol{\vartheta}^*), \end{aligned} \tag{18}$$

in which $\varpi(\boldsymbol{P}) = \sum_{i_1\neq i_2\neq\cdots\neq i_k} \mathsf{E}\{\pi_{i_1}^{e_1}\pi_{i_2}^{e_2}\cdots\pi_{i_k}^{e_k}\}$ with $n(\boldsymbol{P}) = k$ is given in (17), $p(\boldsymbol{y}_i|\boldsymbol{\theta}_i)$ is the probability density function of $N(\boldsymbol{\mu}_i, \boldsymbol{\Lambda}_i\boldsymbol{\Phi}_i\boldsymbol{\Lambda}_i^{\intercal} + \boldsymbol{\Psi}_i)$ and $p(\boldsymbol{Y}|\boldsymbol{P},\boldsymbol{\vartheta}^*)$ is the conditional density

$$p(\boldsymbol{Y}|\boldsymbol{P},\boldsymbol{\vartheta}^*) = \prod_{j=1}^{n(\boldsymbol{P})}\int\prod_{i\in C_j} p(\boldsymbol{y}_i|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta}|\boldsymbol{\mu}^*,\boldsymbol{\Sigma}^*,\boldsymbol{\Phi}^*). \tag{19}$$

In the following derivations, we suppress $\boldsymbol{\vartheta}^*$ in (18) for notation simplicity.

For $r = 1, 2, \ldots, n$, let $\boldsymbol{p}_r$ be the partition of $\{1, 2, \ldots, r\}$, $\boldsymbol{Y}_{(r)} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_r\}$, and $\boldsymbol{P}_{(r)} = \{\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_r\}$. Based on the well-known identity,

$$p(\boldsymbol{Y},\boldsymbol{P}) = p(\boldsymbol{y}_1)\,p(\boldsymbol{p}_1|\boldsymbol{y}_1)\prod_{r=2}^{n} p(\boldsymbol{y}_r|\boldsymbol{Y}_{(r-1)},\boldsymbol{P}_{(r-1)})\,p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)},\boldsymbol{P}_{(r-1)}),$$

the marginal likelihood of $\boldsymbol{Y}$ is given by

$$m(\boldsymbol{Y}) = \sum_{\boldsymbol{P}} \prod_{r=1}^{n} \lambda(r)\, p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)}, \boldsymbol{P}_{(r-1)}),$$

where $\lambda(1) = p(\boldsymbol{y}_1)$, $\lambda(r) = p(\boldsymbol{y}_r|\boldsymbol{Y}_{(r-1)}, \boldsymbol{P}_{(r-1)})$ $(r = 2, 3, \ldots, n)$.

Sequential sampling is implemented by drawing $\boldsymbol{p}_r$ from $p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)}, \boldsymbol{P}_{(r-1)})$ sequentially and to form draws of $\boldsymbol{P}$. The marginal likelihood can be approximated by

$$\widehat{m}(\boldsymbol{Y}) = \frac{1}{M} \sum_{m=1}^{M} \prod_{r=1}^{n} \lambda^{(m)}(r),$$

in which $\lambda^{(m)}(r) = p(\boldsymbol{y}_r|\boldsymbol{Y}_{(r-1)}, \boldsymbol{P}^{(m)}_{(r-1)})$ and $\{\boldsymbol{P}^{(m)} : m = 1, 2, \ldots, M\}$ are the i.i.d. random observations simulated by the sequential sampling algorithm. To draw $\boldsymbol{P}$, we notice that

$$\begin{aligned} p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)}, \boldsymbol{P}_{(r-1)}) &= \frac{p(\boldsymbol{p}_r|\boldsymbol{P}_{(r-1)}, \boldsymbol{Y}_{(r-1)})\, p(\boldsymbol{y}_r|\boldsymbol{P}_{(r)}, \boldsymbol{Y}_{(r-1)})}{\lambda(r)} \\ &= \frac{p(\boldsymbol{Y}_{(r)}|\boldsymbol{P}_{(r)})}{\lambda(r)\, p(\boldsymbol{Y}_{(r-1)}|\boldsymbol{P}_{(r-1)})} \times \frac{p(\boldsymbol{P}_{(r)})}{p(\boldsymbol{P}_{(r-1)})}, \end{aligned} \tag{20}$$

in which $\lambda(r)$ can be considered as the normalizing constant of $p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)}, \boldsymbol{P}_{(r-1)})$.

A key quantity is the $p(\boldsymbol{P}_{(r)})$, the probability of exchangeable partition probability function. Ishwaran and James [37] identified these probabilities via the so-called generalized Chinese restaurant sampling scheme, which is a sequential restaurant 'seating arrangement'. Let $\boldsymbol{p}_r = \{C_{r,1}, C_{r,2}, \ldots, C_{r,n(\boldsymbol{p}_r)}\}$ and $C_{r,j}$ is the j-th cell with size $e_{r,j} = \#C_{r,j}$. It can be shown that

$$\frac{p(\boldsymbol{P}_{(r)})}{p(\boldsymbol{P}_{(r-1)})} = \begin{cases} \dfrac{e_{r-1,j} + c/N}{c + r - 1}, & \text{if } \boldsymbol{p}_r = \boldsymbol{p}_r^j = \{C_{r-1,1}, \ldots, C^*_{r-1,j}, \ldots, C_{r-1,m}\}; \\ \dfrac{c(1 - m/N)}{c + r - 1}, & \text{if } \boldsymbol{p}_r = \boldsymbol{p}_{r-1} \cup \{r\}, \end{cases}$$

where $C^*_{r-1,j} = C_{r-1,j} \cup \{r\}$. Hence,

$$\begin{aligned} &p(\boldsymbol{p}_r|\boldsymbol{Y}_{(r)}, \boldsymbol{P}_{(r-1)}) \\ &= \frac{1}{\lambda(r)} \times \begin{cases} \dfrac{e_{r-1,j} + c/N}{c + r - 1} \dfrac{\displaystyle\int p(\boldsymbol{y}_r|\boldsymbol{\theta}) \prod_{i \in C_{r-1,j}} p(\boldsymbol{y}_i|\boldsymbol{\theta}) G_0(\mathrm{d}\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}{\displaystyle\int \prod_{i \in C_{r-1,j}} p(\boldsymbol{y}_i|\boldsymbol{\theta}) G_0(\mathrm{d}\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}, & \text{if } \boldsymbol{p}_r = \boldsymbol{p}_r^j; \\ \dfrac{c(1 - m/N)}{c + r - 1} \displaystyle\int p(\boldsymbol{y}_r|\boldsymbol{\theta}) G_0(\mathrm{d}\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}, & \text{if } \boldsymbol{p}_r = \boldsymbol{p}_{r-1} \cup \{r\}. \end{cases} \end{aligned}$$

The GWCR algorithm works by building up a sequence of nested partitions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_n$ by assigning labels $\{1, 2, \ldots, n\}$ into sets using a posterior partition rule. Specifically,

(i) Set $\boldsymbol{p}_1 = \{1\}$, compute $\lambda(1) = p(\boldsymbol{y}_1) = \int p(\boldsymbol{y}_1|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta})$;

(ii) Given $\boldsymbol{p}_{r-1} = \{C_{r-1,1}, C_{r-1,2}, \ldots, C_{r-1,m}\}$ derived from $\{1, 2, \ldots, r-1\}$, compute

$$\begin{cases} \rho_r = \dfrac{c(1-m/N)}{c+r-1} \displaystyle\int p(\boldsymbol{y}_r|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta}), \\ \rho_r(j) = \dfrac{e_{r-1,j}+c/N}{c+r-1} \displaystyle\int p(\boldsymbol{y}_r|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|C_{r-1,j}, \boldsymbol{Y}_{(r-1,j)})\mathrm{d}\boldsymbol{\theta}, \qquad j = 1, 2, \ldots, m, \end{cases}$$

in which

$$p(\boldsymbol{\theta}|C_{r-1,j}, \boldsymbol{Y}_{(r-1,j)}) = \frac{\displaystyle\prod_{i\in C_{r-1,j}} p(\boldsymbol{y}_i|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta})}{\displaystyle\int \prod_{i\in C_{r-1,j}} p(\boldsymbol{y}_i|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta})}$$

and the normalizing constant $\lambda(r) = \rho(r) + \displaystyle\sum_{j=1}^{m} \rho(j)$;

(iii) $\boldsymbol{p}_r$ is formed by assigning label r to one of the previous sets $C_{r-1,j}$ with probability $\rho_r(j)/\lambda(r)$, $j = 1, 2, \ldots, m$, or by assigning label r to a new set with probability $\rho_r/\lambda(r)$;

(iv) Up to $r = n$, and compute

$$p_N(\boldsymbol{Y}) = \prod_{r=1}^{n} \lambda(r);$$

(v) Repeat (i)–(iv) B times gives an estimate

$$\widehat{p}_N(\boldsymbol{Y}) = \frac{1}{B}\sum_{b=1}^{B} \Lambda(\boldsymbol{P}^{(b)}).$$

Two quantities involved in generalized weighted Chinese restaurant algorithm are

$$\int p(\boldsymbol{y}_r|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta}), \qquad \text{and} \qquad \int p(\boldsymbol{y}_r|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|C_{r-1,j}, \boldsymbol{Y}_{(r-1,j)})\mathrm{d}\boldsymbol{\theta}, \tag{21}$$

which need assess the high dimensional integrals. This can be solved by Monte Carlo method. More specifically, with M i.i.d. observations $\{\boldsymbol{\theta}^{(m)} : m = 1, 2, \ldots, M\}$ simulated from $G_0(\cdot)$, consistent estimates for (21) are given by

$$\begin{aligned} &\int p(\boldsymbol{y}_r|\boldsymbol{\theta})G_0(\mathrm{d}\boldsymbol{\theta}) \approx \frac{1}{M}\sum_{m=1}^{M} \frac{1}{(\sqrt{2\pi})^p|\boldsymbol{\Sigma}^{(m)}|^{1/2}} \exp\Big\{-\frac{1}{2}d_r^{2(m)}\Big\}, \\ &\int p(\boldsymbol{y}_r|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|C_{r-1,j}, \boldsymbol{Y}_{(r-1,j)})\mathrm{d}\boldsymbol{\theta} \\ &\approx \frac{\displaystyle\sum_{m=1}^{M} |\boldsymbol{\Sigma}^{(m)}|^{-(e_{r-1,j}+1)/2} \exp\Big\{-\frac{1}{2}\Big(\sum_{i\in C_{r-1,j}} d_i^{2(m)} + d_r^{2(m)}\Big)\Big\}}{(\sqrt{2\pi})^p \displaystyle\sum_{m=1}^{M} |\boldsymbol{\Sigma}^{(m)}|^{-e_{r-1,j}/2} \exp\Big\{-\frac{1}{2}\sum_{i\in C_{r-1,j}} d_i^{2(m)}\Big\}}, \end{aligned}$$

in which $d_r^{2(m)} = (\boldsymbol{y}_r - \boldsymbol{\mu}^{(m)})^{\mathsf{T}} \boldsymbol{\Sigma}^{(m)-1} (\boldsymbol{y}_r - \boldsymbol{\mu}^{(m)})$ and $\boldsymbol{\Sigma}^{(m)} = \boldsymbol{\Lambda}^{(m)} \boldsymbol{\Phi}^{(m)} \boldsymbol{\Lambda}^{(m)} + \boldsymbol{\Psi}^{(m)\mathsf{T}}$.

References

[1] Jöreskog K, Sörbom D. *LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language* [M]. New ed. London: Psychology Press, 1993.

[2] Bentler P M, Wu E J C. *EQS 6 for Windows User's Guide* [M]. Encino, CA: Multivariate Software, 2002.

[3] Muthén B O. *LISCOMP: Analysis of Linear Structural Equations with a Comprehensive Measurement Model* [M]. Mooresville: Scientific Software, 1987.

[4] Muthén B O. Latent variable modeling in heterogeneous populations [J]. *Psychometrika*, 1989, **54(4)**: 557–585.

[5] Redner R A, Walker H F. Mixture densities, maximum likelihood and the EM algorithm [J]. *SIAM Rev.*, 1984, **26(2)**: 195–239.

[6] Titterington D M, Smith A F M, Markov U E. *Statistical Analysis of Finite Mixture Distributions* [M]. Chichester: Wiley, 1985.

[7] McLachlan G, Peel D. *Finite Mixture Models* [M]. New York: Wiley, 2000.

[8] Blåfield E. *Jyvaskyla Studies in Computer Science, Economics and Statistics, No. 2: Clustering of Observations from Finite Mixtures with Structural Information* [M]. Finland: Jyvaskyla University, 1980.

[9] Yung Y F. Finite mixtures in confirmatory factor-analysis models [J]. *Psychometrika*, 1997, **62(3)**: 297–330.

[10] Arminger G, Stein P. Finite mixtures of covariance structure models with regressors: loglikelihood function, minimum distance estimation, fit indices, and a complex example [J]. *Sociol. Methods Res.*, 1997, **26(2)**: 148–182.

[11] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm (with discussion) [J]. *J. Roy. Statist. Soc. Ser. B*, 1977, **39(1)**: 1–38.

[12] Jedidi K, Jagpal H S, DeSarbo W S. STEMM: a general finite mixture structural equation model [J]. *J. Classification*, 1997, **14(1)**: 23–50.

[13] McLachlan G J, Peel D, Bean R W. Modelling high-dimensional data by mixtures of factor analyzers [J]. *Comput. Statist. Data Anal.*, 2003, **41(3-4)**: 379–388.

[14] Meng X L, van Dyk D. The EM algorithm – an old folk-song sung to a fast new tune (with discussion) [J]. *J. Roy. Statist. Soc. Ser. B*, 1997, **59(3)**: 511–567.

[15] Hoshino T. Bayesian inference for finite mixtures in confirmatory factor analysis [J]. *Behaviormetrika*, 2001, **28(1)**: 37–63.

[16] Zhu H T, Lee S Y. A Bayesian analysis of finite mixtures in the LISREL model [J]. *Psychometrika*, 2001, **66(1)**: 133–152.

[17] Ferguson T S. A Bayesian analysis of some nonparametric problems [J]. *Ann. Statist.*, 1973, **1(2)**: 209–230.

[18] Antoniak C E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems [J]. *Ann. Statist.*, 1974, **2(6)**: 1152–1174.

[19] Lo A Y. On a class of Bayesian nonparametric estimates: I. density estimates [J]. *Ann. Statist.*, 1984, **12(1)**: 351–357.

[20] Escobar M D. Estimating normal means with a Dirichlet process prior [J]. *J. Amer. Statist. Assoc.*, 1994, **89(425)**: 268–277.

[21] Escobar M D, West M. Bayesian density estimation and inference using mixtures [J]. *J. Amer. Statist. Assoc.*, 1995, **90(430)**: 577–588.

[22] MacEachern S N, Müller P. Estimating mixture of Dirichlet process models [J]. *J. Comput. Graph. Statist.*, 1998, **7(2)**: 223–238.

[23] Walker S G, Damien P, Laud P W, et al. Bayesian nonparametric inference for random distributions and related functions [J]. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 1999, **61(3)**: 485–527.

[24] Müller P, Quintana F A. Nonparametric Bayesian data analysis [J]. *Statist. Sci.*, 2004, **19(1)**: 95–110.

[25] Ishwaran H, Zarepour M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models [J]. *Biometrika*, 2000, **87(2)**: 371–390.

[26] Lee S Y, Lu B, Song X Y. Semiparametric Bayesian analysis of structural equation models with fixed covariates [J]. *Stat. Med.*, 2008, **27(13)**: 2341–2360.

[27] Song X Y, Xia Y M, Lee S Y. Bayesian semiparametric analysis of structural equation models with mixed continuous and unordered categorical variables [J]. *Stat. Med.*, 2009, **28(17)**: 2253–2276.

[28] Yang M G, Dunson D B. Bayesian semiparametric structural equation models with latent variables [J]. *Psychometrika*, 2010, **75(4)**: 675–693.

[29] Chow S M, Tang N S, Yuan Y, et al. Bayesian estimation of semiparametric nonlinear dynamic factor analysis models using the Dirichlet process prior [J]. *Br. J. Math. Stat. Psychol.*, 2011, **64(1)**: 69–106.

[30] Kass R E, Raftery A E. Bayes factors [J]. *J. Amer. Statist. Assoc.*, 1995, **90(430)**: 773–795.

[31] Basu S, Chib S. Marginal likelihood and Bayes factors for Dirichlet process mixture models [J]. *J. Amer. Statist. Assoc.*, 2003, **98(461)**: 224–235.

[32] Chib S. Marginal likelihood from the Gibbs output [J]. *J. Amer. Statist. Assoc.*, 1995, **90(432)**: 1313–1321.

[33] Ishwaran H, James L F, Sun J. Bayesian model selection in finite mixtures by marginal density decompositions [J]. *J. Amer. Statist. Assoc.*, 2001, **96(456)**: 1316–1332.

[34] Laud P W, Ibrahim J G. Predictive model selection [J]. *J. Roy. Statist. Soc. Ser. B*, 1995, **57(1)**: 247–262.

[35] Gelfand A E, Ghosh S K. Model choice: a minimum posterior predictive loss approach [J]. *Biometrika*, 1998, **85(1)**: 1–11.

[36] Ibrahim J G, Chen M H, Sinha D. Criterion-based methods for Bayesian model assessment [J]. *Statist. Sinica*, 2001, **11(2)**: 419–443.

[37] Ishwaran H, James L F. Gibbs sampling methods for stick-breaking priors [J]. *J. Amer. Statist. Assoc.*, 2001, **96(453)**: 161–173.

[38] Müller P, Erkanli A, West M. Bayesian curve fitting using multivariate normal mixtures [J]. *Biometrika*, 1996, **83(1)**: 67–79.

[39] Roeder K, Wasserman L. Practical Bayesian density estimation using mixtures of normals [J]. *J. Amer. Statist. Assoc.*, 1997, **92(439)**: 894–902.

[40] Wang Y, Ng M C Y, So W Y, et al. Prognostic effect of insertion/deletion polymorphism of the ACE gene on renal and cardiovascular clinical outcomes in Chinese patients with type 2 diabetes [J]. *Diabetes Care*, 2005, **28(2)**: 348–354.

[41] Lee S Y, Song X Y. Maximum likelihood analysis of a general latent variable model with hierarchically mixed data [J]. *Biometrics*, 2004, **60(3)**: 624–636.

[42] Dunson D B. Dynamic latent trait models for multidimensional longitudinal data [J]. *J. Amer. Statist. Assoc.*, 2003, **98(463)**: 555–563.

实证因子模型的贝叶斯半参数分析和模型比较

夏业茂　　刘应安

(南京林业大学应用数学系, 南京, 210037)

摘　要: 为了解决多元数据的异质性, 对因子分析模型建立了贝叶斯半参数程序. 方法依赖于有限混合分布空间上先验分布的使用. 分块吉布斯抽样器用以进行后验分析. L_ν测度和贝叶斯因子给出模型比较. 基于广义加权中国餐馆算法, 给出了半参数模型下数据似然的计算. 经验结果显示了方法的有效性.

关键词: 因子分析模型; 有限维Drichlet过程先验; 分块吉布斯抽样器; 模型比较

中图分类号: O212.8; O212.4