Communications for Statistical Applications and Methods
2016, Vol. 23, No. 2, 131–146
http://dx.doi.org/10.5351/CSAM.2016.23.2.131
Print ISSN 2287-7843 / Online ISSN 2383-4757

The inference and estimation for latent discrete outcomes with a small sample

Hyung Choi[a], Hwan Chung[1,a]

[a]Department of Statistics, Korea University, Korea

Abstract

In research on behavioral studies, significant attention has been paid to the stage-sequential process for longitudinal data. Latent class profile analysis (LCPA) is an useful method to study sequential patterns of the behavioral development by the two-step identification process: identifying a small number of latent classes at each measurement occasion and two or more homogeneous subgroups in which individuals exhibit a similar sequence of latent class membership over time. Maximum likelihood (ML) estimates for LCPA are easily obtained by expectation-maximization (EM) algorithm, and Bayesian inference can be implemented via Markov chain Monte Carlo (MCMC). However, unusual properties in the likelihood of LCPA can cause difficulties in ML and Bayesian inference as well as estimation in small samples. This article describes and addresses erratic problems that involve conventional ML and Bayesian estimates for LCPA with small samples. We argue that these problems can be alleviated with a small amount of prior input. This study evaluates the performance of likelihood and MCMC-based estimates with the proposed prior in drawing inference over repeated sampling. Our simulation shows that estimates from the proposed methods perform better than those from the conventional ML and Bayesian method.

Keywords: dynamic data-dependent prior, latent class profile analysis, latent stage-sequential process, maximum posterior estimator, small samples

1. Introduction

Many behavioral and biomedical studies investigate longitudinal stability and change by analyzing collected survey data intended to address sequential patterns of behavioral development. In the area of substance use prevention and treatment, prevention scientists aim to find the best opportunities for intervening in substance use behavior to slow the process of drug dependence. Recently, a number of new developments in methods for the analysis of stage-sequential processes have been derived from latent class analysis (LCA) that includes the latent class profile analysis (LCPA). The LCPA identifies subtypes of the stage-sequential patterns of behavioral development (Chung *et al.*, 2011). In LCPA the measurement model at each time point is an LCA and the stage-sequential patterns are summarized in the form of latent class memberships that vary across multiple time points. A fundamental issue in LCPA involves specifying the relationshipf among class membership over multiple points in a flexible yet parsimonious way. The LCPA can classify a small number of common pathways of latent classes (i.e., latent class profiles) by considering the joint distribution of class membership over time as a mixture of class profiles that are not directly observable. LCPA has been applied to the study of

[1] Corresponding author: Department of Statistics, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea. E-mail: hwanch@korea.ac.kr

Published 31 March 2016 / journal homepage: http://csam.or.kr

early-onset drinking and subsequent drinking behavior among U.S. adolescents (Chung and Anthony, 2013). Four classes of drinking behavior where all drinkers in a class at a certain time point were expected to be homogeneous in terms of drinking behavior. They also showed that the sequences of drinking behavior were grouped into three class profiles, where all individuals in a given profile will have similar sequential pattern of class membership over time. In addition, they provided support for the view that patterns of sequential latent growth depend on the timing of drinking onset.

LCPA parameters can be estimated by maximum likelihood (ML) and Bayesian method using an expectation-maximization (EM) algorithm and Markov chain Monte Carlo (MCMC), respectively. With small samples, however, difficulties in ML and Bayesian inference and estimation can be caused by unusual properties in the likelihood of LCPA. In this study, we will describe and address several problems in small-sample inference for LCPA. We argue that problems can be alleviated with a small amount of prior input when conventional ML and Bayesian estimates behave erratically.

The organization of the rest of this article is: in Sections 2 and 3, we introduce LCPA and show potential complications in small-sample inference with conventional estimation algorithms such as EM and MCMC. To address these problems, a detailed explanation of estimation strategies is provided in Section 4. In Section 5, we evaluate the performance of estimates and intervals from proposed algorithms over repeated sampling. Our simulation shows that proposed methods perform better than the conventional ML and Bayesian method. We then apply proposed methods to alcohol drinking items drawn from the National Longitudinal Survey of Youth 1997 (NLSY97) to assess the performance of each algorithm.

2. Latent class profile analysis

Let $\mathbf{C} = (C_1, \ldots, C_T)$ denote the class membership variables from time $t = 1$ to T, and the i^{th} individual's observation c_{it} can take any integer value from 1 to C (i.e., $c_{it} = 1, 2, \ldots, C$) for $t = 1, 2, \ldots, T$. Let U denote the variable of the latent class profile membership with S nominal categories. The main idea of LCPA is that the relationship among class membership over T points in time can be explained by the assumption that the population consists of unobservable S class profiles. The class memberships $\mathbf{C} = (C_1, \ldots, C_T)$ are conditionally independent given the class profile U. If the i^{th} individual's class memberships over time, $\mathbf{c}_i = (c_{i1}, \ldots, c_{iT})$, and his/her class profile membership s_i are given, the joint probability that they belong to the class sequence $\mathbf{c}_i$ and the class profile s_i is

$$\begin{aligned} P(U = s_i, \mathbf{C} = \mathbf{c}_i) &= P(U = s_i)P(\mathbf{C} = \mathbf{c}_i \mid U = s_i) \\ &= P(U = s_i)\prod_{t=1}^{T} P(C_t = c_{it} \mid U = s_i). \end{aligned} \tag{2.1}$$

Equation (2.1) shows that the class memberships over T points in time $\mathbf{c}_i = (c_{i1}, \ldots, c_{iT})$ are conditionally independent given the class profile s_i. This property, called local independence assumption (Lazarsfeld and Henry, 1968), allows us to draw inferences about discrete latent variables. The class profile membership can be easily identified by an LCA if we can observe the individual's class memberships over T points in time. However, we should classify individual class memberships over time based on item responses in order to identify class profile membership because each individual's class membership over time is not directly observable.

Let $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Mt})$ represent the vectorized discrete M variables measuring latent class membership at time t for $t = 1, 2, \ldots, T$, and let $\mathbf{y}_{it} = (y_{i1t}, \ldots, y_{iMt})$ denote the i^{th} individual's observed values of $\mathbf{Y}_t$, where each response y_{imt} can take any integer value from 1 to r_m for $m = 1, \ldots, M$ and

$t = 1, \ldots, T$. Given the class profile s_i, the contribution of the i^{th} individual in the conditional probability of belonging to class sequencing $\mathbf{c}_i = (c_{i1}, \ldots, c_{iT})$ with responses $\mathbf{y}_i = (\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT})$ would be

$$\begin{aligned} P(\mathbf{C} = \mathbf{c}_i, \mathbf{Y} = \mathbf{y}_i \mid U = s_i) &= \prod_{t=1}^{T} P(C_t = c_{it} \mid U = s_i) P(\mathbf{Y}_t = \mathbf{y}_{it} \mid C_t = c_{it}) \\ &= \prod_{t=1}^{T} P(C_t = c_{it} \mid U = s_i) \prod_{m=1}^{M} P(Y_{mt} = y_{imt} \mid C_t = c_{it}). \end{aligned} \tag{2.2}$$

We assume the following to investigate the relationship among the manifest items in (2.2):

(i) $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Mt})$ are conditionally independent given the class membership C_t.

(ii) The profile membership U is related to items $\mathbf{Y}_t$ only through the class membership C_t.

Here, assumption (ii) implies that U depends on the class memberships over time $\mathbf{C} = (C_1, \ldots, C_T)$ but not on the items $(\mathbf{Y}_1, \ldots, \mathbf{Y}_T)$ conditioning on $\mathbf{C}$. The joint probability that the subject belongs to the class sequencing $\mathbf{c}_i = (c_{i1}, \ldots, c_{iT})$ given the profile s_i and responses $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})$ expressed as

$$\begin{aligned} L_i^* &= P(U = s_i, \mathbf{C} = \mathbf{c}_i, \mathbf{Y} = \mathbf{y}_i) \\ &= P(U = s_i) P(\mathbf{C} = \mathbf{c}_i, \mathbf{Y} = \mathbf{y}_i \mid U = s_i) \\ &= P(U = s_i) \prod_{t=1}^{T} P(C_t = c_{it} \mid U = s_i) \prod_{m=1}^{M} P(Y_{mt} = y_{imt} \mid C_t = c_{it}). \end{aligned} \tag{2.3}$$

Therefore, the i^{th} individual's contribution to the likelihood function of LCPA without considering class. Class profile can be represented by

$$\begin{aligned} L_i &= P(\mathbf{Y} = \mathbf{y}_i) \\ &= \sum_{s_i=1}^{S} \sum_{c_{i1}=1}^{C} \cdots \sum_{c_{iT}=1}^{C} L_i^* \\ &= \sum_{s=1}^{S} \gamma_s \prod_{t=1}^{T} \left[\sum_{c_t=1}^{C} \eta_{c_t|s}^{(t)} \prod_{m=1}^{M} \rho_{mk|c_t}^{(t)} \right]. \end{aligned} \tag{2.4}$$

In (2.4), the following three sets of parameters are estimated:

(i) $\rho_{mk|c_t}^{(t)} = P(Y_{mt} = k \mid C_t = c_t)$: the probability of the response k to the m^{th} item for a given class c_t at t time

(ii) $\eta_{c_t|s}^{(t)} = P(C_t = c_t \mid U = s)$: the conditional probability of belonging to class c_t at time t for a given class profile s

(iii) $\gamma_s = P(U = s)$: the probability of belonging to class profile s.

Here, the ρ-parameters, called *primary measurement parameters*, are generally constrained to be equal across each measurement occasion (i.e., $\rho_{mk|c} = \rho_{mk|c}^{(1)} = \cdots = \rho_{mk|c}^{(T)}$) for $k = 1, \ldots, r_m$ and

$m = 1, \ldots, M$. One difficulty with ρ-parameters that depend on waves is that the definition of the latent class may be unsettled as time advances. Thus, substantive researchers typically constrained the ρ-parameters to be equal across waves in order to make the meaning of the latent class invariant over T points in time. The η-parameters (called *secondary measurement parameters*) describes the relation among a class membership c_t at time t and a class profile s. A class profile can be interpreted by looking at the estimated secondary measurement parameters, which show a specific transitional pattern of class membership over time.

3. Standard estimation algorithms

Parameters in LCPA are easily estimated by ML via EM algorithm or Bayesian method via MCMC. In this part, we will introduce methods with respect to estimating parameters in the LCPA model and show potential complications in small-sample inference with conventional methods.

3.1. EM algorithm

The EM algorithm is a well-used solution for incomplete data with a missing value problem by searching ML through repeated process (Dempster *et al.*, 1977). In LCPA, it is difficult to maximize the observed-data log-likelihood function given in (2.4) directly. However, if latent class and latent class profiles memberships were observed, we could easily maximize the complete-data log-likelihood given in (2.3). Using this idea, EM algorithm maximizes the observed-data log-likelihood function by repeatedly updating the guess of latent memberships (i.e., class and class profile) and maximizing complete-data log-likelihood based on the updates. EM algorithm is an iterative process which has two steps (E-step and M-step) in each iteration. In the E-step, we calculate the expectation of unobservable values for the cross-classification by U and $C_1, \ldots, C_T$ with the current version of model parameters, conditioned on observed data $\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iT}$ for $= 1, \ldots, n$.

$$\begin{aligned} \theta_{i(s,c_1,\ldots,c_T)} &= P(U = s, \mathbf{C} = \mathbf{c}, \mid \mathbf{y}_i) \\ &= \frac{\gamma_s \prod_{t=1}^{T} \eta^{(t)}_{c_t|s} \prod_{m=1}^{M} \rho^{(t)}_{mk|c_t}}{\sum_{s=1}^{S} \gamma_s \prod_{t=1}^{T} \left[\sum_{c_t=1}^{C} \eta^{(t)}_{c_t|s} \prod_{m=1}^{M} \rho^{(t)}_{mk|c_t} \right]}. \end{aligned} \tag{3.1}$$

Note that $\theta_{i(s,c_1,\ldots,c_T)}$ can be easily calculated with the current parameter updates (i.e., The posterior probability given in (3.1) is the function of LCPA model parameters.)

In the M-step, we maximize the complete-data log-likelihood with the expectation of unobservable values computed in the E-step. We update the parameter estimates by

$$\hat{\gamma}_s = \frac{\sum_{i=1}^{n} \theta_{i(s)}}{n}, \qquad \hat{\eta}^{(t)}_{c_t|s} = \frac{\sum_{i=1}^{n} \theta^{(t)}_{i(s,c_t)}}{\sum_{i=1}^{n} \theta_{i(s)}}, \qquad \hat{\rho}_{mk|c} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \theta^{(t)}_{i(c)} I(y_{imt} = k)}{\sum_{i=1}^{n} \sum_{t=1}^{T} \theta^{(t)}_{i(c)}},$$

where $\theta_{is} = \sum_{c_1} \cdots \sum_{c_T} \theta_{i(s,c_1,\ldots,c_T)}$, $\theta^{(t)}_{i(s,c)} = \sum_{c_1} \cdots \sum_{c_{t-1}} \sum_{c_{t+1}} \cdots \sum_{c_T} \theta_{i(s,c_1,\ldots,c_T)}$ and $\theta^{(t)}_{ic} = \sum_s \theta^{(t)}_{i(s,c)}$. The relatively ease of the EM algorithm have made the ML methods popular for the LCPA, especially when the shape of the log-likelihood function is not far from quadratic. However, the likelihood function of the LCPA has some unusual features which can make likelihood-based inferences troublesome. For example, the model parameters may be estimated on the boundary of the parameter space (i.e., zero or one) in small-sample LCPA. Although item-response probabilities close to zero or one are highly desirable from a measurement perspective, the boundary solution is one of main challenges

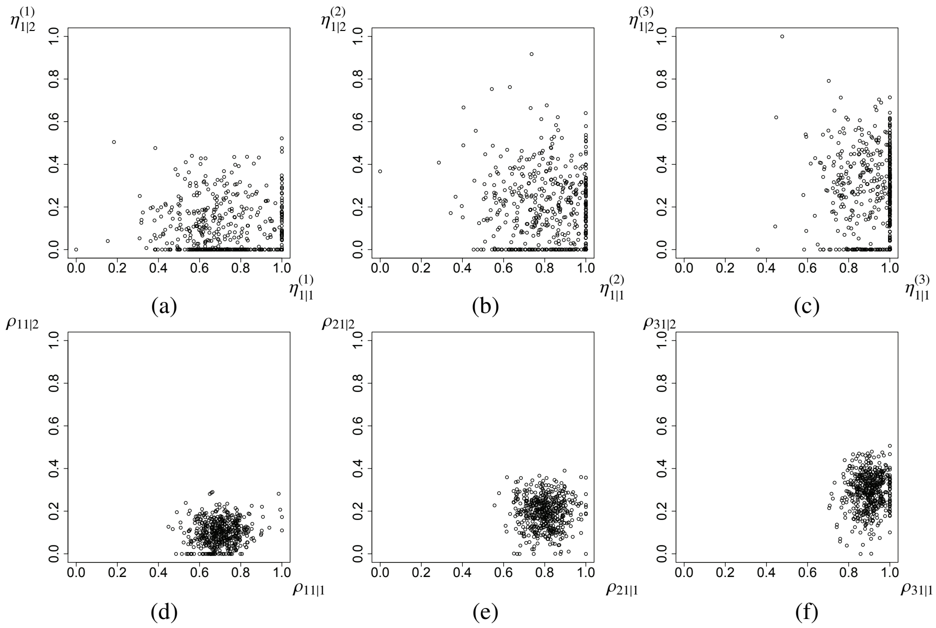

Figure 1: *Maximum-likelihood estimates for (a)* $\boldsymbol{\eta}^{(1)} = (\eta^{(1)}_{1|1},\ \eta^{(1)}_{1|2}) = (0.7,\ 0.1)$, *(b)* $\boldsymbol{\eta}^{(2)} = (\eta^{(2)}_{1|1},\ \eta^{(2)}_{1|2}) = (0.8,\ 0.2)$, *(c)* $\boldsymbol{\eta}^{(3)} = (\eta^{(3)}_{1|1},\ \eta^{(3)}_{1|2}) = (0.9,\ 0.3)$, *(d)* $\boldsymbol{\rho}_1 = (\rho_{11|1},\ \rho_{11|2}) = (0.7,\ 0.1)$, *(e)* $\boldsymbol{\rho}_2 = (\rho_{21|1},\ \rho_{21|2}) = (0.8,\ 0.2)$, *and (f)* $\boldsymbol{\rho}_3 = (\rho_{31|1},\ \rho_{31|2}) = (0.9,\ 0.3)$ *from 500 samples.*

for ML inference. When some of these parameters are estimated on the boundary, it is impossible to obtain standard errors from the inverted Hessian matrix. In such case, standard errors based on asymptotic theory might not portray uncertainty in a useful way. To illustrate, we simulated data from the two-class/two-profile LCPA with three points in time. The two latent classes were measured by three binary items. We drew 500 samples of $n = 50$ and computed ML estimates by an EM algorithm. The true parameters were $\gamma_1 = 0.5$, $\boldsymbol{\eta}^{(1)} = (\eta^{(1)}_{1|1},\ \eta^{(1)}_{1|2}) = (0.7, 0.1)$, $\boldsymbol{\eta}^{(2)} = (\eta^{(2)}_{1|1}, \eta^{(2)}_{1|2}) = (0.8, 0.2)$, $\boldsymbol{\eta}^{(3)} = (\eta^{(3)}_{1|1}, \eta^{(3)}_{1|2}) = (0.9, 0.3)$, $\boldsymbol{\rho}_1 = (\rho_{11|1}, \rho_{11|2}) = (0.7, 0.1)$, $\boldsymbol{\rho}_2 = (\rho_{21|1}, \rho_{21|2}) = (0.8, 0.2)$, and $\boldsymbol{\rho}_3 = (\rho_{31|1}, \rho_{31|2}) = (0.9, 0.3)$.

Figure 1 displays the sampling distributions of the estimates. The non-normal shape of the sampling distribution of the η-estimates shown in Figure 1(a), (b), and (c) suggests that usual large-sample approximations would be inaccurate. Approximately, 15%–25% of the η-estimates converged to the boundary solution with a criterion of 10^{-6}, and these boundary solutions cause the problem in obtaining adequate standard errors for the LCPA model with small samples. The average, root-mean squared error (RMSE), and percent coverage of the ML estimates will be given in Tables 1 and 2 to compare the performance of ML estimates to their counterparts in Section 5.

3.2. MCMC

Given the difficulties associated with likelihood-based inference, MCMC simulates random draws of model parameters for LCPA from a posterior distribution. MCMC may produce estimates and credible regions for model parameters without appealing to large-sample approximations. MCMC treats the class and profile membership of each individual as missing data, and generates the augmented posterior as if class and profile memberships were known. The MCMC is an iterative two-step procedure which can be regarded as a form of data augmentation (Tanner and Wong, 1987) or Gibbs sampling (Gelfand and Smith, 1990). In the first step of MCMC procedure, the imputation or I-step, we compute the posterior probability given in (3.1) and simulate the class and profile membership for each individual with the posterior probability. Let $z_{i(s,c_1,\ldots,c_T)}$ be a binary indicators of the i^{th} individual's class and profile memberships. If the i^{th} individual belongs to profile s and class sequence $(c_1,\ldots,c_T)$, then $z_{i(s,c_1,\ldots,c_T)}$ equals 1 and 0 otherwise. In I-step, we generate a random draw for $z_{i(s,c_1,\ldots,c_T)}$ from a multinomial distribution with the probability $\theta_{i(s,c_1,\ldots,c_T)}$ independently for $i = 1,\ldots,n$. Once the class and profile memberships are imputed, we then calculate marginal counts $n^{(t)}_{(s,c_t)} = \sum_i \sum_{c_1} \cdots \sum_{c_{t-1}} \sum_{c_{t+1}} \cdots \sum_{c_T} z_{i(s,c_1,\ldots,c_T)} = \sum_i z^{(t)}_{i(s,c_t)}$, $n^{(t)}_c = \sum_i \sum_s z^{(t)}_{i(s,c)} = \sum_i z^{(t)}_{i(c)}$, $n_s = \sum_i \sum_c z^{(t)}_{i(s,c)} = \sum_i z_{i(s)}$, and $n_{mk|c} = \sum_i \sum_t z^{(t)}_{i(c)} I(y_{imt} = k)$ for $s = 1,\ldots,S$, $c = 1,\ldots,C$, $t = 1,\ldots,T$, $m = 1,\ldots,M$, and $k = 1,\ldots,r_m$. In the second step, the posterior or P-step, we draw new random values for all parameters independently from the augmented posterior. Here, we may apply the Jeffreys prior and draw new random values for the model parameters from Dirichlet posterior distributions

$$\begin{aligned}
\gamma_1,\ldots,\gamma_S &\sim \text{Dirichlet}\left(n_1 + \frac{1}{2},\ldots,n_S + \frac{1}{2}\right),\\
\eta^{(t)}_{1|s},\ldots,\eta^{(t)}_{C|s} &\sim \text{Dirichlet}\left(n^{(t)}_{1|s} + \frac{1}{2},\ldots,n^{(t)}_{C|s} + \frac{1}{2}\right),\\
\rho_{m1|c},\ldots,\rho_{mr_m|c} &\sim \text{Dirichlet}\left(n_{m1|c} + \frac{1}{2},\ldots,n_{mr_m|c} + \frac{1}{2}\right),
\end{aligned}$$

for $s = 1,\ldots,S$, $t = 1,\ldots,T$, $m = 1,\ldots,M$, and $c = 1,\ldots,C$. This two-step procedure iteratively produces the sequences that converge to a stationary posterior distribution. The details of a Bayesian approach using MCMC algorithm for LCPA can be found in Chung and Anthony (2013).

We can implement the algorithm for a burn-in period to eliminate simulated parameters that rely on the starting values in an archetypal Bayesian approach. Averaging outcome stream of parameters after burn-in produces estimates for posterior means and variances (Tierney, 1994). Previous studies have proposed various methods of choosing the lengths of series and burn-in periods (Gelman and Rubin, 1992; Geweke, 1992; Robert, 1992), and we used time-series plots and autocorrelation functions to visually monitor the outcome stream from the MCMC and to confirm our selection of the length of series and burn-in periods.

The labeling problem becomes more acute with MCMC because the likelihood function of the LCPA has multiple equivalent modes which are invariant to permutations of class and profile labels. This invariant property may cause the most troubling aspect of MCMC for LCPA, called label switching. Because class or profile labels may permute during the MCMC run, the interpretation of the output stream of simulated parameters becomes dubious. Label switching can be particularly problematic in LCPA with small samples. To illustrate, we considered the two-class/two-profile LCPA with three points in time with true parameters given in Figure 1. Figure 2 shows the time-series plots for η parameters over the first 3,000 iterations after 2,000 burn-in iterations. Figure 2 also shows the order of η parameters reversed many times, requiring the tedious reordering of profile labels for obtaining meaningful averages for η parameters. This simulation illustrates the label-switching problem.

4. Inference for a small sample LCPA

From a purely computation standpoint, EM and MCMC are relatively easy to implement for the

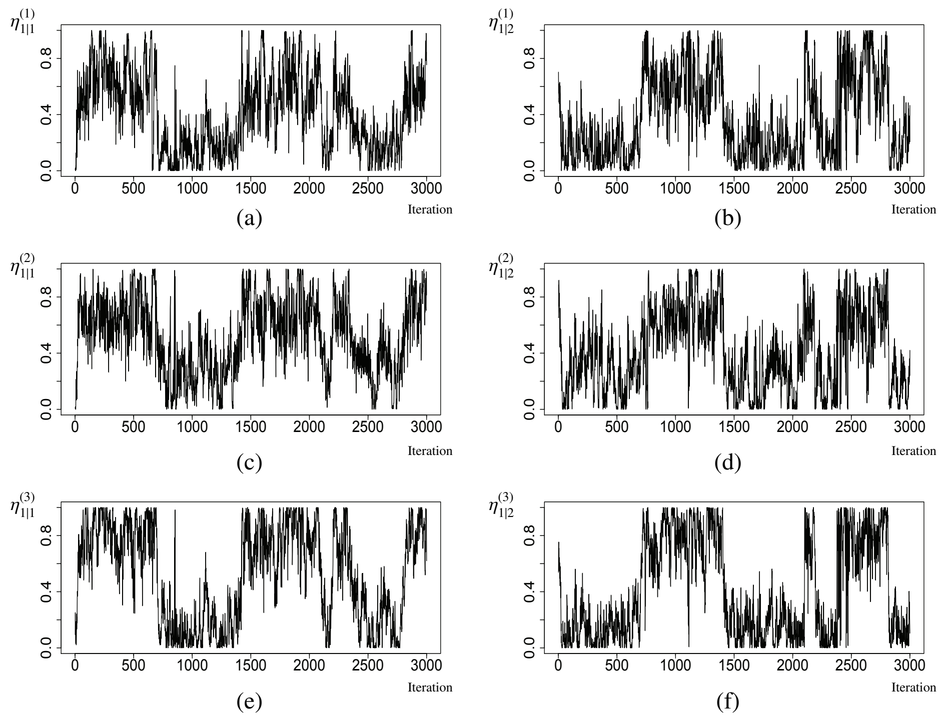

Figure 2: *Time-series plots of (a) $\eta^{(1)}_{1|1}$, (b) $\eta^{(1)}_{1|2}$, (c) $\eta^{(2)}_{1|1}$, (d) $\eta^{(2)}_{1|2}$, (e) $\eta^{(3)}_{1|1}$, and (f) $\eta^{(3)}_{1|2}$ over 3,000 iterations of MCMC.*

LCPA. However, the unusual features of LCPA (e.g., boundary solution and label switching) can make the standard ML and/or Bayesian inferences troublesome, especially for the small sample LCPA. Therefore, we propose two potential solutions to these problems in this section.

4.1. Maximum posterior estimator

The Bayesian approach was applied to the standard ML method for a single binomial variable using Jefrreys prior by Rubin and Schenker (1987). They showed that the maximum posterior estimator (MPE) performed better than the standard maximum likelihood estimator (MLE). In LCPA, it is convenient to choose priors that cause all model parameters to be *a posteriori* independent. One way to achieve this is to impose Dirichlet priors on the joint probabilities of class and profile memberships and item-response probabilities, respectively. Let the vectors of LCPA model parameters given in (2.4) be: $\boldsymbol{\Psi} = (\boldsymbol{\gamma}, \boldsymbol{\eta}^{(1)}_1, \ldots, \boldsymbol{\eta}^{(T)}_S, \boldsymbol{\rho}_{1|1}, \ldots, \boldsymbol{\rho}_{M|C})$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_S)$, $\boldsymbol{\eta}^{(t)}_s = (\eta^{(t)}_{1|s}, \ldots, \eta^{(t)}_{C|s})$ for $s = 1, \ldots, S$ and $t = 1, \ldots, T$, and $\boldsymbol{\rho}_{m|c} = (\rho_{m1|c}, \ldots, \rho_{mr_m|c})$ for $m = 1, \ldots, M$ and $c = 1 \ldots, C$. Then, the Dirichlet priors for model parameters can be given by

$$P(\boldsymbol{\Psi}) \propto P(\boldsymbol{\gamma})\, P\left(\boldsymbol{\eta}^{(1)}_1, \ldots, \boldsymbol{\eta}^{(T)}_S\right) P\left(\boldsymbol{\rho}_{1|1}, \ldots, \boldsymbol{\rho}_{M|C}\right)$$
$$= \left[\prod_{s=1}^{S} \{\gamma_s\}^{\alpha(s)}\right]\left[\prod_{s=1}^{S}\prod_{t=1}^{T}\prod_{c=1}^{C} \left\{\eta^{(t)}_{c|s}\right\}^{\beta(s,c)}\right]\left[\prod_{c=1}^{C}\prod_{m=1}^{M}\prod_{k=1}^{r_m} \{\rho_{mk|c}\}^{\delta(m,k)}\right].$$

The joint posterior for $\mathbf{\Psi}$ given $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ can be expressed as

$$
\begin{aligned}
&P(\mathbf{\Psi} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n) \\
&= P(\mathbf{\Psi}) P(\mathbf{y}_1, \ldots, \mathbf{y}_n \mid \mathbf{\Psi}) \\
&\propto \left[\prod_{s=1}^{S} \{\gamma_s\}^{\sum_i \theta_{i(s)} + \alpha_{(s)}} \right] \times \left[\prod_{s=1}^{S} \prod_{t=1}^{T} \prod_{c=1}^{C} \{\eta_{c|s}\}^{\sum_i \theta^{(t)}_{i(s,c)} + \beta_{(s,c)}} \right] \times \left[\prod_{t=1}^{T} \prod_{c=1}^{C} \prod_{m=1}^{M} \prod_{k=1}^{r_m} \{\rho_{mk|c}\}^{\sum_i \left(\theta^{(t)}_{i(c)} I(y_{imt}=k)\right) + \delta(m,k)} \right], \quad (4.1)
\end{aligned}
$$

where $\alpha_{(s)}$, $\beta_{(s,c)}$ and $\delta_{(m,k)}$ are the hyper-parameters. The updated parameters for MPE maximizing (4.1) are obtained by

$$
\hat{\gamma}_s = \frac{\sum_{i=1}^{n} \theta_{is} + \alpha_{(s)}}{n + \alpha}, \quad \hat{\eta}^{(t)}_{c|s} = \frac{\sum_{i=1}^{n} \theta^{(t)}_{i(s,c)} + \beta_{(s,c)}}{\sum_{i=1}^{n} \theta_{i(s)} + \beta_{(s)}}, \quad \hat{\rho}_{mk|c} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \theta^{(t)}_{ic_t} I(y_{imt} = k) + \delta_{(m,k)}}{\sum_{i=1}^{n} \sum_{t=1}^{T} \theta^{(t)}_{ic_t} + \delta_{(m)}}, \quad (4.2)
$$

where $\alpha = \sum_{s=1}^{S} \alpha_{(s)}$, $\beta_{(s)} = \sum_{c=1}^{C} \beta_{(s,c)}$, and $\delta_{(m)} = \sum_{k=1}^{r_m} \delta_{(m,k)}$. The effect of hyper-parameters $\alpha_{(1)}, \ldots, \alpha_{(S)}$ for γ is to smooth the parameter estimates toward profile sizes. For example, the use of constant hyper-parameters $\omega = \alpha_{(1)} = \cdots = \alpha_{(S)}$ is equivalent to adding the equivalent ω observations to each of profiles. For η, the hyper-parameters $\beta_{(s,1)}, \ldots, \beta_{(s,C)}$ has a flattening effect on the elements of $(\eta^{(t)}_{1|s}, \ldots, \eta^{(t)}_{C|s})$ by adding the fictitious $\beta_{(s)}$ observations to each of classes at time t for $s = 1, \ldots, S$. The hyper-parameters $\delta_{(m,k)}$ could possibly depend on the data. For example, we can select $\delta_{(m,k)} \propto \sum_i \sum_t I(y_{imt} = k)/Tn$ so that the prior distribution smoothes ρ-parameter toward an ML estimate of the raw distribution of the m^{th} item. These hyper-parameters will alleviate the boundary solution problem for the ML method.

4.2. Bayesian estimator with dynamic data-dependent prior

An efficient strategy has been proposed to break the symmetry of the posterior distribution for addressing the label-switching problem in the exponential finite-mixture models (Chung *et al.*, 2004). They pre-classified one (or more) observations (observations with minimum or maximum value) into some mixture components and demonstrated that this method performed well over repeated samples. However, LCPA faces the challenge of determining which individuals to pre-classify for the class and profile memberships. Substantive knowledge can often inform this decision; however, some individuals may be particularly informative for blind selection of the pre-classification. For example, let us consider an LCPA with four profiles with an individual who has posterior probabilities $(\theta_{i(1)}, \theta_{i(2)}, \theta_{i(3)}, \theta_{i(4)}) = (0.01, 0.01, 0.97, 0.01)$ at one iteration during MCMC. The posterior implies that this individual has a 97% chance of belonging to the third profile given their item responses. Therefore, it is most likely that this individual would be imputed as the third profile and not change his or her profile membership across iterations. If the imputed profile membership for this individual changes between iterations, it would be a very strong indication that label switching has occurred. A promising strategy is to assign a particular individual with a high posterior probability to a specific profile and classes over time. A natural criterion is the posterior probability calculated on the basis of the running posterior mean of model parameters $\mathbf{\Psi}$. Here, we propose an automated algorithm using the dynamically adapted priors to select individuals to be assigned to classes and profiles simultaneously.

Let $\mathbf{\Psi}^{(j)}$ be the generated model parameters from the LCPA posterior distribution at the j^{th} MCMC iteration. Furthermore, let $\bar{\mathbf{\Psi}}^{(N)} = N^{-1} \sum_{j=1}^{N} \mathbf{\Psi}^{(j)}$ represent the posterior mean of the model parameters at the N^{th} iteration. We can consider the cumulative posterior probability which is a function of the cumulative posterior mean of model parameters at the N^{th} iteration:

$$\theta_{i(s,c_1,\ldots,c_T)}\left(\bar{\mathbf{\Psi}}^{(N)}\right) = \frac{\bar{\gamma}_s^{(N)} \prod_{t=1}^{T} \bar{\eta}_{c_t|s}^{(t)^{(N)}} \prod_{m=1}^{M} \prod_{k=1}^{r_m} \left\{\bar{\rho}_{mk|c_t}^{(N)}\right\}^{I(y_{imt}=k)}}{\sum_{s=1}^{S} \bar{\gamma}_s^{(N)} \prod_{t=1}^{T} \left[\sum_{c_t=1}^{C} \bar{\eta}_{c_t|s}^{(t)^{(N)}} \prod_{m=1}^{M} \prod_{k=1}^{r_m} \left\{\bar{\rho}_{mk|c_t}^{(N)}\right\}^{I(y_{imt}=k)}\right]}. \tag{4.3}$$

We can select some individuals who have "large" values for the profile s and class sequence $c_1, \ldots, c_t$ and assign them to the respective profile and classes with certainty at the N^{th} iteration. At the N^{th} iteration, for example, we can identify the subject with the largest value of the cumulative posterior probability given in (4.3) by evaluating $\theta_{1(s,c_1,\ldots,c_T)}(\bar{\mathbf{\Psi}}^{(N)}), \ldots, \theta_{n(s,c_1,\ldots,c_T)}(\bar{\mathbf{\Psi}}^{(N)})$. We then pre-classify this individual into respective combination of profile s and classes $c_1, \ldots, c_T$ at the N^{th} iteration. By repeating this procedure independently for all combinations of profile and classes, we can choose $S \times C^T$ subjects to be identified. Pre-classifying a small number of subjects into each combination of profile and class memberships may be sufficient to break symmetry without introducing serious subjectivity if the profile and class memberships are well discriminated. The standard I-step can be easily modified to adopt this procedure: we could decisively put $z_{i(s,c_1,\ldots,c_T)} = 1$ at the N^{th} iteration when the i^{th} subject is selected for profile s and class sequence $c_1, \ldots, c_T$ at that iteration of MCMC. This asymmetric prior to the LCPA likelihood function tends to dampen the nuisance posterior mode while having a small effect on the posterior mode of interest. In situations where some profiles and/or classes are not well differentiated, pre-classifying more individuals might be preferable. For the previous example, the time-series plots for η parameters over the first 3,000 iterations after burn-in 2,000 iterations showed a label switching problem in Figure 2. In contrast, Figure 3 shows that the label switching are disappeared when two individuals are pre-classified into each profile and class.

5. Simulation study

The simulation study is conducted to investigate the performance of the ML, maximum posterior (MP), and Bayesian methods using a dynamic data-dependent prior over repeated samples. In this study, we draw 500 samples with $n = 50$ observations each from two-class (i.e., $C = 2$) and two-profile (i.e., $S = 2$) LCPA with three binary items (i.e., $M = 3$) measured over three time points (i.e., $T = 3$) with parameters for the previous example setting (i.e., $\gamma_1 = 0.5$, $\boldsymbol{\eta}^{(1)} = (\eta_{1|1}^{(1)}, \eta_{1|2}^{(1)}) = (0.7, 0.1)$, $\boldsymbol{\eta}^{(2)} = (\eta_{1|1}^{(2)}, \eta_{1|2}^{(2)}) = (0.8, 0.2)$, $\boldsymbol{\eta}^{(3)} = (\eta_{1|1}^{(3)}, \eta_{1|2}^{(3)}) = (0.9, 0.3)$, $\boldsymbol{\rho}_1 = (\rho_{11|1}, \rho_{11|2}) = (0.7, 0.1)$, $\boldsymbol{\rho}_2 = (\rho_{21|1}, \rho_{21|2}) = (0.8, 0.2)$, $\boldsymbol{\rho}_3 = (\rho_{31|1}, \rho_{31|2}) = (0.9, 0.3)$). For each sample, we estimate model parameters by four different methods: the standard ML method using the EM algorithm, MP method with hyper-parameters defined in (4.1) and (4.2), MCMC method with two and four subjects assigned dynamically across iterations (DYN-2 and DYN-4, respectively) by dynamic data dependent prior. For hyper-parameters in MP, we chose Jeffreys prior, so all hyper-parameters are to be set 1/2 for all parameters. For interval estimates, we compute and invert the Hessian of the log-likelihood (for ML) and posterior log-likelihood (for MP). However, the interval estimates may stray outside the unit interval with the boundary solutions. To solve this problem, we generate the bootstrapping sampling to obtain standard errors for the model parameters, and then apply the normal approximation on the logistic scale (Goodman, 1974). For MCMC, we run 2,000 iterations after 1,000 burn-in periods with two and four subjects assigned to each profile and class at each point in time.

Figure 4 shows the distributions of estimates for η-parameters over the 500 samples from ML, MP, DYN-2, and DYN-4. Comparing distributions of ML estimates (a) in with those of the MP (b), DYN-2 (c), and DYN-4 (d) in Figure 1, we see that MP and Bayesian methods using data-dependent

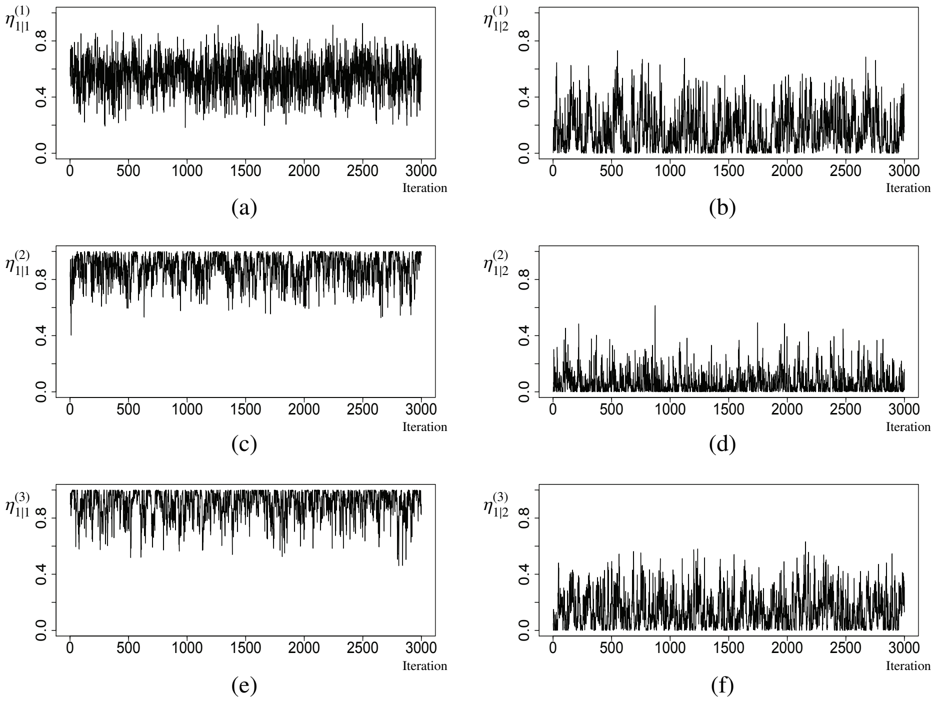

Figure 3: *Time-series plots of (a) $\eta^{(1)}_{1|1}$, (b) $\eta^{(1)}_{1|2}$, (c) $\eta^{(2)}_{1|1}$, (d) $\eta^{(2)}_{1|2}$, (e) $\eta^{(3)}_{1|1}$, and (f) $\eta^{(3)}_{1|2}$ over 3,000 iterations of MCMC with dynamic data-dependent prior.*

prior achieved obvious improvement over ML, especially in η-parameters: MP and Bayesian methods did not converge to the boundary solution.

Table 1 provides the average and RMSE of the estimates of ML, MP, DYN-2, and DYN-4. All methods except ML performed well. Under ML, RMSE values are consistently larger than other methods, especially in the estimates for η-parameters. Table 2 provides the performance of interval estimates and shows the percentage of intervals that covered targets as well as average interval width. Narrower intervals are desirable provided that coverage remains at or above the nominal rate of 95%. ML has poor coverage with wide intervals, comparing with MP, DYN-2, and DYN-4. The intervals from ML for γ and η-parameters have extremely poor coverage, but the intervals are wide. The ML has better coverage for ρ-parameters than γ and η parameters, but MP, DYN-2, and DYN-4 still have higher coverage than ML with narrower intervals for ρ-parameters. Among Bayesian method, MP is conservative in ρ-parameters, showing higher than nominal rates of coverage for ρ-parameters from DYN-2 and DYN-4. The lengths of most of MP intervals for ρ-parameters are in the middle of those from DYN-2 and DYN-4, yet its rates of coverage are higher than those of DYN-2 and DYN-4.

6. An application to adolescent substance use data

Our focus in this section is to apply our approach to a case example. In order to describe the difficulties in inference with small samples in LCPA, we draw data from the National Longitudinal Survey

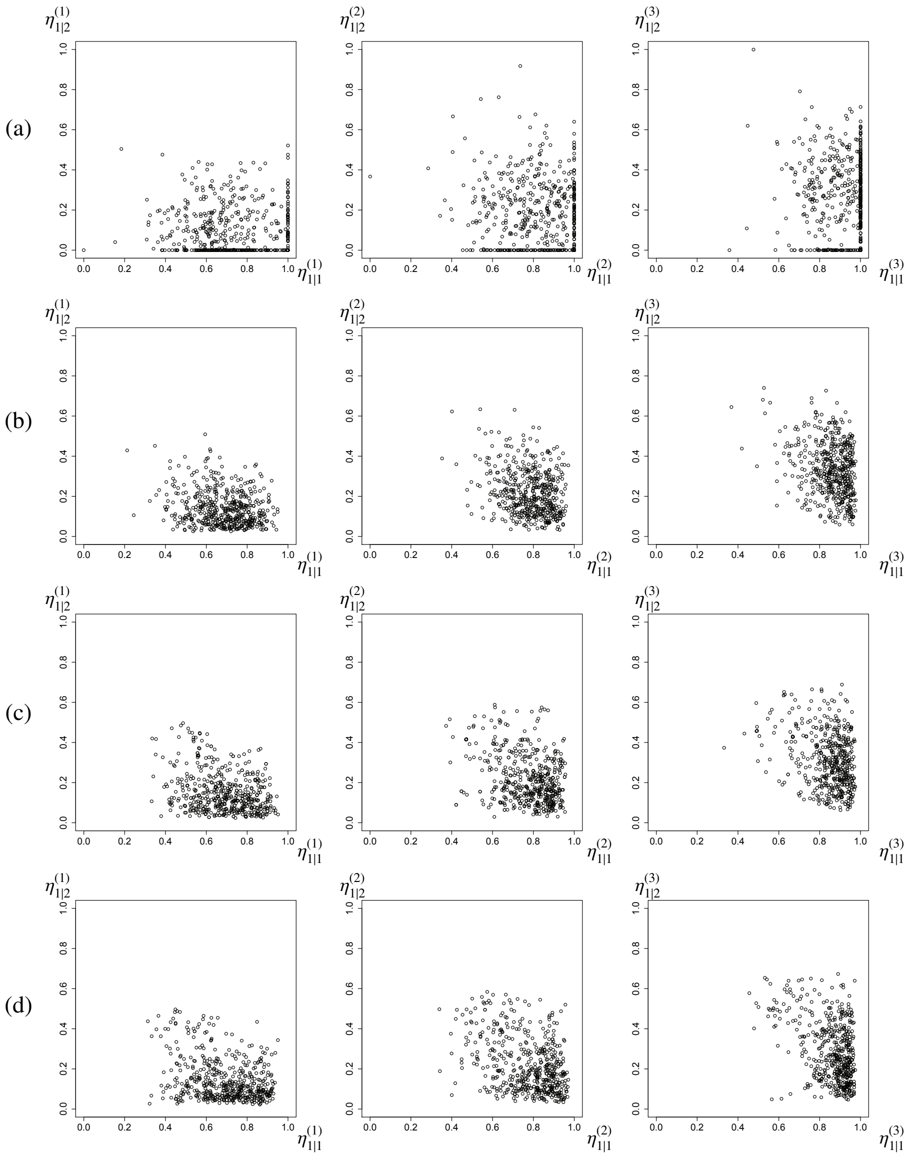

Figure 4: *Distribution of point estimates for η parameters over 500 samples obtained by (a) ML, (b) MP, (c) DYN-2, and (d) DYN-4.*

of Youth 1997 survey that explores the transition from school to work and from adolescence to adulthood in the USA (http://www.bls.gov/nls/nlsy97.htm). The sample considered in this study includes adolescents aged 12–14 years in 1997. Their alcohol drinking behavior were tracked over the three survey years in 1997, 2000, and 2003. In the first survey year, 1997, adolescents were asked if they had ever drunk alcohol. The 14–16 adolescents answered ‘yes’, and they are identified as the early onset drinkers who are of interest for this study. Three self-report items measured drinking behavior in these early onset drinkers: (a) how many days they had had one or more drinks of an alcoholic

Table 1: Average (RMSE) point estimates over 500 repetitions

	True	ML	MP	DYN-2	DYN-4		True	ML	MP	DYN-2	DYN-4
γ_1	0.5	0.502 (0.218)	0.498 (0.128)	0.498 (0.142)	0.506 (0.152)						
$\eta^{(1)}_{1\mid 1}$	0.7	0.731 (0.270)	0.684 (0.193)	0.687 (0.137)	0.691 (0.145)	$\rho_{11\mid 1}$	0.7	0.698 (0.210)	0.704 (0.182)	0.698 (0.142)	0.701 (0.141)
$\eta^{(2)}_{1\mid 1}$	0.8	0.821 (0.266)	0.786 (0.166)	0.777 (0.124)	0.773 (0.139)	$\rho_{21\mid 1}$	0.8	0.801 (0.180)	0.798 (0.154)	0.793 (0.120)	0.795 (0.121)
$\eta^{(3)}_{1\mid 1}$	0.9	0 .913 (0.164)	0.857 (0.144)	0.848 (0.116)	0.860 (0.110)	$\rho_{31\mid 1}$	0.9	0.900 (0.143)	0.898 (0.114)	0.895 (0.088)	0.894 (0.090)
$\eta^{(1)}_{1\mid 2}$	0.1	0.099 (0.167)	0.140 (0.145)	0.151 (0.113)	0.146 (0.115)	$\rho_{11\mid 2}$	0.1	0.098 (0.142)	0.104 (0.116)	0.108 (0.086)	0.105 (0.092)
$\eta^{(2)}_{1\mid 2}$	0.2	0.189 (0.255)	0.218 (0.170)	0.217 (0.123)	0.220 (0.134)	$\rho_{21\mid 2}$	0.2	0.194 (0.182)	0.201 (0.153)	0.205 (0.125)	0.206 (0.121)
$\eta^{(3)}_{1\mid 2}$	0.3	0.263 (0.271)	0.332 (0.195)	0.309 (0.137)	0.286 (0.150)	$\rho_{31\mid 2}$	0.3	0.293 (0.211)	0.294 (0.181)	0.294 (0.144)	0.292 (0.140)

RMSE = root-mean squared error, ML = maximum likelihood, MP = maximum posterior, DYN-2 = two subjects across iterations, DYN-4 = four subjects across iterations.

Table 2: Percent coverage (average width) of nominal 95% interval estimates over 500 repetitions

	ML	MP	DYN-2	DYN-4		ML	MP	DYN-2	DYN-4
γ_1	34.4 (0.605)	83.2 (0.357)	87.0 (0.430)	90.0 (0.447)					
$\eta^{(1)}_{1\mid 1}$	58.6 (0.633)	81.6 (0.516)	92.4 (0.493)	93.0 (0.515)	$\rho_{11\mid 1}$	77.6 (0.333)	92.0 (0.289)	87.6 (0.257)	92.2 (0.291)
$\eta^{(2)}_{1\mid 1}$	49.8 (0.491)	89.2 (0.439)	93.2 (0.431)	92.0 (0.456)	$\rho_{21\mid 1}$	69.0 (0.284)	93.3 (0.244)	88.4 (0.218)	89.8 (0.247)
$\eta^{(3)}_{1\mid 1}$	45.8 (0.318)	81.6 (0.321)	95.0 (0.359)	95.2 (0.375)	$\rho_{31\mid 1}$	71.8 (0.211)	95.0 (0.179)	90.4 (0.166)	93.6 (0.139)
$\eta^{(1)}_{1\mid 2}$	46.0 (0.326)	80.0 (0.321)	94.0 (0.353)	95.2 (0.383)	$\rho_{11\mid 2}$	68.2 (0.210)	94.8 (0.183)	89.4 (0.165)	92.8 (0.185)
$\eta^{(2)}_{1\mid 2}$	51.4 (0.483)	86.8 (0.447)	91.6 (0.423)	93.6 (0.455)	$\rho_{21\mid 2}$	70.0 (0.290)	93.8 (0.240)	84.6 (0.217)	91.6 (0.250)
$\eta^{(3)}_{1\mid 2}$	55.6 (0.636)	79.2 (0.529)	91.4 (0.489)	91.0 (0.506)	$\rho_{31\mid 2}$	72.2 (0.332)	92.4 (0.286)	85.6 (0.254)	92.0 (0.290)

ML = maximum likelihood, MP = maximum posterior, DYN-2 = two subjects across iterations, DYN-4 = four subjects across iterations.

beverage during the last 30 days (*Recent drinking*), (b) how many days they had had five or more drinks on the same occasion during the past 30 days (*Binge drinking*) and (c) how many days they drank immediately before or during school or work hours in the last 30 days (*Drinking at school*). The responses for *Recent drinking*, ranging from 0 to 30 days, were reduced to a three-category indicator: non-drinker (0 days of drinking), occasional drinker (15 days of drinking) or regular drinker (6 or more days of drinking). For *Binge drinking*, respondents who had consumed five or more drinks on the same occasion at least one time in the last 30 days were characterized as binge drinkers. A binary indicator was created for *Drinking at school*: respondents who had consumed alcoholic beverages immediately before or during school or work hours at least once in the last 30 days were placed in a drinking-at-school group, whereas respondents without such drinking were put into a second, non-school-drinking group.

The model parameters of LCPA are estimated based on the model presented by Chung *et al.* (2013) (i.e., an LCPA is specified to include four classes of drinking behavior and three class profiles) without any covariate. We randomly select 200 early onset drinkers among 14–16 adolescents. The

Table 3: Estimated response probabilities to the drinking items for each class (i.e. ρ-parameters) under the four-class/three-profile LCPA from ML, MP, DYN-2, and DYN-4

Method	Class	Probabilities for the following drinking items			
		Recent drinking		Binge drinking	Drinking at school
		Occasional	Regular		
ML	Not current drinkers	0.111	0.000	0.000	0.000
	Light drinkers	0.921	0.079	0.325	0.144
	Occasional binge drinkers	0.666	0.334	0.759	0.458
	Regular binge drinkers	0.353	0.647	0.976	0.135
MP	Not current drinkers	0.054	0.007	0.002	0.002
	Light drinkers	0.916	0.060	0.072	0.122
	Occasional binge drinkers	0.774	0.216	0.819	0.381
	Regular binge drinkers	0.380	0.616	0.965	0.127
DYN-2	Not current	0.025	0.008	0.002	0.002
	Light drinkers	0.942	0.042	0.103	0.099
	Occasional binge drinkers	0.780	0.208	0.771	0.441
	Regular binge drinkers	0.362	0.630	0.979	0.090
DYN-4	Not current drinkers	0.004	0.003	0.002	0.002
	Light drinkers	0.910	0.082	0.018	0.088
	Occasional binge drinkers	0.846	0.148	0.806	0.290
	Regular binge drinkers	0.239	0.757	0.995	0.135

LCPA = latent class profile analysis, ML = maximum likelihood, MP = maximum posterior, DYN-2 = two subjects across iterations, DYN-4 = four subjects across iterations.

point estimates for different estimation methods are reported in Table 3. The probabilities in Table 3 (i.e., ρ-parameters) are the probabilities of a 'yes' response probability for the three drinking items. The label of the four classes can be assigned based on the pattern of these probabilities. The estimates of the four different methods are nearly identical except for *Binge drinking* in 'Light drinkers.'

Table 4 presents the estimated conditional probabilities of class membership at each age group for a particular class profile (i.e., η-parameters) and the marginal probabilities of class profiles (i.e., γ-parameters) from the four different estimation methods. These four methods provide the similar estimated of probabilities in η and γ parameters except in 'Light drinking advancers': there exists a difference in estimated probabilities for 'Light drinkers' class probabilities over three different age groups.

The standard errors for the estimates calculated by ML and MP are not available because the Hessian matrix cannot be inverted. The Bayesian methods, however, have no difficulty in producing interval estimates (not provided here). The standard Bayesian method have label switching problem during an MCMC run, making the output difficult to interpret. Thus, we made attempt to reduce the label switching problem by using data-dependent prior information. We conduct 3,000 iterations after a burn-in period of 1,000 cycles by using a Jeffreys prior. Table 5 displays a summary of the assigned individuals' cumulative posterior probabilities. Here, we construct three sets of estimates that are calculated by Bayesian approach methods using data-dependent prior information: pre-assigning one subject across iterations (DYN-1), two subjects across iterations (DYN-2), and four subjects across iterations (DYN-4). For example, during 3,000 iterations of DYN-1, the average value of the cumulative posterior probabilities for the assigned subjects is close to one (i.e., the range of the cumulative posterior probabilities is $[0.591, 0.997]$), and therefore assigning one subject at each iteration has little impact on the posterior distribution. However, label switching still occurs because the data-dependent prior given in DYN-1 is not sufficient to break the symmetry. By pre-classifying more subjects, the posterior distribution tends to dampen the nuisance mode, but we are introducing more subjectivity (i.e., the range of the cumulative posterior probabilities is wider).

Table 4: Estimated probabilities of belonging to a specific class at a certain time point for each profile and estimated profile prevalence under the four-class/three-profile LCPA from ML, MP, DYN-2, and DYN-4

Method	Profile	Class	Probabilities for the following age groups		
			12–14 years	15–17 years	18–20 years
ML	Non-drinking stayers (44.4%)	Not current drinkers	0.842	1.000	0.556
		Light drinkers	0.024	0.000	0.327
		Occasional binge drinkers	0.134	0.000	0.000
		Regular binge drinkers	0.000	0.000	0.117
	Light drinking advancers (17.5%)	Not current drinkers	0.670	0.000	0.000
		Light drinkers	0.234	1.000	0.689
		Occasional binge drinkers	0.000	0.000	0.000
		Regular binge drinkers	0.096	0.000	0.311
	Regular binge advancers (38.1%)	Not current drinkers	0.653	0.000	0.197
		Light drinkers	0.092	0.000	0.071
		Occasional binge drinkers	0.255	0.122	0.000
		Regular binge drinkers	0.000	0.878	0.731
MP	Non-drinking stayers (42.1%)	Not current drinkers	0.796	0.876	0.546
		Light drinkers	0.082	0.049	0.223
		Occasional binge drinkers	0.074	0.031	0.130
		Regular binge drinkers	0.047	0.044	0.102
	Light drinking advancers (16.1%)	Not current drinkers	0.504	0.141	0.096
		Light drinkers	0.338	0.534	0.690
		Occasional binge drinkers	0.042	0.057	0.091
		Regular binge drinkers	0.080	0.268	0.123
	Regular binge advancers (41.9%)	Not current drinkers	0.630	0.044	0.125
		Light drinkers	0.071	0.129	0.025
		Occasional binge drinkers	0.280	0.158	0.034
		Regular binge drinkers	0.019	0.669	0.816
DYN-2	Non-drinking stayers (40.9%)	Not current drinkers	0.771	0.891	0.522
		Light drinkers	0.096	0.071	0.233
		Occasional binge drinkers	0.091	0.021	0.140
		Regular binge drinkers	0.042	0.017	0.105
	Light drinking advancers (18.0%)	Not current drinkers	0.567	0.070	0.059
		Light drinkers	0.332	0.571	0.771
		Occasional binge drinkers	0.020	0.039	0.066
		Regular binge drinkers	0.081	0.320	0.104
	Regular binge advancers (41.1%)	Not current drinkers	0.599	0.016	0.145
		Light drinkers	0.056	0.106	0.013
		Occasional binge drinkers	0.332	0.258	0.021
		Regular binge drinkers	0.012	0.620	0.822
DYN-4	Non-drinking stayers (44.4%)	Not current drinkers	0.708	0.787	0.538
		Light drinkers	0.155	0.087	0.216
		Occasional binge drinkers	0.131	0.118	0.234
		Regular binge drinkers	0.007	0.008	0.011
	Light drinking advancers (15.5%)	Not current drinkers	0.601	0.026	0.017
		Light drinkers	0.301	0.649	0.779
		Occasional binge drinkers	0.019	0.033	0.152
		Regular binge drinkers	0.080	0.291	0.052
	Regular binge advancers (40.1%)	Not current drinkers	0.608	0.086	0.107
		Light drinkers	0.050	0.016	0.007
		Occasional binge drinkers	0.327	0.373	0.047
		Regular binge drinkers	0.016	0.525	0.838

LCPA = latent class profile analysis, ML = maximum likelihood, MP = maximum posterior, DYN-2 = two subjects across iterations, DYN-4 = four subjects across iterations.

Table 5: Summary of the cumulative posterior probabilities for the assigned subjects

	Modified MCMC		
	DYN-1	DYN-2	DYN-4
Minimum	0.591	0.516	0.533
Maximum	0.997	0.982	0.976

MCMC = Markov chain Monte Carlo, DYN-1 = one subject across iterations, DYN-2 = two subjects across iterations, DYN-4 = four subjects across iterations.

In application study, we applied an LCPA in an investigation of stage sequential patterns of drinking behavior among early onset drinkers, using data from the NLSY97. We found that an LCPA could identify four different patterns of drinking behavior at each measurement occasion; and the sequence of early-onset drinkers' drinking behavior could be summarized by three profiles. LCPA uncovered four common drinking behavior in early onset drinkers over three measurements from early to late adolescence, and the sequences of drinking behaviors were grouped into three sequential patterns representing the most probable progression of early onset drinking behavior.

7. Discussion

Many behavioral scientists try to classify individuals into a small number of groups based on their item-response pattern. The popularity of LCPA is increasing because LCPA can identify subtypes of sequential patterns of discrete latent classes. This research has explored the inferential problems of the LCPA model with small samples. We showed that MP method and Bayesian inference via MCMC may be attractive alternatives to the standard ML method. The simulation study indicated that label switching problem emerged with MCMC and suggested that the problems could be alleviated by proposed strategies using prior information.

The proposed method using the MCMC with dynamic data-dependent prior may have some limitations. It assigns pre-selected observations to different classes and profiles with certainty. The selection of the pre-assigned individuals depends on the posterior probabilities of the class and profile membership given the parameters; subsequently, the amount of impact it has on the mode of interest depends on these probabilities. However, it would be more adversely affected than the mode of interest if the nuisance mode is present. More research is required to investigate the limitations with the pre-classifying technique across LCPA models with different distributions in item response probabilities and larger numbers of classes and profiles.

The LCPA is one of the most useful in drug abuse intervention research when they lead to models that provide an accurate representation of the data. However, model selection is a difficult challenge facing LCPA users. Model selection encompasses selecting the number of latent classes needed to reflect heterogeneity in the data, testing whether assumptions about the model are valid, and assessing if parameter restrictions are reasonable. The first and most crucial step in LCPA is to choose an appropriate number of classes and profiles since model selection has important ramifications for analysis performed with the model. Substantive researchers have several methods at their disposal to evaluate LTA fit depending on the software package used. In addition, a variety of relevant new methods has been developed to assess model fit for finite mixture models. For example, the reversible jump MCMC (RJMCMC) and the Bayesian nonparametric approach are proposed to provide a set of principles for the systematic model selection of LCPA (Chung and Chang, 2012) They explored these two methodologies (i.e., RJMCMC and Dirichlet process) to select the number of latent LCPA components and their performances were empirically evaluated via a simulation study. However, future work should investigate more tailored methodologies specialized for LCPA with small samples.

Acknowledgements

This work was supported by a grant award NRF-2015R1D1A1A01056846 (Hwan Chung) from the National Research Foundation of Korea.

References

Chung H and Anthony JC (2013). A Bayesian approach to a multiple-group latent class-profile analysis: the timing of drinking onset and subsequent drinking behaviors among US adolescents, *Structural Equation Modeling: A Multidisciplinary Journal*, **20**, 658–680.

Chung H, Anthony JC, and Schafer JL (2011). Latent class profile analysis: an application to stage sequential process of early onset drinking behaviours, *Journal of the Royal Statistical Society Series A*, **174**, 689–712.

Chung H and Chang HC (2012). Bayesian approaches to the model selection problem in the analysis of latent stage-sequential process, *Computational Statistics and Data Analysis*, **56**, 4097–4110.

Chung H, Loken E, and Schafer JL (2004). Difficulties in drawing inferences with finite-mixture models: a simple example with a simple solution, *The American Statistician*, **58**, 152–158.

Dempster AP, Laird NM, and Rubin DB (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society Series B*, **39**, 1–38.

Gelfand AE and Smith AFM (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association*, **85**, 398–409.

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

Geweke J (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments, *Bayesian Statistics*, **4**, 169–193,

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

Lazarsfeld PF and Henry NW (1968). *Latent Structure Analysis*, Houghton Mifflin, Boston.

Robert GO (1992). Convergence diagnostics of the Gibbs sampler, *Bayesian Statistics*, **4**, 775–782.

Rubin DB and Schenker N (1987). Logit-based interval estimation for binomial data, *Sociological Methodology*, **17**, 131–144.

Tanner WA and Wong WH (1987). The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528–550.

Tierney L (1994). Markov chains for exploring posterior distributions (with discussion), *Annals of Statistics*, **22**, 1701–1762.

Received November 19, 2015; Revised January 13, 2016; Accepted February 28, 2016