Semiparametric Mixture Modeling for Skewed-longitudinal Data: A Bayesian Approach

Yangxin Huang1,2*, Huahai Qiu1 and Chunning Yan3

1Department of Mathematics, Wuhan Textile University, China
2Department of Epidemiology and Biostatistics, University of South Florida, USA
3School of Management, Shanghai University, China

Abstract

Longitudinal data arise frequently in medical studies and it often happens in longitudinal studies that collected data are observed with the following typical features. (i) Individuals may follow a heterogeneous population (rather than homogeneous population) with more than one mean trajectories; (ii) longitudinal outcomes may suffer from a serious departure of normality in which normality assumption may cause lack of robustness and subsequently lead to invalid inference; and (iii) the response observations may be subject to left censoring due to a limit of detection (LOD). Inferential procedures will become very complicated when one analyzes data with these features together. Recently, a mixture of skew normal distribution has received increasing attention in the treatment of heterogeneous data involving asymmetric behaviors across subclasses, but there are relatively few studies accommodating heterogeneity, non normality and censoring simultaneously arose in longitudinal data setting. This article explores Bayesian semiparametric mixture models of longitudinal measures with an attempt to mediate heterogeneous population, to overcome departures from normality and to treat below LOD as left censoring in the longitudinal response. A real data example is analyzed to demonstrate the proposed methodology for estimating not only model parameters but also class membership probabilities through various scenarios to compare potential models with different specifications of error distributions.

INTRODUCTION

Finite mixture of symmetric (normal) models has been found to be one of the most popular model based approaches to dealing with data in the presence of population heterogeneity in the sense that data intrinsically consist of unlabelled observations, each of which is thought to belong to one of K classes (or components). For a comprehensive survey of mixture models and their applications, the readers are referred to the monographs [1,2] and among others. In recent years, mixtures of asymmetric distributions based models have emerged as a powerful alternative to the traditional normal and t-mixture models [3-5]. However, much of the statistical literature has focused on the development of mixture models for independent data [3-9], but there are relatively few studies exploring modeling longitudinal data [10,11].

Modeling of longitudinal data is an active area of biostatistics and statistics research that has received increasing attention in recent years. A large number of statistical modeling and analysis methods have been suggested for analyzing such data with various features. Researchers may often confront the task of developing inference from samples where longitudinal outcomes for the dependent variable of interest may follow heterogeneous (not homogeneous) characteristics, suffer from a serious departure of normality and be subject to left censoring due to a limit of detection (LOD). For example, in AIDS studies, the infection of human immunodeficiency virus type 1 (HIV-1) is usually assessed by the number of copies of HIV-1 RNA (viral load) in blood plasma, and the change in viral load is an important indicator of HIV disease progression after an application of antiretroviral treatment (ART) [12,13,14]. Modeling such data has many challenges due to the following inherent features of longitudinal data.

In the previous longitudinal modeling, most studies assume that all subjects come from a homogeneous population where large between and within individual variations were
accommodated by random effects and/or time-varying covariates in the models. These typical large between and within individual variations are shown in (Figure 1), viral load trajectory profiles of 6 representative patients in AIDS Clinical Trials Group study 398 (ACTG398) [12] (see Section 2 for details of this study and data), which can be roughly classified into two classes with biological and clinical interpretation: (i) decrease rapidly at the beginning in a short term period and/or then stay stable at a low level later (solid lines); and (ii) decrease at the beginning, but rebound later (dotted lines). We, therefore, can reasonably assume that patients are from a population which consists of two relatively homogeneous classes. Thus, it is motivated to consider a finite mixture of hierarchical models for such data set. By relaxing the homogeneous population assumption, finite mixture models allow for parameter differences across several unobserved classes within a heterogeneous population.

Another common assumption in most studies is that model random errors are normally (symmetrically) distributed, but this assumption may lack robustness against departures from normality and/or outliers and, thus, statistical inference and analysis with normal assumption may lead to misleading results [15,16,17]. For instance, the repeated viral load measurements (in $R$ scale) for 379 subjects enrolled in ACTG398 [12] to be analyzed in this article appear to present a skewed data feature and, thus, a skew-normal (SN) distribution [15-22] should be more appropriate to model this data set. Huang and Dagne [15,16] investigated the biases induced by data symmetry assumption for longitudinal data with skewness, but it is not clear how data with both heterogenous and asymmetric features may influence inferential procedures within a framework of mixture modeling for longitudinal data. It is noted that finite mixture modeling with skew distributions for random error has been investigated in the literature [4, 5], but these publications focused mainly on analyzing independent data. Furthermore, the validity of inference methods relies on an important requirement that variables are “accurately” measured. In practice, however, collected data are often far from “accuracy.” Although longitudinal studies are designed to collect data from every participant in the study at each assessment time, the repeated outcomes may be subject to LOD because of the low sensitivity of current standard assays. For instance, the ACTG398 study was designed to collect data on every individual at each assessment, the response (viral load) measurements may be subject to left censoring due to an LOD. It can be seen from (Figure 1) that for some patients their viral loads are below LOD (50 copies/mL). The proportion of data censored may not be trivial, so failure to account for censoring in data analysis may result in significant biases in the estimates of the fixed-effects and variance components [23]. Ad hoc procedures are sometimes used to adjust for the censoring. For instance, when observations fall below LOD, a common practice is to impute the censored values by either the LOD or some arbitrary value, such as LOD/2 [14,16,24]. To the best of our knowledge, relatively few studies have been conducted on simultaneously accounting for impacts induced by heterogeneity, non normality and left censoring.

It is not clear how these typical data features often observed in longitudinal studies may inherently interact and simultaneously influence inferential procedures. This article proposes a skew-normal (SN) mixture of semiparametric mixed effects (SN-MSPME) models to simultaneously account for longitudinal response with heterogeneity, skewness and left censoring. We demonstrate a Bayesian inferential approach to estimate both model parameters and class membership probabilities based on the SN-MSPME models, where mean functions (components) appear nonlinear. It is noted that the SN distribution reduces to a normal distribution if skewness parameter is zero. Thus, we use an SN distribution to develop the MSPME models and associated statistical methodology, as it can be easily reverted to the normal model case. In what follows, we consider multivariate SN distribution introduced by Sahu et al. [17], which is suitable for a Bayesian inference and briefly discussed in Appendix A. The rest of this article is organized as follows. Section 2 presents SN-MSPME models and associated Bayesian inferential approach in a general form. In Section 3, we describe the data set that motivated this research and discuss the specific mixture model based mean functions of different mixture components for viral load response that are used to illustrate the proposed methodology. In Section 4, we analyze an AIDS data set described in Section 3 to demonstrate the methodology and report the analysis results. Section 5 contains a concluding discussion.

### SETUP OF GENERAL SN-MSPME MODELS AND BAYESIAN INFERENCEAL APPROACH

Denote the number of subjects by $n$ and the number of measurements on the $i$th subject by $n_i$. Let $y_{ij}$ be the value of response for the individual $i$ at time $t_{ij}$ ($i=1,2,...,n; j=1,2,...,n_i$). The observed value $y_{ij}$ is $(q_{ij}, d_{ij})$, where $d_{ij}$ is the censoring indicator such that $y_{ij}$ is observed if $d_{ij} = 0$ and $y_{ij}$ is left censored if $d_{ij} = 1$, i.e. $y_{ij} = q_{ij}$ if $d_{ij} = 0$, and $y_{ij} \leq \rho$ if $d_{ij} = 1$, where $q_{ij}$ denotes the latent response variable that would be measured if the assay did not have an LOD $\rho$ (a known constant). In general, we consider a design matrix including time-independent and/
or time-varying covariates $Z_{ij}$ for the individual $i$ at time $t_i$.

Let the observed data $R = \{(q, d, z_i), i = 1, \ldots, n\}$, where $y_{ij} = (y_{ij1}, \ldots, y_{ijn})^T$, $z_i = (Z_{ij1}, \ldots, Z_{ijn})^T$, and $d_i = (d_{ij1}, \ldots, d_{ijn})^T$. We assume that there are $K$ plausible nonlinear trajectory classes with mean functions $g_k(\cdot)$ ($k = 1, \ldots, K$), which are known to be specified. The true trajectory mean function for the $i^{th}$ subject might be $g_k(\cdot)$ with unknown probability $\pi_i = P(c_i = k)$ which satisfies $\sum_i^{\infty} \pi_i = 1$, where $c_i$ is a latent indicator. In what follows, we present general SN-MSPME models and associated Bayesian inferential approach.

SN-MSPME models

Several studies [10,11] introduced finite mixture models for longitudinal data analysis, where the latent classes correspond to the mixture components and cluster individuals may provide a better inference. However, most finite mixture models for longitudinal data are currently based on linear (polynomial) [10] or piecewise linear [11] mean functions. The partial reason is that the computation for inference can be conveniently carried out because the likelihood function of a model based on these linear mean functions has a closed form [10]. However, in practice, most longitudinal trajectories appear to be nonlinear patterns. When a mixture model is extended to incorporate nonlinear mean functions, which will be conducted in this article, inferential procedures will complicate dramatically because a closed form of likelihood function no longer exists.

An SN semiparametric mixed-effects (SN-SPME) model for individual $i$, given $c_i = k$, can be formulated by

$$
(y_{ij} | c_i = k) = g_k(t_{ij}, A_k^T \beta_y^{(k)} \phi(t_{ij})),
$$

subject to

$$
\begin{align*}
\beta_k^{(k)} &= \phi(t_{ij}) = \frac{1}{\sqrt{2/\pi \sigma^2}} 
(2/\pi \sigma)^{1/2} e_i \Phi_t \left(\begin{array}{c}
\delta_i - \delta_i \left[ Z_{ij}, \beta, \beta_y \right]
\end{array}\right),
\end{align*}
$$

where $\beta_k^{(k)}$ and $\phi(t_{ij})$ are known parametric functions, $w(t)$ and $h(t)$ are unknown smooth fixed-effects and random-effects functions, respectively, and $\delta(t)$ is iid realization of a zero mean stochastic process; $1_{n_i} = (1, \ldots, 1)^T$, $\beta_y$ is a $s_y \times 1$ individual specific time dependent parameter vector, $\beta$ is a $s \times 1$ population parameter vector ($s \geq s_y$); the random error vector $e_i = (e_{i1}, \ldots, e_{in_i})^T$ follows a multivariate SN distribution with unknown variance parameter $\sigma^2$ and skewness parameter $\delta$; $s \times 1$ vector of random effects $b_i^T$ follows the multivariate normal distribution with mean zero and unrestricted covariance matrix $\Sigma$, $s_y \times 1$ vector of random effects $\beta_y^{(k)}$ follows the multivariate normal distribution with mean zero and unrestricted covariance matrix $\Sigma_y$.

Given $c_i = k$, the SPME model model (1) reverts to a commonly used NLME model when the nonparametric parts $w(t)$ and $h(t)$ are constants. To fit model (1), we apply a regression spline method to $w(t)$ and $h(t)$ by using a linear combination of spline basis functions. For instance, $w(t)$ and $h(t)$ can be approximated by a linear combination of basis functions $\Phi(t)$ and $\Phi_y(t)$, respectively. That is,

$$
\begin{align*}
w(t) &= w_\beta \sum_{q=1}^{q_{max}} \beta_q \Phi(t) + \xi_w, \\
h(t) &= h_\beta \sum_{q=1}^{q_{max}} \beta_y^{(k)} \Phi(t) + \xi_h,
\end{align*}
$$

where $\xi_w$ and $\xi_h$ are the vectors of random effects, and $\beta$ is unknown parameters vector.

Similarly, $A_k$ is known $s \times s_y$ indicator matrix, of which diagonal elements are either 0 or 1 and off diagonal elements are all 0. In model (1), we assume that the individual specific parameters $\beta_y^{(k)}$ depend on $Z_{ij}$, a design matrix including time independent and/or time varying covariates, such as CD4 cell counts.

Using $c_i = k$, the SPME model model (1) reverts to a commonly used NLME model when the nonparametric parts $w(t)$ and $h(t)$ are constants. To fit model (1), we apply a regression spline method to $w(t)$ and $h(t)$ by using a linear combination of spline basis functions. For instance, $w(t)$ and $h(t)$ can be approximated by a linear combination of basis functions $\Phi(t)$ and $\Phi_y(t)$, respectively. That is,

$$
\begin{align*}
w(t) &= w_\beta \sum_{q=1}^{q_{max}} \beta_q \Phi(t) + \xi_w, \\
h(t) &= h_\beta \sum_{q=1}^{q_{max}} \beta_y^{(k)} \Phi(t) + \xi_h,
\end{align*}
$$

where $\xi_w$ and $\xi_h$ are the vectors of random effects, and $\beta$ is unknown parameters vector.

Similarly, $A_k$ is known $s \times s_y$ indicator matrix, of which diagonal elements are either 0 or 1 and off diagonal elements are all 0. In model (1), we assume that the individual specific parameters $\beta_y^{(k)}$ depend on $Z_{ij}$, a design matrix including time independent and/or time varying covariates, such as CD4 cell counts.
introducing $A_i$, $A_i$, $\beta$, will set unrelated elements of $\beta$ to 0 in the $k$th trajectory class, respectively. We will illustrate the use of $A_i$ and specify nonlinear mean functions $g_i(\cdot)$ in the HIV dynamic application below. We assume that $e_i$ and $b_i$ are independent of each other.

Similar to discussion in [27], model (4) can be specified conditionally and marginally as follows, respectively.

\[
(y_i | c_i = k) \sim SN_{\gamma_k}(g_i(t_i, A_i, \beta_i), \sqrt{2/\pi \delta^2 \Sigma_k}),
\]

\[
\beta_i = d[Z_{y_i}, \beta, b_i], \quad b_i : N_0(0, \Sigma),
\]

\[
y_i \sim \sum_{k=1}^{K} \pi_k SN_{\gamma_k}(g_i(t_i, A_i, \beta_i) - \sqrt{2/\pi \delta^2 \Sigma_k} \delta I^{\gamma_k}),
\]

\[
\beta_i = d[Z_{y_i}, \beta, b_i], \quad b_i : N_0(0, \Sigma),
\]

The preceding model (6) defines a finite mixture of SN-NLME models which is a approximation from a finite mixture of SN-SPME models. In model (6), the vector of mixture probabilities $\pi = (\pi_1, \pi_2, \ldots, \pi_K)^T$ can be also viewed as the mixture weights of all plausible components under the framework of finite mixture models. Model (6) is identifiable, as long as each of the component models is identifiable and distinguishable from each other. When the component models are identifiable but not distinguishable from each other, certain constraints may be required to make model (6) identifiable [1].

### Bayesian inferential approach

In a longitudinal study, such as the AIDS study described previously, the longitudinal response is usually connected physically or biologically. We propose a Bayesian inferential approach via Markov chain Monte Carlo (MCMC) procedure to estimate all parameters in the mixture model (6) and class membership probabilities simultaneously. The Bayesian modeling approach may pave a way to alleviate the computational burdens and to overcome convergence problems for such complex mixture model setting. Bayesian analysis rests upon computing the conditional posterior distributions of the unknown parameters for inference, given the observed data and weighted by the prior information.

Let $\theta = \{\beta, \sigma^2, \Sigma, \delta\}$ be the collection of unknown parameters except for the mixture weight $\pi$ in the mixture model (6). Under the Bayesian framework, we need to specify prior distributions for all the unknown parameters as follows.

\[
\begin{align*}
\beta & \sim N_0(\tau, \Gamma), \quad \sigma^2 \sim IG(\alpha_0, \omega_0), \\
\Sigma & \sim IW(\Omega, p), \quad \delta \sim N(0, \gamma),
\end{align*}
\]

where $s_i = s_i + p$ and the mutually independent Inverse Gamma (IG), Normal (N) and Inverse Wishart (IW) prior distributions are chosen to facilitate computations. The super parameter matrices $\Gamma$ and $\Omega$ can be assumed to be diagonal for convenient implementation. By its definition, the latent indicating variables $c_i$, $i = 1, \ldots, n$, follow a Categorical distribution (Cat)

\[
c_i \sim \text{Cat}(1, 2, \ldots, K, (\pi_1, \pi_2, \ldots, \pi_K)),
\]

in which $\pi = (\pi_1, \pi_2, \ldots, \pi_K)^T$ follows a Dirichlet distribution (Dir) [7,28],

\[
\pi \sim \text{Dir}(\zeta_1, \zeta_2, \ldots, \zeta_K).
\]

An MCMC scheme for this mixture model is composed of following two basic steps [11]:

(i). Sampling class membership indicators $c_i$, $i = 1, \ldots, n$, conditional on population parameters $\beta$ and individual random effects $b_i$.

Generate $c_i$ ($i = 1, \ldots, n$) from

\[
P(c_i = k | b_i, \beta, y_i) = \frac{\pi_k f(y_i | b_i, c_i = k, \beta)}{\sum_{a \in \text{num}} f(y_i | b_i, c_i = a, \beta)},
\]

where $f(y_i | b_i, c_i = k, \beta)$ is a conditional density function of $y_i$ based on (5). Then, update the probability $\pi$ for next iteration from distribution

\[
\pi \sim \text{Dir}(c_i + \text{num}, \ldots, c_i + \text{num}),
\]

where $\text{num} = \sum f(y_i | b_i, c_i = k, k = 1, \ldots, K)$, in which $I()$ is an indicator function.

(ii). Sampling parameters $\theta$, and individual random-effects $b_i$, conditional on class membership indicator $c = (c_1, \ldots, c_n)^T$.

Following the study by Sahu et al. [17], it can be shown, conditional on $c_i$ determined in step (i), that by introducing the random vector $w = (w_{1n}, \ldots, w_{jn})^T$ based on the stochastic representation for the SN distribution (see Appendix A for details), $y_i$ with random effects $b_i$ in the presence of left censoring can be hierarchically formulated as

\[
y_i | w_i, b_i, c_i \sim N_{\gamma_i}(g_i(t_i, A_i, \beta_i) + \sqrt{2/\pi \delta^2 \Sigma_k} \delta I^{\gamma_k}),
\]

\[
w_i \sim N_{\gamma_i}(0, I), \quad l(w_i > 0),
\]

\[
\sigma \sim N_{\gamma_i}(0, \gamma), \
\]

\[
\delta \sim N(0, \gamma)
\]

Let $f(\cdot | \cdot)$, $F(\cdot | \cdot)$ and $h(\cdot)$ denote a probability density function (pdf), cumulative density function (cdf) and prior density function, respectively. Conditional on the random variables and some unknown parameters, a detectable measurement $y_i$ contributes $f(y_i | b_i, w_i)$, whereas a non-detectable measurement contributes $F(\rho | b_i, w_i) = P(y_i < \rho | b_i, w_i)$ in the likelihood. We assume that $\beta$, $\sigma^2$, $\Sigma$ and $\delta$ are independent of each other, i.e., $h(\theta) = h(\beta)h(\sigma^2)h(\Sigma)h(\delta)$. After we specify the mixture models for the observed data and the prior distributions for the unknown model parameters, we can make statistical inference for the parameters based on their posterior distributions under the Bayesian framework. Thus, the joint posterior density of $\theta$ based on the observed data $g = \{g_i, d, c\}$ and classification indicator can be given by

\[
f(\theta | g) \propto \prod_{i=1}^{n} \prod_{j=1}^{m} f(y_{ij} | b_i, w_{ij})^{c_{ij}}
\]

\[
F(\rho | b_i, w_{ij})^{c_{ij}} f(w_{ij} | w_{ij} > 0) h(b_i) h(\theta).
\]
where \( d_y \) is the censoring indicator such that \( y_y \) is observed if \( d_y = 0 \) and \( y_y \) is left censored if \( d_y = 1 \), i.e. \( y_y = q_y \) if \( d_y = 0 \), and \( y_y \) is treated as left-censoring if \( d_y = 1 \).

In general, the integrals in (13) are of high dimension and do not have a closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of \( \theta \) based on the observed data and class membership. As an alternative, the MCMC procedure can be used to sample population parameters, \( \theta \), and random effects, \( a_i \) and \( b_i \) \((i = 1, ..., n)\), from conditional posterior distributions, based on (13), using the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm. Steps (i) and (ii) are repeated alternatively in iterations of MCMC procedure until convergence is reached. An important advantage of the above representations based on the mixture of hierarchical models with SN distribution is that they are easily implemented using the freely available WinBUGS software [29] interacted with a function called bugs in a package R2WinBUGS of R. The other advantage is that when WinBUGS software is used to implement our modeling approach, it is not necessary to explicitly specify the full conditional posterior distributions for parameters to be estimated. Although their derivations are straightforward by working the complete joint posterior, some cumbersome algebra will be involved. We, thus, omit those here to save space.

Motivating data set and specific components in mixture models

The data set that motivated this research is from an AIDS clinical trial study (ACTG398), which is a randomized, double blind, placebo controlled, with an extension to more than 48 week study of saquinavir, indinavir, or nelfinavir added as second protease inhibitor to the 4 drug class regimen in patients with virologic failure defined by receiving saquinavir, nelfinavir, indinavir, or ritonavir [12]. This study consists of 481 HIV-1 infected patients. The plasma HIV-1 RNA (viral load) is repeatedly measured throughout study on a similar scheme. A log transformation of viral load was used in the analysis in order to stabilize the variation of the measurement errors and to speed up estimation algorithm. The CD4 cell counts were also measured throughout study on a similar scheme. 19.5% observations of viral load were measured below LOD.

As described previously, we may approximately describe the effect of treatment on viral load in two patterns as displayed in (Figure 1). Class (pattern) 1 of the viral load trajectories, where patients have viral load decrease, is used to model suppression of plasma HIV-1 RNA levels that the treatment can be thought successful without serious clinical problems arose, suggesting a confirmed long term virologic response; while class (pattern) 2 of the viral load trajectories, where patients experience viral load increase, results in viral load rebound eventually, implying a virologic failure.

Viral dynamic models can be formulated through a system of ordinary differential equations (ODE) [13,14,30,31,32]. Under some reasonable assumptions and simplifications, a useful approximation of ODE solution, which can be used to capture viral load responses, has been proposed as follows.

\[
y(t) = \log_{\text{base}}(e^{\lambda_1(t)-\lambda_2} + e^{\lambda_2(t)-\lambda_1(t)}),
\]

where \( y(t) \) is the \( \log_{\text{base}} \) scaled plasma HIV-1 RNA levels at time \( t \). \( \lambda_1 \) and \( \lambda_2 \) are called the first and second phase viral decay rates, which may represent the minimum turnover rate of productively infected cells and that of latently or long lived infected cells, respectively [13]. The parameters \( \lambda_1 \) and \( \lambda_2 \) are macro parameters; \( e^{\lambda_1} + e^{\lambda_2} \) are the baseline viral load at time \( t = 0 \) in the two compartment models. It is generally assumed that \( \lambda_1 > \lambda_2 \), which assures that the model is identifiable and is appropriate for empirical studies [14,33]. It was noted that equation (14) can be only applied to the early segment or longer term of the viral load response with decreasing trajectory patterns as shown in (Figure 1) (three solid lines). However, in practice, for some patients the second-phase viral decay rate, \( \lambda_2 \), may vary over time because they depend on some phenomenological parameters that hide with considerable microscopic complexity and change over time [31]. Negative values of the decay rates may correspond to viral increase and lead to viral rebound [14], suggesting that variation in the dynamic parameters, particularly \( \lambda_2 \), may be partially associated with time varying covariates such as repeated CD4 cell counts. Thus, it may not be reasonable to assume that second phase viral decay rate, \( \lambda_2 \), is a constant when viral load is rebounded at the later stage during long term treatment. To model the long-term HIV dynamics, a natural extension of equation (14) is to assume that the second-phase viral decay rate, \( \lambda_2(t) \), changes over time to capture the viral load change including viral rebound. Thus, we introduce an extended function as follows.

\[
y(t) = \log_{\text{base}}(e^{\lambda_1(t)-\lambda_2} + e^{\lambda_2(t)-\lambda_1(t)}),
\]

where the second-phase decay rate, \( \lambda_2(t) \), may be a function of time-varying covariate such as CD4 cell count and/or an unknown smooth function. Based on discussion above, we consider equation (14) with two constant viral decay rates for class 1 trajectories defined in Section 1, and (15) with a time-varying viral decay rate in the second compartment for class 2 trajectories. It is noted that (15) is a natural extension of (14) to consider a time-varying decay rate for capturing viral rebound in class 2 trajectories. Thus, the mean functions of \( K = 2 \) components in the mixture models are specified by

1. Two-compartment equation with two constant decay rates for class 1 trajectories

\[
g_1(t, \lambda_1, \lambda_2) = \log_{\text{base}}(e^{\lambda_1(t) - \lambda_2} + e^{\lambda_2(t) - \lambda_1(t)}),
\]

2. Two-compartment equation with constant and time-varying decay rates for class 2 trajectories

\[
g_2(t, \lambda_1, \lambda_2) = \log_{\text{base}}(e^{\lambda_1(t) - \lambda_2(t)} + e^{\lambda_2(t) - \lambda_1(t)}),
\]

In (16) and (17),
\[ \begin{align*}
\beta_6 = p_6 &= \beta_1 + b_1, \\
\beta_7 = p_7 &= \beta_1 + b_1 + b_3,
\end{align*} \]

where \( z_{a1} \) is the baseline CD4 and \( z_d \) is the value of CD4 cell count at time \( t_a \). The decay rate of the second compartment in (17), \( \beta_{a2} \), is time-varying due to \( z_d \) and unknown nonparametric function \( \phi(t_i) = w(t_i) + h(t_i) \), but other parameters in \( \beta_i \) are time independent. As mentioned previously, the mean functions in different components may involve different subsets of \( \beta_i \); for example, \( g_1(t) \) and \( g_2(t) \) share the same parameters \( \beta_i \), \( \beta_{a2} \), and \( \beta_{a3} \), but have different second-phase decay rate, \( \beta_{a4} \) and \( \beta_{a5} \), respectively. The diagonal indicator matrices, \( A_i \) and \( A_j \), determine which elements of \( \beta_i \) are involved and set other unrelated parameters to be 0 in the mean functions, \( g_1(t) \) and \( g_2(t) \), respectively. With this mixture clustering, our mixture modeling can be used to estimate probabilities of class membership to classify either viral rebound eventually, suggesting a virologic failure (class 2) or viral decrease continuously, indicating a confirmed long-term virologic response (class 1).

We employ the linear combinations of natural cubic splines with percentile-based knots to approximate the nonparametric functions \( w(t) \) and \( h(t) \). Following studies \([26,34]\), we set \( \psi_0(t) = \phi_0(t) = 1 \) and take the same natural cubic splines in the approximations (2) with \( q \leq p \). The values of \( p \) and \( q \) are determined based on the AIC/BIC which suggest the following functions for \( \phi(t_i) \) with \( p = 3 \) and \( q = 1 \) in model (2):

\[ \phi(t_i) = w(t_i) + h(t_i) + \mu_\psi \psi(t_i) + \mu_\phi \phi(t_i) + \xi_{a4}. \]

where \( \mu_\psi = (\mu_\psi, \mu_\psi, \mu_\psi, \mu_\psi, \mu_\psi) \) and \( \xi_{a4} = b_{a4} \).

**ANALYSIS OF AIDS CLINICAL DATA**

**Model implementation**

As discussed previously, the data of viral load outcomes (in log_{10} scale) indicate the asymmetric feature. Thus, it seems plausible to fit a skew distribution to the data. Towards this end, the following three statistical models with specifying different distributions for model error for viral load response in the presence of left-censoring are employed to compare their performance.

- **Model N**: A mixture of semiparametric mixed-effects model where the two mean functions specified by (16) and (17) with the \( N \) distribution for model error.

- **Model SN**: A mixture of semiparametric mixed-effects model where the two mean functions specified by (16) and (17) with the SN distribution for model error.

- **Model SP**: A commonly used semiparametric mixed-effects model where the mean function specified by (17) alone with the SN distribution for model error.

We conducted data analysis with the following three scenarios. First, because a normal distribution is a special case of an SN distribution when the skewness parameter is zero, we investigated how asymmetric distribution for model error (Model SN) contributes to modeling results in estimates of both parameters and class membership probabilities in comparison with a symmetric (normal) distribution for model error (Model N). Second, we estimated the model parameters and class membership probabilities by using the "naive" method (denoted by NM), where left-censoring is ignored and actual observations below LOD in viral load response are used. We used it as a comparison to the modeling approach (denoted by LC) proposed where viral load observations below LOD were treated as left-censoring. This comparison attempted to investigate how the LODs treated as left-censoring in viral load contribute to modeling results. Finally, we further investigated a commonly used semiparametric mixed-effects (SPME) model with SN distribution (Model SP), where the mean function is specified by (17) alone. This SPME model has been widely applied to analyze viral load data \([26,34,35,36]\) in significantly advancing the understanding of pathogenesis of HIV infection and the assessment of effectiveness of ART. We compared this commonly used SPME model (ignoring data feature of heterogeneous population) with the mixture of SPME model proposed in this article to explore how heterogeneous feature influences modeling results.

To carry out the Bayesian inference, we took weakly-informative prior distributions for the parameters in the models. In particular, (i) fixed-effects were taken to be independent normal distribution \( N(0,100) \) for each element of the population parameter vector \( \beta \); (ii) we assume a noninformative inverse Gamma prior distribution \( IG(0.01, 0.01) \), which has mean 1 and variance 100, for variance parameter \( \sigma^2 \); (iii) the prior for the variance-covariance matrix of the random-effects \( \Sigma \) was taken to be inverse Wishart distributions \( IW(\Omega, \rho) \), where the diagonal elements for diagonal variance matrix \( \Omega \) were 5; (iv) for the skewness parameter \( \delta \), we chose independent normal distributions \( N(0,100) \); and (v) we set hyper-parameters of Dirichlet distribution in (9), \( \xi_0 = \xi_1 = 1 \), assuming individuals have equal probabilities of coming from any one of two classes initially.

The MCMC sampler was implemented using WinBUGS software \([29]\) interacted with a function called bugs in a package R2WinBUGS of R and the program code is available from authors upon request. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as trace plots and Gelman-Rubin (GR) diagnostics \([37]\). Figure 2 shows the trace plots, the dynamic version of GR diagnostics and autocorrelation plots based on Model SN as obtained from the WinBUGS software for the representative and important parameters \( \beta_i \) and \( \beta_s \). We observe from trace plots (top panel) that the lines of three different chains mix or cross in trace, implying that convergence...
is ensured. For the plots of GR diagnostics (middle panel) where the three curves are given: the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{W}$) and average within-sample variance ($\hat{W}$), respectively, and the top curve represents their ratio ($\hat{R}$). It is seen that $\hat{R}$ is generally expected to be higher than one at the initial stage of the algorithms, but $\hat{R}$ tends to 1, and $\hat{V}$ and $\hat{W}$ stabilize as the number of iterations increase, indicating that the algorithm has approached convergence. We further monitor convergence using autocorrelation plots (bottom panel) that autocorrelations are very low with a lag being 50, implying that convergence is obtained. With the convergence diagnostics observed, we proposed that, after an initial number of 50,000 burn in iterations of three chains of length 100,000, every 50th MCMC sample (i.e., thin set equal to 50) was retained from the next 50,000 for each chain. Thus, we obtained a total of 3,000 samples of targeted posterior distributions of the unknown parameters for statistical inference.

Results of data analysis

Bayesian modeling approach in conjunction with the mixture of SPME models with different specifications of error distributions was used to fit the viral load data with multiple features simultaneously. Table 1 presents the population posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effects parameters based on the proposed models and two methods (LC and NM). The following findings are obtained for the results of estimated parameters.

In Models N and SN, the findings, particularly for the fixed-effects ($\beta_1, \beta_6$) which are parameters of the first-phase viral decay rate, show that $\beta_1$ is insignificantly estimated, while the estimate of $\beta_6$, the coefficient of baseline CD4 count, is significantly positive since the 95% credible intervals do not contain zero, indicating that the baseline CD4 has positive effect on the first phase viral decay rate. The fixed effects ($\beta_1, \beta_6$), which are parameters related to second-phase viral decay rate, show that these estimates are different from zero. Nevertheless, for the estimate of the coefficient of CD4 covariate $\beta_6$, its estimate is significantly positive. This means that CD4 has a significantly positive effect on the second phase viral decay rate, suggesting that the CD4 covariate may be an important predictor of the second-phase viral decay rate during the treatment. In addition, there is difference in posterior means of the variance (scale

**Figure 2** Convergence diagnostics with three Markov chains as obtained from the WinBUGS software for representative parameters based on Model SN: (a) Trace plots (top panel). (b) Gelman-Rubin (GR) diagnostic plots (middle panel): at the initial stage, the middle (green) and bottom (blue) curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($\hat{W}$), respectively, and the top (red) curve above the dashed horizontal line represents their ratio ($\hat{R}$). (c) Autocorrelation plots (bottom panel).
parameter) $\sigma^2$ (0.239 vs. 0.146) in comparison of (symmetric) Model N with (asymmetric) Model SN. The estimated value of the variance in Model SN is smaller than that in Model N because the former model take into account skewness of the data while the latter does not. The estimate of the skewness parameter (\( \delta \)) of Model SN is 0.436 with 95% CI (0.123, 0.638). This finding suggests that there is a significantly positive skewness in the data and confirms the fact that the distribution of the original data is skewed even after taking log-transformation. Thus, incorporating a skewness parameter in modeling skewed data is recommended.

We now focus on the lower end of the distribution of the viral load where there is left censoring due to LOD. As it was mentioned in the introduction section, the current assay techniques for quantifying HIV-RNA viral load may not give accurate readings below LOD, which in our data is 50 copies/mL. In our analysis, we treated those inaccurate observed viral loads as left-censoring below LOD, which in our data is 50 copies/mL. In our analysis, we treated those inaccurate observed viral loads as left-censoring below LOD, which in our data is 50 copies/mL. The predicted values of the unobserved viral load below LOD are spread out as expected (see Figure 3(b-c)). However, Model N produces the predicted values exceeded the LOD much more than Model SN does. In addition, when we compare Models N and SN in terms of their distributions in predicting viral loads (\( \log_{10} \text{scale} \)) below LOD, we can see that Model SN gives more plausible values ranged within (0.5, 1.8) than Model N does in the sense that the distribution is closely fitted the lower part of the whole distribution of the predicted viral load values based on Model SN as expected, implying that Model SN is the better model. This finding also confirms the conclusion made using other criteria (see below).

To assess the goodness-of-fit for the proposed statistical models, the diagnosis plots for Models N and SN, the residuals versus the fitted values (left panel), the observed values versus the fitted values (center panel), and normal and SN Q-Q plots (right panel), are presented in Figure 4. The residual plots (left panel) indicate that no major systematic patterns are found and in general the models fit the observed data reasonably well. However, it was seen from the plots of the observed values versus the fitted values (middle panel) that Model SN provided better fit to observed data, compared with Model N. This result can be also explained by examining the normal and SN Q-Q plots of the residuals (right panel) that both plots show the existence of outliers, but it is clearly indicated that Model SN only has few positive outliers and, thus, fit observed data better than Model N. This finding is confirmed by their residual sums of squares (RSS), formulated by RSS = \( \sum_{i \in \text{outliers}} (y_{\text{model},i} - y_{\text{obs},i})^2 \), which are 666.8 (Model N) and 406.1 (Model SN), respectively.

**Table 1:** Summary of estimated posterior mean (PM) for parameters of population (fixed-effects), variance, skewness and degree of freedom, and corresponding standard deviation (SD), lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) as well as values of deviance information criterion (DIC), expected predictive deviance (EPD) and residual sum of squares (RSS) based on the proposed modeling approach (LC) and "naive" method (NM).

<table>
<thead>
<tr>
<th>Method</th>
<th>Model</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\sigma^2$</th>
<th>$\delta$</th>
<th>DIC</th>
<th>EPD</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC</td>
<td>N</td>
<td>PM 10.72</td>
<td>16.78</td>
<td>39.03</td>
<td>7.13</td>
<td>-7.22</td>
<td>1.63</td>
<td>0.239</td>
<td>-</td>
<td>13365.2</td>
<td>0.478</td>
<td>666.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$L_{CI}$ 10.53</td>
<td>-2.32</td>
<td>23.34</td>
<td>6.84</td>
<td>-8.57</td>
<td>0.97</td>
<td>0.218</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U_{CI}$ 10.90</td>
<td>35.80</td>
<td>35.32</td>
<td>7.42</td>
<td>-5.97</td>
<td>2.32</td>
<td>0.262</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD 0.09</td>
<td>9.75</td>
<td>3.05</td>
<td>0.15</td>
<td>0.65</td>
<td>0.34</td>
<td>0.011</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>SN</td>
<td>PM 10.74</td>
<td>15.88</td>
<td>31.10</td>
<td>7.12</td>
<td>-8.25</td>
<td>0.24</td>
<td>0.146</td>
<td>0.436</td>
<td>10357.2</td>
<td>0.291</td>
<td>406.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$L_{CI}$ 10.55</td>
<td>-3.00</td>
<td>25.75</td>
<td>6.83</td>
<td>-9.56</td>
<td>0.08</td>
<td>0.078</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U_{CI}$ 10.92</td>
<td>34.10</td>
<td>36.72</td>
<td>7.41</td>
<td>-7.03</td>
<td>0.69</td>
<td>0.240</td>
<td>0.639</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD 0.10</td>
<td>9.61</td>
<td>2.80</td>
<td>0.15</td>
<td>0.64</td>
<td>0.23</td>
<td>0.042</td>
<td>0.222</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>SP</td>
<td>PM 10.73</td>
<td>16.34</td>
<td>29.98</td>
<td>7.01</td>
<td>-9.84</td>
<td>0.21</td>
<td>0.154</td>
<td>0.410</td>
<td>11853.4</td>
<td>0.318</td>
<td>429.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$L_{CI}$ 10.53</td>
<td>-1.74</td>
<td>24.67</td>
<td>6.68</td>
<td>-11.47</td>
<td>0.05</td>
<td>0.091</td>
<td>0.113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U_{CI}$ 10.91</td>
<td>35.02</td>
<td>35.16</td>
<td>7.30</td>
<td>-8.34</td>
<td>0.51</td>
<td>0.242</td>
<td>0.617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD 0.10</td>
<td>9.47</td>
<td>2.72</td>
<td>0.16</td>
<td>0.79</td>
<td>0.16</td>
<td>0.041</td>
<td>0.228</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NM</td>
<td>SN</td>
<td>PM 10.67</td>
<td>19.14</td>
<td>27.04</td>
<td>6.98</td>
<td>-1.98</td>
<td>3.07</td>
<td>0.187</td>
<td>0.601</td>
<td>12112.78</td>
<td>0.374</td>
<td>521.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$L_{CI}$ 10.47</td>
<td>0.64</td>
<td>21.50</td>
<td>6.67</td>
<td>-3.17</td>
<td>2.24</td>
<td>0.142</td>
<td>0.485</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U_{CI}$ 10.87</td>
<td>38.33</td>
<td>32.61</td>
<td>7.27</td>
<td>-0.79</td>
<td>3.79</td>
<td>0.235</td>
<td>0.700</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD 0.10</td>
<td>9.78</td>
<td>2.84</td>
<td>0.15</td>
<td>0.60</td>
<td>0.39</td>
<td>0.024</td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
To select the best model that fits the data adequately, a Bayesian selection criterion, known as deviance information criterion (DIC) suggested by [38], is used. As with other model selection criteria, we caution that DIC is not intended for identification of the “correct” model, but rather merely as a method of comparing a collection of alternative formulations. As an alternative, we also evaluate expected predictive deviance (EPD) formulated by

\[ \text{EPD} = E \left\{ \sum_{ij} (y_{\text{rep},i,j} - y_{\text{obs},i,j})^2 \right\} \]

for model comparison, where the predictive value \( y_{\text{rep},i} \) is a replicate of the observed \( y_{\text{obs},i} \) and the expectation is taken over the posterior distribution of the model parameters \( \theta \) (see [39] in detail). This criterion chooses the model where the discrepancy between predictive values and observed values is the lowest. For mixture models, the structure of DIC does not allow for automatic computation in WinBUGS software. We wrote R codes to calculate estimated DICs based on
As mentioned in Section 1, one of the primary objectives in this data analysis was to cluster all individuals into two classes of viral load trajectories: (i) decrease rapidly at the beginning in a short term period and/or then stay stable at a low level; and (ii) decrease at the beginning, but rebound later. Based on the mixture modeling, we are able to obtain a summary of class membership at both the population and individual levels. Towards this end, the number of components in this analysis is determined empirically based on the viral load trajectory patterns and clinical interpretability. At population level, the MCMC procedure yields samples from the posterior distribution of \( \bar{\lambda} = (\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_t) \) in (11), the population proportion of individuals in each class. The estimates of population proportion and associated 95% CI of \( (\bar{\lambda}_1, \bar{\lambda}_2, \bar{\lambda}_t) \) for two classes are 62.58% (58.94%, 66.04%) and 37.42% (32.97%, 41.06%), respectively. Thus, out of 379 patients, the patterns of changing viral load of 237 and 142 patients followed classes 1 and 2, respectively. This finding indicates that a confirmed virologic response were observed in 62.58% of the patients in class 1. At individual level, the posterior probability of individual \( i \) belonging to the \( k \) th (\( k = 1, 2 \)) class, \( P_a = P[I_C(\text{class } k)] \), can be approximated by \( M \sum_{i=1}^{M} I(C_{ik} = k) \) in which \( C_{ik} \) is class membership of individual \( i \) drawn from the posterior distribution in the \( n \) th MCMC iteration (\( i = 1, \ldots, M \) ), where \( M \) is the iteration number of posterior samples (3,000 here). Table 2 shows individual posterior probabilities for the six representative patients shown in (Figure 1). The probabilities shown in (Table 2) and the classes of trajectories displayed in Figure 1 are matched quite well. The patients 3, 29 and 31 belong to class 1 because their viral load trajectories decrease rapidly at the beginning in a short-term period and/or then stay stable at a low level, with probabilities 97%, 98% and 95% respectively; and the patients 28, 33 and 99 are in class 2 (viral load rebound), with probabilities 89%, 87% and 100% respectively. The probability corresponding to individual patient who is classified as either a confirmed virologic response or a virologic failure (viral load rebound) may help physicians to refine treatment strategy to and point to further research even though the true association described above may be more complicated.

In order to investigate how the left censoring due to LOD in viral load contributes to modeling results, we compared two methods for estimation based on Model SN: the proposed modeling approach (LC), where the viral load below LOD was treated as left censoring and the “naive” method (NM), where left censoring was ignored and the actual observations below LOD in viral load were applied in the mixture modeling. It can be seen from (Table 1) that there are important differences in the parameter estimates, in particular, for the parameters \( \beta_2 \) and \( \beta_3 \), which are directly associated with the time-varying CD4 measurements. The NM may substantially overestimate the covariate CD4 effect (NM: \( \beta_3 = 3.07 \) versus LC: \( \beta_3 = 0.24 \)). The estimated SD for the CD4 effect (\( \beta_2 \)) using LC is smaller than that using NM, probably because LC approach treated viral load observations below LOD as left-censoring. The difference of estimates between LC and NM, due to whether viral load observations below LOD are treated as left censoring or not, indicates that the left censoring due to LOD in viral load should be incorporated in the analysis. Thus, it is important to take the LOD into account when collected data are “inaccurately” measured.

We further investigated a commonly used SPME model with SN distribution (Model SP, where data feature of heterogeneous population is ignored), in which the mean function is specified by (17) alone. We compared it with the mixture of SPME model (Model SN) to explore how heterogeneous feature influences modeling results. We found the important differences in the parameter estimates for the parameters \( \beta_1 \) and \( \beta_2 \), which are associated with the first-phase viral decay rate, and for the parameters \( \beta_3 \), which are associated with the second-

### Table 2: Individual posterior probabilities of belonging to two trajectory classes for the six representative patients

<table>
<thead>
<tr>
<th>Class</th>
<th>Patient ID</th>
<th>( P_a )</th>
<th>( P_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>97%</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>98%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>95%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>11%</td>
<td>89%</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>13%</td>
<td>87%</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>0%</td>
<td>100%</td>
</tr>
</tbody>
</table>


phase viral decay rate. We also obtained estimated DIC values with 10357.2 (Model SN) and 11853.4 (Model SP), respectively, suggesting that the mixture modeling is preferable. Furthermore, one advantage of the mixture of SPME model over the commonly used SPME model is its flexibility to provide not only estimates of all model parameters, but also evaluate class membership probabilities at both population and individual levels, which is helpful for clinicians to develop individualized ART.

**DISCUSSION**

For longitudinal viral load served as outcome measures to evaluate treatment effects of ART in AIDS clinical trials, to understand pathogenesis of HIV infection and to assess risk of disease progression to AIDS, we have developed a Bayesian modeling approach to the finite mixture of SPME models for longitudinal viral load data with features of heterogeneous population, non normality and left censoring due to LOD for estimating both model parameters and class membership probabilities. Along with this line, one advantage of mixture modeling is its flexibility to handle longitudinal data with different characteristics and provide not only estimates of all model parameters, but also model based probabilistic clustering to obtain class membership probabilities at both population and individual levels. This information may help clinicians refine general treatment strategy and develop individualized ART regimens. This kind of mixture modeling approach is important in many biostatistical applications, allowing appropriate inference of parameters while adjusting for heterogeneity, non normality and inaccuracy (due to LOD) of longitudinal data observed. Although this article is motivated by AIDS clinical study, the basic concepts of the developed mixture of SPME models have generally broader applications whenever the relevant technical specifications are met and longitudinal measurements are assumed to arise from two or more identifiable subclasses within a population.

This article have considered two mixture modeling methods (LC and NM) to compare various scenarios with potential models. In particular, (i) we investigated how asymmetric error distribution (Model SN) contributes to inferential results in comparison with a symmetric (normal) error distribution (Model N). Our results based on the proposed mixture modeling approach using skew distribution evidenced that there is potential to gain efficiency and accuracy in estimating certain model parameters and class membership probabilities when the normality assumption does not hold in the data. Thus, it is important to assume a skew distribution in the mixture models for the viral load response in order to achieve reliable results, in particular if the data exhibit the feature of non normality. (ii) We compared LC with NM to investigate how left censoring due to LOD affects modeling results and inference. The findings indicate that it is critical to treat viral load observations below LOD as left censoring when collected data are “inaccurately” measured. (iii) Under the assumption of SN distribution for model error, we further explored how heterogeneous data feature influences modeling results by comparing Model SN with a commonly used SPME model (Model SP), in which the data feature of heterogeneous population is ignored. The results suggest that there are important differences due to whether the data feature of heterogeneous population is ignored or not for inference. The proposed mixture models and methods may have a significant impact on HIV/AIDS research and help improve understanding of the pathogenesis of HIV infection and evaluation of individualized ART regimens.

For inference of mixture models, parameter (or model) identifiability can be an important but difficult problem when a large number of model parameters must be estimated simultaneously. We must ensure each component model to be identifiable to make whole mixture model to be identifiable. To make (16) and (17) to be identifiable, we assume \( \lambda_{ij} > \lambda_{jj} \) in (16) and \( \lambda_{ij} > \lambda_{jj} \) in (17). In practice, if the models are not identifiable, the MCMC algorithm would diverge quickly. In the application considered in this article, the MCMC algorithm was converged without problems as depicted in Figure 2 and, thus, we did not observe potential identifiability problems. As noted by Pauler and Laird [11], to formalize the finite mixture models, the mean functions of component submodels can be similar in form, but varying only mean and/or variance specifications or have entirely different functional forms with parameters of different dimensions and meanings across the component submodels, which is the case conducted in this article; see equations (16) and (17). Instead of specifying different mean functions for classes, alternatively, one can also specify an universal mean function for all classes, for example, the mean function (17), and let data themselves determine the number of clusters and/or shapes of trajectories. Towards this end, the label switching issue may arise and some labelling methods [40,41] must be applied to solve this problem, but significantly additional efforts are needed based on the proposed mixture models of HIV dynamics in our study. We are actively investigating these important issues, and hope that these interesting results could be reported in the near future.

**APPENDIX A. MULTIVARIATE SKEW-NORMAL DISTRIBUTION**

Different versions of multivariate skew distributions have been proposed and used in the literature [17,19,20,21,22]. A new class of distributions by introducing skewness in multivariate elliptically distributions, referred as skew-elliptical (SE) distributions, were developed in publication [17,18]. The class, which is obtained by using transformation and conditioning, which contains many standard families including the multivariate skew-normal (SN) distribution as special case. An \( k \)-dimensional random vector \( \mathbf{Y} \) follows a \( k \)-variate SE distribution if its probability density function (pdf) is given by

\[
 f(\mathbf{y} | \mathbf{\mu}, \mathbf{\Sigma}, \mathbf{\Delta}; m^{(i)}) = 2^\nu f(\mathbf{y} | \mathbf{\mu}, \mathbf{A}; m^{(i)}) P(V > 0),
\]

where \( \mathbf{A} = \mathbf{\Sigma} + \mathbf{\Delta}^2 \), \( \mathbf{\mu} \) is a location parameter vector, \( \mathbf{\Sigma} \) is a \( k \times k \) positive (diagonal) covariance matrix, \( \mathbf{\Delta} = \text{diag}(\delta_1, \delta_2, \ldots, \delta_k) \) is a \( k \times k \) skewness matrix with the skewness parameter vector \( \mathbf{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)^T \); \( V \) follows the elliptical distribution

\[
 E(\mathbf{\Delta} \mathbf{A}^{-1} (\mathbf{y} - \mathbf{\mu}), I_1 - \mathbf{A}^{-1} \mathbf{A}^T; m^{(i)})
\]

and the density generator function

\[
 m^{(i)}(u) = \frac{\Gamma(k/2)}{\pi^{k/2}} \frac{m_i(u)}{\int_0^{\Gamma(k/2)} m_i(u) du}, \text{ with } m_i(u) \text{ being a}
\]
function such that \( \int e^{x^2} m(x)dx \) exists. The function \( m(x) \) provides the kernel of the original elliptical density and may depend on the parameter \( v \). This SE distribution is denoted by \( SE(\mu, \Sigma, \Delta, m(\cdot)) \). One example of \( m(x) \), leading to an important special case used throughout the paper, is \( m(x) = \exp(-x^2 / 2) \). This expression leads to the multivariate SN distribution.

As we know, a normal distribution is a special case of an SN distribution when the skewness parameter is zero. For completeness, this Appendix briefly summarizes the multivariate SN distribution introduced by Sahu et al. [17] to be suitable for a Bayesian inference since it is built using the conditional method. For detailed discussions on properties of SN distribution, see Reference [17]. Assume a \( k \)-dimensional random vector \( Y \) follows a \( k \)-variable SN distribution with location parameter \( \mu \), \( k \times k \) positive (diagonal) covariance matrix \( \Sigma \) and \( k \times 1 \) skewness diagonal matrix \( \Delta = \text{diag} (\delta_1, \delta_2, \ldots, \delta_k) \). A \( k \)-dimensional random vector \( Y \) follows a \( k \)-variable SN distribution, if its pdf is given by

\[
 f(y | \mu, \Sigma, \Delta) = 2^k | A^{-1/2} \| \varphi_i(A^{-1/2} (y - \mu)) P(V > 0),
\]

where \( V \sim \mathcal{N}_k(\Delta^T (y - \mu), I_k - \Delta^T \Delta) \), and \( \varphi_i(\cdot) \) is the pdf of \( \mathcal{N}_i(0, I_1) \). We denote the above distribution by \( SN_k(\mu, \Sigma, \Delta) \). An appealing feature of equation (2) is that it gives independent marginal when \( \Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2) \). The pdf (2) thus simplifies to

\[
 f(y | \mu, \Sigma, \Delta) = \prod_{i=1}^{k} \left[ \frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \varphi_i \left( \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right) \Phi \left( \frac{\delta_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right) \right],
\]

where \( \varphi_i(\cdot) \) and \( \Phi(\cdot) \) are the pdf and cdf of the standard normal distribution, respectively. The mean and covariance matrix are given by \( E(Y) = \mu + \sqrt{2/\pi} \delta \), \( Cov(Y) = \Sigma + (1 - 2/\pi) \Delta^T \). It is noted that when \( \delta = 0 \), the SN distribution reduces to usual normal distribution. In order to have a zero mean vector, we should assume the location parameter \( \mu = -\sqrt{2/\pi} \delta \).

According to the study by Arellano-Valle et al. [20], if \( Y \) follows \( SN_k(\mu, \Sigma, \Delta) \), it can be expressed by a convenient stochastic representation as follows:

\[
 Y = \mu + \Delta | X_0 \rightarrow \Sigma^{1/2} X_1,
\]

where \( X_0 \) and \( X_1 \) are two independent \( N(0, I_1) \) random vectors. Let \( w = | X_0 \|; \) then, \( w \) follows a \( k \)-dimensional standard normal distribution \( N(0, I_k) \) truncated in the space \( w > 0 \). Thus, a two-level hierarchical representation of (4) is given by

\[
 Y \mid w \sim N_k(\mu + \Delta w, \Sigma), \quad w \sim N_1(0, I_1) (w > 0).
\]

REFERENCES

and Data Analysis. 2008; 52: 5033-5045.


