



A Fast Algorithm for Using Semi-Parametric Random Effects Model for Analyzing Longitudinal Data

Taban Baghfalaki^{1, *} and Mojtaba Ganjali²

¹*Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Iran.*

²*Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Iran.*

Received: 2 November 2013; Accepted: 22 September 2014

Editor: Rahim Mahmoudvand

Abstract Mixed effects models are frequently used for analyzing longitudinal data. Normality assumption of random effects distribution is a routine assumption for these models, violation of which leads to model misspecification and misleading parameter estimates. We propose a semi-parametric approach using gradient function for random effect estimation. In the approach, we relax the normality assumption for random effects by estimating the random effects distribution over a pre-specified grid. Unknown parameters of the marginal model are estimated using maximum likelihood method. Some simulation studies and analyzing of a real data set are performed for illustration of the proposed semi-parametric method.

Keywords Directional derivative; Gradient function; Non-parametric approach; Longitudinal data; Mixed effects model.

AMS 2010 subject classifications 62Gxx, 62J99

DOI: 10.19139/soic.v2i4.38

1. Introduction

Longitudinal studies are frequently encountered in many areas such as biology, medical, epidemiology and social sciences. In these studies, individuals are followed over time and some repeated measurements for individuals are collected.

For analyzing longitudinal data, Gaussian mixed effects models [5] are commonly used. These models are flexible and widely applicable, in addition,

*Correspondence to: Department of Statistics, Faculty of Mathematical Sciences, Tarbiat Modares University, Tehran, Iran. Email: t.baghfalaki@gmail.com

many available software can be used for analyzing data using these models. Since, the normality assumption is often violated in practice, the use of multivariate t distribution [9], normal/independent family [6, 11], skew-normal distribution [7, 1] and other parametric distribution assumptions are proposed. Although the fixed effects may be robust to departure of normality assumption [14], the efficiency and validity of inference on random effects may destroy. The nature of random effects may be multimodal or skew and a non-appropriate distributional assumption for random effects may suggest exclusion of important covariate or may lead to unrealistic results. Moreover, less restrictive assumption for random effects distribution may provide considerable insights. The use of non-parametric mixed models [12, 10, 3, 15] is another suggestion when there is departure from the normality assumption.

In this paper, a new non-parametric approach for random effects estimation, based on directional derivative is proposed. The main idea of this approach is to search in each direction for maximizer of the log-likelihood function of the marginal model. Tsonaka et al. [13] apply vertex exchange method (VEM) to obtain semi-parametric maximum likelihood estimates of model parameters. Their approach uses the directional derivative of the log-likelihood function in a pre-specific grid and updates the weights that correspond to at most two grid points in the first step and maximize the marginal likelihood in the second step. These two steps are iterated until convergence takes place. One of the advantage of this approach is its simplicity and applicability in specification of random effects distribution. However, the speed of the algorithm is very slow. Our proposed approach modifies the approach of Tsonaka et al. [13] and increases the speed of it, specially in high dimensional random effects distribution. Similar to Tsonaka et al. (2009)'s algorithm [13], our approach uses the directional derivative of the individual log-likelihood function over a pre-specified grid. But the random effects are estimated by the corresponding value of grid points that maximize the directional derivative of the individual log-likelihood function. Weights of the grid points are estimated by counting the frequency of each grid points. Also, the marginal likelihood is maximized for parameter estimation. These steps are iterated until convergence is achieved.

The advantage of these approaches is that they use the directional derivative, which is an objective measure to evaluate whether the maximum has been reached. However, our approach is faster than the VEM method. The reason is that the weights are updated for all grid points in each iteration.

This paper is organized as follows. In Section 2, the semi-parametric mixed effects model is introduced. Section 3 includes some simulation studies for investigating the performance of the proposed method. Also, the model is used for analyzing a real data set in Section 4. The last section includes some conclusions.

2. The semi-parametric mixed effects model

2.1. Notation and model specification

Let \mathbf{Y}_i , $i = 1, 2, \dots, n$ denote the vector of n_i longitudinal measurements for the i^{th} individual such that $\mathbf{Y}_i = \{y_i(s_{ij}), j = 1, 2, \dots, n_i\}$, where $y_i(s_{ij}) = y_{ij}$ denotes the longitudinal measurement for the i^{th} subject at time s_{ij} . We consider the following linear mixed effect model for describing the longitudinal outcomes:

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

where \mathbf{Y}_i is the longitudinal vector of response variable for the i^{th} subject. $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})'$ is the vector of measurement errors, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a p -dimensional vector of longitudinal fixed-effect parameters. $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})'$ is a q -dimensional vector of random effects and is independent of $\boldsymbol{\varepsilon}_i$. $\mathbf{X}_i = (\mathbf{x}_i(s_{i1}), \dots, \mathbf{x}_i(s_{in_i}))'$ and $\mathbf{Z}_i = (\mathbf{z}_i(s_{i1}), \dots, \mathbf{z}_i(s_{in_i}))'$, where \mathbf{x} and \mathbf{z} are p -dimensional and q -dimensional vectors of explanatory variables, respectively. We assume that $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}, \boldsymbol{\Sigma}_i)$, where we assume that $\boldsymbol{\Sigma}_i = \sigma^2 I_{n_i}$, also let $\boldsymbol{\theta}_y = (\boldsymbol{\beta}', \sigma^2)$.

2.2. Details of the approach

In this paper, we do not make any parametric assumption for the random effects distribution, let \mathbf{b}_i be the random effects such that $\mathbf{b}_i \sim g(\mathbf{b}_i)$ where $g(\cdot)$ belongs to the set of all distribution functions on the parameter space of \mathbf{b}_i . The marginal distribution of \mathbf{Y}_i is given by

$$f(\mathbf{y}_i; \boldsymbol{\theta}_y) = \int f(\mathbf{y}_i|\mathbf{b}_i; \boldsymbol{\theta}_y)dG(\mathbf{b}_i). \tag{2}$$

The nonparametric maximum likelihood estimate of G is discrete with finite support [4, 8]. In the following, the left side of equation (2) is called $\ell_i(G)$ to ensure that the marginal distribution of \mathbf{Y}_i is obtained using G .

We assume that the random effects \mathbf{b}_i is a q -dimensional cube with elements $\boldsymbol{\mu} = \{\mu_{j_1 j_2 \dots j_q}\}$ for $j_1, j_2, \dots, j_q = 1, 2, \dots, C$ with corresponding weights $\boldsymbol{\pi} = \{\pi_{j_1 j_2 \dots j_q}\}$. The structure is a semi-parametric mixed effects model, note that the unknown parameters of this structure are $\boldsymbol{\theta}_y = (\boldsymbol{\beta}, \sigma^2)$.

The algorithm includes three steps. The first step consists of considering some initial values and the other two steps are weights updating and parameter estimations, respectively. The later two steps are repeated iteratively until convergence are achieved. The details of the steps are as follows:

Step 1: Initial values The algorithm needs an initial values for $\boldsymbol{\theta}_y$, say $\boldsymbol{\theta}_y^0$, and an initial value for the weights $\boldsymbol{\pi}$, say $\boldsymbol{\pi}^0$.

- A linear mixed effects model under normal distribution assumption for random effects is considered, the parameter estimates in this stage is used as initial values to have the vector θ_y^0 .
- The grid $\mu = \{\mu_{j_1 j_2 \dots j_q}\}$ for $j_1, j_2, \dots, j_q = 1, 2, \dots, C$ is kept fixed through the algorithm and only the corresponding weights are updated. For the q -dimensional grid with fixed points $\mu_{j_1 j_2 \dots j_q}$, the initial weights $\pi_{j_1 j_2 \dots j_q}^0 = 1/C^q$, $j_1, j_2, \dots, j_q = 1, 2, \dots, C$ are considered.

Step 2: Weight updating Let k be the iteration number. For each point $\mu_{j_1 j_2 \dots j_q}$, the directional derivative $D_i(\hat{G}^{(k)}, G_{\mu_{j_1 j_2 \dots j_q}})$ of $\ell_i(G)$ at $\hat{G}^{(k)}$ in the direction $G_{\mu_{j_1 j_2 \dots j_q}}$ is evaluated. This is given by

$$\begin{aligned}
 D_i(\hat{G}^{(k)}, G_{\mu_{j_1 j_2 \dots j_q}}) &= \frac{f(\mathbf{y}_i | G_{\mu_{j_1 j_2 \dots j_q}}, \theta_y^{(k)})}{f(\mathbf{y}_i | \hat{G}^{(k)}, \theta_y^{(k)})} \\
 &= \frac{f(\mathbf{y}_i | G_{\mu_{j_1 j_2 \dots j_q}}, \theta_y^{(k)})}{\sum_{j_1=1}^C \dots \sum_{j_q=1}^C \pi_{j_1 j_2 \dots j_q} f(\mathbf{y}_i | G_{\mu_{j_1 j_2 \dots j_q}}, \theta_y^{(k)})},
 \end{aligned}
 \tag{3}$$

where $\hat{G}^{(k)}$ is the current estimate of G and $G_{\mu_{j_1 j_2 \dots j_q}}$ is a degenerated distribution at point $\mu_{j_1 j_2 \dots j_q}$.

After directional derivative evolution $\hat{\mathbf{b}}_i = (b_{i1}, \dots, b_{iq})$, the estimated value of random effect for i^{th} individual, is defined as follows:

$$\hat{\mathbf{b}}_i = \operatorname{argmax}_{\mu} D_i(\hat{G}^{(k)}, G_{\mu}),$$

and the corresponding weight of $\pi_{j_1 j_2 \dots j_q}$ is estimated as the frequency of $\mu_{j_1 j_2 \dots j_q}$ in the estimated values of random effects. To ensure that the mean of the random effects is zero, model (1) can be written as

$$\mathbf{Y}_i = \mathbf{X}_i \beta + \mathbf{Z}_i (\hat{\mathbf{b}}_i - \boldsymbol{\pi}' \hat{\mathbf{b}}) + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Step 3: Let $\theta^{(k+1)} = (\beta^{(k+1)}, \sigma^{2(k+1)})'$ is the vector of parameters of $k + 1^{th}$ step, where they are estimated in this step. The marginal distribution is given by $\ell_i(G) = \sum_{j_1=1}^C \dots \sum_{j_q=1}^C \pi_{j_1 j_2 \dots j_q} f(\mathbf{y}_i | G_{\mu_{j_1 j_2 \dots j_q}}, \theta^{(k)})$ and the marginal likelihood is $\ell(G) = \sum_{i=1}^N \ell_i(G)$. The model parameters θ_y are estimated using quasi-Newton method. Note that $\boldsymbol{\pi} = \{\pi_{j_1 j_2 \dots j_q}\}$ and $\mu = \{\mu_{j_1 j_2 \dots j_q}\}$ for $j_1, j_2, \dots, j_q = 1, 2, \dots, C$ in this step are considered as fixed parameters.

The two later steps are repeated iteratively until convergence. The algorithm has converged when $|\hat{\theta}^{k+1} - \hat{\theta}^k| < \varepsilon$, where $|\cdot|$ is a distance measure.

3. Simulation studies

For investigating the proposed method, we perform some simulation studies, 1) with univariate random effect as random intercept, 2) with bivariate random effect as random intercept and random slope.

In the first scenario, the random intercept is sampled from three different distributions: a normal distribution, a uniform distribution and two components normal mixture. Also in the second scenario, the random intercept and the random slope are sampled from three different bivariate distributions: bivariate normal distribution, bivariate uniform distribution and a two components bivariate normal mixture. The details of these simulation studies are given in the following.

3.1. Univariate random effects

The data are simulated from the following linear mixed effect model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + b_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, 3$$

where $\beta_0 = 5$, $\beta_1 = 1$ and $\beta_2 = 1$, $N = 300$, $t_{i1} = 2$, $t_{i2} = 6$. x_i s are generated from a Bernoulli distribution with success probability 0.2 and $\varepsilon_{ij} \sim N(0, \sigma^2)$ where $\sigma^2 = 1$. The random intercept, b_i , is generated from three different distributional assumption: i) a normal distribution with variance $\sigma_b^2 = 4$, ii) a uniform distribution on $[-4, 4]$ and iii) a mixture of two normal distributions, i.e., $0.5 N(-1, 1) + 0.5 N(1, 1)$. Results of this simulation study are presented in Table 1. This table contains estimated value of parameters, standard errors, relative biases and root of mean square errors. The two later criteria are defined as

$$Rel.Bias(\theta) = \frac{1}{N} \sum_{i=1}^N \left(\frac{\hat{\theta}_i}{\theta} - 1 \right), \quad MSE(\theta) = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2$$

where $\hat{\theta}_i$ is the estimated value of θ for the i^{th} sample. In this simulation study, we consider a sequence with length 100 in interval $[-4, 4]$. This table shows that in all scenarios the estimated values of parameters, standard deviations, relative biases and root of MSEs confirm the well performance of the proposed method. Also, figure 1 shows the estimated weights of randomly selected samples for all scenarios, this figure shows that the estimated weights are similar to density plot of the real value: the estimated weights for the normal random intercept (panel a) have unimodal and symmetric form, the estimated weights for the uniform distribution have approximately equal weights at all points of the region. Also, panel (c) shows a bimodal distribution. To ensure zero-mean random effect model, we also consider the following model:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + \hat{b}_i - \hat{\boldsymbol{\pi}}' \hat{\mathbf{b}} + \varepsilon_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, n_i,$$

Table I. Results of simulation studies for mixed model with univariate random effects

| Normal Model | | | | | |
|---------------|-------|-------|-------|-----------|-------|
| Parameter | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.987 | 0.167 | -0.002 | 0.137 |
| β_1 | 1.000 | 0.999 | 0.012 | -0.001 | 0.009 |
| β_2 | 1.000 | 1.021 | 0.407 | 0.021 | 0.300 |
| σ^2 | 1.000 | 1.028 | 0.035 | 0.0280 | 0.036 |
| σ_b^2 | 4.000 | 3.968 | 0.323 | -0.010 | 0.252 |
| Uniform Model | | | | | |
| | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.977 | 0.143 | -0.004 | 0.106 |
| β_1 | 1.000 | 0.999 | 0.013 | -0.001 | 0.010 |
| β_2 | 1.000 | 1.003 | 0.259 | 0.004 | 0.196 |
| σ^2 | 1.000 | 0.996 | 0.031 | -0.004 | 0.023 |
| σ_b^2 | 5.333 | 5.403 | 0.288 | 0.018 | 0.236 |
| Mixture Model | | | | | |
| | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.996 | 0.098 | -0.001 | 0.076 |
| β_1 | 1.000 | 1.002 | 0.013 | 0.001 | 0.010 |
| β_2 | 1.000 | 1.000 | 0.118 | 0.000 | 0.093 |
| σ^2 | 1.000 | 0.972 | 0.029 | -0.027 | 0.033 |
| σ_b^2 | 1.500 | 1.377 | 0.084 | 0.021 | 0.032 |

where $\hat{\mu}_i$ is the estimated value of random intercept for i^{th} individual and $\hat{\pi}$ and $\hat{\mu}$ are the vector of estimated weights and the vector of grid points, respectively. One of the advantages of the proposed method is its ability for considering wide intervals with favorite grid points, but a grid can be chosen for the scaled random effects $\mathbf{b}_i^* = \hat{S}_b^{-1} \mathbf{b}_i$ instead of b_i , where \hat{S}_b is an approximation of the random effects covariance matrix of \mathbf{b}_i (similar to that of Tsonaka et al., 2009). For comparison of the results of the proposed method and the Gaussian model, we also analyze the generated data set using the pure normal model.

3.2. Bivariate random effects

The data are generated from the following linear mixed effects model with random intercept and random slope:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 x_i + b_{0i} + b_{1i} t_j + \varepsilon_{ij}, \quad i = 1, 2, \dots, N, \quad j = 1, 2, 3$$

where $\beta_0 = 5$, $\beta_1 = 1$ and $\beta_2 = 1$, $N = 300$, $t_{i1} = 2$, $t_{i2} = 6$, x_i s are generated from Bernoulli distribution with success probability 0.2 and $\varepsilon_{ij} \sim N(0, \sigma^2)$ where $\sigma^2 = 1$. The vector of random effects $\mathbf{b}_i = (b_{0i}, b_{1i})'$ is generated from three different distributional assumptions: i) a standardized bivariate normal distribution, ii) a bivariate uniform distribution on $[-4, 4]^2$, iii) a mixture of two bivariate normal distribution with mean $(-1, -1)'$ and $(1, 1)'$ and an identity covariance matrix. The results of this simulation study are presented in Table 2. This table shows the well performance of proposed method. Also, Figure 2

Table II. Results of simulation studies for mixed model with bivariate random effects

| Normal Model | | | | | |
|---------------|-------|--------|-------|-----------|-------|
| Parameter | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.997 | 0.114 | -0.001 | 0.091 |
| β_1 | 1.000 | 1.001 | 0.079 | 0.001 | 0.064 |
| β_2 | 1.000 | 1.002 | 0.027 | 0.002 | 0.022 |
| σ^2 | 1.000 | 0.995 | 0.195 | -0.002 | 0.135 |
| d_{11} | 1.000 | 1.364 | 0.150 | 0.363 | 0.366 |
| d_{12} | 0.000 | -0.080 | 0.114 | * | 0.114 |
| d_{22} | 1.000 | 1.011 | 0.090 | 0.011 | 0.071 |
| Uniform Model | | | | | |
| Parameter | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.993 | 0.125 | 0.002 | 0.060 |
| β_1 | 1.000 | 0.985 | 0.111 | -0.045 | 0.107 |
| β_2 | 1.000 | 1.001 | 0.029 | 0.030 | 0.036 |
| σ^2 | 1.000 | 0.956 | 0.087 | -0.058 | 0.150 |
| d_{11} | 5.333 | 5.550 | 0.282 | 0.024 | 0.289 |
| d_{12} | 0.000 | -0.032 | 0.303 | * | 0.187 |
| d_{22} | 5.333 | 5.341 | 0.239 | -0.028 | 0.256 |
| Mixture Model | | | | | |
| Parameter | real | Est. | S.E. | Rel. bias | RMSE |
| β_0 | 5.000 | 4.997 | 0.112 | -0.000 | 0.085 |
| β_1 | 1.000 | 0.994 | 0.077 | -0.005 | 0.062 |
| β_2 | 1.000 | 1.008 | 0.025 | 0.008 | 0.022 |
| σ^2 | 1.000 | 0.924 | 0.094 | -0.075 | 0.103 |
| d_{11} | 1.500 | 0.352 | 0.127 | -0.294 | 0.150 |
| d_{12} | 0.500 | 1.728 | 0.106 | 0.118 | 0.228 |
| d_{22} | 1.500 | 1.522 | 0.109 | 0.014 | 0.089 |

“*”: This is not given, because the real value is 0.000.”

shows estimated values of the weights for grid points under different distribution assumptions for random intercept and random slope.

4. Application

We consider a longitudinal study on 467 HIV infected patients who had failed or were intolerant of zidovudine (AZT) therapy. The aim of the study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). Patients were randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry, when randomization took place, as well as at 2, 6, 12 and 18 months thereafter.

The distribution of CD4 cell counts is right skewed, so we analyze the square root of the CD4 cell counts. Figure 3 shows $\sqrt{CD4}$ trajectories for fifty randomly selected individuals given each drug. Panels of this figure show a sharply increasing degree of missing data over time due to death, dropout, and missed clinic visits. In this figure the profiles of those individuals who remain and those

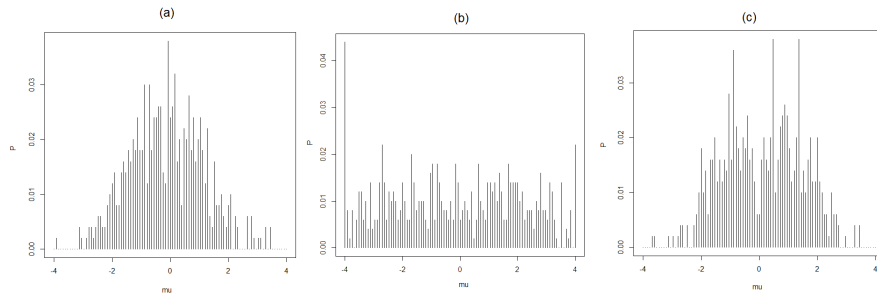


Figure 1. Estimated values of the weights for grid points under different distribution assumptions for random intercepts. (a) normal, (b) uniform, (c) mixture of two normals.

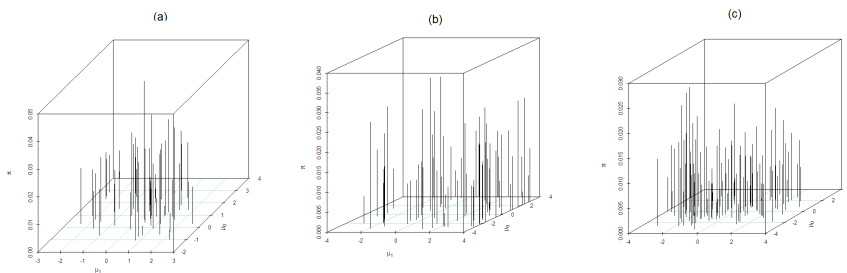


Figure 2. Estimated values of the weights for grid points under different distribution assumptions for random effects. (a) bivariate normal, (b) bivariate uniform, (c) mixture of two bivariate normals.

of individuals who do not remain are indicated using different colors. The missing values belong to people who leave the study for different reasons. This figure underlines that those who do not remain had smaller $\sqrt{CD4}$ than others. More details about this data set can be found in [2].

The following longitudinal model with random intercept is:

$$y_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij} Drug_i + \beta_3 Sex_i + \beta_4 PrevOI_i + \beta_5 Stratum_i + b_i + \sigma \varepsilon_{ij}, \tag{4}$$

where, $\varepsilon_{ij} \sim N(0, 1)$ and $b_i = \mu_i - \pi' \mu$.

In this model, y_{ij} is the square root of the j^{th} CD4 count measurement on the i^{th} individual in the trial, $j : 1, 2, \dots, 5$ and $i : 1, 2, \dots, 467$. Sex_i is a gender indicator (0=female, 1=male). The other three explanatory variables are $Drug_i$ (0=ddC, 1=

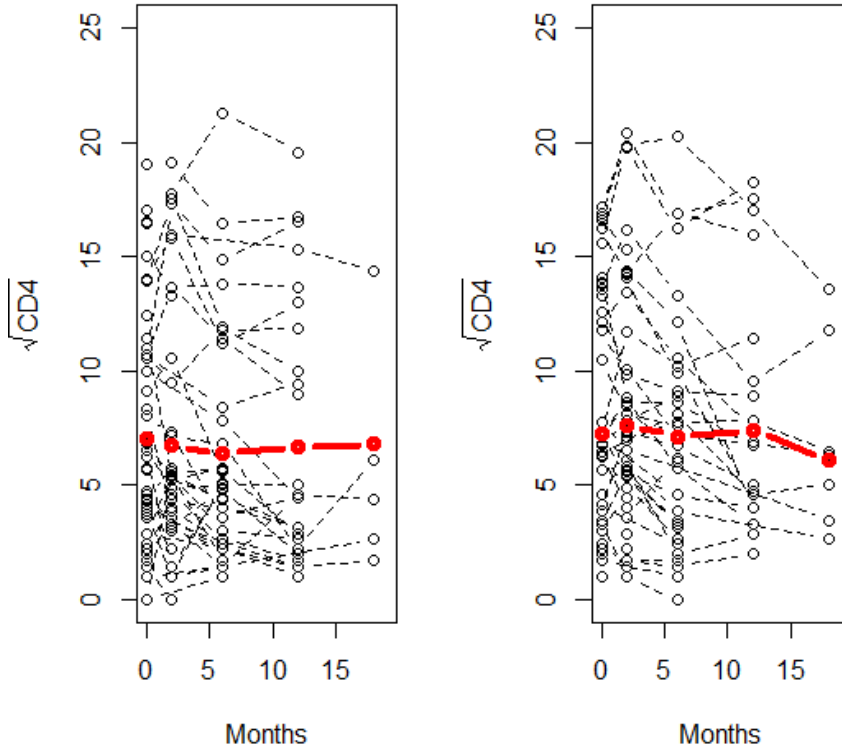


Figure 3. Profiles of $\sqrt{CD4}$ measurements over time for fifty randomly selected individuals from each drug, bold red lines are mean profile for all observed individuals on each drug.

ddi), $PrevOI_i$, previous opportunistic infection (1=AIDS diagnosis, 0=no AIDS diagnosis) and $Stratum_i$ (1=AZT failure, 0=AZT intolerance).

Also, the following longitudinal model with random intercept and random slope is used for analyzing the data set:

$$\begin{aligned}
 y_{ij} &= \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij} Drug_i + \beta_3 Sex_i \\
 &+ \beta_4 PrevOI_i + \beta_5 Stratum_i + b_{0i} + b_{1i} t_{ij} + \sigma \varepsilon_{ij},
 \end{aligned}
 \tag{5}$$

where, $\varepsilon_{ij} \sim N(0, 1)$, $b_{0i} = \mu_{0i} - \pi'_{+1} \mu_0$ and $b_{1i} = \mu_{1i} - \pi'_{+1} \mu_1$.

We consider the proposed method with a grid defined on $[-15, 15]^l$, $l = 1, 2$ and $C = 450$ (we begin with a grid for random effects in $[-5, 5]^l$, $l = 1, 2$, the results show that a wider grid is needed, therefore we refit the model with wider grid). Also, a sensitivity analysis of the results with respect to different value of C for random effects is performed and the results are not sensitive with respect to this change.

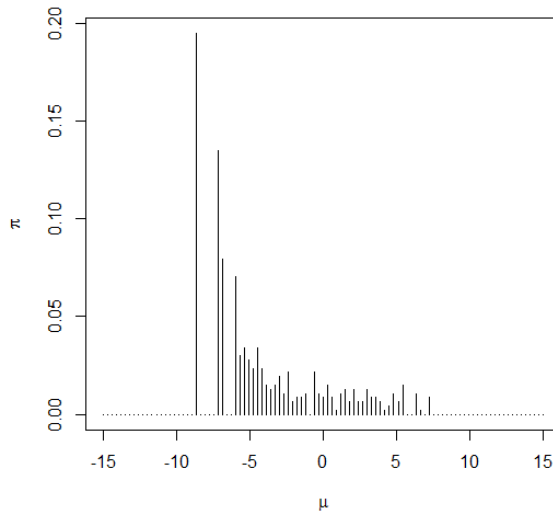


Figure 4. The estimated weights for the univariate random effects for HIV data set.

The results of these models using the proposed approach and normal mixed effects model are summarized in Table 3. In this table, models are indicated by notations SPA and GM, where SPA is used for the proposed semi-parametric approach and GM is the ordinary Gaussian model. The standard deviation of parameters in semi-parametric approach are calculated using Bootstrap approach [13]. The results are compared using AIC (Akaike information criterion), BIC (Bayesian information criterion), HQC (Hannan-Quinn criterion). Let Θ and $\mathbf{Z} = (z_1, \dots, z_N)'$ be the entire model parameters and data, respectively, then $AIC = -2\ell(\hat{\Theta}|\mathbf{Z}) + 2\alpha$, $BIC = -2\ell(\hat{\Theta}|\mathbf{Z}) + \alpha \ln(N)$ and $HQC = -2\ell(\hat{\Theta}|\mathbf{Z}) + 2 \ln(\ln(N))$, where α is the number of unknown parameters and $\hat{\Theta}$ is the value of the parameter estimates at convergence. The smaller the AIC (BIC or HQC), the better the fit of the model. The results show that the semi-parametric model with univariate random effects is the best fitted model for HIV data set. This table shows that time and previous opportunistic

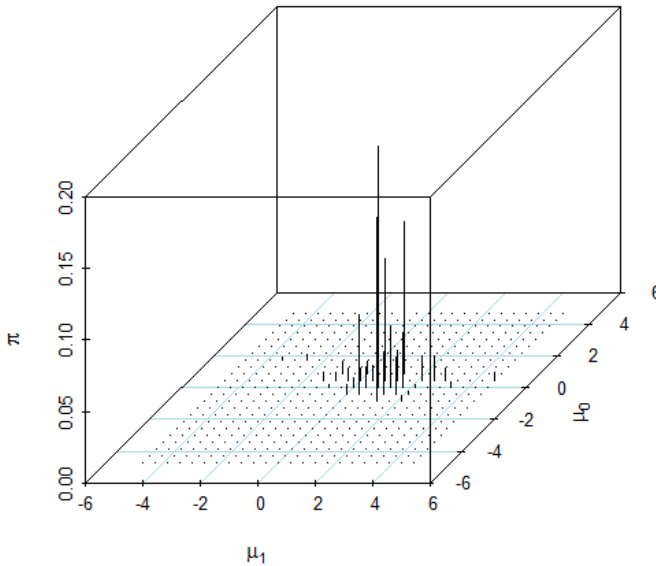


Figure 5. The estimated weights for the bivariate random effects for HIV data set.

infection are significant predictors in the model, such that the increase in time reduces CD4 count. Also, previous opportunistic infection leads to decreasing CD4 count. Figures 3 and 4, respectively, show the estimated weights for the univariate and bivariate random effects for HIV data set. These figures present the weights of the grid points in the semi-parametric fitted models.

5. Conclusion

In this paper, we have developed a semi-parametric mixed effects model for analyzing longitudinal data. The approach uses the directional derivative of each individual for finding the random effects estimation under any distributional assumption and leads to reliable parameter estimates.

The main advantage of this approach is in the use of the directional derivative of the individual log-likelihood. The method uses the relative frequency of grid points as an estimate of their weights, therefore, the speed of the algorithm increases.

Table III. Parameter estimates and standard deviations for the HIV data set. SPM: semi-parametric mixed effect model, PM: parametric model.

| model | Univariate random effects | | Bivariate random effects | |
|-------------------------------------|---------------------------|---------------|--------------------------|---------------|
| | PM | SPA | PA | SPA |
| parameters | Est. (S.E.) | Est. (S.E.) | Est. (S.E.) | Est. (S.E.) |
| Intercept (β_{11}) | 7.983(0.361) | 12.782(1.480) | 8.055(0.350) | 10.312(1.701) |
| Time (β_{12}) | -0.159(0.016) | -0.224(0.028) | -0.199(0.048) | -0.241(0.048) |
| Time \times Drug (β_{13}) | 0.019(0.023) | 0.045(0.035) | 0.055(0.071) | 0.053(0.070) |
| Gender (β_{14}) | -0.149(0.341) | 0.470(0.683) | -0.168(0.326) | -0.201(0.322) |
| PrevOI (β_{15}) | -2.285(0.239) | -1.576(0.598) | -2.327(0.242) | -3.727(0.237) |
| Stratum (β_{16}) | -0.156(0.231) | -0.141(0.234) | -0.105(0.229) | -0.114(0.223) |
| σ^2 | 3.846(0.175) | 1.979(0.074) | 2.904(0.263) | 5.878(0.161) |
| d_{11} | 15.288(1.107) | 17.603(1.405) | 15.619(1.171) | 4.473(0.317) |
| d_{12} | – | – | -0.281(0.139) | -0.078(0.258) |
| d_{22} | – | – | 0.391(0.030) | 0.891(0.956) |
| AIC | 7005.616 | 7111.728 | 7409.634 | 7287.332 |
| BIC | 7034.640 | 7144.952 | 7438.658 | 7329.356 |
| HQC | 6995.248 | 7099.360 | 7399.266 | 7270.964 |

The proposed approach can be easily extended to shared random effects and be used for joint modeling of longitudinal measurements and event time. In these models the association between two process is captured by latent random effects.

REFERENCES

1. Ganjali, M., Baghfalaki, T., and Khazaei, M. (2013). A linear mixed model for analyzing longitudinal skew-normal responses with random dropout. *Journal of the Korean Statistical Society*, **42**(2), 149-160.
2. Goldman, A. I., Carlin, B. P., Crane, L. R., Launer, C., Korvick, J. A., Deyton, L., and Abrams, D. I. (1996). Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology* **11**, 161-169.
3. Ishwaran, H., and Takahara, G. (2002). Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models. *Journal of American Statistical Association*, **97**, 1154-1166.
4. Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, **73**, 805-811.
5. Laird, N. M., and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963-974.
6. Lange, K. L., and J. S. Sinsheimer (1993). Normal/independent distributions and their applications in robust regression. *Journal of the American Statistical Association*, **2**:175-198.
7. Lin, T., and Lee, J. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. *Statistics in Medicine*, **27**, 1490-1507.
8. Lindsay, B. G. (1983). The geometry of mixture likelihoods: A general theory. *The Annals of Statistics*, **11**, 86-94.
9. Pinheiro, J. C., Liu, C. H., and Wu, Y. N. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. *Journal of Computational and Graphical Statistics*, **10**, 249-276.
10. Rice, J. A., and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253-259.

11. Rosa, G. J. M., Padovani, C. R. and Gianola, D. (2003). Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation. *Biometrical Journal*, **45**, 573-590.
12. Shi, M., Weiss, R. E., Taylor, J. M. G. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics*, **45**, 151-163.
13. Tsonaka, R., Verbeke, G., and Lesaffre, E. A. (2009). Semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, **65**, 81-87.
14. Verbeke, G., and Lesaffre, E. (1997). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis*, **23**, 541-556.
15. Zhang, D., and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, **57**, 795-802.