Bayesian Variable Selection and Estimation in Semiparametric Simplex Mixed-Effects Models with Longitudinal Proportional Data

Anmin Tang 1, Xingde Duan 2,* and Yuanying Zhao 3

1 Yunnan Key Laboratory of Statistical Modeling and Data Analysis, Yunnan University, Kunming 650091, China
2 Department of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China
3 College of Mathematics and Information Science, Guiyang University, Guiyang 550005, China

* Correspondence: duanxingde@mail.gufe.edu.cn; Tel.: +86-851-8860-7821

Abstract: In the development of simplex mixed-effects models, random effects in these mixed-effects models are generally distributed in normal distribution. The normality assumption may be violated in an analysis of skewed and multimodal longitudinal data. In this paper, we adopt the centered Dirichlet process mixture model (CDPMM) to specify the random effects in the simplex mixed-effects models. Combining the block Gibbs sampler and the Metropolis–Hastings algorithm, we extend a Bayesian Lasso (BLasso) to simultaneously estimate unknown parameters of interest and select important covariates with nonzero effects in semiparametric simplex mixed-effects models. Several simulation studies and a real example are employed to illustrate the proposed methodologies.

Keywords: simplex distribution; Gibbs sampler; Metropolis–Hastings algorithm; Dirichlet process prior; Bayesian Lasso

1. Introduction

Various mixed-effects models based on simplex distribution have increasingly become popular tools in the analysis of longitudinal continuous proportional data over time in many biological, medical and clinical studies. Under the framework of generalized linear mixed models, see Qiu et al. [1] for information on developing a simplex generalized linear mixed model on the basis of the penalized quasi-likelihood (PQL) and restricted maximum likelihood (REML) inference; see Zhang and Wei [2] for information on using the maximum likelihood estimation combining the stochastic approximation (SA) algorithm and the MCMC method to infer on simplex distribution nonlinear mixed models; see Zhao et al. [3] for information on implementing the MCMC algorithm to obtain the joint Bayesian estimate of simplex distribution nonlinear mixed models from the Bayesian perspective; see Bonat et al. [4] for information on investigating the likelihood analysis for a class of simplex mixed models with logit, probit, complement log–log and Cauchy link functions; see Quintero [5] for information on presenting the sensitivity analysis for variance parameters of random effects in Bayesian simplex mixed models. The random effects in the abovementioned mixed-effects models are assumed to have a multivariate normal distribution. However, in some practical applications, it is questionable for the normal assumption for random effects to analyze the skewed, bimodal and heavy-tailed longitudinal data. Therefore, it is essential to incorporate a semiparametric hierarchical structure via a Dirichlet process prior distribution for the random effects into the simplex mixed-effects models to accommodate longitudinal proportional data.

The nonparametric Bayesian approach based on Dirichlet process (DP) prior for random effects in mixed-effects models has been receiving a lot of attention in recent years.
For example, Kleinman and Ibrahim [6] used a Dirichlet process prior for the general distribution of the random effects in generalized linear mixed model. As a variant of Dirichlet process prior, the truncation approximation Dirichlet process with stick-breaking priors is widely incorporated into various mixed-effects models to specify the general distribution of random effects. For example, Tang and Duan [7] used this approach for a semiparametric Bayesian approach to generalized partial linear mixed model; Tang and Zhao [8] used this approach for nonlinear reproductive dispersion mixed models; Zhao et al. [9] used this approach for a semiparametric Bayesian approach to binomial distribution logistic mixed-effects model. In particular, Duan et al. [10] used a truncated and centered Dirichlet process prior to specify random effects in semiparametric reproductive dispersion mixed model. However, the abovementioned DP with stick-breaking prior for random effects is inappropriate when the underlying density of random effects is continuous. In addition, this type of variant for Dirichlet process prior is rather time-consuming in the calculation process for complicated models. Therefore, to address the above issues, the goal of this paper is to propose a new semiparametric simplex mixed-effects models with the random effects distribution specified by the centered Dirichlet process mixture model (CDPMM).

Although various methodologies have been developed to make statistical inference on the aforementioned simplex mixed-effects models, little work has been performed for the variable selection of simplex mixed-effects models. Classical model-selection methods, such as the step-wise selection method [11], the model comparison via Bayes factor [12], the Akaike information criterion [13] and Deviance information criterion [14], are often used to identify the important covariates in regression analysis; however, these approaches are generally computationally intensive and unstable for complicated models with many covariates. On the other hand, the regularization (penalization) method has increasingly become a popular tool for conducting variable selection in regression analysis. Commonly used regularization methods in the context of linear regression include least absolute shrinkage and selection operator (Lasso) [15], elastic net [16] and adaptive lasso [17]. In addition, Park and Casella [18] proposed the Bayesian version of the Lasso (BLasso) by assigning the conditional Laplace prior of regression coefficients and the gamma distribution of shrinkage parameter under the Bayesian framework. The BLasso procedure has been extended to various complex models including semiparametric structural equation models [19] and semiparametric joint models of multivariate longitudinal and survival data [20]. In particular, Erd et al. [21] pointed out that Bayesian penalization methods perform similarly or sometimes even better than frequentist penalization methods, since Bayesian penalization methods can easily provide credible intervals (CIs) for parameters of interest and obtain the estimate of the penalty parameter by assigning an appropriate prior distribution. Therefore, the other main purpose of this paper is to extend the BLasso procedure to the considered semiparametric simplex mixed-effects models.

The paper is organized as follows: In Section 2, we propose a new semiparametric simplex mixed-effects models with random effects following the centered Dirichlet process mixture model (CDPMM) and incorporate a BLasso procedure into the proposed simplex mixed-effects models. The required conditional distributions are derived in Section 3. Two simulation studies and a real example are used to illustrate the proposed methodologies in Section 4. Some concluding remarks are given in Section 5.

2. Model and Notation

The simplex distribution was firstly proposed by Barndorff-Nielsen and Jørgensen [22], whose probability density function is specified as

\[
p(y; \mu, \sigma^2) = \begin{cases} 
2\pi\sigma^2 (y(1-y))^3)^{-1/2} \exp\left(-\frac{d(y, \mu)}{2\sigma^2}\right), & \text{if } 0 < y < 1 \\
0, & \text{otherwise}
\end{cases}
\]  

(1)
where \( \mu \in (0, 1) \) denotes the mean parameter; \( \sigma^2 > 0 \) represents the dispersion parameter; and \( d(y; \mu) = \frac{(y-\mu)^2}{y(1-y)e^{\mu(1-\mu)}} \). For simplicity of notation, we denote \( y \sim S^-(\mu, \sigma^2) \) if a random variable, \( y_i \), is distributed as a simplex distribution with mean parameter, \( \mu \), and dispersion parameter, \( \sigma^2 \), in the rest of this paper.

In the context of longitudinal data analysis, let \( y_{ij} \) denote the longitudinal percentage outcome for the \( i \)th individual at the \( j \)th follow-up time \( t_{ij} \), and \( 0 < y_{ij} < 1, i = 1, \ldots, n, j = 1, \ldots, n_i \). We assume that, given a \( q \times 1 \) random effects \( b_i \) corresponding to the \( i \)th individual, the responses \( y_{ij} \) are conditionally independent and each \( y_{ij} | b_i \) is distributed as a simplex distribution with conditional means, \( \mu_{ij} = E(y_{ij} | b_i) \), and constant dispersion parameter, \( \sigma^2 \): that is, \( y_{ij} | b_i \sim S^-(\mu_{ij}, \sigma^2) \). Under the framework of GLMM, the conditional mean is linked to explanatory variables and random effects as follows:

\[
f(y_{ij}) = \eta_{ij} = x_{ij}^T \beta + z_{ij}^T b_i,
\]

where an unknown and monotone link function \( f(\cdot) \) is chosen as the logit link; \( x_{ij} \) is a \( (p + 1) \times 1 \) vector of covariates which consist of the constant 1 and time-dependent covariates observed at time point \( t_{ij} \); \( b_i \) is a \( (p + 1) \times 1 \) vector of unknown regression parameters; \( z_{ij} \) is a \( q \times 1 \) vector of time-dependent variables which may include some elements of \( x_{ij} \) corresponding to random effects \( b_i \). In classical random-effects models, the random effects in (2) are generally assumed to be a multivariate normal distribution, which may give rise to biased estimates of parameters or even misleading conclusions. Thus, inspired by Ohlssen and Spiegelhalter [23], we used the DP mixture of normals to specify the random effects: that is, \( b_i \overset{i.i.d.}{\sim} \sum_{g=1}^{\infty} \pi_g N_q(\mu_g, \Omega_g) \) with \( (\mu_g, \Omega_g) \sim \mathcal{P} \), where \( \mathcal{P} \) is an unknown random probability. Clearly, it is rather difficult and inefficient to make Bayesian estimates for regression parameter \( \beta \) and dispersion parameter \( \sigma^2 \) in Equation (2) since an unknown form of \( \mathcal{P} \) is involved. To address the difficulty, the Dirichlet process (DP) prior is usually introduced to approximate \( \mathcal{P} \), i.e., \( \mathcal{P} \sim DP(\tau F_0) \), in which \( F_0 \) is a given base distribution such as multivariate normal distribution that serves as a starting point for constructing the nonparametric distribution, and \( \tau \) is a weight that indicates the researcher’s certainty of \( F_0 \) as the distribution of \( \mathcal{P} \). In particular, Sethuraman [24] showed that the DP prior \( DP(\tau F_0) \) has the stick-breaking prior representation; however, this approach causes a nonzero mean of random effects [25] and a discrete probability distribution of random effects [23]. Generally, the variants of Dirichlet Process proposed by Ishwaran and Zarepour [26] and Yang et al. [25] were regarded as discrete Dirichlet processes (discrete DPs). A discrete DP with stick-breaking prior for random effects is inappropriate when the underlying density of random effects is continuous. Furthermore, violation of zero mean assumption on the random effects may lead to non-identifiability in the aforementioned random effects model. In addition, the discrete DP methods with stick-breaking prior for random effects are generally computationally intensive for the complicated models.

To overcome the above issues, inspired by Ohlssen and Spiegelhalter [23] and Yang et al. [25], we incorporated the following variant of Dirichlet process into the above model in (2) to specify random effects. That is,

\[
b_i \overset{i.i.d.}{\sim} \sum_{g=1}^{\infty} \pi_g N_q(\mu_g, \Omega_g) \quad \text{with} \quad \mu_g = \mu_g' - \sum_{g=1}^{\infty} \pi_g \mu_g' \quad \text{and} \quad (\mu_g', \Omega_g') \overset{i.i.d.}{\sim} F_0,
\]

where \( \pi_g \) is a random probability weight satisfying \( 0 \leq \pi_g \leq 1 \) and \( \sum_{g=1}^{\infty} \pi_g = 1 \). In addition, \( \pi_g \) is assumed to be be independent of \( (\mu_g', \Omega_g') \). This variant of Dirichlet process is referred to as the centered Dirichlet process mixture model (CDPMM). As in Ishwaran
and Zarepour [26], we adopt the following mixture model of the truncated approximation DP for $\mathcal{P}$:

$$b_i \sim \sum_{g=1}^{G} \pi_g N_q(\mu_g, \Omega_g)$$

with $\mu_g = \mu_g^* - \sum_{g=1}^{G} \pi_g \mu_g^*$ and $(\mu_g^*, \Omega_g) \sim \text{F}_0$.

where $G$ is a limited integer satisfying $1 \leq G < \infty$. As for the selection of $G$, Ishwaran and Zarepour [26] pointed out that a moderate value of $G$ such as 25 may be enough to capture a good approximation in practical application. Thus, the value of $G$ is chosen to be 25 in the rest of this paper. Furthermore, the random probability weight, $\pi_g$, is specified by the following stick-breaking procedure:

$$\pi_1 = \theta_1 \text{ and } \pi_g = \theta_g \prod_{i=1}^{g-1} (1 - \theta_i) \text{ for } g = 2, \ldots, G,$$

where $\theta_g \sim \text{Beta}(1, \tau)$ for $g = 1, \ldots, G - 1$, and $\theta_G = 1$ so that $\sum_{g=1}^{G} \pi_g = 1$. The prior distribution for the unknown parameter $\tau$ is chosen as $\tau \sim \Gamma(a_1, a_2)$, such that the posterior distribution for $\tau$ is conjugated. Here, we set the hyperparameters $a_1$ and $a_2$ to be 25 and 5, respectively, such that large value of $\tau$ is generated, which results in more unique $b_i$ values.

It is rather difficult and inefficient to generate observations from posterior distributions of $b_i$ with the above DP prior via MCMC algorithm. Furthermore, a latent variable $L_i \in \{1, \ldots, G\}$ is introduced to solve sample issue since this latent variable can record each $b_i$'s cluster membership and convey its parametric value to the distribution of $b_i$. Let $L = \{L_1, \ldots, L_q\}$, $\pi = \{\pi_1, \ldots, \pi_G\}$, $\mu = \{\mu_1^*, \ldots, \mu_G^*\}$ and $\Omega = \{\Omega_1, \ldots, \Omega_G\}$, in which $\Omega_g = \text{diag}(\omega_g 1, \ldots, \omega_g q)$ for $g = 1, \ldots, G$. As in Ishwaran and Zarepour [26], the hierarchical structure defined in (4) can be written as

$$L_i|\pi \sim \sum_{g=1}^{G} \pi_g \delta_g(\cdot) \text{ and } (\pi, \mu^*, \Omega) \sim f_1(\pi) f_2(\mu^*) f_3(\Omega),$$

where $\delta_g(\cdot)$ denotes a discrete probability measure concentrated at $g$, $f_1(\pi)$ is defined in Equation (5), the prior for $\mu_g^*$ associated with $f_2(\mu^*) = \prod_{g=1}^{G} f_2(\mu_g^*)$ is defined by

$$\mu_g^*|\nu, \Psi \sim N_q(\xi, \Psi), \quad \xi = \mathbf{0}, \Psi = \mathbf{0}, \quad q \sim N_q(\mathbf{0}, \Psi^0), \quad \psi_j \sim \Gamma(c_1, c_2) \text{ for } j = 1, \ldots, q,$$

and the prior for $\omega_g$ related to $f_3(\Omega) = \prod_{g=1}^{G} \prod_{j=1}^{q} f_3(\omega_g)$ is defined by

$$\omega_g^{-1}|\omega_a^a, \omega_j^a \sim \Gamma(\omega_a^a, \omega_j^a) \text{ and } \omega_j^a|\omega_j^b, \omega_j^b \sim \Gamma(\omega_j^a, \omega_j^b),$$

where $\Psi = \text{diag}(\psi_1, \ldots, \psi_q), \Gamma(c_1, c_2)$ denotes the Gamma distribution with parameters $c_1$ and $c_2$, and $\mathbf{0}, \Psi^0, c_1, c_2, \omega_a^a, \omega_j^a$ and $\omega_j^b$ are pre-specified hyperparameters: that is, $\nu = 0_{q \times 1}, \Psi^0 = I_q, c_1 = 11, c_2 = 2.5, \omega_a^a = 3, \omega_j^a = n$ and $\omega_j^b = 10$. Thus, given the values of $L_i$, $\mu^*$ and $\Omega$, the prior for random effect $b_i$ is assumed to be $N_q(\mu_{L_i}, \Omega_{L_i})$ with $\mu_{L_i} = \mu_{L_i}^* - \sum_{g=1}^{G} \pi_g \mu_g^*$.

To estimate the unknown parameters $\beta$ and $\sigma^2$ in Equation (2) from the Bayesian perspective, it is necessary to specify priors for $\beta$ and $\sigma^2$. In order to alleviate the computational burden, the conjugate prior distribution for dispersion parameter $\sigma^2$ is taken to be

$$\sigma^{-2} \sim \Gamma(\sigma_a^2, \sigma_b^2),$$

where the values of hyperparameters $\sigma_a^2$ and $\sigma_b^2$ are taken to be 1 and 0.01, respectively. In this paper, the main goal is to incorporate the Bayesian version of lasso into our proposed
model (2) to conduct parameter estimation and model selection simultaneously. Similar to Park and Casella [18] and Tang et al. [20], the following Laplace prior on $\beta$ is given by

$$\pi(\beta) = \prod_{k=0}^{p} \frac{1}{2} \exp(-\nu |\beta_k|),$$

where $\nu$ is the regularization parameter. Because the mass of the above presented Laplace prior is quite highly concentrated around zero with a distinct peak at zero, posterior means or modes of $\beta_k$'s are shrunk towards zero, which is the key principle in using BLasso method to select the important covariates. Following Robert [15], the Laplace distribution with the form $a \exp(-a|x|)/2$ can be represented as a scale mixture of normal distributions with independent exponentially distributed variance: that is,

$$\frac{a}{2} \exp(-a|x|) = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi u}} \exp\left(-\frac{x^2}{2u}\right) \frac{a^2}{2} \exp\left(-\frac{a^2 u}{2}\right) du, \text{ for } a > 0.$$

Therefore, the aforementioned prior for $\beta$ can be reformulated as the following hierarchical structure:

$$\beta | H_\beta \sim N_p(0, H_\beta) \quad \text{with} \quad H_\beta = \text{diag}(h_{\beta_0}^2, \ldots, h_{\beta_p}^2),$$

$$h_{\beta_0}^2, \ldots, h_{\beta_p}^2 \sim \prod_{j=0}^{p} \frac{\nu^2}{2} \exp\left(-\frac{\nu^2}{2} h_{\beta_j}^2\right)$$

$$\nu^2 \sim \text{Gamma}(\nu_0^2, \nu_0^2)$$

(10)

where the hyperparameters $\nu_0^2$ and $\nu_0^2$ are selected as 1 and 0.1, respectively, which imply diffuse prior. Similar to Park and Casella [18], the posterior distribution for $h_{\beta_k}$ and $\nu^2$ in the hierarchical structure (10) have closed expressions, such that this hierarchical representation greatly simplifies the computation. Therefore, it follows from Equation (10) that the posterior distribution of $\nu^2$ is distributed as the following Gamma distribution

$$\nu^2 | \beta_0, H_\beta \sim \text{Gamma}\left(\nu_0^2 + p + 1, \frac{1}{2} \sum_{k=0}^{p} h_{\beta_k}^2\right).$$

(11)

In addition, the posterior distributions for $h_{\beta_0}^2, \ldots, h_{\beta_p}^2$ are derived as

$$h_{\beta_k}^{-2} | \beta_0, \nu^2 \sim \text{IG}\left(\frac{\nu}{\beta_k}, \nu^2\right) \quad \text{for } k = 0, \ldots, p,$$

(12)

where $\text{IG}(a, b)$ denotes the inverse Gaussian distribution with parameter $a$ and the shape parameter $b$. As for sampling from the inverse Gaussian distribution, Tang et al. [20] gave a detailed procedure.

3. Bayesian Analysis of Model

Let $Y = \{y_{ij} : i = 1, \ldots, n, j = 1, \ldots, n_i\}$, $X = \{x_{ij} : i = 1, \ldots, n, j = 1, \ldots, n_i\}$, $Z = \{z_{ij} : i = 1, \ldots, n, j = 1, \ldots, n_i\}$, and random effects $b = \{b_i : i = 1, \ldots, n\}$.

To obtain joint Bayesian estimates of unknown parameters $\beta$ and $\sigma^2$ and the random effects, as well as to select important covariates in our considered models, a hybrid algorithm combining the block Gibbs sampler and the Metropolis–Hastings algorithm is employed to draw a sequence of random observations from the joint posterior distribution $p(\beta, \sigma^2, b | Y, X, Z)$, as follows. In this hybrid algorithm, observations $\{\beta, \sigma^2, b\}$ are iteratively drawn from the following conditional distributions: $p(\beta | \sigma^2, b, Y, X, Z)$, $p(\sigma^2 | \beta, b, Y, X, Z)$ and $p(b | \beta, \sigma^2, Y, X, Z)$.

Block Gibbs Sampler (A): Conditional distribution related to $\beta$
It follows from Equations (2) and (10) that the conditional distribution \( p(\beta|\sigma^2, b, Y, X, Z) \) is proportional to
\[
\exp \left\{ -\frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n_i} \frac{1}{\sigma_i^2} d(y_{ij}; \mu_{ij}) + (\beta - \beta_0)^T H_\beta^{-1} (\beta - \beta_0) \right\},
\]
which is an unfamiliar distribution. Therefore, we used the well-known Metropolis–Hastings (MH) algorithm to generate observations from the aforementioned conditional distribution as follows. Given the current value \( \beta^{(l)} \), new candidate \( \beta \) is generated from the proposal distribution \( N(\beta^{(l)}, \sigma_\beta^2 \Sigma_\beta) \) and is accepted with probability
\[
\min \left\{ 1, \frac{p(\beta|\sigma^2, b, Y, X, Z)}{p(\beta^{(l)}|\sigma^2, b, Y, X, Z)} \right\},
\]
where
\[
\Sigma_\beta = \left( \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\hat{d}(y_{ij}; \mu_{ij})}{2\sigma_i^2 (f(\mu_{ij}))^2} X_i^T X_i + H_\beta^{-1} \right)^{-1}
\]
with \( f(\mu_{ij}) = \partial f / \partial \mu_{ij} \) and \( \hat{d}(y_{ij}; \mu_{ij}) = E_{y_{ij}} (\partial^2 d(y_{ij}; \mu_{ij}) / \partial \mu_{ij}^2) \big|_{\beta=\beta^{(l)}} \), and the variance coefficient \( \sigma_\beta^2 \) can be chosen, such that the average acceptance rates are approximately 0.25 or more.

**Block Gibbs Sampler (B): Conditional distribution related to \( \sigma^{-2} \)**

The conditional distribution \( p(\sigma^{-2}|\beta, b, Y, X, Z) \) can be derived as
\[
p(\sigma^{-2}|\beta, b, Y, X, Z) \propto (\sigma^{-2})^{0.5 \sum_{i=1}^{n} n_i + \sigma_\beta^2 - 1} \exp \left\{ -\left( 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n_i} d(y_{ij}; \mu_{ij}) + \sigma_\beta^2 \right) \sigma^{-2} \right\},
\]
which can be simplified as
\[
\sigma^{-2}|\beta, b, Y, X, Z \sim \Gamma \left( 0.5 \sum_{i=1}^{n} n_i + \sigma_\beta^2, 0.5 \sum_{i=1}^{n} \sum_{j=1}^{n_i} d(y_{ij}; \mu_{ij}) + \sigma_\beta^2 \right).
\]

Clearly, it is straightforward and efficient to draw observations for \( \sigma^{-2} \) from the Gamma distribution via any statistical software.

**Block Gibbs Sampler (C): Conditional distribution related to \( \theta_b \)**

Let \( \theta_b \) denote all unknown parameters associated with distribution of random effects \( b_i, i = 1, \ldots, n \). \( \theta_b \) can be iteratively sampled by using the following nine steps:

**Step (a).** Conditional distribution of \( \xi \) given \( (\mu^*, \Psi, b) \) is given
\[
\xi | \mu^*, \Psi, b \sim N_q (A_\Psi, \mathbb{B})
\]
where \( \mathbb{B} = (G \Psi^{-1} + (\Psi^0)^{-1})^{-1} \) and \( A_\Psi = \mathbb{B} ((\Psi^0)^{-1} \xi^0 + \Psi^{-1} \sum_{q=1}^{G} \mu^*_{q}) \).

**Step (b).** For \( j = 1, \ldots, q \), the diagonal elements of \( \Psi \) is conditionally distributed as
\[
\psi_{ij}^{-1} | \mu^*, \xi \sim \Gamma \left( c_1 + \frac{C}{2}, c_2 + \frac{1}{2} \sum_{q=1}^{C} (\mu^*_{q} - \xi_j)^2 \right),
\]
where \( \mu^*_{q} \) is the \( j \)th element of \( \mu^*_{q} \) and \( \xi_j \) is the \( j \)th element of \( \xi \).
Step (c). For $j = 1, \ldots, q$, $\omega_j|\Omega$ is conditionally distributed as

$$
\omega_j|\Omega \sim \Gamma\left(\omega_j^\alpha, \omega_j^\beta + \sum_{g=1}^{G} \omega_g^{-1}\right),
$$

where $\omega_g$ is the $j$th diagonal element of $\Omega_g$.

Step (d). Following Ishwaran and Zarepour [26], the conditional distribution of $\tau|\pi$ can be expressed as

$$
\tau|\pi \sim \Gamma\left(a_1 + G - 1, \frac{G}{2} - \sum_{g=1}^{G-1} \log(1 - \nu_g^*)\right),
$$

where $\nu_g^*$ is a random weight sampled from the beta distribution and it is sampled with step (e).

Step (e). It is easily obtained that the conditional distribution of $\pi|L, \tau$ is distributed as the following generalized Dirichlet distribution:

$$
\pi|L, \tau \sim \text{Dir}(\alpha^0, \beta^0, \ldots, \alpha^0_{G-1}, \beta^0_{G-1}),
$$

where $\alpha^0_g = 1 + d_g, \beta^0_g = \tau + \sum_{i=1}^{G} d_i$ for $g = 1, \ldots, G - 1$, and $d_g$ is the number of $L_i$'s (and thus individuals) whose values equal to $g$. Simulating observation from the conditional distribution $\pi|L, \tau$ can be conducted as follows. First, $v_g^*$ is independently generated from a Beta distribution $(a^0_g, b^0_g)$. Then, $\pi_1, \ldots, \pi_G$ are obtained from the following formulae:

$$
\pi_1 = v_1^*, \quad \pi_G = 1 - \sum_{g=1}^{G-1} \pi_g, \quad \text{and} \quad \pi_g = \prod_{i=1}^{g-1} (1 - v_i^*) v_g^*, \quad \text{for} \quad g \neq 1 \quad \text{or} \quad G.
$$

Step (f). Conditional distribution of $\mu|\xi, \Psi, \Omega, L, b$.

Let $L_1, \ldots, L_q$ be the $d$ unique values of $\{L_1, \ldots, L_n\}$ (i.e., unique number of “clusters”), for $g = 1, \ldots, G$; $\mu_g$ is conditionally distributed as follows:

$$
\mu_g|\xi, \Psi \sim N_q(\xi, \Psi) \quad \text{for} \quad g \not\in \{L_1, \ldots, L_q\},
$$

$$
\mu_g|\xi, \Psi, \Omega, L, b \sim N_q(\mathbb{E}_g, \mathbb{V}_g) \quad \text{for} \quad g \in \{L_1, \ldots, L_q\},
$$

where $\mathbb{E}_g = (\Psi^{-1} + \sum_{\{L_i=g\}} \Omega_i^{-1})^{-1}$ and $\mathbb{V}_g = \mathbb{E}_g (\Psi^{-1} + \sum_{\{L_i=g\}} \Omega_i^{-1}) \mathbb{E}_g$ for $g \in \{L_1, \ldots, L_q\}$.

Given $\mu^\prime_g, \mu_g = \mu^\prime_g - \sum_{g=1}^{G} \pi_g \mu^\prime_g, \mu^* = \{\mu_1^*, \ldots, \mu_G^*\}$ and $\mu = \{\mu_1, \ldots, \mu_G\}$.

Step (g). Conditional distribution of $\Omega|\mu, \omega, L, \tau$.

Similar to the notation of step (f), given $g$, for $j = 1, \ldots, q$, the $j$th diagonal element of $\Omega_g$ is conditionally distributed as

$$
\omega_{gj} \sim \Gamma(\omega_j^\alpha, \omega_j^\beta) \quad \text{for} \quad g \not\in \{L_1, \ldots, L_q\},
$$

$$
\omega_{gj} \sim \Gamma\left(\frac{d_j}{2} + \omega_j^\alpha, \omega_j^\beta + \sum_{\{L_i=g\}} \frac{1}{2} (b_i - \mu_g)^2\right) \quad \text{for} \quad g \in \{L_1, \ldots, L_q\},
$$

where $b_i$ is the $j$th element of $b_i$ and $\mu_{gj}$ is the $j$th element of $\mu_g$. Given $\omega_g$, $\Omega_g = \text{diag}(\omega_{g1}, \ldots, \omega_{gG})$ and $\Omega = \{\Omega_1, \ldots, \Omega_G\}$.

Step (h). The conditional distribution of $L_i|\pi, \mu, \Omega, b$ is given by

$$
L_i|\pi, \mu, \Omega, b \overset{iid}{\sim} \text{multinomial}(\pi_g, g = 1, \ldots, G),
$$
where \( \pi_{ig} \) is proportional to \((\pi_g \cdot p(b_i | \mu_g, \Omega_g))\) with \( b_i | \mu_g, \Omega_g \sim N_q(\mu_g, \Omega_g) \), and \( \pi_g \) \((g = 1, \ldots, G)\) are sampled from step (e). Given \( L_i, \mu \) and \( \Omega \), the prior of \( b_i \) is distributed as \( N_q(\mu_{L_i}, \Omega_{L_i}) \), with \( \mu_{L_i} \) and \( \Omega_{L_i} \) being the \( L_i \) elements of sets \( \mu \) and \( \Omega \), respectively.

**Step (i).** The conditional distribution for \( b = \{b_i : i = 1, \ldots, n\} \)

The conditional distribution \( p(b_i | \beta, \sigma^2, Y, X, Z) \) is non-standard and cannot be derived directly via Gibbs sampling for \( i = 1, \ldots, n \). Specifically,

\[
p(b_i | \beta, \sigma^2, Y, X, Z) \propto p(b_i | \mu_{L_i}, \Omega_{L_i}) p(Y_i | \beta, \sigma^2, b_i, X, Z).
\]

where \( Y_i = \{y_{ij} : j = 1, \ldots, n_i\} \), \( p(Y_i | \beta, \sigma^2, b_i, X, Z) = \prod_{j=1}^{n_i} p(y_{ij} | \mu, \sigma^2) \) with \( p(y_{ij} | \mu, \sigma^2) \) specified by Equation (1) and \( \mu \) by Equation (2). The Metropolis–Hastings algorithm used to sample observation \( b_i \) is implemented as follows. At the \( \ell \)th iteration with a current value \( b_i^{(\ell)} \), a new candidate \( b_i \) is drawn from the normal distribution \( N_q(b_i^{(\ell)}, \sigma_b^2 \Sigma_b) \), where \( \Sigma_b = (\Omega_{L_i})^{-1} + \Xi_i^{-1} \) and \( \Xi_i = -\partial^2(\log(p(Y_i | \beta, \sigma^2, b_i, X, Z))) / \partial b_i \partial b_i^T |_{b_i = b_i^{(\ell)}} \). The new \( b_i \) is accepted with probability

\[
\min \left\{ 1, \frac{p(b_i^{(\ell)} | \mu_{L_i}, \Omega_{L_i}) p(Y_i | \beta, \sigma^2, b_i^{(\ell)}, X, Z)}{p(b_i^{(\ell)} | \mu_{L_i}, \Omega_{L_i}) p(Y_i | \beta, \sigma^2, b_i^{(\ell)}, X, Z)} \right\}.
\]

The variance, \( \sigma_b^2 \), can be chosen such that the average acceptance rate is approximately 0.25 or more.

Then, we can obtain a series of sample observations—\( \{ (\beta^{(l)}, \sigma^2^{(l)}, b^{(l)}_{j}) : l = 1, 2, \ldots, q \} \)—via the above iterative process. Then, Bayesian estimates of \( \beta, \sigma^2 \) and \( b_i \) for given \( i \) can be obtained by sample mean as follows:

\[
\hat{\beta} = \frac{1}{q} \sum_{l=1}^{q} \beta^{(l)}, \quad \hat{\sigma}^2 = \frac{1}{q} \sum_{l=1}^{q} \sigma^2^{(l)}, \quad \hat{b}_i = \frac{1}{q} \sum_{l=1}^{q} b^{(l)}_{i}.
\]

Similarly, the consistent estimates of the posterior covariance matrices of \( \beta \) and \( \sigma^2 \) can be obtained via the sample covariance matrices.

**4. Numerical Examples**

To investigate the behavior of our proposed model and the BLasso method under the Bayesian framework, we conduct four simulation studies and a real example related to a prospective ophthalmology study.

**4.1. Simulation Study**

In the first simulation study, we assume that, given the random effects \( b_i = (b_{i1}, b_{i2})^T \), the longitudinal percentage responses, \( y_{ij} \), are conditionally independent and each \( y_{ij} | b_i \) \((i = 1, \ldots, 100, j = 1, \ldots, 6) \) follows the simplex distribution—that is, \( y_{ij} | b_i \sim S^-(\mu_{ij}, \sigma^2) \). The conditional mean parameter \( \mu_{ij} = E(y_{ij} | b_i) \) is specified as follows:

\[
\logit(\mu_{ij}) = x_{ij}^T \beta + z_{ij}^T b_i = \beta_0 + x_{1ij} \beta_1 + x_{2ij} \beta_2 + x_{3ij} \beta_3 + t_{ij} \beta_4 + b_{1i} + t_{ij} b_{2i},
\]

where \( x_{1ij} \) randomly takes 1 or -1 with equal probability—\( x_{2ij} \) and \( x_{3ij} \sim i.i.d N(0, 1) \), \( t_{ij} = 0.2 j \) for \( j = 0, \ldots, 5 \). Moreover, the true values of the parameters are specified as follows: \( \beta = (\beta_0, \ldots, \beta_4)^T = (-0.45, 0.00, 0.45, 0.00, 0.45)^T \). This implies that a covariate corresponding to 0 is unimportant, and that \( \sigma^2 = 1 \). The true distribution of random effect, \( b_i \), is assumed to be

\[
b_{i1} \sim i.i.d N(0, 0.8), b_{i2} = b_{i2}^* - 2 \text{ with } b_{i2}^* \sim \Gamma(4, 2),
\]
where the random effects cover the symmetric and skewed features with mean 0. A total of 500 Monte Carlo replications were conducted on the basis of the above-simulated setup.

In the second simulation study, 500 simulated datasets were generated by using the same setup as specified in the first simulation study except for the distribution of random effects. That is, random effects are distributed as

\[
b_1 \overset{\text{i.i.d}}{\sim} 0.6N(-0.8, 0.5) + 0.4N(1.2, 0.5) \quad \text{and} \quad b_2 \overset{\text{i.i.d}}{\sim} 0.6N(0.8, 0.5) + 0.4N(-1.2, 0.5),
\]

where random effects have bimodal features with 0.

For each dataset generated from the abovementioned two simulation studies, the hybrid algorithm combining the block Gibbs sampler and the Metropolis–Hastings algorithm in conjunction with the BLasso method and the stick-break prior of CDPMM was used to produce Bayesian estimates of parameters and random effects as well as simultaneously select the important covariates. To investigate the convergence for these Bayesian algorithms, we computed the estimated potential scale reduction (EPSR, proposed by Gelman et al. [27]) of parameters via three parallel sequences of observations based on three different initial values. It can be seen from Figure 1 that the EPSR values were less than 1.2 after about 7000 iterations in both simulations for all the test runs. Therefore, \( \ell = 5000 \) observations collected after 7000 iterations were used to compute the simulation results for all replications. Results obtained under simulations 1 and 2 are reported in Table 1, where ‘Bias’ is the difference between the true value and the mean of the estimates based on 500 replications; ‘RMS’ is the root mean square of differences between the true values and their corresponding estimates based on 500 replications. Compared with the Lasso from the frequentist view, the BLasso would not shrink the non-significant elements of \( \beta \) exactly toward 0 since the sampling-based method is involved. Thus, as suggested by Tang et al. [20], the criterion for variable selection is that a coefficient is viewed as 0 if its 95% confidence interval includes zeros. In Table 1, ‘\( F_0 \)’ denotes the proportion that the number of 95% confidence interval for regression parameter including zero in 500 replications is divided by 500. The larger the values of \( F_0 \) corresponding to non-significant regression parameters, and the smaller the values of \( F_0 \) corresponding to significant parameters, the better the performance of the posited model.

![Figure 1](image-url)  
*Figure 1. EPSR values of all parameters against iteration numbers for a randomly selected replication in the first simulation (left panel) and second simulation (right panel).*
Table 1. Bayesian estimates of parameters in the first and second simulation studies.

| Par. | True | Bias | RMS  | F0 (%) | Bias | RMS  | F0 (%) |
|------|------|------|------|--------|------|------|--------|
| \(\beta_0\) | \(-0.45\) | 0.054 | 0.117 | 1.60 | 0.031 | 0.143 | 5.40 |
| \(\beta_1\) | 0.00 | 0.007 | 0.109 | 89.20 | 0.002 | 0.132 | 3.00 |
| \(\beta_2\) | 0.45 | 0.017 | 0.116 | 1.40 | 0.018 | 0.132 | 3.00 |
| \(\beta_3\) | -0.00 | -0.001 | 0.141 | 94.60 | -0.003 | 0.153 | 93.40 |
| \(\beta_4\) | 0.45 | -0.074 | 0.136 | 3.80 | -0.051 | 0.146 | 8.20 |
| \(\sigma^2\) | 1.00 | 0.006 | 0.070 | --- | 0.009 | 0.073 | --- |

Note: Bias denotes the difference between the true value and the mean of the estimates based on 500 replications; RMS denotes the root mean square of differences between the true values and their corresponding estimates based on 500 replications; F0 denotes the proportion of the 95% confidence interval for regression parameter including zero in 500 replications.

Examination of Table 1 indicated that (i) the Bayesian estimates of the unknown parameters \(\beta\) and \(\sigma^2\) were reasonably accurate under the two abovementioned simulation studies since their absolute biases were less than 0.10 and their RMS values were less than 0.16; (ii) BLasso could correctly identify the zero and nonzero coefficients in most cases because the F0 values corresponding to important covariates were less than 10%, whilst the F0 values corresponding to unimportant covariates were near to 90%. On the other hand, to investigate the effectiveness of using the CDPMM prior for the random effects, we introduced the following RMSE (root of mean squared error) criterion in term of random effects,

\[
\text{RMSE}(\hat{b}_i) = \sqrt{\frac{1}{2L} \sum_{m=1}^{2} \sum_{l=1}^{L} (p_m(h_l) - \hat{p}_m(h_l))^2},
\]

where \(p_m(\cdot)\) and \(\hat{p}_m(\cdot)\) denote, respectively, the true density function for random effect \(\hat{b}_{im}\) and kernel density estimation for the estimated values of random effect \(\{\hat{b}_{im} : i = 1, \ldots, n\}\); \(h_l\) is chosen to be the \(\frac{l}{L}\) th quantile of the dataset \(\{\hat{b}_{im} : i = 1, \ldots, n\}\). The sample quantiles for the estimated values of RMSE are reported in Table 2. Furthermore, we chose a typical replication whose RMSE value is equal to the median in the 500 replications. Therefore, on the basis of the selected replication, the estimated densities of \(\hat{b}_1\) and \(\hat{b}_2\) based on the CDPMM prior against their corresponding true densities are plotted in Figures 2 and 3, which indicated that the finite mixture of normal distributions can flexibly capture the symmetric, skewed and bimodal shapes of random effect \(b_i\). From Table 2, based on the results of 500 replication in both simulations, the estimated means and standard deviation (SD) of random effects \(\hat{b}_1\) and \(\hat{b}_2\) is approximate to their corresponding true values, the 25%, 50% and 75% quantiles of RMSE(\(\hat{b}_i\)) values are small and close enough, which indicated that it is robust to apply CDPMM method to estimate random effects. All these findings indicated that (i) our proposed Bayesian procedure could capture the true information of \(b_i\) well, regardless of their true distributions and forms, and (ii) BLasso could identify the true model with a high probability.
Figure 2. Estimated densities versus true densities for random effects $b_1$ and $b_2$ in the first simulation.

Figure 3. Estimated densities versus true densities for random effects $b_1$ and $b_2$ in the second simulation.

Table 2. Estimated means, standard deviations and RMSE quantile of random effects in the first and second simulation studies.

| Random Effects | Simulation 1 | | Simulation 2 | |
|----------------|-------------|---|-------------|---|
|                | Mean | Est Mean | SD | Est SD | Mean | Est Mean | SD | Est SD |
| $b_1$          | 0.000 | 0.040   | 0.894 | 0.831 | 0.000 | -0.030 | 1.100 | 1.032 |
| $b_2$          | 0.000 | 0.063   | 1.000 | 0.903 | 0.000 | 0.031  | 1.100 | 0.995 |

| Quantile of Simulation 1 | | Quantile of Simulation 2 |
|--------------------------|---|--------------------------|
| 5%                       | 0.031 | 0.079                   |
| 25%                      | 0.039 | 0.084                   |
| 75%                      | 0.050 | 0.095                   |

Note: ‘Mean’ denotes true empirical mean of the distribution; ‘Est mean’ denotes mean of the posterior samples; ‘SD’ denotes true empirical standard deviation of the distribution; ‘Est SD’ denotes standard deviation of the posterior samples.
To compare the performance of the CDPMM prior for the random effects with the discrete DP given by Ishwaran and Zarepour [26] and Yang et al. [25], we conducted the following third simulation study. In this simulation study, 500 simulated datasets were generated by using the same setup as specified in the first simulation study, and fitted by the model with the discrete DP for the random effects. In the fourth simulation, we reanalyzed the aforementioned 500 datasets generated in the second simulation by using a parametric Bayesian approach with the random effects distribution specified by a multivariate normal distribution. The aim of this simulation was to compare the semiparametric approach based on the CDPMM prior with the parametric approach based on the Gaussian prior from the Bayesian perspective. Results obtained under the third and fourth simulations are reported in Table 3. Our programs were written in Matlab. It roughly took 119.3 s and 186.9 s in an Intel(R) Xeon(R) Silver 4216 CPU @ 2.10GHz (Intel, Santa Clara, CA, USA) serve to run 12,000 iterations for our proposed CDPMM and discrete DP, respectively; this indicates that the CDPMM method is much more efficient than the discrete DP in our considered simulations. It can be seen from Tables 2 and 3 that (i) our proposed CDPMM and discrete DP methods have the same performance in term of the ‘Bias’ and ‘RMS’ values, but the F0 values corresponding to the non-significant parameters under the proposed CDPMM prior are higher than those under the discrete DP prior; (ii) the RMS values and the correct rates of variable selection based on the F0 values under the semiparametric CDPMM prior are better than those under the parametric Gaussian prior.

Table 3. Bayesian estimates of parameters in the third and fourth simulation studies.

| Par. | Simulation 3 | Simulation 4 |
|------|--------------|--------------|
|      | True Bias RMS F0 (%) | Bias RMS F0 (%) |
| \( \beta_0 \) | \(-0.45\) 0.005 0.101 1.20 | 0.064 0.151 8.40 |
| \( \beta_1 \) | 0.00 0.000 0.104 86.60 | \(-0.009\) 0.130 84.40 |
| \( \beta_2 \) | 0.45 \(-0.001\) 0.123 1.40 | 0.033 0.148 3.80 |
| \( \beta_3 \) | \(-0.00\) 0.005 0.166 89.20 | 0.007 0.185 90.00 |
| \( \beta_4 \) | 0.45 \(-0.003\) 0.119 2.00 | \(-0.068\) 0.143 11.40 |
| \( \sigma^2 \) | 1.00 0.093 0.134 | 0.006 0.078 |

Note: Bias denotes the difference between the true value and the mean of the estimates based on 500 replications; RMS denotes the root mean square of differences between the true values and their corresponding estimates based on 500 replications; F0 denotes the proportion of the 95% confidence interval for regression parameter including zero in 500 replications.

4.2. Real Example

In this section, the application of our proposed approach to the skewed longitudinal proportional data is illustrated by the analysis of a prospective ophthalmology study [28] from the Bayesian perspective. The prospective ophthalmology data used in the study were obtained from the Supplementary Materials of a paper by Song and Tan [29] and were available from https://biometrics.biometricsociety.org/home/archive/supplementary-materials, accessed on 5 september 2022. This prospective ophthalmology study described that the eyes of 31 patients before surgery were injected by three gas concentration levels of \( C_3F_8 \), and all patients after surgery were followed up three–eight times over a three-month period. The outcome variable was the percentage of remaining gas volume relative to the initial volume of gas injected. These longitudinal proportional data from a prospective ophthalmology study were analyzed by Qiu et al. [1], Song and Tan [29] and Song et al. [30], respectively. However, these authors did not consider to conduct variable selection in the analysis of this dataset. Our scientific interest in this study is to investigate the effect of three initial gas concentration levels of \( C_3F_8 \) and time on the percentage of remaining gas volume, while accounting for selecting the important covariates based on the BLasso method. Let the response \( y_{ij} \) denote the percentage of gas left in the eye for patient \( i \) at the
where \( t_{ij} \) is the time covariate for days after the gas injection; \( w_{ij} \) is the covariate of gas concentration levels equal to \(-1, 0\) and \(1\), corresponding to the concentration levels of 15%, 20% and 25%, respectively; and random effects \( b_{i1} \) and \( b_{i2} \) are specified by CDPMM in (4), which characterize the effect fluctuations of interception and logarithmic time among patients.

The abovementioned MCMC algorithm was used to produce the joint Bayesian estimates of parameters and random effects in this real example. In the implementation of MCMC process, the hyperparameter values were taken to be the same as those given in simulation. Similarly, we used the EPSR method given in simulation to investigate the convergence for these algorithms. The EPSR values of all parameters against the iteration numbers was plotted in Figure 4, which indicated that the MCMC algorithm converged after 4000 iterations since their EPSR values were less than 1.2 after about 4000 iterations. Hence, \( \mathcal{L} = 4000 \) observations collected after 4000 iterations were used to calculate Bayesian estimates of parameters and random effects. Results are reported in Table 4 and Figure 5, which indicated that (i) the estimated densities of random effects \( b_{i1} \) and \( b_{i2} \) were bimodal and skewed, which indicated that traditional normality assumption for random effects is inappropriate in this real example; (ii) the square of logarithmic time (\( \log^2(t_{ij}) \)) was detected to be an important covariate with a significantly negative effect on the percentage of gas left in the eye, since its corresponding 95% confidence interval did not include zero. The gas concentration levels \( (w_{ij}) \) and the logarithmic time \( (\log^2(t_{ij})) \) were insignificant at significance level 0.05 because their 95% confidence interval included zero.

Table 4. Bayesian estimates (BEs) and standard deviations (SDs) and 95% credible intervals (CIs) for parameters in the ophthalmology study.

| Par. | BE   | SD  | CI                  |
|------|------|-----|---------------------|
| \( \beta_0 \) | 2.579 | 0.257 | ( 2.200, 3.031)  |
| \( \beta_1 \) | 0.093 | 0.172 | (−0.232, 0.427) |
| \( \beta_2 \) | −0.322 | 0.045 | (−0.412, −0.241) |
| \( \beta_3 \) | 0.368 | 0.281 | (−0.131, 0.697) |
| \( \sigma^2 \) | 9.664 | 1.317 | (7.406, 12.516) |

Figure 4. EPSR values of all parameters against iteration numbers in the ophthalmology study.
5. Conclusions

In this paper, we introduced a new semiparametric simplex mixed-effects model with the random effects following the centered Dirichlet process mixture model (CDPMM). The advantages of the proposed model based on CDPMM are the following: (i) it can capture the features of skewed and bimodal longitudinal proportional data; (ii) it can characterize absolutely continuous distributions for random effects. The novelty of our approach is that we adopted the Blasso procedure to simultaneously estimate parameters of interest, provide credible intervals (CIs) for parameters and conduct both shrinkage and variable selection for our considered models. A hybrid algorithm combining the Gibbs sampler and the MH algorithm was used to simultaneously obtain Bayesian estimates of unknown parameters, random effects and their standard errors and credible intervals. Empirical results show that (i) the proposed semiparametric Bayesian method provides quite accurate estimates of parameters (see Table 1); (ii) the average frequencies of correctly identifying unimportant predictors were near to 90%; (iii) CDPMM can effectively capture the potential features of normal, gamma and mixture normal distributions (see Table 2 and Figures 2, 3 and 5).

Author Contributions: Conceptualization, A.T. and X.D.; methodology, X.D. and A.T.; software, A.T. and Y.Z.; validation, A.T., X.D. and Y.Z.; formal analysis, A.T. and X.D.; investigation, A.T., X.D. and Y.Z.; Preparation of the original work draft, X.D. and A.T.; visualization, A.T. and Y.Z.; supervision, funding acquisition, A.T., X.D. and Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 11961079, No. 12161014, No. 11761016), the Guizhou Provincial Science and Technology Project (KY[2021]134), the Natural Science Research Project of Education Department of Guizhou Province (KY[2021]134), the Project of High Level Creative Talents in Guizhou Province of China, and Guiyang University Multidisciplinary Team Construction Projects in 2021[2021-xk04].

Data Availability Statement: The research data are available on the website: https://biometrics.biometricssociety.org/home/archive/supplementary-materials, accessed on 5 september 2022.

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Qiu, Z.G.; Song, P.X.K.; Tan, M. Simplex mixed-effects models for longitudinal proportional data. *Scand. J. Stat.* 2008, 35, 577–596. [CrossRef]

2. Zhang, W.; Wei, H. Maximum Likelihood Estimation for Simplex Distribution Nonlinear Mixed Models via the Stochastic Approximation Algorithm. *Rocky Mt. J. Math.* 2008, 38, 1863–1875. [CrossRef]

3. Zhao, Y.Y.; Xu, D.K.; Duan, X.D.; Dai, L. Bayesian estimation of simplex distribution nonlinear mixed models for longitudinal data. *Int. J. Appl. Math. Stat.* 2014, 52, 1–10.

4. Bonat, W.H.; Lopes, J.E.; Shimakura, S.E.; Ribeiro, P.J., Jr. Likelihood analysis for a class of simplex mixed models. *Chil. J. Stat.* 2018, 8, 3–7.

5. Quintero, F.O.L. Sensitivity analysis for variance parameters in Bayesian simplex mixed models for proportional data. *Commun. Stat. Simul. Comput.* 2017, 46, 5212–5228. [CrossRef]

6. Kleinman, K.P.; Ibrahim, J.G. A semiparametric Bayesian approach to generalized linear mixed models. *Stat. Med.* 1998, 17, 2579–2596. [CrossRef]

7. Tang, N.S.; Duan, X.D. A semiparametric Bayesian approach to generalized partial linear mixed models for longitudinal data. *Comput. Stat. Data Anal.* 2012, 56, 4348–4365. [CrossRef]

8. Tang, N.S.; Zhao, Y.Y. Semi-parametric Bayesian analysis of nonlinear reproductive dispersion mixed models for longitudinal data. *J. Multivar. Anal.* 2013, 115, 68–83. [CrossRef]

9. Zhao, Y.Y.; Xu, D.K.; Duan, X.D.; Du, J. A semiparametric Bayesian approach to binomial distribution logistic mixed-effects models for longitudinal data. *J. Stat. Comput. Simul.* 2022, 92, 1438–1456.

10. Duan, X.D.; Fung, W.K.; Tang, N.S. Bayesian semiparametric reproductive dispersion mixed models for non-normal longitudinal data: estimation and case influence analysis. *J. Stat. Comput. Simul.* 2017, 87, 1925–1939. [CrossRef]

11. Hocking, R.R. The analysis and selection of variables in linear regression. *Biometrics* 1976, 32, 1–51. [CrossRef]

12. Kass, R.E.; Raftery, A.E. Bayes factors. *J. Am. Stat. Assoc.* 1995, 90, 773–795. [CrossRef]

13. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* 1974, 19, 716–723. [CrossRef]

14. Spiegelhalter, D.J.; Best, N.; Carlin, B.P.; Linde, A. Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. Ser. B* 2002, 64, 583–639. [CrossRef]

15. Robert, T. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B* 1996, 58, 267–288.

16. Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* 2005, 61, 301–320. [CrossRef]

17. Zou, H. The Adaptive Lasso and its Oracle Properties. *J. Am. Stat. Assoc.* 2006, 101, 1418–1429. [CrossRef]

18. Park, T.; Casella, G. The Bayesian Lasso. *J. Am. Stat. Assoc.* 2008, 103, 681–686. [CrossRef]

19. Guo, R.; Zhu, H.; Chow, S.M.; Ibrahim, J.G. Bayesian lasso for semiparametric structural equation models. *Biometrics* 2012, 68, 567–579. [CrossRef]

20. Tang, A.M.; Zhao, X.; Tang, N.S. Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data. *Biom. J.* 2017, 59, 57–78. [CrossRef]

21. Erp, S.V.; Oberski, D.L.; Mulder, J. Shrinkage priors for Bayesian penalized regression. *J. Math. Psychol.* 2019, 89, 31–50. [CrossRef]

22. Barndorff-Nielsen, O.E.; Jørgensen, B. Some parametric models on the simplex. *J. Multivar. Anal.* 1991, 39, 106–116. [CrossRef]

23. Ohlssen, D.I.; Sharples, L.D.; Spiegelhalter, D.J. Flexible random-effects models using Bayesian semiparametric models: applications to institutional comparisons. *Stat. Med.* 2007, 26, 2088–2112. [CrossRef] [PubMed]

24. Sethuraman, J. A constructive definition of Dirichlet priors. *Stat. Sin.* 1994, 4, 639–650.

25. Yang, M.G.; Dunson, D.B.; Baird, D. Semiparametric Bayes hierarchical models with mean and variance constraints. *Comput. Stat. Data Anal.* 2010, 54, 2172–2186. [CrossRef]

26. Ishwaran, H.; Zarepour, M. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *J. Am. Stat. Assoc.* 2000, 87, 371–390. [CrossRef]

27. Gelman, A. Inference and monitoring convergence. In *Markov Chain Monte Carlo in Practice*; Gilks, W.R., Richardson, S., Spiegelhalter, D.J., Eds.; Chapman and Hall: London, UK, 1996.

28. Meyers, S.M.; Ambler, J.S.; Tan, M.; Werner, J.C.; Huang, S.S. Variation of perfluoropropane disappearance after vitrectomy. *Retina* 1992, 4, 359–363. [CrossRef]

29. Song, P.X.K.; Tan, M. Marginal models for longitudinal continuous proportional data. *Biom. J.* 2000, 56, 496–502. [CrossRef]

30. Song, P.X.K.; Qiu, Z.G.; Tan, M. Modelling heterogeneous dispersion in marginal models for longitudinal continuous proportional data. *Biom. J.* 2004, 46, 540–553. [CrossRef]