ROCKY MOUNTAIN
JOURNAL OF MATHEMATICS
Volume 38, Number 5, 2008

MAXIMUM LIKELIHOOD ESTIMATION FOR SIMPLEX DISTRIBUTION NONLINEAR MIXED MODELS VIA THE STOCHASTIC APPROXIMATION ALGORITHM

WENZHUAN ZHANG AND HONGJIE WEI

ABSTRACT. Longitudinal continuous proportional data is common in many fields such as biomedical research, psychological research and so on, e.g., the percent decrease in glomerular filtration rate at different follow-up times from the baseline. As shown in Song and Tan [**16**] such data can be fitted with simplex models. However, the original models of [**16**] for such longitudinal continuous proportional data assumed a fixed effect for every subject. This paper extends the models of Song and Tan [**16**] by adding random effects, and proposes simplex distribution nonlinear mixed models which are one kind of nonlinear reproductive dispersion mixed model. By treating random effects in the models as hypothetical missing data and applying the Metropolis-Hastings (M-H) algorithm, this paper develops the stochastic approximation (SA) algorithm with Markov chain Monte-Carlo (MCMC) method for maximum likelihood estimation in the models. Finally, for ease of comparison, the method is illustrated with the same data from an ophthalmology study on the use of intraocular gas in retinal surgeries in [**16**].

1. Introduction. Dispersion models, which contain a broader class of distributions that accommodate a large number of different data types, were defined in [**9**]. Besides those familiar exponential family distributions, the simplex distribution of Barndorff-Nielsen and Jørgensen [**1**] also represents a special dispersion model for proportional data and is of particular interest in this paper. Based on this distribution, Song and Tan [**16**] developed a marginal simplex model for longitudinal continuous proportional data and assumed a constant dispersion in their model, and this model was used to analyze the eye surgery data in [**14**]; [**15**] further assumed a varying dispersion on the basis of [**16**] and re-analyzed the same surgery data; and Zhang [**18**, page 4] proposed

Keywords and phrases. Simplex distribution, nonlinear mixed models, maximum likelihood estimates, stochastic approximation algorithm, M-H algorithm, longitudinal continuous proportional data.
Received by the editors on July 21, 2007, and in revised form on February 8, 2008.
DOI:10.1216/RMJ-2008-38-5-1863

the Fisher score iterative algorithm to estimate the parameter in the model of [**16**] and obtained a similar result.

For analysis of longitudinal data, see [**2**], mixed models enable the analyst not only to describe the trend over time while taking account of the correlation that exists between successive measurements, but also to describe the variation in baseline measurement and in the rate of change over time. Therefore, mixed models have been widely applied in such fields as biomedical research, psychological research, econometrics and so on, and the methods of parameter estimation in the models have been studied, which is one of the important issues in regression theory. Hartley and Rao [**6**], Harvile [**7**] and McCulloch [**11**] analyzed the issue of parameter estimation about linear mixed models, i.e., variance component models; McCulloch [**12**] and Zhu and Lee [**21**] studied the same issue about generalized linear mixed models; Zong et al. [**22**] further studied the parameter estimation of exponential nonlinear mixed models. Subsequently, Zhang et al. [**19, 20**] studied parameter estimation for nonlinear reproductive dispersion mixed models.

In this paper, we propose a mixed model–simplex distribution nonlinear mixed models (SDNMMs) for longitudinal proportional data in which the simplex distribution was regarded as a random error term. In our model, the random effects are missing data. In most applications, marginal log-likelihood functions based on observed data involve intractable integrals and complicated formulation for such models with latent variables and/or incomplete data, and they lead to complications and difficulties for further statistical analysis. Such is the case with our model. Additionally, it seems to be impossible to conduct a simulation because the sampling algorithm for this distribution is a real challenge needing appropriate and urgent resolution. Inspired by Gu and Kong [**5**] and Zhu and Lee [**21**], we develop a procedure (SA-MCMC) that combines the stochastic approximation (SA) algorithm with the Markov chain Monte-Carlo (MCMC) method for maximum likelihood estimation of our model by treating the random effects as hypothetical missing data and applying the Metropolis-Hastings (M-H) algorithm.

The paper is organized as follows. Section 2 defines simplex distribution nonlinear mixed models (SDNMMs) and introduces some related notations and natures. The SA-MCMC procedure that combines the

SA algorithm and the M-H algorithm is described in Section 3. In Section 4, the proposed models are applied to reanalyze the same eye surgery data in [**14**] and compare the result with that of Song and Tan [**16**].

2. Simplex distribution nonlinear mixed models. Let percentage responses for the ith subject be y_{ij}, observed at time t_{ij}, and $y_{ij} \in (0,1)$, where $j = 1, \ldots, n_i$ and $i = 1, \ldots, I$. Let $\mathbf{b}_i(k \times 1)$ be the ith subject's random effect. $y_{ij} \mid \mathbf{b}_i$, $j = 1, \ldots, n_i$, are mutually independent and follow a simplex distribution of the form

$$p(y_{ij} \mid \mathbf{b}_i; \mu_{ij}, \sigma^2) = \left[2\pi\sigma^2 \left\{y_{ij}(1-y_{ij})\right\}^3\right]^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\, d(y_{ij};\mu_{ij})\right\}, \tag{1}$$

$$d(y_{ij};\mu_{ij}) = \frac{(y_{ij}-\mu_{ij})^2}{y_{ij}(1-y_{ij})\mu_{ij}^2(1-\mu_{ij})^2}, \tag{2}$$

where $\mu_{ij} \in (0,1)$ is the location parameter and $\sigma^2 \in R^+$ is the dispersion parameter.

The SDNMMs are defined by (1) and (2), and the expectation μ_{ij} is followed by

$$\mu_{ij} = f(\mathbf{x}_{ij}, \boldsymbol{\beta}) + \mathbf{z}_{ij}^T \mathbf{b}_i, \tag{3}$$

where $d(\cdot;\cdot)$ is a unique deviance function whose definite field is $(0,1) \times (0,1)$. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a vector of unknown fixed affect parameters. $\mathbf{x}_{ij}$, $p \times 1$, and $\mathbf{z}_{ij}$, $k \times 1$, are design vectors for the fixed and random effects, respectively. The distribution of $\mathbf{b}_i$ is assumed to be normal $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{C}(\boldsymbol{\theta})\mathbf{C}(\boldsymbol{\theta})^T$ via the Cholesty decomposition depends upon $\boldsymbol{\theta}$, an unknown vector.

Let $\bar{\mathbf{b}}_i = \mathbf{C}(\boldsymbol{\theta})^{-1}\mathbf{b}_i$. Then $\bar{\mathbf{b}}_i \sim N(\mathbf{0}, \mathbf{I})$, where $\mathbf{I}$ is the $k \times k$ identical matrix and $i = 1, 2, \ldots, I$. And we can get

$$\mu_{ij} = f(\mathbf{x}_{ij}, \boldsymbol{\beta}) + \mathbf{z}_{ij}^T \mathbf{C}\bar{\mathbf{b}}_i. \tag{4}$$

The following notation will be used in subsequent sections: $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})^T$, $\mathbf{Z}_i = (\mathbf{z}_{i1}, \ldots, \mathbf{z}_{in_i})^T$, $f(\mathbf{X}_i, \boldsymbol{\beta}) =$

$(f(\mathbf{x}_{i1},\boldsymbol{\beta}),\ldots,f(\mathbf{x}_{in_i},\boldsymbol{\beta}))^T$, $\boldsymbol{\mu}_i=(\mu_{i1},\ldots,\mu_{in_i})^T$, $d(\mathbf{y}_i;\boldsymbol{\mu}_i)=\sum_{j=1}^{n_i} d(y_{ij};\mu_{ij})$.

Let $\boldsymbol{\psi}=(\boldsymbol{\beta}^T,\boldsymbol{\theta}^T)^T\in\boldsymbol{\Psi}$ be the vector of unknown parameter. For computational convenience, we suppose that σ^2 is given.

According to the conditional independence of $y_{ij}|\bar{\mathbf{b}}_i$, $j=1,\ldots,n_i$, the joint density of $(\mathbf{y}_i,\bar{\mathbf{b}}_i)$ is given by

$$p(\mathbf{y}_i,\bar{\mathbf{b}}_i)=\Big[\prod_{j=1}^{n_i}p(y_{ij}\mid\bar{\mathbf{b}}_i)\Big](2\pi)^{-k/2}\exp\Big\{-\frac{1}{2}\bar{\mathbf{b}}_i^T\bar{\mathbf{b}}_i\Big\}. \tag{5}$$

Let $\mathbf{Y}_0=(\mathbf{y}_1,\ldots,\mathbf{y}_I)$ and $\bar{\mathbf{b}}=(\bar{\mathbf{b}}_1,\ldots,\bar{\mathbf{b}}_I)$ be the observed-data and the missing-data, respectively, and let $\mathbf{Y}_c=(\mathbf{Y}_0,\bar{\mathbf{b}})$ be the complete-data , $a(y_{ij};\sigma^2)=[2\pi\sigma^2\{y_{ij}(1-y_{ij})\}^3]^{-1/2}$. Then the complete data log-likelihood function of $\boldsymbol{\psi}$ is given by

$$\begin{aligned}L_c(\boldsymbol{\psi}\mid\mathbf{Y}_c)&=\sum_{i=1}^{I}\log p(\mathbf{y}_i,\bar{\mathbf{b}}_i)\\&=\sum_{i=1}^{I}\Big\{\sum_{j=1}^{n_i}\Big[\log a(y_{ij};\sigma^2)-\frac{1}{2\sigma^2}d(y_{ij};\mu_{ij})\Big]\\&\qquad-\frac{k}{2}\log 2\pi-\frac{1}{2}\bar{\mathbf{b}}_i^T\bar{\mathbf{b}}_i\Big\},\end{aligned} \tag{6}$$

where $\mu_{ij}=f(\mathbf{x}_{ij},\boldsymbol{\beta})+\mathbf{z}_{ij}^T\mathbf{C}\bar{\mathbf{b}}_i$.

The observed data log-likelihood function of $\boldsymbol{\psi}$ takes the form

$$\begin{aligned}L_0(\boldsymbol{\psi}\mid\mathbf{Y}_0)=\sum_{i=1}^{I}\Big\{\log\int\Big[\prod_{j=1}^{n_i}p(y_{ij}\mid\bar{\mathbf{b}}_i)\Big](2\pi)^{-k/2}\\\times\exp\Big\{-\frac{1}{2}\bar{\mathbf{b}}_i^T\bar{\mathbf{b}}_i\Big\}d\,\bar{\mathbf{b}}_i\Big\}.\end{aligned} \tag{7}$$

Our objective is to find the maximum likelihood estimate of $\boldsymbol{\psi}$, the value $\widehat{\boldsymbol{\psi}}$ that maximizes $L_0(\boldsymbol{\psi}\mid\mathbf{Y}_0)$. By comparing (6) and (7), it can readily be seen that (6) is simple, in contrast to (7), which involves intractable integrals. When the dimension of $\bar{\mathbf{b}}_i$ is high, the integral in

(7) usually doesn't have an analytic form. Then direct maximization of $L_0(\boldsymbol{\psi} \mid \mathbf{Y}_0)$ is numerically infeasible. Inspired by the stochastic approximation (SA) algorithm in overcoming some difficulties in numerical integration and Monte Carlo integration, we implement the SA-MCMC algorithm for maximum likelihood estimation in the SDNMMs.

It follows from the reasoning in Louis [**10**], $\nabla L_0(\boldsymbol{\psi} \mid \mathbf{Y}_0)$ and $-\nabla^2 L_0(\boldsymbol{\psi} \mid \mathbf{Y}_0)$ are given by

$$\nabla L_0(\boldsymbol{\psi}|\mathbf{Y}_0) = E[\nabla L_c(\boldsymbol{\psi} \mid |\mathbf{Y}_c) \mid \mathbf{b}, \boldsymbol{\psi}], \tag{8}$$

and

$$\begin{aligned} -\nabla^2 L_0(\boldsymbol{\psi} \mid \mathbf{Y}_0) &= E[-\nabla^2 L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)|\bar{\mathbf{b}}, \boldsymbol{\psi}] - E\{[\nabla L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)]^{\otimes 2}|\bar{\mathbf{b}}, \boldsymbol{\psi}\} \\ &\quad + E[\nabla L_c(\boldsymbol{\psi}|\mathbf{Y}_c)|\bar{\mathbf{b}}, \boldsymbol{\psi}]^{\otimes 2}, \end{aligned} \tag{9}$$

where the expectation is taken with respect to the conditional distribution $p(\bar{\mathbf{b}} \mid \mathbf{Y}_0, \boldsymbol{\psi})$; ∇ and ∇^2 are respectively the first and second derivative operators with respect to $\boldsymbol{\psi}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ for any vector $\mathbf{a}$.

For the complete data log-likelihood function of $\boldsymbol{\psi}$, the first and second derivatives can be given by the following theorem.

Theorem 2.1. *According to* (1), (2) *and* (4), *we obtain the first and second derivatives with respect to* $\boldsymbol{\psi}$ *of* $L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)$,

$$\frac{\partial L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta}} = -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \mathbf{e}_i, \tag{10}$$

$$\frac{\partial L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \theta_{tl}} = -\phi \sum_{i=1}^{I} \mathbf{h}_{itl}^T \mathbf{e}_i, \tag{11}$$

$$\frac{\partial^2 L_c(\psi \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\phi \sum_{i=1}^{I} \left\{ [\mathbf{e}_i^T][\mathbf{W}_i] + \mathbf{D}_i^T \mathbf{V}_i \mathbf{D}_i \right\}, \tag{12}$$

$$\frac{\partial^2 L_c(\psi \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \theta_{tl}} = -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \mathbf{V}_i \mathbf{h}_{itl}, \tag{13}$$

$$\frac{\partial^2 L_c(\boldsymbol{\psi} \mid Y_c)}{\partial\theta_{tl}\partial\theta_{sr}} = -\phi \sum_{i=1}^{I} \left\{ \mathbf{h}_{itl,sr}^T \mathbf{e}_i + \mathbf{h}_{itl}^T \mathbf{V}_i \mathbf{h}_{isr} \right\}, \tag{14}$$

where $\phi = 1/(2\sigma^2)$, $\mathbf{D}_i = \partial f(\mathbf{X}_i, \boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$, $\mathbf{e}_i = \partial d(\mathbf{y}_i, \boldsymbol{\mu}_i)/\partial\boldsymbol{\mu}_i$, θ_{tl}, $t = 1, \dots, k$, $l = 1, \dots, t$, *and* θ_{sr}, $s = 1, \dots, k$, $r = 1, \dots, s$, *are the elements* (t, l) *and* (s, r) *of matrix* $\mathbf{C}$, *respectively,* $\dot{\mathbf{C}}(tl) = \partial\mathbf{C}/\partial\theta_{tl}$, $\mathbf{h}_{itl} = \mathbf{Z}_i^T \dot{\mathbf{C}}(tl)\bar{\mathbf{b}}_i$, $\mathbf{V}_i = \partial^2 d(\mathbf{y}_i; \boldsymbol{\mu}_i)/\partial\boldsymbol{\mu}_i\partial\boldsymbol{\mu}_i^T$, $\mathbf{W}_i = \partial^2 f(\mathbf{X}_i, \boldsymbol{\beta})/\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}$, $\ddot{\mathbf{C}}(tl, sr) = \partial^2\mathbf{C}/\partial\theta_{tl}\partial\theta_{sr}$, $\mathbf{h}_{itl,sr} = \mathbf{Z}_i^T \ddot{\mathbf{C}}(tl, sr)\bar{\mathbf{b}}_i$ *and* $[\cdot][\cdot]$ *is the multiplication of two cubic matrixes.*

The proofs of Theorem 2.1 are sketched in the Appendix.

3. The stochastic approximation algorithm.

Gu and Kong [**5**] developed a stochastic approximation-Markov chain Monte Carlo (SA-MCMC) procedure for maximum likelihood estimation of general statistical models with incomplete data. Subsequently, Zhu and Lee [**21**] proposed the SA-MCMC procedure to analyze generalized linear mixed models (GLMMs) and Zhang et al. [**19, 20**] proposed the SA-MCMC procedure to analyze nonlinear reproductive dispersion mixed models (NRDMMs). Here, we propose the SA-MCMC procedure to analyze simplex distribution nonlinear mixed models (SDNMMs).

Let $\boldsymbol{\psi}^{(r)}$ be the rth estimate of $\boldsymbol{\psi}$, and let $\boldsymbol{\Gamma}_r$ be the rth estimate of $-\nabla^2 L_0(\boldsymbol{\psi} \mid \mathbf{Y}_0)$. Given an initial point $\boldsymbol{\psi}^{(0)}$ and an initial matrix $\boldsymbol{\Gamma}_0$, the $r + 1$th improved estimate can be obtained by

$$\begin{cases} \boldsymbol{\Gamma}_{r+1} = \boldsymbol{\Gamma}_r + \gamma_r(\mathbf{H}(\boldsymbol{\psi}^{(r)}) - \boldsymbol{\Gamma}_r), \\ \boldsymbol{\psi}^{(r+1)} = \boldsymbol{\psi}^{(r)} + \gamma_r \boldsymbol{\Gamma}_{r+1}^{-1} \mathbf{S}(\boldsymbol{\psi}^{(r)}), \end{cases} \tag{15}$$

where $\{\gamma_r\}$ is a sequence of positive values such that

$$\sum_{r=1}^{\infty} \gamma_r = \infty \text{ and } \sum_{r=1}^{\infty} \gamma_r^2 < \infty.$$

The choice for $\{\gamma_r\}$ and $\boldsymbol{\Gamma}_0$ are usually $\{1/k\}$ and $\mathbf{0}$, respectively, where $\mathbf{0}$ is a zero matrix.

$\mathbf{S}(\boldsymbol{\psi}^{(r)})$ and $\mathbf{H}(\boldsymbol{\psi}^{(r)})$ are given by

$$\mathbf{S}(\boldsymbol{\psi}^{(r)}) = \frac{1}{m} \sum_{l=1}^{m} \nabla L_c(\boldsymbol{\psi}^{(r)} | \mathbf{Y}_0, \bar{\mathbf{b}}_r^l),$$

and

$$\mathbf{H}(\boldsymbol{\psi}^{(r)}) = \overline{\mathbf{H}}(\boldsymbol{\psi}^{(r)}) - \delta_r \underline{\mathbf{H}}(\boldsymbol{\psi}^{(r)}),$$

where

$$\begin{aligned}
\overline{\mathbf{H}}(\boldsymbol{\psi}^{(r)}) &= \frac{1}{m}\sum_{l=1}^{m}\left[-\nabla^2 L_c(\boldsymbol{\psi}^{(r)}|\mathbf{Y}_0, \bar{\mathbf{b}}_r^l)\right],\\
\underline{\mathbf{H}}(\boldsymbol{\psi}^{(r)}) &= \mathbf{H}_1(\boldsymbol{\psi}^{(r)}) - \mathbf{H}_2(\boldsymbol{\psi}^{(r)}),\\
\mathbf{H}_1(\boldsymbol{\psi}^{(r)}) &= \frac{1}{m}\sum_{l=1}^{m}\left\{\nabla L_c(\boldsymbol{\psi}^{(r)} \mid \mathbf{Y}_0, \bar{\mathbf{b}}_r^l)\right\}^{\otimes 2},\\
\mathbf{H}_2(\boldsymbol{\psi}^{(r)}) &= \left\{\frac{1}{m}\sum_{l=1}^{m}\nabla L_c(\boldsymbol{\psi}^{(r)}|\mathbf{Y}_0, \bar{\mathbf{b}}_r^l)\right\}^{\otimes 2},
\end{aligned}$$

and $\delta_r \in \{0,1\}$, $\bar{\mathbf{b}}_r^l = (\bar{\mathbf{b}}_{1,r}^l, \ldots, \bar{\mathbf{b}}_{I,r}^l)$, $\bar{\mathbf{b}}_{i,r}^l = \{\bar{\mathbf{b}}_{i,r}^1, \ldots, \bar{\mathbf{b}}_{i,r}^m\}$ are observations from the conditional distribution $p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi}^{(r)})$. In practice, we can set $\delta_r = 0$ at the beginning and change δ_r from 0 to 1 as $\boldsymbol{\psi}^{(r)}$ is closed to $\widehat{\boldsymbol{\psi}}$. Zhang et al. [**19**] suggested setting $\delta_r = 0$ and $\gamma_r = 1$ if $\max_i\{|\boldsymbol{\psi}_i^{(r-1)} - \boldsymbol{\psi}_i^{(r-2)}|\} > \tilde{\delta}$, or else to set $\delta_r = 1$ and $\gamma_r = 1/k$, where $\tilde{\delta}$ is a predetermined small value.

According to Bayes's theorem, we can obtain

$$p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi}) = \frac{p(\mathbf{y}_i, \bar{\mathbf{b}}_i)}{\int p(\mathbf{y}_i, \bar{\mathbf{b}}_i)\, d\bar{\mathbf{b}}_i}.$$

It can be readily seen that $\int p(\mathbf{y}_i, \bar{\mathbf{b}}_i) d\bar{\mathbf{b}}_i$ is an constant. So the conditional distribution of $p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi})$ is proportional to

$$\exp\left\{ -\frac{1}{2\sigma^2}\sum_{j=1}^{n_i} d(y_{ij}; \mu_{ij}) - \frac{1}{2}\bar{\mathbf{b}}_i^T \bar{\mathbf{b}}_i \right\}. \tag{16}$$

It can be seen from (16) that it is fairly difficult to simulate observations from $p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi})$ which is nonstandard and complex. The M-H algorithm [**4, 8, 13, 17**] is a well-known MCMC method which has been widely used to simulate observations from target density via the help of a proposal distribution from which it is easy to sample.

Following the methods of [**5, 21**], we choose multivariate normal distribution as the proposal distribution. The following algorithm is implemented to simulate observations from the target density $p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi})$: At the rth iteration of the M-H algorithm with a current value $\bar{\mathbf{b}}_{i,r}^{(t)}$ at the tth, a new candidate $\bar{\mathbf{b}}'_{i,r}$ is generated from $N(\bar{\mathbf{b}}_{i,r}^{(t)}, \tau^2 I)$, and the probability of accepting this new candidate is

$$\alpha = \min\left\{1, \frac{p(\bar{\mathbf{b}}'_{i,r} \mid \mathbf{y}_i, \boldsymbol{\psi}^{(r)})}{p(\bar{\mathbf{b}}_{i,r}^{(t)} \mid \mathbf{y}_i, \boldsymbol{\psi}^{(r)})}\right\},$$

where τ is an unknown parameter and applied to control the accepted rate of candidates from the proposal distribution in the entirely iterative process. Empirically, it can lead to a good result if the accepted rate is between 0.25 and 0.34, see [**3**].

Under some mild conditions, it follows from the reasoning in [**5**], which considered a problem with very similar nature, that $\Gamma_{\mathbf{r}}$ and $\boldsymbol{\psi}^{(r)}$ converge to $-\partial^2 L_o(\widehat{\psi} \mid \mathbf{Y}_0)/\partial\Psi\partial\Psi^T$ and the ML estimate $\widehat{\boldsymbol{\psi}}$, respectively. So, at convergence, $\Gamma_{\mathbf{r}}^{-1}$ can be regarded as an estimate of the covariance matrix of $\widehat{\Psi}$, see [**21**].

According to the above discussion, we propose the following strategy for finding the maximum likelihood estimate of $\boldsymbol{\psi}$.

Step 1. Choose initial values $\boldsymbol{\psi}^{(0)} \in \boldsymbol{\Psi}$;

Step 2. Given the rth iterative values $\boldsymbol{\psi}^{(r)}$, simulate $\{\bar{\mathbf{b}}_{i,r}^1, \dots, \bar{\mathbf{b}}_{i,r}^m\}$ from $p(\bar{\mathbf{b}}_i \mid \mathbf{y}_i, \boldsymbol{\psi}^{(r)})$, $i = 1, \dots, I$, via the M-H algorithm;

Step 3. Update $\boldsymbol{\psi}^{(r)}$ to $\boldsymbol{\psi}^{(r+1)}$ and $\boldsymbol{\Gamma}_{(r)}$ to $\boldsymbol{\Gamma}_{(r+1)}$ by (15);

Step 4. Repeat (2) and (3) until finding r satisfies $\max_i\{|\boldsymbol{\psi}_i^{(r-1)} - \boldsymbol{\psi}_i^{(r-2)}|\} < \delta$ (δ is a predetermined small value). Then set $\widehat{\boldsymbol{\psi}} = \boldsymbol{\psi}^{(r+1)}$. Finally, the estimate of $\boldsymbol{\Sigma}$ can be obtained via $\boldsymbol{\Sigma} = \mathbf{C}(\boldsymbol{\theta})\mathbf{C}(\theta)^T$.

4. Illustration: The ophthalmology study. In this section, we reanalyze the longitudinal proportional data from a prospective ophthalmology study on the use of intraocular gas (C_3F_8) in retinal repair surgeries [**14**], with a special focus on random effect. The outcome variable was the percent of gas left in the eye. The gas was injected into the eye before surgery for a total of 31 patients. The patients were then followed three to eight (an average of five) times over a three-month period, and the volume of the gas in the eye at the follow-up times was recorded as a percentage of the initial gas volume in that eye. An important issue was to estimate the kinetics of the disappearance of the gas, e.g., decay rate of the gas.

To begin with, the population-averaged, i.e., fixed, effects models in both [**16**] and [**18**, page 4] is

$$\operatorname{logit}(\mu_{ij}) = \beta_0 + \log(x_{ij1})\beta_1 + \log^2(x_{ij1})\beta_2 + x_{ij2}\beta_3,$$
$$x_{ij2} = \frac{\text{gas}_{ij} - 20}{5} = \begin{cases} -1 & \text{gas concentration level 15,} \\ 0 & \text{gas concentration level 20,} \\ 1 & \text{gas concentration level 25,} \end{cases}$$

where x_{ij1} is the time covariate of days after the gas injection and gas_{ij} is the covariate of gas concentration levels.

In our model, suppose that the patients have individual difference. The variable b_i denotes the ith patient's random effect, and $b_i \sim N(0, \sigma_1^2)$. Let $\bar{b}_i = b_i/\sigma_1$, where σ_1 is the square root of σ_1^2.

Our model form is given by

$$\operatorname{logit}(\mu_{ij}) = \beta_0 + \log(x_{ij1})\beta_1 + \log^2(x_{ij1})\beta_2 + x_{ij2}\beta_3 + \sigma_1\bar{b}_i.$$

We replace y_{ij} with

$$\tilde{y}_{ij} = \begin{cases} a & y_{ij} = 0, \\ y_{ij} - a & y_{ij} = 1, \\ y_{ij} & \text{else,} \end{cases}$$

where $a > 0$ is a small number to avoid zero denominators.

We choose $a = 0.001$ and run the SA-MCMC algorithm with $m = 20$, $\sigma^2 = 14.2$ [**16**] and $\tau = 4$ (which leads to the accepted rate of candidates 0.294), which started from the starting value $\widetilde{\boldsymbol{\beta}} = (2.6850, 0.0648, -0.3354, 0.3250)^T$ [**16**] and $\sigma_1 = 0.01$. The results of the model are listed in Table 1; meanwhile, the results of Song and Tan's [**16**] model with independent correlation structure are added for ease of comparison. The behaviors of $\boldsymbol{\psi}^{(r)}$ and the accepted rates ω are displayed in Figure 1.

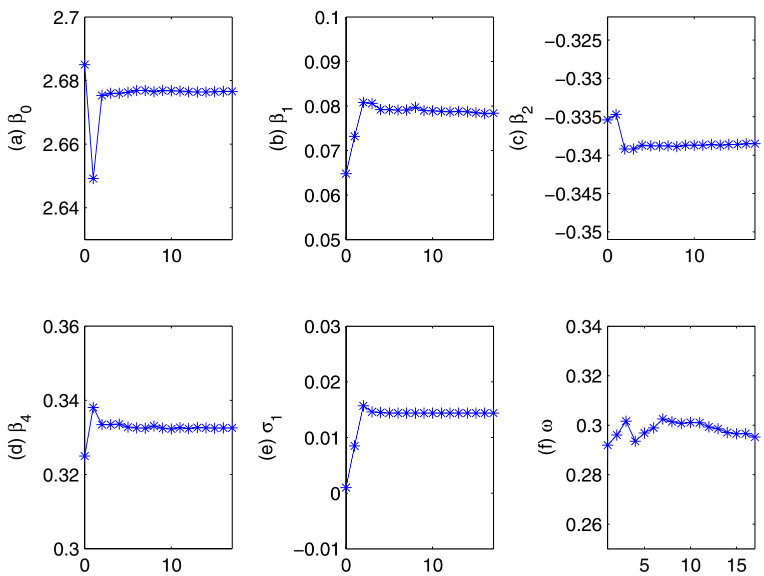

FIGURE 1. $\boldsymbol{\psi}^{(r)}$ and the accepted rate ω at each iteration.

TABLE 1. Results of the eye surgery data.

Parameter	Estimate	SE	SE [**16**]
β_0	2.6766	0.0180	0.3002
β_1	0.0784	0.0221	0.2491
β_2	-0.3385	0.0057	0.0662
β_3	0.3326	0.0122	0.1945
σ_1	0.0144	0.0000	—

APPENDIX

A. Proof of Theorem 2.1. (1) Differentiating (6) with respect to $\boldsymbol{\beta}$, we obtain

$$\frac{\partial L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta}} = -\phi \sum_{i=1}^{I} \left[\sum_{j=1}^{n_i} \left(\frac{\partial \mu_{ij}}{\partial \boldsymbol{\beta}^T} \right)^T \frac{\partial d(y_{ij}; \mu_{ij})}{\partial \mu_{ij}} \right] = -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \mathbf{e}_i.$$

(2) Differentiating (6) with respect to $\boldsymbol{\theta}$, we obtain

$$\begin{aligned}\frac{L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \theta_{tl}} &= -\phi \sum_{i=1}^{I} \left[\sum_{j=1}^{n_i} \frac{\partial \mu_{ij}}{\partial \theta_{tl}} \frac{d(y_{ij};\mu_{ij})}{\partial \mu_{ij}} \right] \\ &= -\phi \sum_{i=1}^{I} \left[\sum_{j=1}^{n_i} \mathbf{z}_{ij}^T \dot{\mathbf{C}}(tl) \bar{\mathbf{b}}_i \frac{\partial d(y_{ij};\mu_{ij})}{\partial \mu_{ij}} \right] \\ &= -\phi \sum_{i=1}^{I} \mathbf{h}_{itl}^T \mathbf{e}_i.\end{aligned}$$

(3) Differentiating (10) with respect to $\boldsymbol{\beta}$, we obtain

$$\begin{aligned}\frac{\partial^2 L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\phi \sum_{i=1}^{I} \frac{\partial}{\partial \boldsymbol{\beta}^T} \left\{ \left(\frac{\partial f(\mathbf{X}_i, \boldsymbol{\beta})}{\boldsymbol{\beta}^T} \right)^T \frac{\partial d(\mathbf{y}_i; \boldsymbol{\mu}_i)}{\boldsymbol{\mu}_i} \right\} \\ &= -\phi \sum_{i=1}^{I} \left\{ \left[\left(\frac{d(\mathbf{y}_i;\boldsymbol{\mu}_i)}{\boldsymbol{\mu}_i} \right)^T \right] \left[\frac{\partial^2 f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right] \right. \\ &\qquad \left. + \left(\frac{\partial f(\mathbf{X}_i,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right)^T \frac{\partial^2 d(\mathbf{y}_i;\boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\beta}^T} \right\} \\ &= -\phi \sum_{i=1}^{I} \left\{ [\mathbf{e}_i^T][\mathbf{W}_i] + \mathbf{D}_i^T \frac{\partial^2 d(\mathbf{y}_i;\boldsymbol{\mu}_i)}{\boldsymbol{\mu}_i} \frac{\mu_i}{\boldsymbol{\beta}^T} \right\} \\ &= -\phi \sum_{i=1}^{I} \left\{ [\mathbf{e}_i^T][\mathbf{W}_i] + \mathbf{D}_i^T \mathbf{V}_i \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}^T} \right\} \\ &= -\phi \sum_{i=1}^{I} \left\{ [\mathbf{e}_i^T][\mathbf{W}_i] + \mathbf{D}_i^T \mathbf{V}_i \mathbf{D}_i \right\}.\end{aligned}$$

(4) Differentiating (10) with respect to $\boldsymbol{\theta}$, we obtain

$$\begin{aligned}\frac{\partial^2 L_c(\boldsymbol{\psi} \mid \mathbf{Y}_c)}{\partial \boldsymbol{\beta} \partial \theta_{tl}} &= -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \frac{\partial^2 d(\mathbf{y}_i;\boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i \partial \theta_{tl}} \\ &= -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \frac{\partial^2 d(\mathbf{y}_i;\boldsymbol{\mu}_i)}{\partial \boldsymbol{\mu}_i \partial \boldsymbol{\mu}_i^T} \frac{\partial \boldsymbol{\mu}_i}{\partial \theta_{tl}} \\ &= -\phi \sum_{i=1}^{I} \mathbf{D}_i^T \mathbf{V}_i \mathbf{h}_{itl}.\end{aligned}$$

(5) Differentiating (11) with respect to $\boldsymbol{\theta}$, we obtain

$$\frac{\partial^2 L_c(\psi \mid \mathbf{Y}_c)}{\partial\theta_{tl}\partial\theta_{sr}} = -\phi \sum_{i=1}^{I} \left[\left(\frac{\mathbf{h}_{itl}}{\theta_{sr}} \right)^T \mathbf{e}_i + h_{itl}^T \frac{\partial^2 d(\mathbf{y}_i; \boldsymbol{\mu}_i)}{\partial\boldsymbol{\mu}_i \partial\theta_{sr}} \right]$$

$$= -\phi \sum_{i=1}^{I} \{ \mathbf{h}_{itl,sr}^T \mathbf{e}_i + \mathbf{h}_{itl}^T \mathbf{V}_i \mathbf{h}_{isr} \}.$$

REFERENCES

1. O.E. Barndorff-Nielsen and B. Jørgensen, *Some parametric models on the simplex*, J. Multivariate Anal. **39** (1991), 106–116.

2. P.J. Diggle, P.J. Heagerty, K.-Y. Liang and S.L. Zeger, *Analysis of longitudinal data*, 2nd ed., Oxford University Press, Oxford, 2002.

3. A. Gelman, G.O. Roberts and W.R. Gilks, *Efficient metropolis jumping rules*, in *Bayesian statistics* 5, J.M. Bernardo et al., eds., Oxford University Press, Oxford, 1996.

4. S. Geman and D. Geman, *Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images*, IEEE Trans. Pattern Anal. Machine Intelligence **6** (1984), 721–741.

5. M.G. Gu and F.H. Kong, *A stochastic approximation algorithm with Markov chain Monte-Carlo method for incomplete data estimation problems*, Proc. Natl. Acad. Sci. **95** (1998), 7270–7274.

6. D.A. Hartley and J.N.K. Rao, *Maximum likelihood estimation for mixed analysis of variance model*, Biometrika **54** (1967), 93–108.

7. D.A. Harville, *Maximum likelihood approaches to variance component estimation and to related problems*, J. Amer. Stat. Assoc. **72** (1977), 320–338.

8. W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their application*, Biometrika **57** (1970), 72–89.

9. B. Jørgensen, *The theory of dispersion models*, Chapman and Hall, London, 1997.

10. T.A. Louis, *Finding the observed information when using the EM algorithm*, Proc. Roy. Stat. Soc. **44** (1982), 226–233.

11. C.E. McCulloch, *Maximum likelihood variance components estimation for binary data*, J. Amer. Stat. Assoc. **89** (1994), 330–335.

12. ———, *Maximum likelihood algorithms for generalized linear mixed models*, J. Amer. Stat. Assoc. **92** (1997), 162–170.

13. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *Equations of state calculations by fast computing machine*, J. Chem. Physics **21** (1953), 1087–1091.

14. S.M. Meyers, J.S. Ambler, M. Tan, J.C. Werner and S.S. Huang, *Variation of perfluorproane disappearance after vitrectomy*, Retina **12** (1992), 359–363.

15. P.X.-K. Song, Z.G. Qiu and M. Tan, *Modelling heterogeneous dispersion in marginal models for longitudinal proportional data*, Biometrical Journal **46** (2004), 540–553.

16. P.X.-K. Song and M. Tan, *Marginal models for longitudinal continuous proportional data*, Biometrics **56** (2000), 496–502.

17. M.A. Tanner, *Tools for statistical Inference: methods for the exploration of posterior distribution and likelihood function*, 2nd ed., Springer-Verlag, New York, 1996.

18. C.L. Zhang, *Statistical diagnostics for the simplex distribution nonlinear models*, Master's dissertation, Southeast University, Nanjing, 2003 (in Chinese).

19. W.Z. Zhang, N.S. Tang and X.R. Wang, *Maximum likelihood estimation in nonlinear reproductive dispersion mixed models and stochastic approximation algorithm*, J. Biomath. **21** (2006), 270–278 (in Chinese).

20. ———, *Approximation to the log-likelihood function in nonlinear reproductive dispersion mixed models*, Appl. Math. J. Chinese Univ. **21** (2006), 321–331 (in Chinese).

21. H.T. Zhu and S.Y. Lee, *Analysis of generalized linear mixed models via a stochastic approximation algorithm with Markov chain Monte-Carlo method*, Statistics Comp. **12** (2002), 175–183.

22. X.P. Zong, J. Zhao, H.B. Wan and B.C. Wei, *Influence analysis on exponential nonlinear models with random effects*, Acta Math. Sci. **23** (2003), 11–22.

School of Mathematics and Statistics, Guizhou College of Finance & Economics, Guiyang, Guizhou 550004, China
Email address: zhwenzh97@hotmail.com

Rubber Research Institute, Chinese Academy of Tropical Agricultural Sciences, Danzhou, Hainan 571737, China
Email address: weihj1010@163.com