Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 154 (2019) 561–565

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology, ICICT 2019

Nested Inverse Gaussian Mixed-Effects Model for Longitudinal Data

Xing De Duan[a], Shi Zhang[a], Wen Zhuan Zhang[a,]*, Ying Wu[b]

[a] School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang, 550025, P. R. of China
[b] School of Mathematics and Statistics, Chuxiong Normal University, Chuxiong, 675000, P. R. of China

Abstract

Following [7], we introduce the nested Inverse Gaussian Mixed-Effects model to analyze right-skewed and continuous longitudinal data. The nested random effects don't follow a specific parameter distribution and rely only on the first two moments assumptions in our model. We apply the truly orthodox best linear unbiased predictor (BLUP) approach to estimate the nested random effects. We derive an optimal estimating equation for the regression parameters under the case of known BLUP of random effects. A real example for Framingham cholesterol data is presented to illustrate our proposed methodology.

Keywords: Estimating equation; Right-skewed data; Tweedie family; Random effects; Moment structures

1. Introduction

Skewed continuous longitudinal data frequently appear in many areas of research. Various skew normal models have been proposed to analyze skewed longitudinal data in recent years. For example, [5] for linear mixed models by substituting the skew-normal assumption of random effects for the normal assumption; [9] for Bayesian partial linear model; [1] for Skew-normal antedependence models; [8] for mixed effects model with the skew-normal and

* Corresponding author. Tel.: +86-139-8411-9979.
E-mail address: zhangwenzhuan@mail.gufe.edu.cn

1877-0509
10.1016/j.procs.2019.06.089

skew-t assumption of the distribution of responses and random effects. However, these approaches are generally computationally intensive.

The inverse Gaussian regression model is a powerful method for analyzing right-skewed continuous data. Following [6] and [7], we consider a class of nested Inverse Gaussian Mixed-Effects Model for right-skewed and continuous longitudinal data. [7] introduced the nested Tweedie mixed model based on an orthodox BLUP approach. Similarly, this orthodox BLUP approach to our models is still computationally simple and efficient. In addition, our approach consolidates conditional and marginal modeling interpretations.

2. Inverse Gaussian Mixed-Effects Model

2.1. The model

The inverse Gaussian distribution has the following probability density function

$$f(y;\mu,\tau^2)=\begin{cases}[2\pi\tau^2 y^3]^{-1/2}\exp\{-\dfrac{(y-\mu)^2}{2\mu^2\tau^2 y}\}, & \text{if } y>0\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

where $\mu>0$ is the mean and $\tau^2>0$ is the shape parameter ([3]). Therefore, we simply denote $Y\sim$) when a random variable Y is distributed as the inverse Gaussian distribution with mean μ and shape parameter τ^2. Also, we have $\mathrm{E}(Y)=\mu$ and $\mathrm{Var}(Y)=\tau^2\mu^3$. According to [7], the inverse Gaussian distribution belongs to the power-variance Tweedie family, thus we have $\mathrm{Tw}_3(\mu,\tau^2)=\mathrm{IG}(\mu,\tau^2)$, where Tw denotes the Tweedie family.

Let Y_{ij} be the outcome of the i th participant measured at time point t_{ij} $(i=1,\cdots$ $\cdots$ .Let $\mathbf{Y}=(Y_{11},\cdots$ $\cdots$ $\cdots$ denote the response vector. Suppose that $\mathbf{W}=(\mathbf{U}^T,\mathbf{V}^T)^T$ is the vector of the positive random effects where $\mathbf{U}$ and $\mathbf{V}$ represent $(U_1,\cdots$ and $(V_{11},\cdots$ $\cdots$ $\cdots$, respectively. In our paper, we call the random effects U_i and V_{ij} for $i=1,\cdots$ $\cdots$ as the first nested and second nested random effect, respectively. Similar to [7], we assume that

(i). $U_1,\cdots$ $\cdots$ are independent and identically distributed with the following first two moments

$$\mathrm{E}(U_i)=1 \text{ and } \mathrm{var}(U_i)=\sigma^2.$$

(ii). Given $\mathbf{U}$, the first two moment structure of $\mathbf{V}$ are given by

$$\mathrm{E}(V_{ij}\mid\mathbf{U})=U_i \text{ and } \mathrm{var}(V_{ij}\mid\mathbf{U})=\nu^2 U_i.$$

(iii). Given $\mathbf{W}$, $Y_{11},\cdots$ $\cdots$ $\cdots$ are conditionally independent with the following probability density function

$$Y_{ij}\mid\mathbf{W}\sim$$ ò $\cdots$ $\cdots$

where $\mu_{ij}=\exp(\mathbf{x}_{ij}^T\beta)$ in which $\mathbf{x}_{ij}=(1,x_{1,ij},\cdots$ is the covariate vector and $\beta=(\beta_0,\beta_1,\cdots$ is the coefficient vector. Clearly, each $Y_{ij}\mid\mathbf{W}$ depends only on μ_{ij}, V_{ij} and ò^2.

2.2. Moment structure

Here, we derive the moment structure of the response as follows. By using the law of total expectation and the law of total variance, the marginal means and variances of the response Y_{ij} are $E(Y_{ij})=\mu_{ij}$ and $Var(Y_{ij})=$ ò$^2\mu_{ij}^3+\mu_{ij}^2(\sigma^2+\nu^2)$, respectively. In addition, it is easy to interpret our conditional and marginal modelling approach ([4]). Therefore, we can obtain the marginal variance-covariance matrix of the response as follows

$$\mathrm{Cov}(Y_{ij}, Y_{i'j'}) = \begin{cases} \dot{\alpha}^2 \mu_{ij}^3 + \mu_{ij}^2(\sigma^2 + \nu^2) & i = i', j = j', \\ \mu_{ij}\mu_{ij'}\sigma^2 & i = i', j \neq j', \\ 0, & i \neq i'. \end{cases} \tag{2}$$

3. Estimating Model Parameters

The explicit expression of the BLUP of the first nested random effects based on [7] can be derived as

$$\hat{U}_i = [1 + \sigma^2 \sum_{j=1}^{n_i} \frac{Y_{ij}}{\epsilon^2 \mu_{ij}^2 + \nu^2 \mu_{ij}}] / [1 + \sigma^2 \sum_{j=1}^{n_i} \frac{1}{\epsilon^2 \mu_{ij} + \nu^2}].$$

Similarly, the BLUP of the second nested random effects has the following explicit expression

$$\hat{V}_{ij} = [\epsilon^2 \hat{U}_i + \nu^2 \mu_{ij}^{-2} Y_{ij}] / [\epsilon^2 + \nu^2 \mu_{ij}^{-1}].$$

3.1. Estimating regression parameters

To make inference on regression parameters β and dispersion parameters σ^2, ν^2 and $\dot{\alpha}^2$, we first construct the estimating equation for β; while the estimating expressions for σ^2, ν^2 and $\dot{\alpha}^2$ are presented in the later section. After a tedious calculation, we can obtain the following estimating equation for the regression parameters β:

$$\psi(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\mathbf{x}_{ij}}{\dot{\alpha}^2 \mu_{ij}^2} (Y_{ij} - \mu_{ij} \hat{V}_{ij}). \tag{3}$$

It follows from [7] that this estimating function for β is optimal. Further, we can obtain the sensitivity matrix $S(\beta)$ for β by calculating $S(\beta) = E\left[\frac{\partial \psi(\beta)}{\partial \beta}\right]$. It follows from [7] that the solution of the estimating equation $\psi(\beta) = 0$ tend to a multivariate normal distribution. To solve $\psi(\beta) = 0$, according to Newton scoring algorithm [2], we implement the following iterative formula

$$\beta^* = \beta - S^{-1}(\beta)\psi(\beta),$$

where the sensitivity matrix $S(\beta)$ has the following explicit expression

$$\begin{aligned} S(\beta) \quad &= -\sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{1}{\epsilon^2 \mu_{ij}} \mathbf{x}_{ij} \mathbf{x}_{ij}^T + \sum_{i=1}^{n} \sum_{j=1}^{n_i} \frac{\nu^2 \omega_{ij}}{\epsilon^2} (\frac{\mathbf{x}_{ij}}{\mu_{ij}})(\frac{\mathbf{x}_{ij}}{\mu_{ij}})^T \\ &+ \sum_{i=1}^{n} \frac{\sigma^2}{1 + \sigma^2 \sum_{j=1}^{n_i} \omega_{ij} \mu_{ij}^{-1}} (\sum_{j=1}^{n_i} \omega_{ij} \mu_{ij}^{-1} \mathbf{x}_{ij})(\sum_{j=1}^{n_i} \omega_{ij} \mu_{ij}^{-1} \mathbf{x}_{ij})^T \end{aligned}$$

where $\omega_{ij} = \mu_{ij} / (\dot{\alpha}^2 \mu_{ij} + \nu^2)$.

3.2. Estimating dispersion parameters

Here, we implement the method of adjusted Pearson estimator to estimate the dispersion parameters σ^2, ν^2 and $\dot{\alpha}^2$. It follows from [7] that the explicit expressions of the unbiased estimators of dispersion parameters are

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}[(\hat{U}_i - 1)^2 + c_i],$$

$$\hat{v}^2 = \frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n}\sum_{j=1}^{n_i}[(\hat{V}_{ij} - \hat{U}_i)^2 + c_{ij} + c_i - 2\hat{\delta}^2 c_i \omega_{ij}],$$

$$\hat{\delta}^2 = \frac{1}{\sum_{i=1}^{n} n_i}\sum_{i=1}^{n}\sum_{j=1}^{n_i}[\frac{(Y_{ij} - \mu_{ij}\hat{V}_{ij})^2}{\mu_{ij}^3} + c_{ij}\mu_{ij}^{-1}],$$

where $c_i = \frac{\sigma^2}{1+\sigma^2\sum_{j=1}^{n_i}\omega_{ij}\mu_{ij}^{-1}}$, and $c_{ij} = \epsilon^2\omega_{ij}(v^2 + \epsilon^2 c_i \omega_{ij})$.

4. Framingham Cholesterol Data

To illustrate our proposed methodologies, we reanalyzed this longitudinal data. The cholesterol levels are measured at 0, 2, 4, 6, 8 and 10 years for 133 participants (60 men and 73 women) ([5]). Our scientific interest of this study is to assess the effects of gender, baseline age and visiting time on the cholesterol levels. More specifically, let Y_{ij} denote the cholesterol level of the i th participant measured at the j th time for $i = 1,\cdots$ and $j = 1,\cdots$, where $Y_{ij} = Y_{ij}/100$. It can be seen from Figure 1 that the right-skewed shape of the cholesterol levels can be captured well via inverse Gaussian distribution; therefore, we considered the following model:

$$\begin{cases} Y_{ij} \mid \mathbf{W} \sim \quad\quad ,\delta^2 V_{ij}^{-2}) \\ \log(\mu_{ij}) = \beta_0 + \beta_1 \mathrm{Sex}_i + \beta_2 \mathrm{Age}_i + \beta_3 \mathrm{t}_{ij}, \end{cases}$$

Where Sex_i takes 1 for male or 0 for female, Age_i denotes the i th participant's baseline age, $t_{ij} = (\mathrm{time} - 5)/10$.

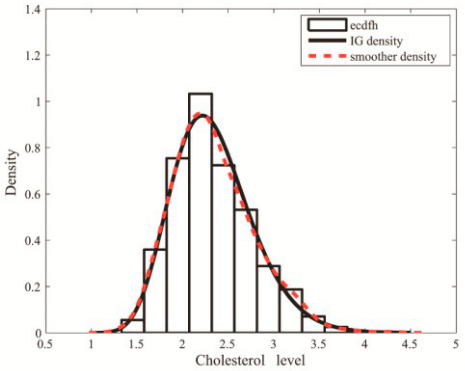

Fig.1. The histogram, smoother density of cholesterol level and density of inverse Gaussian distribution

For comparison, the Framingham cholesterol data are fitted by our proposed model described in Section 2 and the skewed normal mixed model proposed by [5], respectively. Table 1 display the estimates as well as the corresponding standard errors of regression and dispersion parameter. The covariate Age and Time have the positive significant effect on the cholesterol level under both models.

Table 1. Parameter estimates for Framingham cholesterol data.

Parameter	Our proposed model		Lin & Lee	
	Estimate	SE[a]	Estimate	SE
Intercept	0.5447	0.0740	1.2898	0.1731
Sex	-0.0027	0.0267	-0.0382	0.0614
Age	0.0072	0.0017	0.0151	0.0036
Time	0.1215	0.0098	0.2341	0.0273
σ^2	0.0220			
$\grave{o}^2$	0.0008			
ν^2	0.0071			

a: standard error.

5. Conclusion

In this paper, we have introduced the inverse Gaussian regression model with nest random effects for positive and right-skewed continuous longitudinal data. The explicit expressions of BLUPs of random effects and sensitivity matrix for the regression parameters are available in our estimated approach . In addition, our method relaxes the distributional assumption of random effects and depends only on the first two moments assumptions of the random effects; therefore, these assumptions cover a wide range of parametric distributions. The Framingham cholesterol example demonstrated the usefulness of our approach.

Acknowledgements

We are grateful to Professor Renjun Ma for constructive suggestions during the preparation of the manuscription. This work was supported by the National Natural Science Foundation of China (11501073), the Applied Basic Research Project of Yunnan Province (2014FD051, 2016FD087).

References

1. Chang, S. C., and Zimmerman, D. L. Skew-normal antedependence models for skewed longitudinal data. Biometrika 2016; 363-376.
2. Jørgensen, B., Labouriau, R. and Lundbye-Christensen, S. Linear growth curve analysis based on exponential dispersion models. Journal of the Royal Statistical Society Series B (1996); 573-592.
3. Jørgensen, B. The Theory of Dispersion Models. London: Chapman and Hall; 1997.
4. Lee, Y. and Nelder, J.A. Hierarchical generalized linear models (with discussion). Journal of the Royal Statistical Society Series B 1996; 619-678.
5. Lin, T.I. and Lee, J.C. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. Statistics in Medicine 2008; 1490-1507.
6. Ma, R. An orthodox BLUP approach to generalized linear mixed models. PHD Thesis. Department of Statistics, University of British Columbia, Vancouver; 1999.
7. Ma, R., Jørgensen, B. Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. Journal of the Royal Statistical Society Series B 2007; 625-641.
8. Patrick, M., Nadja, K. and Thomas, K. Bayesian Multivariate Distributional Regression With Skewed Responses and Skewed Random Effects. Journal of Computational and Graphical Statistics 2018; 602-611.
9. Tang, Y., Sinha, D., Pati, D., Lipsitz, S. and Lipshultz, S. Bayesian partial linear model for skewed longitudinal data. Biostatistics 2015; 441-453.