The multivariate Gamma-GLG model from the random intercept Gamma model with random effect nonnormal

Lizandra C. Fabio
Federal University of Bahia, Salvador - Bahia, Brazil - lizandrafabio@gmail.com

Francisco J.A. Cysneiros
Federal University of Pernambuco, Brazil- cysneiros@de.ufpe.br

Gilberto A. Paula
University of São Paulo, São Paulo, Brazil - giapaula@ime.usp.br

Abstract

We propose in this paper a random intercept gamma model in which the random effect is assumed to follow a generalized log-gamma (GLG) distribution. This flexibilization in which has been suggested by Fabio et al. (2012) allows distributions for the random effect skew to the right and skew to the left and has the normal distribution as a particular case. For a particular parametrization for the GLG distribution and specifying the adequate link function, we derive a new continuous multivariate distribution named Gamma-GLG and its moments. We also developed an iterative process based on Newton Raphson methods for the parameters estimates of the multivariate model. The deviance function and residual analysis are proposed and an applications with real data is given for illustration.

Keywords: Gamma-GLG multivariate model; Generalized linear models; Generalized log-gamma distribution; Random-effect models; Residual analysis.

1 Introduction

The flexibilization of the random effect distribution in mixed models has received great attention in the last years, particularly since the works by Lee and Nelder (1996, 2001) in which different combinations between the conditional response and appropriate random effect distributions was proposed, but under a hierarchical framework. Under the marginal framework one has, for instance, the works by Molenberghs *et al.* (2007, 2010, 2012) in which different combinations between the normal and appropriate conjugate distributions are considered for modeling the mean in generalized linear mixed models. Recently, Efendi and Molenberghs (2013) suggested the flexibilization of the random effect distribution in multilevel mixed-effect models. Motivations for considering different distributions for the random effects than the normal one are the possibility of bias reduction for the variance component estimates (see for instance, Alonso *et al.*, 2008; Litière *et al.*, 2008) and obtaining closed-form expression for the marginal distribution (see, for example, Fabio *et al.*, 2012).

2 Multivariate Gamma-GLG model

We assume the following hierarchical structure:

(i) $y_{ij}|b_i \overset{\text{ind}}{\sim} \text{Gamma}(u_{ij}, \phi)$,

(ii) $\log({u_{ij}}^{1}) = \eta_{ij} + b_i$ and

(iii) $b_i \overset{\text{iid}}{\sim} \text{GLG}(0, \lambda, \lambda)$.

Performing some algebraic manipulations derive the marginal distribution of $\boldsymbol{y}_i = (\boldsymbol{y}_{i1}^{>}, \ldots, \boldsymbol{y}_{im_i}^{>})^{>}$ given (1) where $\boldsymbol{\theta} = (\boldsymbol{\beta}^{>}, \phi)^{>}$ and $\phi = \lambda \ \ ^{2}$, $\phi > 0$

$$f(\boldsymbol{y}_i;\boldsymbol{\theta}) = \frac{\Gamma\left(\phi(m_i+1)\right)}{\Gamma(\phi)^{(m_i+1)}} \frac{\prod_{j=1}^{m_i} (\delta_{ij} y_{ij}) \ \ ^{1} \delta_{ij}}{\left(1+\sum_{j=1}^{m_i} \delta_{ij} y_{ij}\right) \ \ ^{(m_i+1)}}, \tag{1}$$

in that $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im_i})^{>}$ is the mean vector of the ith cluster (suject) and its observed mean respectively. The pdf (1) is similar to Inverted Direchelet distribution (Kotz *et al.*, 2000 (p. 509) and Wang *et al.*, 2011, (p.175)). The multivariate distribution moments showed in (1) can be deduced from the intercept random effect Gamma-GLG.
Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{im_i})^{>}$ the outcome vector with m_i measured from the ith cluster (subject) for $i = 1, \ldots, n$ and $\boldsymbol{y}_i \sim \mathrm{GGLG}(\mu_{ij}, \phi)$. We propose the following Gamma-GLG model

(i) $\boldsymbol{y}_i \overset{\text{ind}}{\sim} \mathrm{GGLG}(\boldsymbol{\mu}_i, \phi)$

(ii) $\log(\mu_{ij}) = \boldsymbol{x}_{ij}^{>}\boldsymbol{\beta} = \eta_{ij}$

in that $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im_i})^{>}$ is the mean vector for $i = 1, \ldots, n$, $\eta_{ij} = \boldsymbol{x}_{ij}^{>}\boldsymbol{\beta}$ is the linear predictor with $\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijp})^{>}$ containing values of explanatory variables and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{>}$ is the parameter vector and ϕ is the dispersion parameter.
The log-likelihood function for this multivariate model is given by

$$\begin{aligned} \ell(\boldsymbol{\beta}, \phi; \boldsymbol{y}) = & \sum_{i=1}^{n} \log(\Gamma(\phi(m_i+1))) - \sum_{i=1}^{n} (m_i+1)\log(\Gamma(\phi)) + \phi \sum_{i=1}^{n}\sum_{j=1}^{m_i} \log\left(c \ \ \mu_{ij}^{\ \ 1}\right) \\ & + (\phi-1)\sum_{i=1}^{n}\sum_{j=1}^{m_i} \log(y_{ij}) - \phi(m_i+1)\sum_{i=1}^{n} \log\left(1 + c \ \ \sum_{j=1}^{m_i} y_{ij}\mu_{ij}^{\ \ 1}\right), \end{aligned} \tag{2}$$

where $c \ \ = c(\phi) = \phi\frac{\Gamma(\ \ 1)}{\Gamma(\ )}$ for values integer no negativos, $\phi > 0$. The score function and the Fisher information matrix may be obtained for the $\boldsymbol{\beta}$ and ϕ parameters and an interative process can be performed to get the maximum likelihood estimates. Similarly to Fabio *et al.* (2012) we define a deviance as a goodness-of-fit measure, for the Gamma-GLG model. After some algebraic manipulations and by assuming that ϕ is fixed and $y_{ij} > 0$, $\forall i, j$, we can express the deviance as $D(\boldsymbol{y}; \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} d^2(\boldsymbol{y}_i, \widehat{\boldsymbol{\mu}}_i)$, in that

$$d^2(\boldsymbol{y}_i, \hat{\boldsymbol{\mu}}_i, \phi) = 2\phi\left\{\sum_{j=1}^{m_i} \log\left(\frac{\hat{\mu}_{ij}}{y_{ij}}\right) + (m_i+1)\log\left(\frac{1 + c \ \ \sum_{j=1}^{m_i} y_{ij}\hat{\mu}_{ij}^{\ \ 1}}{1+m_i}\right)\right\} \tag{3}$$

3 Application

We will present a comparative study among diabetic groups discussed by Cysneiros and Paula (2004). Three groups were considered for the experiment, control group, diabetic group without complications and diabetic group with hypertension. For each patient the response was a physical task measured in the times 1, 2, 3, 4, 5, 6, 8 and 10 min. Let y_{ij} be the observed physical task for the ith patient of the ith group at the time j. We will assume the multivariate Gamma-GLG model, in that (i) $y_{ij} \overset{\text{ind}}{\sim} \mathrm{GGLG}(\mu_{ij}, \phi)$, (ii) $\log(\mu_{ij}) = \alpha + \beta_i$, for $i = 1, 2, 3$ and $\beta_1 = 0$, with proposal of compare the three groups. We used the software R to perform the maximization of the multivariate model log-likelihood function in (2). In the Figure 1 we can verify the presence of outlier observations from two subjects $\neq 20$ and $\neq 21$. We will suggest o deviance component residual for the ith subjet to show the adequacy of model in this application.

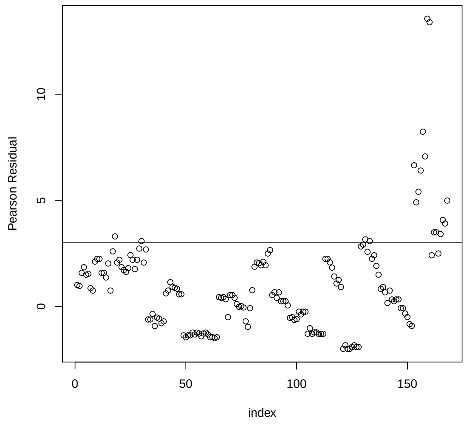

Figure 1: Pearson residual for the observations

References

Alonso, A., Litière, S., and Molenberghs, G. (2008). A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. *Computational Statistics and Data Analysis*, **52**, 4474–4486.

Cysneiros, F. and Paula, G. (2004). One-sided tests in linear models with multivariate t-distribution. *Communications in Statistics, Simulation and Computation*, **33**, 747–771.

Efendi, A. and Molenberghs, G. (2013). A multilevel model for hierarchical, repeated, and overdispersed time-to-event outcomes and its estimation estrategies. *Journal of Biopharmaceutical of Statistics*, **23**, 1420–1434.

Fabio, L., Paula, G. A., and Castro, M. (2012). A Poisson mixed model with nonormal random effect distribution. *Computational Statistics Data Analysis*, **56**, 1499–1510.

Kotz, S., Balakrishnam, N., and Johnson, N. (2000). *Continuous Multivariate Distributions*. Jhon Wiley, New York, 2 edition.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B*, **58**, 619–656.

Lee, Y. and Nelder, J. A. (2001). Hierarchical generalized linear models : a synthesis of generalized linear models, random effect models and structured dispersions. *Biometrika*, **88**, 987–1006.

Litière, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferencial procedures in generalized linear mixed models. *Statistics in Medicine*, **27**, 3125–3144.

Molenberghs, G., Verbeke, G., and Demètrio, C. G. B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, 513–531.

Molenberghs, G., Verbeke, G., Demètrio, C. G. B., and Vieira, A. M. C. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

Molenberghs, G., Verbeke, G., Iddi, S., and Demètrio, G. B. (2012). A combined beta and normal random-effects model for repeated, overdispersed binary and binomial data. *Journal of Multivariate Analysis*, **111**, 94–109.

Wang, k., Tian, G., and Tang, M. (2011). *Direchlet and Related Distributions-Theory, Methods and Applications*. Jhon Wiley, United Kingdon, 1 edition.