Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion

Peer-reviewed author version

RAKHMAWATI, Trias Wahyuni; MOLENBERGHS, Geert; VERBEKE, Geert & FAES, Christel (2016) Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion. In: JOURNAL OF APPLIED STATISTICS, 44 (4), pag. 620-641.

DOI: 10.1080/02664763.2016.1182128
Handle: http://hdl.handle.net/1942/22669

Local Influence Diagnostics for Generalized Linear Mixed Models With Overdispersion

Trias Wahyuni Rakhmawati[1], Geert Molenberghs[1,2], Geert Verbeke[2,1], Christel Faes[1,2]

[1] I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium
[2] I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium

E-mail for correspondence: `triaswahyuni.rakhmawati@uhasselt.be`

Abstract: Since the seminal paper by Cook and Weisberg (1982), local influence, next to case deletion, has gained popularity as a tool to detect influential subjects and measurements for a variety of statistical models. For the linear mixed model the approach leads to easily interpretable and computationally convenient expressions, not only highlighting influential subjects, but also which aspect of their profile leads to undue influence on the model's fit (Verbeke and Lesaffre 1998). Ouwens, Tan, and Berger (2001) applied the method to the Poisson-normal generalized linear mixed model (GLMM). Given the model's non-linear structure, these authors did not derive interpretable components but rather focused on a graphical depiction of influence. In this paper, we consider GLMMs for binary, count, and time-to-event data, with the additional feature of accommodating overdispersion whenever necessary. For each situation, three approaches are considered, based on: (1) purely numerical derivations; (2) using a closed-form expression of the marginal likelihood function; and (3) using an integral representation of this likelihood. The methodology is illustrated in case studies of A Clinical Trial in Epileptic Patients.

Keywords: Boundary condition; Case deletion; GLMM; Combined model; Local Influence.

1 Introduction

Next to linear mixed models (LMM) for hierarchical Gaussian data (Verbeke and Molenberghs 2000), generalized linear mixed models (GLMM) have become a tool for routine use for the analysis of a hierarchical data of a variety of data types over the last twenty years (Molenberghs and Verbeke, 2005). Like with every statistical model, after formulating and fitting a model, an assessment of model fit and a diagnostic analysis is advisable. In this paper, we are concerned with the detection of influential subjects. A large variety of diagnostic tools is available for linear and generalized linear models. Cook and Weisberg and and Chatterjee and Hadi (1988) provide early treatises. In classical linear regression, Cook's distances (Cook

1977a, 1977b, 1979) have been used extensively. Linear mixed models, unlike linear models, generally do not allow for closed-form parameter estimators. Further, residual analysis is not straightforward, given the presence of both fixed-effect and random-effects covariates, so that even uniquely defining residuals is not possible. For these and related reasons, Lesaffre and Verbeke (1998) chose local influence (Cook 1986, Backman, Nachtsheim, and Cook 1987) to examine influence in linear mixed models.
In this study, we extend local influence for the GLMM in several ways. First, we consider outcomes of binary, count, and time-to event type. Second, using the extension proposed by Molenberghs, Verbeke, and Demétrio (2007) and Molenberghs *et al* (2010), we flexibly allow for overdispersion in the GLMM, by introducing conjugate random effects, in addition to normal ones. This model is referred to as the combined model. Third, apart from numerical derivations of local influence, we examine two alternative routes: (a) closed forms for the marginal likelihood such as proposed in Molenberghs *et al* (2010) and (b) the marginal likelihood with integral form. The closed forms in (a) do not always exist; while they are available for the probit-(beta-)normal, Poisson-(gamma-)normal, and Weibull-(gamma-)normal, they are not for the logit-(beta-)normal. Even when they do, they may be somewhat unwieldy and therefore, route (b) is more promising.

2 Local Influence for GLMM

Local influence was presented by Cook (1986) and used by several authors since. The impact of individuals and measurements on the analysis is assessed by comparing standard maximum likelihood estimates with those resulting from slightly perturbing the contribution of an individual or a measurement. Lesaffre and Verbeke (1998) introduced an influence assessment paradigm for the linear mixed model.
Cook (1986) derived a convenient computational scheme.Let $\boldsymbol{\Delta}_i$ be the s-dimensional vector of second-order derivatives of log-likelihood $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$, w.r.t. perturbation ω_i and all components of $\boldsymbol{\theta}$, and evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Also, write Δ for the $s \times r$ matrix with $\boldsymbol{\Delta}_i$ in the ith column. Let $\ddot{L}$ denote the $s \times s$ matrix of second derivatives of $\ell(\boldsymbol{\theta})$, evaluated at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. For any unit vector $\boldsymbol{h}$ in Ω, it follows that: $C_h = 2\left| \boldsymbol{h}'\Delta'\ddot{L}\Delta\boldsymbol{h} \right|$. Various choices for $\boldsymbol{h}$ have received attention. First, as will be done here, one can focus on subject i only, by choosing $\boldsymbol{h} = \boldsymbol{h_i}$, the zero vector with a sole 1 in the ith position. Local influence then is $C_i \equiv C_{h_i} = 2\left| \boldsymbol{\Delta}_i'\ddot{L}\boldsymbol{\Delta}_i \right|$. Lesaffre and Verbeke (1998) showed that local influence C_i can be re-expressed as

$$C_i = 2||\ddot{L}||\ ||\boldsymbol{\Delta_i}||^2 \cos(\varphi_i), \tag{1}$$

where φ_i is the angle between $\mathrm{vec}(-\ddot{L})$ and $\mathrm{vec}(\boldsymbol{\Delta_i}\boldsymbol{\Delta_i}')$.

The integral-based approach can be used as an alternative way to alleviate complexities with the explicit marginal likelihood expressions. The marginal density corresponding to the linear mixed model is defined as: $\widetilde{f}(\boldsymbol{y}_i) = \int \widetilde{f}(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{b}_i)\widetilde{f}(\boldsymbol{b}_i|D)\ d\boldsymbol{b}_i$, with the log-likelihood contribution of the ith individual takes the form: $\ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{N} \widetilde{f}(\boldsymbol{y_i})$.
For count data,the first derivative of log-likelihood contribution for ith subject as followed:

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial \boldsymbol{\beta}} = \sum_{j=1}^{n_i} \{y_{ij} - E(y_{ij}|\boldsymbol{b}_i)\} \boldsymbol{x}_{ij} = \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}, \tag{2}$$

$$\frac{\partial \ell_i(\boldsymbol{\beta}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk}) \left\{ (D^{-1})_{jk} - (D^{-1}D^{-1})_{jk} \sum_{k=1}^{q} \mathrm{Var}(b_{ik}) \right\}, \tag{3}$$

where d_{jk} is a component of D and δ_{jk} is one if j is equal to k, and zero otherwise. Interpretable expressions can now be derived using (1). It showed

$$\begin{aligned} ||\boldsymbol{\Delta}_i||^2 &= \left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)\left(\sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}\right)' \\ &+ \sum_{k,l} \left\{ -\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b}_i) \right\}^2 . \end{aligned}$$

Let $C_i = C_{1i} + C_{2i}$ with:

$$C_{1i} = 2||\ddot{L}||\ ||\boldsymbol{r}_i\boldsymbol{x}_i||^2 \cos(\varphi_i), \tag{4}$$

$$C_{2i} = \frac{1}{2}||\ddot{L}||\ ||(D^{-1})_{kl} - (D^{-1}D^{-1})_{kl}\mathrm{Var}(\boldsymbol{b}_i)||^2 \cos(\varphi_i), \tag{5}$$

where $\boldsymbol{r}_i\boldsymbol{x}_i = \sum_{j=1}^{n_i} r_{ij}\boldsymbol{x}_{ij}$. Note that C_{1i} and C_{2i} are the contributions of subject i to local influence C_i from $\boldsymbol{\beta}$ and D, respectively. Reconstructing the component C_{1i} and C_{2i} leads to the interpretable components that can be described local influence. Hence, the interpretable components of C_i in the case of the Poisson-normal model can be described using the 'length of the fixed effect' ($||\boldsymbol{x}_i\boldsymbol{x}_i'||$), the 'squared length of the residual' ($||\boldsymbol{r_i}||^2$), and the 'squared of random effect variability' ($\mathrm{Var}(\boldsymbol{b}_i)^2$).
In binary cases, the local influence for both probit and logit normal models have been derived.The first derivatives for probit-normal model are:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\xi}} = [I - (\boldsymbol{X}_i\boldsymbol{\xi})^{-1}]\boldsymbol{X}_i, \tag{6}$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = \frac{3}{2}L^{-1}\left(I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right), \tag{7}$$

where $M_i = \left(D^{-1} + Z_i'Z_i\right)^{-1}$. It also follows that

$$||\boldsymbol{\Delta}_i||^2 = [I - (\boldsymbol{X}_i\boldsymbol{\xi})^{-1}]^2\boldsymbol{X}_i\boldsymbol{X}_i' + \sum_{k,l} \frac{9}{4L^2}\left(I_{n_i} - Z_iM_iM_i'(D^{-1}D^{-1})_{jk}Z_i'\right)^2 .$$

Thus, also for this case, the components $||\boldsymbol{X_i}||^2$ and $||\boldsymbol{Z_i}\boldsymbol{Z_i}'||^2$ turn up. Evidently, the same binomial expression is used, but now with $\text{logit}(\lambda_{ij}) = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i$.The derivatives of logit-normal model take the form:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\xi}} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \int \frac{1}{1+\exp(\mu_{ij})} \tilde{\tau}(\boldsymbol{b}_i|\boldsymbol{y}_i) d\boldsymbol{b}_i, \tag{8}$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk}) \left\{ (D^{-1})_{jk} - (D^{-1}D^{-1})_{jk} \text{Var}(\boldsymbol{b}_i) \right\}, \tag{9}$$

where $\mu_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\xi} + \boldsymbol{z}_{ij}'\boldsymbol{b}_i$. It also follows that

$$||\boldsymbol{\Delta}_i||^2 \propto \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right) \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right)' + \sum_{k,l} \left(-\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl} \text{Var}(\boldsymbol{b}_i) \right)^2.$$

Reconstructing the fixed- and random-effects components, respectively, like in the Poisson case, leads to $C_{1i} = 2||\ddot{L}||\ ||\boldsymbol{x}_i||^2 \cos(\varphi_i)$ and C_{2i} as in (5). Hence, the interpretable components of C_i for the logit-normal model can be described using the length of fixed effect ($||\boldsymbol{x}_i||^2$) and the squared random-effects variability, $\text{Var}(\boldsymbol{b}_i)^2$ (i.e., the sum of all variances), in analogy with the Poisson-normal model. The same is true for the Weibull-normal model, as will be seen next.

The first derivative for Weibull case take the form:

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial \boldsymbol{\xi}} = \sum_{j=1}^{n_i} \boldsymbol{x}_{ij} - \lambda \sum_{j=1}^{n_i} y_{ij}^{\rho} \boldsymbol{x}_{ij} \exp(\boldsymbol{\mu}_{ij}), \tag{10}$$

$$\frac{\partial \ell_i(\boldsymbol{\xi}, D)}{\partial d_{jk}} = -\frac{1}{2}(2-\delta_{jk}) \left[(D^{-1})_{jk} - (D^{-1}D^{-1})_{jk} \text{Var}(\boldsymbol{b}_i) \right], \tag{11}$$

where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. It further follows that

$$\begin{aligned} ||\boldsymbol{\Delta}_i||^2 &= \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right) \left(\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \right)' - 2\sum_{j=1}^{n_i} \boldsymbol{x}_{ij} \boldsymbol{Q}_i' + \boldsymbol{Q}_i \boldsymbol{Q}_i' \\ &+ \sum_{k,l} \left\{ -\frac{1}{2}(D^{-1})_{kl} + \frac{1}{2}(D^{-1}D^{-1})_{kl} \text{Var}(\boldsymbol{b}_i) \right\}^2, \end{aligned}$$

where $\boldsymbol{Q}_i = \lambda \sum_{j=1}^{n_i} y_{ij}^{\rho} \boldsymbol{x}_{ij} \exp(\boldsymbol{\mu}_{ij})$.

Like in the Poisson-normal and binary-normal cases, a decomposition $C_i = C_{1i} + C_{2i}$ follows, with $C_{1i} = 2||\ddot{L}||\ \left\{ ||\boldsymbol{x}_i||^2 - 2\boldsymbol{x}_i\boldsymbol{Q}_i + ||\boldsymbol{Q}_i||^2 \right\} \cos(\varphi_i)$ and C_{2i} as in (5). Hence, interpretable components analogous to the earlier settings arise.

3 Application

The Epileptic dataset consisted of 89 patients with 2 different group of treatments, placebo and a new anti-epileptic drug (AED). Patients were followed (double-blind) during 16 weeks (some patients until 27 weeks). The outcome of interest is the number of epileptic seizures experienced during the most recent week. Poisson-normal (P-N) model as well as the combined model with gamma random effect (PGN) have been fitted as follow:

$$\ln(\boldsymbol{\lambda}_{ij}) = \begin{cases} (\beta_{00} + b_i) + \beta_{01} t_{ij} & \text{if placebo} \\ (\beta_{10} + b_i) + \beta_{11} t_{ij} & \text{if treated,} \end{cases} \tag{12}$$

where Y_{ij} represent the number of epileptic seizures patient i experiences during week j, t_{ij} is the time point at which Y_{ij} has been measured and the random intercept $b_i \sim N(0, d)$. Parameter estimates are given in Table 1. Figure 1 contain index plots (versus patient ID) for various local influence analyses. The top row represents local influence for (P-N) model, yet in below rows for (PGN) model. Patients #38, #49, and #62 stand out with large total influence C_i when compared to other patients. Importantly, influences show a major drop when switching from (P-N) to (PGN). To get further insight as to why these subject have higher influence than others, plots with interpretable components are given in Figure 2: 'squared length of the fixed effects' $||\boldsymbol{x}_i \boldsymbol{x}_i'||$, 'squared length of the residual' $||\boldsymbol{r_i}||^2$, and 'random-effect variability' $\mathrm{Var}(b_i)^2$. It is hardly surprising that #38 stands out in terms of $||\boldsymbol{r_i}||^2$. Influences on #49 and #62 are less pronounced.

4 Discussion

It has been showed from this study that the influential subject for hierarchical model can be detect using local influence approach. And it was found that the combined model can be used to reduced the influence effect of the subject. Moreover, the interpretable components can be use as the tools to evaluate in which way the influence subject affect the estimation in modeling process.

References

Cook, R.D (1986). Assessment of Local Influence. *Journal of the Royal Statistical Society, Series B*, **48**, 133 – 169.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570 – 582.

Molenberghs G. *et al* (2010). A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects. *Statistical Science*, **25(3)**, 325 – 347.

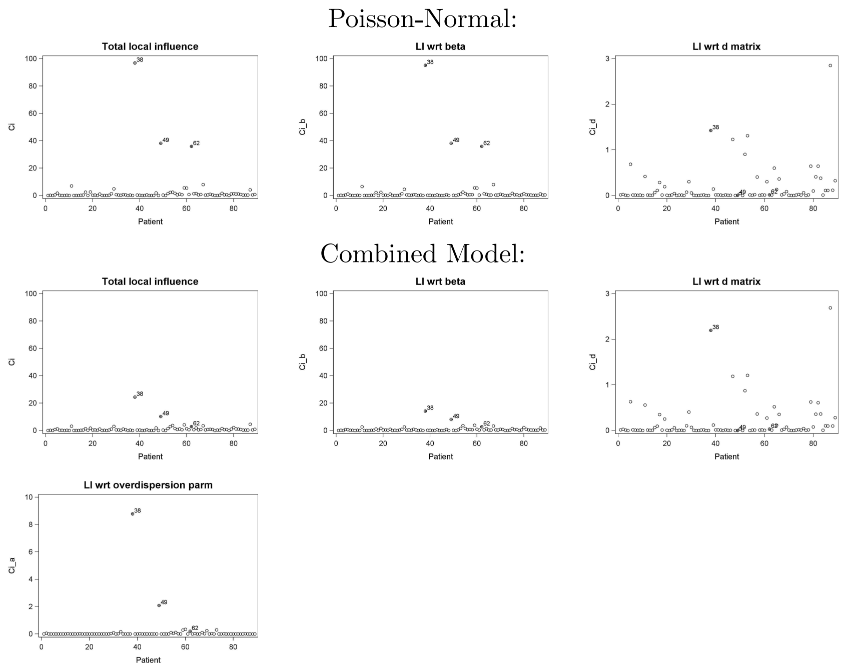

FIGURE 1. Plot of Local Influence

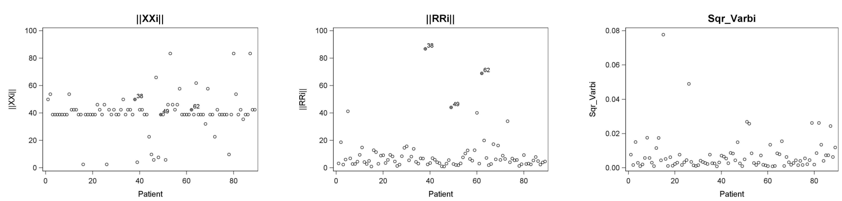

FIGURE 2. Plot of Interpretable components of Local Influence

TABLE 1. Parameter estimates (standard errors) for the P-N and PGN models.

Effect	Par.	P-N	PGN
Interc. plac.	β_{00}	0.818(0.168)	0.911(0.176)
Slope plac.	β_{01}	-0.014(0.004)	-0.025(0.008)
Interc. treat.	β_{10}	0.648(0.170)	0.656(0.178)
Slope treat.	β_{11}	-0.012(0.004)	-0.012(0.007)
Treat. eff.	$\beta_{11}-\beta_{10}$	0.002(0.006)	0.013(0.011)
Treat. eff.	β_{11}/β_{10}	0.840(0.398)	0.475(0.335)
Std. rand. int.	σ	1.076(0.086)	1.063(0.087)
Overdisp. par.	α		2.464(0.211)