Computational Statistics and Data Analysis 82 (2015) 126–136

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Cook's distance for generalized linear mixed models

Luis Gustavo B. Pinho [a], Juvêncio S. Nobre [a,*], Julio M. Singer [b]

[a] Universidade Federal do Ceará, DEMA, Campus do Pici. Fortaleza, CE, 60440-900, Brazil
[b] Universidade de São Paulo, IME, Rua do Matão, 1010. São Paulo, SP, 05508-090, Brazil

ARTICLE INFO

Article history:
Received 22 November 2013
Received in revised form 12 August 2014
Accepted 15 August 2014
Available online 1 September 2014

Keywords:
Diagnostics
GLMM
Influence
Leverage

ABSTRACT

We consider an extension of Cook's distance for generalized linear mixed models with the objective of identifying observations with high influence in the predicted conditional means of the response variable. The proposed distance can be decomposed into factors that help to distinguish between influence on the estimation of fixed effects and on the prediction of random effects. Joint and conditional influence are also considered. A first-order approximation is proposed for more efficient computation and a Monte Carlo simulation is considered to evaluate the efficacy of the proposal. An application to a dataset obtained from the literature is presented to show how such tools can be used in practice.

1. Introduction

By including random effects in the linear predictor, Generalized Linear Mixed Models (GLMMs) constitute a flexible tool to analyze data using distributions in the exponential family. In repeated measures studies, for example, where each unit may contribute with more than one observation, the random effects allow the modeling of individual unit behavior (Zeger et al., 1988). This class of models is also useful to analyze overdispersed data (Breslow, 1984). This, however, is accomplished at the expense of a more complicated maximum likelihood estimation process since it may be necessary to integrate over several dimensions. For details on GLMMs, the reader is referred to Breslow and Clayton (1993), among others.

Whenever statistical models are considered, care should be taken to verify their assumptions and adequacy to the data. Diagnostic tools developed for such purposes may be classified in two broad categories. The first, termed residual analysis, is useful to verify assumptions about the distributions of the random elements and to identify observations (or units) with atypical values. The second, called sensitivity analysis, is employed to evaluate the behavior of the components of the model and predicted values when observations (or units) are perturbed or deleted. In the context of traditional linear models (normal, homoskedastic and independent observations), diagnostic methods have been addressed by many authors, among which we mention Cook (1977), Hoaglin and Welsch (1978), Belsley et al. (1980) and Cook and Weisberg (1982). Extensions and generalizations to linear mixed models are considered in Beckman et al. (1987), Hilden-Minton (1995), Lesaffre and Verbeke (1998), Tan et al. (2001), Demidenko (2004), Demidenko and Stukel (2005), Nobre and Singer (2007), Gumedze et al. (2010) and Nobre and Singer (2011), among others. Diagnostics for GLMMs are still not fully explored; some attempts have been made by Xiang et al. (2002), Zhu and Lee (2003), Tchetgen and Coull (2006) and Abad et al. (2010).

Using an approach similar to the one in Tan et al. (2001), we extend the ideas of Xiang et al. (2002) to allow evaluation of the influence of observations on both the estimation of fixed effects and prediction of random effects separately. This is an important step when GLMMs are used for prediction purposes.

* Corresponding author. Tel.: +55 85 33669155.
E-mail addresses: juvencio@ufc.br, juvenciosantos@gmail.com (J.S. Nobre), jmsinger@ime.usp.br (J.M. Singer).

http://dx.doi.org/10.1016/j.csda.2014.08.008
0167-9473/

This paper is structured as follows. In Section 2, we start with a brief description of a GLMM. In Section 3, we review the ideas underlying Cook's distance and describe some available extensions for linear mixed models and GLMMs. In Section 4, we present our proposal as well as a first-order approximation to speed up the required computation. In Section 5, we illustrate its use with a hypothetical dataset and consider a Monte Carlo simulation to assess its efficacy. A real dataset application is presented in Section 6 and concluding remarks are addressed in Section 7.

2. Generalized linear mixed models

Let Y_{ij} denote the jth observation from unit i, where $1 \leq i \leq m,\ 1 \leq j \leq n_i$ and $\sum_{i=1}^{m} n_i = n$ and assume that every Y_{ij} follows distributions of the same type in the exponential family. Assume further that the probability density function of Y_{ij} depends on a vector of q non-observable random effects, $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{iq})^\top$, through the parameter η_{ij} so that conditionally on $\boldsymbol{b}_i$, we may write

$$Y_{ij}|\,\boldsymbol{b}_i \sim f(y_{ij}|\,\boldsymbol{b}_i, \eta_{ij}, \phi) = \exp\left[\frac{y_{ij}\,\eta_{ij} - c(\eta_{ij})}{a(\phi)} + r(y_{ij}, \phi)\right], \tag{1}$$

where $a(\cdot)$, $c(\cdot)$ and $r(\cdot,\cdot)$ are known functions and ϕ represents a dispersion parameter. The parameter η_{ij} in model (1) relates the expected value of $Y_{ij}|\,\boldsymbol{b}_i$, say μ_{ij}, to a set of explanatory variables by

$$g(\mu_{ij}) = \eta_{ij}, \quad \eta_{ij} = \boldsymbol{x}_{ij}^\top\boldsymbol{\beta} + \boldsymbol{z}_{ij}^\top\boldsymbol{b}_i$$

where $\boldsymbol{x}_{ij}$ is a $p \times 1$ vector of non-stochastic explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression parameters (fixed effects), $\boldsymbol{z}_{ij}$ is a $q \times 1$ vector of non-stochastic variables (often a portion of $\boldsymbol{x}_{ij}$) and $g(\cdot)$ is a monotonic, differentiable link function. In this context, η_{ij} is the linear predictor. We will assume that $\boldsymbol{b}_i$ follows a multivariate normal distribution with null mean vector and a $q \times q$ non-negative definite covariance matrix $\boldsymbol{D}_i$. Usually, $\boldsymbol{D}_i$ is a function of a few covariance parameters. For example, $\boldsymbol{D}_i = \text{diag}\{\sigma_1^2, \ldots, \sigma_q^2\}$. Under certain regularity conditions (Casella and Berger, 2001), which are satisfied by members of the exponential family, we have

$$\mathbb{E}(Y_{ij}|\boldsymbol{b}_i) = \frac{\partial\, c(\eta_{ij})}{\partial\, \eta_{ij}}, \quad \text{and} \quad \text{Var}(Y_{ij}|\boldsymbol{b}_i) = [a(\phi)]^{-1}\frac{\partial^2 c(\eta_{ij})}{\partial\ \eta_{ij}.^2}$$

Letting

$$\begin{aligned}
&\boldsymbol{y}_i = \left(y_{i1}, y_{i2}, \ldots, y_{in_i}\right)^\top, && \boldsymbol{y} = \left(\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top, \ldots, \boldsymbol{y}_m^\top\right)^\top\\
&\boldsymbol{\eta}_i = \left(\eta_{i1}, \eta_{i2}, \ldots, \eta_{in_i}\right)^\top, && \boldsymbol{\eta} = \left(\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top, \ldots, \boldsymbol{\eta}_m^\top\right)^\top\\
&\boldsymbol{X}_i = (\boldsymbol{x}_{i1}; \boldsymbol{x}_{i2}; \ldots; \boldsymbol{x}_{in_i})^\top, && \boldsymbol{X} = (\boldsymbol{X}_1^\top; \boldsymbol{X}_2^\top;\ \ldots; \boldsymbol{X}_m^\top)^\top,\\
&\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}; \boldsymbol{z}_{i2}; \ldots; \boldsymbol{z}_{in_i})^\top, && \boldsymbol{Z} = \bigoplus_{1\leq i\leq m} \boldsymbol{Z}_i,\\
&\boldsymbol{b} = (\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top, \ldots, \boldsymbol{b}_m^\top)^\top,
\end{aligned}$$

the model can be expressed as

$$\begin{aligned}
&\boldsymbol{Y} \sim f(\boldsymbol{y}|\,\boldsymbol{b}, \boldsymbol{\eta}, \phi) = \exp\left[\frac{\boldsymbol{\eta}^\top\boldsymbol{y} - C(\boldsymbol{\eta})}{A(\phi)} + R(\boldsymbol{y}, \phi)\right],\\
&\boldsymbol{b} \sim \mathcal{N}_{mq}\,(\boldsymbol{0}, \boldsymbol{D})\,,
\end{aligned}$$

where $A(\cdot)$, $C(\cdot)$ and $R(\cdot,\cdot)$ are known functions, $\boldsymbol{D} = \bigoplus_{1\leq i\leq m}\boldsymbol{D}_i$ and the operator $\bigoplus$ denotes the direct sum. Given a function $m = m(\theta)$, we denote $\dot{m}_{\mathbf{u}} = \partial m/\partial\mathbf{u}$ and $\ddot{m}_{\mathbf{uv}} = \partial^2 m/\partial\mathbf{u}\partial\mathbf{v}^\top$. When the derivatives are computed at $\mathbf{u} = \widehat{\mathbf{u}}$ and $\mathbf{v} = \widehat{\mathbf{v}}$ we indicate them by $\dot{m}_{\widehat{\mathbf{u}}}$ and $\ddot{m}_{\widehat{\mathbf{u}}\widehat{\mathbf{v}}}$. Using this notation, under the aforementioned regularity conditions we may write $\mathbb{E}(\boldsymbol{Y}|\boldsymbol{b}) = \boldsymbol{\mu} = \dot{C}_{\boldsymbol{\eta}}$ and $\text{Var}(\boldsymbol{Y}|\boldsymbol{b}) = [A(\phi)]^{-1}\ddot{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}$, and

$$g(\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}.$$

Estimation of the fixed effects and prediction of the random effects can be achieved by maximum likelihood methods. Some strategies to accomplish this are described in Breslow and Clayton (1993), McGilchrist (1994) and McCulloch and Searle (2001). A hierarchical approach is considered in Lee and Nelder (1996) and some Bayesian methods are presented in Zeger and Karim (1991) or in Zhao (2006). Although these authors also consider estimation of the variance components of the model, we will omit the details on this topic since it is not our main focus.

The diagnostic method proposed in this paper is based on the estimation method described in McGilchrist (1994). Let

$$l_1 = \sum_{\substack{1\leq i\leq m\\ 1\leq j\leq n_i}}\left[\frac{y_{ij}\,\eta_{ij} - c(\eta_{ij})}{a(\phi)} + r(y_{ij}, \phi)\right]$$

$$l_2 = -\frac{1}{2}\sum_{1\leq i\leq m}\left[q\log(2\pi) + \log|\boldsymbol{D}_i| + \boldsymbol{b}_i^\top\boldsymbol{D}_i^{-1}\boldsymbol{b}_i\right]$$

respectively denote the log-likelihoods of $\boldsymbol{Y}|\boldsymbol{b}$ and $\boldsymbol{b}$, as functions of $\boldsymbol{\beta}$ and $\boldsymbol{b}$. We shall use the term likelihood for l_2 even though it is not a true likelihood, since $\boldsymbol{b}$ is non-observable. The joint (pseudo) log-likelihood of $\boldsymbol{Y}|\boldsymbol{b}$ and $\boldsymbol{b}$ is given by $l = l_1 + l_2$.

Using the first and second order derivatives of l, namely

$$\begin{aligned}
\dot{l}_{\boldsymbol{\beta}} &= A^{-1}(\phi)\boldsymbol{X}^{\top}(\boldsymbol{Y}-\boldsymbol{\mu}),\\
\dot{l}_{\boldsymbol{b}} &= A^{-1}(\phi)\boldsymbol{Z}^{\top}(\boldsymbol{Y}-\boldsymbol{\mu}) - \boldsymbol{D}^{-1}\boldsymbol{b},\\
\ddot{l}_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -A^{-1}(\phi)\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{X},\\
\ddot{l}_{\boldsymbol{b}\boldsymbol{b}} &= -A^{-1}(\phi)\boldsymbol{Z}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z} - \boldsymbol{D}^{-1}, \quad \text{and}\\
\ddot{l}_{\boldsymbol{\beta}\boldsymbol{b}} &= -A^{-1}(\phi)\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z}
\end{aligned}$$

we can use the Newton–Raphson algorithm to maximize l and obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$, the maximum likelihood estimator of $\boldsymbol{\beta}$ and predictor of $\boldsymbol{b}$, respectively. The corresponding iterative process is

$$\widehat{\boldsymbol{\theta}}_{k+1} = \widehat{\boldsymbol{\theta}}_k - \ddot{l}^{-1}_{\widehat{\boldsymbol{\theta}}_k\widehat{\boldsymbol{\theta}}_k}\dot{l}_{\widehat{\boldsymbol{\theta}}_k}, \quad k = 1, 2, \ldots,$$

where $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^{\top}, \widehat{\mathbf{b}}^{\top})^{\top}$. Estimators of the covariance matrix are obtained by maximizing the log-likelihood with respect to covariance parameters. The reader may refer to McGilchrist (1994) and references therein for further details.

3. Some existing diagnostic techniques

To evaluate changes in the estimated vector of parameters when observations are deleted, Cook (1977) proposed one of the most popular measures of influence for standard linear models; the so called Cook's distance for the deletion of the jth observation is defined as

$$CD_j = \frac{\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(j)}\right)^{\top}\left(\boldsymbol{X}^{\top}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(j)}\right)}{p\sigma^2} = \frac{\left(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}}_{(j)}\right)^{\top}\boldsymbol{V}^{-1}\left(\widehat{\boldsymbol{y}} - \widehat{\boldsymbol{y}}_{(j)}\right)}{p\sigma^2},$$

where the subscript (j) indicates deletion of the jth observation and $\sigma^2\boldsymbol{V}$ represents the covariance matrix of $\boldsymbol{Y}$. For linear mixed models, Christensen et al. (1992) and Banerjee and Frees (1997) suggested to use Cook's distance much in the same way as it is used for linear models. However, Tan et al. (2001) showed evidence that their suggestion is not always able to measure correctly the influence of observations in this context. For the conditional homoskedastic model, i.e., for which $\mathrm{Var}(\boldsymbol{Y}|\boldsymbol{b}) = \sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ represents the identity matrix of order n, Tan et al. (2001) proposed the following conditional distance

$$\begin{aligned}
CD^{\mathrm{cond}}_{ij} &= \sum_{i=1}^{m}\frac{\boldsymbol{d}_i^{\top}\boldsymbol{d}_i}{\sigma^2\left[(m-1)q+p\right]},\\
\boldsymbol{d}_i &= \left[\left(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i\right) - \left(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}_{(j)} + \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_{i(j)}\right)\right].
\end{aligned}$$

This distance can be decomposed into the sum of three terms. The first term is designed to detect observations that may have high influence on the estimates of the fixed effects; the second is used to assess the impact of observations on the prediction of random effects and the last term is a measure of the relationship between the changes in fixed and random effects estimators and predictors, respectively. This last term was usually close to zero in several simulation rounds, similar to that was observed by Tan et al. (2001).

For GLMM, Xiang et al. (2002) proposed the conditional distance

$$CD_j = \frac{\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(j)}\right)^{\top}\left[-\ddot{l}_{\hat{\beta}\hat{\beta}}\right]\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(j)}\right)}{p[A(\widehat{\phi})]^{-1}}. \tag{2}$$

Although the initial objective was to identify influential units, it may also be used to identify single observations which may influence the estimates of the fixed effects. To simplify computation, these authors also proposed an approximation for (2). Using simulation, they showed that the approximated distance identifies influential observations efficiently.

4. An extension of Cook's distance for GLMMs

The distance proposed by Xiang et al. (2002) is directed at identifying observations (or units) that may influence $\widehat{\boldsymbol{\beta}}$. This is useful when the fixed effects are the main focus of the analysis. GLMMs, however, are also used for prediction and even though the observations identified by (2) may have an impact on the predicted values, distortions may also occur due to observations which exert strong influence on the predicted random effects. Thus, for the cases where the focus is on prediction, we propose the use of a distance similar to that in Tan et al. (2001) to assess the potential influence of an

observation on the predicted values, namely,

$$CD_{ij} = \frac{\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{(ij)}\right)^{\top} \mathrm{Var}\left(\boldsymbol{Y}|\boldsymbol{b}\right) \left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{(ij)}\right)}{[A(\phi)]^{-1}[(m-1)q+p]} \tag{3}$$

which can be decomposed as

$$CD_{ij} = CD^1_{ij} + CD^2_{ij} + CD^3_{ij}$$

with

$$CD^1_{ij} = \frac{\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(ij)}\right)^{\top} \left[-\ddot{l}_1(\widehat{\boldsymbol{\beta}})\right] \left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(ij)}\right)}{A^{-1}(\phi)[(m-1)q+p]}$$

$$CD^2_{ij} = \frac{\left(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(ij)}\right)^{\top} \left[-\ddot{l}_1(\widehat{\boldsymbol{b}})\right] \left(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(ij)}\right)}{A^{-1}(\phi)[(m-1)q+p]}$$

$$CD^3_{ij} = \frac{2\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(ij)}\right)^{\top} \left[-\ddot{l}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}})\right] \left(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(ij)}\right)}{A^{-1}(\phi)[(m-1)q+p]},$$

where $\ddot{l}_1(\widehat{\boldsymbol{b}}) = \partial^2 l_1/\partial \boldsymbol{b}\partial \boldsymbol{b}^{\top}$ and $\ddot{l}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}) = \partial^2 l_1/\partial \boldsymbol{\beta}\partial \boldsymbol{b}^{\top}$ are evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and $\boldsymbol{b} = \widehat{\boldsymbol{b}}$. The term CD^1_{ij} is a proportional to (2) and is related to the fixed effects; the observations which may influence the predictors of the random effects are identified via CD^2_{ij}; finally, CD^3_{ij} is related to the covariance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$ and is expected to be close to zero, as in Tan et al. (2001). This decomposition is demonstrated in the Appendix.

Because re-estimating the parameters for every deleted observation is not practical, Xiang et al. (2002) use an approach proposed by Pregibon (1981) in a different context and consider a first-order Taylor series expansion of l around $\widehat{\boldsymbol{\beta}}$ to approximate $\left(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(j)}\right)$. This can be directly used to assess the influence on the estimates of the fixed effects; we consider a similar approximation for prediction of the random effects.

The estimator $\widehat{\boldsymbol{\beta}}_{(ij)}$ and predictor $\widehat{\boldsymbol{b}}_{(ij)}$ are such that $\dot{l}_{(ij)}(\widehat{\boldsymbol{\beta}}_{(ij)}) = 0$ and $\dot{l}_{(ij)}(\widehat{\boldsymbol{b}}_{(ij)}) = 0$. Using Taylor expansions around $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}$, we obtain

$$\dot{l}_{(ij)}(\widehat{\boldsymbol{\beta}}) - \ddot{l}_{(ij)}(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(ij)}) \approx 0$$

$$\dot{l}_{(ij)}(\widehat{\boldsymbol{b}}) - \ddot{l}_{(ij)}(\widehat{\boldsymbol{b}})(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(ij)}) \approx 0,$$

which implies

$$(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(ij)}) \approx \left[\ddot{l}_{(ij)}(\widehat{\boldsymbol{\beta}})\right]^{-1} \dot{l}_{(ij)}(\widehat{\boldsymbol{\beta}})$$

$$(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(ij)}) \approx \left[\ddot{l}_{(ij)}(\widehat{\boldsymbol{b}})\right]^{-1} \dot{l}_{(ij)}(\widehat{\boldsymbol{b}})$$

where the derivatives are specified in the previous section. Since it is usually computationally faster to perform some matrix multiplications than to execute optimization routines, we expect some gains with the proposed method.

Plots of the values computed by (3) may be used to identify influential observations with respect to the predicted values $\widehat{\mu}_{ij}$, while plots of CD^1_{ij}, CD^2_{ij} and CD^3_{ij} can be used to identify the nature of this influence. For example, large values of CD_{ij} obtained in a way where the value of CD^2_{ij} is larger than those corresponding to CD^1_{ij} and CD^3_{ij} suggest that the observation ij is only influential with respect to the predicted values of the ith unit.

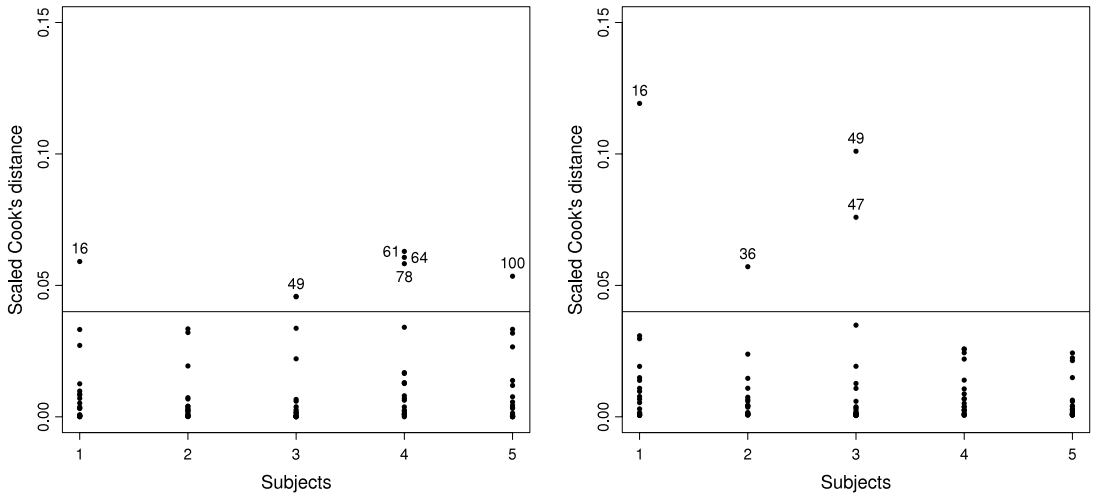
To assess the influence of a unit, instead of removing only individual observations, we delete the set S_i of the observations from the ith unit and compute (3) in the same way as we did with an individual observation. Here, the first order approximations reduce to

$$(\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{(S_i)}) \approx \left[\ddot{l}_{(S_i)}(\widehat{\boldsymbol{\beta}})\right]^{-1} \dot{l}_{(S_i)}(\widehat{\boldsymbol{\beta}})$$

$$(\widehat{\boldsymbol{b}} - \widehat{\boldsymbol{b}}_{(S_i)}) \approx \left[\ddot{l}_{(S_i)}(\widehat{\boldsymbol{b}})\right]^{-1} \dot{l}_{(S_i)}(\widehat{\boldsymbol{b}}),$$

where $l_{(S_i)}$ represents the logarithm of the likelihood function based on a sample that does not include observations from the ith unit. Notice that it would be impossible to assess directly the effects of the deletion of a whole unit in the estimation of the model parameters if the approximation were not used, since it would not be possible to predict the random effects for that unit.

In order to detect correctly the influential observations, one must be aware of masking, that occurs when the influence of an observation is affected by other observations. This may happen when the observations are jointly but not individually influential, as suggested by Chatterjee and Hadi (1986). Atkinson (1986) suggests that masking may also occur when the effects of the influence of an observation are not detected until another observation is deleted. The first case is known as



(a) Measure from Xiang et al. (2002).

(b) Proposed measure.

Fig. 1. (a) Influence on the fixed effects. (b) Influence on the linear predictor.

joint influence and the second, as conditional influence. For joint influence in standard regression models, Lawrance (1995) proposed to observe the changes in the estimation of the model parameters when a set of observations is deleted. For conditional influence, this author suggested that one should evaluate the influence of an observation after the deletion of another. Xiang et al. (2002) used the same approach for assessing the joint and conditional influence with respect to the fixed effects in GLMMs. We use a similar idea to evaluate the influence on linear predictors, and consequently on the predicted conditional means. Let S be a set of observations and define

$$CD_S = \frac{\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{(S)}\right)^\top \mathrm{Var}\left(\boldsymbol{Y}|\boldsymbol{b}\right)\left(\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{(S)}\right)}{[A(\phi)]^{-1}[(m-1)q+p]} \tag{4}$$

where $\widehat{\boldsymbol{\eta}}_{(S)}$ has a similar interpretation as $\widehat{\boldsymbol{\beta}}_{(S)}$. Then, CD_S can be used to identify jointly influential observations. It may even be used for assessing the joint influence of a whole unit as mentioned before. Next, define

$${}_{(ab)}CD_{ij} = \frac{\left(\widehat{\boldsymbol{\eta}}_{(ab)} - \widehat{\boldsymbol{\eta}}_{(ab,ij)}\right)^\top \mathrm{Var}\left(\boldsymbol{Y}|\boldsymbol{b}\right)\left(\widehat{\boldsymbol{\eta}}_{(ab)} - \widehat{\boldsymbol{\eta}}_{(ab,ij)}\right)}{[A(\phi)]^{-1}[(m-1)q+p]}, \tag{5}$$

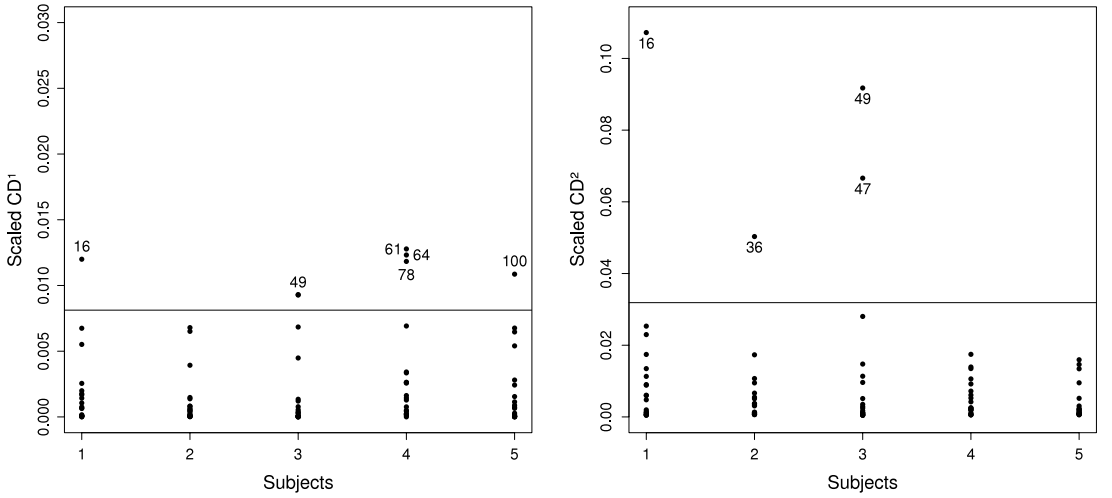
where the left subscript (ab) indicates that a value was obtained in the absence of the bth observation from unit a, and notice that this can be used to identify conditionally influential observations. The quantities (4) and (5) can be decomposed and approximated in the same way as (3).

All the values of (2)–(5) presented in the next two sections use the first-order approximation. Whenever we present (2) or (3) and their decompositions, the values will be scaled by $\sum_{i,j} CD_{(ij)}$ to bring them to the (0, 1) interval and facilitate the comparison of influence in different contexts.

5. Simulation

In this section we consider a hypothetical dataset to illustrate the use of (3). The dataset is simulated from a Poisson GLMM with a logarithmic link function. There are 20 observations from each of 5 different units. The explanatory variable is the same for every unit and consists of equally spaced values ranging from 1 to 3. Random effects are generated from a $\mathcal{N}_5(\mathbf{0}, \mathbf{I}_5)$ distribution. The linear predictor is $\eta_{ij} = (1 + b_i) + 0.5x_{ij}$. All models were fitted in R using the `lmer()` function from the `lme4` package.

The distance proposed in Xiang et al. (2002) will be used here for comparison to (3) even though it is our understanding that their proposal was not designed to detect influence with respect to the linear predictor. Lacking a proper tool for assessing the influence on the predicted means, practitioners may be tempted to use (2) for the wrong purpose. This practice can be misleading, since it will not detect influence with respect to the prediction of random effects. In Fig. 1 we compare (2) and (3) for the hypothetical dataset. A threshold of four times the average distance is used to identify possible influential observations. An alternative is to compute the percentiles of the (2) and (3) and consider as influential the top 20% or 30% values, for example. The chosen threshold should reflect the level of criticism of the decision process. The two extremes are: investigating every point in detail, which may be impractical; not investigating any point in detail. In any case, visual inspection of the values might be valuable. In Fig. 1(a) the potentially influential observations are labeled 16, 49, 61, 64, 78 and 100. In Fig. 1(b) the flagged observations are labeled 16, 36, 47 and 49. Observations 16 and 49 were considered influential by both approaches, suggesting that they are influential with respect to the estimation of the fixed



(a) First component of the proposed measure.

(b) Second component of the proposed measure.

Fig. 2. First two components of the proposed Cook's distance.

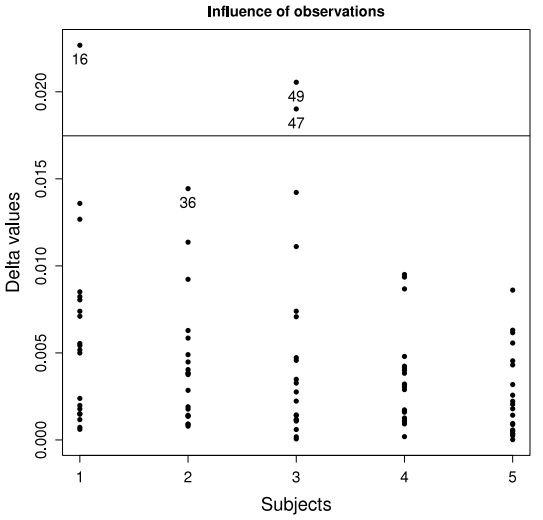


Fig. 3. Values for δ_{ij}.

effects. Observations 36 and 47, on the other hand, are not flagged in Fig. 1(a), possibly indicating they are influential only with respect to the prediction of the random effects. This is in accordance with Fig. 2, where they stand out in the plots of the second component of the proposed distance. Figs. 2(a) and 1(a) are very similar since, as mentioned before, CD_{ij}^1 is proportional to (2).

Let $\delta_{ij} = (|\widehat{\boldsymbol{\eta}} - \widehat{\boldsymbol{\eta}}_{(ij)}|)^\top \mathbf{1}_n/n$, where $\mathbf{1}_n$ is a vector with all n elements equal to 1. A large value of δ_{ij} suggests that the observation ij is influential. In Fig. 3 all values of δ_{ij} for the simulated example are displayed. It requires re-fitting the model for every deleted observation and is therefore impractical, but we use it here to visualize how well (3) detects correctly the influential observations. Here, an observation is considered influential if δ_{ij} is larger than four times the average value, although we recognize that this threshold is arbitrary and should be considered on an *ad hoc* basis.

The four most influential observations were the ones flagged by the proposed distance. The distance proposed by Xiang et al. (2002) detected observations 16 and 49, since they had large influence with respect to the estimation of the fixed effects, but missed observation number 47, which possibly had impact on the prediction of the random effects.

A Monte Carlo simulation with 10 000 replicas was performed to study the efficacy of (3) to identify influential observations. We expect (3) to have a better performance than (2), since the latter is part of the former. Consider the model

$$
\begin{aligned}
&Y_{ij}|b_i \sim \text{Poisson}[\exp(\eta_{ij})],\\
&\eta_{ij} = 1 + 2x_{ij} + z_{ij}b_i,\\
&\boldsymbol{b} \sim \mathcal{N}_5(\mathbf{0}, \mathbf{I}_5),
\end{aligned}
$$

Table 1
Results of the Monte Carlo simulation. The numbers in the third and fourth columns represent the relative frequency in which the modified observations were detected correctly by the two proposals in the total amount of runs, with g representing the portion of modified observations.

n	g	Xiang et al. (2002)	The new proposal
$n = 50$	2%	61.20%	95.30%
	5%	50.30%	91.55%
	10%	34.34%	73.06%
$n = 100$	2%	76.85%	95.05%
	5%	53.66%	84.52%
	10%	36.45%	65.82%
$n = 150$	2%	77.6%	93.37%
	5%	54.97%	83.56%
	10%	36.14%	63.42%

$1 \leq i \leq 5,\ 1 \leq j \leq n/5$, where the value of the explanatory variable for each unit is drawn from a uniform distribution over the [1, 3] interval and each z_{ij} is drawn from a uniform distribution over [0, 1]. Based on this model, we generated samples of sizes $n = 50$, 100 and 150. After generating each sample, we chose some observations at random and replaced the corresponding value of z_{ij} with a random draw from a uniform distribution over [4, 5], so that in the majority of the simulation rounds they should be the ones with largest influence regarding the random effects. For each sample size, we considered simulations with $g = 2\%$, 5% and 10% modified values. The model was fitted and the approximated values of (2) and (3) were computed. We considered an observation y_{ij} to be influential when the corresponding value for CD^1_{ij} or CD^2_{ij} lies among the 30% largest values of CD^1 or CD^2. For (2), we need only to consider CD^1_{ij}, which is proportional to (2). Table 1 shows the relative frequency with which the modified observations were identified correctly by (3) and (2).

This simulation is by no means exhaustive, but clearly suggests that in some situations (3) is better at detecting the correct observations. As the number of modified observations increases, (3) and (2) are less efficient in identifying them. It is worth noticing that the sample size increases without increasing the number of units so that the modified observations are less likely to be influential within the unit and thus, less likely to be detected by CD^2.

6. A practical example

The "third party claims" dataset we considered in this section was analyzed in de Jong and Heller (2008). Third party insurance is a compulsory insurance for drivers in Australia. It insures the owner of the vehicle against injuries caused to others as a result of an accident. This dataset consists of 176 observations of the total number of accidents and insurance claims that occurred in a twelve month period between 1985 and 1986. Each observation corresponds to a local government area, or municipal council area, in the state of New South Wales. These areas are divided into 13 larger territories.

An accident may or may not generate an insurance claim. Since there is always a time lag between the accident and the claim, insurers are interested in predicting the number of claims from the number of accidents. This is an important information for evaluating the balance between risks and monetary reserves of an insurer, and has a great impact on the insurance premium. It is very important to identify observations which possibly distort the predicted values and this can be achieved by the diagnostic procedure described here.

Analysis of this dataset can be seen in, for example, Stasinopoulos and Rigby (2007) and de Jong and Heller (2008, Chapters 6 and 10). The Poisson distribution is a natural choice for this kind of data, however both references show that these data are overdispersed, that is, its variance is higher than it is expected from a Poisson variate. Several alternative approaches are available. In de Jong and Heller (2008) a GLM is used for modeling the response from a negative binomial distribution which can accommodate overdispersion, and in later sections of the book the GAMLSS (see also Stasinopoulos and Rigby, 2007) approach is used to account for the territorial division; a Poisson-inverse-Gaussian (PIG) distribution and a zero-adjusted inverse Gaussian (ZAIG) distribution were also used, each of them with their particular advantages and followed by an interesting discussion.

To illustrate the usefulness of the technique proposed we consider a GLMM with negative binomial response as follows.

$$Y_{ij} \sim \text{NB}(\mu_{ij}, k)$$
$$\log(\mu_{ij}) = (\beta_0 + b_i) + \beta_1 x_{ij},$$

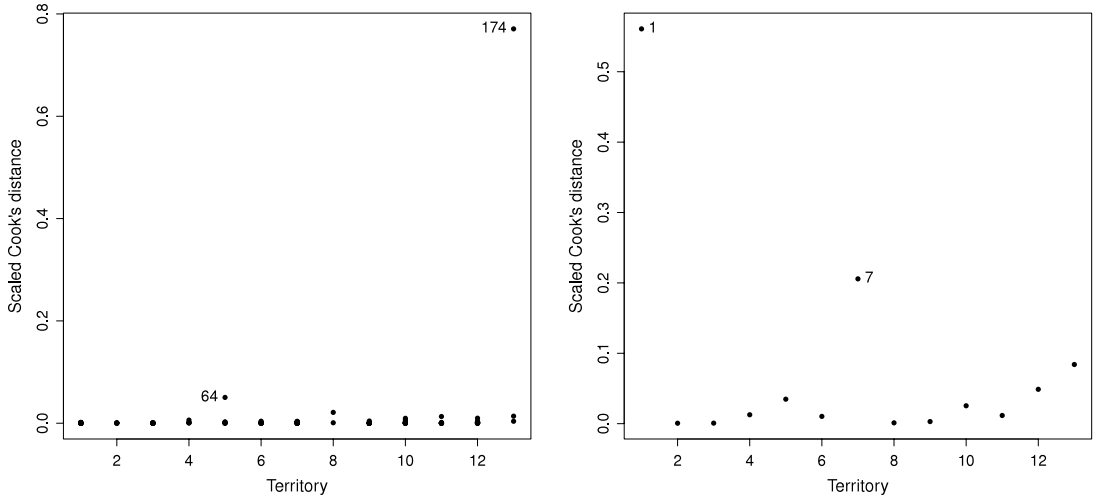
where Y_{ij} and x_{ij} represent, respectively, the claims and logarithm of the number of accidents in the jth area from the ith territory, $b_i \sim \mathcal{N}(\mathbf{0}, \sigma_b^2 \mathbf{I}_{n_i})$, $i = 1, 2, \ldots, 13$, $j = 1, 2, \ldots, n_i$, with $n_1 + n_2 + \cdots n_{13} = 176$. The choice and adequacy of this model is discussed in de Jong and Heller (2008).

The negative binomial distribution has probability mass function given by

$$P(Y = y) = \binom{y + k^{-1}}{k^{-1}} \left(\frac{1}{1 + k\mu}\right)^{k^{-1}} \left(\frac{k\mu}{1 + k\mu}\right)^{y},$$

Table 2
Parameter estimates and 95% bootstrap confidence intervals for the third party claims data.

Situation	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{k}$	$\widehat{\sigma}_b^2$
Complete data	−2.08 (−2.41, −1.76)	1.19 (1.14, 1.24)	0.14 $(8.94 \times 10^{-2}, 0.17)$	0.0305
Excluding Broken Hill	−2.07 (−2.39, −1.76)	1.18 (1.12, 1.25)	0.12 $(8.92 \times 10^{-2}, 0.18)$	0.0127



(a) Proposed Cook's distance. (b) Proposed Cook's distance by subject.

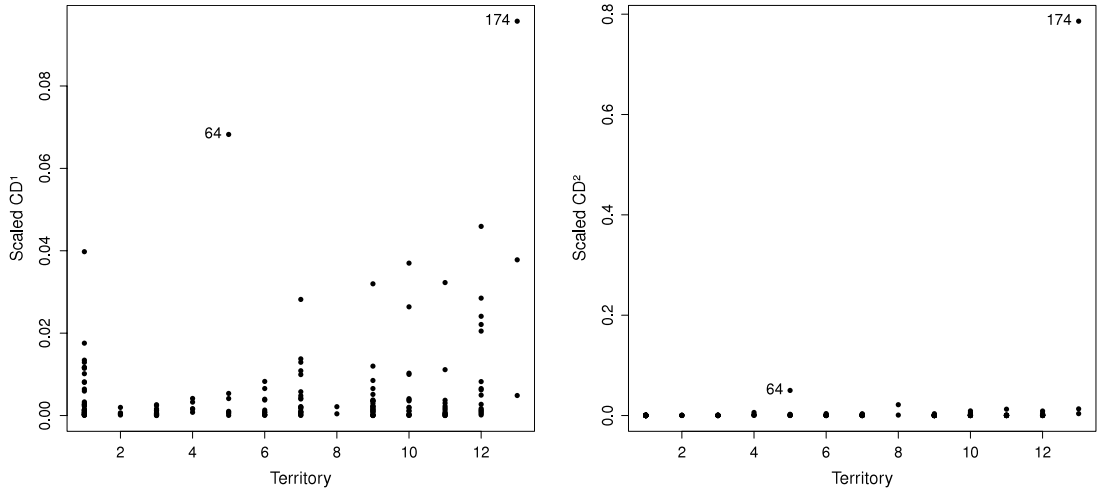
Fig. 4. (a) Potentially influential observations. (b) The first and seventh territories seem to be influential.

with $\binom{a}{b}$ representing the extended binomial coefficient. Then, it follows that $\mathbb{E}[Y] = \mu_{ij}$ and $\ddot{C}_{\eta\eta}$ a diagonal matrix whose entries are given by $\mu_{ij}(1 + ky_{ij})(1 + k\mu_{ij})^{-2}$.

The parameter estimates obtained via the `glmer.nb()` routine from package `lme4` in R are displayed in the first row of Table 2. The numbers within parentheses are the limits of corresponding 95% bootstrap confidence intervals.

In Fig. 4(a) we display the values of (3) corresponding to the fitted model. We investigate the observation labeled 174 further as it was the one considered the most potentially influential observation according to the proposed distance. This observation corresponds to Broken Hill and its influence may be due to the large number of claims (912) relative to the number of accidents (540). The points in Fig. 4(b) suggest that territories 1 and 7 are influential with respect to the prediction of conditional means. Territory 13, which contains observation 174 is the next most potentially influential. In Fig. 5, the first two components of (3) are depicted. Fig. 5(a) shows the scaled version of the distance proposed in Xiang et al. (2002). Observations 64 (Casino) and 174 are highlighted as the two most potentially influential ones regarding the fixed effects. Regarding the influence due to the prediction of random effects, Fig. 5(b) suggests that observation 174 is the most influential by a great margin. From Figs. 4 and 5 it is possible to conclude that, for the sake of prediction, observation 174 is the most influential one and that its influence is mostly due to the prediction of random effects. In Table 2 we also present the estimates of the model parameters when observation 174 is deleted. The estimates for β_0, β_1 and k did not change much, whereas the estimate for σ_b changed much more. Territory 13 has only three observations which are very distinct in the number of claims (912, 75 and 5), thus it is expected for the removal of an observation from this territory to have a great impact on the estimate for σ_b.

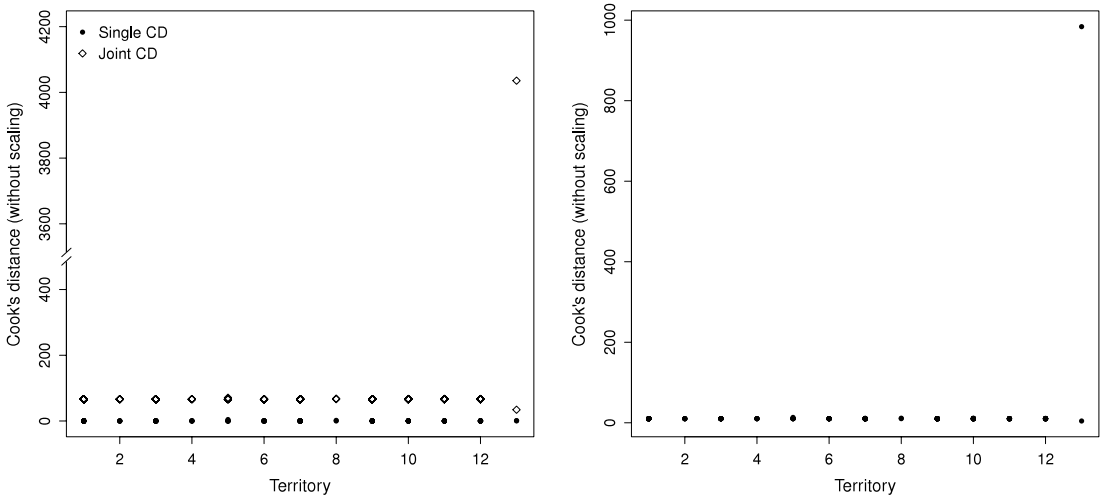
We investigate next if Broken Hill is possibly masking the influence of other areas. In Fig. 6 we display the joint and single conditional influence distances for Broken Hill combined with each one of the remaining observations. In Fig. 6(a), the values for the joint (Broken Hill and the current observation) influence are larger than those for the single influence distance for all areas. Lawrance (1995) called this an enhancing effect. It is worthy noticing that the observations from the thirteenth territory had very different number of claims: 912 for Broken Hill (observation 174), 75 for Central Darling Shire (observation 175) and 5 for an unincorporated area (observation 176). Thus, removing the one with higher number of claims is likely to result in a drastic change in the random effect prediction for that territory. The inclusion of the thirteenth territory as a cluster is arguable. It consists of three very different areas which do not share a sense of unity as seen in the other territories. Fig. 6(b), it is suggested that the observation from Broken Hill had an masking effect on Cook's distance for the observations in the thirteenth territory. This is possibly also due to the same reason in the previous comments regarding the joint influence.



(a) First component of the proposed Cook's distance.

(b) Second component of the proposed Cook's distance.

Fig. 5. First two components of the proposed Cook's distance.



(a) Joint and single influence.

(b) Conditional influence.

Fig. 6. Joint and conditional influence measures based on the new proposal.

7. Concluding remarks

A suitable extension of Cook's distance and a first-order approximation for it was developed inspired by the works of Tan et al. (2001) and Xiang et al. (2002). In a Monte Carlo simulation we evaluated the efficacy of a first-order approximation for detecting observations with significant influence on the linear predictors. The simulation suggests that the proposed method is convenient to detect influential observations not detected by (2). We emphasize that the proposal of Xiang et al. (2002), to our understanding, was intended only to detect influence regarding the estimation of fixed effects, although it seems likely to be used in practice to detect influence with respect to the prediction of random effects. In the real data application, the need for this kind of diagnostic is evident, as it provided a better understanding of the data and may help in a decision process. The idea of joint and conditional influence was used for detecting the masking effect that may occur in influence diagnostics. This is usually overlooked by practitioners. We conclude by reminding the reader that observations flagged as influential by the proposed method (or by similar methods) should not necessarily be deleted from the data. Depending on the chosen threshold, continuously deleting observations could lead to the deletion of the whole sample. If several observations are flagged influential the model may be inadequate for the data. If only a couple of observations are indeed influential, they should be dealt with in a way that depends heavily on the practical situation. In the application section example, a decision maker could treat the thirteenth territory separately. A bootstrap based threshold will be investigated in future works.

The R functions designed to generate the diagnostic plots may be downloaded from https://www.ime.usp.br/~jmsinger/GLMMdiagnostics.zip. The code also supports models whose response variable follows distributions other than the ones

used in this paper. Each supported distribution is actually hard-coded in the file. For using distributions other than those supported, minor modifications are necessary and the instructions are given in the code's comments.

Acknowledgments

The authors thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), Brazil (grant # 308613/2011-2) for partial financial support. We are thankful for the suggestions of the two anonymous referees which greatly improved the manuscript, especially the application section.

Appendix. Decomposition in Section 4

The denominator in (3) was chosen so that our proposal reduces to the one in Tan et al. (2001) when applied to linear mixed models. Starting from the numerator in (3), we proceed as follows.

$$\begin{aligned}
\left(\widehat{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}}_{(ij)}\right)^{\top}\mathrm{Var}\left(\boldsymbol{Y}|\boldsymbol{b}\right)\left(\widehat{\boldsymbol{\eta}}-\widehat{\boldsymbol{\eta}}_{(ij)}\right) &= \left[\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}+\boldsymbol{Z}\widehat{\boldsymbol{b}}\right)-\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{(ij)}+\boldsymbol{Z}\widehat{\boldsymbol{b}}_{(ij)}\right)\right]^{\top}[A(\phi)]^{-1}\ddot{C}_{\eta\eta}\\
&\quad\times\left[\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}+\boldsymbol{Z}\widehat{\boldsymbol{b}}\right)-\left(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{(ij)}+\boldsymbol{Z}\widehat{\boldsymbol{b}}_{(ij)}\right)\right]\\
&= \left(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(ij)}\right)^{\top}[A(\phi)]^{-1}\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(ij)}\right)\\
&\quad+\left(\widehat{\boldsymbol{b}}-\widehat{\boldsymbol{b}}_{(ij)}\right)^{\top}[A(\phi)]^{-1}\boldsymbol{Z}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z}\left(\widehat{\boldsymbol{b}}-\widehat{\boldsymbol{b}}_{(ij)}\right)\\
&\quad+2\left(\widehat{\boldsymbol{\beta}}-\widehat{\boldsymbol{\beta}}_{(ij)}\right)^{\top}[A(\phi)]^{-1}\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z}\left(\widehat{\boldsymbol{b}}-\widehat{\boldsymbol{b}}_{(ij)}\right).
\end{aligned}$$

Now, consider l_1 in Section 2. Use the chain rule to obtain

$$\begin{aligned}
\frac{\partial^2 l_1}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\top}} &= -[A(\phi)]^{-1}\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{X}\\
\frac{\partial^2 l_1}{\partial\boldsymbol{\beta}\partial\boldsymbol{b}^{\top}} &= -[A(\phi)]^{-1}\boldsymbol{X}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z}\\
\frac{\partial^2 l_1}{\partial\boldsymbol{b}\partial\boldsymbol{b}^{\top}} &= -[A(\phi)]^{-1}\boldsymbol{Z}^{\top}\ddot{C}_{\eta\eta}\boldsymbol{Z}.
\end{aligned}$$

The result follows from inserting these expressions, evaluated at $\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{b}=\widehat{\boldsymbol{b}}$, in the previous sum.

References

Abad, A.A., Litière, S., Molenberghs, G., 2010. Testing for misspecification in generalized linear mixed models. Biostatistics 11 (4), 771–786.
Atkinson, A.C., 1986. Masking unmasked. Biometrika 73 (3), 533–541.
Banerjee, M., Frees, E.W., 1997. Influence diagnostics for linear longitudinal models. J. Amer. Statist. Assoc. 92 (439), 999–1005.
Beckman, R.J., Nachtsheim, C.J., Cook, R.D., 1987. Diagnostics for mixed-model analysis of variance. Technometrics 29 (4), 413–426.
Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley-Interscience.
Breslow, N.E., 1984. Extra-Poisson variation in log-linear models. J. R. Stat. Soc. Ser. C 38 (1), 38–44.
Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Amer. Statist. Assoc. 88 (421), 9–25.
Casella, G., Berger, R.L., 2001. Statistical Inference, second ed. Duxbury Press.
Chattejee, S., Hadi, A.S., 1986. Influential observations, high leverage points, and outliers in linear regression. Statist. Sci. 1 (3), 379–393.
Christensen, R., Pearson, L.M., Johnson, W., 1992. Case-deletion diagnostics for mixed models. Technometrics 34 (1), 38–45.
Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics 19 (1), 15–18.
Cook, R.D., Weisberg, S., 1982. Residuals and Influence in Regression. Chapman & Hall.
de Jong, P.d., Heller, G.Z., 2008. Generalized Linear Models for Insurance Data. Cambridge University Press.
Demidenko, E., 2004. Mixed Models: Theory and Applications, first ed. Wiley-Interscience.
Demidenko, E., Stukel, T.A., 2005. Influence analysis for linear mixed-effects models. Stat. Med. 24 (6), 893–909.
Gumedze, F.N., Welham, S.J., Gogel, B.J., Thompson, R., 2010. A variance shift model for detection of outliers in the linear mixed model. Stat. Med. 54 (9), 2128–2144.
Hilden-Minton, J.A., 1995. Multilevel diagnostics for mixed and hierarchical linear models (Ph.D. in Mathematics). University of California, Los Angeles.
Hoaglin, D.C., Welsch, R.E., 1978. The hat matrix in regression and anova. Amer. Statist. 32 (1), 17–22.
Lawrance, A.J., 1995. Deletion influence and masking in regression. J. R. Stat. Soc. Ser. B 57 (1), 181–189.
Lee, Y., Nelder, J.A., 1996. Hierarchical generalized linear models. J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (4), 619–678.
Lesaffre, E., Verbeke, G., 1998. Local influence in linear mixed models. Biometrics 54 (2), 570–582.
McCulloch, C., Searle, S.R., 2001. Generalized, Linear, and Mixed Models, first ed. Wiley-Interscience.
McGilchrist, C.A., 1994. Estimation in generalized linear mixed models. J. R. Stat. Soc. Ser. B Stat. Methodol. 56 (1), 61–69.
Nobre, J.S., Singer, J.d.M., 2007. Residual analysis for linear mixed models. Biom. J. 49 (6), 863–875.
Nobre, J., Singer, J., 2011. Leverage analysis for linear mixed models. J. Appl. Stat. 38 (5), 1063–1072.
Pregibon, D., 1981. Logistic regression diagnostics. Ann. Statist. 9 (4), 705–724.
Stasinopoulos, D.M., Rigby, R., 2007. Generalized additive models for location scale and shape (GAMLSS). J. Stat. Softw. 23 (7), 1–46.
Tan, F.E.S., Ouwens, M.J.N., Berger, M.P.F., 2001. Detection of influential observations in longitudinal mixed effects regression models. J. R. Stat. Soc. Ser. D Stat. 50 (3), 271–284.

Tchetgen, E.J., Coull, B.A., 2006. A diagnostic test for the mixing distribution in a generalised linear mixed model. Biometrika 93 (4), 1003–1010.
Xiang, L., Tse, S., Lee, A.H., 2002. Influence diagnostics for generalized linear mixed models: applications to clustered data. Comput. Statist. Data Anal. 40 (4), 759–774.
Zeger, S.L., Karim, M.R., 1991. Generalized linear models with random effects; a Gibbs sampling approach. J. Amer. Statist. Assoc. 86 (413), 79–86.
Zeger, S.L., Liang, K., Albert, P.S., 1988. Models for longitudinal data: a generalized estimating equation approach. Biometrics 44 (4), 1049–1060.
Zhao, Y., 2006. General design Bayesian generalized linear mixed models. Statist. Sci. 21 (1), 35–51.
Zhu, H., Lee, S., 2003. Local influence for generalized linear mixed models. Canad. J. Statist. 31 (3), 293–309.