Afrika Statistika

Afrika Statistika
Vol. 15 (1), 2020, pages 2247 - 2261.
DOI: http://dx.doi.org/10.16929/as/2020.2247.156

ISSN 2316-090X

Effect of misspecification of random effects distribution on the performance of parameters estimation methods in binary logistic mixed models

Bruno Enagnon Lokonon [1,*], **Marcel Senou**[2] **and Romain Glèlè Kakaï**[1]

[1]Laboratoire de Biomathématiques et d'Estimations Forestières, Université d'Abomey-Calavi, 04 B.P. 1525 Cotonou (Bénin)
[2]Faculté des Sciences Agronomiques (FSA), Université d'Abomey-Calavi, 01BP: 526 Cotonou (Bénin)

Received December 08, 2019; Accepted April 23, 2020

Abstract. We empirically compared a Bayesian estimation method (Integrated Nested Laplace Approximation, INLA) to three classical estimation methods (Penalized Quasi-Likelihood, PQL; Hierarchical Likelihood Method, HLM and Adaptive Gauss-Hermite Quadrature, AGHQ) under six random effect distributions in binary logistic mixed models. Results revealed that AGHQ and HLM had best performance for all distributions considered in the case of fixed effects. For the random effects, classical methods showed best performance for the symmetric distributions (normal, uniform and mixture-normal). AGHQ, HLM and INLA outperform PQL for normal and uniform distributions whatever the sample considered.

Key words: binary multilevel modeling; random effects; non-normality; estimation methods; Bayesian approach.
AMS 2010 Mathematics Subject Classification Objects : 81T80; 65C05; 62F15; 65Y20.

*Corresponding author Bruno Enagnon Lokonon: brunolokonon@gmail.com
Marcel Senou :m_senou2002@yahoo.fr
Romain Glèlè Kakaï : glele.romain@gmail.com

Résumé. Nous avons comparé, à l'aide de la simulation, une méthode d'estimation Bayésienne (INLA) à trois méthodes classiques d'estimation (PQL, HLM et AGHQ) sous six distributions des effets aléatoires dans le cadre des modèles logistiques binaires mixtes. Les résultats ont montré la supériorité de AGHQ et HLM sur les autres pour toutes les distributions considérées dans le cas des effets fixes. Pour les effets aléatoires, les méthodes classiques ont montré les meilleures performances pour les distributions symétriques considérés. AGHQ, HLM et INLA ont donné de meilleurs résultats par rapport à PQL pour les distributions normale et uniforme pour tous les échantillons considérés.

The authors.

(1) Bruno Enagnon Lokonon, Ph.D., is research assistant at Université d'Abomey-Calavi, Bénin.

(2) Marcel Senou, Ph.D., is full professor at Université d'Abomey-Calavi, Bénin.

(3) Romain Glèlè Kakaï, Ph.D., is full professor at Université d'Abomey-Calavi, Cotonou (Bénin).

1. Introduction

Over the years, linear models have been widely used to provide mathematical basis to explain and describe phenomena, extract important information, make future predictions as well as draw inferences (Faraway, 2006). Logistic mixed models are more recently developed to overcome the limits of linear models (Gbur *et al.*, 2012) in the situation where it comes to link categorical outcomes to fixed and random variables. One of the assumptions of the logistic mixed models concerns random effects distribution. The parameter estimates are commonly calculated by maximizing the marginal likelihood obtained by integrating out the random effects. For computational convenience, random effects are assumed to be normally distributed. However, since they are not observed, the validity of this assumption is difficult to verify (McCulloch and Neuhaus, 2011). A natural concern is related to the impact of misspecification of the random effects distribution on the estimators.

For linear mixed models, Verbeke and Lesaffre (1997) showed that the maximum likelihood estimators of the fixed effects and variance components, obtained under the assumption of normal random effects, are consistent and asymptotically normal, even with non-normality of random-effects. However, recent research works suggest that this does not hold for logistic mixed models. Neuhaus *et al.* (1992) conducted a logistic mixed effect model in which gamma, t-Student and normal distributions were considered for the random effects. They estimated the model using a quasi-Newton algorithm and found that the estimated parameters were asymptotically biased. Heagerty and Kurland (2001), Agresti *et al.* (2004) and Litière *et al.* (2008) studied the impact of misspecification on parameters estimated through Gauss–Hermite Quadrature (GHQ) approximation. They found that incorrect assumptions regarding the random effects could lead to substantial bias in the estimates. McCulloch and Neuhaus (2011) and Hernandez *et al.* (2014) used respectively Proc NLMIXED in SAS and Laplace Approximation with lme4 package in their study of random effect misspecification and reached the same conclusions.

As shown, previous studies investigating the impact of misspecification of random effects distribution predominantly focused on one class of parameters estimation methods, the likelihood-based principle and the other classes of estimation methods are rarely used. Limited studies have considered the extended likelihood and the Bayesian approaches to estimate the parameters when the random effects distribution is misspecified. The present study aims to contribute to fill this gap by assessing the impact of random effects distribution misspecification on the likelihood based method, the extended likelihood principle and the Bayesian methods. Adaptive Gauss-Hermite Quadrature, Hierarchical Likelihood Method and Integrated Nested Laplace Approximation were used respectively as likelihood based method, extended likelihood approach and Bayesian estimation method. These methods have been chosen since they are the most improved and the most accurate in each class (Kim *et al.*, 2013; Casals *et al.*, 2015; Lokonon *et al.*, 2019). However, we also include Penalized Quasi-Likelihood (PQL) because it is the simplest and most widely used approximation method (Bolker *et al.*, 2009, Lokonon *et al.*, 2019).

2. Model specification

Let y_{ij} be the i^{th} observation for the j^{th} cluster and Y be a vector of the observations y_{ij} following a Binomial distribution in the random intercept model written as:

$$\text{logit}(p\,(y_{ij} = 1|u_j)) = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2ij} + u_j \quad (1)$$

where, $i = 1, \ldots, n$ (n represents number of observations within cluster j) and $j = 1, \ldots, N$, (N represents is number of clusters). The random intercept u_j $(u_1, \ldots, u_q)$ has zero mean and variance σ^2. The fixed effects were set from previous studies (Hernandez *et al.*, 2014; Hernandez and Giampaoli, 2018): β_0=1, β_1=2 and β_2=1. The between-cluster covariate (x_1) and the within-cluster covariate (x_2) were generated from Standard Normal distribution $N(0, 1)$. The number of clusters and the number of observations per cluster were set respectively as N= 5, 10, 25, 50 and n= 30, 50, 100 in order to obtain sample sizes from 150 to 5000. The variances of the random intercept u_j were set at σ^2=0.5, 1, 2. Variances greater than 2 were not considered because they caused larger values of the random intercept (Hernandez and Giampaoli, 2018).

For the random intercept u_j, six different distributions were considered: normal, uniform, exponential, log-gamma, log-normal and symmetric mixture of two normal densities that were defined as in Hernandez and Giampaoli (2018). The distributions were transformed such that the zero-mean condition was satisfied, and the corresponding variances were equal to the prespecified values for σ^2. The density function for each distribution is plotted in Figure 1 for the case of unit variance. With this choice, we cover a range of densities varying from very symmetric (normal, mixture of two normals and uniform) to very skewed distributions (exponential, log-gamma and log-normal).

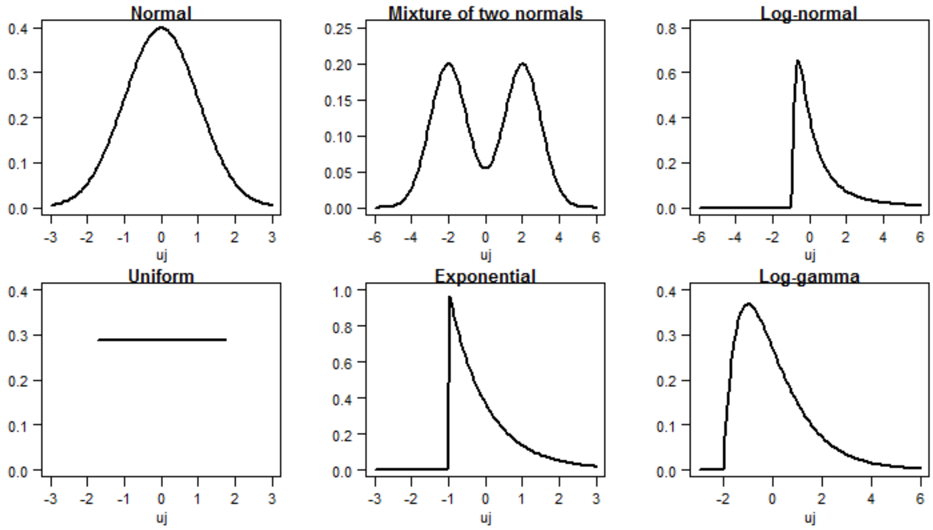

Fig. 1. Random effects distribution considered with zero mean and unit variance.

3. Estimation methods considered

In binary logistic mixed models, parameters are estimated using maximum likelihood estimation. By the local independence assumption, the conditional density of Y given u_j has the form:

$$g\left(y_{ij}|u_j;\beta\right)=\prod_{i=1}^{n}p\left(y_{ij}=1|u_j\right)^{y_{ij}}p\left(y_{ij}=0|u_j\right)^{1-y_{ij}} \tag{2}$$

The multivariate density function of u_j is $f\left(u_j;\Sigma\right)$ and has the following likelihood (Casals *et al.*, 2015):

$$f\left(u_j;\Sigma\right)=\prod_{j=1}^{N}f\left(u_j;\Sigma\right) \tag{3}$$

The model parameters are estimated by maximising the marginal likelihood obtained by integrating the joint distribution of (Y, u_j) over the random effects. The result is the marginal likelihood function given by (Kim *et al.*, 2013):

$$\begin{aligned} l\left(\beta,\Sigma\right) &= \int g\left(y_{ij}|u_j;\beta\right)f\left(u_j;\Sigma\right)du_j \qquad (4)\\ &= \prod_{j=1}^{N}\int\prod_{i=1}^{n}p\left(y_{ij}=1|u_j\right)^{y_{ij}}p\left(y_{ij}=0|u_j\right)^{1-y_{ij}}f\left(u_j;\Sigma\right)du_j \end{aligned}$$

The distributions of both random effects and response variable differ. Thus, the Equation (4) is analytically intractable (Casals *et al.*, 2015). As a result, various approximation methods have been developed with different degrees of accuracy (Capanu *et al.*, 2013). Four of them are considered in this study. These methods were selected due to their robustness, their recent improvement in R software and their accessibility for applied researchers (Casals *et al.*, 2015).

3.1. Adaptative Gauss-Hermite quadrature (AGHQ)

AGHQ is an approximation method that partitions the marginal likelihood (Equation 4) into multiple components (McNeish, 2016). The number of partitions and the accuracy of the approximation are determined by $Q+1$ where Q is the number of quadrature. Equation (4) can be rewritten as:

$$l\left(\beta,\Sigma\right)=\int g\left(y_j|v_j;\beta,\Gamma\right)\Phi(v_j)dv_j \tag{5}$$

where, $u_j=\Gamma v_j,\Gamma\Gamma^{'}=\Sigma$, and v_j has the standard normal density $\Phi(v_j)$. Let n_q denote a vector of quadrature points with the same dimension as u_j and $w(n_q)$ the corresponding weight. The marginal likelihood can then be approximated as follows:

$$l\left(\beta,\Sigma\right)\approx\sum_{q=1}^{Q} g\left(y_j|n_q;\beta,\Gamma\right)w(n_q) \tag{6}$$

3.2. Penalized Quasi-Likelihood (PQL)

Whereas AGHQ approximates the integral of the likelihood function, PQL linearizes its nonlinear components (McNeish, 2016). The general form of logistic mixed effects model is obtained through a linear link $k(.)$ that relates the linear predictor $\eta_{ij}=x_{ij}\beta+z_{ij}u_j$ to the mean of the response variable such that:

$$E\left(y_{ij}|x_{ij},u_j\right)=\mu_j=k^{-1}\left(x_{ij}\beta+z_{ij}u_j\right)=k^{-1}(\eta_{ij}) \tag{7}$$

where k is logit link function.

A first order Taylor series expansion of Equation (7) about $\tilde{\beta}$ and $\tilde{u}_j$, the current, fixed values of β and u_j, has the following form (Codd, 2014):

$$\begin{aligned}k^{-1}(\eta_{ij})\approx\, & k^{-1}\left(x_{ij}\tilde{\beta}+z_{ij}\tilde{u}_j\right)\\ & +\tilde{\Delta}_{ij}\left[x_{ij}\left(\beta-\tilde{\beta}\right)z_{ij}\left(u_j-\tilde{u}_j\right)\right]\end{aligned} \tag{8}$$

where $\tilde{\Delta}_{ij}$, is a diagonal matrix of derivatives of $E\left(y_{ij}|x_{ij},u_j\right)$ assessed at the expansion points $\tilde{\beta}$ and $\tilde{u}_j$. In other words, $\tilde{\Delta}_{ij}=\dfrac{\partial\left(g^{-1}(\eta_{ij})\right)}{\partial\eta_{ij}}|_{\tilde{\beta},\tilde{u}_j}$

The model for the data, y_{ij}, can then be defined as:

$$\begin{aligned} y_{ij} = k^{-1}\left(x_{ij}\tilde{\beta} + z_{ij}\tilde{u}_j\right) \\ + \tilde{\Delta}_{ij}\left[x_{ij}\left(\beta - \tilde{\beta}\right) + z_{ij}\left(u_j - \tilde{u}_j\right)\right] + \varepsilon_i \end{aligned} \quad (9)$$

From Equation (9) a pseudo response vector, $\tilde{y}_{ij}$, can be formed as:

$$\tilde{y}_{ij} = \tilde{\Delta}_{ij}^{-1}\left(y_{ij} - k^{-1}\left(x_{ij}\tilde{\beta} + z_{ij}\tilde{u}_j\right)\right) + x_{ij}\tilde{\beta} + z_{ij}\tilde{u}_j \quad (10)$$

Next, define a weight matrix, $\tilde{W}_{ij}$, as $\tilde{W}_{ij} = \tilde{V}_{ij}^{-1}\tilde{\Delta}_{ij}^2$ where $\tilde{V}_{ij}$ is a matrix of the diagonal elements of var $\left(\varepsilon_{ij}|\tilde{\beta}, \tilde{u}_j\right)$. A linear mixed effects model of the form:

$$\tilde{y}_{ij} = x_{ij}\beta + z_{ij}u_j + \tilde{W}_{ij}^{-1}\varepsilon_{ij}^* \quad (11)$$

can then be fitted to the pseudo data, assuming that $\varepsilon_{ij}^* \sim N\left(0, \Lambda_{ij}^*\right)$. Because the Equation (11) is linear in the random-effects, the likelihood function has a closed-form solution. The model can then be adjusted by pseudo likelihood.

Once the parameter estimates from the model in the Equation (11) are obtained, $\tilde{\beta}$ and $\tilde{u}_j$ are updated and the next iteration is initiated. This method is called penalized quasi-likelihood (PQL) due to the use of Taylor series expansions. It is a doubly iterative procedure, where the Taylor series expansion about the current estimates of $\tilde{\beta}$ and $\tilde{u}_j$ is the first step, and the fitting of a linear mixed model to the pseudo data is the second step. Iteration between these steps continues until the difference between the parameter estimates in successive iterations is sufficiently small.

3.3. Hierarchical likelihood method (HLM)

HLM is also called hierarchical generalized linear model (Collins, 2008). It is an extension of GLMs using hierarchical (h-) likelihood (Lee and Nelder, 2001; Nelder *et al.*, 2006). The h-likelihood is the log joint likelihood of the extended likelihood L_E written as follows (Casals *et al.*, 2015):

$$h = \log\left(L_E(y; \beta, v)\right) = \log\left(f(y; \beta|v)\right) + \log\left(f(v)\right) \quad (12)$$

with $\log\left(f(y; \beta|v)\right)$ being the log of the density function of the response variable and β the parameters. u is a vector of random effects and $v(.)$ is an appropriate link function defining the h-likelihood such that $v = v(u)$. Using the score functions of the h-likelihood, the parameters β and v are estimated as follows:

$$\frac{\partial h}{\partial \beta} = 0, \quad \frac{\partial h}{\partial v} = 0 \quad (13)$$

The adjusted profile h-likelihood is maximized to obtain the variance components as follows:

$$P_{\beta,u} = \left(h + \frac{1}{2}\log\left(2\pi H^{-1}\right)\right)|_{\beta=\hat{\beta},\, u=\hat{u}} \tag{14}$$

where H is a Hessian matrix of the h-likelihood and $\frac{\partial P_{\beta,u}}{\partial \lambda} = 0$, with λ a vector including both variance components (random effects and residuals), $\lambda = (\sigma_u^2, \sigma_e^2)$.

3.4. Integrated Nested Laplace Approximation (INLA)

INLA is a Bayesian procedure and in Equation (4), a prior distribution must be specified for β and u (the random effects). A non-informative normal distribution is defined as prior distribution for β in this study following Casals *et al.* (2015). Let $\gamma = (u, \beta)^T$ be the $G \times 1$ vector of Gaussian parameters. The random component u is supposed to follow a multivariate normal distribution, $u|\Gamma \sim N(0, \Gamma^{-1})$, where the precision matrix $\Gamma = \Gamma(\phi)$ depends on parameters ϕ. Let ϕ also be the vector of the random components with $P(\phi)$ as the prior. The posterior density is given by:

$$\pi\left(\gamma, \phi|y\right) \propto \pi(\gamma|\phi)\pi(\phi)\prod_{i=1}^{m} f\left(y_i|\gamma, \phi\right) \tag{15}$$

where, i=1,...,m (number of observations).

The posterior density is computed such that $\pi\left(\gamma, |y\right) = \int \pi\left(\gamma|\phi, y\right)\pi(\phi|y)d\phi$, where Laplace approximation is applied to carry out the integrations required for the evaluation of $\pi\left(\gamma|\phi, y\right)$. INLA provides a good approximation while reducing computational costs substantially (Rue *et al.*, 2009; Rue *et al.*, 2017).

4. Simulation plan

The simulation was performed in R software in following steps:

Step 1 : Set values for the parameters β_0, β_1, β_2 and σ^2;
Step 2 : Generate the covariates x_1 and x_2 from the Standard Normal distributions and the true distribution of random effects respectively from normal, mixture of normals, uniform, exponential, log-gamma and log-normal distributions;
Step 3 : Set the coefficients and obtain logit such that: $\text{logit}(p\left(y_{ij} = 1|u_j\right)) = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2ij} + u_j$;
Step 4 : Calculate the predicted probabilities of experiencing an event such that: $p_{ij} = \texttt{invlogit}(p\left(y_{ij} = 1|u_j\right))$;
Step 5 : Obtain the binary outcome y_{ij} such that: $y_{ij} = \texttt{rbinom}(n_{ij}, 1, p_{ij})$, where sample size n_{ij} is the combination of the clusters N= 5, 10, 25, 50 and n= 30, 50, 100;
Step 6 : For each combination of n_{ij}, σ^2 and true random effects distribution, run the model (Equation (1)) using the following estimation methods *a)* PQL; *b)* AGHQ; *c)* HLM; *d)* INLA;

Step 7 : Repeat the step 6 S times (S=500).

R packages MASS, lme4, hglm and INLA were respectively used for PQL, AGHQ, HLM and INLA.

5. Comparison criteria

The simulation study included 864 settings, given by 6 distributions, 3 variances for the random intercept, 4 values of N, 3 values of n and 4 estimation methods. The parameters vector in the simulation is given by $\theta = (\beta_0, \beta_1, \beta_2, \sigma^2)^T$. For each simulation setting and estimation method, the empirical bias was calculated for the fixed effects (between-cluster and within-cluster effects) and the random effects as the mean bias over the 500 data sets as follows:

$$\mathrm{B} = \frac{1}{S}\sum_{j=1}^{S}\left(|\beta - \hat{\beta}_j|\right) \tag{16}$$

where $\hat{\beta}_j$ is the estimated parameter, β is the true value and $j = 1, \ldots, S$, S is the number of simulations (S=500). The relative distance (RD) was also used to quantify the impact of the misspecification on the estimates. The RD is defined as:

$$\mathrm{RD} = \frac{||\hat{\theta} - \theta||}{||\theta||} \tag{17}$$

where $\hat{\theta}$ is the estimated parameter vector and θ the true parameter vector ($||\theta|| = \sqrt{\beta_0^2 + \beta_1^2 + \beta_2^2 + \sigma^4}$).

The smaller the values of B and RD, the lower is the impact and better is the estimation method used. Moreover, for each setting, we recorded the computational times with R function $system.time$ and the convergence rate.

6. Results

6.1. Mean bias for AGHQ, INLA, HLM and PQL under varying distribution of random effects

Figure 2 presents boxplots of the mean bias of the four estimation methods for both fixed and random effects according to the random effects distribution. AGHQ and HLM showed the lowest median values of the mean bias for all distributions considered in the case of fixed effects (between and within cluster effects). Moreover, for these estimation methods, the maximum values of the mean bias are smaller compared to the other estimation methods when considering the fixed effects. For random effects, there was no estimation method showing lower median values of the mean bias in all situations. However, except the exponential distribution, the classical methods (PQL, HLM and AGHQ) showed smaller median values of the mean bias compared to the Bayesian estimation method. In addition, the symmetric distributions (normal, uniform and mixture-normal) showed lower mean bias compared to the asymmetric distributions (log-normal, log-gamma, exponential) for all estimation methods considered. Moreover, the dispersion around the median values is less pronounced for classical estimation methods and symmetric distributions in the case of random effects.

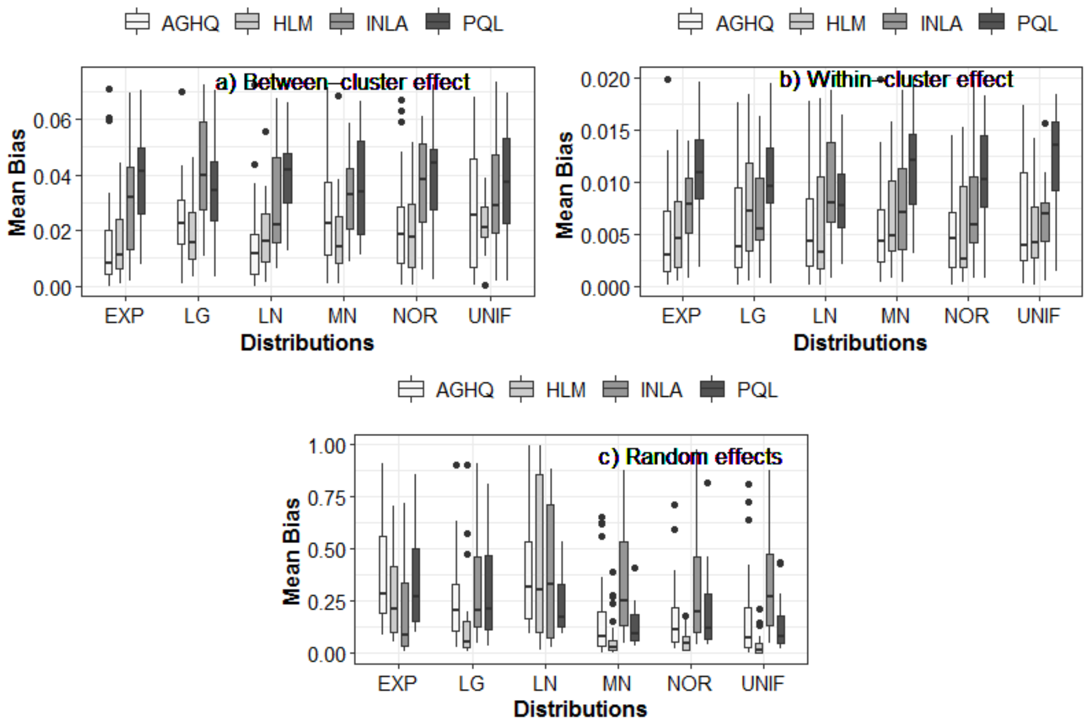

Fig. 2. Boxplots of mean bias for AGHQ, INLA, HLM and PQL under varying distribution of the random effect.
NOR=Normal distribution; UNIF=uniform distribution; LN=Log-normal distribution; LG=Log-gamma distribution; EXP= Exponential distribution; MN=Mixture of two normal distributions

6.2. Relative distance between estimated and true parameter vectors

6.2.1. Case of σ^2=0.5 and n=30

In Figure 3, the median of the relative distance between the estimated parameter vector and the true parameter vector for the estimation methods and the random effects distribution is presented. This figure showed that the relative distance decreases as N increases and this is noted for all estimation methods and random effects distributions. Furthermore, the relative distance is smaller (less than 0.4) when the random effects distribution is symmetric (normal and uniform). Overall, AGHQ and HLM outperformed the other methods in all situations considered. For normal and uniform distributions, AGHQ, HLM and INLA showed the lower relative distances and outperform PQL in all situations. For exponential, log-normal, log-gamma and mixture-normal distributions, the classical estimation methods performed better than the Bayesian method when the number of the clusters is less than 10. On the other hand, for the number of the clusters greater or equal to 10, AGHQ, HLM and INLA performed better than PQL.

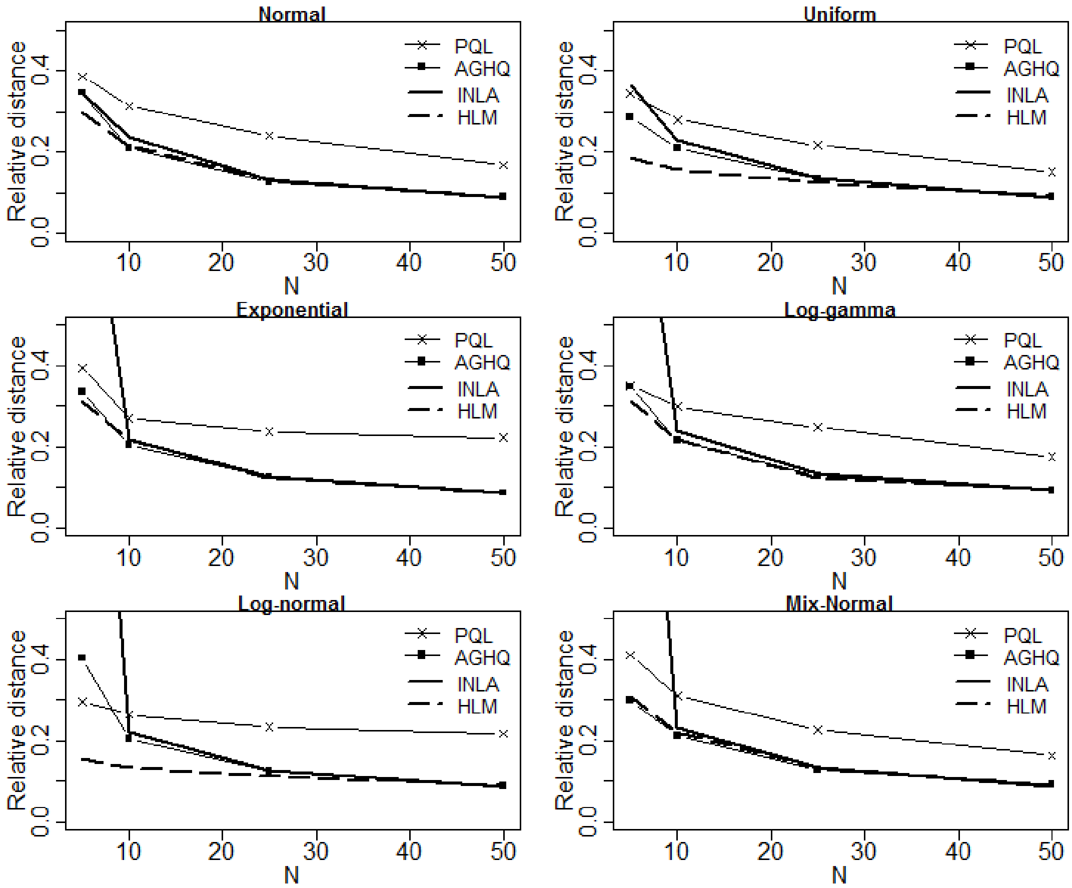

Fig. 3. Median of relative distance between and for σ^2=0.5 and n=30

6.2.2. Case of σ^2=2 and n=30

Figure 4 shows the median of the relative distance between the estimated parameter vector and the true parameter vector for the estimation methods according to the random effects distribution when σ^2=2 and n=30. This figure showed that for all estimation methods and random effects distributions, the relative distance decreases as N increases. A relatively similar pattern to that found when σ^2=0.5 is observed, however, the relative distances are greater than in the case of σ^2=0.5. Uniform and exponential distributions showed similar results where INLA, HLM and AGHQ outperformed PQL. The other distributions showed similar results where PQL, INLA and AGHQ outperformed INLA for N less than 10 while INLA, HLM and

AGHQ outperformed PQL for N greater or equal to 10. Moreover, similar pattern to that found in Figures 3 and 4 is observed for the other combinations between the variances and n. For this reason, these combinations were not presented.

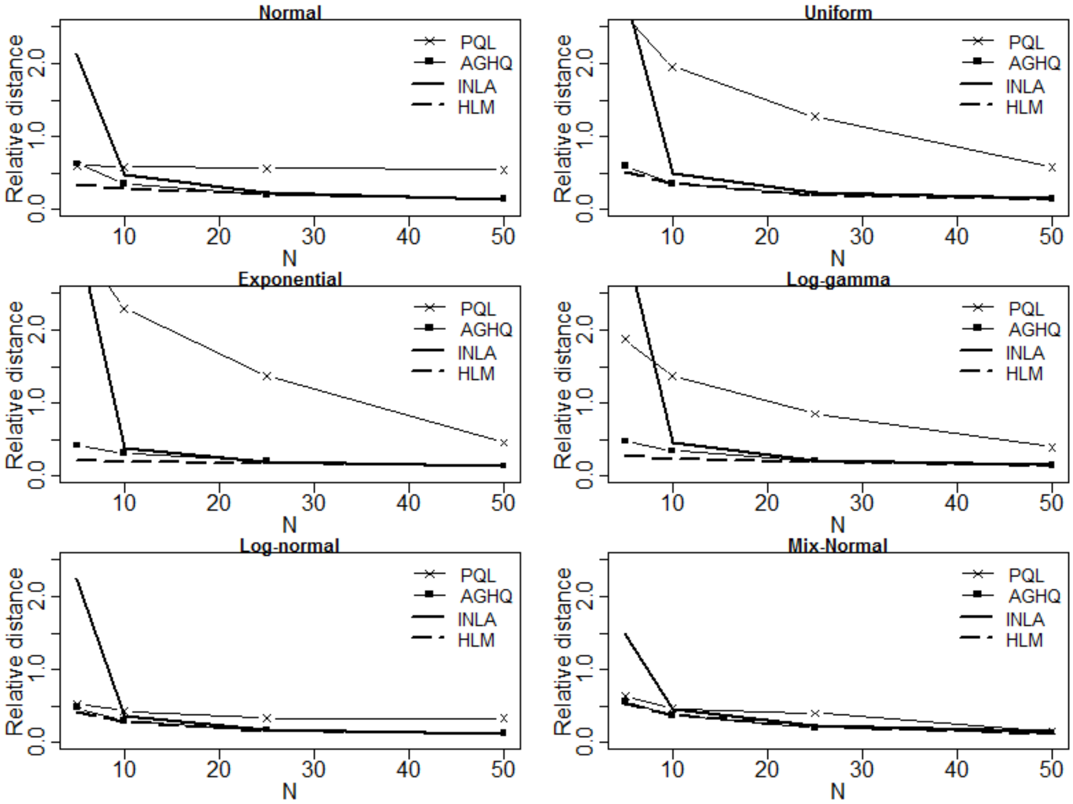

Fig. 4. Median of relative distance between and for σ^2=2 and n=30

6.3. Convergence rate and computation time

Table 1 presents the convergence rate and the computation time (in brackets) of the four methods according to the random intercept distribution. Overall, the classical methods showed relatively low convergence rates for small values of n and N while the Bayesian method presented highest convergence rates (100 %) except one setting where the rate of convergence was 33 %. Regarding the computational time, PQL method requires less time. AGHQ requires relatively lower computational times, whereas the INLA and HLM are the slowest methods.

Table 1. Convergence rate in percentage (%) and computational time in second (in brackets) of the estimation methods (PQL, AGHQ, INLA and HLM)

N; n	PQL						AGHQ					
	NO	UN	LN	LG	EX	MN	NO	UN	LN	LG	EX	MN
5; 30	33.33(0.09)	33.33(0.09)	33.33(0.14)	33.33(0.14)	33.33(0.08)	66.67(0.31)	66.67(0.36)	66.67(0.41)	100(0.36)	66.67(0.35)	66.67(0.33)	66.67(0.41)
5; 50	100(0.10)	33.33(0.10)	66.67(0.15)	33.33(0.17)	33.33(0.08)	66.67(0.18)	66.67(0.40)	100(0.62)	100(0.48)	100(0.50)	100(0.46)	100(1.02)
5; 100	66.67(0.14)	66.67(0.11)	66.67(0.18)	66.67(0.18)	66.67(0.12)	66.67(0.24)	100(0.78)	100(0.92)	66.67(0.71)	100(0.82)	100(0.78)	100(0.79)
10; 30	66.67(0.13)	66.67(0.13)	66.67(0.17)	66.67(0.17)	66.67(0.11)	66.67(0.23)	100(0.60)	100(0.74)	100(0.60)	100(0.60)	100(0.57)	100(0.67)
10; 50	66.67(0.13)	66.67(0.15)	66.67(0.19)	66.67(0.19)	66.67(0.12)	66.67(0.26)	100(0.84)	100(1.06)	100(0.85)	100(0.88)	100(0.80)	100(0.88)
10; 100	66.67(0.19)	66.67(0.17)	66.67(0.25)	66.67(0.25)	66.67(0.16)	66.67(0.35)	100(1. 43)	100(1.69)	100(1.40)	100(1.49)	100(1.49)	100(1.45)
25; 30	66.67(0.17)	66.67(0.16)	66.67(0.24)	66.67(0.27)	66.67(0.16)	66.67(0.29)	100(1. 23)	100(1.48)	100(1.25)	100(1.26)	100(1.19)	100(1.32)
25; 50	66.67(0.22)	66.67(0.20)	100(0.31)	66.67(0.34)	100(0.20)	66.67(0.36)	100(1. 86)	100(2.21)	100(1.90)	100(1.95)	100(1.82)	100(1.88)
25; 100	100(0.34)	66.67(0.33)	100(0.51)	66.67(0.57)	66.67(0.34)	66.67(0.57)	100(3.68)	100(4.17)	100(3.47)	100(3.41)	100(3.55)	100(3.54)
50; 30	100(0.26)	100(0.24)	100(0.41)	66.67(0.56)	100(0.24)	100(0.43)	100(2.33)	100(2.72)	100(2.35)	100(2.36)	100(2.25)	100(2.37)
50; 50	66.67(0.39)	100(0.32)	100(0.59)	66.67(0.80)	100(0.59)	100(0.43)	100(3.55)	100(4.09)	100(3.54)	100(3.78)	100(3.56)	100(3.59)
50; 100	66.67(0.66)	100(0.55)	100(1.01)	66.67(1.48)	100(0.63)	100(0.96)	100(7.37)	100(7.18)	100(6.86)	100(6.63)	100(6.64)	100(7.13)
	INLA						HLM					
	NO	UN	LN	LG	EX	MN	NO	UN	LN	LG	EX	MN
5; 30	33.33(1.03)	100(0.96)	100(0.86)	100(0.88)	100(0.94)	100(0.94)	98.6(2.03)	99.3(2.09)	98.7(3.6)	98.5(5.1)	98.8(12.95)	97.7(1.9)
5; 50	100(2.53)	100(1.13)	100(1.08)	100(1.04)	66.67(1.14)	100(3.69)	97.6(1.22)	98.7(4.53)	99.3(4.1)	99.4(22)	98.77(15.24)	99.4(2.6)
5; 100	100(1.80)	100(1.61)	100(1.44)	100(1.48)	100(1.62)	100(1.43)	98.2(1.22)	99.4(9.19)	99.4(0.6)	99.4(2.1)	99.2(6.2)	99.4(1.2)
10; 30	100 (1.40)	100(1.24)	100(1.11)	100(1.14)	100(1.23)	100(1.21)	99.4(1.36)	99.7(6.27)	99.9(0.6)	99.5(13.9)	99.5(10.4)	99.7(1.1)
10; 50	100(2.14)	100(1.60)	100(1.48)	100(1.51)	100(1.63)	100(1.43)	99.9(0.87)	99.7(0.86)	99.9(0.4)	100(1.5)	100(1.9)	100(1.6)
10; 100	100(2.70)	100(2.46)	100(2.33)	100(2.18)	100(2.52)	100(2.80)	99.8(1.55)	99.6(1.4)	100(0.8)	99.8(1.1)	99.7(1.5)	100(1.8)
25; 30	100(2.33)	100(2.08)	100(1.85)	100(1.90)	100(2.12)	100(1.90)	100(2.02)	100(1.22)	100 (0.9)	100(1.66)	100(1.9)	100(4)
25; 50	100(4.10)	100(3.17)	100(2.77)	100(2.87)	100(3.18)	100(2.77)	100(3.32)	100(2.1)	100(2)	100(2.7)	100(1.4)	100(6)
25; 100	100(6.38)	100(5.61)	100(4.80)	100(4.78)	100(5.59)	100(4.72)	100(10.16)	100(4.95)	100(7.1)	100(3.2)	99.6(12.2)	100(10)
50; 30	100(4.25)	100(3.86)	100(3.27)	100(3.49)	100(3.80)	100(3.35)	100(8.60)	100(5.41)	100(6.6)	100(4.6)	100(5.8)	100(12.3)
50; 50	100(6.96)	100(5.74)	100(4.82)	100(5.18)	100(5.70)	100(10.29)	100(16.46)	100(11.53)	100(11.1)	100(19.1)	100(16.7)	100(19)
50; 100	100(14.99)	100(13.47)	100(11.33)	100(41.09)	100(13.79)	100(11.53)	100(53.96)	100(15.55)	100(23.4)	100(26.8)	100(38.9)	100(20.4)

n=Number of observation per group; N=Number of groups; NO=Normal distribution; UN=uniform distribution; LN=Log-normal distribution; LG=Log-gamma distribution; EX= Exponential distribution; MN=Mixture of two normal distributions

7. Discussion

This study assessed the effect of misspecification of the random effects distribution on the performance of four estimation methods in frame of binary logistic mixed models. We also investigated the impact of the increased random effects variance and varying sample size at both group and individual level on the performance of these methods. A natural concern in using logistic mixed models is misspecifying the model for random effects. For computational convenience, random effects are almost routinely assumed to be normal (McCulloch and Neuhaus, 2011). Many authors have found that likelihood-based inference can be severely affected if the random effects distribution is misspecified (Neuhaus *et al.*, 1992; Agresti *et al.*, 2004; Litière *et al.*, 2008; Hernandez and Giampaoli, 2018). What happens with Bayesian approach? And what methods are less or more sensitive to the misspecification of the random effects distribution? That is what we addressed in this study. Overall, the misspecification of the random effect distribution impacts the performance of the estimation methods according to the simulation conditions.

The four estimation methods considered in our study approximate in different ways the marginal likelihood in order to estimate the parameters. AGHQ use a numerical integration while PQL is based on linearization technique (Codd, 2014). INLA and HLM respectively use Bayesian framework and h-likelihood procedure (Casals *et al.*, 2015). Vonesh (2012) showed that, for fixed effect parameters estimation in the case of binary data, numerical integration methods tend to be more accurate than the linearization methods. Similarly, Collins (2008) found a concordance between h-likelihood procedure and numerical integration in his study. Our study reveals that the results from Collins (2008) and Vonesh (2012) can be extended to non-normal random effects. We also found that the shape of the random effects distribution (symmetric or asymmetric) has an impact on the bias resulting from the distribution misspecification similarly to results from Hernandez *et al.* (2014) indicating that the bias was less when the random effects distribution is symmetric.

The classical estimation methods especially AGHQ and HLM performed better than the Bayesian method as the number of the clusters is less than 10. This result does not confirm those from previous studies stating that classical methods tend to perform well for large number of subjects (Breslow and Lin, 1995) and are less accurate for small clusters. Our result can be explained by the fact that Bayesian approach is more sensitive to non-normality of random effects distribution than the classical methods in the case of small samples. Furthermore, AGHQ, HLM and INLA performed better than PQL for the number of the clusters greater or equal to 10 whatever the distribution considered and this could be explained by the fact that PQL is generally less accurate for fitting binary data (Jang *et al.*, 2007).

8. Conclusions

The present study reveals that the misspecification of the random effects distribution in binary logistic mixed models differently impacts the performance of the estimation methods considered. The choice of an estimation method in binary logistic mixed models should be done based on the characteristics of each data set. In practice, we recommend to users to firstly check the random effects distribution using a diagnostic test as proposed by Efendi *et al.* (2014) and Drikvandi *et al.* (2017) based on graphical checking which can be performed with the function check_model of the R package *performance*. Then, for smaller sample and number of the clusters (less than 10), classical estimation methods (AGHQ and HLM) can be used to estimate the parameters. In contrary, for greater sample (number of clusters greater or equal to 10), Bayesian method (INLA) as well as classical methods (AGHQ and HLM) could be chosen. However, for the big data requesting long time for running, AGHQ could be the first choice since INLA and HLM are slower methods. Moreover, the current study used binary logistic mixed models to compare the estimation methods; it would be useful that future researches compare these methods for multinomial logistic mixed models since several researches often involve categorical outcome variables with more than two levels.

Acknowledgment

The authors are grateful to the referee, the editorial team and the presenter for their helpful suggestions that lead to an improved paper. This research was financially supported by the African Centre of Excellence in Mathematics and Applications (CEA-SMA) of Institute of Mathematics and Physics (IMSP).

References

Agresti, A., Caffo, B. and Ohman-Strickland, P. 2004. Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Comput. Stat. Data Anal.* 47, 639-653.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., White, J. S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology and Evolution* 24, 127-135

Breslow N. E. and Lin X. 1995. Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika*, 82, 81-91.

Capanu M., Gönen M. and Begg C. B. 2013. An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in Medicine*, 32(26),4550-4566.

Casals M., Langohr K., Carrasco J. L. and Rönnegård L. 2015. Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a simulation study. *Sort* 39 (2), 281-308.

Codd C. L. 2014. A Review and Comparison of Models and Estimation Methods for Multivariate Longitudinal Data of Mixed Scale Type. PhD thesis, Ohio State University, USA, 131p.

Collins D. 2008. The performance of estimation methods for generalized linear mixed models. PhD thesis, School of Mathematics & Applied Statistics-Faculty of Informatics, University of Wollongong, Australia, 223 p.

Drikvandi R., Verbeke G. and Molenberghs G. 2017. Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*, 73, 63–71.

Efendi A., Drikvandi R., Verbeke G., Molenberghs G. 2014. A goodness of fit test for the random effects distribution in mixed models. *Statistical Methods in Medicine Research*, 26, 970-983.

Faraway J. J. 2006. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. Second edition. Chapman and Hall/CRC. p. 332.

Gbur E. E., Stroup W. W., McCarter K. S., Durham S., Young L. J., Christman M., West M. and Kramer M. 2012. Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences, edited by American Society of Agronomy Soil Science Society of America, Crop Science Society of America, 33 p.

Heagerty P.J. and Kurland, B.F. 2001. Misspecified Maximum Likelihood Estimates and Generalised Linear Mixed Models. *Biometrika* 88, 973-985.

Hernandez F., Usuga O. and Giampaoli V. 2014. A misspecification simulation study in Poisson mixed model, In: Kneib T, Sobotka F, Fahrenholz J and Irmer H, *Proceedings of the 29th International Workshop on Statistical Modelling, (207-212)* Gottingen, Germany, 14-18 July 2014.

Hernandez F. and Giampaoli V. 2018. The Impact of Misspecified Random Effect Distribution in a Weibull Regression Mixed Model. *stats* 1, 48-76.

Jang M. J., Lee Y., Lawson A. B. and Browne W. J. 2007. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modeling. *Environmetrics*, 18, 809-821.

Kim, Y., Choi, Y.-K., Emery, S. 2013. Logistic regression with multiple random effects: a simulation study of estimation methods and statistical packages. *The American Statistician*, 67, 171-182

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

Lee Y., Nelder J. A. and Pawitan Y. 2006. Generalized linear models with random effects: Unified analysis via H-likelihood.: CRC Press.

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

Lokonon B E., Beh Mba R., Gbeha M and Glèlè Kakaï R. 2019. Parameters Estimation Methods in Generalized Linear Mixed Models Applied in Ecology: A Critical Review. *International Journal of Engineering and Future Technology* 16(3), 12-27.

McCulloch C.E. and Neuhaus J.M. 2011. Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter. *Stat. Sci.* 26, 388-402.

McNeish D. M. 2016. Estimation Methods for Mixed Logistic Models with Few Clusters. *Multivariate Behavioral Research*, 51(6), 790-804.

Neuhaus J.M., Hauck W.W. and Kalbfleisch J.D. 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* 79, 755-762.

Rue H., Martino S. and Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society B*, 71, 319-392.

Rue H., Riebler A., Sorbye S. H., Illian J. B., Simpson D. P. and Lindgren F. K. 2017. Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4, 395-421.

Verbeke G. and Lesaffre E. 1997. The effect of misspecifiying the random-effects distribution in linear mixed models for longitudinal data. *Comput. Stat. Data Anal.* 23, 541-556.

Vonesh E. F. 2012. Generalized linear and nonlinear models for correlated data: Theory and applications using SAS. Cary, NC: SAS Institute Inc.