Statistica Sinica Preprint No: SS-2024-0346	
Title	Gaussian Variational Approximation with Composite Likelihood for Crossed Random Effect Models
Manuscript ID	SS-2024-0346
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0346
Complete List of Authors	Libai Xu, Nancy Reid and Dehan Kong
Corresponding Authors	Libai Xu
E-mails	lbxu@suda.edu.cn
Notice: Accepted author version.	

Gaussian variational approximation with composite likelihood for crossed random effect models

Libai Xu, Nancy Reid and Dehan Kong

Soochow University and University of Toronto

Abstract: Composite likelihood usually ignores dependencies among response components, while variational approximation to likelihood ignores dependencies among parameter components. What both methods have in common is that they essentially break the dependence of random effects. In this paper, we derive a Gaussian variational approximation to the composite log-likelihood function for Poisson and Gamma models with crossed random effects. We present theoretical aspects of the estimates derived from this approximation and support these theories with simulation studies. Specifically, we show the estimates are consistent with a convergence rate $m^{-1/2}+n^{-1/2}$, where m and n denote the number of rows and columns, respectively. We further provide detailed asymptotic normality results under a new regime where $\log m/\log n \to \delta$ for $\delta \in (1/2, 2)$. Additional simulation studies show that our method yields comparable estimation performance and is slightly faster than the Laplace approximation in the package `glmmTMB` and a Gaussian variational approximation to the full log-likelihood function.

Key words and phrases: Gamma regression, generalized linear mixed models, likelihood inference, Poisson regression.

1. Introduction

Generalized linear mixed models (GLMMs) with crossed random effects are useful for the analysis and inference of cross-classified data, such as arise in educational studies (Chung and Cai, 2021; Menictas et al., 2022), psychometric research studies (Baayen et al., 2008; Jeon et al., 2017), and medical studies (Coull et al., 2001), among others.

Suppose $Y_{ij}, i = 1, \ldots, m; j = 1, \ldots, n$ are conditionally independent, given random effects U_i and V_j. A generalized linear model for Y_{ij} has density function

$$f(Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j) = \exp\left[\{Y_{ij}\theta_{ij} - b(\theta_{ij})\}/a(\phi) + c(Y_{ij}, \phi)\right], \qquad (1.1)$$

and we relate this to the covariates and random effects:

$$g(\mu_{ij}) = \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + U_i + V_j, \qquad (1.2)$$

where $\mu_{ij} = E(Y_{ij})$. In (1.2) $\boldsymbol{\beta}$ is a p-vector of fixed-effects parameters, $\mathbf{X}_{ij} = (1, X_{ij2}, ..., X_{ijp})^{\mathrm{T}}$, and U_i and V_j are independent random effects assumed to follow $N(0, \sigma_u^2)$ and $N(0, \sigma_v^2)$ distributions, respectively.

We develop a Gaussian variational approximation (GVA) to a form of composite likelihood (CL) for model (1.1), in order to make the computations more tractable than in a full likelihood approach. We focus on two examples: Poisson regression, where $Y_{ij} \sim \text{Poisson}(\mu_{ij})$, with $\theta_{ij} = \log(\mu_{ij})$, $b(\theta_{ij}) = \exp(\theta_{ij})$, $g(\mu_{ij}) = \theta_{ij}$, and $\phi = 1$, and Gamma regression, where Y_{ij} distributed as a Gamma distribution with shape parameter $\alpha = \phi^{-1}$ and expectation parameter μ_{ij}, so that $\theta_{ij} = -\mu_{ij}^{-1}$, $b(\theta_{ij}) = \log(\mu_{ij})$, and $g(\mu_{ij}) = \log(\mu_{ij})$.

Some approaches to simplifying the likelihood for crossed random effects have been proposed. Penalized quasi-likelihood is discussed in Breslow and Clayton (1993) and Schall (1991), based on Laplace approximation, although for binary data the resulting estimates are not guaranteed to be consistent (Lin and Breslow, 1996). Coull et al. (2001) developed a Monte Carlo Newton-Raphson algorithm based on expectation-maximization (McCulloch, 1997), and Ghosh et al. (2022a) proposed a backfitting algorithm for a Gaussian linear model, where the integral can be evaluated explicitly. This approach was extended to logistic regression in Ghosh et al. (2022b).

CL methods have also been proposed to simplify computation in models with random effects. Renard et al. (2004) and Bartolucci and Lupparelli (2016) proposed pairwise likelihood for nested random effects models, and this approach has been extended to generalized linear models with crossed random effects by Bellio and Varin (2005). Additionally, Bellio et al. (2023) has developed a rapid CL method based on the all-row-column likelihood for crossed random effects probit models.

Variational Bayes is an alternative method for simplifying high dimensional integrals in models with random effects (Blei et al., 2017). Hall et al. (2011) proved the consistency and asymptotic normality of GVA estimators for the simple Poisson mixed model with one predictor. Ormerod and Wand (2012) extended the GVA to GLMMs. Goplerud et al. (2023) developed the partially factorized variational inference for high-dimensional mixed models. Other related variational approaches are developed in Menictas et al. (2022) and Ghandwani et al. (2023) for crossed random effects linear mixed models, in Shi et al. (2022) for extending the simple Poisson mixed model with one predictor to multiple predictors, and in Rijmen and Jeon (2013), Jeon et al. (2017), Hui et al. (2017), and Hui et al. (2019) for GLMMs.

Our approach combines CL with variational inference. We introduce a row-column CL that requires the computation of only one-dimensional integrals and apply a Gaussian variational approximation to this CL. Our method stands out from the approaches described in Ormerod and Wand (2012) and Hall et al. (2011) due to three main innovations: (a) It allows the model to include more complex random effects beyond merely one random effect; (b) The proposed method allows the model to have more than one predictor; (c) We introduce a new asymptotic regime, $\log m/\log n \to \delta$, where $\delta \in (1/2, 2)$, specifically designed for models with crossed random effects. While our theory builds upon the foundational work of Ormerod and Wand (2012) and Hall et al. (2011) on the simple Poisson mixed model, our unique asymptotic regime, $\log m/\log n \to \delta$, where $\delta \in (1/2, 2)$, sets our approach apart. The previous assumption of $n/m \to 0$ does not suit our model's crossed random effects framework. To our knowledge, this is the first attempt to combine variational inference with composite likelihood. This provides some computational savings, but also suggests that further links between VI and CL methods may be useful in other contexts.

The rest of the paper is organized as follows. In §2, we extend the GVA of the likelihood for GLMMs with crossed random effects. In §3, we introduce a row-column CL that requires the computation of only one-dimensional integrals and apply a GVA to this CL. In §4, we prove the consistency and asymptotic normality of the variational estimates for the Poisson and Gamma regression models. In §5, simulations for both models demonstrate that our method is slightly faster than `glmmTMB` and the GVA applied to the full likelihood function. We apply our method to insurance claim data in §6. Finally, we conclude with a discussion in §7. Additional proofs, simulation results, and an analysis of the insurance data are provided in the supplementary material.

2. Gaussian variational approximation

The log-likelihood function for the generalized linear model with crossed random effects is constructed from the marginal distribution of the responses, Y_{ij}, given the covariates $\mathbf{X}_{ij}$, and depends on the parameters $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2)$:

$$\ell(\boldsymbol{\Psi}) \;=\; \log \int_{R^{m+n}} \prod_{i=1}^{m} \prod_{j=1}^{n} f(Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j) p(U_i) p(V_j) dU_i dV_j, \quad (2.3)$$

$p(U_i)$ and $p(V_j)$ are the probability densities of random effects U_i and V_j, respectively. The maximum likelihood estimator is given by $\hat{\boldsymbol{\Psi}} = \arg\max_{\boldsymbol{\Psi}} \ell(\boldsymbol{\Psi})$. To simplify the computation of the $(m+n)$-dimensional integral we apply a GVA to (2.3) by introducing pairs of variational parameters $(\mu_{u_i}, \lambda_{u_i}), i = 1, ..., m$ and $(\mu_{v_j}, \lambda_{v_j}), j = 1, ..., n$. By Jensen's inequality and concavity of the logarithm function:

$$
\begin{aligned}
& \ell(\boldsymbol{\Psi}) \\
= \;& \log \int_{R^{m+n}} \prod_{i=1}^{m}\prod_{j=1}^{n}\left\{ f(Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j) p(U_i) p(V_j) \frac{\varphi(U_i)\varphi(V_j)}{\varphi(U_i)\varphi(V_j)} dU_i dV_j \right\} \\
\geq \;& \int_{R^{m+n}} \sum_{i=1}^{m}\sum_{j=1}^{n} \varphi(U_i)\varphi(V_j) \log f(Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j) dU_i dV_j \\
& + \int_{R^m} \sum_{i=1}^{m} \varphi(U_i) \log p(U_i) dU_i + \int_{R^n} \sum_{j=1}^{n} \varphi(V_j) \log p(V_j) dV_j \\
& - \int_{R^m} \sum_{i=1}^{m} \varphi(U_i) \log \varphi(U_i) dU_i - \int_{R^n} \sum_{j=1}^{n} \varphi(V_j) \log \varphi(V_j) dV_j = \underline{\ell}(\boldsymbol{\Psi}, \boldsymbol{\xi}),
\end{aligned}
$$

where $\varphi(U_i)$ and $\varphi(V_j)$ are Gaussian densities with means μ_{u_i}, μ_{v_j}, and variances λ_{u_i}, λ_{v_j}, respectively. The function $\underline{\ell}(\boldsymbol{\Psi}, \xi)$ is the variational lower bound on $\ell(\boldsymbol{\Psi})$; the variational parameters are $\boldsymbol{\xi} = (\mu_{u_1}, ..., \mu_{u_m}, \mu_{v_1}, ..., \mu_{v_n}, \lambda_{u_i}, ..., \lambda_{u_m}, \lambda_{v_1}, ..., \lambda_{v_n})^{\mathrm{T}}$. Ignoring some constants, this lower bound simplifies to

$$
\begin{aligned}
\underline{\ell}(\boldsymbol{\Psi}, \boldsymbol{\xi}) \;=\;& \sum_{i=1}^{m}\sum_{j=1}^{n} \left[Y_{ij} E_{\varphi(u_i)\varphi(v_j)}(\theta_{ij}) - E_{\varphi(u_i)\varphi(v_j)}\{b(\theta_{ij})\} \right] \\
& + (1/2) \sum_{i=1}^{m} \left\{ \log(\lambda_{u_i}/\sigma_u^2) - (\mu_{u_i}^2 + \lambda_{u_i})/\sigma_u^2 \right\} \\
& + (1/2) \sum_{j=1}^{n} \{ \log(\lambda_{v_j}/\sigma_v^2) - (\mu_{v_j}^2 + \lambda_{v_j})/\sigma_v^2 \}. \qquad (2.4)
\end{aligned}
$$

The GVA estimators maximize the evidence lower bound,

$$
(\widehat{\underline{\boldsymbol{\Psi}}}, \widehat{\underline{\boldsymbol{\xi}}}) = \arg\max_{\boldsymbol{\Psi}, \boldsymbol{\xi}} \underline{\ell}(\boldsymbol{\Psi}, \boldsymbol{\xi}). \qquad (2.5)
$$

The advantage of using the variational lower bound $\underline{\ell}(\boldsymbol{\Psi}, \boldsymbol{\xi})$ over $\ell(\boldsymbol{\Psi})$ is that the former only contains terms $E_{\varphi(u_i)\varphi(v_j)}(\theta_{ij})$ or $E_{\varphi(u_i)\varphi(v_j)}\{b(\theta_{ij})\}$ involving a two-dimensional integral, and in our models these can be evaluated explicitly. For the Poisson regression model,

$$\begin{aligned} E_{\varphi(u_i)\varphi(v_j)}(\theta_{ij}) &= \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mu_{u_i} + \mu_{v_j}, \\ E_{\varphi(u_i)\varphi(v_j)}\{b(\theta_{ij})\} &= \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mu_{u_i} + \lambda_{u_i}/2 + \mu_{v_j} + \lambda_{v_j}/2). \end{aligned}$$

For the Gamma regression model,

$$\begin{aligned} E_{\varphi(u_i)\varphi(v_j)}(\theta_{ij}) &= -\exp(-\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} - \mu_{u_i} + \lambda_{u_i}/2 - \mu_{v_j} + \lambda_{v_j}/2), \\ E_{\varphi(u_i)\varphi(v_j)}\{b(\theta_{ij})\} &= \mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + \mu_{u_i} + \mu_{v_j}. \end{aligned}$$

3. Composite likelihood Gaussian variational approximation

In GLMM with crossed randon effects, computation of the GVA requires mn two-dimensional integrals; for the Poisson and Gamma regression models this simplifies to mn function evaluations. To reduce computation further we develop a version of CL by considering the two random effects U_i and V_j separately.

First, we ignore column random effects V_j; i.e., in (1.2) we replace $\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + U_i + V_j$ with $\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^r + U_i$, and define the row-CL function by

$$CL_1(\boldsymbol{\beta}^r, \sigma_u^2) = \prod_{i=1}^{m} \int_R \prod_{j=1}^{n} f(Y_{ij} \mid \mathbf{X}_{ij}, U_i) p(U_i) dU_i,$$

where $f(Y_{ij}|\mathbf{X}_{ij}, U_i)$ is a conditional probability density function of Y_{ij} given $\mathbf{X}_{ij}$ and U_i. The row-composite log-likelihood function is

$$c\ell_1(\boldsymbol{\beta}^r, \sigma_u^2) = \sum_{i=1}^{m} \log \int_R \prod_{j=1}^{n} f(Y_{ij} \mid \mathbf{X}_{ij}, U_i) p(U_i) dU_i. \tag{3.6}$$

Similarly, ignoring row random effects, we define the column-composite log-likelihood function

$$c\ell_2(\boldsymbol{\beta}^c, \sigma_v^2) = \sum_{j=1}^{n} \log \int_R \prod_{i=1}^{m} f(Y_{ij} \mid \mathbf{X}_{ij}, V_j) p(V_j) dV_j, \tag{3.7}$$

where $f(Y_{ij}|\mathbf{X}_{ij}, V_j)$ is a conditional probability density function of Y_{ij} given $\mathbf{X}_{ij}$ and V_j. The notation $\boldsymbol{\beta}^r = (\beta_1^r, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ and $\boldsymbol{\beta}^c = (\beta_1^c, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ is needed, as the misspecification caused by ignoring the column (or row) random effects changes the intercept from β_1, say, to $\beta_1^r = \beta_1 + \sigma_v^2/2$ for the row-CL and $\beta_1^c = \beta_1 + \sigma_u^2/2$ for the column-CL, for both the Poisson and the Gamma regression models. The other components, $\beta_2, \beta_3, .., \beta_p$, are unchanged.

Based on (3.6) and (3.7), we propose the misspecified row-column composite log-likelihood function by

$$c\ell(\boldsymbol{\Psi}^{rc}) = c\ell_1(\boldsymbol{\beta}^r, \sigma_v^2) + c\ell_2(\boldsymbol{\beta}^c, \sigma_u^2), \tag{3.8}$$

where $\boldsymbol{\Psi}^{rc} = (\beta_1^r, \beta_1^c, \beta_2, \ldots, \beta_p, \sigma_u^2, \sigma_v^2)^{\mathrm{T}}$. This definition of misspecified composite likelihood reduces the computation of an $(m+n)$-dimensional integral in (2.3) to $m+n$ one-dimensional integrals. Bartolucci et al. (2017) proposed a different CL function in which column (or row) random effects are marginalized over instead.

We construct a GVA to the misspecified row-column CL (3.8) using the same variational distributions as in the previous section, leading to

$$
\begin{aligned}
& c\ell(\boldsymbol{\Psi}^{rc}) \\
= \ & \sum_{j=1}^{n} \log \int_R \prod_{i=1}^{m} f(Y_{ij} \mid \mathbf{X}_{ij}, U_i) p(U_i) \varphi(U_i)/\varphi(U_i) dU_i \\
& + \sum_{i=1}^{m} \log \int_R \prod_{j=1}^{n} f(Y_{ij} \mid \mathbf{X}_{ij}, V_j) p(V_j) \varphi(V_j)/\varphi(V_j) dV_j \\
\geq \ & \sum_{j=1}^{n} \sum_{i=1}^{m} \left[Y_{ij} E_{\varphi(U_i)}(\theta_{ij}^r) - E_{\varphi(U_i)}\{b(\theta_{ij}^r)\} \right] \\
& + (1/2) \sum_{i=1}^{m} \left\{ \log(\lambda_{u_i}/\sigma_u^2) - (\mu_{u_i}^2 + \lambda_{u_i})/\sigma_u^2 \right\} \\
& + \sum_{j=1}^{n} \sum_{i=1}^{m} \left[Y_{ij} E_{\varphi(V_j)}(\theta_{ij}^c) - E_{\varphi(V_j)}\{b(\theta_{ij}^c)\} \right] \\
& + (1/2) \sum_{j=1}^{n} \left\{ \log(\lambda_{v_j}/\sigma_v^2) - (\mu_{v_j}^2 + \lambda_{v_j})/\sigma_v^2 \right\} \\
= \ & \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi}).
\end{aligned} \tag{3.9}
$$

For the Poisson regression model,

$$
\begin{aligned}
& \theta_{ij}^r = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^r + U_i, \ E_{\varphi(u_i)}(\theta_{ij}^r) = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^r + \mu_{u_i}, \\
& E_{\varphi(u_i)}\{b(\theta_{ij}^r)\} = \exp(\mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^r + \mu_{u_i} + \lambda_{u_i}/2); \\
& \theta_{ij}^c = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^c + V_j, \ E_{\varphi(v_j)}(\theta_{ij}^c) = \mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^c + \mu_{v_j}, \\
& E_{\varphi(v_j)}\{b(\theta_{ij}^c)\} = \exp(\mathbf{X}_{ij}^{\mathrm{T}} \boldsymbol{\beta}^c + \mu_{v_j} + \lambda_{v_j}/2).
\end{aligned}
$$

For the Gamma regression model,

$$\theta^r_{ij} = -\exp(-\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^r - U_i),\ E_{\varphi(u_i)}(\theta^r_{ij}) = -\exp(-\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^r - \mu_{u_i} + \lambda_{u_i}/2),$$

$$E_{\varphi(u_i)}\{b(\theta^r_{ij})\} = \mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^r + \mu_{u_i};$$

$$\theta^c_{ij} = -\exp(-\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^c - V_j),\ E_{\varphi(v_j)}(\theta^c_{ij}) = -\exp(-\mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^c - \mu_{v_j} + \lambda_{v_j}/2),$$

$$E_{\varphi(v_j)}\{b(\theta^c_{ij})\} = \mathbf{X}^{\mathrm{T}}_{ij}\boldsymbol{\beta}^c + \mu_{v_j}.$$

We define the estimators based on this Gaussian variational approximation of composite likelihood (GVACL) by

$$(\widehat{\underline{\boldsymbol{\Psi}}}^{rc}, \widehat{\underline{\boldsymbol{\xi}}}) = \underset{\boldsymbol{\Psi}^{rc},\boldsymbol{\xi}}{\arg\max} \underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi}). \tag{3.10}$$

To get the estimator $\widehat{\underline{\boldsymbol{\Psi}}}$ from $\widehat{\underline{\boldsymbol{\Psi}}}^{rc}$, we only need to convert $\widehat{\underline{\beta}^r_1}$ and $\widehat{\underline{\beta}^c_1}$ to $\widehat{\underline{\beta}}_1$: see §4, Remark 1. Solving (2.5) and (3.10) by the Newton-Raphson scheme has computational complexity $O(mn)$. The advantages of $\underline{c\ell}(\boldsymbol{\Psi}^{rc}, \boldsymbol{\xi})$ over $\underline{\ell}(\boldsymbol{\Psi}, \boldsymbol{\xi})$ are that the former does not involve interaction terms of u_i and v_j in Hessian matrix. Especially, for Poisson and Gamma regression models with explicit expressions, integrals are no longer involved. The misspecified row-column CL-based variational approximation enhances the efficiency of calculations compared to the variational approximation to the log-likelihood function.

4. Theoretical Properties

In this section, we present our main results on the consistency and convergence rates of the parameter estimates based on the GVACL introduced above. We denote the true value of the parameter with a superscript 0.

The Poisson regression model we study has

$$Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Poisson}(\mu_{ij}),\ \mu_{ij} = \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j), \tag{4.11}$$

and the Gamma regression model has

$$Y_{ij} \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Gamma}(\alpha, \mu_{ij}),\ \mu_{ij} = \exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta}^0 + U_i + V_j), \tag{4.12}$$

where the shape parameter α is assumed known. In both models we assume U_i and V_j are independent, with $U_i \sim N(0, (\sigma_u^2)^0)$, $V_j \sim N(0, (\sigma_v^2)^0)$.

Under assumptions (A1)-(A9) in §S1 of the Supplementary Material, we have the following two theorems.

Theorem 1. *As $m, n \to \infty$, such that $m = O(n^C)$ and $n = O(m^{C_1})$, where constants $C > 0$ and $C_1 > 0$, for both Poisson and Gamma regression models,*

$$\begin{aligned}
\underline{\widehat{\boldsymbol{\beta}}} - \boldsymbol{\beta}^0 &= O_p(m^{-1/2} + n^{-1/2}),\\
\underline{\widehat{\sigma}}_u^2 - (\sigma_u^2)^0 &= O_p(m^{-1/2} + n^{-1/2}),\\
\underline{\widehat{\sigma}}_v^2 - (\sigma_v^2)^0 &= O_p(m^{-1/2} + n^{-1/2}).
\end{aligned}$$

Remark 1. *As noted in §3, the intercept terms for the row-only and column-only CLs are not equal to β_1. In the proof we derive the probability limits of the variational estimates of β_1^r and β_1^c, and show $\underline{\widehat{\beta}}_1 = (\underline{\widehat{\beta_1^r}} - \underline{\widehat{\sigma}}_v^2/2 + \underline{\widehat{\beta_1^c}} - \underline{\widehat{\sigma}}_u^2/2)/2$.*

In addition, we have the following asymptotic normality results.

Theorem 2. *For both Poisson and Gamma regression models,* $\log m/\log n \to \delta$*, where* $\delta \in (1/2, 2)$ *as* m *and* n *diverge, we have*

$$\underline{\widehat{\beta}}_1 - \beta_1^0 = \begin{cases} m^{-1/2}1_3^{\mathrm{T}}Z_1 + o_p(m^{-1/2}), 1/2 < \delta < 1, \\ m^{-1/2}1_3^{\mathrm{T}}Z_1 + n^{-1/2}1_3^{\mathrm{T}}Z_2 + o_p(m^{-1/2} + n^{-1/2}), \delta = 1, \\ n^{-1/2}1_3^{\mathrm{T}}Z_2 + o_p(n^{-1/2}), 1 < \delta < 2, \end{cases}$$

where $1_3 = (1,1,1)^{\mathrm{T}}$*,* Z_1 *and* Z_2 *are independently and normally distributed random variables with mean zero and covariance*

$$\mathbf{\Gamma}_1 = \frac{1}{8}\begin{pmatrix} 2[\exp\{(\sigma_u^2)^0\} - 1] & 2(\sigma_u^2)^0 & -\{(\sigma_u^2)^0\}^2 \\ 2(\sigma_u^2)^0 & 2(\sigma_u^2)^0 & 0 \\ -\{(\sigma_u^2)^0\}^2 & 0 & \{(\sigma_u^2)^0\}^2 \end{pmatrix},$$

and

$$\mathbf{\Gamma}_2 = \frac{1}{8}\begin{pmatrix} 2[\exp\{(\sigma_v^2)^0\} - 1] & 2(\sigma_v^2)^0 & -\{(\sigma_v^2)^0\}^2 \\ 2(\sigma_v^2)^0 & 2(\sigma_v^2)^0 & 0 \\ -\{(\sigma_v^2)^0\}^2 & 0 & \{(\sigma_v^2)^0\}^2 \end{pmatrix}$$

respectively;

$$\underline{\widehat{\sigma}}_u^2 - (\sigma_u^2)^0 = m^{-1/2}Z_3 + o_p(m^{-1/2}),$$

where the random variable Z_3 *follows* $N(0, 2\{(\sigma_u^2)^0\}^2)$*;*

$$\underline{\widehat{\sigma}}_v^2 - (\sigma_v^2)^0 = n^{-1/2}Z_4 + o_p(n^{-1/2}),$$

where the random variable Z_4 *follows* $N(0, 2\{(\sigma_v^2)^0\}^2)$*.*

For Poisson regression, the slope term $\boldsymbol{\beta}_{-1} = (\beta_2, ..., \beta_p)^{\mathrm{T}}$ *satisfies*

$$\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1} \;\; = \;\; (mn)^{-1/2} Z_5 + o_p\{m^{-1} + (mn)^{-1/2} + n^{-1}\},$$

the random variable Z_5 *follows a normal distribution with zero mean and variance*

$$\exp\{-\beta_1^0 - (\sigma_u^2)^0/2 - (\sigma_v^2)^0/2\}\tau_1\tau_2\tau_1 + 1_2^{\mathrm{T}}\boldsymbol{\Gamma}_3 1_2 \tau_1\tau_3\tau_1,$$

where $\tau_1 = \phi(\boldsymbol{\beta}^0_{-1})\{\phi_2(\boldsymbol{\beta}^0_{-1})\phi(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\}^{-1}$, $\tau_2 = \phi_2(\boldsymbol{\beta}^0_{-1}) - \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}/\phi(\boldsymbol{\beta}^0_{-1})$, $\tau_3 = \phi_2(2\boldsymbol{\beta}^0_{-1}) - 2\phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(2\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}/\phi(\boldsymbol{\beta}^0_{-1}) + \phi_1(\boldsymbol{\beta}^0_{-1})\phi_1(\boldsymbol{\beta}^0_{-1})^{\mathrm{T}}\phi(2\boldsymbol{\beta}^0_{-1})/\phi(\boldsymbol{\beta}^0_{-1})^2$, *and*

$$\boldsymbol{\Gamma}_3 = \frac{1}{4}\begin{pmatrix} \exp\{(\sigma_u^2)^0\}[\exp\{(\sigma_v^2)^0\} - 1] & [\exp\{(\sigma_u^2)^0\} - 1][\exp\{(\sigma_v^2)^0\} - 1] \\ [\exp\{(\sigma_u^2)^0\} - 1][\exp\{(\sigma_v^2)^0\} - 1] & \exp\{(\sigma_v^2)^0\}[\exp\{(\sigma_u^2)^0\} - 1] \end{pmatrix};$$

For Gamma regression, the slope term satisfies

$$\underline{\widehat{\boldsymbol{\beta}}}_{-1} - \boldsymbol{\beta}^0_{-1} = (mn)^{-1/2} 1_2^{\mathrm{T}} Z_6 + o_p\{m^{-1} + (mn)^{-1/2} + n^{-1}\},$$

where Z_6 *follows a normal distribution with zero mean and covariance*

$$\widetilde{\boldsymbol{\Sigma}} = \frac{1}{4}\left(\frac{1+\alpha}{\alpha}[\exp\{(\sigma_u^2)^0\} + \exp\{(\sigma_v^2)^0\}] + \frac{2(1-\alpha)}{\alpha}\right)\boldsymbol{\Sigma}^{-1}_{\mathbf{X}_{-1}}, \qquad (4.13)$$

$\boldsymbol{\Sigma}_{\mathbf{X}_{-1}}$ *is the covariance matrix of* $\mathbf{X}_{-1} = (X_2, ..., X_p)^{\mathrm{T}}$.

Remark 2. *We assume above that* m *and* n *diverge at the same rate, whereas Hall et al. (2011) assume* $n/m \to 0$*, to reflect the usual random effects setting, with many groups (random effects) and relatively few observations per group. While Hall et al. (2011)'s asymptotic assumption is* *sufficient to ensure the validity of the asymptotic theory, it is not necessary. In the GLMM with crossed random effects, if $\delta \in (1/2, 1)$, the convergence rate of $\underline{\widehat{\beta}}_1$ in Theorem 1 will be dominated by $m^{-1/2}$, indicating that the column random effects can be disregarded; if $\delta = 1$, the convergence rate of $\underline{\widehat{\beta}}_1$ in Theorem 1 will be dominated by $m^{-1/2}$ and $n^{-1/2}$; if $\delta \in (1, 2)$, the convergence rate of $\underline{\widehat{\beta}}_1$ in Theorem 1 will be dominated by $n^{-1/2}$, indicating that the row random effects can be disregarded.*

5. Simulation Studies

In this section, we perform simulations to evaluate the effectiveness of the proposed approach. We assess the performance of both Poisson and Gamma regression models with crossed random effects as specified in equations (4.11) and (4.12). For both models, we independently generate $\mathbf{X}_{ij,-1}$ from $N(0, I_{8\times 8})$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$, where $I_{8\times 8}$ is an identity matrix. We set $\boldsymbol{\beta}^0 = (-2, 1, -0.5, 1, -1, 0.5, 1, -1, 0.5, 0.8, -0.4, 0.1)$, $(\sigma_u)^0 = 0.5$, and $(\sigma_v)^0 = 0.5$. For the Gamma regression model, there is an additional shape parameter set as $\alpha = 0.8$. We consider the nine different scenarios $(m, n) \in \{(18, 12), (20, 20), (30, 30), (50, 50), (80, 80), (100, 100), (150, 150), (200, 200), (300, 300)\}$ with $\delta = 1$, which imply the nine increasing sample sizes $m * n \in \{216, 400, 900, 2500, 6400, 10000, 22500, 40000, 90000\}$. Additional simulation settings with $\delta = 1.1$ are included in Supplementary Material §S6.3.

We compare the proposed approach with the GVA inference based on

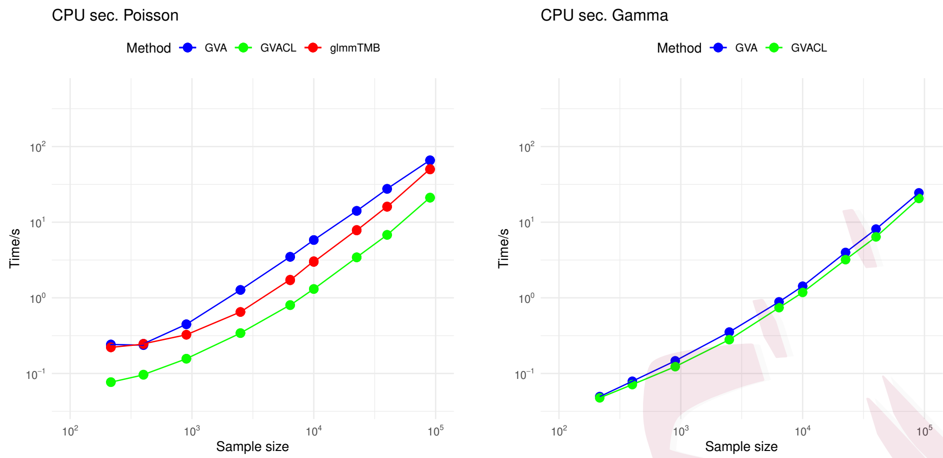

Figure 1: Average time of each simulation in seconds versus sample size $m * n$ under nine different scenarios for the Poisson and Gamma regression model with crossed random effects.

the full log-likelihood function and the Laplace approximation provided in the R package `glmmTMB` (Brooks et al., 2017). We report boxplots of the parameter estimates based on 500 simulations and calculate the average computational times of all methods, using R version 4.1.1 on a laptop equipped with a 1.9 gigahertz Intel Core i7-1370P processor and 64 gigabytes of random access memory. The results are presented in Figures 1-3, S1, and S2. As the package `glmmTMB` cannot accommodate the Gamma regression with an additional fixed shape parameter ($\alpha = 0.8$), the corresponding results for `glmmTMB` are not presented in the right panel of Figure 1. We also calculate actual coverage percentages for the nominally 95% confidence intervals based on the Theorem 2 and the results are presented

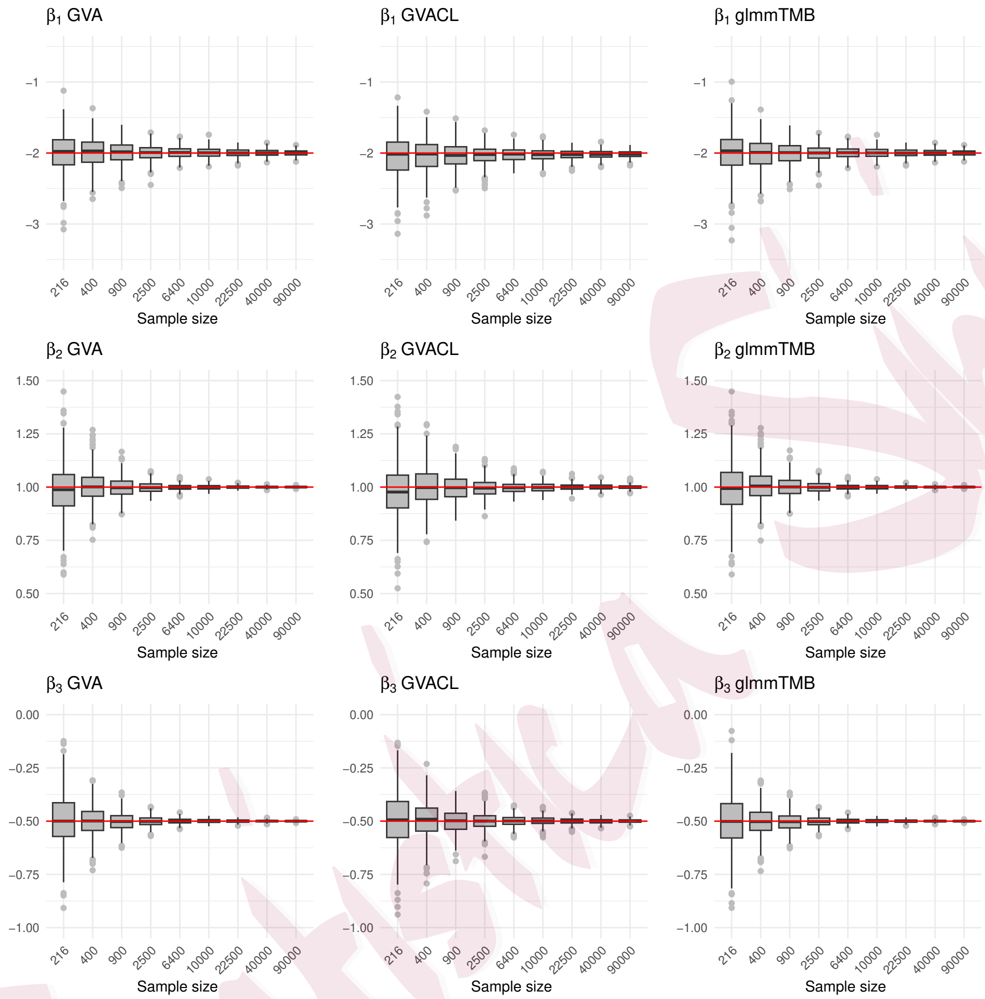

Figure 2: Boxplots of β_1, β_2, and β_3 estimates obtained by GVA, GVACL, and `glmmTMB` for the Poisson regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values.

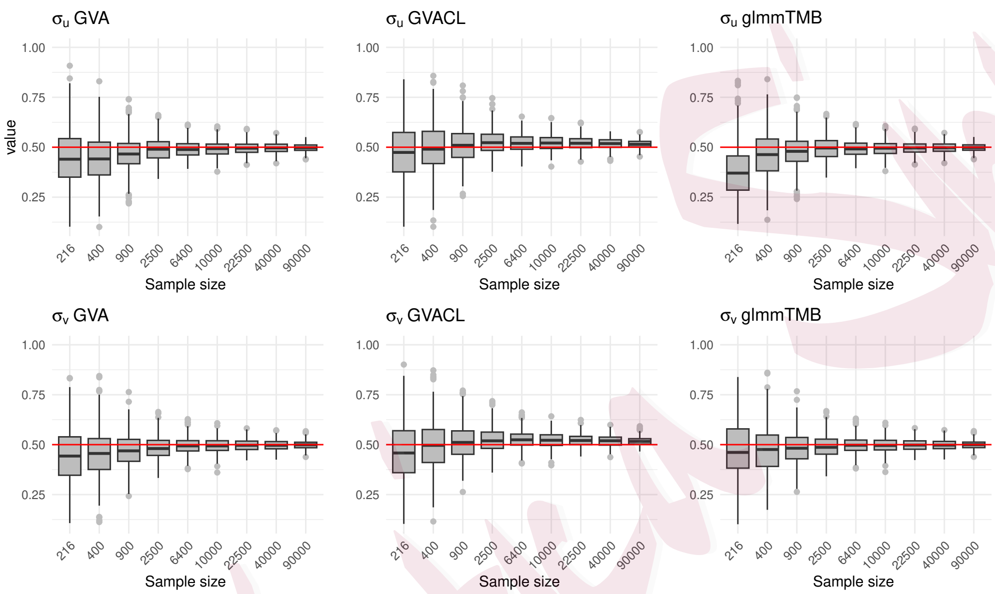

Figure 3: Boxplots of σ_u and σ_v estimates obtained by GVA, GVACL, and `glmmTMB` for Poisson regression model with crossed random effects based on 500 datasets. Red horizontal reference lines represent the true parameter values.

in Figure S3.

In Figures 2, 3, S1, and S2, our results demonstrate that these methods yield similar parameter estimates. However, our proposed approach exhibits a notable advantage in terms of computational costs compared to the GVA to the log-likelihood function and the Laplace approximation in `glmmTMB` from Figure 1.

6. Application

We illustrate our methods with data concerning motor vehicle insurance in Indonesia, in 2014, obtained from the Financial Services Authority (Adam et al., 2021). The dataset consists of 175,381 claim events from 35 distinct areas over a period of 12 months. Each claim event includes information on the claim amount and deductible. However, 17 out of the 35 areas did not have any claim events during the entire 12-month period. Consequently, our analysis focuses on the remaining 18 areas. Table S1 includes the counts of claim events for each area and month.

We randomly selected one claim event from each area and month. The response variable, denoted as Y_{ij}, represents the claim amount, while the explanatory variable, denoted as X_{ij}, represents the deductible, where $i = 1, \ldots, 18$ and $j = 1, \ldots, 12$. A portion of the dataset can be found in Table S4 of the supplementary material. From that table, we can see both X_{ij} and Y_{ij} have large magnitudes, therefore we divide them by 10^7 to avoid singular Hessian matrices in computation (Adam et al., 2021). We then fit a Gamma regression model, introducing random effects for both area and month of occurrence.

Table 1: Insurance data results: "Mean Time(s)" is the average time of each simulation in seconds; "Mean(SD)" reports the mean and sample standard deviations of the model parameters based on the 1000 datasets.

		β_0	β_1	σ_u	σ_v
Method	Mean Time(s)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
GVA	0.19	-0.71 (0.12)	3.17 (1.37)	0.29(0.11)	0.10(0.10)
GVACL	0.13	-0.74 (0.12)	3.60 (1.26)	0.29(0.11)	0.14(0.12)

To evaluate the proposed method and the GVA to the log-likelihood function, we regenerated 1000 datasets using different random seeds at each step of the random event selection process. We then applied these two methods and reported the mean and standard deviation of both estimators across these 1000 datasets.

In addition, we record the average time it takes to fit the model. Table 1 presents the results of this analysis. From Table 1, it is evident that the two methods yield comparable model parameter estimates. However, our proposed approach "GVACL" is faster in terms of computation time.

7. Discussion

In this article, we cover two examples: the Poisson regression and Gamma regression. An interesting future direction is to extend the results to other GLMMs. For example, the logistic regression model with crossed random effects has

$$Y_{ij} = 1 \mid \mathbf{X}_{ij}, U_i, V_j \sim \text{Bernoulli}[1/\{1+\exp(-\mathbf{X}_{ij}\boldsymbol{\beta}^0 - U_i - V_j)\}],$$
$$U_i \sim N(0, (\sigma_u^2)^0), V_j \sim N(0, (\sigma_v^2)^0), U_i \text{ and } V_j \text{ are independent.}$$

Unlike the Poisson and Gamma regression models, there is no analytic solution to the two-dimensional integral

$$\begin{aligned} & E_{\phi(u_i)\phi(v_j)}\{b(\theta_{ij})\} \\ = & \int_{R^2} \log\{1+\exp(\mathbf{X}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + U_i + V_j)\}\phi(U_i)\phi(V_j)dU_i dV_j \end{aligned} \quad (7.14)$$

appearing in (2.4). One solution is to evaluate the integrals using adaptive Gauss-Hermite quadrature (Liu and Pierce, 1994) with a product of $N_1 \times N_2$ quadrature points over the two-dimensional integral in (7.14). The one-dimensional integrals arising in the row-column CL of (3.9) can also be computed by adaptive Gauss-Hermite quadrature. Compared to the Poisson and Gamma cases, the computational time of the proposed method is longer, especially when m and n are large. To avoid evaluating the integrals directly, we derived a lower bound of the variational lower bound for the logistic regression model in Supplementary Material §S6.2. We carried out some limited simulations for the logistic model and reported the results in §S6, Figures S3-S5. We found that the computational time of the proposed method is shorter than that of `glmmTMB` and GVA applied to the log-likelihood function. Proof of consistency and asymptotic normality of logistic regression parameter estimates is a topic for future work.

Variational approximations to the composite log-likelihood function of crossed random effects models establish a connection between two distinct topics in statistics. We introduce the row-CL (3.6) to eliminate the row dependence of responses by disregarding the column random effects. Subsequently, we utilize the GVA to eliminate the column dependence of responses through the variational distributions of row random effects. Similarly, we construct the column-CL (3.7), to eliminate the column dependence of responses by disregarding the random row effects. We then apply the GVA to eliminate the row dependence of responses through the variational distributions of column random effects. Comparing the row-column CL-based GVA and likelihood-based GVA, both approaches share the common goal of breaking the dependence of responses through manipulating random effects for the GLMMs with crossed random effects. Discovering other links between CL and variational approximations is an interesting topic for future research.

Supplementary Material

The online Supplementary Material includes proofs of Theorems 1 and 2, additional simulation results, and additional analyses of the insurance data. The R code for simulations is available at https://github.com/libaixu-

1002/GVACL.

Acknowledgment

The authors are grateful to Dr. Fia Fridayanti Adam for sharing the motor vehicle insurance data, and to the anonymous referees, associate editor, and editor for their careful reading and valuable comments, which greatly improved the overall quality of our original manuscript. This research was partially supported by the Natural Sciences and Engineering Research Council of Canada, the Natural Science Foundation for Youth of Jiangsu Province(SBK20250403335), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (24KJB110024).

References

Adam, F., A. Kurnia, I. Purnaba, I. Mangku, and A. Soleh (2021). Modeling the amount of insurance claim using Gamma linear mixed model with AR (1) random effect. In *Journal of Physics: Conference Series*, Volume 1863, pp. 012027. IOP Publishing.

Baayen, R. H., D. J. Davidson, and D. M. Bates (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language 59*(4), 390–412.

Bartolucci, F., F. Chiaromonte, P. K. Don, and B. G. Lindsay (2017). Composite likelihood inference in a discrete latent variable model for two-way "clustering-by-segmentation" problems. *Journal of Computational and Graphical Statistics 26*(2), 388–402.

Bartolucci, F. and M. Lupparelli (2016). Pairwise likelihood inference for nested hidden Markov chain models for multilevel longitudinal data. *Journal of the American Statistical Association 111*(513), 216–228.

Bellio, R., S. Ghosh, A. B. Owen, and C. Varin (2023). Scalable estimation of probit models with crossed random effects. *arXiv preprint arXiv:2308.15681*.

Bellio, R. and C. Varin (2005). A pairwise likelihood approach to generalized linear models with crossed random effects. *Statistical Modelling 5*(3), 217–227.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association 112*(518), 859–877.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*(421), 9–25.

Brooks, M. E., K. Kristensen, K. J. Van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Machler, and B. M. Bolker (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal 9*(2), 378–400.

Chung, S. and L. Cai (2021). Cross-classified random effects modeling for moderated item calibration. *Journal of Educational and Behavioral Statistics 46*(6), 651–681.

Coull, B. A., J. P. Hobert, L. M. Ryan, and L. B. Holmes (2001). Crossed random effect models for multiple outcomes in a study of teratogenesis. *Journal of the American Statistical Association 96*(456), 1194–1204.

Ghandwani, D., S. Ghosh, T. Hastie, and A. B. Owen (2023). Scalable solution to crossed random effects model with random slopes. *arXiv preprint arXiv:2307.12378*.

Ghosh, S., T. Hastie, and A. B. Owen (2022a). Backfitting for large scale crossed random effects regressions. *The Annals of Statistics 50*(1), 560–583.

Ghosh, S., T. Hastie, and A. B. Owen (2022b). Scalable logistic regression with crossed random effects. *Electronic Journal of Statistics 16*(2), 4604–4635.

Goplerud, M., O. Papaspiliopoulos, and G. Zanella (2023). Partially factorized variational inference for high-dimensional mixed models. *arXiv preprint arXiv:2312.13148*.

Hall, P., J. T. Ormerod, and M. P. Wand (2011). Theory of Gaussian variational approximation for a Poisson mixed model. *Statistica Sinica 21*, 369–389.

Hall, P., T. Pham, M. P. Wand, and S. S. Wang (2011). Asymptotic normality and valid inference for Gaussian variational approximation. *The Annals of Statistics 39*(5), 2502–2532.

Hui, F. K., D. I. Warton, J. T. Ormerod, V. Haapaniemi, and S. Taskinen (2017). Variational approximations for generalized linear latent variable models. *Journal of Computational and Graphical Statistics 26*(1), 35–43.

Hui, F. K., C. You, H. L. Shang, and S. Müller (2019). Semiparametric regression using variational approximations. *Journal of the American Statistical Association 114*, 1765–1777.

Jeon, M., F. Rijmen, and S. Rabe-Hesketh (2017). A variational maximization–maximization algorithm for generalized linear mixed models with crossed random effects. *Psychometrika 82*(3), 693–716.

Lin, X. and N. E. Breslow (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *Journal of the American Statistical Association 91*(435), 1007–1016.

Liu, Q. and D. A. Pierce (1994). A note on Gauss–Hermite quadrature. *Biometrika 81*(3), 624–629.

McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association 92*(437), 162–170.

Menictas, M., G. Di Credico, and M. P. Wand (2022). Streamlined variational inference for linear mixed models with crossed random effects. *Journal of Computational and Graphical Statistics*, 1–17.

Ormerod, J. T. and M. P. Wand (2012). Gaussian variational approximate inference for generalized linear mixed models. *Journal of Computational and Graphical Statistics 21*(1), 2–17.

Renard, D., G. Molenberghs, and H. Geys (2004). A pairwise likelihood approach to estimation in multilevel probit models. *Computational Statistics & Data Analysis 44*(4), 649–667.

Rijmen, F. and M. Jeon (2013). Fitting an item response theory model with random item effects across groups by a variational approximation method. *Annals of Operations Research 206*(1), 647–662.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*(4), 719–727.

Shi, X., X.-S. Wang, and A. Wong (2022). Explicit Gaussian variational approximation for the Poisson lognormal mixed model. *Mathematics 10*(23), 4542.

School of Mathematical Sciences, Soochow University, Suzhou, China

1 Shizi Street, Suzhou 215006, China,

E-mail: lbxu@suda.edu.cn

Department of Statistical Sciences, University of Toronto, Toronto, Canada,

700 University Ave, 9th Floor, Toronto, Ontario M5G 1Z5, Canada.

E-mail: nancym.reid@utoronto.ca and dehan.kong@utoronto.ca