Biometrika (2023), **105**, 1, *pp*. 1–8
Printed in Great Britain Advance Access publication on 31 December 2023

Second term improvement to generalised linear mixed model asymptotics

BY LUCA MAESTRINI
Research School of Finance, Actuarial Studies and Statistics,
The Australian National University, Canberra 2601, Australia
luca.maestrini@anu.edu.au

AISHWARYA BHASKARAN AND MATT P. WAND
School of Mathematical and Physical Sciences, University of Technology Sydney,
P.O. Box 123, Broadway 2007, Australia
aishwarya.bhaskaran@mq.edu.au, matt.wand@uts.edu.au

SUMMARY

A recent article on generalised linear mixed model asymptotics, Jiang *et al.* (2022), derived the rates of convergence for the asymptotic variances of maximum likelihood estimators. If m denotes the number of groups and n is the average within-group sample size then the asymptotic variances have orders m^{-1} and $(mn)^{-1}$, depending on the parameter. We extend this theory to provide explicit forms of the $(mn)^{-1}$ second terms of the asymptotically harder-to-estimate parameters. Improved accuracy of statistical inference and planning are consequences of our theory.

Some key words: Longitudinal data analysis, Maximum likelihood estimation, Sample size calculations.

1. INTRODUCTION

Generalised linear mixed models are a vehicle for regression analysis of grouped data with non-Gaussian responses such as counts and categorical labels. Until recently, the precise asymptotic behaviours of the conditional maximum likelihood estimators were not known for these models. Jiang *et al.* (2022) derived leading term asymptotic variances and showed that they have orders m^{-1} and $(mn)^{-1}$, depending on the parameter, where m is the number of groups and n is the average within-group sample size. The main contribution of this article is to extend the asymptotic variance and covariance approximations to terms in $(mn)^{-1}$ for *all* parameters. This constitutes *second term improvement* to generalised linear mixed model asymptotics. The potential statistical payoffs are improved accuracy of confidential intervals, hypothesis tests, sample size calculations and optimal design.

The essence of generalised linear mixed models is the extension of general linear models via the addition of random effects that allow for the handling of correlations arising from repeated measures. There are numerous types of random effect structures. The most common is the two-level nested structure, corresponding to repeated measures within each of m distinct groups. This version of generalised linear mixed models, with frequentist inference via maximum likelihood and its quasi-likelihood extension, is our focus here.

Suppose that a fixed effects parameter in a two-level generalised linear mixed model is accompanied by a random effect. Jiang *et al.* (2022) showed that the variance of its maximum likelihood estimator, conditional on the predictor data, is asymptotic to $C_1 m^{-1}$ for some deterministic constant C_1 that depends on the true model parameter values. The crux of this article is to extend the asymptotic variance approximation to $C_1 m^{-1} + C_2 (mn)^{-1}$ for an additional deterministic constant C_2. We derive the explicit form of C_2 for two-level nested generalised linear mixed models for both maximum likelihood and maximum quasi-likelihood situations. Even though, in general, C_2 does not have a succinct form it is still usable in that operations such as studentisation are straightforward and result in improvements in statistical utility.

For two-level nested mixed models, $(mn)^{-1}$ is the best possible rate of convergence for the asymptotic variance of the estimator of a model parameter. Such a rate is achieved by maximum likelihood estimators of fixed effects parameters unaccompanied by random effects and dispersion parameters (e.g. Bhaskaran & Wand, 2023). The current article closes the problem of obtaining the precise asymptotic forms of the variances, up to terms in $(mn)^{-1}$, for estimation of *all* model parameters. To achieve this goal, three-dimensional arrays and their combination with regular matrices play a central role. We introduce a new type of array multiplication that streamlines the required manipulations.

2. MODEL DESCRIPTION AND MAXIMUM LIKELIHOOD ESTIMATION

Consider the class of two-parameter exponential family of density, or probability mass, functions with generic form

$$p(y; \eta, \phi) = \exp[\{y\eta - b(\eta) + c(y)\}/\phi + d(y, \phi)]h(y) \tag{1}$$

where η is the *natural parameter* and $\phi > 0$ is the *dispersion parameter*. Examples include the Gaussian density for which $b(x) = \frac{1}{2}x^2$, $c(x) = -\frac{1}{2}x^2$, $d(x_1, x_2) = -\frac{1}{2}\log(2\pi x_2)$ and $h(x) = I(x \in \mathbb{R})$ and the Gamma density function for which $b(x) = -\log(-x)$, $c(x) = \log(x)$, $d(x_1, x_2) = -\log(x_1) - \log(x_2)/x_2 - \log\Gamma(1/x_2)$ and $h(x) = I(x > 0)$. Here $I(\mathcal{P}) = 1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P}) = 0$ if $\mathcal{P}$ is false. The Binomial and Poisson probability mass functions are also special cases of (1) but with ϕ fixed at 1. When (1) is used in regression contexts a common modelling extension for count and proportion responses, usually to account for overdispersion, is to remove the $\phi = 1$ restriction and replace it with $\phi > 0$. In these circumstances

$$\{y\eta - b(\eta) + c(y)\}/\phi + d(y, \phi) \tag{2}$$

is labelled a *quasi-likelihood function* since it is not the logarithm of a probability mass function for $\phi \neq 1$. We use the more general quasi-likelihood terminology for the remainder of this article.

Consider, for observations of the random pairs (X_{ij}, Y_{ij}), $1 \le i \le m$, $1 \le j \le n_i$, generalised linear mixed models of the form,

$$\begin{gathered}Y_{ij}|X_{ij}, U_i \text{ independent having quasi-likelihood function (2) with natural parameter}\\ \left(\beta^0 + \begin{bmatrix} U_i \\ 0 \end{bmatrix}\right)^T X_{ij} \text{ such that the } U_i \text{ are independent } N(0, \Sigma^0) \text{ random vectors.}\end{gathered} \tag{3}$$

The X_{ij} are $d_{\text{F}} \times 1$ random vectors corresponding to predictors. The U_i are $d_{\text{R}} \times 1$ unobserved random effects vectors, where $d_{\text{R}} \le d_{\text{F}}$. Under this set-up the first d_{R} entries of the X_{ij} are partnered by a random effect. The remaining entries correspond to predictors that have a fixed effect only. We assume that the X_{ij} and U_i, for $1 \le i \le m$ and $1 \le j \le n_i$, are totally independent, with the X_{ij} each having the same distribution as the $d_{\mathrm{F}} \times 1$ random vector X and the U_i each having the same distribution as the $d_{\mathrm{R}} \times 1$ random vector U.

For any β ($d_{\mathrm{F}} \times 1$) and Σ ($d_{\mathrm{R}} \times d_{\mathrm{R}}$) that is symmetric and positive definite and conditional on the X_{ij} data, the quasi-likelihood is

$$\ell(\beta,\Sigma) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}[\{Y_{ij}(\beta^T X_{ij} + c(Y_{ij})\}/\phi + d(Y_{ij},\phi)] - \frac{m}{2}\log|2\pi\Sigma|$$
$$+\sum_{i=1}^{m}\log\int_{\mathbb{R}^{d_{\mathrm{R}}}}\exp\left[\frac{1}{\phi}\sum_{j=1}^{n_i}\left\{Y_{ij}\begin{bmatrix}u\\0\end{bmatrix}^T X_{ij} - b\left(\left(\beta+\begin{bmatrix}u\\0\end{bmatrix}\right)^T X_{ij}\right)\right\} - \tfrac{1}{2}u^T\Sigma^{-1}u\right]du.$$

The maximum quasi-likelihood estimator of (β^0,Σ^0) is $(\widehat{\beta},\widehat{\Sigma}) = \mathrm{argmax}_{\beta,\Sigma}\,\ell(\beta,\Sigma)$. In practice computation of $(\widehat{\beta},\widehat{\Sigma})$ can be challenging due to intractable d_{R}-dimensional integrals, although ongoing advances tend to alleviate this problem. We ignore this aspect here and study the theoretical properties of the exact maximum quasi-likelihood estimator rather than approximations to them.

Suppose that $d_{\mathrm{F}} > d_{\mathrm{R}}$ and consider the partition $\beta = [\beta_{\mathrm{A}}^T\ \beta_{\mathrm{B}}^T]^T$ of the fixed effects parameter vector, where β_{A} is $d_{\mathrm{R}} \times 1$ and β_{B} is $(d_{\mathrm{F}} - d_{\mathrm{R}}) \times 1$. The $d_{\mathrm{F}} = d_{\mathrm{R}}$ boundary case is such that β_{B} is null. Also, let $\mathcal{X} \equiv \{X_{ij} : 1 \le i \le m,\ 1 \le j \le n_i\}$. Theorem 1 of Jiang *et al.* (2022) implies that, under some mild conditions, the covariance matrices of $\widehat{\beta}_{\mathrm{A}}$, $\widehat{\beta}_{\mathrm{B}}$ and $\mathrm{vech}(\widehat{\Sigma})$ have leading term behaviour given by

$$\mathrm{Cov}\big(\widehat{\beta}_{\mathrm{A}}|\mathcal{X}\big) = \frac{\Sigma^0\{1+o_p(1)\}}{m},\ \ \mathrm{Cov}\big(\widehat{\beta}_{\mathrm{B}}|\mathcal{X}\big) = \frac{\phi\Lambda_{\beta_{\mathrm{B}}}\{1+o_p(1)\}}{mn}, \tag{4}$$

where $n \equiv \frac{1}{m}\sum_{i=1}^m n_i$, and

$$\mathrm{Cov}\big(\mathrm{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\mathrm{R}}}^{+}(\Sigma^0 \otimes \Sigma^0)D_{d_{\mathrm{R}}}^{+T}\{1+o_p(1)\}}{m}. \tag{5}$$

Here $\Lambda_{\beta_{\mathrm{B}}}$ is a $(d_{\mathrm{F}} - d_{\mathrm{R}}) \times (d_{\mathrm{F}} - d_{\mathrm{R}})$ matrix that depends on β and the (X,U) distribution, $D_{d_{\mathrm{R}}}$ is the matrix of zeroes and ones such that $D_{d_{\mathrm{R}}}\mathrm{vech}(A) = \mathrm{vec}(A)$ for all $d_{\mathrm{R}} \times d_{\mathrm{R}}$ symmetric matrices A and $D_{d_{\mathrm{R}}}^{+} = (D_{d_{\mathrm{R}}}^T D_{d_{\mathrm{R}}})^{-1}D_{d_{\mathrm{R}}}^T$ is the Moore-Penrose inverse of $D_{d_{\mathrm{R}}}$. The theory of Jiang *et al.* (2022) also indicates a degree of asymptotic orthogonality between β_{A} and β_{B} in that $E\big\{(\widehat{\beta}_{\mathrm{A}} - \beta_{\mathrm{A}}^0)(\widehat{\beta}_{\mathrm{B}} - \beta_{\mathrm{B}}^0)^T|\mathcal{X}\big\}$ has $O_p\{(mn)^{-1}\}$ entries, which implies that the correlations between the entries of $\widehat{\beta}_{\mathrm{A}}$ and $\widehat{\beta}_{\mathrm{B}}$ are asymptotically negligible. For Gaussian responses, Lyu & Welsh (2022) considered an extension of (3) for which some entries of X_{ij} are constrained to be constant across all n_i measurements within the ith group. For such constant-within-group predictors they showed that the asymptotic variances of the corresponding fixed effects parameters are of order m^{-1} rather than $(mn)^{-1}$. This type of extension is worthy of future consideration.

The leading term approximations of the variability in $\widehat{\beta}_{\mathrm{A}}$ and $\mathrm{vech}(\widehat{\Sigma})$, given by (4) and (5), are somewhat crude. Unlike the asymptotic covariance of $\widehat{\beta}_{\mathrm{B}}$, they do not show the effect of the average within-group sample size n. In the next section we investigate their second term improvements.

3. TWO-TERM ASYMPTOTIC COVARIANCE RESULTS

We define the two-term asymptotic covariance matrix problem to be the determination of the *unique deterministic* matrices M_β and M_Σ such that, under reasonably mild conditions,

$$\mathrm{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{M_\beta\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\mathrm{Cov}\big(\mathrm{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\mathrm{R}}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\mathrm{R}}}^{+T}}{m} + \frac{M_\Sigma\{1+o_p(1)\}}{mn}.$$

An example for which a solution to the two-term asymptotic covariance problem can be expressed relatively simply is the $d_{\mathrm{F}}=2$, $d_{\mathrm{R}}=1$ Poisson quasi-likelihood special case of (3), with parameters $\beta=(\beta_0,\beta_1)$ and $\Sigma=\sigma^2$ and predictor variable $[1\ X]^T$ for a scalar random variable X. Define

$$a_1\big(\beta_0,\beta_1,\sigma^2\big) \equiv e^{\beta_0+\sigma^2/2}\big[E(X^2e^{\beta_1 X})E(e^{\beta_1 X}) - \{E(Xe^{\beta_1 X})\}^2\big] \quad \text{and}$$

$$a_2(\beta_1,\sigma^2) \equiv \frac{e^{\sigma^2}E\big(X^2e^{\beta_1 X}\big)E\big(e^{\beta_1 X}\big) + \big(1-e^{\sigma^2}\big)\{E\big(Xe^{\beta_1 X}\big)\}^2}{E\big(e^{\beta_1 X}\big)}.$$

Then the two-term covariance matrix of $(\widehat{\beta}_0,\widehat{\beta}_1)$ is

$$\mathrm{Cov}\left(\begin{bmatrix}\widehat{\beta}_0\\ \widehat{\beta}_1\end{bmatrix}\middle|\,\mathcal{X}\right) = \frac{1}{m}\begin{bmatrix}(\sigma^2)^0 & 0\\ 0 & 0\end{bmatrix} + \frac{\phi\{1+o_p(1)\}}{a_1\big(\beta_0^0,\beta_1^0,(\sigma^2)^0\big)\,mn}\begin{bmatrix}a_2\big(\beta_1^0,(\sigma^2)^0\big) & -E\big(Xe^{\beta_1^0 X}\big)\\ -E\big(Xe^{\beta_1^0 X}\big) & E\big(e^{\beta_1^0 X}\big)\end{bmatrix}.$$

Studentisation of the two-term asymptotic covariance matrix for obtaining confidence intervals and Wald hypothesis tests is straightforward. For example, $E(X^2e^{\beta_1^0 X})$ can be replaced by the estimator $(mn)^{-1}\sum_{i=1}^m\sum_{j=1}^{n_i}X_{ij}^2e^{\widehat{\beta}_1 X_{ij}}$.

The remainder of this section is concerned with the *theoretical* problem of obtaining the forms of M_β and M_Σ for model (3) in general. The score asymptotic approximation approach used in Jiang *et al.* (2022) requires higher numbers of terms to obtain valid two-term covariance matrix approximations. Some of these terms can only be expressed using three-dimensional arrays rather than with matrices. Succinct statement of M_β and M_Σ is only possible with well-designed nested function notation. A novel notation for multiplicative combining of three-dimensional arrays with compatible matrices is also beneficial. The next subsection focusses on these notational aspects.

3.1. *Notation for the Main Result*

Let $\mathcal{A}$ be a $d_1\times d_2\times d_3$ array and M be a $d_1\times d_2$ matrix. Then we let

$$\mathcal{A}\bigstar M \quad \text{denote the } d_3\times 1 \text{ vector with } t\text{th entry given by} \quad \sum_{r=1}^{d_1}\sum_{s=1}^{d_2}(\mathcal{A})_{rst}(M)_{rs}. \tag{6}$$

Next, for $U\sim N(0,\Sigma^0)$, define

$$\Omega_{\mathrm{AA}}(U) \equiv E\Big\{b''\big((\beta_{\mathrm{A}}^0+U)^TX_{\mathrm{A}} + (\beta_{\mathrm{B}}^0)^TX_{\mathrm{B}}\big)X_{\mathrm{A}}X_{\mathrm{A}}^T|U\Big\},$$

$$\Omega_{\mathrm{AB}}(U) \equiv E\Big\{b''\big((\beta_{\mathrm{A}}^0+U)^TX_{\mathrm{A}} + (\beta_{\mathrm{B}}^0)^TX_{\mathrm{B}}\big)X_{\mathrm{A}}X_{\mathrm{B}}^T|U\Big\}$$

$$\text{and}\quad \Omega_{\mathrm{BB}}(U) \equiv E\Big\{b''\big((\beta_{\mathrm{A}}^0+U)^TX_{\mathrm{A}} + (\beta_{\mathrm{B}}^0)^TX_{\mathrm{B}}\big)X_{\mathrm{B}}X_{\mathrm{B}}^T|U\Big\}.$$

Also let $\Omega'_{\mathrm{AAA}}(U)$ be the $d_{\mathrm{R}} \times d_{\mathrm{R}} \times d_{\mathrm{R}}$ array with (r,s,t) entry equal to

$$E\Big\{b'''\big((\beta^0_{\mathrm{A}}+U)^T X_{\mathrm{A}} + (\beta^0_{\mathrm{B}})^T X_{\mathrm{B}}\big)(X_{\mathrm{A}})_r(X_{\mathrm{A}})_s(X_{\mathrm{A}})_t|U\Big\}$$

and $\Omega'_{\mathrm{AAB}}(U)$ be the $d_{\mathrm{R}} \times d_{\mathrm{R}} \times (d_{\mathrm{F}} - d_{\mathrm{R}})$ array with (r,s,t) entry equal to

$$E\Big\{b'''\big((\beta^0_{\mathrm{A}}+U)^T X_{\mathrm{A}} + (\beta^0_{\mathrm{B}})^T X_{\mathrm{B}}\big)(X_{\mathrm{A}})_r(X_{\mathrm{A}})_s(X_{\mathrm{B}})_t\big|U\Big\}.$$

Define the random vectors:

$$\psi_1(U) \equiv \mathrm{vech}(\Sigma^0 - UU^T),\;\; \psi_2(U) \equiv \Omega'_{\mathrm{AAA}}(U)\bigstar\Omega_{\mathrm{AA}}(U)^{-1},\;\; \psi_3(U) \equiv \Omega'_{\mathrm{AAB}}(U)\bigstar\Omega_{\mathrm{AA}}(U)^{-1}$$
$$\text{and}\;\; \psi_4(U) \equiv D^+_{d_{\mathrm{R}}}\mathrm{vec}\Big(\Omega_{\mathrm{AA}}(U)^{-1}(\Sigma^0)^{-1}\big\{\Sigma^0 - UU^T - \Sigma^0\psi_2(U)U^T\big\}\Big).$$

Then define the random matrices:

$$\begin{aligned}
&\Psi_5(U) \equiv \Omega_{\mathrm{AA}}(U)^{-1}\Omega_{\mathrm{AB}}(U), \quad \Psi_6(U) \equiv \Omega_{\mathrm{BB}}(U) - \Psi_5(U)^T\Omega_{\mathrm{AB}}(U),\\
&\Psi_7(U) \equiv UU^T(\Sigma^0)^{-1}\Omega_{\mathrm{AA}}(U)^{-1}, \quad \Psi_8(U) \equiv D^+_{d_{\mathrm{R}}}\big[(UU^T)\otimes\{\Omega_{\mathrm{AA}}(U)^{-1}\}\big]D^{+T}_{d_{\mathrm{R}}}\\
\text{and}\quad &\Psi_9(U) \equiv \psi_1(U)\psi_4(U)^T + \psi_4(U)\psi_1(U)^T.
\end{aligned}$$

Lastly, define the expectation matrices:

$$\begin{aligned}
\Lambda_{\mathrm{AA}} &\equiv E\Big\{\Psi_7(U) + \Psi_7(U)^T - \Omega_{\mathrm{AA}}(U)^{-1} + \Omega_{\mathrm{AA}}(U)^{-1}\psi_2(U)U^T + U\psi_2(U)^T\Omega_{\mathrm{AA}}(U)^{-1}\Big\},\\
\Lambda_{\mathrm{AB}} &\equiv E\Big\{UU^T(\Sigma^0)^{-1}\Psi_5(U) + U\psi_2(U)^T\Psi_5(U) - U\psi_3(U)^T\Big\} \quad \text{and}\\
\Delta &\equiv E\Big(\Big[\Psi_5(U)^T\big\{(\Sigma^0)^{-1}U + \psi_2(U)\big\} - \psi_3(U)\Big]\psi_1(U)^T\Big).
\end{aligned}$$

3.2. *Assumptions for the Main Result*

The main result depends on the following sample size asymptotic assumptions: the number of groups m diverges to ∞; the within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \le i \le m$; the ratio n/m converges to zero. The last of these conditions is in keeping with the number of groups being large compared with the within-group sample sizes, as often arises in practice. For our asymptotics it ensures that, for the harder-to-estimate parameters, the asymptotic variances of the maximum likelihood estimators have leading terms of the form $C_1 m^{-1} + C_2(mn)^{-1}$. In addition, it ensures that the Fisher information is sufficiently dominant for obtaining asymptotic variances.

We also assume that the (X,U) joint distribution is such that all required convergence in probability limits that appear in the deterministic order $(mn)^{-1}$ terms are justified. The determination of sufficient conditions on the (X,U) distribution that guarantee the validity of the main result is a tall order, involving the determination of at least eighteen additional moment-type conditions for results similar to Lemma A1 of Jiang *et al.* (2022), and beyond the scope of this article.

3.3. *Statement of the Main Result*

Using the notation presented in § 3.1 and under the assumptions described in § 3.2, and assuming $d_{\mathrm{F}} > d_{\mathrm{R}}$ we have

$$\mathrm{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{\phi}{mn}\begin{bmatrix}\Lambda_{\mathrm{AA}}^{-1} & \Lambda_{\mathrm{AA}}^{-1}\Lambda_{\mathrm{AB}}\\ \Lambda_{\mathrm{AB}}^T\Lambda_{\mathrm{AA}}^{-1} & \Lambda_{\mathrm{AB}}^T\Lambda_{\mathrm{AA}}^{-1}\Lambda_{\mathrm{AB}} + E\big\{\Psi_6(U)\big\}\end{bmatrix}^{-1}\{1+o_p(1)\}$$

$$\text{and } \mathrm{Cov}\big(\mathrm{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\mathrm{R}}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\mathrm{R}}}^{+T}}{m} \qquad (7)$$

$$+\frac{\phi}{mn}\Big(2E\big\{\Psi_9(U) - 2\Psi_8(U)\big\} + \Delta^T\big[E\{\Psi_6(U)\}\big]^{-1}\Delta\Big)\{1+o_p(1)\}.$$

For the $d_{\mathrm{F}} = d_{\mathrm{R}}$ boundary case the first term of $\mathrm{Cov}\big(\widehat{\beta}|\mathcal{X}\big)$ is simply Σ^0/m. A supplement to this article contains a full derivation of (7).

In the Gaussian response special case we have $b''(x) = 1$ and $b'''(x) = 0$ and the main result reduces to the following succinct form:

$$\mathrm{Cov}\big(\widehat{\beta}|\mathcal{X}\big) = \frac{1}{m}\begin{bmatrix}\Sigma^0 & O\\ O & O\end{bmatrix} + \frac{\phi\big\{E(XX^T)\big\}^{-1}\{1+o_p(1)\}}{mn} \quad \text{and}$$

$$\mathrm{Cov}\big(\mathrm{vech}(\widehat{\Sigma})|\mathcal{X}\big) = \frac{2D_{d_{\mathrm{R}}}^{+}(\Sigma^0\otimes\Sigma^0)D_{d_{\mathrm{R}}}^{+T}}{m} + \frac{4\phi D_{d_{\mathrm{R}}}^{+}\big[\Sigma^0\otimes\{E(X_{\mathrm{A}}X_{\mathrm{A}}^T)\}^{-1}\big]D_{d_{\mathrm{R}}}^{+T}\{1+o_p(1)\}}{mn}.$$

We are not aware of any previous appearances of this result in the linear mixed model literature.

4. UTILITY OF THE SECOND TERM IMPROVEMENTS

The second term improvements of (7) have ready and straightforward applications to confidence intervals, Wald hypothesis tests and sample size calculations. Optimal design is another possible utility, but would require second term improvements of the type of theory given in § 5 of Jiang *et al.* (2022). In order to understand potential practical impacts of second term improvements to generalised linear mixed model asymptotics, we present an illustration on sample size calculations for a Poisson-response mixed model. In the supplement we report results from simulation exercises involving confidence intervals for the parameters of a logistic regression model. These assess the improvements afforded by our two-term asymptotic covariance expressions against the theory of Jiang *et al.* (2022) and also serve as a comparison to existing software.

Consider the following $d_{\mathrm{F}} = d_{\mathrm{R}} = 2$ Poisson quasi-likelihood special case of (3):

$$Y_{ij}|X_{ij}, U_{0i}, U_{1i},\ 1\le i\le m,\ 1\le j\le n,\ \text{independently distributed as}$$

$$\text{Poisson}\big[\exp\big\{\beta_0^0 + U_{0i} + (\beta_1^0 + U_{1i})X_{ij}\big\}\big] \text{ where } \begin{bmatrix}U_{0i}\\ U_{1i}\end{bmatrix}^T \text{ are independent } N(0,\Sigma^0) \qquad (8)$$

random vectors and the X_{ij}'s are independently drawn from $X \sim \text{Bernoulli}(p)$.

Suppose the model above is expected to be adopted in a study involving m subjects. Now suppose we would like to determine the required number of subjects m to detect a possibly positive effect of the binary predictor X at a global level by testing

$$H_0\ :\ \beta_1^0 = 0 \quad \text{versus} \quad H_1\ :\ \beta_1^0 > 0 \qquad (9)$$

with a significance level α and at least P power. If $\beta_1^a > 0$ is a particular alternative value of β_1^0 and $\hat{\beta}_1$ an estimate of β_1^0, then standard arguments lead to the sample size being the solution to

$$\frac{\beta_1^a}{\sqrt{\text{Var}(\hat{\beta}_1)}} = \Phi^{-1}(\alpha) + \Phi^{-1}(1-P) \tag{10}$$

where Φ^{-1} is the $N(0,1)$ quantile function. Application of the main result (7) to model (8) and the derivations provided in the supplement lead to the two-term asymptotic variance of $\hat{\beta}_1$ being

$$\text{Asy.Var}(\hat{\beta}_1) = \frac{\Sigma_{11}^0}{m} + \frac{\phi c(\beta^0, \Sigma^0)}{mn} \quad \text{with}$$

$$c(\beta^0, \Sigma^0) \equiv \frac{1}{p}\exp\left\{-\beta_0^0 - \beta_1^0 + \frac{1}{2}(\Sigma_{00}^0 + 2\Sigma_{01}^0 + \Sigma_{11}^0)\right\} + \frac{1}{1-p}\exp\left\{-\beta_0^0 + \frac{\Sigma_{00}^0}{2}\right\}.$$

where $\beta^0 = [\beta_0^0\ \beta_1^0]$ and $\Sigma^0 = \text{vech}^{-1}\left([\Sigma_{00}^0\ \Sigma_{01}^0\ \Sigma_{11}^0]^T\right)$. This can be used to replace $\text{Var}(\hat{\beta}_1)$ in (10), providing the following lower bound for the number of subjects required to achieve at least P power in test (9) at the α significance level:

$$m = \left\lceil \frac{1}{(\beta_1^a)^2}\left\{\Sigma_{11}^0 + \frac{\phi c(\beta^0, \Sigma^0)}{n}\right\}\{\Phi^{-1}(\alpha) + \Phi^{-1}(1-P)\}^2 \right\rceil \tag{11}$$

where, for any $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

We conducted a simulation exercise aimed at understanding whether the number of subjects m chosen according to the two-term asymptotic variance of $\hat{\beta}_1$ leads to the advertised power for hypothesis tests. The simulation study involved producing 1,000 replicates corresponding to (8) with $p = 0.5$ and average group size $n = 20$ for various combinations of β_0^0, β_1^0 and Σ^0 according to Table 1. We then fitted model (8) to each simulated dataset via the glmmTMB package (Brooks *et al.*, 2023) in R, and assumed $\phi = 1$, $\alpha = 0.05$ and $P = 0.9$. Table 1 shows the empirical estimates of P and corresponding 95% confidence intervals resulting from both the fully asymptotic theory of Jiang *et al.* (2022) and our two-term asymptotic results. For this illustration we see that the sample size formula (11) performs very well with regards to the actual power delivered, with the true power value P falling inside of all the confidence intervals. When the other parameters are held fixed, the required number of subjects decreases for increasing values of $\exp(\beta_0^0)$, larger β_1^0 values or smaller within-group variation. On the other hand, the minimum number of subjects values obtained by plugging the one-term asymptotic variance of $\hat{\beta}_1$ in (10) are substantially different from those computed using (11) and produced empirical estimates of power that are well below the advertised level P for all the β_0^0, β_1^0 and Σ^0 combinations.

Simulation results such as those summarised by Table 1 provide an appreciation for the practical utility of our second term improvement to generalised linear mixed models asymptotics.

ACKNOWLEDGEMENTS

We are grateful to Mauro Bernardi, Francis Hui, Alessandra Salvan and Nicola Sartori for advice related to this research, and to the associate editor and referee who reviewed our manuscript for their effort in improving the paper. This research was supported by the Australian Research Council Discovery Project DP230101179.

	$\beta_1^0 = 0.3$, $\Sigma^0 = \text{vech}^{-1}([0.5\ 0.1\ 0.25]^T)$					
	$\beta_0^0 = -1.5$		$\beta_0^0 = -0.5$		$\beta_0^0 = 0.5$	
	1-term var.	2-term var.	1-term var.	2-term var.	1-term var.	2-term var.
Minimum m:	24	130	24	63	24	39
Power estimate:	37.0	88.0	58.5	90.1	75.8	89.8
Power conf. int.:	$(34.0, 40.0)$	$(86.0, 90.0)$	$(55.4, 61.6)$	$(88.2, 92.0)$	$(73.1, 78.5)$	$(87.9, 91.7)$

	$\beta_0^0 = -1.5$, $\Sigma^0 = \text{vech}^{-1}([0.5\ 0.1\ 0.25]^T)$					
	$\beta_1^0 = 0.2$		$\beta_1^0 = 0.4$		$\beta_1^0 = 0.5$	
	1-term var.	2-term var.	1-term var.	2-term var.	1-term var.	2-term var.
Minimum m:	54	304	14	71	9	44
Power estimate:	32.8	88.5	40.5	89.7	41.6	89.8
Power conf. int.:	$(29.9, 35.7)$	$(86.5, 90.5)$	$(37.5, 43.5)$	$(87.8, 91.6)$	$(38.5, 44.7)$	$(87.9, 91.7)$

	$\beta_0^0 = -1.5$, $\beta_1^0 = 0.3$					
	$\Sigma^0 = \text{vech}^{-1}([0.5\ -0.1\ 0.25]^T)$		$\Sigma^0 = \text{vech}^{-1}([1\ 0.2\ 0.5]^T)$		$\Sigma^0 = \text{vech}^{-1}([1\ -0.2\ 0.5]^T)$	
	1-term var.	2-term var.	1-term var.	2-term var.	1-term var.	2-term var.
Minimum m:	24	121	48	200	48	173
Power estimate:	33.5	88.4	46.5	92.2	47.1	91.2
Power conf. int.:	$(30.6, 36.4)$	$(86.4, 90.4)$	$(43.4, 49.6)$	$(90.5, 93.9)$	$(44.0, 50.2)$	$(89.4, 93.0)$

Table 1. *The results from the illustrative sample size calculation and corresponding checks of empirical power (as a percentage) for the simulation study described in the text with* $n = 20$ *and for various combinations of* β_0^0, β_1^0 *and* Σ^0 *values. The values of the minimum number of subjects* m *correspond to an advertised power of 90% and are calculated using both Asy.Var*$(\hat{\beta}_1) = \Sigma_{11}^0/m$ *('1-term var.') and Asy.Var*$(\hat{\beta}_1) = \Sigma_{11}^0/m + \phi c(\beta^0, \Sigma^0)/(mn)$ *('2-term var.'). The 95% confidence intervals of power are also provided.*

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online contains derivational details and additional simulation results.

REFERENCES

BHASKARAN, A. AND WAND, M.P. (2023). Dispersion parameter extension of precise generalized linear mixed model asymptotics. *Statist. Probab. Lett.*, **193**, Article 109691.

BROOKS, M., BOLKER, B., KRISTENSEN, K., MAECHLER, M., MAGNUSSON, A., SKAUG, H. NIELSEN, A., BERG, C., VAN BENTHAM, K. (2023). glmmTMB 1.1.7: Generalized linear mixed models using Template Model Builder. R package.

JIANG, J., WAND, M.P. & BHASKARAN, A. (2022). Usable and precise asymptotics for generalized linear mixed model analysis and design. *J. R. Statist. Soc., Ser. B*, **84**, 55–82.

LYU, Z. & WELSH, A.H. (2022). Increasing cluster size asymptotics for nested error regression models. *J. Statist. Plan. Inf.*, **217**, 52–68.

[*Received on* 30 $March$ 2023. *Editorial decision on* 13 $November$ 2023]