Statistics and Probability Letters 193 (2023) 109691

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Dispersion parameter extension of precise generalized linear mixed model asymptotics

Aishwarya Bhaskaran *, Matt P. Wand

University of Technology Sydney, Australia

ARTICLE INFO

Article history:
Received 1 March 2022
Received in revised form 24 August 2022
Accepted 18 September 2022
Available online 28 September 2022

Keywords:
Longitudinal data analysis
Maximum likelihood estimation
Multilevel models
Studentization

ABSTRACT

We extend a recent asymptotic normality theorem for generalized linear mixed models to include the dispersion parameter. The maximum likelihood estimators of all model parameters have asymptotically normal distributions with asymptotic mutual independence between fixed effects, covariance and dispersion parameters.

1. Introduction

We extend the main theorem of Jiang et al. (2022) to include conditional maximum likelihood estimation of the dispersion parameter. The essence of the new findings is that the dispersion parameter maximum likelihood estimator has a simple asymptotic normal distribution, identical to that for the generalized linear model case, which is amenable to practical inference. Moreover, we establish asymptotic orthogonality between the dispersion parameter and the other model parameters.

The motivations and benefits of precise asymptotics for generalized linear mixed models are described in Section 1 of Jiang et al. (2022). As mentioned there, books such as Faraway (2016), Jiang and Nguyen (2021), McCulloch et al. (2008) and Stroup (2013) provide summaries and access to the large literature on generalized linear mixed models. Fig. 1 provides visualization of a data set that potentially benefits from generalized linear mixed model analysis. A variable of interest is mathematics achievement score for 7185 United States of America school students across 160 schools. In Fig. 1 each panel corresponds to a school and mathematics achievement is plotted against socio-economic status. The color-coding is according to sex and minority categorical variables. The data are from a 1980s survey known as "High School and Beyond" and an early source is Coleman et al. (1982). The Fig. 1 data is stored in the data frame `MathAchieve` within the package `nlme` (Pinheiro et al., 2022) of the `R` computing environment (R Core Team, 2022).

Estimation of the dispersion parameter, denoted here by ϕ, was not considered by Jiang et al. (2022) for a few reasons. One is that ϕ is often treated as a nuisance parameter, with the fixed and random effects being of primary interest. Another is that maximum likelihood does not apply for the most common families: Bernoulli and Poisson. Instead, for these two families, quasi-likelihood is required for the $\phi \neq 1$ extension. A third reason is that, even when maximum likelihood estimation is available for situations such as Gamma response models, it is common to use the simpler Pearson estimator instead. Nevertheless, maximum likelihood estimation of ϕ is viable in important generalized linear mixed model contexts and a full treatment of precise asymptotics requires its inclusion.

* Corresponding author.
E-mail address: aishwarya.bhaskaran@student.uts.edu.au (A. Bhaskaran).

https://doi.org/10.1016/j.spl.2022.109691
0167-7152/

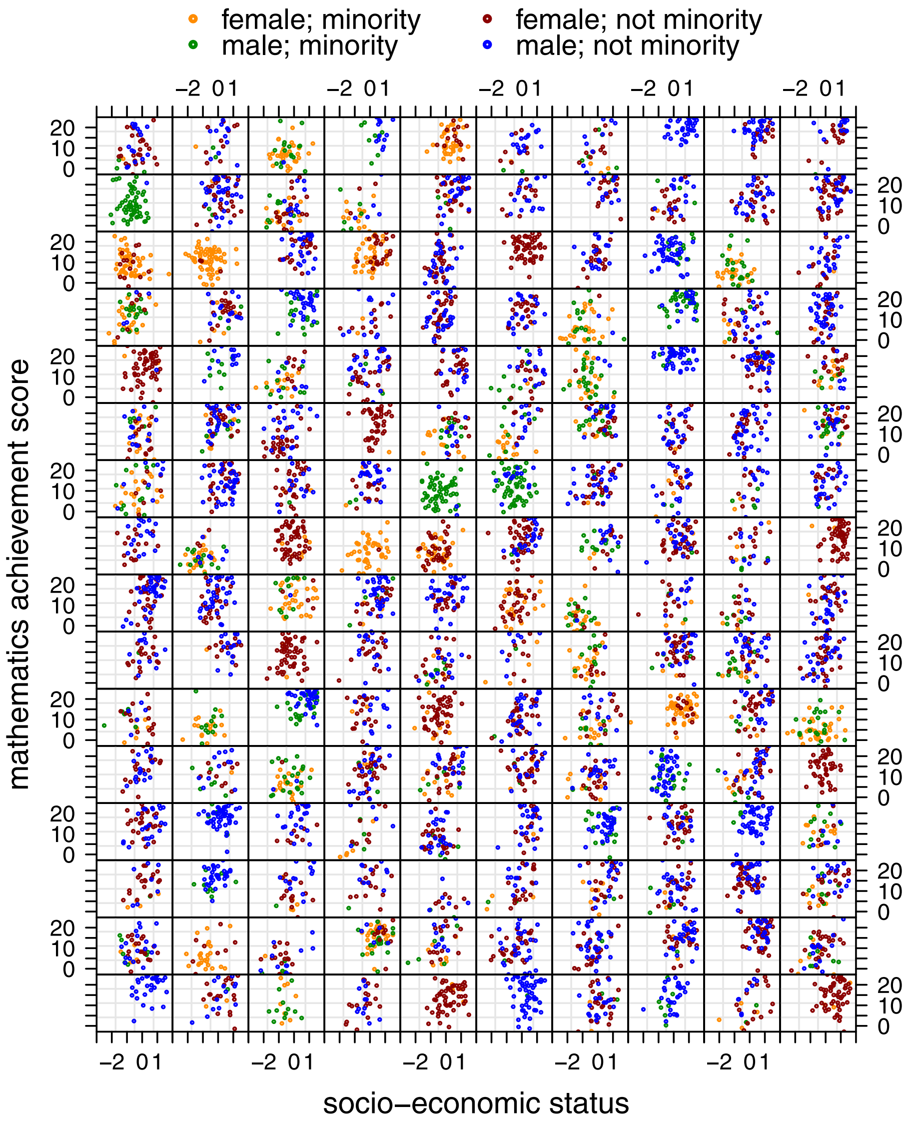

Fig. 1. A visualization of the data set concerning mathematics achievement described in the text. Each panel corresponds to one of 160 schools and plots a mathematics achievement score against a socio-economic status index for students in that school. The color-coding corresponds to the sex and minority status categorical variables.

Section 2 describes the data, generalized linear mixed models set-up and maximum likelihood estimation of the model parameters. The focal point of this article is the asymptotic normality theorem stated in Section 3 with proof provided in supplemental material. Asymptotically valid likelihood-based inference for the dispersion parameter is discussed in Section 4. Sections 2–4 are confined to reproductive exponential families, which covers the common cases arising in generalized linear mixed model applications. The extension to general exponential families is discussed in Section 6.

2. Model description and maximum likelihood estimation

Consider the class of two-parameter reproductive linear exponential family density functions with generic form

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

where η is the *natural parameter* and $\phi>0$ is the *dispersion parameter*. The definition of a reproductive exponential family is given in, for example, Jørgensen (1987). Not all two-parameter linear exponential family density functions are reproductive. However, those families commonly used in applications of generalized linear and mixed models are reproductive and Sections 2–4 are restricted to this case. We discuss the general case briefly in Section 6.

Table 1
Specific two-parameter exponential families and their b, c, d, e and h functions.

Family	$b(\eta)$	$c(y)$	$d(\phi)$	$e(y)$	$h(y)$
Gaussian	$\frac{1}{2}\eta^2$	$-\frac{1}{2}y^2$	$\frac{1}{2}\log(\phi)$	$\frac{1}{2}\log(2\pi)$	1
Gamma	$-\log(-\eta)$	$\log(y)$	$\log\big(\phi^{1/\phi}\,\Gamma(1/\phi)\big)$	$\log(y)$	$I(y>0)$
Inverse Gaussian	$-(-2\eta)^{1/2}$	$-1/(2y)$	$\frac{1}{2}\log(\phi)$	$\frac{1}{2}\log(2\pi y^3)$	$I(y>0)$

Table 1 gives some explicit examples of the b, c, d, e and h functions appearing in (1). In Table 1 the following notation is used: $I(\mathcal{P})=1$ if the condition $\mathcal{P}$ is true and $I(\mathcal{P})=0$ if $\mathcal{P}$ is false. Theoretical results given in Blæsild and Jensen (1985) imply that the three families listed in Table 1 are the *only* possibilities for $p(y;\eta,\phi)$, even if $-d(\phi)-e(y)$ is relaxed to be a general bivariate function of (ϕ,y). Therefore, without loss of generality, we can assume that $p(y;\eta,\phi)$ is one of the three forms given by Table 1.

In this article we study generalized linear mixed models of the form, for observations of the random triples $(\boldsymbol{X}_{\mathrm{A}ij},\boldsymbol{X}_{\mathrm{B}ij},Y_{ij})$, $1\le i\le m$, $1\le j\le n_i$,

$$\begin{array}{l} Y_{ij}|\boldsymbol{X}_{\mathrm{A}ij},\boldsymbol{X}_{\mathrm{B}ij},\boldsymbol{U}_i \text{ independent having density function (1) with natural}\\ \text{parameter } (\boldsymbol{\beta}_{\mathrm{A}}^0+\boldsymbol{U}_i)^T\boldsymbol{X}_{\mathrm{A}ij}+(\boldsymbol{\beta}_{\mathrm{B}}^0)^T\boldsymbol{X}_{\mathrm{B}ij} \text{ such that the } \boldsymbol{U}_i \text{ are independent}\\ N(\mathbf{0},\,\boldsymbol{\Sigma}^0) \text{ random vectors.}\end{array} \tag{2}$$

The $\boldsymbol{U}_i$ are $d_{\mathrm{A}}\times 1$ unobserved random effects vectors. The $\boldsymbol{X}_{\mathrm{A}ij}$ are $d_{\mathrm{A}}\times 1$ random vectors corresponding to predictors that are partnered by a random effect. The $\boldsymbol{X}_{\mathrm{B}ij}$ are $d_{\mathrm{B}}\times 1$ random vectors corresponding to predictors that have a fixed effect only. Let $\boldsymbol{X}_{ij}\equiv(\boldsymbol{X}_{\mathrm{A}ij}^T,\boldsymbol{X}_{\mathrm{B}ij}^T)^T$ denote the combined predictor vectors. We assume that the $\boldsymbol{X}_{ij}$ and $\boldsymbol{U}_i$, for $1\le i\le m$ and $1\le j\le n_i$, are totally independent, with the $\boldsymbol{X}_{ij}$ each having the same distribution as the $(d_{\mathrm{A}}+d_{\mathrm{B}})\times 1$ random vector $\boldsymbol{X}=(\boldsymbol{X}_{\mathrm{A}}^T,\boldsymbol{X}_{\mathrm{B}}^T)^T$ and the $\boldsymbol{U}_i$ each having the same distribution as the random vector $\boldsymbol{U}$.

For any $\boldsymbol{\beta}_{\mathrm{A}}$ ($d_{\mathrm{A}}\times 1$), $\boldsymbol{\beta}_{\mathrm{B}}$ ($d_{\mathrm{B}}\times 1$), $\boldsymbol{\Sigma}$ ($d_{\mathrm{A}}\times d_{\mathrm{A}}$) that is symmetric and positive definite and $\phi>0$, conditional on the $\boldsymbol{X}_{ij}$ data, the log-likelihood is

$$\begin{aligned}\ell(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\phi)&=-\tfrac{m}{2}\log|2\pi\,\boldsymbol{\Sigma}|\\ &+\sum_{i=1}^m\sum_{j=1}^{n_i}\big[\{Y_{ij}(\boldsymbol{\beta}_{\mathrm{A}}^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij})+c(Y_{ij})\}/\phi-d(\phi)-e(Y_{ij})\big]\\ &+\sum_{i=1}^m\log\int_{\mathbb{R}^{d_{\mathrm{A}}}}\exp\Bigg[\sum_{j=1}^{n_i}\{Y_{ij}\boldsymbol{u}^T\boldsymbol{X}_{\mathrm{A}ij}-b\big((\boldsymbol{\beta}_{\mathrm{A}}+\boldsymbol{u})^T\boldsymbol{X}_{\mathrm{A}ij}+\boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}ij}\big)\}/\phi-\tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{u}\Bigg]d\boldsymbol{u}.\end{aligned} \tag{3}$$

The conditional maximum likelihood estimator of $(\boldsymbol{\beta}_{\mathrm{A}}^0,\boldsymbol{\beta}_{\mathrm{B}}^0,\boldsymbol{\Sigma}^0,\phi^0)$ is

$$(\widehat{\boldsymbol{\beta}}_{\mathrm{A}},\widehat{\boldsymbol{\beta}}_{\mathrm{B}},\widehat{\boldsymbol{\Sigma}},\widehat{\phi})=\underset{\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\phi}{\operatorname{argmax}}\ \ell(\boldsymbol{\beta}_{\mathrm{A}},\boldsymbol{\beta}_{\mathrm{B}},\boldsymbol{\Sigma},\phi).$$

Even though conditional maximum likelihood provides a natural estimator for ϕ based on data modeled according to (2), the relevant literature and software is such that alternative approaches are common. For the generalized linear model special case, Section 8.3.6 of McCullagh and Nelder (1989) expresses a preference for the natural moment-based estimator of ϕ, which is sometimes referred to as the Pearson estimator. The functions glm() for generalized linear models and glmer() within the R package lme4 (Bates et al., 2015) for generalized linear mixed models each use Pearson estimation of the dispersion parameter. In Section 2 of Cordeiro and McCullagh (1991) some alternative estimators for the dispersion parameters are proposed, motivated by bias correction and computational convenience considerations. Jo and Lee (2017) compare the efficiencies of dispersion parameter estimators for Gamma generalized linear models and recommend conditional maximum likelihood estimation compared with the Pearson and Cordeiro and McCullagh (1991) estimators.

3. Main result

Given the addition of ϕ to the set of parameters being estimated, compared with the Jiang et al. (2022) set-up, our aim is to extend the asymptotic normality result for this enlarged estimation problem. We start by repeating definitions and conditions from Section 3 of Jiang et al. (2022). Let

$$n\equiv\frac{1}{m}\sum_{i=1}^m n_i=\text{average of the within-group sample sizes},$$

$$\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U})\equiv E\left\{b''\Big((\boldsymbol{\beta}_{\mathrm{A}}^0+\boldsymbol{U})^T\boldsymbol{X}_{\mathrm{A}}+(\boldsymbol{\beta}_{\mathrm{B}}^0)^T\boldsymbol{X}_{\mathrm{B}}\Big)\begin{bmatrix}\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{B}}^T\\ \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{A}}^T & \boldsymbol{X}_{\mathrm{B}}\boldsymbol{X}_{\mathrm{B}}^T\end{bmatrix}\middle|\,\boldsymbol{U}\right\},$$

$$\boldsymbol{\Lambda}_{\boldsymbol{\beta}_{\mathrm{B}}} \equiv \left(E\left[\left\{\text{lower right } d_{\mathrm{B}} \times d_{\mathrm{B}} \text{ block of } \boldsymbol{\Omega}_{\boldsymbol{\beta}_{\mathrm{B}}}(\boldsymbol{U})^{-1}\right\}^{-1}\right]\right)^{-1}$$

and $\|\boldsymbol{v}\| \equiv (\boldsymbol{v}^T\boldsymbol{v})^{1/2}$ denote the Euclidean norm of a column vector $\boldsymbol{v}$. For a symmetric matrix $\boldsymbol{M}$ let $\lambda_{\min}(\boldsymbol{M})$ denote the smallest eigenvalue of $\boldsymbol{M}$. Also, let $\boldsymbol{D}_d$ denote the matrix of zeroes and ones such that $\boldsymbol{D}_d\text{vech}(\boldsymbol{A}) = \text{vec}(\boldsymbol{A})$ for all $d \times d$ symmetric matrices $\boldsymbol{A}$. The Moore–Penrose inverse of $\boldsymbol{D}_d$ is $\boldsymbol{D}_d^+ = (\boldsymbol{D}_d^T\boldsymbol{D}_d)^{-1}\boldsymbol{D}_d^T$. Let d' and d'' denote the first and second derivatives of the d function.

The theorem relies on the following assumptions:

(A1) The number of groups m diverges to ∞.

(A2) The within-group sample sizes n_i diverge to ∞ in such a way that $n_i/n \to C_i$ for constants $0 < C_i < \infty$, $1 \le i \le m$. Also, $n/m \to 0$ as m and n diverge.

(A3) The distribution of $\boldsymbol{X}$ is such that

$$E\left[\frac{E\left[\max\left(1, \|\boldsymbol{X}\|\right)^8 \max\left\{1, b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U})^T\boldsymbol{X}_{\mathrm{A}} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}}\right)\right\}^4 \middle| \boldsymbol{U}\right]}{\min\left\{1, \lambda_{\min}\left(E\{\boldsymbol{X}_{\mathrm{A}}\boldsymbol{X}_{\mathrm{A}}^T b''\left((\boldsymbol{\beta}_{\mathrm{A}} + \boldsymbol{U})^T\boldsymbol{X}_{\mathrm{A}} + \boldsymbol{\beta}_{\mathrm{B}}^T\boldsymbol{X}_{\mathrm{B}}\right)|\boldsymbol{U}\}\right)\right\}^2}\right] < \infty \tag{4}$$

for all $\boldsymbol{\beta}_{\mathrm{A}} \in \mathbb{R}^{d_{\mathrm{A}}}$, $\boldsymbol{\beta}_{\mathrm{B}} \in \mathbb{R}^{d_{\mathrm{B}}}$ and $\boldsymbol{\Sigma}$ a $d_{\mathrm{A}} \times d_{\mathrm{A}}$ symmetric and positive definite matrix.

Theorem 1. *Assume that conditions (A1)–(A3) hold. Then*

$$\sqrt{m}\begin{bmatrix} \widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A^0 \\ \sqrt{n}\left(\widehat{\boldsymbol{\beta}}_B - \boldsymbol{\beta}_B^0\right) \\ \text{vech}(\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}^0) \\ \sqrt{n}(\widehat{\phi} - \phi^0) \end{bmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \\ \boldsymbol{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}^0 & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \phi^0\boldsymbol{\Lambda}_{\boldsymbol{\beta}_B} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & 2\boldsymbol{D}_{d_A}^+(\boldsymbol{\Sigma}^0 \otimes \boldsymbol{\Sigma}^0)\boldsymbol{D}_{d_A}^{+T} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} & \dfrac{1}{2d'(\phi^0)/\phi^0 + d''(\phi^0)} \end{bmatrix}\right).$$

A proof of Theorem 1 is given in Section S.2 of the supplemental material.

Some remarks concerning Theorem 1 are:

1. The diagonal blocks appearing in the Multivariate Normal covariance matrix of Theorem 1 correspond to asymptotic covariances of: (1) fixed effects partnered by a random effect, (2) fixed effects not partnered by a random effect, (3) random effects covariance parameters and (4) the dispersion parameter. For the first and third of these types of parameters the asymptotic variances have order m^{-1}. The remaining parameters, including the dispersion parameter, have order $(mn)^{-1}$ asymptotic variances.
2. Theorem 1 reveals asymptotic orthogonality between ϕ and $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma})$. This is in addition to the asymptotic orthogonality between the components of $(\boldsymbol{\beta}_{\mathrm{A}}, \boldsymbol{\beta}_{\mathrm{B}}, \boldsymbol{\Sigma})$, established by Theorem 1 (Jiang et al., 2022).
3. The asymptotic distribution of $\widehat{\phi}$ is the same as that arising for the generalized linear model special case of (2) when there are no random effects. In other words, the asymptotic behavior of $\widehat{\phi}$ is not impacted by the extension from generalized linear models to generalized linear mixed models.
4. After obtaining the $2d'(\phi)/\phi+d''(\phi)$ expressions for the specific d functions of Table 1 and simplifying, the asymptotic variances of $\widehat{\phi}$ become:

$$\text{Asy.Var}(\widehat{\phi}) = \begin{cases} \dfrac{2(\phi^0)^2}{mn} & \text{for the Gaussian and the Inverse Gaussian families,} \\[2ex] \dfrac{(\phi^0)^4}{\{\text{trigamma}(1/\phi^0) - \phi^0\}\, mn} & \text{for the Gamma family.} \end{cases} \tag{5}$$

5. The trigamma function has the following asymptotic expansion:

$$\text{trigamma}(x) = \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \cdots .$$

where the coefficients are simple functions of Bernoulli numbers. It follows that

$$\frac{(\phi^0)^4}{\text{trigamma}(1/\phi^0) - \phi^0} \approx 2(\phi^0)^2 \quad \text{for small values of } \phi^0.$$

This connects the asymptotic variance results of (5) for low values of the dispersion parameter.

6. The moment condition in (A3) is sufficient but not necessary for Theorem 1 to hold. A recently discovered erratum has resulted in the replacement of the second moment in the second numerator factor in (4) by the fourth moment, compared with (A3) of Jiang et al. (2022). For various special cases of (1) and (2) mathematical analysis can be used

to replace (A3) by a simpler moment condition. The Gaussian response case is such that $b'' = 1$ and relatively simple arguments show that (A3) can be replaced by

$$E(\|\boldsymbol{X}\|^8) < \infty \quad \text{and} \quad \text{no entry of } \boldsymbol{X}_{\rm A} \text{ is the zero degenerate random variable.}$$

In this case it is clear that (A3) holds for sufficiently light-tailed $\boldsymbol{X}$ distributions but fails if, for example, an entry of $\boldsymbol{X}_{\rm B}$ has a heavy-tailed distribution such as the t distribution with a low degrees of freedom parameter. A more involved example involves Poisson responses with $\boldsymbol{X}_{\rm A} = 1$, corresponding to random intercepts. In Section S.3 of the supplemental material it is shown that (A3) can be replaced by the moment generating function existence condition

$$E\{\exp(\boldsymbol{t}^T\boldsymbol{X}_{\rm B})\} < \infty \text{ for all } \boldsymbol{t} \in \mathbb{R}^{d_{\rm B}}$$

which is satisfied by, for example, $\boldsymbol{X}_{\rm B}$ having a Multivariate Skew-Normal distribution (Azzalini and Dalla Valle, 1996). The arguments in Section S.3 of the supplemental material give a flavor of what is involved to simplify (A3). Other cases require more detailed mathematical analysis.

4. Asymptotically valid inference

An immediate consequence of Theorem 1 is

$$\sqrt{mn}\left\{2d'(\phi^0)/\phi^0 + d''(\phi^0)\right\}^{1/2}\left(\widehat{\phi} - \phi^0\right) \xrightarrow{\mathcal{D}} N(0, 1).$$

This asymptotic normality result still holds when the unknown quantities on the left-hand side are replaced by consistent estimators, often referred to as *studentization*. Hence an asymptotically valid $100(1-\alpha)\%$ confidence interval for ϕ^0 is

$$\widehat{\phi} \pm \Phi^{-1}(1 - \tfrac{1}{2}\alpha)\left[\left\{2d'(\widehat{\phi})/\widehat{\phi} + d''(\widehat{\phi})\right\}mn\right]^{-1/2} \tag{6}$$

where Φ is the $N(0, 1)$ cumulative distribution function. It follows that Theorem 1 provides simple closed form asymptotically valid inference for ϕ^0. For the specific families listed in Table 1, (6) simplifies to

$$\widehat{\phi} \pm \Phi^{-1}(1 - \tfrac{1}{2}\alpha)\sqrt{\frac{2\widehat{\phi}^2}{mn}} \quad \text{for the Gaussian and Inverse Gaussian families}$$

and

$$\widehat{\phi} \pm \Phi^{-1}(1 - \tfrac{1}{2}\alpha)\sqrt{\frac{\widehat{\phi}^4}{\{\text{trigamma}(1/\widehat{\phi}) - \widehat{\phi}\}mn}} \quad \text{for the Gamma family.} \tag{7}$$

A simulation study was run to assess the actual coverages of Theorem 1-based confidence intervals for the dispersion parameter for the $d_{\rm A} = 1$ and $d_{\rm B} = 3$ Gamma mixed model. In this scalar random effects case, $\boldsymbol{\beta}_{\rm A}$ and $\boldsymbol{\Sigma}$ are replaced by the scalar parameter symbols β_0 and σ^2. The true parameter vector $(\beta_0^0, \boldsymbol{\beta}_{\rm B}^0, (\sigma^2)^0, \phi^0)$ had the following four settings:

setting A:	$(-2.78, -1.55, 0, 0.98, 0.25, 0.54)$,
setting B:	$(-4.06, -2.41, 0.16, -3.93, 0.52, 1.92)$,
setting C:	$(-8.55, 3.13, -7.82, -0.23, 1.27, 0.86)$,
setting D:	$(-14.45, 8.78, 0.41, -3.32, 1.88, 2.11)$

and the $\boldsymbol{X}_{{\rm B}ij}$s were generated independently from the uniform distribution over the unit cube. The number of groups m varied over the set $\{50, 100, 150, 200, 250, 300, 350, 400\}$ and the sample size within each group was n fixed at $m/5$. For each of the possible combinations of the true parameter vector and the sample size pair we simulated 1000 replications. Established generalized linear mixed model software, such the `glmer()` function within the `R` package lme4 (Bates et al., 2015), does not use maximum likelihood to estimate ϕ. For this $d_{\rm A} = 1$ case the likelihood surface (3) was relatively straightforward to evaluate exactly using univariate quadrature and maximize using a derivative-free optimization algorithm such as the Nelder–Mead simplex method (Nelder and Mead, 1965) with `glmer()` starting values. After obtaining the maximum likelihood estimates, we computed 95% confidence intervals based on (7) with $\alpha = 0.05$.

Fig. 2 summarizes the empirical coverage results. The shaded regions around the line segments in Fig. 2 indicate plus and minus two standard errors for the sample proportions. It is seen that the actual coverage percentages are in keeping with the advertized level of 95% across all settings and sample sizes. There is a tendency for the empirical coverages to get closer, on average, to 95% as m and n in increase, as expected for statistical inference based on leading term asymptotics.

5. Analysis of the mathematics achievement data

As an illustration of the asymptotically valid inference results given in the previous section we conducted an analysis of the mathematics achievement data shown in Fig. 1 using a Gaussian response version of (2). Specifically, we considered

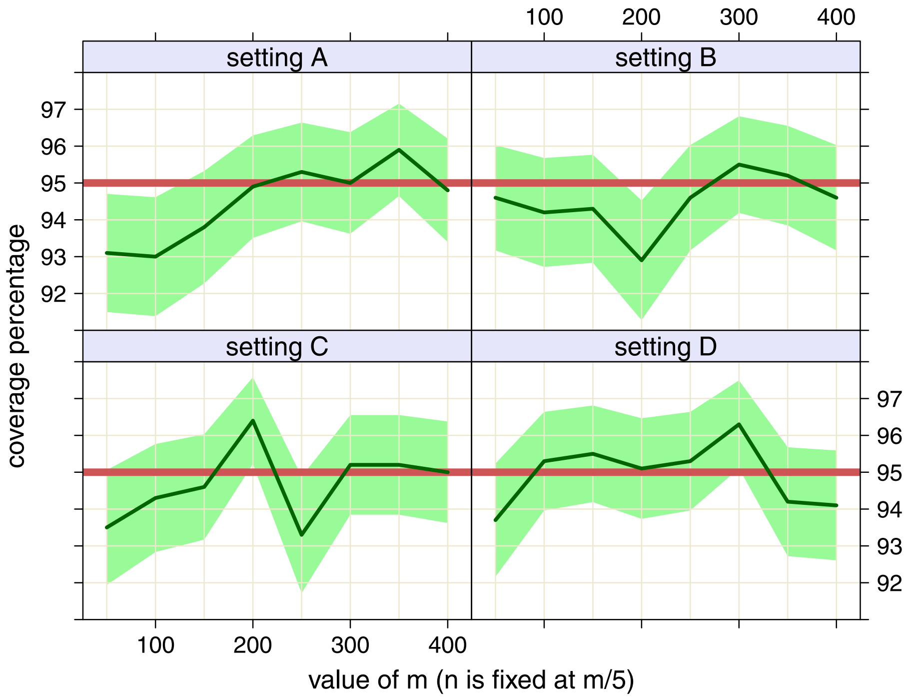

Fig. 2. Actual coverage percentage of nominally 95% confidence intervals for ϕ^0 in $d_A = 1$ and $d_B = 3$ Gamma mixed model with four different true parameter vectors. The nominal percentage is shown as a thick horizontal line. The percentages are based on 1000 replications. The values of m are 50, 100, 150, 200, 250, 300, 350, 400. The value of n is fixed at $m/5$. The shaded regions correspond to plus and minus two standard errors of the sample proportions.

the model

$$\begin{aligned}&\texttt{MathAchieve}_{ij}|U_{0i},U_{1i} \overset{\text{ind.}}{\sim} N\Big(\beta_0^0 + U_{0i} + (\beta_1^0 + U_{1i})\,\texttt{SES}_{ij} + \beta_2^0\,\texttt{isMale}_{ij}\\ &\qquad\qquad + \beta_3^0\,\texttt{isMinority}_{ij}, \phi^0\Big),\\ &\begin{bmatrix} U_{0i}\\ U_{1i}\end{bmatrix} \overset{\text{ind.}}{\sim} N\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}(\sigma_0^0)^2 & \rho\sigma_0^0\sigma_1^0\\ \rho\sigma_0^0\sigma_1^0 & (\sigma_1^0)^2\end{bmatrix}\right), \quad 1\le i\le m, \quad 1\le j\le n_i,\end{aligned} \tag{8}$$

where $\overset{\text{ind.}}{\sim}$ stands for "independently distributed as" and n_i is the number of students in the ith school. In (8)$\texttt{MathAchieve}_{ij}$ and $\texttt{SES}_{ij}$ denote, respectively, the mathematics achievement and socio-economic status scores for the jth student in the ith school. In addition

$$\texttt{isMale}_{ij} = \begin{cases} 0 & \text{if the } j\text{th student in the } i\text{th school is female,}\\ 1 & \text{if the } j\text{th student in the } i\text{th school is male}\end{cases}$$

and $\texttt{isMinority}_{ij}$ is defined similarly for minority status. The n_i values range over 14 to 67.

Table 2 shows the estimates of the fixed effects and standard deviation parameters based on conditional maximum likelihood estimation as described in Section 2. Also shown in Table 2 are approximate 95% confidence intervals based on Theorem 1 and studentization. The confidence interval for β_2^0 indicates a statistically significant elevation in mean mathematics achievement score for the male student population. The confidence interval for β_3^0 indicates a statistically significant lowering for minority students. The confidence intervals for σ_0^0 and σ_1^0 are both within the positive half-line, which indicates significant heterogeneities in the intercepts and slopes of the social economic status effects. The last row of Table 2 provides an estimate and 95% confidence interval for the within-school error standard deviation.

An examination of the residuals revealed reasonable accordance with model assumptions but some heteroscedasticity. More delicate modeling, involving variance function extensions of (2) and other response families may lead to model fit improvements.

Table 2
Maximum likelihood estimates and approximate 95% confidence intervals of the fixed effects parameters and standard deviation parameters for the fit of the model (8) to the mathematics achievement data.

Parameter	Estimate	95% confid. interv.
β_0^0	12.93	(12.55, 13.31)
β_1^0	2.097	(1.875, 2.319)
β_2^0	1.219	(0.9003, 1.537)
β_3^0	−2.999	(−3.404, −2.594)
σ_0^0	1.903	(1.682, 2.101)
σ_1^0	0.4964	(0.4386, 0.5481)
$\sqrt{\phi^0}$	5.982	(5.883, 6.079)

6. General two-parameter exponential family extension

We now turn attention to general two-parameter exponential families. If the restriction to reproductive exponential families is removed then (1) should be replaced by

$$p(y;\eta,\phi)=\exp\Big[\{y\eta-b(\eta)+c(y)\}/\phi-\widetilde{d}(y,\phi)\Big]h(y) \tag{9}$$

where $\widetilde{d}(y,\phi)$ is some bivariate function of y and ϕ that is not necessarily additive in its arguments. Section 2.1 of Jørgensen (1987) describes a procedure for generating versions of (9) from any distribution possessing a moment generating function. An example of a non-reproductive version of (9) given there is such that

$$b(\eta)=\log\Big(-\eta-\sqrt{\eta^2-1}\Big),\quad c(y)=0,\quad \widetilde{d}(y,\phi)=-\log\Big\{I_{1/\phi}(y/\phi)\big/(y\phi)\Big\}\ \ \text{and}\ \ h(y)=I(y>0)$$

where I_ν denotes the modified Bessel function of the first kind with index ν (e.g. Section 8.431 of Gradshteyn and Ryzhik, 1994).

For this more general set-up, Theorem 1 arguments still apply with $\widetilde{d}(y,\phi)$ replacing $d(\phi)+e(y)$ and the lower right block of the covariance matrix on the right-hand side of Theorem 1 statement generalizes to

$$(\phi^0)^4\Big/E\left(\left[\frac{\partial^2}{\partial\psi^2}\widetilde{d}\left(Y,\frac{1}{\psi}\right)\right]_{\psi=\frac{1}{\phi^0}}\right) \tag{10}$$

where Y is a random variable having the same distribution as the Y_{ij}s. Method of moments estimation of the expectation over Y leads to the following extension of (6) for approximate $100(1-\alpha)$% confidence intervals for ϕ^0:

$$\widehat{\phi}\pm\Phi^{-1}(1-\tfrac{1}{2}\alpha)\widehat{\phi}^2\left(\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\frac{\partial^2\widetilde{d}\left(Y_{ij},1/\psi\right)}{\partial\psi^2}\right]_{\psi\,=\,1/\widehat{\phi}}\right)^{-1/2}. \tag{11}$$

The practical relevance of (10) and (11) is much lower than for the reproductive exponential family special case and they have been included for completeness.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to Gordon Smyth for advice related to this research. This research was partially supported by Australian Research Council grant DP180100597.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spl.2022.109691.

References

Azzalini, A., Dalla Valle, A., 1996. The multivariate skew-normal distribution. Biometrika 83, 715–726.

Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48.

Blæsild, P., Jensen, J.L., 1985. Saddlepoint formulas for reproductive exponential models. Scand. J. Stat. 12, 193–202.

Coleman, J.S., Hoffer, T., Kilgore, S.B., 1982. High School Achievements: Public, Catholic and Other Private Schools Compared. Basic, New York.

Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. J. R. Stat. Soc. Ser. B Stat. Methodol. 53, 629–643.

Faraway, J.J., 2016. Extending the Linear Model with R, second ed. CRC Press, Boca Raton, Florida.

Gradshteyn, I.S., Ryzhik, I.M., 1994. Tables of Integrals, Series, and Products, fifth ed. Academic Press, San Diego, California.

Jiang, J., Nguyen, T., 2021. Linear and Generalized Linear Mixed Models and Their Applications, second ed. Springer, New York.

Jiang, J., Wand, M.P., Bhaskaran, A., 2022. Usable and precise asymptotics for generalized linear mixed model analysis and design. J. R. Stat. Soc. Ser. B Stat. Methodol. 84, 55–82.

Jo, S., Lee, W., 2017. Comparing the efficiency of dispersion parameter estimators in gamma generalized linear models. Korean J. Appl. Stat. 30, 95–102.

Jørgensen, B., 1987. Exponential dispersion models (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 49, 127–162.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, second ed. Chapman and Hall, London.

McCulloch, C.E., Searle, S.R., Neuhaus, J.M., 2008. Generalized, Linear, and Mixed Models, second ed. John Wiley & Sons, New York.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7, 308–313.

Pinheiro, J., Bates, D., R Core Team, 2022. Nlme: linear and nonlinear mixed effects models. R package version 3.1-157. https://CRAN.R-project.org/package=nlme.

R Core Team, 2022. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. https://www.R-project.org.

Stroup, W.W., 2013. Generalized Linear Mixed Models. CRC Press, Boca Raton, Florida.