Fisher information under Gaussian quadrature models

Antonio Hermes Marques da Silva Júnior*[1,2], Jochen Einbeck†[2] and Peter S. Craig‡[2]

[1]Departamento de Estatística, Universidade Federal do Rio Grande do Norte, Natal, Brazil
[2]Department of Mathematical Sciences, Durham University, Science Laboratories, South Road, DH1 3LE Durham City, UK

Abstract

This paper develops formulae to compute the Fisher information matrix for the regression parameters of generalised linear models with Gaussian random effects. The Fisher information matrix relies on the estimation of the response variance under the model assumptions. We propose two approaches to estimate the response variance: the first is based on an analytic formula (or a Taylor expansion for cases where we cannot obtain the closed-form) and the second is an empirical approximation using the model estimates via the EM process. Further, simulations under several response distributions and a real data application involving a factorial experiment are presented and discussed. In terms of standard errors and coverage probabilities for model parameters, the proposed methods turn out to behave more reliably than the 'disparity rule' or direct extraction of results from the generalised linear model fitted in the last EM iteration.

Keywords: Gaussian random effects, Generalised linear models, overdispersion, standard errors.

*a.h.marques-da-silva-jun@durham.ac.uk
†jochen.einbeck@durham.ac.uk
‡p.s.craig@durham.ac.uk

1 Introduction

Consider a generalised linear model with Gaussian random effects (GLMwRE) for a data set containing n independent observations of a response variable, given by $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, and corresponding observations on p explanatory variables, given by $\boldsymbol{x}_i^\top = (x_{i1}, \ldots, x_{ip})^\top$, for $i = 1, \ldots, n$. The linear predictor for the i-th observation, η_i, has the form

$$\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i, \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the vector of regression parameters, $z_i \sim \mathcal{N}(0, 1)$ is an unobserved random effect, and $\sigma > 0$. The relationship between y_i and η_i is given by the conditional mean $\mu_i = \mathrm{E}[y_i|z_i]$ and the monotonic and differentiable *link function*, $g(\,\cdot\,)$ such that $\mu_i = g^{-1}(\eta_i)$.

By definition, $\boldsymbol{y}$ is a vector of independent random variables and each y_i, $i = 1, \ldots, n$ has a distribution in an exponential family with dispersion parameter. Thus, the probability density function of y_i can be written as

$$f(y_i|\theta_i, \tau) = \exp[\tau^{-1}\{y_i\theta_i - b(\theta_i)\} + c(y_i, \tau)], \tag{2}$$

where $\theta_1, \ldots, \theta_n$ are unknown parameters, $\tau > 0$ is a *dispersion* parameter common to all observations, and $b(\,\cdot\,)$ and $c(\,\cdot\,,\,\cdot\,)$ are known functions. The parameter estimation procedure requires the probability density function in (2) to be differentiable with respect to θ_i and τ.

In (2), θ_i is related to μ_i, and consequently to η_i, through two useful properties of an exponential dispersion family:

$$\mathrm{E}[y_i|z_i] = b'(\theta_i) \quad \text{and} \quad \mathrm{Var}[y_i|z_i] = \tau V_i = \tau b''(\theta_i), \tag{3}$$

where $V_i = V(\mu_i)$ and $V(\mu_i) = d\mu_i/d\theta_i = b''(\theta_i)$. The function $V(\mu_i)$ is called the *variance function*.

In the special case where $\sigma = 0$, the GLMwRE reduces to an ordinary generalised linear model (GLM). If the data are subject to unobserved heterogeneity, for instance due to the absence of predictor variables which would be important for explaining variation in the response, then the mean-variance relationship (3) becomes incorrect for a GLM. This problem is known as *overdispersion*, and leads, among other issues, to incorrect standard errors for model parameters. The random intercept term in the random effects model (1) models this excess variation.

In matrix notation,

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \sigma\boldsymbol{z} \tag{4}$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\top$, $g(\boldsymbol{\mu}) = (g(\mu_1), \ldots, g(\mu_n))^\top$, $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^\top$, $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^\top$ and $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$. The likelihood function for the GLMwRE is

$$L^*(\boldsymbol{\beta}, \sigma, \tau) = \prod_{i=1}^{n} \int f(y_i|\boldsymbol{\beta}, \sigma, \tau, z_i)\phi(z_i)dz_i \tag{5}$$

where $\phi(\,\cdot\,)$ is the standard normal density and $f(\,\cdot\,)$ is the response density. However, the integral in (5) usually has no analytic solution and, therefore, is approximated using the respectively Gaussian quadrature weights π_k and the quadrature points $\tilde{z}_k$, $k = 1, \ldots, K$, see e.g. Abramowitz and Stegun [1972]. Hence, the likelihood (5) can be approximated by

$$L(\boldsymbol{\beta}, \sigma, \tau) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f(y_i|\boldsymbol{\beta}, \sigma, \tau, \tilde{z}_k) \equiv \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k f_{ik}, \tag{6}$$

which is the likelihood for a per-observation K-component mixture of response distributions. According to Laird [1978], the approximation (6) becomes accurate already for a small integer K. Thus, in the subsequent theoretical development, we shall assume that this mixture model is in fact the true model so that $L(\boldsymbol{\beta}, \sigma, \tau)$ is the true likelihood. Gaussian quadrature provides a convenient solution to the problem of solving the integral (5) due to its computational simplicity and its straightforward extendibility to nonparametric random effect distributions [Aitkin et al., 2009].

It is noted at this occasion that the 'Adaptive Gaussian Quadrature' [Pinheiro and Bates, 1995] used in R function `glmer` [Bates et al., 2014], which reduces to the Laplace approximation when only a single support point is used, is different to the methodology being used here, which follows the approach laid out in Aitkin [1996] and subsequent publications. Further discussion of this matter is provided in the Concluding remarks in Section 5.

The linear predictor for the k-th component of the i-th observation is

$$g(\mu_{ik}) = \eta_{ik} = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma\tilde{z}_k, \tag{7}$$

where $\mu_{ik} = \mathrm{E}[y_i|z_i = \tilde{z}_k]$. We now introduce nK pseudo-observations

$$\ddot{\boldsymbol{y}} = (\underbrace{\boldsymbol{y}^\top, \boldsymbol{y}^\top, \ldots, \boldsymbol{y}^\top}_{K \text{ times}})^\top$$

so that (7) can be formulated in stacked form as

$$g(\ddot{\boldsymbol{\mu}}) = \ddot{\boldsymbol{\eta}} = \ddot{\mathbf{X}}\boldsymbol{\beta} + \sigma\ddot{\boldsymbol{z}} = \ddot{\mathbf{Z}}\boldsymbol{\gamma} \tag{8}$$

where

$$\begin{aligned}
\ddot{\boldsymbol{\mu}} &= (\mu_{11}, \ldots, \mu_{n1}, \ldots, \mu_{1K}, \ldots, \mu_{nK})^\top, \\
\ddot{\boldsymbol{\eta}} &= (\eta_{11}, \ldots, \eta_{n1}, \ldots, \eta_{1K}, \ldots, \eta_{nK})^\top, \\
\ddot{\mathbf{X}} &= (\underbrace{\mathbf{X}^\top, \mathbf{X}^\top, \ldots, \mathbf{X}^\top}_{K \text{ times}})^\top, \\
\ddot{\boldsymbol{z}} &= (\underbrace{\tilde{z}_1, \tilde{z}_1, \ldots, \tilde{z}_1}_{n \text{ times}}, \ldots, \underbrace{\tilde{z}_K, \tilde{z}_K, \ldots, \tilde{z}_K}_{n \text{ times}})^\top, \\
\ddot{\mathbf{Z}} &= \begin{pmatrix} \ddot{\mathbf{X}} & \ddot{\boldsymbol{z}} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma \end{pmatrix}.
\end{aligned}$$

The log–likelihood function for the GLMwRE is

$$\ell = \log L = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k f_{ik}\right), \tag{9}$$

and it turns out that equating the first partial derivatives to zero, that is $\partial\ell/\partial\boldsymbol{\gamma} = 0$, one obtains precisely the single–distribution score equations [Aitkin et al., 2009] for the GLM, but summed over $k = 1, \ldots, K$ and weighted by

$$\omega_{ik} = \frac{\pi_k f_{ik}}{\sum_{l=1}^{K} \pi_l f_{il}}. \tag{10}$$

Each ω_{ik} can be interpreted as the *posterior probability* that observation y_i came from component k. Alternating between this estimation step and an update step for the w_{ik} leads to an EM algorithm:

E-step Calculate weights ω_{ik} according to (10);

M-step Update the parameter estimates by fitting the GLM (8) with weights ω_{ik}.

Inference for the dispersion parameter τ, which we consider as a nuisance parameter, is not of primary interest in this paper. One can estimate τ in any EM iteration through

$$\hat{\tau} = \frac{1}{n} \sum_i \sum_k w_{ik} \frac{(y_i - \hat{\mu}_{ik})^2}{V(\hat{\mu}_{ik})},$$

using the current component mean estimates $\hat{\mu}_{ik} = g^{-1}(\boldsymbol{x}_i^\top \hat{\boldsymbol{\beta}} + \hat{\sigma}\tilde{z}_k)$ and weights w_{ik}. The estimate $\hat{\tau}$ can be used at all occasions where τ appears henceforth in this manuscript. See for instance Einbeck and Hinde [2006] for details.

For practical applications, it is very important to have reliable inferential tools for the regression parameters, $\boldsymbol{\beta}$. This is relevant, for instance, for the construction of confidence intervals or the assessment of strength of effects through hypothesis testing. Such inferences rely on the standard errors of the parameter estimates, $\hat{\boldsymbol{\beta}}$, which, in turn, can be computed via the Fisher information matrix. Therefore, the ability to compute this matrix accurately is paramount for most subsequent inferential procedures.

In this article, we describe a way to obtain the Fisher information matrix for the GLMwRE approximated by Gaussian quadrature. It is not the first attempt at doing so; Friedl and Kauermann [2000] tailored the method by Oakes [1999] to the specific structure of a random effect model. Their approximation makes use of the final estimates from the EM process but relies on a Monte-Carlo sandwich correction. We compute analytic forms of the Fisher information matrix where possible, and give approximate expressions in those cases where analytic forms are not available.

The article unfolds as follows: Section 2 outlines the theory underlying the score vector and the Fisher information matrix. The Section 3 presents the response variance estimators. Section 4 provides applications, where we compare the standard errors computed using our formulae against Monte Carlo, a heuristic approximation known as the ‘disparity rule’, and the result of a Laplace approximation. Finally, Section 5 contains concluding remarks and directions for future research and implementation.

2 The score vector and the Fisher information matrix

Writing the log–likelihood (9) as $\ell = \log L(\boldsymbol{\gamma})$, the total score vector for $\boldsymbol{\gamma}$, $\mathbf{U} = \mathbf{U}(\boldsymbol{\gamma})$, is

$$\mathbf{U} = \frac{\partial \ell}{\partial \boldsymbol{\gamma}} = \sum_{i=1}^{n} \frac{1}{\sum_{l=1}^{K} \pi_l f_{il}} \sum_{k=1}^{K} \pi_k \frac{\partial f_{ik}}{\partial \boldsymbol{\gamma}}.$$

By logarithmic differentiation, we find

$$\mathbf{U} = \sum_{i=1}^{n} \frac{\displaystyle\sum_{k=1}^{K} \pi_k f_{ik} \frac{\partial \log f_{ik}}{\partial \boldsymbol{\gamma}}}{\displaystyle\sum_{l=1}^{K} \pi_l f_{il}} = \tau^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \omega_{ik} \left\{ \frac{d\mu_{ik}}{d\eta_{ik}} \frac{(y_i - \mu_{ik})}{V_{ik}} \begin{pmatrix} \boldsymbol{x}_i \\ \tilde{z}_k \end{pmatrix} \right\}, \quad (11)$$

where ω_{ik} is given by (10) and $V_{ik} = V(\mu_{ik})$. In matrix notation, we can rewrite $\mathbf{U}$ as

$$\mathbf{U} = \ddot{\mathbf{Z}}^\top \mathbf{D}(\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\mu}}), \quad (12)$$

where $\mathbf{D}$ is the diagonal matrix with diagonal entries $d_{11}, \ldots, d_{n1}, \ldots, d_{1K}, \ldots, d_{nK}$ given by

$$d_{ik} = \tau^{-1} \frac{d\mu_{ik}}{d\eta_{ik}} \frac{\omega_{ik}}{V_{ik}}.$$

Similarly, denote by $\mathbf{K} = \mathbf{K}(\boldsymbol{\gamma})$ the GLMwRE Fisher information matrix for $\boldsymbol{\gamma}$. Then $\mathbf{K} = \mathrm{Var}[\mathbf{U}]$ and, from (12), we have

$$\begin{aligned} \mathbf{K} &= \mathrm{Var}\left[\ddot{\mathbf{Z}}^\top \mathbf{D}(\ddot{\boldsymbol{y}} - \ddot{\boldsymbol{\mu}})\right] \\ &= \ddot{\mathbf{Z}}^\top \mathbf{D} \mathrm{Var}[\ddot{\boldsymbol{y}}] \mathbf{D} \ddot{\mathbf{Z}} \\ &= \ddot{\mathbf{Z}}^\top \mathbf{D} \ddot{\boldsymbol{\Upsilon}} \mathbf{D} \ddot{\mathbf{Z}}, \end{aligned}$$

where $\ddot{\boldsymbol{\Upsilon}} = \mathrm{Var}[\ddot{\boldsymbol{y}}]$ is the unconditional variance-covariance matrix for $\ddot{\boldsymbol{y}}$. Since the observations in the GLMwRE are independent,

$$\begin{aligned} \mathrm{Cov}(y_i, y_i) &= \mathrm{Var}(y_i), \quad \forall i \in \{1, \ldots n\}, \text{ and} \\ \mathrm{Cov}(y_i, y_j) &= 0, \quad \forall i \neq j \in \{1, \ldots n\}, \end{aligned}$$

one finds for the K copies in $\ddot{\boldsymbol{y}}$ that

$$\begin{aligned} \mathrm{Cov}(y_i^{(k)}, y_i^{(l)}) &= \mathrm{Var}(y_i), \quad \forall i \in \{1, \ldots n\}, k, m \in \{1, \ldots K\}, \text{ and} \\ \mathrm{Cov}(y_i^{(k)}, y_j^{(l)}) &= 0 \quad \forall i \neq j \in \{1, \ldots n\}, k, m \in \{1, \ldots K\}. \end{aligned}$$

Therefore,

$$\ddot{\boldsymbol{\Upsilon}} = \underbrace{\begin{pmatrix} \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \\ \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon} & \boldsymbol{\Upsilon} & \cdots & \boldsymbol{\Upsilon} \end{pmatrix}}_{K \text{ times}} \left.\vphantom{\begin{pmatrix} a \\ a \\ a \\ a \end{pmatrix}}\right\} K \text{ times},$$

where $\boldsymbol{\Upsilon} = \text{diag}(\upsilon_i)$ and $\upsilon_i = \text{Var}(y_i)$. Here, the response variances play an important role and the following Section 3 develops the necessary formulae.

3 Response variance

Recall that, in model (1), the z_i follow a standard normal distribution. That is, though they are approximated by a discrete set $\tilde{z}_1, \ldots \tilde{z}_K$ for estimation purposes, they are random in nature, so that the unconditional mean and variance of y_i are

$$\text{E}[y_i] = \text{E}[\text{E}[y_i|z_i]] = \text{E}[\mu_i] \tag{13}$$

and

$$\begin{aligned}\text{Var}(y_i) &= \text{E}[\text{Var}[y_i|z_i]] + \text{Var}[\text{E}[y_i|z_i]] \\ &= \tau\text{E}[V(\mu_i)] + \text{Var}[\mu_i]. \end{aligned} \tag{14}$$

The remaining task is to determine $\text{E}[V(\mu_i)]$ and $\text{Var}[\mu_i]$. This can be achieved either approximately, by use of the Gaussian quadrature rule, or analytically, based on explicit expressions depending on the response distribution and link function. We explain both approaches below.

3.1 Estimation via analytic expressions

We derived the analytic form of $\text{E}[V(\mu_i)]$ and $\text{Var}[\mu_i]$ for Normal, Gamma, Poisson, Binomial and Inverse Gaussian response distribution and a wide range of commonly used link functions. The resulting expressions for $\text{Var}(y_i)$ are summarized in Table 1. Some combinations of distribution and link function required the use of a Taylor expansion, which is indicated by a $\circledast$. All such expansions were made to third order. Of course, for practical use, $\boldsymbol{\beta}$ and σ need to be replaced by their corresponding estimates.

We do not give all derivations in detail, but restrict ourselves to two exemplary derivations. Firstly, suppose a GLMwRE with normal response and identity link function.

$$\mu_i = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i$$

and $V(\mu) = 1$. Therefore, $\mathrm{E}[y_i] = \boldsymbol{x}_i^\top \boldsymbol{\beta}$ and the response variance is

$$\begin{aligned} \mathrm{Var}(y_i) &= \tau \mathrm{E}[1] + \mathrm{Var}[\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i] \\ &= \tau + \sigma^2. \end{aligned}$$

However, there are cases in which there is no analytical solution for $\mathrm{E}[V(\mu_i)]$ and $\mathrm{Var}[\mu_i]$. In such cases, an approximate solution can be obtained by expanding $V(\mu_i)$ and μ_i by Taylor series around $z_i = 0$. Therefore, secondly, consider a GLMwRE with Gamma response and inverse link. For this configuration, $V(\mu) = \mu^2$ and

$$\mu_i = \frac{1}{\eta_i} = \frac{1}{\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i}.$$

Thus,

$$\begin{aligned} \mathrm{E}[V(\mu_i)] &= \mathrm{E}\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-2}\right], \quad \text{and} \\ \mathrm{Var}[\mu_i] &= \mathrm{Var}[(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-1}]. \end{aligned}$$

By Taylor expansion around 0, we have

$$(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-1} \approx (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-1} - (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} \sigma z_i + (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-3} \sigma^2 z_i^2 - (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^3 z_i^3,$$

and

$$(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma z_i)^{-2} \approx (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} - 2(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-3} \sigma z_i + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2 z_i^2 - 4(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-5} \sigma^3 z_i^3.$$

Therefore, after some algebra, we have the response variance as

$$\begin{aligned} \mathrm{Var}(y_i) \approx \tau &\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2\right] + \\ &+ (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4} \sigma^2 + 8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-6} \sigma^4 + 15(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-8} \sigma^6. \end{aligned}$$

Conceptually, it is clear that this approximation can only yield useful results if $\sigma < |\boldsymbol{x}_i^\top \boldsymbol{\beta}|$ as otherwise the expansion will diverge. For practical purposes, we found empirically that $\sigma < 0.4 \times |(\boldsymbol{x}_i^\top \boldsymbol{\beta})|$ should be fulfilled in order to enable a reasonably accurate approximation.

Table 1: Variance of response under Gaussian quadrature models.

Response Distribution	Link function	$\mathrm{Var}(y_i)$
Normal	identity	$\tau + \sigma^2$
	log	$\tau + \exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \sigma^2\}(\exp\{\sigma^2\} - 1)$
	inverse[⊛]	$\tau + (\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4}\sigma^2 + 8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-6}\sigma^4 + 15(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-8}\sigma^6$
Gamma	identity	$(\tau + 1)\sigma^2 + \tau(\boldsymbol{x}_i^\top \boldsymbol{\beta})^2$
	log	$\exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \sigma^2\}[(\tau + 1)\exp\{\sigma^2\} - 1]$
	inverse[⊛]	$\tau\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-2} + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4}\sigma^2\right] +$
		$+(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4}\sigma^2 + 8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-6}\sigma^4 + 15(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-8}\sigma^6$
Poisson	log	$(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})\exp\{\sigma^2/2\}\times$
		$\times[1 + (\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})\exp\{\sigma^2/2\}(\exp\{\sigma^2\} - 1)]$
	identity	$\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma^2$
	square root	$(\boldsymbol{x}_i^\top \boldsymbol{\beta})^2 + 4(\boldsymbol{x}_i^\top \boldsymbol{\beta})^2\sigma^2 + \sigma^2 + 2\sigma^4$
Binomial	logit[⊛]	$\dfrac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} + 1} - \dfrac{(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})^2}{(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} + 1)^2} -$
		$-\dfrac{[(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}]\sigma^2}{2(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} + 1)^3} +$
		$+\dfrac{[(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})^3 - (\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})^2]\sigma^2}{(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} + 1)^4} -$
		$-\dfrac{(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\})^2 - \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}]^2\sigma^4}{4(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} + 1)^6}$
	probit[⊛]	$\Phi(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \dfrac{(\boldsymbol{x}_i^\top \boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{2} - \Phi^2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) +$
		$+(\boldsymbol{x}_i^\top \boldsymbol{\beta})\sigma^2\phi(\boldsymbol{x}_i^\top \boldsymbol{\beta})\Phi(\boldsymbol{x}_i^\top \boldsymbol{\beta}) -$
		$-\dfrac{(\boldsymbol{x}_i^\top \boldsymbol{\beta})^2\sigma^4\phi^2(\boldsymbol{x}_i^\top \boldsymbol{\beta})}{4}$
	cauchit[⊛]	$\dfrac{1}{4} - \dfrac{1}{\pi^2}\left\{\arctan(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \dfrac{(\boldsymbol{x}_i^\top \boldsymbol{\beta})\sigma^2}{[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^2 + 1]^2}\right\}^2$
	log	$\exp\left\{\boldsymbol{x}_i^\top \boldsymbol{\beta} + \dfrac{\sigma^2}{2}\right\} - \exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \sigma^2\}$
	comp. log-log[⊛]	$\exp\{-\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\} + \dfrac{\exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}]\sigma^2}{2\exp\{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\}} -$
		$-\exp\{-2\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\} - \dfrac{[\exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta})\} - \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}]\sigma^2}{\exp\{2\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\}} -$
		$-\dfrac{[\exp\{4(\boldsymbol{x}_i^\top \boldsymbol{\beta})\} - 2\exp\{3(\boldsymbol{x}_i^\top \boldsymbol{\beta})\} + \exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta})\}]\sigma^4}{4\exp\{2\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\}}$
Inv. Gaussian	$1/\mu^2$[⊛]	$\tau\left[\dfrac{1}{(\boldsymbol{x}_i^\top \boldsymbol{\beta})\sqrt{\boldsymbol{x}_i^\top \boldsymbol{\beta}}} + \dfrac{15\sigma^2}{8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^3\sqrt{\boldsymbol{x}_i^\top \boldsymbol{\beta}}}\right] +$
		$+\dfrac{\sigma^2}{4(\boldsymbol{x}_i^\top \boldsymbol{\beta})^3} + \dfrac{\sigma^4}{2(\boldsymbol{x}_i^\top \boldsymbol{\beta})^5} + \dfrac{375\sigma^6}{256(\boldsymbol{x}_i^\top \boldsymbol{\beta})^7}$
	inverse[⊛]	$\tau\left[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-3} + 6(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-5}\sigma^2\right] +$
		$+(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-4}\sigma^2 + 8(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-6}\sigma^4 + 15(\boldsymbol{x}_i^\top \boldsymbol{\beta})^{-8}\sigma^6$
	identity	$\tau[(\boldsymbol{x}_i^\top \boldsymbol{\beta})^3 + 3(\boldsymbol{x}_i^\top \boldsymbol{\beta})\sigma^2] + \sigma^2$
	log	$\tau\exp\left\{3(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \dfrac{9\sigma^2}{2}\right\} + \exp\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta} + \sigma^2)\} -$
		$-\exp\left\{2(\boldsymbol{x}_i^\top \boldsymbol{\beta}) + \dfrac{\sigma^2}{2}\right\}$

[⊛] Approximated via Taylor expansion.

3.2 Estimation via Gaussian Quadrature

Approximating

$$\mathrm{E}[V(\mu_i)] \approx \sum_{k=1}^{K} V_{ik}\pi_k, \qquad \mathrm{Var}[\mu_i] = \sum_{k=1}^{K} \mu_{ik}^2 \pi_k - \left(\sum_{k=1}^{K} \mu_{ik}\pi_k \right)^2,$$

one has via (14)

$$\mathrm{Var}(y_i) \approx \tau \sum_{k=1}^{K} V_{ik}\pi_k + \sum_{k=1}^{K} \mu_{ik}^2 \pi_k - \left(\sum_{k=1}^{K} \mu_{ik}\pi_k \right)^2. \tag{15}$$

An advantage of this expression is that it extends to nonparametric maximum likelihood estimation (NPML) of random effect models [Aitkin et al., 2009] by substituting μ_k and z_k with their estimates from the final EM iteration. In the context of Gaussian quadrature, which is the focus of this manuscript, we found (15) to behave very similarly to the analytic expressions above, as demonstrated in the following section.

4 Examples

We now provide two examples, using simulated and real data, to illustrate the use of the Fisher information matrix for the computation of standard errors of the regression parameter estimates. The first example involves four simulated data scenarios, one each for models with Poisson, Gamma, Normal and Inverse Gaussian responses. The second example illustrates the application of the Inverse Gaussian distribution to a real dataset with 30 observations. The results for the real data example can be reproduced with code available in the supplementary material. As reference, we use the standard errors obtained by two procedures: (i) via Monte Carlo with 10,000 replicates, (ii) via the heuristic formula

$$\mathrm{se}(\hat{\beta}_j) = \frac{|\hat{\beta}_j|}{\sqrt{\Delta\,\mathrm{disp}_j}},$$

where $\Delta\,\mathrm{disp}_j$ is the change in disparity (-2ℓ) when omitting the explanatory variable x_j [Aitkin et al., 2009, p. 439]. A natural limitation of this formula is that it is not possible to compute the standard error for σ or the intercept term. The values given in column (iii) of Tables 2 to 5 are the standard errors of $\hat{\boldsymbol{\gamma}}$ in the GLM fit of the last EM iteration. The results (iv) and (v) are the standard errors obtained using the analytic formula for variance (or the Taylor expansion) from Section 3.1 and the approximation from Section 3.2, respectively. In all of (i) to (v), the actual model fitting was carried out using R function `alldist` [Einbeck et al., 2013], using $K = 3$ throughout. For comparative purposes, we also provide the standard errors (vi) and parameter estimates $\hat{\boldsymbol{\gamma}}^*$ produced by the `glmer` function [Bates et al., 2014], using the default option `nAGQ=1` which implies a Laplace Approximation for the integral in (5).

4.1 Simulated data example

For each considered response distribution, we simulate 10,000 data sets of size $n = 90$ based on the following linear predictor

$$\eta_i = \beta_0 + \beta_1 x_i + \beta_{2i} + \sigma z_i, \quad i = 1, \ldots, n,$$

with the intercept $\beta_0 = 1$ in the case of Poisson, Gamma and Normal response, and $\beta_0 = 1.5$ for Inverse Gaussian. The covariate x is generated from $\mathcal{U}(0, 1)$ with coefficient $\beta_1 = -1$ for Poisson, Gamma and Normal cases and $\beta_1 = -0.125$ for Inverse Gaussian. The β_{2i} represent the coefficients of a factor with three levels, which $\beta_{2i} = (i \mod 3) - 1$ for Poisson, Gamma and Normal cases and $\beta_{2i} = 0.125 \times \{(i \mod 3) - 1\}$ for Inverse Gaussian. The random effect term is generated from $\mathcal{N}(0, 1)$ and the amount of variability due to the random effects is controlled by σ with value 0.125 for all models. We choose τ equal 1 for Gaussian and Gamma model, and equal 1/64 for the Inverse Gaussian model. The link functions are log for Poisson and Gamma, identity for Normal and inverse for Inverse Gaussian. We opt for a different set of parameter values for the Inverse Gaussian model due to the inverse link constraint $\eta_i \neq 0$, and the larger value of τ offers a balance for μ_i^3 in $\mathrm{Var}(y_i) = \mathrm{E}[\tau \mu_i^3] + \mathrm{Var}[\mu_i]$. We resample a new dataset for cases where `alldist` or `glmer` did not fit the model. For the Normal model, `lmer` is used instead of `glmer`.

Tables 2, 3, 4 and 5 display, respectively, the average value of the 10,000 values of $\hat{\boldsymbol{\gamma}}$ as well as the average standard errors of $\hat{\boldsymbol{\gamma}}$ for models fitted to the simulated response distributions Poisson, Gamma, Normal and Inverse Gaussian.

Table 2: Estimated fixed effects and respective standard errors (Poisson model with log link)

	γ	$\widehat{\gamma}$	se($\hat{\gamma}$) (i)	(ii)	(iii)	(iv)	(v)	(vi)	$\widehat{\gamma}^*$
β_0	1	0.98608	0.17855	—	0.17658	0.17149	0.17658	0.18070	0.98629
β_1	-1	-1.00175	0.24900	0.23640	0.24903	0.24019	0.24903	0.25527	-1.00199
β_{22}	1	1.00395	0.16835	0.16769	0.16607	0.16157	0.16607	0.16924	1.00445
β_{23}	-1	-1.01731	0.27644	0.23448	0.27395	0.27098	0.27396	0.27595	-1.01726
σ	0.125	0.12582	0.07255	—	0.07192	0.06912	0.07192	—	0.08803

Standard errors for $\hat{\gamma}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (15); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for γ obtained via `glmer`.

For the standard errors of the regression parameters, we see from columns (iv) and (v) of Tables 2 to 5 that the values obtained using our proposed methods are slightly below those obtained by Monte Carlo resampling (i). The standard errors (ii) using the disparity rule offer numbers close to (i) in the Poisson, Gamma and Normal examples. However, (ii) shows rather small standard error estimates for the Inverse Gaussian example. The standard errors (iii) taken from the generalised linear model fit of the last EM iteration are quite accurate for the Poisson model, but are underestimating the standard error for the Gamma, Normal and Inverse Gaussian models. We did not observe much difference between the approaches (iv) and (v) using the Fisher information, though the standard errors using (v) were slightly more accurate in general, especially for the Inverse Gauss scenario where a Taylor expansion was used for the analytic formula (iv). The standard errors using `glmer` (vi) were usually higher than than those of (i), (iv) and (v), except for the inverse Gaussian model, and were reasonably consistent with overall results. However, it is observed that `glmer` struggles to estimate the σ parameter correctly, sometimes underestimating (Poisson model) but mostly overestimating, and does not provide a value for the standard error of $\hat{\sigma}$ at all. We further note that, for the Gamma model, the average of `glmer` estimates for β_0 is less than half of the true value, which might indicate an identifiability issue.

Tables 6 and 7 show the estimated coverage probabilities for the Poisson

Table 3: Estimated fixed effects and respective standard errors (Gamma model with log link)

	γ	$\widehat{\gamma}$	se($\hat{\gamma}$) (i)	(ii)	(iii)	(iv)	(v)	(vi)	$\widehat{\gamma}^*$
β_0	1	0.96764	0.27011	—	0.14468	0.23456	0.24826	0.29173	0.40759
β_1	-1	-0.99366	0.38621	0.33437	0.20773	0.33435	0.35646	0.41886	-0.98002
β_{22}	1	0.99877	0.26766	0.26031	0.14359	0.23107	0.24639	0.28946	1.00816
β_{23}	-1	-0.99716	0.26561	0.25669	0.14350	0.23092	0.24624	0.28927	-0.98762
σ	0.125	0.12427	0.11265	—	0.05980	0.09665	0.10261	—	1.25017

Standard errors for $\hat{\gamma}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (15); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for γ obtained via `glmer`.

and Gamma models, respectively. In each table, the numbers show the results of estimated coverage probability (C.P.) computed through confidence intervals which use the standard error estimates (i), (ii), (iii), (iv), (v) and (vi) already discussed. Our intention here is to show two rather different scenarios, where the first (Poisson model) exemplifies well behaved numbers of C.P. and, in the second (Gamma model), an extreme case where we are able to note an evident contrast between the methods on the C.P.s.

Assuming that a specific method to compute the standard errors is reasonably suitable to compute the confidence intervals, we overall expect values close to the usual true confidence levels (C.L.) of 90%, 95% and 99% on average. Thus, we observe that for the Poisson model in Table 6, all five methods are acceptable according to our criteria, except for the disparity rule in (ii). However, for the Gamma model in Table 7, we note that the Monte-Carlo values in (i) are rather close to the true confidence levels, followed by the estimates via Fisher information matrix in (v) and (iv). The values computed using the disparity rule in (ii), the last EM iteration in (iii) and, especially, from `glmer` output in (vi) are overall smaller than the confidence levels.

4.2 Real data example

As a real data example, we take a subsample of the data from a 5×2 factorial experiment given by Ostle and Mensing [1963]. This subsample

Table 4: Estimated fixed effects and respective standard errors (Normal model with identity link)

	γ	$\widehat{\gamma}$	se($\hat{\gamma}$) (i)	(ii)	(iii)	(iv)	(v)	(vi)	$\widehat{\gamma}^*$
β_0	1	0.99425	0.27235	—	0.15096	0.25531	0.25903	0.26715	0.99440
β_1	-1	-0.99892	0.38068	0.32482	0.21190	0.35837	0.36360	0.37497	-0.99938
β_{22}	1	1.00644	0.26113	0.24809	0.14772	0.24983	0.25348	0.26142	1.00670
β_{23}	-1	-0.99233	0.26064	0.24702	0.14786	0.25007	0.25372	0.26171	-0.99250
σ	0.125	0.12676	0.11043	—	0.06150	0.10400	0.10553	—	0.79139

Standard errors for $\hat{\gamma}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (15); and
(vi) Laplace approximation (`lmer` output).
* shows the estimates for γ obtained via `lmer`.

is provided in the `R` library `mdscore` [da Silva-Júnior et al., 2014], using the syntax `data(strength)`. It is of interest to investigate how the impact strength of an insulating material is affected by the lot (I, II, III, IV, V) of the material and the type of specimen cut (lengthwise and crosswise). Previous analysis of the original dataset is given in Shuster and Miura [1972] and for the same subsample in da Silva-Júnior et al. [2014]. Shuster and Miura [1972] argued that the impact strength, which is a positive quantity, could be described by a failure time model (where 'strength' plays the role of 'time'), so that an Inverse Gaussian response distribution is deemed adequate. In our analysis, we adopt this reasoning, and hence assume that the impact strength measurements of a given replicate corresponding to the i-th cut and j–th lot are independently distributed as Inverse Gaussian distributions with means μ_{ij} and a fixed dispersion parameter. Suppose the linear predictor in the inverse link scale corresponds to the two–way interaction model

$$\mu_{ij}^{-1} = \tau_0 + \tau_i + \beta_j + (\tau\beta)_{ij} + \sigma z, \quad i = 1, 2, \; j = 1, 2, ..., 5, \qquad (16)$$

where $\tau_1 = 0$, $\beta_1 = 0$, $(\tau\beta)_{11} = \cdots = (\tau\beta)_{15} = (\tau\beta)_{21} = 0$, and z is a random effect that has Gaussian distribution.

Again, the estimate $\hat{\gamma}$ was obtained using `alldist`, and columns (ii) to (v) of Table 8 report the standard errors of $\hat{\gamma}$ obtained using the different techniques. Additionally, column (i) reports Monte-Carlo standard errors for $\hat{\gamma}$ by generating 10,000 new samples of size 30 responses based on (16),

Table 5: Estimated fixed effects and respective standard errors (Inv. Gaussian model with inverse link)

	γ	$\widehat{\gamma}$	se($\hat{\gamma}$) (i)	(ii)	(iii)	(iv)	(v)	(vi)	$\widehat{\gamma}^*$
β_0	1.5	1.50068	0.03923	—	0.02197	0.02801	0.03771	0.02686	1.54060
β_1	-0.125	-0.12495	0.05492	0.01756	0.03127	0.03961	0.05366	0.03140	-0.12421
β_{22}	0.125	0.12508	0.03986	0.01699	0.02243	0.02879	0.03849	0.03743	0.12406
β_{23}	-0.125	-0.12523	0.03848	0.01702	0.02153	0.02702	0.03694	0.04122	-0.12429
σ	0.125	0.12490	0.01615	—	0.00910	0.01166	0.01562	—	0.19321

Standard errors for $\hat{\gamma}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance;
(v) Fisher information matrix using approximation (15); and
(vi) Laplace approximation (`glmer` output).
* shows the estimates for γ obtained via `glmer`.

taking $\hat{\gamma}$ as "true" parameter values, and refitting the model for each one. It is further noted that, for this data set and model specification, the `glmer` attempt to fitting the model (16) failed to converge in our trials even when we relax the tolerances and the algorithm stopping criteria.

The numbers in (iv) and (v), for the fixed effects, are very slightly smaller than their counterparts in (i) and (ii). However, and contrary to our simulations presented in Subsection 4.1, the numbers in (iv) and (v) for $\hat{\sigma}$ are rather large. This might be due to misspecification of the random effects distribution. Finally, the numbers in (iii) are considerably smaller than their counterparts in (i), (ii), (iv) and (v).

5 Concluding remarks

Our simulation experiments indicate that the two presented approaches to compute the Fisher information matrix offer similar results for the standard errors of model parameters in GLMwRE's. While the resulting standard errors tend to be slightly underestimated as compared to their Monte Carlo counterparts, they are overall still more reliable than the disparity rule or the GLM standard errors obtained from the last EM iteration, which have been commonly used until now [Aitkin et al., 2009]. Standard errors based on EM are generally known to underestimate the true values, since they ignore the uncertainty inherent in the EM procedure. Also, while the standard errors

Table 6: Estimated coverage probabilities (Poisson model with log link)

	C.L. (%)	C.P. (%) $\beta_0 = 1$	$\beta_1 = -1$	$\beta_{22} = 1$	$\beta_{23} = -1$	$\sigma = 0.125$
(i)	90.00	89.96	90.26	90.19	90.30	89.94
	95.00	94.86	94.94	94.97	94.89	94.93
	99.00	98.90	98.97	98.86	98.69	98.94
(ii)	90.00	—	87.30	89.72	83.85	—
	95.00	—	92.67	94.56	90.26	—
	99.00	—	97.11	98.48	96.06	—
(iii)	90.00	89.86	90.46	89.96	90.87	89.80
	95.00	94.91	95.19	94.88	95.71	95.05
	99.00	99.05	99.11	98.90	99.17	99.08
(iv)	90.00	88.87	89.12	88.96	90.48	87.71
	95.00	94.07	94.31	94.23	95.44	93.21
	99.00	98.80	98.77	98.66	99.09	98.26
(v)	90.00	89.86	90.46	89.96	90.87	89.80
	95.00	94.91	95.19	94.88	95.71	95.05
	99.00	99.05	99.11	98.90	99.17	99.08
(vi)	90.00	89.80	90.46	90.03	90.86	—
	95.00	94.93	94.92	94.85	95.69	—
	99.00	99.00	99.08	98.83	99.18	—

Coverage probabilities of confidence intervals for $\hat{\gamma}$ computed using the standard errors obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with analytic variance;
(v) Fisher information matrix using approximation (15); and
(vi) `glmer` output.

computed using the disparity rule appeared reasonable for the examples, this methodology is not applicable for the intercept and the random effects parameter estimators.

The `glmer` alternative offers generally good standard errors (except for the σ parameter), which are however based on a different model estimation methodology at first place. The difference to the approach considered herein can be understood by following the classification given in Fahrmeir and Tutz [2001], who distinguish *direct* and *indirect* methods for maximizing integral

Table 7: Estimated coverage probabilities (Gamma model with log link)

	C.L. (%)	C.P. (%) $\beta_0 = 1$	$\beta_1 = -1$	$\beta_{22} = 1$	$\beta_{23} = -1$	$\sigma = 0.125$
(i)	90.00	90.20	90.10	89.88	89.88	90.05
	95.00	94.88	94.74	94.96	94.89	94.73
	99.00	98.60	98.99	99.07	98.95	98.73
(ii)	90.00	—	80.33	86.35	86.58	—
	95.00	—	85.53	91.62	91.45	—
	99.00	—	91.05	96.36	96.08	—
(iii)	90.00	62.39	62.56	62.43	62.46	61.78
	95.00	70.70	70.94	70.59	71.07	70.77
	99.00	83.03	83.34	82.49	83.12	83.02
(iv)	90.00	84.06	84.13	83.66	84.13	83.40
	95.00	90.41	90.48	90.22	90.29	89.32
	99.00	96.52	96.80	96.89	96.72	95.15
(v)	90.00	86.63	87.04	86.38	86.94	86.74
	95.00	92.42	92.54	92.33	92.54	92.31
	99.00	97.65	97.92	97.93	98.03	97.76
(vi)	90.00	39.44	79.00	78.44	79.17	—
	95.00	50.37	83.99	83.35	83.90	—
	99.00	68.89	87.75	87.80	88.04	—

Coverage probabilities of confidence intervals for $\hat{\gamma}$ computed using the standard errors obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance;;
(v) Fisher information matrix using approximation (15); and
(vi) `glmer` output.

(5). The 'Adaptive Gaussian Quadrature' (AGQ) [Pinheiro and Bates, 1995] used in `glmer` is a direct method, where derivatives of the marginal likelihood (5) with respect to the model parameters are taken first, and then Gauss-Hermite integration is carried out to find a numerical approximation of the score vector. The approach discussed in this paper is an indirect method, in which the integral is approximated first, and then the EM algorithm is used to iteratively find the expected probabilities of component membership, and maximize the corresponding expected complete log-likelihood (again by

Table 8: Estimated fixed effects and respective standard errors for strength data

	$\hat{\gamma}$	se($\hat{\gamma}$)				
		(i)	(ii)	(iii)	(iv)	(v)
τ_0	1.01704	0.07042	—	0.03197	0.06869	0.06832
τ_2	0.32828	0.10564	0.11340	0.04873	0.10462	0.10413
β_2	0.03201	0.10043	0.09876	0.04557	0.09780	0.09728
β_3	0.35915	0.10711	0.11543	0.04904	0.10531	0.10482
β_4	0.14128	0.10273	0.10293	0.04676	0.10037	0.09986
β_5	0.82348	0.11757	0.15159	0.05359	0.11513	0.11468
$(\tau\beta)_{22}$	-0.40636	0.14657	0.15279	0.06637	0.14247	0.14175
$(\tau\beta)_{23}$	-0.10864	0.15825	0.15968	0.07322	0.15726	0.15661
$(\tau\beta)_{24}$	-0.35020	0.14937	0.15481	0.06841	0.14689	0.14619
$(\tau\beta)_{25}$	-0.19501	0.17043	0.17270	0.07879	0.16928	0.16867
σ	0.00887	0.02119	—	0.01131	0.37348	0.37174

Standard errors for $\hat{\gamma}$ obtained via
(i) Monte Carlo;
(ii) disparity rule;
(iii) GLM fit in last EM iteration (`summary.glmmGQ` output);
(iv) Fisher information matrix with Taylor expansion of the analytic variance; and
(v) Fisher information matrix using approximation (15).

taking appropriate derivatives). This approach is computationally simpler than the direct approach, as the component probabilities in each iteration are completely known, while in the direct approach they depend on the true parameter vector [Fahrmeir and Tutz, 2001, p. 306]. The difference between the two approaches is most obvious by considering the case $K = 1$: in the approach considered herein, this corresponds to a fixed effect model, while in the context of AGQ it corresponds to the Laplace approximation. Due to this very different way in which the support points affect the estimation, we did not attempt to compare the direct and the indirect approach using a 'matching' number of components.

As pointed out by a referee, in generalised linear models the likelihood may be quite skewed around the MLE, and so the suitability of information–based standard errors for representing the precisions of parameter inference may be questioned. From this perspective, a Bayesian approach to the problem might be suggested. The reported coverage probabilities give some evidence that, despite this possible conceptual concern, the approach taken is still feasible, and that efforts to compute the standard errors accurately are useful and important.

In this manuscript we have discussed random effect models fitted via Gaussian Quadrature in the spirit of Aitkin [1996]. This model class is by construction easily extendible to the case of an unspecified random effect distribution, where the masses and support points are estimated simultaneously with the model parameters through the EM algorithm. This technique is known as 'Nonparametric Maximum Likelihood' (NPML, Aitkin et al. [2009]). The methods introduced in this manuscript could be extended to this scenario.

We finally note that, while the procedures outlined in this paper give standard errors for both fixed effects and random effects parameter estimators, we consider the standard errors produced for $\hat{\sigma}$ not to be fully reliable for any of the techniques investigated. The estimation of the standard error of $\hat{\sigma}$ suffers from identifiability issues, and is also sensitive to the correctness of the Normality assumption of the random effects. Accurate estimation of this standard error is a very hard problem which requires further attention.

Acknowledgements

We gratefully acknowledge grant n° 9622/13-6 from CAPES foundation through Brazil's Science Without Borders Program. The authors are grateful to Professor John Hinde, NUI Galway, for helpful comments on an earlier draft version, and to the Associate Editor for the careful handling of this manuscript.

References

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Number 55. Courier Dover Publications, 1972.

M. Aitkin. A general maximum likelihood analysis of overdispersion in generalized linear models. *Statistics and Computing*, 6(3):251–262, 1996. ISSN 0960-3174.

M. A. Aitkin, B. Francis, J. Hinde, and R. Darnell. *Statistical Modelling in R.* Oxford University Press Oxford, 2009.

D. Bates, M. Maechler, B. Bolker, and S. Walker. *lme4: Linear mixed-effects models using Eigen and S4.*, 2014. URL `http://CRAN.R-project.org/package=lme4`. R package version 1.1-7.

A. H. M. da Silva-Júnior, D. N. da Silva, and S. L. P. Ferrari. mdscore: An R package to compute improved score tests in generalized linear models. *Journal of Statistical Software*, 61(2):1–16, 2014. URL `http://www.jstatsoft.org/v61/c02/`.

J. Einbeck and J. Hinde. A note on NPML estimation for exponential family regression models with unspecified dispersion parameter. *Austrian Journal of Statistics*, 35(2&3):233–243, June 2006.

J. Einbeck, R. Darnell, and J. Hinde. *npmlreg: Nonparametric maximum likelihood estimation for random effect models.*, 2013. URL `http://CRAN.R-project.org/package=npmlreg`. R package version 0.45-4.

L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York, 2001.

H. Friedl and G. Kauermann. Standard errors for EM estimates in generalized linear models with random effects. *Biometrics*, 56(3):761–767, 2000.

N. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978.

D. Oakes. Direct calculation of the information matrix via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 61:479–482, 1999.

B. Ostle and R. W. Mensing. *Statistics in Research*. Iowa State University Press, Ames, 1963.

J. Pinheiro and D. Bates. Approximations to the log–likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4:12–35, 1995.

J. J. Shuster and C. Miura. Two-way analysis of reciprocals. *Biometrika*, 59:478–481, 1972.