HAL open science

A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature

Sebastian Ueckert, France Mentré

To cite this version:

Sebastian Ueckert, France Mentré. A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. Computational Statistics and Data Analysis, 2016, 10.1016/j.csda.2016.10.011 . inserm-01397584

HAL Id: inserm-01397584

https://inserm.hal.science/inserm-01397584

Submitted on 16 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature

Sebastian Ueckert, France Mentré

INSERM, IAME, UMR 1137, F-75018 Paris, France; Univ Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France

Abstract

The design of experiments for discrete mixed effect models is challenging due to the unavailability of a closed-form expression for the Fisher information matrix (FIM), on which most optimality criteria depend. Existing approaches for the computation of the FIM for those models are all based on approximations of the likelihood. A new method is presented which is based on derivatives of the exact conditional likelihood and which uses Monte Carlo (MC) simulations as well as adaptive Gaussian quadrature (AGQ) to integrate those derivatives over the data and random effects. The method is implemented in R and evaluated with respect to the influence of the tuning parameter, the accuracy of the FIM approximation, and computational complexity. The accuracy evaluation is performed by comparing the expected relative standard errors (RSE) from the MC/AGQ FIM with RSE obtained in a simulation study with four different discrete data models (two binary, one count and one repeated time-to-event model) and three different estimation algorithms. Additionally, the results from the MC/AGQ FIM are compared with expected RSE from a marginal quasi-likelihood (MQL) approximated FIM. The comparison resulted in close agreement between the MC/AGQ-based RSE and empirical RSE for all models investigated, and better performance of MC/AGQ than the MQL approximated FIM for variance parameters. The MC/AGQ method also proved to be well suited to calculate the expected power to detect a group effect for a model with binary outcomes.

Keywords: optimal design, Fisher information matrix, generalized linear mixed model, nonlinear mixed effect models, discrete data

1. Introduction

Mixed effect models allow to naturally capture data features arising in inherently longitudinal experiments with heterogeneity between experimental blocks. These design properties are found in many practical contexts (most notably clinical trials), which has

Email address: sebastian.ueckert@inserm.fr (Sebastian Ueckert)

led to an increasing utilization of mixed effect models for both continuous (often normally distributed) and discrete responses. The design of experiments for the discrete response mixed effect case is especially challenging due to the lack of a closed-form expression for the Fisher information matrix (FIM), for both discrete response generalized linear mixed effect models (GLMEM) and for discrete response nonlinear mixed effect models (NLMEM) (this is similar to the normal case, but here effective approximations are available [1]) . In this work, we describe a new method for evaluation of the FIM for this class of models.

There is a relativly large body of work on calculation of the FIM for GLMM. However, most articles describe methods that are based on approximate inferential methods, such as quasi-likelihood and generalized estimating equations, and are response type-specific. Examples include the work by Tekle et al., which focuses on the case of longitudinal binary data [2], and Niaparast's article, which is limited to Poisson models with random intercepts [3]. Also based on generalized estimating equations inference, but more general in terms of the distribution of the responses, is the approach presented by Woods and van de Ven [4]. For likelihood estimation of general GLMM with random intercepts, FIM approximations have been presented and evaluated in detail by Waite and Woods using marginal quasi-likelihood (MQL), penalized quasi-likelihood (PQL) or new complete enumeration-based methods, as well as Monte Carlo (MC) approximations thereof [5, 6].

For discrete NLMEM, there is considerably less preexisting work concerning the calculation of the FIM. Ogungbenro and Aarons describe a method based on generalized estimating equations and the MQL approximation, adapted to binary, ordinal and count responses [7]. Nyberg et al. presented a method utilizing a second-order approximation of the likelihood and applied it to binary and count responses [8].

Here we take a slightly different approach: rather than deriving the FIM for an approximate likelihood, we determine an expression for the FIM based on the exact conditional likelihood and subsequently compute the resulting expressions using MC simulations and adaptive Gaussian quadrature (AGQ). Provided there are a sufficient number of quadrature grid points, our method evaluates the exact likelihood, it is adapted for likelihood estimation of general discrete mixed effect models (GLMM as well as NLMEM) and requires only the specification of the expression for the conditional likelihood. The proposed method is an extension of the work by Nguyen and Mentré which describes the application of an MC/AGQ-based approach to approximate the FIM for NLMEM with normally distributed responses [9]. In their work, Nguyen and Mentré show the superiority of the MC/AGQ-based approach to a linearization-based calculation of the FIM when the model nonlinearity is increasing [9], as measured, for example, through the absolute value of the second derivative [10]. This finding, together with the performance of AGQ when used for parameter estimation in NLMEM [11, 12], was our motivation to extend the approach of Nguyen and Mentré and apply it to discrete response mixed effect models.

We evaluate our method by comparing the expected relative standard errors (RSE) from the MC/AGQ-based FIM with the parameter precisions obtained by repeatedly simulating responses from a model and subsequently re-estimating the model parameters. We will refer to this procedure as clinical trial simulation (CTS) as our primary interests are models originating from the analysis of clinical trials. The use of CTS is a common approach to evaluate FIM approximations for NLMEM, due to the lack of an analytic reference expression [7, 8, 9]. However, rather than limiting the compari-

son to a reference value obtained with one estimation algorithm, as is generally done, we obtain reference values with three different estimation algorithms: the Laplace approximation [13], importance sampling [14] and stochastic approximation expectation maximization (SAEM) [15]. These algorithms represent different approaches to handling the analytically intractable NLMEM likelihood as well as its maximization. The Laplace approximation-based estimation algorithm uses a second-order approximation of the likelihood and a gradient-based algorithm for maximization. Importance sampling approximates the likelihood through MC sampling and performs maximization using the expectation maximization algorithm. SAEM relies on Markov-Chain MC sampling in combination with the expectation maximization algorithm. All three algorithms are commonly used in practice and have proven suitable for a wide range of models [12, 16, 17, 18, 19].

We structured our paper in three parts. Part one (section 2) describes the derivation of an expression of the FIM for discrete mixed effect models as well as the proposed way of approximating this expression using MC simulations and AGQ. In part two (section 3), the proposed method is evaluated in regard to the influence of tuning parameters, its performance in comparison with CTS or an MQL approximation, and the computational complexity of the method. Part three (section 4) presents an application example that uses the MC/AGQ method to calculate the expected power to detect a group effect for a clinical trial with binary outcomes under cost constraints.

2. Fisher information matrix for discrete nonlinear mixed effect models

2.1. Notation

2.1.1. Design.

Let ξ_i denote the vector of design variables for individual i ($i = 1, \ldots, N$). In most cases, a limited number K of groups of subjects with identical design variables $\xi_{\{k\}}$ exist. Hence, the population design, i.e. the superset of design variables from all subjects, can be efficiently written as the set of pairs $\Xi = \{[\xi_{\{1\}}, N_{\{1\}}], \ldots, [\xi_{\{K\}}, N_{\{K\}}]\}$ where $\xi_{\{k\}}$ is the design variable for group k and $N_{\{k\}}$ is the number of subjects in that group.

2.1.2. Nonlinear mixed effect models.

This work considers NLMEM for discrete data where the conditional probability for observation j from subject i can be written as

$$P(y_{ij}|b_i) = h(y_{ij}, \xi_i, g(\mu, b_i, z_i)). \tag{1}$$

Here h is a function describing the probability of y_{ij} given b_i for a vector of design variables ξ_i and a vector of subject-specific parameters modeled through the nonlinear function g. The function g depends on the vector of fixed effects parameters μ, the vector of random effects b_i and the vector of subject-specific covariates z_i. The random effects are assumed to follow a multivariate normal distribution with mean zero and covariance matrix Ω, i.e. $b_i \sim N(0, \Omega)$. Furthermore, let Ψ be the vector of all model parameters, i.e. $\Psi = (\mu, \Omega_u)^T$ where Ω_u is a vector containing all unique elements of Ω.

Generally, observations are assumed independent conditionally on the random effect, i.e. the joint conditional probability for the vector of observations for subject i ($y_i =$

$(y_{i1}, \ldots, y_{in_i})^T)$ is equal to

$$P(y_i|b_i) = \prod_{j=1}^{n_i} h(y_{ij}, \xi_i, g(\mu, b_i, z_i)). \tag{2}$$

The likelihood $L(y_i, \Psi)$ is a marginal likelihood, given as the integral

$$L(y_i, \Psi) = \int \prod_{j=1}^{n_i} h(y_{ij}, \xi_i, g(\mu, b_i, z_i)) p(b_i) db_i \tag{3}$$

$$= \int l(y_i, \Psi, b_i) db_i, \tag{4}$$

where $p(b_i)$ is the p.d.f. of a multivariate normal distribution (mean 0 and covariance matrix Ω) and $l(y_i, \Psi, b_i)$ the joint likelihood of observations and random effects.

2.2. Fisher information matrix

The information matrix $\mathcal{I}(\Psi, \Xi)$ for the parameters Ψ and the population design Ξ is given by the sum over all elementary matrices $\mathcal{I}(\Psi, \xi_i)$, i.e.,

$$\mathcal{I}(\Psi, \Xi) = \sum_{i=1}^{N} \mathcal{I}(\Psi, \xi_i) \tag{5}$$

or in the case of K groups of subjects of size $N_{\{k\}}$ ($N = \sum N_{\{k\}}$) with identical design

$$\mathcal{I}(\Psi, \Xi) = \sum_{k=1}^{K} N_{\{k\}} \mathcal{I}(\Psi, \xi_{\{k\}}). \tag{6}$$

The elementary information matrix $\mathcal{I}(\Psi, \xi_i)$ is defined as

$$\mathcal{I}(\Psi, \xi_i) = \mathbf{E}_{y_i} \left[\frac{\partial \log L(y_i, \Psi)}{\partial \Psi} \frac{\partial \log L(y_i, \Psi)}{\partial \Psi}^{\mathsf{T}} \right] \tag{7}$$

$$= \mathbf{E}_{y_i} \left[\frac{\partial L(y_i, \Psi)}{\partial \Psi} \frac{\partial L(y_i, \Psi)}{\partial \Psi}^{\mathsf{T}} (L(y_i, \Psi))^{-2} \right]. \tag{8}$$

Under fairly general conditions the differentiation and integration operators can be exchanged and the derivative of the likelihood in equation (8) can be calculated as

$$\frac{\partial L(y_i, \Psi)}{\partial \Psi} = \frac{\partial}{\partial \Psi} \int l(y_i, \Psi, b_i) db_i = \int \frac{\partial}{\partial \Psi} l(y_i, \Psi, b_i) db_i. \tag{9}$$

Performing the differentiation under the integral sign w.r.t. to μ and Ω_u results in (the details of this derivation are given in the appendix)

$$\frac{\partial l(y_i, \Psi)}{\partial \mu} = \int \prod_{j=1}^{n_i} h(\cdot) \left(\sum_{j=1}^{n_i} \frac{\partial \log h(\cdot)}{\partial g} \frac{\partial g(\cdot)}{\partial \mu} \right) p(b_i) db_i \tag{10}$$

$$\frac{\partial l(y_i, \Psi)}{\partial \Omega_u} = \int \prod_{j=1}^{n_i} h(\cdot) \left(\Omega^{-\frac{1}{2}} b_i \Omega^{-\frac{1}{2}} \frac{\partial \Omega^{\frac{1}{2}}}{\partial \Omega_u} \Omega^{-\frac{1}{2}} b_i - \mathrm{Tr} \left(\frac{\partial \Omega^{\frac{1}{2}}}{\partial \Omega_u} \Omega^{-\frac{1}{2}} \right) \right) p(b_i) db_i \tag{11}$$

The derivative of the Cholesky factor $\Omega^{\frac{1}{2}}$ w.r.t. Ω_u is given by the matrix [20]

$$\frac{\partial \Omega^{\frac{1}{2}}}{\partial \Omega_u} = \Omega^{\frac{1}{2}} \Lambda \left(\Omega^{-\frac{1}{2}} \frac{\partial \Omega}{\partial \Omega_u} (\Omega^{-\frac{1}{2}})^{\mathsf{T}} \right), \tag{12}$$

with Λ defined as the function

$$\Lambda(M) = \begin{cases} M_{ij} & \text{if } i > j \\ \frac{1}{2} M_{ij} & \text{if } i = j \\ 0 & \text{if } i < j. \end{cases} \tag{13}$$

2.3. Proposed algorithm

The challenge in the calculation of the FIM is the evaluation of the integrals. First, the integration required to calculate the expectation in equation (8) and, second, the integration over the random effects in equations (3), (10) and (11). We propose to solve the former using MC integration (section 2.3.1) and the latter using AGQ (section 2.3.2), and refer to our method as MC/AGQ.

2.3.1. Monte Carlo integration.

The expectation over the observations in equation (8) was approximated through MC integration for each elementary design ξ_i, by first simulating S vectors of random effects b_s, subsequently simulating observations y_s according to ξ_i for each random effect b_s and finally approximating the FIM as the average over all samples, i.e.,

$$\mathbf{E}_{y_i} \left[\frac{\partial L(y_i, \Psi)}{\partial \Psi} \frac{\partial L(y_i, \Psi)}{\partial \Psi}^{\mathsf{T}} (L(y_i, \Psi))^{-2} \right] \approx \frac{1}{S} \sum_{s=1}^{S} \frac{\partial L(y_{i,s}, \Psi)}{\partial \Psi} \frac{\partial L(y_{i,s}, \Psi)}{\partial \Psi}^{\mathsf{T}} (L(y_{i,s}, \Psi))^{-2} \tag{14}$$

2.3.2. Adaptive Gaussian quadrature.

To compute the integrals in equations (3), (10), and (11) over the random effects we use numerical quadrature. First, we note that all three equations can be written as

$$\int_{\mathbb{R}} f(b_i, \cdot) p(b_i) db_i \tag{15}$$

with

$$f(b_i, \cdot) = \prod_{j=1}^{n_i} h(\cdot) \tag{16}$$

for equation (3),

$$f(b_i, \cdot) = \prod_{j=1}^{n_i} h(\cdot) \left(\sum_{j=1}^{n_i} \frac{\partial \log h(\cdot)}{\partial g} \frac{\partial g(\cdot)}{\partial \mu} \right) \tag{17}$$

for equation (10) and

$$f(b_i,\cdot) = \prod_{j=1}^{n_i} h(\cdot)\left(\Omega^{-\frac{1}{2}} b_i \Omega^{-\frac{1}{2}} \frac{\partial \Omega^{\frac{1}{2}}}{\partial \Omega_u} \Omega^{-\frac{1}{2}} b_i - \mathrm{Tr}\left(\frac{\partial \Omega^{\frac{1}{2}}}{\partial \Omega_u}\Omega^{-\frac{1}{2}}\right)\right) \tag{18}$$

for equation (11). Second, we introduce the transformation $\eta_i = \Omega^{-\frac{1}{2}} b_i$ to yield

$$\int f(b_i,\cdot)p(b_i)db_i = \int f(\Omega^{\frac{1}{2}}\eta_i,\cdot)p(\Omega^{\frac{1}{2}}\eta_i)|\Omega^{\frac{1}{2}}|d\eta_i = \int \tilde{f}(\eta_i,\cdot)\phi(\eta_i)d\eta_i, \tag{19}$$

where $\phi(\cdot)$ denotes the p.d.f. of the standard multivariate normal distribution and $\tilde{f}(\eta_i,\cdot) = f(\Omega^{\frac{1}{2}}\eta_i,\cdot)$. The right-hand side of equation (19) can then be computed using Gaussian quadrature with appropriately chosen quadrature nodes $a_{q_1,\ldots,q_d}$ and weights $w_{q_1,\ldots,q_d}$ as

$$\int \tilde{f}(\eta_i,\cdot)\phi(\eta_i)d\eta_i \approx \sum_{q_1=1}^{Q}\cdots\sum_{q_d=1}^{Q} w_{q_1,\ldots,q_d}\cdot \tilde{f}(a_{q_1,\ldots,q_d},\cdot). \tag{20}$$

Both nodes and weights can be calculated as described, for example, by Golub and Welsch [21]. In general, the higher the order Q, the better the approximation will be [22]. Alternatively, the integral approximation can be improved by using AGQ, which takes the shape of the integrand ($\tilde{f}(\eta_i,\cdot)\phi(\eta_i)$) into account. Nguyen and Mentré, for example, propose to center the AGQ grid at the simulated random effect value $\eta_s = \Omega^{-\frac{1}{2}} b_s$ [9]. Here, we extend their approach by centering the grid at the mode and scaling it by the second derivative of the joint likelihood. The nodes $a^*_{q_1,\ldots,q_d}$ and weights $w^*_{q_1,\ldots,q_d}$ for this adaptation can be calculated by first determining the mode $\hat{\eta}_i$ of the joint likelihood under the transformation $\eta_i = \Omega^{-\frac{1}{2}} b_i$, i.e.,

$$\hat{\eta}_i = \arg\max_{\eta_i} \log \tilde{l}(y_i,\Psi,\eta_i) \tag{21}$$

with $\tilde{l}(y_i,\Psi,\eta_i) = l(y_i,\Psi,\Omega^{\frac{1}{2}}\eta_i)$. Then computing the nodes according to

$$a^*_{q_1,\ldots,q_d} = \hat{\eta}_i + \left[-\frac{\partial^2}{\partial \eta_i^2}\log \tilde{l}\Big|_{\eta_i=\hat{\eta}_i}\right]^{-\frac{1}{2}} a_{q_1,\ldots,q_d} \tag{22}$$

and the weights according to

$$w^*_{q_1,\ldots,q_d} = \left|\left[-\frac{\partial^2}{\partial \eta_i^2}\log \tilde{l}\Big|_{\eta_i=\hat{\eta}_i}\right]^{-\frac{1}{2}}\right| \prod_{k=1}^{d} w_{q_k}\frac{\phi(a^*_{q_k})}{\phi(a_{q_k})}, \tag{23}$$

where $\frac{\partial^2}{\partial \eta_i^2}\log \tilde{l}|_{\eta_i=\hat{\eta}_i}$ denotes the second derivative of the joint log-likelihood w.r.t. η_i evaluated at $\hat{\eta}_i$. The latter can be calculated using a finite difference approximation.

Substituting the adaptive nodes $a^*_{q_1,\ldots,q_d}$ and weights $w^*_{q_1,\ldots,q_d}$ in equation (20) yields the AGQ formula

$$\int \tilde{f}(\eta_i,\cdot)\phi(\eta_i)d\eta_i \approx \sum_{q_1=1}^{Q}\cdots\sum_{q_d=1}^{Q} w^*_{q_1,\ldots,q_d}\cdot \tilde{f}(a^*_{q_1,\ldots,q_d},\cdot) \tag{24}$$

which was used to compute the integrals in equations (3), (10), and (11).

The FIM approximated with S MC samples and Q quadrature points will be denoted $\mathcal{I}_{S,Q}$ in the following sections.

3. Evaluation of the algorithm

The MC/AGQ FIM method described above was implemented in R version 3.1.1 [23] using the statmod package [24] for the calculation of the Gaussian quadrature nodes and weights. The resulting implementation was then evaluated using four different evaluation examples regarding the influence of algorithm tuning parameters, the performance of the FIM approximation obtained in comparison with results from CTS as well as an MQL-based FIM, and the computational complexity of the algorithm.

3.1. Examples

All algorithm evaluations were performed for four different evaluation examples of discrete mixed models, representing a selection of different discrete data types. A sample of the data simulated for each example as part of this evaluation is displayed in Figure 1.

3.1.1. Example 1 – Binary response.

The model was based on an example used by Nyberg et al. [8] as well as Ogungbenro and Aarons [7] to evaluate alternative implementations of the FIM for discrete NLMEM. For this example, the observations are binary responses ($y_{ij} \in \{0, 1\}$) for different dose levels (D_j). The model of the observations was given as

$$h(y_{ij}, \xi, g(\mu, b_i, z_i)) = \begin{cases} \text{expit}(g_1 + g_2 D_j) & \text{if } y_{ij} = 1 \\ 1 - \text{expit}(g_1 + g_2 D_j) & \text{if } y_{ij} = 0 \end{cases} \tag{25}$$

with

$$\text{expit}(x) = \frac{1}{1 + e^{-x}} \tag{26}$$

and

$$g(\mu, b_i, z_i) = (g_1, g_2)^\mathsf{T} = (\mu_1 + b_i, \mu_2)^\mathsf{T}, \tag{27}$$

where μ_1 and μ_2 are fixed effects for intercept and slope of the dose relationship and b_i is an intercept random effect. The parameter values used for this example were $\mu = (\mu_1, \mu_2) = (-0.5, 5)$ for the fixed effects and $\omega_1^2 = 0.1$ for the variance of the random effect. A study design with $N = 50$ individuals and $n = 30$ observations per individual was used ($\xi = (D_1, \ldots, D_n)^\mathsf{T}$). The 30 observations were split across three dose levels with

$$D_j = \begin{cases} 0 & j = 1, \ldots, 10 \\ 0.25 & j = 11, \ldots, 20 \\ 0.45 & j = 21, \ldots, 30. \end{cases} \tag{28}$$

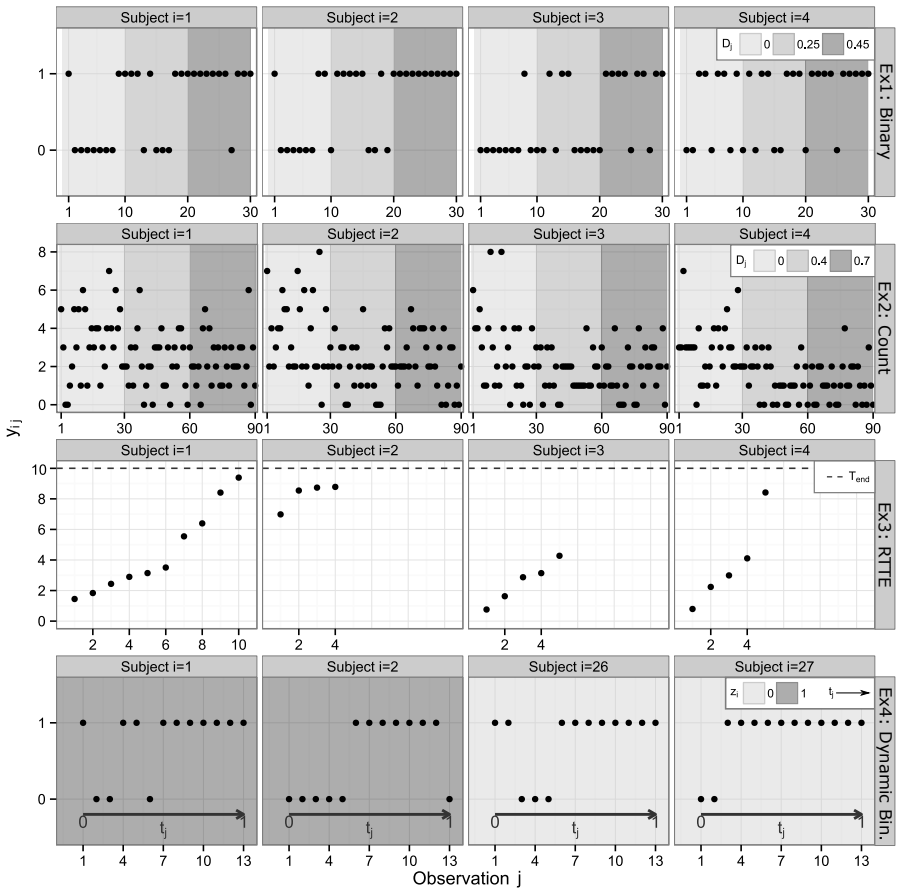

Figure 1: Examples of simulated observations y_{ij} for four subjects for each of the four models (see section 3.1). The figure shows, in a separate row for each of the four examples, the observation index on the x axis, the simulated value on the y axis as well as the association of the design variables through shading, lines, or arrows.

3.1.2. Example 2 – Count response.

The second example models count type data and was also used previously to evaluate the performance of an FIM approximation for discrete data [8, 7]. The observations are non-negative integer values ($y_{ij} \in \{0, 1, 2, \ldots\}$) representing counts for different dose levels (D_j). The probability for each observation was described using the Poisson distribution with dose-dependent Poisson parameter λ_{ij}, i.e.,

$$h(y_{ij}, \xi, g(\mu, b_i, z_i)) = \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \exp(-\lambda_{ij}). \tag{29}$$

with

$$\log \lambda_{ij} = g_1 \left(1 - \frac{D_j}{D_j + g_2}\right) \tag{30}$$

and

$$g(\mu, b_i, z_i) = (g_1, g_2)^\mathsf{T} = (\mu_1 \exp(b_{1i}), \mu_2 \exp(b_{2i}))^\mathsf{T}, \tag{31}$$

where μ_1 and b_{1i}, and μ_2 and b_{2i} are fixed and random effects of the dose relationship. The parameter values used for this example were $\mu = (\mu_1, \mu_2) = (1, 0.5)$ for the fixed effects and $\omega_1^2 = \omega_2^2 = 0.09$ for the variances of the random effects. The design considered for this example included $N = 20$ subjects with $n = 90$ observations ($\xi = (D_1, \ldots, D_n)^\mathsf{T}$), split across the three dose levels with

$$D_j = \begin{cases} 0 & j = 1, \ldots, 30 \\ 0.4 & j = 31, \ldots, 60 \\ 0.7 & j = 61, \ldots, 90. \end{cases} \tag{32}$$

3.1.3. Example 3 – Repeated time-to-event response.

The third example described repeated time-to-event (RTTE) type data (which strictly speaking are not discrete, but were included here to show the versatility of the approach). For each subject, the observations y_i consisted of the event times between 0 and the end of the study T_{end} ($\{y_{i,j} \in \mathbb{R} | 0 \leq y_{ij} \leq T_{end}\}$). The model assumed a constant, subject-specific hazard. The conditional probability for y_i was given as

$$h(y_{ij}, \xi, g(\mu, b_i, z_i)) = \begin{cases} (g \exp(-g(y_{ij} - y_{i(j-1)}))) & \text{if } j < n_i \\ \exp(-g(T_{end} - y_{in_i})) & \text{if } j = n_i \end{cases} \tag{33}$$

with

$$y_{i0} = 0 \tag{34}$$

and

$$g(\mu, b_i, z_i) = \mu_1 \exp(b_i). \tag{35}$$

The parameters for the hazard used in this example were $\mu_1 = 0.5$ for the fixed effect and $\omega_1^2 = 0.1$ for the variance of the random effect. A study design with $N = 50$ subjects and a study length of $T_{end} = 10$ time units was assumed.

3.1.4. Example 4 – Dynamic binary response.

In the final, fourth, example a dynamic model for a binary response was considered. The observations y_i consisted of binary responses ($y_{ij} \in \{0, 1\}$) recorded over time in two different groups. The model describes an increasing probability of one response over time with group-specific slopes, i.e.

$$h(y_{ij}, \xi, g(\mu, b_i, z_i)) = \begin{cases} \text{expit}(g_1 + g_2 t_j) & \text{if } y_{ij} = 1 \\ 1 - \text{expit}(g_1 + g_2 t_j) & \text{if } y_{ij} = 0 \end{cases} \tag{36}$$

with $\xi = (t_1, \ldots, t_n)^\mathsf{T}$ and

$$g(\mu, b_i, z_i) = (g_1, g_2)^\mathsf{T} = (\mu_1 + b_{i1}, (\mu_2 + b_{i2})(1 - \mu_3 z_i))^\mathsf{T}, \tag{37}$$

where μ_1 and μ_2 are the baseline and slope fixed effects, b_{1i} and b_{2i} are the associated random effects, μ_3 is the group effect, t_j is time, and $z_i \in \{0, 1\}$ is the group indicator. The parameter values $\mu = (\mu_1, \mu_2, \mu_3) = (-1, 4, 0.4)$ were used for the fixed effects and $(\omega_1^2, \omega_2^2) = (0.5, 4)$ for the variances of the random effects. The design consisted of two equal sized groups with $N_1 = 25$ subjects $z_i = 0$ in group 1 and $N_2 = 25$ subjects $z_i = 1$ in group 2 ($N = N_1 + N_2 = 50$), as well as $n = 13$ observations per subject. Observations were assumed to be taken between 0 and 1 time units with equal spacing between observations.

This example was also used to evaluate the performance of the MC/AGQ method when varying design variables. Starting from the reference scenario described above, three different scenarios were studied: scenario S_1 varied the total study size $N = N_1 + N_2$, scenario S_2 varied the number of observations per subject n, and scenario S_3 varied both study size and number of observations per subject while keeping the total number of observations $n_{tot} = N \cdot n$ constant. All three scenarios assumed equal spacing between observations between 0 and 1 time units.

3.2. Evaluation of MC/AGQ tuning parameters

The two most important tuning parameters of the MC/AGQ algorithm are the number of MC samples, S in equation (14), and the number of quadrature grid points Q (equation (24)). While Q affects the accuracy of the approximation of the FIM, S influences its precision.

3.2.1. Methods – Evaluation of MC/AGQ tuning parameters

The influence of both parameters was evaluated for each of the four examples using three criteria of the FIM, the Pth root of the determinant of the FIM where P is the total number of parameters, i.e.,

$$\phi_1(S, Q) = \sqrt[P]{|\mathcal{I}_{S,Q}(\Psi, \Xi)|}, \tag{38}$$

the trace of the fixed effect parameter part of the inverse of the FIM

$$\phi_2(S, Q) = \text{Tr}((\mathcal{I}_{S,Q}^{-1}(\Psi, \Xi))_{(\mu)}), \tag{39}$$

and the trace of the variance of the random effect part of the inverse of the FIM

$$\phi_3(S,Q) = \mathrm{Tr}((\mathcal{I}_{S,Q}^{-1}(\Psi,\Xi))_{(\Omega_u)}). \tag{40}$$

The three criteria were normalized by their respective values obtained with a sufficiently large number of grid points, i.e.

$$\phi_{rel,k}(S,Q) = \frac{\phi_k(S,Q)}{\phi_k(10^4,10)}. \tag{41}$$

For the evaluation of Q, $\phi_{rel,k}(10^4,Q)$ was studied for $Q = 1,3,5,10$ and for the evaluation of S, the standard deviation of $\phi_{rel,k}(S,3)$ from 100 repetitions of MC/AGQ with $S = 100,250,500,1000,2000,5000$ was evaluated. The value of $Q = 3$ for the latter was chosen based on the results from the Q evaluation.

3.2.2. Results – Evaluation of MC/AGQ tuning parameters

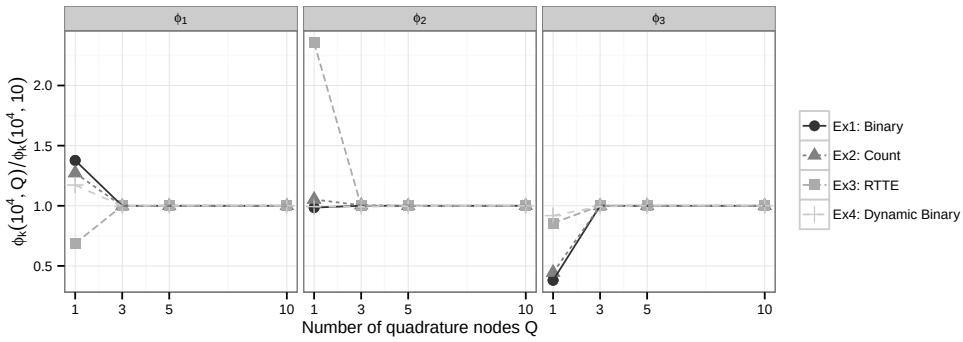

Figure 2: Influence of the number of quadrature nodes Q on the normalized determinant of the Fisher information matrix (FIM) (ϕ_1, left panel), the trace of the fixed effect parameter part of the inverse of the FIM (ϕ_2, middle panel), and the trace of the variance of the random effect part of the inverse of the FIM (ϕ_3, right panel) for all four models. All criteria are shown relative to the values obtained with 10 quadrature nodes and were obtained with $S = 10,000$ Monte Carlo samples.

The influence of the number of quadrature grid points on the three evaluation criteria is illustrated in Figure 2. With 1 grid point, results differed by more than a factor of 2.5 from the reference, but when using 3 or more grid points all evaluation criteria were virtually identical to the reference.

Figure 3 displays the impact of the number of MC samples S on the standard deviation of the three evaluation criteria. Clearly, the influence of S differs between criteria. For example, to obtain a standard deviation of less than 10% only 500 MC samples are sufficient for the trace of the fixed effect parameter part of the inverse of the FIM, but 1,000 are required for the normalized FIM and even 2,000 for the trace of the variance of the random effect part of the inverse of the FIM.

3.3. Performance of the MC/AGQ-based FIM approximation

To evaluate the performance of the MC/AGQ-based FIM approximation, the expected RSE based on the predicted FIM (ratio of the square root of the diagonal terms

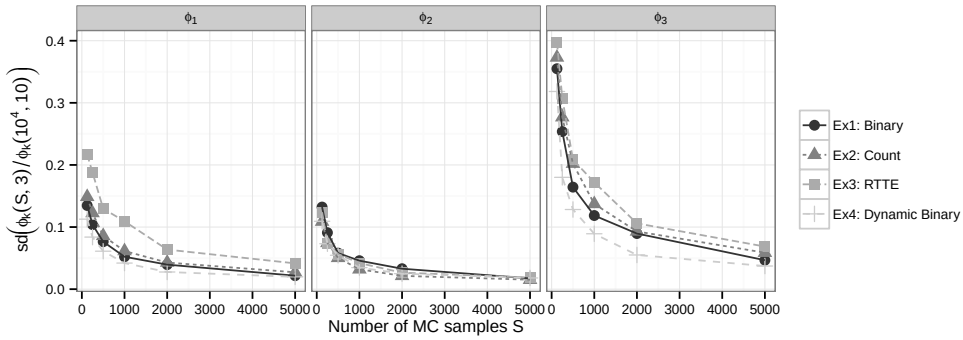

Figure 3: Influence of the number Monte Carlo (MC) samples S on the standard deviation (sd) of the normalized determinant of the Fisher information matrix (FIM) (ϕ_1, left panel), the sd of the trace of the fixed effect parameter part of the inverse of the FIM (ϕ_2, middle panel), and the sd of the trace of the variance of the random effect part of the inverse of the FIM (ϕ_3, right panel) for all four models. The sd for all criteria was obtained with $Q = 3$ quadrature nodes.

of $\mathcal{I}^{-1}$ and the parameter value) from the algorithm were compared with the empirical RSE from CTS for all four evaluation examples. Where applicable (examples 1, 2, and 4), the results from the MC/AGQ-based FIM were also compared with the expected RSE obtained with an MQL-based FIM approximation presented by Ogungbenro and Aarons [7] (the details of this approximation are given in the appendix).

3.3.1. Methods – Clinical trial simulations

For each example, $M_C = 1000$ datasets were simulated using R. Each dataset was subsequently used to estimate the population parameters using the SAEM estimation algorithm in MONOLIX version 4.3.3 as well as the Laplace and importance sampling (IMP) estimation algorithms in NONMEM version 7.3. SAEM in MONOLIX and Laplace in NONMEM were run with their default settings (except for example 4, where MONOLIX was run with 5 chains), IMP in NONMEM was run with the option AUTO=1. The 1000 sets of parameter estimates from each algorithm were used to calculate the empirical RSE according to

$$\mathrm{RSE}(\hat{\psi}) = \frac{1}{\psi}\sqrt{\frac{1}{M_C}\sum_{m=1}^{M_C}\left(\hat{\psi}_m - \mathbf{E}(\hat{\psi})\right)^2}, \tag{42}$$

where $\hat{\psi}_m$ is the parameter estimate, ψ is the true parameter value that was used to simulate the data, and $\mathbf{E}(\hat{\psi})$ is the empirical mean ($\mathbf{E}(\hat{\psi}) = 1/M_C \sum_{m=1}^{M_C} \hat{\psi}_m$). Additionally, a 95% confidence interval for the empirical RSE of each parameter and estimation algorithm was determined through a bootstrap procedure. The bootstrap procedure determined the 2.5% and 97.5% percentiles of the empirical RSE distribution, by sampling (with repetition) 1000 vectors of size 1000 from the set of final parameter estimates and subsequent calculation of the RSE according to equation (42).

The set of 1000 estimates from each estimation algorithm was also used to calculate the observed FIM as the inverse of the empirical variance-covariance matrix of estimates.

Table 1: Combination of design variables (total study size N, number of observations per subject n and total number of observations $n_{tot} = N \cdot n$) studied with example 4 and method of evaluation used.

Scenario	Method of evaluation	Design variables		
		N	n	$n_{tot} = N \cdot n$
S_1: *Total study size*				
	MC/AGQ FIM	20 - 80	13	260 - 1040
	CTS	20, 50, 80	13	260, 650, 1040
S_2: *Observations per subject*				
	MC/AGQ FIM	50	5 - 20	250 - 1000
	CTS	50	5, 13, 30	250, 650, 1000
S_3: *Constant total number of observations*				
	MC/AGQ FIM	20 - 60	4 - 12	240
	CTS	20, 40, 60	12, 6, 4	240

For the observed FIM, the normalized determinant was calculated as specified before and compared with the normalized determinant of the expected FIM.

For the second part of example 4, three specific designs from each scenario were evaluated through CTS for the evaluation of the performance of the MC/AGQ method when varying design variables. The range of design variables studied with the MC/AGQ FIM and with CTS under the three scenarios is listed in Table 1.

3.3.2. Methods – MC/AGQ-based FIM approximation

Based on the results of section 3.2 and the precision achieved with CTS, the MC/AGQ algorithm was run with $S = 2000$ MC samples and $Q = 3$ adaptive quadrature grid points, i.e. the comparison was performed for $\mathcal{I}_{3,2000}$. The expected standard error (SE) was calculated as the square root of the diagonal terms of $\mathcal{I}_{3,2000}^{-1}$ and the expected RSE was defined as the ratio of SE and the parameter value. Additionally, the Pth root of the determinant of the FIM, with P equal to the total number of parameters, was calculated as defined in equation (38).

3.3.3. Results – Performance of the MC/AGQ-based FIM approximation

For all four examples, Figure 4 compares the expected RSE, obtained through the MC/AGQ and MQL FIM, with the empirical RSE, obtained through CTS with three different estimation algorithms. The differences in empirical RSE between estimation algorithms are minor for all models, with larger deviations only for the variance of the random effects for the count and dynamic examples. Concerning the evaluation of the RSE predictions, it should be noted that based on the Cramer-Rao inequality, the FIM predicted RSE are expected to be equal to or lower than the empirical RSE. Since rather rich designs were used for the evaluation, a close agreement between the two quantities is expected for a good approximation of the FIM.

Generally, the empirical RSE are in good agreement with the MC/AGQ FIM predicted RSE. For all fixed effect parameters as well as for the random effect variances of the binary example and RTTE examples the agreement is very high. For the random effect variances ω_2^2 of the count and dynamic examples, the deviations between the MC/AGQ

FIM predicted RSE and the empirical RSE are estimation algorithm-dependent and low overall. Also, the agreement between the normalized determinant of the expected and observed FIM, shown in Table 2, is high overall. Only the determinant of the expected FIM for the RTTE model is lower than the determinant of the observed FIM.

For fixed effect parameters, the agreement between the empirical RSE from CTS and the expected RSE from the MQL approximated FIM is also very high and essentially identical to the MC/AGQ-based results. For random effect variances, however, the performance of the MQL approximated FIM is clearly inferior. Especially for ω_2^2 of the count and both ω^2 parameters of the dynamic example the agreement with CTS is weak. Most importantly, the MQL approximation could not be applied to the RTTE example.

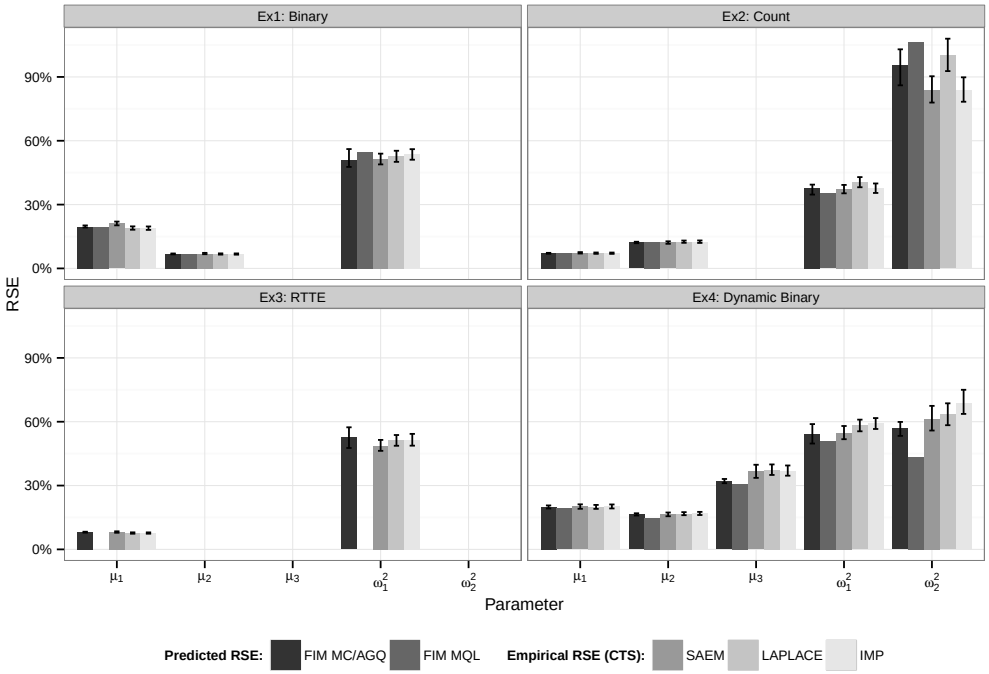

Figure 4: Relative standard errors (RSE) for the four simulated examples as predicted through the MC/AGQ Fisher information matrix (FIM) with $Q = 3$ and $S = 2000$, in comparison with RSE predicted from a marginal quasi-likelihood approximation (MQL) of the FIM as well as the empirical RSE obtained through clinical trial simulations (CTS) using three different estimation algorithms. Error bars indicate a 95% confidence interval of the RSE obtained by repetition for MC/AGQ and bootstrap for CTS.

For fixed effects, Figure 5 illustrates the performance of the MC/AGQ method when varying design variables of example 4 in three different scenarios: total study size, observations per subject, and constant total number of observations. The figure shows good agreement between FIM predicted RSE and the empirical RSE from CTS for the whole range of design variables. Nevertheless, for less informative designs (small study size or few observations per subject) the predictions from FIM are optimistic (as expected from the asymptotic nature of the FIM). Still, even for a study size of 20 subjects (i.e. 10

Table 2: P^{th} root of the determinant of the expected (E) and observed (O) Fisher information matrix, with P equal to the total number of parameters in the model.

	Method	Binary	Count	RTTE	Dynamic Binary
E	MC/AGQ	85.2	385.8	501.5	8.1
E	MQL	81.8	273.3	-	9.1
O	SAEM	80.7	300.6	516.4	7.1
O	LAPLACE	84.3	263.3	540.4	6.9
O	IMP	83.7	298.7	541.8	6.6

subjects in each group) there is rather good agreement between MC/AGQ FIM predicted RSE and CTS-based RSE.

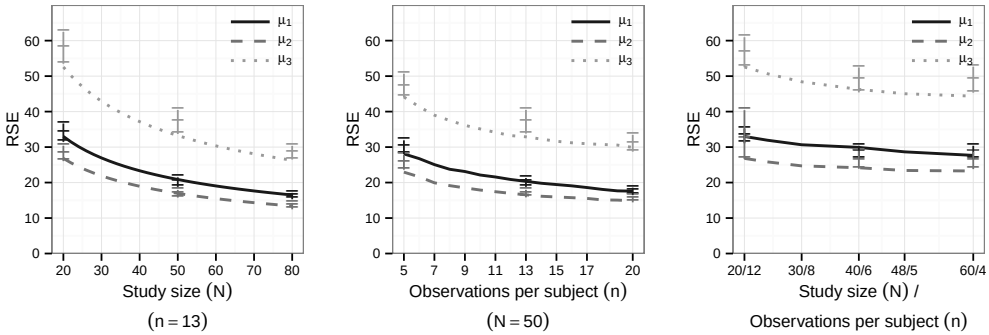

Figure 5: Expected relative standard errors (RSE) from MC/AGQ ($Q = 3, S = 2000$) for the fixed effects of example 4 (longitudinal binary) as a function of varying total study sizes (left, $n = 13$ observations per subject), varying number of observations per subject (middle, study size of $N = 50$) and varying both with constant number of total observations (right, $n_{tot} = 240$ observations). The median of the error bars indicates the empirical RSE obtained through clinical trial simulations and the whiskers indicate their 95% confidence interval obtained by bootstrap.

3.4. Computational complexity evaluation

For a practical use of an algorithm, runtime or computational complexity is one of the most important properties. In this section, the computational complexity of the MC/AGQ algorithm is compared with CTS and an MQL-based approximation of the FIM.

3.4.1. Methods – Computational complexity evaluation

The comparison of CTS and MC/AGQ in terms of time is complicated by the difference in implementation, i.e., while the three estimation algorithms used for the CTS are implemented in fast low-level programming languages (`C++` for MONOLIX and `FORTRAN 90` for NONMEM), the two FIM-based methods are implemented in R, which is comparatively slow. Therefore, rather than comparing algorithm runtimes, the computational complexity was compared based on the number of evaluations of the conditional

probability function for the observations given in equation 1. A formal evaluation of this quantity (big O notation) is challenging since some of the algorithms require multiple optimization steps, which is why the actual number of evaluations was recorded and used for the comparison. The details of how this was done for the MQL method are given in Appendix B.

Algorithm settings were set to the values used to obtain the results visualized in Figure 4[1]. Furthermore, for this comparison the CTS were run with full finite difference gradient evaluation (as used for MC/AGQ and the MQL approximation) as the number of conditional probability function evaluations is also dependent on the method used to calculate the gradients (especially for the Laplace algorithm).

3.4.2. Results – Computational complexity evaluation

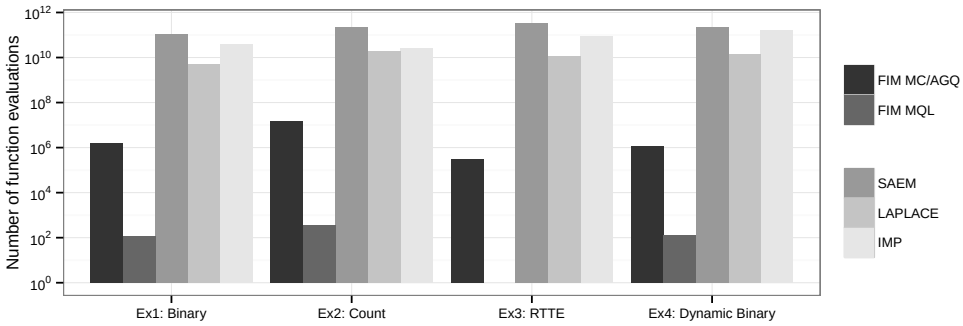

Figure 6: Number of conditional probability evaluations for the observations required to obtain relative standard errors with a precision as shown in Figure 4.

The results of the computational complexity comparison is visualized in Figure 6. While the absolute number of evaluations is a function of the model (number of fixed and random effects) as well as the design (number of subjects and observations per subject), and hence differed between examples, the relative difference between methods was the same across examples. The MQL-based evaluation of the FIM clearly required the fewest evaluations (approximately 10^2), followed by MC/AGQ (10^6 to 10^7 evaluations). The CTS-based methods required a significantly larger number of conditional probability evaluations (between 10^{10} and 10^{12}).

4. Application of the MC/AGQ method

Section 3.3.3 also highlighted the ability of the MC/AGQ method to predict the RSE for different design variables. This section illustrates a potential application, by studying the influence of the design on power to detect a group effect under cost constraints [25]

[1]The number of conditional probability evaluations for the SAEM algorithm was obtained with the corresponding NONMEM method. While differences between MONOLIX and NONMEM are expected, the order of magnitude is likely to be the same.

(p. 16). The design variables considered in this example are study size (N) and number of observations per subject (n).

In this case, the example is theoretical, but the chosen setting easily extends to clinical studies with binary outcomes such as when performing a responder/non-responder analysis or when studying the ACR20 response in rheumatoid arthritis [26](ACR20 is a binary endpoint and a response of one corresponds to a 20% improvement on the American College of Rheumatology scale) .

4.1. Method

Model formulation and parameter values from evaluation example 4 are reutilized here. Hence, a null and alternative hypothesis for the test of a group effect on the slope can be formulated as

$$H_0: \ \mu_3 = 0 \quad H_1: \mu_3 \neq 0, \tag{43}$$

where μ_3 is the group effect parameter. In this setting, the expected power for a Wald test is a function of the design (Ξ) given by

$$\pi(\Xi) = 1 - F_{\chi^2}(\chi^2_{1-\alpha,k}, k, \lambda(\Xi)), \tag{44}$$

where F_{χ^2} is the cumulative distribution function of the non-central chi-square distribution, $\chi^2_{1-\alpha,k}$ is the $1-\alpha$ quantile of the chi-square distribution with $k = 1$ degrees of freedom [27], and the non-centrality parameter $\lambda(\Xi)$ is given through the expected Wald statistic

$$\lambda(\Xi) = \frac{\mu_3^2}{\mathrm{SE}\,(\mu_3)^2}, \tag{45}$$

where $\mathrm{SE}\,(\mu_3)$ is the expected estimation SE of μ_3 as determined from the inverse of the FIM for a specific design. Similar to example 4, observations are assumed to be taken between 0 and 1 time units with equal spacing between observations, but varying the number of subjects (N) and number of observations per subject (n).

Study costs are assumed to involve recruitment costs (c_r) and observation costs (c_o). For fixed c_r and c_o, the overall study costs are therefore a function of the number of subjects (N) and the number of observations per subject (n), i.e.,

$$c(N, n) = N \cdot n \cdot c_o + N \cdot c_r. \tag{46}$$

For the sake of this example, a cost constraint of 12,000 units will be considered with an expected recruitment cost of $c_r = 100$ units per subject and costs of $c_o = 10$ units per observation. Assuming a true group effect of $\mu_3 = 0.4$, the design is constrained to have at least 80% power ($\alpha = 0.05$) and costs of at most 12,000 units. The two 'best' designs will be determined: the most powerful design (within the given cost constraint) ξ_π and the design with lowest costs (and at least 80% power) ξ_c.

4.2. Results

The influence of the number of subjects and the number of observations per subject on the expected power to detect a group effect as well as on the expected study costs is

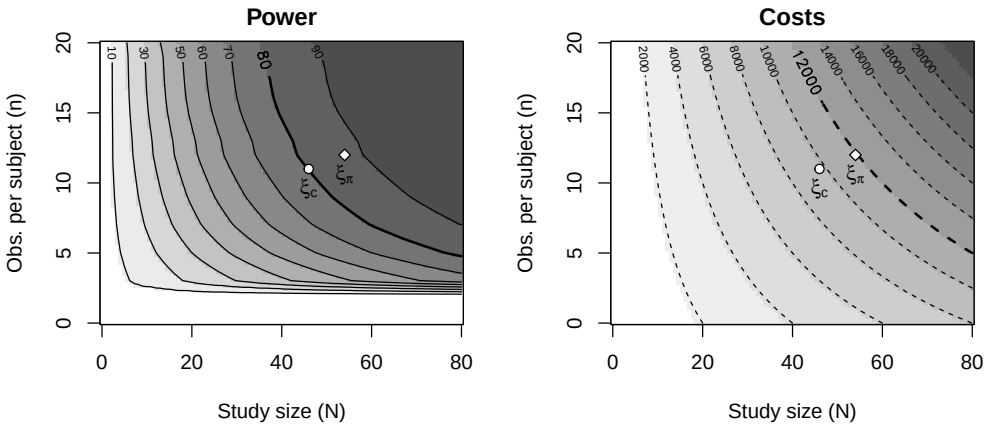

Figure 7: Contour plots of the expected power to detect a group effect (left panel) and expected study costs (right panel) as a function of study size (N) and number of observations per subject (n) for the trial designs from the application example. The design constraints (at least 80% power and costs of at most 12,000) are indicated as contours plotted with thicker lines and the designs with maximal power ξ_π as well as with minimal costs are ξ_c are indicted through $\diamond$ and $\bigcirc$, respectively.

Table 3: Study size (N), observations per subject (n), power (π) and costs (c) for the most powerful (ξ_π) and the least expensive design (ξ_c) from the application example

	N	n	π	c
ξ_π	54	12	87.8	11,880
ξ_c	46	11	80.2	9,660

visualized in Figure 7. The graph shows a contour plot of power and cost as a function of the number of subjects (N) and the number of observations per subject (n). It illustrates that specific power and cost levels are obtainable with multiple combinations of the two design variables. For example, an expected power of 80% can be achieved with a study size of 40 and 20 observations per subject ($40 \cdot 20 = 800$ observations in total) or with 80 subjects and 5 observations per subject ($80 \cdot 5 = 400$ in total). This illustrates that by balancing study size and number of observations per subject, the total number of observations can be reduced by a factor of 2, while maintaining the level of expected power. This non-trivial relationship between the two design variables and expected power is the result of the dependence between observations from the same subject, introduced through the random effects. The evaluation of the study design is further complexified in the present example through the cost constraint, resulting in the two 'best' designs ξ_π and ξ_c. Both designs, are annotated in Figure 7 and their details are listed in Table 3.

5. Discussion

In this work a new method for the evaluation of the FIM for discrete mixed effect models was presented. Unlike previous approaches, this method does not use any likelihood approximation and is based on the derivatives of the exact conditional likelihood.

MC simulations and AGQ are subsequently used for the numerical computation of the FIM.

The evaluation of the algorithm was performed through a simulation study as well as in comparison with a competing approximation due to the lack of a closed-form solution for discrete NLMEM. For the simulation study, we selected three diverse estimation algorithms to obtain more agnostic reference values. For the four models investigated in this work the expected RSE were in very good agreement with the CTS results for most parameters. Notable deviations between empirical RSE and MC/AGQ-based RSE were found only for the RSE obtained for the random effect variance ω_2^2 of the count example and the RSE for ω_2^2 of the dynamic example. For these parameters, larger differences between the three estimation algorithms could be observed, indicating that estimation might be challenging. Furthermore, it should be noted that the deviations found in this investigation are of minor importance from a practical perspective as their magnitude was high only for parameters with already high RSE. The computational complexity evaluation revealed the drastically lower effort needed by the MC/AGQ method compared with CTS. This difference in computational effort is a major advantage of optimal experimental design methodology over CTS and, hence, is essential for adoption in practice.

The second comparison performed in this manuscript compared the MC/AGQ-based approximation of the FIM with the MQL-based approximation suggested by Ogungbenro and Aarons [7]. Advantages of the MC/AGQ method were visible only for random effect variances. However, it is important to highlight that two of the four examples were taken from the publication of Ogungbenro and Aarons, and a good performance of the method for those examples is therefore not surprising. For more complex examples with more parameters and sparse design the advantages in accurate prediction of RSE by the MC/AGQ method are expected to be greater. This has been shown for NLMEM with normally distributed observations, for example, by Nguyen and Mentré [9], but remains to be verified for discrete NLMEM. The computational effort for the MQL-based approximation of the FIM is several orders of magnitude lower than for the MC/AGQ-based approximation. It should be noted, however, that the MQL-based approximation requires a different formula for different types of observation (see section Appendix B) to account for the difference in marginal distribution (most importantly the marginal variance). This calculation of the distribution of the observations is also where a big part of the computational effort of the MC/AGQ method goes. One can therefore interpret the reduction in computational effort for the MQL-based method as being 'bought' by replacing machine through human effort. The universal applicability of the MC/AGQ method is in our opinion also its biggest advantage. It merely requires formulas for data simulation and likelihood calculation to apply optimal design-based methods.

Considering the agreement between FIM predictions and CTS for the first two examples, the approach presented appears also to perform similarly to the method by Nyberg et al. [8], the second existing approach for calculating the FIM for discrete data NLMEM (examples 1 and 2 are shared between the evaluations). However, the method presented here requires considerably less computational effort than the second order-based approximation by Nyberg et al.. Another advantage of the MC/AGQ method is the possibility of increasing the number of quadrature nodes to improve the approximation of the marginal likelihood integral. As a result, a more precise prediction of the parameter precision than through the approximation by Nyberg et al. is expected for more complex examples.

The MC/AGQ method is in its overall concept identical to the one presented by Nguyen and Mentré [9] for normally distributed observations. However, it represents a considerable improvement in the sense that it is adopted to GLMM as well as discrete data NLMEM. Furthermore, our method uses an improved AGQ method by employing both centering and scaling, compared with only centering as done by Nguyen and Mentré. Finally, instead of centering the quadrature grid at the simulated value of the random effect parameters, we center the grid at the mode of the joint likelihood. This results in better approximation of the integrals over the random effects and allows drastic reduction in the number of quadrature grid points. In an evaluation for one of the examples from the work of Nguyen and Mentré, our algorithm achieved a comparable precision with 100 times (10 in each random effect dimension) fewer quadrature points than the algorithm presented by Nguyen and Mentré (the details of this evaluation are presented in Appendix C).

An application example was used to highlight the utility of the proposed algorithm. It utilizes the expected parameter precision from the MC/AGQ approximated FIM to calculate the expected power to detect a group effect and combines it with a cost evaluation. Despite the relatively simple structure of the model and the restriction to two design variables, the example illustrates the complexity of designing an experiment for discrete mixed effect models and stresses the usefulness of carefully studying the design with the FIM.

One limitation of the approach presentedis a consequence of the curse of dimensionality. The number of model evaluations increases exponentially with the number of random effects in the model, as the same number of quadrature grid points is needed in each dimension of the random effect vector. While this is still manageable for models with 2 and 3 random effects, the computational effort quickly becomes infeasible for more complex models. Possible alternative integration techniques that avoid this drawback are discussed in the following paragraph, but it is important to note that for categorical data models the number of random effects is often low due to the lower amount of information in the data. A second issue of this approach is its stochastic nature due to the MC sampling involved. This makes it harder to employ deterministic optimization algorithms such as Fedorov-Wynn [28, 29]. A possible solution is either the increase of the number of MC samples or the use of a stochastic optimization algorithm (e.g., simulated annealing [30] or stochastic gradient [31]).

Future perspectives for this work involve the investigation of alternative approaches to calculate i) the expectation over the data (equation (14)) and ii) the integrals over the random effects (equation (20)) more efficiently. For the former, variance reduction techniques such as Latin hypercube or importance sampling are envisionable [32]. A promising alternative technique for the evaluation of the integrals over the random effects is the use of quasi-Monte Carlo sampling [33]. This integration method avoids the exponential growth of the number of quadrature points as the number of random effects increases and has a higher accuracy than classical MC methods with same number of function evaluations.

Finally, it will be important to make the methods developed in this work available to potential users, thereby it is hoped further increasing the use of optimal experimental design for discrete data mixed effect models.

Acknowledgements

This work was supported by the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115156, resources of which are composed of financial contributions from the European Unions Seventh Framework Program (FP7/2007 2013) and EFPIA companies in-kind contribution. The DDMoRe project is also supported by financial contribution from academic and SME partners. This work does not necessarily represent the view of all DDMoRe partners.

This work has also received funding from the European Unions 7th Framework Programme for research, technological development and demonstration under Grant Agreement no. 602552.

The authors thank Thu Thuy Nguyen for help regarding the results presented in the appendix.

Bibliography

[1] J. Nyberg, C. Bazzoli, K. Ogungbenro, A. Aliev, S. Leonov, S. Duffull, A. C. Hooker, F. Mentré, Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies, British Journal of Clinical Pharmacology 79 (1) (2015) 6–17. doi:10.1111/bcp.12352.

[2] F. B. Tekle, F. E. S. Tan, M. P. F. Berger, Maximin D-optimal designs for binary longitudinal responses, Computational Statistics & Data Analysis 52 (12) (2008) 5253–5262. doi:10.1016/j.csda.2008.04.037.

[3] M. Niaparast, On optimal design for a Poisson regression model with random intercept, Statistics & Probability Letters 79 (6) (2009) 741–747. doi:10.1016/j.spl.2008.10.035.

[4] D. C. Woods, P. van de Ven, Blocked designs for experiments with correlated non-normal response, Technometrics 53 (2) (2011) 173–182. doi:10.1198/TECH.2011.09197.

[5] T. W. Waite, Design of experiments with mixed effects and discrete responses plus related topics, Thesis for the degree of doctor of philosophy, University of Southampton (2012).

[6] T. W. Waite, D. C. Woods, Designs for generalized linear models with random block effects via information matrix approximations, Biometrika 102 (3) (2015) 677–693. doi:10.1093/biomet/asv005.

[7] K. Ogungbenro, L. Aarons, Population Fisher information matrix and optimal design of discrete data responses in population pharmacodynamic experiments, Journal of Pharmacokinetics and Pharmacodynamics 38 (4) (2011) 449–469. doi:10.1007/s10928-011-9203-7.

[8] J. Nyberg, M. O. Karlsson, A. C. Hooker, Population optimal experimental design for discrete type data, in: PAGE 18 (2009), no. Abstr 1468.
URL `www.page-meeting.org/?abstract=1468`

[9] T. T. Nguyen, F. Mentré, Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature, Computational Statistics & Data Analysis 80 (2014) 57–69. doi:10.1016/j.csda.2014.06.011.

[10] G. K. Smyth, Nonlinear Regression, in: Encyclopedia of Environmetrics, John Wiley & Sons, Ltd, 2006.

[11] E. L. Plan, A. Maloney, F. Mentré, M. O. Karlsson, J. Bertrand, Performance comparison of various maximum likelihood nonlinear mixed-effects estimation methods for dose-response models, The AAPS Journal 14 (3) (2012) 420–432. doi:10.1208/s12248-012-9349-2.

[12] S. Jönsson, M. Kjellsson, M. Karlsson, Estimating bias in population parameters for some models for repeated measures ordinal data using NONMEM and NLMIXED, Journal of Pharmacokinetics and Pharmacodynamics 31 (4) (2004) 299–320.

[13] J. Geweke, Bayesian inference in econometric models using monte carlo integration, Econometrica: Journal of the Econometric Society (1989) 1317–1339.

[14] L. Tierney, J. B. Kadane, Accurate approximations for posterior moments and marginal densities, Journal of the American Statistical Association 81 (393) (1986) 82–86.

[15] E. Kuhn, M. Lavielle, Coupling a stochastic approximation version of EM with an MCMC procedure, ESAIM: Probability and Statistics 8 (2004) 115–131. doi:10.1051/ps:2004007.

[16] E. Plan, A. Maloney, I. Trocóniz, M. Karlsson, Performance in population models for count data, part i: maximum likelihood approximations, Journal of Pharmacokinetics and Pharmacodynamics 36 (4) (2009) 353–366. doi:10.1007/s10928-009-9126-8.
[17] K. E. Karlsson, E. L. Plan, M. O. Karlsson, Performance of three estimation methods in repeated time-to-event modeling, The AAPS Journal 13 (1) (2011) 83–91. doi:10.1208/s12248-010-9248-3.
[18] R. M. Savic, F. Mentré, M. Lavielle, Implementation and evaluation of the SAEM algorithm for longitudinal ordered categorical data with an illustration in pharmacokinetics–pharmacodynamics, The AAPS Journal 13 (1) (2010) 44–53. doi:10.1208/s12248-010-9238-5.
[19] R. Savic, M. Lavielle, Performance in population models for count data, part II: a new SAEM algorithm, Journal of Pharmacokinetics and Pharmacodynamics 36 (4) (2009) 367–379. doi:10.1007/s10928-009-9127-7.
[20] S. Särkkä, Bayesian Filtering and Smoothing, Cambridge University Press, 2013.
[21] G. H. Golub, J. H. Welsch, Calculation of gauss quadrature rules, Mathematics of Computation 23 (106) (1969) 221–230.
[22] G. Molenberghs, G. Verbeke, The generalized linear mixed model (GLMM), in: Models for Discrete Longitudinal Data, Springer Series in Statistics, Springer New York, 2005, pp. 265–280.
[23] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2014).
[24] G. Smyth, Y. Hu, P. Dunn, B. Phipson, Y. Chen, statmod: Statistical Modeling, R package version 1.4.20 (2014).
[25] V. V. Fedorov, S. L. Leonov, Optimal Design for Nonlinear Response Models, CRC Press, 2013, Ch. Optimal Design under Constraints, pp. 111–126.
[26] B. D. Lacroix, M. R. Lovern, A. Stockis, M. L. Sargentini-Maier, M. O. Karlsson, L. E. Friberg, A pharmacodynamic markov mixed-effects model for determining the effect of exposure to certolizumab pegol on the ACR20 score in patients with rheumatoid arthritis, Clinical Pharmacology & Therapeutics 86 (4) (2009) 387–395. doi:10.1038/clpt.2009.136.
[27] R. F. Engle, Wald, likelihood ratio, and Lagrange multiplier tests in econometrics, in: Zvi Griliches and Michael D. Intriligator (Ed.), Handbook of Econometrics, Vol. Volume 2, Elsevier, 1984, pp. 775–826.
[28] V. V. Fedorov, Theory of Optimal Experiments, Academic Press, 1972.
[29] H. P. Wynn, Results in the theory and construction of D-optimum experimental designs, Journal of the Royal Statistical Society. Series B (Methodological) 34 (2) (1972) 133–147.
[30] S. B. Duffull, S. Retout, F. Mentré, The use of simulated annealing for finding optimal population designs, Computer Methods and Programs in Biomedicine 69 (1) (2002) 25–35. doi:10.1016/S0169-2607(01)00178-X.
[31] M. Foracchia, A. Hooker, P. Vicini, A. Ruggeri, POPED, a software for optimal experiment design in population kinetics, Computer Methods and Programs in Biomedicine 74 (1) (2004) 29–46. doi:10.1016/S0169-2607(03)00073-7.
[32] J. S. Liu, Monte Carlo Strategies in Scientific Computing, Springer Series in Statistics, Springer New York, New York, NY, 2004.
[33] J. Pan, R. Thompson, Quasi-Monte Carlo estimation in generalized linear mixed models, Computational Statistics & Data Analysis 51 (12) (2007) 5765–5775. doi:10.1016/j.csda.2006.10.003.
[34] J. R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Third Edition, John Wiley & Sons, 2007, Ch. Some important differentials, pp. 167–172.

Appendix A. Derivative of the joint-likelihood

With the definition of the joint-likelihood as

$$l(y_i, \Psi, b_i) = \prod_{j=1}^{n_i} h(y_{ij}, \xi_i, g(\mu, b_i, z_i))p(b_i)$$

the derivative w.r.t. μ is

$$\begin{aligned}\frac{\partial}{\partial\mu}l(y_i,\Psi,b_i) &= \prod_{j=1}^{n_i} h(\cdot)\left(\sum_{j=1}^{n_i}\frac{\partial\log h(\cdot)}{\partial\mu}\right)p(b_i)\\ &= \prod_{j=1}^{n_i} h(\cdot)\left(\sum_{j=1}^{n_i}\frac{\partial\log h(\cdot)}{\partial g}\frac{\partial g(\cdot)}{\partial\mu}\right)p(b_i)\end{aligned}$$

and w.r.t. Ω_u is

$$\begin{aligned}\frac{\partial}{\partial\Omega_u}l(y_i,\Psi,b_i) =& \frac{\partial}{\partial\Omega_u}\left(\prod_{j=1}^{n_i} h(y_{ij},\xi_i,g(\mu,b_i,z_i))\phi(\Omega^{-\frac{1}{2}}b_i)|\Omega^{-\frac{1}{2}}|\right)\\ =& \prod_{j=1}^{n_i} h(\cdot)\left(|\Omega^{-\frac{1}{2}}|\operatorname{Tr}\left(\Omega^{\frac{1}{2}}\frac{\partial\Omega^{-\frac{1}{2}}}{\partial\Omega_u}\right)\phi(\Omega^{-\frac{1}{2}}b_i)+|\Omega^{-\frac{1}{2}}|(-\phi(\Omega^{-\frac{1}{2}}b_i)\Omega^{-\frac{1}{2}}b_i)\frac{\partial\Omega^{-\frac{1}{2}}}{\partial\Omega_u}b_i\right)\\ =& \prod_{j=1}^{n_i} h(\cdot)\left(\operatorname{Tr}\left(\Omega^{\frac{1}{2}}(-\Omega^{-\frac{1}{2}})\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}\right)+\left(-\Omega^{-\frac{1}{2}}b_i\right)(-\Omega^{-\frac{1}{2}})\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}b_i\right)\\ &\cdot\phi(\Omega^{-\frac{1}{2}}b_i)|\Omega^{-\frac{1}{2}}|\\ =& \prod_{j=1}^{n_i} h(\cdot)\left(\Omega^{-\frac{1}{2}}b_i\Omega^{-\frac{1}{2}}\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}b_i-\operatorname{Tr}\left(\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}\right)\right)\phi(\Omega^{-\frac{1}{2}}b_i)|\Omega^{-\frac{1}{2}}|\\ =& \prod_{j=1}^{n_i} h(\cdot)\left(\Omega^{-\frac{1}{2}}b_i\Omega^{-\frac{1}{2}}\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}b_i-\operatorname{Tr}\left(\frac{\partial\Omega^{\frac{1}{2}}}{\partial\Omega_u}\Omega^{-\frac{1}{2}}\right)\right)p(b_i)\end{aligned}$$

where $h(\cdot)=h(y_{ij},\xi_i,g(\mu,b_i,z_i))$, $g(\cdot)=g(\mu,b_i,z_i)$ and $\phi(\cdot)$ denotes the p.d.f. of the standard multivariate normal distribution with mean 0 and variance matrix I [34] (p. 23).

Appendix B. Methods – MQL approximation of the FIM

Following Ogungbenro and Aarons [7], the elementary information matrix $\mathcal{I}(\Psi,\xi_i)$ was evaluated as

$$\mathcal{I}(\Psi,\xi_i)=\begin{pmatrix} J_i^\mathsf{T}V_i^{-1}J_i & 0\\ 0 & \frac{1}{2}\left[\operatorname{Tr}\left\{Z_{i,(k)}^\mathsf{T}V^{-1}Z_{i,(l)}(Z_{i,(k)}^\mathsf{T}V_i^{-1}Z_{i,(l)})^\mathsf{T}\right\}\right]_{0\le k,l\le r}\end{pmatrix}$$

with

$$V_i=Z_i\Omega Z_i^\mathsf{T}+W_i$$

where $Z_{i,(l)}$ denotes column l of Z_i. The matrices J_i, Z_i and W_i are all model-dependent and for the binary examples 1 and 4 are defined as

$$J_{ij} = \frac{\partial h(y_{ij}, \xi, g(\mu, b_i, z_i))}{\partial \mu}\Big|_{\substack{y_{ij}=1 \\ b_i=0}}$$
$$Z_{ij} = \frac{\partial h(y_{ij}, \xi, g(\mu, b_i, z_i))}{\partial b_i}\Big|_{\substack{y_{ij}=1 \\ b_i=0}}$$
$$W_i = \mathrm{diag}(p_{i1}(1-p_{i1}), \ldots, p_{ij}(1-p_{ij}))$$

with

$$p_{ij} = h(y_{ij}, \xi, g(\mu, b_i, z_i))\Big|_{\substack{y_{ij}=1 \\ b_i=0}}.$$

For the count model in example 3 these matrices were defined as

$$J_{ij} = \frac{\partial \lambda_{ij}}{\partial \mu}\Big|_{b_i=0}$$
$$Z_{ij} = \frac{\partial \lambda_{ij}}{\partial b_i}\Big|_{b_i=0}$$
$$W_i = \mathrm{diag}(\lambda_{i1}, \ldots, \lambda_{ij}).$$

The derivatives needed to calculate J_{ij} and Z_{ij} were approximated using forward finite differencing according to the formula

$$\frac{\partial f(x)}{\partial x} = \frac{f(x+h) - f(x)}{h}$$

with $h = 10^{-6}$.

For the computational complexity comparison, the number of function evaluations of h for the binary and λ for the count example were considered for the MQL method, i.e. the number of evaluations needed to calculate W_i as well as J_{ij} and Z_{ij} using the finite difference method were added together.

Appendix C. Evaluation of the AGQ adaptation strategy

For the MC/AGQ method presented in this manuscript, the quadrature grid was centered at the mode and scaled by the second derivative of the joint likelihood. This is in contrast to the work of Nguyen and Mentré where the simulated value of the random effect was used for the centering and no scaling was performed [9]. The influence of the adaptation strategy on the RSE prediction for the example M3S of the work by Nguyen and Mentré is compared in Figure C.1. The example used the model

$$y_{ij} = 5 + \frac{\mu_1 \exp(b_1) d_j^3}{\mu_2 \exp(b_2)^3 + d_j^3} + \varepsilon_{ij} \tag{C.1}$$

where ε_{ij} was assumed to follow a normal distribution with mean 0 and standard deviation σ. The parameter values were $\mu = (30, 500)$, $\omega_1^2 = \omega_2^2 = 0.09$, and $\sigma = 2$. The

design consisted of two doses d per individual with $d = (300, 1000)$. Figure C.1 compares the three adaptation strategies: i) centering and scaling at the mode (CS MODE), ii) centering only at the mode (C MODE), and iii) centering at the simulated value of the random effect (C SIM). In the figure $\phi_{rel,k}$ for $k = 1, 2, 3$ (see section 3.2) is plotted as a function of the number of quadrature grid points Q.

Figure C.1 highlights the advantage of centering the quadrature grid at the mode rather than at the simulated value. While with CS MODE and C MODE stable results for ϕ_2 and ϕ_3 are already obtained with 3 grid points, C SIM requires more than 30. Figure C.1 also illustrates, most clearly for ϕ_1, how scaling further improves the quality of the approximation. For this criterion, in order to reach the same quality in prediction as CS MODE, both approaches without scaling require more than 10 times more quadrature grid points.

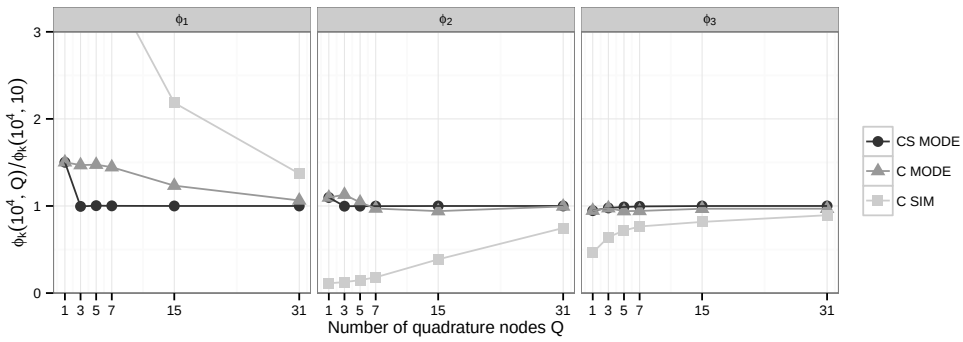

Figure C.1: Influence of the number of quadrature nodes Q and the adaptation strategy (CS MODE – centering and scaling at the mode, C MODE – centering only at the mode, and C SIM – centering at the simulated value of the random effect) on the normalized determinant of the Fisher information matrix (FIM) (ϕ_1, left panel), the trace of the fixed effect parameter part of the inverse of the FIM (ϕ_2, middle panel), and the trace of the variance of the random effect part of the inverse of the FIM (ϕ_3, right panel) for example M3S of Nguyen and Mentré [9].