Journal of Statistical Planning and Inference 141 (2011) 1394–1402

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Sequential D-optimal designs for generalized linear mixed models

Sanjoy K. Sinha [a,*], Xiaojian Xu [b]

[a] School of Mathematics and Statistics, Carleton University, Ottawa, Ontario, Canada
[b] Department of Mathematics, Brock University, St. Catharines, Ontario, Canada

ARTICLE INFO

Article history:
Received 2 February 2010
Received in revised form
1 August 2010
Accepted 19 October 2010
Available online 26 October 2010

Keywords:
D-optimality
Generalized linear model
Longitudinal data
Mixed model
Optimal design

ABSTRACT

We discuss the construction of D-optimal sequential designs for the analysis of longitudinal data or repeated measurements using generalized linear mixed models (GLMMs). We investigate the performance of the design through a simulation study, which indicates that the proposed design can be very successful in improving the efficiency of the ML estimators in GLMMs relative to some common competitors. Our simulations also suggest that the usual normal-theory inference procedures remain valid under the sequential sampling schemes. We also present an example using real data obtained from a clinical study.

1. Introduction

The optimal choice of designs is important in many clinical studies. A common goal of the optimal design of an experiment is to estimate the model parameters as precisely as possible, while minimizing the cost associated with sampling individuals. For example, in a case–control study, patients may be observed repeatedly over a certain period of time. The repeated responses from a single patient are naturally correlated, and mixed effects models are commonly used to describe the correlation structure. The goal of the design technique is to determine optimal sampling times for patients so that precise estimates of the parameters can be obtained for a model defining the relationship between the response variable and other covariates in the study.

Optimal designs in the settings of linear and nonlinear mixed effects models were investigated by a number of authors. Stroud et al. (2001) studied optimal designs in population pharmacokinetic studies. They discussed the choice of optimal sampling times in a population pharmacokinetic study of the anticancer agent paclitaxel conducted by the Cancer and Leukemia Group B (CALGB). The authors proposed a simulation-based method for Bayesian optimal design in the CALGB study. Ouwens et al. (2002) considered optimal selection and allocation of time points in repeated measured experiments. They proposed a maxmin criterion for locally D-optimal designs, and demonstrated that for a large class of symmetric designs, the smallest relative efficiency over the model parameter space could be substantial. Han and Chaloner (2004) investigated a Bayesian experimental design for the analysis of nonlinear mixed effects models. The authors formulated the design problem as a decision of choosing the values of the predictor(s) based on a criterion that minimizes the posterior risk. Schmelter (2007) considered an approximate optimal design for the estimation of fixed effects parameters in a certain class of mixed models.

Dragalin et al. (2008) proposed an adaptive design for dose-finding in clinical trials with combination of two drugs when several responses can be observed simultaneously on each subject. They considered modeling the distribution of bivariate

* Corresponding author.
E-mail addresses: sinha@math.carleton.ca (S.K. Sinha), xxu@brocku.ca (X. Xu).

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.10.008

binary responses using the bivariate probit model, and described the adaptive designs in the framework of the optimal design theory. Adewale and Xu (2010) discussed robust static designs for generalized linear models with protection for possible departures from the usual model assumptions, such as inaccuracy in an assumed linear predictor, overdispersion, and misspecification of the link function.

Although the choice of optimal designs in the framework of linear and nonlinear mixed models for continuous responses has been studied extensively in the literature, very few studies have been done for optimal designs in the context of generalized linear mixed models (GLMM), and our contribution of constructing D-optimal sequential designs will break completely new ground. GLMMs are commonly used in the analysis of clustered correlated discrete binary and count data including longitudinal data or repeated measurements. GLMMs are also useful for accommodating the overdispersion often observed among nonnormally distributed responses and for modeling the dependence among responses inherent in longitudinal or repeated measures data by incorporating random effects (Stiratelli et al., 1984; Zeger et al., 1988; Breslow and Clayton, 1993). It is usually assumed that the random effects have a multivariate normal distribution whose variance components are to be estimated from the data. A full maximum likelihood (ML) analysis based on the joint marginal likelihood of the responses can be used for estimating both fixed and random effects parameters in GLMMs.

In this article, we propose and explore techniques for the design of experiments, where the design issue is formulated as a decision of choosing the values of the predictor(s) for GLMMs. The mean response in GLMMs is generally nonlinear, and the usual measures of performance of a design depend on the parameters being estimated. A sequential approach is then naturally suggested—one should choose design points so as to maximize a measure of performance evaluated at the estimates obtained from observations made at previous design points. See Fedorov (1972, Chapter 4).

Chaudhuri and Mykland (1993) considered such problems in the framework of linear and nonlinear models for independent data in which a static initial design was to be augmented by a sequentially chosen set of design points. Subsequent design points were to be chosen so as to maximize the determinant of the Fisher information matrix for both fixed and random effects parameters, evaluated at the current parameter estimates. Assuming that the fitted response was a member of the chosen parametric family and that the random errors were homoscedastic, they demonstrated that a sequence of the maximum likelihood estimators of the regression parameters was asymptotically normal, and that the sequence of designs was asymptotically D-optimal, in the sense of maximizing the determinant of the true Fisher information matrix.

Sinha and Wiens (2002) also studied design problems in the framework of an approximate nonlinear regression model, and investigated sequential design methodologies when the fitted model was possibly of an incorrect parametric form. The authors considered approximating the mean response in the underlying regression model $y = E(y|\mathbf{x}) + \varepsilon$ as $E(y|\mathbf{x}) \approx f(\mathbf{x}, \boldsymbol{\theta}_0)$, for a certain function f and parameter vector $\boldsymbol{\theta}_0$. The random errors ε were assumed to be uncorrelated, with variances possibly dependent on $\mathbf{x}$. They fitted the approximate model $y = f(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon$ by ordinary least squares, or more generally, by M-estimation, but chose design points sequentially using the I-optimality criterion with the possibility of heteroscedasticity in mind, and with an awareness of the approximate nature of the fitted model.

In this paper, we extend the D-optimal sequential design methodologies to the case of a generalized linear mixed model for dependent data. We provide the Fisher information for the MLEs of both fixed effects parameters and random effects parameters. An adaptive method of solving the maximum likelihood estimating equations is given, and some comparative studies of our resulting designs and commonly used uniform designs are presented for both cross-sectional and longitudinal cases. The resulting designs offer better precision in the estimation for both fixed effects parameters and random effects parameters.

The paper is organized as follows. In Section 2, we introduce the model and notation to define the D-optimal design criterion for GLMMs. In Section 3, we present a simple example to illustrate the computational issues inherent in the choice of sequential designs. In Section 4, we present results from a simulation study, and demonstrate that the proposed design is useful in improving the efficiency of the ML estimators in GLMMs. We also investigate the usual normal-theory inference properties of the ML estimators under sequential designs in the simulations. Section 5 presents an application of the proposed design using some real data from a clinical experiment. Section 6 concludes the paper with some discussion.

2. Model and notation

Suppose conditional on the vector of random effects $\mathbf{u}$, the elements of the response vector $\mathbf{y} = (y_1, \ldots, y_n)^t$ are independently distributed and follow a distribution in the exponential family:

$$f_{y_i|u}(y|\mathbf{u}, \boldsymbol{\beta}, \phi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi)\right\} \tag{1}$$

for some functions a, b and c. Here the canonical parameter $\theta_i = \mathbf{x}_i^t\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}$, with $\mathbf{x}_i^t$ being the ith row of the design matrix $\mathbf{X}$ for the fixed effects and with $\mathbf{z}_i^t$ being the ith row of the design matrix $\mathbf{Z}$ for the random effects. We further assume that the vector of random effects $\mathbf{u}$ follows a distribution:

$$\mathbf{u} \sim f_u(\mathbf{u}|\boldsymbol{\alpha}) \tag{2}$$

depending on parameters $\boldsymbol{\alpha}$. For (1) and (2), the classical likelihood function can be defined as

$$L(\boldsymbol{\beta}, \phi, \boldsymbol{\alpha}|\mathbf{y}) = \int \prod_{i=1}^{n} f_{y_i|u}(y_i|\mathbf{u}, \boldsymbol{\beta}, \phi) f_u(\mathbf{u}|\boldsymbol{\alpha})\, d\mathbf{u}. \tag{3}$$

The ML estimators of the parameters $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\alpha}$ can be obtained by maximizing this likelihood function using suitable numerical techniques.

For simplicity, here we consider $\phi = 1$, as this is the case for binary and Poisson regression. Note that when the marginal distribution of $\mathbf{y}$ can be defined as a mixture as in (3), the classical ML estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ take the form

$$E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} = \mathbf{0} \tag{4}$$

and

$$E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\middle|\mathbf{y}\right\} = \mathbf{0}, \tag{5}$$

respectively, where the expectation is taken with respect to the conditional distribution of $\mathbf{u}$ given $\mathbf{y}$ (see McCulloch and Searle, 2001, for details). The ML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be obtained by solving the above equations using some iterative method such as the Newton–Raphson method as described in McCulloch (1997).

The observed Fisher information can be obtained in a matrix form as

$$\mathbf{I}_o(\boldsymbol{\beta},\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{I}_{o11}(\boldsymbol{\beta},\boldsymbol{\alpha}) & \mathbf{I}_{o12}(\boldsymbol{\beta},\boldsymbol{\alpha}) \\ \mathbf{I}_{o21}(\boldsymbol{\beta},\boldsymbol{\alpha}) & \mathbf{I}_{o22}(\boldsymbol{\beta},\boldsymbol{\alpha}) \end{bmatrix}, \tag{6}$$

where

$$\mathbf{I}_{o11}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t} = E\left\{\frac{\partial^2 \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\} + E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\} - E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\}, \tag{7}$$

$$\mathbf{I}_{o22}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\partial^2 \log L}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^t} = E\left\{\frac{\partial^2 \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\} + E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\} - E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\}, \tag{8}$$

and

$$\mathbf{I}_{o12}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \mathbf{I}_{o21}^t(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}^t} = E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\ \frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\} - E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\}. \tag{9}$$

It can be shown that for the exponential family (1),

$$E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} = \mathbf{X}^t[\mathbf{y} - E\{\boldsymbol{\mu}(\boldsymbol{\beta},\mathbf{u})|\mathbf{y}\}] \tag{10}$$

and

$$E\left\{\frac{\partial^2 \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\} = -E\{\mathbf{X}^t\mathbf{W}(\boldsymbol{\beta},\mathbf{u})\mathbf{X}|\mathbf{y}\}, \tag{11}$$

where $\mathbf{W}(\boldsymbol{\beta},\mathbf{u}) = \mathrm{diag}\{\mathrm{var}(y_i|\mathbf{u})\}$. The expected Fisher information can be obtained by taking the marginal expectations of the expressions (6)–(9) with respect to the response vector $\mathbf{y}$. After simplification, we can show that

$$E\left\{-\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t}\right\} = E\left[E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\}\right], \tag{12}$$

$$E\left\{-\frac{\partial^2 \log L}{\partial \boldsymbol{\alpha}\partial \boldsymbol{\alpha}^t}\right\} = E\left[E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\}\right], \tag{13}$$

and

$$E\left\{-\frac{\partial^2 \log L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\alpha}^t}\right\} = E\left[E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{u}(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t}\middle|\mathbf{y}\right\}\right]. \tag{14}$$

Similarly to Sinha and Wiens (2002), we adopt a sequential approach to the D-optimal design problem. We assume that the experimenter initially obtains data $\{y_i; i = 1,\ldots,n_0\}$ from a group of n_0 individuals measured at n_0 locations (or design points) $\{\mathbf{x}_j; j = 1,\ldots,n_0\}$ determined in advance. Then subsequent data are obtained at locations $\mathbf{x}_j (j = n_0+1, n_0+2,\ldots)$ intended to maximize the determinant of the expected Fisher information matrix. The algorithm for choosing the sequential D-optimal designs can be described as follows:

1. For the initial data $\{(y_i,\mathbf{x}_i); i = 1,\ldots,n_0\}$, find the ML estimates of the model parameters $\boldsymbol{\gamma} = (\boldsymbol{\beta}',\boldsymbol{\alpha}')'$. Call these initial estimates $\hat{\boldsymbol{\gamma}}_0 = (\hat{\boldsymbol{\beta}}_0'\hat{\boldsymbol{\alpha}}_0)'$.

2. Using these initial estimates, evaluate the Fisher information matrix

$$\mathbf{I}(\hat{\gamma}_0) = E\left(-\frac{\partial^2 \log L}{\partial \gamma \partial \gamma^t}\right)_{\gamma = \hat{\gamma}_0}. \tag{15}$$

3. Choose a new design point $\mathbf{x}^*_{n_0+1}$ by maximizing the determinant of the Fisher information matrix $\mathbf{I}(\mathbf{x}_{n_0+1}|\hat{\gamma}_0)$ with respect to $\mathbf{x}_{n_0+1}$:

$$\mathbf{x}^*_{n_0+1} = \underset{\mathbf{x}_{n_0+1}}{\operatorname{argmax}} |\mathbf{I}(\mathbf{x}_{n_0+1}|\hat{\gamma}_0)|. \tag{16}$$

4. Update the parameter estimates for the augmented data obtained at the new design point $\mathbf{x}^*_{n_0+1}$. Obtain the next sequential design point based on the new set of estimates, and so on.

We choose n_1 design points $\mathbf{x}^*_{n_0+1}, \ldots, \mathbf{x}^*_{n_0+n_1}$ sequentially using this algorithm. In the next section, we present a simple example to illustrate the computational issues involving the calculation of the expected Fisher information.

3. Illustrative example: a binary mixed model

Suppose in a dose-finding clinical study, the experimenter initially obtains n_0 repeated measurements from each of k patients treated with n_0 different doses $\{x_j; j = 1, \ldots, n_0\}$ of a drug over a certain period determined in advance. Assume that the n_0 repeated measurement $\{y_{ij}, j = 1, \ldots, n_0\}$ on the ith patient follow a simple binary mixed model with a single random effect and a single fixed effect:

$$\begin{aligned} &y_{ij}|u_i \sim \text{independent Bernoulli}(p_{ij}), \quad i = 1, \ldots, k; \ \ j = 1, \ldots, n_0, \\ &\theta_{ij} = \log\{p_{ij}/(1-p_{ij})\} = \beta_0 + \beta_1 x_j + u_i, \\ &u_i \sim \text{independent } N(0, \sigma^2). \end{aligned} \tag{17}$$

In this setup, we have $E(y_{ij}|u_i) = \mu_{ij}(\boldsymbol{\beta}, u_i) = \exp(\beta_0 + \beta_1 x_j + u_i)/\{1 + \exp(\beta_0 + \beta_1 x_j + u_i)\}$ and $\text{var}(y_{ij}|u_i) = \mu_{ij}(\boldsymbol{\beta}, u_i)\{1 - \mu_{ij}(\boldsymbol{\beta}, u_i)\}$. To estimate the regression parameters $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, we can solve the ML estimating equations (4) by an iterative method. The Newton–Raphson iterative equation takes the form

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + [E\{\mathbf{X}^t \mathbf{W}(\boldsymbol{\beta}^{(m)}, \mathbf{u})\mathbf{X}|\mathbf{y}\}]^{-1} \mathbf{X}^t [\mathbf{y} - E\{\boldsymbol{\mu}(\beta^{(m)}, \mathbf{u})|\mathbf{y}\}] \tag{18}$$

for $m = 1, 2, 3, \ldots$. The variance component σ^2 can be estimated simultaneously from the iterative equation

$$\sigma^{2(m+1)} = \frac{1}{k} \sum_{i=1}^{k} E(u_i^2|\mathbf{y}_i), \tag{19}$$

where $\mathbf{y}_i$ is the response vector $(y_{i1}, \ldots, y_{in_0})^t$ for the ith subject.

To calculate the Fisher information matrix, we can show that

$$E\left(-\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t}\right) = E\left[E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t}\middle|\mathbf{y}\right\}\right] = E[\mathbf{X}^t[\mathbf{y} - E\{\boldsymbol{\mu}(\boldsymbol{\beta}^{(m)}, \mathbf{u})|\mathbf{y}\}][\mathbf{y} - E\{\boldsymbol{\mu}(\boldsymbol{\beta}^{(m)}, \mathbf{u})|\mathbf{y}\}]^t \mathbf{X}], \tag{20}$$

$$E\left(-\frac{\partial^2 \log L}{\partial (\sigma^2)^2}\right) = E\left[\left[E\left\{\frac{\partial \log f_u(\mathbf{u}|\sigma^2)}{\partial \sigma^2}\middle|\mathbf{y}\right\}\right]^2\right] = \left(\frac{k}{2\sigma^4}\right)^2 E\left[\left\{\frac{1}{k}\sum_{i=1}^{k} E(u_i^2|\mathbf{y}_i) - \sigma^2\right\}^2\right] = \left(\frac{1}{2\sigma^4}\right)^2 \sum_{i=1}^{k} \text{var}\{E(u_i^2|\mathbf{y}_i)\}, \tag{21}$$

and

$$E\left(-\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \sigma^2}\right) = E\left[E\left\{\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\middle|\mathbf{y}\right\} E\left\{\frac{\partial \log f_u(\mathbf{u}|\sigma^2)}{\partial \sigma^2}\middle|\mathbf{y}\right\}\right] = E\left[\mathbf{X}^t[\mathbf{y} - E\{\boldsymbol{\mu}(\boldsymbol{\beta}^{(m)}, \mathbf{u})|\mathbf{y}\}] \frac{k}{2\sigma^4}\left\{\frac{1}{k}\sum_{i=1}^{k} E(u_i^2|\mathbf{y}_i) - \sigma^2\right\}\right]. \tag{22}$$

For the given initial data $\{(y_{ij}, x_j); i = 1, \ldots, k; j = 1, \ldots, n_0\}$ at n_0 design points (or doses), we find the initial ML estimates $\hat{\gamma}_0 = (\hat{\boldsymbol{\beta}}_0', \hat{\sigma}_0^2)'$ of the model parameters $\gamma = (\boldsymbol{\beta}', \sigma^2)'$. Based on these initial estimates, we then obtain n_1 new design points x_j $(j = n_0 + 1, n_0 + 2, \ldots, n_0 + n_1)$, with $n = n_0 + n_1$, by using the sequential D-optimal design scheme as described in Section 2. We can consider obtaining repeated observations at the n_1 new locations from a fixed set of k patients (or subjects) determined in advance. Alternatively, additional data can be obtained from a group of new patients measured repeatedly at the n_1 sequential design points.

4. Simulation study

We ran a series of simulations to explore the properties of the ML estimators obtained under the sequential design scheme. We compare two classes of designs:

1. The sequentially determined adaptive designs described in the previous section. The performance is analyzed after the addition of k observations, at a chosen location x_j^*, for $j = n_0+1, \ldots, n_0+n_1$. We choose x_j^* by numerically maximizing the determinant of the Fisher information matrix $\mathbf{I}(x_j|\gamma)$.
2. The "uniform" designs, with the initial design augmented by an additional k observations at each of $n-n_0$ locations. In these designs, the experimental points are uniformly distributed throughout the design space. In the case of uniform designs, when $n = n_0+n_1$ the locations are equally spaced over [0,3]; for smaller values of n they form a subset of these sites. Thus these designs are sequential but nonadaptive.

Data were generated from the binary mixed model (17) with the values of the parameters fixed at $(\beta_0,\beta_1) = (-2,2)$ and $\sigma^2 = 0.5$. The initial data were obtained from $k = 50$ individuals (referred to as clusters) with each individual being observed at four equally spaced design points $x = 0,1,2$, and 3. Then under the sequential design scheme, data were augmented by adding binary responses from $k = 50$ clusters observed at each sequentially chosen design point from the design space $x \in [0,3]$. Fig. 1 exhibits a plot of the determinant of the Fisher information against the design x for a typical set of initial data generated from the mixed model (17). It is clear that the next sequential design point chosen by the D-optimal design scheme is x=0.6. We obtained additional data at three new locations under both sequential and uniform design schemes.

In Fig. 2, we exhibit the plot of the average, over 500 simulation runs, of the determinant of the Fisher information against the number of sequential points at which additional data were obtained. It is clear from the figure that the sequential design out-performed the uniform design, as expected.

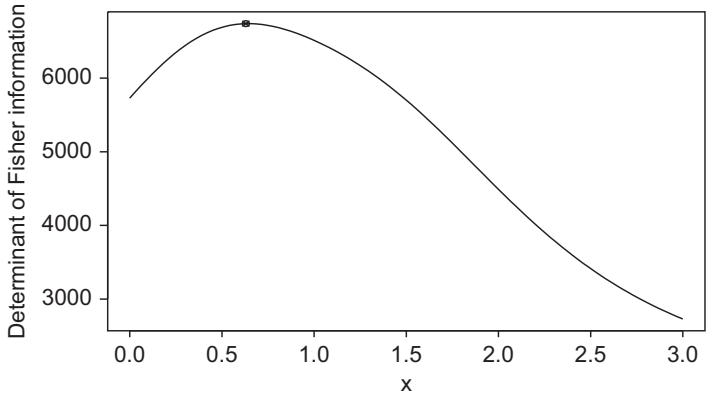

Fig. 1. Determinant of Fisher information versus design x for data from a binary mixed model. Next design point is at x=0.6.

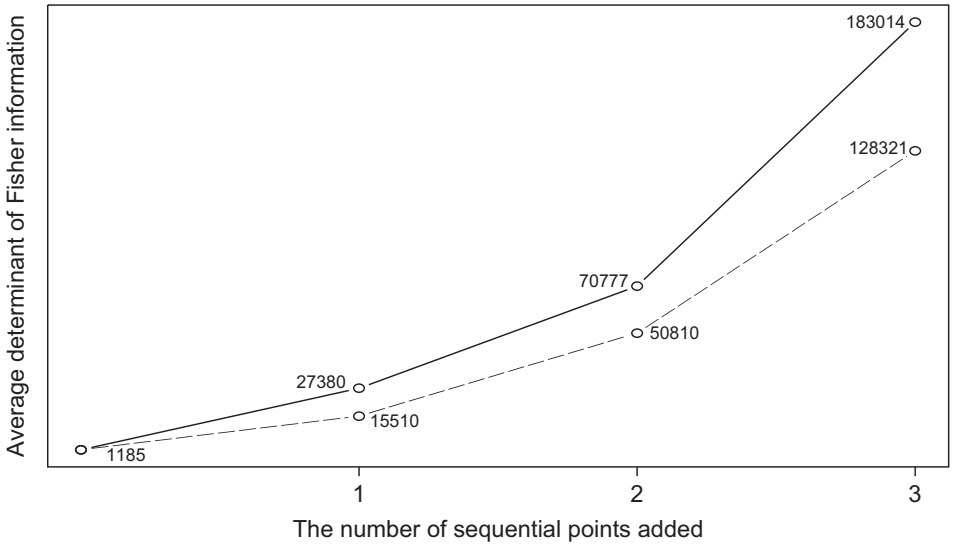

Fig. 2. Average (over 500 simulation runs) of the determinant of Fisher information under sequential (—) and uniform (- - -) design schemes for a binary mixed model.

In Fig. 3, we exhibit the histogram of the sequentially chosen design points over the 500 runs. We observe that the location of the first sequential design is approximately 0.75, whereas the locations of the second and third sequential designs are 1.75 and 0.5, respectively.

We also study the asymptotic normality properties of the estimators obtained under the two design schemes. Table 1 reports the empirical coverages (over 500 simulation runs) and their standard errors, and the mean lengths and standard errors of these means for 95% individual normal confidence intervals on the model parameters. It is clear that the coverage probabilities are similar under both designs, but the sequential design generally provides smaller lengths of the confidence intervals. The coverage probability for the variance component σ^2 appears to be smaller than the nominal 95% confidence level under both designs, which is mainly due to the bias in the estimator of σ^2 for small sample sizes considered here. We find (not shown here) that when the number of clusters, k, increases, the empirical coverages for σ^2 move closer to the nominal 95% confidence level.

We also studied the properties of the ML estimators under the D-optimal and uniform design schemes in the case of a Poisson mixed model for count data. We consider a simple Poisson mixed model in the form

$$y_{ij}|u_i \sim \text{independent Poisson}(\lambda_{ij}), \quad i=1,\ldots,k;\ j=1,\ldots,n_0,$$

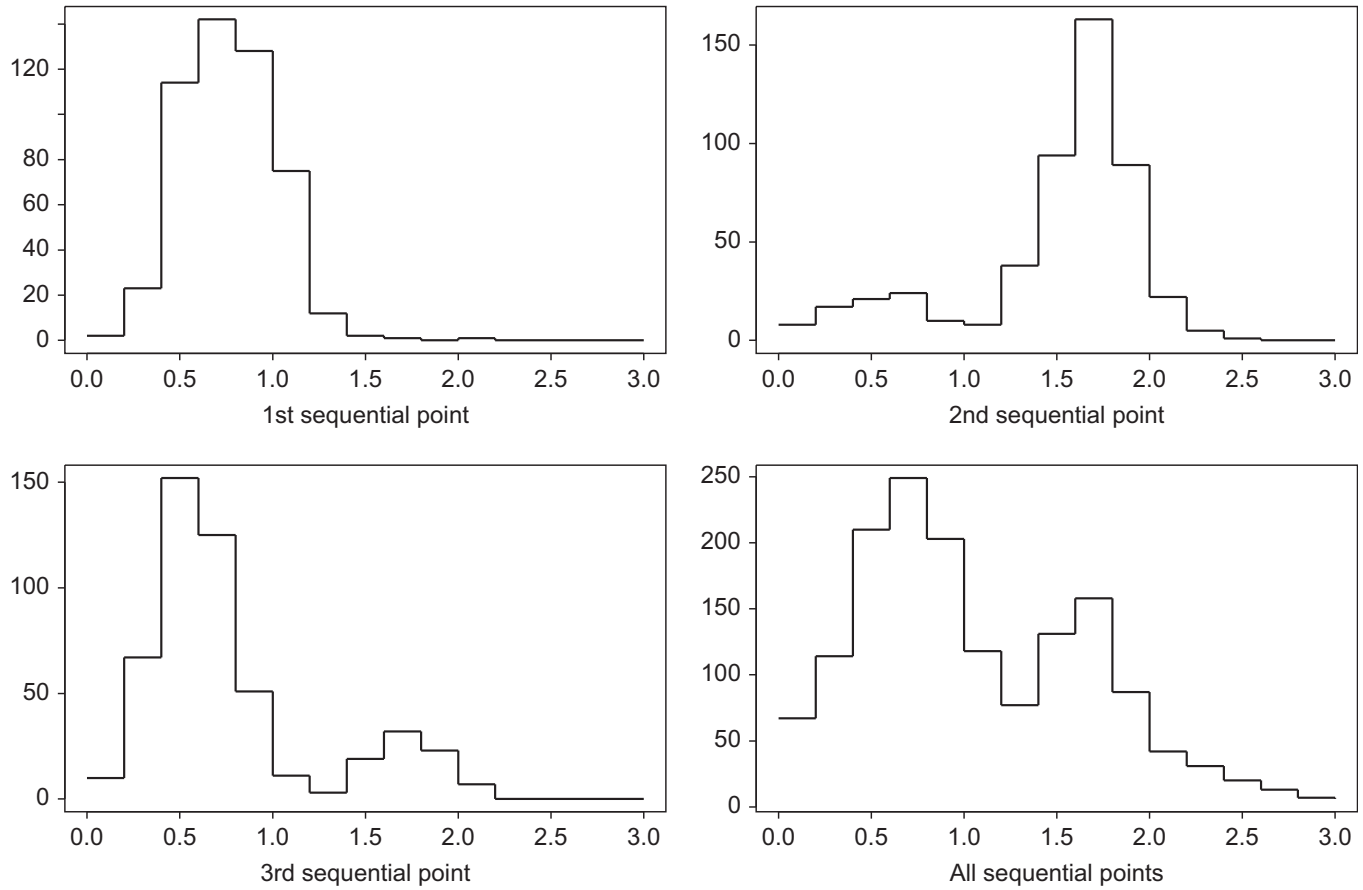

Fig. 3. Histogram of optimal sequential design points for a binary mixed model chosen over 500 simulation runs.

Table 1
Empirical coverages and lengths of 95% confidence intervals on the parameters of a binary mixed model (standard errors in parentheses).

Design	$n-n_0$	β_0		β_1		σ^2	
		Coverage	Length	Coverage	Length	Coverage	Length
D-optimal	1	0.960 (0.0088)	1.473 (0.0121)	0.952 (0.0096)	1.214 (0.0108)	0.880 (0.0145)	2.584 (0.0815)
	2	0.962 (0.0086)	1.238 (0.0072)	0.960 (0.0088)	1.002 (0.0059)	0.882 (0.0144)	1.297 (0.0214)
	3	0.978 (0.0066)	1.081 (0.0058)	0.968 (0.0079)	0.895 (0.0055)	0.894 (0.0138)	0.983 (0.0164)
Uniform	1	0.944 (0.0103)	1.594 (0.0151)	0.936 (0.0109)	1.230 (0.0138)	0.854 (0.0158)	2.839 (0.1027)
	2	0.948 (0.0099)	1.341 (0.0103)	0.938 (0.0108)	1.007 (0.0086)	0.856 (0.0157)	1.657 (0.0462)
	3	0.958 (0.0090)	1.174 (0.0076)	0.948 (0.0099)	0.896 (0.0068)	0.886 (0.0142)	1.161 (0.0283)

Augmented data were obtained from sequential D-optimal and uniform designs.

$$\theta_{ij} = \log(\lambda_{ij}) = \beta_0 + \beta_1 z_i + \beta_2 x_j + \beta_3 z_i x_j + u_i,$$

$$u_i \sim \text{independent } N(0,\sigma^2), \tag{23}$$

where z_i is a subject-specific binary covariate, which is 1 if individual i has a particular disease, and is 0 if the individual is from an otherwise healthy group. We consider $k=50$ individuals in the study in which 25 individuals correspond to the disease group ($z_i=0$), and the remaining 25 individuals correspond to the healthy group ($z_i=1$). The covariate x_j represents the jth dose of a drug that is given to each individual.

Data were generated from the binary mixed model (23) with the values of the parameters fixed at $(\beta_0,\beta_1,\beta_2,\beta_3) = (0.25, 0.25, -0.25, -0.10)$ and $\sigma^2 = 0.25$. The initial data were obtained from $k = 50$ individuals with each individual being observed at $n_0=2$ design points $x=0$ and 10. Then under the sequential design scheme, data were augmented by adding Poisson responses from $k=50$ clusters observed at each sequentially chosen design point from the design space $x \in [0,10]$. Fig. 4 exhibits a plot of the determinant of the Fisher information against the design x for a typical set of initial data generated from the Poisson mixed model (23). It is clear that the next sequential design point chosen by the D-optimal design scheme is $x=2$, whereas the uniform design scheme chooses the design point $x=5$.

As Poisson mixed models use count data, fitting such models requires much more computation as compared to binary mixed models. So in the simulations, we use 100 replications of datasets. As before, for each replication, we first generate initial data and find the initial estimates of the model parameters. Based on these initial estimates, we choose a single design point based on the D-optimal and the uniform design schemes. Fig. 5 displays the histogram of the D-optimal design points chosen over the 100 simulation runs. We observe that the D-optimal design points are centered at the values 0, 2, and 6,

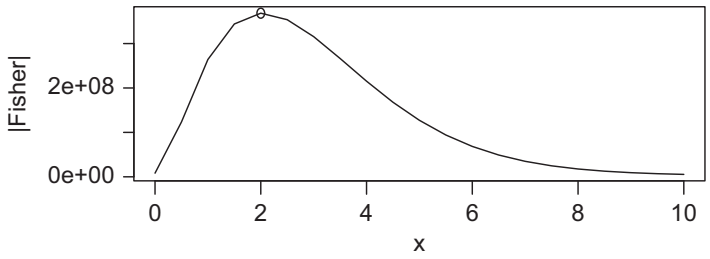

Fig. 4. Determinant of Fisher information versus design x for a typical set of count data from a Poisson mixed model. Next design point is at $x=2$.

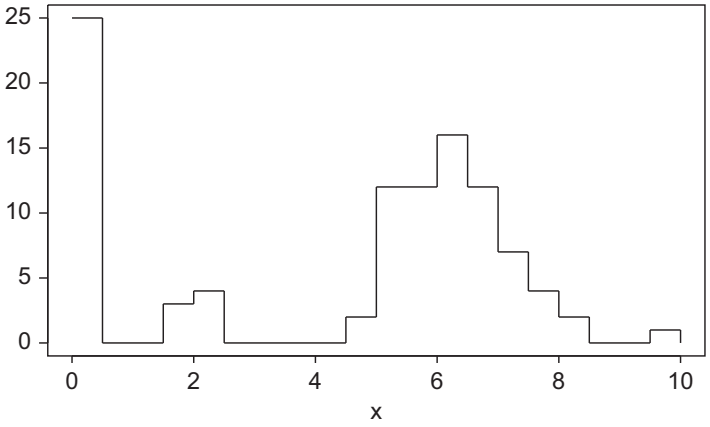

Fig. 5. Histogram of optimal design points for a Poisson mixed model chosen over 100 simulation runs.

Table 2
Empirical coverages and lengths of 95% confidence intervals on the parameters of a Poisson mixed model (standard errors in parentheses).

Coefficients	D-optimal		Uniform	
	Coverage	Length	Coverage	Length
β_0	0.88 (0.0325)	0.7484 (0.0063)	0.87 (0.0336)	0.7920 (0.0100)
β_1	0.89 (0.0313)	0.9747 (0.0058)	0.92 (0.0271)	1.0330 (0.0117)
β_2	0.88 (0.0325)	0.2103 (0.0019)	0.82 (0.0384)	0.3496 (0.1530)
β_3	0.89 (0.0313)	0.3740 (0.0197)	0.89 (0.0313)	0.6271 (0.2113)
σ^2	0.89 (0.0313)	0.5652 (0.0040)	0.94 (0.0237)	0.5927 (0.0180)

Augmented data were obtained from D-optimal and uniform designs.

whereas the uniform design chooses the value 5. The average of the determinants of Fisher information matrices under the D-optimal design was obtained as exp(22.98), whereas the average under the uniform design was obtained as exp(22.76). Here also the D-optimal design outperforms the uniform design, as expected.

Table 2 reports the empirical coverages and their standard errors, and the mean lengths and standard errors of these means for 95% individual normal confidence intervals on the model parameters. We observe that both designs provide coverages that are generally lower than the nominal 95% confidence level. But here also the D-optimal design provides uniformly smaller length as compared to the uniform designs. We find (not shown here) that the empirical coverages move closer to the nominal level for larger sample size.

5. An example: epilepsy data

Thall and Vail (1990) presented and analyzed data from a clinical trial of 59 epileptics who were randomized to receive either the antiepileptic drug progabide (trt=1) or a placebo (trt=0), as an adjuvant to standard chemotherapy. In the initial experiment, the number of seizures during the two weeks before each of the four clinic visits was reported. The covariates considered in the analysis were the number of baseline seizure count recorded in the preceding eight-week period, age in years at entry into the trial, the binary indicators for the progabide (treatment) group, and visit for the four clinic visits.

Here we revisit the epilepsy data and consider analyzing them using a binary response variable y, which is defined to be 0 if the number of seizures during a two-week period is less than or equal 10, and y is 1, otherwise. The covariates considered in the analysis are: baseline seizures, "base", which is defined to be 0 if the number of baseline seizures in the preceding eight-week period is less than or equal to 40, and it is 1, otherwise; binary indicators, "trt", for the treatment group; and "visit" (=2, 4, 6, and 8) representing the times (in weeks) of the four clinic visits. We consider a simple binary logistic regression model for the conditional mean response, $p_{it} = E(y_{it}|u_i)$, in the form

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = \beta_0 + \beta_1\ \text{base}_i + \beta_2\ \text{trt}_i + \beta_3\ \text{visit}_t + u_i, \tag{24}$$

for $i=1,\ldots,k$ with $k=59$ patients, and $t=1,\ldots,4$ time points. Here u_i represent the subject-specific random effects for the patients, which are assumed to be independent normal with mean 0 and variance component σ^2. The ML estimates of the model parameters are obtained as $(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\sigma}) = (-3.849, 6.912, -0.635, -0.286, 3.121)$ with the corresponding standard errors, s.e.$(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2,\hat{\beta}_3,\hat{\sigma}) = (1.439, 1.862, 1.214, 0.137, 0.945)$. The seizure count appears to be decreasing over time. The treatment (drug progabide) also reduces the number of seizures, but its effect is not statistically significant.

In the case of a recommendation by the analysts for further experimentation over a period of time, a natural question may arise: When should the next visit be taken place in order to obtain more efficient estimates of the model parameters? Fig. 6 provides estimate of the determinant of the Fisher information over the design space, visit $\in$ [9,25], in weeks. Assuming that the given model remains valid over this design space, the experimenter can plan the next visit after 23 weeks. However, if the experimenter wishes to take the measurements earlier, for example over the design space, visit $\in$ [9,16], then week 16 may be the optimal choice as indicated by the curve in Fig. 6.

An important issue in a sequential design scheme involves a stopping rule—the experimenter needs to determine how many new data points should be generated in order to obtain reliable estimates of the model parameters. This may depend on the experimenter's point of view as well as on available experimental facilities. Under adequate experimental facilities, the experimenter may generate new data based on the sequentially chosen design points until he/she is confident in the fitted

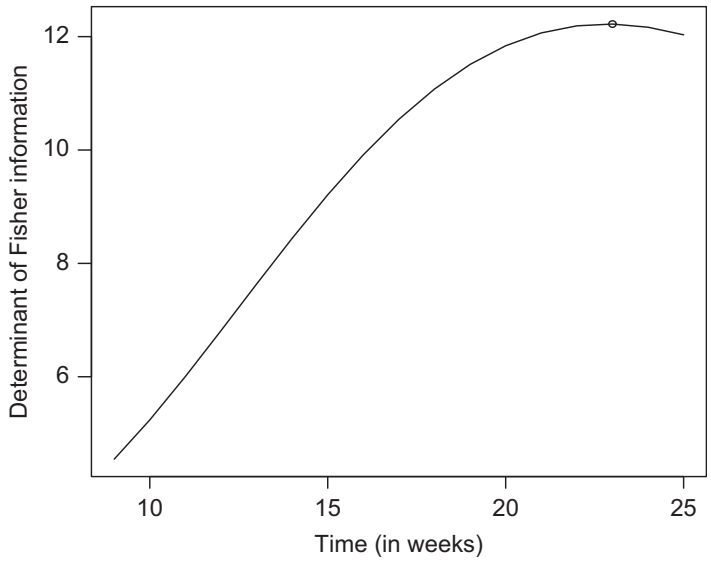

Fig. 6. Choice of optimal time for epilepsy data. Next visit at Time=23 weeks.

model, and has obtained satisfactory estimates of the model parameters. A workable rule may be to continue sampling until the maximized Fisher information, anticipated at the next design point, reaches above a preassigned threshold.

6. Discussion

We have studied the performance of the locally D-optimal sequential designs for analyzing generalized linear mixed models. We have demonstrated that one could attain considerable gain in efficiency from the maximum likelihood estimators when data are augmented with the sequential design scheme rather than the much simpler uniform design scheme. Although, in the simulations, we have explored the sequential design in the setting of repeated measurements from a fixed group of patients, this sequential approach can also be applied to other settings of GLMMs for dependent data. For example, we can consider obtaining additional data from new subjects (or patients) at each sequentially chosen design points. In such cases, patients from a given medical practice (or cluster) can be correlated, and GLMMs can be used to describe the correlation structure within a cluster.

We have explored the sequential design scheme in the case of a one-dimensional design space. However, the method can be extended for choosing multidimensional sequential designs in a similar manner. For example, similarly to Dragalin et al. (2008), we can consider adaptive sequential points over a two-dimensional design space for selecting dose combination of two drugs when bivariate binary responses representing efficacy and toxicity of a drug are available. Here we can obtain two-dimensional sequential designs by maximizing the determinant of the Fisher information matrix from a bivariate binary regression model for the two response variables.

Acknowledgements

The authors thank two reviewers for their useful comments and suggestions that led to significant improvements in the presentation. The research of both authors is partially supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Adewale, A.J., Xu, X., 2010. Robust designs for generalized linear models with possible overdispersion and misspecified link functions. Computational Statistics and Data Analysis 54, 875–890.
Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9–25.
Chaudhuri, P., Mykland, P., 1993. Nonlinear experiments: optimal design and inference based on likelihood. Journal of the American Statistical Association 88, 538–546.
Dragalin, V., Fedorov, V.V., Wu, Y., 2008. Adaptive designs for selecting drug combinations based on efficacy-toxicity response. Journal of Statistical Planning and Inference 138, 352–373.
Fedorov, V.V., 1972. Theory of Optimal Experiments. Academic Press, New York.
Han, C., Chaloner, K., 2004. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. Biometrics 60, 25–33.
McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association 92, 162–170.
McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models. John Wiley, New York.
Ouwens, M.J.N.M., Tan, F.E.S., Berger, M.P.F., 2002. Maximin D-optimal designs for longitudinal mixed effects models. Biometrics 58, 735–741.
Schmelter, T., 2007. The optimality of single-group designs for certain mixed models. Metrika 65, 183–193.
Sinha, S.K., Wiens, D.P., 2002. Robust sequential designs for nonlinear regression. The Canadian Journal of Statistics 30, 601–618.
Stiratelli, R., Laird, N., Ware, J., 1984. Random effects models for serial observations with binary responses. Biometrics 40, 961–971.
Stroud, J.R., Muller, P., Rosner, G.L., 2001. Optimal sampling times in population pharcokinetics studies. Applied Statistics 50, 345–359.
Thall, P.F., Vail, S.C., 1990. Some covariance models for longitudinal count data with overdispersion. Biometrics 46, 657–671.
Zeger, S.L., Liang, K.Y., Albert, P.S., 1988. Models for longitudinal data: a generalized estimating equation approach. Biometrics 44, 1049–1060.