Robust Design for Generalized Linear Mixed Models with Different Types of Misspecifications

Xiaojian Xu

Department of Mathematics and Statistics
Brock University
St. Catharines, Ontario, Canada

Joint research with Sanjoy Sinha, Carleton University

Aug. 8, 2017

BIRS Workshop on Latest Advances in the Theory and Applications of Design and Analysis of Experiments. Banff, Canada, Aug. 6-11, 2017

Outline

- Introduction
- Assumed Statistical Models
- Possible Departures
- Optimal Sequential Designs
- Simulations
- Discussion and Conclusion

Introduction

- Generalized linear mixed models (GLMMs) are often used for analyzing clustered correlated and repeated measures data.
- We explore techniques for design of experiments, where the optimal choices of predictors' values can be made in order to produce the most accurate estimation or prediction through a GLMM.
- We also consider the situations when the fitted GLMM is possibly of an incorrect parametric form.

Assumed Model and Notation

- Conditional on $\mathbf{u}$, elements of response vector $\mathbf{y} = (y_1, \ldots, y_N)^t$ are assumed independent and follow a distribution in exponential family:

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

  for some functions a, b and c.

- Canonical parameter $\theta_i = \mathbf{x}_i^t\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}$.
- $\mathbf{x}_i^t$ is ith row of the design matrix $\mathbf{X}$ for fixed effects and $\mathbf{z}_i^t$ is ith row of the design matrix $\mathbf{Z}$ for random effects.
- Further assume $\mathbf{u}$ follows a distribution:

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

  depending on parameters $\boldsymbol{\alpha}$.

Likelihood Function

- Likelihood function can be expressed as

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

- For simplicity, consider $\phi = 1$, as used in binary and Poisson regression.

Maximum Likelihood (ML) Estimation

- ML estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ take the form:

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

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

- Expectation is with respect to the conditional distribution of $\mathbf{u}|\mathbf{y}$.
- ML estimators are obtained using an iterative method.

Fisher Information

- Observed Fisher Information matrix is

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

  where

$$\begin{aligned} \mathbf{I}_{o11}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= -\frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \\ &= -E\left\{ \left. \frac{\partial^2 \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right| \mathbf{y} \right\} \\ &\quad -E\left\{ \left. \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t} \right| \mathbf{y} \right\} \\ &\quad +E\left\{ \left. \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right| \mathbf{y} \right\} E\left\{ \left. \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t} \right| \mathbf{y} \right\} \end{aligned}$$

Fisher Information

$$
\begin{aligned}
\mathbf{I}_{o22}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= -\frac{\partial^2 \log L}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^t} \\
&= -E\left\{ \left. \frac{\partial^2 \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^t} \right| \mathbf{y} \right\} \\
&\quad -E\left\{ \left. \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t} \right| \mathbf{y} \right\} \\
&\quad +E\left\{ \left. \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \right| \mathbf{y} \right\} E\left\{ \left. \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t} \right| \mathbf{y} \right\}
\end{aligned}
$$

Fisher Information

$$
\begin{aligned}
\mathbf{I}_{o12}(\boldsymbol{\beta},\boldsymbol{\alpha}) &= \mathbf{I}^t_{o21}(\boldsymbol{\beta},\boldsymbol{\alpha}) = -\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}^t} \\
&= -E\left\{ \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t} \middle| \mathbf{y} \right\} \\
&\quad + E\left\{ \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \middle| \mathbf{y} \right\} E\left\{ \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t} \middle| \mathbf{y} \right\}
\end{aligned}
$$

Fisher Information

- For the exponential family,

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

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

- $\mathbf{W}(\boldsymbol{\beta},\mathbf{u}) = \mathrm{diag}\{\mathrm{var}(y_i|\mathbf{u})\}$.

Fisher Information

- Expected Fisher Information matrix has components

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

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

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

Possible departures

- (M1) Misspecified linear predictors
- (M2) Overdispersion
- (M3) Imprecisions in the assumed link function
- (M4) Misspecified distribution for the random effects
- (M5) Combined departures

(M1) Misspecified linear predictors

- The covariates included in the systematic component of the model, the linear predictor, may not reflect the influence of covariates correctly.
- Possibly due to the use of a wrongly specified functional form of the covariates in the model or an omission of essential covariates.
- Impact - biased estimators
- For instance, 'substantial bias in the conditionally specified regression point estimators can result from using a simple random intercepts model when either the random effects distribution depends on measured covariates or there are autoregressive random effects' - Heagerty and Kurland (2001)

(M1) Misspecified linear predictors

- The unknown model belongs to a class

$$E\left(Y_i|\mathbf{u}\right) = \mu\left(\eta_{T,i}\right), \;\; \eta_{T,i} = \mathbf{x}_i^t\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + h(\mathbf{x}), \quad (1)$$

  where $h(\mathbf{x})$ is a 'small' contaminant, $\mathbf{u}\sim f_u(\mathbf{u}|\boldsymbol{\alpha})$, and $E\left(\mathbf{u}\right) = 0$.
- However, fits a GLMM model with $\mu^u\left(\eta_i|\mathbf{u}\right) = \mu\left(\mathbf{x}_i^t\beta + \mathbf{z}_i^t\mathbf{u}\right)$.
- Define the fitted marginal mean response

$$\mu\left(\mathbf{x}_i, \beta\right) = E\left(\mu^u\left(\eta_i|\mathbf{u}\right)\right) = \int_{\mathbf{u}} \mu\left(\mathbf{x}_i^t\beta + \mathbf{z}_i^t\mathbf{u}\right) f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u}$$

- Then, true marginal mean response:

$$E(Y_i|\mathbf{x}_i) = E\left(\mu^u\left(\eta_{T,i}|\mathbf{u}\right)\right) \;\; = \mu\left(\mathbf{x}_i, \beta\right) + d_i^{(1)}(\mathbf{x}_i, \boldsymbol{\beta}),$$

  where

$$d_i^{(1)}(\mathbf{x}_i, \boldsymbol{\beta}) = \int \left[\mu'(\mathbf{x}_i^t\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u})h(\mathbf{x}) + o(h(\mathbf{x}))\right] f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u}$$

(M1) Misspecified linear predictors

- We define "true" $\boldsymbol{\beta}_0$ through minimization of integrated squared discrepancy

$$\boldsymbol{\beta}_0 = \arg\min \int_S \left[d^{(1)}(\mathbf{x}, \boldsymbol{\beta}) \right]^2 d\mathbf{x}. \tag{2}$$

This implies that

$$\int_S \mathbf{t}(\mathbf{x}, \boldsymbol{\beta}_0) d^{(1)}(\mathbf{x}, \boldsymbol{\beta}_0) d\mathbf{x} = \mathbf{0},$$

with $\mathbf{t}(\mathbf{x}, \boldsymbol{\beta}) = \partial\mu(\mathbf{x}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$. Then, the true marginal mean is given by

$$E(Y_i|\mathbf{x}_i) = \mu_i(\mathbf{x}_i, \boldsymbol{\beta}_0) + d^{(1)}(\mathbf{x}_i, \boldsymbol{\beta}_0).$$

(M1) Misspecified linear predictors

- *Result 1: For a GLMM model with possible (M1) type misspecification, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_N$ of the model parameter vector $\boldsymbol{\beta}_0$* has the following property:

$$\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0 \sim \mathcal{N}\left(\mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\mathbf{b}_N^{(1)}(\boldsymbol{\beta}_0), \mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\mathbf{Q}_N^{(1)}(\boldsymbol{\beta}_0)\mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\right),$$

where

$$\mathbf{M}_N(\boldsymbol{\beta}) = \sum_{i=1}^{N} E\left[\frac{\partial}{\partial\boldsymbol{\beta}} E\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right],$$

$$\mathbf{b}_N^{(1)}(\boldsymbol{\beta}) = \sum_{i=1}^{N} d_i^{(1)}(\mathbf{x}_i, \boldsymbol{\beta})\mathbf{x}_i,$$

$$\mathbf{Q}_N^{(1)}(\boldsymbol{\beta}) = \sum_{i=1}^{N} E\left[E\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right. \\ \left. E^t\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right].$$

Relationship among departures (M1)-(M5)

- In summary, among (M1)-(M5), the overdispersion problem (M2) and misspecified random effects distribution problem (M4) mainly contribute to the variance part of the response, and the resulting mean responses approximately remain the same or relatively small;
- With some approximation, a (M2) problem can be treated as a (M4) problem.
- The link function misspecification problems (M3) can be all cast as linear predictor misspecification (M1) problems;
- A (M5) problem can be viewed as a combination of (M1) and (M4) problems.

(M5) Combined departures

- Assuming any GLMM with a random intercept, where overdispersion can be accommodated, any types of departures (M1)-(M4) and any possible combinations of them can be treated as a general case:
- Formally, with (M5), the true but unknown model belongs to a class of alternative models

$$E\left(Y_i|\mathbf{u}\right) = \mu\left(\eta_{T,i}\right),\ \eta_{T,i} = \mathbf{x}_i^t\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + h(\mathbf{x}), \tag{3}$$

$$\mathbf{u} \sim f_u^T\left(\mathbf{u}|\gamma\right),\ \text{and } E\left(\mathbf{u}\right) = 0. \tag{4}$$

- Fitting a GLMM with $\mu^u\left(\eta_i|\mathbf{u}\right) = \mu\left(\mathbf{x}_i^t\beta + \mathbf{z}_i^t\mathbf{u}\right),\ \mathbf{u}\sim f_u(\mathbf{u}|\boldsymbol{\alpha})$.

(M5) Combined departures

- We define $\mu(\mathbf{x}_i, \beta, \boldsymbol{\alpha}) = E(\mu^u(\eta_i|\mathbf{u})) = \int_{\mathbf{u}} \mu(\mathbf{x}_i^t \beta + \mathbf{z}_i^t \mathbf{u}) f_u(\mathbf{u}|\boldsymbol{\alpha}) d\mathbf{u}$ as the fitted marginal mean response. Then, the true marginal mean response:

$$
\begin{aligned}
E(Y_i|\mathbf{x}_i) &= \int \mu\left(\mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u} + h(\mathbf{x})\right) f_u^T(\mathbf{u}|\gamma)\, d\mathbf{u} \\
&= \mu(\mathbf{x}_i, \beta, \boldsymbol{\alpha}) + d^{(3)}(\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}),
\end{aligned}
$$

where

$$
\begin{aligned}
d^{(3)}(\mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\alpha}) &= \int \mu(\mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u}) \left[f_u^T(\mathbf{u}|\gamma) - f_u(\mathbf{u}|\boldsymbol{\alpha}) \right] d\mathbf{u} \\
&\quad + \int \left[\mu'(\mathbf{x}_i^t \boldsymbol{\beta} + \mathbf{z}_i^t \mathbf{u}) h(\mathbf{x}) + o(h(\mathbf{x})) \right] f_u^T(\mathbf{u}|\gamma)\, d\mathbf{u}.
\end{aligned}
$$

(M5) Combined departures

- In this case, we define 'true' $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$ as

$$\boldsymbol{\theta}_0 = \begin{pmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\alpha}_0 \end{pmatrix} = \arg\min \int_S \left[d^{(3)}(\mathbf{x}, \beta, \boldsymbol{\alpha}) \right]^2 d\mathbf{x}.$$

This implies

$$\int_S \mathbf{t}(\mathbf{x}, \boldsymbol{\theta}_0) d^{(3)}(\mathbf{x}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) d\mathbf{x} = \mathbf{0}.$$

Then, the true marginal mean is given by

$$E(Y_i|\mathbf{x}_i) = \mu_i(\mathbf{x}_i, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0) + d^{(3)}(\mathbf{x}_i, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0).$$

(M5) Combined departures

- *Result 3: For a GLMM model with possible (M5) type misspecification, the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_N$ of the model parameter vector $\boldsymbol{\beta}_0$* has the following property:

$$\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0 \sim \mathcal{N}\left(\mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\mathbf{b}_N^{(3)}(\boldsymbol{\beta}_0), \mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\mathbf{Q}_N^{(3)}(\boldsymbol{\beta}_0)\mathbf{M}_N^{-1}(\boldsymbol{\beta}_0)\right),$$

where

$$\mathbf{M}_N(\boldsymbol{\beta}) = \sum_{i=1}^{N} E\left[\frac{\partial}{\partial\boldsymbol{\beta}} E\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right], \ \mathbf{b}_N^{(2)}(\boldsymbol{\beta}) = \sum_{i=1}^{N} d_i^{(3)}(\mathbf{x}_i, \boldsymbol{\beta})\mathbf{x}_i,$$

and

$$\mathbf{Q}_N^{(3)}(\boldsymbol{\beta}) = \sum_{i=1}^{N} E_T\left[E\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\} E^t\left\{(y_i - \mu_i^u(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right].$$

Relationship among departures (M1)-(M5)

- The typical focus of the design for a generalized linear model is usually the prediction or estimation of mean response structure rather than variability cross clusters.
- Therefore, we will focus on optimal and robust designs for GLMMs in protection of possible imprecision in assumed linear predictor: Departure (M1).
- The design construction procedure will be the same for all of (M1)-(M5).
- Hereafter we adapt a general notation of $d(\mathbf{x}_i, \boldsymbol{\beta}_0)$ for the discrepancy between $E(Y_i|\mathbf{x}_i)$ and $\mu(\mathbf{x}_i, \boldsymbol{\beta}_0)$ instead of using $d^{(1)}(\mathbf{x}_i, \boldsymbol{\beta}_0)$ as we did for (M1).

Example: Binary model

- Suppose that an experimenter fits a logistic regression model by

$$\text{logit}\{\mu^u(x, u, \boldsymbol{\beta})\} = \beta_0 + \beta_1 x + u$$

over the range $x \in [0, 3]$; however, the true model is

$$\text{logit}\{\mu^{*u}(u, x, \boldsymbol{\alpha})\} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + u,$$

where $u \sim \mathcal{N}(0, 1)$.
Figure 1 (left) shows plots of marginal means $\mathrm{E}[\mu^{*u}(u, x, \boldsymbol{\alpha}_0)]$ with $\boldsymbol{\alpha}_0 = (-2, 1, 0.5)'$ and $\mathrm{E}\{\mu^u(x, u, \boldsymbol{\beta}_0)\}$ with $\boldsymbol{\beta}_0 = (-2.60, 2.23)'$ obtained by minimizing the squared distance

$$D^2(\boldsymbol{\beta}) = \int_0^3 \{\mathrm{E}(\mu^{*u}(u, x, \boldsymbol{\alpha}_0)) - \mathrm{E}(\mu^u(x, u, \boldsymbol{\beta}))\}^2 \, dx$$

between two marginal response functions.

Example: Poisson model

- Consider a Poisson regression model

$$\log\{\mu^u(x, u, \boldsymbol{\beta})\} = \beta_0 + \beta_1 x + u$$

over the range $x \in [0, 3]$, whereas the true mean response model is

$$\log\{\mu^{*u}(u, x, \boldsymbol{\alpha})\} = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + u$$

Figure 1 (right) shows plots of marginal means $\mathrm{E}[\mu^{*u}(u, x, \boldsymbol{\alpha}_0)]$ with $\boldsymbol{\alpha}_0 = (-1, 0.2, 0.2)'$ and $\mathrm{E}\{\mu^u(x, u, \boldsymbol{\beta}_0)\}$ with $\boldsymbol{\beta}_0 = (-1.80, 1.04)'$, the values minimize the overall squared discrepancy.

Example: Misspecified binary and Poisson models

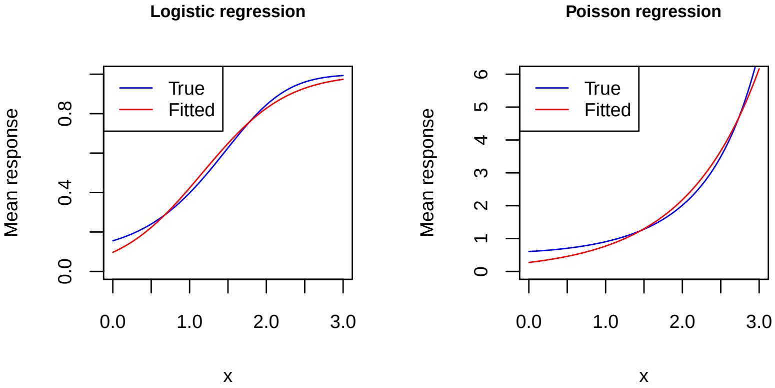

Figure: 1. True versus fitted models

Asymptotics

- ML estimator $\hat{\beta}_N$ follows the property:

$$\hat{\beta}_N - \beta_0 \sim \mathcal{N}\left(\mathbf{M}_N^{-1}(\beta_0)\mathbf{b}_N(\beta_0), \mathbf{M}_N^{-1}(\beta_0)\mathbf{Q}_N(\beta_0)\mathbf{M}_N^{-1}(\beta_0)\right),$$

where

$$\mathbf{M}_N(\beta) = \sum_{i=1}^{N} \mathrm{E}\left[\frac{\partial}{\partial \beta}\mathrm{E}\left\{(y_i - \mu_i^u(\mathbf{x}_i, \beta, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right]$$

$$\mathbf{Q}_N(\beta) = \sum_{i=1}^{N} \mathrm{E}\left[\mathrm{E}\left\{(y_i - \mu_i^u(\mathbf{x}_i, \beta, \mathbf{u}))\mathbf{x}_i | y_i\right\}\mathrm{E}'\left\{(y_i - \mu_i^u(\mathbf{x}_i, \beta, \mathbf{u}))\mathbf{x}_i | y_i\right\}\right]$$

$$\mathbf{b}_N(\beta) = \sum_{i=1}^{N} d_i(\mathbf{x}_i, \beta)\mathbf{x}_i.$$

Integrated mean squared error (IMSE)

- Define asymptotic IMSE by

$$
\begin{aligned}
&\int_S \mathrm{E}\left[\left\{\mu(\mathbf{x},\hat{\boldsymbol{\beta}}_N) - \mathrm{E}(y|\mathbf{x})\right\}^2\right] d\mathbf{x} \\
\approx\ & \int_S \mathrm{E}\left\{\left[(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}_0)'\mathbf{t}(\mathbf{x},\boldsymbol{\beta}_0) - d(\mathbf{x},\boldsymbol{\beta}_0)\right\}^2\right] d\mathbf{x} \\
=\ & \mathrm{trace}\{\mathrm{MSE}_N(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)\} + \int_S d^2(\mathbf{x},\boldsymbol{\beta}_0) d\mathbf{x},
\end{aligned}
$$

where $\mathbf{A}(\boldsymbol{\beta}) = \int_S \mathbf{t}(\mathbf{x},\boldsymbol{\beta})\mathbf{t}'(\mathbf{x},\boldsymbol{\beta}) d\mathbf{x}$.

Example: Estimation of discrepancy function d

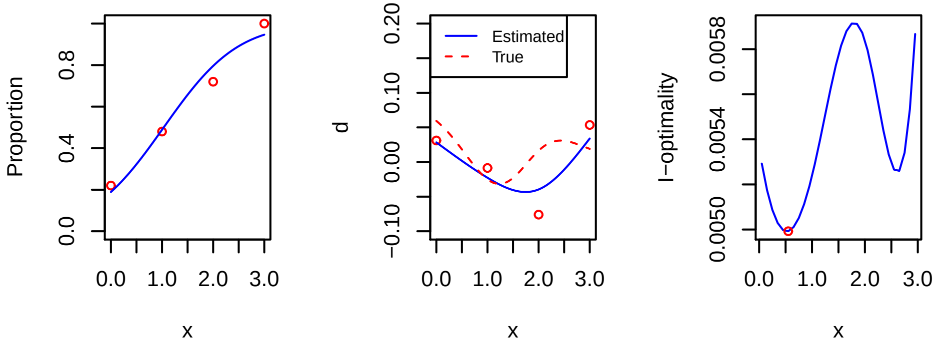

Figure: 2. Exact and estimated discrepancies, with the same parameter values as Figure 1 (left).

Integrated mean squared error (IMSE)

- We consider two design criteria, where design points are obtained by minimizing
  I-optimality criterion:
  $$\mathcal{L}_I = \text{trace}\{\text{MSE}_N(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)\}$$
  D-optimality criterion:
  $$\mathcal{L}_D = [\det\{\text{MSE}_N(\boldsymbol{\beta}_0)\}]^{1/p}$$

Introduction

- Usual measures of performance of a design in GLMMs depend on parameters being estimated.
- Commonly used methods: minimax (maximin), Bayesian (min-average), and sequential
- We adopt a sequential approach for choosing a design point that maximizes a measure of performance evaluated at estimates obtained from previous observations.
- The performance of the proposed design is assessed by simulations.
- Optimal sequential designs for nonlinear regression were studied earlier by Chaudhuri & Mykland (1993) and Sinha & Wiens (2002).
- More recently, Bayesian approaches considered: for nonlinear regression by Dror & Steinberg (2008); for Poisson regression by Zhang & Ye (2014), and for logistic regression with a random intercept by Maram & Jafari (2016).

I-optimal sequential design

1. Find ML estimates $\hat{\gamma}_0 = (\hat{\beta}_0^t, \hat{\alpha}_0^t)^t$ for initial $\{(y_i, \mathbf{x}_i);\ i = 1, \dots, N_0\}$.
2. Compute
$$\mathcal{L}_I = \text{trace}\{\text{MSE}_{N_0}(\hat{\beta}_0)\mathbf{A}(\hat{\beta}_0)\}$$
3. Choose a new design point $\mathbf{x}^*_{N_0+1}$ from
$$\mathbf{x}^*_{N_0+1} = \arg\min_{\mathbf{x}_{N_0+1}} \left[\text{trace}\{\text{MSE}_{N_0+1}(\hat{\beta}_0)\mathbf{A}(\hat{\beta}_0)\}\right]$$
4. Update estimates for augmented data obtained at $\mathbf{x}^*_{N_0+1}$. Obtain next sequential design point based on new set of estimates.
5. Choose N_1 design points $\mathbf{x}^*_{N_0+1}, \dots, \mathbf{x}^*_{N_0+N_1}$ sequentially.

Simulation Study - Logistic

- We study empirical properties using simulations
- Data were generated from a "true" binary mixed model:

$$y_{ij}|u_i \sim \text{ind. Bernoulli}(p_{ij}), i = 1, \ldots, k; j = 1, \ldots, n_0$$
$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 x_j + \alpha_2 x_j^2 + u_i$$
$$u_i \sim \text{ind. } \mathcal{N}(0, \sigma_u^2).$$

- Parameters were fixed at $\sigma_u^2 = 1$, $\alpha_0 = -2$, $\alpha_1 = 0.5$ and $\alpha_2 = (0.5, 1)$
- Fitted model: $\text{logit}(p_{ij}^*) = \beta_0 + \beta_1 x_j + u_i$
- Initial data are based on $n_0 = 7$ design points equally spaced in $[0, 3]$.
- Two design points were chosen sequentially using both I-optimal and D-optimal criteria.

Simulation Study - Logistic

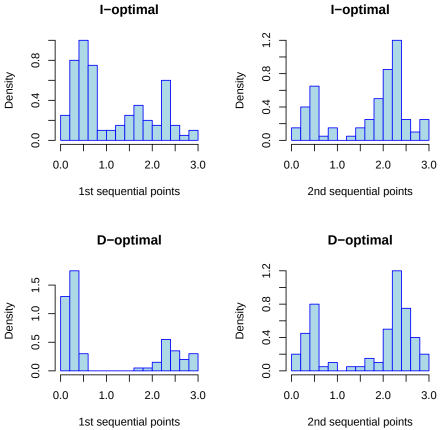

Figure: 3. Histogram of sequential points for the binary model. $k = 25$ clusters, $\alpha_2 = 0.5$.

Simulation Study - Logistic

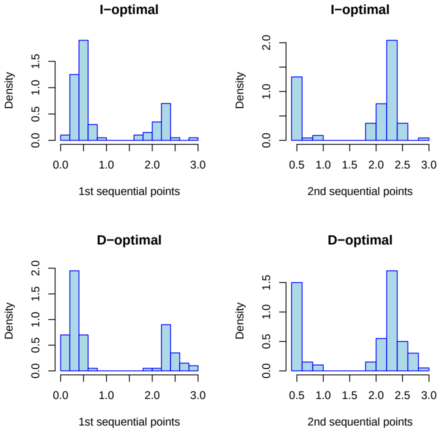

Figure: 4. Histogram of sequential points for the binary model. $k = 50$ clusters, $\alpha_2 = 0.5$.

Simulation Study - Logistic

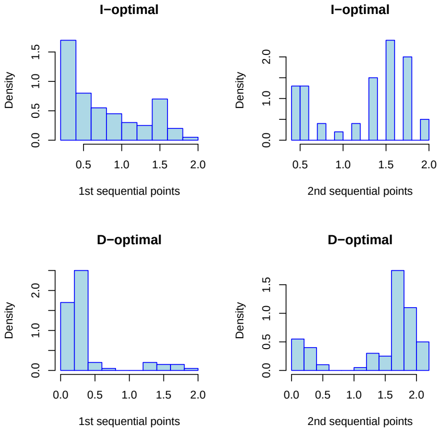

Figure: 5. Histogram of sequential points for the binary model. $k = 25$ clusters, $\alpha_2 = 1$.

Simulation Study - Logistic

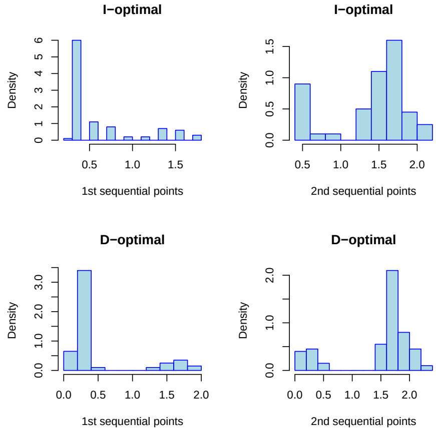

Figure: 6. Histogram of sequential points for the binary model. $k = 50$ clusters, $\alpha_2 = 1$.

Simulation Study - Logistic

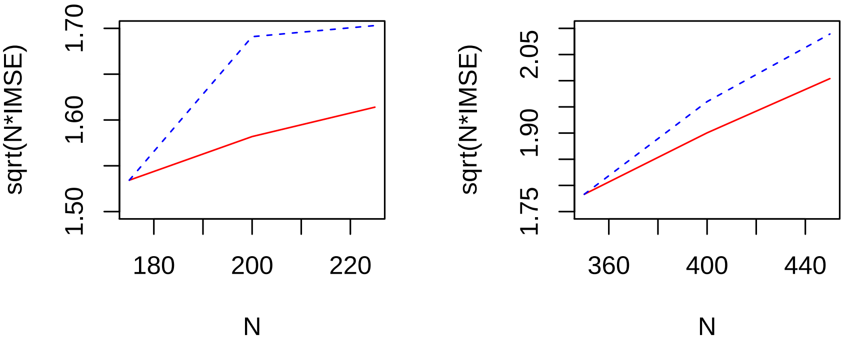

Figure: 7. IMSE for $\alpha_2 = 0.5$. ——: I-optimal, ——: D-optimal.

Simulation Study - Logistic

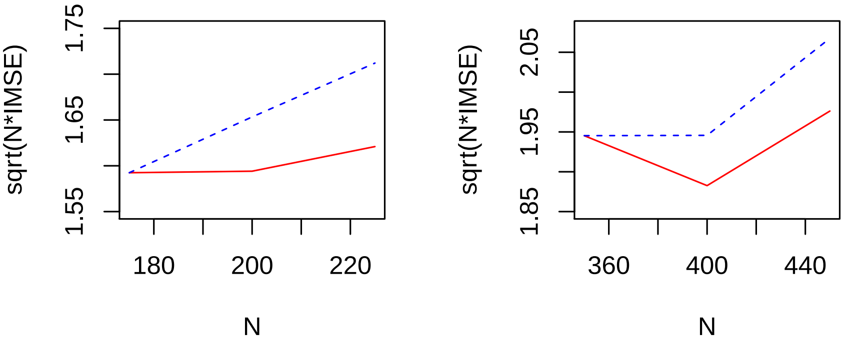

Figure: 8. IMSE for $\alpha_2 = 1$. ——:I-optimal, ——: D-optimal.

Simulation Study - Poisson

- Data were generated from a "true" Poisson mixed model:

$$y_{ij}|u_i \sim \text{ind. Poisson}(p_{ij}), i = 1, \ldots, k; j = 1, \ldots, n_0$$
$$\log(p_{ij}) = \alpha_0 + \alpha_1 x_j + \alpha_2 x_j^2 + u_i$$
$$u_i \sim \text{ind. } \mathcal{N}(0, \sigma_u^2).$$

- Parameters were fixed at $\sigma_u^2 = 0.5$, $\alpha_0 = -1$, $\alpha_1 = 0.2$ and $\alpha_2 = (0.2, 0.4)$
- Fitted model: $\log(p_{ij}^*) = \beta_0 + \beta_1 x_j + u_i$
- Initial data are based on $n_0 = 7$ design points equally spaced in $[0, 3]$.
- Two design points were chosen sequentially using both I-optimal and D-optimal criteria.

Simulation Study - Poisson

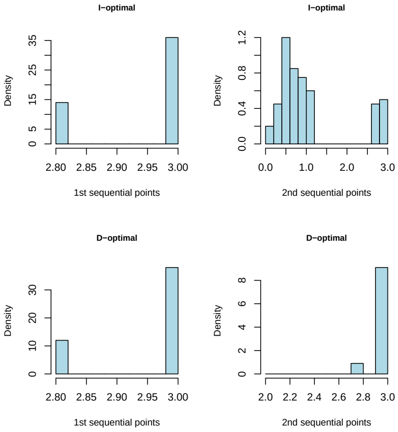

Figure: 9. Histogram of sequential points for the Poisson model. $k = 25$ clusters, $\alpha_2 = 0.2$.

Simulation Study - Poisson

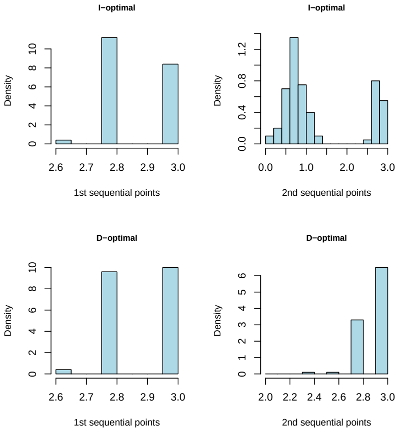

Figure: 10. Histogram of sequential points for the Poisson model. $k = 50$ clusters, $\alpha_2 = 0.2$.

Simulation Study - Poisson

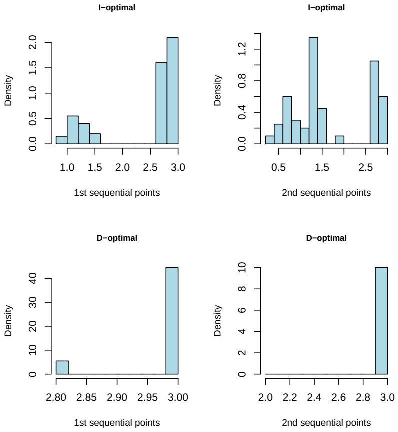

Figure: 11. Histogram of sequential points for the Poisson model. $k = 25$ clusters, $\alpha_2 = 0.4$.

Simulation Study - Poisson

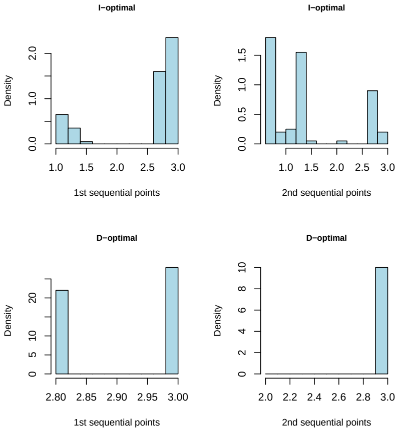

Figure: 12. Histogram of sequential points for the Poisson model. $k = 50$ clusters, $\alpha_2 = 0.4$.

Simulation Study - Poisson

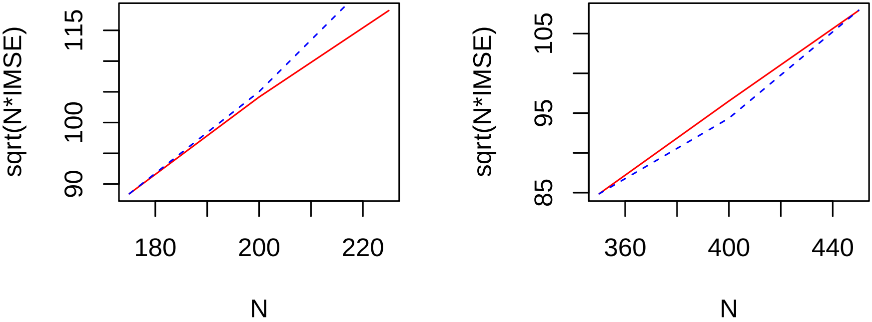

Figure: 13. IMSE for $\alpha_2 = 0.2$. ——-: I-optimal, ——-: D-optimal.

Simulation Study - Poisson

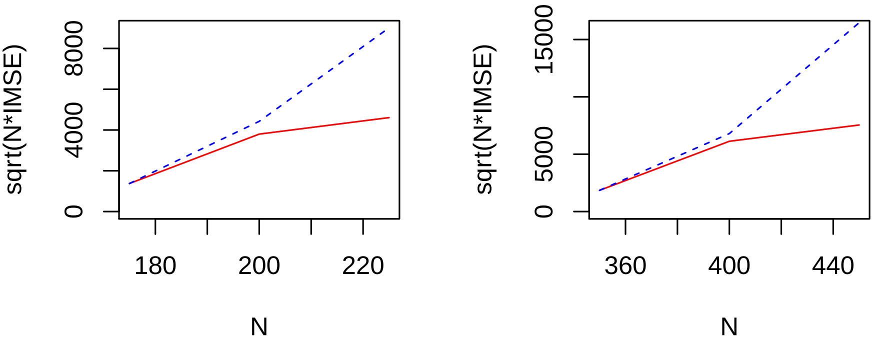

Figure: 14. IMSE for $\alpha_2 = 0.4$. ——:I-optimal, ——: D-optimal.

Simulation Study - Poisson - Efficiency comparision

- Compare four classes of designs:
- 1. 2. Our proposed I- and D-optimal sequential designs.
- 3. Conventionally used 'uniform' designs: the experimental points are uniformly distributed throughout the design space. 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.
- 4. The classical sequential D-optimal designs without consideration of model departures: using the same procedure as developed previously. However, in these designs, the design sequential points are obtained by maximizing the determinant of the information matrix evaluated at the estimates attained from the previous step.

Simulation Study - Poisson - Efficiency comparison

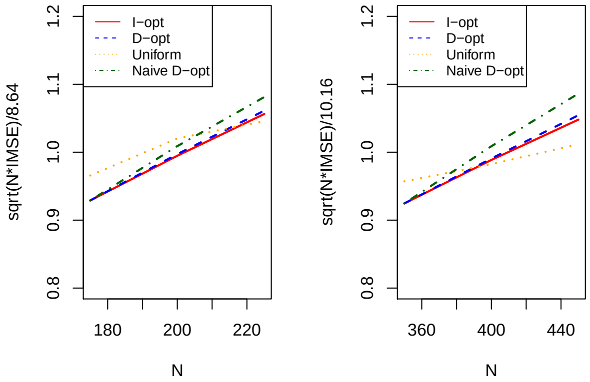

Figure: 15. Efficiency comparison. IMSEs for $\alpha_2 = 0.2$. Left: k=25. Right: k=50.

Simulation Study - Poisson - Efficiency comparison

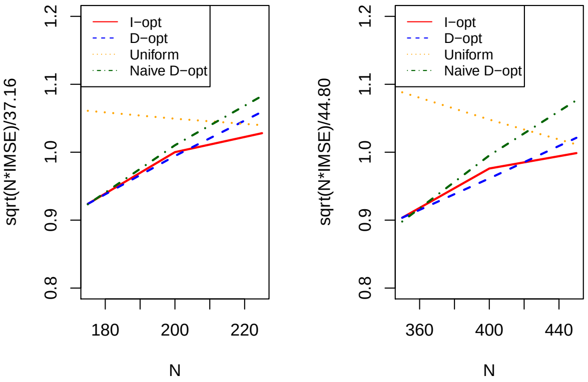

Figure: 16. Efficiency comparison. IMSEs for $\alpha_2 = 0.4$. Left: k=25. Right: k=50.

Conclusions

- I-optimal design appears to be more efficient than D-optimal design for all cases under misspecified logistic models and for most cases under misspecified Poisson models.
- Both design criteria require intensive computation.
- Some practical approximation methods can be found in Sinha and Xu (2016) to reduce the computational difficulties.
- The efficiencies of both resulting I- and D-optimal sequential designs are higher than their corresponding classical D-optimal sequential designs for all cases considered (about 3-20 percents higher in the example).

References

- 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. (2006). Adaptive designs for dose-finding based on efficacy-toxicity response. *Journal of Statistical Planning and Inference*, **136**, 1800–1823.
- Dror, H. A. & Steinberg, D. M. (2008). Sequential Experimental Designs for Generalized Linear Models. *Journal of the American Statistical Association*, **92**, 162–170.
- Heagerty, P.J. and Kurland, B.F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika, 88 (4), 973-985.
- Maram, P. P. & Jafari, H. (2016). Bayesian D-optimal design for logistic regression model with exponential distribution for random intercept. *Journal of Communications in Statistics - Theory and Methods*, **43**, 1234–1247.

References

- McCulloch, C. E. (1997). Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **103**, 288–298.
- Zhang, Y. & Ye, K. (2014). Bayesian D-optimal design for Poisson regression models. *Journal of Statistical Computation and Simulation*, **86**, to appear.
- Sinha, S. K. & Wiens, D. P. (2002). Robust sequential designs for nonlinear regression. *The Canadian Journal of Statistics*, **30**, 601–618.
- Sinha, S. K. & Xu, X. (2011). Sequential optimal designs for generalized linear mixed models. *Journal of Statistical Planning and Inference*, **141**, 1394–1402.
- Sinha, S. K. & Xu, X. (2016). Sequential Designs for Repeated Measures Experiments. *Journal of Statistical Theory and Practice*, **10(3)**, 497-514.