Optimal Block Designs for Generalized Linear Mixed Models

Tim Waite

1

1

1

1

1

1

1

1

University of Southampton, UK

DAE2012, Athens, GA

18 October, 2012

Joint work with

Dave Woods & Tim Waterhouse

1

1

2

2

2

2

Eli Lilley, USA

Block designs for GLMMs

18 Oct 2012
Outline

- Motivation and background
- Information approximations for GLMMs
- Comparisons of design performance
- Bayesian designs
Motivation

Need to design efficient experiments with non-normal response
E.g. binary (0/1) response:
- bioassay
- crystallography
- engineering

Units divide into homogeneous blocks
- batches
- repeated observations on an individual
Generalized Linear Mixed Models

Assume $m_i$ units in $i$th block, $i = 1, \ldots, n$.

Associated with $i$th block, vector of random effects $u_i \sim N_r(0, G)$.

For $j$th unit in $i$th block

$$y_{ij} | u_i \sim \pi[\mu(x_{ij} | u_i), \varphi V(x_{ij} | u_i)]$$

where

- $\pi[\mu, \varphi V]$ a distribution from exponential family
- $g\{\mu(x | u)\} = \nu(x | u)$
  - $\nu(x | u) = f^T(x)\beta + z^T(x)u$, \hspace{1em} \text{linear predictor: fixed and random parts}
  - $\eta(x) = f^T(x)\beta$, \hspace{1em} \text{fixed part}
- $x_{ij} \in \mathcal{X} = [-1, 1]^q$ is a treatment vector; $\beta$ holds $p$ unknown regression parameters
- $f : \mathcal{X} \to \mathbb{R}^p$, $z : \mathcal{X} \to \mathbb{R}^r$ are known vectors of functions
We focus mainly on a special case.

**Logistic random intercept**

- $\pi$ is Bernoulli, $\varphi = 1$, $V = \mu(1 - \mu)$
- $g(\mu) = \log\{\mu/(1 - \mu)\}$
- $\mathbf{u}_i = u_i \sim N(0, \sigma^2)$ is scalar, $\mathbf{z}(\mathbf{x}) = 1$ so that
  \[
  \nu(\mathbf{x}|u) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + u.
  \]
- blocks have additive (random) effect
Approximate block designs

For simplicity, assume all blocks of identical fixed size $m_i = m$.

Let $\zeta_i = (x_{i1}, x_{i2}, \ldots, x_{im}) \in \mathcal{X}^m$ be the treatments used in $i$th block.

- distinguish $\zeta_i$ and $\zeta_j$ only up to permutations
- let $b$ be number of distinct $\zeta_i$

Can assume that $\zeta_k, k = 1, \ldots, b$, are distinct.

Write the design as

$$\xi = \begin{cases} \zeta_1 & \cdots & \zeta_b \\ w_1 & \cdots & w_b \end{cases},$$

(1)

where $w_k = \text{prop}^n$ blocks using $\zeta_k$. Clearly $w_k > 0$ and $\sum_{k=1}^{b} w_k = 1$.

We relax the constraint that $nw_k$ is an integer - approximate theory.
Assume we are interested in estimating $\beta$ only, and that $\sigma^2$ is known.\footnote{Standard assumption, e.g.: Cheng (1995), Goos & Vandeboek (2001), Niaparast (2008)}

We seek $D$-optimal designs, which maximize

$$
\psi_D(\xi) = \log |M_\beta(\xi; \theta)|,
$$

where

$$
M_\beta(\xi; \theta) = \sum_{k=1}^{b} w_k M_\beta(\zeta_k; \theta)
$$

and

$$
M_\beta(\zeta_k; \theta) = E_{y_k} \left\{ - \frac{\partial^2 \log p(y_k|\zeta_k, \theta)}{\partial \beta \partial \beta^T} \right\},
$$

and $p(y|\zeta, \theta)$ is probability of observing $y \in \{0, 1\}^m$ from block $\zeta \in \mathcal{X}^m$.

For large $n$, approximately

$$
\hat{\beta} \sim N_p[\beta, M_\beta(\xi; \theta)^{-1}/n].
$$

\[1\text{Standard assumption, e.g.: Cheng (1995), Goos & Vandeboek (2001), Niaparast (2008)}\]
Complete enumeration

Information matrix can be written using sum over outcomes

\[ M_\beta(\zeta; \theta) = F^T \left\{ \sum_{y \in \{0,1\}^m} \frac{1}{p_y} \left( \frac{\partial p_y}{\partial \eta} \right) \left( \frac{\partial p_y}{\partial \eta} \right)^T \right\} F, \]

where

- \( F \) is the model matrix for \( \zeta = (x_1, \ldots, x_m) \in \mathcal{X}^m \)

Logistic random intercept

- Letting \( \eta_j = \eta(x_j) \), and \( h \) be the logistic function,

\[
p_y(\eta, \sigma^2) = \int_{-\infty}^{\infty} \prod_{j=1}^{m} \left[ y_j h(\eta_j + \sigma u) + (1 - y_j)\left\{1 - h(\eta_j + \sigma u)\right\} \right] \phi(u) \, du
\]

- The derivative is

\[
\frac{\partial p_y}{\partial \eta_j} = (2y_j - 1) \int_{-\infty}^{\infty} h'(\eta_j + \sigma u) \prod_{j' \neq j} \left[ y_{j'} h(\eta_{j'} + \sigma u) \right] \phi(u) \, du + (1 - y_j)\left\{1 - h(\eta_j + \sigma u)\right\} \phi(u) \, du
\]
Approximations

- fast enough to compare a small set of candidate designs
- too slow to optimize over continuous factor space

Main idea

Use cheap approximations as a surrogate for $M_B(\xi; \theta)$ in the optimization stage.

We have a suite of available approximations.

Two questions:
- relative performance
- absolute performance
Approximations

Early GLMM literature - methods for estimation

- iteratively linearize the model around the current fitted predictor
- weighted least squares
- first order MQL, PQL; also MQL2

Idea: use approximations of \( \text{var}(\hat{\beta}) \) from these methods.

- .. to give approximations to \( M_\beta(\xi; \theta) \)
- some design papers on this
  - mainly restricted numbers of factors / predictor structure
- no assessment of choice of approximation or its impact

Design: Moerbeek et al. (2001), Moerbeek & Maas (2005), Tekle et al. (2008)
Approximations

Expressions for MQL and PQL - logistic random intercept

\[
M_{\text{MQL}}(\zeta_i; \theta) = F_i^T \left\{ \sigma^2 J + \text{diag} \left( \frac{e^{-f^T(x_{ij})\beta}}{1 + e^{-f^T(x_{ij})\beta}} \right)^2 \right\}^{-1} F_i
\]

\[
M_{\text{PQL}}(\zeta_i; \theta) = F_i^T \left\{ \sigma^2 J + \text{diag} \left( 2 + 2e^{\sigma^2/2} \cosh (f^T(x_{ij})\beta) \right) \right\}^{-1} F_i,
\]

where \( F_i \) is model matrix corresponding to \( \zeta_i \), and \( J \) is a matrix of 1s.

- MQL crudely sets \( u \approx 0 \), PQL keeps track of estimates of \( u \)
- MQL2: based on higher order Taylor expansion, similar complexity

[N.B. our PQL is not the same as in Tekle et al. (2008)]
We define adjusted MQL (AMQL), by

\[ M_{\text{AMQL}}(\xi; \theta) = M_{\text{MQL}}(\xi; \theta_{\text{adj}}) \]

\[ \beta_{\text{adj.}} = \beta (1 + c^2 \sigma^2)^{-1/2} \]

\[ \theta_{\text{adj.}} = (\beta_{\text{adj.}}, \sigma^2)^T, \]

where \( c = 15\sqrt{3}/(16\pi) \).

Related to this: for logistic random intercept model,

\[ P(y_{ij} = 1) = \int_{-\infty}^{\infty} h\{f^T(x_{ij})\beta + \sigma u\} \phi(u) \, du \]

\[ \approx h\left(f^T(x_{ij})\beta_{\text{adj}}\right), \]

see e.g. Breslow and Clayton (1993).

Marginal model is approximately logistic with attenuated parameters, \( \beta_{\text{adj.}} \).
Related problems and approaches

Other estimation methods can have more tractable properties.

**Generalized estimating equations**
- Woods & Van de Ven (2011)
- Binary response in blocks

**Quasi-likelihood**
- Niaparast (2009), Niaparast & Schwabe (2012)
- Model: Poisson response GLMM
Optimal designs for estimating $\beta$ depend on the unknown parameters...

...overcome this with a pseudo-Bayesian $D$-optimal design $\xi^*$, maximizing

$$
\Psi_\alpha(\xi) = \begin{cases} 
\frac{1}{\alpha} \log \int_{\Theta} |M(\xi; \theta)|^\alpha dG(\theta) & \alpha \neq 0 \\
\int_{\Theta} \log |M(\xi; \theta)| dG(\theta) & \alpha = 0 
\end{cases},
$$

where $G$ is a distribution across the parameter space $\Theta$.

- concave for $\alpha \leq p^{-1}$ (Firth & Hinde, 1997)
- $\alpha = 0$ is usually recommended (Chaloner & Verdinelli, 1995)
- in our first example, we choose $\alpha = p^{-1}$ to ensure a finite objective function for prior distributions with unbounded support
Parameter dependence

Approximation of objective function
is via numerical quadrature (Gotwalt et al., 2009)

Alternatives include
- Monte Carlo methods (e.g. Waterhouse et al., 2008)
- Latin Hypercube Sampling (e.g. Woods and Van de Ven, 2011)

Assessment of designs
via simulation from $G(\theta)$ and design efficiencies (Woods et al., 2006)
For a given design $\xi$, define efficiency in terms of locally optimal $\xi^+$

$$\text{eff}(\xi; \theta) = \left[\frac{|M(\xi; \theta)|}{|M(\xi^+; \theta)|}\right]^{1/p}$$
Example

Set $m = 4$, and $\mathbf{x} = (x_1, x_2)^T$.

Model: $\pi = \text{Bernoulli}$, $g = \logit$

$$
\nu(\mathbf{x}|u) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u
$$

Prior:

$$
\begin{pmatrix}
\beta_0 \\
\beta_1 \\
\beta_2 \\
\log \sigma^2
\end{pmatrix}
\sim N
\left( 
\begin{pmatrix}
0 \\
1 \\
2 \\
0
\end{pmatrix},
\begin{pmatrix}
.5 & .5 \\
.5 & .5
\end{pmatrix}
\right)
$$

Moderate value of $\sigma^2$. 
Example: optimal designs

Tim Waite (U. Southampton)
Example: comparisons

How to compare

Use complete enumeration to evaluate $M(\xi; \theta)$

- calculate $\Psi_\alpha(\xi)$
- relative pseudo-Bayesian efficiency

$$\text{eff}(\xi_1; \xi_2) = \frac{\int_\Theta |M(\xi_1; \theta)|^{1/p} dG(\theta)}{\int_\Theta |M(\xi_2; \theta)|^{1/p} dG(\theta)}$$

$$= \exp \frac{1}{p} \left\{ \Psi(\xi_1) - \Psi(\xi_2) \right\} ,$$

Results

Using AMQL design as reference

<table>
<thead>
<tr>
<th>Design</th>
<th>MQL</th>
<th>PQL</th>
<th>MQL2</th>
<th>AMQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative Efficiency (%)</td>
<td>98.8</td>
<td>97.7</td>
<td>98.6</td>
<td>100</td>
</tr>
</tbody>
</table>
Approximate efficiency distributions

Using 1000 samples from prior

![Graph showing efficiency distributions for different models: MQL, PQL, MQL2, and AMQL. The x-axis represents efficiency, and the y-axis represents density. The graph illustrates the comparison of these models through their efficiency distributions.]
For $m = 2$, we can compute $D$-optimal designs using full ML. Recall

$$M(\zeta; \theta) = F^T W(\eta, \sigma^2) F,$$

where $W(\eta, \sigma^2)$ is a function of $p_y$ and $\partial p_y / \partial \eta_k$, $y \in \{0, 1\}^2$, $k = 1, 2$.

**Precomputation**

- fix $\sigma^2$, consider $f(\eta) = p_y(\eta, \sigma^2)$ as function $\mathbb{R}^2 \rightarrow \mathbb{R}$
- use quadrature to calculate $f(\eta)$...
- ...for every $\eta$ in a grid $\mathcal{G} \subset [-20, 20]^2$
- approximate $f(\eta)$ for $\eta \notin \mathcal{G}$
- ...using interpolation, e.g. bilinear\(^1\)

\(^1\)We use interp.surface in R package fields (Furrer et al., 2012)
effectively we have precomputed all relevant integrals
  - can even produce Bayesian designs
- doesn’t scale well to larger block size
Absolute performance: example

Set \( m = 2 \) so we can precompute.

**2-factor model**

Model: \( \nu(x|u) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \).

Prior: \((\beta_0, \beta_1, \sigma^2)^T = (0, 5, 5)^T\) \(\beta_2 \sim U(0, 10)\).

Seek designs maximizing

\[
\Psi_0(\xi) = E_\theta \{ \log |M_\theta(\xi; \theta)| \}.
\]

In this case, use a very crude approximation to integral, set

\[
\theta_s = (0, 5, s, 5), \quad s = 0, 1, \ldots, 10,
\]

\[
\Psi_0(\xi) \approx \sum_{s=0}^{10} \frac{1}{11} \log |M_\beta(\xi; \theta_s)|
\]
Absolute performance: example

Design points under ML and MQL

weights = .37, .29
weights = .23, .17
weights = .40, .21

Tim Waite (U. Southampton)
Absolute performance: example

weight = .33
Absolute performance: example

Local efficiencies

![Graph showing local efficiencies with lines for ML, AMQL, MQL, PQL, and Factorial, with axes for $\beta_2$ and Efficiency]
Summary

- for small $\sigma^2$
  - MQL, MQL2, PQL give similar designs
  - performance comparable to AMQL
- for larger $\sigma^2$
  - AMQL is closest to ML
  - MQL outperforms PQL

Current and future work

- prediction of random effects (HGLMs)
- estimation of variance components
- more complex random effects structures
- case studies (e.g. pharmaceutical)
Gotwalt, C., Jones, B. and Steinberg, D. (2009), *Technometrics*, 51, 88–95
Why $\alpha = 1/p$?

In our first example, we use prior distributions with unbounded support.

For logistic RI model, it can be shown that

$$\log |M(\xi; \theta)| \to -\infty$$

- as $\sigma^2 \to \infty$
- as $\beta \to \infty$ within $A \subseteq \mathbb{R}^p$, having $P(A) \approx 1$

Therefore there are possible convergence issues with $\Psi_\alpha$ when $\alpha \leq 0$, e.g.

$$\Psi_0(\xi) = E_\beta \{ \log |M(\xi; \theta)| \} .$$

If $\alpha > 0$, there are no such issues with $\Psi_\alpha$. 
Optimization

Algorithms

1. Transform to unconstrained problem, as in Atkinson et al. (2007, OUP)
   - use general purpose quasi-Newton or simplex algorithm
   - consider omission of blocks with very small weight (e.g. $10^{-5}$)

2. Co-ordinate optimization
   - similar to Meyer and Nachtsheim (1995, Technometrics)
   - factor values: optimize holding all other co-ordinates fixed
   - weights: if $w \rightarrow w'$ is proposal, multiply other weights by $(1 - w')/(1 - w)$
   - Also incorporate consolidation step
     - if $M(\zeta_i; \theta) - M(\zeta'_i; \theta) \approx 0_{p \times p}$, $i < i'$, then transfer weight from $\zeta_{i'}$ to $\zeta_i$

Both algorithms greedy: prone to becoming stuck in suboptimal attractor states
   - multiple random initializations
   - co-ordinate optimization seems to be slightly better