Optimal Block Designs for Generalized Linear Mixed Models

Tim Waite[1]

tww1g08@soton.ac.uk

Joint work with

Dave Woods[1] & Tim Waterhouse[2]

[1]University of Southampton, UK
[2]Eli Lilley, USA

DAE2012, Athens, GA
18 October, 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 ith block, $i = 1, \ldots, n$.

Associated with ith block, vector of random effects $\mathbf{u}_i \sim N_r(\mathbf{0}, G)$.

For jth unit in ith block

$$y_{ij}|\mathbf{u}_i \sim \pi[\,\mu(\mathbf{x}_{ij}|\mathbf{u}_i),\, \varphi V(\mathbf{x}_{ij}|\mathbf{u}_i)\,]$$

where

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

Generalized Linear Mixed Models

We focus mainly on a special case.

Logistic random intercept

- π 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

$$\nu(\mathbf{x}|u) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + u\,.$$

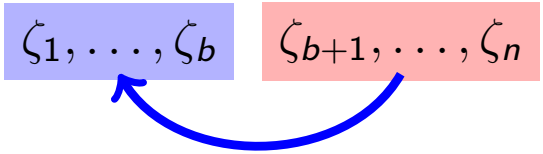
- blocks have additive (random) effect

Approximate block designs

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

Let $\zeta_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{im}) \in \mathcal{X}^m$ be the treatments used in ith block.

- distinguish ζ_i and ζ_j only up to permutations
- let b be number of distinct ζ_i



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

Write the design as

$$\xi = \left\{ \begin{array}{ccc} \zeta_1 & \ldots & \zeta_b \\ w_1 & \ldots & w_b \end{array} \right\}, \tag{1}$$

where w_k = prop^{n} blocks using ζ_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.

Optimal designs

Assume we are interested in estimating $\boldsymbol{\beta}$ only, and that σ^2 is known.[1]

We seek D-optimal designs, which maximize

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

where

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

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

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

For large n, approximately

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

[[1]Standard assumption, e.g.: Cheng (1995), Goos & Vandebroek (2001), Niaparast (2008)]

Complete enumeration

Information matrix can be written using sum over outcomes

$$M_{\boldsymbol{\beta}}(\zeta;\boldsymbol{\theta}) = F^T \left\{ \sum_{\mathbf{y}\in\{0,1\}^m} \frac{1}{p_{\mathbf{y}}} \left(\frac{\partial p_{\mathbf{y}}}{\partial \boldsymbol{\eta}}\right) \left(\frac{\partial p_{\mathbf{y}}}{\partial \boldsymbol{\eta}}\right)^T \right\} F\,,$$

where

- F is the model matrix for $\zeta = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \in \mathcal{X}^m$

Logistic random intercept

- Letting $\eta_j = \eta(\mathbf{x}_j)$, and h be the logistic function,

$$p_{\mathbf{y}}(\boldsymbol{\eta}, \sigma^2) = \int_{-\infty}^{\infty} \prod_{j=1}^{m} \Big[y_j h(\eta_j + \sigma u) + (1 - y_j)\{1 - h(\eta_j + \sigma u)\} \Big] \phi(u)\, du$$

- The derivative is

$$\frac{\partial p_{\mathbf{y}}}{\partial \eta_j} = (2y_j - 1) \int_{-\infty}^{\infty} h'(\eta_j + \sigma u) \prod_{j' \neq j} \Big[y_{j'} h(\eta_{j'} + \sigma u) + (1 - y_{j'})\{1 - h(\eta_{j'} + \sigma u)\} \Big] \phi(u)\, du$$

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 $M_{\boldsymbol{\beta}}(\xi; \boldsymbol{\theta})$ 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 var($\hat{\boldsymbol{\beta}}$) from these methods..

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

Analysis: Breslow & Clayton (1993), Rodriguez & Goldman (1995), Goldstein & Rasbash (1996)
Design: Moerbeek et al. (2001), Moerbeek & Maas (2005), Tekle et al. (2008)

Approximations

Expressions for MQL and PQL - logistic random intercept

$$M_{MQL}(\zeta_i;\boldsymbol{\theta}) = F_i^T \left\{ \sigma^2 J + \text{diag}\left(e^{-\mathbf{f}^T(\mathbf{x}_{ij})\boldsymbol{\beta}} / \left(1 + e^{-\mathbf{f}^T(\mathbf{x}_{ij})\boldsymbol{\beta}}\right)^2 \right)^{-1} \right\}^{-1} F_i$$

$$M_{PQL}(\zeta_i;\boldsymbol{\theta}) = F_i^T \left\{ \sigma^2 J + \text{diag}\left(2 + 2e^{\sigma^2/2} \cosh\left(\mathbf{f}^T(\mathbf{x}_{ij})\boldsymbol{\beta}\right) \right) \right\}^{-1} F_i \,,$$

where F_i is model matrix corresponding to ζ_i, and J is a matrix of 1s.

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

[N.B. our PQL is not the same as in Tekle et al. (2008)]

Binary data: attenuation

We define adjusted MQL (AMQL), by

$$M_{\text{AMQL}}(\xi;\boldsymbol{\theta}) = M_{\text{MQL}}(\xi;\boldsymbol{\theta}_{\text{adj.}})$$

$$\boldsymbol{\beta}_{\text{adj.}} = \boldsymbol{\beta}\left(1 + c^2\sigma^2\right)^{-1/2}$$

$$\boldsymbol{\theta}_{\text{adj.}} = \left(\boldsymbol{\beta}_{\text{adj.}}, \sigma^2\right)^T,$$

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

Related to this: for logistic random intercept model,

$$P(y_{ij} = 1) = \int_{-\infty}^{\infty} h\{\mathbf{f}^T(\mathbf{x}_{ij})\boldsymbol{\beta} + \sigma u\}\,\phi(u)\,du$$

$$\approx h\left(\mathbf{f}^T(\mathbf{x}_{ij})\boldsymbol{\beta}_{\text{adj}}\right),$$

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

Marginal model is approximately logistic with attenuated parameters, $\boldsymbol{\beta}_{\text{adj.}}$

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

Parameter dependence

Optimal designs for estimating $\boldsymbol{\beta}$ depend on the unknown parameters...
...overcome this with a pseudo-Bayesian D-optimal design ξ^*, maximizing

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

where G is a distribution across the parameter space Θ.

- 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 $G(\boldsymbol{\theta})$ and design efficiencies (Woods et al., 2006)
For a given design ξ, define efficiency in terms of locally optimal ξ^+

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

Relative performance

Example

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

Model:

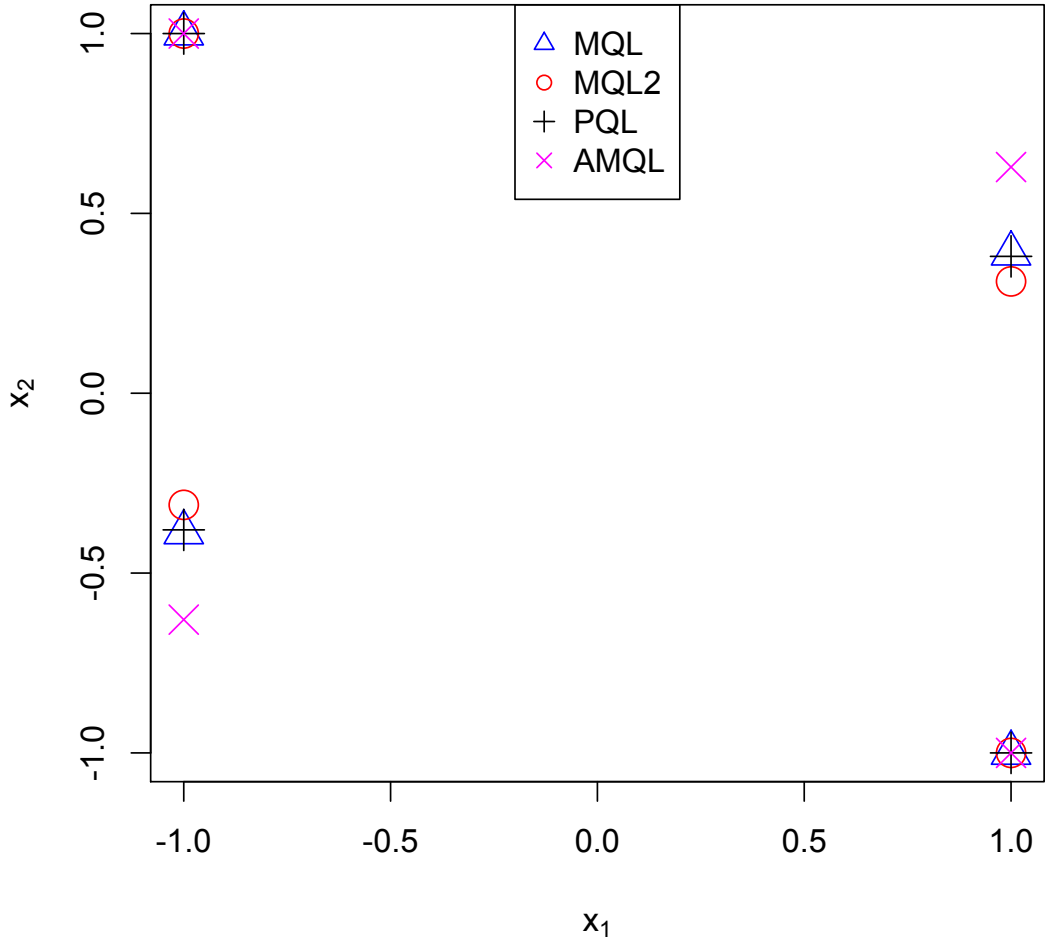
$$\pi = \text{Bernoulli}, \quad g = \text{logit}$$
$$\nu(\mathbf{x}|u) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

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 σ^2.

Example: optimal designs



Example: comparisons

How to compare

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

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

$$\begin{aligned}\mathrm{eff}(\xi_1; \xi_2) &= \frac{\int_\Theta |M(\xi_1; \boldsymbol{\theta})|^{1/p}\, dG(\boldsymbol{\theta})}{\int_\Theta |M(\xi_2; \boldsymbol{\theta})|^{1/p}\, dG(\boldsymbol{\theta})} \\ &= \exp \frac{1}{p}\{\Psi(\xi_1) - \Psi(\xi_2)\}\,,\end{aligned}$$

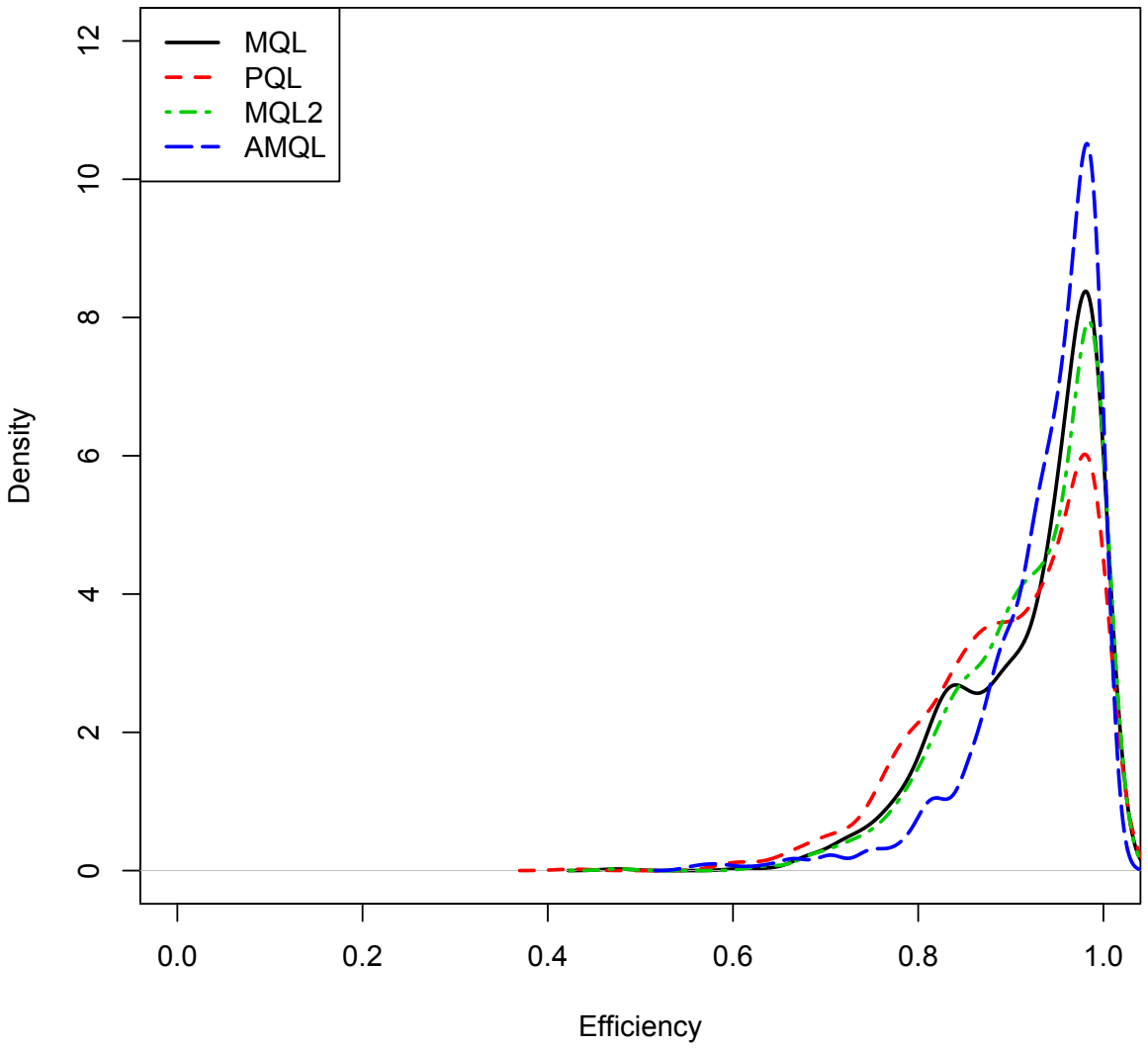
Results

Using AMQL design as reference

Design	MQL	PQL	MQL2	AMQL
Relative Efficiency (%)	98.8	97.7	98.6	100

Approximate efficiency distributions

Using 1000 samples from prior



Absolute performance

For $m = 2$, we can compute D-optimal designs using full ML. Recall

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

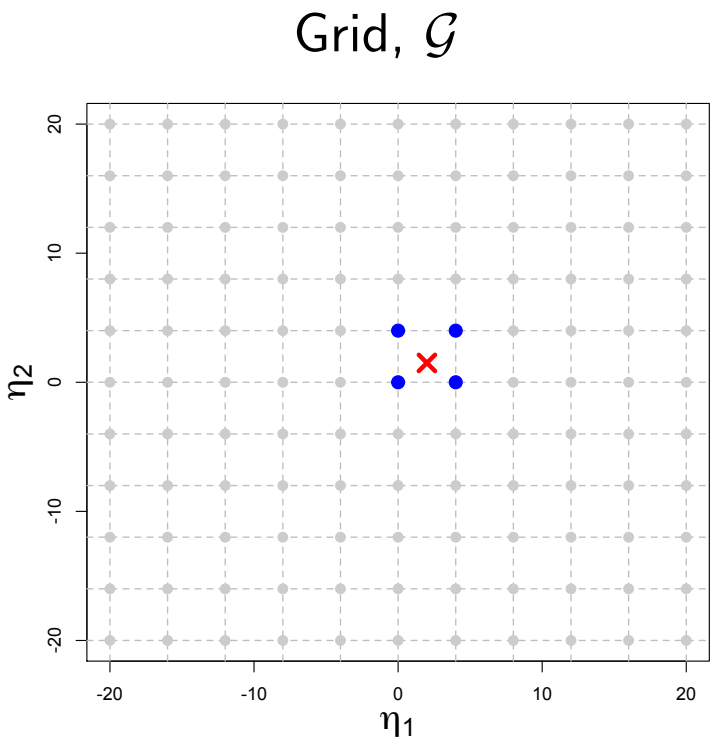
where $W(\boldsymbol{\eta}, \sigma^2)$ is a function of $p_{\mathbf{y}}$ and $\partial p_{\mathbf{y}} / \partial \eta_k$, $\mathbf{y} \in \{0, 1\}^2$, $k = 1, 2$.

Precomputation

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

[1] We use `interp.surface` in `R` package `fields` (Furrer et al., 2012)

Precomputation



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

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

Seek designs maximizing

$$\Psi_0(\xi) = E_{\boldsymbol{\theta}}\{\log|M_{\boldsymbol{\theta}}(\xi; \boldsymbol{\theta})|\}\,.$$

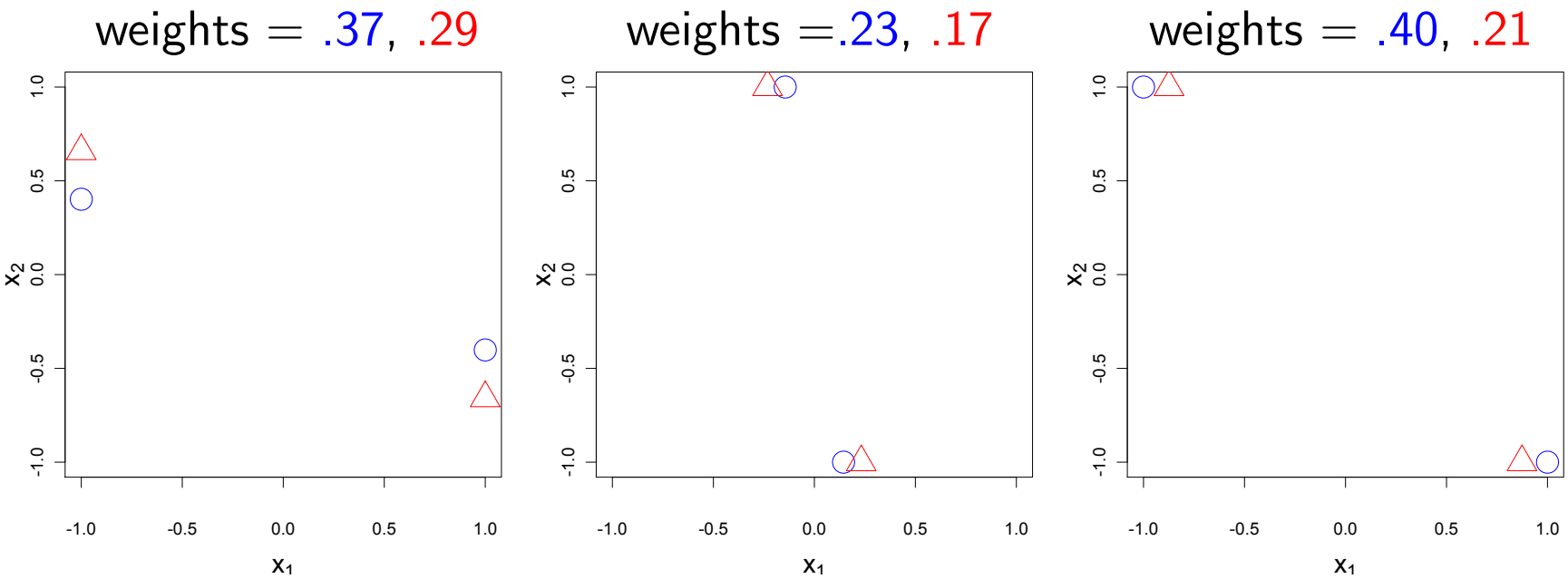
In this case, use a very crude approximation to integral, set

$$\boldsymbol{\theta}_s = (0, 5, s, 5), \qquad s = 0, 1, \ldots, 10,$$

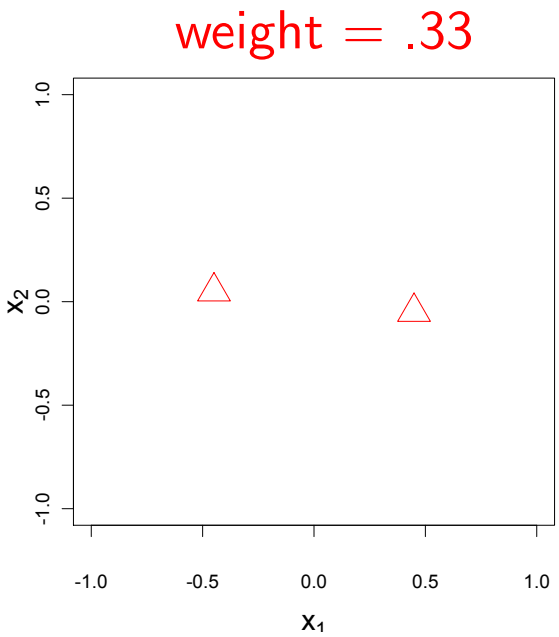
$$\Psi_0(\xi) \approx \sum_{s=0}^{10} \frac{1}{11} \log|M_{\boldsymbol{\beta}}(\xi; \boldsymbol{\theta}_s)|$$

Absolute performance: example

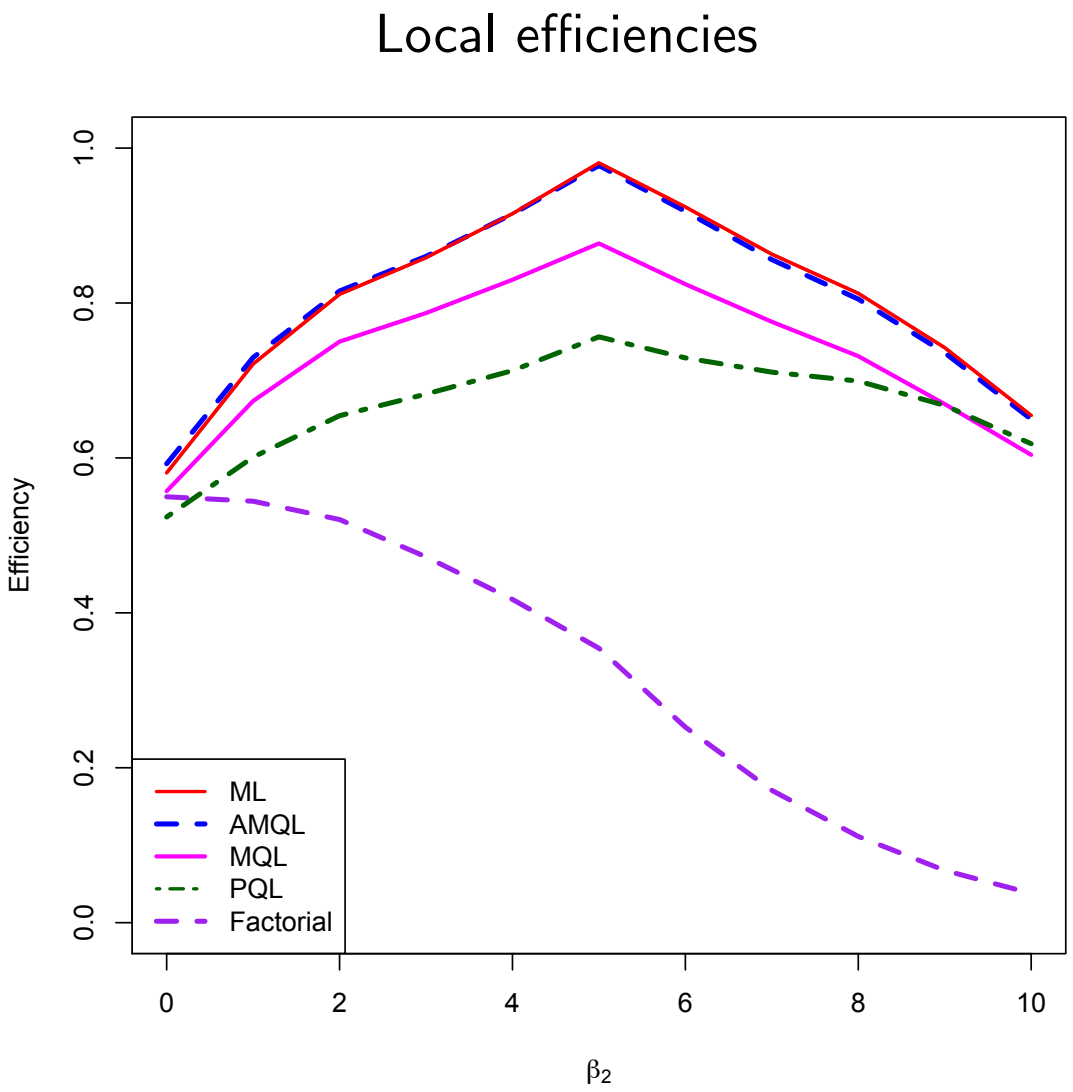
Design points under ML and MQL



Absolute performance: example



Absolute performance: example



Summary

- for small σ^2
  - MQL, MQL2, PQL give similar designs
  - performance comparable to AMQL
- for larger σ^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)

References

- Breslow, N. and Clayton, D. (1993), *JASA*, 88, 9–25
- Chaloner, K. and Verdinelli, I. (1995), *Stat. Sci.*, 10, 273–304
- Cheng, C. (1995), *Stat. Sin.*, 5, 485–497
- Firth, D. and Hinde, J. (1997), *JRSSB*, 59, 793–797
- Furrer, R., Nychka, D. and Stain, S. (2012), `R` *package* `fields`, *CRAN*.
- Goldstein, H. and Rasbash, J. (1996), *JRSSA*, 159, 505–513
- Goos, P. and Vandebroek, M. (2001), *CSDA*, 37, 433–453
- Gotwalt, C., Jones, B. and Steinberg, D. (2009), *Technometrics*, 51, 88–95
- Moerbeek, M., Van Breukelen, J. and Berger, M. (2001), *JRSSD*, 50, 17–30
- Moerbeek, M. and Maas, C. (2005), *Comm. Stat. Th. Meth.*, 34, 1151–1167
- Niaparast, M. (2009), *Stat. Prob. Lett.*, 79, 741–747
- Niaparast, M. and Schwabe, R. (2012), *JSPI*, in press.
- Tekle, F., Tan, F. and Berger, M. (2008), *CSDA*, 52, 5253–5262
- Waterhouse, T., Woods, D., Eccleston, J. and Lewis, S. (2008), *JSPI*, 132–144
- Woods, D. and Van de Ven, P. (2011), *Technometrics*, 53, 173–182
- Woods, D., Lewis, S., Eccleston, J. and Russell, K. (2006), *Technometrics*, 53, 173–182

Parameter dependence

Why $\alpha = 1/p$?

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

For logistic RI model, it can be shown that

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

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

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

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

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

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 $w \to w'$ is proposal, multiply other weights by $(1 - w')/(1 - w)$
   - Also incorporate consolidation step
     - if $M(\zeta_i; \boldsymbol{\theta}) - M(\zeta_i'; \boldsymbol{\theta}) \approx 0_{p \times p}$, $i < i'$, then transfer weight from $\zeta_{i'}$ to ζ_i

Both algorithms greedy: prone to becoming stuck in suboptimal attractor states

- multiple random initializations
- co-ordinate optimization seems to be slightly better