Block designs for non-normal data via conditional and marginal models

Dave Woods

D.Woods@southampton.ac.uk
www.southampton.ac.uk/~davew

Supported by EPSRC
Outline

• Introduction and motivation

• Designs for GEE models
  • with Peter van de Ven (VU University)

• Conditional and marginal modelling paradigms

  • Designs for GLMMs
    • with Tim Waite (Southampton) & Tim Waterhouse (Eli Lilly)
Outline

- Introduction and motivation
- Designs for GEE models
  - with Peter van de Ven (VU University)
- Conditional and marginal modelling paradigms
- Designs for GLMMs
  - with Tim Waite (Southampton) & Tim Waterhouse (Eli Lilly)
Motivation

Increasing recognition of the need to design for experiments with non-normal response

1. Science - crystallography
2. Engineering - aeronautics
3. Medicine - clinical trials

Often, experimental units are heterogeneous

- Which treatments should be run?
- Allocation of treatments to blocks?
- Analysis of data using an appropriate model that takes account of blocking?
Aim: to investigate whether or not a salt forms from a reaction between an acid and a base

Controllable variables

- rate of agitation during mixing
- volume of composition
- temperature
- evaporation rate

Outcome

Salt formed (1) or not (0)
Engine bearings (1)

Bearing experiment - Goodrich
Study of the factors affecting cracking of engine bearing coatings

Controllable variables
- Spray distance
- Spray angle
- Sweep speed

Outcome
Each bearing passes (1) or fails (0) a visual inspection
Engine bearings (2)

Units and response

- Experimental unit: pair of bearings
- Response: number of bearings which pass: 0, 1 or 2
  - binomial with $m = 2$

Blocks

- $N = 16$ runs allocated to four sessions of four runs
- Sessions are nuisance variables of no intrinsic meaning
- Little interest in modelling session-to-session variability

Reasonable assumptions

- Sessions are a random sample
- Observations within blocks (sessions) are more similar than observations in different blocks
If the observations were independent...

... we could model the response using a GLM (McCullagh & Nelder, 1989)

Probit regression

- $Y \sim \text{Binomial}(m, \pi); \mu = m\pi$
- $g(\pi) = \Phi^{-1}(\pi) = \eta = X\beta$

where $\Phi$ is the standard normal c.d.f.

Multi-variable designs

- Woods et al. (2006), Gotwalt et al. (2009)
  - Bayesian design via numerical integration
- Dror & Steinberg (2006), Russell et al. (2009)
  - K-means or model-based clustering of locally optimal designs
- Dror & Steinberg (2008)
  - sequential design via Bayesian $D$-optimality
Block designs under linear models

If the observations were normally distributed...

... we could use linear model designs for random block effects

Available methods include

- Goos & Vandebroek (2001)
  - algorithmic search

- Cook & Nachtsheim (1989), Trinca & Gilmour (2001)
  - allocation to blocks of factorial treatments from an unblocked design

- Cheng (1995)
  - minimal support designs
Models for correlated non-normal data

Conditional (random effect) models

\[ g(\pi) = X\beta + Z\tau \]

where \( Z \) is a binary matrix allocating treatments to blocks and \( \tau \) is a vector of random effects

- Generalized Linear Mixed Models (Breslow & Clayton. 1993; Ouwens et al., 2006)
- Hierarchical Generalized Linear Models (Lee & Nelder, 1996)

Marginal models

- Quasi-likelihood (Niaparast, 2009)
- Comparison with conditional models via an industrial example
  - Robinson et al. (2004)
Generalized Estimating Equations (1)

GEE models consist of . . .

- A marginal mean and variance relationship for each observation, taken from an appropriate GLM
- A “working correlation” matrix \( R(\alpha) \); observations from different blocks are independent

GEE models don’t include . . .

- a multivariate probability distribution for the observations
  \( \Rightarrow \) no likelihood function

GEE models are appropriate when . . .

- The blocks are a random sample from some larger population
- The blocks are not of interest
- The aim of the experiment is estimation of the marginal (population-averaged) effects of the variables
Generalized Estimating Equations (1)

GEE models consist of . . .

- A marginal mean and variance relationship for each observation, taken from an appropriate GLM
- A “working correlation” matrix $R(\alpha)$; observations from different blocks are independent

GEE models don’t include . . .

- a multivariate probability distribution for the observations
  $\Rightarrow$ no likelihood function

GEE models are appropriate when . . .

- The blocks are a random sample from some larger population
- The blocks are not of interest
- The aim of the experiment is estimation of the marginal (population-averaged) effects of the variables
Generalized Estimating Equations (1)

GEE models consist of . . .

- A marginal mean and variance relationship for each observation, taken from an appropriate GLM
- A “working correlation” matrix $R(\alpha)$; observations from different blocks are independent

GEE models don’t include . . .

- a multivariate probability distribution for the observations
  $\Rightarrow$ no likelihood function

GEE models are appropriate when . . .

- The blocks are a random sample from some larger population
- The blocks are not of interest
- The aim of the experiment is estimation of the marginal (population-averaged) effects of the variables
Generalized Estimating Equations (2)

Estimating equations for $\beta$

$$X' \Delta V^{-1} (Y - \mu) = 0_{q \times 1}$$

where $X$ is the model matrix, $\Delta = \text{diag}(\partial \mu_i/\partial \eta_i)$, $V = A^{1/2} R(\alpha) A^{1/2}$ and $A = \text{diag}[\text{Var}(Y_i)]$

- $R(\alpha)$ adjusts the weight given to each observation – it is not the actual correlation, more a measure of average dependance.

- Ancillary estimating equations for $\alpha$, or fix prior to analysis (Chaganty & Joe, 2004, suggest $\alpha \approx 0.2$ for moderate dependence)

Asymptotic distribution

$\hat{\beta} \sim \text{MVN}(\beta, M^{-1})$, where

$$M = X' \Delta V^{-1} \Delta X (X' \Delta V^{-1} \Sigma V^{-1} \Delta X)^{-1} X' \Delta V^{-1} \Delta X,$$

and $\Sigma$ is the true var-cov matrix for $Y$. 
Generalized Estimating Equations (2)

Estimating equations for $\beta$

$$X' \Delta V^{-1} (Y - \mu) = 0_{q \times 1}$$

where $X$ is the model matrix, $\Delta = \text{diag}(\partial \mu_i / \partial \eta_i)$, $V = A^{1/2} R(\alpha) A^{1/2}$ and $A = \text{diag}[\text{Var}(Y_i)]$

- $R(\alpha)$ adjusts the weight given to each observation – it is not the actual correlation, more a measure of average dependance

- Ancillary estimating equations for $\alpha$, or fix prior to analysis (Chaganty & Joe, 2004, suggest $\alpha \approx 0.2$ for moderate dependence)

Asymptotic distribution

$$\hat{\beta} \sim \text{MVN}(\beta, M^{-1}),$$

where

$$M = X' \Delta V^{-1} \Delta X (X' \Delta V^{-1} \Sigma V^{-1} \Delta X)^{-1} X' \Delta V^{-1} \Delta X,$$

and $\Sigma$ is the true var-cov matrix for $Y$
Bearing experiment

Marginal model
Probit regression for 16 runs with $m = 2$

- $E(Y_i) = \mu_i = 2\Phi(x_i'\beta)$
- $\text{Var}(Y_i) = 2\Phi(x_i'\beta)[1 - \Phi(x_i'\beta)] \quad i = 1, \ldots, 16$

Exchangeable working correlation
Within-block ($b = 1, \ldots, 4$)

$$R_b(\alpha) = (1 - \alpha)I_{4 \times 4} + \alpha J_{4 \times 4}$$

Working correlation matrix

$$R(\alpha) = I_{4 \times 4} \otimes R_b(\alpha)$$
Design selection criterion (1)

Prior to data collection . . .

. . . assume \( \Sigma = A^{1/2} R(\alpha) A^{1/2} \). Hence,

\[
M = X' \Delta [A^{1/2} R(\alpha) A^{1/2}]^{-1} \Delta X
\]

and design performance depends on \( \beta \) and \( \alpha \) through \( \Delta, A \) and \( R \)

To overcome this dependency . . .

. . . a Bayesian \( D \)-optimal design \( d^* \) maximizes

\[
\Psi_D(d; \mathcal{B}, \alpha) = \int_{\mathcal{B}} \log \psi_D(d; \beta, \alpha) \, dF(\beta) \quad (\ast)
\]

where \( F \) is a probability distribution across the parameter space \( \mathcal{B} \), and

\[
\psi_D(d; \beta, \alpha) = \det (M)
\]
Design selection criterion (2)

Calculation of objective function (\(\star\))

... is via averaging the local objective function \(\phi_D\) over a (21 point) Latin Hypercube Sample in \(\mathcal{B}\)

Alternatives include

- Monte Carlo methods (e.g. Waterhouse et al., 2008)
- Quadrature (e.g. Gotwalt at al., 2009)

Assessment of designs ...

... via simulation from \(\mathcal{B}\) and design efficiencies

For a given design \(d\), define efficiency in terms of locally optimal \(d^+\)

\[
\text{eff}(d) = \left(\frac{\phi_D(d)}{\phi_D(d^+)}\right)^{(1/q)}
\]

for \(q\) model parameters
Block designs

For a fixed block size $k$ ...

- an *exact* design finds the optimal selection of $N$ design points, \( x_i \in \mathcal{X} \), and their allocation to blocks
- a *continuous* design finds an optimal set of blocks and the proportion of experimental effort that should be used on each

We concentrate on...

... exact designs, appropriate for the bearing example

However, for continuous designs...

... a General Equivalence Theorem can be established
Methods of finding exact designs (1)

We consider two methods for exact designs

1. Algorithmic search of design space
   Use computer optimization to find the best exact design

2. Allocation of optimal GLM design to blocks
   Via an interchange algorithm
Methods of finding designs (2)

Comparison using . . .
Probit regression

• $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3$

• Design: 16 binomial runs ($m = 2$) in 4 blocks; $\alpha = 0.2$

Prior distribution $F(\beta)$
Independent uniform distributions with ranges

\[
\begin{align*}
\beta_0 & \sim (-2, 0) \\
\beta_1 & \sim (1, 3) \\
\beta_2 & \sim (0, 2) \\
\beta_3 & \sim (-2, 0) \\
\beta_{12}, \beta_{13}, \beta_{23} & \sim (-0.5, 1.5)
\end{align*}
\]
Algorithmic search (1)

Simulated annealing
For a given number and sizes of blocks...

- perform a continuous optimization for each design point
- adjust points by random perturbation
- accept adjustment
  - if the performance of the design is improved or
  - with probability inversely proportional to the decrease in design performance
- reduce ("cool") the acceptance probability and perturbation size, to end with a greedy algorithm
Algorithmic search (2)

$\alpha = 0.2$
Allocation of unblocked design (1)

Find an optimal design...

... for the corresponding GLM

• Particularly useful for locally optimal designs where analytic results may be available
  • Russell et al. (2009), Yang et al. (2011)

Assign treatments optimally to blocks...

... using an interchange algorithm
Allocation of unblocked design (2)

$\alpha = 0.2$
Comparison (1)

Efficiencies of robust designs...

...relative to locally optimal designs for 10,000 vectors from $B$
Empirical distribution of relative efficiencies
Comparison (3)

Comparison of designs from the three methods

• Algorithmic and allocated designs perform similarly
• Designs are fairly robust for all values of $\alpha$: median 0.70-0.76; lower quartile 0.64-0.70
• For larger values of $\alpha$, the algorithmic design has a small but consistent advantage; reversed for small $\alpha$
• For each $\alpha$, the relative efficiency distribution has a long right tail
• Designs are very robust to $\alpha$

Median efficiency of designs found for $\alpha = 0.2$

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic</td>
<td>0.67</td>
<td>0.70</td>
<td>0.71</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>Allocated</td>
<td>0.68</td>
<td>0.70</td>
<td>0.70</td>
<td>0.72</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Comparison (3)

Comparison of designs from the three methods

- Algorithmic and allocated designs perform similarly
- Designs are fairly robust for all values of $\alpha$: median 0.70-0.76; lower quartile 0.64-0.70
- For larger values of $\alpha$, the algorithmic design has a small but consistent advantage; reversed for small $\alpha$
- For each $\alpha$, the relative efficiency distribution has a long right tail
- Designs are very robust to $\alpha$

Median efficiency of designs found for $\alpha = 0.2$

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic</td>
<td>0.67</td>
<td>0.70</td>
<td>0.71</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>Allocated</td>
<td>0.68</td>
<td>0.70</td>
<td>0.70</td>
<td>0.72</td>
<td>0.74</td>
</tr>
</tbody>
</table>
Effectiveness of blocking

Efficiencies of different allocations of unblocked optimal design
Robustness of design performance to the working correlation assumption (1)

Upper bound on pairwise correlation...

... between binary observations $y_j$ and $y_k$

$$r = \min \left\{ \sqrt{ \frac{p_j p_k}{(1 - p_j)p_k} } , \sqrt{ \frac{(1 - p_j)p_k}{p_j(1 - p_k)} } \right\}$$

where $p_j = g^{-1}(\eta_j)$ (see Joe, 1997)

Set pairwise correlation to be

$$\rho = \min \{ r, \gamma \} ,$$

where $0 < \gamma < 1$ is a bound on the size of the correlation, and let $R^*$ be the resulting unstructured correlation matrix

Then an asymptotic covariance matrix for $\mathbf{Y}$ using the bounds is

$$\Sigma^* = A^{1/2} R^* A^{1/2}$$
Robustness of design performance to the working correlation assumption (2)

Compare designs under two objective functions

- Assume true covariance is equal to the “working covariance”

\[ \phi_D = |X' \Delta V^{-1} \Delta X| \]

- Assume true covariance is unstructured and derived from the bounds

\[ \psi_D = |X' \Delta V^{-1} \Delta X \left( X' \Delta V^{-1} \Sigma^* V^{-1} \Delta X \right)^{-1} X' \Delta V^{-1} \Delta X| \]
Robustness of design performance to the working correlation assumption (3)
Robustness of design performance to the working correlation assumption (4)
Comparison of marginal and conditional approaches

Example
Consider the following (marginal, conditional) distribution for the observations

- Poisson distribution: $E(Y) = \text{Var}(Y) = \mu; Y|\tau \sim \text{Poisson}(\mu)$
- Log link: $\eta = e^\mu$
- Quadratic linear predictor: $\eta = (\tau+) 5x + x^2; x \in [-1, 1]$

We find designs for blocks of size two for

- GEE with

\[ R_b = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix} \]

- Quasi-likelihood with covariance structure derived from $\tau \sim N(0, I\sigma^2)$
  - available in closed-form for the Poisson (Niaparast, 2009)

- GLMM (MQL) with $\tau \sim N(0, I\sigma^2)$
Comparison of marginal and conditional approaches

Example
Consider the following (marginal, conditional) distribution for the observations

- Poisson distribution: \( E(Y) = \text{Var}(Y) = \mu; \ Y|\tau \sim \text{Poisson}(\mu) \)
- Log link: \( \eta = e^\mu \)
- Quadratic linear predictor: \( \eta = (\tau +) 5x + x^2; \ x \in [-1, 1] \)

We find designs for blocks of size two for

- GEE with

\[
R_b = \begin{pmatrix}
1 & \alpha \\
\alpha & 1
\end{pmatrix}
\]

- Quasi-likelihood with covariance structure derived from \( \tau \sim N(0, I\sigma^2) \)
  - available in closed-form for the Poisson (Niaparast, 2009)

- GLMM (MQL) with \( \tau \sim N(0, I\sigma^2) \)
Designs

For $\alpha = 0.5$ and $\sigma^2 = 0.5$

$$\xi^*(\text{GEE}) = \begin{cases} (0.02, 0.84) & (0.72, 1) & (1, 0.26) \\ 0.38 & 0.35 & 0.27 \end{cases}$$

$$\xi^*(\text{Quasi}) = \xi^*(\text{GLMM}) = \begin{cases} (0.10, 0.88) & (0.75, 1) \\ 0.5 & 0.5 \end{cases}$$

NB: average intrablock correlation for Quasi and GLMM design is $\sim 0.49$
Optimality and efficiency

(a) GEE

(b) Quasi

(c) GLMM

(d) Efficiencies

<table>
<thead>
<tr>
<th>Design</th>
<th>Approximation</th>
<th>GEE</th>
<th>Quasi</th>
<th>GLMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEE</td>
<td>1</td>
<td>0.87</td>
<td>0.87</td>
<td></td>
</tr>
<tr>
<td>Quasi</td>
<td>0.90</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>GLMM</td>
<td>0.90</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Summary

In conclusion

- Design & analysis methods for blocked experiments with a non-normal response
- It is important to take account of blocks when designing the experiment
- Methods can also be applied to other working correlation structures – autoregressive, nearest neighbour
  - designs are also quite robust to the assumed structure
- Generalises to include uncertainty in linear predictor & link function
Ongoing work

Conditional models

• Optimal design for GLMMs - with Tim Waterhouse (Eli Lilly) and Tim Waite (Southampton)

• Optimal design for HGLMs - with Peter van de Ven and Tim Waite

Compare designs found for each model, and assess robustness to modelling method
Selected references

Conditional models

Mixed effects model

\[ g(E(Y|\tau)) = X\beta + Z\tau \]

with \( \tau \sim [0, \Sigma_{\tau}] \)

Full probabilistic treatment...

... but marginal likelihood typically requires (numerical) integration

\[ l = \log \int f(y|\tau)f(\tau)\,d\tau \]

Appropriate when...

... blocks have meaning (e.g. patients in clinical trial), the inter-block variability is of interest and subject-specific inferences are required

Note...

... conditional and marginal models are equivalent for Normally distributed observations
Quasi-likelihood

Estimating equations imply...
... an associated quasi-likelihood

\[ Q = \int_{\mu} X' \Delta W^{-1}(Y - t) \, dt + \int_{\mu} X' \Delta W^{-1}(h - t) \, dt \]

By definition, \( \partial Q/\partial \mu \) are the estimating equations, which mimic properties of the score functions for maximum likelihood estimation

Flattening constant...
... can be viewed as a quasi-prior distribution on \( \mu \) but no distributional assumptions are required
References (1)

References (2)

References (3)