Design of experiments for generalized linear models with random block effects

Tim Waite

timothy.waite@manchester.ac.uk

School of Mathematics
University of Manchester

Outline

- Background, motivation and challenges
- Generalized linear mixed models and approximations to the information matrix
- Design selection and assessment
- Illustrative examples throughout

Joint work with Prof. Dave Woods (Southampton).
Acknowledgments: EPSRC, Tim Waterhouse (Eli Lilly).

Background

- Experiments are used to investigate the impact of interventions ("treatments") on a process or system, by applying treatments to a number of "units"
- Design of experiments concerns the selection of treatments (settings of the controllable variables) to be applied
- Usually, the aim of the experiment is to collect data to answer scientific questions through the estimation of a statistical model
- The design can be selected to maximize information gained for a given level of resource ('optimal design')
  - Information is typically measured with respect to uncertainty about the model and its parameters

Motivation

(i) Increasing recognition of the need to design for experiments with non-normal response (Woods et al., 2006)

e.g. for Generalized Linear Models in areas such as

- Science - crystallography
- Engineering - aeronautics

(ii) Often, experimental units are heterogeneous (batches or repeated observations)

Current design methods for GLMs may be inefficient ...
... see Woods & van de Ven (2011)

- Include blocking or grouping to increase precision
- Design = selection of 'treatments' + allocation to blocks
- Analyse data using an appropriate model that accounts for blocking

Engine bearings

Aeronautics - Goodrich

Aim: to study the factors affecting cracking of engine bearing coatings

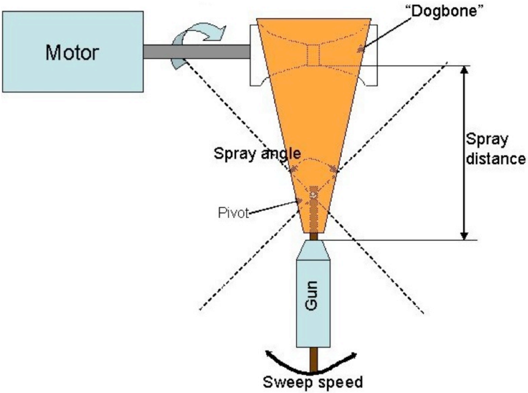
Controllable variables

- Spray distance
- Spray angle
- Sweep speed

Outcome

Each bearing passes (1) or fails (0) a visual inspection

Blocks - sessions (mornings and afternoons)



Approximate block designs

Assume q variables with treatment vector $\mathbf{x} \in \mathcal{X} = [-1, 1]^q$ and, for simplicity, all blocks have identical fixed size m

The design can be defined as

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

where

- $\zeta_k = (\mathbf{x}_{k1}, \mathbf{x}_{k2}, \dots, \mathbf{x}_{km}) \in \mathcal{X}^m$ are **distinct sets** of m treatments - support blocks
- w_k = prop^n of blocks whose units receive the treatments in set ζ_k

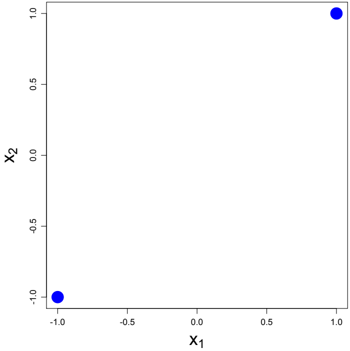
Clearly $w_k > 0$ and $\sum_{k=1}^{b} w_k = 1$

Approximate block designs

Example: $q = 2$ variables: x_1, x_2; two distinct treatment-sets of size $m = 2$

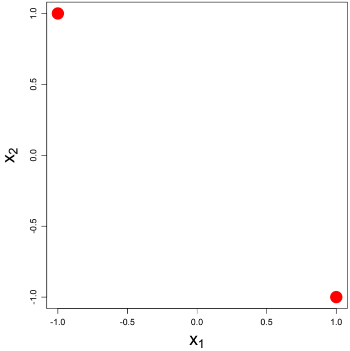
$\zeta_1 = \big((-1,-1),(1,1)\big)$

$w_1 = .5$



$\zeta_2 = \big((1,-1),(-1,1)\big)$

$w_2 = .5$



Generalized Linear Mixed Models

For jth unit in ith block $(i = 1, \ldots, n;\ j = 1, \ldots, m)$

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

where

- $\pi[\mu, \varphi V]$ a distribution from exponential family
- Associated with ith block, vector of random effects $\mathbf{u}_i \sim N_r(\mathbf{0}, G)$
- $g\{\mu(\mathbf{x}|\mathbf{u})\} = \nu(\mathbf{x}|\mathbf{u})$
  - $\nu(\mathbf{x}|\mathbf{u}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \mathbf{z}^T(\mathbf{x})\mathbf{u}$ *linear predictor: fixed & random parts*
  - $\eta(\mathbf{x}) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ *fixed part*
- $\mathbf{f} : \mathcal{X} \to \mathbb{R}^p$, $\mathbf{z} : \mathcal{X} \to \mathbb{R}^r$ are known vectors of functions
- $\boldsymbol{\beta}$ holds p unknown regression parameters

Generalized Linear Mixed Models

We focus on the special case of a random intercept

- $\mathbf{u}_i = u_i \sim N(0, \sigma^2)$ is scalar, $\mathbf{z}(\mathbf{x}) = 1$, and hence

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

- Blocks have an additive (random) effect on the scale of the linear predictor

D-optimal designs

Initially, we seek a D-optimal design, ξ^*, which maximizes

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

where $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T$, and $M_{\boldsymbol{\beta}}(\xi; \boldsymbol{\theta})$ is the information matrix for $\boldsymbol{\beta}$

- we assume a particular 'guess' for $\boldsymbol{\theta}$ – locally D-optimal design

For large n, approximately

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

See, for example, Atkinson, Donev & Tobias (2007)

Optimal designs

- Optimal designs need to be found using numerical optimization ...

  ... and hence many evaluations of the information matrix and its determinant are required

- Naïve evaluation of the information matrix will therefore be computationally infeasible ...

  ... and more efficient approximations are required

We will consider analytical and computational approximations

Information matrix approximations

Strongest results for the binary response case

- logistic regression with random intercept, $\mathbf{u}_i = u_i \sim N(0, \sigma^2)$

1. Outcome-enumeration methods
    - naïve numerical 'brute force'
    - asymptotic strong dependence (large σ^2)
    - interpolated – well-suited for Bayesian design
2. Methods based on marginal model approximations
    - marginal quasi-likelihood (MQL), generalized estimating equations (GEE)
    - attenuation-adjusted MQL and GEE
        - adjustment critical for performance of the designs

Information matrix

For a GLMM, we can write

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

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

$$= F^T E_{\mathbf{y}}\left\{p(\mathbf{y}|\zeta,\boldsymbol{\theta})^{-2}\left(\frac{\partial p(\mathbf{y}|\zeta,\boldsymbol{\theta})}{\partial\boldsymbol{\eta}}\right)\left(\frac{\partial p(\mathbf{y}|\zeta,\boldsymbol{\theta})}{\partial\boldsymbol{\eta}}\right)^T\right\} F$$

- $p(\mathbf{y}|\zeta,\boldsymbol{\theta})$ is the probability of observing $\mathbf{y} \in \mathbb{R}^m$ from block $\zeta = (\mathbf{x}_1,\ldots,\mathbf{x}_m)$ (the marginal likelihood)
- $F = [\mathbf{f}(\mathbf{x}_1),\ldots,\mathbf{f}(\mathbf{x}_m)]^T$ (the model matrix)
- $\boldsymbol{\eta} = [\eta(\mathbf{x}_1),\ldots,\eta(\mathbf{x}_m)]^T$ (vector of fixed parts of linear predictors)

Outcome enumeration

For binary data and small blocks, the information matrix can be evaluated by outcome enumeration

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

- Evaluate integrals using Gauss-Hermite quadrature
- Computationally expensive, but can be used for assessment of other methods
- For GLMMs with other conditional response distributions (e.g. Poisson), Monte Carlo approximations are possible (but even more expensive)

Quasi-likelihood approximations (I)

Marginal quasi-likelihood (MQL)

$$M_\beta^{\text{marg}}(\xi, \boldsymbol{\theta}) = \sum_{k=1}^{b} w_k\, F_k^T V_k^{-1} F_k$$

- F_k, Z_k are respectively the fixed and random effects model matrices for ζ_k
- $V_k = \mathcal{V}(\zeta_k, \boldsymbol{\theta})$ is determined from $\mathcal{V}(\zeta, \boldsymbol{\theta}) = W(\zeta, \boldsymbol{\theta})^{-1} + Z(\zeta) G Z(\zeta)^T$
- $W(\zeta, \boldsymbol{\theta})$ is the diagonal matrix with entries $v(\mathbf{x}_1; \mathbf{0}, \boldsymbol{\beta}), \ldots, v(\mathbf{x}_m; \mathbf{0}, \boldsymbol{\beta})$

- arises from small-**u** Taylor expansion of the model (Breslow & Clayton, 1993)
- derivation relies on intermediate approximation

$$E(y_{ij}) \approx g^{-1}\{f^T(\mathbf{x}_{ij})\boldsymbol{\beta}\}$$

- for design using similar methods, see Moerbeek & Maas (2005)

Quasi-likelihood approximations (II)

Adjusted marginal quasi-likelihood (AMQL)

For logistic models, a better approximation is

$$E(y_{ij}) \approx g^{-1}\{f^T(\mathbf{x}_{ij})\boldsymbol{\beta}_{\text{adj}}\}$$

where $$\boldsymbol{\beta}_{\text{adj.}} = \boldsymbol{\beta}\,(1 + c^2\sigma^2)^{-1/2}, \quad \boldsymbol{\theta}_{\text{adj.}} = (\boldsymbol{\beta}_{\text{adj.}}, \sigma^2)^T\,,$$

and $c = 15\sqrt{3}/(16\pi)$ see e.g. Breslow and Clayton (1993)

Marginal model for mean is approximately logistic with attenuated parameters

To obtain efficient designs using MQL, we found it critical to adjust for this attenuation when using MQL. Define:

$$M_{\boldsymbol{\beta}}^{\text{AMQL}}(\xi; \boldsymbol{\theta}) = M_{\boldsymbol{\beta}}^{\text{marg}}(\xi; \boldsymbol{\theta}_{\text{adj}})$$

Example 1: binary data (1)

$m = 4$, $\mathbf{x} = (x_1, x_2)$

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

Compare locally optimal designs from AMQL method and outcome enumeration.

Parameter scenarios

Want to investigate performance of methods as intra-block correlation increases
- better to specify parameter scenarios on the marginal scale $(\boldsymbol{\beta}_{\text{adj}}, \sigma^2)$ in this case

Example

Efficiency, $\text{eff}(\xi; \boldsymbol{\theta}) = \{|M_{\boldsymbol{\beta}}(\xi; \boldsymbol{\theta})| / \sup_{\xi'} |M_{\boldsymbol{\beta}}(\xi'; \boldsymbol{\beta})|\}^{1/p} \times 100\%$

		σ^2				
β_{adj}^T	Approx	2	5	10	20	50
(0,1,1)	AMQL	100	100	100	100	100
(0,3,2)	AMQL	99.9	100.0	99.9	99.4	95.2
(1,2,3)	AMQL	99.1	96.6	92.1	84.8	**73.5**
(1,4,4)	AMQL	100.0	99.9	99.4	98.2	97.2
(1,3,3)	AMQL	99.5	100.0	99.3	97.9	95.1
(1,2,2)	AMQL	100.3	100.2	98.5	96.1	92.6
(2,1,3)	AMQL	99.1	96.6	92.1	84.8	**78.0**

Table: Efficiencies (%) of adjusted MQL designs, compared to the naïve outcome enumeration design

Outcome enumeration - asymptotic

Asymptotic approximations (large σ^2)

- Study the limit $\sigma^2 \to \infty$, with appropriate regularity conditions on the design and on $\boldsymbol{\beta}$. Focus on a single block $\zeta = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$.
  - $\boldsymbol{\beta} = \boldsymbol{\beta}_{\text{att}}\sqrt{1 + c^2\sigma^2}$, with $\boldsymbol{\beta}_{\text{att}}$ fixed
  - allow the $\mathbf{x}_j$ to vary
  - For all j, either $\eta_j^* = \mathbf{f}^T(\mathbf{x}_j)\boldsymbol{\beta}_{\text{att}}$ is constant or there is l with η_l^* constant and $\eta_j^* - \eta_l^* = o(\sigma^{-1})$.
- Approximations can be derived for the likelihood and derivatives, and substituted into our outcome enumeration formula.
- The above framework gives a better approximation to $M_{\boldsymbol{\beta}}(\zeta; \boldsymbol{\theta})$ than $\mathbf{x}_j$ fixed when $\eta_l^* \approx \eta_j^*$ $(j \neq l)$

Example

Efficiency, $\text{eff}(\xi;\boldsymbol{\theta}) = \{|M_{\boldsymbol{\beta}}(\xi;\boldsymbol{\theta})| / \sup_{\xi'} |M_{\boldsymbol{\beta}}(\xi';\boldsymbol{\beta})|\}^{1/p} \times 100\%$

		σ^2				
β_{adj}^T	Approx	2	5	10	20	50
(0,1,1)	AMQL	100	100	100	100	100
	Asymptotic	-	-	-	100.0	94.8
(0,3,2)	AMQL	99.9	100.0	99.9	99.4	95.2
	Asymptotic	-	-	-	96.4	97.4
(1,2,3)	AMQL	99.1	96.6	92.1	84.8	**73.5**
	Asymptotic	-	-	-	93.5	**98.3**
(1,4,4)	AMQL	100.0	99.9	99.4	98.2	97.2
	Asymptotic	-	-	-	97.1	98.2
(1,3,3)	AMQL	99.5	100.0	99.3	97.9	95.1
	Asymptotic	-	-	-	97.1	98.2
(1,2,2)	AMQL	100.3	100.2	98.5	96.1	92.6
	Asymptotic	-	-	-	95.1	97.9
(2,1,3)	AMQL	99.1	96.6	92.1	84.8	**78.0**
	Asymptotic	-	-	-	95.0	**98.0**

Table: Efficiencies (%) of adjusted MQL and asymptotic (large σ^2) designs, compared to the naïve outcome enumeration design

Interpolation (1)

Suppose that $\lambda : \mathcal{D} \rightarrow \mathbb{R}$ is an expensive-to-evaluate function.

Often it is faster to use a cheaper numerical approximation, constructed as follows:

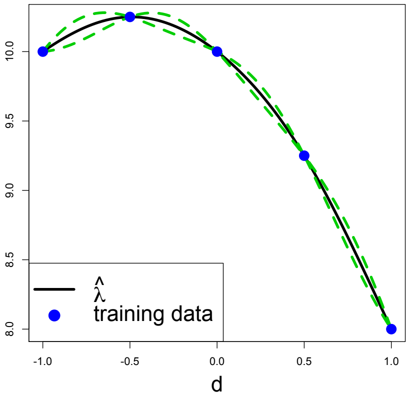
- Evaluate $\lambda(\mathbf{d}_1), \ldots, \lambda(\mathbf{d}_n)$ for a set of training points, $\mathbf{d}_i \in \mathcal{D}$
- Fit a model, $\hat{\lambda}$, that interpolates the data
- We say that $\hat{\lambda}$ is an emulator of λ

Then $\hat{\lambda}(\mathbf{d})$ can be used to predict $\lambda(\mathbf{d})$ for $\mathbf{d}$ outside of the training set.

[For multivariate $\mathbf{d}$, we use space-filling designs and Gaussian process models, c.f. computer experiments literature, e.g. Santner et al., 2003]

Interpolation (2)

Toy example



An emulator built using a training set of 5 points

Parameter dependence

Optimal designs for estimating β depend on the unknown parameters...
...overcome this with a pseudo-Bayesian D-optimal design ξ^*, maximizing

$$\Psi_D(\xi) = \int_\Theta \log |M(\xi; \boldsymbol{\theta})| \, d\mathcal{G}(\boldsymbol{\theta})$$

where $\mathcal{G}$ is a distribution across the parameter space Θ

Approximation of objective function
is via numerical quadrature

Outcome enumeration - interpolated

For random intercept logistic regression, we can write the information matrix for block ζ as

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

where $W(\boldsymbol{\eta},\sigma^2)$ is a function of $p_{\mathbf{y}} = p(\mathbf{y}|\zeta,\boldsymbol{\theta}) = p(\mathbf{y}|\boldsymbol{\eta},\sigma^2)$ and $\partial p_{\mathbf{y}}/\partial\eta_j$, $\mathbf{y} \in \{0,1\}^m$, $j = 1,\ldots,m$.

To approximate the information matrix, we build an emulator for W.

- As the emulator is in terms of $\boldsymbol{\eta}$, we can efficiently produce Bayesian designs
- To obtain a highly accurate approximation, we use a large training set and compactly-supported covariance functions

Example 2: binary data (1)

Set $m = 4$, and $\mathbf{x} = (x_1, x_2)$ (two variables)

Model:

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

Prior:

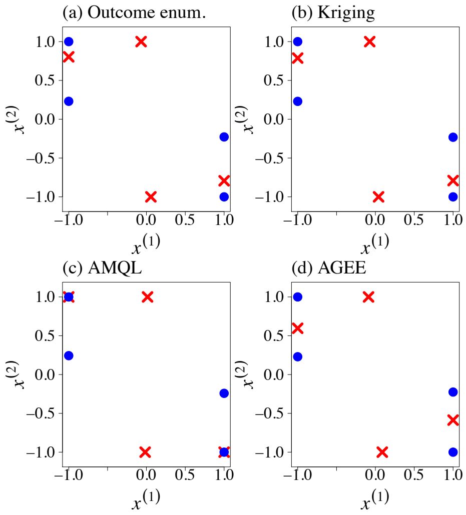
$$\beta_0 \sim U(-0.5, 0.5) \qquad \beta_1 \sim U(3, 5)$$
$$\beta_2 \sim U(0, 10) \qquad \sigma^2 = 5$$

Find pseudo-Bayesian D-optimal designs, using a Latin hypercube sample, $\boldsymbol{\beta}^{(1)}, \ldots \boldsymbol{\beta}^{(50)}$ from $[-0.5, 0.5] \times [3, 5] \times [0, 10]$ to approximate the integral:

$$\Psi_D(\xi) \approx \sum_{s=1}^{50} \frac{1}{50} \log |M_{\boldsymbol{\beta}}(\xi; \boldsymbol{\theta}_s)|$$

with $\boldsymbol{\theta}_s = (\boldsymbol{\beta}^{(s)T}, 5)^T$.

Example 2: binary data (2)



Example 2: binary data (3)

	Block weights			
Design method	•	×	Bayes efficiency (%)	Time (s)
Outcome enum.	0·744	0·256	100·00	$1{\cdot}65\times10^{7}$
Interpolation	0·749	0·251	99·96	$5{\cdot}19\times10^{6}$
AMQL	0·748	0·252	99·79	$1{\cdot}80\times10^{5}$
AGEE	0·466	0·534	97·94	$2{\cdot}20\times10^{5}$

$$\text{Bayes-eff}(\xi;\boldsymbol{\theta}) = \left[\exp\left\{\Psi_D(\xi)\right\} / \exp\left\{\Psi_D(\xi^\star)\right\}\right]^{1/p}$$

Conclusions

- A range of further examples and approximations studied
- For small σ^2
  - Most approximations give similar designs
  - Performance is comparable to outcome enumeration
- Simulation results suggest small-sample ranking and performance of designs is consistent with these (large n) asymptotic results
- For (very much) larger σ^2
  - Both AMQL and AGEE can perform poorly, compared to outcome enumeration
  - Asymptotic approximations may be more effective

Waite, T.W., & Woods, D.C. (2015). Designs for generalized linear models with random block effects via information matrix approximations. Biometrika, accepted (arXiv:1412.4355).

Related work

- Further design problems
  - More complex random effect structures (e.g. split-plots)
  - Estimation of variance components
  - Prediction of random effects (HGLMs)
- Extensions of computational methodology (e.g. interpolation) to design for other nonlinear models
  - e.g., see Overstall & Woods (2015), (tech. rep., arXiv:1501.00264)
- Applications, e.g. in Biostatistics

References

Atkinson, A.C., Donev, A.N. and Tobias, R.D. (2007). Optimum Design of Experiments, with SAS. OUP.

Breslow, N. & Clayton, D. (1993), *JASA*, 88, 9–25

Chaganty, N.R. & Joe, H. (2004), *JRSSB*, 66, 851-860.

Chaloner, K. & Verdinelli, I. (1995), *Stat. Sci.*, 10, 273–304

Furrer, R., Nychka, D. & Stain, S. (2012), `R` *package* `fields`, *CRAN*.

Godambe, V.P. (Ed.) (1991). Estimating Equations. OUP.

Liang, K.-Y. & Zeger, S.L. (1986), *Biometrika*, 73, 13–22

Moerbeek, M. & Maas, C. (2005), *Comm. Stat. Th. Meth.*, 34, 1151–1167

Müller, P. & Parmigiani, G. (1995) *JASA*, 90, 1322–1330

Niaparast, M. & Schwabe, R. (2013), *JSPI*, 143, 296–306

Santner, T.J., Williams, B.J. & Notz, W.I. (2003). The Design and Analysis of Computer Experiments. Springer.

Schonlau, M. and Welch, W.J. (2006). Chapter in Dean, A. and Lewis, S. (2006), Screening. Springer.

Tekle, F., Tan, F. & Berger, M. (2008), *CSDA*, 52, 5253–5262

Wedderburn, R.W.M. (1974), *Biometrika*, 61, 439–447

Woods, D. & van de Ven, P. (2011), *Technometrics*, 53, 173–182

Woods, D., Lewis, S., Eccleston, J. & Russell, K. (2006), *Technometrics*, 48, 284-292

Back-up slides

Further details of asymptotics

Define two partitions of the indices $\{1, \ldots, m\}$:

- $\mathcal{S}_{0,1} = \{i : y_i = 0, 1 \text{ respectively}\}$
- $\mathcal{N}(j)$, $\mathcal{Z}(j)$, $\mathcal{P}(j)$ the set of indices l such that $\lim_{\sigma\to\infty}[\eta_l^* - \eta_j^*] < 0$, $= 0$ and > 0, respectively.

Two important subclasses of outcomes $\mathbf{y} = (y_1, \ldots, y_m)^T$:

- increasing : $\exists j'$ with

$$\eta_j^* < \eta_{j'}^* \implies y_j = 0$$
$$\eta_j^* > \eta_{j'}^* \implies y_j = 1$$

- quasi-increasing : $\exists j'$ with $\mathcal{N}(j') \subseteq \mathcal{S}_0$, $\mathcal{P}(j') \subseteq \mathcal{S}_1$.
- contributions to $M(\zeta; \boldsymbol{\theta})$ from these classes of $\mathbf{y}$ are $O(\sigma^{-1})$ and $O(\sigma^{-2})$
- contributions from other outcomes negligible: $O(\sigma^{-1}e^{-\sigma B})$, $B > 0$.

Theorem: approximation of the likelihood

Suppose that the outcome is quasi-increasing:

(i) If $|\mathcal{S}_0 \cap \mathcal{Z}(j')| = 0$ or $|\mathcal{S}_1 \cap \mathcal{Z}(j')| = 0$, as $\sigma^2 \to \infty$,

$$P(Y|\theta,\zeta) = \max\left\{0, \Phi\left(-\max_{j\in\mathcal{S}_0}\{\eta_j/\sigma\}\right) - \Phi\left(-\min_{j\in\mathcal{S}_1}\{\eta_j/\sigma\}\right)\right\} + O(\sigma^{-1})$$

(ii) If $|\mathcal{S}_0 \cap \mathcal{Z}(j')| \geq 1$ and $|\mathcal{S}_1 \cap \mathcal{Z}(j')| \geq 1$, then as $\sigma^2 \to \infty$,

$$P(Y|\theta,\zeta) = \frac{\phi(\eta_{j'}/\sigma)}{\sigma}\int_{-\infty}^{\infty}\{1-h(t)\}^{|\mathcal{S}_0\cap\mathcal{Z}(j')|}h(t)^{|\mathcal{S}_1\cap\mathcal{Z}(j')|}dt + \sum_{l\in\mathcal{Z}(j')} O(\Delta_{lj'}/\sigma) + O(\sigma^{-2})$$

where $\Delta_{lj} = \eta_l - \eta_j$, and h is the logistic function.