Optimal design for generalised linear mixed models

Tim Waterhouse

School of Physical Sciences, University of Queensland

with

John Eccleston, School of Physical Sciences, UQ

Steve Duffull, School of Pharmacy, UQ

Motivation

- Pharmacodynamic study

Motivation

- Pharmacodynamic study
  - Drug with binary response
  - Need to design cross-over trials to test effect of various dose levels
  - Logistic regression
  - D-optimal designs

Motivation

- Pharmacodynamic study
  - Drug with binary response
  - Need to design cross-over trials to test effect of various dose levels
  - Logistic regression
  - D-optimal designs
- What about subject effect?

Overview

- Need information matrix for logistic regression model with random coefficients

Overview

- Need information matrix for logistic regression model with random coefficients
- Longford (1994) presents an approximation based on Taylor expansion of log-likelihood

Overview

- Need information matrix for logistic regression model with random coefficients
- Longford (1994) presents an approximation based on Taylor expansion of log-likelihood
- Numerical investigation showed large discrepancies

Overview

- Need information matrix for logistic regression model with random coefficients
- Longford (1994) presents an approximation based on Taylor expansion of log-likelihood
- Numerical investigation showed large discrepancies
- Aim: to develop a more accurate approximation

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}$$

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}$$

- Exact designs: where $w_i = r_i/N$, with r_i replications at x_i

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} x_1 & x_2 & \ldots & x_n \\ w_1 & w_2 & \ldots & w_n \end{Bmatrix}$$

- Exact designs: where $w_i = r_i/N$, with r_i replications at x_i
- Approximate designs: where weights $w_i \in [0,1]$ sum to 1, represent proportion of experimental effort at each point

Optimal design

- Given the form of the model, how do we best select design ξ to efficiently estimate parameter vector θ?

Optimal design

- Given the form of the model, how do we best select design ξ to efficiently estimate parameter vector θ?
- Cramer-Rao Lower Bound:
  Var($\hat{\theta}$) $\geq$ (Fisher information matrix)$^{-1}$

Optimal design

- Given the form of the model, how do we best select design ξ to efficiently estimate parameter vector θ?
- Cramer-Rao Lower Bound:
  Var($\hat{\theta}$) $\geq$ (Fisher information matrix)$^{-1}$
- D-optimality (Smith, 1918; Box and Lucas, 1959): maximise determinant of Fisher information matrix

D-optimality

- Fisher information matrix:

$$M_F(\theta, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\theta; y)}{\partial\theta\partial\theta^T}\right]$$

D-optimality

- Fisher information matrix:

$$M_F(\theta, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\theta; y)}{\partial\theta\partial\theta^T}\right]$$

- For nonlinear models (including logistic regression), $M_F(\theta, \xi)$ involves unknown θ

D-optimality

- Fisher information matrix:

$$M_F(\theta, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\theta; y)}{\partial\theta\partial\theta^T}\right]$$

- For nonlinear models (including logistic regression), $M_F(\theta, \xi)$ involves unknown θ
- So D-optimal design depends on θ, such designs are *locally* optimal

Logistic regression

- Fixed effects model:

$$\eta_j = \text{logit}\{P(y_j = 1)\} = \log\{p_j/(1 - p_j)\} = x_j\theta$$

so that

$$\eta = X\theta$$

Logistic regression

- Fixed effects model:

$$\eta_j = \mathrm{logit}\{P(y_j = 1)\} = \log\{p_j/(1-p_j)\} = x_j\theta$$

so that

$$\eta = X\theta$$

- Fisher information matrix:

$$M_F(\theta, \xi) = X^T W X$$

where

$$W = \mathrm{diag}\{w_j p_j (1-p_j)\}$$

Logistic regression example

- Model (Atkinson & Haines, 1996):

$$\eta = \mathrm{logit}\{p\} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$(x_1, x_2) \in [-1,\, 1] \times [-1,\, 1]$

Logistic regression example

- Model (Atkinson & Haines, 1996):

$$\eta = \text{logit}\{p\} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$(x_1, x_2) \in [-1,\ 1] \times [-1,\ 1]$

- Locally D-optimal design for $\theta = (3, 3, 1)^{\mathrm{T}}$:

$$\xi = \begin{Bmatrix} (-1, -1) & (-1, 1) & (-0.068, -1) \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}$$

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$
- Alternatively, $\theta_j = \beta + b_j$, where $b_j \sim \mathrm{N}(0, \Sigma)$

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$
- Alternatively, $\theta_j = \beta + b_j$, where $b_j \sim \mathrm{N}(0, \Sigma)$
- New model, p parameters, $q \leq p$ of these random:

$$\eta_{ij} = \mathrm{logit}\{p_{ij}\} = x_{ij}\beta + z_{ij}b_j$$

where $b_j \sim \mathrm{N}(0, \Sigma)$ is vector of length q,
z_{ij} are the relevant subsets of vectors x_{ij}

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$
- Alternatively, $\theta_j = \beta + b_j$, where $b_j \sim \mathrm{N}(0, \Sigma)$
- New model, p parameters, $q \leq p$ of these random:

  $$\eta_{ij} = \mathrm{logit}\{p_{ij}\} = x_{ij}\beta + z_{ij}b_j$$

  where $b_j \sim \mathrm{N}(0, \Sigma)$ is vector of length q,
  z_{ij} are the relevant subsets of vectors x_{ij}
- Assume elements of b_j are uncorrelated, i.e.
  $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_q^2\}$

Random coefficients

- Parameter vector for jth subject: $\theta_j \sim \mathrm{N}(\beta, \Sigma)$
- Alternatively, $\theta_j = \beta + b_j$, where $b_j \sim \mathrm{N}(0, \Sigma)$
- New model, p parameters, $q \leq p$ of these random:

$$\eta_{ij} = \mathrm{logit}\{p_{ij}\} = x_{ij}\beta + z_{ij}b_j$$

  where $b_j \sim \mathrm{N}(0, \Sigma)$ is vector of length q,
  z_{ij} are the relevant subsets of vectors x_{ij}
- Assume elements of b_j are uncorrelated, i.e. $\Sigma = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_q^2\}$
- Parameters: $\psi = (\beta^{\mathrm{T}}, \sigma_1^2, \ldots, \sigma_q^2)^{\mathrm{T}}$

Log-likelihood

$$\ell = \sum_j \log \int \underset{\mathbb{R}^q}{\cdots} \int P_j(b_j)\Phi_q(b_j)\mathrm{d}b_j$$

where

$$\begin{aligned} P_j(b_j) &= \prod_{i=1}^{n_j} \{p_{ij}(b_j)\}^{y_{ij}} \{1 - p_{ij}(b_j)\}^{1-y_{ij}} \\ p_{ij}(b_j) &= \mathrm{logit}^{-1}\{x_{ij}\beta + z_{ij}b_j\} \end{aligned}$$

Log-likelihood

$$\ell = \sum_j \log \int \underset{\mathbb{R}^q}{\cdots} \int P_j(b_j)\Phi_q(b_j)\mathrm{d}b_j$$

where

$$\begin{aligned} P_j(b_j) &= \prod_{i=1}^{n_j} \{p_{ij}(b_j)\}^{y_{ij}} \{1 - p_{ij}(b_j)\}^{1-y_{ij}} \\ p_{ij}(b_j) &= \mathrm{logit}^{-1}\{x_{ij}\beta + z_{ij}b_j\} \end{aligned}$$

- Gaussian quadrature
- Taylor expansion

Taylor approximation to log-likelihood

From Longford (1994),

$$\ell \;\approx\; \sum_{j=1}^{N} \left[\log\{P_j(0)\} + \frac{1}{2} e_j^{\mathrm{T}} H_j e_j - \frac{1}{2} \log |G_j| \right]$$

where

$$\begin{aligned} G_j &= I_q + \Sigma^{1/2} Z_j^{\mathrm{T}} W_j Z_j \Sigma^{1/2}, \quad \text{and} \\ H_j &= W_j Z_j \Sigma^{1/2} G_j^{-1} \Sigma^{1/2} Z_j^{\mathrm{T}} W_j \\ W_j &= \mathrm{diag}\{w_j\}, \quad w_j = (w_{1j}, \ldots, w_{n_j j})^{\mathrm{T}} \\ w_{ij} &= p_{ij}(0)(1 - p_{ij}(0)) \end{aligned}$$

etc.

Approximate information matrix

- Longford (1994) ignores dependence of w_{ij} on on β to give

$$M_F(\psi, \xi_j) \approx \begin{bmatrix} A_j & 0 \\ 0 & B_j \end{bmatrix},$$

where

$$A_j = X_j^{\mathrm{T}} V_j^{-1} X_j,$$

and the (u, v) element of B_j is

$$B_j^{(u,v)} = \frac{1}{2}\mathrm{tr}\left(V_j^{-1} \frac{\partial V_j}{\partial \sigma_u^2} V_j^{-1} \frac{\partial V_j}{\partial \sigma_v^2} \right)$$

where $V_j = W_j^{-1} + Z_j \Sigma Z_j^{\mathrm{T}}$

Less approximate information matrix?

- What if we acknowledge the dependence of w_{ij} on on β?

Less approximate information matrix?

- What if we acknowledge the dependence of w_{ij} on on β?
- $M_F(\psi, \xi_j) \approx$ 10 pages of LaTeX

Brute force numerical methods

- Recall $M_F(\psi, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\psi;y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right]$

Brute force numerical methods

- Recall $M_F(\psi, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\psi; y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right]$
- For an exact design, we can calculate the expectation by the following sum:

$$\mathrm{E}\left[\frac{\partial^2 \ell(\psi; y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right] = \sum_{y_{ij}} \frac{\partial^2 \ell(\psi; y_{ij})}{\partial\psi\partial\psi^{\mathrm{T}}} \ell(\psi; y_{ij})$$

where the derivatives are evaluated numerically

Brute force numerical methods

- Recall $M_F(\psi, \xi) = -\mathrm{E}\left[\frac{\partial^2 \ell(\psi;y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right]$
- For an exact design, we can calculate the expectation by the following sum:

$$\mathrm{E}\left[\frac{\partial^2 \ell(\psi;y)}{\partial\psi\partial\psi^{\mathrm{T}}}\right] = \sum_{y_{ij}} \frac{\partial^2 \ell(\psi;y_{ij})}{\partial\psi\partial\psi^{\mathrm{T}}} \ell(\psi;y_{ij})$$

  where the derivatives are evaluated numerically
- ℓ may be evaluated
  - by Gaussian quadrature, or
  - Taylor approximation

Summary of methods

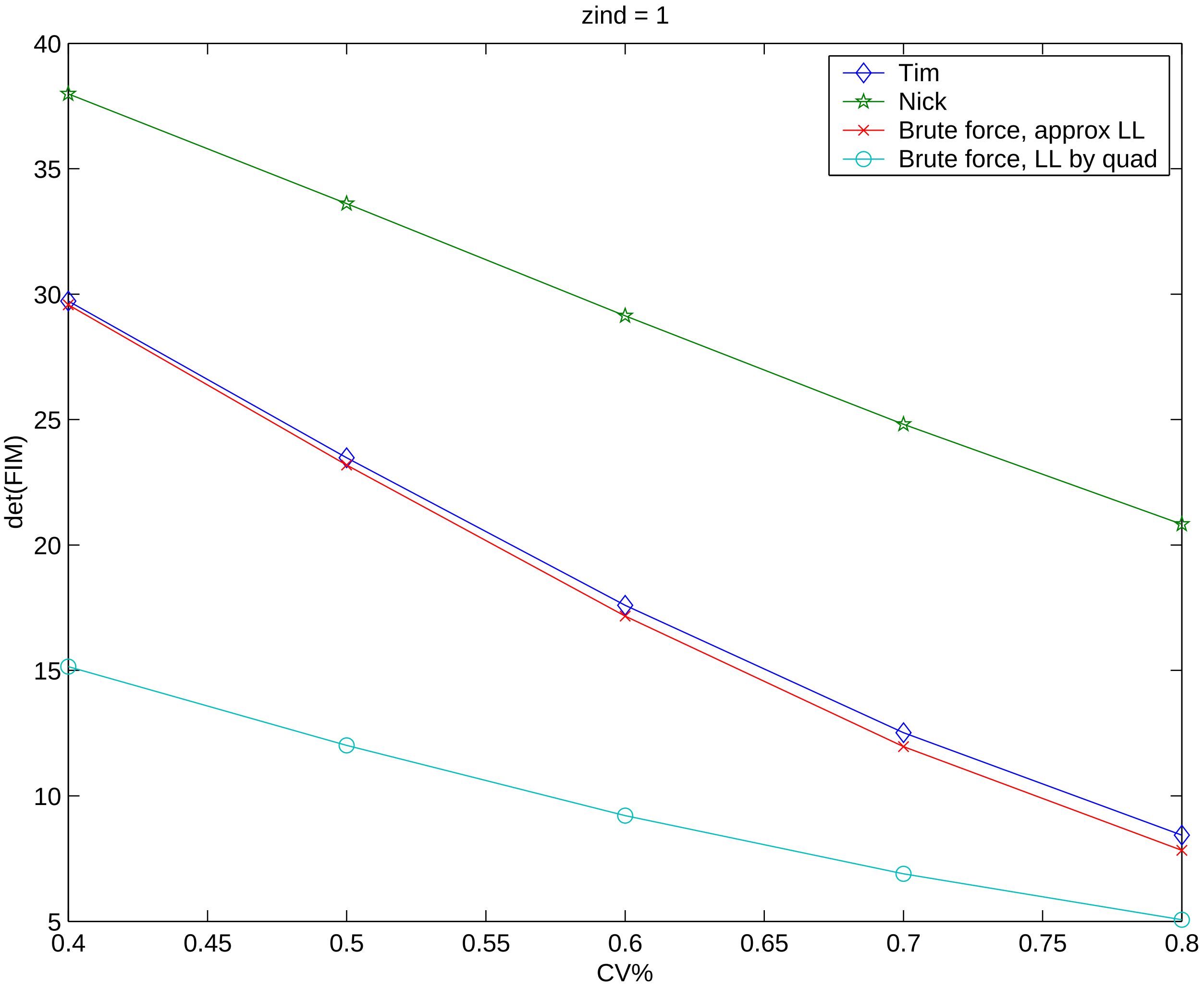
	Derivatives	Log-likelihood	w_{ij} depends on β?
Nick [Longford]	Analytic	Approx.	No
Tim [me]	Analytic	Approx.	Yes
Brute force 1	Numerical	Approx.	N/A
Brute force 2	Numerical	Quadrature	N/A

Testing. . .

$$\text{logit}\{p_j\} = (\beta_0 + b_j) + \beta_1 x_1 + \beta_2 x_2$$

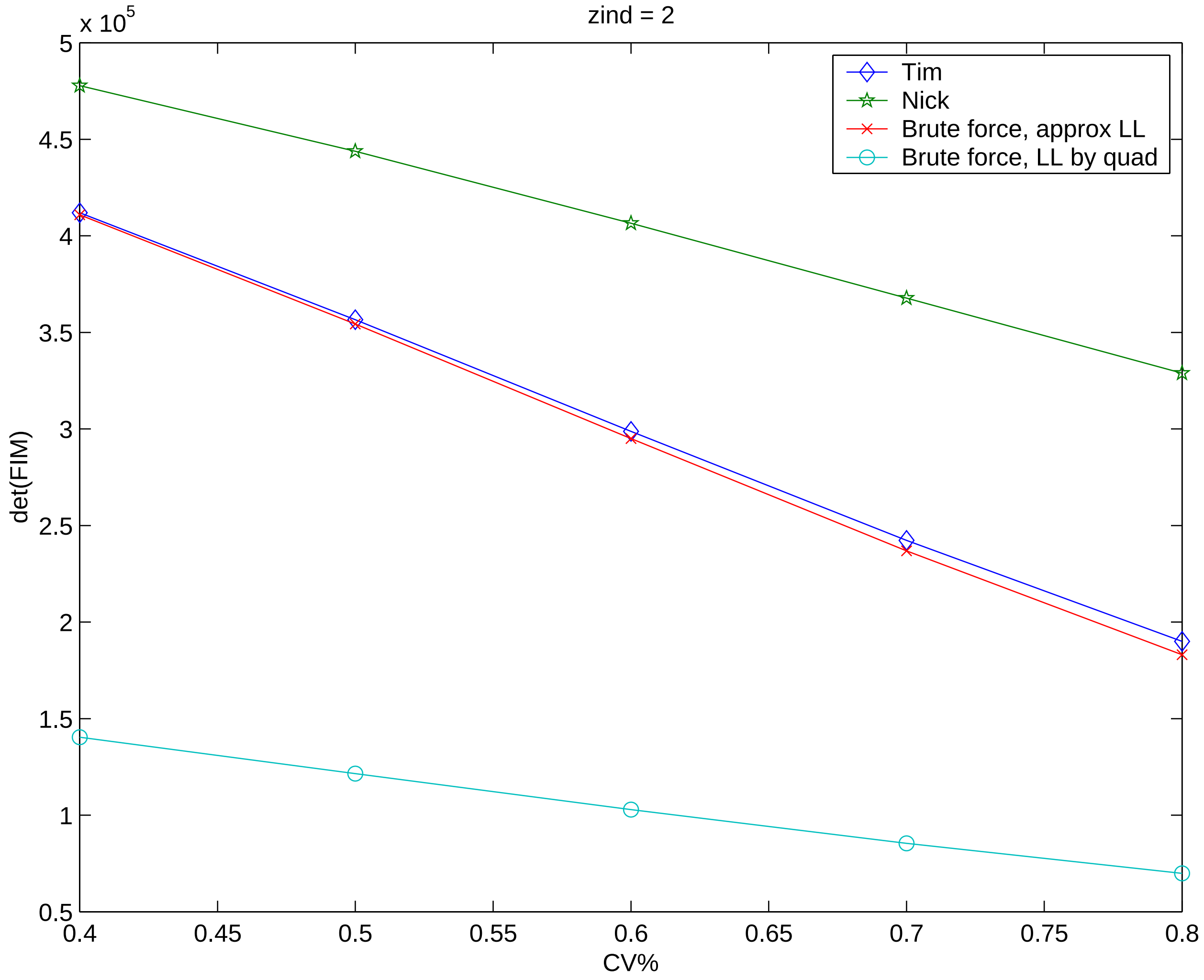
Testing. . .

$$\text{logit}\{p_j\} = (\beta_0 + b_j) + \beta_1 x_1 + \beta_2 x_2$$



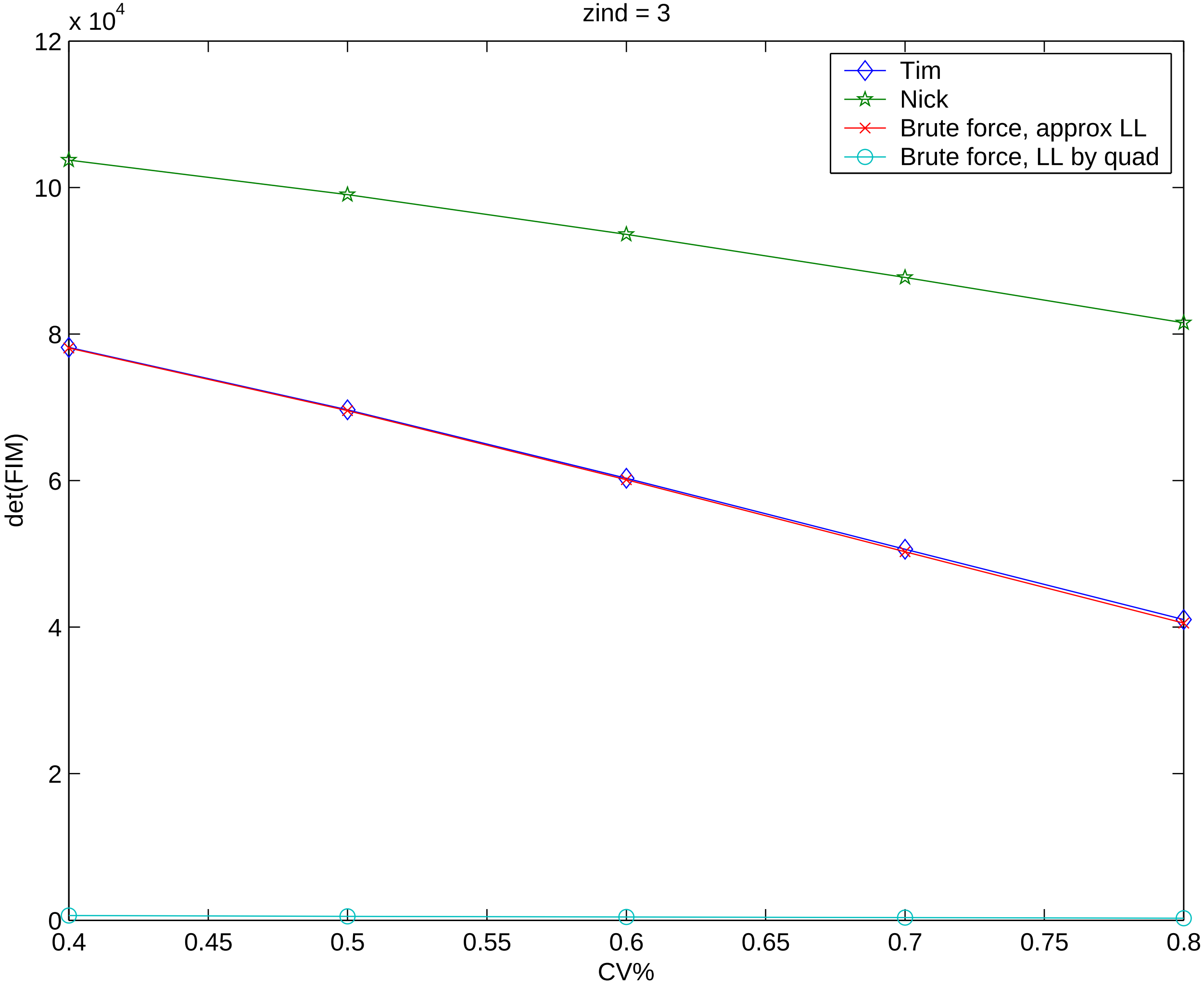
Testing. . .

$$\mathrm{logit}\{p_j\} = \beta_0 + (\beta_1 + b_j)x_1 + \beta_2 x_2$$



Testing. . .

$$\text{logit}\{p_j\} = \beta_0 + \beta_1 x_1 + (\beta_2 + b_j) x_2$$



Discussion

- Acknowledging the dependence of w_{ij} on β gives a closer approximation to the true information matrix

Discussion

- Acknowledging the dependence of w_{ij} on β gives a closer approximation to the true information matrix
- The Taylor approximation to the log-likelihood still adds a large degree of error

Discussion

- Acknowledging the dependence of w_{ij} on β gives a closer approximation to the true information matrix
- The Taylor approximation to the log-likelihood still adds a large degree of error
- Higher order Taylor expansion would make algebra even more unmanageable

Discussion

- Acknowledging the dependence of w_{ij} on β gives a closer approximation to the true information matrix
- The Taylor approximation to the log-likelihood still adds a large degree of error
- Higher order Taylor expansion would make algebra even more unmanageable
- Next step: optimal design!