Optimal design for discrimination and estimation in generalised linear models

John Eccleston and Tim Waterhouse

School of Physical Sciences, University of Queensland

Generalised linear model

- Response vector: $\boldsymbol{Y}$, $\mathrm{E}(Y_j) = \pi_j$

Generalised linear model

- Response vector: $\boldsymbol{Y}$, $\mathrm{E}(Y_j) = \pi_j$
- Linear predictor: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$

Generalised linear model

- Response vector: $\boldsymbol{Y}$, $\mathrm{E}(Y_j) = \pi_j$
- Linear predictor: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$
- Link function: $\eta_j = g(\pi_j)$

Generalised linear model

- Response vector: $\boldsymbol{Y}$, $\mathrm{E}(Y_j) = \pi_j$
- Linear predictor: $\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}$
- Link function: $\eta_j = g(\pi_j)$
- eg. Logistic regression: Bernoulli response, logit link function

$$\eta_j = \mathrm{logit}\{P(y_j = 1)\} = \log\left\{\frac{\pi_j}{1 - \pi_j}\right\}$$

Example: Nested logistic regression models

$$
\begin{aligned}
M_1: \qquad & \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\
M_2: \qquad & \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2 \\
& \qquad\qquad -1 \leq x_1, x_2 \leq 1
\end{aligned}
$$

Talk outline

- Approximate designs

Talk outline

- Approximate designs
- Optimal designs
  - for model discrimination
  - for parameter estimation

Talk outline

- Approximate designs
- Optimal designs
  - for model discrimination
  - for parameter estimation
- Evaluation of optimal designs

Talk outline

- Approximate designs
- Optimal designs
    - for model discrimination
    - for parameter estimation
- Evaluation of optimal designs
- Hybrid designs for compromising over criteria

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}$$

- $\boldsymbol{\xi}_i$ are support points, w_i are design weights

Experimental design

- An experimental design on n points:

$$\xi = \begin{Bmatrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{Bmatrix}$$

- $\boldsymbol{\xi}_i$ are support points, w_i are design weights
- Exact designs: where $w_i = r_i/N$, with r_i replications at $\boldsymbol{\xi}_i$

Experimental design

- An experimental design on n points:

$$\xi = \left\{ \begin{matrix} \boldsymbol{\xi}_1 & \boldsymbol{\xi}_2 & \dots & \boldsymbol{\xi}_n \\ w_1 & w_2 & \dots & w_n \end{matrix} \right\}$$

- $\boldsymbol{\xi}_i$ are support points, w_i are design weights
- Exact designs: where $w_i = r_i/N$, with r_i replications at $\boldsymbol{\xi}_i$
- Approximate designs: where weights $w_i \in [0, 1]$ sum to 1, represent proportion of experimental effort at each point

Designs for model discrimination

- T-optimality for nonlinear models (Atkinson & Fedorov, 1975): assume model 2 true, choose ξ to maximise $\sum_{i=1}^{n} w_i\{\eta_2(\boldsymbol{\xi}_i) - \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1)\}^2$

Designs for model discrimination

- T-optimality for nonlinear models (Atkinson & Fedorov, 1975): assume model 2 true, choose ξ to maximise $\sum_{i=1}^{n} w_i\{\eta_2(\boldsymbol{\xi}_i) - \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1)\}^2$
- T-optimality for GLMs (Ponce de Leon & Atkinson, 1992)

Designs for model discrimination

- T-optimality for nonlinear models (Atkinson & Fedorov, 1975): assume model 2 true, choose ξ to maximise $\sum_{i=1}^{n} w_i \{\eta_2(\boldsymbol{\xi}_i) - \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1)\}^2$
- T-optimality for GLMs (Ponce de Leon & Atkinson, 1992)
  - For nested GLMs, assume larger model M_2 is true, with $\boldsymbol{\pi}_2$ known

Designs for model discrimination

- T-optimality for nonlinear models (Atkinson & Fedorov, 1975): assume model 2 true, choose ξ to maximise $\sum_{i=1}^{n} w_i\{\eta_2(\boldsymbol{\xi}_i) - \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1)\}^2$
- T-optimality for GLMs (Ponce de Leon & Atkinson, 1992)
  - For nested GLMs, assume larger model M_2 is true, with $\boldsymbol{\pi}_2$ known
  - Choose design ξ to maximise the deviance of the fit of M_1 to the expected response under M_2, $\boldsymbol{\pi}_2$

Designs for model discrimination

- T-optimality for nonlinear models (Atkinson & Fedorov, 1975): assume model 2 true, choose ξ to maximise $\sum_{i=1}^{n} w_i\{\eta_2(\boldsymbol{\xi}_i) - \eta_1(\boldsymbol{\xi}_i, \boldsymbol{\theta}_1)\}^2$
- T-optimality for GLMs (Ponce de Leon & Atkinson, 1992)
  - For nested GLMs, assume larger model M_2 is true, with $\boldsymbol{\pi}_2$ known
  - Choose design ξ to maximise the deviance of the fit of M_1 to the expected response under M_2, $\boldsymbol{\pi}_2$

$$\xi^{*T} = \arg\max_{\xi} D(\boldsymbol{\pi}_2; \hat{\boldsymbol{\pi}}_1), \text{ where}$$

$$D(\boldsymbol{\pi}_2; \hat{\boldsymbol{\pi}}_1) = 2\sum_{i=1}^{n} w_i \left[\pi_{2i} \log\left(\frac{\pi_{2i}}{\hat{\pi}_{1i}}\right) + (1 - \pi_{2i}) \log\left(\frac{1 - \pi_{2i}}{1 - \hat{\pi}_{1i}}\right)\right]$$

Example: T-optimal design

- Recall

$$
\begin{aligned}
M_1: &\qquad \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\
M_2: &\qquad \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2 \\
&\qquad\qquad\qquad\qquad -1 \leq x_1, x_2 \leq 1
\end{aligned}
$$

Example: T-optimal design

- Recall

$$M_1: \quad \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2$$
$$M_2: \quad \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2$$
$$-1 \leq x_1, x_2 \leq 1$$

- Assume M_2 is true, with

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' = (1, 1, 1, 0.5)'$$

Example: T-optimal design

- Recall

$$
\begin{aligned}
M_1: &\quad \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\
M_2: &\quad \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2 \\
&\quad -1 \le x_1, x_2 \le 1
\end{aligned}
$$

- Assume M_2 is true, with

$$
\begin{aligned}
\boldsymbol{\beta}_2 &= (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' \\
&= (1, 1, 1, 0.5)'
\end{aligned}
$$

- Find ξ^{*T} using a search routine such as simulated annealing

Example: T-optimal design

- Recall

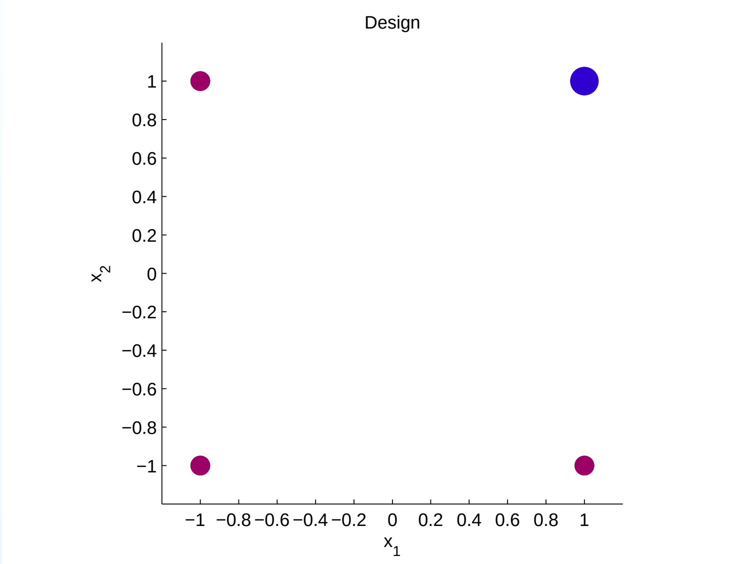
$$
\begin{aligned}
M_1: \quad & \text{logit}(\pi_1) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 \\
M_2: \quad & \text{logit}(\pi_2) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_{i1}x_2 \\
& -1 \leq x_1, x_2 \leq 1
\end{aligned}
$$

- Assume M_2 is true, with

$$
\begin{aligned}
\boldsymbol{\beta}_2 &= (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' \\
&= (1, 1, 1, 0.5)'
\end{aligned}
$$

- Find ξ^{*T} using a search routine such as simulated annealing
- intermission. . .

Example: T-optimal design



x_1	-1.0000	-1.0000	1.0000	1.0000
x_2	-1.0000	1.0000	-1.0000	1.0000
w	0.1973	0.1973	0.1973	0.4081

$$\begin{aligned}\hat{\boldsymbol{\beta}}_1 &= (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})' \\ &= (0.8679, 0.8679, 0.8679)'\end{aligned}$$

Designs for parameter estimation

- D-optimal designs
  - Maximise determinant of information matrix
  - Minimise volume of confidence ellipsoid of parameter estimates

Designs for parameter estimation

- D-optimal designs
  - Maximise determinant of information matrix
  - Minimise volume of confidence ellipsoid of parameter estimates
- Information matrix for model u in logistic regression example:

$$\boldsymbol{M}_u(\boldsymbol{\beta}_u, \xi) = \boldsymbol{X}_u' \boldsymbol{W}_u \boldsymbol{X}_u$$

where $\boldsymbol{\eta}_u = \boldsymbol{X}_u \boldsymbol{\beta}_u$ and $\boldsymbol{W}_u = \mathrm{diag}(w_i \pi_{ui}(1 - \pi_{ui}))$.

Designs for parameter estimation

- D-optimal designs
  - Maximise determinant of information matrix
  - Minimise volume of confidence ellipsoid of parameter estimates
- Information matrix for model u in logistic regression example:

$$\boldsymbol{M}_u(\boldsymbol{\beta}_u, \xi) = \boldsymbol{X}_u' \boldsymbol{W}_u \boldsymbol{X}_u$$

  where $\boldsymbol{\eta}_u = \boldsymbol{X}_u \boldsymbol{\beta}_u$ and $\boldsymbol{W}_u = \mathrm{diag}(w_i \pi_{ui}(1 - \pi_{ui}))$.
- Product optimality (Atkinson & Cox, 1974, Woods *et al.*, 2005): maximise product of determinants, scaled for # parameters

$$\xi^{*P} = \arg\max_{\xi} |\boldsymbol{M}_1(\boldsymbol{\beta}_1, \xi)|^{1/p_1} |\boldsymbol{M}_2(\boldsymbol{\beta}_2, \xi)|^{1/p_2}$$

Example: Product optimal design

- Parameters for M_2 are already specified for T-optimal design

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' = (1, 1, 1, 0.5)'$$

Example: Product optimal design

- Parameters for M_2 are already specified for T-optimal design

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' = (1, 1, 1, 0.5)'$$

- Parameters for M_1 are taken from T-optimal design

$$\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})' = (0.8679, 0.8679, 0.8679)'$$

Example: Product optimal design

- Parameters for M_2 are already specified for T-optimal design

$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' = (1, 1, 1, 0.5)'$$

- Parameters for M_1 are taken from T-optimal design

$$\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})' = (0.8679, 0.8679, 0.8679)'$$

- Find ξ^{*P} using a search routine such as simulated annealing

Example: Product optimal design

- Parameters for M_2 are already specified for T-optimal design

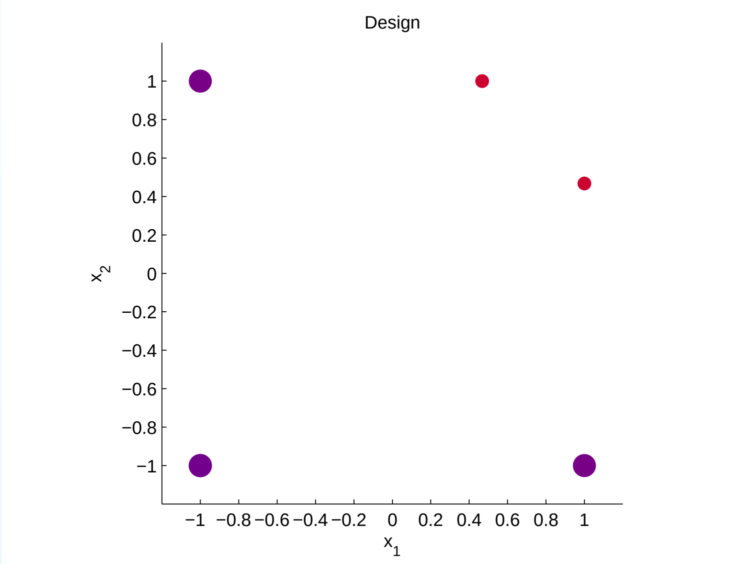
$$\boldsymbol{\beta}_2 = (\beta_{20}, \beta_{21}, \beta_{22}, \beta_{23})' = (1, 1, 1, 0.5)'$$

- Parameters for M_1 are taken from T-optimal design

$$\hat{\boldsymbol{\beta}}_1 = (\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12})' = (0.8679, 0.8679, 0.8679)'$$

- Find ξ^{*P} using a search routine such as simulated annealing
- intermission…

Example: Product optimal design



x_1	-1.0000	-1.0000	0.4676	1.0000	1.0000
x_2	-1.0000	1.0000	1.0000	-1.0000	0.4676
w	0.2745	0.2655	0.0973	0.2655	0.0973

Evaluation: Model discrimination

- For design ξ, N_{sim} sets of simulated random binomial data generated from $\text{Bin}([w_i N_{\text{sample}}], \pi_{2i})$ distribution, where $[x]$ is the integer nearest to x

Evaluation: Model discrimination

- For design ξ, N_{sim} sets of simulated random binomial data generated from $\mathrm{Bin}([w_i N_{\mathrm{sample}}], \pi_{2i})$ distribution, where $[x]$ is the integer nearest to x
- N_{sample} is total sample size, chosen in an ad hoc manner to be large enough to be able to reasonably fit both models to the data, but not so large that discriminating between them becomes a trivial

Evaluation: Model discrimination

- For design ξ, $N_{\rm sim}$ sets of simulated random binomial data generated from $\mathrm{Bin}([w_i N_{\rm sample}], \pi_{2i})$ distribution, where $[x]$ is the integer nearest to x
- $N_{\rm sample}$ is total sample size, chosen in an ad hoc manner to be large enough to be able to reasonably fit both models to the data, but not so large that discriminating between them becomes a trivial
- Power = (# times M_2 is chosen by LRT)/$N_{\rm sim}$

Evaluation: Model discrimination

- For design ξ, N_{sim} sets of simulated random binomial data generated from $\text{Bin}([w_i N_{\text{sample}}], \pi_{2i})$ distribution, where $[x]$ is the integer nearest to x
- N_{sample} is total sample size, chosen in an ad hoc manner to be large enough to be able to reasonably fit both models to the data, but not so large that discriminating between them becomes a trivial
- Power = (# times M_2 is chosen by LRT)/N_{sim}
- To compare power of these designs, $N_{\text{sample}} = 200$ was chosen after preliminary investigation

Evaluation: Model discrimination

- For design ξ, N_{sim} sets of simulated random binomial data generated from $\text{Bin}([w_i N_{\text{sample}}], \pi_{2i})$ distribution, where $[x]$ is the integer nearest to x
- N_{sample} is total sample size, chosen in an ad hoc manner to be large enough to be able to reasonably fit both models to the data, but not so large that discriminating between them becomes a trivial
- Power = (# times M_2 is chosen by LRT)/N_{sim}
- To compare power of these designs, $N_{\text{sample}} = 200$ was chosen after preliminary investigation
- $N_{\text{sim}} = 5000$ gave reasonably stable estimates of power

Evaluation: Parameter estimation

- Individual D-optimal designs were found for each model:

$$\xi^{*D_u} = \arg\max_{\xi} |\boldsymbol{M}_u(\boldsymbol{\beta}_u, \xi)|^{1/p_u}, \quad u = 1, 2$$

- D-efficiencies for design ξ under model u calculated by

$$D^u_{\text{eff}}(\xi) = \frac{|\boldsymbol{M}_u(\boldsymbol{\beta}_u, \xi)|^{1/p_u}}{|\boldsymbol{M}_u(\boldsymbol{\beta}_u, \xi^{*D_u})|^{1/p_u}}$$

Design evaluation

Design	Power	D_{eff}^1	D_{eff}^2
ξ^{*T}	0.6914	0.8452	0.9168
ξ^{*P}	0.5358	0.9592	0.9879

Design evaluation

Design	Power	D_{eff}^1	D_{eff}^2
ξ^{*T}	0.6914	0.8452	0.9168
ξ^{*P}	0.5358	0.9592	0.9879

- How do we compromise between the criteria?

Compromise designs

- For nonlinear models, we considered ‘conditional’ criteria: find T-optimal design, add support points which maximise product criterion (Waterhouse et al., 2004)

Compromise designs

- For nonlinear models, we considered ‘conditional’ criteria: find T-optimal design, add support points which maximise product criterion (Waterhouse et al., 2004)
- Later tried reversing the process, finding product optimal design, adding points to maximise T-optimality criterion
    - Similar designs to original method

Compromise designs

- For nonlinear models, we considered ‘conditional’ criteria: find T-optimal design, add support points which maximise product criterion (Waterhouse et al., 2004)
- Later tried reversing the process, finding product optimal design, adding points to maximise T-optimality criterion
  - Similar designs to original method
- Computationally simpler ‘hybrid’ designs showed similar results yet again

Hybrid designs

Given T-optimal and product optimal designs,

$$\xi^{*T} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*T} & \ldots & \boldsymbol{\xi}_{n_T}^{*T} \\ w_1^{*T} & \ldots & w_{n_T}^{*T} \end{matrix} \right\} \quad \text{and} \quad \xi^{*P} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*P} & \ldots & \boldsymbol{\xi}_{n_P}^{*P} \\ w_1^{*P} & \ldots & w_{n_P}^{*P} \end{matrix} \right\},$$

Hybrid designs

Given T-optimal and product optimal designs,

$$\xi^{*T} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*T} & \dots & \boldsymbol{\xi}_{n_T}^{*T} \\ w_1^{*T} & \dots & w_{n_T}^{*T} \end{Bmatrix} \quad \text{and} \quad \xi^{*P} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*P} & \dots & \boldsymbol{\xi}_{n_P}^{*P} \\ w_1^{*P} & \dots & w_{n_P}^{*P} \end{Bmatrix},$$

we define a 'hybrid' of the two designs as

$$\xi_\alpha^{\text{hybrid}} = \begin{Bmatrix} \boldsymbol{\xi}_1^{*P} & \dots & \boldsymbol{\xi}_{n_P}^{*P} & \boldsymbol{\xi}_1^{*T} & \dots & \boldsymbol{\xi}_{n_T}^{*T} \\ \alpha w_1^{*P} & \dots & \alpha w_{n_P}^{*P} & (1-\alpha) w_1^{*T} & \dots & (1-\alpha) w_{n_T}^{*T} \end{Bmatrix},$$

Hybrid designs

Given T-optimal and product optimal designs,

$$\xi^{*T} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*T} & \ldots & \boldsymbol{\xi}_{n_T}^{*T} \\ w_1^{*T} & \ldots & w_{n_T}^{*T} \end{matrix} \right\} \quad \text{and} \quad \xi^{*P} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*P} & \ldots & \boldsymbol{\xi}_{n_P}^{*P} \\ w_1^{*P} & \ldots & w_{n_P}^{*P} \end{matrix} \right\},$$

we define a ‘hybrid’ of the two designs as

$$\xi_\alpha^{\text{hybrid}} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*P} & \ldots & \boldsymbol{\xi}_{n_P}^{*P} & \boldsymbol{\xi}_1^{*T} & \ldots & \boldsymbol{\xi}_{n_T}^{*T} \\ \alpha w_1^{*P} & \ldots & \alpha w_{n_P}^{*P} & (1-\alpha) w_1^{*T} & \ldots & (1-\alpha) w_{n_T}^{*T} \end{matrix} \right\},$$

where $\alpha \in (0,\, 1)$ represents the importance placed on parameter estimation

Hybrid designs

Given T-optimal and product optimal designs,

$$\xi^{*T} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*T} & \ldots & \boldsymbol{\xi}_{n_T}^{*T} \\ w_1^{*T} & \ldots & w_{n_T}^{*T} \end{matrix} \right\} \quad \text{and} \quad \xi^{*P} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*P} & \ldots & \boldsymbol{\xi}_{n_P}^{*P} \\ w_1^{*P} & \ldots & w_{n_P}^{*P} \end{matrix} \right\},$$

we define a 'hybrid' of the two designs as

$$\xi_\alpha^{\text{hybrid}} = \left\{ \begin{matrix} \boldsymbol{\xi}_1^{*P} & \ldots & \boldsymbol{\xi}_{n_P}^{*P} & \boldsymbol{\xi}_1^{*T} & \ldots & \boldsymbol{\xi}_{n_T}^{*T} \\ \alpha w_1^{*P} & \ldots & \alpha w_{n_P}^{*P} & (1-\alpha) w_1^{*T} & \ldots & (1-\alpha) w_{n_T}^{*T} \end{matrix} \right\},$$

where $\alpha \in (0,\, 1)$ represents the importance placed on parameter estimation

- intermission. . .

Discussion

- T-optimal designs are not as efficient as product optimal designs in terms of parameter estimation

Discussion

- T-optimal designs are not as efficient as product optimal designs in terms of parameter estimation
- Conversely, product optimal designs are not as effective as T-optimal designs in terms of model discrimination

Discussion

- T-optimal designs are not as efficient as product optimal designs in terms of parameter estimation
- Conversely, product optimal designs are not as effective as T-optimal designs in terms of model discrimination
- Hybrid designs offer a suitable trade-off between criteria, with no further optimisation

Discussion

- T-optimal designs are not as efficient as product optimal designs in terms of parameter estimation
- Conversely, product optimal designs are not as effective as T-optimal designs in terms of model discrimination
- Hybrid designs offer a suitable trade-off between criteria, with no further optimisation
- Concept of hybrid designs can be easily generalised to handle any number of type of criteria

References

1. Atkinson A.C. and Fedorov V.V. (1975). The design of experiments for discriminating between two rival models. *Biometrika* **62**, 57–70.
2. Ponce de Leon A.C. and Atkinson A.C. (1992). The design of experiments to discriminate between two rival generalized linear models. In L. Fahrmeir, B. Francis, R. Gilchrist, and G. Tutz, eds., *Advances in GLIM and Statistical Modelling*, vol. 78 of *Lecture Notes in Statistics*, pp. 159–164, Springer-Verlag, New York.
3. Atkinson A.C. and Cox D.R. (1974). Planning experiments for discriminating between models. *Journal of the Royal Statistical Society. Series B (Methodological)* **36**, 321–348.
4. Waterhouse T.H., Eccleston J.A., and Duffull S.B. (2004). On optimal design for discrimination and estimation. In J. Antoch, ed., *COMPSTAT 2004 - Proceedings In Computational Statistics: 16th Symposium Held In Prague, Czech Republic, 2004*, pp. 1963–1970, Physica-Verlag.
5. Woods D.C., Lewis S.M., Eccleston J.A. and Russell K.G. (2005). Designs for generalized linear models with several variables and model uncertainty. *Technometrics*, to appear.