UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper M08/06
Methodology

Continuous Optimal Designs For Generalised Linear Models Under Model Uncertainty

David C. Woods, Susan M. Lewis

Abstract

We propose a general design selection criterion for experiments where a generalised linear model describes the response. The criterion allows for several competing aims, such as parameter estimation and model discrimination, and also for uncertainty in the functional form of the linear predictor, the link function and the unknown model parameters. A general equivalence theorem is developed for this criterion. In practice, an exact design is required by experimenters and can be obtained by numerical rounding of a continuous design. We derive bounds on the performance of an exact design under this criterion which allow the efficiency of a rounded continuous design to be assessed.

CONTINUOUS OPTIMAL DESIGNS FOR GENERALISED LINEAR MODELS UNDER MODEL UNCERTAINTY

David C. Woods* and Susan M. Lewis

University of Southampton

Summary

We propose a general design selection criterion for experiments where a generalised linear model describes the response. The criterion allows for several competing aims, such as parameter estimation and model discrimination, and also for uncertainty in the functional form of the linear predictor, the link function and the unknown model parameters. A general equivalence theorem is developed for this criterion. In practice, an exact design is required by experimenters and can be obtained by numerical rounding of a continuous design. We derive bounds on the performance of an exact design under this criterion which allow the efficiency of a rounded continuous design to be assessed.

Key words: exponential family; general equivalence theorem; logistic regression; nonlinear regression; optimal design.

1. Introduction

Generalised linear models (GLMs; see McCullagh & Nelder, 1989) are an important empirical modelling tool which have found application in a wide variety of experiments in medicine, science and technology (Collett, 2002; Myers et al., 2002). We consider an experiment on n treatments, or combinations of variable values, with the ith treatment replicated m_i times and $\sum_{i=1}^{n} m_i = N$. Each treatment is

*Author to whom correspondence should be addressed.

Southampton Statistical Sciences Research Institute, School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. Email: D.Woods@soton.ac.uk

Acknowledgments. This work was supported by the Engineering and Physical Sciences Research Council grant EP/C008863/1.

represented as a k-vector $\boldsymbol{x}_i \in \mathcal{X} \subset \mathbb{R}^k$ $(i = 1, \ldots, n)$, that is $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik})'$. Under the assumption of exchangeability of experimental units, the response Y_{ij}, obtained when the ith treatment is applied to the jth unit receiving that treatment $(j = 1, \ldots, m_i;\ i = 1, \ldots, n)$, is assumed to follow a distribution from the exponential family with the following components:

(i) a distribution for the response,

(ii) a *linear predictor* $\eta_i = f(\boldsymbol{x}_i)'\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is a p-vector of unknown model parameters and $f(\boldsymbol{x}_i)$ is a vector of known functions of the k explanatory variables, whose values for the ith run are held in $\boldsymbol{x}_i$,

(iii) a *link function* that relates the mean response from the ith support point to the linear predictor, $g(\mu_i) = \eta_i$.

Widely applied examples of GLMs include logistic regression for binary data with $g(\mu_i) = \log\{\mu_i/(1-\mu_i)\}$, and log-linear models for count data with $g(\mu_i) = \log(\mu_i)$.

As GLMs are nonlinear in the model parameters $\boldsymbol{\beta}$, the performance of a design under any model-based criterion will depend on the values of $\boldsymbol{\beta}$. Most research in the design of experiments for GLMs has focused on *locally* optimal designs for given values of $\boldsymbol{\beta}$ (Atkinson, 2006) or on robust designs for one or two variables (Chaloner & Larntz, 1989; Sitter, 1992; King & Wong, 2000). Recent work (Woods et al., 2006; Dror & Steinberg, 2006; Gotwalt et al., 2008) has extended these methods to multi-variable experiments through the application of a model-robust, or compromise, design criterion implemented in computationally intensive algorithms to find exact designs. Woods et al. (2006) and Dror & Steinberg (2006) also investigated robustness to the functional form of $f(\cdot)$ and the choice of link function $g(\cdot)$. There has been a parallel development of designs for the problem of discriminating between given GLMs (López-Fidalgo et al., 2007; Waterhouse et al., 2008).

In this paper we propose a general model-robust criterion which extends previous criteria by encompassing not only uncertainty in the model form and parameters, but also competing aims of an experiment, such as estimation and model discrimination.

We establish necessary and sufficient conditions for a *continuous*, or approximate, design to be optimal under this criterion. Such a design, $\xi \in \Xi$, relaxes the assumption that m_i must be integer ($i = 1, \ldots, n$) and is represented as a probability measure over a compact design space $\mathcal{X}$ with finite support. The support of ξ defines the set of distinct points of the design, and the non-zero image of the ith element specifies the proportion, $0 < \omega(\boldsymbol{x}_i) \leq 1$, of experimental effort assigned to the ith point. The design is expressed as

$$\xi = \left\{ \begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \\ \omega(\boldsymbol{x}_1) & \omega(\boldsymbol{x}_2) & \cdots & \omega(\boldsymbol{x}_n) \end{array} \right\}, \tag{1}$$

where $\sum_{i=1}^{n} \omega(\boldsymbol{x}_i) = 1$. Continuous designs are a common tool in optimal design theory (Atkinson et al., 2007, ch.9), because they are mathematically more tractable than exact designs and provide a continuous convex optimization problem for design criteria satisfying mild assumptions. To obtain a realisable exact design, the image of ξ must be scaled by N, the total number of runs, and the $N\omega(\boldsymbol{x}_i)$ values rounded to integer values as necessary; see, for example, Fedorov & Hackl (1997, p.53).

In Section 2, the design selection criterion is outlined and a general equivalence theorem established. This provides necessary and sufficient conditions for a continuous design to be optimal. A numerical example is presented in Section 3, and Section 4 discusses the performance of exact designs through bounds on the size of the objective function. The approach taken can be applied to other non-linear models, for example, those derived from mathematical theory in Physics or Chemistry.

2. Design selection criterion

We consider a general class of design selection criteria that explicitly takes account of uncertainties in the model and potentially differing objectives of an experiment. For a given distribution, a GLM is defined by the triple $s = (g, f, \boldsymbol{\beta})$ of link function, form of linear predictor and vector of model parameters. We represent uncertainty in the model s through sets $\mathcal{G}$, $\mathcal{N}$ and $\mathcal{B}$ of possible link functions, linear predictors and model parameters, respectively.

The conflicting aims of the experiment are represented through a set $\mathcal{C}$ of objective functions, each of which corresponds to a criterion which assesses the usefulness of a design for a given task, for example, parameter estimation or model discrimination. Each element of $\mathcal{C}$ is a *local* objective function ϕ_c: $(\xi, s) \to \mathbb{R}$ where $\xi \in \Xi$ and $s \in \mathcal{S} = \mathcal{G} \times \mathcal{N} \times \mathcal{B}$. Further, we assume that every ϕ_c is a "larger the better" objective function, that $\mathcal{C}$ and $\mathcal{S}$ do not depend on the design ξ and that every $\phi_c \in \mathcal{C}$ is defined for all $s \in \mathcal{S}$. If $\mathcal{C}$ and $\mathcal{S}$ are uncountably infinite sets, we assume ϕ_c is continuous with respect to c and s.

The criterion for design selection considered in this paper is based on the compromise, or portmanteau, objective function defined through the Stieltjes integral

$$\Phi(\xi) = \int_{\mathcal{C}} \int_{\mathcal{S}} \phi_c(\xi, s) \, \mathrm{d}G(s) \, \mathrm{d}H(c) \,, \tag{2}$$

where $G(\cdot)$ and $H(\cdot)$ are distribution functions chosen to reflect the relative importance of each model, and the relative importance of each local objective function, respectively. A Φ-optimal design $\xi^\star$ is such that

$$\xi^\star = \arg \max_{\xi \in \Xi} \Phi(\xi) \,. \tag{3}$$

Special cases of (2) have been applied to GLMs by a variety of authors. Chaloner & Larntz (1989) used a single objective function and allowed uncertainty in $\boldsymbol{\beta}$; Woods et al. (2006) and Dror & Steinberg (2006) investigated the use of a single objective function and uncertainty in all three aspects of the model s. For linear models similar criteria, for a single objective function were considered by Läuter (1974) and Cook & Nachtsheim (1982). Atkinson (2008) and Waterhouse et al. (2008) found locally optimal designs under a portmanteau criterion for parameter estimation and model discrimination for linear models and GLMs respectively.

The Φ-criterion may be viewed from a Bayesian perspective. Then $G(\cdot)$ summarises the prior belief across the model space of the experimenters, and (2) is the preposterior expectation for each element ϕ_c of $\mathcal{C}$, averaged across $\mathcal{C}$ with respect to

$H(\cdot)$.

In order to develop a general equivalence theorem, we follow Whittle (1973) and Chaloner & Larntz (1989) in formulating the theorem directly in terms of the measure ξ. For linear models, equivalence theorems are usually formulated in terms of a compact set of information matrices. The same is true for nonlinear models when only locally optimal designs are considered. For our problem, however, the information matrices are dependent on s, and we do not wish to restrict to local objective functions which are convex functions of an information matrix.

Suppose that for all $s \in \mathcal{S}$ and $\xi \in \Xi$, $\phi_c(\xi, s)$ is a concave, continuous and differentiable function of ξ with continuous derivatives; see also Chaloner & Larntz (1989). We also assume that there exists at least one measure ξ such that $\Phi(\xi) < \infty$, and that if $\xi_1 \to \xi_2$ in weak convergence, then $\Phi(\xi_1) \to \Phi(\xi_2)$.

Define the Fréchet directional derivative of $\Phi(\xi)$ as

$$\Psi(\xi_1, \xi_2) = \lim_{\alpha \to 0^+} \frac{\Phi\{(1-\alpha)\xi_1 + \alpha\xi_2\} - \Phi(\xi_1)}{\alpha}, \tag{4}$$

where ξ_1 and ξ_2 are measures and $0 \leq \alpha \leq 1$. Then $\tilde{\xi} = (1-\alpha)\xi_1 + \alpha\xi_2$ is also a measure, and (4) is the derivative of Φ at ξ_1 in the direction of ξ_2. The following general equivalence theorem can be proved which provides necessary and sufficient conditions for a design to be Φ-optimal.

Theorem 1: The following three conditions are equivalent

1. $\Phi(\xi^\star) = \max_{\xi \in \Xi} \Phi(\xi)$,
2. $\max_{\boldsymbol{x} \in \mathcal{X}} \Psi(\xi^\star, \xi_{\boldsymbol{x}}) \leq 0$, where $\xi_{\boldsymbol{x}}$ is the measure with point mass at $\boldsymbol{x}$,
3. $\Psi(\xi^\star, \xi_{\boldsymbol{x}}) = 0$ for all $x \in \text{support}(\xi^\star)$.

The proof is analogous to that of Whittle (1973). The required directional derivative is provided by

$$
\begin{aligned}
\Psi(\xi, \xi_{\boldsymbol{x}}) &= \lim_{\alpha \to 0^+} \frac{1}{\alpha} \int_{\mathcal{C}} \int_{\mathcal{S}} \left\{ \phi_c[(1-\alpha)\xi + \alpha \xi_{\boldsymbol{x}}, s] - \phi_c(\xi, s) \right\} \, \mathrm{d}G(s) \, \mathrm{d}H(c) \\
&= \int_{\mathcal{C}} \int_{\mathcal{S}} \psi_c(\xi, \xi_{\boldsymbol{x}}) \, \mathrm{d}G(s) \, \mathrm{d}H(c) \,,
\end{aligned}
$$

where

$$
\psi_c(\xi, \xi_{\boldsymbol{x}}) = \lim_{\alpha \to 0^+} \frac{1}{\alpha} \left\{ \phi_c[(1-\alpha)\xi + \alpha \xi_{\boldsymbol{x}}, s] - \phi_c(\xi, s) \right\}
$$

is the directional derivative for objective function c. Note that if one or more of $\mathcal{C}$, $\mathcal{G}$, $\mathcal{N}$ and $\mathcal{B}$ consists of continuous random variables, then the corresponding distribution function must be differentiable.

3. Numerical example

In this section, a Φ-optimal design is found numerically for a special case of the criterion, and its optimality confirmed using the necessary and sufficient conditions. Consider an experiment with four variables, x_1-x_4, with each variable scaled so that $\mathcal{X} = [-1, 1]^4$. A logistic regression model is used for the response with $\eta = \beta_0 + \sum_{l=1}^{4} \beta_l x_l$. A single local D-optimality objective function is considered, namely,

$$
\phi(\xi, s) = \log |M(\xi, \boldsymbol{\beta})|^{1/5} \,,
$$

where $M(\xi, \beta)$ is the standardised information matrix for design ξ and model parameters $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_4)$. For logistic regression, the information matrix is given by

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

where $X(\xi)$ is the $n \times 5$ model matrix for design ξ, and $W(\xi, \boldsymbol{\beta})$ is a diagonal matrix with entries $\omega(\boldsymbol{x}_i) v(\boldsymbol{x}_i)$ with $v(\boldsymbol{x}_i) = \mu_i(1-\mu_i)$. The mean response, μ_i, at the ith support point depends on $\boldsymbol{\beta}$ and $\boldsymbol{x}_i$ through the linear predictor.

Table 1: *Values of the parameters for the example*

Model	β_0	β_1	β_2	β_3	β_4
1	1.6	2	1.6	1.6	2
2	0	2.4	2	2.8	2.8
3	-1.6	1.2	2.8	1.2	1.6
4	-0.8	2.8	1.2	2	1.2
5	0.8	1.6	2.4	2.4	2.4

Objective function (2) then becomes

$$\Phi(\xi) = \int_{\mathcal{B}} \log |M(\xi, \boldsymbol{\beta})|^{1/5} \, \mathrm{d}G(\boldsymbol{\beta}) \,. \tag{5}$$

Suppose that $G(\boldsymbol{\beta})$ is a five-point discrete distribution defined on the parameter values given in Table 1 and with equal weight given to each point. The five points were selected as a Latin hypercube sample (McKay et al., 1979).

In general, an optimal design is not unique and, for compromise or Bayesian designs, there is no upper bound on the number of support points (see, for example Atkinson et al., 2007, ch.18). In addition, any weighted average of two optimal design measures will itself be an optimal design. Hence to find a Φ-optimal design for the example, a numerical search was used (see Woods, 2008) where the maximum number of support points was set to a suitably large number, and then decreased to find the optimal design with the smallest support. The design obtained is shown in Table 2 and has 16 support points.

The Φ-optimality of this design may be confirmed numerically from the general equivalence theorem, using condition 2, by evaluating the directional derivative of (5) across the design region. This derivative is given by

TABLE 2: A Φ-*optimal design for four variables in 16 runs*

Run	x_1	x_2	x_3	x_4	w	Run	x_1	x_2	x_3	x_4	w
1	-1	-1	1	1	0.071	9	1	-1	-1	1	0.111
2	1	-1	1	-1	0.088	10	-1	-1	0.03	1	0.058
3	-0.16	1	1	-1	0.038	11	-1	0.03	-1	1	0.003
4	-1	1	-1	-0.14	0.067	12	1	-0.05	-1	1	0.027
5	1	-0.85	1	1	0.018	13	0.17	-1	1	-1	0.067
6	-1	1	-1	1	0.095	14	1	1	-1	-0.18	0.020
7	-1	1	1	-1	0.124	15	1	1	-1	-1	0.142
8	1	-1	-0.39	-1	0.025	16	-1	0.43	1	1	0.045

$$\begin{aligned}\Psi(\xi, \xi_{\boldsymbol{x}}) &= \frac{1}{p}\int_{\mathcal{B}} \left[v(\boldsymbol{x})f(\boldsymbol{x})'M^{-1}(\xi, \boldsymbol{\beta})f(\boldsymbol{x}) - p\right] \mathrm{d}G(\boldsymbol{\beta}) \\ &= \frac{1}{25}\sum_{i=1}^{5} v(\boldsymbol{x})f(\boldsymbol{x})'M^{-1}(\xi, \boldsymbol{\beta}_i)f(\boldsymbol{x}) - 1\,.\end{aligned}$$

Figure 1 shows two different projections of $\Psi(\xi^\star, \xi_{\boldsymbol{x}})$ into the x_1-x_2 plane and illustrates that the selected support points are those points for which $\psi(\xi^\star, \xi_{\boldsymbol{x}})$ has its maximum value of 0. The other projections are similar, illustrating that the support points of $\xi^\star$ form the level set for $\Psi(\xi^\star, \xi_{\boldsymbol{x}}) = 0$.

4. Assessing the performance of exact designs

In practice, exact designs (i.e. having integer replication of each support point) are required. Let $\Delta(N)$ denote the set of all such designs with N runs. Then $\delta(N) \in \Delta(N)$ is a measure which has integer non-zero image with $\sum_{i=1}^{n} \omega(\boldsymbol{x}_i) = N$.

In addition, for any continuous design $\xi \in \Xi$, let $\tilde{\xi}(N)$ denote the measure on $\mathcal{X}$ which has identical support to ξ with the ith element of the support having image

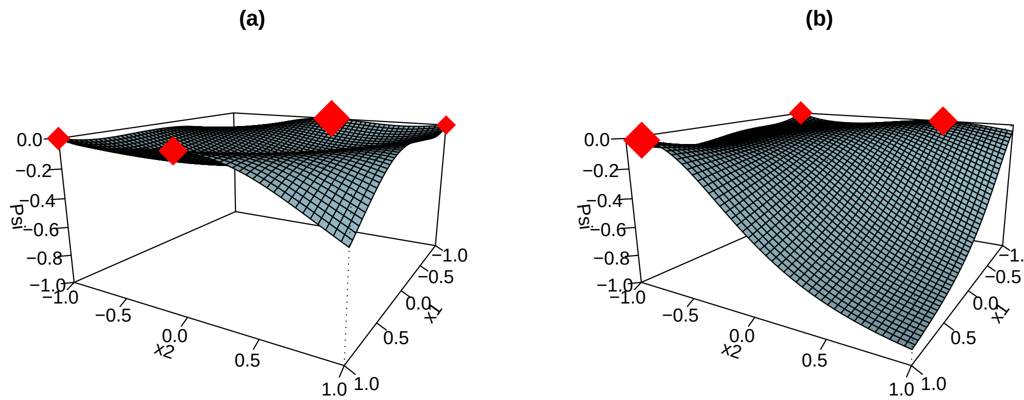

Figure 1: $\Psi(\xi^\star, \boldsymbol{x})$ surface in the x_1-x_2 plane and design points ($\Diamond$) for the numerical example: (a) $x_3 = -1$, $x_4 = 1$; (b) $x_3 = 1$, $x_4 = 1$. The area of the diamonds is proportion to $\omega(\boldsymbol{x})$.

$\tilde{w}(\boldsymbol{x}_i) = Nw(\boldsymbol{x}_i)$. Further, let $\tilde{\Xi}(N) = \{\tilde{\xi}; \xi \in \Xi\}$. Note that $\tilde{\xi}(N)$ is an exact design only when it has integer non-zero image.

We extend the domain of Φ in (2) to $\tilde{\Xi}(N)$, where $\Delta(N) \subset \tilde{\Xi}(N)$. Let $\delta^\star(N)$ and $\tilde{\xi}^\star(N)$ be the elements having maximum values of Φ in $\Delta(N)$ and $\tilde{\Xi}(N)$ respectively. Then $\delta^\star(N)$ is the Φ-optimal exact design in N runs.

The following theorem establishes bounds on the value of $\Phi(\delta^\star(N))$ when local objective function ϕ_c is a monotonically non-decreasing function with respect to N, i.e. $\phi_c(\tilde{\xi}(N_1), s) \geq \phi_c(\tilde{\xi}(N_2), s)$ for $N_1 \geq N_2$.

Theorem 2: If $\Phi(\tilde{\xi}(N))$ is monotonically non-decreasing with respect to N, then $\Phi(\delta^\star(N))$ is bounded below by

$$\Phi(\delta^\star(N)) \geq \Phi(\tilde{\xi}^\star(N-n))\,, \text{for } N \geq n\,,$$

and bounded above by

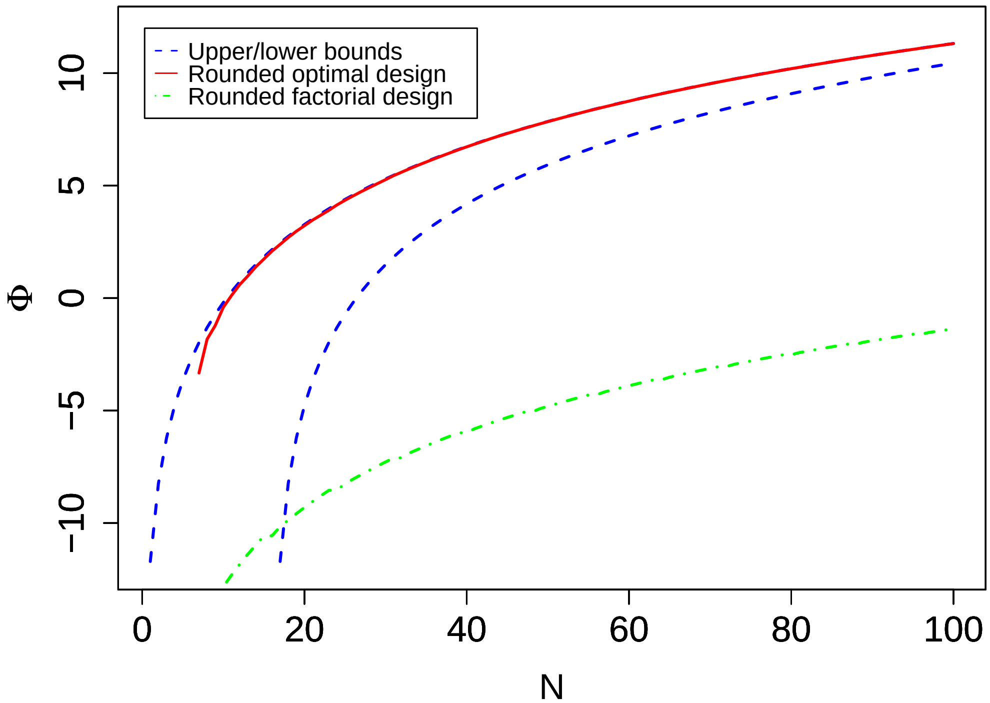

Figure 2: Bounds on $\Phi(\delta^{\star}(N))$ and values of $\Phi(\delta^{\star}_R(N))$ and $\Phi(\delta_F(N))$ for the numerical example.

$$\Phi(\delta^{\star}(N)) \leq \Phi(\tilde{\xi}^{\star}(N))\,, \text{for } N \geq 0\,.$$

Proof: Let $\xi_+(N) \in \Delta(N)$ be the measure with the same support as $\xi \in \Xi$ and ith element of the support having image $\omega_+(\boldsymbol{x}_i) = [N\omega(\boldsymbol{x}_i)]^+$ where $[u]^+$ is the smallest integer greater than u. Then $\sum_{i=1}^n \omega_+(\boldsymbol{x}_i) \geq \sum_{i=1}^n N\omega(\boldsymbol{x}_i)$ and, by the monotonicity of Φ,

$$\Phi(\tilde{\xi}^{\star}(N-n)) \leq \Phi(\xi^{\star}_+(N-n)) \leq \Phi(\delta^{\star}(N)) \quad \text{for } N \geq n\,,$$

as $\xi^{\star}_+(N-n)$ is an exact design on $N-n$ runs having the same support as $\xi^{\star}$. The upper bound follows directly from the fact that $\Delta(N) \subset \tilde{\Xi}(N)$.

Figure 2 shows the bounds for $0 \leq N \leq 100$ for the numerical example of

Section 3 and objective function (5). It also shows the values of the objective function for an exact design, $\delta_R^\star(N)$, obtained by numerical rounding of an optimal continuous design $\xi^\star$, and the performance of a 2_{IV}^{4-1} fractional factorial, $\delta_F(N)$. For this design, the replication for each of the eight support points was chosen to give the largest value of $\Phi(\delta_f(N))$ while ensuring as equal replication as possible. These designs can only be assessed for $N \geq 5$, when they have sufficient distinct runs to estimate the model parameters. Clearly, $\delta_R^\star(N)$ performs well, nearly attaining the upper bound for $N \geq 5$. This is in contrast to the poor performance of $\delta_F(N)$.

References

Atkinson, A. C. (2006). Generalized linear models and response transformation. In *Response Surface Methodology* (ed. A. I. Khuri), 173–202. Singapore: World Scientific.

— (2008). DT-optimum designs for model discrimination and parameter estimation. *J. Stat. Plan. Infer.* **138**, 56–64.

Atkinson, A. C., Donev, A. N. & Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS.* Oxford: Oxford Science Publications, 2nd edn.

Chaloner, K. & Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Stat. Plan. Infer.* **21**, 191–208.

Collett, D. (2002). *Modelling Binary Data.* Boca Raton, Fl: Chapman and Hall, 2nd edn.

Cook, R. D. & Nachtsheim, C. J. (1982). Model robust, linear-optimal designs. *Technometrics* **24**, 49–54.

Dror, H. A. & Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520–529.

Fedorov, V. V. & Hackl, P. (1997). *Model-Oriented Design of Experiments.* New York: Springer.

Gotwalt, C. M., Jones, B. A. & Steinberg, D. M. (2008). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Tech. rep.* 0801, Tel Aviv University.

King, J. & Wong, W. K. (2000). Minimax D-optimal designs for the logistic model. *Biometrics* **56**, 1263–1267.

Läuter, E. (1974). Experimental design in a class of models. *Mathematische Operationsforschung und Statistik* **5**, 379–396.

López-Fidalgo, J., Trandafir, C. & Tommasi, C. H. (2007). An optimal experimental design criterion for discriminating between non-Normal models. *J. Roy. Statist. Soc. Ser. B* **69**, 231–242.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models.* New York: Chapman and Hall, 2nd edn.

McKay, M. D., Beckman, R. J. & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239–245.

Myers, R. H., Montgomery, D. C. & Vining, G. G. (2002). *Generalized Linear Models with Applications in Engineering and the Sciences.* New York: Wiley.

Sitter, R. R. (1992). Robust designs for binary data. *Biometrics* **48**, 1145–1155.

Waterhouse, T. H., Woods, D. C., Eccleston, J. A. & Lewis, S. M. (2008). Design selection criteria for discrimination/estimation for nested models and a binomial response. *J. Statist. Plan. Infer.* **138**, 132–144.

Whittle, P. (1973). Some general points in the theory of optimal experimental design. *J. Roy. Statist. Soc. Ser. B* **35**, 123–130.

Woods, D. C. (2008). Robust designs for binary data: applications of simulated annealing. *Working Paper M08/03*, Southampton Statistical Sciences Research Institute.

Woods, D. C., Lewis, S. M., Eccleston, J. A. & Russell, K. G. (2006). Designs for generalized linear models with several variables and model uncertainty. *Technometrics* **48**, 284–292.