Research Report

Department of Statistics

No. 2006:8

Sequential Designs for Binary Data with the purpose to Maximize the Probability of Response

Ellinor Fackle Fornius

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Sequential Designs for Binary Data with the purpose to Maximize the Probability of Response

Ellinor Fackle Fornius
Department of Statistics
Stockholm University

Abstract

Two kinds of sequential designs are proposed for finding the point that maximizes the probability of response assuming a binary response variable and a quadratic logistic regression model. One is a parametric optimal design approach and the other one is a nonparametric stochastic approximation approach. The suggested sequential designs are evaluated and compared in a simulation study. In summary the parametric approach performed very well whereas its competitor failed in some cases.

Keywords: Sequential design, c-optimal design, Stochastic approximation, Quadratic logistic model, Response Surface Methodology.

1 Introduction

Response Surface Methodology (RSM) concerns design and analysis of experiments with the purpose to study the behavior of a response variable. A response function can be used to describe the connection between a response variable and the control variable(s). This function, which is often a polynomial, is assumed to approximate the true relationship in a restricted region of interest. Experiments are then performed to gain information about the response function. RSM treats all stages of the experimental design including the initial choices of which control variable(s) to consider at what levels, the magnitude of the region of interest and the complexity of the approximating model, through the conduct of the experiment to the analysis of the data. Reconsiderations will often have to be made along this path making RSM an iterative procedure. For a full review of these techniques, see Box and Draper (1987). One of the main interests for RSM is to determine the optimum operating conditions, i.e. to locate the point of maximum or minimum response. Examples on response variables are different quality aspects of a product, such as yield or strength.

In many applications the response variable is binary, the outcome of the experiment is either response or nonresponse. A product can be good or defect, an answer to a test is right or wrong and a drug has an effect or not. In such cases the response function is describing the relationship between the probability of response and the control variable(s). Generalized Linear Models (GLM) is a class of models that are suitable for binary data and one of the candidate models is logistic regression which is the one used in this paper.

The theory of optimal design provides methods to determine at what levels of the control variable(s) to run an experiment, see for example Atkinson and Donev (1992). Based on a criterion function that reflects some specific information aspect an optimal design can be chosen. It is possible to tailor-make optimal designs for the purpose of estimating the optimum of a response function. However, the optimal design generally depends on the unknown true model parameters for GLMs. Of course there would be no point in performing experiments if the true parameters were known. One approach to solve this problem is to use sequential designs. Starting with a preliminary estimate or guess of the parameters a locally optimal design can be constructed. This design is then used to update the parameter estimates, which in turn leads to another locally optimal design, new parameter estimates and so on. Sequential designs for binary data with the purpose to estimate a percentile of the response curve are treated in Wu (1985).

The theory of stochastic approximation provides an alternative method for estimation of the optimum point of a response function. Stochastic approximation started with the work of Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) and

is a nonparametric sequential approach. The design points are determined successively according to a recursive scheme such that the resulting sequence will converge to the point of optimum response. A stochastic approximation method is evaluated for binary data in the case of estimating a percentile of the response curve in Wu (1985). Wu (1986) goes through the connections and differences between the stochastic approximation method and a parametric approach based on maximum likelihood estimation.

A quadratic logistic regression model is considered in this paper and two kinds of sequential designs are proposed for the purpose of estimating the point of optimum response. The first sequential design is an adaptive nonparametric stochastic approximation approach based on the recursion of Kiefer and Wolfowitz. Adaptive means that all of the information obtained from the previous steps in the recursion is used to choose the next design point. The second approach is parametric, the model parameters are estimated at each step and an optimal design is derived assuming the current parameter estimates to be true. The performances of the two sequential designs are evaluated through a simulation study. The paper is organized according to the following. The model is described in the next section. Section 3 starts with a review of the Robbins-Monro and the Kiefer-Wolfowitz methods, followed by the details of the proposed nonparametric approach. The problem of finding optimal designs for the quadratic logistic regression model is treated before the details of the parametric approach are given. The results of the simulation study can be found in section 5. The last section contains some concluding remarks.

2 Model

The quadratic logistic regression model belongs to the class of Generalized Linear Models. GLMs are generally defined by three parts; the random component specifying the distribution of the responses, the systematic component containing the linear predictor and the link function relating the expected value of the response variable to the linear predictor, see McCullagh and Nelder (1989). The link function is assumed to be monotonic and differentiable. In the current logistic regression model the responses are Bernoulli distributed, $Y_i \sim Bern(\pi_i),\ i = 1, \ldots, n$. The linear predictor is given by

$$\begin{aligned} \eta_i &= \mathbf{x}_i^T \boldsymbol{\beta} = \begin{bmatrix} 1 & x_i & x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \\ &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \end{aligned}$$

and the logit link function is

$$g(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i.$$

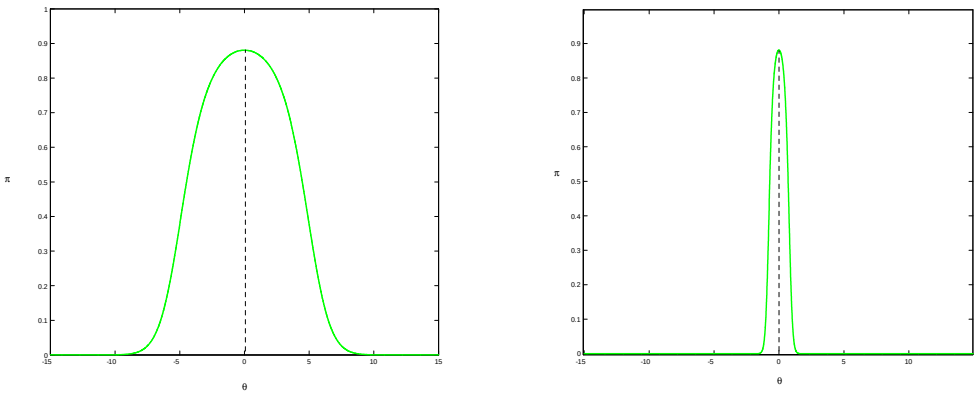
The probability of obtaining a response is given by

$$P\left(Y_i = 1\right) = \pi_i = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

The optimum operating conditions for the quadratic logistic model is located at the point of maximum or minimum response given by θ, see Figure 1. As an example consider manufacturing a product where defect items have to be discarded and the probability of a good product is maximized for the right adjustment of machine speed. The optimum point is found as the solution to $\frac{\partial}{\partial x}\eta = 0$ (because the link function is monotonic), that is

$$\theta = -\frac{\beta_1}{2\beta_2}.$$

Because $\frac{\partial^2}{\partial^2 x}\eta = 2\beta_2$ the sign of β_2 determines whether the response curve has a maximum or a minimum. Furthermore, $\eta\left(\theta\right) = \beta_0 - \beta_1^2/4\beta_2$ so that a larger β_0 means that the optimum point is closer to 1. Given a certain height of the response curve the relative width of the curve is determined by the size of the parameter β_2 and the larger $|\beta_2|$ the more narrow is the response curve. Two specific parameter sets are considered in this paper, both are given in Table 1 and displayed in Figure 1. Because β_2 is negative both optimum points are maxima. Since $\beta_1 = 0$ the optimum point is given by $\theta = 0$ in both cases. The height of the response curve is also the same for the two models. However, the response curve is wider for model A than for model B.



$\boldsymbol{\beta} = \begin{bmatrix} 2 & 0 & -0.1 \end{bmatrix}^T$ $\boldsymbol{\beta} = \begin{bmatrix} 2 & 0 & -4 \end{bmatrix}^T$

Figure 1: Two response curves $\pi(x)$ for the quadratic logistic model with maximum at $\theta = 0$.

Table 1: Two parameter sets.

Model	Parameter set
A	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$
B	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$

3 Sequential designs

Under comparison are a nonparametric sequential approach and two variants of a parametric sequential approach with the purpose of estimating the maximum of the response function $\pi(x) = P(Y = 1|x)$. It is equivalent to search for the optimum operating conditions in the RSM context.

3.1 A nonparametric sequential design

Let $f(x)$ denote a response function that is unknown. Robbins and Monro (1951) gives a stochastic approximation method for finding the solution $x = \theta$ to the equation $f(x) = d$, where d is a constant. It is assumed that $f(x) = d$ has a unique solution θ and that for every x observations can be made on a random variable $Y(x)$ such that $E[Y(x)] = f(x)$. Starting at an arbitrary x_1 consecutive observations are made on $Y(x)$ at $x_2, x_3, ...$ in such a way that x_r converges to θ as $r \to \infty$. At x_r the next design point, x_{r+1}, is chosen according to the following scheme

$$x_{r+1} = x_r + a_r (y_r - d), \tag{1}$$

where $\{a_r\}$ is a fixed sequence of decreasing positive constants satisfying $\sum_{r=1}^{\infty} a_r^2 < \infty$. The choice of the constants $\{a_r\}$ is important for the performance of the sequence $\{x_r\}$. In the sense of achieving minimal asymptotic variance it is optimal to set a_r equal to $-[rf'(\theta)]^{-1}$, see e.g. Chung (1954) and Sacks (1958). However, $f'(\theta)$ is generally unknown and needs to be estimated. One possible estimator is the least squares estimator of the slope in the regression Y on x. An adaptive version of the Robbins-Monro procedure where the estimate of $f'(\theta)$ is updated in each step by using the least squares estimator is given by

$$\begin{aligned} x_{r+1} &= x_r - \left(r\widehat{\beta}_r\right)^{-1}(y_r - d), \\ \text{where} \quad \widehat{\beta}_r &= \frac{\sum_{i=1}^{r} y_i (x_i - \overline{x}_r)}{\sum_{i=1}^{r} (x_i - \overline{x}_r)^2}, \quad \overline{x}_r = \frac{\sum_{i=1}^{r} x_i}{r}. \end{aligned} \tag{2}$$

This procedure is described by Anbar (1978) and was reviewed and evaluated for binary data in Wu (1985). The adaptive procedure (2) was proven to be asymptotically equivalent to the nonadaptive Robbins-Monro procedure (1) with $a_r = -[rf'(\theta)]^{-1}$ by Anbar (1978) and Lai and Robbins (1979).

The Robbins-Monro procedure was further developed by Kiefer and Wolfowitz (1952) and extended to the case of estimating the maximum of a response function. The unknown response function $f(x)$ has its maximum at the point θ and $f(x)$ is assumed to be strictly increasing (decreasing) for $x < \theta$ $(x > \theta)$. Furthermore, it is assumed that observations can be taken on the random variable $Y(x)$ at any level x and that $E[Y(x)] = f(x)$. The principle is the same as for the Robbins-Monro procedure, that is successive observations are made on $Y(x)$ according to a specified scheme. The difference is that observations are made in pairs at each step. Starting at an arbitrary x_1, the following $x_2, x_3, ...$ are obtained by making observations at $x_r \pm c_r$, i.e. both $Y(x_r - c_r)$ and $Y(x_r + c_r)$ are observed at each step. The sequence $\{x_r\}$ is defined as

$$x_{r+1} = x_r + a_r \frac{y(x_r + c_r) - y(x_r - c_r)}{2c_r} = x_r + a_r z_r \tag{3}$$

and converges to θ as $r \to \infty$. $\{a_r\}$ and $\{c_r\}$ are preassigned infinite sequences of positive numbers such that

$$\begin{aligned} c_r &\to 0, \\ \textstyle\sum_{r=1}^{\infty} a_r &= \infty, \\ \textstyle\sum_{r=1}^{\infty} a_r c_r &< \infty, \\ \textstyle\sum_{r=1}^{\infty} a_r^2 c_r^{-2} &< \infty. \end{aligned}$$

For example $a_r = r^{-1}$ and $c_r = r^{-1/3}$ satisfy these conditions. The random variable $Z_r = [Y(x_r + c_r) - Y(x_r + c_r)] / (2c_r)$ can be viewed as an approximation to the derivative of the response function at x_r. This reduces the problem to find the maximum of $f(x)$ to that of finding the solution to the equation $f'(x) = 0$. The recursion (3) can then be thought of as a special case of the Robbins-Monro method for finding the solution $x = \theta$ to $f'(x) = 0$ by making successive observations on Z.

In this paper the procedure of Kiefer and Wolfowitz is combined with the adaptive Robbins-Monro procedure (2) described above for estimating the maximum of the response function $\pi(x)$, abbreviated as KW hereafter. At each step m observations are taken at $x_r \pm c_r$ so that the numerical derivative becomes $Z_r = [Y(x_r + c_r) - Y(x_r + c_r)] / (2mc_r)$ where $Y(x_r + c_r) \sim bin(m, \pi(x_r + c_r))$ and $Y(x_r - c_r) \sim bin(m, \pi(x_r - c_r))$. A graphical illustration is given in Figure 2. Starting at an arbitrary x_1, the value on x_r is then updated via the adaptive Robbins-Monro procedure

$$x_{r+1} = x_r - \left(r\widehat{\beta}_r\right)^{-1} z_r$$

where $\widehat{\beta}_r$ is the least squares estimator of the slope in the regression Z on x. The estimate of the optimum point at the $r:th$ step will then be given by

$$\widehat{\theta}_r = x_r.$$

Kiefer and Wolfowitz put some conditions on the response function that prevents it from being too steep or too flat (outside the neighborhood of θ). If the curve is too steep it may cause unduly large changes in x. If the curve is too flat (at a distance from θ) it is impossible to know in which direction to take the next step. They comment however that it will be sufficient if the conditions are fulfilled in an interval $[D_1,\ D_2]$. No observations will be taken outside the limits of this interval. The flatness of the logistic response curve may pose some problems unless there is knowledge about an appropriate interval before the experiment is started. In some situations it may be the case that such information is not available. The success of this procedure is also dependent upon good choices of the starting values x_1, β_1 and c_1. If the starting point is too far off there will be nearly zero probability of obtaining a response so that no information can be gained. The starting value for c_r should be small enough to avoid unduly large changes in x and large enough to enable the sequence to move away from a bad starting point. The relation between the number of observations taken at each step (m) and the number of steps (r) is another aspect that probably will be important.

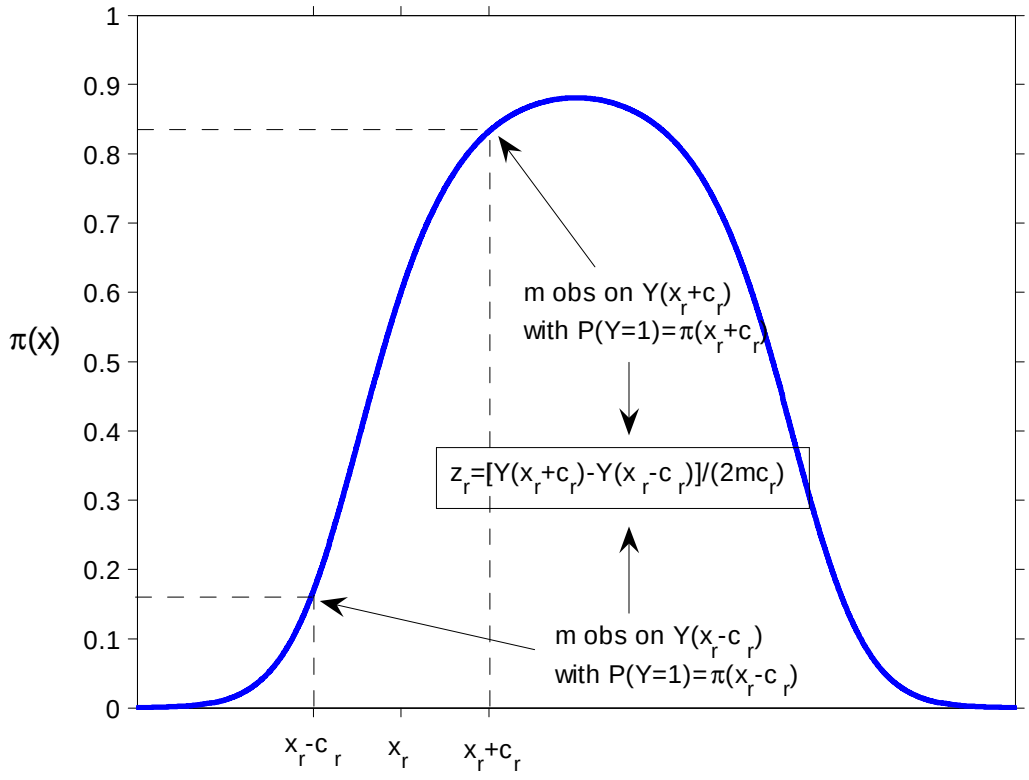


Figure 2: A graphical illustration of the KW approach.

3.2 c-optimal sequential designs

A design point represents a specific level of the control variable, e.g. a specific adjustment of the machine speed in the manufacturing experiment. An experimental design specifies the number of design points (n), the number of replicates at each design point (m) and the placing of the design points. Let a design denoted as ξ be summarized as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{array} \right\}, \qquad \sum_{i=1}^{n} w_i = 1$$

where the design weight $w_i = m_i/N$ is the proportion of the total number of observations allocated to the design point x_i. The amount of information in a design ξ is reflected by the size of the standardized information matrix $\mathbf{M}(\xi)$. Let $\mathbf{X}$ be the $n \times p$ matrix with $\mathbf{x}_i^T$ as rows and let $\mathbf{W}$ be the $n \times n$ diagonal matrix containing the design weights. Also let $\mathbf{V}$ be a diagonal matrix with GLM weights defined as

$$v(x_i) = \frac{1}{Var(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2.$$

The standardized information matrix for any GLM can then be expressed as

$$\mathbf{M}(\xi) = \mathbf{X}^T \mathbf{W} \mathbf{V} \mathbf{X}.$$

A criterion function $\Psi\{\mathbf{M}(\xi)\}$ is a function of the information matrix that represents some specific information aspect of a design. The criterion function should be chosen to match the aim of the experiment. The optimal design is then the choice of points and corresponding weights that optimize the selected criterion function. The criterion function is commonly related in some way to the precision of the parameter estimates, such as the size of a confidence region or the sum of the variances of the parameter estimates. It can concern estimation of all parameters, a subset or some function of the parameters.

In the case of estimating the point of optimum response; $\theta = -\beta_1/(2\beta_2)$, interest is in a function of two of the parameters. Estimating θ with minimum variance would be desirable justifying the following criterion function to be minimized

$$\begin{aligned} \Psi\{\mathbf{M}(\xi)\} &= \mathbf{c}^T \mathbf{M}^{-1}(\xi)\, \mathbf{c}, \\ \text{with } \mathbf{c}^T &= \begin{pmatrix} 0 & -\frac{1}{2\beta_2} & \frac{\beta_1}{2\beta_2^2} \end{pmatrix} \end{aligned}$$

because $var\left(\widehat{\theta}\right) \propto \mathbf{c}^T \mathbf{M}^{-1}(\xi)\, \mathbf{c}$. A design that minimizes $\mathbf{c}^T \mathbf{M}^{-1}(\xi)\, \mathbf{c}$ is a so called c-optimal design. The General Equivalence Theorem (Kiefer and Wolfowitz (1959) and Kiefer (1961)) states that the following holds for a c-optimal design ξ^*

$$v(x) \left\{ \mathbf{x}^T \mathbf{M}^{-1}(\xi^*)\, \mathbf{c} \right\}^2 \leq \mathbf{c}^T \mathbf{M}^{-1}(\xi^*)\, \mathbf{c} \tag{4}$$

for all x. This theorem provides an opportunity to check whether a suggested design is c-optimal or not.

The GLM weights for the logistic model are $v(x_i) = m_i\pi_i(1-\pi_i)$ so that the standardized information matrix is given by

$$\mathbf{M}(\xi,\boldsymbol{\beta}) = \begin{bmatrix} 1 & 1 & .. & 1 \\ x_1 & x_2 & .. & x_n \\ x_1^2 & x_2^2 & .. & x_n^2 \end{bmatrix} diag \begin{bmatrix} w_1 \\ w_2 \\ : \\ w_n \end{bmatrix} diag \begin{bmatrix} m_1\pi_1(1-\pi_1) \\ m_2\pi_2(1-\pi_2) \\ : \\ m_n\pi_n(1-\pi_n) \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ : & : & : \\ 1 & x_n & x_n^2 \end{bmatrix}.$$

The notation $\mathbf{M}(\xi,\boldsymbol{\beta})$ is used to stress the fact that the information matrix depends on the unknown model parameters. Assuming that a design consists of n points the minimization of $\Psi\{\mathbf{M}(\xi,\boldsymbol{\beta})\} = \mathbf{c}^T\mathbf{M}^{-1}(\xi,\boldsymbol{\beta})\,\mathbf{c}$ can be achieved with numerical methods. Sometimes the problem with a singular information matrix causes trouble in the derivation of c-optimal designs. Atkinson and Donev (1992) show that this problem can be worked out by adding a small number ε to the diagonal elements of $\mathbf{M}(\xi,\boldsymbol{\theta})$ before inversion. A 2-point design was assumed for the two parameter sets presented in Table 1 in the previous section, $\boldsymbol{\beta}_A$ and $\boldsymbol{\beta}_B$. Due to the symmetry property of the response curve that $\pi(x+\theta) = \pi(-x+\theta)$ one might expect that the optimal design is symmetric as well. The minimization of $\Psi\{\mathbf{M}(\xi,\boldsymbol{\beta})\}$ for the models A and B was carried out with the help of the program Mathcad 12, resulting in two sets of weights and design points. After checking the condition given by (4) the c-optimal designs were verified and found to be

$$\xi_A^* = \begin{Bmatrix} -5.2529 & 5.2529 \\ 0.5 & 0.5 \end{Bmatrix}$$

$$\xi_B^* = \begin{Bmatrix} -0.8306 & 0.8306 \\ 0.5 & 0.5 \end{Bmatrix}$$

As expected both designs consist of two points with equal weights that are symmetric around the maximum $\theta = 0$. It may also be noted that the optimal designs are distinct for the two parameter sets. c-optimal designs were derived for several other parameter sets and all of them turned out to be equally weighted 2-point designs with the points symmetric around θ. Henceforth it is assumed that the c-optimal designs consist of two points with equal weight.

The proposed sequential design is an attempt to deal with the problem of parameter dependence for the logistic model. The idea is to combine the features of c-optimality with the updating knowledge feature of a sequential design. Two points are taken at a certain distance from each other, a distance that does not necessarily decrease as opposed to the KW approach. This parametric sequential approach, called COPT, can be described by the following steps:

1. Choose an initial design $\xi_{(initial)}$
2. Estimate the parameters $\widehat{\boldsymbol{\beta}}^{(initial)}$
3. Minimize $\widehat{\mathbf{c}}^T\mathbf{M}^{-1}\left(\xi,\widehat{\boldsymbol{\beta}}^{(initial)}\right)\widehat{\mathbf{c}}$ to find the locally c-optimal design given $\widehat{\boldsymbol{\beta}}^{(initial)} \rightarrow \xi_1^* = \left\{ \begin{array}{cc} x_1 & x_2 \\ 0.5 & 0.5 \end{array} \right\}$
4. Take m observations at the design points in ξ_1^* and estimate the parameters $\widehat{\boldsymbol{\beta}}^{(1)}$
5. Minimize $\widehat{\mathbf{c}}^T\mathbf{M}^{-1}\left(\xi,\widehat{\boldsymbol{\beta}}^{(1)}\right)\widehat{\mathbf{c}}$ to find the locally c-optimal design given $\widehat{\boldsymbol{\beta}}^{(1)} \rightarrow \xi_2^* = \left\{ \begin{array}{cc} x_3 & x_4 \\ 0.5 & 0.5 \end{array} \right\}$

etc.

The parameters are estimated with the maximum likelihood method. Maximum likelihood estimates for GLMs can be obtained with the method of scoring, which is iterative, as described in for example Dobson (2002). The estimate of the optimum point after r steps is then given by

$$\widehat{\theta}_r = -\frac{\widehat{\beta}_1^{(r)}}{2\widehat{\beta}_2^{(r)}}.$$

If the initial design contains N^* observations and $2m$ observations are taken at each step there are $N^* + 2mr = N^* + N = N_{tot}$ observations in total after r steps. The observed standardized information matrix can then be expressed as

$$\mathbf{M}_{obs}\left(\xi_{obs},\widehat{\boldsymbol{\beta}}\right) = \begin{bmatrix} 1 & 1 & .. & 1 \\ x_1 & x_2 & .. & x_{N_{tot}} \\ x_1^2 & x_2^2 & .. & x_{N_{tot}}^2 \end{bmatrix} diag \begin{bmatrix} w_1 \\ w_2 \\ : \\ w_{N_{tot}} \end{bmatrix} diag \begin{bmatrix} m\widehat{\pi}_1\left(1-\widehat{\pi}_1\right) \\ m\widehat{\pi}_2\left(1-\widehat{\pi}_2\right) \\ : \\ m\widehat{\pi}_{N_{tot}}\left(1-\widehat{\pi}_{N_{tot}}\right) \end{bmatrix} \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ : & : & : \\ 1 & x_{N_{tot}} & x_{N_{tot}}^2 \end{bmatrix}$$

where $\widehat{\pi}_i = \left(e^{\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}}\right) / \left(1 + e^{\mathbf{x}_i^T\widehat{\boldsymbol{\beta}}}\right)$. Another version of the previous sequential design, called COPT2, makes use of the observed information matrix in the following way. $\mathbf{M}\left(\xi,\widehat{\boldsymbol{\beta}}\right)$ is replaced by a weighted information matrix computed as

$$\mathbf{M}_W\left(\xi,\widehat{\boldsymbol{\beta}}\right) = \frac{N_{tot}}{N_{tot}+2m}\mathbf{M}_{obs}\left(\xi_{obs},\widehat{\boldsymbol{\beta}}\right) + \frac{2m}{N_{tot}+2m}\mathbf{M}\left(\xi,\widehat{\boldsymbol{\beta}}\right)$$

and $\widehat{\mathbf{c}}^T\mathbf{M}_W^{-1}\left(\xi,\widehat{\boldsymbol{\beta}}\right)\widehat{\mathbf{c}}$ is minimized instead. $\mathbf{M}_{obs}$ is the information matrix from all observations up to this point and ξ_{obs} is the design consisting of these observations. This is potentially an improvement because it also takes into account the information from the preceding steps. Besides the changed information matrix everything is the same as for COPT.

4 A simulation study

The efficiencies of the suggested sequential designs are to be evaluated in a simulation study. There are many questions to bring clarity about including: Is any of the approaches superior? Is there an optimal choice of c for the KW approach? Is the weighted version of the information matrix preferred over the unweighted? What is best, taking few steps with many observations at each step, or taking many steps with fewer observations? Are the performances robust to misspecifications of the model?

All three approaches: KW, COPT and COPT2, are started from the same initial design before they take separate paths. Two different initial designs are used, see Table 2. Design 1 is constructed to be better than Design 2 for estimation of the optimum point in models like A and B where $\theta = 0$. The c-optimal designs are characterized by two symmetric points and design 1 adopts this property, although it is nothing like a c-optimal design considering the number of points. It might be a reasonable strategy at the initial stage of the experiment not to focus on just two points. Design 2 is constructed based on the same principle but pushed aside, intended to represent a beforehand assumption that θ is close to 2. The initial designs consist of a fairly large amount of observations ($N^* = 150$) to avoid the problem of non-existing maximum likelihood estimates, which is also a motive for choosing many design points.

Table 2: Two initial designs.

Design 1	-5	-3	-1	-0.5	0.5	1	3	5
	$\frac{5}{150}$	$\frac{10}{150}$	$\frac{20}{150}$	$\frac{40}{150}$	$\frac{40}{150}$	$\frac{20}{150}$	$\frac{10}{150}$	$\frac{5}{150}$
Design 2	-3	-1	1	1.5	2.5	3	5	7
	$\frac{5}{150}$	$\frac{10}{150}$	$\frac{20}{150}$	$\frac{40}{150}$	$\frac{40}{150}$	$\frac{20}{150}$	$\frac{10}{150}$	$\frac{5}{150}$

In the beginning observations are made on the response variable Y at the design points specified by the initial design. Initial maximum likelihood estimates $\widehat{\boldsymbol{\beta}}^{(initial)}$ are then calculated. Both the COPT and COPT2 approaches are now ready to start from step 3 (as described in section 3.2 above) with the minimization of $\widehat{\mathbf{c}}^T \mathbf{M}^{-1}\left(\xi, \widehat{\boldsymbol{\beta}}^{(initial)}\right)\widehat{\mathbf{c}}$. The *fminsearch* function in Matlab is used to carry out this task. The KW approach needs at least two observations on $Z = [Y(x+c) - Y(x+c)]/(2mc)$ before a least squares estimate of the slope parameter in the Z on x regression can be obtained. As discussed before the KW approach is also sensitive to the choices of x_1 and β_1. Some effort is therefore put in selecting good starting values. A first estimate of the optimum point is

$$\widehat{\theta}_{initial} = -\frac{\widehat{\beta}_1^{(initial)}}{2\widehat{\beta}_2^{(initial)}},$$

and the first design point in KW is set equal to this point:

$$x_1 = \widehat{\theta}_{initial}.$$

This point plus the two points given by $0.7\pi\left(\widehat{\theta}_{initial}\right)$, as depicted in Figure 3, comprise a start design for the KW approach. There is no elaborate reason for choosing 0.7 specifically, other than that it seemed to work good in comparison with several alternatives. Pairs of observations are made at these three design points that will result in three observations on Z which then can be used to obtain the starting value β_1. For $r = 2, 3, ...$ the design points are obtained according to (3) from section 3.1. Several values on c_r are tested, namely $c_r = cr^{-1/3}$ with $c = 1, 3, 5, 7, 9, 11, 13, 15$. Three different number of observations to be made at every design point are considered: $m = 5$, $m = 10$ and $m = 20$.

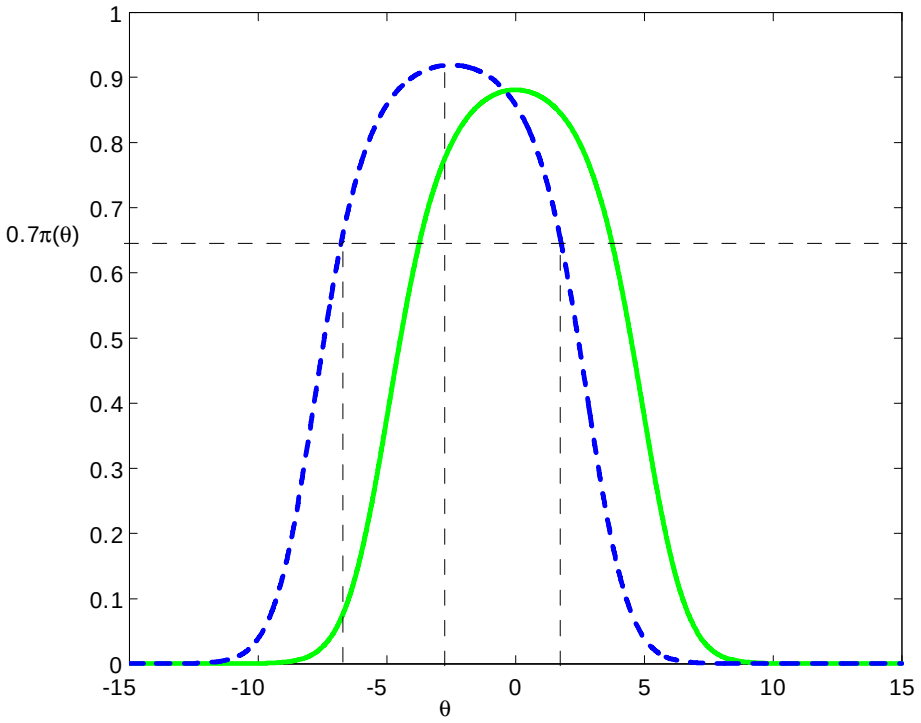


Figure 3: A representation of the choice of first design points for the KW approach. The true response curve is given by the solid line and the estimated curve is given by the broken line.

To examine the effects of a misspecified model another model with a different linear predictor is also evaluated. Model C defined according to

$$\begin{aligned} \eta_i &= -abs\left(x_i\right), \\ \pi_i &= \frac{e^{\eta_i}}{1+e^{\eta_i}}, \end{aligned}$$

is displayed in Figure 4 together with models A and B. The optimum point has the same location as for the logistic models but the response curve is lower at the maximum and it has a different shape. The responses are generated according to this model while the COPT and COPT2 approaches still (incorrectly) assume that the three parameter logistic model is true. The KW approach is nonparametric apart from the initial stage where the (inaccurate) logistic assumption remains unchanged.

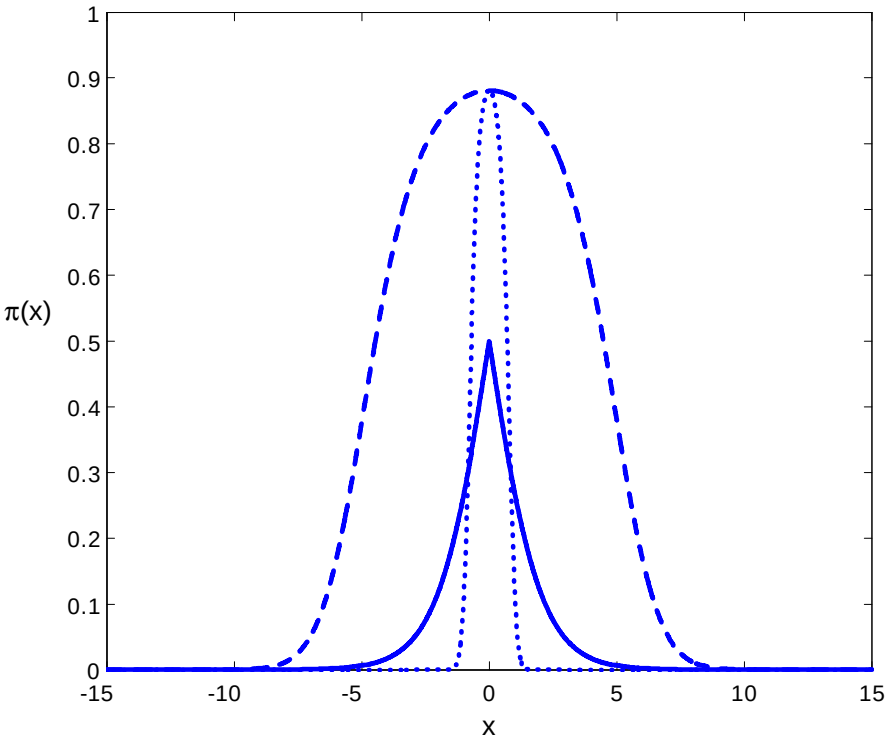


Figure 4: Models A (broken line), B (dotted line) and C (solid line).

4.1 Simulation results

The simulation results are presented in Tables 3 to 7. Each table shows the mean squared error (mse) of $\widehat{\theta}$ based on 500 simulated samples. The mse of $\widehat{\theta}$ before any of the sequential procedures is started, which is based on the initial maximum likelihood estimates of the parameters, is presented in the second column. In the following columns the mse of $\widehat{\theta}$ is given after in turn $N = 200$, $N = 600$ and $N = 1000$ observations in the sequential designs, in addition to the $N^* = 150$ observations that was made at the initial stage. If $\widehat{\theta}_{initial}$ is a "bad" point that is far from θ it may cause difficulties. For the parametric approaches there may be numerical problems related to the parameter estimation. These problems appear in

either that the maximum number of iterations is reached or that the computations results in NaN. NaN stands for "not a number" and turns up in Matlab when the computations do not produce a numerical result. Being at a position where the probability to obtain a response is nearly zero it is impossible for the KW approach to know in which direction to take next step; for such cases this approach never leaves the starting point. It can also happen that the KW sequence ends up at such a position even though the starting point is "good". $\widehat{\theta}$ is considered deteriorated when $\left|\widehat{\theta}\right| > 10$. The number of deteriorated starting points is shown in brackets in the second column. The simulation samples that failed, either because of numerical problems or because $\widehat{\theta}$ is deteriorated are discarded and the number of such samples is shown in brackets. In each table there are four sections; $a - d$ where a, b and c contain the results for $m = 20$, $m = 10$ and $m = 5$. As a reference section d contains the mse of $\widehat{\theta}$ based on taking all $N^* + N$ observations at the initial design and making one maximum likelihood estimate of θ. The $N^* + N$ observations are divided among the initial design points according to the same distribution as for the case with N^* observations. This nonsequential alternative is tested to make certain that one cannot do just as good without bothering about any sequential issues.

Model A, design 1. The mse for the two parametric approaches; COPT and COPT2, are remarkably similar for all N. The choice between different values of m does not seem to have any impact on the mse. For the KW approach it is evident that the constant c is important, the best choice of c varies from 7 to 11. If c is too small or too large there is a greater risk that the sequence degenerates to a point that is too far off from where it cannot recover. This happens more often when m is small. The best KW approach has about twice the mse of the COPT approaches for $m = 20$ and $m = 10$ and up to three times for $m = 5$. Both the sequential procedures beat the alternative of making all observations at one time.

Model A, design 2. The number of "bad" starting points has increased compared to the previous case because initial design 2 is inferior to design 1. The COPT approaches manage to sort out some of these problematic cases whereas the number stays the same or even increases for the KW approach. The mse for COPT and COPT2 are still similar. Although the mse is greater compared to when design 1 was used at the initial stage it is about the same size as before for all N. For the KW method the best choice of c seems to be somewhere between 9 and 13. The mse for the KW approach does not differ that much compared to when initial design 1 was used for the well-behaved cases. However, it should be noted that the problematic cases have increased in number compared to the previous case. The COPT approaches are preferred over the KW approach for all N. The sequential procedures are still performing better than the non-sequential alternative for all N with mse in view, while there exist no failing cases when all observations are made at the same time. All three alternatives for m give about the same results.

Model B, design 1. The mse is lower even at the initial stage for this model and

it decreases even more as N increases for the COPT approaches. The KW scheme remains at $\widehat{\theta}_{initial}$ almost every time which is also reflected in the fact that the mse is not changing. The COPT routines are associated with the lowest mse followed by the non-sequential alternative. The results are also the same regardless of the size of m.

Model B, design 2. Another problem that can occur is that of nonexisting maximum likelihood estimates. If the data pattern is such that the responses are separated from the nonresponses it is not possible to obtain any maximum likelihood estimates, see Albert and Anderson (1984). It is only when data are overlapped, that is it is not possible to separate the two different response types, that parameter estimates exist. For design 2 there are mainly two data points (± 1) where both responses and nonresponses can be observed, in all other design points the probability of a response is nearly zero. Almost all samples will therefore be separated, only about 2 percent of the samples are overlapped which is the reason for not presenting these results. This shows how important it is to take the problems with non-existing estimates seriously.

Model C, design 1. The mse of $\widehat{\theta}$ at the initial design is lower when compared to model A but higher in comparison with model B. The COPT and COPT2 procedures continue to perform equally well. As more observations are made the mse decreases although the reduction is not as large as for model A, so the mse is a little higher than for the other two models for all N. It would be surprising if the parametric approaches were completely unaffected by the misspecification of the model. The nonparametric approach on the other hand, that would be expected to handle the erroneous model the best, works no good. For $m = 20$ the mse even increases with N for some constants c and for some c it is unchanged. For $m = 10$ or $m = 5$ the mse is at best slightly reduced. The response curve for model C is lower and has a different shape which results in that z becomes equal to zero more often. As a consequence it is more common that the sequence gets stuck at some point. The descending order of performance turned out to be as follows; parametric sequential > parametric nonsequential >nonparametric sequential.

Model C, design 2. The fact that design 2 is poor is reflected in a higher mse at the initial stage. COPT and COPT2 recover to almost the same level as when design 1 was used in the following steps. The results for the nonparametric procedure show the same tendency that the mse sometimes gets even higher as N grows or just decreases a little. The same order of performance that puts the parametric sequential design before the parametric nonsequential followed by the nonparametric method prevails.

Table 3: Simulation results of three sequential designs for estimating the optimum point in model A with initial design 1, mse based on 500 samples.

Design	initial design	$N = 200$	$N = 600$	$N = 1000$
		a) m=20		
$COPT$	0.41 (1)	0.020 (1)	0.0075 (0)	0.0044 (0)
$COPT2$	0.41 (1)	0.023 (1)	0.0071 (0)	0.0043 (0)
$KW, c = 1$	0.41 (1)	1.25 (1)	0.87 (1)	0.84 (1)
$KW, c = 3$	0.41 (1)	0.27 (1)	0.11 (1)	0.088 (1)
$KW, c = 5$	0.41 (1)	0.075 (1)	0.043 (1)	0.037 (1)
$KW, c = 7$	0.41 (1)	0.040 (1)	0.020 (1)	0.017 (1)
$KW, c = 9$	0.41 (1)	0.086 (1)	0.015 (1)	0.010 (1)
$KW, c = 11$	0.41 (1)	0.24 (1)	0.023 (1)	0.011 (1)
$KW, c = 13$	0.41 (1)	0.35 (1)	0.072 (1)	0.021 (1)
$KW, c = 15$	0.41 (1)	0.40 (1)	0.17 (1)	0.057 (1)
		b) m=10		
$COPT$	0.55 (1)	0.024 (0)	0.0084 (0)	0.0048 (0)
$COPT2$	0.55 (1)	0.022 (0)	0.0071 (0)	0.0044 (0)
$KW, c = 1$	0.55 (1)	2.40 (10)	1.73 (14)	1.91 (14)
$KW, c = 3$	0.55 (1)	0.21 (4)	0.16 (4)	0.14 (4)
$KW, c = 5$	0.55 (1)	0.10 (5)	0.060 (5)	0.056 (5)
$KW, c = 7$	0.55 (1)	0.047 (2)	0.029 (2)	0.025 (2)
$KW, c = 9$	0.55 (1)	0.034 (2)	0.019 (2)	0.013 (2)
$KW, c = 11$	0.55 (1)	0.072 (2)	0.015 (2)	0.011 (2)
$KW, c = 13$	0.55 (1)	0.20 (1)	0.020 (1)	0.011 (1)
$KW, c = 15$	0.55 (1)	0.33 (1)	0.040 (1)	0.014 (1)
		c) m=5		
$COPT$	0.51 (1)	0.021 (0)	0.0075 (0)	0.0045 (0)
$COPT2$	0.51 (1)	0.023 (0)	0.0071 (0)	0.0047 (0)
$KW, c = 1$	0.51 (1)	4.66 (35)	3.67 (40)	3.87 (46)
$KW, c = 3$	0.51 (1)	0.47 (3)	0.27 (3)	0.24 (3)
$KW, c = 5$	0.51 (1)	0.15 (8)	0.11 (8)	0.085 (8)
$KW, c = 7$	0.51 (1)	0.10 (4)	0.052 (4)	0.050 (4)
$KW, c = 9$	0.51 (1)	0.048 (3)	0.030 (3)	0.025 (3)
$KW, c = 11$	0.51 (1)	0.099 (3)	0.022 (3)	0.016 (3)
$KW, c = 13$	0.51 (1)	0.16 (1)	0.043 (1)	0.041 (1)
$KW, c = 15$	0.51 (1)	0.10 (2)	0.017 (2)	0.012 (2)
		d) all N observations at the initial design		
	0.45 (1)	0.14 (0)	0.055 (0)	0.037 (0)

Table 4: Simulation results of three sequential designs for estimating the optimum point in model A with initial design 2, mse based on 500 samples.

Design	initial design	$N = 200$	$N = 600$	$N = 1000$
		a) m=20		
$COPT$	2.67 (14)	0.021 (14)	0.016 (10)	0.0043 (9)
$COPT2$	2.67 (14)	0.022 (13)	0.0081 (9)	0.0046 (8)
$KW, c = 1$	2.67 (14)	1.92 (32)	1.18 (32)	1.07 (32)
$KW, c = 3$	2.67 (14)	0.55 (29)	0.29 (29)	0.27 (29)
$KW, c = 5$	2.67 (14)	0.25 (23)	0.21 (22)	0.20 (22)
$KW, c = 7$	2.67 (14)	0.27 (20)	0.18 (20)	0.18 (20)
$KW, c = 9$	2.67 (14)	0.12 (20)	0.022 (20)	0.013 (20)
$KW, c = 11$	2.67 (14)	0.68 (16)	0.19 (16)	0.17 (16)
$KW, c = 13$	2.67 (14)	0.98 (18)	0.25 (18)	0.18 (18)
$KW, c = 15$	2.67 (14)	1.38 (16)	0.42 (16)	0.23
		b) m=10		
$COPT$	2.86 (23)	0.022 (10)	0.0080 (4)	0.0042 (4)
$COPT2$	2.86 (23)	0.023 (11)	0.0068 (2)	0.0044 (2)
$KW, c = 1$	2.86 (23)	4.02 (48)	3.14 (53)	2.59 (57)
$KW, c = 3$	2.86 (23)	0.25 (42)	0.19 (42)	0.17 (42)
$KW, c = 5$	2.86 (23)	0.099 (32)	0.27 (31)	0.27 (31)
$KW, c = 7$	2.86 (23)	0.067 (24)	0.038 (23)	0.040 (23)
$KW, c = 9$	2.86 (23)	0.11 (24)	0.030 (24)	0.020 (24)
$KW, c = 11$	2.86 (23)	0.11 (21)	0.022 (21)	0.018 (21)
$KW, c = 13$	2.86 (23)	0.25 (23)	0.035 (23)	0.017 (23)
$KW, c = 15$	2.86 (23)	0.55 (23)	0.052 (23)	0.021 (23)
		c) m=5		
$COPT$	2.26 (21)	0.10 (8)	0.0080 (7)	0.0044 (7)
$COPT2$	2.26 (21)	0.021 (9)	0.0069 (9)	0.0041 (9)
$KW, c = 1$	2.26 (21)	4.99 (66)	5.02 (75)	5.08 (82)
$KW, c = 3$	2.26 (21)	0.60 (33)	0.31 (34)	0.30 (34)
$KW, c = 5$	2.26 (21)	0.14 (32)	0.10 (32)	0.099 (32)
$KW, c = 7$	2.26 (21)	0.22 (27)	0.25 (27)	0.24 (27)
$KW, c = 9$	2.26 (21)	0.10 (28)	0.080 (28)	0.076 (28)
$KW, c = 11$	2.26 (21)	0.053 (26)	0.033 (26)	0.025 (26)
$KW, c = 13$	2.26 (21)	0.17 (26)	0.019 (26)	0.016 (26)
$KW, c = 15$	2.26 (21)	0.16 (26)	0.024 (26)	0.020 (26)
		d) all N observations at the initial design		
	2.67 (12)	0.87 (1)	0.21 (0)	0.11 (0)

Table 5: Simulation results of three sequential designs for estimating the optimum point in model B with initial design 1, mse based on 500 samples.

Design	initial design	$N = 200$	$N = 600$	$N = 1000$
		a) m=20		
$COPT$	0.0019 (1)	0.00041 (0)	0.00015 (1)	0.00010 (1)
$COPT2$	0.0019 (1)	0.00040 (0)	0.00015 (1)	$9.8 \cdot 10^{-5}$ (1)
$KW, c = 1$	0.0019 (1)	0.020 (1)	0.24 (1)	0.30 (1)
$KW, c = 3$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0018 (1)
$KW, c = 5$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
$KW, c = 7$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
$KW, c = 9$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
$KW, c = 11$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
$KW, c = 13$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
$KW, c = 15$	0.0019 (1)	0.0019 (1)	0.0019 (1)	0.0019 (1)
		b) m=10		
$COPT$	0.0020 (1)	0.00046 (1)	0.00018 (1)	$9.3 \cdot 10^{-5}$ (1)
$COPT2$	0.0020 (1)	0.00048 (1)	0.00016 (1)	$9.8 \cdot 10^{-5}$ (1)
$KW, c = 1$	0.0020 (1)	0.39 (3)	0.71 (3)	0.76 (3)
$KW, c = 3$	0.0020 (1)	0.0020 (1)	0.0021 (1)	0.0014 (1)
$KW, c = 5$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
$KW, c = 7$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
$KW, c = 9$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
$KW, c = 11$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
$KW, c = 13$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
$KW, c = 15$	0.0020 (1)	0.0020 (1)	0.0020 (1)	0.0020 (1)
		c) m=5		
$COPT$	0.0019 (3)	0.00041 (3)	0.00015 (3)	0.00011 (3)
$COPT2$	0.0019 (3)	0.00048 (3)	0.00017 (3)	$9.9 \cdot 10^{-5}$ (3)
$KW, c = 1$	0.0019 (3)	0.13 (6)	0.20 (7)	0.21 (7)
$KW, c = 3$	0.0019 (3)	0.0019 (3)	0.0027 (3)	0.0048 (3)
$KW, c = 5$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0018 (3)
$KW, c = 7$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0019 (3)
$KW, c = 9$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0019 (3)
$KW, c = 11$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0019 (3)
$KW, c = 13$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0019 (3)
$KW, c = 15$	0.0019 (3)	0.0019 (3)	0.0019 (3)	0.0019 (3)
		d) all N observations at the initial design		
	0.0020 (3)	0.00080 (0)	0.00043 (0)	0.00024 (0)

Table 6: Simulation results of three sequential designs for estimating the optimum point in model C with initial design 1, mse based on 500 samples.

Design	initial design	$N = 200$	$N = 600$	$N = 1000$
		a) m=20		
$COPT$	0.13 (0)	0.035 (0)	0.014 (0)	0.0087 (0)
$COPT2$	0.13 (0)	0.036 (0)	0.015 (0)	0.0096 (0)
$KW, c = 1$	0.13 (0)	0.98 (4)	1.34 (5)	1.31 (6)
$KW, c = 3$	0.13 (0)	0.19 (0)	0.21 (0)	0.19 (0)
$KW, c = 5$	0.13 (0)	0.21 (0)	0.24 (0)	0.22 (0)
$KW, c = 7$	0.13 (0)	0.14 (0)	0.17 (0)	0.16 (0)
$KW, c = 9$	0.13 (0)	0.13 (0)	0.12 (0)	0.10 (0)
$KW, c = 11$	0.13 (0)	0.13 (0)	0.13 (0)	0.12 (0)
$KW, c = 13$	0.13 (0)	0.13 (0)	0.13 (0)	0.13 (0)
$KW, c = 15$	0.13 (0)	0.13 (0)	0.13 (0)	0.13 (0)
		b) m=10		
$COPT$	0.12 (0)	0.030 (0)	0.013 (0)	0.0084 (0)
$COPT2$	0.12 (0)	0.032 (0)	0.013 (0)	0.0090 (0)
$KW, c = 1$	0.12 (0)	1.14 (6)	1.35 (8)	1.49 (9)
$KW, c = 3$	0.12 (0)	0.21 (2)	0.23 (3)	0.30 (3)
$KW, c = 5$	0.12 (0)	0.10 (0)	0.34 (2)	0.46 (2)
$KW, c = 7$	0.12 (0)	0.10 (0)	0.060 (1)	0.052 (1)
$KW, c = 9$	0.12 (0)	0.12 (0)	0.097 (0)	0.082 (0)
$KW, c = 11$	0.12 (0)	0.12 (0)	0.10 (0)	0.091 (0)
$KW, c = 13$	0.12 (0)	0.12 (0)	0.11 (0)	0.11 (0)
$KW, c = 15$	0.12 (0)	0.12 (0)	0.12 (0)	0.11 (0)
		c) m=5		
$COPT$	0.13 (0)	0.033 (0)	0.012 (0)	0.0079 (0)
$COPT2$	0.13 (0)	0.034 (0)	0.014 (0)	0.0082 (0)
$KW, c = 1$	0.13 (0)	2.70 (13)	2.95 (18)	2.86 (19)
$KW, c = 3$	0.13 (0)	0.63 (3)	1.01 (4)	1.09 (4)
$KW, c = 5$	0.13 (0)	0.25 (0)	0.54 (3)	0.61 (3)
$KW, c = 7$	0.13 (0)	0.11 (0)	0.15 (0)	0.14 (1)
$KW, c = 9$	0.13 (0)	0.12 (0)	0.10 (0)	0.084 (0)
$KW, c = 11$	0.13 (0)	0.13 (0)	0.10 (0)	0.098 (0)
$KW, c = 13$	0.13 (0)	0.13 (0)	0.12 (0)	0.11 (0)
$KW, c = 15$	0.13 (0)	0.13 (0)	0.12 (0)	0.11 (0)
		d) all N observations at the initial design		
	0.13 (0)	0.067 (0)	0.031 (0)	0.022 (0)

Table 7: Simulation results of three sequential designs for estimating the optimum point in model C with initial design 2, mse based on 500 samples.

Design	initial design	$N = 200$	$N = 600$	$N = 1000$
		a) m=20		
$COPT$	0.59 (2)	0.047 (0)	0.015 (0)	0.010 (0)
$COPT2$	0.59 (2)	0.042 (0)	0.015 (0)	0.0097 (0)
$KW, c = 1$	0.59 (2)	2.23 (9)	2.09 (9)	2.04 (9)
$KW, c = 3$	0.59 (2)	0.63 (6)	0.61 (6)	0.69 (6)
$KW, c = 5$	0.59 (2)	0.51 (5)	0.46 (5)	0.41 (5)
$KW, c = 7$	0.59 (2)	0.66 (2)	0.56 (2)	0.58 (2)
$KW, c = 9$	0.59 (2)	0.49 (3)	0.51 (3)	0.48 (3)
$KW, c = 11$	0.59 (2)	0.54 (2)	0.53 (2)	0.46 (2)
$KW, c = 13$	0.59 (2)	0.57 (2)	0.54 (2)	0.52 (2)
$KW, c = 15$	0.59 (2)	0.59 (2)	0.55 (2)	0.49 (2)
		b) m=10		
$COPT$	0.63 (0)	0.044 (0)	0.015 (0)	0.0091 (0)
$COPT2$	0.63 (0)	0.046 (0)	0.016 (0)	0.010 (0)
$KW, c = 1$	0.63 (0)	2.63 (11)	2.82 (14)	2.82 (14)
$KW, c = 3$	0.63 (0)	0.76 (4)	0.90 (4)	1.09 (4)
$KW, c = 5$	0.63 (0)	0.54 (3)	0.55 (4)	0.55 (4)
$KW, c = 7$	0.63 (0)	0.53 (0)	0.44 (0)	0.44 (1)
$KW, c = 9$	0.63 (0)	0.41 (1)	0.32 (1)	0.27 (1)
$KW, c = 11$	0.63 (0)	0.38 (2)	0.28 (3)	0.24 (3)
$KW, c = 13$	0.63 (0)	0.53 (1)	0.46 (1)	0.42 (1)
$KW, c = 15$	0.63 (0)	0.54 (1)	0.49 (1)	0.46 (1)
		c) m=5		
$COPT$	0.52 (2)	0.045 (0)	0.016 (0)	0.0095 (0)
$COPT2$	0.52 (2)	0.035 (0)	0.014 (0)	0.0088 (0)
$KW, c = 1$	0.52 (2)	5.00 (16)	5.19 (19)	5.32 (20)
$KW, c = 3$	0.52 (2)	1.54 (7)	1.98 (10)	2.05 (10)
$KW, c = 5$	0.52 (2)	0.71 (2)	0.89 (2)	0.92 (4)
$KW, c = 7$	0.52 (2)	0.36 (2)	0.45 (2)	0.50 (2)
$KW, c = 9$	0.52 (2)	0.30 (4)	0.24 (4)	0.47 (5)
$KW, c = 11$	0.52 (2)	0.56 (2)	0.50 (2)	0.48 (2)
$KW, c = 13$	0.52 (2)	0.50 (2)	0.44 (2)	0.40 (2)
$KW, c = 15$	0.52 (2)	0.62 (2)	0.60 (2)	0.66 (2)
		d) all N observations at the initial design		
	0.65 (1)	0.20 (0)	0.070 (0)	0.040 (0)

5 Concluding remarks

The parametric approach based on constructing a c-optimal design at each step was superior in all cases that were examined here. It even excelled the nonparametric approach in the case with an incorrect model. The results suggest that the problems of parameter dependence can be worked out by using any of the two parametric approaches. The effect of having the wrong beforehand idea of the location of the optimum point almost disappeared after taking 200 observations sequentially. The sequential parametric approaches always outperformed the nonsequential strategy of maximum likelihood estimation of θ based on making all observations at once. The potential improvement from using the observed information matrix in the construction of designs failed to appear. In fact the performances of COPT and COPT2 were almost identical.

Choosing a good constant c is of crucial importance for the success of the KW approach. In this setting a good choice turned out to be choosing c somewhere around 10. The main advantage with the nonparametric approach is its simplicity, that is not having to make any distributional assumptions. Opposite to what might be expected it was more sensitive to the misspecified model than the parametric approaches. There were several cases where the KW approach was outperformed by the nonsequential alternative of making all observations at the initial stage. It was dependent upon good initial estimates and thereby on a good initial design in the sense of avoiding a large number of degenerated estimates. There were also many cases for which the KW approach did not leave the starting point at all. Perhaps an interval with appropriate limits that prevents the sequence to degenerate is required.

Non-existing maximum likelihood estimates may cause problems as was apparent by the lack of results for model B in combination with design 2. There were no such problems for model A in combination with design 2 and the only difference between these two response curves is the scale on the control variable. This implies that considering the measurement scale of the control variable might be worthwhile.

The different combinations of the number of observations taken at each step, m, and the number of steps, r, did not have that large impact on the results for neither of the approaches, though the problems with degenerated cases tended to be more frequent for $m = 5$.

No general conclusions should be drawn but the empirical results imply that there are gains to be made by using a sequential c-optimal approach for finding the point of optimum response. More extensive evaluations and evaluations of theoretical nature are required.

References

[1] Albert, A. and Anderson, J.A. (1984). "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika* 71, 1-10.

[2] Anbar, D. (1978). "A Stochastic Newton-Raphson Method," *Journal of Statistical Planning and Inference* 2, 153-163.

[3] Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs*. Oxford University Press, New York

[4] Box, G.E.P. and Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces*. Wiley, New York.

[5] Chung, K.L. (1954). "On a Stochastic Approximation Method," *The Annals of Mathematical Statistics* 25, 463-483.

[6] Dobson, A.J. (2002). *An Introduction to Generalized Linear Models*. Chapman and Hall, London.

[7] Kiefer, J. and Wolfowitz, J. (1952). "Stochastic Estimation of the Maximum of a Regression Function," *The Annals of Mathematical Statistics* 23, 462-466.

[8] Kiefer, J. and Wolfowitz, J. (1959). "Optimum Designs in Regression Problems," *The Annals of Mathematical Statistics* 30, 271-294.

[9] Kiefer, J. (1961). "Optimum Designs in Regression Problems, II," *The Annals of Mathematical Statistics* 32, 298-325.

[10] Lai, T. L., and Robbins, H. (1979). "Adaptive Design and Stochastic Approximation," *Annals of Statistics* 7, 1196-1221.

[11] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd edition. Chapman and Hall, London.

[12] Robbins, H. and Monro, S. (1951). "A Stochastic Approximation Method," *The Annals of Mathematical Statistics* 22, 400-407.

[13] Sacks, J. (1958). "Asymptotic Distribution of Stochastic Approximation Procedures," *The Annals of Mathematical Statistics* 29, 373-405

[14] Wu, C. F. J. (1985). "Efficient Sequential Designs With Binary Data," *Journal of the American Statistical Association* 80, 974-984.

[15] Wu, C. F. J. (1986). "Maximum Likelihood Recursion and Stochastic Approximation in Sequential Designs," *Adaptive Statistical Procedures and Related Topics* 8, 298-313.

91-7155-245-6