Research Report

Department of Statistics

No. 2005:2

D-optimal Designs for Quadratic Logistic Regression Models

Ellinor Fackle Fornius

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

D-optimal Designs for Quadratic Logistic Regression Models

Ellinor Fackle Fornius
Department of Statistics
Stockholm University

Abstract

D-optimal designs are derived for certain quadratic logistic regression models. The performance of the D-optimal designs regarding maximum likelihood estimation of the model parameters and estimation of the optimum of the response function is studied for different sample sizes. Comparisons are made with a couple of non-optimal designs. There were found to be disagreements between the asymptotic distribution and the small sample distribution of the maximum likelihood estimator. The designs are also evaluated as to what extent they suffer from the problem of non-existence of the maximum likelihood estimator. The probability that the maximum likelihood estimate exists is compared for the various designs. Non-existence proved to be a substantial problem for these quadratic logistic models.

Keywords: D-optimal design, Logit model, Response Surface Methodology, Maximum likelihood estimator, Separation.

1 Introduction

The statistical methods for designing and analyzing the outcome of experiments where interest is in a response variable that is affected by one or several variables are known as *Response Surface Methodology* (RSM). A thorough exposition of these techniques is given in the book by Box and Draper (1987). RSM is traditionally used for finding optimum operating conditions in the industry. It is now common in many different fields like physical, chemical, biological, clinical and social sciences.

The principal objective of RSM is to explore the unknown relationship between the response/output variable and the control/input variables. For example, the response variable might be the yield of a chemical process and the control variables might be temperature and pressure, or the response variable might be the reaction time of an individual and the control variables might then be dose of alcohol and amount of sleep. RSM attempts to answer questions about how the response variable behaves when the levels of the control variables are changed, e.g. what happens to the yield when the temperature and pressure levels are varied. The true unknown function that describes this relationship is often locally approximated by a polynomial in the region of interest. Thus, this approximating function is not expected to be valid outside the limited region of interest. One common application for RSM is to find the optimum operating conditions, e.g. to find out for what levels of temperature and pressure the yield is maximized.

RSM is essentially a sequential procedure where the experimental design is gradually updated as investigation proceeds. Initially choices have to be made regarding the model, the number of replicates, the levels of the control variables and the size and location of the region of interest. The objective is that the procedure is such that the right conclusions can be drawn even if the initial experimental design is poor and that the path to arrive there is as short as possible, for details see Box and Draper (1987).

It is often assumed that the responses are normally distributed and an ordinary regression model is used. However, there are many situations where this is not the case. An example is a binary response variable that assumes one of two possible values, "success" or "failure". *Generalized Linear Models* (GLMs) is a class of models that can be used in such situations, which include the linear regression model as a special case but also allows discrete data. Examples on GLMs also include logistic, probit and log-linear models. GLMs are treated in several books, e.g. Dobson (2002) and McCullagh and Nelder (1989). For an overview of the use of Generalized linear models in RSM, see the article by Khuri (2001). The problem of finding the optimum operating conditions when the response variable is binary is about determining the levels of the control variables that give the maximum/minimum of the probability of "success"/"failure".

When estimating the parameters in the model that is assumed to describe the relationship between the response and the control variables the aim is to choose a design that gives the highest possible precision in these estimates, i.e. to choose an optimal design. The problem of finding an optimal design is treated in, for example, Atkinson and Donev (1992), Silvey (1980) and Fedorov and Hackl (1997). A problem with finding optimal designs for GLMs is that the optimal design generally depends on the true parameters. If the true parameters are known it is possible to find an optimal design but then at the same time there is no need for estimation at all. Another problem that can arise with maximum likelihood estimation of the parameters in the logistic model and small samples is that the maximum likelihood estimate does not always exist.

The theoretical results about optimal designs for GLMs are asymptotic results. In practice experiments are often restricted by time and money constraints. Furthermore, the true parameters are generally unknown in practice. The purpose of this paper is to derive optimal designs for a quadratic logistic model given four different sets of true parameters and to examine the properties of these designs in small samples. In addition, comparisons will be made with the properties of some non-optimal designs. This paper is organized as follows. GLM, the model and the sets of parameters are outlined in section 2. Section 3 presents the derivation of the optimal designs. The results of the maximum likelihood estimation are given in section 4. A concluding discussion is found in the final section.

2 Model

A logistic regression model with one control variable and a quadratic term is examined in this paper. The logistic model belongs to the generalized linear models which generally are characterized by three components: (1) The distribution of each of the independent response variables $Y_1, ..., Y_N$ belongs to the exponential family (2) The linear predictor $\eta = \mathbf{x}^T\boldsymbol{\beta}$ is a linear (in β) combination of k control variables $(x_1, x_2, ..., x_k)$ and p parameters and (3) The link function $g(\mu)$ specifies the relationship between the expected value of the response variable $(E(Y) = \mu)$ and the linear predictor. This is a monotonic and differentiable function.

The response variable in a logistic model is binary. Success/failure, broken/not broken and pass a test/not pass a test are examples on outcomes of binary response variables. The responses are independent and Bernoulli distributed

$$Y_i \sim bern(\pi_i) = bin(1, \pi_i),$$

with the logit link function

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

and

$$\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}.$$

The linear predictor for the logistic model with one control variable and a quadratic term is given by

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

The probability of "success", π, is a function of the control variable x. The levels of the control variable are set by the experimenter. For example, consider manufacturing a food product, then x might be the quantity of an additive which affects π, the probability that the product is good. If the quantity is too small or too large the probability of a good product is small but for some quantity π is maximized. In this case interest is in determining the value of the control variable for which π is maximized, i.e. determining the optimum operating conditions. For some applications interest is instead in minimizing the probability of an undesirable outcome. Because of the fact that the link function is monotonic optimizing π is equivalent to optimizing η, thus the optimum point x_m is obtained according to

$$\frac{\partial}{\partial x}\eta = 0 \Rightarrow x_m = -\frac{\beta_1}{2\beta_2}.$$

The response curve that describes π as a function of x is symmetric around this point. Taking the second derivative of η

$$\frac{\partial^2}{\partial x^2}\eta = 2\beta_2,$$

shows that whether $\pi(x)$ has a maximum or a minimum is determined by the sign of the parameter β_2. The parameter β_0 determines the height of the curve in the optimum point. The linear predictor in the optimum point is given by

$$\eta(x_m) = \beta_0 + \beta_1 x_m + \beta_2 x_m^2 = \beta_0 - \frac{\beta_1^2}{4\beta_2}.$$

If the response curve has a maximum ($\beta_2 < 0$) a larger β_0 means that the maximum of $\pi(x)$ is closer to 1 and if there is a minimum ($\beta_2 > 0$) a larger β_0 means that the minimum of π is closer to 1. Furthermore, the size of β_2 determines the relative width of the response curve for a given height and a larger absolute value of β_2 means a more narrow curve. Thus, the parameters determine the shape of the function $\pi(x)$. The four sets of parameters that are examined in this paper are given in Table 1 and displayed in Figure 1. The sets are chosen to represent different variations of the shape of the response curve. The response curve is considered to be high when the maximum value of $\pi(x)$ is close to 1 and low when the maximum value of $\pi(x)$ is close to 0. Given the scale on x the response curves have different widths. Two of the curves are named wide and two are named narrow. The meaning of these labels should be understood in a relative sense.

Table 1: Four parameter sets labeled according to their associated characteristics of $\pi(x)$.

Type of response curve	**Parameter set**
"High-wide"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$
"High-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$
"Low-wide"	$\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -0.1 \end{pmatrix}^T$
"Low-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$

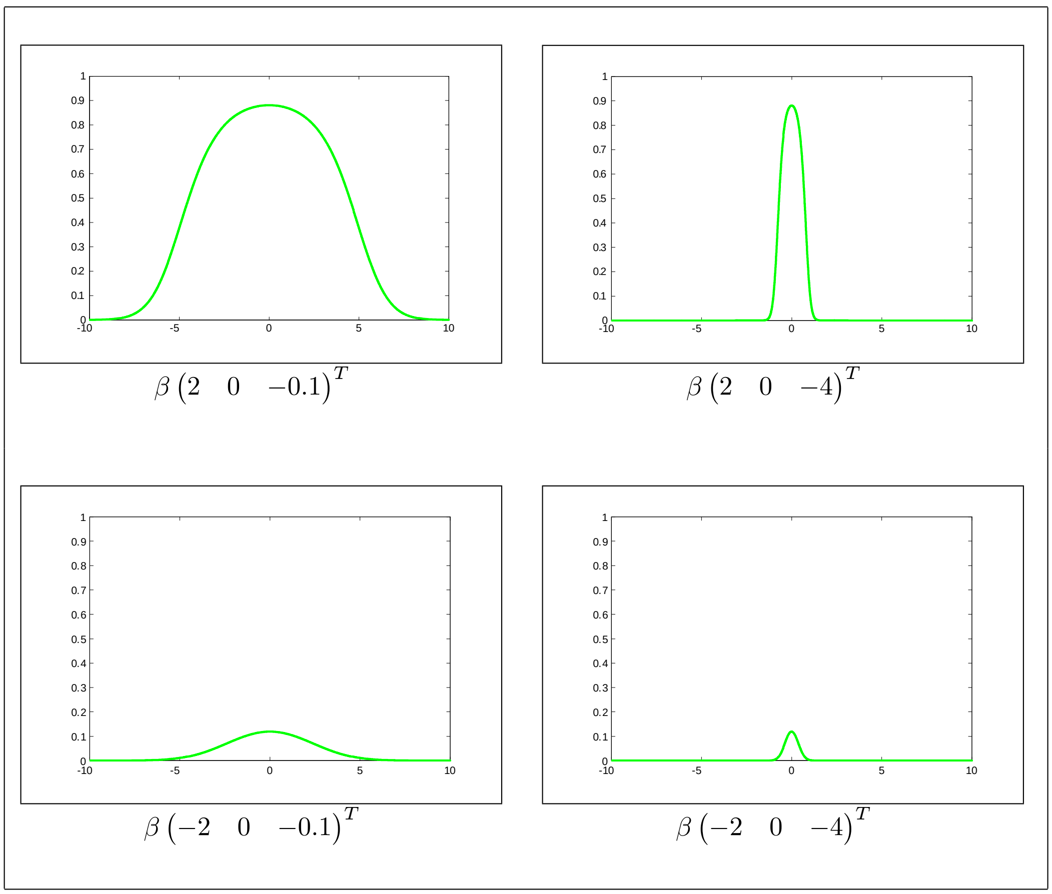

Figure 1: The probability π is plotted against the control variable x for the four parameter sets in Table 1.

The maximum likelihood estimates of the model parameters are found by using the method of scoring (see e.g. Dobson, 2002, Section 4.3). Let $\mathbf{x}_i^T$ be a row-vector of control variables and let $\mathbf{X}$ be the $N \times p$ matrix with $\mathbf{x}_i^T$ as rows. Also let $\mathbf{V}$ be the $N \times N$ diagonal matrix with weights

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

The asymptotic sampling distribution of the MLE is normal with covariance $\mathbf{I}^{-1} = \left(\mathbf{X}^T \mathbf{V} \mathbf{X}\right)^{-1}$. That is, the MLE is a consistent and asymptotically efficient estimator of $\boldsymbol{\beta}$. Variance estimates can be obtained by the diagonal elements in

$$\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}^T \widehat{\mathbf{V}} \mathbf{X}\right)^{-1},$$

where $\widehat{\mathbf{V}}$ is the same as $\mathbf{V}$ but with the MLE of $\boldsymbol{\beta}$ used instead of $\boldsymbol{\beta}$. The maximum likelihood estimator of the optimum point x_m is given by

$$\hat{x}_m = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}.$$

This estimator is also consistent with asymptotic variance

$$\begin{aligned} V\left(g\left(\hat{\boldsymbol{\beta}}\right)\right) &= V(\hat{x}_m) = \left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right)^T \mathbf{V}\left(\hat{\boldsymbol{\beta}}\right) \left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right) \\ &= \left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right)^T \mathbf{I}^{-1} \left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right), \end{aligned}$$

with

$$\left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right)^T = \begin{pmatrix} 0 & -\frac{1}{2\beta_2} & \frac{\beta_1}{2\beta_2^2} \end{pmatrix}.$$

An estimate of $V\left(g\left(\hat{\boldsymbol{\beta}}\right)\right)$ can be obtained as follows.

$$\widehat{V}\left(g\left(\hat{\boldsymbol{\beta}}\right)\right) = \widehat{V}(\hat{x}_m) = \left(\frac{\partial \widehat{x}_m}{\partial \widehat{\boldsymbol{\beta}}}\right)^T \widehat{\mathbf{V}}\left(\hat{\boldsymbol{\beta}}\right) \left(\frac{\partial \widehat{x}_m}{\partial \widehat{\boldsymbol{\beta}}}\right).$$

3 Design

When designing an experiment the objective is to obtain as much information as possible. A criterion function, $\Psi(\xi, \boldsymbol{\theta})$, can be used to reflect the amount of information in a design and therefore to decide which design is the best. An optimal design is constructed by selecting the levels of the control variables, the design points, and the proportions of the total number of observations to be allocated to the design points, the design weights, in a way that the criterion function is optimized. Different criterion functions are used for different situations depending on the aim of the experiment. For example, when the aim is to explore the relationship between the response and the control variables, i.e. to estimate the model parameters, the criterion function is different from when the aim is to find the optimum operating conditions. Thus, the optimal design for these two situations will also be different.

A design can be denoted as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_n \\ w_1 & w_2 & \cdots & w_n \end{array} \right\},$$

$$w_i \geq 0, \qquad \sum_{i=1}^{n} w_i = 1,$$

where $x_1, x_2, ..., x_n$ represent the design points and $w_1, w_2, ..., w_n$ represent the corresponding design weights. The information obtained from an observation at the design point x_i is given by

$$\mathbf{m}\left(\boldsymbol{\theta}, x_i\right) = v\left(x_i\right) \mathbf{x}_i \mathbf{x}_i^T.$$

when the model belongs to the GLMs. The standardized information matrix for a design can be expressed as the weighted sum of the information from the individual design points in the $p \times p$ matrix

$$\mathbf{M}\left(\xi, \boldsymbol{\theta}\right) = \sum_{i=1}^{n} w_i \mathbf{m}\left(\boldsymbol{\theta}, x_i\right).$$

The Fisher information for the design is given by

$$\mathbf{I}\left(\xi, \boldsymbol{\theta}\right) = N \mathbf{M}\left(\xi, \boldsymbol{\theta}\right)$$

and the asymptotic covariance matrix of the MLE of $\boldsymbol{\theta}$ is given by the inverse of the Fisher information matrix

$$\mathbf{I}^{-1}\left(\xi, \boldsymbol{\theta}\right) = \frac{1}{N} \mathbf{M}^{-1}\left(\xi, \boldsymbol{\theta}\right).$$

Different designs will thus lead to different asymptotic sampling distributions of the MLE. The standardized predictor variance for a GLM is defined as

$$d\left(x, \xi\right) = v\left(x\right) \mathbf{x}^T \mathbf{M}^{-1}\left(\xi, \boldsymbol{\theta}\right) \mathbf{x} = tr\left[\mathbf{m}\left(\boldsymbol{\theta}, x\right) \mathbf{M}^{-1}\left(\xi, \boldsymbol{\theta}\right)\right].$$

One reasonable criterion function when interest is in estimating the model parameters with high precision is $\Psi\left(\xi, \boldsymbol{\beta}\right) = \ln\left|\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})\right|$, the optimal design is then found by minimizing $\ln\left|\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})\right|$. The square root of $\left|\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})\right|$ is proportional to the volume of the confidence region for the parameters which is thus minimized. Such a design is called D-optimal. If interest instead is in estimating the optimum point x_m a reasonable choice of criterion function is $\Psi\left(\xi, \boldsymbol{\beta}\right) = \mathbf{c}^T \mathbf{M}^{-1}(\xi, \boldsymbol{\beta}) \mathbf{c}$, with $\mathbf{c} = \left(\frac{\partial x_m}{\partial \boldsymbol{\beta}}\right)$, which is the asymptotic variance of $\hat{x}_m$. Minimizing this criterion function results in a c-optimal design.

The General Equivalence Theorem, see Kiefer and Wolfowitz (1959) and Kiefer (1961), can be used to check optimality of a suggested design. If the design is optimal it is known from the theorem that the maximum of $d(x, \xi)$ should be equal to the number of parameters in the model. The maxima will also appear at the design points. Furthermore, it is known that there exists a D-optimal design with $p \leq n \leq \frac{p(p+1)}{2}$ design points.

For the logistic model with linear predictor $\eta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ the standardized information matrix for a particular design is given by the 3×3 matrix

$$\begin{aligned} \mathbf{M}(\xi, \boldsymbol{\beta}) &= \sum_{i=1}^{n} w_i \mathbf{m}(\boldsymbol{\beta}, x_i) = \sum_{i=1}^{n} w_i v(x_i) \mathbf{x}_i \mathbf{x}_i^T = \\ & \sum_{i=1}^{n} w_i \pi(x_i)(1 - \pi(x_i)) \begin{bmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 & x_i^3 \\ x_i^2 & x_i^3 & x_i^4 \end{bmatrix} \end{aligned}$$

The D-optimal design is found by minimizing $\ln |\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})|$, or equivalently maximizing $\ln |\mathbf{M}(\xi, \boldsymbol{\beta})|$, which can be accomplished with numerical methods. To start with the number of design points n is not known. Assume for example p points, minimize $|\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})|$ and plot the standardized predictor variance, $d(x, \xi)$. This plot will show whether the design is optimal or not. For the non-optimal case it can give a hint of the optimal number of design points by looking at the number of peaks of the function $d(x, \xi)$.

The D-optimal design depends on the true parameter vector $\boldsymbol{\beta}$ for the logistic model. Because of the symmetry property that $\pi(x + x_m) = \pi(-x + x_m)$ there is a possibility that the optimal design is also symmetric. Therefore to start with the assumption that the D-optimal design consists of $p = 3$ points, where one point is x_m and the other two are symmetric around this point, is made. For all of the four parameter sets presented in the previous section $\beta_1 = 0$ so that $x_m = -\beta_1/(2\beta_2) = 0$. It is also assumed that the design weights are equal to 1/3. This design, denoted as ξ_3, is given by

$$\xi_3 = \left\{ \begin{array}{ccc} -x + x_m & x_m & x + x_m \\ 1/3 & 1/3 & 1/3 \end{array} \right\} = \left\{ \begin{array}{ccc} -x & 0 & x \\ 1/3 & 1/3 & 1/3 \end{array} \right\}.$$

The standardized information matrix will now be

$$\mathbf{M}(\xi_3, \boldsymbol{\beta}) = \frac{1}{3} \left(2v(x) \begin{bmatrix} 1 & 0 & x^2 \\ 0 & x^2 & 0 \\ x^2 & 0 & x^4 \end{bmatrix} + v(0) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)$$

where

$$\begin{aligned} v(-x) &= v(x) = \frac{\exp\{\beta_0 + \beta_1 x + \beta_2 x^2\}}{(1 + \exp\{\beta_0 + \beta_1 x + \beta_2 x^2\})^2} \\ v(0) &= \frac{\exp\{\beta_0\}}{(1 + \exp\{\beta_0\})^2}. \end{aligned}$$

The next step is to find the value of x that maximizes $|\mathbf{M}(\xi_3, \boldsymbol{\beta})|$, which here was done using Mathcad 11.0. The resulting 3-point designs for the four sets of true parameters are given in Table 2.

Table 2: Suggested 3-point designs.

Type of response curve	**True parameters**	**Design**
"High-wide"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$	$\xi_3 = \begin{Bmatrix} -5.5398 & 0 & 5.5398 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$
"High-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$	$\xi_3 = \begin{Bmatrix} -0.8759 & 0 & 0.8759 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$
"Low-wide"	$\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -0.1 \end{pmatrix}^T$	$\xi_3 = \begin{Bmatrix} -3.9819 & 0 & 3.9819 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$
"Low-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -4 \end{pmatrix}^T$	$\xi_3 = \begin{Bmatrix} -0.6296 & 0 & 0.6296 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$

It turns out that these suggested designs are D-optimal when the curve describing $\pi(x)$ is low, i.e. for the two parameter sets named "low-wide" and "low-narrow". This can be seen from the plots of the standardized predictor variance, $d(x, \xi_3)$ given in Figure 2. $d(x, \xi_3) \leq 3$ and the maxima are attained at the design points for these two models, which is in line with the General Equivalence Theorem. However, $d(x, \xi_3) > 3$ when $\pi(x)$ is high. As can be seen from the plots in Figure 2 it seems like the D-optimal designs in these cases consist of 4 symmetric points. Another design with 4 points that is symmetric around x_m is given by ξ_4, $x_m = 0$ because $\beta_1 = 0$ in the current special cases.

$$\begin{aligned} \xi_4 &= \begin{Bmatrix} -x_2 + x_m & -x_1 + x_m & x_1 + x_m & x_2 + x_m \\ w_2/2 & w_1/2 & w_1/2 & w_2/2 \end{Bmatrix} \\ &= \begin{Bmatrix} -x_2 & -x_1 & x_1 & x_2 \\ w_2/2 & w_1/2 & w_1/2 & w_2/2 \end{Bmatrix} \end{aligned}$$

The standardized information matrix for ξ_4 is given by

$$\mathbf{M}(\xi_4, \boldsymbol{\beta}) = w_1 v(x_1) \begin{bmatrix} 1 & 0 & x_1^2 \\ 0 & x_1^2 & 0 \\ x_1^2 & 0 & x_1^4 \end{bmatrix} + w_2 v(x_2) \begin{bmatrix} 1 & 0 & x_2^2 \\ 0 & x_2^2 & 0 \\ x_2^2 & 0 & x_2^4 \end{bmatrix}.$$

Now the problem is to find the values of x_1, x_2, w_1 and w_2 that makes $|\mathbf{M}(\xi_4, \boldsymbol{\beta})|$ take on its largest value. Doing so results in the two designs shown in Table 3 with standardized predictor variance $d(x, \xi_4)$ as in Figure 3. An examination of these plots shows that $d(\mathbf{x}, \xi_4) \leq p = 3$ in both cases and that the maxima are attained at the design points. Hence, these designs are D-optimal.

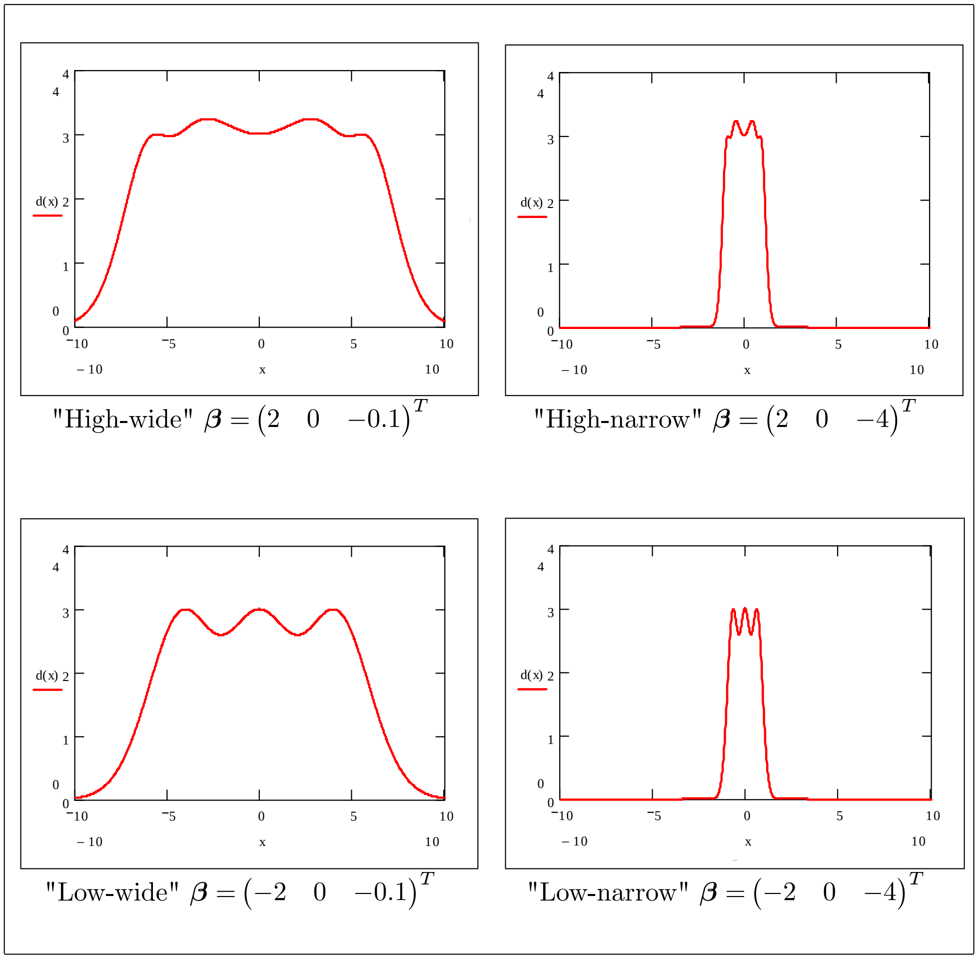

Figure 2: The standardized predictor variance $d(x, \xi_3)$ for the four designs given in Table 2.

Table 3: Suggested 4-point designs.

Type of response curve	**True parameters**	**Design**
"High-wide"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$	$\xi_4 = \begin{Bmatrix} -5.7185 & -2.7017 & 2.7017 & 5.7185 \\ 0.3138 & 0.1862 & 0.1862 & 0.3138 \end{Bmatrix}$
"High-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$	$\xi_4 = \begin{Bmatrix} -0.9042 & -0.4272 & 0.4272 & 0.9042 \\ 0.3138 & 0.1862 & 0.1862 & 0.3138 \end{Bmatrix}$

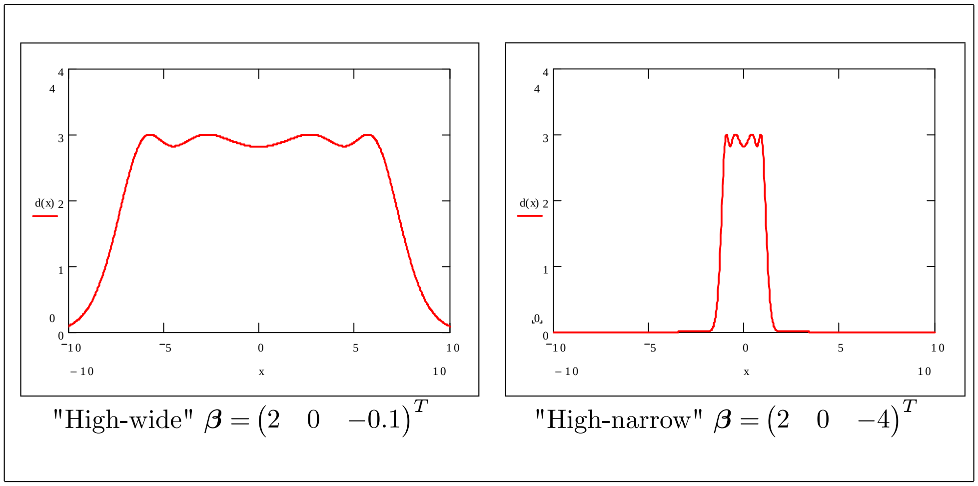

Figure 3: The standardized predictor variance $d(x, \xi_4)$ for the two designs given in Table 4.

For the parameter sets where $\pi(x)$ is low the D-optimal designs consist of 3 points and for the parameter sets where $\pi(x)$ is high the D-optimal designs consist of 4 points indicating that more points are needed to estimate a higher curve. For the two cases when the shape of the curve $\pi(x)$ is narrow it can be noted that the design points are closer to the optimum point $x_m = 0$. Furthermore, regardless of the number of design points the optimal designs are symmetric around the optimum point in all four cases.

When deriving c-optimal designs a problem sometimes arises with a singular information matrix. A way to circumvent this is to add a small number ε to the diagonal elements of $\mathbf{M}(\xi, \boldsymbol{\beta})$ before inversion, see Section 10.3 in Atkinson and Donev (1992). Using Mathcad 11.0 $\mathbf{c}^T\mathbf{M}^{-1}(\xi, \boldsymbol{\beta})\,\mathbf{c}$ was minimized and the resulting c-optimal designs consist of 2 points that are symmetric around the optimum point for all four sets of true parameters, see Table 4.

Table 4: c-optimal designs.

Type of response curve	**True parameters**	**Design**
"High-wide"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$	$\left\{\begin{matrix} -5.2529 & 5.2529 \\ 1/2 & 1/2 \end{matrix}\right\}$
"High-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$	$\left\{\begin{matrix} -0.8306 & 0.8306 \\ 1/2 & 1/2 \end{matrix}\right\}$
"Low-wide"	$\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -0.1 \end{pmatrix}^T$	$\left\{\begin{matrix} -3.3089 & 3.3089 \\ 1/2 & 1/2 \end{matrix}\right\}$
"Low-narrow"	$\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -4 \end{pmatrix}^T$	$\left\{\begin{matrix} -0.5232 & 0.5232 \\ 1/2 & 1/2 \end{matrix}\right\}$

4 Sampling distribution of the maximum likelihood estimator

The D-optimal designs shown in the previous section concern the asymptotic sampling distribution of the parameter estimators. In practice the sample sizes are often small due to time and money constraints raising the question how these designs will work in a small sample setting. Furthermore, the optimal designs depend on the true parameters which are unknown. The properties of the MLE when the D-optimal designs are used are examined for different sample sizes and compared to the properties of the MLE when some non-optimal designs are used.

For each of the four sets of true parameters presented in section 2 three designs are considered. One is the D-optimal design which depends on the parameters and thus is different in all four cases. In practice the true parameters are not known and in most cases more than the optimal number of points are taken to hopefully get some good points. Sometimes there is an understanding of where $P(Y = 1)$ is appreciably greater than zero and less than one. If so, an interval can be specified where it is believed that both $Y = 1$ and $Y = 0$ may be observed. This interval may then be used to determine where the design points shall be. One 7-point design (ξ_7) and one 8-point design (ξ_8) are used here, the design with 8 points is symmetric around the optimum point whereas the 7-point design is not. All designs are given in Table 5. The c-optimal designs are derived to be optimal when it comes to estimating the optimum point x_m. However, it is not possible to estimate the model parameters with a c-optimal design because two points are not sufficient to estimate three parameters. This is the reason for not considering the c-optimal designs here when the small sample distribution of the MLE is studied.

Table 5: D-optimal designs, one 7-point design and one 8-point design.

D-optimal	"High-wide" $\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -0.1 \end{pmatrix}^T$	$\xi^* = \begin{Bmatrix} -5.7185 & -2.7017 & 2.7017 & 5.7185 \\ 0.3138 & 0.1862 & 0.1862 & 0.3138 \end{Bmatrix}$
D-optimal	"High-narrow" $\boldsymbol{\beta} = \begin{pmatrix} 2 & 0 & -4 \end{pmatrix}^T$	$\xi^* = \begin{Bmatrix} -0.9042 & -0.4272 & 0.4272 & 0.9042 \\ 0.3138 & 0.1862 & 0.1862 & 0.3138 \end{Bmatrix}$
D-optimal	"Low-wide" $\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -0.1 \end{pmatrix}^T$	$\xi^* = \begin{Bmatrix} -3.9819 & 0 & 3.9819 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$
D-optimal	"Low-narrow" $\boldsymbol{\beta} = \begin{pmatrix} -2 & 0 & -4 \end{pmatrix}^T$	$\xi^* = \begin{Bmatrix} -0.6296 & 0 & 0.6296 \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}$
7-point	$\xi_7 = \begin{Bmatrix} -4 & -1 & 0.5 & 1 & 1.5 & 3 & 6 \\ 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 & 1/7 \end{Bmatrix}$	
8-point	$\xi_8 = \begin{Bmatrix} -5 & -2 & -0.75 & -0.25 & 0.25 & 0.75 & 2 & 5 \\ 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 \end{Bmatrix}$	

For every combination of true parameter values and design four sample sizes are examined, two small samples $N = 10$ and $N = 20$ and two larger, $N = 50$ and $N = 100$. The exact sampling distribution of the MLE is obtained for the small sample sizes by generating all possible samples, the parameter estimate for each sample is then weighted with the probability of obtaining the current sample. For the larger sample sizes simulations are performed instead because the number of possible samples grows very large. Given a set of true parameter values and design, response values are generated. For every x the probability of a "success", $P(Y_j = 1) = \pi_j$, is calculated. Uniform random numbers $U_j,\ j = 1, ..., N$ are generated and if $U_j < \pi_j$ 1 is assigned to the response variable and 0 otherwise. These response values are, together with the values of the control variable given by the design, used to estimate the parameters.

The proportions of the sample to be allocated to the design points are given by the design weights. The number of observations to be taken at each design point are thus given by $n_i = w_i N$. However, adjustments has to be made to n_i because the number of observations need to be integer values. The resulting designs will then be approximations to the designs given in Table 5. In Table 6 the number of observations per design point are shown.

Table 6: Number of observations taken at each design point.

N	**3-point design**	**4-point design**	**7-point design**	**8-point design**
10	3/4/3	3/2/2/3	1/2/1/2/1/2/1	2/1/1/1/1/1/1/2
20	7/6/7	6/4/4/6	3/3/3/2/3/3/3	2/3/2/3/2/3/2/3
50	17/16/17	16/9/9/16	7/7/7/8/7/7/7	7/6/6/6/6/6/6/7
100	33/34/33	31/19/19/31	14/14/15/14/15/14/14	12/13/12/13/12/13/12/13

4.1 Non-existence of the MLE

For certain data the maximum likelihood estimation procedure does not converge and there exist no MLEs. Depending on the pattern of the data points a data set can be categorized as belonging to one of three types of data configurations; complete separation, quasi-complete separation or overlap, as described in Albert and Anderson (1984). It is only when data belong to the third configuration that the MLE is finite and unique. The responses are binary and the data can thus be divided into two response groups, one including the points where $Y_i = 1$ and one including the points where $Y_i = 0$. If there is a vector that correctly allocates all observations to their respective response group complete separation is present, that is, if there is a vector $\boldsymbol{\alpha}$ so that $\boldsymbol{\alpha}^T \mathbf{x}_i > 0$ for all $Y_i = 1$ and $\boldsymbol{\alpha}^T \mathbf{x}_i < 0$ for all $Y_i = 0$. Quasi-complete separation occurs when there is a vector $\boldsymbol{\alpha}$ such that $\boldsymbol{\alpha}^T \mathbf{x}_i \geq 0$ for all $Y_i = 1$ and $\boldsymbol{\alpha}^T \mathbf{x}_i \leq 0$ for all $Y_i = 0$. If the data configuration is neither complete separation nor quasi-complete separation the data points are overlapped and the

MLE exists and is unique. Examples on the three data configurations are shown in Figure 4.

The problem of separation occurs for the logistic model considered here, to what extent depends on the sample size, the number of design points and the true parameters. When the sample size is $N = 10$ divided among three points the number of possible samples is equal to

$$(n_1 + 1) \cdot (n_2 + 1) \cdot (n_3 + 1) = 4 \cdot 5 \cdot 4 = 80.$$

The only way for the data to be overlapped and thereby for the MLE to exist is to have both response types ($Y = 1$ and $Y = 0$) observed at all three points. In a design point where three observations are taken there are two variants where both response types are represented: 1 or 2 ones out of 3. When four observations are taken at a design point there are three such cases: 1, 2 or 3 ones out of 4. In total there are $2 \cdot 3 \cdot 2 = 12$ distinct samples where the MLE exists. Let Z_i be the number of ones observed at design point x_i, that is $Z_i \sim bin\,(n_i, \pi_i)$ where $\pi_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$. The probability that the MLE exists can then be computed as

$$P\,(MLE) = [P\,(Z_1 = 1) + P\,(Z_1 = 2)] \cdot [P\,(Z_2 = 1) + P\,(Z_2 = 2) + P\,(Z_2 = 3)] \cdot [P\,(Z_3 = 1) + P\,(Z_3 = 2)]\,.$$

This probability is shown for all combinations of design and true parameters for $N = 10$ in section a) of Table 7. The probability to obtain a MLE is essentially zero for the "low-narrow" model irrespective of what design is used. It is not possible to estimate the parameters with only 10 observations in this case. For the "low-wide" model this probability is at best around 10 percent when one of the two non-optimal designs are used. The reason why the non-optimal designs perform better in this respect is that the D-optimal design consists of only 3 points which makes it more difficult to obtain overlapped data compared to the designs with more points. This is particularly the case for the wide models because for these models $P(Y = 1)$ is appreciably greater than zero for a wider range of x values than for the narrow models. As a consequence there are more points to choose from for which there are a possibility of observing both zeros and ones and thus avoiding complete or quasi-complete separation in the data. This is also reflected in the pattern for the "high-wide" model where the probability of obtaining a MLE is approximately 50 % for the non-optimal designs compared to 35 % for the D-optimal design. For the "high-narrow" model on the other hand the D-optimal design outperforms the other two having 35 % chance of obtaining parameter estimates as against 11 % (the 8-point design) and 1 % (the 7-point design). The low models are more problematic compared to the high models because $P(Y = 1)$ is low and therefore there will be many points where only zeros are observed.

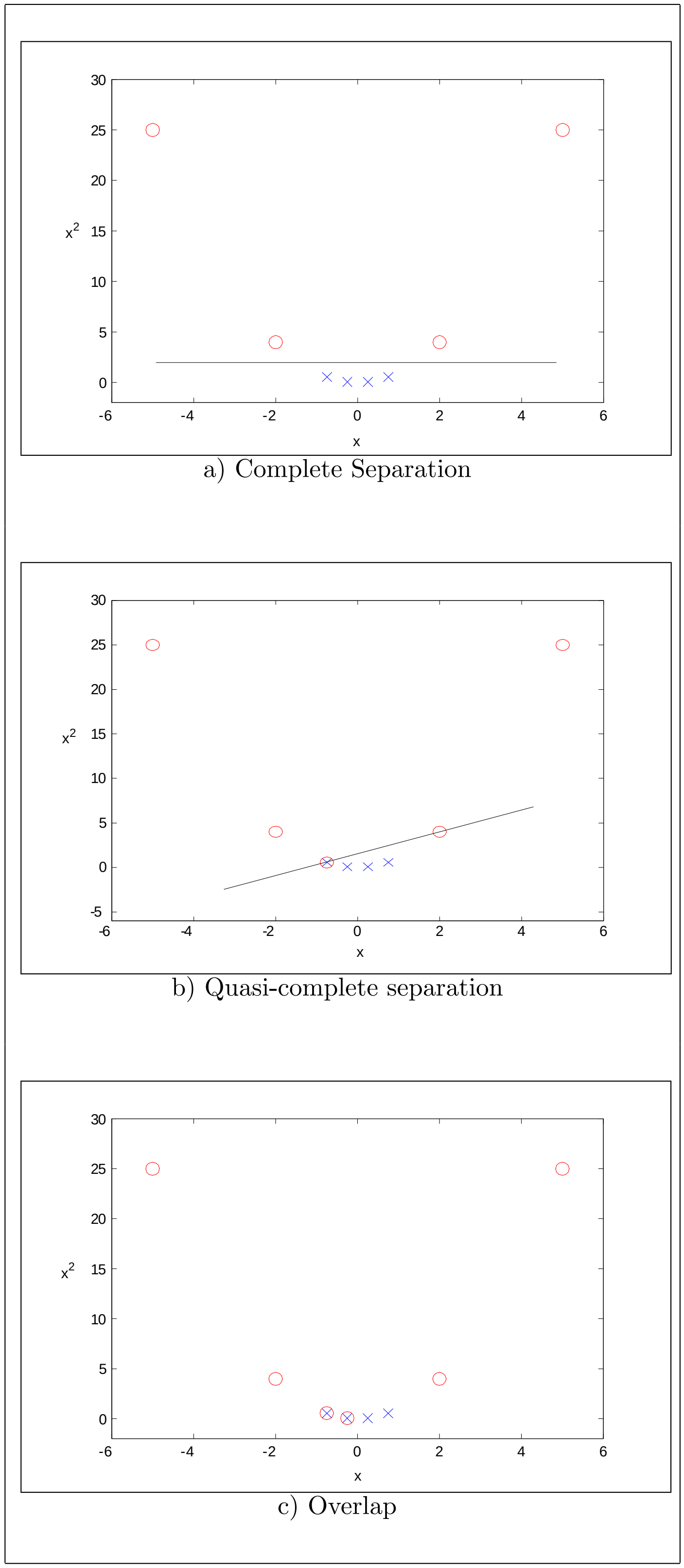

Figure 4: Examples on data configurations for the 8-point design, if there are only zeros observed at a point a ring is displayed, if there are only ones observed a cross is displayed and if both response types are present a cross in a ring is displayed.

Table 7: The number of possible samples, the number of samples where the MLE exists and the probability that the MLE exists for N=10 and N=20.

Type of response curve	Design	# samples	# samples where the MLE exists	P(MLE exists)
		a) N = 10		
	D-opt	144	68	0.35
"High-wide"	7-p	648	544	0.53
	8-p	576	488	0.49
	D-opt	144	68	0.35
"High-narrow"	7-p	648	544	0.014
	8-p	576	488	0.11
	D-opt	80	12	$2.5 \cdot 10^{-3}$
"Low-wide"	7-p	648	544	0.11
	8-p	576	488	0.096
	D-opt	80	12	$2.5 \cdot 10^{-3}$
"Low-narrow"	7-p	576	488	$2.4 \cdot 10^{-5}$
	8-p	192	156	$2.6 \cdot 10^{-3}$
		b) N = 20		
	D-opt	1225	927	0.74
"High-wide"	7-p	12288	11944	0.82
	8-p	20736	20384	0.84
	D-opt	1225	927	0.76
"High-narrow"	7-p	12288	11944	0.046
	8-p	20736	20384	0.43
	D-opt	448	180	0.016
"Low-wide"	7-p	12228	11944	0.28
	8-p	20736	20384	0.36
	D-opt	448	180	0.016
"Low-narrow"	7-p	12228	11944	$4.4 \cdot 10^{-5}$
	8-p	20736	20384	0.016

In section b) of Table 7 the probabilities that the MLE exist when $N = 20$ are presented. The probability to obtain parameter estimates has increased but is still extremely low for the "low-narrow" model. For the "low-wide" model these probabilities have increased about three times for the two non-optimal designs which still are preferable. The D-optimal design has the highest probability for the "high-narrow" model and the three designs are almost equivalent for the "high-wide" model.

For $N = 50$ the probability of obtaining parameter estimates can be estimated by the percentage share of the simulations where the MLE existed. These are shown in column a) of Table 8. There are hardly ever problems with non-existence for the "high-wide" model with 99 % existence for all three designs. The same thing applies to the "high-narrow" model if the D-optimal design is used. The situation is not quite as good for the "low-wide" model where the 80 % associated with the 8-point design is the highest proportion. Nevertheless it is a lot better than for the "low-narrow" case where existence in 11 % of the times (for the D-optimal design) is the maximum.

Table 8: The percentage share of the simulations where the MLE existed for N=50 and N=100.

Type of response curve	**Design**	**% of the simulation samples where the MLE existed**	
		a) N = 50	**b) N = 100**
	D-opt	99	100
"High-wide"	7-p	99	100
	8-p	99	100
	D-opt	99	100
"High-narrow"	7-p	34	69
	8-p	84	98
	D-opt	12	35
"Low-wide"	7-p	73	97
	8-p	80	98
	D-opt	11	35
"Low-narrow"	7-p	0.04	0.2
	8-p	8	21

For $N = 100$ the percentage shares of the simulations where the MLE existed are given in column b) of Table 8. There are no longer any problems with non-existence for the "high-wide" model for any of the designs. If the D-optimal or 8-point design is chosen for the "high-narrow" model the problem is also avoided. By choosing one of the non-optimal designs for the "low-wide" model almost 100 % existence of the MLE can be expected. When the D-optimal design was used for the "low-narrow" model existence occurred in only 35 % of the times even though the sample size is quite large.

4.2 Results

When the sample sizes are small, i.e. when $N = 10$ and $N = 20$, all possible samples are generated and parameter estimates are calculated when possible. The mean squared error, mean and variance given that the MLE exists are displayed in Table 9. Furthermore, an estimate of the variance of the MLE, given by the diagonal elements in $\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right)$ found in section 2, is calculated. The mean and variance of this estimator, given that the MLE exists, are also presented in the table. In the column $\mathbf{AV}\left(\widehat{\boldsymbol{\beta}}\right)$ the theoretical approximate variance derived using asymptotic results, obtained from the diagonal in $\mathbf{I}^{-1}(\xi, \boldsymbol{\beta})$, can be found.

The results for the smallest sample size $N = 10$ are summarized as follows. The mean squared error given that the MLE exists is not consistently lower for the D-optimal design compared to the two non-optimal designs, e.g. for the "high-narrow" model. In order to maximize the probability of obtaining a MLE for the "high-wide" model the 7-point design is preferable. The downside is that the mean squared error given that the MLE exists is larger than for the D-optimal design. For the "high-narrow" model the D-optimal design that has the highest probability of existence does not have the lowest mean squared error for two of the parameters. The mean squared error given that the MLE exists is smallest for the D-optimal design and the "low-wide" model, but the probability of existence is practically zero. Of the other two designs the best choice is the 7-point design, however, the probability of existence is only 11 % so this is still no good choice. The results for the "low-narrow" model are less important because the probability that the MLE exists is virtually zero. It can be noted that the bias (given that the MLE exists) of $\widehat{\beta}_1$ is small whenever a symmetric design (the D-optimal or the 8-point design) is used for all models. Apart from these cases, the bias is large. The variance of the MLE (given that the MLE exists) is always less than the theoretical approximate variance when the D-optimal design is used, that is the asymptotic results are not applicable for such small samples. For the non-optimal designs the variance given that the MLE exists is sometimes larger and sometimes smaller depending on the true parameters. The mean of the variance estimator (given that the MLE exists) is almost always overestimating the variance.

When $N = 20$ the probability of existence has increased although it is still very low for the "low-narrow" model and quite low for the "low-wide" model. The mean squared error given that the MLE exists has decreased. There is no design that consistently has a lower mean squared error than the others even though the D-optimal design more often is favored. A higher probability of obtaining an estimate can sometimes be traded for a higher mean squared error, like for the "low-wide" model. The bias (given that the MLE exists) is quite large throughout, with the exception of estimation of β_2 for the "high-wide" model regardless of which design is used, and of β_1 when the D-optimal design is used. The bias has increased for $\widehat{\beta}_1$ and the 8-point design compared to when $N = 10$. There is no clear-cut pattern as to when the variance given that the MLE exists is larger than or less than the theoretical approximate variance. The mean of the variance estimator given that the MLE exists is with one exception an overestimation of the variance. In addition, the variance of this variance estimator (given that the MLE exists) is very large for some cases for the non-optimal designs, especially for the 7-point design and the "low-wide" model.

Simulations are performed for the larger sample sizes $N = 50$ and $N = 100$. MLEs of the parameters are obtained for each of the true parameter-design-sample size combinations which is repeated 5000 times. The simulation results concerning parameter estimation when $N = 50$ and $N = 100$ are summarized in Table 10. The mean squared error for the simulation samples where the maximum likelihood estimation procedure converged is computed and shown in the second column. The average and sample variance of the MLE in the simulation samples where the MLE existed are given in the third and fourth columns. The next column contains the theoretical approximate variance derived using asymptotic results, abbreviated as $\mathbf{AV}\left(\widehat{\boldsymbol{\beta}}\right)$. An estimate of the variance of the MLE, given by the diagonal elements in $\widehat{\mathbf{V}}\left(\widehat{\boldsymbol{\beta}}\right)$, is calculated for each of the simulation samples where the MLE existed. The average and sample variance of these variance estimates are found in the last two columns.

When $N = 50$ almost all of the 5000 simulations resulted in overlapped data with an existing MLE for the "high-wide" model no matter what design was used and for the "high-narrow" model together with the D-optimal design. 80 % existence was the best that was achieved for the "low-wide" model (the 8-point design) which might be tolerable contrary to 11 % existence that was the maximum for the "low-narrow" model (the D-optimal design). The pattern of the simulation mean squared error based on the times where the MLE existed is similar to the pattern of the mean squared error (given that the MLE exists) observed when $N = 20$. That is, although none of the designs entirely outperforms the others, the D-optimal design is more often associated with the lowest mean squared error. The bias is still quite large, in most cases it has decreased but there are examples where the opposite is true, e.g. $\widehat{\beta}_1$ and the D-optimal design. The simulation sample variance of the MLE exceeds but is quite close to the theoretical approximate variance when the percentage share of existence was as high as 99 %. The variance estimator also worked well when the MLE existed in 99 % of the times, it slightly underestimated the simulation sample variance in those cases. For almost all the rest it was an overestimate of the sample variance and the higher proportion of existence the closer it came. The simulation sample variance of this variance estimator was particularly large for the non-optimal designs and both the low models.

When the sample size is $N = 100$ the D-optimal design was the design with the highest proportion of existence in three out of four cases. For the "low-wide" model on the other hand, both the non-optimal designs managed to estimate the parameters in almost 100 % of the times compared to only 35 % for the D-optimal design. The mean squared error for the simulation samples where the MLE existed is lowest for the D-optimal design besides for $\widehat{\beta}_0$ in both the wide models and it is lowest for all parameters in the "high-narrow" model. The 8-point design is instead associated with the lowest mean squared error for two parameters in the "low-narrow" model. The MLE is still biased (with the exception of $\widehat{\beta}_2$ for the "high-wide" model and the D-optimal or 8-point design) although it has decreased. The bias is generally smallest for $\widehat{\beta}_1$ when the D-optimal or 8-point design is used. For all three designs and the "high-wide" model and for the two symmetric designs and the "high-narrow" model the simulation sample variance agrees well with the theoretical approximate variance. Furthermore, the variance estimator comes close to the simulation sample variance. In spite of the fact that the proportion where the MLE existed was approximately 100 % for the two non-optimal designs and the "low-wide" case, the simulation sample variance exceeds the theoretical approximate variance with a fairly large amount. However, the variance estimator succeeds quite well in these two cases.

The mean and variance of the optimum point, $\hat{x}_m$, given that the MLE exists are computed for $N = 10$ and $N = 20$ and the average and simulation sample variance of $\hat{x}_m$ are computed for $N = 50$ and $N = 100$. Occasionally the estimate of β_2 comes close to zero which causes $\hat{x}_m$ and with that the mean or simulation average of $\hat{x}_m$ to degenerate. When $\hat{x}_m$ falls outside $\hat{x}_m \pm 10 \cdot \sqrt{AV(\widehat{x}_m)}$, where $AV(\widehat{x}_m)$ is the approximate theoretical variance, that estimate is discarded. The probability that $\hat{x}_m$ is not degenerated according to this definition is displayed in the second column of Table 11. For the two smallest sample sizes this probability is calculated by summing the probabilities to obtain each of the samples where $\hat{x}_m$ is not degenerated. For the two largest sample sizes this probability is estimated by the percentage share of the samples (where the MLE existed) where $\hat{x}_m$ was not degenerated. The mean and variance of $\hat{x}_m$ given that the MLE exists and that $\hat{x}_m$ is not degenerated is found in the third and fourth columns of Table 11. The fifth column contains the theoretical approximate variance of $\hat{x}_m$ derived using asymptotic results. The variance estimator $\widehat{V}(\widehat{x}_m)$ is computed according to the equations given in section 2 and its mean and variance are given in last two columns. The corresponding simulation results given that the MLE existed and that $\hat{x}_m$ was not degenerated are presented in Table 12 for $N = 50$ and $N = 100$.

The probability of obtaining degenerated estimates is higher for the wide models because β_2 is closer to zero, as can be seen from Table 11 for $N = 10$. The differences between the designs are small in this respect. The differences when it comes to bias and variance of $\hat{x}_m$ given that the MLE exists and that $\hat{x}_m$ is not degenerated was on the other hand large. The D-optimal design had the smallest bias (for all but the "low-wide" model) and variance irrespective of the true parameters. The 8-point design comes in a good second place when grading is made by size of bias, it is not far away from the D-optimal design. A symmetric design is as might be expected better at estimating the optimum point. The variance of $\hat{x}_m$ was smallest for the D-optimal design and in three cases smaller for the 8-point than for the 7-point design. The variance exceeded the theoretical approximate variance for the high models and vice versa for the low models. The variance estimator does not work well at all. For all models it overestimates the variance of $\hat{x}_m$ given that the MLE exists and that $\hat{x}_m$ is not degenerated and it only came fairly close for the 8-point design and the narrow models. It needs to be pointed out that the probability that the MLE exists is extremely low for many cases making some of these results not that meaningful.

The probability that $\hat{x}_m$ is not degenerated has increased somewhat when the sample size is increased from $N = 10$ to $N = 20$. The D-optimal design is now associated with the lowest bias and variance given that the MLE exists and that $\hat{x}_m$ is not degenerated in all cases. The 8-point design is still in a second place but has now fallen behind. The bias has also increased for the D-optimal design and the high models. The variance of $\hat{x}_m$ given that the MLE exists and that $\hat{x}_m$ is not degenerated was higher than the theoretical approximate variance for the two high models and lower for the two low models. The variance estimator is still no good for any of the models, although it is somewhat closer for the 7-point design together with the narrow models. These patterns remain when $N = 50$ apart from that the simulation sample variance of $\hat{x}_m$ now is closer to the theoretical approximate variance, especially for the D-optimal design, and that the variance estimator has improved for the "high-wide" model. For the largest sample size the same grading prevails, i.e. that the D-optimal design was better than the non-optimal designs. The variance of $\hat{x}_m$ is quite close to the theoretical approximate variance for the high models. The variance estimator also comes close to the simulation sample variance for these models.

Table 9: Results of maximum likelihood estimation of the parameters in the quadratic logit model for N=10 and N=20.

N	Design $(P(MLE))$	MSE	$E(\hat{\boldsymbol{\beta}})$	$V(\hat{\boldsymbol{\beta}})$	$AV(\hat{\boldsymbol{\beta}})$	$E(\widehat{\mathbf{V}}(\hat{\boldsymbol{\beta}}))$	$V(\widehat{\mathbf{V}}(\hat{\boldsymbol{\beta}}))$
	a) True response curve: "high-wide" model $\beta = (2 \quad 0 \quad -0.1)^T$						
N = 10	D (0.35)	1.53 0.018 $2.66 \cdot 10^{-3}$	1.10 $-2.34 \cdot 10^{-18}$ −0.064	0.72 0.018 $1.39 \cdot 10^{-3}$	2.67 0.025 $3.87 \cdot 10^{-3}$	2.19 0.03 $3.81 \cdot 10^{-3}$	0.063 $1.10 \cdot 10^{-4}$ $5.20 \cdot 10^{-7}$
	7 − P (0.53)	0.63 0.42 0.017	1.82 −0.16 −0.090	0.60 0.40 0.017	1.28 0.074 $5.70 \cdot 10^{-3}$	1.60 0.57 0.03	1.41 1.07 $5.38 \cdot 10^{-3}$
	8 − P (0.49)	0.71 0.25 0.019	1.41 $-4.16 \cdot 10^{-17}$ −0.093	0.36 0.25 0.019	1.31 0.062 $5.16 \cdot 10^{-3}$	1.30 0.31 0.028	0.094 0.40 0.022
N = 20	D (0.74)	0.79 0.016 $1.43 \cdot 10^{-3}$	1.89 $-1.26 \cdot 10^{-17}$ −0.10	0.78 0.016 $1.43 \cdot 10^{-3}$	1.34 0.013 $1.94 \cdot 10^{-3}$	1.40 0.019 $2.45 \cdot 10^{-3}$	0.19 $1.16 \cdot 10^{-4}$ $8.13 \cdot 10^{-7}$
	7 − P (0.82)	0.60 0.17 $7.82 \cdot 10^{-3}$	2.14 −0.097 −0.11	0.58 0.16 $7.74 \cdot 10^{-3}$	0.64 0.037 $2.85 \cdot 10^{-3}$	1.02 0.17 $9.64 \cdot 10^{-3}$	0.73 0.46 0.013
	8 − P (0.84)	0.47 0.16 0.014	2.05 −0.019 −0.11	0.47 0.16 0.014	0.65 0.031 $2.58 \cdot 10^{-3}$	0.82 0.14 0.015	0.16 0.14 0.038
	b) True response curve: "high-narrow" model $\beta = (2 \quad 0 \quad -4)^T$						
N = 10	D (0.35)	1.53 0.73 4.26	1.10 $-4.24 \cdot 10^{-17}$ −2.58	0.72 0.73 2.23	2.68 1.01 6.20	2.19 1.21 6.09	0.063 0.18 1.33
	7 − P (0.014)	4.20 0.51 12.20	$-9.79 \cdot 10^{-4}$ −0.38 −0.51	0.19 0.37 0.023	7.87 3.26 13.00	1.68 1.07 0.65	0.35 1.83 0.041
	8 − P (0.11)	1.53 1.66 0.14	0.86 $-3.31 \cdot 10^{-16}$ −1.56	0.23 1.66 0.14	4.04 2.73 19.18	1.65 3.71 2.72	0.011 1.89 1.78
N = 20	D (0.76)	0.79 0.64 2.29	1.89 $-3.89 \cdot 10^{-16}$ −4.01	0.78 0.64 2.29	1.34 0.50 3.10	1.40 0.75 3.91	0.19 0.19 2.08
	7 − P (0.28)	1.46 0.33 5.28	1.08 0.012 −1.77	0.61 0.33 0.30	3.93 1.63 6.49	1.97 1.13 1.97	1.35 40.83 2.43
	8 − P (0.36)	0.68 1.63 2.88	1.43 −0.14 −2.94	0.35 1.61 1.76	2.02 1.37 9.59	1.26 1.99 6.26	0.13 1.02 19.89
	c) True response curve: "low-wide" model $\boldsymbol{\beta} = (-2 \quad 0 \quad -0.1)^T$						
N = 10	D $(2.5 \cdot 10^{-3})$	1.47 $1.59 \cdot 10^{-3}$ 0.014	−0.88 $2.11 \cdot 10^{-17}$ 0.014	0.23 $1.59 \cdot 10^{-3}$ $1.00 \cdot 10^{-3}$	2.86 0.36 0.034	1.28 0.047 $8.06 \cdot 10^{-3}$	0.015 $4.34 \cdot 10^{-19}$ $2.44 \cdot 10^{-7}$
	7 − P (0.11)	2.34 1.36 0.15	−0.86 0.36 −0.35	1.05 1.22 0.088	2.21 0.55 0.060	2.35 2.23 0.35	5.58 11.96 0.14
	8 − P (0.096)	3.10 2.01 0.69	−0.43 $-5.65 \cdot 10^{-16}$ −0.63	2.63 2.01 0.41	1.66 0.63 0.060	1.40 2.31 0.86	0.21 3.96 2.08
N = 20	D (0.016)	0.73 $1.95 \cdot 10^{-3}$ $6.60 \cdot 10^{-3}$	−1.31 $3.36 \cdot 10^{-17}$ −0.026	0.25 $1.95 \cdot 10^{-3}$ $1.12 \cdot 10^{-3}$	1.43 0.18 0.017	1.07 0.036 $6.49 \cdot 10^{-3}$	0.044 $8.19 \cdot 10^{-6}$ $7.31 \cdot 10^{-4}$
	7 − P (0.28)	1.21 2.37 0.45	−1.52 0.61 −0.46	0.98 2.01 0.32	1.11 0.27 0.030	2.61 6.56 1.46	26.52 696.44 41.63
	8 − P (0.43)	0.88 1.66 0.69	−1.36 −0.044 −0.68	0.48 1.66 0.69	0.83 0.31 0.030	0.88 2.04 1.52	0.19 7.66 13.61
	d) True response curve: "low-narrow" model $\boldsymbol{\beta} = (-2 \quad 0 \quad -4)^T$						
N = 10	D $(2.5 \cdot 10^{-3})$	1.47 0.064 22.52	−0.88 $-1.45 \cdot 10^{-16}$ 0.57	0.23 0.064 1.60	2.86 14.4 54.56	1.28 1.89 12.91	0.015 $1.33 \cdot 10^{-15}$ 0.62
	7 − P $(2.4 \cdot 10^{-5})$	3.45 0.34 12.45	−0.15 −0.37 −0.47	0.026 0.20 0.011	103.06 139.39 329.69	1.59 0.96 0.63	0.18 1.12 $7.24 \cdot 10^{-3}$
	8 − P $(2.6 \cdot 10^{-3})$	6.11 2.68 4.93	0.47 $-5.46 \cdot 10^{-16}$ −1.80	0.022 2.68 0.094	6.32 30.32 133.93	1.59 4.69 3.69	$8.87 \cdot 10^{-3}$ 0.91 0.73
N = 20	D (0.016)	0.73 0.078 10.56	−1.31 $-3.19 \cdot 10^{-16}$ −1.04	0.25 0.078 1.79	1.43 7.21 27.28	1.07 1.42 10.38	0.044 0.013 1.87
	7 − P $(4.4 \cdot 10^{-5})$	1.37 0.85 11.71	−0.98 −0.05 −0.59	0.32 0.85 0.10	1.43 7.21 27.28	1.52 2.10 0.70	1.78 30.67 1.62
	8 − P (0.016)	1.54 1.70 5.55	−0.85 0.35 −1.84	0.22 1.58 0.88	3.16 15.16 66.97	0.96 3.52 6.13	0.016 1.71 36.13

Table 10: Simulation results of maximum likelihood estimation of the parameters in the quadratic logit model for N=50 and N=100.

	Design (% MLE)	mse	$\bar{\hat{\beta}}$	$s^2_{\hat{\beta}}$	$AV(\hat{\beta})$	$\widehat{V}(\hat{\beta})$	$s^2_{\widehat{V}(\hat{\beta})}$
a) True response curve: "high-wide" model $\beta = (2 \quad 0 \quad -0.1)^T$							
N = 50	D (99)	0.72 $7.34 \cdot 10^{-3}$ $1.14 \cdot 10^{-3}$	2.17 $2.18 \cdot 10^{-4}$ −0.11	0.69 $7.34 \cdot 10^{-3}$ $1.06 \cdot 10^{-3}$	0.54 $5.03 \cdot 10^{-3}$ $7.75 \cdot 10^{-4}$	0.69 $6.33 \cdot 10^{-3}$ $1.01 \cdot 10^{-3}$	0.12 $1.26 \cdot 10^{-5}$ $1.94 \cdot 10^{-6}$
	7 − P (99)	0.39 0.031 $2.50 \cdot 10^{-3}$	2.16 −0.030 −0.11	0.36 0.030 $2.35 \cdot 10^{-3}$	0.26 0.015 $1.14 \cdot 10^{-3}$	0.35 0.025 $2.03 \cdot 10^{-3}$	0.068 $5.03 \cdot 10^{-3}$ $4.73 \cdot 10^{-6}$
	8 − P (99)	0.38 0.020 $1.73 \cdot 10^{-3}$	2.12 $-1.35 \cdot 10^{-4}$ −0.11	0.36 0.020 $1.65 \cdot 10^{-3}$	0.26 0.012 $1.03 \cdot 10^{-3}$	0.34 0.016 $1.37 \cdot 10^{-3}$	0.041 $1.72 \cdot 10^{-4}$ $1.19 \cdot 10^{-6}$
N = 100	D (100)	0.33 $2.93 \cdot 10^{-3}$ $4.79 \cdot 10^{-4}$	2.09 $2.56 \cdot 10^{-4}$ −0.10	0.32 $2.93 \cdot 10^{-3}$ $4.57 \cdot 10^{-4}$	0.27 $2.51 \cdot 10^{-3}$ $3.87 \cdot 10^{-4}$	0.29 $2.79 \cdot 10^{-3}$ $4.28 \cdot 10^{-4}$	$8.17 \cdot 10^{-3}$ $3.68 \cdot 10^{-7}$ $9.20 \cdot 10^{-9}$
	7 − P (100)	0.16 $9.50 \cdot 10^{-3}$ $8.57 \cdot 10^{-4}$	2.08 −0.011 −0.11	0.16 $9.38 \cdot 10^{-3}$ $8.25 \cdot 10^{-4}$	0.13 $7.42 \cdot 10^{-3}$ $5.70 \cdot 10^{-4}$	0.14 $8.57 \cdot 10^{-3}$ $7.22 \cdot 10^{-3}$	$2.38 \cdot 10^{-3}$ $3.84 \cdot 10^{-6}$ $1.57 \cdot 10^{-7}$
	8 − P (100)	0.15 $7.72 \cdot 10^{-3}$ $6.53 \cdot 10^{-4}$	2.05 $1.42 \cdot 10^{-3}$ −0.10	0.14 $7.72 \cdot 10^{-3}$ $6.53 \cdot 10^{-4}$	0.13 $6.15 \cdot 10^{-3}$ $5.16 \cdot 10^{-4}$	0.14 $6.96 \cdot 10^{-3}$ $5.83 \cdot 10^{-4}$	$2.01 \cdot 10^{-3}$ $4.14 \cdot 10^{-6}$ $2.54 \cdot 10^{-8}$
b) True response curve: "high-narrow" model $\beta = (2 \quad 0 \quad -4)^T$							
N = 50	D (99)	0.67 0.26 1.70	2.15 $-1.26 \cdot 10^{-3}$ −4.32	0.60 0.29 1.60	0.54 0.20 1.24	0.68 0.25 1.59	0.10 0.020 0.46
	7 − P (34)	0.86 0.14 1.27	1.87 −0.083 −3.43	0.85 0.14 0.94	1.57 0.65 2.60	1.56 0.48 2.20	0.27 0.030 0.52
	8 − P (84)	0.44 0.82 2.72	1.99 −0.020 −4.11	0.44 0.82 2.71	0.81 0.55 3.84	0.89 0.74 4.31	0.16 0.19 4.66
N = 100	D (100)	0.31 0.11 0.74	2.08 $4.97 \cdot 10^{-4}$ −4.16	0.30 0.11 0.71	0.27 0.10 0.62	0.29 0.11 0.68	$8.64 \cdot 10^{-3}$ $5.75 \cdot 10^{-4}$ 0.025
	7 − P (69)	0.76 0.18 0.96	2.10 $-1.77 \cdot 10^{-3}$ −4.03	0.75 0.18 0.96	0.79 0.33 1.30	0.87 0.36 1.40	0.12 0.020 0.26
	8 − P (98)	0.47 0.32 2.21	2.14 0.017 −4.30	0.45 0.32 2.12	0.40 0.27 1.92	0.50 0.30 2.12	0.084 $5.05 \cdot 10^{-3}$ 1.02
c) True response curve: "low-wide" model $\beta = (-2 \quad 0 \quad -0.1)^T$							
N = 50	D (12)	0.39 $4.50 \cdot 10^{-3}$ $5.75 \cdot 10^{-3}$	−1.99 $-1.91 \cdot 10^{-4}$ −0.037	0.39 $4.50 \cdot 10^{-3}$ $1.78 \cdot 10^{-3}$	0.57 0.072 $6.82 \cdot 10^{-3}$	0.68 0.030 $4.59 \cdot 10^{-3}$	0.090 $2.46 \cdot 10^{-5}$ $1.53 \cdot 10^{-6}$
	7 − P (73)	1.09 2.32 0.50	−2.04 0.49 −0.44	1.09 2.08 0.38	0.44 0.11 0.012	1.53 2.96 0.65	9.05 106.93 5.79
	8 − P (80)	0.42 1.15 0.82	−1.88 −0.026 −0.55	0.41 1.15 0.62	0.33 0.13 0.012	0.50 1.05 0.82	0.12 2.91 5.08
N = 100	D (35)	0.34 $5.71 \cdot 10^{-3}$ $2.86 \cdot 10^{-3}$	−2.11 $-1.94 \cdot 10^{-3}$ −0.066	0.33 $5.71 \cdot 10^{-3}$ $1.72 \cdot 10^{-3}$	0.29 0.036 $3.41 \cdot 10^{-3}$	0.35 0.026 $3.06 \cdot 10^{-3}$	0.042 $3.63 \cdot 10^{-5}$ $7.89 \cdot 10^{-7}$
	7 − P (97)	0.72 1.51 0.34	−2.09 0.30 −0.33	0.71 1.42 0.28	0.22 0.055 $5.95 \cdot 10^{-3}$	0.66 1.30 0.30	3.26 42.62 2.39
	8 − P (98)	0.23 0.44 0.33	−1.99 $-2.73 \cdot 10^{-3}$ −0.35	0.23 0.44 0.27	0.17 0.063 $6.02 \cdot 10^{-3}$	0.22 0.33 0.24	0.022 0.50 1.01
d) True response curve: "low-narrow" model $\beta = (-2 \quad 0 \quad -4)^T$							
N = 50	D (11)	0.37 0.14 8.15	−1.95 $2.68 \cdot 10^{-3}$ −1.64	0.36 0.14 2.55	0.57 2.88 10.91	0.66 1.20 7.23	0.083 0.037 3.69
	7 − P (0.04)	0.79 2.77 11.52	−2.56 0.58 −0.62	0.47 2.43 0.099	20.61 27.68 65.94	2.10 3.87 0.78	3.10 21.88 0.48
	8 − P (8)	0.46 1.72 4.64	−1.60 −0.078 −2.16	0.30 1.71 1.25	1.26 6.06 26.79	0.69 2.88 6.06	0.012 1.17 16.19
N = 100	D (35)	0.34 0.25 4.55	−2.10 0.012 −2.67	0.33 0.25 2.79	0.29 1.44 5.46	0.35 1.05 4.89	0.041 0.055 1.92
	7 − P (0.2)	0.64 0.14 9.13	−1.83 −0.33 −1.03	0.62 0.027 0.33	10.31 13.94 32.97	0.94 0.49 1.05	0.024 $3.09 \cdot 10^{-3}$ 0.20
	8 − P (21)	0.33 1.78 3.74	−2.02 0.067 −2.58	0.33 1.77 1.72	0.63 3.03 13.39	0.52 2.42 6.06	0.021 1.42 11.94

Table 11: Results of maximum likelihood estimation of the optimum point in the quadratic logit model for N=10 and N=20.

	Design ($P(MLE)$)	$\mathbf{P}(\widehat{x}_m)$	$\mathbf{E}(\widehat{x}_m)$	$\mathbf{V}(\widehat{x}_m)$	$\mathbf{AV}(\widehat{x}_m)$	$\mathbf{E}\left(\widehat{V}(\widehat{x}_m)\right)$	$\mathbf{E}\left(\widehat{V}(\widehat{x}_m)\right)$
a) True response curve: "high-wide" model $\beta = (2 \quad 0 \quad -0.1)^T$, $x_m = 0$							
N = 10	D (0.35)	0.94	$1.31 \cdot 10^{-16}$	1.40	0.63	6.18	170.6
	7 − P (0.53)	0.90	−0.32	14.23	1.86	345.85	$1.55 \cdot 10^{6}$
	8 − P (0.49)	0.95	$-6.38 \cdot 10^{-16}$	6.90	1.54	81.17	$1.10 \cdot 10^{5}$
N = 20	D (0.74)	0.99	$2.55 \cdot 10^{-4}$	0.59	0.31	$3.79 \cdot 10^{27}$	$3.38 \cdot 10^{58}$
	7 − P (0.82)	0.95	−0.12	4.03	0.93	42.16	$1.92 \cdot 10^{5}$
	8 − P (0.84)	0.96	0.076	3.40	0.77	59.30	$7.49 \cdot 10^{7}$
b) True response curve: "high-narrow" model $\beta = (2 \quad 0 \quad -4)^T$, $x_m = 0$							
N = 10	D (0.35)	0.94	$4.56 \cdot 10^{-17}$	0.035	0.016	0.15	0.11
	7 − P (0.014)	1	−0.44	0.12	0.051	1.55	0.22
	8 − P (0.11)	1	$-1.55 \cdot 10^{-17}$	0.13	0.043	0.29	$7.08 \cdot 10^{-3}$
N = 20	D (0.76)	0.99	$-5.31 \cdot 10^{-5}$	0.015	$7.86 \cdot 10^{-3}$	$4.17 \cdot 10^{27}$	$4.09 \cdot 10^{58}$
	7 − P (0.046)	1	−0.032	0.049	0.025	0.14	0.069
	8 − P (0.43)	1	−0.022	0.051	0.021	0.074	$1.86 \cdot 10^{-3}$
c) True response curve: "low-wide" model $\beta = (-2 \quad 0 \quad -0.1)^T$, $x_m = 0$							
N = 10	D $\left(2.5 \cdot 10^{-3}\right)$	0.99	$-3.82 \cdot 10^{-16}$	0.070	9.01	15.37	26.85
	7 − P (0.11)	0.99	0.86	9.59	13.73	$1.24 \cdot 10^{3}$	$3.39 \cdot 10^{8}$
	8 − P (0.096)	0.97	$2.80 \cdot 10^{-16}$	6.24	15.63	92.72	$1.05 \cdot 10^{7}$
N = 20	D (0.016)	1	$1.65 \cdot 10^{-15}$	1.30	4.51	78.97	$1.42 \cdot 10^{4}$
	7 − P (0.28)	0.99	0.49	21.00	6.87	$2.61 \cdot 10^{3}$	$2.01 \cdot 10^{8}$
	8 − P (0.36)	0.98	−0.056	5.51	7.82	278.24	$4.66 \cdot 10^{7}$
d) True response curve: "low-narrow" model $\beta = (-2 \quad 0 \quad -4)^T$, $x_m = 0$							
N = 10	D $\left(2.5 \cdot 10^{-3}\right)$	0.99	$1.00 \cdot 10^{-16}$	$1.74 \cdot 10^{-3}$	0.23	0.38	0.017
	7 − P $\left(2.4 \cdot 10^{-5}\right)$	1	−0.44	0.072	2.18	1.71	0.047
	8 − P $\left(2.6 \cdot 10^{-3}\right)$	1	$-2.60 \cdot 10^{-17}$	0.19	0.47	0.28	0.012
N = 20	D (0.016)	1	$-1.03 \cdot 10^{-15}$	0.032	0.11	1.97	8.88
	7 − P $\left(4.4 \cdot 10^{-5}\right)$	1	−0.29	0.32	1.09	1.07	0.19
	8 − P (0.016)	1	0.089	0.12	0.24	0.28	0.024

Table 12: Simulation results of maximum likelihood estimation of the optimum point in the quadratic logit model for N=50 and N=100.

	Design (% *MLE*)	$\widehat{x}_m$ %	$\overline{\widehat{x}_m}$	$s^2_{\widehat{x}_m}$	$AV(\widehat{x}_m)$	$\overline{\widehat{V}(\widehat{x}_m)}$	$s^2_{\widehat{V}(\widehat{x}_m)}$
a) True response curve: "high-wide" model $\boldsymbol{\beta} = (\mathbf{2} \quad \mathbf{0} \quad \mathbf{-0.1})^T$, $x_m = 0$							
N = 50	**D** (99)	100	$-4.13 \cdot 10^{-3}$	0.16	0.13	0.16	0.028
	7 − P (99)	99	-0.18	0.78	0.37	1.86	106.54
	8 − P (99)	100	$-5.40 \cdot 10^{-3}$	0.50	0.31	1.25	420.49
N = 100	**D** (100)	100	$1.30 \cdot 10^{-3}$	0.069	0.063	0.068	$6.10 \cdot 10^{-4}$
	7 − P (100)	100	-0.095	0.28	0.19	0.32	0.66
	8 − P (100)	100	$2.09 \cdot 10^{-3}$	0.19	0.15	0.20	0.072
b) True response curve: "high-narrow" model $\boldsymbol{\beta} = (\mathbf{2} \quad \mathbf{0} \quad \mathbf{-4})^T$, $x_m = 0$							
N = 50	**D** (99)	100	$-8.44 \cdot 10^{-4}$	$4.06 \cdot 10^{-3}$	$3.14 \cdot 10^{-3}$	$5.99 \cdot 10^{-3}$	0.020
	7 − P (34)	100	-0.013	$4.83 \cdot 10^{-3}$	0.010	0.014	$3.43 \cdot 10^{-4}$
	8 − P (84)	100	$-1.79 \cdot 10^{-3}$	0.016	$8.54 \cdot 10^{-3}$	0.018	$1.95 \cdot 10^{-4}$
N = 100	**D** (100)	100	$1.06 \cdot 10^{-3}$	$1.68 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$1.72 \cdot 10^{-3}$	$5.28 \cdot 10^{-7}$
	7 − P (69)	100	$2.73 \cdot 10^{-3}$	$3.40 \cdot 10^{-3}$	$5.09 \cdot 10^{-3}$	$5.89 \cdot 10^{-3}$	$9.68 \cdot 10^{-6}$
	8 − P (98)	100	$1.29 \cdot 10^{-3}$	$5.58 \cdot 10^{-3}$	$4.27 \cdot 10^{-3}$	$6.27 \cdot 10^{-3}$	$2.36 \cdot 10^{-5}$
c) True response curve: "low-wide" model $\boldsymbol{\beta} = (\mathbf{-2} \quad \mathbf{0} \quad \mathbf{-0.1})^T$, $x_m = 0$							
N = 50	**D** (12)	100	0.039	1.59	1.80	141.68	$4.76 \cdot 10^{4}$
	7 − P (73)	99	0.29	4.63	2.75	221.35	$8.16 \cdot 10^{6}$
	8 − P (80)	100	0.018	3.57	3.13	187.39	$1.61 \cdot 10^{7}$
N = 100	**D** (35)	98	-0.020	0.79	0.90	80.03	$1.32 \cdot 10^{5}$
	7 − P (97)	99	0.074	1.75	1.37	17.95	$9.70 \cdot 10^{4}$
	8 − P (98)	100	-0.027	1.65	1.56	17.10	$8.57 \cdot 10^{4}$
d) True response curve: "low-narrow" model $\boldsymbol{\beta} = (\mathbf{-2} \quad \mathbf{0} \quad \mathbf{-4})^T$, $x_m = 0$							
N = 50	**D** (11)	100	$4.38 \cdot 10^{-3}$	0.036	0.045	3.22	27.19
	7 − P (0.04)	100	0.17	1.37	0.44	1.04	0.18
	8 − P (8)	100	-0.014	0.097	0.095	0.21	0.026
N = 100	**D** (35)	98	$3.92 \cdot 10^{-3}$	0.022	0.023	1.62	68.91
	7 − P (0.2)	100	-0.25	0.054	0.22	0.40	0.29
	8 − P (21)	100	$8.90 \cdot 10^{-3}$	0.066	0.047	0.13	0.019

5 Discussion

The foremost important conclusion to be drawn is that non-existence is a big problem for this three parameter logistic model and the sets of true parameters that were examined here, especially for small samples. How severe the problem is depends on the true parameters and the design. The non-optimal designs considered here were sometimes better than the D-optimal design in this respect, due to the larger number of design points. The models where the response curve $\pi(x)$ is low were more problematic, in particular the "low-narrow" model where existence practically never occurred for the smallest sample size ($N = 10$) and only in 35 % of the times for the largest sample size ($N = 100$). The practical consequence is that large samples demanding big time and money efforts need to be taken and yet there may be a large risk of not obtaining estimates, depending on the true parameters.

It turned out to be quite a large discrepancy between the asymptotic sampling distribution of the MLE and its small sample distribution given that the MLE exists. Even when existence was 100 % and the sample size was $N = 100$ the MLE remained biased. It was only when the probability that the MLE exists was close to 100 % (as for the high models and the two largest sample sizes) and the D-optimal design was used that the simulation sample variance was close to the theoretical approximate variance. This is a problem because the construction of the optimal designs is based on the theoretical approximate variance. The accuracy of the variance estimator was also dependent on the proportion of existence of the MLE.

The parameter dependence makes things troublesome for these kind of models. However, when estimating the parameters the D-optimal design was not consistently outperforming the non-optimal designs, although it was preferable more often. When estimating the optimum point and for the smaller sample sizes in particular the difference between the D-optimal designs and the non-optimal designs was more distinct. These results imply that choosing one of the non-opitmal designs does not have to be disastrous, at least not when it comes to parameter estimation. Such a design might perform equally well or it might even be an improvement. Yet again, the choice of design did have a large impact on the probability of obtaining a MLE. The ideal would be to combine a D-optimal design with more points in a way that maximizes the probability of existence of the MLE. Some kind of sequential procedure could be required considering the parameter dependence of this problem. Further study is needed to find ways to deal with the problems of non-existence of the MLE in this quadratic logistic model.

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

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

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

[5] Fedorov, V. and Hackl, P. (1997). *Model-oriented Design of Experiment*. Springer, New York.

[6] Khuri, A. (2001). "An Overview Of The Use Of Generalized Linear Models In Response Surface Methodology," *Nonlinear Analysis* 47, 2023-2034.

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

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

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

[10] Silvey, S.D., (1980). *Optimal Design*. Chapman and Hall, London.