OPTIMAL DESIGN OF EXPERIMENTS FOR THE QUADRATIC LOGISTIC MODEL

Ellinor Fackle Fornius

Licentiate Dissertation
Department of Statistics
Stockholm University
2006

Licentiate Dissertation
Department of Statistics
Stockholm University
S-106 91 Stockholm

ISBN 91-7155-245-6

Abstract

Optimal design of experiments for binary data is the topic of this thesis. A particular logistic model containing a quadratic linear predictor and one control variable is considered. The problem that the optimal designs for this model depends on the unknown true parameters is in focus. The aim of the first paper is to examine the small sample performance of the optimal designs and to investigate the loss from using non-optimal designs. D-optimal designs are derived for different parameter sets and compared to a couple of non-optimal designs in a simulation study. The evaluations are made in terms of mean squared error of the maximum likelihood parameter estimator. Another problem with this model is the occurrence of certain data patterns for which no maximum likelihood estimates can be obtained. The designs differed considerably in this respect and this problem also turned out to be parameter dependent. When it comes to the small sample distribution of the maximum likelihood estimator it was demonstrated to be quite different from the asymptotic distribution. The aim of the second paper is to find a solution to the parameter dependence issue. Two sequential approaches are proposed and tested in a simulation study. The purpose is to find the conditions that maximize the probability of response. The first approach is an optimal design approach where c-optimal designs are updated sequentially. The second approach is a stochastic approximation approach which is a nonparametric approach, that is no distributional assumptions has to be made. The two approaches are compared in terms of mean squared error. Based on the simulation results the c-optimal design approach was consistently favored.

Keywords: Response Surface Methodology, D-optimality, c-optimality, Maximum likelihood estimator, Sequential design, Stochastic approximation.

Acknowledgements

Above all, my supervisor Professor Hans Nyquist deserves all the credit. Hans is the perfect supervisor: full of ideas, gifted with exceptional pedagogic skills and always willing to help.

My colleagues/friends at the Department of Statistics have contributed in many different ways, directly or indirectly, I wish to thank you all.

In particular I would like to thank my roommate Jessica Franzén for all the fun and for much good advice concerning statistical issues as well as how to find the right phrases.

Special thanks to Daniel Bruce for *fminsearch,* for sorting out numerous SWP problems and for many fruitful discussions about optimal design matters.

I am also very grateful to Marie Grufberg and Johanna Öhman for being such good friends.

To my family members: "mormor Gertrud", "mamma Eva", "lillebror Mikael" and "lillasyster Lill", thank you for always being there for me and for making my life full of joy.

To my husband, Ola Lindberg, thank you for your unfailing encouragement, thank you for making me happy and I love you.

Contents

1 Introduction

Experimentation constitutes a cornerstone of the empirical sciences. With the help of experiments it is possible to answer questions, to test hypotheses and ultimately to either confirm or refute theories. An experiment is a controlled study in which observations are made and data collected that forms the basis for analysis and subsequent conclusions. The quality of the analysis depends directly on the experimental design. A well-designed experiment allows valid conclusions to be drawn. Statistical methods are of vital importance to achieve this ambition. Available as an alternative is to perform an observational study in which the researcher does not make any interventions but merely observes existing states. The advantage with experiments over observational studies is the opportunity to control the experimental conditions and to determine which variables to include. Observational studies are vulnerable in the sense that the interpretations may be distorted by important variables that are not measured or even unknown.

A common objective is to gain knowledge about a process or system, such as a manufacturing process or a biomedical system, that is affected by one or more controllable variables and possibly a number of uncontrollable variables. A response variable reflects some observable aspect of the output that is of interest, like the yield of a chemical process or whether a product is defect or not. The person conducting the experiment, the experimenter, can vary the levels of the control variable(s) systematically in order to investigate how it influences the response variable(s). Response Surface Methodology (RSM) treats the statistical methods for design and analysis of such experiments. The search for optimum operating conditions in a manufacturing process is a frequent application.

RSM rests on the notion that there are many equally satisfactory paths that lead to correct conclusions and that the learning process is iterative. It is often wise to split the resources into several stages instead of devoting everything to one large experiment. For instance, it may very well happen that a screening experiment points in a different direction than anticipated beforehand. For example, moving from current towards optimum operating conditions is usually accomplished in more than one step.

The theory of optimal design provides an approach that enables the experimental design to be customized to a specific inferential goal. Planning and performing experiments require resources, optimal experimental design is about getting good value for the time and money invested. By carefully considering certain choices before the experiment is conducted the information obtained can be maximized (given a cost constraint) or the expenses can be minimized (for a desired precision). The choices that have to be made include deciding on which variables to examine, at what levels to make observations and the corresponding proportion of observations

as well as how the resources should be divided between different stages of the experiment.

A pioneer in the design of experiments area was Sir R. A. Fisher. He introduced statistical principles to the experimental design in the studies of agricultural systems. The work of Fisher starting in the 1930's laid the foundation to the statistical experimental design, see for example Fisher (1935). The applications were mainly in the agricultural, biological and medical fields. In the 1950's the development of RSM caused statistical design of experiments to enter the industrial field of application. RSM originated with the paper by Box and Wilson (1951) on which an extensive growth followed during the next decades. The book by Box and Draper (1987) gives a comprehensive treatment of the RSM techniques.

The optimal design theory was initiated in the late 1950's where many contributions are attributable to J. Kiefer. For instance, the alphabetical terminology referring to the optimality criteria was introduced by Kiefer (1959) and the General Equivalence Theorem is owing to Kiefer and Wolfowitz (1959) and Kiefer (1961). Fedorov (1972), Silvey (1980), Pazman (1986), Atkinson and Donev (1992), Pukelsheim (1993) and Fedorov and Hackl (1997) are all classic books that cover optimal design of experiments.

The attention was predominantly concentrated to linear models in the beginning but the research has been extended to concern non-linear models including Generalized Linear Models. The crux of the problem of finding optimal designs for the latter cases is that the optimal design generally depends on the unknown model parameters. It may seem discouraging that the construction of an optimal design to estimate the parameters requires the very same parameters to be known prior to the experiment. However, there exist several approaches to solve this problem. The most straightforward solution is to base the construction on a best guess, obtained from earlier experiments or based on expert knowledge, which obviously risks being poor if the guess is poor. Such a design is called a locally optimal design and was introduced by Chernoff (1953). A natural development of this approach is the so called optimum on average designs (also known as Bayesian designs) which assume a prior distribution on the parameters instead of focusing on just one guess, see e.g. Fedorov and Hackl (1997). Another alternative is the sequential construction of designs which is the approach that is adopted in this thesis. The idea is that the parameter estimates and the optimal design are updated stepwise, that is a locally optimal design is derived at each step based on the parameter estimates obtained in the previous step.

Much of the research concerning optimal experimental design for GLMs has been devoted to the logistic two-parameter model. A common application is the dose-response models that relates the control variable, e.g. the dose of a drug, to the probability of response, e.g. that a test subject is cured. The interest is often to find the dose associated with some specific response rate, referred to as the effective

dose. An overview of optimal designs for the two-parameter logistic model can be found in Myers et al. (1994). Wu (1985) presents various sequential designs for binary data, that for instance are suitable to find the effective dose.

This thesis treats optimal design of experiments for the logistic model. After a summary of the subject matters that are relevant for the thesis, two papers are appended; paper I with the title "D-optimal Designs for Quadratic Logistic Regression Models" and paper II with the title "Sequential Designs for Binary Data with the purpose to Maximize the Probability of Response". In common for the two papers is that a particular logistic model involving three parameters is considered. Furthermore, both of the papers deal with the problem of parameter dependence but in different ways. In paper I the aim is to examine the extent of the adverse effects from using non-optimal designs in a simulation study. In addition, the focus is on the performances of optimal as well as non-optimal designs in small samples. In paper II two sequential approaches are proposed as a solution to the parameter dependence issue. The performances of the sequential designs and their abilities to find the optimum operating conditions are evaluated through simulations. The outline of the summary is as follows. Chapter 2 introduces the concept of generalized linear models and reviews the theory of optimal design of experiments. Chapter 3 goes through Response Surface Methodology, it starts with a general introduction followed by the definition of the logistic model. Summaries of the two papers are given in the next chapter. Conclusions together with some suggestions for future research areas appear in the last chapter.

2 Design of experiments for Generalized Linear Models

2.1 Generalized Linear Models

The concept of a generalized linear model unifies the analysis of a wide variety of statistical models. The standard linear regression model with the responses being continuous and normally distributed can be viewed as a special case. However, both continuous and discrete data as well as several other probability distributions fit into this framework. Consider for example an experiment where the outcome is one out of two possibilities, e.g. success or failure, explode or not explode, resulting in a binary response variable. Logit and probit regression are commonly used models for situations of that kind. Generalized Linear Models were introduced in Nelder and Wedderburn (1972), a comprehensive book-length treatment is provided by McCullagh and Nelder (1989) and an introduction is given by Dobson (2002). The class of GLMs shares the following characteristics.

- *The random component* specifies the distribution of the independent response variables $Y_1, ..., Y_N$. This can be any distribution that is a member of the exponential family and the response variable may be both discrete or continuous.

- *The linear predictor* defined as

$$\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}, \quad i = 1, ..., N$$

  where $\mathbf{x}_i$ is a $p \times 1$ vector containing the control variables and $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector. The vector $\mathbf{x}_i$ may include non-linear elements, in contrast to the parameter vector $\boldsymbol{\beta}$. For instance, a model with an intercept, a quadratic term and a cubic term yields $\mathbf{x}_i = \begin{bmatrix} 1 & x_i & x_i^2 & x_i^3 \end{bmatrix}^T$.

- *The link function*

$$g(\mu_i) = \eta_i \quad i = 1, ..., N$$

  defines the connecting link between the mean response, $E(Y_i) = \mu_i$, and the linear predictor, η_i. The only limitation is that g should be a differentiable and monotonic function. Examples on standard GLMs and corresponding link functions are shown in Table 1.

Table 1: Some GLM examples.

Regression model	**Response**	**Distribution**	**Link function**	**Inverse link function**
Linear	Continuous	Normal	identity: $\eta = \mu$	$\mu = \eta$
Logistic	Binary	Binomial	logit: $\eta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp^\eta}{1+\exp^\eta}$
Probit	Binary	Binomial	probit: $\eta = \Phi^{-1}(\mu)$	$\mu = \Phi(\eta)$
Poisson	Counts	Poisson	log: $\eta = \ln \mu$	$\mu = \exp^\eta$

The mean response μ_i is a function of the linear predictor, see Table 1, that in turn is a function of the control variables and the model parameters. Furthermore, the variance of Y_i generally depends on the mean response μ_i. For example, in the case of a binomial response variable $V(Y_i) = N_i \mu_i (1 - \mu_i)$, where N_i is the number of observations at x_i, and for a Poisson response variable $V(Y_i) = \mu_i$, whereas in the standard linear regression model $V(Y_i) = \sigma^2$. The function

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

is called the GLM weight. For binomial Y_i and the logit link the GLM weight becomes $v(x_i) = N_i \mu_i (1 - \mu_i)$, for a Poisson response and log link $v(x_i) = \mu_i$ and

for a normally distributed response it is constant, $v(x_i) = 1/\sigma^2$. This is what makes the construction of optimal designs for GLMs generally more complicated than for linear models. An optimal design often attempts to maximize the precision, or equivalently minimize the variance, of the parameter estimates, e.g. to make good predictions of the response. The GLM weight enters as term into the variance associated with estimating the model parameters. Since $v(x_i)$ depends on the true parameters through μ_i the optimal design will also depend on the true parameters.

2.2 Optimal design

The experimenter has to decide on what levels of the control variable that should be used in the experiment, such as the amounts of a certain fertilizer in an agricultural experiment, the quantities of a new drug in a biomedical experiment or the temperatures in a manufacturing experiment. A design point is a particular level of the control variable, like 150 degrees or 5 ml of the drug. The locations of the design points have a direct influence on the amount of information that can be extracted from the experiment. Other issues related to the experimental design refer to the number of design points and the allocation of observations to the points. In the case of sequential experimentation additional questions arise, for example are a few large batches preferred over many batches of smaller size?

The purpose of conducting experiments is to increase knowledge, to gain as much information as possible. The core of the theory of optimal design is therefore the information matrix which mathematically summarizes the amount of information. Optimal designs are derived by maximizing some function of the information matrix or minimizing some function of its inverse.

2.2.1 The information matrix

Let ξ denote a design formulated according to

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ N_1 & N_2 & \dots & N_n \end{array} \right\}, \qquad \sum_{i=1}^{n} N_i = N,$$

where N_i is the number of observations taken at the design point x_i. Alternatively the design can be formulated as

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{array} \right\}, \quad w_i \geq 0, \quad \sum_{i=1}^{n} w_i = 1,$$

where the design weights, $w_i = N_i/N$, specify the allocation to the design points. The restriction that N_i should be integer needs to be imposed for the design to be realizable. The fulfilment of this condition is what distinguishes an exact design from a continuous design. It is typically more practical to handle continuous designs calculation-wise although such designs are not guaranteed to be feasible. However,

a continuous design can often serve as a satisfactory approximation, provided that N is not too small. Let the $N \times p$ design matrix be given by

$$\mathbf{X} = \begin{bmatrix} \left.\begin{matrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_1^T \end{matrix}\right\} & N_1 \text{ times} \\ \left.\begin{matrix} \mathbf{x}_i^T \\ \vdots \\ \mathbf{x}_i^T \end{matrix}\right\} & N_i \text{ times} \\ \left.\begin{matrix} \mathbf{x}_n^T \\ \vdots \\ \mathbf{x}_n^T \end{matrix}\right\} & N_n \text{ times} \end{bmatrix}$$

where the exact appearance of the $1 \times p$ vector $\mathbf{x}_i^T$ is determined by the model. Each of the N observations made according to the design ξ enters as a row in $\mathbf{X}$, one time each. If there are any replicates some of the rows will be exactly the same. The design matrix contains information about the location of the design points in the design space χ. The information matrix for a GLM

$$\mathbf{I}(\xi, \boldsymbol{\beta}) = \mathbf{X}^T \mathbf{V} \mathbf{X}$$

is composed of the design matrix and the $N \times N$ GLM weight matrix

$$\mathbf{V} = diag \left[\underbrace{v(x_1) \ldots v(x_1)}_{N_1 \text{ times}} \quad \ldots \quad \underbrace{v(x_i) \ldots v(x_i)}_{N_i \text{ times}} \quad \ldots \quad \underbrace{v(x_n) \ldots v(x_n)}_{N_n \text{ times}} \right].$$

That is, smaller weights are assigned to less informative locations/points and vice versa. The weights are determined by the model parameters $\boldsymbol{\beta}$, as argued in the previous section. Again, there may be elements in $\mathbf{V}$ that are identical if replicates are made at any design point. The information matrix reduces to

$$\mathbf{I}(\xi) = \frac{1}{\sigma^2} \mathbf{X}^T \mathbf{X}$$

for the linear regression model, in which case it suffices to consider the design matrix when deriving optimal designs (not the true parameters). Furthermore, it is customary to deal with the standardized information matrix that reveals the average information per observation. Adding more points to a design will never decrease the total information, worst case scenario is that it remains the same. Unless a candidate point causes the average information to increase it would pay off more to make replications at an already existing point. Seeking to maximize average information assures the inclusion of only highly informative design points. The standardized information matrix from a design is given by

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

The design matrix $\mathbf{X}$ is here of dimension $n \times p$ and the GLM weight matrix $\mathbf{V}$ of dimension $n \times n$. The $n \times n$ design weight matrix $\mathbf{W}$ contains the proportions of observations on the diagonal according to

$$\mathbf{W} = diag\begin{bmatrix} w_1 & ... & w_i & ... & w_n \end{bmatrix}.$$

The standardized information matrix can also be expressed as the weighted sum of the information obtained from the individual design points

$$\mathbf{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i v(x_i) \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^{n} \mathbf{m}(\boldsymbol{\beta}, x_i).$$

Moreover, the standardized variance associated with making predictions of the response at $\mathbf{x}$ is defined as

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

for a GLM. $d(\mathbf{x}, \xi)$, known as the standardized predictor variance, plays an important role in the construction of D-optimal designs as will be seen in the D-optimality section.

2.2.2 Maximum likelihood estimation

The inferential goal of an experiment is often to get high precision/low variance when estimating the model parameters. Maximum likelihood estimates for a GLM are obtained by solving the equations

$$\frac{\partial l}{\partial \beta_j} = U_j = \sum_{i=1}^{N} \frac{(y_i - \mu_i) x_{ij}}{V(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) = 0 \qquad j = 1, 2, ..., p$$

where l is the logarithm of the likelihood function $L(\boldsymbol{\beta}; \mathbf{y})$. The solutions to these equations generally have to be found numerically. The method of scoring (see e.g. Dobson, 2002) is an iterative method that is useful to compute the estimates. Given a guess $\mathbf{b}^{(m-1)}$ of the parameter vector $\boldsymbol{\beta}$ a new guess $\mathbf{b}^{(m)}$ is obtained by

$$\mathbf{b}^{(m)} = \mathbf{b}^{(m-1)} + \left[\mathbf{I}^{(m-1)}\right]^{-1} \mathbf{U}^{(m-1)}.$$

$\mathbf{I}^{(m-1)}$ is the information matrix and $\mathbf{U}^{(m-1)}$ is the vector of scores (U_j), both evaluated at $\mathbf{b}^{(m-1)}$. This can be rewritten as

$$\mathbf{b}^{(m)} = \left(\mathbf{X}^T \mathbf{V} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{V} \mathbf{z},$$

where $\mathbf{z}$ stands for a $N \times 1$ vector with elements

$$z_i = \mathbf{x}_i^T \mathbf{b}^{(m-1)} + (y_i - \mu_i) \left(\frac{\partial \mu_i}{\partial \eta_i}\right).$$

and $\mathbf{X}$ is the $N \times p$ design matrix. $\mathbf{V}$ and $\mathbf{z}$ are evaluated at $\mathbf{b}^{(m-1)}$. Iterations are continued until a termination criterion is reached. It can for example be that

the relative difference between $\mathbf{b}^{(m)}$ and $\mathbf{b}^{(m-1)}$ is less than a predetermined small number. $\mathbf{b}^{(m)}$ is then taken as the MLE of $\boldsymbol{\beta}$ denoted by $\widehat{\boldsymbol{\beta}}$. The asymptotic sampling distribution of $\widehat{\boldsymbol{\beta}}$ is normal with covariance

$$\mathbf{I}^{-1}(\xi, \boldsymbol{\beta}) = \left(\mathbf{X}^T \mathbf{V} \mathbf{X}\right)^{-1} = N^{-1} \mathbf{M}^{-1}(\xi, \boldsymbol{\beta}).$$

A problem that sometimes turns up is that of nonexisting maximum likelihood estimates. The data pattern of a sample determines whether the MLE exists or not. These issues are treated in Albert and Anderson (1984). As an extreme example consider a binary response experiment in which only zeros (nonresponse) are observed at all design points. In such case it is intuitive that no parameter estimates can be obtained.

2.2.3 Optimality criteria

Ideally the design that maximizes the entire standardized information matrix $\mathbf{M}(\xi, \boldsymbol{\beta})$ would be chosen as the optimal design. Unfortunately, such an optimization task is generally not doable. Some appropriate function of $\mathbf{M}(\xi, \boldsymbol{\beta})$ will instead be the subject of optimization. Let $\Psi\{\mathbf{M}(\xi, \boldsymbol{\beta})\}$ denote a criterion function that agrees well with the inferential goal of the experiment. Generally a design ξ^* is said to be Ψ-optimal if

$$\xi^* = \arg \min_{x_i, w_i, n} \Psi\{\mathbf{M}(\xi, \boldsymbol{\beta})\}$$

Let $\bar{\xi}$ be a design with design weight 1 at the design point x and let ξ' be defined as

$$\xi' = (1 - \alpha)\xi + \alpha \bar{\xi}$$

for $0 \leq \alpha \leq 1$. The directional derivative of $\Psi(\xi, \boldsymbol{\beta})$ in the direction $\bar{\xi}$ is given by

$$\phi(\mathbf{x}, \xi) = \lim_{\alpha \to 0^+} \frac{1}{\alpha} \left[\Psi\{\mathbf{M}(\xi', \boldsymbol{\beta})\} - \Psi\{\mathbf{M}(\xi, \boldsymbol{\beta})\}\right]$$

A design is Ψ-optimal if and only if the minimum of the directional derivative $\phi(\mathbf{x}, \xi^*) \geq 0$ for all $x \in \chi$. This result is stated in The General Equivalence Theorem (Kiefer and Wolfowitz (1959) and Kiefer (1961)) together with two equivalent conditions on ξ^*:

1. *The design* ξ^* *minimizes* $\Psi\{\mathbf{M}(\xi, \boldsymbol{\beta})\}$
2. *The minimum of* $\phi(\mathbf{x}, \xi^*) \geq 0$
3. *The derivative* $\phi(\mathbf{x}, \xi^*)$ *achieves its minima at the points of the design*

see e.g. Atkinson and Donev (1992). Various optimality criteria have been proposed, some popular examples are listed below.

D-optimality The perhaps most widely used optimality criteria is that of D-optimality, where the optimal design is found by minimizing

$$\Psi\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\}=\ln\left|\mathbf{M}^{-1}(\xi,\boldsymbol{\beta})\right|.$$

Equivalently a D-optimal design can be found by maximizing

$$\Psi^{*}\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\}=\ln\left|\mathbf{M}(\xi,\boldsymbol{\beta})\right|$$

because $|\mathbf{M}^{-1}|=|\mathbf{M}|^{-1}$. Numerical methods and the aid of a computer program are required to solve this task. Mathcad and Matlab were used in the work with this thesis. The rationale of this criterion function is that the square root of $|\mathbf{M}^{-1}(\xi,\boldsymbol{\beta})|$ is proportional to the volume of the asymptotic confidence region for the parameters and it is desirable to have this region as small as possible. When $\Psi\left(\xi,\boldsymbol{\beta}\right)=|\mathbf{M}^{-1}(\xi,\boldsymbol{\beta})|$ it can be shown that

$$\phi\left(\mathbf{x},\xi\right)=p-d\left(\mathbf{x},\xi\right),$$

see for example Silvey (1980), with the resulting alternative formulation of the conditions on ξ^{*} in The General Equivalence Theorem

1. *The design* ξ^{*} *minimizes* $\Psi\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\}$
2. $d\left(\mathbf{x},\xi^{*}\right)\leq p$
3. $d\left(\mathbf{x},\xi^{*}\right)$ *achieves its maxima at the points of the design*

The practical implication of this is that the optimality of a suggested design can easily be verified or disproved. A graphical examination of a plot of $d\left(\mathbf{x},\xi\right)$ reveals whether a design is optimal or not. In the case of a non-optimal design the appearance of the curve can give a clue about the optimal number of design points.

Sometimes interest is in s linear combinations of the p parameters, a situation that often arises in medical experiments when interest is in making comparisons between a control group and several treatment groups. The asymptotic covariance matrix for the linear combination $\mathbf{A}^{T}\boldsymbol{\beta}$ is given by

$$\mathbf{A}^{T}\mathbf{M}^{-1}\left(\xi,\boldsymbol{\beta}\right)\mathbf{A}$$

where $\mathbf{A}$ is a $p\times s$ matrix. This special case of D-optimality, referred to as D_{A}-optimality occurs when

$$\Psi\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\}=\ln\left|\mathbf{A}^{T}\mathbf{M}^{-1}\left(\xi,\boldsymbol{\beta}\right)\mathbf{A}\right|$$

is minimized.

Another special case, D_{s}-optimality, comes about in situations where there are s parameters of interest and $p-s$ nuisance parameters, or when interest is in model checking. Consider for example that the parameter vector is partitioned as $\boldsymbol{\beta}=\left[\boldsymbol{\beta}_{s}\quad\boldsymbol{\beta}_{p-s}\right]$ and that the aim is to find the optimal design to estimate $\boldsymbol{\beta}_{s}$. It is analogous to D_{A}-optimality if $\mathbf{A}$ is set equal to $\left[\mathbf{I}_{s}\quad\mathbf{0}\right]$, $\mathbf{I}_{s}$ being the $s\times s$ identity matrix and $\mathbf{0}$ a $(p-s)\times s$ matrix with zeros.

c-optimality c-optimality is an appropriate criterion when the aim is to estimate some function $h(\boldsymbol{\beta})$ of the model parameters with minimum variance. Such situations often arise in the RSM context when interest is in estimating the point of optimum response. Consider for example manufacturing a food product that needs to be prepared in oven, an experiment might then be conducted to find the baking time that maximizes the probability of a good product. The criterion function to be minimized is then

$$\Psi\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\} = \mathbf{c}^T\mathbf{M}^{-1}\left(\xi,\boldsymbol{\beta}\right)\mathbf{c}$$

where $\mathbf{c}$ is a $p \times 1$ vector. Specifically, when the target function is nonlinear in $\boldsymbol{\beta}$, the asymptotic variance of $h\left(\widehat{\boldsymbol{\beta}}\right)$ is given by

$$\begin{aligned} V\left(h\left(\widehat{\boldsymbol{\beta}}\right)\right) &= \mathbf{c}^T\mathbf{V}\left(\widehat{\boldsymbol{\beta}}\right)\mathbf{c}, \\ \mathbf{c} &= \frac{\partial h\left(\boldsymbol{\beta}\right)}{\partial \boldsymbol{\beta}}. \end{aligned}$$

Hence, it follows that c-optimality is a natural criterion when the purpose of the experiment is accurate estimation of $h(\boldsymbol{\beta})$.

A-optimality Even if the volume of a confidence ellipsoid is small (as strived for using the D-optimality criterion) all variances of the parameter estimates are not necessarily small. As an alternative A-optimality strives for minimizing the sum of the variances of the parameter estimates. The diagonal elements of the inverse of the standardized information matrix are proportional to the asymptotic variances of the MLE of $\boldsymbol{\beta}$. The design that minimizes

$$\Psi\left\{\mathbf{M}\left(\xi,\boldsymbol{\beta}\right)\right\} = tr\left[\mathbf{M}^{-1}(\xi,\boldsymbol{\beta})\right]$$

is called A-optimal.

It needs to be pointed out that these optimal design criteria rely on asymptotics for the GLM models. In practice all designs are limited regarding the number of observations. As a consequence the success of a particular design depends on the degree of agreement between the asymptotic sampling distribution and the current small sampling distribution.

3 Response Surface Methodology

The main aim of RSM is to study the nature of the relationship between the control variables and the response variable. A response surface (or response curve in the case of two dimensions) is used to depict this relationship. Except for some rare

occasions where the exact nature of the relationship is known allowing a mechanistic model to be formulated, RSM mostly deals with approximate empirical models.

In general a response variable Y is observed in an experiment and the relationship between Y and the control variables $\mathbf{x} = \begin{bmatrix} x_1 & ... & x_k \end{bmatrix}^T$ is assumed to have the functional form

$$y = f(\mathbf{x}, \boldsymbol{\theta}) + \varepsilon.$$

Observing the response may be associated with measurement and/or observational errors. Furthermore, there may be errors due to variations in the experimental setting. The term ε captures all kinds of errors. Hence, the mean response expressed as

$$E(Y) = f(\mathbf{x}, \boldsymbol{\theta}),$$

forms the response surface. Often the exact true functional form $f(\mathbf{x}, \boldsymbol{\theta})$ is unknown and needs to be approximated. A function $g(\mathbf{x}, \boldsymbol{\beta})$ is taken as a local approximation to $f(\mathbf{x}, \boldsymbol{\theta})$. The approximation is local because it is restricted to be valid only in a limited region of interest, R. The function $g(\mathbf{x}, \boldsymbol{\beta})$ is usually in the form of a polynomial. The use of a polynomial stems from making a Taylor expansion of $f(\mathbf{x}, \boldsymbol{\theta})$, a polynomial of degree d corresponds to truncating the Taylor's series after the $d : th$ order term. The higher degree of complexity of the polynomial, the better the approximation. Given a certain degree d, a smaller R will increase the closeness of the approximating function to the true function. Of course there has to be a balance between the complexity and a reasonable dimension, the higher degree of complexity the more parameters to be estimated.

It is common that an RSM experiment is performed stepwise. One important application is to determine the conditions of the control variables that maximize or minimize the response function, for example to find the optimum operating conditions for a manufacturing process. Starting at current operating conditions it may be adequate to fit a lower order model at the first stage to find the direction towards the optimum operating conditions. The method of steepest ascent (descent) is a sequential procedure that aims at taking steps towards the maximum increase (decrease) in the response. Assume that two control variables are believed to influence the response variable and that a first order model is fitted to begin with. A contour plot of the first order response surface may look like Figure 1, the steepest ascent is then found in the direction of the arrow. The steps are proportional to the parameter estimates, the actual step length is obtained by normalizing the vector $\widehat{\boldsymbol{\beta}}$ and then multiplying it with an appropriate step size. Observations on the response variable are made along this path until the response decreases, when a new first order model is fitted, a new direction of steepest ascent is located and a new path is followed. Experimentation proceeds in this manner until a lack of fit test indicates that the first order approximation no longer will do. This usually happens when the maximum of the response surface is nearby. A higher order model is then fitted and analyzed.

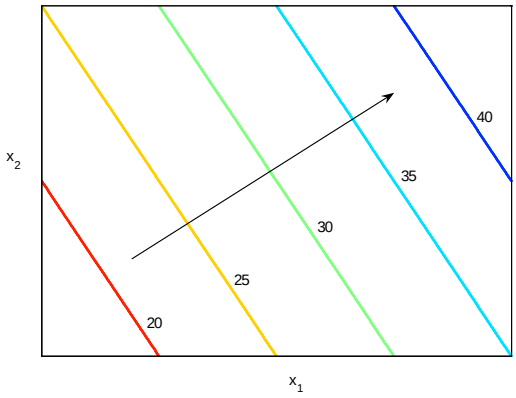

Figure 1: A contour plot of a first order response surface, the arrow points in the direction of steepest ascent.

RSM was initially concerned with linear regression models. However, there is nothing that prevents a response surface being fitted for a GLM model. In such case the mean response

$$E(Y) = \mu$$

that forms the response surface is connected to the linear predictor η via the inverse link function. Some examples on inverse link functions can be found in Table 1 in section 2.1. In the special case of a linear model the connection between the mean response and the linear predictor is direct, that is $\mu = \eta$. This thesis deals with a second order logistic regression model for binary data. Consider a binary response variable having the Bernoulli distribution, such as a food product that is either good or defect. The probability of a good product depends on the values of the control variables. The response surface is then given by

$$E(Y) = \pi(\mathbf{x}),$$

where $\pi(\mathbf{x})$ is the probability of obtaining a response. See Figure 2 for an example of a second order response surface with two control variables.

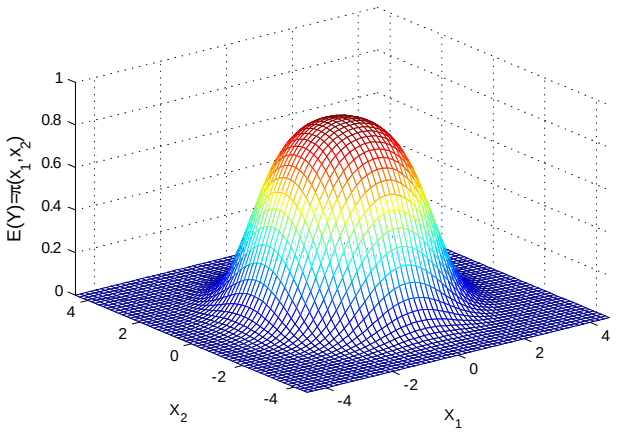

Figure 2: A second order response surface of the probability of response π.

Say for example that the control variable is baking time and a second order response curve approximately describes the relationship between the probability of a good product and the baking time. One reasonable model is the quadratic logistic model with the following link function

$$logit : \eta_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x}_i^T\boldsymbol{\beta} = \left[\begin{array}{ccc} 1 & x_i & x_i^2 \end{array}\right]\cdot\left[\begin{array}{ccc} \beta_0 & \beta_1 & \beta_2 \end{array}\right]^T = \beta_0+\beta_1 x_i+\beta_2 x_i^2,$$

such that mean response can be expressed as

$$\mu_i = \pi\left(x_i\right) = \frac{\exp\left\{\eta_i\right\}}{1+\exp\left\{\eta_i\right\}} = \frac{\exp\left\{\beta_0+\beta_1 x_i+\beta_2 x_i^2\right\}}{1+\exp\left\{\beta_0+\beta_1 x_i+\beta_2 x_i^2\right\}}.$$

Other link functions may also come into question to model this situation, the probit model for example agrees closely with the logit in the range $0.1 \leq \pi \leq 0.9$. An investigator may want to perform an experiment with the purpose to locate the point of maximum of the response curve in order to decide on the optimum baking time. Maximizing the response function is equivalent to maximizing the linear predictor since $\mu = \pi\left(x\right)$ is a monotonic function of η. The point of maximum response (θ) is easily found as the maximum of a second order polynomial in this case, that is

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

and θ is a nonlinear function of the model parameters. A design that allows estimation of a function of the parameters with high precision is found by using the c-optimality criterion, see section 2.2.3. Thus, a c-optimal design will be appropriate when the aim of the experiment is to estimate θ. The applications are not limited to maximizing the probability of response $\pi\left(x\right)$, optimizing any general

function $h(x)$ may as well be the objective. For instance, let $c(x)$ be the production cost and $h(x) = c(x)/\pi(x)$ be the cost per produced unit that is saleable. Interest may then be in finding the value of x that minimizes $h(x)$.

For linear response surface models there are several standard designs that are widely used, for example factorial designs and central composite designs. All of these designs are described in Box and Draper (1987). These are generally not applicable to GLM models due to the dependence of the variance function on the mean response and thereby on the model parameters. An overview of the usage of GLMs for response surfaces with the connection to optimal design theory is provided by Khuri (2001). The optimal designs derived using a criterion function are also associated with the parameter dependence problem. This issue calls for a solution. The approach of updating the design sequentially is a possibility explored in the second paper of this thesis. When an experiment is carried out in practise the sample size is inevitably restricted, a fact that may cause problems with using an optimal design if there are dissimilarities between the small sample and the asymptotic distributions. Another subject that needs to be examined concerns the size of the penalty in the form of lost precision from using a non-optimal design instead of an optimal design.

4 Summary of papers

4.1 Paper I: D-optimal Designs for Quadratic Logistic Regression Models

Paper I concerns the quadratic logistic model for binary data that was introduced in the previous chapter. Locally D-optimal designs are derived for four different parameter sets. The parameter sets yield four response curves with maxima that are either "high" or "low". The term "high" refers to the maximum value of $\pi(x)$ being close to 1 and "low" to the maximum value being close to 0. The D-optimal designs consist of three points with equal design weights for the "low" response curves. The optimal designs for estimation of the "high" response curves require four points and with the design weights now being different. A common feature for the designs is that the design points are placed symmetric around the point of maximum response.

There are two problems with these locally optimal designs, the first is the parameter dependence and the second one is the fact that the derivation of the designs involves the asymptotic information matrix. A simulation study is performed with the purpose to address these problems. In practise, when the parameters are unknown, a strategy might be to include more than the optimal number of design points to increase the chance that some points are good. Two non-optimal designs composed

of 7 and 8 design points are compared to the D-optimal designs. The comparison is made in terms of mean squared error to get an idea about the loss from not using the optimal designs. The evaluations are made both for the inferential goal of parameter estimation as well as for the pursuit of the optimum operating conditions. One of the aims with this paper is to make an assessment of how well the asymptotics work for various sample sizes, $N = 10, 20, 50$ and 100 are used in the simulations. Yet another predicament is that the maximum likelihood parameter estimate does not always exist, the probability for this to occur is computed for all designs. The most important points about the simulation results are summarized as follows.

- The non-existence of the MLE came about as a severe problem for all cases that were investigated in this paper. The probability that the MLE exists was dependent on the choice of design and the true parameters besides the sample size. From this viewpoint the non-optimal designs sometimes benefited from the many design points and were favored in several cases. The fact that the MLE is impaired by the lack of existing estimates reveals another aspect that needs to be taken into account when designing experiments for this particular logistic model.

- In terms of mean squared error the optimal designs were superior compared to the non-optimal designs for the most cases. However, there were also cases where the performances were equal and cases where the loss from choosing a non-optimal design was quite small. The differences were the largest for estimation of the point of optimum response in small samples.

- There were generally rather large disagreements between the small sample variances of the parameter estimators and their asymptotic analogue. This may cause problems since the asymptotic information matrix is the basis for the construction of the optimal designs. Although the differences diminished when N was increased they were still appreciable for several cases when $N = 100$ and the response curve was "low". However, for N exceeding 50 in the cases of a "high" response curve the differences were quite small. When it came to estimation of the point of optimum response there were also discrepancies but to a somewhat lesser extent.

4.2 Paper II: Sequential Designs for Binary Data with the purpose to Maximize the Probability of Response

Paper II deals with the same quadratic logistic model as in the first paper. Sequential estimation of the optimum operating conditions is proposed as a solution to the parameter dependence issue. Two sequential strategies are adopted, one is a parametric approach and the other one is nonparametric, for the case when the aim is to find the point of maximum response $x = \theta$. The advantage of the

nonparametric approach is to avoid having to make any distributional assumptions whereas the parametric approach makes more use of the information.

The method of the parametric approach is to derive c-optimal designs sequentially based on the assumption that the most recent parameter estimates are true. Hence, the idea is that improving the parameter estimates stepwise will prevent the negative effects of the parameter dependence. Locally c-optimal designs consisting of two equally weighted designs points are computed at each stage, using two versions of the standardized information matrix. The first is the regular version that reflects the information in the candidate design and the second one is a weighted version that also takes into account the information from the already observed points. The nonparametric sequential design provides an adaptive stochastic approximation approach. In accordance with a recursive scheme successive observations are made on the response variable in a way that the sequence of design points $x_1, x_2, ...$ converges to θ. This approach is adaptive in the sense that it makes use of the information gained so far by adjusting the step from x_r to x_{r+1}.

A simulation study is set up to investigate how the two approaches compare to each other and to a nonsequential approach. Different sample sizes ranging from $N = 200$ to $N = 1000$ as well as different batch sizes are explored. To study the effects of misspecifications in the model a different kind of linear predictor is also tested. The results of the simulations are really promising for the c-optimal sequential designs which consistently outperformed the other alternatives. The choice between the two versions did not matter as they were almost identical. The success of the stochastic approximation approach was limited to some cases and it was sometimes inferior to the nonsequential approach. In addition, it was more sensitive to the initial design and to the misspecified model. Furthermore, it required the specification of a constant that proved to have a large impact on the results. There were not any large differences pertaining to the division between the number of batches and the batch size for any of the approaches.

5 Conclusions and suggestions for future research

In Paper I it was concluded that non-existence of the MLE is a critical problem for the quadratic logistic model in association with small samples. The extent of these problems turned out to be dependent on the true parameters and the design. It was also concluded that the small sample distribution of the parameter estimator (given that the MLE exists) many times differed from what could be anticipated based on the asymptotic results. In Paper II the conclusion was that the proposed sequential approach for estimating the point of optimum response overcomes the problem of parameter dependence, at least for larger sample sizes. Some of the questions and problems that remain unsolved are mentioned below.

- First of all, solutions to the problem of nonexistence for the maximum likelihood estimator are demanded. Perhaps some alternative estimator that handles more data types can be tested for this logistic model. Heinze and Schemper (2002) proposed an estimator based on a modification to the score function originally suggested by Firth (1993). They concluded, based on a simulation study and the analysis of two clinical data sets, that it provides a solution to these problems for logistic regression.

- The risk of coming across non-existence has to be taken into account when it comes to the design of experiments. Consider facing the situation of having already performed an experiment and that no estimates can be computed. Good strategies for choosing additional points need to be developed for such situations. Designs that are optimal with respect to maximizing the probability to obtain estimates would also be of interest.

- More work needs to be done concerning the implications of the departures from the asymptotic results in small samples.

- More extensive evaluations are needed for the sequential designs, regarding for example other models and smaller samples. Theoretical evaluations are desired but the problem is that they are complicated by the dependency between the sequential steps.

- It remains to formally establish that the c-optimal design for the quadratic logistic model consists of two design points, an assumption that was made in Paper II. Ford et al. (1992) shows how c-optimal (and D-optimal) designs can be geometrically constructed for the case when η is a linear function consisting of two parameters and one control variable. It would be interesting to extend their method to the case when η is a quadratic function.

- This thesis has only dealt with models including one control variable. Adding more control variables would obviously increase the number of applications. RSM is often concerned with processes that involve several variables.

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] Box, G.E.P. and Wilson, K.B. (1951). "On the Experimental Attainment of Optimum conditions," *Journal of the Royal Statistical Society. Series B* 13, 1-45.

[5] Chernoff, H. (1953). "Locally Optimal Designs for Estimating Parameters," *Annals of Mathematical Statistics* 24, 586-602.

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

[7] Fedorov, V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.

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

[9] Ford, I. and Torsney, B and Wu, J. (1992). "The Use of a Canonical Form in the Construction of Locally Optimal Designs for Non-Linear Problems," *Journal of the Royal Statistical Society. Series B* 54, 569-583.

[10] Firth D. (1993). "Bias Reduction of Maximum Likelihood Estimates," *Biometrika* 20, 272-319.

[11] Fisher R.A. (1935). *The Design of Experiments*. Oliver & Boyd Ltd., Edinburgh.

[12] Heinze, G. and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression," *Statistics in Medicine* 21, 2409-2419.

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

[14] Kiefer, J. (1959). "Optimum Experimental Design," *Journal of the Royal Statistical Society. Series B* 21, 272-319.

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

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

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

[18] Myers, W.R. and Myers, R.H. and Carter, W.H. Jr. (1994). "Some Alphabetic Optimal Designs for the Logistic Regression Model," *Journal of Statistical Planning and Inference* 42, 57-77.

[19] Nelder, J.A. and Wedderburn, R.W.M. (1972). "Generalized Linear Models," *Journal of the Royal Statistical Society. Series A* 135, 370-384

[20] Pazman, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

[21] Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

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

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