A MODEL FOR BIVARIATE BERNOULLI VARIABLES

–OPTIMAL DESIGN AND INFERENCE

Daniel Bruce

Department of Statistics
Stockholm University

2006

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

ISBN 91-7155-213-8

Abstract

This thesis presents a model for dependent and equally distributed Bernoulli variables. The model, which incorporates explanatory variables, is a multivariate generalized linear models (MGLM).
In the particular case with just two dependent variables, the model can be further divided into three different cases according to symmetry properties and independency between the variables. Different cases can be distinguished using different parameter restrictions. Locally D-optimal designs are given for each case. The special case with symmetry properties for two dependent variables is examined in detail. D-optimal designs for the special case have 2, 3, or 4 design points, where the number of design points are determined by the log-odds ratio. The log-odds ratio also partly explain the appearance of the probability distribution of the response variables. For some $2-$point designs and some $4-$point designs a general expression for D-optimal designs is given. These proposed designs are found to have a high efficiency. In case the variables are independent, a general expression for D-optimal designs is derived.

Score tests and likelihood ratio tests are derived for testing if the Bernoulli variables are independent. Test statistics for two particular situations are outlined in detail. Numerical illustrations of these test statistics are presented in three examples, including one with visual impairment data.
An optimal design for maximization of the local asymptotic power of the score test is proposed. The asymptotically power function based on the proposed design is compared with power function for finite sample sizes using a small simulation experiment. The finite sample power of the proposed design resembles the properties of the asymptotic power as long as the log-odds ratio is negative. The simulation study also indicate that there can be problems when computing the test statistic for large values of the log-odds ratio. Since the locally optimal design depends on the unknown parameters, the robustness of the design is examined. The examination shows that the optimal design is robust against fairly incorrect parameter values.

Keywords: Cox bivariate binary model, D-optimality, E-optimality, power maximization, efficiency, log-odds ratio.

Contents

1 Introduction **1**

2 Overview of models for bivariate Bernoulli variables **2**

3 The modified Cox model **6**

4 Optimal designs **11**
- 4.1 D-optimal designs 12
- 4.2 E-optimal designs 14

5 Bivariate symmetric model **16**
- 5.1 Locally D-optimal designs 22
  - 5.1.1 Locally D-optimal designs when the log-odds ratio is large . 24
  - 5.1.2 Locally D-optimal designs when the log-odds ratio is large negative 27

6 Bivariate independent model **28**

7 Testing for independency **34**
- 7.1 Models without covariates 34
- 7.2 Models with covariates 36

8 Optimal designs for testing for independency of Bernoulli variables **40**
- 8.1 Small sample performances of the test; a simulation experiment . . 41
- 8.2 Robustness of locally optimal designs 45

9 Conclusions and suggestions for further research **48**

1 Introduction

This thesis considers a model for dependent Bernoulli variables. Especially, situations with pairs of Bernoulli variables are studied. A symmetric model for dependent Bernoulli variables is given. The presented model is particularly adequate for matched-pairs data, see e.g. Agresti (2002). The key property in the model is that the marginal probabilities are equal for all variables. In this model data follow a multinomial distribution. An advantage with the model is that the expressions for the likelihood as well as the Fisher information matrix are relatively uncomplicated. Consequently, parameter estimators are obtained quite readily even when there are many dependent variables. Examples with dependent Bernoulli variables exist in several sciences including both observational and experimental studies.

One example of an application is visual impairment data. The probability for visual impairment on the left eye is assumed to be equal to the probability of visual impairment on the right eye. There is also a dependency between the eyes. This kind of data have been studied by Rosner (1984), Tielsch et al. (1991), and Liang et al. (1992).
Another type of application is certain longitudinal studies. Suppose that some binary property is measured at two different occasions and that these measurements are possibly dependent. For example, respondents answer the same dichotomous question at two different occasions. The two answers have to be modeled as dependent although the marginal probability of "success" is the same at both times.
Consider an experiment where fry (of fish) are studied. The goal of the experiment is to determine how different types of food (treatments) affect some property of the fish. Fry that are assigned to a certain treatment are therefore kept isolated. The fry can be further divided by other explanatory variables (covariates). It is clear that the responses from the fry that are kept in the same isolation box are dependent. The model with equal marginal probabilities can be applied if the fry within a treatment are homogeneous.
In the fourth example groups of plants grow in common soil. Different batches of plants are then exposed to different amount of some fertilizer. Since the plants share soil the condition of each plant is dependent of the other plants within the same batch. If the response variable is binary or coded as binary the model applies to this kind of experiment.
A last example involves a company that produces a certain product. Employees perform repeated measurements on the product to investigate whether the product is flawless or not. The company wants to investigate if these employees come to the same result in every measurement. The probability that the product is flawless is constant for different measurements. Hence, a model for equally distributed dependent Bernoulli variables is suitable for this special situation.

The example above with fry allocated to different treatments is one example of a design of an experiment. A design of an experiment includes choosing the treatments and choosing the corresponding number of observations to be allocated to each treatment. The design is important since all analysis is based on the design. A design that generates the most precise estimators of the parameters in the model, according to some criterion, is referred to as an optimal design. Optimal designs for univariate binary data have received a lot of attention the past twenty years. Less has been written on optimal experimental designs for multivariate binary data. Still a large number of models for multivariate binary variables have been proposed.

The main aim of this thesis is to describe optimal designs for dependent Bernoulli variables using a modified Cox model. In the bivariate case, different properties for the probability distribution of the variables are explored. If possible, analytical results for locally D-optimal designs are derived for the bivariate model. The thesis also addresses test procedures for testing independency between the variables. Another aim is to propose and motivate optimal designs for maximizing the power of these tests. A limitation with the thesis is that only one class of models is considered.

Chapter 2 gives a brief overview over different models for bivariate Bernoulli variables. The modified Cox model is presented in Chapter 3. Expressions for the likelihood function, the score functions and the information matrix are derived. Chapter 4 contains an introduction to optimal designs. The different design criteria used throughout the thesis are illustrated by examples. A symmetric bivariate model is outlined in Chapter 5. Different symmetry properties for the probability distribution are given together with an example of D-optimal designs and some general results about D-optimal designs. Chapter 6 deals with the special situation when the variables are independent. Likelihood ratio tests and score tests for testing for independency between the Bernoulli variables are discussed in Chapter 7. In Chapter 8, an expression for a design that maximize the power of the score test is derived. The performance in small samples for the score test based on the optimal design is examined in a simulation experiment. Robustness of the optimal design is also examined in Chapter 8. Finally, Chapter 9 summarizes the results of the thesis and discusses further research.

2 Overview of models for bivariate Bernoulli variables

Let S_1 and S_2 denote two possibly dependent binary variables. Moreover let x_i be a covariate associated to the ith response unit, $i = 1, 2, \ldots, N$. Several ways of modelling the joint distribution of S_1 and S_2 as a function of x has been proposed. A summary of different approaches was given already in Cox (1972). Bonney (1987) presents general loglinear multivariate logistic models for arbitrary number of dependent binary variables. In the case of two variables, let η_1 and η_2 be

$$\begin{aligned}\eta_1 &= \ln \frac{P(S_1 = 1 \mid x)}{P(S_1 = 0 \mid x)} = \alpha + \beta x \\ \eta_2 &= \ln \frac{P(S_2 = 1 \mid S_1, x)}{P(S_2 = 0 \mid S_1, x)} = \alpha + \gamma_1 Z_{21} + \beta x,\end{aligned}$$

where

$$Z = 2S_1 - 1.$$

If S_1 and S_2 are independent then $\gamma_1 = 0$. Based on

$$P(S_1, S_2 \mid x) = \prod_{i=1}^{2} \frac{e^{\eta_i S_i}}{1 + e^{\eta_i}},$$

the probability of the four possible outcomes of (S_1, S_2), $(1,1)$, $(1,0)$, $(0,1)$, and $(0,0)$ are

$$\begin{aligned}\pi_{11}(x) &= \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \frac{e^{\alpha+\gamma_1+\beta x}}{1 + e^{\alpha+\gamma_1+\beta x}} \\ \pi_{10}(x) &= \frac{e^{\alpha+\beta x}}{1 + e^{\alpha+\beta x}} \frac{1}{1 + e^{\alpha+\gamma_1+\beta x}} \\ \pi_{01}(x) &= \frac{1}{1 + e^{\alpha+\beta x}} \frac{e^{\alpha-\gamma_1+\beta x}}{1 + e^{\alpha-\gamma_1+\beta x}} \\ \pi_{00}(x) &= \frac{1}{1 + e^{\alpha+\beta x}} \frac{1}{1 + e^{\alpha-\gamma_1+\beta x}},\end{aligned}$$

respectively. Thus, the bivariate probability distribution of (S_1, S_2) can be expressed as products of ordinary logistic functions. Therefore the log-likelihood function and the information matrix can be obtained quite readily. This model is referred to as the unsaturated model by Bonney (1987). The saturated model include different α and different β for η_1 and η_2.

Murtaugh and Fisher (1990) utilizes the bivariate logistic cumulative distribution function (cdf) given by Gumbel (1961). Define

$$\begin{aligned}\eta_1 &= \alpha_1 + \beta_1 x \\ \eta_2 &= \alpha_2 + \beta_2 x.\end{aligned}$$

The bivariate probability distribution of (S_1, S_2) can then be modelled using the Gumbel cdf

$$F_{U,V}(u, v) = \frac{1}{1 + e^{-u}} \frac{1}{1 + e^{-v}} \left[1 + \frac{\gamma e^{-u-v}}{(1 + e^{-u})(1 + e^{-v})}\right],$$

where U and V are continuous latent variables such that

$$\begin{aligned}S_1 = 1 &\text{ iff } U \leq \eta_1 \\ S_2 = 1 &\text{ iff } V \leq \eta_2\end{aligned}$$

The parameter γ incorporates the possible dependency between S_1 and S_2 in the model. Using $F_{U,V}(u,v)$ the probabilities

$$\begin{aligned}
\pi_{11}(x) &= \frac{1}{1+e^{-\eta_1}}\frac{1}{1+e^{-\eta_2}} + \frac{\gamma e^{-\eta_1-\eta_2}}{\left(1+e^{-\eta_1}\right)^2\left(1+e^{-\eta_2}\right)^2} \qquad (1)\\
\pi_{10}(x) &= \frac{1}{1+e^{-\eta_1}} - \frac{1}{1+e^{-\eta_1}}\frac{1}{1+e^{-\eta_2}} - \frac{\gamma e^{-\eta_1-\eta_2}}{\left(1+e^{-\eta_1}\right)^2\left(1+e^{-\eta_2}\right)^2}\\
\pi_{01}(x) &= \frac{1}{1+e^{-\eta_2}} - \frac{1}{1+e^{-\eta_1}}\frac{1}{1+e^{-\eta_2}} - \frac{\gamma e^{-\eta_1-\eta_2}}{\left(1+e^{-\eta_1}\right)^2\left(1+e^{-\eta_2}\right)^2}\\
\pi_{00}(x) &= 1 - \frac{1}{1+e^{-\eta_1}} - \frac{1}{1+e^{-\eta_2}} + \frac{1}{1+e^{-\eta_1}}\frac{1}{1+e^{-\eta_2}}\\
&\quad + \frac{\gamma e^{-\eta_1-\eta_2}}{\left(1+e^{-\eta_1}\right)^2\left(1+e^{-\eta_2}\right)^2}
\end{aligned}$$

are obtained. It follows directly that S_1 and S_2 are independent if and only if $\gamma = 0$. As Murtaugh and Fisher (1990) point out, the marginal probabilities of S_1 and S_2 are logistic in η_1 and η_2, respectively. The likelihood function follows directly from (1). Maximum likelihood estimation of $(\alpha_1, \beta_1, \alpha_2, \beta_2, \gamma)$ are conducted by numerical maximization of the likelihood function. Heise and Myers (1996) and Dragalin and Fedorov (2006) also use the Gumbel model in bivariate logistic regression models.

In addition Murtaugh and Fisher (1990) and Dragalin and Fedorov (2006) use the Cox bivariate binary model, given in Cox and Snell (1989), to model dependent binary variables. In this model

$$\begin{aligned}
\pi_{11}(x) &= \frac{e^{\eta_2}}{1+e^{\eta_{10}}+e^{\eta_{01}}+e^{\eta_2}}\\
\pi_{10}(x) &= \frac{e^{\eta_{10}}}{1+e^{\eta_{10}}+e^{\eta_{01}}+e^{\eta_2}}\\
\pi_{01}(x) &= \frac{e^{\eta_{01}}}{1+e^{\eta_{10}}+e^{\eta_{01}}+e^{\eta_2}}\\
\pi_{00}(x) &= \frac{1}{1+e^{\eta_{10}}+e^{\eta_{01}}+e^{\eta_2}},
\end{aligned}$$

where

$$\begin{aligned}
\eta_2 &= \alpha_2 + \beta_2 x\\
\eta_{10} &= \alpha_{10} + \beta_{10} x\\
\eta_{01} &= \alpha_{01} + \beta_{01} x
\end{aligned}$$

The marginal probabilities of S_1 and S_2 are not logistic in η_2, η_{10}, and η_{01}. Instead the conditional probabilities are logistic in η_2, η_{10}, and η_{01}, see e.g. Murtaugh and Fisher (1990). Throughout the thesis, the Cox binary model will be referred to as the Cox model.

This thesis presents a modified version of the Cox model. Assume that S_1 and S_2 are equally distributed, i.e. S_1 and S_2 have the same marginal distribution. Thus, S_1 and S_2 are exchangeable under this restriction, see Galambos (1988) for a definition of exchangeability. By imposing the restrictions

- $\eta_{10} = \eta_{01}$
- $e^{\eta_{10}} + e^{\eta_{01}} = e^{\eta_1}$

the bivariate probability distribution of (S_1, S_2) becomes

$$\begin{aligned}
\pi_{11}(x) &= \frac{e^{\eta_2}}{1+e^{\eta_1}+e^{\eta_2}} \\
\pi_{10}(x) &= \frac{1}{2}\frac{e^{\eta_1}}{1+e^{\eta_1}+e^{\eta_2}} \\
\pi_{01}(x) &= \frac{1}{2}\frac{e^{\eta_1}}{1+e^{\eta_1}+e^{\eta_2}} \\
\pi_{00}(x) &= \frac{1}{1+e^{\eta_1}+e^{\eta_2}},
\end{aligned}$$

where

$$\begin{aligned}
\eta_1 &= \alpha_1 + \beta_1 x \\
\eta_2 &= \alpha_2 + \beta_2 x.
\end{aligned}$$

The number of response categories is reduced from four to three, which yields a trinomial response for $S = S_1 + S_2$. Moreover the number of parameters is reduced from six to four. The parameter restrictions for independence as well as the expression for the likelihood function are changed compared to the Cox bivariate binary model. This modified Cox model is outlined in Chapter 3.

Another class of models are loglinear models. A loglinear model for two binary variables is defined by

$$\ln N\pi_{ij} = \lambda + \alpha_i + \beta_j + (\alpha\beta)_{ij}\,, \quad i, j = 0, 1,$$

where $N\pi_{ij}$ is the expected frequency under the current model. The model is analogous to a model for analysis of variance. To model the probabilities π_{11}, π_{10}, π_{01}, and π_{00} a four factor model is required. Agresti (2002) points out that loglinear models focuses on association and interaction in the joint distribution of categorical response variables. Logit models are preferable if a single categorical response variable depends on explanatory variables. This thesis focus on the latter situation where the probability of the different outcomes of S depend on an explanatory variable, x. Loglinear models are presented in Agresti (2002) , Bishop et al. (1975), and Christensen (1997).

The modified version of the Cox model can be represented in terms of a model often referred to as multinomial logistic model. Models for multinomial responses can be categorized depending on the type of data. Zocchi and Atkinson (1999) argues that there are different models for nominal, ordinal and hierarchical data. Agresti (2002) divides the models in a similar way. Models for nominal data have been explored by Fahrmeir and Tutz (2001), Puu (2003), and Agresti (2002). This kind of models is sometimes called simple multinomial logit models. When there is an ordering between the outcomes of a response, several models exist. Agresti (2002), Dobson (2002), Fahrmeir and Tutz (2001), and Zocchi and Atkinson (1999) present some models, examples include cumulative logit model, proportional odds model and continuation-ratio logit model. The continuation-ratio logit model is further explored in Fan (1999) and Fan and Chaloner (2004). The models for ordered responses are especially useful for efficacy-toxicity responses where a natural order among the different responses exist.
All the models above have used the same link function, the logit link. Other link functions such as probit link and complementary log-log link are discussed in Agresti (2002), Dobson (2002), and Fahrmeir and Tutz (2001).

Another type of models utilizes the odds ratio as a measure of the dependency between S_1 and S_2. This type of models are based on the cross-ratio model, see e.g. Dale (1986). Palmgren (1991), Le Cassie and Van Houwelingen (1994), and Appelgren (2004) use this model for bivariate binary responses. Let $\pi_{1\cdot} = \pi_{11} + \pi_{10}$ and $\pi_{\cdot 1} = \pi_{11} + \pi_{01}$ denote the marginal probabilities $P(S_1 = 1)$ and $P(S_2 = 1)$, respectively. Moreover, let Ω denote the odds ratio between S_1 and S_2, defined as

$$\Omega = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}}.$$

Using the expression from Palmgren (1991)

$$\pi_{11} = \begin{cases} \frac{1}{2}(\Omega - 1)^{-1}\left\{a - \sqrt{a^2 + b}\right\} & \text{if } \Omega \neq 1 \\ \pi_{1\cdot}\pi_{\cdot 1} & \text{if } \Omega = 1 \end{cases},$$

where

$$\begin{aligned} a &= 1 + (\pi_{1\cdot} + \pi_{\cdot 1})(\Omega - 1) \\ b &= -4\Omega(\Omega - 1)\pi_{1\cdot}\pi_{\cdot 1}. \end{aligned}$$

The other probabilities π_{10}, π_{01}, and π_{00} follow from the marginal probabilities $\pi_{1\cdot}$ and $\pi_{\cdot 1}$. These probabilities can be associated with covariates using the bivariate logistic regression model given by McCullagh and Nelder (1989). One example is obtained if

$$\begin{aligned} \ln \frac{\pi_{1\cdot}}{1 - \pi_{1\cdot}} &= \eta_1 = \alpha_1 + \beta_1 x \\ \ln \frac{\pi_{\cdot 1}}{1 - \pi_{\cdot 1}} &= \eta_2 = \alpha_2 + \beta_2 x \\ \ln \Omega &= \eta_{12} = \alpha_{12} + \beta_{12} x. \end{aligned}$$

In this model S_1 and S_2 are independent if and only if $\ln \Omega = 0$.

3 The modified Cox model

Let $S_1, S_2, \ldots, S_k$ denote k identically distributed binary variables. Let

$$S = \sum_{i=1}^{k} S_i,$$

and

$$P(S = s) = \pi_s \quad \text{for } s = 0, 1, \ldots, k.$$

Note that since this thesis focuses on identically distributed variables the number of response categories for $S_1, S_2, \ldots, S_k$ is reduced from 2^k to $k + 1$.

A model for S can be viewed as a multivariate generalized linear model (MGLM). In a MGLM the response variable, the linear predictor, and the link function are vector-valued functions, see Fahrmeir and Tutz (2001). The response vector is denoted Y,

$$Y = \begin{pmatrix} Y_1 & Y_2 & \ldots & Y_k \end{pmatrix}^T,$$

where

$$Y_i = \begin{cases} 1, \text{ if } S = i \\ 0, \text{ otherwise} \end{cases} \quad \text{for } i = 1, 2, \ldots, k.$$

Hence, the expected value of Y is

$$\mu = E\left[\begin{pmatrix} Y_1 & Y_2 & \ldots & Y_k \end{pmatrix}^T\right] = \begin{pmatrix} \pi_1 & \pi_2 & \ldots & \pi_k \end{pmatrix}^T.$$

In a logit model, one of the response categories is chosen to be reference category. Because the way Y is defined, the event $S = 0$ is chosen to be reference category. Given the reference category, the logit link function $g(\pi_1, \pi_2, ..., \pi_k)$ is

$$g(\pi_1, \pi_2, \ldots, \pi_k)^T = \begin{pmatrix} \ln \frac{\pi_1}{\pi_0} & \ln \frac{\pi_2}{\pi_0} & \ldots & \ln \frac{\pi_k}{\pi_0} \end{pmatrix}^T = \eta,$$

where η is the linear predictor.

$$\eta = \begin{pmatrix} \eta_1 & \eta_2 & \ldots & \eta_k \end{pmatrix}^T = \begin{pmatrix} \alpha_1 + \beta_1 x & \alpha_2 + \beta_2 x & \ldots & \alpha_k + \beta_k x \end{pmatrix}^T = \mathbf{x}\theta,$$

where

$$\mathbf{x} = \begin{pmatrix} 1 & 0 & \ldots & 0 & x & 0 & \ldots & 0 \\ 0 & \ddots & & \vdots & 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 & \vdots & & \ddots & 0 \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 & x \end{pmatrix}$$

and

$$\theta = \begin{pmatrix} \alpha_1 & \ldots & \alpha_k & \beta_1 & \ldots & \beta_k \end{pmatrix}^T.$$

$\mathbf{x}$ is a $(k \times 2k)$ matrix and θ is a size $2k$ vector. The probabilities $\pi_0, \pi_1, \ldots, \pi_k$ as a function of x are

$$\begin{aligned}\pi_0 &= \frac{1}{1+\sum_{i=1}^{k} e^{\eta_i}} \\ \pi_j &= \frac{e^{\eta_j}}{1+\sum_{i=1}^{k} e^{\eta_i}} \quad \text{for } j = 1, \ldots, k.\end{aligned}$$

No simple and direct interpretation of the parameters exist. The parameters α_k and β_k in η_k are interpreted from the expression $\eta_k = \ln \frac{\pi_k}{\pi_0}$. Thus it is difficult to interpret how different parameters affect the joint probability function of $S_1, S_2, \ldots, S_k$.

The probability function for one observation of Y is

$$P\left(Y = y \mid \theta\right) = \pi_1^{y_1} \pi_2^{y_2} \ldots \pi_k^{y_k} \left(1 - \pi_1 - \pi_2 - \ldots - \pi_k\right)^{(1-y_1)(1-y_2)\ldots(1-y_k)}.$$

Using the expression for the probability distribution of Y it follows that the distribution of Y belongs to the exponential family of distributions. Moreover, the likelihood function for a whole sample is

$$L\left(\theta|\mathbf{y}\right) = \prod_{i=1}^{N} \left\{\pi_{1i}^{y_{1i}} \pi_{2i}^{y_{2i}} \ldots \pi_{ki}^{y_{ki}} \left(1 - \pi_{1i} - \pi_{2i} - \ldots - \pi_{ki}\right)^{(1-y_{1i})(1-y_{2i})\ldots(1-y_{ki})}\right\}.$$

The loglikelihood function is

$$l\left(\theta|\mathbf{y}\right) = \sum_{i=1}^{N} \left(y_{1i}\eta_{1i} + \ldots + y_{ki}\eta_{ki} - \ln\left(1 + e^{\eta_{1i}} + \ldots + e^{\eta_{ki}}\right)\right).$$

The score function for one observation can be derived using the chain rule,

$$u\left(\theta\right) = \left(\frac{\partial \eta}{\partial \theta}\right)^T \left(\frac{\partial \pi}{\partial \eta}\right)^T \left(\frac{\partial l}{\partial \pi}\right)^T$$

The derivatives are given by

$$\left(\frac{\partial \eta}{\partial \theta}\right)^T = \begin{pmatrix} \frac{\partial \eta_1}{\partial \alpha_1} & \frac{\partial \eta_1}{\partial \alpha_2} & \cdots & \frac{\partial \eta_1}{\partial \alpha_k} & \frac{\partial \eta_1}{\partial \beta_1} & \frac{\partial \eta_1}{\partial \beta_2} & \cdots & \frac{\partial \eta_1}{\partial \beta_k} \\ \frac{\partial \eta_2}{\partial \alpha_1} & \ddots & & \frac{\partial \eta_2}{\partial \alpha_k} & \frac{\partial \eta_2}{\partial \beta_1} & \ddots & & \frac{\partial \eta_2}{\partial \beta_k} \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ \frac{\partial \eta_k}{\partial \alpha_1} & \frac{\partial \eta_k}{\partial \alpha_2} & \cdots & \frac{\partial \eta_k}{\partial \alpha_k} & \frac{\partial \eta_k}{\partial \beta_1} & \frac{\partial \eta_k}{\partial \beta_2} & \cdots & \frac{\partial \eta_k}{\partial \beta_k} \end{pmatrix}^T = \mathbf{x}^T$$

$$\left(\frac{\partial \pi}{\partial \eta}\right)^T = D = \begin{pmatrix} \frac{\partial \pi_1}{\partial \eta_1} & \frac{\partial \pi_2}{\partial \eta_1} & \cdots & \frac{\partial \pi_k}{\partial \eta_1} \\ \frac{\partial \pi_1}{\partial \eta_2} & \frac{\partial \pi_2}{\partial \eta_2} & & \frac{\partial \pi_k}{\partial \eta_2} \\ \vdots & & \ddots & \vdots \\ \frac{\partial \pi_1}{\partial \eta_k} & \frac{\partial \pi_2}{\partial \eta_k} & \cdots & \frac{\partial \pi_k}{\partial \eta_k} \end{pmatrix} = \begin{pmatrix} \pi_1\left(1-\pi_1\right) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_k \\ -\pi_1\pi_2 & \pi_2\left(1-\pi_2\right) & & -\pi_2\pi_k \\ \vdots & & \ddots & \\ -\pi_1\pi_k & -\pi_2\pi_k & & \pi_k\left(1-\pi_k\right) \end{pmatrix}$$

$$\left(\frac{\partial l}{\partial \pi}\right)^T = \begin{pmatrix} \frac{y_1}{\pi_1} - \frac{(1-y_1)(1-y_2)\ldots(1-y_k)}{(1-\pi_1-\pi_2-\ldots-\pi_k)} \\ \vdots \\ \frac{y_k}{\pi_k} - \frac{(1-y_1)(1-y_2)\ldots(1-y_k)}{(1-\pi_1-\pi_2-\ldots-\pi_k)} \end{pmatrix}.$$

The matrix D is symmetric. Moreover D is equal to $\mathrm{Var}(Y)$. Using the fact that

$$D\left(\frac{\partial l}{\partial \pi}\right)^T = (y-\mu)$$

yields the score function for a whole sample

$$u.(\theta) = \begin{pmatrix} u_{\alpha_1}.(\theta) \\ \vdots \\ u_{\alpha_k}.(\theta) \\ u_{\beta_1}.(\theta) \\ \vdots \\ u_{\beta_k}.(\theta) \end{pmatrix} = \sum_{i=1}^{N} \mathbf{x}_i^T (y_i - \mu_i) = \sum_{i=1}^{N} \begin{pmatrix} (y_{1i} - \pi_{1i}) \\ \vdots \\ (y_{ki} - \pi_{ki}) \\ x_i (y_{1i} - \pi_{1i}) \\ \vdots \\ x_i (y_{ki} - \pi_{ki}) \end{pmatrix}.$$

The Fisher information for one observation is derived using the score function.

$$\begin{aligned} I(\theta, x) &= E\left[u(\theta)\, u^T(\theta)\right] \\ &= E\left[\mathbf{x}^T (y-\mu)(y-\mu)^T \mathbf{x}\right] \\ &= \mathbf{x}^T D \mathbf{x} \end{aligned}$$

$$= \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \dots & -\pi_1\pi_k & x\pi_1(1-\pi_1) & -x\pi_1\pi_2 & \dots & -x\pi_1\pi_k \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & & \vdots & -x\pi_1\pi_2 & x\pi_2(1-\pi_2) & & \vdots \\ \vdots & & \ddots & -\pi_{k-1}\pi_k & \vdots & & \ddots & -x\pi_{k-1}\pi_k \\ -\pi_1\pi_k & -\pi_2\pi_k & \dots & \pi_k(1-\pi_k) & -x\pi_1\pi_k & -x\pi_2\pi_k & \dots & x\pi_k(1-\pi_k) \\ x\pi_1(1-\pi_1) & -x\pi_1\pi_2 & \dots & -x\pi_1\pi_k & x^2\pi_1(1-\pi_1) & -x^2\pi_1\pi_2 & \dots & -x^2\pi_1\pi_k \\ -x\pi_1\pi_2 & x\pi_2(1-\pi_2) & & \vdots & -x^2\pi_1\pi_2 & x^2\pi_2(1-\pi_2) & & \vdots \\ \vdots & & \ddots & -x\pi_{k-1}\pi_k & \vdots & & \ddots & -x^2\pi_{k-1}\pi_k \\ -x\pi_1\pi_k & -x\pi_2\pi_k & \dots & x\pi_k(1-\pi_k) & -x^2\pi_1\pi_k & -x^2\pi_2\pi_k & \dots & x^2\pi_k(1-\pi_k) \end{pmatrix}$$

$$= \mathbf{x}^* \otimes D,$$

where

$$\mathbf{x}^* = \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}.$$

In general, there is no problem to numerically obtain maximum likelihood estimators for models with $k \geq 3$ variables. Nevertheless analytical results are hard to derive for such models. For example when $k = 3$ the dependence structure between the variables is much more complicated than for a model with $k = 2$ variables. For the bivariate case when $k = 2$, the joint probability distribution for (S_1, S_2) is given in Table 1. Table 1 also includes marginal probabilities for S_1 and S_2.

		S_2 0	1	
S_1	0	$\pi_{00} = \dfrac{1}{1+e^{\eta_1}+e^{\eta_2}}$	$\pi_{01} = \dfrac{e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$	$\pi_{0\cdot} = \dfrac{2+e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$
	1	$\pi_{10} = \dfrac{e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$	$\pi_{11} = \dfrac{e^{\eta_2}}{1+e^{\eta_1}+e^{\eta_2}}$	$\pi_{1\cdot} = \dfrac{2e^{\eta_2}+e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$
		$\pi_{\cdot 0} = \dfrac{2+e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$	$\pi_{\cdot 1} = \dfrac{2e^{\eta_2}+e^{\eta_1}}{2\left(1+e^{\eta_1}+e^{\eta_2}\right)}$	1

Table 1: Joint probability function for (S_1, S_2) and marginal distributions for S_1 and S_2.

This modified Cox model is a special case of the original Cox model described in the Chapter 2. Assuming a bivariate model and that $\pi_{10} = \pi_{01}$, the Cox model reduces to the modified Cox model if

$$\pi_{10} + \pi_{01} = \pi_1$$

The relationship between the models can also be expressed in terms of the linear predictors

$$\eta_{10} = \eta_{01} = \eta_1 - \ln 2.$$

Recall that the linear predictors for each model is given by

The original Cox model	The modified Cox model
$\eta_{10} = \alpha_{10} + \beta_{10}x$	$\eta_1 = \alpha_1 + \beta_1 x$
$\eta_{01} = \alpha_{01} + \beta_{01}x$	$(\eta_2 = \alpha_2 + \beta_2 x)$
$(\eta_2 = \alpha_2 + \beta_2 x)$.	

Using the expression for the linear predictors above yields the following connection between the parameters in the models.

$$\begin{cases} \alpha_{10} = \alpha_{01} = \alpha_1 - \ln 2 \\ \beta_{10} = \beta_{01} = \beta_1 \end{cases}$$

Since there are only two dependent variables (S_1, S_2), the (marginal) odds ratio can be given by just one expression.

Property 1 *The odds ratio for* $S_1 = 1$ *is* $4e^{\eta_2 - 2\eta_1}$.

Proof. Denote the odds ratio for $S_1 = 1$ by Ω.

$$\Omega = \frac{\Omega_{1|1}}{\Omega_{1|0}} = \frac{\frac{\pi_{11}}{\pi_{01}}}{\frac{\pi_{10}}{\pi_{00}}} = \frac{\pi_{11}\pi_{00}}{\pi_{10}\pi_{01}} = \frac{\frac{1}{1+e^{\eta_1}+e^{\eta_2}}\frac{e^{\eta_2}}{1+e^{\eta_1}+e^{\eta_2}}}{\frac{e^{\eta_1}}{2(1+e^{\eta_1}+e^{\eta_2})}\frac{e^{\eta_1}}{2(1+e^{\eta_1}+e^{\eta_2})}} = 4e^{\eta_2-2\eta_1}.$$

■

Hence the log-odds ratio is

$$\ln\Omega = \ln 4 + \alpha_2 - 2\alpha_1 + x\left(\beta_2 - 2\beta_1\right).$$

In general the log-odds ratio depends on the value of x. A model for (S_1, S_2) contains the four parameters,

$$\theta^T = \begin{pmatrix}\alpha_1 & \alpha_2 & \beta_1 & \beta_2\end{pmatrix}.$$

The parameters can be interpreted using $\ln\frac{\pi_1}{\pi_0}$ and $\ln\frac{\pi_2}{\pi_0}$ as described above. Another way of interpreting the parameters is to use the expression for the log-odds ratio. For example, the effect of the covariate on the log-odds ratio is controlled by β_1 and β_2.

To see how the probability distribution of S changes with x, four plots with different parameter values are shown in Figure 1. Although the plots differ a lot, they share some general properties. Since β_1 and β_2 is larger than zero, $P(S = 0)$ decreases with x and $P(S = 2)$ increases with x.

4 Optimal designs

A design is defined as

$$\xi = \left\{\begin{array}{cccc} x_1 & x_2 & \dots & x_r \\ w_1 & w_2 & \dots & w_r \end{array}\right\},$$

where $x_1, x_2, \dots, x_r$ are called design points and $w_1, w_2, \dots, w_r$ are the corresponding design weights. The design weights determine the proportion of observations to be taken at the different design points. A design is optimal according to a specific criterion if it minimizes the corresponding criterion function. The choice of optimality criterion is controlled by the objectives of the experiment. These objectives are usually connected to the precision in the parameter estimators. Optimal designs for nonlinear models depend in general on the true and unknown values of the parameters. Typically, the Fisher information matrix, denoted by $I.(\theta, x)$, depends on the parameter vector θ as well as the vector of covariates, x. A design that is optimal for a given set of parameter values is therefore referred to as a locally optimal design. Thus, a locally optimal design is optimal in case the true value on the parameter vector equals the particular value chosen when determining the design. However, if the true value of the parameter vector is different from that chosen for determining the design, there is no guarantee that the design has any favourable properties.

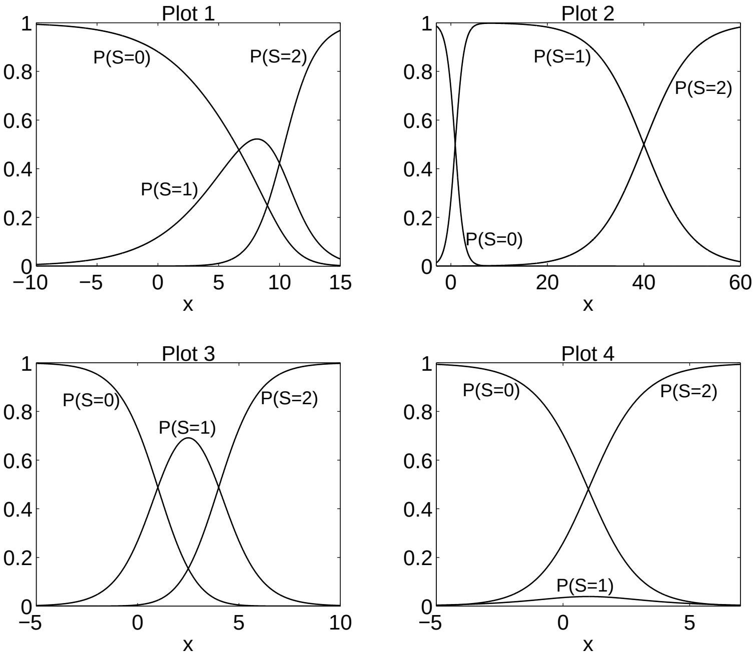


Figure 1: Four examples of different probability distributions for S. The parameters are $\theta_1^T = (-2, -9, 0.3, 1), \theta_2^T = (-1, -9, 1.1, 1.3), \theta_3^T = (-1, -5, 1, 2)$, and $\theta_4^T = (-3, -1, 0.5, 1)$, respectively.

Optimal design theory uses the standardized information matrix, denoted

$$M(\theta, x) = \frac{I_{\cdot}(\theta, x)}{N},$$

rather than the Fisher information matrix. In order to stress that the information matrices depend on the design, the information matrices are denoted

$$I_{\cdot}(\theta, \xi) \quad \text{and} \quad M(\theta, \xi),$$

respectively. Optimal designs for nonlinear models in general is treated in e.g. Silvey (1980), Atkinson and Donev (1992), Atkinson and Haines (1996), and Fedorov and Hackl (1997).

This thesis uses D-optimality and E-optimality as criteria for an optimal design. These optimality criteria have in common that their respective criterion function is a function of the standardized information matrix, and consequently also a function of the unknown parameters θ. Therefore, the derived optimal designs in this thesis are locally optimal designs. Furthermore, there is in general no closed form formula that defines an optimal design, it must be numerically determined using some routines for function optimization. In this thesis routines in Mathcad and Matlab have been used. A description of the criterion for D-optimality is given below together with a brief description of the criterion function for E-optimality.

4.1 D-optimal designs

The criterion function for D-optimality is

$$\psi\left\{M\left(\theta,\xi\right)\right\}=\ln\left(\left|M^{-1}\left(\theta,\xi\right)\right|\right).$$

Given regularity conditions, see e.g. Casella and Berger (2002), the variance of the maximum likelihood estimator $\widehat{\theta}_{mle}$ is asymptotically equal to the inverse of the Fisher information matrix, $I.\left(\theta,\xi\right)^{-1}$. The determinant of the inverse of the standardized information matrix in the criterion function is interpreted as a generalized volume of a confidence ellipsoid of the parameters. A smaller value of the criterion function therefore leads to better precision in the parameter estimators.

A design can be evaluated using the standardized variance of the predicted response, $d(x,\xi)$, see Puu (2003)

$$d(x,\xi)=tr\left\{M\left(\theta,\xi\right)^{-1}I\left(\theta,x\right)\right\}\quad\forall x\epsilon\mathfrak{X}$$

$\mathfrak{X}$ is called design region and specifies the possible values of x. Let

$$\overline{\xi}=\left\{\begin{array}{c}x\\1\end{array}\right\}$$

denote the design which puts unit mass at the point x. The directional derivative of $\psi\left\{M\left(\theta,\xi\right)\right\}$ in the direction of $\overline{\xi}$ is

$$\phi\left(x,\xi\right)=\lim_{\alpha\longrightarrow0+}\frac{1}{\alpha}\left[\psi\left\{\left(1-\alpha\right)M\left(\theta,\xi\right)+\alpha M\left(\theta,\overline{\xi}\right)\right\}-\psi\left\{M\left(\theta,\xi\right)\right\}\right],$$

see Atkinson and Donev (1992). Silvey (1980) showed that the directional derivative of the criterion function for a D-optimal design is

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

where p is the number of parameters in the model. According to the general equivalence theorem, (Kiefer (1959) and Kiefer and Wolfowitz (1960)), the design ξ^* is optimal if

$$\min_{x\epsilon\mathfrak{X}}\phi\left(x,\xi^*\right)\geqslant0.$$

Hence a design, ξ^*, that satisfy

$$d(x,\xi^*)\leqslant p\quad\forall x\epsilon\mathfrak{X}$$

is D-optimal. Furthermore, the general equivalence theorem state that, $d(x,\xi^*)=p$ at the design points. This inequality with $d(x,\xi)$ is very useful in order to verify if a design is D-optimal or not.

Example 1 To illustrate a D-optimal design, Figure 2 shows the probability distribution for an example of the bivariate model. The parameters are $\alpha_1 = -1$, $\alpha_2 = -9$, $\beta_1 = 1.1$, and $\beta_2 = 1.3$, respectively. In Figure 2, $d(x, \xi^*)$ for the locally D-optimal design

$$\xi^* = \left\{ \begin{array}{cccc} -0.4719 & 2.3431 & 32.3787 & 47.6213 \\ 0.2514 & 0.2598 & 0.2422 & 0.2466 \end{array} \right\}$$

is also included. Note that $d(x, \xi^*) = 4 = p$ at the design points. This result is in line with the general equivalence theorem, see Atkinson and Donev (1992).

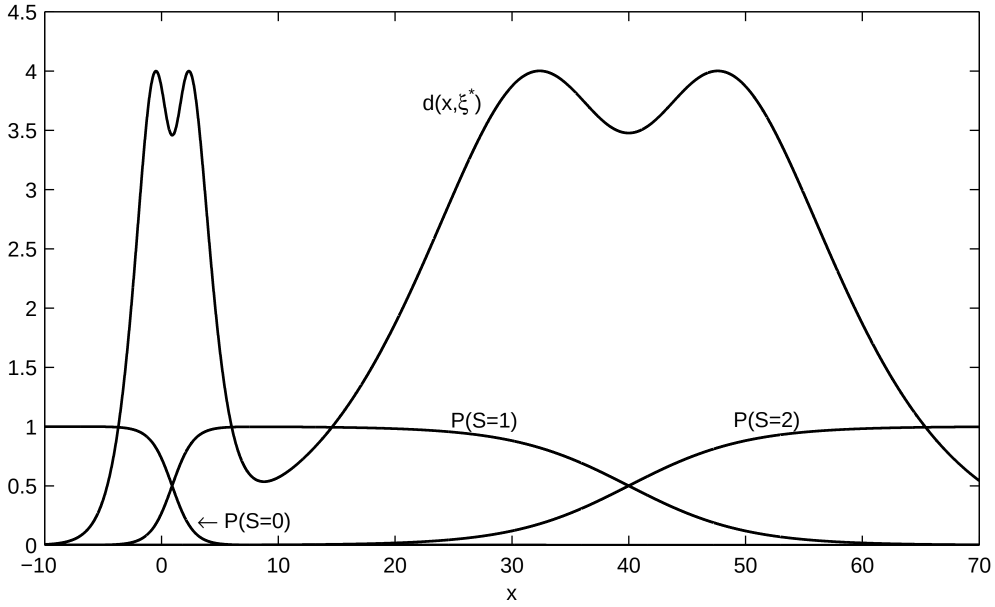


Figure 2: The probability distribution for S given $\alpha_1 = -1, \alpha_2 = -9, \beta_1 = 1.1$ and $\beta_2 = 1.3$. The standardized variance of the predicted response for design $\xi^*, d(x, \xi^*)$.

4.2 E-optimal designs

E-optimal designs minimize the variance of the worst estimated linear contrast, $a^T\theta$. An E-optimal design satisfy

$$\min \max_{i=1,\ldots,p} \frac{1}{\lambda_i},$$

where $\frac{1}{\lambda_i}$ is an eigenvalue to $M(\theta, \xi)^{-1}$. The E-optimal design is interpreted as the design that minimizes the length of the long axis of the confidence ellipsoid of the parameters.

The directional derivative of the criterion function for an E-optimal design is

$$\phi(x, \xi) = \lambda_{\min} - r^T I(\theta, x) r, \tag{2}$$

where $\lambda_{\min}$ is the smallest eigenvalue to $M(\theta,\xi)$ and r^T is the corresponding eigenvector. The expression in (2) can be used to verify that a design is E-optimal.

Example 2 Assume that the parameters are $\alpha_1 = -2$, $\alpha_2 = -5.3863$, $\beta_1 = 1$, and $\beta_2 = 2$, respectively. The corresponding locally E-optimal design, denoted ξ^*, is a 3−point design with unequal design weights.

$$\xi^* = \left\{ \begin{array}{ccc} 0.1741 & 2.2049 & 5.7469 \\ 0.4414 & 0.3706 & 0.1880 \end{array} \right\}$$

The directional derivative for ξ^* is given in Figure 3. Figure 3 shows that $\phi(x,\xi^*)$ achieves its minimum at the design points. and that the minimum value is equal to zero. ξ^* is therefore E-optimal, according to the general equivalence theorem.

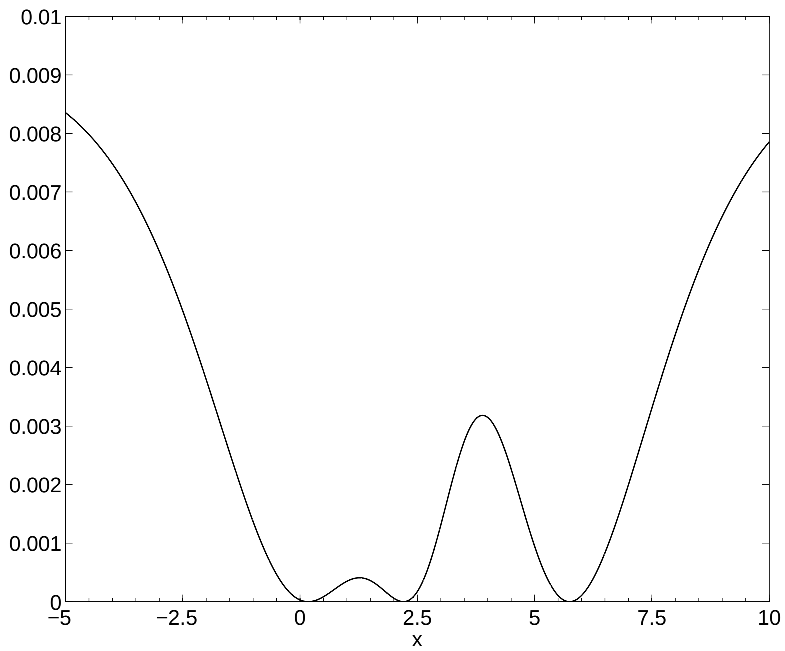


Figure 3: Directional derivative function, $\phi(x,\xi^*)$, for the E-optimal design ξ^*. The parameters are $\alpha_1 = -2, \alpha_2 = -5.3863, \beta_1 = 1$ and $\beta_2 = 2$, respectively.

Many articles have been written about optimal designs for multinomial models. Most authors address the situation where the binary responses "efficacy" (yes/no) and "toxicity" (yes/no) have a joint distribution. Heise and Myers (1996) derive locally D-optimal designs for this situation using the Gumbel model described above. The locally D-optimal designs are derived for different values on the parameters. The results show that the design points often are symmetrically allocated about some ratio of the parameters. They also study locally Q-optimal designs. The

Q-optimal designs minimize the predicted variance of the response ("efficacy","no toxicity"). Dragalin and Fedorov (2006) present locally D-optimal designs based on Cox bivariate binary model. They include a penalty function in the criterion function in order to avoid situations where the covariate attains the wrong values. Other authors have used the trinomial model with response categories "no response", "efficacy", and "adverse reaction". Puu (2003) considers locally D- and D_A-optimal designs for a multinomial logit model.
Appelgren (2004) derives locally D-optimal designs for the bivariate logistic regression model, (McCullagh and Nelder (1989)). He studied both models with independent margins and models with dependent margins. Results show that the parameters for the margins are most important for the location of the design points. The locally D-optimal designs have two, three, or four design points.

Since the D-optimal designs are only locally optimal designs, the designs can have a low efficiency given wrong parameter values. An optimal in average design, also known as Bayesian design, average the criterion function over a "prior" distribution of the parameters. In general D-optimal in average designs are therefore more robust than locally D-optimal designs. Optimal in average designs in general are treated in Chaloner and Larntz (1989), Pettersson (2001), and Pettersson and Nyquist (2003). Zocchi and Atkinson (1999) derive D-optimal in average designs and compare them with locally D-optimal designs for a trinomial model. Fan (1999) and Fan and Chaloner (2004) consider locally D-optimal designs, D-optimal in average designs, and locally c-optimal designs for a continuation-ratio logit model. They derive an analytical expression for a design which is approximately locally D-optimal. The number of design points as well as the location of the design points depend on the values of some of the parameters.

5 Bivariate symmetric model

In general, a model for two variables does not have any symmetry properties. However it is possible to define a model with symmetry properties where S_1 and S_2 are still dependent. In such a model, $\beta_2 = 2\beta_1$. The model is a particular case of the modified Cox model, but because $\beta_2 = 2\beta_1$, some properties are more specific. Compared with the modified Cox model for two variables the response variable and the link function are left unchanged. The linear predictor changes to

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \alpha_1 + \beta_1 x \\ \alpha_2 + 2\beta_1 x \end{pmatrix} = \mathbf{x}\theta,$$

where

$$\mathbf{x} = \begin{pmatrix} 1 & 0 & x \\ 0 & 1 & 2x \end{pmatrix}$$

and

$$\theta^T = \begin{pmatrix} \alpha_1 & \alpha_2 & \beta_1 \end{pmatrix}.$$

Based on the expression derived for the modified Cox model, the log-odds ratio is

$$\ln \Omega = \ln 4 + \alpha_2 - 2\alpha_1.$$

Note that the odds ratio depends neither on β_1 nor on x.

The probability distribution for S depends on the parameters α_1, α_2, and β_1 and the covariate x. In Figure 4 $P(S = 0)$, $P(S = 1)$ and $P(S = 2)$ are plotted for four different combinations of α_1, α_2 and β_1.

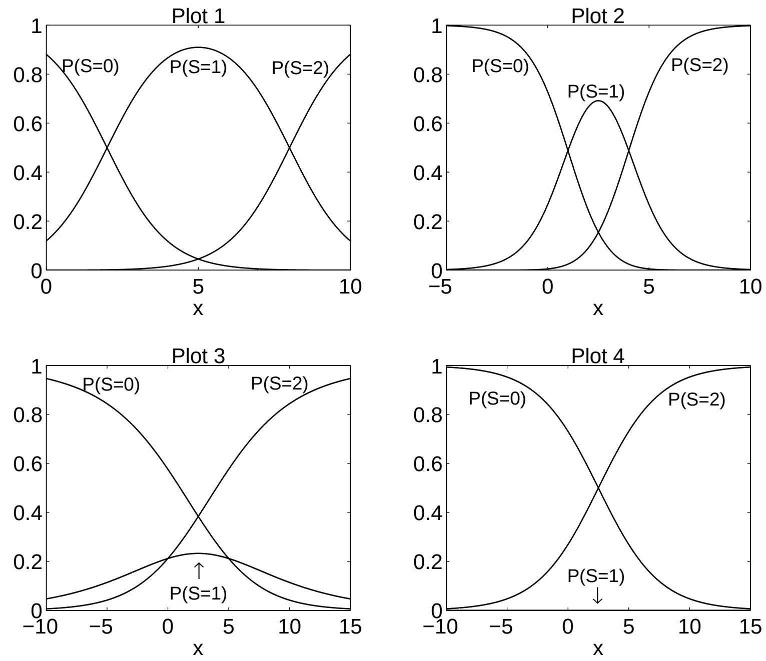


Figure 4: Probability distribution for S as a function of x. The parameters are $\theta_1^T = (-2, -10, 1), \theta_2^T = (-1, -5, 1), \theta_3^T = (-1, -1, 0.2)$, and $\theta_4^T = (-10, -1, 0.2)$, respectively. The corresponding log-odds ratios are $\ln\Omega_1 = -4.61$, $\ln\Omega_2 = -1.61$, $\ln\Omega_3 = 2.39$, and $\ln\Omega_4 = 20.39$, respectively.

The maximum value of $P(S = 1)$ decreases as the log-odds ratio increases. In Plot 4 with a log-odds ratio of over 20, $P(S = 1)$ is not possible to see since it is so close to zero. Since $\beta_1 > 0$, $P(S = 0)$ decreases with x and $P(S = 2)$ increases with x.

Define x_0 as

$$x_0 = \arg\max_{x \in \mathfrak{X}} P(S = 1).$$

Property 2

$$x_0 = \frac{-\alpha_2}{2\beta_1}$$

Proof.

$$\frac{dP(S=1)}{dx} = \frac{\beta_1 e^{\eta_1} - \beta_1 e^{\eta_1+\eta_2}}{(1+e^{\eta_1}+e^{\eta_2})^2}$$

Equating to zero yields

$$x_0 = \frac{-\alpha_2}{2\beta_1}$$

By applying standard calculus technique one can show that x_0 is a global maximum of $P(S=1)$. ■

The term x_0 is important in obtaining D-optimal designs and to show the symmetry properties for this model.

Property 3

$$P(S=1; x=x_0) = \frac{1}{1+\sqrt{\Omega}}$$

Proof.

$$P(S=1; x=x_0) = \frac{e^{\alpha_1+\beta_1 x_0}}{1+e^{\alpha_1+\beta_1 x_0}+e^{\alpha_2+2\beta_1 x_0}} = \frac{e^{\alpha_1-\frac{\alpha_2}{2}}}{2+e^{\alpha_1-\frac{\alpha_2}{2}}} = \frac{1}{1+\sqrt{\Omega}}$$

■

The value of $P(S=1; x=x_0)$ depends only on the odds ratio Ω. Thus a very large log-odds ratio gives a very small $P(S=1; x=x_0)$ and vice versa.

The following Property shows that $P(S=0)$ and $P(S=2)$ are symmetric around x_0 in the sense that $P(S=0; x=x_0-k) = P(S=2; x=x_0+k)$.

Property 4

$$P(S=0; x=x_0-k) = P(S=2; x=x_0+k) \quad \textit{for all } k$$

Proof.

$$P(S=0; x=x_0-k) = \frac{1}{1+e^{\alpha_1+\beta_1(x_0-k)}+e^{\alpha_2+2\beta_1(x_0-k)}} = \frac{1}{1+e^{\alpha_1-\frac{\alpha_2}{2}-k\beta_1}+e^{-2k\beta_1}}$$

$$P(S=2; x=x_0+k) = \frac{e^{\alpha_2+2\beta_1(x_0+k)}}{1+e^{\alpha_1+\beta_1(x_0+k)}+e^{\alpha_2+2\beta_1(x_0+k)}} = \frac{1}{1+e^{\alpha_1-\frac{\alpha_2}{2}-k\beta_1}+e^{-2k\beta_1}}$$

Hence $P(S=0)$ and $P(S=2)$ are symmetric around x_0. ■

For the current model S_1 and S_2 are not independent in general. Expressions for the covariance and the correlation between S_1 and S_2 are derived below.

Property 5

$$Cov\,(S_1, S_2) = \frac{4e^{\eta_2} - e^{2\eta_1}}{4\,(1+e^{\eta_1}+e^{\eta_2})^2}$$

Proof.

$$\begin{aligned} Cov\left(S_1, S_2\right) &= \frac{e^{\eta_2}}{1+e^{\eta_1}+e^{\eta_2}} - \pi_{1\cdot}\pi_{\cdot 1} \\ &= \frac{4e^{\eta_2} - e^{2\eta_1}}{4\left(1+e^{\eta_1}+e^{\eta_2}\right)^2}, \end{aligned}$$

since $\beta_2 = 2\beta_1$. ■

Property 6

$$Corr\left(S_1, S_2\right) = \frac{4e^{\alpha_2} - e^{2\alpha_1}}{\left(2+e^{\alpha_1+\beta_1 x}\right)\left(2e^{\alpha_2}+e^{\alpha_1-\beta_1 x}\right)}$$

Proof.

$$\begin{aligned} Corr\left(S_1, S_2\right) &= \frac{Cov\left(S_1, S_2\right)}{\sqrt{\pi_{1\cdot}\pi_{0\cdot}}\sqrt{\pi_{\cdot 1}\pi_{\cdot 0}}} \\ &= \frac{\frac{4e^{\eta_2}-e^{2\eta_1}}{4\left(1+e^{\eta_1}+e^{\eta_2}\right)^2}}{\frac{\left(2+e^{\eta_1}\right)\left(2e^{\eta_2}+e^{\eta_1}\right)}{4\left(1+e^{\eta_1}+e^{\eta_2}\right)^2}} \\ &= \frac{4e^{\alpha_2} - e^{2\alpha_1}}{\left(2+e^{\alpha_1+\beta_1 x}\right)\left(2e^{\alpha_2}+e^{\alpha_1-\beta_1 x}\right)}, \end{aligned}$$

since $\beta_2 = 2\beta_1$. ■
The variance of S is given by

Property 7

$$Var\left(S\right) = \frac{4e^{\eta_2} + e^{\eta_1}\left(1+e^{\eta_2}\right)}{\left(1+e^{\eta_1}+e^{\eta_2}\right)^2}.$$

Proof.

$$\begin{aligned} Var\left(S\right) &= \pi_{1\cdot}\pi_{0\cdot} + \pi_{\cdot 1}\pi_{\cdot 0} + 2\frac{4e^{\eta_2} - e^{2\eta_1}}{4\left(1+e^{\eta_1}+e^{\eta_2}\right)^2} \\ &= \frac{4e^{\eta_2} + e^{\eta_1}\left(1+e^{\eta_2}\right)}{\left(1+e^{\eta_1}+e^{\eta_2}\right)^2} \end{aligned}$$

■
Consider the following expression for the covariance between S_1 and S_2.

$$Cov\left(S_1, S_2\right) = \sum_{s_1=0,1}\sum_{s_2=0,1}\left(s_1 - \pi_{1\cdot}\right)\left(s_2 - \pi_{\cdot 1}\right)P\left(S_1 = s_1, S_2 = s_2\right)$$

It was stated previously that $P(S=0)$ is a decreasing function in x and $P(S=2)$ is an increasing function in x if $\beta_1 > 0$. For very small x, $P\left(S_1 = s_1, S_2 = s_2\right)$ is close to zero for all s_1, s_2 except for $s_1 = s_2 = 0$. Since $\pi_{1\cdot}$ is close to zero for very small x, $Cov\left(S_1, S_2\right)$ is close to zero for very small x. In a similar way $Cov\left(S_1, S_2\right)$

is close to zero for very large x. Hence, the correlation tends to zero when x tends to minus or plus infinity.

Figure 5 includes four plots of the correlation between S_1 and S_2 for the same parameter values as in Figure 4. In Plot 1 and Plot 2 the parameter values generate a negative log-odds ratio, hence the correlation is also negative. Plot 4 is based on parameter values that generate a large log-odds ratio. The correlation is therefore close to one for values of x in the interval -30 to 30. In all plots the correlation tends to zero for very large and for very small values on x.

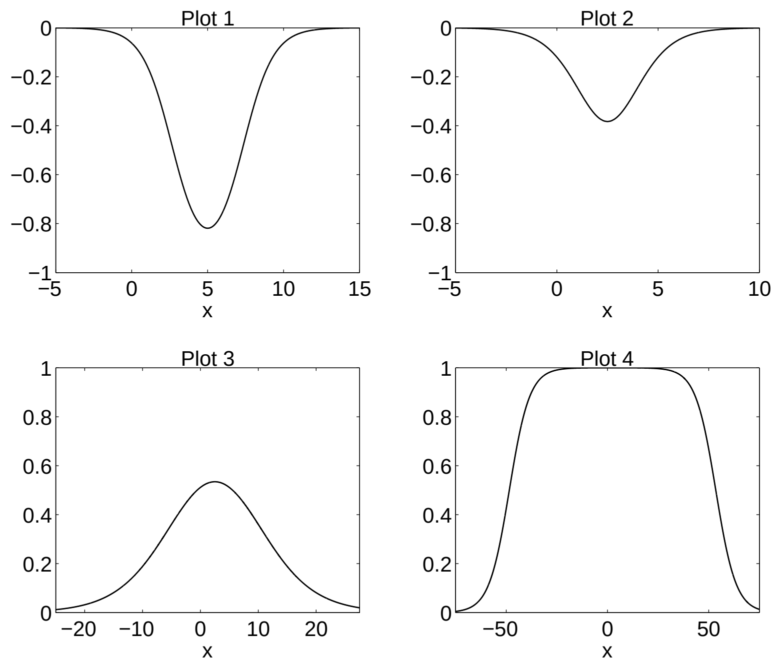


Figure 5: Correlation between S_1 and S_2 for different sets of parameter values. The parameters are $\theta_1^T = (-2, -10, 1), \theta_2^T = (-1, -5, 1), \theta_3^T = (-1, -1, 0.2)$, and $\theta_4^T = (-10, -1, 0.2)$, respectively. The corresponding log-odds ratios are $\ln\Omega_1 = -4.61$, $\ln\Omega_2 = -1.61$, $\ln\Omega_3 = 2.39$, and $\ln\Omega_4 = 20.39$, respectively.

Property 8 *$Corr\,(S_1, S_2)$ has a global minimum or maximum at $x = x_0$*

Proof.

$$\frac{d}{dx}Corr\,(S_1, S_2) = \frac{\beta_1 \left(e^{2\alpha_1} - 4e^{\alpha_2}\right)\left(e^{\alpha_1 - \beta_1 x}\left(2 + e^{\alpha_1 + \beta_1 x}\right) - e^{\alpha_1 + \beta_1 x}\left(2e^{\alpha_2} + e^{\alpha_1 - \beta_1 x}\right)\right)}{\left(2 + e^{\alpha_1 + \beta_1 x}\right)^2 \left(2e^{\alpha_2} + e^{\alpha_1 - \beta_1 x}\right)^2}$$

Equating to zero yields

$$x = x_0 = \frac{-\alpha_2}{2\beta_1}.$$

By applying standard calculus technique one can show that x_0 is a global minimum or maximum of $Corr\left(S_1, S_2\right)$. ■

The conditional probability that $S_1 = 1$ given that $S_2 = 1$ is derived below.

Property 9

$$P\left(S_1 = 1 \mid S_2 = 1\right) = \frac{2}{2 + e^{\alpha_1 - \alpha_2 - \beta_1 x}}$$

Proof.

$$\begin{aligned} P\left(S_1 = 1 \mid S_2 = 1\right) &= \frac{P\left(S_1 = 1, S_2 = 1\right)}{P\left(S_2 = 1\right)} \\ &= \frac{\frac{e^{\eta_2}}{1 + e^{\eta_1} + e^{\eta_2}}}{\frac{2e^{\eta_2} + e^{\eta_1}}{2\left(1 + e^{\eta_1} + e^{\eta_2}\right)}} \\ &= \frac{2}{2 + e^{\alpha_1 - \alpha_2 - \beta_1 x}} \end{aligned}$$

■

If β_1 is positive the conditional probability that $S_1 = 1$ given that $S_2 = 1$ tends to one when x tends to infinity and to zero when x tends to minus infinity. Figure 6 presents the conditional probability, $P\left(S_1 = 1 \mid S_2 = 1\right)$, for the same parameter values as in Figure 4 and Figure 5.
Assuming that β_1 is positive, $P(S = 2)$ is an increasing function in x and consequently $P\left(S_1 = 1 \mid S_2 = 1\right)$ increases with x. This property is illustrated in Figure 6. The conditional probability in Plot 4 is close to one for so small values on x as -30. This is explained by a strong dependence between S_1 and S_2. Plot 1 and Plot 2 are generated with a larger β_1 compared with Plot 3 and Plot 4. Therefore the conditional distribution increases more rapidly in Plot 1 and Plot 2.

Since the distribution of the response variable for the modified Cox model belongs to the exponential family, it follows immediately that also the response variable under the symmetric model is an exponential family. The likelihood function is basically the same likelihood function as for the modified Cox model,

$$L\left(\theta|\mathbf{y}\right) = \prod_{i=1}^{N} \left\{\pi_{1i}^{y_{1i}} \pi_{2i}^{y_{2i}} \left(1 - \pi_{1i} - \pi_{2i}\right)^{(1-y_{1i})(1-y_{2i})}\right\}.$$

The loglikelihood function is

$$l\left(\theta|\mathbf{y}\right) = \sum_{i=1}^{N} \left(y_{1i}\eta_{1i} + y_{2i}\eta_{2i} - \ln\left(1 + e^{\eta_{1i}} + e^{\eta_{2i}}\right)\right).$$

Also the score function is similar to the score function under the modified Cox model. The only difference is that the matrix $\mathbf{x}$ is different.

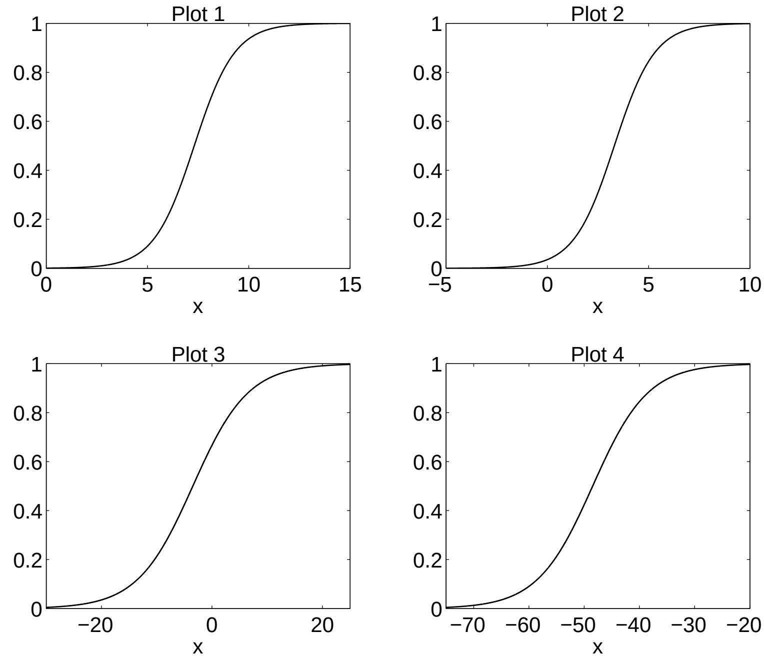


Figure 6: $P(S_1 = 1 \mid S_2 = 1)$ for different sets of parameter values. The parameters are $\theta_1^T = (-2, -10, 1), \theta_2^T = (-1, -5, 1), \theta_3^T = (-1, -1, 0.2)$, and $\theta_4^T = (-10, -1, 0.2)$, respectively. The corresponding log-odds ratios are $\ln\Omega_1 = -4.61$, $\ln\Omega_2 = -1.61$, $\ln\Omega_3 = 2.39$, and $\ln\Omega_4 = 20.39$, respectively.

$$\begin{pmatrix} u_{\alpha_1 \cdot}(\theta) \\ u_{\alpha_2 \cdot}(\theta) \\ u_{\beta_1 \cdot}(\theta) \end{pmatrix} = \sum_{i=1}^{N} \begin{pmatrix} (y_{1i} - \pi_{1i}) \\ (y_{2i} - \pi_{2i}) \\ x_i (y_{1i} - \pi_{1i} + 2y_{2i} - \pi_{2i}) \end{pmatrix} = \sum_{i=1}^{N} \mathbf{x}_i^T (y_i - \mu_i)$$

The Fisher information for one observation is partly the same compared with the modified Cox model.

$$I(\theta, x) = \mathbf{x}^T D \mathbf{x}$$

$$= \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & x(\pi_1(1-\pi_1) - 2\pi_1\pi_2) \\ -\pi_1\pi_2 & \pi_2(1-\pi_2) & x(2\pi_2(1-\pi_2) - \pi_1\pi_2) \\ x(\pi_1(1-\pi_1) - 2\pi_1\pi_2) & x(2\pi_2(1-\pi_2) - \pi_1\pi_2) & x^2(\pi_1(1-\pi_1) - 4\pi_1\pi_2 + 4\pi_2(1-\pi_2)) \end{pmatrix}$$

5.1 Locally D-optimal designs

The example of a D-optimal design for the modified Cox model indicates that there are no symmetry properties in the design. For instance, no design weights are equal. For the symmetric model in this chapter the number of design points and the design weights change with different parameter values. Nevertheless D-optimal designs have some symmetry properties under this model.

#	α_1	α_2	β_1	$\ln\Omega$	x_0	D-optimal design
1	-2	-10	1	-4.61	5	$\left\{\begin{array}{cccc} 1.176 & 3.680 & 6.320 & 8.824 \\ 0.269 & 0.231 & 0.231 & 0.269 \end{array}\right\}$
2	-1	-5	1	-1.61	2.5	$\left\{\begin{array}{ccc} 0.132 & 2.5 & 4.868 \\ 0.255 & 0.49 & 0.255 \end{array}\right\}$
3	-1	-1	0.2	2.39	2.5	$\left\{\begin{array}{cc} -1.487 & 6.487 \\ 0.5 & 0.5 \end{array}\right\}$
4	-10	-1	0.2	20.39	2.5	$\left\{\begin{array}{cc} -0.889 & 5.889 \\ 0.5 & 0.5 \end{array}\right\}$

Table 2: D-optimal design for different sets of parameter values. The parameter values are the same as in Figure 4, Figure 5 and Figure 6.

Example 1 Four D-optimal designs are presented in Table 2.
Although the number of design points are different for the designs there are some general properties. The design points are placed symmetric around x_0. In a plot over $d(x,\xi)$, the function should have maximum points at the design points. $d(x,\xi)$ at these maximum points should also be equal to 3. These properties are illustrated in Figure 7 where it is also possible to see x_0 as a local minimum or maximum point.

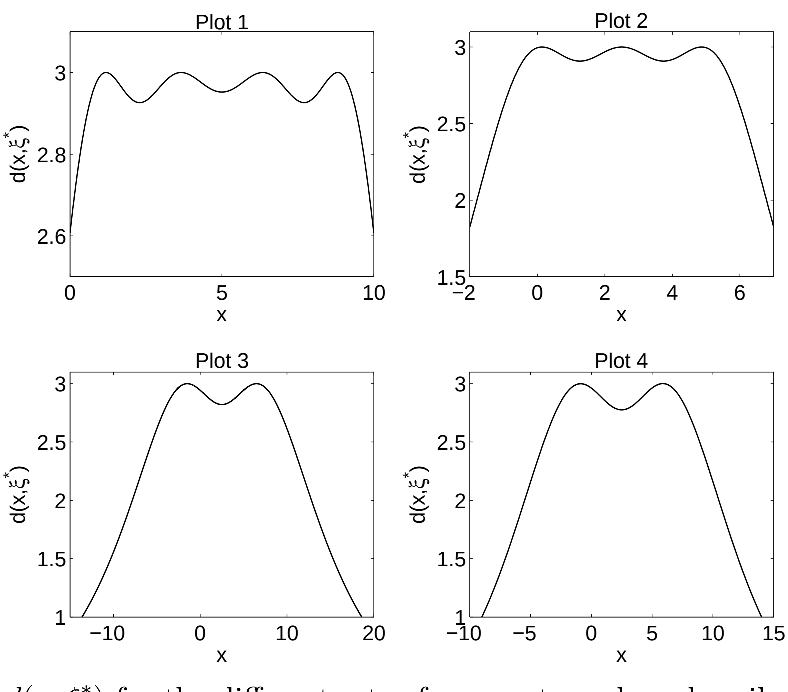


Figure 7: $d(x,\xi^*)$ for the different sets of parameter values described in Table 2.

The results in the example are in line with Fan (1999) and Fan and Chaloner (2004). They also found optimal designs with two, three or four design points although they had a different model. Fan (1999) argues that the differences in the number of design points and design weights can be explained by certain differences and ratios between parameters. The differences in the number of design points can also be explained using the log-odds ratio. Several plots have shown that the value of the log-odds ratio determines the number of design points in the D-optimal design. For example when the log-odds ratio is large it is sufficient with two design points. This is because it is sufficient to gather most information about the model using only two design points. When the log-odds ratio decreases, $P(S = 1; x = x_0)$ increases and a design with two design points is no longer optimal. The design points giving the most information about $P(S = 0)$ do not give any information about $P(S = 2)$ and vice versa. Therefore the optimal design contains three or more design points.

The relationship between the log-odds ratio and the number of design points is illustrated in Figure 8. The result has been derived by determining the D-optimal design for a number of different values on the parameters. Given two vectors of values on α_1 and α_2, the log-odds ratio is determined for each combination of (α_1, α_2). Depending on the log-odds ratio, a $2-$point, $3-$point, or $4-$point design is then derived. The optimality of the proposed design is checked by calculating $\max(d(x, \xi))$.

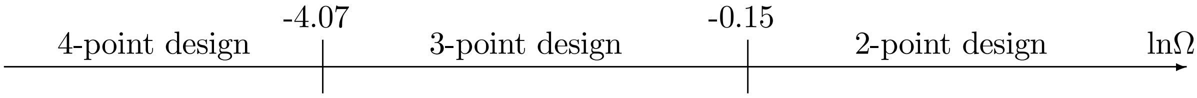


Figure 8: Number of design points given the log-odds ratio.

Using Figure 8 as a starting point it is possible to derive some more results on D-optimal designs.

5.1.1 Locally D-optimal designs when the log-odds ratio is large

For models with a large log-odds ratio it is reasonable to assume that a D-optimal design has two design points. A design is in general given by

$$\xi = \left\{ \begin{array}{cc} x_0 - \frac{c}{\beta_1} & x_0 + \frac{c}{\beta_1} \\ 0.5 & 0.5 \end{array} \right\}.$$

The design points are placed symmetric around x_0 with equal design weights. Equal design weights are assumed because $P(S = 0)$ and $P(S = 2)$ are symmetric around x_0. The question is if there is a way to find an expression for c. In this section the particular case when $\alpha_2 = 0$ and $\beta_1 = 1$ is examined. The restriction on α_2 and β_2 simplifies the coming calculations considerable. The standardized information matrix for a $2-$point design under the restriction is

$$M(\alpha_1, c) = \frac{1}{2}\left(I(\alpha_1, -c) + I(\alpha_1, c)\right).$$

The determinant of M is

$$|M(\alpha_1, c)| = \frac{c^2 e^{\alpha_1 - 6c}\left(e^{\alpha_1} + e^{\alpha_1+2c} + 4e^c\right)}{\left(1 + e^{\alpha_1 - c} + e^{-2c}\right)^5}$$

Using this expression the derivative of the determinant of M with respect to c is

$$\begin{aligned}\frac{d|M(\alpha_1, c)|}{dc} &= \{ce^{\alpha_1-4c}[2\left(1+e^{\alpha_1-c}+e^{-2c}\right)\left[e^{-2c}\left(e^{\alpha_1}+e^{\alpha_1+2c}+4e^c\right)-c\left(3e^{\alpha_1-2c}+2e^{\alpha_1}+10e^{-c}\right)\right] \\ &\quad +5ce^{-3c}\left(e^{\alpha_1}+2e^{-c}\right)\left(e^{\alpha_1}+e^{\alpha_1+2c}+4e^c\right)]\}/\left(1+e^{\alpha_1-c}+e^{-2c}\right)^6.\end{aligned}$$

There is no general expression for c. Nevertheless an asymptotic result can be derived. If $\alpha_2 = 0$ then

$$\ln\Omega \to \infty \quad \text{when} \quad \alpha_1 \to -\infty.$$

Setting

$$\frac{d|M(\alpha_1, c)|}{dc} = 0$$

yields,

$$\begin{aligned}2\left(1+e^{\alpha_1-c}+e^{-2c}\right)\left[e^{-2c}\left(e^{\alpha_1}+e^{\alpha_1+2c}+4e^c\right)-c\left(3e^{\alpha_1-2c}+2e^{\alpha_1}+10e^{-c}\right)\right] \\ +5ce^{-3c}\left(e^{\alpha_1}+2e^{-c}\right)\left(e^{\alpha_1}+e^{\alpha_1+2c}+4e^c\right)] = 0.\end{aligned}$$

Let $\alpha_1 \to -\infty$ and it follows that

$$\begin{aligned}c &= \frac{2\left(1+e^{-2c}\right)}{5\left(1-e^{-2c}\right)}. \\ c &\approx 0.6778\end{aligned}$$

In Figure 9 the value of c is plotted as a function of the log-odds ratio. For the same c and the same log-odds ratio

$$\max_{x\in\mathfrak{X}} d(x, \xi(c))$$

is plotted. $d(x, \xi(c))$ is used since it is an easy way to verify if a design is D-optimal or not. In the plot over c note that when the log-odds ratio becomes smaller c increases. When the log-odds ratio is less than -0.15 a $2-$point design is no longer optimal. That is why $d(x, \xi(c))$ is larger than three when the log-odds ratio is less than -0.15.

For a log-odds ratio of approximately 10 and larger c is constant. This value of c is around 0.6778. One of the examples shown previously had a log-odds ratio of around 20. It is possible to verify that c is around 0.6778 in that example[1].

[1] In the example $a_1 = -10$ and $a_2 = -1$. This gives a log-odds ratio, $ln\Omega = ln4 + 19$. The symmetry point $x_0 = \frac{1}{0.4} = 2.5$. The design points are located $5.889 - 2.5 = 3.389$ from x_0. So in this example c is equal to $3.389\beta_1 \approx 0.6778$.

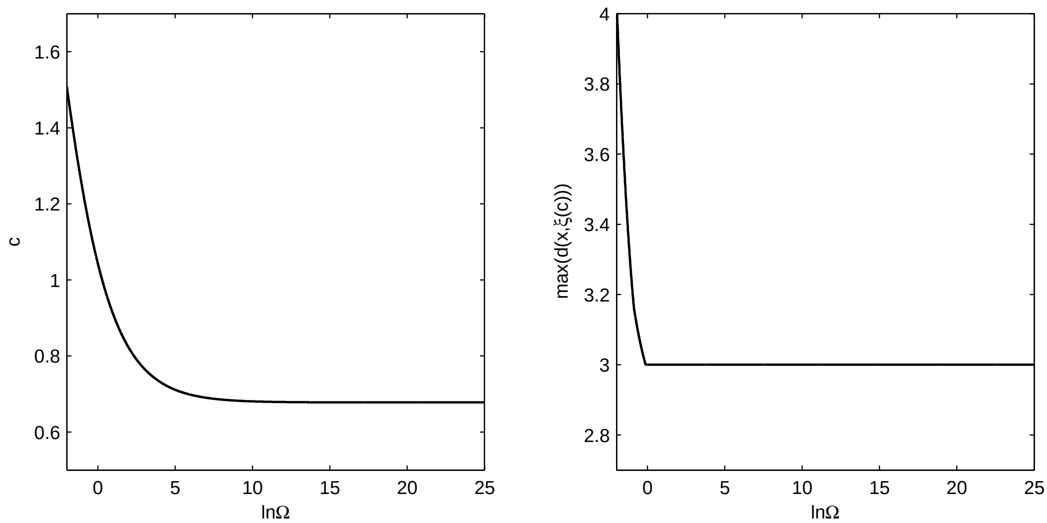


Figure 9: a) The value of c that maximizes $\mid M(\alpha_1, \xi) \mid$ for different values of the log-odds ratio. b) Maximum of $d(x, \xi(c))$ as a function of both c and the log-odds ratio.

Designs based on $c = 0.6778$ can be evaluated using D-efficiency, see Atkinson and Donev (1992).

$$D_{eff} = \left(\frac{|M(\theta, \xi(c))|}{|M(\theta, \xi^*)|} \right)^{\frac{1}{p}}$$

Figure 10 presents the D-efficiency for designs with $c = 0.6778$ given different parameter values (different log-odds ratios).

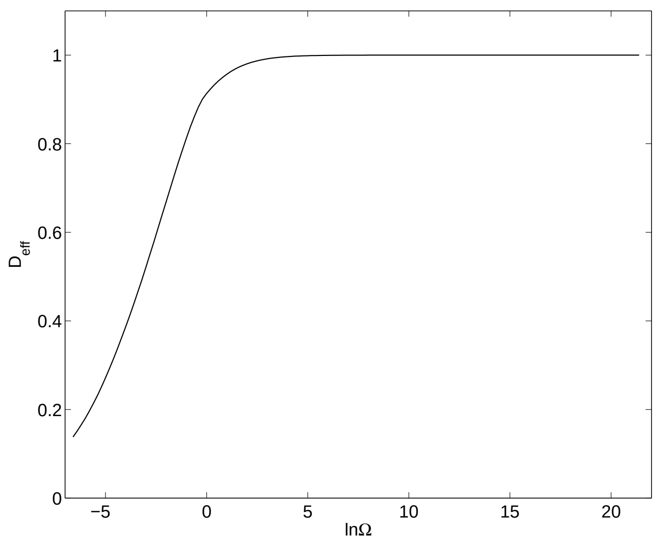


Figure 10: D-efficiency for designs with c=0.6778 for different parameter values.

For parameter values with a log-odds ratio of approximately larger than five the designs are optimal or almost optimal. When the log-odds ratio is negative a 2−point design is not optimal and the D-efficiency decreases rapidly.

5.1.2 Locally D-optimal designs when the log-odds ratio is large negative

For parameters with a large negative log-odds ratio a 4−point design is optimal. A plot of the probability distribution of S as a function of x for such parameter values is given in Figure 11.

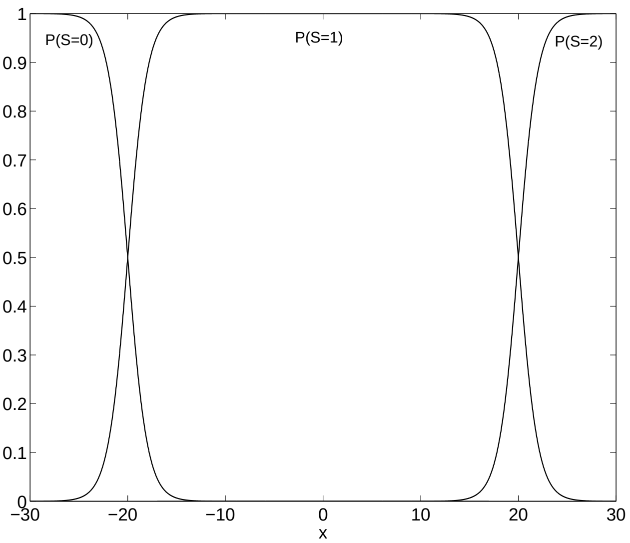


Figure 11: An example of the probability distribution of S when the log-odds ratio is large negative. The parameters are $\theta^T = (20, 0, 1)$ and the log-odds ratio is $\ln\Omega = -38.61$.

Based on Figure 11 it is reasonable that the design points are located around two symmetry points. These symmetry points are where $P(S = 0)$ equals $P(S = 1)$ and where $P(S = 1)$ equals $P(S = 2)$. When $P(S = 0)$ equals $P(S = 1)$, $\eta_1 = 0$ and hence

$$\alpha_1 + \beta_1 x = 0 \quad \text{so that} \quad x = \frac{-a_1}{\beta_1}.$$

In the same way, when $P(S = 1)$ equals $P(S = 2)$, $\eta_1 = \eta_2$ and hence

$$\alpha_1 + \beta_1 x = \alpha_2 + 2\beta_1 x \quad \text{so that} \quad x = \frac{a_1 - \alpha_2}{\beta_1}.$$

The proposed 4−point design is therefore

$$\xi = \left\{ \begin{array}{cccc} \frac{-\alpha_1}{\beta_1} - \frac{c}{\beta_1} & \frac{-\alpha_1}{\beta_1} + \frac{c}{\beta_1} & \frac{\alpha_1-\alpha_2}{\beta_1} - \frac{c}{\beta_1} & \frac{\alpha_1-\alpha_2}{\beta_1} + \frac{c}{\beta_1} \\ 0.25 & 0.25 & 0.25 & 0.25 \end{array} \right\}. \tag{3}$$

This 4−point design has a more complex expression for the determinant of M compared with the 2−point design. In the particular case when $\alpha_2 = 0$ and $\beta_1 = 1$ one can numerically find the value of c that maximizes $|M|$. M is then

$$M(\alpha_1, c) = \frac{1}{4}\left\{I(-\alpha_1, -c) + I(-\alpha_1, c) + I(\alpha_1, -c) + I(\alpha_1, c)\right\}.$$

In Figure 12 the value of c that maximizes $|M(\alpha_1, c)|$ is plotted versus $\ln \Omega$.
For designs with a log-odds ratio less than -10, $c \approx 1.2229$. A design based on (3) with $c = 1.2229$ is an approximation of the locally D-optimal design. The approximation is evaluated using D-efficiency. The D-efficiency for different parameter values are shown in Figure 13.

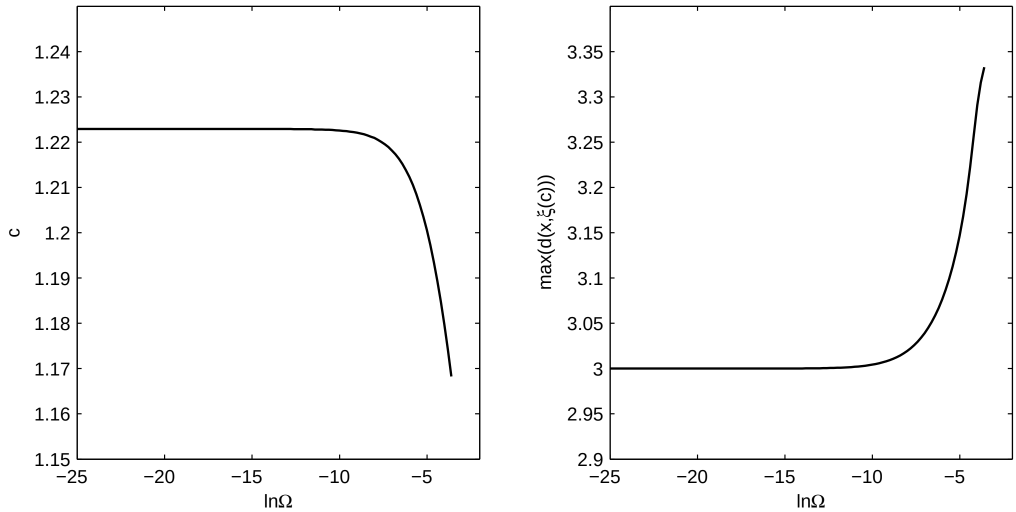


Figure 12: a) The value of c that maximizes $| M(\alpha_1, \xi) |$ for different values of the log-odds ratio. b) Maximum of $d(x, \xi(c))$ as a function of both c and the log-odds ratio.

If the log-odds ratio is approximately -10 or smaller $c = 1.2229$ yields very efficient designs. A 3−point design is D-optimal when the value of the log-odds ratio is between -4.07 and -0.15. Nevertheless the D-efficiency for the proposed design does not decrease when the log-odds ratio is between -4.07 and -0.15. The explanation is that two design points are almost equal for the proposed 4−point design. Hence, the proposed design has almost the same D-efficiency as a D-optimal 3−point design. For positive log-odds ratios the D-efficiency decreases fast.

6 Bivariate independent model

In this chapter it is assumed that S_1 and S_2 are independent. Independence between S_1 and S_2 is equivalent to a log-odds ratio that is zero. Given the expression for the log-odds ratio an independent model has the restrictions

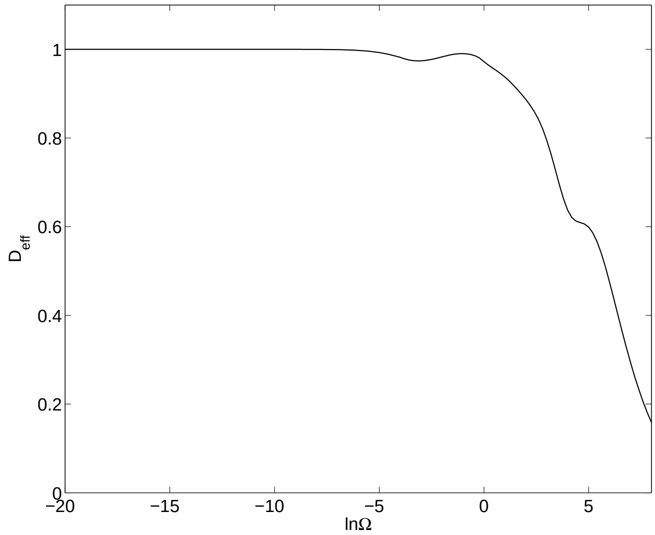


Figure 13: D-efficiency for designs with c=1.229 for different parameter values.

- $\alpha_2 = 2\alpha_1 - \ln 4$,
- $\beta_2 = 2\beta_1$.

Because S_1 and S_2 are independent the distribution for S is simply

$$S \sim Bin\,(2, \pi)\,,$$

where π is the common probability for observing a "success". In this model S is the response variable, and using the restrictions for the model

$$\pi = \frac{e^{\eta_1}}{2 + e^{\eta_1}}.$$

The variance of S is

$$Var(S) = 2\pi\,(1 - \pi)\,.$$

Figure 14 presents an example of the probability distribution for S as a function of x.
In the remaining part of this chapter different properties of the probability distribution for S are analyzed. As for the previous model x_0 is important when deriving symmetry properties for this model.

Property 10

$$x_0 = \frac{\ln 2 - \alpha_1}{\beta_1}$$

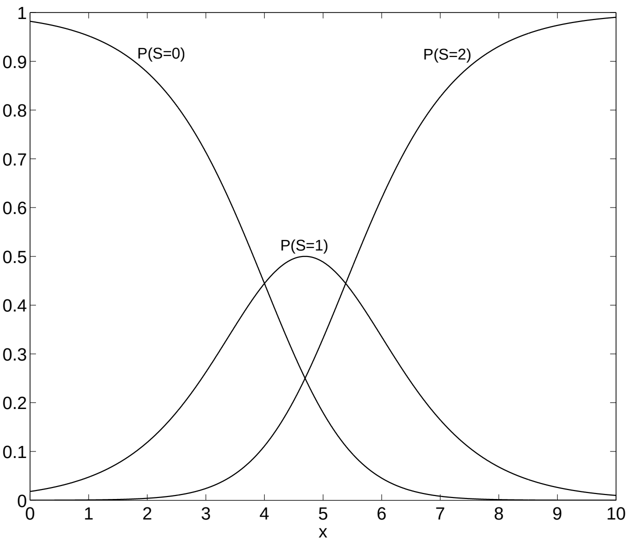


Figure 14: Probability distribution for S with parameters $\alpha_1 = -4$ and $\beta_1 = 1$.

Proof. Differentiate $P(S = 1)$ with respect to x.

$$\frac{dP(S=1)}{dx} = \frac{8\beta_1 e^{\eta_1} - 4\beta_1 e^{2\eta_1}}{(2 + e^{\eta_1})^3}$$

Equating to zero yields

$$2 = e^{\eta_1}$$

and hence

$$x_0 = \frac{\ln 2 - \alpha_1}{\beta_1}.$$

By applying standard calculus technique it is easy to verify that x_0 is a global maximum of $P(S = 1)$. ■

Property 11

$$P(S = 1; x = x_0) = \frac{1}{2}$$

Proof.

$$P(S = 1; x = x_0) = \frac{4e^{\eta_1}}{(2 + e^{\eta_1})^2} = \frac{1}{2}$$

■

Hence the maximum value of $P(S = 1)$ is always equal to $\frac{1}{2}$ in this model.
As for the previous model, $P(S = 0)$ and $P(S = 2)$ are symmetric around x_0 in the sense that $P(S = 0; x = x_0 - k) = P(S = 2; x = x_0 + k)$

Property 12

$$P(S = 0; x = x_0 - k) = P(S = 2; x = x_0 + k) \quad \textit{for all } k$$

Proof.

$$P(S=0;x=x_0-k)=\frac{4}{\left(2+e^{\alpha_1+\beta_1(x_0-k)}\right)^2}=\frac{1}{\left(1+e^{-k\beta_1}\right)^2}$$

$$P(S=2;x=x_0+k)=\frac{e^{2(\alpha_1+\beta_1(x_0+k))}}{\left(2+e^{\alpha_1+\beta_1(x_0+k)}\right)^2}=\frac{1}{\left(1+e^{-k\beta_1}\right)^2}$$

Hence $P(S=0;x=x_0-k)=P(S=2;x=x_0+k)$. ∎

Since S has a binomial distribution the probability distribution for S belongs to the exponential family of distributions. The likelihood is well determined

$$L\left(\alpha_1,\beta_1|\mathbf{s}\right)=\prod_{i=1}^{N}\binom{2}{s_i}\pi_i^{s_i}\left(1-\pi_i\right)^{2-s_i},$$

$$\ln L\left(\alpha_1,\beta_1|\mathbf{s}\right)=\sum_{i=1}^{N}\left\{\ln\binom{2}{s_i}+s_i\ln\left(\frac{\pi_i}{(1-\pi_i)}\right)+2\ln\left(1-\pi_i\right)\right\}.$$

The score function is determined by applying the chain rule.

$$\begin{pmatrix}u_{\alpha_1\cdot}(\theta)\\ u_{\beta_1\cdot}(\theta)\end{pmatrix}=\sum_{i=1}^{N}\begin{pmatrix}(s_i-2\pi_i)\\ x_i\left(s_i-2\pi_i\right)\end{pmatrix}$$

The Fisher information is

$$\begin{aligned}I_{\cdot}(\theta,x)&=E\left[u(\theta)\,u^T(\theta)\right]\\&=\sum_{i=1}^{N}2\pi_i\left(1-\pi_i\right)\begin{pmatrix}1&x_i\\x_i&x_i^2\end{pmatrix}.\end{aligned}$$

Locally D-optimal designs for independent binary variables are easily found, see for example Kalish and Rosenberger (1978).

Theorem 1 *For the independent model in this chapter the locally D-optimal design is*

$$\xi^*=\left\{\begin{array}{cc}x_0-\frac{c}{\beta_1}&x_0+\frac{c}{\beta_1}\\0.5&0.5\end{array}\right\},$$

where c is the solution to the equation

$$c=\frac{e^c+1}{e^c-1}.$$

$$c\approx 1.5434$$

Proof. In order to show that the locally D-optimal design is given by ξ^* the value of c that maximizes $|M(\theta,\xi^*)|$ is first found. Second, it must be shown that the design that maximizes $|M(\theta,\xi^*)|$ is the locally D-optimal design. This is done by checking that maximum of $d(x,\xi^*)$ is achieved at the design points and that $d(x,\xi^*) \le 2$ for all x.
The standardized information matrix for the design ξ^* is

$$M(\theta,\xi^*) = \frac{2e^c}{(1+e^c)^2}\begin{pmatrix} 1 & \frac{\ln 2-\alpha_1}{\beta_1} \\ \frac{\ln 2-\alpha_1}{\beta_1} & \frac{1}{\beta_1^2}\left((\ln 2-\alpha_1)^2+c^2\right)\end{pmatrix}.$$

$$|M(\theta,\xi^*)| = \frac{4c^2e^{2c}}{\beta_1^2(1+e^c)^4}$$

$$\frac{d|M(\theta,\xi^*)|}{dc} = \frac{8ce^c(1+e^c+c-ce^c)}{\beta_1^2(5+10e^c+10e^{2c}+5e^{3c}+e^{4c})}$$

Equating to zero yields

$$c = \frac{e^c+1}{e^c-1}.$$

The solution to the equation, $c \approx 1.5434$ maximizes $|M(\theta,\xi^*)|$.
The standardized variance of the predicted response

$$d(x,\xi^*) = \left(\frac{\partial\mu}{\partial\theta}\right)^T M(\theta,\xi^*)^{-1}\left(\frac{\partial\mu}{\partial\theta}\right) Var(S),$$

where

$$\left(\frac{\partial\mu}{\partial\theta}\right) = Var(S)\begin{pmatrix}1\\x\end{pmatrix}$$

and

$$M(\theta,\xi^*)^{-1} = \frac{(1+e^c)^2}{2e^c}\frac{\beta_1}{c^2}\begin{pmatrix}\frac{1}{\beta_1^2}\left((\ln 2-\alpha_1)^2+c^2\right) & -\frac{\ln 2-\alpha_1}{\beta_1} \\ -\frac{\ln 2-\alpha_1}{\beta_1} & 1\end{pmatrix}$$

$$d(x,\xi^*) = \begin{pmatrix}1 & x\end{pmatrix}\begin{pmatrix}\frac{1}{\beta_1^2}\left((\ln 2-\alpha_1)^2+c^2\right) & -\frac{\ln 2-\alpha_1}{\beta_1} \\ -\frac{\ln 2-\alpha_1}{\beta_1} & 1\end{pmatrix}\begin{pmatrix}1\\x\end{pmatrix}\frac{4e^{\alpha_1+\beta_1 x}}{(2+e^{\alpha_1+\beta_1 x})^2}\frac{(1+e^c)^2}{2e^c}\frac{\beta_1^2}{c^2}$$

Which simplifies to

$$d(x,\xi^*) = \left[\left(\frac{\ln 2-\alpha_1}{\beta_1}-x\right)^2+\frac{c^2}{\beta_1^2}\right]\frac{4e^{\alpha_1+\beta_1 x}}{(2+e^{\alpha_1+\beta_1 x})^2}\frac{(1+e^c)^2}{2e^c}\frac{\beta_1^2}{c^2}.$$

Next it is shown that $d(x,\xi^*)$ achieves its maximum at the design points.

$$\frac{d}{dx}d(x,\xi^*) = \left\{\frac{4e^{\alpha_1+\beta_1 x}}{(2+e^{\alpha_1+\beta_1 x})^2}\left[2x-\frac{2(\ln 2-\alpha_1)}{\beta_1}\right]\right.$$

$$+\frac{4\beta_1 e^{\alpha_1+\beta_1 x}\left(2-e^{\alpha_1+\beta_1 x}\right)}{\left(2+e^{\alpha_1+\beta_1 x}\right)^3}\left[\left(\frac{\ln 2-\alpha_1}{\beta_1}-x\right)^2+\frac{c^2}{\beta_1^2}\right]\left\{\frac{\left(1+e^c\right)^2}{2e^c}\frac{\beta_1^2}{c^2}\right.$$

$$=\left\{\frac{\beta_1\left(2-e^{\alpha_1+\beta_1 x}\right)}{\left(2+e^{\alpha_1+\beta_1 x}\right)}\left[\left(\frac{\ln 2-\alpha_1}{\beta_1}-x\right)^2+\frac{c^2}{\beta_1^2}\right]\right.$$

$$+2x-\frac{2\left(\ln 2-\alpha_1\right)}{\beta_1}\frac{4e^{\alpha_1+\beta_1 x}}{\left(2+e^{\alpha_1+\beta_1 x}\right)^2}\left\{\frac{\left(1+e^c\right)^2}{2e^c}\frac{\beta_1^2}{c^2}\right.$$

Substituting $\ln 2-\alpha_1-\beta_1 x=\pm c$ and using the expression $c=\frac{e^c+1}{e^c-1}$ yields

$$\begin{aligned} d(x,\xi^*) &= \left\{\frac{\beta_1\left(1-e^c\right)}{\left(1+e^c\right)}\frac{2c^2}{\beta_1^2}+\frac{2c}{\beta_1}\right\}\frac{2e^c}{\left(1+e^c\right)^2}\frac{\left(1+e^c\right)^2}{2e^c}\frac{\beta_1^2}{c^2} \\ &= \left\{\frac{-2c}{\beta_1}+\frac{2c}{\beta_1}\right\}\frac{2e^c}{\left(1+e^c\right)^2}\frac{\left(1+e^c\right)^2}{2e^c}\frac{\beta_1^2}{c^2}=0 \end{aligned}$$

Hence, ξ^* satisfy the condition that $d(x,\xi^*)$ attains its maximum at the design points.
Next it is shown that $d(x,\xi^*)\le 2$. The maximum value of $d(x,\xi^*)$ is obtained by inserting $x=\frac{\ln 2-\alpha_1\pm c}{\beta_1}$ in $d(x,\xi^*)$,

$$d\left(\frac{\ln 2-\alpha_1\pm c}{\beta_1},\xi^*\right)=\frac{2c^2}{\beta_1^2}\frac{8e^c}{4\left(1+e^c\right)^2}\frac{\left(1+e^c\right)^2}{2e^c}\frac{\beta_1^2}{c^2}=2$$

Thus, the proposed design is locally D-optimal. ■

Example 1 For the parameters $\alpha_1=-4$ and $\beta_1=1$ a D-optimal design is given by

$$\xi^*=\left\{\begin{array}{cc} 3.1497 & 6.2365 \\ 0.5 & 0.5 \end{array}\right\}.$$

In Figure 15, $d(x,\xi^*)$ is plotted. Note that $d(x,\xi^*)$ has two maximum points appearing at the two design points 3.1497 and 6.2365. If ξ^* is a D-optimal design the criterion function, ψ, has minimum value for this design. However ξ^* is also optimal if the determinant of the standardized information matrix, $|M(\theta,\xi^*)|$, has maximum points at the design points. Figure 15 shows that this is true for the current example.
The plot shows $|M(\theta,\xi)|$ for the design

$$\xi=\left\{\begin{array}{cc} \ln 2+4-c & \ln 2+4+c \\ 0.5 & 0.5 \end{array}\right\}.$$

The maximum value of $|M(\theta,\xi)|$ is attained for $c\approx 1.543$. Thus the optimal design points are approximately 3.150 and 6.236.

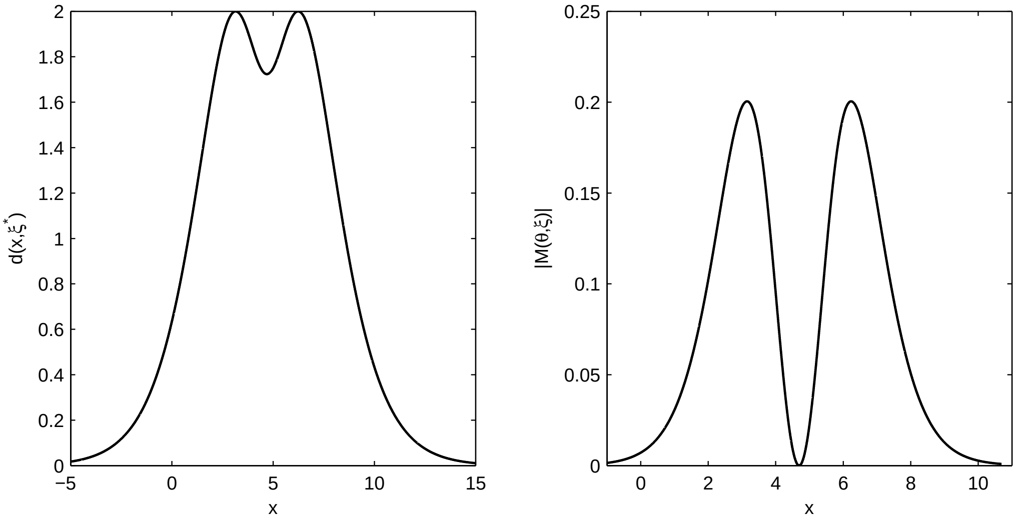


Figure 15: a) $d(x, \xi)$ for parameters $\alpha_1 = -4$ and $\beta_1 = 1$. b) Determinant of the information matrix for different choices of design points given the same parameters.

7 Testing for independency

The previous chapters show that a model for independent variables is less complex and thus easier to estimate. Locally D-optimal designs are also easier to derive. Test procedures for testing the null hypothesis that the variables S_1 and S_2 are independent are therefore of interest. In this thesis the score test and the likelihood ratio test are considered. This chapter is based on Bruce and Nyquist (2005).

7.1 Models without covariates

If no covariate is included in the model, the log likelihood as a function of π_0, π_1, and π_2 is

$$\ell(\pi; \mathbf{y}) = \sum_{i=1}^{N} y_{0i} \ln \pi_0 + y_{1i} \ln \pi_1 + y_{2i} \ln \pi_2, \tag{4}$$

$$y_{ji} = \begin{cases} 1, \text{ if } S_i = j \\ 0, \text{ otherwise} \end{cases} \quad \text{for } j = 0, 1, 2 \text{ and } i = 1, 2, \ldots, N.$$

The test statistic for the score test is defined as

$$T_S = u^T_{\cdot}(\widetilde{\pi})\, I^{-1}(\widetilde{\pi})\, u_{\cdot}(\widetilde{\pi}), \tag{5}$$

where $\widetilde{\pi} = (\widetilde{\pi}_1, \widetilde{\pi}_2)^T$ is the estimated vector of probabilities under the null hypothesis. The hypothesis of independence implies the restrictions $\pi_0 = (1 - \pi_{1\cdot})^2$, $\pi_1 = 2\pi_{1\cdot}(1 - \pi_{1\cdot})$, and $\pi_2 = \pi_{1\cdot}^2$, where $\pi_{1\cdot} = P(S_1 = 1) = P(S_2 = 1)$ is the marginal probability to observe a "success". Under the hypothesis of independence, the maximum likelihood estimator of $\pi_{1\cdot}$ is evidently

$$\widetilde{\pi}_{1\cdot} = (2N)^{-1} \sum_{i=1}^{N} (y_{1i} + 2y_{2i}) = \frac{r_1 + 2r_2}{2N}, \tag{6}$$

where r_j is the number of observed pairs that result in $y_j = 1$. Hence, the estimator $\widetilde{\pi}_{1\cdot}$ equals the total number of observed "successes" divided by the number of observed variables. Maximum likelihood estimators of π_0, π_1, and π_2 are accordingly

$$\widetilde{\pi}_0 = (1 - \widetilde{\pi}_{1\cdot})^2, \ \widetilde{\pi}_1 = 2\widetilde{\pi}_{1\cdot}(1 - \widetilde{\pi}_{1\cdot}), \text{ and } \widetilde{\pi}_2 = \widetilde{\pi}_{1\cdot}^2. \tag{7}$$

By deriving expressions for the scores and the information matrix from (4) and inserting these expressions in (5), the score test statistic becomes

$$T_S = \sum_{j=0}^{2} \frac{(r_j - N\widetilde{\pi}_j)^2}{N\widetilde{\pi}_j}. \tag{8}$$

The test statistic coincides with the χ^2-test statistic for testing the goodness of fit of a trinomial distribution with probabilities restricted as described above. Asymptotically, T_S has a χ^2 distribution with 1 degree of freedom, the approximation being good provided the expected frequencies, $N\widetilde{\pi}_j$, $j = 0, 1, 2$, are sufficiently large.
The test statistic for the likelihood ratio test is defined as

$$\begin{aligned} T_{LR} &= 2\left(\ell(\widehat{\pi}; \mathbf{y}) - \ell(\widetilde{\pi}; \mathbf{y})\right) \\ &= 2\sum_{j=0}^{2} r_j \ln \frac{\widehat{\pi}_j}{\widetilde{\pi}_j}, \end{aligned}$$

where $\widehat{\pi}_j = \frac{r_j}{N}$ is the unrestricted maximum likelihood estimator of π_j, $j = 0, 1, 2$.

This simple case generalizes straightforwardly to the case with several, say K, groups with N_k observations in each group. The distribution of the trinomial response vector in each group is here defined by the vector $(\pi_{0k}, \pi_{1k}, \pi_{2k})^T$, $k = 1, 2, \ldots, K$, of probabilities. The test statistic for the score test now becomes

$$T_S = \sum_{k=1}^{K} \sum_{j=0}^{2} \frac{(r_{jk} - N_k\widetilde{\pi}_{jk})^2}{N_k\widetilde{\pi}_{jk}}, \tag{9}$$

where r_{jk} is the observed frequency of category j, $j = 0, 1, 2$, in group k, $k = 1, 2, \ldots, K$,

$$\widetilde{\pi}_{0k} = (1 - \widetilde{\pi}_{1\cdot k})^2, \widetilde{\pi}_{1k} = 2\widetilde{\pi}_{1\cdot k}(1 - \widetilde{\pi}_{1\cdot k}), \widetilde{\pi}_{2k} = \widetilde{\pi}_{1\cdot k}^2, \tag{10}$$

and

$$\widetilde{\pi}_{1\cdot k} = (r_{1k} + 2r_{2k}) / (2N_k). \tag{11}$$

Similarly, the test statistic for the likelihood ratio test becomes

$$T_{LR} = 2\sum_{k=1}^{K}\sum_{j=0}^{2} r_{jk} \ln \frac{\widehat{\pi}_{jk}}{\widetilde{\pi}_{jk}}. \tag{12}$$

where the unrestricted estimator is

$$\widehat{\pi}_{jk} = \frac{r_{jk}}{N_k}.$$

The test statistics T_S and T_{LR} are asymptotically equivalent and has a χ^2 distribution with K degrees of freedom, asymptotically. Here it is important for the approximation to be good that each $N_k\widetilde{\pi}_{jk}$ is sufficiently large.

7.2 Models with covariates

A more structured model is obtained if the vector of probabilities π is governed by a vector of explanatory variables x. Assume that a modified bivariate Cox model as described in Chapter 3 is used. The vector valued linear predictor, $\eta = (\eta_1, \eta_2)^T$ is then

$$\eta_j = x_j^T \theta_j,\ \ j = 1, 2,$$

where x_j and θ_j are vectors of explanatory variables and associated parameters used for determining the probability π_j. Denoting the maximum likelihood estimator of the parameter vector θ under H_0 by $\widetilde{\theta}$, the score test statistic becomes

$$T_S = u^T.\left(\widetilde{\theta}\right) I^{-1}\left(\widetilde{\theta}\right) u.\left(\widetilde{\theta}\right). \tag{13}$$

The test statistic can be calculated using the previously derived expressions for the score vector and the information matrix.
Denote the maximum likelihood estimator without restrictions by $\widehat{\theta}$. The likelihood ratio test statistic is then obtained by evaluating the log likelihood function at $\widetilde{\theta}$ and $\widehat{\theta}$

$$T_{LR} = 2\left(l\left(\widehat{\beta};\mathbf{y}\right) - l\left(\widetilde{\beta};\mathbf{y}\right)\right). \tag{14}$$

Both the expression for the likelihood under H_0 and the expression for the unrestricted likelihood have been given above. The score test statistic and the likelihood ratio test statistic have the same χ^2-distribution asymptotically.

Suppose that data exist for K independent groups with N_k observations in each group. Then the two vectors of explanatory variables are identical and consist of dummy variables $x_1 = x_2 = (d_1, d_2, \ldots, d_K)^T$, where each d_k is either 1 or 0, indicating if an observation comes from response group k or not, respectively, $j = 1, 2, \ldots, K$. In this case the model reduces to the case with K response groups discussed above and the test statistics for independence are (9) and (12), respectively.

Example 1

The data for this example are taken from Liang et al. (1992). 5199 people are subject to a visual examination, measuring if the left eye and/or the right eye has a visual impairment or not. The outcome for each eye is binary, where " + " indicates visual impairment and " − " no visual impairment. Age is used as explanatory variable, see Table 3. In Table 3 there are, for example, 3627 out of 3958 people in age $40 - 70$ that have no visual impairment.

Left	Right	Age: 40 − 70	Age: 71+	Total
−	−	3627	913	4540
+	−	122	89	211
−	+	133	104	237
+	+	76	135	211
	Total	3958	1241	5199

Table 3: Joint distribution of visual impairment for both eyes, for the two age groups $40 - 70$ and over 70, respectively. Data are taken from Liang et al. (1992).

The probability that the left eye is visually impaired is assumed to be equal to the probability that the right eye is visually impaired. This assumption is reasonable since the risk of visual impairment (in percent) is similar for the left and the right eye in both groups.

Let S_1 and S_2 be Bernoulli variables for visual impairment of the left eye and the right eye, respectively. The elements of the response vector $y_i = (y_{1i}, y_{2i})^T$, $i = 1, 2, \ldots, 5199$, are the corresponding indicator variables. The vector of explanatory variables consists of the dummy variables d_1 and d_2 denoting the two age groups. The link function is therefore

$$\eta = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \beta_{11} d_1 + \beta_{12} d_2 \\ \beta_{21} d_1 + \beta_{22} d_2 \end{pmatrix}.$$

Suppose now that primary interest is in the possible dependency between S_1 and S_2. In this model S_1 and S_2 are independent if the parameter restrictions

$$\beta_{2j} = 2\beta_{1j}, \ j = 1, 2$$

are satisfied. As stated previously the score test statistic is given by (9). The test statistic has a χ^2 distribution with 2 degrees of freedom, asymptotically. The observed test statistic for the data material in Table 3 becomes, using (9),

$$T_S \approx 751.22.$$

Hence the hypothesis about independence is rejected since the critical value on 5% level is 5.991. The observed likelihood ratio test statistic is derived from (12). The value on the test statistic is

$$T_{LR} \approx 465.35,$$

so the hypothesis about independence is rejected when using the likelihood ratio test as well.

Another model is used when the linear predictors consist of an intercept and a single explanatory variable, z, the same variable in both linear predictors, so that $\eta_j = x^T\theta_j$, $x = (1, z)^T$ and $\theta_j = \left(\alpha_j, \beta_j\right)^T$, $j = 1, 2$. In this model, the explanatory variable z may influence the success probabilities π_1 and π_2 differently. In this situation S_1 and S_2 are independent if $\beta_2 = 2\beta_1$ and $\alpha_2 = 2\alpha_1 - \ln 4$. Asymptotically, both T_S and T_{LR} are χ^2-distributed with 2 degrees of freedom. The test statistics are now derived for two artificially created data materials.

Example 2

Data consists of 100 pairs of Bernoulli variables. Each pair is associated with a single covariate, z, ranging between zero and ten, see Figure 16. Ignoring the covariate z, the observed frequencies for $S = 0, 1, 2$ are $49, 17$, and 34, respectively. By only looking at the observed frequencies S_1 and S_2 seem to be dependent. The goodness of fit test given in (8) confirms this. The observed test statistic when the covariate is ignored is

$$\chi^2_{obs} \approx 42.54.$$

Clearly, the conclusion based only on this test would be that S_1 and S_2 are dependent. However, it is not sufficient to look at observed marginal frequencies only. When testing for independency one has to study how the probabilities $\pi_0(z), \pi_1(z)$, and $\pi_2(z)$ change when taking account of the covariate, z. The relative low frequency of pairs where $S = 1$ is explained by the fact that many observations are taken at z-values where $\pi_1(z)$ is small.
The score test statistic and the likelihood ratio test statistic, given in (13) and (14), take covariates into account in the test procedures. The observed test statistics for the two tests become

$$T_S \approx 0.0340$$

and

$$T_{LR} \approx 0.0338,$$

respectively. Because the critical value at the 5% level is 5.991, the hypothesis of independence can not be rejected in neither of the tests.
Another good indicator of the possible dependency between S_1 and S_2 is the estimated probability distribution of S, given in Figure 16.
The probability distribution closely resembles the appearance of a distribution for independent Bernoulli variables, see Chapter 6. A model for independent data has several characteristic properties. Two of these properties are clearly shown in Figure 16. First the maximum value of $\widehat{\pi}_1(z)$ is close to 0.5, and secondly $\widehat{\pi}_1(z)$ is a symmetric function around $\arg\max_z P(S = 1)$. This example emphasizes the importance of including existing covariates in the analysis.

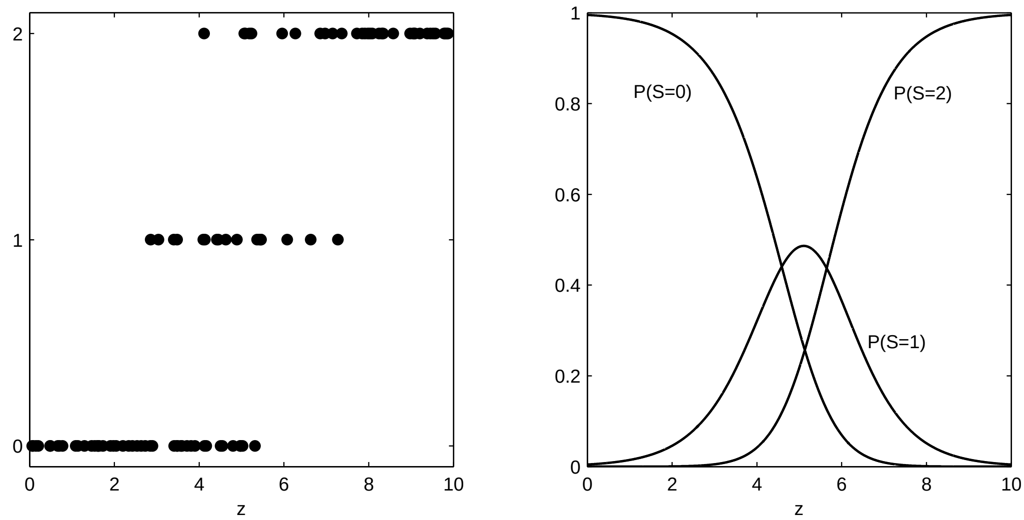


Figure 16: a) The 100 observations on S for Example 2. b) Probability distribution for S as a function of z. The parameters used are the estimator of the unrestricted likelihood, $\widehat{\theta}$.

Example 3

The data in the third example have a similar structure as the data in Example 2. Data consist of 100 pairs of Bernoulli variables, where each pair is associated with a single covariate. Thus, the same model can be fitted to this data material as to the previous data material. Figure 17 shows that the data in Example 3 resemble the data in Example 2. Nevertheless, the observed score test statistic and the observed likelihood ratio test statistic are given by

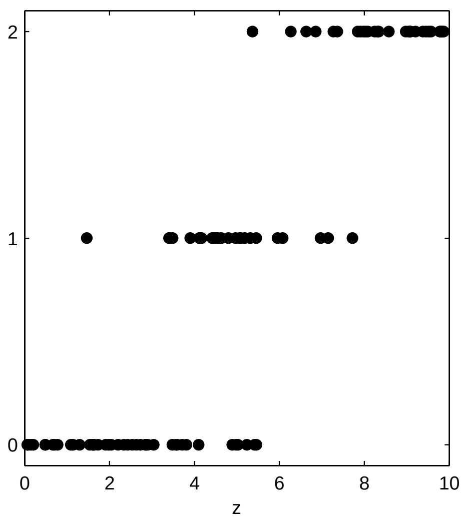
$$T_S \approx 6.3066$$

and

$$T_{LR} \approx 7.7791,$$

respectively. The hypothesis of independence is rejected in both tests because the observed test statistics exceed the critical value at the 5% level. Figure 17 presents the probability distribution of S based on the estimator of the unrestricted likelihood, $\widehat{\theta}$.

The probability distribution does not share the properties that independent Bernoulli variables would have generated. The maximum value of $\widehat{\pi}_1(z)$ is relatively far from 0.5 and $\widehat{\pi}_1(z)$ is not a symmetric function around $\arg\max_z P(S=1)$.



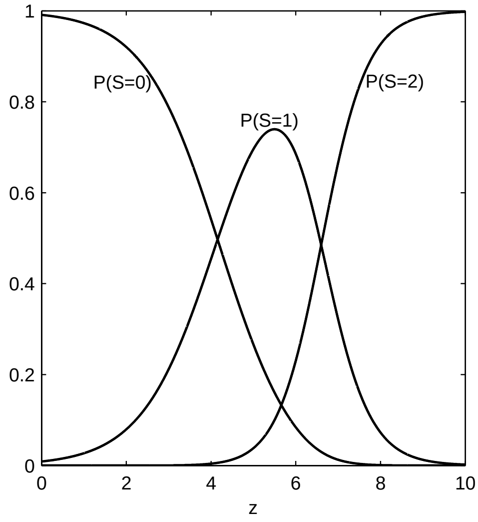


Figure 17: a) The 100 observations on S for Example 3. b) Estimated probability distribution for S as a function of z.

8 Optimal designs for testing for independency of Bernoulli variables

The previous chapter showed that an available covariate that is important in explaining the probability distribution of S should be included when testing for independency between S_1 and S_2. Examples showed that the response probabilities, π_0, π_1, and π_2, clearly depend on the covariate in this case. In an experimental study the values of the covariate can be controlled. Therefore it is interesting to find values of the covariate so that properties of the test are optimized. In particular, different sets of values of the covariate generate different power of the test. In this framework, a favourable power function can be generated if the values of the covariate, i.e. the design, are chosen in an appropriate way. This chapter considers the problem of finding optimal designs such that the local asymptotic power of the score test is maximized.

For determining an approximation to the power of the test at an alternative hypothesis close to H_0, let θ be the true value of the parameter vector and θ_0 the value under H_0. Let further $\delta = \sqrt{N}\,(\theta - \theta_0)$ be fixed so that θ converges to θ_0 as N tends to infinity. The first-order expansion of $\frac{1}{\sqrt{N}}u.\,(\theta_0)$ around θ is

$$\frac{1}{\sqrt{N}}u.\,(\theta_0) = \frac{1}{\sqrt{N}}u.\,(\theta) + \frac{1}{N}H\sqrt{N}\,(\theta_0 - \theta)\,,$$

where H is the matrix of second-order derivatives of the log likelihood function. Now, the first term in this expression converges to a normally distributed random variable with zero mean and variance $M\,(\theta_0)$, and the second term converges to

$M(\theta_0)\,\delta$. Hence, the distribution of $\frac{1}{\sqrt{N}}u.(\theta_0)$ is approximately normal with expectation $M(\theta_0)\,\delta$ and variance $M(\theta_0)$, in large samples. This makes that the distribution of T_S can be approximated in large samples by a noncentral χ^2-distribution with 2 degrees of freedom and non-centrality parameter

$$\varphi = \delta^T M(\theta_0)\,\delta. \tag{15}$$

The power of the test is found as the probability that T_S exceeds the critical value T_c. Obviously, the power depends on φ and is smallest in the direction δ in which $\delta^T M(\theta_0)\,\delta$ is minimized. The smallest possible value of φ is the smallest eigenvalue of $M(\theta_0)$ and δ is the eigenvector associated to the smallest eigenvalue.

If an experiment is to be conducted in order to test H_0, it is reasonable to select a design that makes the power of the test as large as possible. Furthermore, the smallest power is in the direction of the eigenvector associated to the smallest eigenvalue of $M(\theta_0)$. If no direction is of particular interest, a design that maximizes the smallest eigenvalue of $M(\theta_0)$ is proposed. This design is recognized as an E-optimal design, see Chapter 4 for a description of E-optimal designs.
Unfortunately, the E-optimal design for maximizing the smallest local power depends on the unknown parameter vector, so only a locally optimal design can be determined. As an example, with $\alpha_1 = -2$ and $\beta_1 = 1$, and accordingly $\alpha_2 = -4 - \ln 4 \approx -5.3863$ and $\beta_2 = 2$, the 3−point design

$$\xi^* = \left\{ \begin{array}{ccc} 0.1741 & 2.2049 & 5.7469 \\ 0.4414 & 0.3706 & 0.1880 \end{array} \right\}$$

is obtained. The probabilities π_0, π_1, and π_2 are estimated at each design point as in (10) and (11) and then using the test statistic (9), treating the design points as different groups. The asymptotic power for a test based on the design ξ^* is shown in Figure 18.
It should be noted that this design maximizes the smallest power. This means that there may be other designs that yield a stronger power at some parameter values but at some other parameter values yield a smaller power than the E-optimal design. On the other hand, there is no design that dominates the E-optimal design in that it provides a larger asymptotic power than that for the E-optimal design for all parameter values.

8.1 Small sample performances of the test; a simulation experiment

As mentioned previously it is of interest to test hypotheses about α_2 and β_2 using the score test. The proposed design for this test procedure is a locally E-optimal. A corresponding power function for small samples can be evaluated using simulation. This section presents a numerical illustration of the power of this score test. Values of the parameters α_1 and β_1 are chosen arbitrary to $\alpha_1 = -2$ and $\beta_1 = 1$. The

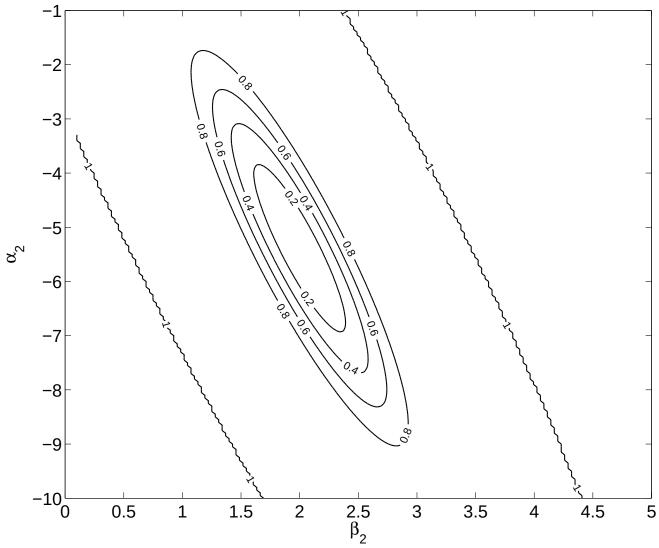


Figure 18: Contour plot of the asymptotic power for testing dependency. The power is given for different alternative hypothesis, i.e. different values of α_2 and β_2.

power of the score test will depend on the alternative hypothesis. Values of α_2 and β_2 far away from their restriction under H_0 will in general give a large power. In order to examine the power for different alternative hypotheses, α_2 and β_2 are varied over an interval of values. Sample size, N, is chosen to 50, 100, and 400, respectively. Using simulation a data set is created 5000 times for each value of α_2 and β_2. The score test statistic is then calculated based on the simulated data using the formulas described above. Power is determined as the percentage of score test statistics larger than the critical value T_c. The significance level is chosen to be 5% in all studied cases.

For testing the hypothesis

$$\begin{aligned} H_0 &: \begin{cases} \alpha_2 = 2\alpha_1 - \ln 4 \\ \beta_2 = 2\beta_1 \end{cases} \\ H_1 &: \alpha_2 \neq 2\alpha_1 - \ln 4 \quad \text{or} \quad \beta_2 \neq 2\beta_1 \end{aligned}$$

the proposed design points are E-optimal and depend only on α_1 and β_1. The score test statistic, based on the simulated data, is calculated using (9), (10), and (11). Contour plots of the power as a function of α_2 and β_2 for different sample sizes are given in Figure 19, Figure 20 and Figure 21.

When calculating the power based on simulation two interesting phenomena occur, both of which may be explained in terms of the log-odds ratio between S_1 and S_2. For large values on α_2 and β_2, $\ln \Omega$ is large which implies that $\widetilde{\theta}$ is often equal to one. As a direct consequence the score test statistic, T_S in (9), can not be computed. Due to this, the right side of the contour plots in Figure 19 and

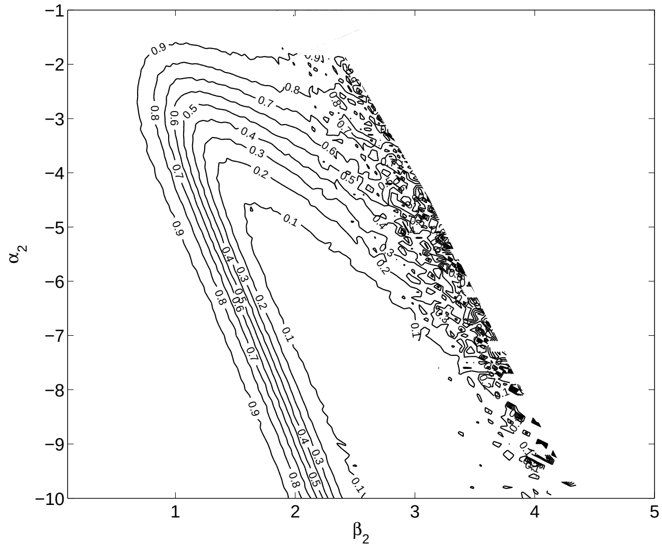


Figure 19: Contour plot of the simulated power as a function of α_2 and β_2 for N=50.

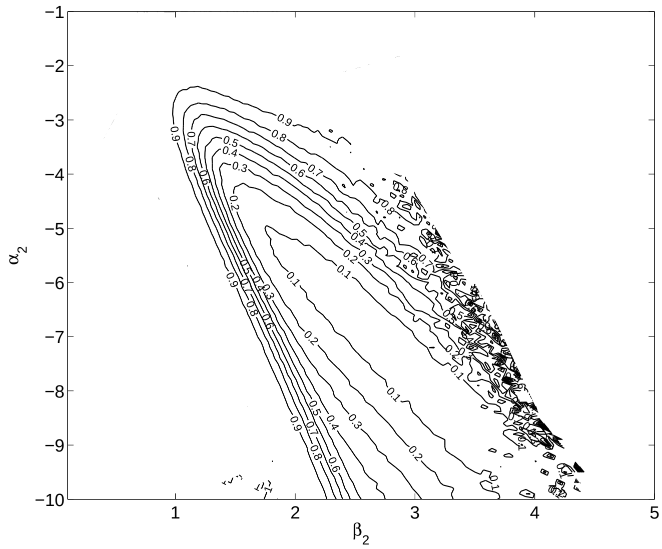


Figure 20: Contour plot of the simulated power as a function of α_2 and β_2 for N=100.

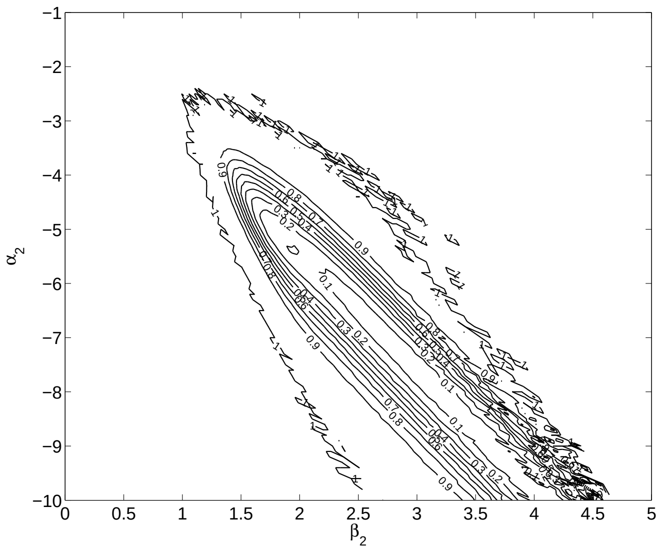


Figure 21: Contour plot of the simulated power as a function of α_2 and β_2 for N=400.

Figure 20 as well as the lower right corner of the contour plot in Figure 21 are not smooth. These computational problems are more extensive in small samples, ($N = 50$ and $N = 100$). The power for large values on α_2 and β_2 is not calculated, since large computational problems exist together with the fact that S_1 and S_2 are very dependent. Therefore a test for independency is not meaningful for these parameter values. Because the computational problems for some values of α_2 and β_2, the power is not based on 5000 replicates for all (α_2, β_2) in the plots.
Under H_0 the log-odds ratio is equal to zero since S_1 and S_2 are assumed to be independent. Some combinations of α_2 and β_2 outside H_0 generate values on the log-odds ratio that are close to zero. In those cases the estimated frequencies $N\widetilde{\pi}_j$ in (9) are similar to the observed frequencies r_j for $j = 0, 1, 2$. This results in low power, despite the fact that α_2 and β_2 are far from their restriction under the null hypothesis. The problem with low power is present in all three contour plots. It can be seen as an area with low power starting from ($\alpha_2 \approx -5, \beta_2 \approx 2$) and going in the direction where α_2 gets smaller and β_2 gets larger. Note that the log-odds ratio in general depends on x. The log-odds ratio may be close to zero for one design point but not necessary for the other design points.

The score test statistic in small samples is not completely comparable to the asymptotic distribution of the score test statistic. An important difference is that the asymptotic power of the test is based only on the cdf of the noncentral χ^2-distribution. The expression for the test statistic in small samples is based on observed data. Therefore, the described situation where the test statistic can not be computed can only occur in finite samples.

8.2 Robustness of locally optimal designs

As stated before, the optimal designs depend on the parameters α_1 and β_1. For that reason it is of great interest to study how well the optimal designs perform against incorrect guesses of the parameter values. Especially it is important that the designs generate a good power function for the test they are suppose to optimize, regardless of incorrect guesses of the parameter values. The robustness is investigated by calculating the power of the test for different values of the parameters. A design is considered robust if the power of the test for fairly incorrect guesses of the parameter values is close to the power generated by the correct values of the parameters. Note that all power functions are based on the optimal design under consideration. It is assumed through out the chapter that the designs are optimal for $\alpha_1 = -2$ and $\beta_1 = 1$. Robustness is evaluated for a number of different alternative hypotheses since power functions also depend on the alternative hypothesis. Because a complete robustness examination of the optimal design is extensive, only a sample of plots is shown here.

Consider the E-optimal design for testing both restrictions $\alpha_2 = 2\alpha_1 - \ln 4$ and $\beta_2 = 2\beta_1$. Assume that the alternative hypothesis is given by $\alpha_2 = 2\alpha_1 - \ln 4 + a$ and $\beta_2 = 2\beta_1 + b$ where a and b are constants. Since the power, given α_1 and β_1, is an even function of a and b, only positive values of a and b are considered. The evaluation of the robustness of the E-optimal design involves α_1, β_1, a, and b, making the analysis somewhat immense. Comparing the power as a function of a and b separately, or as a function of α_1 and β_1 separately, does not give a complete description of the robustness of the design. The evaluation of the robustness utilizes the relative power for different (α_1, β_1) with respect to $(\alpha_1 = -2, \beta_1 = 1)$. Relative power is used because it gives a direct measure of the robustness.

Figure 22, Figure 23, Figure 24, and Figure 25 show the relative power as a function of both α_1 and β_1 given some values on a and b. Note that the relative power in all figures is equal to one in the point $(\alpha_1 = -2, \beta_1 = 1)$.
In general, the design is robust along the diagonal where $(\alpha_1 > -2; \beta_1 < 1)$ and where $(\alpha_1 < -2; \beta_1 > 1)$. The contour plots in all figures are parallel to this diagonal, verifying that the design is robust for these parameter values. Some values of α_1 and β_1 generate a relative power larger than 2, showing that the design is very robust for these parameter values. But on the other hand, the relative power decreases fast when the value on α_1 or β_1 change slightly in the wrong direction. This kind of sensitivity is demonstrated in several of the Figures, e.g. in Plot 2 in Figure 22, Figure 23, and Figure 24 (when $b = 0.20$). The design is least robust when $\alpha_1 < -2$ and $\beta_1 < 1$, simultaneously. From Figure 22, Figure 23, Figure 24, and Figure 25 it is hard to conclude how the alternative hypothesis, i.e. a and b affects the robustness. Especially for alternative hypothesis quite close to the null hypothesis no clear relation between power and alternative hypothesis exist.

Figure 26 and Figure 27 illustrate in more detail how the alternative hypothesis

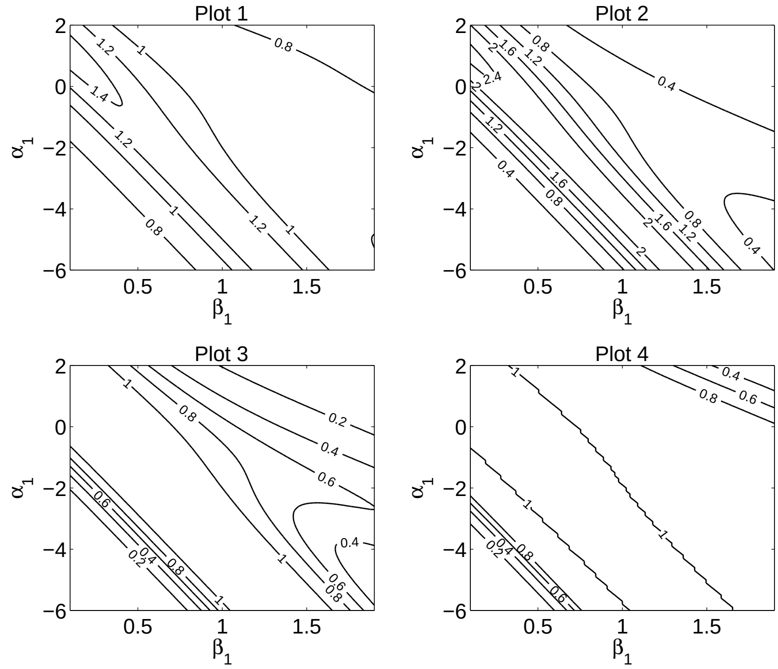


Figure 22: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The alternative hypotheses are $(a = 0.05; b = 0.05)$, $(a = 0.05; b = 0.2)$, $(a = 0.05; b = 0.5)$, and $(a = 0.05; b = 1.5)$, respectively.

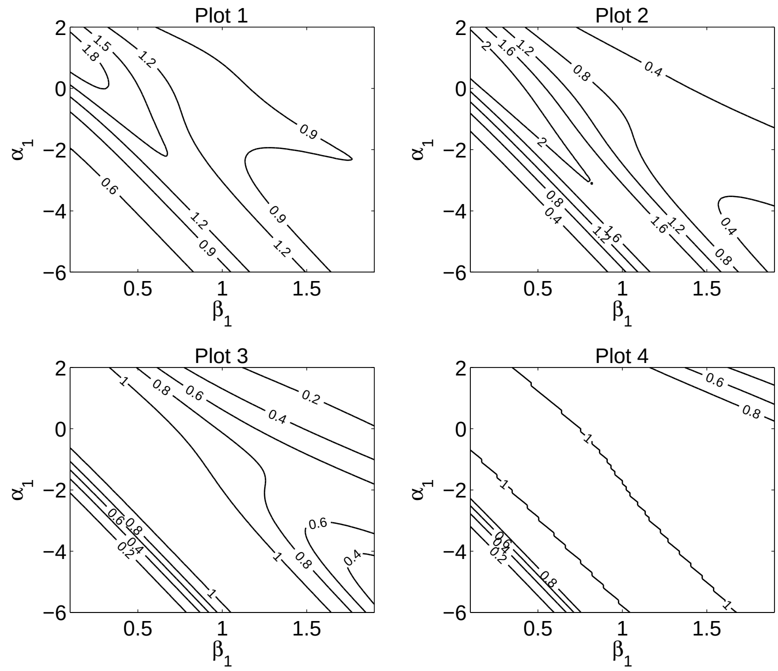


Figure 23: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The alternative hypotheses are $(a = 0.2; b = 0.05)$, $(a = 0.2; b = 0.2)$, $(a = 0.2; b = 0.5)$, and $(a = 0.2; b = 1.5)$, respectively.

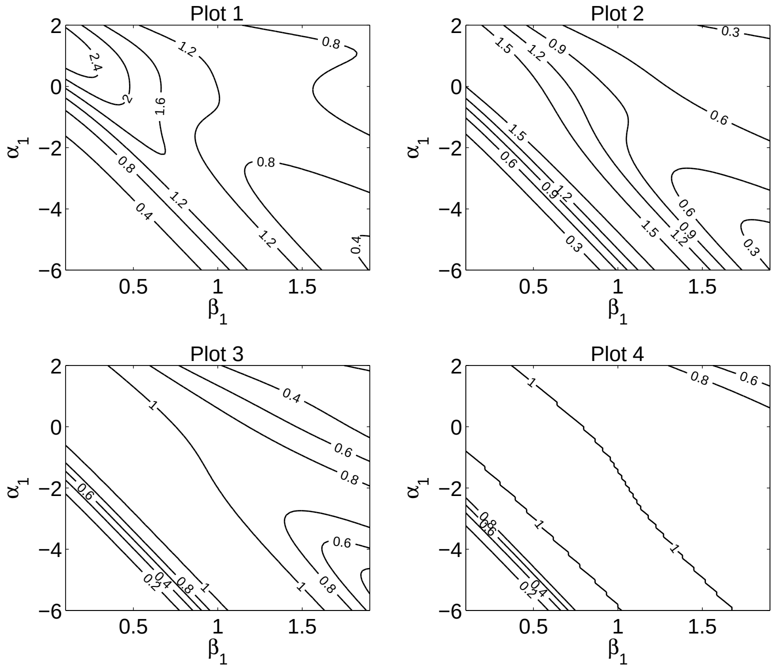


Figure 24: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The alternative hypotheses are $(a = 0.5; b = 0.05)$, $(a = 0.5; b = 0.2)$, $(a = 0.5; b = 0.5)$, and $(a = 0.5; b = 1.5)$, respectively.

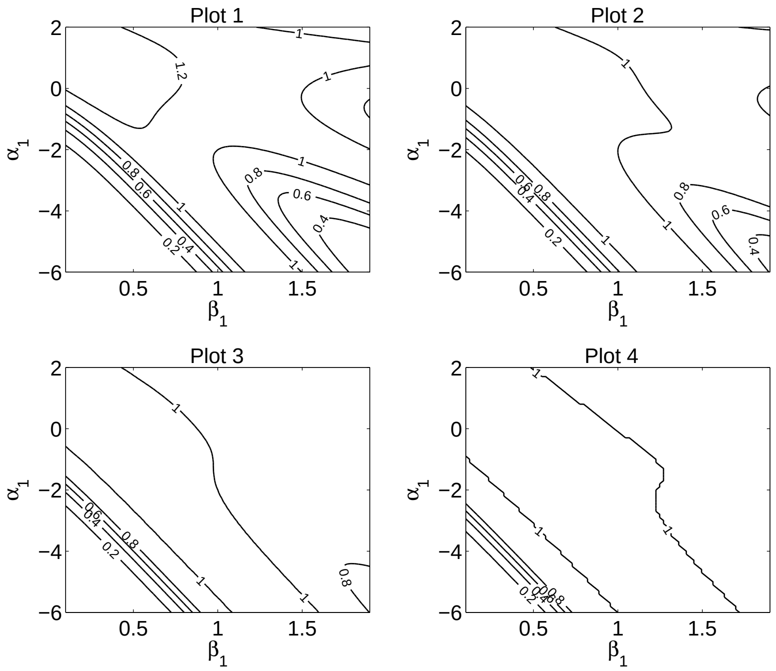


Figure 25: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The alternative hypotheses are $(a = 1.5; b = 0.05)$, $(a = 1.5; b = 0.2)$, $(a = 1.5; b = 0.5)$, and $(a = 1.5; b = 1.5)$, respectively.

affects the robustness. Figure 26 and Figure 27 show the relative power as a function of both a and b given some values on α_1 and β_1.
Based on Figure 26 and Figure 27 the design is robust since the relative power is sufficiently large for most values on a and b. It should be noted though, that only one of the parameters is incorrect at the same time in these figures compared to the previous figures.

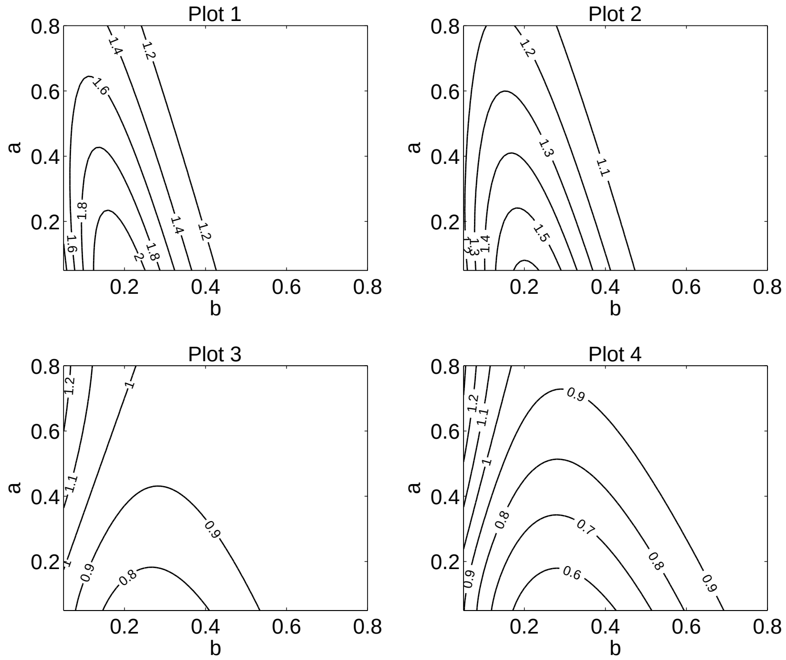


Figure 26: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The parameter values are $(\alpha_1 = -4, \beta_1 = 1)$, $(\alpha_1 = -3, \beta_1 = 1)$, $(\alpha_1 = -1, \beta_1 = 1)$, and $(\alpha_1 = 0, \beta_1 = 1)$, respectively.

9 Conclusions and suggestions for further research

A multinomial logit model for equally distributed Bernoulli variables is examined in this thesis. In Chapter 3 the modified Cox model for k variables is introduced. Expressions for the likelihood function, the score function and the information matrix are derived. How well data are fitted in this model depends on several factors and assumptions. The model is only applicable to data where equally distributed Bernoulli variables exist. For a similar model Agresti (2002) argues that data fit poorly if the marginal distributions of the Bernoulli variables differ substantially. Hence, it remains to investigate how poorly the model fits when the assumption is not valid.
Furthermore all observations with a certain value on the covariates are assumed to be homogeneous in the sense that they have the same parameter values. A further

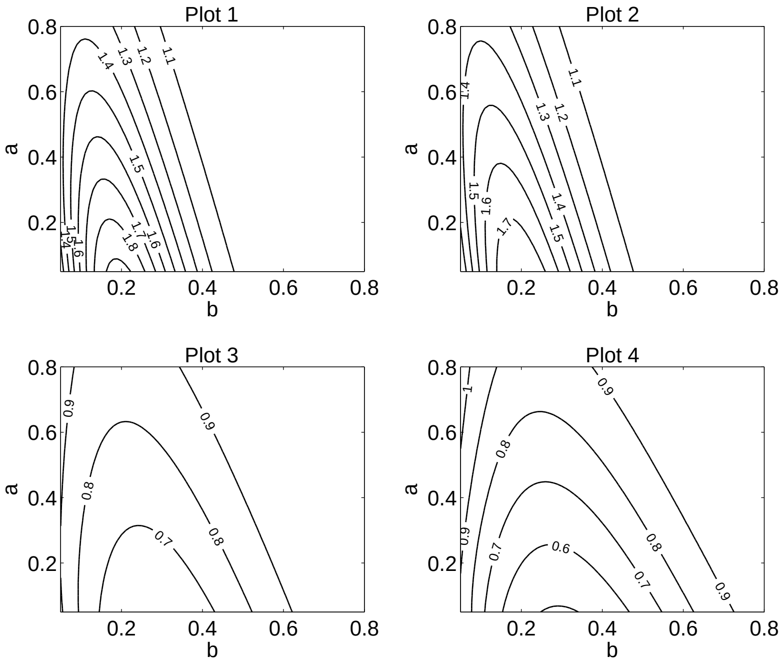


Figure 27: Contour plots of the relative power with respect to $\alpha_1 = -2$ and $\beta_1 = 1$. The parameter values are $(\alpha_1 = -2, \beta_1 = 0.5)$, $(\alpha_1 = -2, \beta_1 = 0.8)$, $(\alpha_1 = -2, \beta_1 = 1.2)$, and $(\alpha_1 = -2, \beta_1 = 1.5)$, respectively.

development would be to include, for example random effect parameters in the linear predictor. These parameters would then account for variations among observations. Parameters for modelling the heterogeneity among individuals are included in the general model for dependent binary responses given in Agresti (1997).

Chapter 5 presents a model for two Bernoulli variables, where the log-odds ratio between the variables is constant. Different symmetry properties for the probability distribution are given together with some general results about locally D-optimal designs. All these results can be interpreted in terms of the log-odds ratio. For situations where the log-odds ratio is large negative or large positive a general expression for D-optimal designs is given. The theoretical foundation of these general expressions would be more solid if some analytical results were derived. Foremost, to derive analytical results that show how D-optimal designs for different parameter vectors with the same log-odds ratio are related. Moreover, to derive analytical results that show how the information matrix is affected when the values of the parameters change. In this context, Fan (1999) and Puu (2003) derived analytical results for the determinant of the information matrix. These results are used to derive general expressions for locally D-optimal designs. Although they worked with other models, it is of interest to derive similar results for the model in this thesis. When the variables are independent, the model has several symmetry properties. Since the information matrix in this case has a relatively simple expression, a general expression for locally D-optimal designs can be derived.

Tests for determining if the variables are independent is considered in Chapter 7. Score tests and likelihood ratio tests for different situations are derived.
In Chapter 8 a design that maximizes the local asymptotic power of the score test is proposed. A small sample study indicates that the locally optimal design performs well as long as the log-odds ratio is negative. Problem occurs, though, for large values of the log-odds ratio. The problem is related to the fact that the expected frequency for some response categories is small. This affects the performance of the test statistic in small samples. If the Bernoulli variables are strongly correlated the value of the test statistic might not exist. On the other hand, other test procedures for testing independence in 2×2 contingency tables have the same problem when small expected cell frequencies appear, see Agresti (2002) and Haberman (1988).
Chapter 8 also contains a section where the robustness of the locally optimal design is examined. The design is fairly robust against incorrect parameter values. It should be stated though, that the investigation about robustness is not comprehensive since only one parameter setup is examined, i.e. $\alpha_1 = -2$ and $\beta_1 = 1$. A complete investigation of the robustness of the designs would require an examination with a very large number of parameter values and alternative hypotheses. The robustness could also be evaluated by creating values of the test statistic using simulations instead of relying on asymptotic results. At least in situations where the correlation between the variables is strong, the power function based on simulations does not resemble the asymptotic power function completely. One drawback with simulations is that the procedure would be very time consuming.

It is clear from Chapter 6 and Chapter 8 that if two variables are independent certain parameter restrictions must hold. Analogously, parameter restrictions for $k = 3,\ 4, ...$ jointly dependent variables could be derived. These restrictions can then be used in tests for independency as well as in deriving optimal designs that maximizes the power of such tests.
Finally a comment about the robustness of the various locally optimal designs derived in this thesis. According to Zocchi and Atkinson (1999) D-optimal designs for logistic models are dependent on the parameters. Changes in the parameter values result in different design points, different design weights and sometimes even different number of design points. Consequently a nonoptimal design can have a low efficiency. It is therefore of interest to study D-optimal in average designs for the models in this thesis.

References

Agresti, A. (1997). "A Model for Repeated Measurements of a Multivariate Binary Response." *J. Amer. Statist. Assoc.*, 92, 437, 315–321.

— (2002). *Categorical data analysis*. Wiley Series in Probability and Statistics, 2nd ed. New York: Wiley-Interscience [John Wiley & Sons].

Appelgren, J. (2004). "Locally D-optimal Designs for Bivariate Logistic Regression." Department of Statistics, University of Umeå. Statistical Studies No. 32.

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs*. Oxford science publications.

Atkinson, A. C. and Haines, L. M. (1996). "Designs for nonlinear and generalized linear models." In *Design and analysis of experiments*, vol. 13 of *Handbook of Statist.*, 437–475. Amsterdam: North-Holland.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. The MIT Press, Cambridge, Mass.-London. With the collaboration of Richard J. Light and Frederick Mosteller.

Bonney, G. E. (1987). "Logistic Regression for Dependent Binary Observations." *Biometrics*, 43, 4, 951–973.

Bruce, D. and Nyquist, H. (2005). "Testing for Dependency of Bernoulli Variables." Research Report 2005:3, Department of Statistics, Stockholm University.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*. 2nd ed. Duxbury.

Chaloner, K. and Larntz, K. (1989). "Optimal Bayesian Design applied to Logistic Regression Experiments." *Journal of Statistical Planning and Inference*, 21, 2, 191–208.

Christensen, R. (1997). *Log-linear models and logistic regression*. Springer Texts in Statistics, 2nd ed. New York: Springer-Verlag.

Cox, D. R. (1972). "The Analysis of Multivariate Binary Data." *Applied Statistics*, 21, 2, 113–120.

Cox, D. R. and Snell, E. J. (1989). *Analysis of binary data*, vol. 32 of *Monographs on Statistics and Applied Probability*. 2nd ed. London: Chapman & Hall.

Dale, J. R. (1986). "Global Cross-Ratio Models for Bivariate, Discrete, Ordered Responses." *Biometrics*, 42, 4, 909–917.

Dobson, A. J. (2002). *An introduction to generalized linear models*. Chapman & Hall/CRC Texts in Statistical Science Series, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.

Dragalin, V. and Fedorov, V. (2006). "Optimal Designs and Limiting Optimal Designs for a Trinomial Response." *Journal of Statistical Planning and Inference*. To be published.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate statistical modelling based on generalized linear models*. Springer Series in Statistics, 2nd ed. New York: Springer-Verlag. With contributions by Wolfgang Hennevogl.

Fan, S. K. (1999). "Multivariate Optimal Designs." Ph.D. thesis. School of Statistics, University of Minnesota.

Fan, S. K. and Chaloner, K. (2004). "Optimal Designs and Limiting Optimal Designs for a Trinomial Response." *Journal of Statistical Planning and Inference*, 126, 1, 347–360.

Fedorov, V. V. and Hackl, P. (1997). *Model-oriented design of experiments*, vol. 125 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

Galambos, J. (1988). *Advanced probability theory*, vol. 3 of *Probability: Pure and Applied*. New York: Marcel Dekker Inc.

Gumbel, E. J. (1961). "Bivariate logistic distributions." *J. Amer. Statist. Assoc.*, 56, 335–349.

Haberman, S. J. (1988). "A warning on the use of chi-squared statistics with frequency tables with small expected cell counts." *J. Amer. Statist. Assoc.*, 83, 402, 555–560.

Heise, M. A. and Myers, R. H. (1996). "Optimal Designs for Bivariate Logistic Regression." *Biometrics*, 52, 2, 613–624.

Kalish, L. and Rosenberger, J. (1978). "Optimal Designs for the estimation of the Logistic Function." Tech. Rep. 33, Pennsylvania State University, University Park, Pennsylvania.

Kiefer, J. (1959). "Optimum experimental designs." *J. Roy. Statist. Soc. Ser. B*, 21, 272–319.

Kiefer, J. and Wolfowitz, J. (1960). "The equivalence of two extremum problems." *Canad. J. Math.*, 12, 363–366.

Le Cassie, S. and Van Houwelingen, J. (1994). "Logistic Regression for Correlated Binary Data." *Applied Statistics*, 43, 1, 95–108.

Liang, K.-Y., Zeger, S. L., and Qaqish, B. (1992). "Multivariate Regression Analyses for Categorical Data." *Journal of the Royal Statistical Society, Series B*, 54, 1, 3–40.

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

Murtaugh, P. A. and Fisher, L. D. (1990). "Bivarate Binary Models of Efficacy and Toxicity in Dose-ranging Trials." *Comm. Statist. Theory Methods*, 19, 6, 2003–2020.

Palmgren, J. (1991). "Approaches to Modelling Bivariate Binary Responses; An Empirical Adventure." National Public Health Institute, Helsinki.

Pettersson, H. (2001). "Optimum in Average and Minimax Designs for estimation of Generalized Linear Models." Ph.D. thesis. Department of Statistics. University of Umeå.

Pettersson, H. and Nyquist, H. (2003). "Computation of Optimum in Average Designs for Experiments with Finite Design Space." *Communications in Statistics. Simulation and Computation*, 32, 1, 205–221.

Puu, M. (2003). "Optimum Experimental Designs for Generalized Linear Models with Multinomial Response." Department of Statistics, University of Umeå. Statistical Studies No. 31.

Rosner, B. (1984). "Power and sample size for a collection of 2×2 tables." *Biometrics*, 40, 4, 1025–1035.

Silvey, S. D. (1980). *Optimal design*. London: Chapman & Hall.

Tielsch, J., Sommer, A., Katz, J., Quigley, H., and Ezrine, S. (1991). "Socioeconomic Status and Visual Impairment among urban Americans." *Archives of Ophthalmology*, 109, 5, 637–641.

Zocchi, S. S. and Atkinson, A. C. (1999). "Optimum Experimental Desings for Multinomial Logistics Models." *Biometrics*, 55, 2, 437–444.

ISBN 91-7155-213-8