Numerical Construction of Maximin Optimal Designs for Binary Response Models

Stefanie Biedermann
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: stefanie.biedermann@ruhr-uni-bochum.de

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum
Germany
email: holger.dette@ruhr-uni-bochum.de
FAX: +49 2 34 70 94 559

October 6, 2003

Abstract

For the binary response model, we determine optimal designs which are robust with respect to misspecifications of the unknown parameters. We propose a maximin approach and provide a numerical method to identify the best two point designs for the commonly applied link functions. This method is broadly applicable and can be extended to designs with a given number (≥ 2) of support points and further link functions. The results are illustrated for the logistic and probit model, for which the maximin optimal designs are found explicitly.

Keywords and Phrases: Binary response model, robust optimal design, maximin D-optimality, Bayesian D-optimality, prior distribution

1 Introduction

We consider the common binary response model where a subject is administered a stimulus at a dose level $x \in I\!R$. The response X is a binary random variable with success probability $p(x, \vartheta)$, i.e. $X \sim Bin(1, p(x, \vartheta))$, where $x \in I\!R$ is the explanatory variable and ϑ is an unknown parameter. In this article, we deal with the following parametrization of a two parameter binary response model,

$$(1.1) \qquad p(x, \vartheta) = F(\beta(x - \mu)), \qquad \vartheta = (\mu, \beta)^T, \ \mu \in I\!R, \ \beta \in I\!R^+,$$

where F denotes a known distribution function with density f. For this model the Fisher information for the parameter ϑ at a point x is given by

$$I(x,\vartheta) = h(\beta(x-\mu)) \begin{pmatrix} \beta^2 & -\beta(x-\mu) \\ -\beta(x-\mu) & (x-\mu)^2 \end{pmatrix},$$

where the function $h(z)$ is defined as

$$h(z) = \frac{f^2(z)}{F(z)(1-F(z))}. \tag{1.2}$$

An (approximate) design ξ is a probability measure with finite support on $I\!R$, i.e. the observations are taken at the support points of the measure proportional to the corresponding masses. The Fisher information of the design ξ is given by

$$M(\xi,\vartheta) = \int I(x,\vartheta)\, d\xi(x), \tag{1.3}$$

and an optimal design maximizes a real-valued function of the Fisher information matrix, which is usually referred to as an optimality criterion [see e.g. Silvey (1980)]. A typical example is D-optimality where the determinant of the Fisher information is maximized with respect to the design ξ, thus minimizing the (first order approximation of the) volume of the ellipsoid of concentration for the parameter ϑ.

Much effort has been devoted to the problem of finding good designs for the binary response model [see Kalish and Rosenberger (1978), Abdelbasit and Plackett (1983), Minkin (1987), Sitter and Fainaru (1997) among many others]. Since the Fisher information and thus the D-optimal design depends on the unknown parameter a D-optimal design cannot be implemented directly in practice. Following Chernoff (1953), numerous authors assume that an initial guess of ϑ is available and determine so-called locally optimal designs [see Ford, Torsney and Wu (1992), Sitter and Wu (1993)]. However, misspecifications of the parameter can lead to poor results in the subsequent data analysis. A more robust alternative is to assume sufficient knowledge of ϑ to specify a prior distribution for this parameter and to average the respective optimality criteria over the plausible values of ϑ defined by the prior. This leads to so-called Bayesian optimality criteria [see e.g. Chaloner and Larntz (1989)].

As an alternative for the construction of robust designs, we propose a maximin approach based on the D-optimality criterion, which only requires the specification of a certain range for the unknown parameter. We feel that this is a more realistic scenario since practitioners will often have difficulties to specify a prior distribution for the unknown parameter ϑ, especially if this is multidimensional. So far, maximin D-optimal designs have been found only within a very restricted class of designs, i.e. equidistant designs with equal weights and symmetric about the mean of the μ-interval [see Sitter (1992)]. The main difficulty for the construction of optimal designs with respect to maximin optimality criteria is the non-differentiability of this type of criteria. We solve this problem by exploiting the close relation between Bayes- and maximin optimality criteria [see Dette, Haines and Imhof (2003)]. This allows us to obtain the maximin D-optimal designs as the limit of some Bayesian optimal designs.

In Section 2, we describe the above optimality criteria, the method for the construction of maximin optimal designs and its implementation considering minimally supported designs. Several examples of link functions are given, for which the method works without modifications. Section 3 deals with extensions of the method to the other commonly applied link functions as well as designs with more than two support points. In Section 4, the minimally supported designs from Section 2 are taken up again and investigated with respect to global optimality (i.e. optimality within the class of all designs). Finally, some of the more technical arguments are deferred to an appendix.

2 Maximin D-optimal two point designs in the binary response model

In the following, we motivate and define the optimality criterion under consideration in this article. As indicated in the introduction, D-optimal designs depend on the unknown model parameters and are not necessarily robust with respect to their misspecification. There are two non-sequential approaches to construct optimal designs based on the D-criterion, which are more robust. On the one hand, it can be reasonable to assume that some prior knowledge about the parameter ϑ is available in advance, which can be specified by a probability distribution on the parameter space Θ. In such cases it makes sense to choose a design that maximizes a function of the determinant of the information matrix $M(\xi, \vartheta)$ after averaging out the plausible values of ϑ by a prior distribution, say $\tilde{\pi}$. Since the use of standardized criteria is recommended to avoid different scaling [see Dette and Wong (1996) or Dette (1997)] this leads to a Bayesian optimality criterion based on the D-efficiencies

$$\mathrm{eff}_D(\xi, \vartheta) = \left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{1/2},$$

where ξ_ϑ denotes the locally D-optimal design for the parameter ϑ. The criterion function $\tilde{\Psi}_p(\xi)$ is then defined as

$$\tilde{\Psi}_p(\xi) = \left[\int_\Theta \left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{p/2} d\tilde{\pi}(\vartheta) \right]^{\frac{1}{p}}, \quad -\infty < p < 0, \tag{2.1}$$

and a Bayesian $\tilde{\Psi}_p$-optimal design ξ with respect to the prior $\tilde{\pi}$ maximizes this expression over the set of all approximate designs on the design space [see Dette and Wong (1996)]. Equivalently, we will use a simplified notation for this criterion, where each prior distribution $\tilde{\pi}$ is identified with an associated prior π defined by

$$d\pi(\vartheta) = |M(\xi_\vartheta, \vartheta)|^{-q} \, d\tilde{\pi}(\vartheta)$$

replacing the expression $p/2$ by q. A monotone transformation of the standardized Bayesian D-optimality criterion function can then be written as

$$\Psi_q(\xi) = \left[\int_\Theta |M(\xi, \vartheta)|^q \, d\pi(\vartheta) \right]^{\frac{1}{q}}.$$

If, on the other hand, the specification of a prior distribution π on the parameter space is not possible in advance, it is sensible to construct the design for the protection of the experiment against the worst possible case, i.e. against the values of $\vartheta \in \Theta$ minimizing the D-efficiency. In this case the optimal design maximizes the minimal D-efficiency with respect to Θ which yields the standardized maximin D-optimality criterion

$$\Psi_{-\infty}(\xi) = \inf_{\vartheta \in \Theta} \left[\left(\frac{|M(\xi, \vartheta)|}{|M(\xi_\vartheta, \vartheta)|} \right)^{1/2} \right] \tag{2.2}$$

[see Müller (1995) or Imhof (2001)]. Throughout this article, we call a design maximizing the above function $\Psi_{-\infty}$-optimal (with respect to Θ), where Θ denotes the space of possible values for the parameter ϑ. Obviously, in the maximin case it is not necessary to have a sufficient amount of prior knowledge to specify certain preferable values for ϑ as it is in the Bayesian case. The only "parameter" the experimenter has to choose in advance is the "region of uncertainty" Θ as a subset of $\mathbb{R} \times \mathbb{R}^+$. Note that the standardized maximin criterion is obtained in the limit from the Bayesian criterion as q tends to $-\infty$. For this reason, the notation $\Psi_{-\infty}$ is consistent with the above definitions. It is easy to see that in the binary response model with parametrization (1.1) the value of the optimal determinant of the Fisher information matrix $|M(\xi_\vartheta, \vartheta)|$ does not depend on the value of the parameter vector $\vartheta \in \Theta$, i.e. the standardization of the criterion function is constant with respect to ϑ. Therefore we will use the notation "maximin D-optimal" equivalently to "standardized maximin D-optimal" as well as the analogous notations for the Bayesian case in the following.
A computational advantage of the Bayesian optimality criterion is its Fréchet differentiability. As a consequence, standard numerical methods for the determination of Bayesian D-optimal designs can easily be adapted [see e.g. Chaloner and Larntz (1989)]. The determination of maximin D-optimal designs, however, is a substantially more complex problem. Usually the structure of the particular design criterion under consideration has to be used for the construction of algorithms [see Sitter (1992) or Fandom Noubiap and Seidel (2000)]. We will now discuss a general strategy to obtain maximin optimal designs as limits from Bayesian optimal designs.
From Theorem 3.1 in Dette, Haines and Imhof (2003) we obtain a close connection between Bayesian and maximin optimal designs, which we will use to construct maximin D-optimal designs from their Bayesian counterparts. The first part of the following theorem gives a special case of the above-mentioned result, so that it is applicable for our purpose. Part two, moreover, deals with a simplification with respect to the choice of the prior π, which will be exploited in the following to simplify our method.

Theorem 1

1. *Let Θ be compact and π be an arbitrary prior on Θ where* $\text{supp}(\pi) = \Theta$. *Suppose that the standardized Bayesian D-optimal designs ξ_q within a class of designs Δ with respect to π converge weakly to some limit $\xi^* \in \Delta$ as $q \to -\infty$. Then ξ^* is standardized maximin D-optimal within the class Δ.*

2. *Let*

$$\mathcal{N}(\xi) = \Big\{\vartheta \in \Theta \mid \Psi_{-\infty}(\xi) = \Big(\frac{|M(\xi,\vartheta)|}{|M(\xi_\vartheta,\vartheta)|}\Big)^{1/2}\Big\}.$$

If a superset E of $\mathcal{N}(\xi^*)$ is known, the above result still holds if we choose a prior π on E with the condition $\mathrm{supp}(\pi) = \Theta$ *replaced by* $\Theta \supset \mathrm{supp}(\pi) = E \supseteq \mathcal{N}(\xi^*)$.

By a standard argument in design theory [see Silvey (1980), Lemma 5.1.3] it can be shown that Bayesian and maximin D-optimal two point designs for the binary response model are equally weighted. Thus, in the following we will restrict our attention to uniform distributions as far as two point designs are considered. The following theorem gives a condition on the function $h(z)$ so that $\mathcal{N}(\xi)$ is a subset of the boundary of Θ for all two point designs ξ with equal weights.

Theorem 2 *Let Θ be convex. If the function $h(z)$ in (1.2) is log-concave on $\mathbb{R}$ then for all equally weighted two point designs ξ the determinant $|M(\xi,\vartheta)|$ is a log-concave function in ϑ.*

It follows from Ford, Torsney and Wu (1992) that the value of the optimal determinant of the Fisher information matrix $M(\xi_\vartheta,\vartheta)$ does not depend on the value of the parameter $\vartheta \in \Theta$, and as a consequence the standardization of the criterion function is constant with respect to ϑ. For this reason, log-concavity of the function $h(z)$ implies log-concavity of the standardized determinant function $\frac{|M(\xi,\vartheta)|}{|M(\xi_\vartheta,\vartheta)|}$ and thus $\mathcal{N}(\xi^*) \subset \partial\Theta$, where $\partial\Theta$ denotes the boundary of the region of uncertainty.

Table 1: *Results on the log-concavity of the function $h(z)$ for several common link functions $F(z)$ ($s(z) = sign(z)$).*

link function	$F(z)$	$h(z)$	log-concavity of $h(z)$
Double Exponential	$\frac{1+s(z)}{2} - \frac{s(z)}{2}e^{-\lvert z\rvert}$	$\frac{1}{2e^{\lvert z\rvert}-1}$	no
Double Reciprocal	$\frac{1+s(z)}{2} - \frac{s(z)}{2}(\frac{1}{1+\lvert z\rvert})$	$\frac{1}{(1+\lvert z\rvert)^2(2\lvert z\rvert+1)}$	no
Complementary Log-Log	$1-e^{-e^z}$	$\frac{e^{2z}}{-1+e^{e^z}}$	yes
Logit	$\frac{1}{(1+e^{-z})}$	$\frac{e^z}{(1+e^z)^2}$	yes
Probit	$\Phi(z)$	$\frac{\phi^2(z)}{\Phi(z)(1-\Phi(z))}$	yes
Skewed Logit, $m \geq 1$	$\frac{1}{(1+e^{-z})^m}$	$\frac{m^2}{(1+e^z)^2(-1+(1+e^{-z})^m)}$	yes
Skewed Logit, $m < 1$	$\frac{1}{(1+e^{-z})^m}$	$\frac{m^2}{(1+e^z)^2(-1+(1+e^{-z})^m)}$	no

Some of the commonly applied link functions and the information concerning the log-concavity of the corresponding functions $h(z)$ are displayed in Table 1. It is interesting to note that the size of the parameter m in the skewed logit model determines if the function $h(z)$ is log-concave. The next theorem gives sufficient conditions for the function $h(z)$ and the parameter region Θ such that the D-optimal two point design is unique.

Theorem 3 *If Θ is compact and $h(z)$ is log-concave on $I\!R$ then the maximin D-optimal two point design is unique.*

We are now ready to outline a numerical method to calculate the maximin D-optimal two point designs for binary response models with parametrization (1.1). From a practical point of view it is reasonable to assume that the experimenter can specify a certain range for the position of either parameter before the experiment. This information leads to a rectangular subset of $I\!R \times I\!R^+$ for Θ, i.e. $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. The following result relates maximin D-optimal designs with respect to various rectangles and yields a substantial simplification of the optimization problem.

Lemma 4 *Let ξ^* denote the maximin D-optimal design for the binary response model (1.1) with respect to the parameter space $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$ and $c > 0$, $\Delta \in I\!R$. The standardized maximin D-optimal design ξ^{**} with respect to the parameter space*

$$\tilde{\Theta} = \Big[\frac{\mu_1 + \Delta}{c}, \frac{\mu_2 + \Delta}{c}\Big] \times [c\beta_1, c\beta_2]$$

is given by

$$\xi^{**}(\{x\}) = \xi^*(\{cx - \Delta\}).$$

From Theorem 2 it follows that if the function $h(z)$ is log-concave then for all equally weighted two point designs ξ the minimum of the D-efficiency of ξ over $\vartheta \in \Theta$ can only be attained in the vertices of Θ, i.e.

$$\mathcal{N}(\xi) \subseteq \{(\mu_1, \beta_1), (\mu_1, \beta_2), (\mu_2, \beta_1), (\mu_2, \beta_2)\}.$$

We can now use Theorem 1 for the construction of maximin D-optimal two point designs, where E denotes the set of the vertices of Θ and a prior π on E can be chosen arbitrarily with positive weight for each vertex. Theorem 1 suggests that for all such priors π the maximin D-optimal two point design can be calculated as the limit of the Bayesian D-optimal two point designs with respect to π where q tends to $-\infty$.
For computational reasons, we choose π to be the uniform distribution on E. Thus the Bayesian D-optimal two point design ξ_q (with respect to the prior π) can easily be computed numerically maximizing the function

$$\Big(\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{1}{4}(|M(\xi, \mu_i, \beta_j)|)^q\Big)^{\frac{1}{q}}.$$

We start with an initial value for $q < 0$, e.g., $q_0 = -100$, calculate ξ_{q_0} and $\xi_{q_1} := \xi_{q_0 \cdot 10}$ and compare the respective values of $x_1(q_j)$ and $x_2(q_j)$, $j = 0, 1$. If both absolute distances are less than some threshold value we propose to use ξ_{q_1} as an approximation to the maximin D-optimal two point design. Else we go on computing $\xi_{q_2} = \xi_{q_1 \cdot 10}$, compare the values of $x_1(q_j)$ and $x_2(q_j)$ and so on. Usually four (or even less) steps of this procedure are sufficient to obtain the maximin D-optimal two point design with high accuracy.
It is not possible to show that the Bayesian D-optimal two point designs ξ_q are unique in all circumstances but in case of convergence we will always get the unique maximin D-optimal two

point design [see Theorems 1 and 3]. If the sequence of Bayesian D-optimal designs does not converge as $q \to -\infty$ we propose to change the prior π and start the procedure again.
Recall that in the case of a log-concave function $h(z)$ the worst case set $\mathcal{N}(\xi)$ is a subset of the vertices of the parameter region Θ for all uniform two point designs ξ if the set Θ of parameters specified by the experimenter is a closed rectangle. If the condition of log-concavity of $h(z)$ is not fulfilled the above algorithm cannot be applied without modifications since we have no knowledge about the set $\mathcal{N}(\xi)$. The necessary modifications of our method in situations without log-concavity will be described in the next section.
Below, we give some examples of maximin D-optimal (two point) designs obtained by our method. We calculated the maximin D-optimal designs by the algorithm described in the previous paragraph. The sequence of designs obtained by this algorithm converged to a limit in all cases under consideration in our study. We restrict ourselves to the logit and probit link functions only, to keep the length of this article in acceptable limits. The Maximin D-optimal two point designs for the logistic regression model with respect to several representative situations (concerning different parameter regions Θ; see column 1-4) are shown in the 5th and 6th column of Table 2. The next two columns of this table give the corresponding locally D-optimal designs with respect to parameter values for μ and β, which seem plausible from a heuristic point of view, i.e. μ and β are chosen as the arithmetic means of the intervals $[\mu_1, \mu_2]$ and $[\beta_1, \beta_2]$, respectively. In the two right panels of Table 2 the corresponding designs for the probit model are listed.

Table 2: *Maximin D-optimal two point designs for the logit and probit link functions with respect to several parameter spaces. For comparison: The corresponding locally D-optimal designs with parameters $\mu = (\mu_1 + \mu_2)/2$, $\beta = (\beta_1 + \beta_2)/2$.*

Θ				ξ^* (logit)		local (logit)		ξ^* (probit)		local (probit)	
μ_1	μ_2	β_1	β_2	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
-1	1	$\frac{2}{3}$	$\frac{3}{2}$	-1.295	1.295	-1.425	1.425	-0.698	0.698	-1.051	1.051
-1	1	1	2	-1.018	1.018	-1.029	1.029	-0.505	0.505	-0.759	0.759
0	1	1	2	-0.507	1.507	-0.529	1.529	-0.064	1.064	-0.259	1.259
-0.2	0.2	1	1.5	-1.242	1.242	-1.235	1.235	-0.889	0.889	-0.910	0.910
-0.5	0.5	1	1.5	-1.202	1.202	-1.235	1.235	-0.746	0.746	-0.910	0.910
-0.5	0.5	1	2	-1.007	1.007	-1.029	1.029	-0.564	0.564	-0.759	0.759

It is interesting to see that the locally D-optimal designs (with respect to the center of the rectangle) in the logistic regression model are rather close to the maximin D-optimal two point designs. Thus a practical recommendation for the optimal design of experiment in the logistic regression model is to use the locally D-optimal design for the center $((\mu_1+\mu_2)/2, (\beta_1+\beta_2)/2))$ of the parameter region $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$, if Θ is not too large. On the other hand the situation in the probit model is rather different. Here we observe larger differences between the locally and maximin D-optimal designs and the locally D-optimal designs with respect to the center of the rectangle can be rather inefficient. We finally remark again that the maximin optimal designs given in Table 2 are optimal within the class of all two point designs. The discussion if these designs are optimal within the class of all designs is deferred to Section 4.

3 Extensions

There are some "mild" conditions under which the algorithm is applicable without modifications. We can summarize them as the region of uncertainty Θ being a closed rectangle, which is reasonable from a practical point of view, and log-concavity of the function $h(z)$, which is satisfied for many but not all of the commonly applied link functions. In the following, we will briefly discuss how the method can be modified if the above conditions are not met. In these cases only the prior distribution on the parameter space Θ in the Bayesian optimality criterion has to be modified appropriately and the maximin optimal designs can again be obtained as limits from Bayesian optimal designs.

(a) Assume that $\Theta \subset \mathbb{R} \times \mathbb{R}^+$ is a convex, compact set, but not a rectangle (or polygon). If the function $h(z)$ is log-concave we can choose π as the uniform distribution on the boundary of the set Θ. For computational reasons, this distribution should be approximated by an equidistant grid on the boundary of the set Θ.

(b) There are several cases where two point designs are not optimal and an efficient design ξ with more than two support points is of importance. A typical example for such a situation appears if the optimal two point design is far from the globally optimal design (for a precise formulation see the next section), or if the design should also be used for model checking. Then even if the function $h(z)$ is log-concave the function $|M(\xi, \vartheta)|$ will in general not be log-concave with respect to ϑ for designs ξ with more than two support points. In this case we cannot specify a subset of the parameter space Θ containing $\mathcal{N}(\xi^*)$ and thus have to use the uniform distribution on Θ or its approximation on a grid as a prior in the Bayesian optimality criterion Ψ_q.

(c) If the function $h(z)$ is not log-concave and there is no other way to show that the minimum of $|M(\xi, \vartheta)|$ over Θ is attained at the boundary of Θ we also have to use the uniform distribution on Θ or its approximation by a discrete prior π.

Note that in all these cases the numerical procedure is applicable, where the only modification consists in the choice of the support for the prior distribution π on Θ. We calculate Bayesian Ψ_q-optimal designs and obtain the maximin optimal designs as limit as $q \to -\infty$. However, an additional amount of numerical calculations is required in such a situation because of the more complicated structure of the optimality criterion caused by the larger support of the prior π.

4 Global Optimality and Efficiency

The designs found in Section 2 are maximin D-optimal within the class of all two point designs only. In this section, we will carefully analyze how they perform compared to the corresponding maximin D-optimal designs within the class of all designs. For the sake of brevity we call these designs also globally optimal. A powerful tool for checking optimality of a design is an equivalence theorem, which can be found in Dette, Haines and Imhof (2003).

Theorem 5 *A design ξ^* is maximin D-optimal with respect to Θ if and only if there exists a prior π^* supported on the set $\mathcal{N}(\xi^*)$ such that the inequality*

$$d(\xi^*, x) = \int_{\mathcal{N}(\xi^*)} trace\{I(x, \vartheta) M^{-1}(\xi^*, \vartheta)\}\, d\pi^*(\vartheta) \leq 2 \tag{4.1}$$

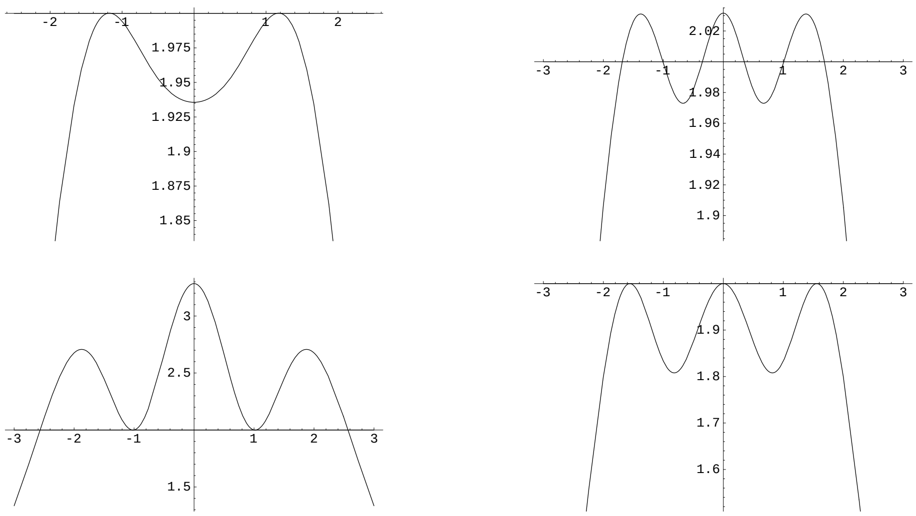
holds for all $x \in I\!R$.

Following Dette, Haines and Imhof (2003) we call the prior π^* least favourable distribution. Checking optimality of the two point designs from Table 2 by the equivalence theorem indicates that designs corresponding to "small" parameter spaces Θ tend to be more efficient than designs corresponding to a larger amount of uncertainty about the position of the unknown parameter. In our examples, the maximin D-optimal two point designs corresponding to the parameter spaces $\Theta = [-0.2, 0.2] \times [1, 1.5]$ and $\Theta = [-0.5, 0.5] \times [1, 1.5]$ turn out to be already optimal within the class of all designs. To give an illustration, Figure 1 shows the function $d(\xi^*, x)$ from the equivalence theorem for some two point designs and a three point design for the logistic regression model.
In the last section, the necessary modifications of our method for determining maximin optimal designs with more than two support points were given. In the cases, where the two-point designs from Table 2 are not globally optimal we calculated the optimal three point designs for the logit example. These designs are listed in Table 3.

Table 3: *Maximin D-optimal three point designs for the logit link function with respect to various parameter spaces.*

Θ				ξ^*					
μ_1	μ_2	β_1	β_2	x_1	x_2	x_3	w_1	w_2	w_3
-1	1	$\frac{2}{3}$	$\frac{3}{2}$	-1.889	0	1.889	0.331	0.338	0.331
-1	1	1	2	-1.559	0	1.559	0.281	0.438	0.281
0	1	1	2	-0.655	0.5	1.655	0.415	0.170	0.415
-0.5	0.5	1	2	-1.155	0	1.155	0.415	0.170	0.415

Figure 1: *The function $d(\xi^*, x)$ from the equivalence theorem for the logistic regression model. Top left: The maximin D-optimal two point design with respect to $\Theta = [-0.5, 0.5] \times [1, 1.5]$ (globally $\Psi_{-\infty}$-optimal). Top right: The maximin D-optimal two point design with respect to $\Theta = [-0.5, 0.5] \times [1, 2]$ (almost globally $\Psi_{-\infty}$-optimal). Bottom left: The maximin D-optimal two point design with respect to $\Theta = [-1, 1] \times [1, 2]$ (far from global $\Psi_{-\infty}$-optimality). Bottom right: The maximin D-optimal three point design with respect to $\Theta = [-1, 1] \times [1, 2]$ (globally $\Psi_{-\infty}$-optimal).*



Applying the equivalence theorem yields that all three point designs listed in Table 3 are maximin D-optimal within the class of all designs. Note that the optimal designs for the rectangles $\Theta = [-0.5, 0.5] \times [1, 2]$ and $\Theta = [0, 1] \times [1, 2]$ are the same except for a shift by 0.5, which is the difference of the means of the respective μ-intervals [see Lemma 4].
A similar result can be obtained for the corresponding maximin D-optimal three point designs in the probit model (see Table 4). Only for the parameter region $\Theta = [-1, 1] \times [1, 2]$, three support points are not sufficient for global maximin D-optimality (see Figure 2). The corresponding maximin D-optimal four point design, which is also optimal within the class of all designs, is given in Table 5. The optimality check by the equivalence theorem for the design corresponding to the parameter region $\Theta = [-1, 1] \times [1, 2]$ is carried out in Figure 2.

Considering the maximin D-optimal three point designs for the respective link functions, it is remarkable that the designs corresponding to the probit model allocate almost half of the weight to the support point in the center. In the logistic case, however, the situation appears to be rather different. If the maximin D-optimal two point design is already close to the maximin D-optimal (three point) design, the outer support points of the maximin D-optimal three point design are fairly close to those of the maximin D-optimal two point design and are assigned a large amount of weight. If, on the other hand, the region of uncertainty is larger, so that

Table 4: *Maximin D-optimal three point designs for the probit link function with respect to various parameter spaces.*

Θ				ξ^*					
μ_1	μ_2	β_1	β_2	x_1	x_2	x_3	w_1	w_2	w_3
-1	1	$\frac{2}{3}$	$\frac{3}{2}$	-1.436	0	1.436	0.262	0.476	0.262
-1	1	1	2	-1.223	0	1.223	0.255	0.490	0.255
0	1	1	2	-0.484	0.5	1.484	0.273	0.454	0.273
-0.5	0.5	1	2	-0.984	0	0.984	0.273	0.454	0.273

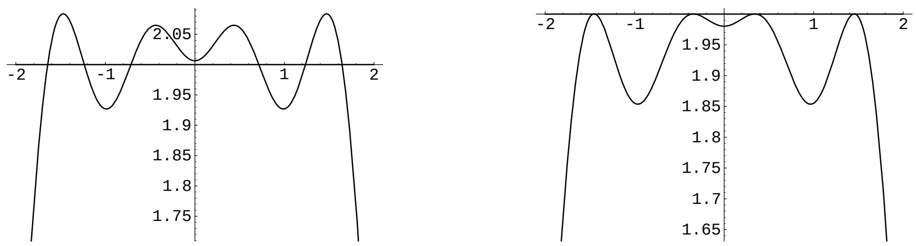
Table 5: *Maximin D-optimal four point design for the probit link function with respect to the parameter space* $\Theta = [-1, 1] \times [1, 2]$.

Θ				ξ^*							
μ_1	μ_2	β_1	β_2	x_1	x_2	x_3	x_4	w_1	w_2	w_3	w_4
-1	1	1	2	-1.442	-0.319	0.319	1.442	0.223	0.277	0.277	0.223

almost four support points are needed for maximin D-optimality, the optimal design assigns more weight to the center point.

The final goal of this section is to investigate the performance of the maximin D-optimal designs (with respect to various parameter spaces Θ) obtained by our method, in particular the performance of the maximin D-optimal two point designs. An obvious action we can take to achieve this aim is to have a close look at the D-efficiencies of the respective designs with respect to different values of ϑ within the particular parameter space. Note that the criterion value $\Psi_{-\infty}(\xi^*)$ of the maximin D-optimal (k point) design ξ^* itself represents the maximum of the minimal D-efficiencies within the corresponding class of designs. It is thus natural from the

Figure 2: *The function* $d(\xi^*, x)$ *from the equivalence theorem in the probit model. Left: The maximin D-optimal three point design with respect to* $\Theta = [-1, 1] \times [1, 2]$ *(almost globally* $\Psi_{-\infty}$*-optimal). Right: The maximin D-optimal four point design with respect to* $\Theta = [-1, 1] \times [1, 2]$ *(globally* $\Psi_{-\infty}$*-optimal).*



choice of our optimality criterion to start our study with the minimal D-efficiencies over the parameter space Θ of the maximin D-optimal two (and more) point designs calculated above. It is illustrated in Table 6 that (even for the globally maximin D-optimal designs) the minimal D-efficiency decreases with larger parameter spaces Θ, which is intuitively clear. Moreover, the loss in D-efficiency caused by a restriction to a k point design is significantly larger in the probit model compared to the logit case, particularly if only the maximin D-optimal two point designs are considered. If, on the other hand, a third design point is added, the increase in minimal D-efficiency is remarkable in the probit model. For the logistic regression, this increase is substantially smaller.

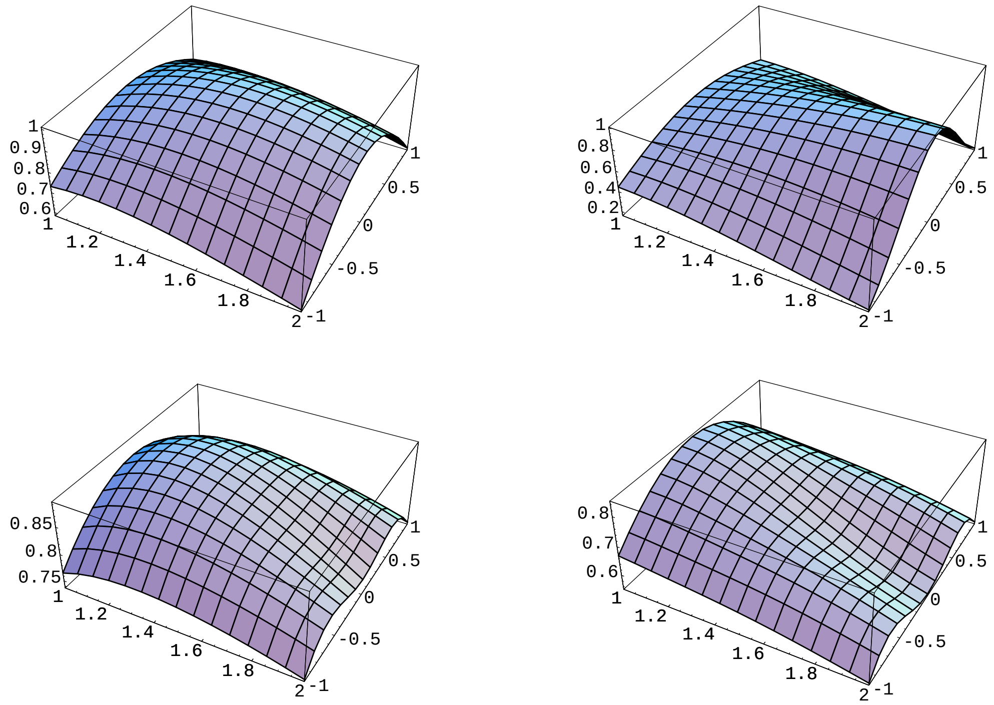
Table 6: *Minimal D-efficiencies of the maximin D-optimal k point designs ξ from the examples of Table 2 (in parentheses: the number of support points of ξ). If there is no entry in the table, the two or three-point design is already globally optimal.*

Θ				minimal D-efficiency of ξ				
μ_1	μ_2	β_1	β_2	logit (2)	probit (2)	logit (3)	probit (3)	probit (4)
-1	1	$\frac{2}{3}$	$\frac{3}{2}$	73.4%	38.2%	78.9%	66.0%	–
-1	1	1	2	59.4%	17.9%	74.0%	54.1%	55.7%
0	1	1	2	84.0%	65.2%	84.5%	73.1%	–
-0.2	0.2	1	1.5	95.8%	93.2%	–	–	–
-0.5	0.5	1	1.5	91.3%	78.7%	–	–	–
-0.5	0.5	1	2	84.0%	65.2%	84.5%	73.1%	–

Note that the D-efficiencies from Table 6 are the minimal D-efficiencies with respect to the respective parameter regions Θ representing the worst case scenario. If the true value of $\vartheta \in \Theta$ is not an element of the support $\mathcal{N}(\xi)$ of the least favourable distribution the D-efficiency of ξ with respect to ϑ can be significantly larger. Figure 3 shows the D-efficiency of the minimax optimal designs if ϑ varies in the specified set Θ. It can be seen that the D-efficiencies of the respective designs increase rapidly if the "true" value has some distance to the support $\mathcal{N}(\xi)$ of the least favourable distribution.

Table 7 gives the D-efficiencies of the maximin optimal designs corresponding to the situation in Figure 3 with respect to some representative "true" values of $\vartheta \in [-1,1] \times [1,2]$. Since the minimal D-efficiencies (at least for two point designs) always occur on the vertices of the parameter region Θ we have already compared the D-efficiencies with respect to the vertices of Θ in Table 6. These results indicate that the maximin D-optimal designs yield reasonable D-efficiencies over a broad range of Θ.

Figure 3: *The D-efficiencies of the maximin D-optimal two point designs with respect to $\Theta = [-1,1] \times [1,2]$ and the corresponding globally optimal designs. Top left: The maximin D-optimal two point design for the logistic regression model. Top right: The maximin D-optimal two point design for the probit model. Bottom left: The globally maximin D-optimal design for the logit model. Bottom right: The globally maximin D-optimal design for the probit case.*



Note that for a fixed value $\vartheta \notin \mathcal{N}(\xi)$ the globally maximin D-optimal designs ξ do not necessarily perform more efficiently than the maximin D-optimal two point designs. On the contrary, the maximin D-optimal two point designs in both models show even higher D-efficiencies towards the center of the parameter region Θ than the corresponding globally optimal designs. A reason for this observation can be found in the structure of the considered optimality criterion. The globally maximin D-optimal designs ξ protect the experiment against the worst case with respect to Θ, i.e. the true value of the parameter vector ϑ is in the set $\mathcal{N}(\xi)$. The optimality of ξ on this set is "bought" at the expense of lower D-efficiencies with respect to parameter values in $\Theta \setminus \mathcal{N}(\xi)$. Since we cannot eliminate the possibility that the "true" parameter is a support point of the least favourable distribution π^*, it is still recommended to use the globally maximin D-optimal designs ξ when Θ is large, because optimal designs with respect to this criterion yield some protection against the worst case scenario, which the maximin D-optimal two point designs cannot accomplish for large parameter regions Θ.

In the following, we consider another somewhat related indicator for the performance of a design ξ with a restricted number of support points, i.e. its efficiency with respect to the criterion

Table 7: *Some D-efficiencies of the maximin D-optimal two point and globally maximin D-optimal designs ξ with respect to the parameter region $\Theta = [-1, 1] \times [1, 2]$.*

True ϑ		D-Efficiency of ξ			
μ	β	logit (2)	probit (2)	logit (global)	probit (global)
0	1.5	100%	87.6%	87.2%	72.0%
-0.5	1.25	90.9%	66.9%	86.7%	77.3%
0.5	1.25	90.9%	66.9%	86.7%	77.3%
-0.5	1.75	89.2%	69.1%	82.7%	67.7%
0.5	1.75	89.2%	69.1%	82.7%	67.7%

function $\Psi_{-\infty}$,

$$\mathrm{eff}_{\Psi_{-\infty}}(\xi) = \frac{\Psi_{-\infty}(\xi)}{\max_\eta \Psi_{-\infty}(\eta)},$$

where the maximum is taken over the set of all designs. For the maximin D-optimal two point designs $\Psi_{-\infty}$-efficiencies are listed in Table 8.

Table 8: $\Psi_{-\infty}$-*efficiencies of the Maximin D-optimal two point designs ξ, which are not globally optimal.*

Θ				$\mathrm{eff}_{\Psi_{-\infty}}(\xi)$	
μ_1	μ_2	β_1	β_2	logit	probit
-1	1	$\frac{2}{3}$	$\frac{3}{2}$	92.9%	57.9%
-1	1	1	2	80.2%	32.1%
0	1	1	2	99.3%	89.2%
-0.5	0.5	1	2	99.3%	89.2%

For the logistic regression model the efficiencies taken from Table 8 are quite high, so that for moderate regions Θ it seems sufficient to use the corresponding maximin D-optimal two point design provided that the data should not be used for a goodness-of-fit test of the model. However, the use of the maximin D-optimal two point design in the probit model can only be recommended for small parameter spaces. As pointed out before, the implementation of our method is much easier if only two point designs are required. Furthermore, it may reduce costs for the experimenter, if the number of support points is as low as possible (depending on the kind of experiment to be carried out). If, on the other hand, the parameter space specified by the experimenter is larger, it is recommended to use the modified algorithm and calculate the globally maximin D-optimal design [see Section 3]. Another situation, in which the determination of maximin D-optimal designs with more than two support points is necessary, occurs if the design should also be used for model checking.

5 Appendix: Proofs

Proof of Theorem 1: The first part is a special case of Theorem 3.1 from Dette, Haines and Imhof (2003) where the criterion function $\psi(M(\xi,\vartheta))$ is given by the (square root of the) determinant of the Fisher information, which is obviously continuous in each argument.
The proof of the second part follows along the same lines as the proof of the above-mentioned theorem where the auxiliary parameter ϑ_0 is chosen from E instead of Θ. If the set E is discrete it is not necessary (and possible) to define open neighborhoods within the set E around the elements $\vartheta_E \in E$. Instead of its neighborhood the respective ϑ_E itself can be inserted in the proof of Theorem 3.1 from Dette, Haines and Imhof (2003). $\Box$

Proof of Theorem 2: Let, for the sake of typographical brevity, $\delta(x)$ denote the second derivative of log $h(z)$ evaluated at the point $\beta(x-\mu)$. Using the formula

$$|M(\xi,\vartheta)| = \frac{1}{4} h(\beta(x_1-\mu))h(\beta(x_2-\mu))(x_2-x_1)^2\beta^2$$

for the determinant of the information matrix $M(\xi,\vartheta)$ of a two point design ξ with equal weights at the support points $x_1,\ x_2$ we calculate the eigenvalues of the Hessian H of $\log|M(\xi,\vartheta)|$ with respect to μ and β as

$$\lambda_{1,2} = \frac{H_{11}+H_{22}}{2} \pm \sqrt{\frac{(H_{11}+H_{22})^2}{4} - H_{11}H_{22} + H_{12}^2},$$

where $H_{ij},\ i,j=1,2$, denote the entries of H. Since

$$\begin{aligned} H_{11} &= \beta^2\,\delta(x_1) + \beta^2\,\delta(x_2), \quad H_{22} \;=\; -\frac{2}{\beta} + (x_1-\mu)^2\,\delta(x_1) + (x_2-\mu)^2\,\delta(x_2), \\ H_{12} &= -\frac{h'(\beta(x_1-\mu))}{h(\beta(x_1-\mu))} - \beta(x_1-\mu)\,\delta(x_1) - \frac{h'(\beta(x_2-\mu))}{h(\beta(x_2-\mu))} - \beta(x_2-\mu)\,\delta(x_2) \end{aligned}$$

we see immediately that the smaller eigenvalue is less than zero if $\delta(x)<0$ for all $x \in I\!R$, i.e. $h(z)$ log-concave in $z \in I\!R$. The product of the eigenvalues $\lambda_1\lambda_2$ turns out to be

$$-2(\delta(x_1)+\delta(x_2))+\beta^2[(x_1-\mu)^2\,\delta^2(x_1)+(x_2-\mu)^2\,\delta^2(x_2)+\delta(x_1)\delta(x_2)\{(x_1-\mu)^2+(x_2-\mu)^2\}]>0$$

if, again, $\delta(x)<0$ for all $x \in I\!R$. Hence the larger eigenvalue is also negative and $\log|M(\xi,\vartheta)|$ is a concave function of ϑ for all equally weighted two point designs ξ if the function $h(z)$ is log-concave. $\Box$

Proof of Theorem 3: For the sake of typographical brevity, we define the D-efficiency function $\Phi(\xi,\vartheta) := (|M(\xi,\vartheta)|/|M(\xi_\vartheta,\vartheta)|)^{1/2}$. It is straightforward to show that for every equally weighted two point design ξ the function $|M(\xi,\vartheta)|$ and thus the square root of the standardized version $\Phi(\xi,\vartheta)$ is log-concave with respect to the design points $x_1,\ x_2$ for all $\vartheta \in \Theta$ if $h(z)$ is log-concave. The arguments are the same as in the proof of Theorem 2. (Note that the maximin D-optimal two point design must have equal weights [see Silvey (1980)].)

Define two different designs $\xi^{(1)}$, $\xi^{(2)}$ with support points $x_i^{(1)}$, $x_i^{(2)}$, $i = 1, 2$, respectively, and construct a further design $\xi^{(1,2)}$ by $x_i^{(1,2)} = (x_i^{(1)} + x_i^{(2)})/2$, $i = 1, 2$. For fixed values of ϑ we have

$$\Phi(\xi^{(1,2)}, \vartheta) > \min\{\Phi(\xi^{(1)}, \vartheta), \Phi(\xi^{(2)}, \vartheta)\} \tag{A.1}$$

because of the log-concavity of $\Phi(\xi, \vartheta)$ in x_1, x_2. Let now $\xi^{(1)}$, $\xi^{(2)}$ be two maximin D-optimal two point designs with optimal criterion value $\Psi^*_{-\infty}$, i.e. $\min_{\vartheta\in\Theta} \Phi(\xi^{(1)}, \vartheta) = \min_{\vartheta\in\Theta} \Phi(\xi^{(2)}, \vartheta) = \Psi^*_{-\infty}$. In particular, we have

$$\Phi(\xi^{(1)}, \vartheta) \geq \Psi^*_{-\infty}, \ \ \Phi(\xi^{(2)}, \vartheta) \geq \Psi^*_{-\infty} \qquad \forall\, \vartheta \in \Theta. \tag{A.2}$$

The above minima exist because Θ is assumed to be compact, which also implies that the set $\mathcal{N}(\xi^{(1,2)})$ is not empty. From the optimality of the value $\Psi^*_{-\infty}$ we obtain for all $\vartheta^* \in \mathcal{N}(\xi^{(1,2)})$ the inequality $\Phi(\xi^{(1,2)}, \vartheta^*) \leq \Psi^*_{-\infty}$. But from (A.1) and (A.2) it follows that

$$\Phi(\xi^{(1,2)}, \vartheta^*) > \min\{\Phi(\xi^{(1)}, \vartheta^*), \Phi(\xi^{(2)}, \vartheta^*)\} \geq \Psi^*_{-\infty},$$

which is a contradiction to the assumption that there exist more than one $\Psi_{-\infty}$-optimal two point design. □

Proof of Lemma 4. Since the value of the determinant of the locally D-optimal design $|M(\xi_\vartheta, \vartheta)|$ does not depend on $\vartheta \in \Theta$, the standardized maximin D-optimal design with respect to $\tilde{\Theta} = [\frac{\mu_1+\Delta}{c}, \frac{\mu_2+\Delta}{c}] \times [c\beta_1, c\beta_2]$ can be obtained by maximizing

$$\begin{aligned}
&\min_{\substack{\beta\in[c\beta_1, c\beta_2] \\ \mu\in\left[\frac{\mu_1+\Delta}{c}, \frac{\mu_2+\Delta}{c}\right]}} \left| \int h(\beta(x-\mu)) \begin{pmatrix} \beta^2 & -\beta(x-\mu) \\ -\beta(x-\mu) & (x-\mu)^2 \end{pmatrix} d\xi(x) \right| \\
= &\min_{\substack{\beta\in[\beta_1, \beta_2] \\ \mu\in[\mu_1, \mu_2]}} \left| \int h(c\beta(x - \frac{\mu+\Delta}{c})) \begin{pmatrix} c^2\beta^2 & -c\beta(x - \frac{\mu+\Delta}{c}) \\ -c\beta(x-\frac{\mu+\Delta}{c}) & (x - \frac{\mu+\Delta}{c})^2 \end{pmatrix} d\xi(x) \right| \\
= &\min_{\substack{\beta\in[\beta_1, \beta_2] \\ \mu\in[\mu_1, \mu_2]}} \left| \int h(\beta(x-\mu)) \begin{pmatrix} \beta^2 & -\beta(x-\mu) \\ -\beta(x-\mu) & (x-\mu)^2 \end{pmatrix} d\tilde{\xi}(x) \right|
\end{aligned}$$

where the design $\tilde{\xi}$ is obtained from ξ by the relation

$$\tilde{\xi}(\{x\}) = \xi(\{cx - \Delta\}).$$

The assertion of Lemma 4 is now obvious. □

Acknowledgements: The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2) is gratefully acknowledged. The authors are grateful to Isolde Gottschlich, who typed parts of this paper with considerable technical expertise.

References

[1] Abdelbasit, K.M. and Plackett, R.L. (1983). Experimental design for binary data. *Journal of the American Statistical Association*, 78, 90-98.

[2] Chaloner, K. and Larntz, K. (1989). Optimal Bayesian experimental design applied to logistic regression. *Journal of Statistical Planning and Inference*, 21, 191-208.

[3] Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Annals of Mathematical Statistics*, 24, 586-602.

[4] Dette, H. (1997). Designing experiments with respect to standardized optimality criteria. *Journal of the Royal Statistical Society, Ser. B*, 59, No.1, 97-110.

[5] Dette, H., Haines, L. and Imhof, L. (2003). Maximin and Bayesian optimal designs for regression models. http://www.ruhr-uni-bochum.de/mathematik3/preprint.htm

[6] Dette, H. and Wong, W.K. (1996). Optimal Bayesian designs for models with partially specified heteroscedastic structure. *Annals of Statistics*, 24, No. 5, 2108-2127.

[7] Fandom Noubiap, R. and Seidel, W. (2000). A minimax algorithm for constructing optimal symmetrical balanced designs for a logistic regression model. *Journal of Statistical Planning and Inference*, 91, No. 1, 151-168.

[8] Ford, I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society*, 54, 569-583.

[9] Imhof, L. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, 29, No. 2, 561-576.

[10] Kalish, L.A. and Rosenberger, J.L. (1978). Optimal designs for the estimation of the logistic function. *Technical Report 33*, Department of Statistics, Pennsylvania State University.

[11] Minkin, S. (1987). Optimal designs for binary data. *Journal of the American Statistical Association*, 82, 1098-1103.

[12] Müller, C. H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *Journal of Statistical Planning and Inference*, 44, 117-132.

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

[14] Sitter, R.R. (1992). Robust Designs for Binary Data. *Biometrics,* 48, 1145-1155.

[15] Sitter, R.R. and Fainaru, I. (1997). Optimal designs for the logit and probit models for binary data. *Canadian Journal of Statistics*, 25, No. 2, 175-190.

[16] Sitter, R.R. and Wu, C.F.J. (1993). Optimal designs for binary response experiments: Fieller-, D- and A-criteria. *Scandinavian Journal of Statistics,* 20, 329-341.