The Canadian Journal of Statistics
Vol. 34, No. 4, 2006, Pages ???-???
La revue canadienne de statistique

Some robust design strategies for percentile estimation in binary response models

Stefanie BIEDERMANN, Holger DETTE and Andrey PEPELYSHEV

Key words and phrases: Binary response model; robust optimal design; c-efficiency; percentile estimation; multi-objective designs.
MSC 2000: Primary 62K05; secondary 62J12.

Abstract: For the problem of percentile estimation of a quantal response curve, the authors determine multi-objective designs which are robust with respect to misspecifications of the model assumptions. They propose a maximin approach based on efficiencies and provide designs that are simultaneously efficient with respect to the particular choice of various parameter regions and link functions. Furthermore, the authors deal with the problems of designing model and percentile robust experiments and give various examples of such designs, which are calculated numerically.

1. INTRODUCTION

We consider an experiment where the response is dichotomous and the probability of toxicity p increases with the log-dose level x of a drug. A $100p$th percentile x_p is the log-dose level corresponding to a desired probability of toxicity p, i.e. for a sample space of doses $\mathcal{X} \subset \mathbb{R}$ the percentile $x_p \in \mathcal{X}$ corresponds to a target probability of toxicity $p \in (0,1)$. Estimation may be facilitated by assuming that the response function follows some probability distribution from a location scale family. If we let the probability of toxicity at log-dose level x be modelled as $F(\beta x + \mu)$, i.e.

$$p = p(x, \vartheta) = F(\mu + \beta x), \qquad \vartheta = (\mu, \beta)^T,\ \mu \in \mathbb{R},\ \beta \in \mathbb{R}^+, \tag{1}$$

where μ and β are unknown parameters and F is a strictly increasing distribution function, then we obtain the following expression for the $100p$th percentile

$$x = x_p = \frac{F^{-1}(p) - \mu}{\beta}. \tag{2}$$

As proposed in the literature, we consider log-doses instead of original doses to avoid possible negative dose levels as the optimal support points of the design problem.

An (approximate) design ξ is a probability measure with finite support on $\mathbb{R}$, i.e. the observations are taken at the support points of the measure proportional to the corresponding masses. 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). In the present context, the Fisher information for the parameter ϑ depends on the unknown value of ϑ itself, which complicates the determination of optimal designs substantially. Much effort has been devoted to the problem of finding good designs for the estimation of the parameter ϑ in the binary response model (see, e.g., Chaloner & Larntz 1989; Sitter & Wu 1993; among many others).

The goal of this article is to provide designs ξ, which are on the one hand efficient for estimating the 100pth percentile x_p and on the other hand robust against misspecifications of the model assumptions. Wu (1988) shows that designing an experiment optimal for percentile estimation is equivalent to calculating the c-optimal design for the estimation of a linear combination of the parameters μ and β with respect to a particular choice of the vector c. Locally (in the sense of Chernoff 1953) c-optimal designs for the above model can be derived from the article of Ford, Torsney & Wu (1992). Since the Fisher information and thus the c-optimal designs depend on the unknown parameter these designs cannot be implemented directly in practice. Assuming that a good initial guess for the unknown parameter ϑ is available is often not realistic in real applications. Moreover, misspecifications of the parameter ϑ for the construction of a design can lead to a loss of efficiency (see, e.g., López-Fidalgo & Wong 2000) and as a consequence 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 optimality criterion (or an appropriately standardized version thereof) over the plausible values of ϑ defined by the prior. This leads to so-called Bayesian optimality criteria (see, e.g., Chaloner & Larntz 1989).

As an alternative for the construction of robust designs, we propose a maximin approach based on c-efficiencies, which only requires the specification of a certain range for the unknown parameter. It appears to us 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 two-dimensional. A (non-standardized) maximin approach has been used by Sitter (1992), King & Wong (2000) and Fandom Noubiap & Seidel (2000) for the problem of parameter estimation in binary response models. In contrast to these authors, we discuss a minimax concept based on efficiencies, because as pointed out in Dette (1997) the consideration of non-standardized quantities can be rather misleading. The optimal designs with respect to our approach are calculated numerically using the Nelder-Mead algorithm in a standard MATLAB environment with the aid of the corresponding non-standard equivalence theorems.

In Section 2, we describe the model in detail and derive the connection between optimality for percentile estimation and c-optimality. In Section 3, the robust optimality criterion used in the following is illustrated, and maximin efficient designs are given for several choices of spaces Θ for the unknown parameter ϑ. These results are then compared to the corresponding Bayesian optimal designs with respect to the uniform prior distribution on Θ. While the problem of robustness of an optimal design with respect to the unknown parameter is caused by the nonlinear structure of the binary response model, there are other issues of robustness which arise in applications, because a specific model or a particular percentile is chosen, for which the determined design should be efficient. For this reason, Section 4 extends the maximin efficiency approach, dealing with maximin efficient designs for percentile estimation for competing link functions. In Section 5, we present some results on designs, which are maximin efficient for estimating several percentiles simultaneously. A somewhat related approach is given in Zhu & Wong (2000), who derive designs, which are (non-standardized) Bayesian c-optimal (with respect to several prior distributions for the unknown parameter ϑ) for the estimation of $x_{.5}$, the $ED50$, under the constraint that the Bayesian efficiencies for estimating $x_{.25}$ and $x_{.75}$ separately will not be lower than some given values. Furthermore, Zhu & Wong (2001) provide Bayesian compound optimal designs for estimating sets of symmetrical quantiles. We demonstrate the usefulness of our approach in Section 6, where we apply our designs towards redesigning a real data example (see Rosenberger & Grill 1997), and compare their performance with several designs used in practice. The concluding Section 7 summarizes our results and gives a detailed discussion of the various aspects of robustness in the design of experiment for the binary response setup. The proofs of our results are finally deferred to an appendix.

2. LOCALLY OPTIMAL DESIGNS

Before we present our results on maximin efficient designs for percentile estimation, we give a short summary of the results on the corresponding locally optimal designs, which were already established by Wu (1988). We consider the common binary response model where a subject is administered a stimulus at a log-dose level $x \in I\!R$. The response is a binary random variable with success probability $p = p(x, \vartheta)$, where $x \in I\!R$ is the explanatory variable and ϑ is an unknown parameter. In this article, we deal with parameterization (1) of a two parameter binary response model, where F denotes a so-called link function, i.e. a known strictly increasing distribution function with density f. Having chosen a distribution family for F, one obtains maximum likelihood estimators $\hat{\mu}$, $\hat{\beta}$ for the unknown parameters and $\hat{x}_p = (F^{-1}(p) - \hat{\mu})/\hat{\beta}$ for the 100pth percentile x_p. The information matrix for a single log-dose x associated with $\hat{\mu}$ and $\hat{\beta}$ is given by

$$I(x, \vartheta) = h(z) \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix},$$

where $z = \beta x + \mu$, and the function $h(z)$ is defined as

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

The Fisher information of a design ξ is thus given by

$$M(\xi, \vartheta) = \int I(x, \vartheta)\, d\xi(x)$$

and also called information matrix in the design literature. The notations $I(x, \vartheta)$, $M(\xi, \vartheta)$ indicate the dependence of the Fisher information on the particular value of the unknown parameter ϑ. By inverting the information matrix $M(\xi, \vartheta)$ of the underlying design ξ, we obtain a first-order approximation to the variance-covariance matrix of the maximum likelihood estimate $(\hat{\mu}, \hat{\beta})^T$. If n observations are taken according to an approximate design (using an appropriate rounding procedure), we derive an approximation to the asymptotic variance of $\hat{x}_p$ as

$$Var(\hat{x}_p) \approx \frac{1}{n} (\nabla x_p)^T M^-(\xi, \vartheta)(\nabla x_p), \quad \nabla x_p = \left(\frac{\partial x_p}{\partial \mu}, \frac{\partial x_p}{\partial \beta} \right)^T,$$

where the gradient of x_p is given by $c = c_p = -(1, x_p)^T/\beta$ and M^- denotes a generalized inverse of the matrix M. Here and throughout this paper, we assume that the linear combination $c_p^T \vartheta$ is estimable by the design, i.e. $c_p \in \mathrm{range}(M(\xi, \vartheta))$ for all $\vartheta \in \Theta$, which guarantees that the approximation for the variance of $\hat{x}_p$ is independent of the choice of the generalized inverse (see, e.g., Pukelsheim 1993, Chapter 3). An optimal design for estimating the 100pth percentile minimizes $Var(\hat{x}_p)$ or its approximation. Thus the problem of finding a (locally) optimal design for estimating x_p is equivalent to the problem of finding a locally c_p-optimal design, i.e. a design which minimizes

$$c_p^T M^-(\xi, \vartheta) c_p. \tag{3}$$

These designs are called locally optimal, since the minimization of (3) depends on the parameter $\vartheta = (\mu, \beta)^T$ (see Chernoff 1953).

Wu (1988) shows that the c_p-optimal design problem for the binary response model (1) can be solved simultaneously for all parameters. There exist values $p_1 < p_2$ depending on the link function F such that for $p_1 \le p \le p_2$ the locally c_p-optimal design is the one-point design concentrating at the 100pth percentile x_p defined in (2), whereas for $p < p_1$ or $p > p_2$ the locally c_p-optimal design is supported on the two percentiles x_{p_1} and x_{p_2}, while the corresponding weights can be calculated from an explicit formula depending on p, p_1, p_2 and the particular choice of F. For the logistic link function we obtain the values $p_2 = 1 - p_1 = .917$. The normal link function yields $p_2 = 1 - p_1 = .942$. Note that while the design points x_{p_1} and x_{p_2} depend on the unknown parameter ϑ, the values p_1 and p_2 do only depend on the link function F under consideration.

The most commonly applied link functions are the logistic distribution and the normal distribution, and we will restrict ourselves to these two link functions in the following. Other links, which are also used in practice, are the complementary log-log, the double exponential, the double reciprocal and the skewed logistic with several choices of skewing parameters (see Ford, Torsney & Wu 1992) and can be treated by similar methods.

3. EFFICIENT DESIGNS UNDER A GIVEN MODEL ASSUMPTION

A general objection to one-point designs is that they do not allow the estimation of both parameters μ and β with the same data. In the context of sequential experiments, however, such designs were shown to be robust to model assumptions (see Wu 1985). A disadvantage of collecting data sequentially is that they exhibit dependencies, which can lead to problems when it comes to data analysis.

Thus, we consider a non-sequential way of obtaining efficient designs for percentile estimation, which are robust against misspecifications of the unknown parameter $\vartheta = (\mu, \beta)^T$. Our approach is motivated by our experience that in many cases it is possible to specify a certain range, $\Theta \subset \mathbb{R} \times \mathbb{R}^+$ say, for the unknown parameters. Following Müller (1995), we propose to use designs, which maximize the minimal c_p-efficiency, $\min \text{eff}_\Theta(\xi)$, over the set Θ, thus protecting the experiment against the worst case scenario. In other words, we propose to maximize the expression

$$\min \text{eff}_\Theta(\xi) = \min_{\vartheta \in \Theta} \text{eff}_{c_p}(\xi, \vartheta) = \min_{\vartheta \in \Theta} \frac{c_p^T M^-(\xi^*_{c_p}, \vartheta) c_p}{c_p^T M^-(\xi, \vartheta) c_p}$$

with respect to the choice of the design ξ, where $\xi^*_{c_p}$ denotes the locally c_p-optimal design and $M^-(\xi^*_{c_p}, \vartheta)$ is a generalized inverse of the information matrix $M(\xi^*_{c_p}, \vartheta)$. Since the locally c_p-optimal designs are one point designs in many cases, this definition takes singular Fisher information matrices into account. Note that the vector c_p and the design $\xi^*_{c_p}$ depend on the parameter ϑ, although this dependence is not reflected in our notations. Throughout this section a maximin efficient design will be denoted by ξ^*_M.

A powerful tool for checking the optimality of a given design are equivalence theorems. The equivalence theorem for local c_p-optimality, which is given below, is well-known in the design literature. The proof can, for example, be found in Pukelsheim (1993).

THEOREM 1. *A design $\xi^*_{c_p} = \xi^*_{c_p}(\vartheta)$ is locally c_p-optimal if and only if there exists a generalized inverse $G(\xi^*_{c_p}, \vartheta)$ of $M(\xi^*_{c_p}, \vartheta)$ such that the following inequality holds*

$$\frac{c_p^T G(\xi^*_{c_p}, \vartheta) I(x, \vartheta) G(\xi^*_{c_p}, \vartheta) c_p}{c_p^T M^-(\xi^*_{c_p}, \vartheta) c_p} \leq 1 \qquad \forall\, x \in \mathbb{R}. \tag{4}$$

Moreover, there is equality in (4), when x is a support point of the design $\xi^*_{c_p}$.

The equivalence theorem corresponding to the maximin c_p-efficiency criterion can be obtained as a special case of Theorem 3.3 b) in Dette, Haines & Imhof (2006).

THEOREM 2. *Let*

$$\mathcal{N}(\xi) = \{\tilde{\vartheta} \in \Theta \mid \mathit{eff}_{c_p}(\xi, \tilde{\vartheta}) = \min_{\vartheta \in \Theta} \mathit{eff}_{c_p}(\xi, \vartheta)\},$$

be the subset of Θ consisting of those values of ϑ, for which the c_p-efficiency of a design ξ takes its minimal value over Θ. A design ξ^*_M is maximin c_p-efficient if and only if for each $\vartheta \in \mathcal{N}(\xi^*_M)$

there exists a generalized inverse $G(\xi_M^*, \vartheta)$ of $M(\xi_M^*, \vartheta)$ and a prior π^* on $\mathcal{N}(\xi_M^*)$ such that the following inequality is valid

$$\int_{\mathcal{N}(\xi_M^*)} \frac{c_p^T G(\xi_M^*, \vartheta) I(x, \vartheta) G(\xi_M^*, \vartheta) c_p}{c_p^T M^-(\xi_M^*, \vartheta) c_p} d\pi^*(\vartheta) \leq 1 \qquad \forall\, x \in I\!R. \tag{5}$$

Again, there is equality in (5), when x is a support point of the design ξ_M^*.

Following Dette, Haines & Imhof (2006), we call any prior π^* in (5) least favourable distribution. The maximin efficient designs given in the following discussion were found numerically making use of the Nelder-Mead algorithm in the standard MATLAB environment (optimization toolbox), since methods that use only function evaluations such as the simplex search of Nelder and Mead are most suitable for problems where derivative information is not available. We first calculated the optimal k-point design (usually starting with $k = 2$), and checked the optimality of this design by Theorem 2. If the optimal design supported at k points proved to be not optimal among all designs, we computed the optimal $(k + 1)$-point design and so on. More computational details can be obtained from the third author. Moreover, we intend to make computer codes available online in the medium-term within the framework of some more comprehensive design project so that practitioners will be enabled to apply the maximin efficient designs proposed in this article.

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. $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. In the following, we will thus consider only parameter regions Θ of this form. Lemma 1 and Lemma 2 below give considerable simplifications of the maximin optimal design problem for parameter regions of the form $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. For conciseness of presentation, the proofs of these lemmata have been deferred to the appendix. The first result describes a symmetry property of the maximin efficient designs for estimating the percentiles x_p and x_{1-p} in case of an underlying symmetric distribution function F.

LEMMA 1. *Let F be a symmetric distribution function, i.e. $F(x) = 1 - F(-x)$. Let, furthermore, $\xi = \{x_1, \ldots, x_n; w_1, \ldots, w_n\}$ denote the maximin efficient design for estimating the percentile x_p with respect to the parameter space $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. Then the design $\{-x_1, \ldots, -x_n; w_1, \ldots, w_n\}$ is maximin efficient for estimating the percentile x_{1-p} with respect to $[-\mu_2, -\mu_1] \times [\beta_1, \beta_2]$.*

Our next result refers to the relationship between the maximin efficient designs for estimating the percentile x_p with respect to parameter regions Θ of a certain form, i.e. if the β-intervals are scaled by a positive parameter γ the maximin efficient designs for the new parameter spaces can be obtained by rescaling the "old" designs. This result shows that maximin efficient designs are equivariant with respect to different scalings of the explanatory variable x. Moreover, it also enables us to restrict ourselves to consider only β-intervals with one fixed parameter value, e.g. $\beta_1 = 1$.

LEMMA 2. *Let $\xi = \{x_1, \ldots, x_n; w_1, \ldots, w_n\}$ denote the maximin efficient design for estimating the percentile x_p with respect to the parameter region $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. Then the design $\xi_\gamma = \{x_1/\gamma, \ldots, x_n/\gamma; w_1, \ldots, w_n\}$ is maximin efficient for estimating the percentile x_p with respect to the parameter region $[\mu_1, \mu_2] \times [\gamma\beta_1, \gamma\beta_2]$ for all choices of $\gamma > 0$.*

It is obvious that maximin efficient designs must be supported on more than one point, since a one-point design does not allow the estimation of the percentile x_p with respect to different unknown parameters. From a heuristic point of view it is also clear that if the length $\mu_2 - \mu_1$ and $\beta_2 - \beta_1$ of the parameter intervals is small then the maximin efficient designs for estimating

Table 1: *Maximin efficient designs for percentile estimation in the logistic model with respect to various "small" parameter regions $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$. In these cases the optimal designs are supported at two points.*

p	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	$\min \mathrm{eff}_\Theta(\xi_M^*)$	$b(\xi_M^*)$
.2	-.1	.1	1	1.1	-1.381	.903	.938	.062	.8888	.9342
.4	-.1	.1	1	1.1	-.751	.189	.621	.379	.8986	.9329
.5	-.1	.1	1	1.1	-.423	.423	.500	.500	.9121	.9385
.6	-.1	.1	1	1.1	-.189	.751	.379	.621	.8986	.9329
.7	-.1	.1	1	1.1	-.363	.992	.149	.851	.8891	.9303
.8	-.1	.1	1	1.1	-.903	1.381	.062	.938	.8888	.9342
.9	-.1	.1	1	1.1	-1.914	1.964	.054	.946	.9230	.9609
.9	-.2	.2	1	1.2	-1.664	1.831	.094	.906	.8359	.9183
.9	-.3	.3	1	1.3	-1.375	1.756	.124	.877	.7509	.8769
.95	-.1	.1	1	1.1	-2.254	2.232	.111	.890	.9728	.9949
.95	-.2	.2	1	1.2	-2.104	2.061	.141	.859	.9108	.9758
.95	-.3	.3	1	1.3	-1.963	1.923	.171	.829	.8349	.9494

percentiles x_p with $p_1 \le p \le p_2$ take the form $\{x_1, x_2; w, 1-w\}$, where $x_1 \approx x_p$ and $w \approx 1$ (see Wu 1988). In other words, the maximin efficient design ξ_M^* is close to the locally c_p-optimal designs with respect to $(\mu, \beta)^T \in [\mu_1, \mu_2] \times [\beta_1, \beta_2]$.

These presumptions are confirmed by our numerical results, which are given in Tables 1 and 2. Table 1 presents some maximin efficient designs with respect to several relatively small parameter intervals for the logistic distribution

$$F(x) = 1/(1+e^{-x}),\ \mathcal{X} = I\!R,$$

whereas Table 2 gives the corresponding designs for the normal distribution

$$F(x) = \int_{-\infty}^{x} e^{-u^2/2} du/\sqrt{2\pi},\ \mathcal{X} = I\!R.$$

For "large" parameter regions Θ we obtain maximin efficient designs with more than two support points (see Tables 3 and 4). This phenomenon is widely observed in the literature as far as Bayesian or maximin optimal design problems are concerned (see, e.g., Dette & Biedermann 2003; Dette & Wong 1996). The expression

$$b(\xi_M^*) = \int_\Theta \mathrm{eff}_{c_p}(\xi_M^*, \vartheta)\, d\vartheta / \lambda(\Theta)$$

in the tables denotes the value of the integrated (with respect to the uniform distribution) c_p-efficiency of the maximin efficient design over the parameter region Θ and is given for comparison with Bayesian designs with respect to the uniform prior, which will be discussed below. Here $\lambda(\Theta)$ denotes the Lebesgue measure of Θ, i.e. $\lambda(\Theta) = (\mu_2 - \mu_1)(\beta_2 - \beta_1)$.

We observe from Tables 1, 2, 3 and 4 that the support points of the optimal designs are much more spread on the real axis in case of an underlying logistic distribution than they are for the normal distribution. For several parameter regions Θ and percentiles x_p, the maximin c_p-efficient designs with respect to the normal model even require more support points than the respective designs for the logistic link function (see Tables 3 and 4). Furthermore, the minimal efficiencies

Table 2: *Maximin efficient designs for percentile estimation in the normal model with respect to various "small" parameter regions* $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$.

p	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	$\min \mathrm{eff}_\Theta(\xi_M^*)$	$b(\xi_M^*)$
.2	-.1	.1	1	1.1	-.891	.615	.906	.094	.8663	.9185
.4	-.1	.1	1	1.1	-.558	.261	.616	.384	.8848	.9224
.5	-.1	.1	1	1.1	-.382	.382	.500	.500	.8960	.9269
.6	-.1	.1	1	1.1	-.260	.558	.384	.616	.8848	.9225
.7	-.1	.1	1	1.1	-.371	.682	.189	.811	.8741	.9193
.8	-.1	.1	1	1.1	-.615	.891	.094	.906	.8663	.9185
.9	-.1	.1	1	1.1	-1.066	1.202	.059	.941	.8744	.9305
.9	-.2	.2	1	1.2	-.878	1.162	.108	.892	.7541	.8665
.9	-.3	.3	1	1.3	-.680	1.148	.150	.850	.6447	.8087
.95	-.1	.1	1	1.1	-1.371	1.399	.076	.924	.9243	.9718
.95	-.2	.2	1	1.2	-1.216	1.290	.123	.877	.8076	.9221
.95	-.3	.3	1	1.3	-1.090	1.213	.165	.835	.6893	.8670

Table 3: *Maximin efficient designs for percentile estimation in the logistic model with respect to various "moderate" and "large" parameter regions* $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$.

p	μ_1	μ_2	β_1	β_2	x_1	x_2	x_3	x_4	w_1	w_2	w_3	w_4	$\min \mathrm{eff}_\Theta(\xi_M^*)$	$b(\xi_M^*)$
.5	-1	1	1	2	-1.812	-.675	.675	1.812	.144	.356	.356	.144	.4948	.5941
.5	-2	2	1	4	-2.264	-.518	.518	2.264	.181	.319	.319	.181	.2519	.3919
.8	-1	1	1	2	-.126	.986	2.333		.286	.294	.420		.4736	.5898
.8	-2	2	1	4	-1.105	-.100	.858	2.741	.075	.255	.365	.305	.2609	.4045
.9	-.4	.4	1	1.4	-1.666	.184	1.876		.082	.122	.796		.6787	.8136
.9	-1	1	1	2	.118	1.051	2.591		.197	.307	.496		.4756	.5826
.9	-2	2	1	4	-.121	.420	1.200	3.111	.164	.221	.307	.308	.3101	.4133

Table 4: *Maximin efficient designs for percentile estimation in the normal model with respect to various "moderate" and "large" parameter regions* $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$.

p	μ_1	μ_2	β_1	β_2	x_1	x_2	x_3	x_4	w_1	w_2	w_3	w_4	$\min \mathrm{eff}_\Theta(\xi_M^*)$	$b(\xi_M^*)$
.5	-1	1	1	2	-1.804	-.613	.613	1.804	.125	.375	.375	.125	.4272	.5262
.5	-2	2	1	4	-1.507	-.452	.452	1.507	.260	.240	.240	.260	.1118	.3007
.8	-1	1	1	2	-.684	.027	1.031	2.273	.065	.349	.357	.229	.3877	.4804
.8	-2	2	1	4	-.994	.001	.712	1.946	.174	.291	.275	.260	.1211	.3045
.9	-.4	.4	1	1.4	-.974	.504	1.665		.024	.440	.537		.5611	.6560
.9	-1	1	1	2	-.381	.161	1.022	2.165	.041	.306	.359	.294	.3674	.4537
.9	-2	2	1	4	-.644	.065	.743	1.973	.158	.258	.275	.309	.1283	.2967

are higher in the logistic case, particularly for larger parameter regions Θ. These phenomena are probably due to differences in tail-behaviour of the link functions. We also observe that the maximin efficient designs yield very good minimal efficiencies if the space $\Theta = [\mu_1, \mu_2] \times [\beta_1, \beta_2]$ is not too large. For example, if the 95% percentile has to be estimated in the logistic model, the minimal efficiency over the range $[-.3, .3] \times [1, 1.3]$ of the maximin efficient design is 83.49%. On the other hand, a larger uncertainty with respect to the location of the unknown parameters $(\mu, \beta)^T$ yields (as expected) a loss in minimal efficiency. Consider for example the problem of estimating the median in a logistic model where $\Theta = [-2, 2] \times [1, 4]$ is used for the construction of the maximin efficient design. In this case, the maximin efficient design is supported at four points and exhibits a minimal efficiency of 25.19%. The reason for this low efficiency is that the maximin optimal design is a compromise between the rather extreme situations appearing in Θ. For example if $\vartheta = (-2, 1)^T$ or $\vartheta = (2, 1)^T$, the locally optimal design advises the experimenter to take only observations at the point $x = 2$ or $x = -2$, respectively, and the corresponding efficiencies for the other situation are both obtained as 19.63%. Note that this value is already smaller than the minimal efficiency of the maximin optimal design, although only the value of μ was misspecified, whereas the efficiency of the maximin optimal design over the whole parameter region Θ is at least 25.19%. Moreover, this value corresponds to the worst case with respect to Θ and for many parameter values of Θ the efficiency of the maximin optimal design is considerably higher. Imagine that the true value of ϑ is given by $\vartheta = (0, 2)^T$ or $\vartheta = (0, 1)^T$. Then the maximin optimal design achieves efficiencies 5.87% and even 72.05%, respectively.

In the remaining part of this section, we want to address the application of Bayesian designs when the experimenter is in the situation that he can only specify a parameter region Θ, in which the unknown parameter ϑ will be located, but he has not enough knowledge in advance of the experiment to assign plausible weights to the elements of Θ. In this case, it is reasonable to give each point in Θ the same weight, i.e. to choose the uniform prior $d\vartheta/\lambda(\Theta)$ on the parameter region, where $\lambda(\Theta)$ denotes the Lebesgue measure of Θ. The Bayesian c_p-efficient design ξ_B^* with respect to the uniform prior $d\vartheta/\lambda(\Theta)$ thus maximizes the integral of the c_p-efficiencies with respect to $d\vartheta/\lambda(\Theta)$, i.e.

$$\xi_B^* = \arg\max_{\xi} \int_{\Theta} \mathrm{eff}_{c_p}(\xi, \vartheta)\, d\vartheta/\lambda(\Theta).$$

For a general Bayesian criterion the uniform prior $d\vartheta/\lambda(\Theta)$ has to be replaced by an arbitrary prior π_Θ on Θ. The equivalence theorem corresponding to the above problem is given as a special case of Theorem 3.3 a) in Dette, Haines & Imhof (2006) and formulated for a general prior on Θ.

THEOREM 3. *A design ξ_B^* is Bayesian c_p-efficient with respect to a prior π_Θ if and only if for each $\vartheta \in \Theta$ there exists a generalized inverse $G(\xi_B^*, \vartheta)$ of $M(\xi_B^*, \vartheta)$ such that the following inequality holds for all $x \in \mathbb{R}$*

$$\int_{\Theta} \frac{c_p^T G(\xi_B^*, \vartheta) I(x, \vartheta) G(\xi_B^*, \vartheta) c_p}{c_p^T M^-(\xi_B^*, \vartheta) c_p}\, \mathit{eff}_{c_p}(\xi_B^*, \vartheta)\, d\pi_\Theta(\vartheta) \le \int_{\Theta} \mathit{eff}_{c_p}(\xi_B^*, \vartheta)\, d\pi_\Theta(\vartheta). \tag{6}$$

Moreover, there is equality in (6) for each support point of the Bayesian c_p-efficient design ξ_B^*.

The situation described above is exactly that way, for which we recommend to use the maximin c_p-efficiency criterion for the design of the experiment and it is of some interest to compare the designs with respect to the different concepts. In the following, we give some optimal designs with respect to the Bayesian c_p-efficency criterion with uniform prior distribution on Θ and compare these with the maximin designs given in Tables 1, 3, 2 and 4. Table 5 displays several Bayesian c_p-efficient designs for the logistic model with respect to uniform priors on various parameter regions

Table 5: *Bayesian c_p-efficient designs for percentile estimation in the logistic model with respect to the uniform prior distributions on various parameter regions Θ.*

p	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	$b(\xi_B^*)$	$\min \mathrm{eff}_\Theta(\xi_B^*)$
.9	-.1	.1	1	1.1	-2.010	2.025	.025	.975	.9715	.8947
.9	-.2	.2	1	1.2	-1.845	1.898	.043	.957	.9406	.7765
.9	-.3	.3	1	1.3	-1.688	1.798	.058	.942	.9093	.6622
.9	-.4	.4	1	1.4	-1.535	1.717	.070	.930	.8788	.5583
.9	-1	1	1	2	-.655	1.457	.113	.887	.7283	.1838

Table 6: *Bayesian c_p-efficient designs for percentile estimation in the normal model with respect to the uniform prior distributions on various parameter regions Θ.*

p_1	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	$b(\xi_B^*)$	$\min \mathrm{eff}_\Theta(\xi_B^*)$
.9	-.1	.1	1	1.1	-1.134	1.210	.026	.974	.9483	.8287
.9	-.2	.2	1	1.2	-1.003	1.157	.047	.953	.8988	.6707
.9	-.3	.3	1	1.3	-.884	1.116	.066	.934	.8524	.5354
.9	-.4	.4	1	1.4	-.775	1.082	.082	.918	.8093	.4238
.9	-1	1	1	2	-.298	.945	.172	.829	.6097	.1037

Θ, where the expression $b(\xi_B^*)$ denotes the optimal criterion value, i.e.

$$b^*(\xi_B^*) = \int_\Theta \mathrm{eff}_{c_p}(\xi_B^*, \vartheta)\, d\vartheta / \lambda(\Theta),$$

and

$$\min \mathrm{eff}_\Theta(\xi_B^*) = \min_{\vartheta\in\Theta} \mathrm{eff}_{c_p}(\xi_B^*, \vartheta)$$

is the minimal c_p-efficiency of ξ_B^* in Θ. For the sake of brevity, we restrict ourselves to the problem of estimating the 90% percentile. The corresponding designs with respect to the normal model are given in Table 6.

Comparing the Bayesian c_p-efficient designs from Tables 5 and 6 with their maximin counterparts, we first notice that the Bayesian optimal designs have always two support points even for relatively large parameter regions Θ. Moreover, one support point is assigned a very small weight so that the Bayesian optimal designs are close to the locally c_p-optimal designs on the corresponding parameter region Θ, which are supported on one point only. This phenomenon is also observed for the maximin efficient designs, if the specified parameter space Θ is "small" (see Tables 1 and 2), but not for "larger" parameter spaces (see Tables 3 and 4). As expected from the definitions of the respective optimality criteria, the averaged efficiencies b^* of the Bayesian optimal designs are somewhat higher than the averaged efficiencies of the maximin c_p-efficient designs. Turning to the comparison of minimal c_p-efficiencies, things change dramatically, especially for large parameter regions.

We consider as a representative example the optimal designs in the logistic model for estimating the percentile x_p with $p = .9$ and assume that the experimenter is able to specify the parameter region $\Theta = [-1, 1] \times [1, 2]$. The averaged efficiency of the Bayesian optimal design is given by the value $b(\xi_B^*) = .7283$, indicating high c_p-efficiencies for this design at most points in the parameter region Θ, whereas the integrated efficiency of the corresponding maximin optimal design is given by $b(\xi_M^*) = .5826$, which is of course smaller than the result for the Bayesian design, but still in acceptable limits. The result for the minimal c_p-efficiency of the Bayesian design, $\min \mathrm{eff}_\Theta(\xi_B^*) = .1838$, however, reveals that there are regions in Θ, in which the Bayesian optimal design does not facilitate efficient estimation of the $100p$th percentile x_p, i.e. the Bayesian design is not robust against every value $\vartheta \in \Theta$. The corresponding maximin efficient design, in contrast, achieves a minimal efficiency of .4756, which is significantly higher than .1838. In this example, the Bayesian optimal design is supported on two points with much weight assigned to 1.457, i.e. it is close to the locally optimal designs for parameters satisfying the equation $1.457\beta = F^{-1}(.9) - \mu$. For such parameters, we thus expect the design to perform highly efficient, whereas for parameter combinations from Θ, which are located far away from the straight line described above, the efficiency of the Bayesian optimal design will be considerably smaller. Consider for example the parameter combinations $\mu = 0$, $\beta = 1.5$ and $\mu = -1$, $\beta = 1$ from the parameter space Θ, respectively. For the first parameter combination, which is located almost on the straight line, the Bayesian optimal design has efficiency .972, whereas its efficiency with respect to the second parameter combination turns out to be the unsatisfactory value of .189. The maximin optimal design, which is much more spread on the real axis and thus not close to any locally optimal design at all, yields efficiencies .648 and .489 for the above parameters, respectively. Summarizing this result, we notice that for many values ϑ in the parameter region Θ the Bayesian optimal design will be slightly more efficient than the corresponding maximin optimal design, but there are elements $\vartheta \in \Theta$, for which efficient percentile estimation is almost impossible when using the Bayesian optimal design. It is also worthwhile to mention that maximin efficient designs often have more support points than the corresponding Bayesian optimal designs with respect to the uniform distribution on Θ. This is an important advantage because additional support points facilitate to check the model assumption by goodness-of-fit tests (e.g. testing against a binary response model with three parameters).

For small parameter regions, the differences between Bayesian and maximin c_p-efficient designs are relatively small, so it will make not much difference which criterion the experimenter chooses. The larger the assumed parameter region (i.e. the higher the experimenter's uncertainty about the position of the parameter ϑ), the stronger the use of the maximin c_p-efficiency criterion is recommended.

4. EFFICIENT DESIGNS FOR PERCENTILE ESTIMATION IN CASE OF MODEL UNCERTAINTY

Besides the robustness problems of experimental designs with respect to the misspecification of some unknown non-linear model parameters, there are other issues of robustness, which have to be taken into account for the design of experiment. In particular the design also depends on the link function in (1), but in general it will be difficult for the experimenter to specify one distribution F as the appropriate link function in advance. For example, it might not be obvious if the normal distribution or a more heavy-tailed link function such as the logistic link would fit the data adequately before even collecting them. Misspecification of the quantal response model, however, may result in a large bias when estimating the percentiles of interest. It is thus sensible to choose the experimental conditions as in some sense "good" for several choices of links. Following Läuter (1974), we assume that the experimenter can define preferences α_j with respect to certain link functions F_j, where the weights α_j add up to one. Define $M_{(j)}(\xi, \vartheta)$ as the information matrix with respect to the quantal response model with link function F_j and $\xi^*_{c_p(j)}$ as the corresponding locally $c_p(j)$-optimal design. Then an appropriate optimality criterion is of the form

$$\sum_j \alpha_j \min_{\vartheta \in \Theta} \frac{c_p(j)^T M^-_{(j)}(\xi^*_{c_p(j)}, \vartheta) c_p(j)}{c_p(j)^T M^-_{(j)}(\xi, \vartheta) c_p(j)} \to \max_\xi . \tag{7}$$

Table 7: *Maximin efficient designs for percentile estimation with respect to the optimality criterion (7) if the normal and the logistic model are assigned weights $\alpha_1 = \alpha_2 = 1/2$.*

p	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	u^*	u_1^*	u_2^*
.5	-.1	.1	1	1.1	-.398	.398	.5	.5	.9027	.9105	.8948
.6	-.1	.1	1	1.1	-.025	1.163	.645	.355	.8343	.8591	.8096
.7	-.1	.1	1	1.1	.006	1.299	.421	.579	.7628	.8144	.7112
.8	-.1	.1	1	1.1	-.314	1.422	.199	.801	.6900	.7709	.6091
.9	-.1	.1	1	1.1	-2.040	1.143	.174	.826	.6782	.5641	.7923
.9	-.2	.2	1	1.2	-1.739	1.071	.202	.798	.5924	.4814	.7033

The equivalence theorem, by which the optimality of a design with respect to criterion (7) can be checked, is given below and can be derived by combining parts a) and b) of Theorem 3.3 in Dette, Haines & Imhof (2006).

THEOREM 4. *A design ξ^* is optimal with respect to the criterion defined in (7) if and only if there exist prior distributions π_j^* on $\mathcal{N}_j(\xi^*)$ and for each $\vartheta \in \mathcal{N}_j(\xi^*)$ generalized inverses $G_{(j)}(\xi^*, \vartheta)$ of $M_{(j)}(\xi^*, \vartheta)$ such that the inequality*

$$\sum_j \alpha_j \int_{\mathcal{N}_j(\xi^*)} \frac{c_p(j)^T G_{(j)}(\xi^*, \vartheta) I(x, \vartheta) G_{(j)}(\xi^* \vartheta) c_p(j)}{c_p(j)^T M_{(j)}^-(\xi^*, \vartheta) c_p(j)}\, \mathit{eff}_{c_p(j)}(\xi^*, \vartheta)\, d\pi_j^*(\vartheta) \leq u^* \qquad (8)$$

holds for all $x \in I\!R$. The expression $\mathit{eff}_{c_p(j)}(\xi, \vartheta)$ denotes the c_p-efficiency of the design ξ with respect to the binary response model with link function F_j, $\mathcal{N}_j(\xi^*)$ is given by the set

$$\mathcal{N}_j(\xi^*) = \{\tilde{\vartheta} \in \Theta \mid \mathit{eff}_{c_p(j)}(\xi^*, \tilde{\vartheta}) = u_j^*\},$$

and the values u^* and u_j^* are defined as

$$u^* = \sum_j \alpha_j u_j^*, \qquad u_j^* = \min_{\vartheta \in \Theta} \mathit{eff}_{c_p(j)}(\xi^*, \vartheta).$$

Moreover, there is equality in (8) for all support points of the design ξ^*.

We consider the situation, in which both the normal and the logistic link function are assigned weight $\alpha_j = 1/2$, $j = 1, 2$, i.e. the experimenter has no preference with respect to one of these models, thus designing the experiment "equally well" for both choices of link functions. If the experimenter considers the logistic link function more likely, he could e.g. assign weight 3/4 to this model and only 1/4 to the normal model. Some results for different values of p with respect to several parameter regions Θ are given in Table 7. Again, the optimality of these designs with respect to the criterion (7) has been verified by Theorem 4.

We observe from Table 7 that for some percentiles the minimal efficiencies u_1^* and u_2^* in both models and thus their weighted means u^* are substantially smaller compared to the results in Tables 1 and 2 which assume a fixed link function, whereas for other percentiles the loss in efficiency is much less severe. For example for $p = .5$, i.e. estimation of the median, the optimal designs in the sense of (7) are still highly efficient, whereas optimal designs for the estimation of the 80%-percentile exhibit significantly lower efficiencies u_1^*, u_2^* and u^*. These results correspond to intuition, because the logistic and the normal distribution function with the same parameters are "close together" in the area around the mean but differ significantly elsewhere (see Figure 1). The maximin optimal designs for estimating the 50%-percentiles in both models separately are thus also closer together than the corresponding designs for the 80%-percentiles (see Tables 1 and 2) yielding a higher efficiency for the compromise design with respect to the criterion (7), which is in some sense situated between the two individual designs.

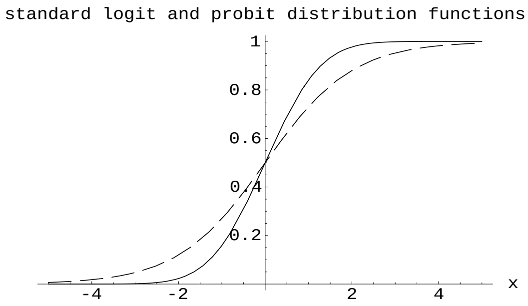

Figure 1: The distribution functions of the logistic and the normal model with mean zero and variance 1. Dashed line: logistic distribution. Solid line: normal distribution.

Using (7) as an optimality criterion, we thus gain model robustness of the design at the expense of efficiency with respect to the "true" model. There is, however, no robust alternative to this approach, particularly when keeping in mind that in most applications data are not "purely" either normal or logistic but something in between or even from a completely different (unknown) distribution family.

In practice, sometimes data from previous studies are available so that the form of the "true" link function F can be assumed to be in a small neighbourhood of some distribution function F_0. We can incorporate this situation in the above settings by choosing the functions F_j in the optimality criterion (7) as $F_j(z) = (1-\gamma)F_0(z) + \gamma G_j(z)$ where γ is some small constant, e.g. $\gamma = .05$, and G_j are several different distribution functions. It is also feasible to choose different values γ_j for each F_j.

A further robustification of this approach would be to use a maximin criterion not only with respect to the parameters, but also for the link function F instead of the rather Bayesian criterion discussed above. For this optimality criterion, calculations become extremely hard, but are still feasible as long as a finite class of distribution functions is chosen. We do not provide examples here for the sake of brevity.

5. MAXIMIN EFFICIENT DESIGNS FOR ESTIMATING SEVERAL PERCENTILES AT THE SAME TIME

In this section, we consider a different aspect in the construction of robust designs, which is motivated by the observation that the experimenter may have several quantiles of interest. The goal here is to design the experiment efficiently for estimating several percentiles $x_p,\ p \in P$, simultaneously, where P is a subset of the interval $(0,1)$. We thus choose a weight function ζ on P from the class of distribution functions, according to our preferences with respect to p. For example, if we want to estimate the percentiles x_{p_1} and x_{p_2} "equally well", the uniform weight function, which assigns weights 1/2 and 1/2 to p_1 and p_2, respectively, will be appropriate. In

Table 8: *Maximin efficient designs for the estimation of several percentiles simultaneously in the logistic model,* $\zeta = \{p_1, p_2; 1/2, 1/2\}$.

p_1	p_2	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	u^*	$u^*_{p_1}$	$u^*_{p_2}$
.5	.6	-.1	.1	1	1.1	-.504	.972	.496	.504	.7699	.7608	.7790
.6	.7	-.1	.1	1	1.1	-.535	1.106	.314	.686	.7570	.7460	.7680
.7	.8	-.1	.1	1	1.1	-.779	1.341	.188	.812	.7481	.7341	.7620
.8	.9	-.1	.1	1	1.1	-1.364	1.667	.142	.858	.7642	.7291	.7993
.5	.7	-.1	.1	1	1.1	-.704	1.213	.417	.583	.6659	.6569	.6750
.5	.8	-.1	.1	1	1.1	-.881	1.409	.394	.606	.5955	.5984	.5926
.5	.9	-.1	.1	1	1.1	-1.110	1.608	.371	.629	.5746	.5258	.6235

general, the optimality criterion is then given by the function

$$\int_P \min_{\vartheta\in\Theta} \frac{c_p^T M^-(\xi^*_{c_p}, \vartheta) c_p}{c_p^T M^-(\xi, \vartheta) c_p}\, d\zeta(p) \to \max_\xi . \tag{9}$$

Obviously, this maximization problem is extremely hard to tackle in case of a continuous weight function ζ. For calculations in practice, we therefore recommend to approximate a continuous weight function ζ on P by a discrete one. For a discrete weight function, however, we are in the same situation as in Section 4 with the "prior" over the different possible link functions replaced by a function allocating weights to the percentiles to be estimated. The corresponding equivalence theorem, which is given below, is thus a generalization of Theorem 4 to the case of continuous prior distributions.

THEOREM 5. *Let ζ be a weight function on $P \subset (0,1)$. A design ξ^* is optimal with respect to the criterion (9) if and only if for each $p \in P$ there exists a prior distribution π_p^* on the set $\mathcal{N}_p(\xi^*)$ and for each $\vartheta \in \mathcal{N}_p(\xi^*)$ a generalized inverse $G(\xi^*, \vartheta)$ of the matrix $M(\xi^*, \vartheta)$ such that the inequality*

$$\int_P \int_{\mathcal{N}_p(\xi^*)} \frac{c_p^T G(\xi^*, \vartheta) I(x, \vartheta) G(\xi^*, \vartheta) c_p}{c_p^T M^-(\xi^*, \vartheta) c_p}\, \mathit{eff}_{c_p}(\xi^*, \vartheta) d\pi_p^*(\vartheta)\, d\zeta(p) \le u^*$$

holds for all $x \in \mathbb{R}$, where $\mathcal{N}_p(\xi^*)$ and u^* are defined by

$$\mathcal{N}_p(\xi^*) = \{\tilde\vartheta \in \Theta \mid \mathit{eff}_{c_p}(\xi^*, \tilde\vartheta) = u_p^*\},$$

$$u^* = \int_P u_p^*\, d\zeta(p), \qquad u_p^* = \min_{\vartheta\in\Theta} \mathit{eff}_{c_p}(\xi^*, \vartheta).$$

Again, there is equality if x is a support point of the design ξ^*.

In our numerical study, we consider situations, in which either two (see Tables 8 and 9) or three (see Table 10) different percentiles are of equal interest to the experimenter. Table 8 shows some representative results for the logistic model, whereas Tables 9 and 10 give optimal designs with respect to the criterion (9) for the normal model. For the sake of brevity, only situations are displayed, where the optimal designs are supported at two points.

As expected, we observe from Tables 8 and 9 a decline in minimal efficiency compared to the results in Tables 1 and 2, respectively, if the experiment is designed for estimating two different

Table 9: *Maximin efficient designs for the estimation of several percentiles simultaneously in the normal model,* $\zeta = \{p_1, p_2; 1/2, 1/2\}$.

p_1	p_2	μ_1	μ_2	β_1	β_2	x_1	x_2	w_1	w_2	u^*	$u^*_{p_1}$	$u^*_{p_2}$
.5	.6	-.1	.1	1	1.1	-.439	.704	.487	.514	.7863	.7760	.7966
.6	.7	-.1	.1	1	1.1	-.474	.773	.326	.674	.7724	.7599	.7849
.7	.8	-.1	.1	1	1.1	-.607	.907	.210	.790	.7562	.7423	.7702
.8	.9	-.1	.1	1	1.1	-.867	1.101	.151	.849	.7462	.7249	.7675
.5	.7	-.1	.1	1	1.1	-.575	.855	.417	.583	.6967	.6853	.7081
.5	.8	-.1	.1	1	1.1	-.692	.985	.389	.611	.6254	.6260	.6248
.5	.9	-.1	.1	1	1.1	-.800	1.091	.388	.612	.5840	.5895	.5785

Table 10: *Maximin efficient designs for the estimation of several percentiles simultaneously in the normal model,* $\zeta = \{p_1, p_2, p_3; 1/3, 1/3, 1/3\}$, $[\mu_1, \mu_2] = [-.1, .1]$, $[\beta_1, \beta_2] = [1, 1.1]$.

p_1	p_2	p_3	x_1	x_2	w_1	w_2	u^*	$u^*_{p_1}$	$u^*_{p_2}$	$u^*_{p_3}$
.5	.6	.7	-.522	.734	.412	.589	.7355	.7092	.8317	.6655
.6	.7	.8	-.580	.822	.293	.707	.7223	.7097	.8224	.6348
.7	.8	.9	-.770	.965	.203	.797	.7115	.6923	.8057	.6365
.5	.7	.9	-.742	.901	.348	.652	.6278	.6093	.7526	.5216

percentiles x_{p_1} and x_{p_2}, where the decrease is even more significant if the distance between p_1 and p_2 is large. The last-mentioned effect obviously emerges from the fact that the optimal designs with respect to one percentile at a time are close if the distance between p_1 and p_2 is small and hence the optimal design for estimating both percentiles at the same time, which is in some sense situated between the other two designs, is not too far away from both of them. In general, all of the efficiencies displayed in Tables 8 and 9 are still significantly larger than 50% and thus in acceptable limits.

Table 10, finally, displays a somewhat further decline in minimal efficiency if the design is constructed for the efficienct estimation of three different percentiles simultaneously. We observe that the minimal efficiencies $u^*_{p_1}$, $u^*_{p_2}$, $u^*_{p_3}$ of these designs with respect to the estimation of one particular percentile are highest for the percentile corresponding to the value of p_2, which lies between the values p_1 and p_3. This effect probably occurs since designs, which are efficient for the simultaneous estimation of three percentiles with respect to values p_1, p_2 and p_3, will be compromise designs between the designs for the estimation of one of these percentiles at a time. The design that is maximin optimal for estimating the $100p_2$th percentile is in some sense located between the other two designs and the compromise design will therefore be close.

We must, however, keep in mind that the values presented in Tables 8-10 are only minimal efficiencies, i.e. they give only the efficiencies for the worst case scenario with respect to Θ. If other values from Θ are the "true" parameters in our model, the efficiency of the designs will be significantly higher. Consider, for example, the parameter combination $\mu = 0$, $\beta = 1.05$, i.e. the midpoint of the rectangle Θ. Assume, furthermore, the percentiles corresponding to $p_1 = .8$ and

$p_2 = .9$ are equally important and the logistic link function describes the model appropriately. Then the value of the overall efficiency of the optimal design with respect to the criterion (9) is given by .829, which is considerably higher compared to its minimal efficiency .764 from Table 8. The efficiencies with respect to one percentile at a time are given by .797 and .860, respectively, which are also higher than the corresponding minimal values .729 and .799.

Analogously to the discussion at the end of the previous section, one can also define a maximin criterion with respect to a class of percentiles to be estimated. We calculated some optimal designs with respect to this criterion, which was computationally possible even if an uncountable set (interval) for p was chosen. These designs will be discussed in the following section, where we apply our designs towards redesigning a real data example.

6. DATA EXAMPLE

In this section, we demonstrate the usefulness of maximin efficient designs relative to designs used in practice for some real application. We redesign a psychophysical experiment given in Rosenberger & Grill (1997), who propose a sequential procedure for estimating the three quartiles of the underlying quantal response curve, which was supposed to be logistic. Zhu & Wong (2001) used this example to compare the performance of Bayesian median optimal designs with respect to the uniform priors on different parameter regions with the original study design. To make our results comparable with the designs found by Rosenberger & Grill (1997) and Zhu & Wong (2001) we changed the parametrization of model (1) to $p(x, \vartheta) = F(\beta(x - \alpha))$ for this particular example. Following Zhu & Wong (2001), we chose the three parameter regions $\Theta_1 = [28.5, 37.5] \times [.12, .54]$, $\Theta_2 = [28.5, 37.5] \times [.1, 2]$ and $\Theta_3 = [28.5, 37.5] \times [1, 5]$, and calculated the maximin median efficient designs with respect to these regions. Moreover, we calculated the corresponding maximin efficient designs for the estimation of the $ED25$ and the $ED75$ simultaneously with equal weights to compare these designs with the median optimal designs. These designs and their minimal efficiencies on the respective parameter spaces are displayed in Table 11. For comparison, we have added the minimal efficiencies $\min \mathrm{eff}_\Theta(\xi_B^*)$ of the corresponding Bayesian optimal designs. Since the maximin and Bayesian optimal designs with respect to Θ_3 both work with a completely misspecified region for β if one assumes that (35.56, .38) is the true value of ϑ (as we will do in the subsequent simulation study based on the prior estimates by Zhu & Wong 2001) we also calculated the minimal efficiencies of the maximin and Bayesian optimal design when guarding against the union of ranges, i.e. $\beta \in [.12, 5]$. For this scenario, we obtained almost exactly the same minimal efficiencies as for Θ_3, so that we do not consider these designs any further.

We observe from Table 11 that for the moderate parameter region Θ_1 the maximin efficient designs have reasonable minimal efficiencies whereas for large regions of parameter uncertainty the minimal efficiencies can be as small as 1%. The minimal efficiencies of the Bayesian optimal designs are even considerably smaller. Examining these minimal efficiencies, one has to keep in mind that these values correspond to the worst case within the respective parameter region. For some specific parameter values, we obtain the following result (see Table 12). For brevity, we restrict ourselves to present only results for estimating the median.

For these particular values, the efficiencies of the maximin and Bayesian optimal designs are considerably higher than the respective minimal efficiencies although some parameter values are not even contained in the regions Θ_1, Θ_2 or Θ_3. The optimal designs with respect to Θ_3, for example, are highly efficient (in particular the maximin design) although this region comprises only one out of the four values for β considered in Table 12. Only if $\beta = 2$, the region Θ_1 is too far away from this value so that the efficiencies go down.

We also calculated the maximin efficient designs with respect to Θ_1, Θ_2 and Θ_3 and an interval of $[.25, .75]$ with respect to a maximin criterion for the percentiles to be estimated. These designs do not differ substantially from the respective maximin efficient designs for estimating the $ED25$

Table 11: *Maximin efficient designs for estimating the median (first three rows) as well as the 25%- and the 75%-percentiles simultaneously with equal weights (rows 4 - 6) in the logistic model for parameter spaces* Θ_1, Θ_2, Θ_3, *respectively. For comparison: The last column shows the minimal efficiencies of the corresponding Bayesian optimal designs.*

Θ	x_1	x_2	x_3	x_4	x_5	x_6	w_1	w_2	w_3	w_4	w_5	$\min \mathrm{eff}_\Theta$	$\min \mathrm{eff}_\Theta(\xi_B^*)$
Θ_1	28.08	33.00	37.92	–	–	–	.38	.24	.38	–	–	.43	.35
Θ_2	28.60	30.50	32.21	33.79	35.50	37.40	.19	.16	.14	.14	.16	.165	.10
Θ_3	28.61	30.36	32.12	33.88	35.64	37.39	.16	.17	.17	.17	.17	.013	$2 \cdot 10^{-4}$
Θ_1	13.55	26.36	30.92	35.08	39.64	52.45	.09	.25	.16	.16	.25	.273	.062
Θ_2	27.11	29.71	31.90	34.10	36.29	38.89	.23	.13	.14	.14	.13	.0915	.011
Θ_3	28.50	30.30	32.10	33.90	35.70	37.50	.17	.17	.17	.17	.17	.0073	10^{-5}

Table 12: *Some selected efficiencies of maximin and Bayesian optimal designs for estimating the median.*

design	$\alpha = 35.56, \beta = .38$	$\alpha = 30, \beta = .5$	$\alpha = 32, \beta = .2$	$\alpha = 32, \beta = 2$
maximin Θ_1	.567	.474	.816	.013
Bayesian Θ_1	.604	.543	.833	.001
maximin Θ_2	.539	.450	.842	.181
Bayesian Θ_2	.524	.434	.841	.199
maximin Θ_3	.521	.430	.841	.204
Bayesian Θ_3	.369	.274	.793	.149

Table 13: *Simulated standard deviations of the maximum likelihood estimates for the three quartiles where data were generated according to different designs with $n = 132$, $\alpha = 35.56$, $\beta = .38$, 10,000 runs.*

design	std $ED25$	std $ED50$	std $ED75$
maximin Θ_1 (.5)	18	16	22
maximin Θ_1 (.25, .75)	21	18	23
Bayesian Θ_1	19	15	20
maximin Θ_2 (.5)	16	16	29
maximin Θ_2 (.25, .75)	17	16	23
Bayesian Θ_2	17	15	24
maximin Θ_3 (.5)	16	17	30
maximin Θ_3 (.25, .75)	16	17	29
Bayesian Θ_3	16	16	28
sequential	33	19	27
uniform 9	16	19	35
uniform 5	16	18	32

and the $ED75$ simultaneously with equal weights. We therefore expect a similar performance of these designs, so that it seems not necessary to include them in the subsequent simulation study.

To compare the performance of our designs with the designs proposed by Rosenberger & Grill (1997) and Zhu & Wong (2001), we estimate by simulation the standard deviations of the estimates of the three quartiles, where the data are generated according to the different designs. The sample size is $N = 132$, the true parameter values for the simulation study are assumed to be $\alpha = 35.56$ and $\beta = .38$, and 10,000 runs are carried out. The standard deviations for the Bayesian and the sequential designs are taken from Rosenberger & Grill (1997) and Zhu & Wong (2001). We further include two different uniform designs in our study since equal allocation rules are widely used in practice; see, e.g., Zhu, Ahn & Wong (1998). On the one hand, we use the uniform design on the 9 levels 29, 30, $\ldots 37$ in the initial design used by Rosenberger & Grill (1997). As it is unlikely that practitioners will adopt a design on too many different doses to estimate three parameters, we also add the uniform design on the 5 levels $29, 31, 33, 35, 37$, respectively, to the study.

Table 13 displays the simulated standard deviations of the maximum likelihood estimates for the three quartiles where data were generated according to the different designs under comparison. By "sequential", we denote the design proposed by Rosenberger & Grill (1997), whereas Bayesian Θ stands for the Bayesian optimal designs for estimating the median with respect to the uniform prior on Θ found by Zhu & Wong (2001). Maximin Θ (.5) and maximin Θ (.25, .75) are finally the maximin optimal designs with respect to Θ for estimating the median ($ED50$) and the 25%- and the 75%-percentiles simultaneously with equal weights.

We learn from Table 13 that for this particular example the Bayesian optimal designs perform slightly better than the maximin efficient designs when it comes to estimating the $ED75$. For the two other quartiles the results are about equal. Comparing the two different types of maximin efficient designs (on the one hand for estimating the median and on the other hand for estimating the $ED25$ and the $ED75$ simultaneously), we obtain very similar results, in particular for the larger parameter regions Θ_2 and Θ_3, which is not surprising since these designs are quite similar (see Table 11). An intuitive explanation for this result is that for large regions of uncertainty the optimal support points of a maximin median efficient design are so widespread that the other quartiles can also be estimated well. The uniform designs perform quite well for the estimation of the $ED25$ and the median, but poorly for estimating the $ED75$, whereas the sequential design exhibits a particular loss of efficiency when estimating the $ED25$. The slight superiority of the Bayesian optimal designs might intuitively be explained by the position of the "true" parameter, which is almost in the center of the different parameter regions (except in the situation of Θ_3) where the c_p-efficiencies are usually higher than closer to the boundary or even outside the specified region as is confirmed by the results presented in Table 12. For a "true" parameter value close to the boundary, one would expect the maximin designs to be superior. In general, the non-sequential optimal designs exhibit a better performance than the sequential design.

7. CONCLUSIONS

In this article, we discuss several aspects of robust design of experiment for estimating the effective log-dose level in binary response models. In general, optimal designs minimize an approximation of the variance of the maximum likelihood estimate for the $100p$th percentile x_p and depend sensitively on the unknown parameter, the link function and the quantile of interest. Several strategies based on the concept of maximin efficiency are discussed to obtain designs which are on the one hand efficient for the effective log-dose level and on the other hand robust with respect to these assumptions. It is demonstrated that robust designs with respect to the new criteria can be calculated numerically and that the optimality of these designs can be checked by non-standard equivalence theorems. The most important robustification is obtained assuming that the experimenter is only able to specify a certain range Θ for the unknown parameters and it is demonstrated that the maximin efficient designs are efficient for all values in Θ, provided that this set is not too large. These designs are then further robustified with respect to the assumptions on the link function (Section 4) and the particular quantiles x_p chosen by the experimenter.

It is also –theoretically and numerically– possible to combine the approaches of Sections 4 and 5 to achieve a further robustification of the designs, i.e. to define an optimality criterion, which is robust with respect to misspecifications of the parameters, the binary response model and at the same time efficient for the estimation of several percentiles. This criterion corresponds in some sense to a combination of maximin- and Bayesian criteria over the parameter space $\tilde{\Theta} = (\Theta \times \mathcal{F} \times \mathcal{P})$, where $\mathcal{F}$ denotes the class of reasonable link functions and $P \subset (0, 1)$ a set of percentiles of interest. One should, however, keep in mind that adding extra dimensions to the possible alternatives, against which the design of the experiment has to be robustified, has the same effect on the efficiency of the resulting designs as enlarging its individual components, i.e. the (minimal) efficiencies will decrease as can already be seen from the examples in Tables 7-10. Robustification of designs will always be bought at the expense of a loss in efficiency. The experimenter should therefore carefully deliberate about the question which amount of robustness is necessary for an appropriate data analysis.

The discussion at the end of Section 3, however, shows that there is no robust alternative to the maximin approach with respect to the parameter region Θ. In particular, it turns out that the less information about the approximate position of the unknown parameter ϑ is available (i.e. the larger the region Θ is chosen), the less robust against particular values of $\vartheta \in \Theta$ the Bayesian c_p-efficient design (with respect to the uniform prior on Θ) will be. Consequently, the more uncertainty about the parameter ϑ exists, the stronger the maximin approach is recommended.

A further type of robustification, which was suggested by one of the referees, would be to robustify the designs against model departures of the form $F(\mu + \beta x + k(x))$ where $k(x)$ is some small perturbation of the linear term, e.g. $k(x) = \varepsilon x^2$ for some small value of ε. Considering such designs, however, would be so complex that it might well amount to a new paper. To the knowledge of the authors there is only one article on (locally D-)optimal designs for multi-stage models, i.e. models of the form $F(b_0 + b_1 x + b_2 x^2 + \ldots)$ (see Fang, Wiens & Wu (2005)). This topic might well be the subject of future research.

We would finally like to mention again that it is our intention to make computer codes available online in the medium-term within the framework of some more comprehensive design project. This will enable practitioners to apply the maximin efficient designs proposed in this article.

APPENDIX

Proof of Lemma 1. Combining the equalities

$$c_p = c_p(\mu,\beta) = -(1,x_p)^T/\beta = -(1, \frac{F^{-1}(p)-\mu}{\beta})^T/\beta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} c_{1-p}(-\mu,\beta)$$

and

$$\begin{aligned} M(\xi,\mu,\beta) &= \sum_i w_i \frac{f(\beta x_i+\mu)^2}{F(\beta x_i+\mu)(1-F(\beta x_i+\mu))} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} M(\{-x_i, w_i\}, -\mu, \beta) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

the assertion of Lemma 1 is obvious. □

Proof of Lemma 2. Combining the equalities

$$c_p(\mu,\beta) = -(1,x_p)^T/\beta = -(1, \frac{F^{-1}(p)-\mu}{\beta})^T/\beta = \gamma \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} c_p(\mu,\gamma\beta)$$

and

$$\begin{aligned} M(\xi,\mu,\beta) &= \sum_i \frac{f(\beta x_i+\mu)^2}{F(\beta x_i+\mu)(1-F(\beta x_i+\mu))} \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} M(\xi_\gamma, \mu, \gamma\beta) \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix} \end{aligned}$$

we obtain the statement of Lemma 2. □

ACKNOWLEDGEMENTS

The support of the Deutsche Forschungsgemeinschaft (SFB 475, Komplexitätsreduktion in multivariaten Datenstrukturen, Teilprojekt A2, Sachbeihilfe 436 RUS) is gratefully acknowledged. The authors would also like to thank two referees and the Assiociate Editor for their constructive comments and helpful suggestions, which led to a substantial improvement of an earlier version of this article.

REFERENCES

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

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

H. Dette & S. Biedermann. (2003). Robust and efficient designs for the Michaelis-Menten model. *Journal of the American Statistical Association*, 98, 679–686.

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

H. Dette, L. Haines & L. Imhof (2006). Maximin and Bayesian optimal designs for linear and non-linear regression models. Under revision for: *Statistica Sinica*.

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

R. Fandom Noubiap & W. Seidel (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.

Z. Fang, D. P. Wiens & Z. Wu (2005). Locally D-optimal designs for multistage models and heteroscedastic polynomial regression models. *Journal of Statistical Planning and Inference*, in press.

I. Ford, B. Torsney & C. F. J. Wu (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.

J. King & W. K. Wong (2000). Minimax D-optimal designs for the logistic model. *Biometrics*, 56, 1263–1267.

E. Läuter (1974). Experimental design in a class of models. *Mathematische Operationsforschung und Statistik*, 5, 379–398.

J. López-Fidalgo & W. K. Wong (2000). A comparative study of MV- and SMV-optimal designs for binary response models. *In: Advances in stochastic simulation methods (ed. N. Balakrishnan)*, Birkhäuser, Boston, USA, 135–151.

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

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

W. F. Rosenberger & S. E. Grill (1997). A sequential design for psychophysical experiments: an application to estimating timing of sensory events. *Statistics in Medicine*, 16, 2245–2260.

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

R. R. Sitter (1992). Robust Designs for Binary Data. *Biometrics,* 48, 1145–1155.

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

C. F. J. Wu (1985). Efficient sequential designs with binary data. *Journal of the American Statistical Association*, 80, 974–984.

C. F. J. Wu (1988). Optimal design for percentile estimation of a quantal response curve. *Optimal design and analysis of experiments.* Amsterdam: Elsevier (North Holland).

W. Zhu, H. Ahn & W. K. Wong (1998). Multiple-Objective Optimal Designs for the logistic Model. *Communications in Statistics: Theory and Methods*, 27, 1581–1592.

W. Zhu & W. K. Wong (2000). Multiple-objective designs in a dose-response experiment. *Journal of Biopharmaceutical Statistics*, 10, No. 1, 1–14.

W. Zhu & W. K. Wong (2001). Bayesian optimal designs for estimating a set of symmetric quantiles. *Statistics in Medicine*, 20, 123–137.