Maximin and Bayesian optimal designs for regression models

Holger Dette
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
holger.dette@ruhr-uni-bochum.de

Linda M. Haines
School of Mathematics,
Statistics and Information Technology,
University of Natal,
Pietermaritzburg, South Africa
haines@nu.ac.za

Lorens A. Imhof
Aachen University
Institut für Statistik
und Wirtschaftsmathematik
52056 Aachen, Germany
imhof@stochastik.rwth-aachen.de

June 4, 2003

Abstract

For many problems of statistical inference in regression modelling, the Fisher information matrix depends on certain nuisance parameters which are unknown and which enter the model nonlinearly. A common strategy to deal with this problem within the context of design is to construct maximin optimal designs as those designs which maximize the minimum value of a real valued (standardized) function of the Fisher information matrix, where the minimum is taken over a specified range of the unknown parameters. The maximin criterion is not differentiable and the construction of the associated optimal designs is therefore difficult to achieve in practice. In the present paper the relationship between maximin optimal designs and a class of Bayesian optimal designs for which the associated criteria are differentiable is explored. In particular, a general methodology for determining maximin optimal designs is introduced based on the fact that in many cases these designs can be obtained as weak limits of appropriate Bayesian optimal designs.

AMS Subject Classification: Primary 62K05 ; Secondary 62F15

Keywords and Phrases: maximin optimal designs, Bayesian optimal designs, nonlinear regression models, parameter estimation, least favourable prior

1 Introduction

In many practical problems in regression modelling the Fisher information for the parameters of interest depends on certain unknown nuisance parameters. Within the context of design this problem translates into that of maximizing a concave function of the information matrix which depends on the unknown parameters and clearly this cannot be achieved directly. Over the last forty years a number of strategies have been developed to address this design problem. Specifically, in 1953, Chernoff suggested the simple but elegant expedient of adopting a best guess for the unknown parameters and termed the resultant designs locally optimal. The main disadvantage to such an approach is that if the unknown parameters are misspecified the resulting optimal designs can be highly inefficient within the true model setting.
A more robust approach to this problem is to, in some sense, quantify the uncertainty in those parameters and to incorporate this additional information into the formulation of suitable optimality criteria. This has been achieved in practice through the introduction of the concepts of Bayesian and of maximin optimality. In particular Bayesian optimality criteria are based on criteria in classical design theory and many of the results from that theory, such as those relating to equivalence theorems and numerical procedures, can immediately be transferred into the Bayesian context [see e.g. Pronzato and Walter (1985), Chaloner and Larntz (1989), Chaloner (1993), Chaloner and Verdinelli (1995)]. For maximin optimality, designs which maximize the minimum of a function of the Fisher information matrix over a range of parameter values are sought [see e.g. Pronzato and Walter (1985), Müller (1995), Dette (1997), Müller and Pázman (1998)]. The resultant designs, termed maximin optimal designs, are particularly attractive from a practical point of view in that the experimenter is only required to specify an appropriate range for the unknown parameters. The major problem lies in the construction of these designs in the sense that the maximin optimality criterion is not differentiable and that as a consequence results, both algebraic and numeric, are elusive. Indeed there have been few reports of maximin optimal designs in the literature and strategies for their construction are somewhat ad hoc [see e.g. Wong (1992), Haines (1995), Imhof (2001)].
In the present study a general approach to obtaining maximin optimal designs as the limits of a particular class of Bayesian optimal designs is introduced and explored. Roughly speaking the powerful tools for constructing Bayesian optimal designs for which the associated criteria are differentiable can be used to obtain maximin optimal designs for which the corresponding criteria are not differentiable. In particular the approach avoids the calculation of a least favourable prior distribution. Although interest is centered primarily on the construction of maximin optimal designs for nonlinear regression models the approach is quite general and can be applied to other optimal design problems with a similar structure.
The paper is organized in the following way. In Section 2 some preliminary definitions are given and Bayesian optimality criteria analogous to Kiefer's (1974) Φ_p-criteria are introduced. The main results of the study are then presented in Section 3. In particular it is shown that under fairly general conditions the Bayesian optimal designs converge weakly to maximin optimal designs, a result which mirrors the limiting relationship of the corresponding optimality criteria. Furthermore, the relationship between Bayesian and maximin optimal designs is explored and powerful equivalence theorems and other associated results are presented. Several applications of this methodology are illustrated in Section 4 and some broad conclusions are given in Section 5. For ease of reading the proofs of all lemmas and theorems in the paper are included in an appendix.

2 Preliminaries

Consider a regression model which depends, possibly nonlinearly, on the parameters θ from a parameter space $\Theta \subset \mathbb{R}^k$ and on explanatory variables x varying in a compact design space $\mathcal{X} \subset \mathbb{R}^\ell$ equipped with a σ-field, which contains all one point sets. An approximate design ξ for this model is a probability measure on the design space $\mathcal{X}$ with finite support $x_1, \dots, x_n$ and weights $w_1, \dots, w_n$ representing the relative proportion of total observations taken at the corresponding design points [see e.g. Kiefer (1974)]. Let Ξ denote the class of all approximate designs and $\Delta \subset \Xi$ some subset of that class. Then, very broadly, an optimality criterion can be specified as

$$\psi : \Delta \times \Theta \to [0, \infty) ,$$

where the function $\psi(\xi, \theta)$ is continuous in the sense that, if a sequence of designs $\xi_n \in \Delta$ converges weakly to a design $\xi \in \Delta$ as $n \to \infty$, then

$$\lim_{n \to \infty} \psi(\xi_n, \theta) = \psi(\xi, \theta)$$

for all $\theta \in \Theta$. Additionally, for fixed $\xi \in \Delta$, the function $\psi(\xi, \theta)$ is assumed to be continuous in θ. Examples of such a criterion include, *inter alia*, D- and c-optimality [Pukelsheim (1993)].
In the present study attention is focussed on optimality criteria which accommodate uncertainty in the unknown parameters and, specifically, on criteria based on functions of the form $\psi(\xi, \theta)$. To this end it is first necessary to consider a single, fixed parameter value $\theta \in \Theta$ and to introduce a locally ψ-optimal design over the class of designs Δ as a design $\xi_\theta^* \in \Delta$ for which the condition

$$\psi(\xi_\theta^*, \theta) \geq \psi(\xi, \theta)$$

holds for all $\xi \in \Delta$. A standardized maximin ψ-optimal design in the class Δ can then be defined as a design which maximizes the criterion

$$\Psi_{-\infty}(\xi) = \inf_{\theta \in \Theta} \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)} \tag{2.1}$$

over all $\xi \in \Delta$ [see Dette (1997)] and a Bayesian ψ-optimal design with respect to a prior distribution π on the parameter space Θ as a design which maximizes

$$\Psi_0(\xi) = \exp \int_\Theta \log \psi(\xi, \theta) d\pi(\theta) \tag{2.2}$$

over the set Δ [see e.g. Pronzato and Walter (1985) or Chaloner and Larntz (1989)]. More generally, for fixed q such that $-\infty < q < 0$ a Bayesian Ψ_q-optimal design for a prior distribution π on Θ can be defined as a design $\xi \in \Delta$ maximizing the criterion

$$\Psi_q(\xi) = \left[\int_\Theta \left\{ \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)} \right\}^q d\pi(\theta) \right]^{\frac{1}{q}} \tag{2.3}$$

over the subclass of designs Δ [see Dette and Wong (1996)]. Note that the Bayesian ψ-optimality criterion (2.2) is obtained from (2.3) in the limit as $q \to 0$ and that the standardized maximin criterion (2.1) is recovered as $q \to -\infty$ provided the support of the prior π coincides with the parameter space Θ, i.e. $\mathrm{supp}(\pi) = \Theta$.

Remark 2.1. Note that a "relative" metric is used in the definition of the optimality criteria (2.1) and (2.3) because for different values of θ the values of the criteria $\psi(\xi, \theta)$ may not be not comparable. For example if $\psi(\xi, \theta) \geq \psi(\xi, \theta_0)$ for all $\theta \in \Theta$ and for any design ξ, a non-standardized maximin approach would always yield the locally optimal design $\xi^*_{\theta_0}$. Indeed the advantages of standardization are emphasized in Silvey (1980, pages 57-61) and clearly illustrated in Dette (1997). On the other hand all the results of this paper remain true for non-standardized criteria by simply omitting the terms $\psi(\xi^*_\theta, \theta)$ from the corresponding standardized criteria. An application of the non-standardized case is given in Example 4.5.

3 Bayesian and standardized maximin optimal designs

For every fixed design ξ, the criterion value $\Psi_q(\xi)$ converges to the value of the maximin criterion $\Psi_{-\infty}(\xi)$ as $q \to -\infty$. It is therefore tempting to surmise that this convergence is mirrored in the corresponding optimal designs themselves. The main result of the present study shows that, under fairly general conditions, standardized maximin ψ-optimal designs can indeed be obtained as weak limits of Bayesian Ψ_q-optimal designs as $q \to -\infty$. Although it often turns out that a maximin optimal design is also Bayesian optimal with respect to a least favourable prior distribution, the approach does not require that such a distribution is known. Moreover, no convexity assumption, either on the criterion or on the set of designs, is made.

Theorem 3.1. *Let Θ be compact and let π denote an arbitrarily chosen prior distribution on Θ with $\operatorname{supp}(\pi) = \Theta$. Suppose that the optimality criterion $\psi : \Delta \times \Theta \to (0, \infty)$ is continuous in each argument. Suppose that for every $q < 0$, ζ_q is a Bayesian Ψ_q-optimal design in the class of designs Δ with respect to the prior π and suppose also that the designs ζ_q converge weakly to some design $\zeta^* \in \Delta$ as $q \to -\infty$. Then the design ζ^* is standardized maximin ψ-optimal.*

Remark 3.2. For an efficient application of the results in this section it is important to note that Theorem 3.1 does not depend on the particular prior in the Bayesian optimality criterion. Therefore specific priors, which either allow an explicit calculation of the Bayesian Ψ_q-optimal designs and the corresponding limit as $q \to -\infty$ or which simplify the numerical construction of the Bayesian designs, can be used.

In practice it may well be possible to use Theorem 3.1 to construct a maximin ψ-optimal design over a class of designs Δ which is not necessarily convex such as, for example, a class of designs based on a fixed number of support points. Then the global optimality or otherwise of this design over a class of designs which is convex and which contains Δ, such as the class of all approximate designs Ξ, can be confirmed by invoking the general equivalence theorem formulated in Theorem 3.3 below.

Suppose that the Fisher information matrix for the parameter $\theta \in \Theta$ of a design $\xi \in \Delta$ can be expressed as

$$M(\xi, \theta) = \int_{\mathcal{X}} f(x, \theta) f^T(x, \theta) d\xi(x) \in \mathbb{R}^{\ell_\theta \times \ell_\theta}$$

where $f(x, \theta) \in \mathbb{R}^{\ell_\theta}$ is a vector-valued function appropriate to the specified regression model and the dimension l_θ may depend on θ. Then the criterion of interest has the form

$$\psi(\xi, \theta) = \phi_\theta\{C_\theta(\xi)\} \tag{3.1}$$

where $\phi_\theta(\cdot)$ is an information function in the sense defined by Pukelsheim (1993, page 119) and

$$C_\theta(\xi) = C_{K_\theta}(\xi,\theta) = \Big\{K_\theta^T M^-(\xi,\theta) K_\theta\Big\}^{-1}.$$

Here $K_\theta \in \mathbb{R}^{\ell_\theta \times s_\theta}$ represents a matrix of full column rank $s_\theta \le \ell_\theta$, $M^-(\xi,\theta)$ denotes a generalized inverse of $M(\xi,\theta)$, and it is assumed that $\xi \in \Delta$ is feasible, that is $\mathcal{R}(K_\theta) \subset \mathcal{R}(M(\xi,\theta))$ for all $\theta \in \Theta$.

An Equivalence Theorem for Bayesian Ψ_q-optimal and standardized maximin ψ-optimal designs based on criteria of the form (3.1) is now introduced and holds strictly for classes of designs Δ which are convex. The formulation adopted here is that of Pukelsheim (1993) and relies on the definition of the polar function of $\phi_\theta(\cdot)$ given by

$$\phi_\theta^\infty(D) = \inf_C \Big\{ \frac{\mathrm{tr}(CD)}{\phi_\theta(C)} \Big| C > 0 \Big\}$$

where C and D are nonnegative definite matrices. The proof of the next theorem follows essentially the same arguments as those presented in Pukelsheim (1993), Chapter 11, and is therefore omitted. An alternative formulation of a similar result using directional derivatives can be found in Fedorov (1980).

THEOREM 3.3. *Assume that the criterion $\psi(\xi,\theta)$ has the form (3.1) and that the class of designs Δ is convex. Assume also that a design denoted $\xi^* \in \Delta$ satisfies the condition $\mathcal{R}(K_\theta) \subset \mathcal{R}(M(\xi^*,\theta))$ for all $\theta \in \Theta$.*

(a) The design ξ^* is Bayesian Ψ_q-optimal in the class Δ with respect to a prior π on Θ if and only if for each $\theta \in \Theta$ there exists a nonnegative definite matrix D_θ which solves the polarity equation

$$\phi_\theta\{C_{K_\theta}(\xi^*)\}\phi_\theta^\infty(D_\theta) = \mathrm{tr}\{\mathrm{C}_{\mathrm{K}_\theta}(\xi^*)\mathrm{D}_\theta\} = 1 \tag{3.2}$$

and a generalized inverse of $M(\xi^*,\theta)$, say G_θ, such that the inequality

$$\int_\Theta \Big\{\frac{\psi(\xi^*,\theta)}{\psi(\xi_\theta^*,\theta)}\Big\}^q \mathrm{tr}\{\mathrm{M}(\eta,\theta)\mathrm{B}(\xi^*,\theta)\}\, \mathrm{d}\pi(\theta) \;-\; \int_\Theta \Big\{\frac{\psi(\xi^*,\theta)}{\psi(\xi_\theta^*,\theta)}\Big\}^{\mathrm{q}} \mathrm{d}\pi(\theta) \;\le\; 0 \tag{3.3}$$

holds for all $\eta \in \Delta$, where $B(\xi^*,\theta) = G_\theta K_\theta C_\theta(\xi^*) D_\theta C_\theta(\xi^*) K_\theta^T G_\theta$.

(b) Let

$$\mathcal{N}(\xi^*) := \Big\{\theta \in \Theta \mid \Psi_{-\infty}(\xi^*) = \frac{\psi(\xi^*,\theta)}{\psi(\xi_\theta^*,\theta)}\Big\}$$

denote the set of all parameter values in Θ, for which the minimum in (2.1) is attained. Then the design ξ^* is standardized maximin ψ-optimal in the class Δ if and only if there exists a prior π_ω on the set $\mathcal{N}(\xi^*)$, for each $\theta \in \mathrm{supp}(\pi_\omega)$ a nonnegative definite matrix D_θ satisfying (3.2) and a generalized inverse of $M(\xi^*,\theta)$, say G_θ, such that the inequality

$$\int_{\mathcal{N}(\xi^*)} \mathrm{tr}\{\mathrm{M}(\eta,\theta)\mathrm{B}(\xi^*,\theta)\}\, \mathrm{d}\pi_\omega(\theta) \;-\; 1 \;\le\; 0 \tag{3.4}$$

holds for all $\eta \in \Delta$.

Note that in the case of differentiability the left hand side of the inequality (3.3) is the directional derivative of the optimality criterion at the point ξ^* in the direction of η [see Silvey (1980)]. The more general formulation of Theorem 3.3 is required for non-differentiable criteria and singular information matrices. Morover, the second part of this theorem in effect states that the standardized maximin ψ-optimal design ξ^* coincides with the Bayesian Ψ_0-optimal design for the prior distribution π_w defined on the set $\mathcal{N}(\xi^*)$. The prior π_w is usually referred to as the least favourable or "worst" prior, a term which is borrowed from Bayesian decision theory [see Berger (1985), page 360]. The next result provides insight into the nature of the set $\mathcal{N}(\xi^*)$.

Lemma 3.4. *Suppose that the parameter space Θ comprises at least two points and that the class of designs Δ is convex. Then, for the standardized maximin ψ-optimal design $\xi^* \in \Delta$, the cardinality of the set $\mathcal{N}(\xi^*)$ defined in Theorem 3.3 is at least* 2.

In summary therefore, suppose that a candidate standardized maximin D-optimal design, say ξ_c^*, is available. Then the global optimality or otherwise of this design over a class of designs Δ which is convex can be confirmed by invoking Theorem 3.3 together with Lemma 3.4. The next two results follow directly from Theorem 3.3 and Lemma 3.4. The proofs are straightforward and are therefore omitted.

Lemma 3.5. *The Bayesian Ψ_q-optimal design ξ^* with respect to the prior π is Bayesian $\Psi_{q'}$-optimal with respect to the prior $\tilde{\pi}'$, where*

$$d\tilde{\pi}'(\theta) = \left(\frac{|\, M(\xi^*, \theta)\, |}{|\, M(\xi_\theta^*, \theta)\, |} \right)^{q-q'} d\pi(\theta)$$

and q and q' are such that $-\infty < q, q' \leq 0$.

Theorem 3.6. *The standardized maximin ψ-optimal design ξ^* is Bayesian Ψ_q-optimal with respect to the least favourable prior π_w on the set $\mathcal{N}(\xi^*)$ for all $q \leq 0$. Conversely, if the design ξ^* is Bayesian Ψ_q-optimal for all q such that $-\infty < q \leq 0$, then it is standardized maximin ψ-optimal.*

Note that Theorem 3.6 is closely related to results in Section 5.4 of Pshenichnyi (1971), which show that in cases, where optimization can be performed over a finite dimensional space, a maximin optimal design is also optimal with respect to a compound criterion of the form $\sum_{i=1}^N \lambda_i \psi(\xi, \theta_i)$. Here the quantities λ_i and θ_i are not known and correspond to the least favourable prior distribution [see also Cook and Fedorov (1995) for a similar relation in the context of constrained optimization]. The applicability of Theorem 3.6, or the corresponding equivalent formulation in Pshenichnyi (1971), is limited because in practice it is not easy to construct the unknown least favourable prior distribution. However, in cases, where optimization can be performed over a finite dimensional space, these results can be used to derive bounds on the number of support points of this distribution [see Example 4.5].

4 Applications

4.1 Nonlinear models. Consider a nonlinear model for which the response variable y follows a distribution from an exponential family with

$$E(y|x) = \eta(x, \theta) \qquad \text{and} \qquad \mathrm{Var}(y|x) = \sigma^2(x) \;, \tag{4.1}$$

where x represents an explanatory variable in the design space $\mathcal{X} \subset \mathbb{R}^\ell$ and θ is a vector of unknown parameters in the space $\Theta \subset \mathbb{R}^k$. If $\eta(x, \theta)$ is continuously differentiable with respect to θ, then the Fisher information matrix for θ at a single point x is given by

$$I(x, \theta) = \frac{1}{\sigma^2(x)} \left\{ \frac{\partial \eta(x,\theta)}{\partial \theta} \right\} \left\{ \frac{\partial \eta(x,\theta)}{\partial \theta} \right\}^T$$

and the information matrix for a design ξ belonging to a specified class of designs Δ can be expressed as $M(\xi, \theta) = \int_{\mathcal{X}} I(x, \theta)\, d\xi(x)$ [see e.g. Silvey (1980)]. For this model setting it is usual to consider criteria $\phi\{M(\xi, \theta)\}$, which are concave functions of the Fisher information matrix. Then the Bayesian Ψ_q-optimality criterion with respect to a prior π on Θ and the standardized maximin optimality criterion is obtained as a special case from (2.1) and (2.3) and the general theory with $K_\theta = I_k$ and $\psi(\xi, \theta) = \phi(M(\xi, \theta))$. The following corollary specifies the fairly general conditions under which Theorem 3.1 holds for the nonlinear models considered in this section. Note that, in the statement of the theorem, the set of all nonnegative definite matrices of order $k \times k$ is denoted $\mathrm{NND}(k)$.

Corollary 4.1. *Consider the nonlinear model specified by (4.1) and a local optimality criterion of the form $\psi(\xi, \theta) = \phi\{M(\xi, \theta)\}$, where $\phi(\cdot)$ is a continuous function from $\mathrm{NND}(k)$ to $[0, \infty)$. Let Θ be compact and let π represent any prior distribution on Θ for which $\mathrm{supp}(\pi) = \Theta$. Suppose that $\psi(\xi, \theta) > 0$ on $\Delta \times \Theta$ and that $I(x, \theta)$ is bounded and continuous on $\mathcal{X} \times \Theta$. Then, as $q \to -\infty$, the weak limit of the Bayesian Ψ_q-optimal designs with respect to the prior π in the class of designs Δ is a standardized maximin ϕ-optimal design, provided it belongs to Δ.*

Example 4.2. Consider the one-parameter logistic regression model with probability of success $1/\{1 + \exp(-(x - \theta))\}$ and $x \in I\!R$. Note that the information on θ at an observation x is given by

$$I(x, \theta) = \frac{\exp(-(x - \theta))}{\{1 + \exp(-(x - \theta))\}^2}$$

and is bounded and continuous. Note also that the locally D-optimal one-point design is located at $x^*_\theta = \theta$ with $I(x^*_\theta, \theta) = 1/4$. Suppose now that a parameter space of the form $\Theta = [-a, a]$ with $a > 0$ is of interest and that single-point standardized maximin D-optimal designs over that space are to be constructed. For a uniform prior on Θ, the one-point Bayesian Ψ_q-optimal design, say x_q, maximizes the criterion

$$\Psi_q(x) = \left\{ \frac{1}{2a} \int_{-a}^{a} \left[\frac{4 \exp(-(x - \theta))}{\{1 + \exp(-(x - \theta))\}^2} \right]^q d\theta \right\}^{\frac{1}{q}} \quad \text{for } -\infty < q < 0$$

and it is straightforward to show, either algebraically or by symmetry arguments, that $x_q = 0$ for all such q. Thus, since the conditions specified in Corollary 4.1 are satisfied for this example, it follows trivially that the one-point standardized maximin D-optimal design is given by $x^* = 0$.

Consider now invoking Theorem 3.3 in order to determine whether or not the design putting all observations at the point $x^* = 0$ is globally maximin D-optimal. The efficiency of this design relative to the locally optimal design $x^*_\theta = \theta$ is given by $4 \exp(\theta)/(1 + \exp(\theta))^2$ with minima at the end points of the interval $[-a, a]$ and thus the set $\mathcal{N}(\xi^\star)$ comprises the points $-a$ and a. Consider now a worst prior which puts equal weights on these points. Then the left hand side of (3.4) reduces to

$$\frac{1}{2}\left\{\exp(-x)\left(\frac{1+\exp(a)}{1+\exp(-(x-a))}\right)^2+\exp(-x)\left(\frac{1+\exp(-a)}{1+\exp(-(x+a))}\right)^2\right\}-1$$

and it can be shown numerically that this derivative is less than or equal to zero for all $x \in \mathbb{R}$ provided $0 < a \leq \ln(2+\sqrt{3})$. It therefore follows that the single-point design $x^* = 0$ is globally standardized maximin D-optimal on the parameter space $[-a, a]$ provided a satisfies this inequality, a result in accord with the finding of Haines (1995).

4.2 Model robust and discrimination designs. It is not uncommon for a practitioner to identify a set of plausible models, rather than a single model, as being appropriate for a particular data set. In order to accommodate such model uncertainty within the context of optimal design, criteria which are robust to the choice of model have been developed [see e.g. Läuter (1974)] and certain of these are explored here. To be specific, consider a class of linear models with means

$$E(y|x) = g(x,\theta) = \beta_0 f_0(x,\theta) + \ldots + \beta_{\ell_\theta} f_{\ell_\theta}(x,\theta),$$

where x belongs to some design space $\mathcal{X}$ and the regression functions $f_i(x,\theta), i = 0,\ldots,\ell_\theta$, are known, and with constant variances, σ^2. Each model is indexed by a parameter θ taken from a finite set of indices Θ and the class of such models is denoted $\mathcal{F} = \{g(x,\theta) \mid \theta \in \Theta\}$. Note that in many applications the models in the set $\mathcal{F}$ are nested but this is not necessary for the development of the robust design criteria described here.
The Fisher information matrix for the regression parameters $(\beta_0,\ldots,\beta_{\ell_\theta})$ in the model specified by $g(x,\theta)$ at a design $\xi \in \Delta$ can be expressed as

$$M(\xi,\theta) = \frac{1}{\sigma^2}\Big(\int_{\mathcal{X}} f_i(x,\theta) f_j(x,\theta) d\xi(x)\Big)_{i,j=0}^{\ell_\theta}$$

for $\theta \in \Theta$. Thus an optimal design which is robust to

choice of model over the class $\mathcal{F}$ should maximize an appropriate real valued function of the matrices $\{M(\xi,\theta) \mid \theta \in \Theta\}$ over the set of designs Δ [see e.g. Läuter (1974)]. In particular, suppose that a prior π on the index set Θ puts probability $\pi(\theta)$ on the parameter θ, where $\pi(\theta) \geq 0$ and $\sum_{\theta\in\Theta}\pi(\theta) = 1$. Suppose also that for each model $g(x,\theta)$ in the class $\mathcal{F}$, a criterion of the form $\psi(\xi,\theta) = \phi_\theta\{M(\xi,\theta)\}$, where $\phi_\theta(\cdot)$ is an information function, is of interest and that ξ^*_θ is the locally ϕ_θ-optimal design associated with this criterion. Then, following Läuter (1974), a Ψ_q-optimal robust design with respect to the prior π for the class of models $\mathcal{F}$ maximizes the criterion

$$\Psi_q(\xi) = \left[\sum_{\theta\in\Theta}\pi(\theta)\left\{\frac{\phi_\theta\{M(\xi,\theta)\}}{\phi_\theta\{M(\xi^*_\theta,\theta)\}}\right\}^q\right]^{\frac{1}{q}} \tag{4.2}$$

over the set of designs Δ. Furthermore, following Dette (1997), a standardized maximin optimal robust design for the class $\mathcal{F}$ maximizes the function

$$\min_{\theta\in\Theta}\left\{\frac{\phi_\theta\{M(\xi,\theta)\}}{\phi_\theta\{M(\xi^*_\theta,\theta)\}}\right\}$$

again over the set Δ. In view of Theorem 3.1, the standardized maximin robust designs can be found as weak limits of Ψ_q-optimal robust designs. These ideas are illustrated by means of the following example, which discusses the problem of identifying the degree of a polynomial regression.

EXAMPLE 4.3. Consider the class of nested polynomial models with means

$$g(x,\theta) = \beta_0 + \beta_1 x + \ldots + \beta_\theta x^\theta ,$$

where $x \in \mathcal{X} = [-1,1]$ and $\theta \in \Theta = \{1,\ldots,d\}$. Note that the regression functions are given by $f_i(x,\theta) = x^i$, $i = 0,\ldots,\theta$, and that the information matrix for the model of degree θ can be expressed as $M(\xi,\theta) = (\int_{\mathcal{X}} x^{i+j} d\xi(x))_{i,j=0}^{\theta}$. In order to obtain efficient designs for identifying the appropriate degree of the polynomial regression, Spruill (1990) proposed that a function of the criteria

$$\psi(\xi,\theta) = \phi_\theta\{M(\xi,\theta)\} = \frac{|M(\xi,\theta)|}{|M(\xi,\theta-1)|}$$

for $\theta \in \{1,\ldots,d\}$ should be maximized. Suppose now that a uniform prior π is placed on the index set Θ, i.e. $\pi(\theta) = \frac{1}{d}$ for $\theta \in \{1,\ldots,d\}$. Then the Ψ_q-optimal (discrimination) design with respect to the prior π, say ξ_q^*, maximizes (4.2) and can be characterized explicitly in terms of its canonical moments [see Dette and Studden (1997)]. In particular, by using results in Dette (1994) it can be shown that the canonical moments $(p_1,\ldots,p_{2d})$ of the Ψ_q-optimal discrimination design ξ_q^* are given by $p_{2d} = 1$, $p_{2j-1} = \frac{1}{2}$ for $j = 1,\ldots,d$, and by the system of equations

$$2^{2(d-\ell)} \left\{ \prod_{i=\ell+1}^{d-1} p_{2i}^{1+1/q} q_{2i}^{1-1/q} \right\} (1-p_{2\ell})^{1-1/q} (2p_{2\ell}-1)^{1/q} \;=\; 1 \;, \quad \ell = 1,\ldots,d-1 \;,$$

where $q_{2i} = 1 - p_{2i}$ and $\prod_d^{d-1}$ is interpreted as unity. As $q \to -\infty$ this latter system reduces to the recursion $p_{2\ell} = 1 - 2^{-2(d-\ell)} \prod_{i=\ell+1}^{d-1} (p_{2i} q_{2i})^{-1}$ and consequently ξ_q^* converges weakly to the design ξ^* with canonical moments $p_{2d} = 1, p_{2j-1} = \frac{1}{2}$ for $j = 1,\ldots,d$, and

$$p_{2\ell} = \frac{d-\ell+2}{2(d-\ell)+2}$$

for $\ell = 1,\ldots,d-1$. It now follows immediately from Theorem 3.1 that the design ξ^* is standardized maximin optimal. Moreover, by invoking Corollary 4.3.3 in Dette and Studden (1997), it is readily shown that the design ξ^* puts equal masses at the zeros of the ultraspherical polynomial $C_{d-1}^{(2)}(x)$ [see Szegö (1975)] and masses 1.5 times larger at the boundary points $+1$ and -1. The Ψ_q-optimal and the maximin optimal discrimination designs described here are in fact globally optimal in the sense that they are optimal over the class of all approximate designs, Ξ. Thus there exists a least favourable prior π_w on the index set Θ for which ξ^* is Ψ_0-optimal. Furthermore this prior can be obtained explicitly from the canonical moments of the optimal design ξ^* by invoking Theorem 6.2.3 of Dette and Studden (1997) and puts weights

$$\pi_w(\theta) = \frac{2(d-\theta+1)}{d(d+1)}$$

on the parameters $\theta \in \{1,\ldots,d\}$. For example, consider the case of $d = 4$. Then the standardized maximin optimal discrimination design has masses $\frac{1}{4}$, $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}$ and $\frac{1}{4}$ at the points -1, $\sqrt{3/8}$, 0, $\sqrt{3/8}$ and 1, respectively. The least favourable prior associates weights $2/5, 3/10, 1/5$ and $1/10$ with the polynomial models of degree $1, 2, 3$ and 4, respectively.

4.3 Designs for estimating nonlinear functions. Consider the homoscedastic linear regression model with mean

$$E(y|x) = \beta_0 f_0(x) + \beta_1 f_1(x) + \ldots + \beta_d f_d(x) \ . \tag{4.3}$$

Suppose that the parameters β belong to a space $\mathcal{B}$ and that a nonlinear function of those parameters, denoted $h(\beta)$, is of interest. Then the approximate asymptotic variance of the estimate of such a function is proportional to

$$\theta(\beta)^T M^-(\xi)\theta(\beta),$$

where $\theta(\beta)$ represents the vector of derivatives of $h(\beta)$ with respect to β, $M^-(\xi)$ is a generalized inverse of the information matrix and $\theta(\beta) \in \mathcal{R}(M(\xi))$. Optimal designs which in some sense minimize this variance are now sought. For ease of notation, consider the induced parameter space $\Theta = \{[\theta(\beta)^T\theta(\beta)]^{-\frac{1}{2}}\theta(\beta) : \beta \in \mathcal{B}\}$. Then an appropriate optimality criterion can be formulated as

$$\psi(\xi, \theta) = \begin{cases} \left\{\theta^T M^-(\xi)\theta\right\}^{-1} & \text{for } \theta \in \mathcal{R}(M(\xi)) \\ 0 & \text{otherwise} \end{cases}$$

and the locally optimal design ξ_θ^* maximizes this criterion. The design problem so described occurs, for example, when the turning point of a quadratic regression function is of interest [see e.g. Chaloner (1989)] and also in the context of constructing optimal extrapolation designs for an interval [see e.g. Spruill (1987)]. The definitions of Bayesian Ψ_q-optimal and of standardized maximin ψ-optimal designs based on the above criterion follow directly from the general formulations given in Section 2. Furthermore Theorem 3.1 holds under the conditions specified in the following corollary.

Corollary 4.4. *Let π denote any prior distribution on Θ with* $\operatorname{supp}(\pi) = \Theta$. *Suppose that the functions $f_0(x), \ldots, f_d(x)$ in model (4.3) are continuous and bounded and that the locally optimal criterion value $\psi(\xi_\theta^*, \theta)$ is continuous in θ. Then the weak limit of the Bayesian Ψ_q-optimal designs in the class of designs Δ as $q \to -\infty$ is a standardized maximin ψ-optimal design, provided the limiting design belongs to Δ and is non-singular, i.e. its Fisher information matrix is non-singular.*

Example 4.5. Consider the polynomial regression model $\beta_0 + \beta_1 x + \ldots + \beta_d x^d$ on the interval $[-1, 1]$ and suppose that interest centres on estimating the function $h(\beta) = \frac{1}{2}(\beta_{d-1}^2 + \beta_d^2)$. It is well known [see Pukelsheim (1993)] that there always exists a symmetric optimal design relating to this problem. The asymptotic variance of the estimate of $h(\beta)$ for such a design is given by $\delta^{-1}(\xi, \beta)$, where

$$\delta(\xi, \beta) = \delta(\xi, \beta_{d-1}, \beta_d) = \Big(\beta_{d-1}^2 v_1(\xi) + \beta_d^2 v_2(\xi)\Big)^{-1},$$

and $v_1(\xi)$ and $v_2(\xi)$ are the diagonal elements of the inverse of the information matrix $M^{-1}(\xi) = \big\{\big(\int_{[-1,1]} x^{i+j} d\xi(x)\big)_{i,j=0}^d\big\}^{-1}$ corresponding to β_{d-1} and β_d respectively. Suppose now that a prior distribution π which is defined on a compact set $\mathcal{B} \subset \mathbb{R}^2\backslash\{0\}$ is placed on the parameters β_{d-1} and β_d. If no standardization is used a Bayesian Ψ_q-optimal design for this prior maximizes the criterion

$$\Psi_q(\xi) = \left\{\int_{\mathcal{B}} \{\delta(\xi, \beta_{d-1}, \beta_d)\}^q \, d\pi(\beta_{d-1}, \beta_d)\right\}^{1/q}. \tag{4.4}$$

Moreover, it is well known that there exists an optimal design with at most $d+1$ support points. Therefore it follows from Pshenichnyi (1971) that there exists a least favourable distribution associated with the maximin design which is based on at most $2d+2$ support points. The following two examples involve different parameter sets $\mathcal{B}$ and are introduced in order to demonstrate the potential applications of Theorem 3.1.

(a) Suppose that the prior π on the parameters β_{d-1} and β_d is defined on the upper half of the unit circle, i.e. $\mathcal{B} = \{(\beta_{d-1}, \beta_d) \mid \beta_{d-1}^2 + \beta_d^2 = 1,\ \beta_d \geq 0\}$. Then the maximin criterion of interest reduces to

$$\min_{\beta_{d-1}^2+\beta_d^2=1, \beta_d \geq 0} \delta(\xi, \beta_{d-1}, \beta_d). \tag{4.5}$$

Note that no standardization in (4.5) is used here since this is implicit in the definition of the set $\mathcal{B}$. The Bayesian Ψ_q-optimality criterion (4.4) can be maximized by invoking arguments similar to those presented in Studden (1989). Specifically the resultant optimal design has canonical moments of odd order given by $p_{2i-1} = \frac{1}{2}$ for $i = 1, \ldots, d$, while those of even order maximize

$$\Psi_q(\xi) = \prod_{j=1}^{d-1} q_{2j-2}p_{2j} \Big\{ \int_0^1 \Big(1 - \beta_d^2 + \frac{\beta_d^2}{q_{2d-2}p_{2d}}\Big)^{-q} d\pi(\beta_d) \Big\}^{1/q} \tag{4.6}$$

Suppose now that the prior distribution on β_d is taken to be $d\pi(\beta_d) = 2\beta_d d\beta_d$. Then the integral in (4.6) can be evaluated explicitly and it follows that $p_{2i} = \frac{1}{2}$ for $i = 1, \ldots, d-2$, $p_{2d} = 1$, and that p_{2d-2} is given by the unique solution of the equation

$$1 - z - q(1 - 2z) = (1 - q(1 - z))(1 - z)^{-q+1}$$

in the interval $(0, 1)$. It is easy to see that this solution converges to $1/2$ as $q \to -\infty$, and consequently that the required maximin optimal design, i.e. the design ξ^* maximizing (4.5), has canonical moments $p_i = 1/2$ for $i = 1, \ldots, 2d-1$ and $p_{2d} = 1$. Then it follows immediately from Corollary 4.3.3 of Dette and Studden (1997) that ξ^* puts masses $1/d$ at the roots of the Chebyshev polynomial of the second kind $U_{d-1}(x)$ and masses $1/2d$ at the points -1 and 1. Thus the maximin optimal design in fact coincides with the D_1-optimal design for polynomial regression of degree d on the interval $[-1, 1]$.

(b) This second example illustrates the application of Theorem 3.1 to the numerical construction of maximin optimal designs and also involves no standardization of the criterion. Specifically, suppose that the set $\mathcal{B}$ is chosen to be $[1, 2] \times [2, 3]$ and that the prior distribution on $\mathcal{B}$ has density proportional to $\beta_{d-1}\beta_d d\beta_{d-1} d\beta_d$. In this case the Bayesian Ψ_q-optimality criterion relating to (4.4) can be maximized numerically in terms of canonical moments. It follows algebraically that $p_i = \frac{1}{2}$ for $i = 1, \ldots, 2d-3$, $p_{2d-1} = \frac{1}{2}$, $p_{2d} = 1$, while the canonical moment p_{2d-2} depends on the parameter q and can easily be obtained numerically. Values of p_{2d-2} for selected values of q are given in the following table

q	-5	-10	-20	-50	-100	-200	$-\infty$
p_{2d-2}	0.5834	0.5776	0.5801	0.5834	0.5846	0.5852	0.5858

The Bayesian Ψ_q- and the maximin optimal designs corresponding to these canonical moments can then be obtained from the results in Studden (1989). In particular the support points

$-1 = x_0 < x_1 < \ldots < x_{d-1} < x_d = 1$ of these designs correspond to the roots of the polynomial $(x^2-1)\Big\{U_{d-1}(x)+(2p_{2d-2}-1)U_{d-3}(x)\Big\}$ and the masses are given by $\frac{p_{2d-2}}{2(1+p_{2d-2}(d-2))}$ for the points -1 and 1 and by

$$\xi^*(x_j) = \Big[d-1-\frac{2p_{2d-2}U_{d-2}(x_j)}{U_d(x_j)+(2p_{2d-2}-1)U_{d-2}(x_j)}\Big]^{-1}$$

for the interior support points $x_j, j = 1, \ldots, d-1$. For example, for polynomial regression of degree 5 on the interval $[-1, 1]$, the requisite maximin optimal design can be obtained numerically and puts masses $0.1885, 0.2053$ and 0.1062 on the points $\pm 0.2880, \pm .7900$ and ± 1 respectively.

5 Conclusions

This study provides a cohesive approach to the construction of maximin optimal designs for a broad range of nonlinear model settings. It is demonstrated that under fairly general conditions Bayesian Ψ_q-optimal designs converge to standardized maximin optimal designs. Before the results can be implemented however, Bayesian Ψ_q-optimal designs for the model, the optimality criterion and the class of designs of interest must be constructed. In some cases such designs are available in the literature. Then, in implementing Theorem 3.1, it is necessary to find the requisite maximin optimal design as the limit of the appropriate Bayesian Ψ_q-optimal designs.

On the other hand for many nonlinear model settings it is possible that Bayesian Ψ_q-optimal designs cannot be obtained in an explicit algebraic form. In such cases these Bayesian optimal designs can usually be calculated numerically for a range of increasingly negative q values and the limiting and hence the standardized maximin optimal design identified, at least approximately. Specific choices of the prior distribution, as for example discrete approximations of the uniform distribution, can simplify this numerical construction substantially.

A secondary but nevertheless important feature of the present study is the suite of results for convex classes of designs presented in Section 3 and based on Theorem 3.3. These results provide considerable insight into the nature of standardized maximin optimal designs and their relation to the Bayesian Ψ_q-optimal designs and in addition provide tools for confirming the global optimality or otherwise of candidate designs. However it should immediately be emphasized that, while a standardized maximin optimal design is globally optimal provided it is Bayesian Ψ_0-optimal for some least favourable prior, the identification of that prior is not straightforward.

Acknowledgements. The authors are grateful to Isolde Gottschlich who typed most parts of this paper with considerable technical expertise. The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, Reduction for complexity in multivariate data structures), and of the University of Natal and the National Research Foundation of South Africa is gratefully acknowledged. The authors are also grateful to two unknown referees for their helpful comments which led to a substantial improvement of an earlier version of this paper.

References

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

Chaloner, K. (1989). Bayesian design for estimating the turning point of a quadratic regression. *Commun. Statist. Theor. & Meth.* **18**, 1385-1400.

Chaloner, K. (1993). A note on optimal Bayesian design for nonlinear problems. *J. Statist. Plann. Inf.* **37**, 229-235.

Chaloner, K. and Larntz, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inf.* **21**, 191-208.

Chaloner, K. and Verdinelli, I. (1995). Bayesian experimental design: a review. *Statistical Science* **10**, 273-304.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* **24**, 586-602.

Cook, D., Fedorov, V.V. (1995). Constrained optimization of experimental designs. *Statistics* **26**, 129-178.

Dette, H. (1994). Discrimination designs for polynomial regression on compact intervals. *Ann. Statist.* **22**, 890-903.

Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.

Dette, H. and Studden, W.J. (1997). *The Theory of Canonical Moments with Applications in Statistics, Probability, and Analysis.* Wiley, New York.

Dette, H. and Wong, W.K. (1996). Optimal Bayesian designs for models with partially specified heteroscedastic structure. *Ann. Statist.* **24**, 2108-2127.

Fedorov, V.V. (1980). Convex design theory. *Math. Operat. Statist.* **11**, 403-413.

Haines, L.M. (1995). A geometric approach to optimal design for one-parameter non-linear models. *J. Roy. Statist. Soc. Ser. B* **57**, 575-598.

Hewitt, E. and Stromberg, K. (1965). *Real and Abstract Analysis.* Springer, New York.

Imhof, L.A. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Ann. Statist.* **29**, 561-576.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849-879.

Läuter, E. (1974). Experimental design in a class of models. *Math. Operat. Statist.* **5**, 379-398.

Müller, C.H. (1995). Maximin efficient designs for estimating nonlinear aspects in linear models. *J. Statist. Plann. Inf.* **44**, 117-132.

Müller, C.H. and Pázman, A. (1998). Applications of necessary and sufficient conditions for maximin efficient designs. *Metrika* **48**, 1-19.

Pshenichnyi, B. (1971). *Necessary Conditions for an Extremum.* Marcel Dekker, New York.

Pronzato, L. and Walter, E. (1985). Robust experiment design via stochastic approximation. *Math. Biosci.* **75**, 103-120.

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

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

Spruill, M.C. (1987). Optimal designs for interpolation. *J. Statist. Plann. Inf.* **16**, 219-229.

Spruill, M.C. (1990). Good designs for polynomial extrapolation. *J. Statist. Plann. Inf.* **26**, 149-159.

Studden, W.J. (1989). Note on some ϕ_p-optimal designs for polynomial regression. *Ann. Statist.* **17**, 618-623.

Szegö, G. (1975). *Orthogonal Polynomials.* American Mathematical Society, Providence, Rhode Island.

Wong, W. K. (1992). A unified approach to the construction of minimax designs. *Biometrika* **79**, 611-619.

APPENDIX

PROOF OF THEOREM 3.1. Note first that the continuity of ψ implies that the normalizing function

$$\nu(\theta) := \psi(\xi_\theta^*, \theta)\ , \quad \theta \in \Theta,$$

is lower semicontinuous. For if $\theta \in \Theta$ and $\{\theta_j\}_{j=1}^\infty \subset \Theta$ is a sequence that converges to θ, then

$$\liminf_{j\to\infty} \nu(\theta_j) = \liminf_{j\to\infty} \psi(\xi_{\theta_j}^*, \theta_j) \geq \liminf_{j\to\infty} \psi(\xi_\theta^*, \theta_j) = \psi(\xi_\theta^*, \theta) = \nu(\theta).$$

Let $\epsilon > 0$, and let $\theta_0 \in \Theta$ be such that

$$\frac{\psi(\zeta^*, \theta_0)}{\nu(\theta_0)} \ \leq \ \Psi_{-\infty}(\zeta^*) + \epsilon.$$

Then, since ν is lower semicontinuous and ψ is continuous, there is a relatively open neighborhood $U \subset \Theta$ of θ_0 such that

$$\frac{\psi(\zeta^*, \theta)}{\nu(\theta)} \leq \Psi_{-\infty}(\zeta^*) + 2\epsilon \qquad \text{for all } \theta \in U.$$

As $\mathrm{supp}(\pi) = \Theta$, $\pi(U) > 0$. Since ζ_q converges weakly to ζ^*,

$$\frac{\psi(\zeta_q, \theta)}{\nu(\theta)} \ \to \ \frac{\psi(\zeta^*, \theta)}{\nu(\theta)}$$

for every $\theta \in \Theta$. It therefore follows from Egorov's theorem [see e.g. Hewitt and Stromberg (1965), page 158] that there exist a measurable set $V \subset \Theta$ with $\pi(V) > 1 - \frac{1}{2}\pi(U)$ and a number $q_0 < 0$ such that

$$\left| \frac{\psi(\zeta_q, \theta)}{\nu(\theta)} - \frac{\psi(\zeta^*, \theta)}{\nu(\theta)} \right| \leq \epsilon \qquad \text{for all } \theta \in V \text{ and all } \ -\infty < q \leq q_0.$$

Thus for all $-\infty < q \leq q_0$,

$$\{\Psi_q(\zeta_q)\}^q \geq \int_{U\cap V} \left\{ \frac{\psi(\zeta_q, \theta)}{\nu(\theta)} \right\}^q d\pi(\theta) \geq \{\Psi_{-\infty}(\zeta^*) + 3\epsilon\}^q \, \pi(U \cap V).$$

Obviously, $\pi(U \cap V) > 0$, and it follows that

$$\limsup_{q\to-\infty} \Psi_q(\zeta_q) \leq \{\Psi_{-\infty}(\zeta^*) + 3\epsilon\} \limsup_{q\to-\infty} \{\pi(U \cap V)\}^{\frac{1}{q}} = \Psi_{-\infty}(\zeta^*) + 3\epsilon.$$

As $\epsilon > 0$ was arbitrary, one has

$$\limsup_{q\to-\infty} \Psi_q(\zeta_q) \leq \Psi_{-\infty}(\zeta^*).$$

Consequently, if $\xi \in \Delta$ is any competing design, then

$$\Psi_{-\infty}(\xi) = \lim_{q \to -\infty} \Psi_q(\xi) \le \limsup_{q \to -\infty} \Psi_q(\zeta_q) \le \Psi_{-\infty}(\zeta^*).$$

This proves that ζ^* is indeed a standardized maximin optimal design in the class Δ. $\square$

Proof of Lemma 3.4. Let ξ^* denote the standardized maximin optimal design and assume that $\mathcal{N}(\xi^*) = \{\theta_0\}$ is a singleton. Then the Equivalence Theorem 3.3 for standardized maximin optimality shows that ξ^* is locally D-optimal for the parameter θ_0 in the class Δ. Therefore

$$1 = \frac{\psi(M(\xi^*, \theta_0))}{\psi(M(\xi^*_{\theta_0}, \theta_0))} = \min\Big\{ \frac{\psi(M(\xi^*, \theta))}{\psi(M(\xi^*_\theta, \theta))} \;\Big|\; \theta \in \Theta \Big\} \;\le\; 1,$$

which implies $\mathcal{N}(\xi^*) = \Theta$ contradicting the hypothesis that $\#\mathcal{N}(\xi^*) = 1$. $\square$

Proof of Corollary 4.1. The assumption that $I(x, \theta)$ is continuous and bounded implies that for every fixed θ, the criterion

$$\psi(\xi, \theta) = \phi\{M(\xi, \theta)\}$$

is continuous in ξ. The assumption also implies by Lebesgue's convergence theorem that for every ξ, $\psi(\xi, \theta)$ is continuous in θ. The assertion now follows from Theorem 3.1. $\square$

Proof of Corollary 4.4. Let ζ_q be the Ψ_q-optimal designs in the class Δ, so that as $q \to -\infty$, ζ_q converges weakly to a non-singular design ζ^*. Then

$$\lim_{q \to -\infty} M(\zeta_q) = M(\zeta^*),$$

in particular $M(\zeta_q)$ is non-singular for $q \le q_0$, say. Hence $\psi(\zeta_q, \theta)$ converges to $\psi(\zeta^*, \theta)$ for each θ. Thus for $q \le q_0$,

$$\frac{\psi(\zeta_q, \theta)}{\psi(\xi^*_\theta, \theta)}$$

is continuous and for $q \to -\infty$, converges to

$$\frac{\psi(\zeta^*, \theta)}{\psi(\xi^*_\theta, \theta)}.$$

An inspection of the proof of Theorem 3.1 shows that this is sufficient to ensure that ζ^* is a standardized maximin optimal design. $\square$