Bayesian and maximin optimal designs for heteroscedastic 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 KwaZulu-Natal,
Pietermaritzburg, South Africa
haines@nu.ac.za

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

November 21, 2003

Key words and phrases: Bayesian design; D-optimal design; maximin design; polynomial regression; standardized criterion.

MSC 2000: Primary 62K05.

Abstract

The problem of constructing standardized maximin D-optimal designs for weighted polynomial regression models is addressed. In particular it is shown that, by following the broad approach to the construction of maximin designs introduced recently by Dette, Haines and Imhof (2003), such designs can be obtained as weak limits of the corresponding Bayesian Φ_q-optimal designs. The approach is illustrated for two specific weighted polynomial models and also for a particular growth model.

1. Introduction

There has been considerable interest in recent years in the construction of optimal designs for weighted polynomial regression models (see e.g. Haines, Dette and Imhof (1999), Imhof (2001), and references therein). The essential problem within this setting is that the Fisher information matrix for the regression parameters depends, not on the regression parameters themselves, but rather on the unknown parameters describing the heteroscedasticity of the model. There are three broad approaches to this problem, all based on constructing designs which in some sense optimize a function of the information matrix. In particular it is common, following Chernoff (1953), to assume a "best guess" for the unknown parameters and to construct a "locally" optimal design. Such designs are however not necessarily robust to the choice of parameter value. A more flexible approach involves invoking the Bayesian paradigm and averaging an appropriate criterion based on the information matrix over a prior distribution on the parameters (Chaloner and Verdinelli (1995)). This approach is also somewhat restrictive in the sense that the resultant designs can depend quite strongly on the choice of prior distribution. Alternatively, and less stringently, it is possible to adopt a maximin strategy and, specifically, to consider maximizing the minimum of a function of the information matrix taken over a specified range of the unknown parameters (Silvey (1980), p. 59))

The maximin approach is, arguably, the most appealing of the three approaches described above. However maximin criteria are not differentiable, the attendant Equivalence Theorems are difficult to invoke in practice and as a consequence the problem of constructing the associated maximin optimal designs is a challenging one (see e.g. Wong (1992), Müller (1995) and Wiens (1998)). A number of methods for constructing maximin designs have been proposed in the literature. In particular Sitter (1992), King and Wong (2000) and Fandom Noubiap and Seidel (2000) present algorithms for the construction of maximin designs which can be implemented numerically. Further Haines (1995), Imhof (2001) and Biedermann and Dette (2003a) derive explicit expressions for maximin designs for certain model settings but the underlying arguments are mathematically intricate and also case-specific. A particularly attractive approach to the construction of maximin designs, and one that promises to be broadly applicable, is that introduced recently by Dette, Haines and Imhof (2003). Specifically, these authors demonstrate that, under fairly general conditions, the Bayesian Φ_q-optimal designs introduced by Dette and Wong (1996) converge to the corresponding maximin optimal designs as the index q approaches minus infinity, a convergence which mirrors the well-known convergence of the associated criteria. The approach has been illustrated with examples involving nonlinear models and model robust and discrimination designs by Dette, Haines and Imhof (2003) and implemented numerically for binary response models by Biedermann and Dette (2003b).

The aim of the present paper is to demonstrate that the broad approach of Dette, Haines and Imhof (2003) to the construction of maximin designs can be applied to the weighted polynomial regression model setting, and in particular to two such models for which the efficiency functions capturing the heteroscedasticity have not been widely studied. The relevant definitions and an extension to the theorem fundamental to the work of Dette, Haines and Imhof (2003) are presented in Section 2. The construction of Bayesian Φ_q-optimal designs for the weighted polynomial regression models of interest is described in Section 3 and the construction of the associated maximin D-optimal designs is discussed in Section 4. In particular explicit expressions for both the Bayesian Φ_q-optimal and the maximin D-optimal designs are presented. In Section 5 a further example, based on a growth model for which the optimal design problem coincides with that for a specific weighted polynomial regression model, is introduced in order to demonstrate the wider applicability of the method. Finally the results are summarized and some broad conclusions drawn in Section 6.

2. Preliminaries

Consider the weighted polynomial regression model of degree d

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_d x^d + \epsilon \tag{2.1}$$

where y is the response corresponding to an explanatory variable x taken from some design space $\mathcal{X} \subset \mathbb{R}$, $\beta = \{\beta_0, \ldots, \beta_d\}$ is a vector of unknown parameters and ϵ is a random term with mean 0 and variance $\sigma^2/\lambda(x,\theta)$, with $\lambda(x,\theta)$ an efficiency function depending on a single parameter θ. An approximate design ξ for model (2.1) is a probability measure on the design space $\mathcal{X}$ with finite support $x_1, \ldots, x_n$ and associated weights $w_1, \ldots, w_n$ respectively. The information matrix for the regression parameters $\beta = \{\beta_0, \ldots, \beta_d\}$ depends only on the parameter θ and has the form

$$M(\xi,\theta) = \int_{\mathcal{X}} f(x) f^T(x) \lambda(x,\theta)\, d\xi(x)$$

where $f(x) = \{1, x, x^2, \ldots, x^d\}$. In the present study interest centres primarily on two efficiency functions, the one

$$\lambda(x,\theta) = (1+x^2)^{-\theta}$$

with $x \in I\!R$ and $\theta > d$ introduced by Dette, Haines and Imhof (1999) in the context of locally optimal designs and the other

$$\lambda(x,\theta) = (1+x)^{-\theta}$$

with $x \in [0,\infty)$ and $\theta > 2d$ not used previously in the literature as an efficiency function.

Optimal designs are those designs which in some sense maximize the information matrix. In the present context interest focusses on Bayesian Φ_q-optimal and standardized maximin D-optimal designs for model (2.1) and these are defined as follows. Let ξ_θ^* denote the locally D-optimal design, i.e. that design which maximizes the determinant of $M(\xi,\theta)$ for a specific parameter value θ. Then the Bayesian Φ_q-optimal design is the design which maximizes the criterion

$$\Phi_q(\xi) = \{\int_\Theta \left\{ \frac{\mid M(\xi,\theta) \mid}{\mid M(\xi_\theta^*,\theta) \mid} \right\}^q d\pi(\theta)\}^{1/q} = \{\int_\Theta \mid M(\xi,\theta) \mid^q d\tilde{\pi}(\theta)\}^{1/q}$$

where $-\infty < q \leq \frac{1}{d+1}$ and $d\tilde{\pi}(\theta) = \mid M(\xi_\theta^*,\theta) \mid^{-q} d\pi(\theta)$, with $\pi(\theta)$ a prior distribution placed on the parameter $\theta \in \Theta$ (Dette and Wong, 1996). Note that for $q = 0$ the Bayesian D-optimality criterion is recovered. Further the standardized maximin D-optimal design is that design which maximizes

$$\min_{\theta\in\Theta} \frac{\mid M(\xi,\theta) \mid}{\mid M(\xi_\theta^*,\theta) \mid} = \min_{\theta\in\Theta} R(\xi,\theta)$$

where the parameter θ is assumed to belong to some specified parameter space Θ, an assumption less stringent than that of invoking a prior distribution.

Bayesian Φ_q-optimal and standardized maximin D-optimal designs are intimately related. In particular, Dette and Wong (1996) observed, and indeed it is clear, that the criterion $\Phi_q(\xi)$ converges to the standardized maximin D-optimal criterion as $q \to -\infty$. Further, and more importantly from the point of view of design construction, Dette, Haines and Imhof (2003) presented a powerful result which mirrors this convergence in criterion with convergence in design. Their result is couched in very general terms and relates to an optimality criterion $\psi(\xi, \theta)$ with designs ξ belonging to some design space Δ not necessarily convex and with θ some unknown parameter in a possibly non-linear model. In this section a generalization of this result is presented, where the prior distribution is allowed to depend on q. Thus for each q the prior may be chosen to simplify the calculation of the corresponding Bayesian design. Moreover, conditions are specified which ensure that the Bayesian designs do have a limit, whereas in Dette, Haines and Imhof (2003), the existence of a limit design was part of the assumptions.

Throughout this paper, Δ is the set of all competing designs endowed with the Prohorov metric [see Billingsley (1999)]. The convergence theorem applies to general Bayesian Ψ_q- and standardized maximin ψ-optimal designs. The definitions of these designs are the same as those for Bayesian Φ_q- and standardized maximin D-optimal designs given above but with $|M(\xi, \theta)|$ replaced by the more general criterion $\psi(\xi, \theta)$. Thus for $\xi \in \Delta$ let

$$\Psi_q(\xi) = \left[\int_\Theta \left\{ \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)} \right\}^q d\pi_q(\theta) \right]^{\frac{1}{q}} \quad (-\infty < q < 0), \qquad \Psi_{-\infty}(\xi) = \min_{\theta \in \Theta} \frac{\psi(\xi, \theta)}{\psi(\xi_\theta^*, \theta)},$$

where ξ_θ^* is the locally $\psi(\cdot, \theta)$-optimal design in Δ.

THEOREM 2.1. *Let Θ be compact. Suppose the optimality criterion $\psi : \Delta \times \Theta \to (0, \infty)$ is continuous in each argument. For every $q < 0$, let π_q denote an arbitrarily chosen prior distribution on Θ and let ζ_q be a Bayesian Ψ_q-optimal design with respect to the prior π_q. Suppose that the following conditions hold.*

a) There is at most one standardized maximin ψ-optimal design.

b) The class of Bayesian designs $\{\zeta_q : q < 0\}$ is tight and its closure is contained in Δ.

c) There is a finite measure π on Θ with supp$(\pi) = \Theta$ such that for every measurable subset $T \subset \Theta$ with $\pi(T) > 0$,

$$\liminf_{q \to -\infty} \pi_q(T) > 0.$$

It then follows that the Bayesian designs ζ_q converge weakly to a design ζ^* in Δ and that ζ^* is a standardized maximin optimal design.

The proof is given in the Appendix. The result can be used to obtain the maximin optimal designs as limits ($q \to -\infty$) from the Bayesian optimal designs, which are usually easier to calculate because of the differentiability of the Bayesian optimality criteria. In contrast to the result of Dette, Haines and Imhof (2003) Theorem 2.1 guarantees the convergence of the sequence of Bayesian optimal designs. Moreover, the introduction of priors depending on the parameter q gives the statistician extra flexibility to simplify the numerical or analytical calculations of the Bayesian optimal designs.

If condition a) in Theorem 2.1 is not satisfied, there may exist sequences of Bayesian optimal designs that converge to different limits. It follows from the proof of Theorem 2.1 that any such limit design is a standardized maximin optimal design. Condition b) is satisfied if the design space $\mathcal{X}$ is compact and Δ is the set of all probability measures. In applying Theorem 2.1, conditions a) and b) can be disregarded if it is known in advance that the Bayesian designs converge to some design in Δ. Condition c) is met if all priors have densities with respect to a dominating finite measure π and the densities are uniformly bounded away from zero. Condition c) is trivially met if all the prior distributions π_q are the same with common support Θ. In this sense Theorem 2.1 contains the result of Dette, Haines and Imhof (2003) as a special case.

The following example shows that some condition on the priors π_q must be imposed. For arbitrary priors π_q with $\mathrm{supp}(\pi_\mathrm{q}) = \Theta$ the Bayesian Ψ_q-optimal designs need not converge to a standardized maximin design. In fact, they need not converge at all, and even if they do, they can converge to any prescribed design.

Example 2.1. Consider the non-linear homoscedastic regression model $E[Y(x)] = \exp(-\theta x)$, $x \geq 0$, with $\Theta = \{\theta_1, \theta_2\}$, $0 < \theta_1 < \theta_2$. Let Δ be the class of one-point designs and $\psi(\xi, \theta) = \int x^2 \exp(-2\theta x)\, d\xi(x)$. Any prior π_q with $\mathrm{supp}(\pi_\mathrm{q}) = \Theta$ is of the form $\pi_q(\theta_1) = \alpha_q$, $\pi_q(\theta_2) = 1 - \alpha_q$, where $0 < \alpha_q < 1$. The Bayesian Ψ_q-optimal design point $x = x_q$ is the unique solution in $(\theta_2^{-1}, \theta_1^{-1})$ of

$$\left\{ \frac{\theta_1}{\theta_2} e^{(\theta_2 - \theta_1)x} \right\}^{2q} = \left(\frac{1}{\alpha_q} - 1 \right) \frac{\theta_2 x - 1}{1 - \theta_1 x}. \tag{2.2}$$

Now fix any point $x^* \in (\theta_2^{-1}, \theta_1^{-1})$. Then for every $q < 0$ there is a unique $\alpha_q \in (0, 1)$ such that (2.2) holds with $x = x^*$. For this choice of priors, every Bayesian Ψ_q-optimal design is the unit mass at x^*, and so the limit, as $q \to -\infty$, is the unit mass at x^*, too. As x^* was arbitrary, the limit design will in general not be a standardized maximin optimal design. An obvious modification of the argument yields prior distributions such that the corresponding Bayesian designs do not converge at all.

The applicability of Theorem 2.1 to the setting of model (2.1) is somewhat intricate and is introduced later as a formal result emanating from the present study.

3. Bayesian Φ_q-optimal designs

3.1 $(d+1)$-point Bayesian Φ_q-optimal designs

Observe first that a $(d+1)$-point optimal design necessarily puts equal masses at its support points $x_1, \ldots, x_{d+1}$ (see e.g. Silvey (1980), p.43). It then follows that the determinant of the information matrix for β at such a design ξ can be written as

$$\mid M(\xi, \theta) \mid = \frac{1}{(d+1)^{d+1}} \mid X_R \mid^2 \prod_{i=1}^{d+1} \lambda(x_i, \theta),$$

where X_R is the Vandermonde matrix with ith row $\{1 \; x_i \; x_i^2 \; \ldots \; x_i^d\}$ and hence that

$$\Phi_q(\xi) = \frac{1}{(d+1)^{d+1}} \mid X_R \mid^2 \{ \int_\Theta \prod_{i=1}^{d+1} \lambda(x_i, \theta)^q d\tilde{\pi}(\theta) \}^{1/q}.$$

Dette and Wong (1996) derived $(d+1)$-point Bayesian Φ_q-optimal designs for a range of efficiency functions using arguments based on canonical moments. Their approach is not completely general however and specifically does not hold for the case where $\lambda(x, \theta) = (1 + x^2)^{-\theta}$ and $\lambda(x, \theta) = (1 + x)^{-\theta}$. The following theorems were derived using results from the theory of differential equations.

The first lemma and theorem relate to the efficiency function $(1 + x^2)^{-\theta}$ and introduce support points at the roots of an ultraspherical polynomial.

Lemma 3.1. *Suppose that $\lambda(x, \theta) = (1 + x^2)^{-\theta}$ with $x \in \mathbb{R}$ and $\theta > d$. Then the locally D-optimal design ξ_θ^* in the class of all approximate designs Ξ puts equal weights on the roots $x_1, \ldots, x_{d+1}$ of the ultraspherical polynomial $C_{d+1}^{(-\theta-\frac{1}{2})}(\sqrt{-x^2})$ and (i)*

$$\prod_{i=1}^{d+1} (1 + x_i^2) \;\; = \;\; \prod_{j=1}^{d} \frac{(d - 2\theta - j)^2}{(2d + 1 - 2\theta - 2j)^2},$$

(ii)

$$\mid M(\xi_\theta^*, \theta) \mid \;\; = \;\; \prod_{j=1}^{d} j^j \prod_{j=1}^{d} \frac{(2\theta - 2j + 1)^{2\theta - 2j + 1}}{(2\theta - j + 1)^{2\theta - j + 1}}.$$

Theorem 3.1. *Consider model (2.1) with $\lambda(x, \theta) = (1 + x^2)^{-\theta}$, $x \in \mathbb{R}$ and $\theta > d$. Assume that the condition $\int_\Theta a^{-q\theta} d\tilde{\pi}(\theta) < \infty$ holds for all $a > 1$. Then the $(d+1)$-point Bayesian Φ_q-optimal design with respect to the prior π puts equal weights on the roots of the ultraspherical polynomial*

$$C_{d+1}^{(F(qz)-\frac{1}{2})}(\sqrt{-x^2})$$

where

$$F(qz) = -\frac{\displaystyle\int_\Theta \theta e^{-\theta q z} d\tilde{\pi}(\theta)}{\displaystyle\int_\Theta e^{-\theta q z} d\tilde{\pi}(\theta)},$$

the measure $\tilde{\pi}$ is given by

$$d\tilde{\pi}(\theta) = \Big(\prod_{j=1}^{d} \frac{(2\theta - j + 1)^{2\theta-j+1}}{(2\theta - 2j + 1)^{2\theta-2j+1}}\Big)^q d\pi(\theta),$$

and z is a solution to the equation

$$z = 2\sum_{j=1}^{d} \{\ln(d + 2F(qz) - j) - \ln(2d + 1 + 2F(qz) - 2j)\} \ .$$

The next lemma and theorem relate to the efficiency function $(1+x)^{-\theta}$ and involve support points at the roots of a Jacobi polynomial.

LEMMA 3.2. *Suppose that $\lambda(x,\theta) = (1+x)^{-\theta}$ with $x \in [0,\infty)$ and $\theta > 2d$. Then the locally D-optimal design ξ_θ^* puts equal weights on the roots $x_1, \dots, x_{d+1}$ of the Jacobi polynomial $xP_d^{(1,-\theta-1)}(2x+1)$ and*

(i)

$$\prod_{i=1}^{d+1}(1 + x_i) = \prod_{j=1}^{d} \frac{(\theta - j + 1)}{(\theta - d - j)},$$

(ii)

$$\mid M(\xi_\theta^*, \theta) \mid = \prod_{j=1}^{d} j^{2j} \frac{(\theta - d - j)^{\theta-d-j}}{(\theta - j + 1)^{\theta-j+1}}.$$

THEOREM 3.2. *Consider model (2.1) with $\lambda(x,\theta) = (1+x)^{-\theta}$, $x \in [0,\infty)$ and $\theta > 2d$. Assume that the condition $\displaystyle\int_\Theta a^{-q\theta} d\tilde{\pi}(\theta) < \infty$ holds for all $a > 1$. The $(d+1)$-point Bayesian Φ_q-optimal design with respect to the prior π puts equal weights on the roots of the Jacobi polynomial*

$$xP_d^{(1,F(qz)-1)}(2x+1),$$

where the function $F(\cdot)$ is defined in Theorem 3.1, the measure $\tilde{\pi}$ is given by

$$d\tilde{\pi}(\theta) = \Big(\prod_{j=1}^{d} \frac{(\theta - j + 1)^{\theta-j+1}}{(\theta - d - j)^{\theta-d-j}}\Big)^q d\pi(\theta)$$

and z satisfies the equation

$$z = \sum_{j=1}^{d} \{\ln(-F(qz) - j + 1) - \ln(-F(qz) - d - j)\}.$$

The proofs of the above lemmas and theorems are deferred to the Appendix. Similar proofs can also be readily formulated for a range of efficiency functions including $\lambda(x, \theta) = e^{-\theta x}$ and $e^{-\theta x^2}$ and indeed Biedermann and Dette (2003a) present such a proof for a two-parameter efficiency function.

Observe that for Bayesian D-optimality with $q = 0$, $-F(qz) = E_\pi(\theta)$ and that otherwise the Bayesian Φ_q-optimal design coincides with the locally D-optimal design for a best guess of the parameter $\theta_0 = -F(qz)$. Observe also that it follows immediately from standard arguments that the solutions to the equations involving z in Theorems 3.1 and 3.2 are unique for $q \leq 0$ (Dette and Wong, 1996).

3.2 Approximate designs

Suppose now that the approximate design $\xi^\star$ is Bayesian Φ_q-optimal over the class of all possible design measures. Then the following equivalence theorem holds (Dette and Wong, 1996).

THEOREM 3.3. *A design $\xi^\star$ is Bayesian Φ_q-optimal if and only if the condition*

$$\int_\Theta \mid M(\xi^\star, \theta) \mid^q \lambda(x, \theta) f^T(x) M^{-1}(\xi^\star, \theta) f(x) d\tilde{\pi}(\theta) \leq (d+1) \int_\Theta \mid M(\xi^\star, \theta) \mid^q d\tilde{\pi}(\theta)$$

where $f(x) = \{1, x, \ldots, x^d\}$ holds for all $x \in \mathcal{X}$. Equality is attained at the support points of $\xi^\star$.

It thus follows that if a $(d+1)$-point Bayesian Φ_q-optimal design complies with the conditions of this theorem then it is optimal over all approximate designs. Otherwise Bayesian Φ_q-optimal designs based on more than $d+1$ points are sought and these cannot be derived algebraically, at least in general. In practice, such designs are obtained numerically and the global optimality or otherwise confirmed by invoking Theorem 3.3.

3.3 Example

Consider the weighted quadratic regression model with efficiency function $\lambda(x, \theta) = (1 + x)^{-\theta}$, where $x \in [0, \infty)$ and $\theta > 4$. It then follows from Theorem 3.2 that a three-point Bayesian Φ_q-optimal design puts equal masses at the support points given by the zeros of the polynomial $xP_2^{(1, F(qz)-1)}(2x+1)$ where z satisfies the equation

$$z = \ln \frac{-F(qz)(-F(qz) - 1)}{(-F(qz) - 3)(F(qz) - 4)} \tag{3.1}$$

and specifically at the points

$$0 \text{ and } \frac{3(\theta_0 - 3) \pm \sqrt{3(\theta_0 - 1)(\theta_0 - 3)}}{(\theta_0 - 3)(\theta_0 - 4)}$$

where $\theta_0 = -F(qz)$. Values of z and thus of $F(qz)$ can only be found numerically, even in the most straightforward cases. A selection of three-point Bayesian Φ_q-optimal designs with priors for θ uniformly distributed on intervals of the form $[\theta_{min}, \theta_{max}]$ are presented in Table 1.

Table 1: Three-point Bayesian Φ_q-optimal designs for uniform priors on $[\theta_{min}, \theta_{max}]$

	prior on $[5, 6]$		prior on $[5, 10]$		prior on $[5, 15]$	
q	θ_0	design	θ_0	design	θ_0	design
0	5.5	0, 0.4508, 3.5492	7.5	0, 0.2624, 1.4519	10	0, 0.1727, 0.8273
-1	5.4989	0, 0.4510, 3.5519	7.3848	0, 0.2688, 1.5038	9.4662	0, 0.1863, 0.9114
-10	5.4790	0, 0.4543, 3.6026	7.1109	0, 0.2855, 1.6432	8.8095	0, 0.2062, 1.0413

It is easy to show numerically, by invoking Theorem 3.3, that the three-point Bayesian Φ_q-optimal designs for the uniform prior on $[5, 6]$ with $q = 0, -1$ and -10 and for the uniform prior on $[5, 10]$ with $q = 0$ and -1 are globally optimal over the set of all possible approximate designs but that the remaining designs given in Table 1 are not. In the latter cases the Bayesian Φ_q-optimal designs are based on four points of support. For example for the uniform prior on $[5, 15]$ with $q = -1$ the Bayesian Φ_q-optimal design has support at the points $0, 0.1569, 0.6461$ and 2.0659 with attendant weights $0.3355, 0.2883, 0.2807$ and 0.1055 respectively.

4. Standardized maximin *D*-optimal designs

4.1 $(d + 1)$-point maximin D-optimal designs

Consider first $(d + 1)$-point designs which comprise equally weighted support points. In this case Imhof (2001) and Biedermann and Dette (2003a) have derived standardized maximin D-optimal designs for polynomial models of the form (2.1) with selected efficiency functions but the underlying mathematics is intricate. The results presented here are obtained in a more straightforward and broadly applicable manner by introducing a corollary to Theorem 2.1. The corollary is first stated in some generality and is then applied to the polynomial models of interest in the present study. Its proof is given in the Appendix.

COROLLARY 4.1. *Consider the heteroscedastic regression model*

$$y = f_0(x)\beta_0 + f_1(x)\beta_1 + f_2(x)\beta_2 + \ldots + f_d(x)\beta_d + \epsilon,$$

where ϵ is a random variable with mean zero and variance $\sigma^2/\lambda(x, \theta)$. Suppose that the regression functions $f_0, f_1, \ldots, f_d$ and the efficiency function $\lambda(x, \theta)$ of this model are continuous. Consider a local optimality criterion of the form $\psi(\xi, \theta) = \phi\{M(\xi, \theta)\}$, where $M(\xi, \theta)$ is the *information matrix and ϕ is a continuous non-negative function on the set of non-negative definite $(d+1)\times(d+1)$-matrices. Let Θ be compact and supp$(\pi)=\Theta$. Suppose that $\psi(\xi,\theta)>0$ on $\Delta\times\Theta$ and that one of the following conditions is met.*

(i) The design space $\mathcal{X}$ is compact.

(ii) $\mathcal{X}=[a,\infty)$, $a\in\mathbb{R}$, and

$$\lim_{x\to\infty}\left\{f_j(x)\max_{\theta\in\Theta}\sqrt{\lambda(x,\theta)}\right\}=0,\qquad j=0,1,\ldots,d.$$

(iii) $\mathcal{X}=\mathbb{R}$ and

$$\lim_{x\to-\infty}\left\{f_j(x)\max_{\theta\in\Theta}\sqrt{\lambda(x,\theta)}\right\}=\lim_{x\to\infty}\left\{f_j(x)\max_{\theta\in\Theta}\sqrt{\lambda(x,\theta)}\right\}=0,\qquad j=0,1,\ldots,d.$$

Then the weak limit of Bayesian Ψ_q-optimal designs in the class Δ for $q\to-\infty$ is a standardized maximin optimal design, provided the limit design belongs to Δ.

This corollary can now be invoked to prove the following two theorems. The first theorem, Theorem 4.1, relates to the efficiency function $\lambda(x,\theta)=(1+x^2)^{-\theta}$ and the proof is given in the Appendix.

Theorem 4.1. *The $(d+1)$-point standardized maximin D-optimal design for model (2.1) with $\lambda(x,\theta)=(1+x^2)^{-\theta}$, $x\in\mathbb{R}$, $\theta\in\Theta=[\theta_{\min},\theta_{\max}]$ and $\theta_{\min}>d$ puts equal weights on the roots of the ultraspherical polynomial*

$$C_{d+1}^{\left(-\theta_0-\frac{1}{2}\right)}(\sqrt{-x^2}),$$

where θ_0 falls in the interior of Θ and satisfies the equation

$$2\sum_{j=1}^{d}\{\ln(d-2\theta_0-j)-\ln(2d+1-2\theta_0-2j)\}=\frac{-\ln\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}}$$

with

$$m(\theta)=\prod_{j=1}^{d}\frac{(2\theta-2j+1)^{2\theta-2j+1}}{(2\theta-j+1)^{2\theta-j+1}}.$$

The next theorem relates to the efficiency function $\lambda(x,\theta)=(1+x)^{-\theta}$ and the proof follows in a manner similar to that of Theorem 4.1.

Theorem 4.2. *The $(d+1)$-point standardized maximin D-optimal design for model (2.1) with $\lambda(x,\theta)=(1+x)^{-\theta}$, $x\in[0,\infty)$, $\theta\in\Theta=[\theta_{\min},\theta_{\max}]$ and $\theta_{\min}>2d$ puts equal weights on the roots of the Jacobi polynomial*

$$xP_d^{(1,-\theta_0-1)}(2x+1),$$

where θ_0 falls in the range Θ and satisfies the equation

$$\sum_{j=1}^{d}\{\ln(\theta_0 - j + 1) - \ln(\theta_0 - j - d)\} = \frac{-\ln(m(\theta_{\max})/m(\theta_{\min})))}{\theta_{\max} - \theta_{\min}}$$

with

$$m(\theta) = \prod_{j=1}^{d} \frac{(\theta - d - j)^{\theta - d - j}}{(\theta - j + 1)^{\theta - j + 1}}.$$

Note that similar proofs are available for the results stated in Imhof (2001) for $\lambda(x, \theta) = e^{-\theta x}$ and $\lambda(x, \theta) = e^{-\theta x^2}$. Note also that the equations involving θ_0 in Theorems 4.1 and 4.2 have unique solutions in the range Θ. These equations must be solved numerically however, at least in general.

4.2 Approximate designs

Consider now designs which are standardized maximin D-optimal over all possible design measures. Results for these designs follow immediately from the general design theory given in Pukelsheim (1993) and the related results for designs minimizing the maximum variance of the parameter estimates presented in Dette and Sahm (1998). The following theorem is relevant for heteroscedastic polynomial models of the form (2.1) and, since it is a special case of other more general results (see e.g. Dette, Haines and Imhof, 2003), is stated without proof.

THEOREM 4.3. *The design $\xi^\star$ is standardized maximin D-optimal if and only if there exists a prior distribution $\pi_w(\theta)$ supported on the set of parameter values $\mathcal{N}(\xi^\star)$ such that the condition*

$$\int_{\mathcal{N}(\xi^\star)} \lambda(x, \theta) f(x)^T M^{-1}(\xi^\star, \theta) f(x) d\pi_w(\theta) \leq (d + 1)$$

holds for all $x \in \mathcal{X}$.

It follows immediately from Theorem 4.3 that the maximin D-optimal design $\xi^\star$ coincides with the Bayesian D-optimal design for the prior $\pi_w(\theta)$ defined on the set $\mathcal{N}(\xi^\star)$. The prior $\pi_w(\theta)$ is usually referred to as the least favourable or "worst" prior, a term borrowed from Bayesian decision theory (Berger, 1980, p. 360). It should be emphasized however that Theorem 4.3 is attractive theoretically but difficult to invoke in practice. Specifically it is not easy to construct the prior $\pi_w(\theta)$. Thus, at least in general, standardized maximin D-optimal designs over all approximate designs are difficult to obtain both algebraically and numerically.

In the case of $(d+1)$-point designs for model (2.1) with efficiency functions $\lambda(x, \theta) = (1 + x^2)^{-\theta}$ and $\lambda(x, \theta) = (1 + x)^{-\theta}$ it is possible to identify a candidate worst prior. Specifically it is easy to see that the criterion $R(\xi, \theta)$ for ξ a $(d+1)$-point design is unimodal and hence that, if the $(d+1)$-point standardized maximin D-optimal design is globally optimal, then the cardinality of the set $\mathcal{N}(\xi^\star)$ is 2. This in turn implies that a candidate least favourable prior $\pi_w(\theta)$ can be formulated as putting weights w and $1-w$ on the parameter values θ_{min} and θ_{max} respectively such that

$$w\theta_{min} + (1-w)\theta_{max} = \theta_0$$

with θ_0 specified in Theorem 4.1 or Theorem 4.2. Thus if a $(d+1)$-point standardized maximin D-optimal design is available, then it is possible to check whether or not the associated Bayesian D-optimal design with prior $\pi_w(\theta)$ is optimal over all possible designs by invoking Theorem 3.3 and hence to ascertain whether or not the maximin design is indeed globally optimal.

4.3. Example

Consider again the example presented in Section 3.3. Three-point standardized maximin D-optimal designs for values of θ falling in the range $\Theta = [\theta_{min}, \theta_{max}]$ have support points at the zeros of the polynomial $xP_2^{(1,-\theta_m-1)}(2x+1)$ where θ_m satisfies

$$\prod_{i=1}^{3}(1+x_i) = \frac{\theta_m(\theta_m-1)}{(\theta_m-3)(\theta_m-4)} = \left[\frac{m(\theta_{max})}{m(\theta_{min})}\right]^{-\frac{1}{(\theta_{max}-\theta_{min})}}$$

with

$$m(\theta) = \frac{2^4(\theta-3)^{\theta-3}(\theta-4)^{\theta-4}}{\theta^\theta(\theta-1)^{\theta-1}}.$$

Note that θ_m necessarily falls in the specified range and in fact is given explicitly by

$$\theta_m = \frac{7c-1+\sqrt{1+34c+c^2}}{2(c-1)}.$$

where $c = [m(\theta_{min})/m(\theta_{max})]^{\frac{1}{(\theta_{max}-\theta_{min})}}$.

Table 2: Three-point standardized maximin D-optimal designs with equally weighted support points.

Range Θ	θ_m	design, ξ	candidate worst prior, π_w	$\min_\theta \text{eff}(\xi,\theta)$
$\{5,6\}$	5.4665	0, 0.4563, 3.6350	5 6 .5335 .4665	0.9721
$\{5,10\}$	7.0301	0, 0.2909, 1.6893	5 10 .5940 .4060	0.7569
$\{5,15\}$	8.6996	0, 0.2100, 1.0667	5 15 .6300 .3700	0.5586

Three-point standardized maximin D-optimal designs for the ranges of θ associated with the priors considered earlier for the Bayesian Φ_q-optimal designs are given in Table 2. It follows immediately from observations in the previous subsection that these maximin designs coincide with three-point Bayesian D-optimal designs associated with prior distributions on θ for which $E(\theta) = \theta_m$. It also follows that a standardized three-point maximin D-optimal design is optimal over the set of all possible designs provided the prior which puts weights w and $(1-w)$ on the parameter values θ_{min} and θ_{max} respectively where w satisfies

$$w\theta_{min} + (1-w)\theta_{max} = \theta_m$$

is the least favourable prior or, in other words, provided the three-point Bayesian D-optimal design associated with this prior is itself optimal over the set of all possible designs. The candidate least favourable priors for the designs of Table 2 are included in that table together with the the minimal D-efficiency

$$\min_{\theta \in \Theta} \text{eff}(\xi, \theta) = \min_{\theta \in \Theta} \Big(\frac{\mid M(\xi,\theta) \mid}{\mid M(\xi_\theta^*, \theta) \mid} \Big)^{1/(d+1)} .$$

over the set Θ and the candidate worst (two point) prior. It is easy to show numerically by invoking Theorem 4.3 that only the maximin design for $\Theta = [5, 6]$ is optimal over the set of all possible designs and that in the other two cases maximin designs which are universally optimal are based on four or more points of support. For the interval $\Theta = [5, 10]$ the standardized maximin D-optimal design has masses 0.32, 0.26, 0.27 and 0.15 at the four points 0, 0.21, 0.89, and 4.49, respectively. The minimum D-efficiency on the interval $[5, 10]$ is 0.8402, while the least favourable prior has masses 0.45, 0.40 and 0.15 at the points 5, 7.06 and 10, respectively. Thus we only observe a minor improvement of the best two point design. However, in the case $\Theta = [5, 15]$ the improvement by using designs with more than three support points is more visible. Here the standardized maximin D-optimal design is a 5 point design with masses 0.32, 0.23, 0.28, 0.07 and 0.11 at the points 0, 0.14, 0.54, 1.62 and 3.91, respectively. The minimum D-efficiency is 0.7910 and the least favourable prior has masses 0.36, 0.32 and 0.32 at the points 5, 8.42 and 15 respectively.

5. Further Applications

In order to demonstrate the potential application of the above results to other related problems, consider finding the standardized maximin D-optimal designs for the non-linear growth model

$$y(x) = x^v \exp(-\theta x) \sum_{k=0}^{d-1} \beta_k x^k + \epsilon, \qquad x \in [0, \infty), \tag{5.1}$$

where the error ϵ is assumed to be normally distributed with mean 0 and variance σ^2 and the parameter v is assumed fixed and greater than or equal to 0. Assume that $\theta \in \Theta = [\theta_{\min}, \theta_{\max}]$, $0 < \theta_{\min} < \theta_{\max}$, and that optimization is restricted to the set of $(d+1)$-point designs. Then the determinant of the Fisher matrix of a design ξ is proportional to $\psi(\xi, \theta) = |M(\xi, \theta)|$, where $M(\xi, \theta) = \int_0^\infty f(x) f^T(x) x^{2v} \exp(-2\theta x)\, d\xi(x)$ and $f(x) = (1, x, \ldots, x^d)^T$. Because this is precisely the design problem for a heteroscedastic polynomial regression model with efficiency function $x^{2v}\exp(-2\theta x)$, Corollary 4.1 is applicable here and the $(d+1)$-point standardized maximin D-optimal design can be obtained as a limit of the associated $(d+1)$-point Bayesian Φ_q-optimal designs.

Let $\pi(\theta)$ be any prior distribution on the parameter θ with $\mathrm{supp}(\pi)=\Theta$. Then Dette and Wong (1996) have shown that the $(d+1)$-point Bayesian Φ_q-optimal design puts equal mass at the zeros of the polynomial

$$xL_d^{(1)}\left\{-2F(qz_q)x\right\}, \qquad \text{if } v=0,$$

and

$$L_{d+1}^{(2v-1)}\left\{-2F(qz_q)x\right\}, \qquad \text{if } v>0.$$

Here $L_n^{(\alpha)}(x)$ is the generalized Laguerre polynomial of degree n orthogonal with respect to $x^\alpha\exp(-x)\,dx$, $x\geq 0$, and $z=z_q$ is the unique solution of

$$z=-\frac{(d+1)(d+2v)}{F(qz)},$$

where

$$F(x)=-\frac{\int_\Theta \theta e^{-\theta x}\theta^{(d+1)(d+2v)q}\,d\pi(\theta)}{\int_\Theta e^{-\theta x}\theta^{(d+1)(d+2v)q}\,d\pi(\theta)}.$$

To determine the weak limit of these Bayesian Φ_q-optimal designs, that is, to determine the limit of $F(qz)$ for $q\to-\infty$, the following lemma is invoked. The proof of this lemma is given in the Appendix.

Lemma 5.1. *Let π be a prior distribution with support $\Theta=[\theta_{\min},\theta_{\max}]$, $0<\theta_{\min}<\theta_{\max}$. Let*

$$F_q(x)=-\frac{\int \theta e^{-\theta x}g(\theta)^{-q}\,d\pi(\theta)}{\int e^{-\theta x}g(\theta)^{-q}\,d\pi(\theta)}, \qquad q<0, x\in\mathbb{R},$$

where g is a continuous log-convex function on Θ. For every $q<0$ let z_q be such that

$$z_q=h\{-F_q(qz_q)\}, \tag{5.2}$$

where h is a strictly decreasing function on Θ. If there exists $t^*\in\Theta$ such that

$$h(t^*)=-\frac{\ln\{g(\theta_{\max})/g(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}}, \tag{5.3}$$

then $\lim_{q\to-\infty}F_q(qz_q)=-t^*$.

Consider now taking $g(\theta)=\theta^{-(d+1)(d+2v)}$, which is log convex, and $h(t)=(d+1)(d+2v)/t$, which is decreasing. Then it follows immediately from Lemma 5.1 and from the inequality $\ln x\leq x-1$ for $x>0$ that

$$\lim_{q\to-\infty}F(qz_q)=-t^*=-\frac{\theta_{\max}-\theta_{\min}}{\ln\left(\theta_{\max}/\theta_{\min}\right)},$$

provided $\theta_{\min}\leq t^*\leq\theta_{\max}$. The following theorem therefore holds.

Theorem 5.1. *The $(d+1)$-point standardized maximin D-optimal design for growth model (5.1) with $\theta \in [\theta_{\min}, \theta_{\max}]$ puts equal masses at the zeros of*

$$xL_d^{(1)}\left\{2\frac{\theta_{\max}-\theta_{\min}}{\ln\left(\theta_{\max}/\theta_{\min}\right)}x\right\}, \qquad \text{if } v=0,$$

and

$$L_{d+1}^{(2v-1)}\left\{2\frac{\theta_{\max}-\theta_{\min}}{\ln\left(\theta_{\max}/\theta_{\min}\right)}x\right\}, \qquad \text{if } v>0.$$

Note that this result extends Theorem 5.1 of Imhof (2001), where only the case of $v=0$ is considered.

6. Conclusions

The main aim of the present study has been to construct standardized maximin D-optimal designs for weighted polynomial regression models by invoking the general approach to the construction of maximin designs introduced recently in the paper by Dette, Haines and Imhof (2003). The relevant theory is developed from the fundamental result of that paper and is applicable to heteroscedastic regression models in general. The theory is illustrated for two specific weighted polynomial regression models, the one with an efficiency function $\lambda(x,\theta) = (1+x^2)^{-\theta}$ introduced by Dette, Haines and Imhof (1999) and the other with an efficiency function $\lambda(x,\theta) = (1+x)^{-\theta}$ not studied previously. Bayesian Φ_q-optimal designs for these models are constructed using tools based on the theory of differential equations. This feature is of interest in itself since the method of construction for such designs developed by Dette and Wong (1996) and based on canonical moments does not hold in these cases. Standardized maximin D-optimal designs for the weighted polynomial regression models of interest are then constructed as weak limits of the corresponding Bayesian Φ_q-optimal designs. In addition the maximin D-optimal design for a specific growth model, or equivalently for the weighted polynomial regression model with efficiency function $\lambda(x,\theta) = x^v e^{-\theta x}$, is constructed similarly. In all cases explicit expressions for the Bayesian Φ_q-optimal designs and for the standardized maximin D-optimal designs are obtained. More generally, this study highlights the usefulness and broad applicability of the approach of Dette, Haines and Imhof (2003) to the construction of maximin optimal designs.

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

KwaZulu-Natal and the National Research Foundation of South Africa is gratefully acknowledged. The authors are also grateful to Andrey Pepelyshev for useful comments on an earlier version and computational assistance.

Appendix : Proofs

PROOF OF THEOREM 2.1. It will be shown that if $\{q(j)\}_{j=1}^{\infty}$ is any sequence of negative numbers with $\lim_{j\to\infty} q(j) = -\infty$ such that the sequence $\left\{\zeta_{q(j)}\right\}_{j=1}^{\infty}$ converges weakly to some probability measure ζ^* on $\mathcal{X}$, then ζ^* is a maximin optimal design in Δ. The tightness assumption ensures, by Prohorov's theorem [Billingsley (1999), Theorem 5.1, page 59], that there does exist such a sequence $\{q(j)\}_{j=1}^{\infty}$. Thus the existence of a maximin design will follow. Moreover, the assumption that there is at most one maximin optimal design implies that every convergent sequence $\left\{\zeta_{q(j)}\right\}$ converges to the same limit. Now another application of Prohorov's theorem [more precisely, Billingsley (1999), Corollary, page 59] will show that the entire family $\{\zeta_q\}$ converges to the maximin optimal design.

Thus suppose that $\{q(j)\}_{j=1}^{\infty} \subset (-\infty, 0)$ is such that $q(j) \to -\infty$ and $\zeta_{q(j)} \to \zeta^*$ as $j \to \infty$. Since the closure of $\{\zeta_q : q < 0\}$ is contained in Δ, $\zeta^* \in \Delta$. Now assume that ζ^* is not maximin optimal. Then there exists a better design $\xi' \in \Delta$ and $\epsilon > 0$ such that

$$m_{\xi'} := \inf_{\theta\in\Theta} \frac{\psi(\xi', \theta)}{\psi(\xi^*_\theta, \theta)} > 3\epsilon + \inf_{\theta\in\Theta} \frac{\psi(\zeta^*, \theta)}{\psi(\xi^*_\theta, \theta)}.$$

Clearly,

$$\Psi_q(\zeta_q) \geq \Psi_q(\xi') \geq m_{\xi'} \qquad \text{for all } q < 0. \tag{A.1}$$

The function $\psi(\zeta^*, \theta)/\psi(\xi^*_\theta, \theta)$ is upper semicontinuous in θ. Indeed, if $\theta \in \Theta$ and $\{\theta_k\}_{k=1}^{\infty} \subset \Theta$, $\theta_k \to \theta$, then

$$\limsup_{k\to\infty} \frac{\psi(\zeta^*, \theta_k)}{\psi(\xi^*_{\theta_k}, \theta_k)} \leq \limsup_{k\to\infty} \frac{\psi(\zeta^*, \theta_k)}{\psi(\xi^*_\theta, \theta_k)} = \frac{\psi(\zeta^*, \theta)}{\psi(\xi^*_\theta, \theta)}.$$

It follows that the set

$$U := \left\{\theta \in \Theta : \frac{\psi(\zeta^*, \theta)}{\psi(\xi^*_\theta, \theta)} < m_{\xi'} - 2\epsilon\right\}$$

is a non-empty relatively open subset of Θ. In particular, by assumption c), $\pi(U) > 0$. Since $\zeta_{q(j)} \to \zeta^*$, $\psi(\zeta_{q(j)}, \theta)/\psi(\xi^*_\theta, \theta)$ converges pointwise for every $\theta \in U$ to $\psi(\zeta^*, \theta)/\psi(\xi^*_\theta, \theta)$ as $j \to \infty$. Also, $\pi(U) < \infty$. Hence by Egorov's theorem [Hewitt and Stromberg (1965), Theorem 11.32, page 158], there is a measurable subset $T \subset U$ such that $\pi(T) > \frac{1}{2}\pi(U) > 0$ and $\psi(\zeta_{q(j)}, \theta)/\psi(\xi^*_\theta, \theta)$ converges uniformly on T to $\psi(\zeta^*, \theta)/\psi(\xi^*_\theta, \theta)$. There exists therefore $j_0 \in \mathbb{N}$ such that

$$\frac{\psi(\zeta_{q(j)}, \theta)}{\psi(\xi^*_\theta, \theta)} < m_{\xi'} - \epsilon \qquad \text{for all } \theta \in T,\ j \geq j_0.$$

Hence, for $j \geq j_0$,

$$\Psi_{q(j)}(\zeta_{q(j)}) \leq \left[\int_T \left\{\frac{\psi(\zeta_{q(j)}, \theta)}{\psi(\xi^*_\theta, \theta)}\right\}^{q(j)} d\pi_{q(j)}(\theta)\right]^{\frac{1}{q(j)}} \leq (m_{\xi'} - \epsilon)\left\{\pi_{q(j)}(T)\right\}^{\frac{1}{q(j)}}.$$

In view of hypothesis c), $\liminf_{j\to\infty} \pi_{q(j)}(T) > 0$, and we obtain that

$$\limsup_{j\to\infty} \Psi_{q(j)}(\zeta_{q(j)}) \le m_{\xi'} - \epsilon,$$

which contradicts (A.1). Therefore it follows that the design ζ^* is indeed maximin optimal. $\Box$

PROOF OF LEMMA 3.1. It was shown in Theorem 3.1 of Dette, Haines and Imhof (1999) that for $\theta > d$ the locally D-optimal design has equal masses at the roots $-1 < x_1 < \ldots < x_{d+1} < 1$ of the polynomial

$$C_{d+1}^{(-\theta-1/2)}(\sqrt{-x^2}).$$

For a proof of the representations in (i) we put $\lambda = -\theta - 1/2$ and note that

$$C_{d+1}^{(\lambda)}(\sqrt{-x^2}) = C_{d+1}^{(\lambda)}(ix) = c_{d+1} \prod_{\ell=1}^{d+1} (x - x_\ell) \tag{A.2}$$

where c_{d+1} denotes the leading coefficient of the ultraspherical polynomial, i.e.

$$c_{d+1} = (2i)^{d+1} \binom{d+\lambda}{d+1} \tag{A.3}$$

[see e.g. Szegö (1975), formula (4.7.9)]. Therefore the identity (A.2) implies

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2) = \prod_{\ell=1}^{d+1}(i-x_\ell)(-i-x_\ell) = \frac{C_{d+1}^{(\lambda)}(1)C_{d+1}^{(\lambda)}(-1)}{c_{d+1}^2} = (-1)^{d+1}\frac{\left\{C_{d+1}^{(\lambda)}(1)\right\}^2}{c_{d+1}^2},$$

where the last identity follows from the symmetry of ultraspherical polynomials [see Szegö (1975), formula (4.7.4)]. ¿¿From formula (4.7.3) in the same reference and (A.3) we therefore obtain

$$\begin{aligned}\prod_{\ell=1}^{d+1}(1+x_\ell^2) &= \Big\{\frac{1}{2^{d+1}} \prod_{j=1}^{d+1} \frac{j}{d+1+\lambda-j} \cdot \frac{d+2\lambda+1-j}{j}\Big\}^2 \\ &= \Big\{\prod_{j=1}^{d} \frac{d-2\theta-j}{2d+1-2\theta-2j}\Big\}^2, \end{aligned} \tag{A.4}$$

which proves the assertion (i) of Lemma 3.1.

For a proof of part (ii) we note that

$$|M(\xi_\theta,\theta)| = \Big(\frac{1}{d+1}\Big)^{d+1} \prod_{\ell=1}^{d+1}(1+x_i^2) \prod_{1\le \ell<k\le d+1} (x_\ell - x_k)^2$$

and it is therefore sufficient to calculate the value of the last factor. To this end we put again $\lambda = -\theta - 1/2$ and define

$$P_{d+1}(x) = \frac{C_{d+1}^{(\lambda)}(ix)}{c_{d+1} i^{d+1}} \tag{A.5}$$

as the ultraspherical polynomial with parameter λ, argument ix and leading coefficient 1, then it follows by a straight forward calculation

$$\prod_{1\le \ell<k\le d+1}(x_\ell - x_k)^2 = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1}\prod_{k\neq \ell}(x_\ell - x_k) = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1} P'_{d+1}(x_\ell). \tag{A.6}$$

¿¿From formula (4.7.27) in Szegö (1975) we have for any $\ell \in \{1,\dots,d+1\}$

$$\begin{aligned}(1+x^2)\frac{d}{dx}P_{d+1}(x)\Big|_{x=x_\ell} &= \frac{1+x^2}{i^{d+1}c_{d+1}}\frac{d}{dx}C^{(\lambda)}_{d+1}(ix)\Big|_{x=x_\ell} \\ &= \frac{i(d+2\lambda)}{i^d c_{d+1}}C^{(\lambda)}_d(ix_\ell) = \frac{(d+2\lambda)}{c_{d+1}}c_d P_d(x_\ell) = \frac{d+1}{2}\frac{d+2\lambda}{d+\lambda}P_d(x_\ell),\end{aligned}$$

where the notation (A.3) is used in the last equality. Observing that the recursive relation for the polynomial $P_j(x)$ is given by

$$P_{k+1}(x) = xP_k(x) + \frac{(k-1+2\lambda)k}{4(k+\lambda-1)(k+\lambda)}P_{k-1}(x)$$

$(P_{-1}(x)=0, P_0(x)=1)$ it now follows from formula (6.71.2) in Szegö (1975) that $[a_n = 1, c_n = -(n-1)(n-2+2\lambda)/4(n-2+\lambda)(n-1+\lambda)]$

$$\begin{aligned}&\prod_{\ell=1}^{d+1}(1+x_\ell^2)\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2 = (-1)^{d(d+1)/2}\prod_{\ell=1}^{d+1}\Big(\frac{d+1}{2}\Big)\Big(\frac{d+2\lambda}{d+\lambda}\Big)P_d(x_\ell) \\ = \;& (-1)^{d(d+1)/2}\Big(\frac{d+1}{2}\Big)^{d+1}\Big(\frac{d+2\lambda}{d+\lambda}\Big)^{d+1}\prod_{j=1}^{d+1}\Big\{\frac{(j-1)(j-2+2\lambda)}{4(j-2+\lambda)(j-1+\lambda)}\Big\}^{j-1} \\ = \;& \prod_{j=1}^{d+1} j^j \cdot \prod_{j=1}^{d}\frac{(2\theta-j+1)^{j+1}}{(2\theta-2j+1)^{2j+1}}.\end{aligned}$$

Combining this identity with (A.4) yields

$$\prod_{\ell=1}^{d+1}(1+x_\ell^2)^{-\theta}\prod_{1\le\ell<k\le d+1}(x_\ell-x_k)^2 = \prod_{j=1}^{d+1} j^j \cdot \prod_{j=1}^{d}\frac{(2\theta-2j+1)^{2\theta-2j+1}}{(2\theta-j+1)^{2\theta-j+1}}$$

and the assertion (ii) of Lemma 3.1 follows. □

Proof of Theorem 3.1. The determinant of the information matrix for the parameters β of model (2.1) from a $(d+1)$-point design ξ which puts equal masses at the support points $x_1,\dots,x_{d+1}$ can be written as

$$\mid M(\theta,\xi)\mid = \mid X_R \mid^2 \prod_{i=1}^{d+1}(1+x_i^2)^{-\theta}$$

where X_R is the Vandermonde matrix with ith row $\{1\ x_i\ x_i^2\ \ldots\ x_i^d\}$. Thus

$$(d+1)^{d+1}\cdot\Phi_q(\xi) = \mid X_R\mid^2 \left\{\int_\Theta \prod_{i=1}^{d+1}(1+x_i^2)^{-q\theta}d\tilde{\pi}(\theta)\right\}^{1/q}$$

and differentiating $\ln\Phi_q(\xi)$ with respect to x_j and setting the result to 0 in turn gives

$$\frac{2}{\mid X_R\mid}\frac{\partial \mid X_R\mid}{\partial x_j} - \frac{2x_j}{1+x_j^2}\frac{\displaystyle\int_\Theta \theta\prod_{i=1}^{d+1}(1+x_i^2)^{-q\theta}d\tilde{\pi}(\theta)}{\displaystyle\int_\Theta \prod_{i=1}^{d+1}(1+x_i^2)^{-q\theta}d\tilde{\pi}(\theta)} = 0.$$

It then follows by arguments similar to those used in Dette, Haines and Imhof (1999) that the required points are the roots of the polynomial $\prod_{j=1}^{d+1}(x-x_j)$, which satisfies the differential equation

$$(1+x^2)f^{''}(x)+2xF(qz)f^{'}(x)-(d+1)(d+2F(qz))f(x)=0,$$

where $f(x)$ is a polynomial of degree $d+1$ in x, $z=\sum_{i=1}^{d+1}\ln(1+x_i^2)$ and

$$F(qz) = -\frac{\displaystyle\int_\Theta \theta e^{-\theta qz}d\tilde{\pi}(\theta)}{\displaystyle\int_\Theta e^{-\theta qz}d\tilde{\pi}(\theta)}$$

with $-F(qz)>d$. The support points of the Bayesian Φ_q-optimal design are thus the roots of the ultraspherical polynomial

$$C_{d+1}^{(F(qz)-\frac{1}{2})}(\sqrt{-x^2})$$

and $z=\sum_{i=1}^{d+1}\ln(1+x_i^2)$ is obtained by invoking expression for $\prod_{i=1}^{d+1}(1+x_i^2)$ given in Lemma 3.1. Note that for $q=0$ the criterion corresponds to Bayesian D-optimality and that for a one point prior the locally D-optimal design is recovered. □

Proof of Lemma 3.2. It follows by similar arguments as given in Dette, Haines and Imhof (1999) that the locally D-optimal design is supported at $d+1$ points including the point 0, say $0=x_1<x_2<\ldots<x_{d+1}$, and that the supporting polynomial $f(x)=\Pi_{i=1}^{d+1}(x-x_i)$ is a solution of the differential equation

$$x(1+x)y''(x)-\theta xy'(x)+(d+1)(\theta-d)y(x)=0.$$

The polynomial solution of this equation is given by the hypergeometric series

$$xF(-d,d+1-\theta,2,-x)$$

which is proportional to the Jacobi polynomial

$$f(x)=xP_d^{(1,-\theta-1)}(2x+1)$$

[see formula (4.21.2) in Szegö (1975)]. The assertions (i) and (ii) of Lemma 3.2 now follow by similar arguments as given in the proof of Lemma 3.1, which are omitted for the sake of brevity. □

Proof of Theorem 3.2. This follows along the lines stated in the proof of Theorem 3.1. □

Proof of Corollary 4.1. Under any of the three conditions, $f_i(x)f_j(x)\lambda(x,\theta)$ is bounded and uniformly continuous on $\mathcal{X} \times \Theta$ for $i, j = 0, \ldots, d$. Thus for every fixed $\theta \in \Theta$, each entry of the information matrix $M(\xi, \theta)$ is continuous in ξ. Moreover, by Lebesgue's convergence theorem, for every fixed design $\xi \in \Delta$, each entry of $M(\xi,\theta)$ is continuous in θ. Therefore, the local criterion $\psi(\xi,\theta) = \phi\{M(\xi,\theta)\}$ is continuous in both arguments. The assertion now follows from Theorem 2.1 with $\pi_q = \pi$ for all $q < 0$. Note that conditions a) and b) of Theorem 2.1 do not have to be verified, since the existence of the limit design is an assumption of Corollary 4.1. Condition c) is trivially satisfied here. □

Proof of Theorem 4.1. For $j = 0, \ldots, d$,

$$\lim_{x\to\pm\infty} x^j \max_{\theta\in\Theta} \sqrt{\lambda(x,\theta)} = \lim_{x\to\pm\infty} \frac{x^j}{(1+x^2)^{\frac{1}{2}\theta_{\min}}} = 0.$$

Thus condition (iii) of Corollary 4.1 is satisfied and the maximin design can therefore be obtained as the limit of Φ_q-optimal designs. Suppose d is even; the case where d is odd is similar. Let π be any prior distribution with support Θ and write

$$F_q(x) = -\frac{\int \theta e^{-\theta x} m(\theta)^{-q}\, d\pi(\theta)}{\int e^{-\theta x} m(\theta)^{-q}\, d\pi(\theta)}, \qquad q < 0,\ x \in \mathbb{R},$$

$$h(\theta) = 2\sum_{j=1}^{\frac{d}{2}} \left\{\ln(2\theta - 2j + 2) - \ln(2\theta - d - 2j + 1)\right\}, \qquad \theta \in \Theta.$$

If z_q denotes the solution of the equation $z_q = h(-F_q(qz_q))$, then, by Theorem 3.1, the Bayesian Φ_q-optimal $(d+1)$-point design puts equal weights on the roots of the polynomial

$$C_{d+1}^{\left(F_q(qz_q)-\frac{1}{2}\right)}(\sqrt{-x^2}).$$

It remains to show that $\lim_{q\to-\infty} F_q(qz_q) = -\theta_0$. To see that there indeed exists $\theta_0 \in \mathrm{int}\Theta$ as defined in the theorem, set

$$H(x,\theta) = \sum_{j=1}^{\frac{d}{2}} \ln \frac{(2x - d - 2j + 1)^{2\theta - d - 2j + 1}}{(2x - 2j + 2)^{2\theta - 2j + 2}}, \qquad x, \theta \in \Theta.$$

Then

$$\frac{\partial H(x,\theta)}{\partial x} = 2(d+1)\sum_{j=1}^{\frac{d}{2}} \frac{\theta - x}{(2x-d-2j+1)(x-j+1)},$$

so that $H(\cdot,\theta_{\max})$ is strictly increasing and $H(\cdot,\theta_{\min})$ is strictly decreasing on Θ. Hence, in view of Lemma 3.1 (ii),

$$\begin{aligned}\frac{-\ln\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}} &= \frac{H(\theta_{\min},\theta_{\min}) - H(\theta_{\max},\theta_{\max})}{\theta_{\max}-\theta_{\min}} \\ &< \frac{H(\theta_{\min},\theta_{\min}) - H(\theta_{\min},\theta_{\max})}{\theta_{\max}-\theta_{\min}} = h(\theta_{\min})\end{aligned}$$

and, similarly,

$$\frac{-\ln\{m(\theta_{\max})/m(\theta_{\min})\}}{\theta_{\max}-\theta_{\min}} > h(\theta_{\max}).$$

This ensures the existence of θ_0. It is easily verified that $h(t)$ is strictly decreasing and that $\ln m(\theta)$ is convex. It now follows by Lemma 5.1 that $\lim_{q\to-\infty} F_q(qz_q) = -\theta_0$, which completes the proof of Theorem 4.1. $\square$

Proof of Lemma 5.1. Set

$$G_q(t) = t + F_q\{qh(t)\}, t \in \Theta$$

and $t_q = -F_q(qz_q)$. It follows from (5.2) that $G_q(t_q) = 0$. As $F_q'(x) \geq 0$ for all x, G_q is strictly increasing. Thus t_q is the only zero of the function G_q. It has to be shown that $\lim_{q\to-\infty} t_q = t^*$.

Assume first that $t^* < \theta_{\max}$. Let $\varepsilon > 0$ be such that $t^* + \varepsilon \leq \theta_{\max}$. Setting

$$\phi(\theta) = \exp\{\theta h(t^*+\varepsilon)\} g(\theta),$$

one has

$$F_q\{qh(t^*+\varepsilon)\} = -\frac{\int_\Theta \theta\phi(\theta)^{-q}\, d\pi(\theta)}{\int_\Theta \phi(\theta)^{-q}\, d\pi(\theta)}. \tag{A.7}$$

As $\theta h(t^*+\varepsilon) + \ln g(\theta)$ is convex, so is $\phi(\theta)$. Since h is strictly decreasing, we obtain by (5.3),

$$\frac{\phi(\theta_{\max})}{\phi(\theta_{\min})} = \exp\{(\theta_{\max}-\theta_{\min})h(t^*+\varepsilon)\}\frac{g(\theta_{\max})}{g(\theta_{\min})} < \exp\{(\theta_{\max}-\theta_{\min})h(t^*)\}\frac{g(\theta_{\max})}{g(\theta_{\min})} = 1.$$

Thus $\phi(\theta_{\max}) < \phi(\theta_{\min})$, and so $\phi(\theta) < \phi(\theta_{\min})$ for all $\theta > \theta_{\min}$. Consequently, we have for every $\theta_0 \in \mathrm{int}\Theta$,

$$\lim_{q\to-\infty} \frac{\left\{\int_{[\theta_0,\theta_{\max}]} \phi(\theta)^{-q}\, d\pi(\theta)\right\}^{-\frac{1}{q}}}{\left\{\int_\Theta \phi(\theta)^{-q}\, d\pi(\theta)\right\}^{-\frac{1}{q}}} = \frac{\max_{\theta\in[\theta_0,\theta_{\max}]}\phi(\theta)}{\max_{\theta\in\Theta}\phi(\theta)} < 1,$$

and so

$$\lim_{q\to-\infty}\frac{\int_{[\theta_0,\theta_{\max}]}\phi(\theta)^{-q}\,d\pi(\theta)}{\int_\Theta\phi(\theta)^{-q}\,d\pi(\theta)}=0,\qquad \lim_{q\to-\infty}\frac{\int_{[\theta_{\min},\theta_0)}\phi(\theta)^{-q}\,d\pi(\theta)}{\int_\Theta\phi(\theta)^{-q}\,d\pi(\theta)}=1.$$

In view of (A.7) this yields

$$\begin{aligned}G_q(t^*+\varepsilon)&=t^*+\varepsilon-\frac{\int_\Theta\theta\phi(\theta)^{-q}\,d\pi(\theta)}{\int_\Theta\phi(\theta)^{-q}\,d\pi(\theta)}\\&\geq t^*+\varepsilon-\frac{\theta_0\int_{[\theta_{\min},\theta_0)}\phi(\theta)^{-q}\,d\pi(\theta)+\theta_{\max}\int_{[\theta_0,\theta_{\max}]}\phi(\theta)^{-q}\,d\pi(\theta)}{\int_\Theta\phi(\theta)^{-q}\,d\pi(\theta)}.\end{aligned}$$

It follows that

$$\liminf_{q\to-\infty}G_q(t^*+\varepsilon)\geq t^*+\varepsilon-\theta_0$$

for all $\theta_0\in\mathrm{int}\Theta$. Thus

$$\liminf_{q\to-\infty}G_q(t^*+\varepsilon)\geq\varepsilon,$$

so that $G_q(t^*+\varepsilon)>0$ for $q\leq q_0=q_0(\varepsilon)$, say. Since G_q is increasing, this implies that $t_q<t^*+\varepsilon$ for $q\leq q_0$. As $\varepsilon>0$ was arbitrarily small, $\limsup_{q\to-\infty}t_q\leq t^*$, which is trivially true if $t^*=\theta_{\max}$. A similar argument shows that $\liminf_{q\to-\infty}t_q\geq t^*$, completing the proof of Lemma 5.1. $\square$

References

Berger J.O. (1980). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

Billingsley P. (1999). *Convergence of Probability Measures.* Wiley, New York.

Biedermann S. and Dette H. (2003a). A note on maximin and Bayesian D-optimal designs in weighted polynomial regression. *Mathematical Methods of Statistics*, in press. http://ruhr-uni-bochum.de/mathematik3/preprint.htm

Biedermann S. and Dette H. (2003b). Numerical construction of maximin optimal designs for binary response models. Submitted. http://ruhr-uni-bochum.de/mathematik3/preprint.htm

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.

Dette H., Haines L.M. and Imhof L. (1999). Optimal designs for rational models and weighted polynomial regression. *Annals of Statistics*, **27**, 1272-1293.

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

Dette H. and Sahm M. (1998). Minimax optimal designs in nonlinear regression models. *Statistica Sinica*, **8**, 1249–1264.

Dette H. and Wong W.K. (1996). Optimal Bayesian designs for models with partially specified heteroscedastic structure. *Annals of Statistics*, **24**, 2108–2127.

Fandom Noubiap R. and Seidel W. (2000). A minimax algorithm for constructing optimal symmetric balanced designs. *J. Statist. Plann. Inf.*, **91**, 151–168.

Haines L.M. (1995). A geometric approach to optimal designs for one-parameter non-linear models. *Journal of the Royal Statistical Society*, Ser. B, **57**, 575–598.

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

Imhof L. (2001). Maximin designs for exponential growth models and heteroscedastic polynomial models. *Annals of Statistics*, **29**,561-576.

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

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

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

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

Sitter R.R. (1992). Robust designs for binary data. *Biometrics*, **48**, 1145–1155.

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

Wiens D.P. (1998). Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. *J. Amer. Statist. Assoc.*, **93**, 1440-1450.

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