

MAXIMIN POWER DESIGNS IN TESTING LACK OF FIT

Douglas P. Wiens¹

June 25, 2018

Abstract In a previous article (Wiens, 1991) we established a maximin property, with respect to the power of the test for Lack of Fit, of the absolutely continuous uniform ‘design’ on a design space which is a subset of \mathbb{R}^q with positive Lebesgue measure. Here we discuss some issues and controversies surrounding this result. We find designs which maximize the minimum power, over a broad class of alternatives, in discrete design spaces of cardinality N . We show that these designs are supported on the entire design space. They are in general not uniform for fixed N , but are asymptotically uniform as $N \rightarrow \infty$. Several examples with N fixed are discussed; in these we find that the approach to uniformity is very quick.

AMS 2010 Subject Classifications: Primary 62K05; Secondary 62F35, 62J05.

Key words and phrases F-test, finite design space, maximin, minimax, non-centrality, power.

1 Introduction

In Wiens (1991), henceforth referred to as [W], we studied the uniform ‘design’, as applied to design spaces S_c that are subsets of \mathbb{R}^q – intervals, hypercubes, etc. – with positive Lebesgue measure. We call such design spaces *continuous*, to distinguish them from the finite, discrete design spaces considered in this article.

The uniform design on S_c is the absolutely continuous measure, with constant density $1/\int_{S_c} d\mathbf{x}$. Of course such a design must be approximated in order to implement it in an actual experiment. A contribution of [W] was that, in a sense made precise there and detailed in §2 below, the uniform design possesses an optimality property in the class of all designs on S_c – it maximizes the minimum power of the standard F-test for Lack of Fit (LOF) of a fitted linear regression model, with the minimum taken over a broad class of alternatives.

The theory in [W] has been adapted to justify the use of discrete uniform designs in numerous applications in the sciences. For its application to drug combination studies see the series of papers Tan, Fang, Tian and Houghton (2001), Fang, Ross, Sausville and Tan (2008), Tan, Fang and Tian (2009), Fang, Tian, Li and Tan (2009) and Fang, Chen, Pei, Grant and Tan (2015). The ideas in [W] have gained traction in the theory of artificial neural networks – see Zhang, Liang, Jiang, Yu and Fang (1998) – and reduced support vector machines – see Lee and Huang (2007).

¹Department of Mathematical and Statistical Sciences; University of Alberta, Edmonton, Alberta; Canada T6G 2G1. e-mail: doug.wiens@ualberta.ca

The theory has been extended to nonparametric regression models – Xie and Fang (2000) – and, also allowing for heteroscedasticity, by Biedermann and Dette (2001) and Bischoff and Miller (2006).

The continuous nature of S_c in this context has been controversial. Indeed Bischoff (2010) argues that it allows for classes of alternative regression models – as used both in [W] and in Biedermann and Dette (2001) – that are too broad for the optimality property to be asymptotically meaningful (when the continuous uniform design is viewed as the limit of discrete uniform designs); he proposes a restricted interpretation.

That the richness of classes of alternatives as in [W] makes discrete designs inadmissible was noted in Wiens (1992, p. 355), where we state ‘*Our attitude is that an approximation to a design which is robust against more realistic alternatives is preferable to an exact solution in a neighbourhood which is unrealistically sparse.*’ This remains our view. Nonetheless, in this article we suggest an alternate approach that we feel is less controversial. We take a *finite* design space $S = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ – here N can be arbitrarily large, allowing for at least a close approximation of the space of interest in an anticipated application. We obtain exact designs for small values of N – these are non-uniform – and show that the maximin designs are asymptotically uniform, as $N \rightarrow \infty$. Theory and examples show that this limit is approached very quickly.

In the next section we outline the mathematical framework, provide a reduction of the maximin problem to a simpler minimax problem, and prove the asymptotic optimality of the uniform design. Some solutions with N fixed are given in §3. Proofs are in the Appendix. The computing code is available from the author’s personal web site.

2 Preliminaries

As far as possible we use notation as in [W], to which we refer the reader for background material relating the standard F-test for lack of fit to properties of the design. We denote by λ the uniform probability measure on S , viz. $\lambda(\mathbf{x}_i) = 1/N$, $i = 1, \dots, N$. To facilitate comparisons with [W] we now write $\lambda_c(\mathbf{x})$ for the continuous uniform design on a continuous design space S_c .

For a design ξ on S we write $\xi_i = \xi(\mathbf{x}_i)$. An implementable design with n observations requires that $n\xi_i$ be an integer; we shall loosen this restriction and allow ξ to be any probability distribution on S . In particular, we include λ as a possible design.

For p -dimensional regressors $\mathbf{z}(\mathbf{x})$ we entertain a class of departures

$$E[Y(\mathbf{x})] = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta} + f(\mathbf{x}) \quad (f \in \mathcal{F}_\eta^+), \quad (1)$$

(the ‘full’ models, in LOF terminology) from the fitted (‘reduced’) regression model

$$E[Y(\mathbf{x})] = \mathbf{z}'(\mathbf{x})\boldsymbol{\theta}. \quad (2)$$

In [W] we defined this class of functions on S_c by (i) $\int_{S_c} f^2(\mathbf{x}) d\lambda_c(\mathbf{x}) \geq \eta^2$ and (ii) $\int_{S_c} \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\lambda_c(\mathbf{x}) = \mathbf{0}_{p \times 1}$. We now adopt an analogous definition of \mathcal{F}_η^+ on the discrete design space S , viz.,

$$\int_S f^2(\mathbf{x}) d\lambda(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N f^2(\mathbf{x}_i) \geq \eta^2, \quad (3a)$$

$$\int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\lambda(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) = \mathbf{0}_{p \times 1}. \quad (3b)$$

Condition (3a) enforces a separation between the fitted and alternate models, so that the test has positive power, and (3b) ensures the identifiability of the regression parameters under (1), via

$$\boldsymbol{\theta} \stackrel{def}{=} \arg \min_{\mathbf{t}} \sum_{i=1}^N (E[Y(\mathbf{x}_i)] - \mathbf{z}'(\mathbf{x}_i) \mathbf{t})^2.$$

This defines $\boldsymbol{\theta}$ uniquely, in the presence of (3b) and the requirement, made here, that the matrix $\mathbf{Z}_{N \times p} = [\mathbf{z}(\mathbf{x}_1) : \dots : \mathbf{z}(\mathbf{x}_N)]'$ be of full column rank. We write $f_i = f(\mathbf{x}_i)$ and define $\mathbf{f} = (f_1, \dots, f_N)'$ and $\mathbf{D}_\xi = \text{diag}(\xi_1, \dots, \xi_N)$. Define as well

$$\begin{aligned} \mathbf{b}_{f,\xi} &= \int_S \mathbf{z}(\mathbf{x}) f(\mathbf{x}) d\xi(\mathbf{x}) = \sum_{i=1}^N \mathbf{z}(\mathbf{x}_i) f(\mathbf{x}_i) \xi_i = \mathbf{Z}' \mathbf{D}_\xi \mathbf{f}, \\ \mathbf{B}_\xi &= \int_S \mathbf{z}(\mathbf{x}) \mathbf{z}'(\mathbf{x}) d\xi(\mathbf{x}) = \sum_{i=1}^N \mathbf{z}(\mathbf{x}_i) \mathbf{z}'(\mathbf{x}_i) \xi_i = \mathbf{Z}' \mathbf{D}_\xi \mathbf{Z}, \end{aligned}$$

and assume that \mathbf{B}_ξ is non-singular. Then as at (2.2) of [W] the non-centrality parameter (NCP) of the F-statistic for testing the LOF of the fitted model (2), with alternatives of the form (1), and using a design ξ , is proportional to

$$\mathcal{B}(f, \xi) = \mathbf{f}' \mathbf{D}_\xi \mathbf{f} - \mathbf{b}_{f,\xi}' \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi}.$$

The power of the test is an increasing function of the NCP, as long as the F-statistic is stochastically increasing in this parameter. This monotonicity is well known to hold in finite samples under a Gaussian error distribution, and is at least asymptotically valid otherwise, under mild conditions.

In its alternate form

$$\mathcal{B}(f, \xi) = \sum_{i=1}^N (f(\mathbf{x}_i) - \mathbf{z}'(\mathbf{x}_i) \mathbf{B}_\xi^{-1} \mathbf{b}_{f,\xi})^2 \xi_i,$$

we see that $\mathcal{B}(f, \xi)$ is the $\mathcal{L}_2(\xi)$ distance from f to the nearest function of the form (2). Thus $\min_f \mathcal{B}(f, \xi)$ is a natural measure of the discrepancy between the ‘full’

and ‘reduced’ models being compared, with larger values leading to a more effective test.

It is reasonable to think that the following conjecture should hold. If true, it extends the main result of [W] to discrete design spaces.

Conjecture *For any design ξ on S we have that $\min_{\mathcal{F}_\eta^+} \mathcal{B}(f, \xi) \leq \eta^2 = \min_{\mathcal{F}_\eta^+} \mathcal{B}(f, \lambda)$, so that the uniform design λ maximizes the minimum power of the F -test of LOF.*

Despite its plausibility this conjecture turns out to be false – a first counterexample is furnished in Example 1 below.

To express the conjecture more explicitly, we first characterize the classes \mathcal{F}_η^+ parametrically. Write the qr-decomposition of \mathbf{Z} as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix},$$

where the columns of $\mathbf{Q}_1 : N \times p$ form an orthogonal basis for $\text{col}(\mathbf{Z})$, the column space of \mathbf{Z} , and the columns of $\mathbf{Q}_2 : N \times (N - p)$ form an orthogonal basis for the orthogonal complement $\text{col}(\mathbf{Z})^\perp = \text{col}(\mathbf{Q}_1)^\perp$. Then $\mathbf{Q} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix}$ is an orthogonal matrix. Condition (3b) requires \mathbf{f} to lie in $\text{col}(\mathbf{Q}_2)$, i.e. $\mathbf{f} = \eta\sqrt{N}\mathbf{Q}_2\mathbf{d}$ for some $\mathbf{d}_{N-p \times 1}$, and then (3a) requires $\|\mathbf{d}\| \geq 1$ for $f \in \mathcal{F}_\eta^+$.

With

$$\begin{aligned} \mathbf{P} &\stackrel{\text{def}}{=} \sqrt{N}\mathbf{D}_\xi^{1/2}\mathbf{Q} = \left[\sqrt{N}\mathbf{D}_\xi^{1/2}\mathbf{Q}_1 : \sqrt{N}\mathbf{D}_\xi^{1/2}\mathbf{Q}_2 \right] \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{P}_1 & \mathbf{P}_2 \end{bmatrix}, \\ \mathbf{H} &\stackrel{\text{def}}{=} \mathbf{P}_1(\mathbf{P}_1'\mathbf{P}_1)^{-1}\mathbf{P}_1' = \mathbf{D}_\xi^{1/2}\mathbf{Q}_1(\mathbf{Q}_1'\mathbf{D}_\xi\mathbf{Q}_1)^{-1}\mathbf{Q}_1'\mathbf{D}_\xi^{1/2}, \end{aligned}$$

we find that

$$\mathcal{B}(f, \xi) = \eta^2 \mathbf{d}'\mathbf{P}_2'(\mathbf{I}_N - \mathbf{H})\mathbf{P}_2\mathbf{d}.$$

We denote by Ξ the set of all designs on S , and by ch_{\min} and ch_{\max} the minimum and maximum eigenvalues of a matrix. The conjecture then asks that we solve

$$\max_{\xi \in \Xi} \min_{\|\mathbf{d}\| \geq 1} \mathcal{B}(f, \xi) = \eta^2 \max_{\xi \in \Xi} ch_{\min}[\mathbf{P}_2'(\mathbf{I}_N - \mathbf{H})\mathbf{P}_2],$$

and show that the solution is $\xi = \lambda$.

The problem is greatly simplified by the following theorem.

Theorem 1 *The following are equivalent:*

(a) *A design ξ_0 is maximin with respect to the power of the test of LOF, in that*

$$\xi_0 = \arg \max_{\xi \in \Xi} ch_{\min}[\mathbf{P}_2'(\mathbf{I}_N - \mathbf{H})\mathbf{P}_2].$$

(b) *A design ξ_0 places positive mass at each point of S and is a minimax design within the set Ξ_+ of all such designs, in that*

$$\xi_0 = \arg \min_{\xi \in \Xi_+} ch_{\max}[\mathbf{Q}_2'\mathbf{D}_\xi^{-1}\mathbf{Q}_2].$$

If $\xi \in \Xi_+$ then $ch_{\min}[\mathbf{P}_2'(\mathbf{I}_N - \mathbf{H})\mathbf{P}_2] = N/ch_{\max}[\mathbf{Q}_2'\mathbf{D}_\xi^{-1}\mathbf{Q}_2]$.

To obtain the required maximin/minimax design we are to minimize

$$\mathcal{L}(\xi) \stackrel{\text{def}}{=} ch_{\max} [\mathbf{Q}_2' \mathbf{D}_\xi^{-1} \mathbf{Q}_2],$$

over the set Ξ_+ of designs that place positive mass on each point of S . Note that $\mathcal{L}(\lambda) = N$; the following theorem shows that we cannot expect much improvement on this, for large N .

Theorem 2 *For any design $\xi \in \Xi_+$, $\mathcal{L}(\xi) \geq N - p$, so that*

$$1 - \frac{p}{N} \leq \frac{\min_{\xi \in \Xi_+} \mathcal{L}(\xi)}{\mathcal{L}(\lambda)} \leq 1.$$

In a sense made precise by Theorem 2 the uniform design is asymptotically optimal, as $N \rightarrow \infty$. The following example shows that this optimality does not in general hold for finite N .

Example 1. Suppose $N = 2$, $p = 1$, $\mathbf{Q}_2 = (\alpha, \beta)'$ with $\alpha^2 + \beta^2 = 1$ and $\alpha, \beta > 0$. Then there is only one eigenvalue, given by

$$\mathbf{Q}_2' \mathbf{D}_\xi^{-1} \mathbf{Q}_2 = \frac{\alpha^2 + (\beta^2 - \alpha^2) \xi_1}{\xi_1 (1 - \xi_1)}.$$

This is minimized by $\xi_{0,1} = \alpha / (\alpha + \beta) = 1 - \xi_{0,2}$, with $\mathcal{L}(\xi_0) = (\alpha + \beta)^2 = 1 + 2\alpha\beta$. This improves on λ , for which $\mathcal{L}(\lambda) = 2 \geq 1 + 2\alpha\beta$ for all α, β . There is strict inequality unless $\alpha = \beta = 1/\sqrt{2}$, in which case $\xi_0 = \lambda$.

3 Minimax designs for fixed N

The set Ξ_+ is not closed, and this poses technical difficulties which will become evident. Thus we shall first minimize instead over the closed, convex set Ξ_ε of designs that place mass of at least $\varepsilon > 0$ on each point of S . In most cases it turns out that the minimax design lies in the interior of this set, so that the restriction to Ξ_ε is moot and the solution holds for all of Ξ_+ .

When the maximum eigenvalue of $\mathbf{Q}_2' \mathbf{D}_\xi^{-1} \mathbf{Q}_2$ is simple, the minimax design has a parametric form.

Theorem 3 *Denote by $\mathbf{q}_1', \dots, \mathbf{q}_N'$ the rows of \mathbf{Q}_2 . For a positive constant μ and a vector $\mathbf{x}_{N-p \times 1}$ define a design by*

$$\xi_{0,i}(\mu, \mathbf{x}) = \max \left(\frac{|\mathbf{q}_i' \mathbf{x}|}{\sqrt{\mu}}, \varepsilon \right). \quad (4)$$

(i) If μ and \mathbf{x} satisfy

$$\sum_{i=1}^N \xi_{0,i}(\mu, \mathbf{x}) = 1, \quad (5)$$

$$\mathbf{x} = \text{eigenvector belonging to } ch_{\max} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_0(\mu, \mathbf{x})}^{-1} \mathbf{Q}_2 \right], \quad (6)$$

and if this maximum eigenvalue is simple, then ξ_0 minimizes $ch_{\max} [\mathbf{Q}'_2 \mathbf{D}_{\xi}^{-1} \mathbf{Q}_2]$ in Ξ_{ε} .

(ii) If $\mathbf{q}'_i \mathbf{x} \neq 0$ for all i , then we may take $\varepsilon = 0$ in (4) and the solution in Ξ_+ is $\xi_{0,i} = |\mathbf{q}'_i \mathbf{x}| / \sqrt{\mu}$, with $\sqrt{\mu} = \sum_{i=1}^N |\mathbf{q}'_i \mathbf{x}|$ and $\mathcal{L}(\xi_0) = \mu$.

Example 1 continued. In this example Theorem 3(ii) applies. With $q_1 = \alpha$, $q_2 = \beta$, we have $x = 1$, $\mu = (|\alpha| + |\beta|)^2$, and $\xi_{0,1} = |\alpha| / (|\alpha| + |\beta|) = 1 - \xi_{0,2}$, extending the solution obtained earlier, when we took $\alpha, \beta > 0$.

Example 2. Here we consider cubic regression ($p = 4$) with $N = 7$ and (randomly generated) design space

$$S = \{-0.4240, 0.0522, 0.2069, 0.3358, 0.4145, 0.4594, 0.5628\}.$$

Then the regressors are $\mathbf{z}(x) = (1, x, x^2, x^3)'$. We implement Theorem 3 via a constrained nonlinear minimizer in MATLAB. The minimax design is found to be

$$\xi_0 = \{0.0086, 0.1356, 0.1901, 0.1707, 0.1781, 0.1863, 0.1306\}.$$

The eigenvalues of $\mathbf{Q}'_2 \mathbf{D}_{\xi_0}^{-1} \mathbf{Q}_2$ are $\{5.9799, 5.6917, 5.5114\}$, so that Theorem 3(i) applies; we then check numerically that (ii) does as well. The maximizing eigenvector (6) is $\mathbf{x} = (0.4223, 0.3314, -0.8437)'$, and $\mu = \mathcal{L}(\xi_0) = 5.9799$.

It turns out to be quite rare for Theorem 3 to hold – in most cases the maximum eigenvalue is not simple. This is of course expected asymptotically, since all eigenvalues are then equal, and is illustrated for fixed N in the following example.

Example 3. Take $N = 3$, $p = 1$, $\mathbf{Q} = \mathbf{I}_3$, so that $\varepsilon > 0$ and $\mathbf{Q}_2 = [\mathbf{0}; \mathbf{I}_2]'$. Then $\mathbf{Q}'_2 \mathbf{D}_{\xi}^{-1} \mathbf{Q}_2 = \text{diag}(1/\xi_2, 1/\xi_3)$, and $\mathcal{L}(\xi) = 1/\min(\xi_2, \xi_3)$. The problem of maximizing the minimum of (ξ_2, ξ_3) subject to $\xi_2 + \xi_3 = 1 - \varepsilon$ leads to $\xi_{0,2} = \xi_{0,3} = (1 - \varepsilon)/2$, so that the maximum eigenvalue is not simple and Theorem 3 does not apply.

When Theorem 3 does not apply we minimize $\mathcal{L}(\xi)$ directly over Ξ_+ , using randomly generated designs on S as starting values, which are then used in a constrained nonlinear minimizer. While this might not be feasible for very large design spaces, we have found that the limiting behaviour implied by Theorem 2 is approached very quickly – for all but very small values of N the designs are, to the limits of numerical accuracy, uniform on S . See Figures 1 and 2 for illustrations.

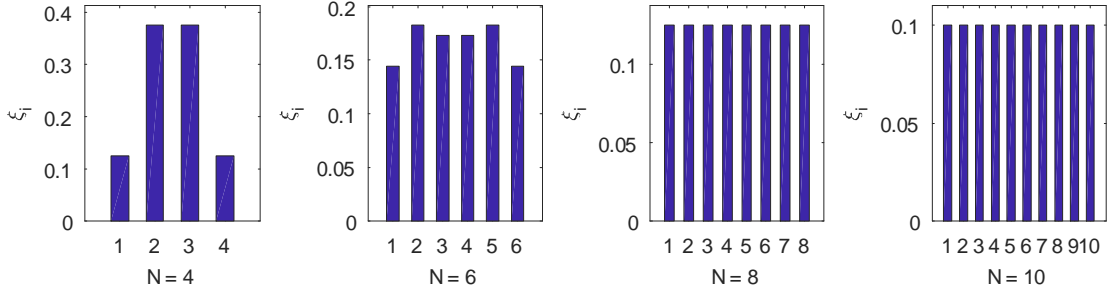


Figure 1: Maximin LOF designs for testing the fit of a quadratic model ($p = 3$); $S = \{-1 + (i - 1)/(N - 1) | i = 1, \dots, N\}$; $N = 4, 6, 8, 10$.

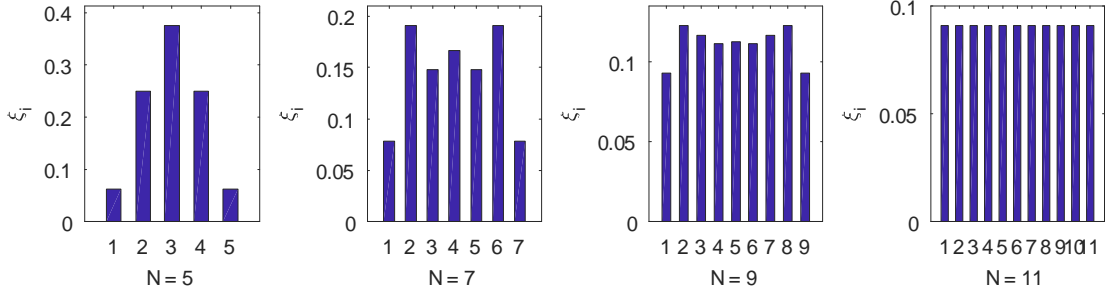


Figure 2: Maximin LOF designs for testing the fit of a cubic model ($p = 4$); $S = \{-1 + (i - 1)/(N - 1) | i = 1, \dots, N\}$; $N = 5, 7, 9, 11$.

3.1 An open problem

Our results have pointed out the limitations of *any* design, supported on a proper subset of S , in providing robustness against alternatives in \mathcal{F}_η^+ . Indeed, it is brought out in the proof of Theorem 1 that if the design is not supported on all of S then there will be departures in \mathcal{F}_η^+ for which the NCP is zero. Such departures may be pathological and inconsequential, but against them the power of the test is no greater than the size. This is a reflection of the richness of \mathcal{F}_η^+ ; an interesting open problem is to find a smaller but still realistic class in which this difficulty is avoided.

4 Appendix: Proofs

Proof of Theorem 1: Note that

$$\mathbf{P}_2' (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2 = N \left\{ \mathbf{Q}_2' \mathbf{D}_\xi \mathbf{Q}_2 - \mathbf{Q}_2' \mathbf{D}_\xi \mathbf{Q}_1 (\mathbf{Q}_1' \mathbf{D}_\xi \mathbf{Q}_1)^{-1} \mathbf{Q}_1' \mathbf{D}_\xi \mathbf{Q}_2 \right\}, \quad (\text{A.1})$$

and that

$$\mathbf{Q}'\mathbf{D}_\xi\mathbf{Q} = \begin{pmatrix} \mathbf{Q}'_1 \\ \mathbf{Q}'_2 \end{pmatrix} \mathbf{D}_\xi \begin{pmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_1 & \mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_2 \\ \mathbf{Q}'_2\mathbf{D}_\xi\mathbf{Q}_1 & \mathbf{Q}'_2\mathbf{D}_\xi\mathbf{Q}_2 \end{pmatrix},$$

so that

$$\begin{aligned} |\mathbf{D}_\xi| &= |\mathbf{Q}'\mathbf{D}_\xi\mathbf{Q}| = |\mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_1| \cdot \left| \mathbf{Q}'_2\mathbf{D}_\xi\mathbf{Q}_2 - \mathbf{Q}'_2\mathbf{D}_\xi\mathbf{Q}_1 (\mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_1)^{-1} \mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_2 \right| \\ &= \left(\frac{1}{N} \right)^{N-p} \cdot |\mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_1| \cdot |\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2|, \end{aligned}$$

using (A.1). Note that $\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2 \succeq 0$, since $\mathbf{I}_N - \mathbf{H}$ is idempotent, hence non-negative definite. The eigenvalues of $\mathbf{Q}'_1\mathbf{D}_\xi\mathbf{Q}_1$, assumed non-singular, are positive. Hence if, and only if, all ξ_i are strictly positive is $|\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2| > 0$, equivalently $ch_{\min} [\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2] > 0$. Thus a maximin design places positive mass on every point in S – otherwise it is ‘beaten’ by such a design.

We then have that

$$(\mathbf{Q}'\mathbf{D}_\xi\mathbf{Q})^{-1} = \mathbf{Q}'\mathbf{D}_\xi^{-1}\mathbf{Q} = \begin{pmatrix} * & * \\ * & \left(\frac{\mathbf{P}'_2(\mathbf{I}_N - \mathbf{H})\mathbf{P}_2}{N} \right)^{-1} \end{pmatrix},$$

so that $\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2 = N (\mathbf{Q}'_2\mathbf{D}_\xi^{-1}\mathbf{Q}_2)^{-1}$. Thus

$$ch_{\min} [\mathbf{P}'_2 (\mathbf{I}_N - \mathbf{H}) \mathbf{P}_2] = N ch_{\min} [(\mathbf{Q}'_2\mathbf{D}_\xi^{-1}\mathbf{Q}_2)^{-1}] = \frac{N}{ch_{\max} [\mathbf{Q}'_2\mathbf{D}_\xi^{-1}\mathbf{Q}_2]}.$$

□

Proof of Theorem 2: Define $\mathbf{K} = \mathbf{Q}_2\mathbf{Q}'_2$, an idempotent matrix whose diagonal elements $\{k_{ii} \mid i = 1, \dots, N\}$ lie in $[0, 1]$ and sum to $rk(\mathbf{Q}_2) = N - p$. Since the average eigenvalue of a positive definite matrix cannot exceed the maximum, we have that

$$\mathcal{L}(\xi) \geq \frac{1}{N-p} tr [\mathbf{Q}'_2\mathbf{D}_\xi^{-1}\mathbf{Q}_2] = \frac{1}{N-p} tr [\mathbf{D}_\xi^{-1}\mathbf{K}] = \sum_{i=1}^N \xi_i^{-1} \frac{k_{ii}}{N-p}.$$

We now view $\{k_{ii}/(N-p)\}$ as a probability distribution and apply Jensen’s Inequality to the convex function $f(\xi) = \xi^{-1}$ to obtain

$$\sum_{i=1}^N \xi_i^{-1} \frac{k_{ii}}{N-p} \geq \frac{1}{\sum_{i=1}^N \xi_i \frac{k_{ii}}{N-p}} \geq \frac{1}{\sum_{i=1}^N \frac{\xi_i}{N-p}} = N-p.$$

□

Proof of Theorem 3: (i) Put $\xi_t = (1 - t)\xi_0 + t\xi_1$ for any ξ_1 and $t \in [0, 1]$. For an undetermined multiplier μ put

$$F(t, \mu) = ch_{\max} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_t}^{-1} \mathbf{Q}_2 \right] + \mu (\mathbf{1}' \mathbf{D}_{\xi_t} \mathbf{1} - 1).$$

Since $\mathbf{1}' \mathbf{D}_{\xi_t} \mathbf{1} - 1 \equiv 0$ for designs ξ_t , it suffices to show that $F(t, \mu)$ is minimized unconditionally at $t = 0$ for fixed μ and any $\xi_1 \in \Xi_\varepsilon$, and that the side conditions ensuring that $\xi_0 \in \Xi_\varepsilon$ are satisfied. (That all $\xi_{0,i} \geq \varepsilon$ is obtained without the use of a multiplier.)

Note that $F(t, \mu)$ is convex in t : we have that $\mathbf{D}_{\xi_t}^{-1} \preceq (1 - t) \mathbf{D}_{\xi_0}^{-1} + t \mathbf{D}_{\xi_1}^{-1}$ by the convexity of matrix inversion, so that

$$\begin{aligned} ch_{\max} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_t}^{-1} \mathbf{Q}_2 \right] &\leq ch_{\max} \left[\mathbf{Q}'_2 \left((1 - t) \mathbf{D}_{\xi_0}^{-1} + t \mathbf{D}_{\xi_1}^{-1} \right) \mathbf{Q}_2 \right] \\ &\leq ch_{\max} \left[\mathbf{Q}'_2 \left((1 - t) \mathbf{D}_{\xi_0}^{-1} \right) \mathbf{Q}_2 \right] + ch_{\max} \left[\mathbf{Q}'_2 \left(t \mathbf{D}_{\xi_1}^{-1} \right) \mathbf{Q}_2 \right] \\ &= (1 - t) ch_{\max} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_0}^{-1} \mathbf{Q}_2 \right] + t ch_{\max} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_1}^{-1} \mathbf{Q}_2 \right]. \end{aligned}$$

A necessary and sufficient condition for a minimum at $t = 0$ is then that $(d/dt) F(t, \mu)|_{t=0} \geq 0$; this must hold for all $\xi_1 \in \Xi_\varepsilon$. With $\delta = \delta(\xi_0)$ and $\mathbf{x} = \mathbf{x}(\xi_0)$ being the maximum eigenvalue (assumed simple) and corresponding eigenvector of unit norm of $\mathbf{Q}'_2 \mathbf{D}_{\xi_0}^{-1} \mathbf{Q}_2$ we have, using Theorem 1 of Magnus (1985), that

$$\frac{d}{dt} \delta(\xi_t)|_{t=0} = \mathbf{x}'(\xi_0) \frac{d}{dt} \left[\mathbf{Q}'_2 \mathbf{D}_{\xi_t}^{-1} \mathbf{Q}_2 \right]_{t=0} \mathbf{x}(\xi_0),$$

whence

$$\begin{aligned} \frac{d}{dt} F(t, \mu)|_{t=0} &= \left\{ \mathbf{x}'(\xi_0) \mathbf{Q}'_2 \frac{d}{dt} \left[\mathbf{D}_{\xi_t}^{-1} \right] \mathbf{Q}_2 \mathbf{x}(\xi_0) + \mu \mathbf{1}' \frac{d}{dt} \mathbf{D}_{\xi_t} \mathbf{1} \right\}_{t=0} \\ &= -\mathbf{x}'(\xi_0) \mathbf{Q}'_2 \mathbf{D}_{\xi_0}^{-1} [\mathbf{D}_{\xi_1} - \mathbf{D}_{\xi_0}] \mathbf{D}_{\xi_0}^{-1} \mathbf{Q}_2 \mathbf{x}(\xi_0) + \mu \mathbf{1}' [\mathbf{D}_{\xi_1} - \mathbf{D}_{\xi_0}] \mathbf{1} \\ &= \sum_{i=1}^N (\xi_{1,i} - \xi_{0,i}) \left[\mu - \left(\frac{\mathbf{q}'_i \mathbf{x}(\xi_0)}{\xi_{0,i}} \right)^2 \right]. \end{aligned}$$

If (5) and (6) hold then $\xi_0 \in \Xi_\varepsilon$ and for any $\xi_1 \in \Xi_\varepsilon$ the final line above is

$$\frac{d}{dt} F(t, \mu)|_{t=0} = \mu \sum_{\xi_{0,i}=\varepsilon} \frac{(\xi_{1,i} - \varepsilon)}{\varepsilon^2} \left[\varepsilon^2 - \left(\frac{\mathbf{q}'_i \mathbf{x}(\xi_0)}{\sqrt{\mu}} \right)^2 \right] \geq 0,$$

since $|\mathbf{q}'_i \mathbf{x}(\xi_0)|/\sqrt{\mu} \leq \varepsilon$ when $\xi_{0,i} = \varepsilon$, and the summands vanish when $\xi_{0,i} = |\mathbf{q}'_i \mathbf{x}(\xi_0)|/\sqrt{\mu} > \varepsilon$.

(ii) If we have that $\mathbf{q}'_i \mathbf{x} \neq 0$ for all i , then the minimizing design lies in the interior of Ξ_+ and so it is the solution in Ξ_ε for any $\varepsilon \leq \min_{i=1,\dots,N} |\mathbf{q}'_i \mathbf{x}|$, and hence for all

of Ξ_+ . In this case (4) and (5) become $\xi_{0,i} = |\mathbf{q}'_i \mathbf{x}| / \sqrt{\mu}$, with $\sqrt{\mu} = \sum_{i=1}^N |\mathbf{q}'_i \mathbf{x}|$, and then (6) yields

$$\mathcal{L}(\xi_0) = \mathbf{x}' \mathbf{Q}'_2 \mathbf{D}_{\xi_0}^{-1} \mathbf{Q}_2 \mathbf{x} = \sum_{i=1}^N \frac{(\mathbf{q}'_i \mathbf{x})^2}{\xi_{0,i}} = \sqrt{\mu} \sum_{i=1}^N |\mathbf{q}'_i \mathbf{x}| = \mu.$$

□

Acknowledgements

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada. We thank the two anonymous referees for their incisive comments.

References

- Biedermann, S., Dette, H. (2001), “Optimal designs for testing the functional form of a regression via nonparametric estimation techniques,” *Statistics and Probability Letters*, 52, 215-224.
- Bischoff, W. (2010), “An improvement in the lack-of-fit optimality of the (absolutely) continuous uniform design in respect of exact designs,” *mODa*, 33-40.
- Bischoff, W., Miller, F. (2006), “Optimal designs which are efficient for lack of fit tests,” *The Annals of Statistics*, 34, 2015-2025.
- Fang, H. B., Chen, X., Pei, X. Y., Grant, S., Tan, M. (2015), “Experimental design and statistical analysis for three-drug combination studies,” *Statistical Methods in Medical Research*, 0962280215574320.
- Fang, H. B., Ross, D. D., Sausville, E., Tan, M. (2008), “Experimental design and interaction analysis of combination studies of drugs with log-linear dose responses,” *Statistics in Medicine*, 27, 3071-3083.
- Fang, H. B., Tian, G. L., Li, W., Tan, M. (2009), “Design and sample size for evaluating combinations of drugs of linear and loglinear dose-response curves,” *Journal of Biopharmaceutical Statistics*, 19, 625-640.
- Fang, K. T., Li, R., Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, CRC Press.
- Lee, Y. J., Huang, S. Y. (2007), “Reduced support vector machines: A statistical theory,” *IEEE Transactions on Neural Networks*, 18, 1-13.

- Magnus, J. R.: On Differentiating Eigenvalues and Eigenvectors. *Econometric Theory* 1, 179-191 (1985).
- Tan, M., Fang, H. B., Tian, G. L., Houghton, P. J. (2003), "Experimental design and sample size determination for testing synergism in drug combination studies based on uniform measures," *Statistics in Medicine*, 22, 2091-2100.
- Tan, M. T., Fang, H. B., Tian, G. L. (2009), "Dose and sample size determination for multi-drug combination studies," *Statistics in Biopharmaceutical Research*, 1, 301-316.
- Wiens, D. P. (1991), "Designs for approximately linear regression: two optimality properties of uniform designs," *Statistics and Probability Letters*, 12, 217-221.
- Wiens, D. P. (1992), "Minimax designs for approximately linear regression," *Journal of Statistical Planning and Inference*, 31, 353-371.
- Xie, M. Y., Fang, K. T. (2000), "Admissibility and minimaxity of the uniform design measure in nonparametric regression model," *Journal of Statistical Planning and Inference*, 83, 101-111.
- Zhang, L., Liang, Y. Z., Jiang, J. H., Yu, R. Q., Fang, K. T. (1998), "Uniform design applied to nonlinear multivariate calibration by ANN," *Analytica Chimica Acta*, 370, 65-77.