A NOTE ON THE OPTIMALITY OF UNIFORM DESIGNS IN TESTING LACK OF FIT
Douglas P. Wiens

July 7, 2017

Abstract In a previous article we established a maximin property, with respect to the test for Lack of Fit, of the absolutely continuous uniform ‘design’ on a design space which is a dense set in $\mathbb{R}^q$. Here we discuss some issues and controversies surrounding this result. We then prove an analogous result – both more applicable and less controversial – pertaining to a finite design space.

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

Key words and phrases Bias, F-test, maximin, power, regression.

1 Introduction

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

The uniform design on $S$ is the absolutely continuous design, with constant density $1/\int_S dx$. 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, the uniform design possesses an optimality property in the class of all designs on $S$ – 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 dense nature of $S$ 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) – which 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 which we feel is both more applicable and less controversial. We take a finite design space \( S = \{x_1, \ldots, x_N\} \) – here \( N \) can be arbitrarily large, allowing for at least a close approximation of the space of interest in an anticipated application – and show that the optimality results derived earlier hold in this setting and in finite samples.

In §2 we outline the mathematical framework and state our main result as Theorem 1, and then discuss issues of implementation. The proof of Theorem 1 is given in §3.

2 Statement of the optimality property

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 \( \lambda \) the uniform probability measure on \( S \), viz. \( \lambda(x_i) = 1/N \); \( i = 1, \ldots, N \).

For a design \( \xi \) on \( S \) we write \( \xi_i = \xi(x_i) \). An implementable design with \( n \) observations requires that \( n\xi_i \) be an integer; we shall loosen this restriction and allow \( \xi \) to be any probability distribution on \( S \). In particular, we include \( \lambda \) as a possible design. See the discussion in Remark 2 below.

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

\[
E[Y(x)] = z^T(x) \theta + f(x) \quad (f \in \mathcal{F}_p^+),
\]

(1)

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

\[
E[Y(x)] = z^T(x) \theta.
\]

(2)

The class \( \mathcal{F}_p^+ \) is as at (i), (ii) on p. 218 of [W], written here in a manner that reflects the discreteness of the design space:

\[
\int_S f^2(x) d\lambda(x) = \frac{1}{N} \sum_{i=1}^{N} f^2(x_i) \geq \eta^2,
\]

(3a)

\[
\int_S z(x) f(x) d\lambda(x) = \frac{1}{N} \sum_{i=1}^{N} z(x_i) f(x_i) = 0_{p \times 1}.
\]

(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 \( \theta \overset{\text{def}}{=} \arg \min_t \sum_{i=1}^N \left( \mathbb{E} [Y (x_i)] - z^T (x_i) t \right)^2 \). This defines \( \theta \) uniquely, in the presence of (3b) and the requirement, made here, that the matrix \( Z_{N \times p} = [z (x_1) : \cdots : z (x_N)]^T \) be of full column rank. We write \( f_i = f (x_i) \) and define \( f = (f_1, ..., f_N)^T \) and \( D_{\xi} = \text{diag} (\xi_1, ..., \xi_N) \). Define as well

\[
    b_{f,\xi} = \int_S z (x) f (x) d\xi (x) = \sum_{i=1}^N z (x_i) f (x_i) \xi_i = Z^T D_{\xi} f, \\

    B_{\xi} = \int_S z (x) z^T (x) d\xi (x) = \sum_{i=1}^N z (x_i) z^T (x_i) \xi_i = Z^T D_{\xi} Z,
\]

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

\[
    B (f, \xi) = f^T D_{\xi} f - b_{f,\xi}^T B_{\xi}^{-1} b_{f,\xi}. \tag{4}
\]

The power of the test is an increasing function of the non-centrality parameter, 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

\[
    B (f, \xi) = \sum_{i=1}^N \left( f (x_i) - z^T (x_i) B_{\xi}^{-1} b_{f,\xi} \right)^2 \xi_i,
\]

we see that \( B (f, \xi) \) is the \( L_2 (\xi) \) distance from \( f \) to the nearest function of the form (2). Thus \( \min_f B (f, \xi) \) is a natural measure of the discrepancy between the ‘full’ and ‘reduced’ models being compared, with larger values leading to a more effective test.

We prove

**Theorem 1** With notation as above, for any design \( \xi \) on \( S \) we have that

\[
    \min_{f \overset{\lambda}{\in} \mathcal{D}_n} B (f, \xi) \leq \eta^2 = \min_{f \overset{\lambda}{\in} \mathcal{D}_n} B (f, \lambda), \tag{5}
\]

so that the uniform design \( \lambda \) maximizes the minimum power of the F-test of \( \text{LOF} \). The inequality in (5) is strict unless \( \xi = \lambda \).
Remarks:

1. Although it is not used in the proof of Theorem 1, in order that the Sum of Squares for Lack of Fit, appearing in the numerator of the F-statistic, have positive degrees of freedom it is necessary that the number $c$ of locations $x_i$ at which $\xi_i > 0$ exceed $p$. The F-statistic is then on $c - p$ and $n - c$ degrees of freedom. That $n$ exceed $c$ is achieved by replicating the design.

2. Unless $N$ is quite small it is not feasible to employ an exactly uniform design $\lambda$. The implication of Theorem 1, and the manner in which it – more precisely, the continuous version of Theorem 1 as given in [W] – has often been interpreted in the literature is that one should aim for a space-filling design. Such designs are discussed in the computer experimentation literature – see, e.g., Santner, Williams and Notz (2003) or Fang, Li and Sudjianto (2005) – and are generally used without replication, since there is no random error to control.

3. Uniform designs are particularly apt in computer experimentation, and in other areas where the control of the bias is of particular concern; this is because the bias of the least squares estimate $\hat{\theta}$ is $E[\hat{\theta} - \theta] = B_\xi^{-1}b_{f,\xi}$, and this vanishes if $\xi = \lambda$.

3 Proof of Theorem 1

Note that $B_\lambda = N^{-1}I_N$ and that by (3b) $b_{f,\lambda} = 0$, so that $B(f,\lambda) = N^{-1}f^Tf$; this combined with (3a) yields the equality in (5).

To establish the inequality it suffices to construct a vector $f_{N \times 1}$ satisfying (3a) with equality and (3b), i.e.,

(a) $f^Tf = \eta^2 N$,

(b) $Z^Tf = 0_{p \times 1}$,

and also satisfying

(c) $f^TD_\xi f \leq \eta^2$.

Then by (a) and (b), $f \in F_\eta^+$ and by (4) together with (c),

$$B(f,\xi) \leq f^TD_\xi f \leq \eta^2.$$
\( Q_{N \times N} = [Q_1; Q_2] \) is orthogonal. Then by (b), \( f \in \text{col}(Q_1) = \text{col}(Q_2) \), so that using (a) as well, \( f = \eta \sqrt{N} Q_2 d \) for some \( d_{N-p \times 1} \) with \( \|d\| = 1 \).

Now to satisfy (c), and thus complete the proof, we need only show the existence of a \( d \), with unit norm, such that \( \eta^2 N d^T Q_2^T D_\xi Q_2 d \leq \eta^2 \). Equivalently,

\[
ch_{\text{min}} \left[ Q_2^T (ND_\xi) Q_2 \right] \leq 1,
\]

where \( ch_{\text{min}} [\cdot] \) denotes the minimum eigenvalue; we can take \( d \) to be the corresponding eigenvector.

We can assume that the positive semidefinite matrix \( Q_2^T (ND_\xi) Q_2 \) is in fact positive definite, else its minimum eigenvalue is 0 and there is nothing to prove. Denote by \( D_\xi^{1/2} \) the diagonal matrix with diagonal elements \( \sqrt{\xi_i} \). Recall that the nonzero eigenvalues of a matrix product \( AA^T \) coincide with those of \( A^T A \). Finally, note that, with respect to the Loewner ordering by positive semidefiniteness, \( I_N = QQ^T = Q_1 Q_1^T + Q_2 Q_2^T \succeq Q_2 Q_2^T \). Then

\[
ch_{\text{min}} \left[ Q_2^T (ND_\xi) Q_2 \right] = N ch_{\text{min}} \left[ Q_2^T D_\xi^{1/2} \cdot D_\xi^{1/2} Q_2 \right] \\
= N ch_{\text{min}} \left[ D_\xi^{1/2} Q_2 \cdot Q_2^T D_\xi^{1/2} \right] \\
\leq N ch_{\text{min}} \left[ D_\xi^{1/2} \cdot D_\xi^{1/2} \right] \\
= N \min_i \xi_i \\
\leq 1,
\]

and the final inequality is strict unless \( \xi = \lambda \). \( \square \)

**Acknowledgements**

This work was carried out with the support of the Natural Sciences and Engineering Research Council of Canada.

**References**


design and statistical analysis for three-drug combination studies,” Statistical
Methods in Medical Research, 0962280215574320.

and interaction analysis of combination studies of drugs with log-linear dose
responses,” Statistics in Medicine, 27, 3071-3083.

evaluating combinations of drugs of linear and loglinear dose-response curves,”

Fang, K. T., Li, R., Sudjianto, A. (2005), Design and Modeling for Computer
Experiments, CRC Press.


Santner, T. J., Williams, B. J., Notz, W. I. (2003), The Design and Analysis of

sign and sample size determination for testing synergism in drug combination
studies based on uniform measures,” Statistics in Medicine, 22, 2091-2100.

for multi-drug combination studies,” Statistics in Biopharmaceutical Research,
1, 301-316.

properties of uniform designs,” Statistics and Probability Letters, 12, 217-221.


Xie, M. Y., Fang, K. T. (2000), “Admissibility and minimaxity of the uniform de-
sign measure in nonparametric regression model,” Journal of Statistical Plan-
ning and Inference, 83, 101-111.

sign applied to nonlinear multivariate calibration by ANN,” Analytica Chimica Acta, 370, 65-77.