Journal of Statistical Planning and Inference 139 (2009) 23 – 29

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Two-level supersaturated designs for 2^k runs and other cases

Neil A. Butler*

School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

ARTICLE INFO

Available online 23 May 2008

Keywords:
Effect sparsity
$E(s^2)$-optimality
Factor sparsity
Fractional factorial
Hadamard matrix
Minimax

ABSTRACT

Two-level supersaturated designs are constructed for $n=2^k$ $(k \geqslant 5)$ runs and m factors where $n+3 \leqslant m \leqslant 5(n-4)$. The designs so formed are shown to have a maximum absolute correlation between factors of $\frac{1}{4}$ and to be efficient in terms of $E(s^2)$, particularly when the number of factors m is approximately double the number of runs n or greater. Thus, supersaturated designs with favourable properties are found for much higher numbers of runs than would be possible solely using algorithms.

1. Introduction

Supersaturated designs (Satterthwaite, 1959; Booth and Cox, 1962; Lin, 1993, 1995) are designs with at least as many factor main effects as there are experimental runs. The designs are particularly useful when there are many factors and when the cost of experimentation is expensive. The objective of experimentation using supersaturated designs is to determine the so-called active factors which have the most substantial effect on the response of interest. However, when using supersaturated designs, it is important to be aware that the complex aliasing structure between the factors makes it much more likely that the results of the analysis could be misinterpreted. Therefore, follow-up experiments (Meyer et al., 1996; Lewis and Dean, 2001) should usually be carried out on the factors that are highly correlated with the response to check that they represent real effects.

In choosing a two-level supersaturated design, the aim is to minimise in some sense the pairwise correlations between factors, or strictly speaking between factor main effects. The two main criteria introduced by Booth and Cox (1962) are the minimisation of $E(s^2)$, where the average squared correlation between factors is minimised, and minimax, where the maximum absolute correlation between factors is minimised. Work on the minimax criterion includes Lin (1993), Wu (1993), Cheng (1997), Butler (2005) and Ryan and Bulutoglu (2007). Work on $E(s^2)$-optimality includes Nguyen (1996), Cheng (1997), Tang and Wu (1997), Liu and Zhang (2000), Butler et al. (2001), Bulutoglu and Cheng (2004), Liu and Dean (2004), Eskridge et al. (2004), Bulutoglu (2007), Koukouvinos et al. (2007, 2008) and Nguyen and Cheng (2008).

In this paper, two-level supersaturated designs are constructed for $n=2^k$ runs $(k \geqslant 5)$ with m factors where $n+3 \leqslant m \leqslant 5(n-4)$ except for the 11 cases where $3(n-4) < m < 3n$. Designs are also constructed for n a multiple of 64. The designs utilise some 16-run designs that are efficient in terms of $E(s^2)$ and minimax previously given by Butler (2005), Ryan and Bulutoglu (2007) and Liu and Zhang (2000). The designs so formed are shown to be appealing in terms of both $E(s^2)$ and minimax with a maximum absolute correlation between factors equal to $\frac{1}{4}$. Moreover, the methods in this paper allow much larger supersaturated designs, in terms of the number of runs, to be found than would be possible with search algorithms.

* Tel.: +44 115 951 4949; fax: +44 115 951 4951.
E-mail address: neil.butler@nottingham.ac.uk.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.05.013

2. Background

Two-level supersaturated designs for n runs and $m \geqslant n$ factors are specified by an $n \times m$ design matrix X. The design matrix X has all elements equal to ± 1, and has an equal number of $+1$'s and -1's in each column so that factors are orthogonal to the overall mean.

The absolute pairwise correlation between factors i and j with factor columns x_i and x_j is $|s_{ij}|/n$ where $s_{ij} = x_i' x_j$. Note that factors are orthogonal if $s_{ij} = 0$. The $E(s^2)$-optimality criterion involves minimising

$$E(s^2) = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} s_{ij}^2,$$

where, for identifiability purposes, designs are not allowed to have completely aliased factors or columns. It turns out that $E(s^2)$-optimal designs tend to rely heavily on Hadamard matrices. An $n \times n$ Hadamard matrix H has all elements equal to ± 1 and has orthogonal rows and columns so that $H'H = HH' = nI_n$; see, for example, Dey and Mukerjee (1999). A Hadamard matrix is said to be normalized if all the elements in the first row and first column of the matrix equal $+1$. Note that Hadamard matrices only exist if $n = 1, 2$ or n is a multiple of 4, and the most well known in statistics are the Plackett and Burman (1946) designs.

$E(s^2)$-optimal designs are easily constructed from Hadamard matrices when the number of runs n is a multiple of 4 and the number of factors is $m = q(n-1)$ where q is an integer. In these cases, $E(s^2)$-optimality is achieved by designs with

$$X_0 = (F_1, \ldots, F_q), \tag{1}$$

where each F_i is an $n \times (n-1)$ matrix formed by excluding the first column of an $n \times n$ normalized Hadamard matrix. However, it must be checked that the designs do not have any completely aliased columns. The designs attain a lower bound LB for $E(s^2)$

$$\mathrm{LB} = \frac{(m-n+1)n^2}{(n-1)(m-1)}$$

given by Nguyen (1996) and Tang and Wu (1997). Also

$$X_0 X_0' = qnI_n - qJ_n,$$

where I_n and J_n are the $n \times n$ identity matrix and matrix of ones, respectively.

Designs for n runs and $m = q(n-1) + r$ factors $(0 < r < n-1)$ can be formed by combining the design (1) with r extra columns (excluding the first column) from an additional Hadamard matrix; see Tang and Wu (1997). Again it needs to be checked that no factor columns are completely aliased. Such a design is called here a Hadamard constructed design. The following result appeared in Butler (2005) without proof but is proved here in the Appendix.

Theorem 1. *A Hadamard constructed design for n runs and $m = q(n-1) + r$ factors $(0 \leqslant r < n-1)$ has*

$$\sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} s_{ij}^2 = [m(q-1) + r(q+1)]n^2. \tag{2}$$

Hadamard constructed designs are $E(s^2)$-optimal when $\min(r, n-1-r) \leqslant 2$ (Cheng, 1997), but they are not generally $E(s^2)$-optimal for other values of r (Butler et al., 2001). However, we shall see that Hadamard constructed designs often perform reasonably well under the $E(s^2)$ criterion, motivating the following definition.

Definition. A design is said to be as $E(s^2)$-efficient as Hadamard constructed designs if

$$\sum_{i=1}^{m} \sum_{j=1, j \neq i}^{m} s_{ij}^2 \leqslant [m(q-1) + r(q+1)]n^2. \tag{3}$$

3. 16-Run designs

Butler (2005) found some two-level supersaturated designs for 16 runs with a maximum absolute correlation of $r_{\max} = \frac{1}{4}$ which are as $E(s^2)$-efficient as Hadamard constructed designs. The construction of these designs is based on the following definitions. Define the 16×16 `regular design' Hadamard matrix

$$H_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$