An Algorithm for Constructing a D-Optimal 2^K Factorial Design for Linear Model Containing Main Effects and One-Two Factor Interaction

E. O. Effanga
Department of Mathematics,Statistics and Comp. Science, University of Calabar
P.M.B. 1115, Calabar, Cross River State, Nigeria
E-mail: effanga2005@yahoo.com

C. E. Onwukwe
Department of Mathsematics, Statistics and Comp. Science, University of Calabar
P.M.B. 1115, Calabar, Cross River State, Nigeria
E-mail: ceonwukwe@yahoo.com

Received: December 15, 2010 Accepted: January 6, 2011 doi:10.5539/jmr.v3n2p200

Abstract

In this paper an algorithm to construct a D-optimal 2^k factorial design based on the work of (Hedayat & Pesotan, 2007) is developed and coded in a high level computer programming language, JAVA. Our algorithm is able to generate all possible square matrices of order (k + 2) from a 2^k by (k + 2) matrix, select all possible g(k, 1) design matrices of order (k + 2), and hence select a D-optimal design matrix. Furthermore, the computational formulas for the estimation of parameters for the 2^2 and 2^3 designs are derived. The results obtained by our algorithm agree with the theoretical results derived in.

Keywords: Algorithm, Factorial design, D-optimal, Linear model, g(k, 1) design

1. Introduction

A 2^k factorial design is a design involving k-factors each at two levels, low '0' level and high '1' level. As the number of factors k increases, the number of runs required for a complete replicate of the design rapidly outgrows the resources of the experimenters, (Montgomery, 1976).

If certain high order interactions are assumed to be negligible, then information on the main effects and low order interaction may be obtained by running only a fraction of the complete factorial experiments, (Montgomery, 1976). Such designs are widely used in factors screening experiments and many scientific investigations.

The problem is, given k factors each at two levels, to specify a design with at least $N = k + 2$ level combinations so that N parameters can be estimated unbiased. A number of approaches are available to tackle this problem. (Taguchi, 1959) suggested a graph aided procedure to help construct a suitable fraction. (Greenfield, 1976) suggested a symmetric tree search procedure to select an appropriate design. (Wu/Chen, 1992) also suggested a variant of graph aided procedure. (Hedayat/Pesotan, 1992) Studied 2^k factorial design for main effects and selected two - factor interaction. They introduce the concepts of g(k, e) - design and g(k, e) - matrix to study designs of 2^k experiments which can unbiasedly estimate the mean, the main effects, and a specified two - factor interactions appearing in an orthogonal model. (Hedayat/Pesotan, 2007) Provide tools for the construction of a D- optimal 2^k factorial design for a linear model containing main effects and one two-factor interaction.

The purpose of this paper is to develop an algorithm and computer software to implement the work of (Hedayat/Pesotan, 2007).

2. The Basic Design Principles Underlying the Development of our Algorithm

2.1 The Linear Model

A 2^k factorial design is a set of level combinations, $i = (i_1, i_2, ..., i_k),\ i = 1, 2, ..., 2^k$, where $i_j = 0$ or 1, $j = 1, 2, ..., k$. Let y_i be the response corresponding to the i^{th} combination level. Then the general linear model containing the main effects and one two - factor interaction as given in (Hedayat/Pesotan, 2007) is

$$y_i = \mu + \sum x_{ij}\alpha_j + x_{ij}x_{ir}\alpha_{jr} + e_i, \quad i = 1, 2, \cdots, N, \tag{2.1}$$

where

α_i := Main effect of factor j

α_{jr} := Interaction effect of factor j and factor r

μ := Overall mean

$N = 2^k$:= number of level cobinatimons

$x_{ij} = 1$, if factor j appears in the i^{th} combination level

$x_{ij} = -1$, if factor j do not appear in the ith combination level

e_i := random error which is assumed to be normal with mean 0 and constant variance.

2.2 The g(k, 1) - Design

In In (Hedayat/Pesotan, 2007), a 2^k factorial design is called a g(k, 1) - design if and only if:

(i) It is capable of providing an unbiased estimate for each of the parameters provided by the linear model (2.1),

(ii) It is saturated, that is, it contains (k+2) level combinations which corresponds to the number of parameters in the model.

A nonsingular square matrix X of order (k+2) whose entries are -1 and 1 is called a g(k, 1) - matrix if and only if it is in the form,

$$X = (1|X_1|W), \tag{2.2}$$

where $w_i = x_{ij}x_{ir},\ i = 1, 2, \cdots, k+2$; the sub-matrix X_1 is called the core of X.

If -1 is replaced by 0 in the matrix X, a g(k, 1) - design is produced. Thus constructing a g(k, 1) - design is equivalent to constructing a g(k, 1) - matrix.

2.3 The D-Optimal Design

Let Ω be a class of all n - square matrices. Then an n - square matrix M is said to be D - optimal in Ω if $|M| \geq |Q|$, for any other n - square matrix Q in Ω.

3. Index of interacting Columns of Core Matrix X_1

Let r and s be two columns of X_1 which interact to give W. The product $w = ros$ is called a shur product of r and s. Let f_{++} be the frequency of ++ pair between r and s, f_{+-} the frequency of + - pair between r and s, f_{-+} the frequency of - + pair between r and s, andf_{--} the frequency of - - pair between r and s. Then the index of r and s is defined as

$$i(r, s) = \max\{f_{++}, f_{+-}, f_{--}\} \tag{3.1}$$

The following lemma is in (Hedayat/Pesotan, 2007).

Lemma Let G (k, 1) be a class of all design matrices g(k, 1). Let T(r, s) be a design matrix, then with the interacting columns r and s, $i(r, s) \geq 1$.

3.1 The Design of Algorithm

The sequence of steps for our algorithm is as follows:

Step 0:

Input the number of factors k, the number of combination level N, and N by (k+2) matrix of -1 and 1 entries X,

Step 1:

Compute

$$N^* = \binom{N}{k+2}$$

Select all possible N^* square sub-matrices $A_n = (1\ B_n\ W_n)$ of order (k+2), $n = 1, 2, \cdots, N^*$, from X, where $w_n = r_n o s_n,\ r_n$ and s_n are the interacting columns of B_n.

Step 2:

Set n: = 1.

Is $n \leq N^*$?

If Yes: Compute $\det(A_n)$ and go to step 3.

If No: Go to step 6.

Step 3:

Is $\det(A_n) = 0$?

If Yes: A_n is not a design matrix. Set n: $= n + 1$ and return to to step 2.

If No: Go to step 4.

Step 4:

Compute the frequencies $f_{++},\ f_{+-},\ f_{-+}$, and f_{--} of the interacting columns r and s.

Compute the index $i^n(r,\ s)$ of the interacting columns r and s.

Step 5:

Is $i^n(r,\ s) \geq 1$?

If Yes: A_n is a design matrix. Set n: $= n + 1$ and return to to step 2.

If No: A_n is not a design matrix. Set n: $= n + 1$ and return to to step 2.

Step 6: Compute

$$\det(A_n^*) = \max\{\det(A_n) :\ \det(A_n) \neq 0,\ 1 \leq n \leq N^*\}$$

A_n^* is a D-optimal 2^k design matrix.

Stop.

4. Illustration

4.1 Example 1: D-Optimal 2^2 factorial Design

There are 4 combination levels in a 2^2 factorial design as shown below:

$< Table\ 1 >$

The linear model for this design is as follows:

$$y_i = \mu + x_{1i}\alpha_1 + x_{2i}\alpha_2 + x_{1i}x_{2i}\alpha_{12} + e_i,\ i = 1, 2, 3, 4$$

That is,

$$\begin{aligned} y_1 &= \mu - \alpha_1 - \alpha_2 + \alpha_{12} + e_1 \\ y_2 &= \mu + \alpha_1 - \alpha_2 - \alpha_{12} + e_2 \\ y_3 &= \mu - \alpha_1 + \alpha_2 - \alpha_{12} + e_3 \\ y_4 &= \mu + \alpha_1 + \alpha_2 + \alpha_{12} + e_4 \end{aligned}$$

Therefore the normalized matrix X is as follows:

$$X = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

In step 0 of our algorithm, $k = 2$ and X is a 4 by 4 matrix.

In step 1, $N = 1$, so we have only one 4 by 4 sub-matrix that can be selected from X, and that is X itself.

In step 2, $n = 1 = N^*$, so we compute $\det(A_1) = \det(X) = 2^4$.

In step 3, $\det(A_1) \neq 0$, so we go to step 4.

In step 4, we compute the index of the interacting columns, column 1 and column 2, of the core matrix as $i(1,\ 2) = 1$.

The matrix X is a g(2, 1)-matrix since $i(1,\ 2) = 1$ and $\det(A_1) \neq 0$.

The unbiased estimate of the parameters $\mu,\ \alpha_1,\ \alpha_2,\ \alpha_{12}$ is the solution of the following system of linear equations:

$$\begin{aligned}
y_1 &= \mu - \alpha_1 - \alpha_2 + \alpha_{12} \\
y_2 &= \mu + \alpha_1 - \alpha_2 - \alpha_{12} \\
y_3 &= \mu - \alpha_1 + \alpha_2 - \alpha_{12} \\
y_4 &= \mu + \alpha_1 + \alpha_2 + \alpha_{12}
\end{aligned}$$

That is,

$$\hat{\mu} = \frac{1}{4}\left(y_1 + y_2 + y_3 + y_4\right)$$

$$\hat{\alpha}_1 = \frac{1}{4}\left(y_2 + y_4 - y_1 - y_3\right)$$

$$\hat{\alpha}_2 = \frac{1}{4}\left(y_3 + y_4 - y_1 - y_2\right)$$

$$\hat{\alpha}_{12} = \frac{1}{4}\left(y_1 + y_4 - y_2 - y_3\right).$$

4.2 Example 1: D-Optimal 2^3 *Factorial Design*

There are 8 combination levels in a 2^3 factorial design as shown in:

$< Table\ 2 >$

The linear model for this design is

$$y_i = \mu + x_{1i}\alpha_1 + x_{2i}\alpha_2 + x_{3i}\alpha_3 + x_{1i}x_{2i}\alpha_{12} + e_i,\ 1 \le i \le 8,$$

$$\begin{aligned}
y_1 &= \mu - \alpha_1 - \alpha_2 - \alpha_3 + \alpha_{12} + e_1 \\
y_2 &= \mu + \alpha_1 - \alpha_2 - \alpha_3 - \alpha_{12} + e_2 \\
y_3 &= \mu - \alpha_1 + \alpha_2 - \alpha_3 - \alpha_{12} + e_3 \\
y_4 &= \mu - \alpha_1 - \alpha_2 + \alpha_3 + \alpha_{12} + e_4 \\
y_5 &= \mu + \alpha_1 + \alpha_2 - \alpha_3 + \alpha_{12} + e_5 \\
y_6 &= \mu + \alpha_1 - \alpha_2 + \alpha_3 - \alpha_{12} + e_6 \\
y_7 &= \mu - \alpha_1 + \alpha_2 + \alpha_3 - \alpha_{12} + e_7 \\
y_8 &= \mu + \alpha_1 + \alpha_2 + \alpha_3 + \alpha_{12} + e_8
\end{aligned}$$

Therefore the normalized matrix X is as follows:

$$X = \begin{pmatrix}
1 & -1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 \\
1 & 1 & 1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 \\
1 & -1 & 1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1
\end{pmatrix}$$

We need a 5 by 5 sub-matrix A of X to find the unbiased estimate of the five parameters. There are 56 such sub-matrices of which 36 are g(3, 1)-matrices each with determinant 25 and index interacting column $i(1,\ 2) = 1$.

The matrix

$$X = \begin{pmatrix}
1 & -1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 \\
1 & -1 & 1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 \\
1 & 1 & 1 & -1 & 1
\end{pmatrix}$$

is a D - optimal 2^5 design matrix, and the corresponding unbiased estimates of the parameters $\mu,\ \alpha_1,\ \alpha_2,\ \alpha_3$ and α_{12} is the solution of the following system of linear equations:

$$\begin{aligned}
y_1 &= \mu - \alpha_1 - \alpha_2 - \alpha_3 + \alpha_{12} + e_1\\
y_2 &= \mu + \alpha_1 - \alpha_2 - \alpha_3 - \alpha_{12} + e_2\\
y_3 &= \mu - \alpha_1 + \alpha_2 - \alpha_3 - \alpha_{12} + e_3\\
y_4 &= \mu - \alpha_1 - \alpha_2 + \alpha_3 + \alpha_{12} + e_4\\
y_5 &= \mu + \alpha_1 + \alpha_2 - \alpha_3 + \alpha_{12} + e_5.
\end{aligned}$$

That is,

$$\begin{aligned}
\hat{\mu} &= \frac{1}{4}\left(y_5 + 2y_4 + 2y_3 + 2y_2 - 3y_1\right)\\
\hat{\alpha}_1 &= \frac{1}{4}\left(y_5 + 2y_2 - y_3 - 3y_1\right)\\
\hat{\alpha}_2 &= \frac{1}{4}\left(y_5 + 2y_3 - y_2 - 3y_1\right)\\
\hat{\alpha}_3 &= \frac{1}{2}\left(y_4 - y_1\right)\\
\hat{\alpha}_{12} &= \frac{1}{4}\left(y_5 - y_3 - y_2 - y_1\right).
\end{aligned}$$

5. Conclusion

We have studied the work of (Hedayat/Pesotan, 2007) carefully, and then developed an algorithm to implement it. The algorithm so developed is coded in a high level programming language, JAVA. Our algorithm is used in constructing a D-optimal g(2, 1) and g(3, 1)-designs. We observed that a D-optimal design is not unique for $k > 2$. In particular, for $k = 3$ there are 36 g(3, 1) D-optimal designs.

The unbiased estimates of the parameters in the linear model for the two designs are obtained, respectively. Since the D-optimal design is not unique for $k > 2$, the unbiased estimates of the parameters varies from one design to another. One should therefore seek for the D-optimal design that will minimize the sum of squares of the errors.

The program code for our algorithm is available for the interested reader by contacting any of the authors.

References

A. A. Greenfield. (1976). Selection of Defining Contrast in Two-level Experiments, *Appl. Statist.* 25, 64 – 67.

A. S. Hedayat and H. Pesotan. (1992). Two-level Factorial Designs for Main Effects and Selected Two-factor Interactions *Statist. Sinica*, 2, 453 – 464.

A. S. Hedayat and H. Pesotan. (1997). Designs for Two-level Factorial Experiments with Linear Model Containing Effects and Selected Two-factor Interaction *J. Statist. Plann. Inference*, 64, 109 – 124.

A. S. Hedayat and H. Pesotan. (2007). Tools for Constructing Optimal Two-level Factorial Designs fora Linear Model Containing Main Effects and Two-factor Interactions *J. Statist. Plann. Inference*, 137, 1452-1463.

C. F. J. Wu and Y. Chen. (1992). Graph Aided Method for Planning Two-level Experiments when Certain Interactions are Important *Technometrics*, 34, 164-175.

D. C. Montgomery. (1976). *Design and Analysis of Experiments*, John Willey & sons, New York.

G. Taguchi. (1959). Linear Graphs for Orthogonal Arrays and their Applications to Experimental Designs with the Aid of Various Techniques *Reports of Statistical Applications Research Union of Japanese Scientists and Engineers*, 6, 1-43.

Table 1. Levels Combination in 2^2 Factorial Design

Level Combination	Factor A	Factor B	Response
1	0	0	$y_1 = (1)$
2	1	0	$y_2 = a$
3	0	1	$y_3 = b$
4	1	1	$y_4 = ab$

Table 2. Levels Combination in 2^3 Factorial Design

Level Combination	Factor A	Factor B	Factor C	Response
1	0	0	0	$y_1 = (1)$
2	1	0	0	$y_2 = a$
3	0	1	0	$y_3 = b$
4	0	0	1	$y_4 = c$
5	1	1	0	$y_5 = ab$
6	1	0	1	$y_6 = ac$
7	0	1	1	$y_7 = bc$
8	1	1	1	$y_8 = abc$