Statistics and Applications
Volume 6, Nos.1 & 2, 2008 (New Series), pp.125-140

On searching probability of two-factor interaction[1]

Kashinath Chatterjee[1], Sudhir Gupta[2] and Angshuman Sarkar[1]
[1] *Department of Statistics, Visva-Bharati, Santiniketan, India*
[2] *Division of Statistics, Northern Illinois University, Illinois, USA*

Abstract

The probability of correct model identification, the searching probability, is important for assessing search designs. We show that the expression for searching probability provided by Shirakura, Takashi and Srivastava (1996) significantly over-estimates the true searching probability. We first present two classes of balanced designs for identifying one non-negligible two-factor interaction along with their searching probabilities obtained using the expression given by Shirakura, Takashi and Srivastava (1996). Then, we present a simulation study which shows that their expression substantially over-estimates the true searching probability, particularly for moderate effect size relative to error variance. We, thus, recommend that searching probability should be evaluated through simulation unless error variance is expected to be small relative to effect size. Some results on searching probability for identifying two non-negligible two-factor interactions are also presented.

Key words: Main-effect plus k plan; Model identification; Search designs; Two-factor interaction.

1 Introduction

Identification of correct model is essential for optimization of processes and systems in engineering and the sciences. Fractional factorials provide an important tool for evaluating several factors simultaneously. Under the assumption of effect sparsity, frequently a resolution III plan is utilized for carrying out the experiment. However, in practice

[1]Dedicated to Professor Aloke Dey on the occasion of his retirement.

such an assumption may not hold and it is likely that few interactions are also present in addition to the main effects. Assuming hierarchy of effects, non-negligible effects are usually taken to be the two-factor interactions but it is seldom known to the experimenter which of the two-factor interactions may be non-zero. Srivastava (1975) introduced search designs for identifying few non-zero effects, e.g. interactions, and estimating them in addition to the estimation of a given set of effects the experimenter is interested in, e.g main effects.

A plan is said to be a *main-effect plus k* (MEP.k) plan if it can estimate all the main effects and can identify and estimate at most k non-negligible interactions. Srivastava (1975) gave a necessary condition for a plan to be MEP.k. This condition is also sufficient under a deterministic model. Since then many researchers have contributed to this area, especially for MEP.1 and MEP.2 plans for two-level factors, that can identify and estimate one interaction and two interactions respectively from the set of all two- and three-factor interactions. Ghosh (1980) constructed a class of MEP.1 plans for $2^p - 1$ two-level factors, $p \geq 2$. For experiments having between 3 and 8 factors, Ghosh (1981) provided another MEP.1 plan; see also Ohinishi and Shirakura (1985) for some new designs and related results. Shirakura and Tazawa (1991) proposed another class of MEP.1 plans for two-level factors. Shirakura (1991) gave an MEP.1 plan for seven factors, and proved that the minimum number of runs for an MEP.1 plan for seven factors is 15. He also presented a method of constructing MEP.2 plans for $2^p - 1$ two-level factors, $p \geq 3$, using BIB designs. Srivastava (1992) gave an MEP.2 plan for a 2^8 experiment. Mukerjee and Chatterjee (1994) and Chatterjee *et al.* (2002) gave some further methods of constructing MEP.2 plans.

Keeping in view the main objective of search designs, it is important that search designs should be able to identify true non-negligible effects with high probability. Searching probability of designs was, therefore, investigated by Shirakura *et. al.* (1996). They studied the properties of sum of squares of error, the main criterion used for identifying non-zero effects, and developed an expression for searching probability using its upper bound. They also compared the search probabilities of designs proposed by Ohinishi and Shirakura (1985) and, Shirakura and Tazawa (1991) for various effect sizes relative to error variance. The authors also presented a new plan by adding two runs to the plan given by Shirakura (1991) and compared its efficiency with the plans of Ohinishi and Shirakura (1985) in the light of search probability. Ghosh and Teschmacher (2002) introduced the concept of search probability matrix and defined two new criteria for search designs. For 2^4 experiments, they compared three 12-run plans and three 11-run plans with respect to their criteria. They noted that search probability lies between 1/2 and 1.0. Of course, as our simulation results show, search probability could be near zero for small effect sizes relative to error variance. For further results on search designs for two-level factors, the reader is referred to the above mentioned papers and the references cited therein.

The present paper is organized as follows. Section 2 gives the notations and preliminaries. Section 3 presents two classes of balanced designs for searching and estimating one two-factor interaction in addition to estimating the general mean and all main effects. Corresponding searching probabilities obtained through the expression developed by Shirakura *et. al.* (1996) are also presented in this section. In section 4, searching probabilities are computed using simulation for the two classes of designs. The simulated searching probabilities are compared with those presented in section 3. One of the two classes of designs presented, namely D_2, is also capable of searching and estimating two non-negligible two-factor interactions, in addition to estimating the general mean and all main effects. Searching probability obtained using simulation for D_2, for the case of two two-factor interactions, is also presented in section 4.

2 Notations and preliminaries

We consider m two-level factors $F_1, \ldots, F_m$, where the levels of each factor are coded as $0, 1$. Consider a design d consisting of N treatment combinations $a_{i1}a_{i2}\ldots a_{im}$, $1 \le i \le N$, where $a_{ij} \in \{0,1\}$ for every i, j. We consider the case where three-factor and higher order interactions are negligible and it is known, *a priori* , that at most one or two two-factor interactions may be non-zero in addition to the main effects. Let θ_{kl} denote the interaction effect between factors F_k and F_l, $1 \le k < l \le m$, $W = \{\theta_{kl}\}$, the collection of all possible $g = m(m-1)/2$ two-factor interaction effects, $h_s = \{\theta_{k_1l_1}, \theta_{k_2l_2}, ..., \theta_{k_sl_s}\}$, $(k_i, l_i) \ne (k_j, l_j)$, $i \ne j = 1, 2, ..., s$, and $H(s) = \{h_s\}$, the collection of all sets of s two-factor interaction effects. Note that our attention will be mostly restricted to $s = 1$.

The full linear model for a design d with N runs that includes all main effects and two-factor interactions is given by

$$y = 1_N\mu + \sum_{j=1}^{m} Z_j\theta_j + \sum_{\theta_{kl}\in W} Z_{kl}\theta_{kl} + e, \tag{1}$$

where y is the $N\times 1$ observational vector, 1_N is the $N\times 1$ vector of 1's, μ is the general mean, θ_j is the main effect of F_j with corresponding design matrix $Z_j,\ 1\le j\le m$, θ_{kl} is the two-factor interaction between factors F_k and F_l with corresponding design matrix Z_{kl}, $1\le k< l\le m$, and e is the vector of random errors. It is assumed that $e\sim N(0,\sigma^2 I_N)$, where I_N is the identity matrix of order N.

Let $M(kl)$ denote the model containing the general mean, all the main effects, and θ_{kl}, the two-factor interaction between the factors F_k and F_l. If we assume that $M(kl)$ is the true model, then (1) reduces to

$$y = 1_N\mu + Z\theta + Z_{kl}\theta_{kl} + e,$$

where $Z=[Z_1,\ldots,Z_m]$ and $\theta=(\theta_1,\ldots,\theta_m)'$. Following Srivastava (1975), a necessary condition for design d to estimate μ, all the elements of θ and identify and estimate at most one two-factor interaction is given by

$$\mathrm{rank}[1_N,\, Z,\, Z_{k_1l_1},\, Z_{k_2l_2}] = m+3, \tag{2}$$

for every $(k_1,l_1)\ne(k_2,l_2)$, $1\le k_1<l_1\le m$, $1\le k_2<l_2\le m$.

The best linear unbiased estimators of $\beta=(\mu,\,\theta',\,\theta_{kl})'$ and y are given by

$$\hat\beta = (X_{kl}'X_{kl})^{-1}X_{kl}'y, \text{ and } \hat y = X_{kl}\hat\beta,$$

respectively, where $X_{kl}=[1_N,\, Z,\, Z_{kl}]$. The sum of squares due to error $SSE(kl)$ is then given by

$$SSE(kl) = y'(I_N - Q(kl))y,$$

where $Q(kl)=X_{kl}(X_{kl}'X_{kl})^{-1}X_{kl}'$.

Since we have no idea regarding which of the element of W is non-negligible, following Srivastava (1975), the sum of squares due to error corresponding to each of the models $M(kl), 1\le k<l\le m$, say,

$SSE(12)$, $SSE(13)$,..., $SSE((m-1)m)$ is evaluated. Then $M(kl)$ is selected as the true model if

$$SSE(kl) = \min_{1 \le k^* < l^* \le m} SSE(k^* l^*).$$

Even if a design d satisfies the necessary condition (2), the above procedure based on minimization of error sum of squares may not identify the correct model with certainty when $\sigma^2 > 0$. Therefore, Shirakura *et al.* (1996) studied searching probability, i.e. the probability of correct model identification, for the search procedure based on minimizing the error sum of squares. Shirakura *et al.* (1996) gave the following expression for the searching probability of design d,

$$\mathrm{P}_d = \min_{1 \le k < l \le m} \; \min_{1 \le k^* < l^* \le m, (k^*, l^*) \ne (k,l)} \mathrm{P}(h_{kl} > h_{k^* l^*}),$$

where $h_{kl} = y'(Q(kl) - Q)y$, $Q = X(X'X)^{-1}X'$, and $X = [1_N, Z]$. After consideration simplification, P_d can be expressed as

$$\mathrm{P}_d = \min_{1 \le k < l \le m} \; \min_{1 \le k^* < l^* \le m, (k^*, l^*) \ne (k,l)} [1 - \Phi(\lambda\sqrt{1 - a_{klk^*l^*}}\,) - \Phi(\lambda\sqrt{1 + a_{klk^*l^*}}\,) + 2\Phi(\lambda\sqrt{1 - a_{klk^*l^*}}\,)\Phi(\lambda\sqrt{1 + a_{klk^*l^*}}\,)], \quad (3)$$

where $\rho = \theta_{kl}/\sigma$ is the effect size relative to error variance, $\Phi(x)$ is the distribution function of the standard normal distribution $N(0,1)$,

$$a_{klk^*l^*} = \frac{Z'_{kl}(I_N - Q)Z_{k^*l^*}}{\sqrt{Z'_{kl}(I_N - Q)Z_{kl}}\sqrt{Z'_{k^*l^*}(I_N - Q)Z_{k^*l^*}}}, \quad (4)$$

and

$$\lambda = \sqrt{\frac{Z'_{kl}(I_N - Q)Z_{kl}}{2}} \;\; \rho.$$

3 Designs for searching one two-factor interaction

Two new classes of balanced designs for estimating the general mean, all main effects and searching and estimating one possible non-negligible two-factor interactions are presented in this section. By balance we mean that the designs are orthogonal arrays of strength one. Since the probability distribution of searching probability is singular, in general, unbalanced designs do not afford a simple expression for searching probability.

One class of designs, D_1, is for any $m \geq 5$, while the other, D_2, is for $m = 2^p - 1$, $p \geq 3$. The plan of D_1 having $N = 2(m+1)$ runs is as follows,

$$D_1 \quad = \quad \left[\begin{array}{c} 0'_m \\ I_m \\ 1_m 1'_m - I_m \\ 1'_m \end{array} \right],$$

where 0_m is an $m \times 1$ vector with all elements zero, 1_m is an $m \times 1$ vector with all elements unity, and I_m is the identity matrix of order m. We first derive the search probability for design D_1 using expression (3) before presenting design D_2. The following Lemmas 1 - 4 are needed for derivation of the search probability.

Lemma 1.
(i) $1'_N Z = 0'_m$,
(ii) $1'_N Z_{kl} = N - 8$, $1 \leq k < l \leq m$,
(iii) $Z'Z = 8I_m + (N-8)1_m 1'_m$,
(iv) $Z'Z_{kl} = 0_m$, $1 \leq k < l \leq m$.

Proof. With respect to the factors F_k, F_l, F_{k^*} and F_{l^*}, the treatment combinations 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111 appear in design D_1 with frequencies shown in Table 1 below. Moreover, the table provides the entries of the columns Z_k, Z_l, Z_{k^*}, Z_{l^*}, Z_{kl} and $Z_{k^*l^*}$ corresponding to the above treatment combinations. The lemma then follows easily.

Table 1

Treatment combination (1)	Frequency (2)	Z_k (3)	Z_l (4)	Z_{k^*} (5)	Z_{l^*} (6)	Z_{kl} (7)	Z_{kl^*} (8)	Z_{ll^*} (9)	$Z_{k^*l^*}$ (10)
0000	N/2-4	-1	-1	-1	-1	1	1	1	1
0001	1	-1	-1	-1	1	1	-1	-1	-1
0010	1	-1	-1	1	-1	1	1	1	-1
0100	1	-1	1	-1	-1	-1	1	-1	1
1000	1	1	-1	-1	-1	-1	-1	1	1
1110	1	1	1	1	-1	1	-1	-1	-1
1101	1	1	1	-1	1	1	1	1	-1
1011	1	1	-1	1	1	-1	1	-1	1
0111	1	-1	1	1	1	-1	-1	1	1
1111	N/2-4	1	1	1	1	1	1	1	1

Lemma 2. *For* $1 \leq k < l \leq m, 1 \leq k^* < l^* \leq m, (k,l) \neq (k^*, l^*)$,

$$Z'_{kl} Z_{k^*l^*} \quad = \quad \begin{cases} N - 16, & \textit{if } k, k^*, l, \textit{ and } l^* \textit{ are all distinct}, \\ N - 8, & \textit{otherwise}. \end{cases}$$

Proof. The lemma follows from columns (2), (7)-(10) of Table 1.

Lemma 3. *For* $1 \le k < l \le m$, $Z'_{kl}(I_N - Q)Z_{kl} = (16N - 64)/N$.

Proof. We have $Z'_{kl}QZ_{kl} = Z'_{kl}X(X'X)^{-1}X'Z_{kl}$. Also,

$$Z'_{kl}X = Z'_{kl}\begin{pmatrix} 1_N, & Z \end{pmatrix} = \begin{pmatrix} Z'_{kl}1_N, & Z'_{kl}Z \end{pmatrix} = \begin{pmatrix} N-8, & 0'_m \end{pmatrix}.$$

Thus,

$$\begin{aligned} Z'_{kl}(I_N - Q)Z_{kl} &= Z'_{kl}Z_{kl} - \begin{pmatrix} N-8, & 0'_m \end{pmatrix} \begin{bmatrix} \frac{1}{N} & 0'_m \\ 0_m & (8I_m + (N-8)1_m1'_m)^{-1} \end{bmatrix} \begin{pmatrix} N-8 \\ 0_m \end{pmatrix} \\ &= N - \frac{(N-8)^2}{N} \\ &= \frac{16N - 64}{N}. \end{aligned}$$

Hence the lemma.

Lemma 4. *For* $1 \le k < l \le m, 1 \le k^* < l^* \le m, (k,l) \ne (k^*, l^*)$, *we have*

$$Z'_{kl}(I_N - Q)Z_{k^*l^*} = \begin{cases} -\frac{64}{N}, & \text{if } k, k^*, l, \text{ and } l^* \text{ are all distinct,} \\ \frac{8(N-8)}{N}, & \text{otherwise.} \end{cases}.$$

Proof. The lemma follows from Lemmas 2 and 3.
Finally, the searching probability for design D_1 is given by the following theorem, the proof of which follows from expressions (3) and (4), and Lemmas 3 - 4.

Theorem 1 *The searching probability of one non-negligible two-factor interaction for design* D_1 *is* $P_d = min\{P_1, P_2\}$, *where*

$$\begin{aligned} P_1 &= 1 - \Phi(2\rho) - \Phi\left(2\rho\sqrt{\frac{3N-16}{N}}\right) + 2\Phi(2\rho)\Phi\left(2\rho\sqrt{\frac{3N-16}{N}}\right), \\ P_2 &= 1 - \Phi(2\rho\sqrt{2}) - \Phi\left(2\rho\sqrt{\frac{2(N-8)}{N}}\right) + 2\Phi(2\rho\sqrt{2})\Phi\left(2\rho\sqrt{\frac{2(N-8)}{N}}\right). \end{aligned}$$

We now present the other class of balanced search designs, namely D_2, for $m = 2^p - 1$, $p \ge 3$. Let H denote a Hadamard matrix of order 2^p, $p \ge 3$. Without loss of generality, suppose the first column of H is a column of all ones and H_1 is obtained from H by deleting its first column. Let H_0 be obtained by replacing -1's by zeros in H_1. Then, the plan of D_2 with $m = 2^p - 1$ and $N = 3(m+1)$ is obtained by appending H_0 to the plan of design D_1, i.e.

$$D_2 = \begin{bmatrix} H_0 \\ 0'_m \\ I_m \\ 1_m 1'_m - I_m \\ 1'_m \end{bmatrix}.$$

In fact, the plan of D_2 is obtained by adding one run to the unbalanced design of Mukerjee and Chatterjee (1994). Without loss of generality, let the columns of H_1 be arranged such that the first p columns, denoted by h_1, h_2, ..., h_p, are such that its remaining 2^p-1-p columns can be obtained by taking the Hadamard product of two or more of these p columns. Let Ω be the set of all non-null binary p-tuples $x = (x_1 \ldots x_p)$, $x_i = 0$ or 1, $i = 1, ..., p$, and let h_x denote a typical column of H_1. Then, we can establish a one-to-one relationship with the elements of Ω and the columns of H_1, i.e. $h_x = h_{x_1} * h_{x_2} * ... * h_{x_p}$, where $*$ denotes the Hadamard product, and

$$h_{x_i} = \begin{cases} h_i & \text{if } x_i = 1 \\ 1_{m+1} & \text{if } x_i = 0 \end{cases}.$$

Lemma 5.
(i) $1'_N Z = 0'_m$,
(ii) $1'_N Z_{kl} = 2N/3 - 8$, $1 \leq k < l \leq m$,
(iii) $Z'Z = (N/3+8)I_m + (2N/3-8)1_m 1'_m$.

Proof. With respect to factors F_k and F_l, the treatment combinations 00, 01, 10, 11 appear in design D_2 with frequencies as shown in Table 2 below. The table also provides entries of the columns Z_k, Z_l and Z_{kl} respectively corresponding to the above treatment combinations.

Table 2

Treatment combination	Frequency	Z_k	Z_l	Z_{kl}
00	5N/12-2	-1	-1	1
01	N/12+2	-1	1	-1
10	N/12+2	1	-1	-1
11	5N/12-2	1	1	1

The lemma then follows easily.

Lemma 6. *For any $1 \leq k < l \leq m$, with respect to the first $N/3$ treatment combinations of design D_2, let the first $N/3$ elements of Z_k and Z_l be given by the columns $h_{x_{11}} * \ldots * h_{x_{1p}}$ and $h_{x_{21}} * \ldots * h_{x_{2p}}$ respectively. Then $Z_j' Z_{kl} = N/3$ for exactly one column, say $j = t$, and $Z_j' Z_{kl} = 0$ for $j \neq t$. Further, the first $N/3$ elements of Z_t are given by the column $h_{x_1} * \ldots * h_{x_p}$, where for $1 \leq i \leq p$, $x_i = (x_{1i} + x_{2i})$, mod (2).*

Proof. Case 1. It is readily seen from Table 2 that $Z_j' Z_{kl} = 0$ for $j = k$ or l.
Case 2. For all values of $j \neq k, l$, except one value, say t, the treatment combinations $000, 001, 010, 011, 100, 101, 110, 111$ with respect to the factors F_j, F_k and F_l appear in design D_2 with frequencies shown in Tabe 3. Corresponding elements of the columns Z_j, Z_k, Z_l and Z_{kl} are also shown in Table 3,

Table 3

Treatment combination	Frequency	Z_j	Z_k	Z_l	Z_{kl}
000	$a_1 + N/3$-3	-1	-1	-1	1
001	$a_2 + 1$	-1	-1	1	-1
010	$a_3 + 1$	-1	1	-1	-1
011	$a_4 + 1$	-1	1	1	1
100	$a_5 + 1$	1	-1	-1	1
101	$a_6 + 1$	1	-1	1	-1
110	$a_7 + 1$	1	1	-1	-1
111	$a_8 + N/3$-3	1	1	1	1

where $a_1 + a_4 + a_6 + a_7 = a_2 + a_3 + a_5 + a_8$. It is then easy to observe that $Z_j' Z_{kl} = 0$.
Case 3. Finally, for $j = t$, it can be verified that the first $N/3$ elements of the columns Z_t and Z_{kl} are identical and the remaining elements of the two columns are as shown in Table 1. Thus, $Z_t' Z_{kl} = N/3$. Hence the lemma.

Lemma 7. *For $1 \leq k < l \leq m, 1 \leq k^* < l^* \leq m, (k, l) \neq (k^*, l^*)$, with respect to the first $N/3$ treatment combinations of D_2, let the elements of Z_k, Z_l, Z_{k^*}, and Z_{l^*} be given by the columns $h_{x_{11}} * \ldots * h_{x_{1p}}$, $h_{x_{21}} * \ldots * h_{x_{2p}}$, $h_{x_{31}} * \ldots * h_{x_{3p}}$, and $h_{x_{41}} * \ldots * h_{x_{4p}}$ respectively. Consider the following three cases:*

(a) all of k, k^*, l, and l^* are not distinct,

(b) k, k^*, l, *and* l^* *are all distinct and for* $1 \le i \le p$, $(x_{1i} + x_{2i}) = (x_{3i} + x_{4i})$, *mod* (2), *and*

(c) k, k^*, l, *and* l^* *are all distinct and for* $1 \le i \le p$, $(x_{1i} + x_{2i}) \neq (x_{3i} + x_{4i})$, *mod* (2).

Then,

$$Z'_{kl}Z_{k^*l^*} = \begin{cases} 2N/3 - 8, & \text{for (a)} \\ N - 16, & \text{for (b)} \\ 2N/3 - 16 & \text{for (c)} \end{cases} .$$

Proof. The proof of case (a) follows from Table 2. Now, if (b) holds, then the first $N/3$ elements of Z_{kl} and $Z_{k^*l^*}$ are identical and their remaining elements are as shown in Table 1. Thus, $Z'_{kl}Z_{k^*l^*} = N - 16$. Finally, if (c) holds, then the first $N/3$ elements of Z_{kl} and $Z_{k^*l^*}$ are orthogonal to each other and their remaining elements are as shown in Table 3. Thus, $Z'_{kl}Z_{k^*l^*} = 2N/3 - 16$. Hence the lemma.

Lemma 8. *For* $1 \le k < l \le m$, $Z'_{kl}(I_N - Q)Z_{kl} = c_N$, *where*

$$c_N = N - \frac{(2N-24)^2}{9N} - \frac{N^2\{(N+24) + (m-1)(2N-24)\}}{3(N+24)\{(N+24) + m(2N-24)\}} .$$

Proof. Let for the first $N/3$ treatment combinations of the design d, the elements of Z_k and Z_l be given by the columns $h_{x_{11}} * \ldots * h_{x_{1p}}$ and $h_{x_{21}} * \ldots * h_{x_{2p}}$ respectively, and let $x_{1i} + x_{2i} = x_i$, *mod* (2), $1 \le i \le p$. Then, for the first $N/3$ treatment combinations, the elements of Z_{kl} are given by the column $h_{x_1} * \ldots * h_{x_p}$. We have $Z'_{kl}QZ_{kl} = Z'_{kl}X(X'X)^{-1}X'Z_{kl}$. Now, $Z'_{kl}X = Z'_{kl}[1_N, Z] = (Z'_{kl}1_N, Z'_{kl}Z) = (2N/3 - 8, \delta'_{kl})$, where exactly one element of δ_{kl} equals $N/3$ and its remaining elements are zeros. The non-zero element of δ_{kl} occurs at the position occupied by the column $h_{x_1} * \ldots * h_{x_p}$ in the matrix Z. Therefore,

$$Z'_{kl}(I_N - Q)Z_{kl} = Z'_{kl}Z_{kl} - (2N/3 - 8, \quad \delta'_{kl}) \begin{bmatrix} \frac{1}{N} & 0'_m \\ 0_m & A^{-1} \end{bmatrix} \begin{pmatrix} 2N/3 - 8 \\ \delta_{kl} \end{pmatrix},$$

where $A = (N/3 + 8)I_m + (2N/3 - 8)1_m1'_m$. The lemma then follows by simplifying the above expression using $A^{-1} = (3/(N+24))I_m - \{3(2N-24)/((N+24)((N+24) + (2N-24)m))\}1_m1'_m$.

Lemma 9. *For* $1 \leq k < l \leq m, 1 \leq k^* < l^* \leq m, (k,l) \neq (k^*, l^*)$,

$$Z'_{kl}(I_N - Q)Z_{k^*l^*} = \begin{cases} g_{1N}, & \text{if (a) of Lemma 7 holds,} \\ g_{2N}, & \text{if (b) of Lemma 7 holds, and} \\ g_{3N} & \text{if (c) of Lemma 7 holds,} \end{cases}$$

where

$$\begin{aligned} g_{1N} &= \left(\frac{2N}{3} - 8\right) - \left(\frac{2N}{3} - 8\right)^2 /N + \frac{N^2(2N-24)}{3(N+24)\{(N+24)+m(2N-24)\}}, \\ g_{2N} &= (N-16) - \left(\frac{2N}{3} - 8\right)^2 /N - \frac{N^2\{(N+24)+(m-1)(2N-24)\}}{3(N+24)\{(N+24)+m(2N-24)\}}, \\ g_{3N} &= \left(\frac{2N}{3} - 16\right) - \left(\frac{2N}{3} - 8\right)^2 /N + \frac{N^2(2N-24)}{3(N+24)\{(N+24)+m(2N-24)\}}. \end{aligned}$$

Proof. For the sake of brevity, we indicate here the proof of $k = k^*$. Other cases can be established similarly.

$$\begin{aligned} Z'_{kl}(I_N - Q)Z_{kl^*} &= Z'_{kl}Z_{kl^*} - (2N/3 - 8,\ \delta'_{kl}) \begin{bmatrix} \frac{1}{N} & 0'_m \\ 0_m & A^{-1} \end{bmatrix} \begin{pmatrix} 2N/3 - 8 \\ \delta_{kl^*} \end{pmatrix} \\ &= (\frac{2N}{3} - 8) - (\frac{2N}{3} - 8)^2/N + \frac{N^2(2N-24)}{3(N+24)\{(N+24)+m(2N-24)\}}. \end{aligned}$$

Finally, the searching probability for design D_2 is given by Theorem 2, the proof of which follows after consideration simplification of expression (3) using Lemmas 8 and 9.

Theorem 2 *The searching probability of one non-negligible two-factor interaction for design* D_2 *is* $P_d = min\{P_1, P_2, P_3\}$, *where*

$$\begin{aligned} P_1 &= 1 - \Phi\left(\rho\sqrt{(c_N - g_{1N})/2}\right) - \Phi\left(\rho\sqrt{(c_N + g_{1N})/2}\right) + \\ &\quad 2\Phi\left(\rho\sqrt{(c_N - g_{1N})/2}\right)\Phi\left(\rho\sqrt{(c_N + g_{1N})/2}\right), \\ P_2 &= 1 - \Phi\left(\rho\sqrt{(c_N - g_{2N})/2}\right) - \Phi\left(\rho\sqrt{(c_N + g_{2N})/2}\right) + \\ &\quad 2\Phi\left(\rho\sqrt{(c_N - g_{2N})/2}\right)\Phi\left(\rho\sqrt{(c_N + g_{2N})/2}\right), \\ P_3 &= 1 - \Phi\left(\rho\sqrt{(c_N - g_{3N})/2}\right) - \Phi\left(\rho\sqrt{(c_N + g_{3N})/2}\right) + \\ &\quad 2\Phi\left(\rho\sqrt{(c_N - g_{3N})/2}\right)\Phi\left(\rho\sqrt{(c_N + g_{3N})/2}\right). \end{aligned}$$

We now present searching probabilities for the two classes of designs in Tables 4 and 5 for number of factors $m = 7, 15, 31$, for 8 different values of ρ. It is observed that searching probability is invariant to different values of the pair (θ_{kl}, σ) for a given ρ. In fact, the plan of D_1 is obtained by adding two runs to the unbalanced search design given by Shirakura and Tazawa (1991). Our motivation was that the addition of two extra runs would result in higher search probabilities compared to the design of Shirakura and Tazawa (1991). Although search probabilities did improve, the gain was not found to be substantial.

Table 4
Searching probabilities for design D_1 using the expression of Shirakura *et al.* (1996)

m	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$	$\rho = 1.2$	$\rho = 1.4$	$\rho = 1.6$
7	0.5666	0.7138	0.8504	0.9347	0.9750	0.9915	0.9974	0.9993
15	0.5735	0.7288	0.8627	0.9401	0.9765	0.9917	0.9974	0.9993
31	0.5766	0.7349	0.8670	0.9417	0.9768	0.9918	0.9974	0.9993

Table 5
Searching probabilities for design D_2 using the expression of Shirakura *et al.* (1996)

m	$\rho = 0.2$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.8$	$\rho = 1.0$	$\rho = 1.2$	$\rho = 1.4$	$\rho = 1.6$
7	0.5954	0.7793	0.9122	0.9721	0.9926	0.9983	0.9997	1.0000
15	0.6224	0.8217	0.9357	0.9804	0.9951	0.9990	0.9998	1.0000
31	0.6506	0.8494	0.9459	0.9841	0.9964	0.9994	0.9999	1.0000

It is clear from the above two tables that searching probability for designs D_1 and D_2 lies between 0.5 and 1.0. Also, it is an increasing function of the number of factors. A similar increasing pattern is generally observed for the design of Shirakura and Tazwa (1991) as well. This increasing pattern in searching probability is counter intuitive. The possible number of models among which the correct model is to be searched for increases with an increase in the number of factors. It would thus seem that searching probability should decrease as the number of possible models becomes larger. Therefore, the searching probability was evaluated using simulation, the results of which are discussed in the next section.

4 Searching probability through simulation

For the searching probability as defined in section 2 for the method proposed by Srivastava (1975), we consider the event, $E^{kl}_{k^*l^*} = \{SSE(kl) < SSE(k^*l^*)\}$, for all possible $(k,l) \neq (k^*,l^*) \in H(1)$. Then, the searching probability can be defined by

$$P_{dT} = \min_{(k,l)\in H(1)} \mathrm{P}\left(\bigcap_{(k^*,l^*)(\neq(k,l))\in H(1)\}} E^{kl}_{k^*l^*}\right). \qquad (5)$$

Here we calculate P_{dT} using simulation for the two designs D_1 and D_2. The algorithm used for simulation is as follows. Considering a particular model $M(kl)$, we generate an observational vector y. Using this observational vector y, we calculate $SSE(k^*l^*)$ for all $\theta_{k^*l^*} \in W$. These two steps are repeated $10,000$ times and let p_{kl} be the proportion of times $SSE(kl) < SSE(k^*l^*)$, where $\theta_{k^*l^*}(\neq \theta_{kl}) \in W$. Finally, the minimum value of p_{kl} provides an estimate of P_{dT}, where the minimum is taken over all possible models for $\theta_{kl} \in W$.

The values of P_{dT} were calculated for the values of m and ρ considered in Tables 4 and 5. The values of P_{dT} were also observed to be invariant to different choices of the pair (θ_{kl}, σ) for a given ρ. These results are presented in the tables below.

Table 6
Simulated search probabilities for design D_1

m	$\rho=0.2$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=1.0$	$\rho=1.2$	$\rho=1.4$	$\rho=1.6$
7	0.0495	0.1503	0.2986	0.4956	0.6987	0.8397	0.9325	0.9674
15	0.0000	0.0397	0.1506	0.3691	0.6125	0.7932	0.9296	0.9502
31	0.0000	0.0000	0.0908	0.2419	0.4933	0.7874	0.8293	0.8710

Table 7
Simulated search probabilities for design D_2

m	$\rho=0.2$	$\rho=0.4$	$\rho=0.6$	$\rho=0.8$	$\rho=1.0$	$\rho=1.2$	$\rho=1.4$	$\rho=1.6$
7	0.0537	0.2237	0.5431	0.7621	0.8891	0.9643	0.9843	0.9998
15	0.0108	0.1324	0.3358	0.6464	0.8631	0.9496	0.9700	0.9987
31	0.0000	0.0226	0.1344	0.3452	0.7871	0.8710	0.8968	0.9980

A comparison of Tables 4 and 5 with the above simulation results show that the expression (3) given by Shirakura *et al.* (1996) substantially over-estimates the true searching probability P_{dT}, unless effect size is not small relative to error variance. In fact, Shirakura *et al.* (1996) approximate searching probability with the upper bound given by

$$P_d = \min_{1 \leq k < l \leq m} \min_{1 \leq k^* < l^* \leq m, (k^*, l^*) \neq (k,l)} \mathrm{P}\left(E_{k^* l^*}^{kl}\right).$$

In general, the upper bound may not be close to the true searching probability. The above simulation results also show that it is possible for searching probability to be less than 0.5, even close to zero. Thus, the observation of Ghosh and Teschmacher (2002) that searching probability lies between 0.5 and 1.0 does not hold for all search designs.

Although the number of runs in design D_2 is considerable larger than design D_1, it does not afford a substantial advantage over design D_1 in terms of searching probability for $s = 1$. A major advantage of D_2 is that it is capable of searching and estimating two possibly present two-factor interactions in addition to estimating the general mean and all the main effects. For design D_2, an expression for search probability similar to (3) for $s = 2$ is not presently available. Therefore, we evaluate searching probability for design D_2 for the case of $s = 2$ using simulation. Define an event, $E_{h_2^*}^{h_2} = \{SSE(h_2) < SSE(h_2^*)\}$, for all possible $h_2 \neq h_2^* \in H(2)$. As discussed in section 2, according to the searching method proposed by Srivastava (1975), the searching probability of a design for $s = 2$ is defined by

$$P_{dT2} = \min_{h_2 \in H(2)} \mathrm{P}\left(\bigcap_{h_2^*(\neq h_2) \in H(2)} E_{h_2^*}^{h_2}\right).$$

Using a similar simulation procedure as described above, search probabilities for $m = 7, 15$, and $\theta_{kl}/\sigma = \theta_{k'l'}/\sigma = \rho$, are presented in the table below, where θ_{kl} and $\theta_{k'l'}$ denote two possibly non-negligible two-factor interactions for $s = 2$.

Table 8

Simulated search probabilities of design D_2 for $s = 2$.

ρ	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6
m=7	0.0186	0.0805	0.2833	0.5776	0.8233	0.9410	0.9800	0.9976
m=15	0.0038	0.0667	0.1987	0.4949	0.7193	0.8320	0.8941	0.9276

Acknowledgements : The authors thank a referee for critical reading of the paper and for making useful comments.

References

Chatterjee, K., Deng, L. Y. and Lin, D.K.J. (2002). Two-level search design for main-effect plus two plan. *Metrika* **54**, 233-245.

Ghosh, S.(1980). On main effect plus two plans for 2^m factorials. *Ann. Statist.* **8**, 922-930.

Ghosh, S.(1981). On some new search designs for 2^m factorials. *J. Statist. Plann. Inf.* **5**, 381-389.

Ghosh, S. and Teschmacher, L. (2002). Comparisons of search designs using search probabilities. *J. Statist. Plann. Inf.* **B 104**, 439-458.

Mukerjee, R. and Chatterjee, K. (1994). An application of Hadamard matrices for the construction of main effect plus two plans for 2^m factorials. *Utilitas Math.* **45**, 213-218.

Ohnishi, T. and Shirakura, T.(1985). Search designs for 2^m factorial experiments. *J. Statist. Plann. Inf.* **11**, 241-245.

Shirakura, T. (1991). Main effect plus one or two plans for 2^m factorials. *J. Statist. Plann. Inf.* **27**, 65-75.

Shirakura, T. and Tazawa, S. (1991). Series of main effect plus one or two plans for 2^m factorials when three-factor and higher order interactions are negligible. *J. Japan Statist. Soc.* **21**, 211-219.

Shirakura, T., Takahashi, T. and Srivastatava, J.N. (1996). Searching probabilities for non zero effects in search designs for the noisy case. *Ann. Stat.* **24**, 2560-2568.

Srivastava, J.N. (1975). Designs for searching non negligible effects. *A Survey of Statistical Design and Linear Models* (J.N. Srivastava, ed.), North-Holland:Amsterdam, 507-519.

Srivastava, J.N. (1992). A 2^8 factorial main effect plus two plan with good revealing power. *Sankhya* **54**, 461-474.

Kashinath Chatterjee
Department of Statistics
Visva-Bharati, Santiniketan, India
Email:kashinathchatterjee@gmail.com

Sudhir Gupta
Division of Statistics
Northern Illinois University, Illinois
Email : sudhir@math.niu.edu

Angshuman Sarkar
Department of Statistics
Visva-Bharati, Santiniketan, India
Email: angshuman.sarkar@gmail.com