Journal of Pharmacognosy and Phytochemistry
Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2016; 5(4): 370-374
Received: 15-05-2016
Accepted: 17-06-2016

Basudev Kole
Department of Statistics Mathematics and Computer Application, Bihar Agricultural College, Bihar Agricultural University, Sabour, Bhagalpur Bihar, India

Corresponding Author:
Basudev Kole
Department of Statistics Mathematics and Computer Application, Bihar Agricultural College, Bihar Agricultural University, Sabour, Bhagalpur Bihar, India

Construction of optimal Multi-level supersaturated designs

Basudev Kole

Abstract
A new method of constructing multi-level supersaturated designs has been proposed. The method of construction is based on the association between the rows of the design. An algorithm has been developed to construct such designs. The designs constructed by this method are both f_{NOD} - optimal and χ^2 - optimal. A catalogue of 11 optimal multi-level supersaturated designs have also prepared. AMS Subject Classification: 62K15, 62K05, 62K99

Keywords: SSD, optimality criteria, supersaturated design

1. Introduction

Supersaturated Designs (SSDs) are fractional factorial designs in which the number of design runs is smaller than the degrees of freedom for the main effects and the intercept term. These are helpful in screening experiments where experimentation is expensive and the number of factors is large. SSDs were initiated by Satterthwaite (1959) [27] and studied further by Watson (1961). Booth and Cox (1962) [3] were amongst the earliest researchers to study SSDs in a systematic manner. Lin (1993) revisited construction of SSDs and since then there has been a burst of activity in obtaining efficient two-level SSDs. Among others some important references are Lin (1995) [21], Nguyen (1996) [25], Li and Wu (1997) [18] and Gupta *et al*. (2008) [12-13]. These authors have restricted their construction to balanced two-level SSDs.

Although two-level SSDs are considered in screening experiments, designs with multi-level factors are often requested in agricultural, industrial and scientific experimentation for exploring nonlinear effects of the factors. It is undesirable to reduce the factor levels to two as it may result in severe loss in information. Research on multi-level SSDs include Yamada and Lin (1999) [30], Fang *et al*. (2000, 2002, 2004) [6, 7, 8], Lu and Sun (2001) [23], Lu and Zheng (2003) [24], Aggarwal and Gupta (2004) [1], Xu and Wu (2005) [16], Georgiou *et al*. (2006) [11], Liu *et al*. (2007) [22], and Chen and Liu (2008) [5]. In an experimental situation when several factors have same number of levels and one or two factors have different number of levels than the rest of the factors, mixed level SSDs become useful. Some recent references on mixed level SSDs are Fang *et al*. (2003) [7], Li *et al*. (2004), Koukouvinos and Mantas (2005) [16], Ai *et al*. (2007), Tang *et al*. (2007), Gupta *et al*. (2008a, 2009) [12-13] and Chen and Liu (2008) [5].

Heavlin and Finnegan (1993) [10] were first to propose generation of an SSD through computeralgorithm. In their column wise exchange algorithm, levels of one or more factors are changed. Since then a number of computer algorithms have become available to generate efficient SSDs (see e.g., Nguyen (1996) [25], Li and Wu (1997) [18], Lejeune (2003) [17], Ryan and Bulutoglu (2007) [26] and Gupta *et al*. (2008b, 2009 and 2010)] [12-13, 15]. All these algorithms depend on column-wise exchange of symbols to generate efficient SSDs.

An SSD (two-level, multi-level or mixed - level) is said to be balanced if every level (or symbol) of each factor (or column) appears equally often in the design runs. A balanced multi-level SSD is said to be $v(<m)$ associate design when the coincidences between the levels of the m factors in any two rows of the design is v.

This paper proposes an algorithm to generate v associate balanced multi-level SSDs. $E(f_{NOD})$ and $E(\chi^2)$ criteria have been used to measure the non-orthogonality of the generated SSDs. Lower bound to $E(f_{NOD})$ given by Fang *et al*. (2004) [8] and to $E(\chi^2)$ given by Ai *et al*. (2007) [28] have been used to measure the design optimality. A simple method of construction of $E(f_{NOD})$ and $E(\chi^2)$ optimal multi-level and mixed-level SSDs has also been given.

Some preliminaries are given in Section 2. Lower bounds to the two criteria and the lemma used are also given in Section 2. In Section 3 relation between v, number of association between any pair of rows, and m, number of columns for a given SSD $(n; q^m)$, are obtained. The algorithm developed along with its implementation is also given in Section 3.

All the designs generated by the algorithm are both f_{NOD} -optimal and χ^2 -optimal. The algorithm also ensures that no two columns of the design are fully aliased in the sense that no column of the design can be obtained from any other column by permuting levels. Some concluding remarks and a catalogue of optimal v associate multi-level SSDs is given in Section 4.

2. Some preliminaries

Let $d(n; q_1, q_2, \cdots, q_m)$ denote a mixed-level SSD D having n runs and m factors with levels $q_1, q_2, \ldots, q_m$. The design D can be represented as an $n \times m$ array in which each row represents a design run, each column represents a factor and the u^{th} column takes values from a set $Q_u = \{1, 2, \cdots, q_u\}$, $u = 1, 2, \cdots, m.$ Let $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^u, \cdots, \mathbf{x}^m$ denote the columns of D, with $\mathbf{x}^u$ being an n-dimensional column vector consisting of symbols from the set Q_u, $u = 1, 2, \cdots, m.$ The $n \times m$ matrix D can be written as $\mathbf{D} = [\mathbf{x}^1 \quad \mathbf{x}^2 \cdots \mathbf{x}^j \cdots \mathbf{x}^m]$.

If the design has elements $1, 2, \cdots, q_u$ at $\mathbf{x}^u$ such that q_u elements appear in this column equally often (n/q_u times each), then we call it as a balanced SSD. Thus for a balanced mixed level SSD $q_1, \cdots, q_m$ are positive divisors of n. In the literature, these designs have also been termed as U- type SSD. We shall, however, call these designs as balanced designs. Further a balanced design is called an orthogonal design, if for every pair of design columns, all of their level combinations appear equally often.

Fang *et al*. (2003) [7] used $\text{Ave}(f^2)$ to define a measure of column non-orthogonality for mixed-level SSDs, and called it as $E(f_{NOD})$. They defined f_{NOD}^{ij} as

$$f_{NOD}^{ij} = \sum_{u=1}^{q_i} \sum_{v=1}^{q_j} \left(n_{uv}^{ij} - \frac{n}{q_i q_j} \right)^2$$

where n_{uv}^{ij} is the number of (u, v)-pairs in $(\mathbf{x}^i, \mathbf{x}^j)$, and $n/(q_i q_j)$ stands for the average frequency of the level-combinations in each pair of columns $\mathbf{x}^i$ and $\mathbf{x}^j$. Here the subscript *NOD* stands *for non-orthogonality of the design*, and q_i and q_j denote the levels of the factors in columns x^i and x^j, respectively.

A criterion $E(f_{NOD})$ for overall non-orthogonality of the design is defined as minimizing

$$E(f_{NOD}) = \sum_{1 \le i < j \le m} f_{NOD}^{ij} \Big/ \binom{m}{2}.$$

For a mixed-level $(n; q_1, q_2, \cdots, q_m)$-design D, let $\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^m$ be its m columns. For every pair of columns $(\mathbf{x}^i, \mathbf{x}^j)$, Yamada and Lin (1999) [30] defined the following χ^2 value:

$$\chi^2(\mathbf{x}^i, \mathbf{x}^j) = \frac{q_i q_j}{n} \sum_{u=1}^{q_i} \sum_{v=1}^{q_j} \left(n_{uv}^{ij} - \frac{n}{q_i q_j} \right)^2$$

Obviously, the value $\chi^2(\mathbf{x}^i, \mathbf{x}^j)$ measures the non-orthogonality between two columns $\mathbf{x}^i$ and $\mathbf{x}^j$. Then the $E(\chi^2)$ value defined as

$$E(\chi^2) = \frac{2}{m(m-1)} \sum_{1 \le i < j \le m} \chi^2(\mathbf{x}^i, \mathbf{x}^j)$$

can be used to evaluate the overall non-orthogonality between the columns of D. An SSD is called $E(\chi^2)$-optimal if it minimizes the value of $E(\chi^2)$.

For a balanced mixed level $(n; q_1, q_2, \cdots, q_m)$ -SSD D, Fang *et al*. (2003) [7] gave a lower bound on E(f_{NOD}). Fang *et al*. (2004) [8] further improved the lower bound on E(f_{NOD}) and is given in the following theorem.

Theorem 2.1. [Fang *et al*., 2004] [8]. For any balanced SSD-$(n; q_1 \times q_2 \times \cdots \times q_m)$

$$E(f_{NOD}) \ge \frac{n(n-1)}{m(m-1)} \left[(\gamma + 1 - \psi)(\psi - \gamma) + \psi^2 \right] + C(n, q_1, \ldots\ldots\ldots, q_m) = L[E(f_{NOD})]$$

Where

$$C(n; q_1, q_2, \cdots, q_m) = \frac{nm}{m-1} - \frac{1}{m(m-1)} \left(\sum_{i=1}^{m} \frac{n^2}{q_i} + \sum_{i,j=1, j \ne i}^{m} \frac{n^2}{q_i q_j} \right)$$

depends on X only through $n, q_1, q_2, \cdots, q_m$.

Here $\psi = \frac{\sum_{i=1}^{m} n/q_i - m}{(n-1)}$, $\gamma = [\psi]$, and $[x]$ denotes the largest integer in x.

For a balanced design, Ai *et al*. (2007) [28] obtained lower bound on the value of $E(\chi^2)$ as given in Theorem 2.2.

Theorem 2.2. (Ai, Fang and He, 2007) [2]. For any balanced SSD- $\left(n; q_1 \times q_2 \times \cdots \times q_m\right)$,

$$E\left(\chi^2\right) \geq \frac{1}{m(m-1)(n-1)}\left(nm - \sum_{k=1}^{m} q_k\right)^2 + C_1\left(n; q_1, q_2, \cdots, q_m\right) = L\left[E\left(\chi^2\right)\right]$$

Where

$$C_1\left(n; q_1, q_2, \cdots, q_m\right) = \frac{1}{m(m-1)}\left[\begin{matrix}\left(\sum_{k=1}^{m} q_k\right)^2 - \\ n\sum_{k=1}^{m} q_k\end{matrix}\right] - n.$$

Lower bound L[$E(\chi^2)$], given by Ai *et al*. (2007) [28] and lower bound L[E(f_{NOD})], given by Fang *et al*. (2004) [8], can be used to ensure the optimality of the generated designs. For a design d generated through the proposed methods we define f_{NOD} -efficiency and χ^2 -efficiency as f_{NOD} -efficiency= L[E(f_{NOD})] /E_d(f_{NOD}), χ^2-efficiency= L[$E(\chi^2)$] / $E_d(\chi^2)$
where E_d(f_{NOD}) and $E_d(\chi^2)$ are the values of E(f_{NOD}) and $E(\chi^2)$, respectively, for design d. A design with efficiency one is an optimal design.
A mixed-level SSD with $q_1 = q_2 = q_3 = \ldots\ldots = q_m = q$ is a multi-level SSD. Therefore, the two criteria of non-orthogonality defined above can easily be used for multi-level SSDs also.

3. Construction of optimal multi-level SSDs

In this section we have given some method of constructing optimal multi-level SSDs. These method is a computer algorithm which generates *v* associate balanced multi-level SSDs.

Theorem 3.1 Consider a design d ϵ D(n, q^m) then for a given *v*, the association between any pair of rows, the number of factors (or columns) that need to be included in the design will satisfy the relation m=(n(n-1)/2)(v/ y) where y=(z(z-1)/2)q, z = n/q and $v = \sum_{j=1}^{m} I_j\,(a,b)$ and $I_j(a,b)$ is the is the indicator function defined as

$$I_j(a,b) = \begin{cases} 1, if\ the\ j^{th}\ entries\ of\ rows\ a\ and\ b\ are\ same \\ 0, if\ Otherwise \end{cases}$$

Proof: In every column of the design D(n, q^m), all the *q* levels appear, say *f*=*n*/*q* times each. Therefore, the number of association for each column will be $qf(f-1)/2$ and hence the total number of associations in the *m* columns of the design is $f(f-1)/2$. On the other hand if *v* denotes the association between any of the $n(n-1)/2$ pairs of rows in the design *d* then the total number of associations in the design is $vn(n-1)/2$. The proof is thus complete.

Proposed algorithm for constructing SSD- $(n; q^m, v)$

Step1: *Generate a balanced first column for given parameters*

1. Input *q* (the number of levels), *n* (the number of runs) and *v* (the number of associations allowed), where *n*=*tq* (*t*, a positive integer).
2. The algorithm checks the associations between all the $\xi = n(n-1)$ rows for this column. Algorithm also keeps these associations in separate place for further use.

Step 2: *Generation of another column*

1. Generate a random balanced column for the given parameters.
2. The algorithm checks the associations between the rows for this column.
3. Appends these associations with the earlier set of associations and calculate whether number of associations for any two rows is less than or equal to *v*.
4. If the number is less than or equal to *v* algorithm updates the associations after appending the associations for this column with the earlier associations and goes to Step 3; otherwise the algorithm goes back to (i) above.

Step 3: *Selection of new columns*

1. Check whether the new column generated in Step 2 is fully aliased with other columns of the design.
2. If the new column is aliased with any other columns of the design, discard this column otherwise juxtapose the new column with the earlier columns.
3. Repeat Step 2 for another new column unless the algorithm terminates.

Step 4: *Termination rule*

The algorithm terminates with the generation of an optimal design *d* when the number of columns generated satisfy the relationship with other design parameters given in Theorem 3.1.

Step 5: *Design parameters*

The algorithm calculates E_d(f_{NOD}) and $E_d(\chi^2)$ for the optimal design *d* using f_{NOD} and χ^2 criterion respectively.

4.1. Implementation of the algorithm

Suppose we want to construct an SSD (6; 3^5, 1). Theorem 3.1 shows that the relation between the number of columns with number of association is satisfied therefore we can generate an optimal SSD for this parameter using this method.
Using Step 1 of the algorithm the randomly generated balanced first column of the design **X** is [3 1 2 3 1 2]' having associations in (1, 4), (2, 5) and (3, 6) rows.
Using Step 2 of the algorithm randomly generated balanced column is [3 2 1 3 1 2] ' having associations in (1, 4), (2, 6) and (3, 5) rows. Appending these associations with the earlier set of associations we get number of associations for 1st row and 4th row is 2>v=1.
Therefore using (iv) of Step 2 the algorithm goes back to (i) of Step 2 and finally get the column [1 2 2 3 1 3] ' having associations in (1, 5), (2, 3) and (4, 6) rows where the number of associations after appending these associations are less than or equal to *v*.
Invoking Step 3 of the algorithm shows that this column in not aliased with other columns of X therefore this column is accepted. Using (iii) of this Step we get another 3 columns as [3 3 1 1 2 2] ', [1 3 2 3 2 1] ' and [1 2 1 3 3 2] '.
Now Step 4 of the algorithm shows that the number of columns generated satisfy the relation given in Theorem 3.1 therefore the algorithm terminates with the following optimal

$$SSD(6: 3^5, 1) = \begin{bmatrix} 3 & 1 & 3 & 1 & 1 \\ 1 & 2 & 3 & 3 & 2 \\ 2 & 2 & 1 & 2 & 1 \\ 3 & 3 & 1 & 3 & 3 \\ 1 & 1 & 2 & 2 & 3 \\ 2 & 3 & 2 & 1 & 2 \end{bmatrix}.$$

Finally using Step 5 the algorithm calculates $E_d(f_{NOD})$ =2.00 and $E_d(\chi^2)$ =3.00 for the generated optimal design.

4. Discussion

This article introduces a method of construction for optimal multi-level SSDs. The method is based on the association of levels in the rows of the design. For a design constructed through this method has a fixed number of associations and the design is both f_{NOD} and χ^2 -optimal. We have also defined an upper bound of columns for a given number of runs, given number of levels and number of associations.

One computer algorithm has also proposed to generate optimal SSDs using this method. The algorithm has constructed in such a way that no two columns of the designs generated are fully aliased. Using the algorithm, a catalogue of optimal designs generated is prepared and is given below in Table 4.1. The designs available in the catalogue have number of levels as 3, 4, 5 and number of runs5, 8, 9 and 10. The catalogue gives 11 designs in the restricted parametric range. The layout of these designs is available with the author.

Table 4.1: Catalogue of *v* associate optimal balanced multi-level SSDs

Design	$E(f_{NOD})$	$E(\chi^2)$
SSD(6; 3^5; 1)	2.00	3.00
SSD(6; 3^{10}; 2)	2.67	4.00
SSD(6; 3^{15}; 3)	2.86	4.29
SSD(8; 4^7; 1)	4.00	8.00
SSD(8; 4^{14}; 2)	4.62	9.23
SSD(8; 4^{21}; 3)	4.80	9.60
SSD(9; 3^8; 2)	2.57	2.57
SSD(9; 3^{12}; 3)	3.27	3.27
SSD(10; 5^9; 1)	6.00	15.00
SSD(10; 5^{18}; 2)	6.59	16.47
SSD(10; 5^{27}; 3)	6.77	16.92

References

1. Aggarwal ML, Gupta S. A new method of construction of multi-level supersaturated designs, Journal of Statistical Planning and Inference. 2004; 121:127-134.
2. Mingyao AI, Fang, Kai, Tai, Shuyuan He. E(χ^2)-optimal mixed-level supersaturated designs. Journal of Statistical Planning and Inference. 2007; 137:306-316.
3. Booth KHV, Cox DR. Some systematic supersaturated designs. Technometrics. 1962; 4:489-495.
4. Bulutoglu DA, Cheng CS. Construction of E(s^2)-optimal supersaturated designs. Annals of Statistics. 2004; 32:1662-1678.
5. Chen J, Liu MQ. Optimal mixed-level supersaturated design with general number of runs. Statistics and Probability Letters. 2008; 78:2496-2502.
6. Fang KT, Ge G, Liu MQ. Uniform supersaturated design and its construction, Science in China. 2002; 45:1080-1088.
7. Fang KT, Lin DKJ, Liu MQ. Optimal mixed-level supersaturated design. Metrika. 2003; 58:279-291.
8. Fang KT, Ge G, Liu MQ, Qin H. Combinatorial constructions for optimal supersaturated designs, Discrete Mathematics. 2004; 279:191-202.
9. Fang KT, Lin DKJ, Ma CX. On the construction of multi–level supersaturated designs. Journal of Statistical Planning and Inference. 2000; 86:239-252.
10. Heavlin WD, Finnegan GP. Adapting Experiments with Sequentially Processed Factors. Section of the Physical and Engirzeering Sciences. American Statistical Association, 1993, 44-49.
11. Georgiou S, Koulouvinos C, Mantas P. On multi-level supersaturated designs. Journal of Statistical Planning and Inference. 2006; 136:2805-2819.
12. Gupta VK, Prasad R, Bhar LM, Kole B. Supersaturated Designs for Asymmetrical factorial Experiments. Journal of Statistical Theory and practice. 2008a; 2:95-108.
13. Gupta VK, Prasad R, Kole B, Bhar LM. Computer-aided construction of efficient two level supersaturated designs. Journal of the Indian Society of Agricultural Statistics. 2008b; 62(2):183-194.
14. Gupta VK, Singh P, Kole B, Prasad R. Construction of efficient unbalanced mixed-level supersaturated designs. Statistics and Probability Letters. 2009; 79(22):2359-2366.
15. Gupta VK, Singh P, Kole B, Parsad R. Computer-aided construction of efficient multi-level supersaturated designs. Journal of Statistical Theory and Practice. 2010; 4:221-231.
16. Koukouvinos C, Mantas P. Construction of some E(f_{NOD}) optimal mixed-level supersaturated designs. Statistics and Probability Letters. 2005; 74(4):312-321.
17. Lejeune Miguel A. A coordinate-columnwise exchange algorithm for the construction of supersaturated, saturated, and non-saturated experimental designs. Modern mathematical, management, and statistical sciences. II: Advances in theory & application. Amer. J Math. Management Sci. 2003; 23:109-142.
18. Li WW, Wu CFJ. Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. Technometrics. 1997; 39:171-179.
19. Li PF, Liu MQ, Zhang RC. Some theory and the construction of mixed-level supersaturated designs. Statistics and Probability Letters. 2004; 69:105-116.
20. Lin DKJ. A new class of supersaturated designs. Technometrics. 1993; 35:28-31.
21. Lin DKJ. Generating systematic supersaturated designs. Technometrics. 1995; 37:213-225.
22. Liu Y, Liu M, Zhang R. Construction of multi-level supersaturated design via Kronecker product. Journal of Statistical Planning and Inference. 2007; 137:2984-2992.
23. Lu X, Sun Y. Supersaturated design with more than two levels. Chinese Annals of Mathematics Series B. 2001; 22:183-194.
24. Lu X, Hu W, Zheng Y. A systematic procedure in the construction of multi-level supersaturated design, Journal of Statistical Planning and Inference. 2003; 115:287-310.
25. Nguyen NK. An algorithmic approach to constructing supersaturated designs. Technometrics. 1996; 38:69-73.
26. Ryan KJ, Bulutoglu DA. E(s^2)-Optimal supersaturated designs with good minimax properties. Journal of Statistical Planning and Inference. 2007; 137:2250-2262.
27. Satterwaite F. Random balance Experimentation. Technometrics. 1959; 1:111-137.
28. Tang Yu, Ai M, Ge G, Fang KT. Optimal Mixed-level Supersaturated Designs and a New Class of Combinatorial Designs. Journal of Statistical Planning and Inference. 2007; 137:2249-2301.

29. Xu H, Wu CFJ. Construction of Optimal Multilevel Supersaturated Designs. UCLA, Department of Statistics, Electronic Publication. (http:// repositories. cdlib. orgluclastat), 2003.
30. Yamada S, Lin DKJ. Three level supersaturated designs. Statistics and Probability Letters. 1999; 45:31-39.