Journal of Statistical Planning and Inference 157–158 (2015) 100–107

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Extended mixed-level supersaturated designs

Hong Qin[a,*], Kashinath Chatterjee[b], Sutapa Ghosh[b]

[a] Faculty of Mathematics and Statistics, Central China Normal University, China
[b] Department of Statistics, Visva-Bharati University, Santiniketan, India

ARTICLE INFO

Article history:
Received 12 August 2013
Received in revised form 11 August 2014
Accepted 11 August 2014
Available online 16 September 2014

MSC:
62K15
62K10
62K99

Keywords:
Supersaturated design
$E(\chi^2)$ optimality criterion
Lower bound

ABSTRACT

This paper considers the study of the optimality of the extended design generated by adding few runs to an existing $E(\chi^2)$-optimal mixed-level supersaturated design. This paper covers the work of Gupta et al. (2010, 2012) on extended two-level and s-level supersaturated designs as two special cases. A lower bound to $E(\chi^2)$ has been obtained for the proposed designs. A small example is presented here attaining the lower bound.

1. Introduction

It is well cited in the literature that for an expensive experimentation with large number of factors, supersaturated designs (SSDs) are helpful. An SSD is a fractional factorial design whose run size is not large enough even for estimating the main effects represented by the columns of the design matrix. The design and analysis rely on the assumption that the number of relatively important effects are small.

Satterthwaite (1959) initiated the notion of SSD through random balance designs. Booth and Cox (1962) proposed an algorithm to construct systematic SSDs. Since then there has been a burst of activity in obtaining two-level SSDs. For an excellent review, one may refer to Kole et al. (2010).

There are many experimental settings where it is not desirable to reduce the factor levels to two as it would result in severe loss of information. In that case, designs with multi-level factors are required. The problem of generating optimal/efficient multi-level SSDs has been studied extensively in the literature. Some useful references on multi-level SSDs include Yamada and Lin (1999), Yamada et al. (1999), Fang et al. (2000), Lu and Sun (2001), Lu et al. (2003), Xu and Wu (2005), Georgious et al. (2003, 2006a,b), Liu et al. (2007) and Gupta et al. (2010).

It is to be remarked that we frequently come across situations where factors having different levels come into play. In that case mixed-level SSDs are required. Some useful references on mixed-level SSDs are Yamada and Matsui (2002), Yamada and Lin (2002), Fang et al. (2003, 2004), Li et al. (2004), Koukouvinos and Mantas (2005), Ai et al. (2007), Tang et al. (2007) and Gupta et al. (2009a,b, 2010).

Literature review on mixed-level SSDs reveal that $E(\chi^2)$, due to Yamada and Lin (1999), is a popular criterion which measures the overall non-orthogonality of the design. An SSD is said to be balanced if the levels of each factor appears

* Corresponding author.
E-mail address: qinhong@mail.ccnu.edu.cn (H. Qin).

http://dx.doi.org/10.1016/j.jspi.2014.08.004

equally often in the entire design. However, unbalanced two-level, multi-level and mixed-level SSDs have also been studied in the literature (see e.g., Gupta et al., 2009a,b, 2010).

Suppose that an experimenter begins an experiment with an $E(\chi^2)$-optimal column balanced SSD involving m factors having different levels and N runs. After the experiment is over or during the experimentation, some additional resources may be available and the experimenter can afford to include r, more runs to the chosen $E(\chi^2)$-optimal column balanced SSD. A natural question arises regarding the choice of the additional runs. How does the experimenter choose the runs and then augment these runs with the chosen design so as to get an extended $E(\chi^2)$-optimal SSD. Gupta et al. (2010, 2012) considered the same issue for two-level and s-level factors respectively.

The paper is organized as follows. In Section 2, a brief description of the notations used are given. Section 3 provides the main result related to the lower bound to $E(\chi^2)$ for the extended design. Also, as for illustration, an example of extended SSD is presented in this section. Concluding remarks are made in Section 4. Section 5 deals with proof of main results and some important lemmas for proving these main results of this paper.

2. Notations and preliminaries

Let $\mathcal{D}(N, s_1 \times s_2 \times \cdots \times s_m)$ be a class of N-run supersaturated designs with m factors $F_1, F_2, \ldots, F_m$ each at $s_1, s_2, \ldots, s_m$ levels. Let q be the highest common factor (HCF) of $s_1, s_2, \ldots, s_m$ and for $1 \le j \le m,\ s_j = qs_j^*$. The $v = \prod_{j=1}^m s_j$ treatment combinations are represented by m-tuples $a_1a_2\cdots a_m$, and are lexicographically ordered, where $a_j \in \{0, 1, \ldots, s_j - 1\}$ denote the levels of the factor $F_j,\ 1 \le j \le m$. Let $\Delta(m, 1) = \{1, 2, \ldots, m\}$, and $\Delta(m, 2) = \{kl, 1 \le k < l \le m\}$. For $1 \le j \le m, 0 \le \alpha \le s_j - 1$, let n_α^j be the number of times the factor F_j appears at the level α and $n_{\alpha\beta}^{kl}$ be the number of times the factors F_k and F_l appear at levels α and β respectively. Let

$$\phi(k) = s_k \sum_{\alpha=0}^{s_k-1} (n_\alpha^k)^2, \quad k \in \Delta(m, 1) \quad \text{and} \quad \phi(kl) = s_k s_l \sum_{\alpha=0}^{s_k-1} \sum_{\beta=0}^{s_l-1} (n_{\alpha\beta}^{kl})^2, \quad kl \in \Delta(m, 2).$$

It is to be remarked that if d is column balanced, then $\phi(k) = N^2,\ 1 \le k \le m$. Also, let 1_t be the $t \times 1$ vector with all elements unity, I_t be an identity matrix of order t and $P_j = [p_j(0), p_j(1), \ldots, p_j(s_j-1)]$ is an $(s_j-1)\times s_j$ matrix satisfying $P_jP_j^T = s_jI_{s_j-1}$ and $P_j1_{s_j} = 0$. It is easy to note that, for $1 \le j \le m, P_j^TP_j = s_jI_{s_j} - 1_{s_j}^T1_{s_j}$. This implies, for $1 \le \alpha, \beta \le s_j - 1, p_j(\alpha)^Tp_j(\beta) = s_j\delta_{\alpha\beta} - 1$, where δ is the Kronecker delta. Let Z_j be the $N \times (s_j - 1)$ matrix with rows $p_j(a_{ij})^T, 1 \le i \le N, 1 \le j \le m$. Then, the design matrix (excluding the column of ones) will be

$$X = [Z_1\ Z_2 \cdots Z_m]. \tag{1}$$

For any $d \in \mathcal{D}(N, s_1 \times s_2 \times \cdots \times s_m)$, let $c_{d1}, c_{d2}, \ldots, c_{dm}$ be its m columns. Along the line of Yamada and Matsui (2002), let us define the following χ^2 statistic

$$E(\chi^2(d)) = \frac{1}{m(m-1)} \sum_{j_1=1}^m \sum_{j_2(\neq j_1)=1}^m \chi^2(c_{dj_1}, c_{dj_2}),$$

where for $1 \le j_1 < j_2 \le m$,

$$\chi^2(c_{dj_1}, c_{dj_2}) = \sum_{\alpha=0}^{s_{j_1}-1} \sum_{\beta=0}^{s_{j_2}-1} \left(\frac{s_{j_1}s_{j_2}}{N}\right) \left(n_{\alpha\beta}^{j_1j_2} - \frac{N}{s_{j_1}s_{j_2}}\right)^2.$$

The following definition provides a notion of $E(\chi^2)$-optimal SSD.

Definition 1. An SSD $d \in \mathcal{D}(N, s_1 \times s_2 \times \cdots \times s_m)$ will be said to be $E(\chi^2)$-optimal if it has the minimum $E(\chi^2)$-value as stated above.

Now, we can write

$$\begin{aligned} E(\chi^2(d)) &= \frac{1}{Nm(m-1)} \sum_{j_1=1}^m \sum_{j_2(\neq j_1)=1}^m \sum_{\alpha=0}^{s_{j_1}-1} \sum_{\beta=0}^{s_{j_2}-1} \left(n_{\alpha\beta}^{j_1j_2} - \frac{N}{s_{j_1}s_{j_2}}\right)^2 s_{j_1}s_{j_2} \\ &= \frac{1}{Nm(m-1)} \sum_{j_1=1}^m \sum_{j_2(\neq j_1)=1}^m s_{j_1}s_{j_2} \left[\sum_{\alpha=0}^{s_{j_1}-1} \sum_{\beta=0}^{s_{j_2}-1} (n_{\alpha\beta}^{j_1j_2})^2 - \frac{N^2}{s_{j_1}s_{j_2}}\right] \\ &= \frac{1}{Nm(m-1)} \sum_{j_1=1}^m \sum_{j_2(\neq j_1)=1}^m \phi(j_1j_2) - N. \end{aligned} \tag{2}$$

The $E(\chi^2)$-optimal criterion is to choose some mixed-level SSDs such that their $E(\chi^2(d))$-values are minimized.