Statistica Sinica **12**(2002), 931-939

$E(s^2)$-OPTIMALITY AND MINIMUM DISCREPANCY IN 2-LEVEL SUPERSATURATED DESIGNS

Min-Qian Liu and Fred J. Hickernell

Tianjin University and Hong Kong Baptist University

Abstract: Supersaturated experimental designs are often assessed by the $E(s^2)$ criterion, and some methods have been found for constructing $E(s^2)$-optimal designs. Another criterion for assessing experimental designs is discrepancy, of which there are several different kinds. The discrepancy measures how much the empirical distribution of the design points deviates from the uniform distribution. Here it is shown that for 2-level supersaturated designs the $E(s^2)$ criterion and a certain discrepancy share the same optimal designs.

Key words and phrases: Hamming distance, reproducing kernel, uniformity.

1. Introduction

In the context of factorial designs, there has been recent interest in the study of *supersaturated designs*. Whenever the run size of a design is insufficient for estimating all the main effects represented by the columns of the design matrix, the design is called supersaturated. In industrial statistics and other scientific experiments, especially in their preliminary stages, very often there are a large number of factors to be studied and the run size is limited because of cost. However, in many situations only a few factors are believed to have significant effects. Under this assumption of *effect sparsity* (Box and Meyer (1986)), supersaturated designs can be used effectively, allowing the simultaneous identification of the active factors. Most studies have focused on 2-level supersaturated designs, although recent work by Yamada and Lin (1999) and Fang, Lin and Ma (2000) has considered multi-level designs. Booth and Cox (1962), in the first systematic construction of supersaturated designs, proposed the $E(s^2)$ criterion defined below.

Throughout this paper let $\boldsymbol{X}$ be an $n \times m$ matrix of a factorial design with elements ± 1, and rows and columns, identified with the runs and factors respectively. Also assume that each column of $\boldsymbol{X}$ has the same number of ± 1 elements, and no two columns are proportional to each other. When $n < m + 1$, the design is supersaturated. The commonly used $E(s^2)$ criterion for comparing supersaturated designs is

$$E(s^2) = \frac{2}{m(m-1)} \sum_{1 \le i < j \le m} s_{ij}^2, \tag{1.1}$$

where s_{ij} is the (i,j)th entry of $\boldsymbol{X}^T\boldsymbol{X}$. The quantity $E(s^2)$ is a measure of non-orthogonality under the assumption that *only two out of the* m factors are active. So, it should be minimized. That is, an $E(s^2)$-*optimal* design minimizes $E(s^2)$ over all possible supersaturated designs of the same size.

After Booth and Cox (1962), there was not much work on the subject of supersaturated designs until Lin (1993). Other recent work focusing on constructions of supersaturated designs includes Wu (1993), Lin (1995), Nguyen (1996), Tang and Wu (1997), Cheng (1997), Li and Wu (1997), Yamada and Lin (1997), Lu and Meng (2000) and Liu and Zhang (2000). These papers described different methods for constructing $E(s^2)$-optimal supersaturated designs. Cheng (1997) gave a unified treatment of the optimality result of Tang and Wu (1997) and the optimality of Lin's half Hadamard matrices.

Another measure used in constructing designs is the *discrepancy*, $D(\boldsymbol{X})$ (Fang and Wang (1994), Fang and Mukerjee, Fang (2000), Lin, Winker and Zhang (2000), Fang, Ma and Winker (2002)). The discrepancy measures how much the empirical distribution of the design points departs from the uniform distribution (Hickernell (1999a)), and so minimum discrepancy designs are often called uniform designs. The discrepancy has been used to construct space-filling designs in computer experiments (Bates, Buck, Riccomagno and Wynn (1996)), and to construct designs for evaluating multiple integrals (Niederreiter (1992)). Fang and Wang (1994) gave the details of uniform designs and Hickernell (1999b) described several examples of discrepancies.

This article shows that the $E(s^2)$ criterion shares the same optimal designs with a certain discrepancy. The next section defines discrepancy in general and constructs a particular discrepancy on the discrete domain $\{-1,1\}^m$. Section 2 also shows that for 2-level factorial designs both $E(s^2)$ and the discrepancy defined on $\{-1,1\}^m$ can be expressed in terms of the *Hamming distances* between any two runs of the design. These expressions in terms of Hamming distances lead to lower bounds on $E(s^2)$ and the discrepancy for 2-level supersaturated designs in Theorem 3 of Section 3. In certain cases considered in Theorem 4 these lower bounds can be achieved. Moreover, in these cases the $E(s^2)$-optimal designs are the same as the minimum discrepancy designs, even though the discrepancy is *not* equivalent to the $E(s^2)$ criterion.

2. Discrepancy and Its Relationship with $E(s^2)$

Most discrepancies that are easy to compute are defined in terms of a symmetric, positive definite kernel function, $K(\boldsymbol{x},\boldsymbol{w})$, defined on $\mathcal{X}\times\mathcal{X}$, where the experimental domain, $\mathcal{X}$, is a measurable subset of $\boldsymbol{R}^m$. Let F denote some fixed probability distribution on $\mathcal{X}$, e.g., the uniform distribution. Let $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)^T$ denote any design matrix with n runs (rows) and m factors (columns), where each design point $\boldsymbol{x}_i$ is in $\mathcal{X}$. Let $F_{\boldsymbol{X}}$ denote the empirical

distribution of this design, i.e., $F_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} 1_{[\boldsymbol{x}_i,+\infty]}(\boldsymbol{x})$, where $1_{\{\cdot\}}(\boldsymbol{x})$ is the indicator function. Then the discrepancy of the design $\boldsymbol{X}$ with respect to F, $\mathcal{X}$, and K, denoted by $D(\boldsymbol{X}; K)$, is defined as (Hickernell, 1999a)

$$D^2(\boldsymbol{X}; K) = \int_{\mathcal{X}^2} K(\boldsymbol{x}, \boldsymbol{w}) d(F - F_{\boldsymbol{X}})(\boldsymbol{x}) d(F - F_{\boldsymbol{X}})(\boldsymbol{w})$$

$$= \int_{\mathcal{X}^2} K(\boldsymbol{x}, \boldsymbol{w}) dF(\boldsymbol{x}) dF(\boldsymbol{w}) - \frac{2}{n}\sum_{i=1}^{n}\int_{\mathcal{X}} K(\boldsymbol{x}, \boldsymbol{x}_i) dF(\boldsymbol{x}) + \frac{1}{n^2}\sum_{i,j=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}_j). \quad (2.1)$$

For a fixed number of points, n, a design with low discrepancy is preferred.

Many discrepancies have been defined using the unit cube domain, $\mathcal{X} = [0,1]^m$. In situations where the variables in the experimental domain have a continuous range of values, it may be reasonable to define the experimental domain as $[0,1]^m$ and define the discrepancy there. However, in other situations one restricts the number of possible levels for each factor to a finite number. For example, a factor may have only two values (on and off) or three values (low, medium and high). In these situations it makes more sense to represent the experimental domain as a discrete set, e.g., $\mathcal{X} = \{-1, 1\}^m$ for two levels or $\mathcal{X} = \{0, \ldots, q-1\}^m$ for q levels. A discrepancy of the form (2.1) can be defined directly on such a discrete domain. Note that for the discrete domain with q levels on each of m factors, a perfect design (in terms of coverage) would be a full factorial design with q^m points that samples every possible point once. For such a design the discrepancy defined with respect to the discrete domain vanishes.

For a 2-level factorial design with m factors let the experimental domain be $\mathcal{X}_d = \{-1, 1\}^m$, and let F_d be the uniform distribution corresponding to $\mathcal{X}_d$, i.e., F_d assigns probability 2^{-m} to each member of this set. Define the following kernel on $\mathcal{X}_d^2$:

$$K_d(\boldsymbol{x}, \boldsymbol{w}) = \prod_{k=1}^{m} [1 + \beta \hat{K}_d(x_k, w_k)], \quad \text{for any} \quad \boldsymbol{x}, \boldsymbol{w} \in \mathcal{X}_d, \quad (2.2a)$$

$$\hat{K}_d(x, w) = \begin{cases} 1, & x = w, \\ \rho, & x \neq w, \end{cases} \quad (2.2b)$$

β some positive number and $-1 \leq \rho < 1$. Then the discrepancy of a 2-level design with an $n \times m$ design matrix $\boldsymbol{X}$, whose rows are elements of $\mathcal{X}_d$, is

$$D^2(\boldsymbol{X}; K_d) = -\left[1 + \frac{\beta(1+\rho)}{2}\right]^m + \frac{1}{n^2}\sum_{i,j=1}^{n}\prod_{k=1}^{m}[1 + \beta \hat{K}_d(x_{ik}, x_{jk})]. \quad (2.3)$$

Remark 1. Note that the definition of K_d is independent of the values of the two levels. Thus, one may change the experimental domain from $\mathcal{X}_d = \{-1, 1\}^m$ to

$\mathcal{X}_d = \{0,1\}^m$, for example, and use the same definition of K_d to get an equivalent discrepancy.

Next we show the relationship between discrepancy and $E(s^2)$ via Hamming distance. For an $n \times m$ design matrix $\boldsymbol{X}$ with rows $\boldsymbol{x}_i^T$, the *Hamming distance* between the ith and jth rows is defined as the number of places where they differ, i.e., $d_{ij} = \sum_{k=1}^m 1_{\{x_{ik} \neq x_{jk}\}}$. Note that $d_{ii} = 0$. Also, note that for supersaturated designs, since the number of ± 1 elements in each column must be the same, it follows that

$$\sum_{j=1}^n d_{ij} = \frac{mn}{2}. \tag{2.4}$$

The discrepancy $D(\boldsymbol{X}; K_d)$ in (2.3) and $E(s^2)$ may be written in terms of Hamming distance.

Theorem 1. *For a supersaturated design $\boldsymbol{X}$ with n runs and m 2-level factors,*

$$E(s^2) = -\frac{(m+1)n^2}{m-1} + \frac{4}{m(m-1)} \sum_{1 \le i \neq j \le n} d_{ij}^2, \tag{2.5}$$

$$D^2(\boldsymbol{X}; K_d) = -\left[1 + \frac{\beta(1+\rho)}{2}\right]^m + \frac{(1+\beta)^m}{n^2}\left[n + \sum_{1 \le i \neq j \le n} \left(\frac{1+\rho\beta}{1+\beta}\right)^{d_{ij}}\right]. \tag{2.6}$$

Proof. From the definition of $E(s^2)$ in (1.1) and some basic matrix identities it follows that

$$\begin{aligned} E(s^2) &= \frac{\operatorname{tr}(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{X}) - mn^2}{m(m-1)} = \frac{\operatorname{tr}(\boldsymbol{X}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{X}^T) - mn^2}{m(m-1)} \\ &= \frac{1}{m(m-1)}\left[m^2 n - mn^2 + \sum_{1 \le i \neq j \le n} (m - 2d_{ij})^2\right] \\ &= \frac{1}{m(m-1)}\left[m^2n^2 - mn^2 + 4\sum_{1 \le i \neq j \le n} d_{ij}^2 - 4m \sum_{1 \le i \neq j \le n} d_{ij}\right]. \end{aligned}$$

Substituting (2.4) into the above equation yields (2.5).

To derive (2.6) note that, for any runs $\boldsymbol{x}_i^T$ and $\boldsymbol{x}_j^T$, one has $x_{ik} \neq x_{jk}$ for d_{ij} values of k and $x_{ik} = x_{jk}$ for $m - d_{ij}$ values of k. Thus,

$$\prod_{k=1}^m [1 + \beta \hat{K}_d(x_{ik}, x_{jk})] = (1+\beta)^{m-d_{ij}}(1+\rho\beta)^{d_{ij}} = (1+\beta)^m \left(\frac{1+\rho\beta}{1+\beta}\right)^{d_{ij}}.$$

Substituting this into (2.3) gives (2.6).

3. Bounds on $E(s^2)$ and the Discrepancy

Although (2.5) and (2.6) differ, there is enough similarity that under certain conditions the same design minimizes both. The following lemma and theorem make this statement precise.

Lemma 2. *Let $\boldsymbol{d}$ be an integer vector containing M elements, and let $f(\boldsymbol{d})$ be an arbitrary function of $\boldsymbol{d}$ satisfying:* (i) *$f(\boldsymbol{d})$ is invariant if any two elements of $\boldsymbol{d}$ are interchanged, and* (ii) *$f(\boldsymbol{d})$ is strictly convex. For any integer α define $f_{\min} = \min\{f(\boldsymbol{d}) : \mathbf{1}^T\boldsymbol{d} = \alpha\}$, $\gamma = \lfloor \alpha/M \rfloor$, and $L = M(\gamma+1) - \alpha$. Then $f_{\min} = f(\boldsymbol{d}_{\text{opt}})$, where $\boldsymbol{d}_{\text{opt}}$ has L elements with the value γ and $M-L = \alpha - M\gamma$ elements with the value $\gamma+1$.*

Proof. First, it is shown that the elements of $\boldsymbol{d}_{\text{opt}} = (d_{\text{opt},i})$ must satisfy

$$|d_{\text{opt},i} - d_{\text{opt},j}| \le 1 \quad \text{for all } i, j. \tag{3.1}$$

Suppose that (3.1) is false, i.e., there exists an i and j such that $d_{\text{opt,i}} - d_{\text{opt,j}} > 1$. Let $\boldsymbol{d}_1$ be obtained from $\boldsymbol{d}_{\text{opt}}$ by exchanging the ith and jth elements, and note that $f(\boldsymbol{d}_1) = f(\boldsymbol{d}_{\text{opt}}) = f_{\min}$. Let $\lambda = 1/(d_{\text{opt,i}} - d_{\text{opt,j}})$, and let $\boldsymbol{d}_2 = (1-\lambda)\boldsymbol{d}_{\text{opt}} + \lambda\boldsymbol{d}_1$. The vector $\boldsymbol{d}_2$ also has integer elements and satisfies $\mathbf{1}^T\boldsymbol{d} = \alpha$. By the strict convexity assumption $f(\boldsymbol{d}_2) < (1-\lambda)f(\boldsymbol{d}_{\text{opt}}) + \lambda f(\boldsymbol{d}_1) = f_{\min}$, a contradiction.

Given (3.1) it follows that the elements of $\boldsymbol{d}_{\text{opt}}$ must take on the integer values γ and $\gamma+1$. The condition $\mathbf{1}^T\boldsymbol{d} = \alpha$ determines the value of γ and the number of times γ appears in $\boldsymbol{d}_{\text{opt}}$.

Theorem 3. *Let $\boldsymbol{X}$ be a 2-level supersaturated design with n runs and m factors, let $\gamma = \lfloor mn/[2(n-1)] \rfloor$, and suppose that $\rho\beta > -1$. Then the following lower bounds hold.*

$$E(s^2) \ge \frac{n^2(m-n+1)}{(m-1)(n-1)} + \frac{4n(n-1)}{m(m-1)}\left[\gamma+1-\frac{mn}{2(n-1)}\right]\left[\frac{mn}{2(n-1)}-\gamma\right], \tag{3.2a}$$

$$\begin{aligned} D^2(\boldsymbol{X}; K_d) \ge & -\left[1+\frac{\beta(1+\rho)}{2}\right]^m + \frac{(1+\beta)^m}{n} \\ & + \frac{(1+\beta)^{m-\gamma}(1+\rho\beta)^\gamma(n-1)}{n}\left\{\left[\gamma+1-\frac{mn}{2(n-1)}\right]\right. \\ & \left. + \left[\frac{mn}{2(n-1)}-\gamma\right]\frac{1+\rho\beta}{1+\beta}\right\}. \end{aligned} \tag{3.2b}$$

Proof. The proof proceeds in a straightforward manner by applying Lemma 2. The result of this lemma is a lower bound, since there is no guarantee from the lemma that there exist designs $\boldsymbol{X}$ whose Hamming distances satisfy condition

(3.1). Discrepancy and the $E(s^2)$ criterion are functions of $\boldsymbol{d} = (d_{ij})_{i<j}$. In this case the parameters in Lemma 2 may be identified as $\alpha = mn^2/4$, $M = n(n-1)/2$, $\gamma = \lfloor mn/[2(n-1)] \rfloor$. Thus, under condition (2.4), Lemma 2 implies that

$$\begin{aligned}\sum_{1\le i\ne j\le n} d_{ij}^2 &\ge 2\{L\gamma^2 + (M-L)(\gamma+1)^2\} = 2\{M\gamma^2 + (M-L)(2\gamma+1)\}\\ &= 2\{M\gamma^2 + (\alpha - M\gamma)(2\gamma+1)\} = 2\{-M\gamma(\gamma+1) + \alpha(2\gamma+1)\}\\ &= 2M\left\{\left[\gamma + 1 - \frac{\alpha}{M}\right]\left[\frac{\alpha}{M} - \gamma\right] + \frac{\alpha^2}{M^2}\right\}.\end{aligned}$$

Substituting this expression into (2.5) and substituting the values of M and α gives (3.2a).

To prove (3.2b), note from (2.6) that $D(\boldsymbol{X}; K_d)$ is a strictly convex function of $\boldsymbol{d} = (d_{ij})_{i<j}$ provided $1 + \rho\beta > 0$, which has been assumed. Substituting γ for L values of d_{ij} and $\gamma + 1$ for $M - L$ values of d_{ij} leads to (3.2b), after some straightforward algebraic manipulation.

The lower bound on $E(s^2)$ in (3.2a) is a generalization of those obtained by Nguyen (1996), Tang and Wu (1997), and Cheng (1997). The bound on the discrepancy is new. It is interesting to note that if a design $\boldsymbol{X}$ can attain one of the lower bounds above, then it attains both of them. In other words, an $E(s^2)$-optimal design is also uniform (minimal discrepancy) for the discrepancy $D(\boldsymbol{X}; K_d)$, provided that $\rho\beta > -1$.

For some values of m and n one can show that the lower bounds in Theorem 3 are attainable, i.e., one knows how to construct $E(s^2)$-optimal and minimum discrepancy designs.

Theorem 4. *Let $\boldsymbol{X}$ be a 2-level design with n runs and m factors, where each column has the same number of ± 1 elements. Suppose that $\rho\beta > -1$, and that $m = c(n-1)+e$ for $e = -1, 0$ or 1. Also, suppose that either* (a) *n is a multiple of 4 and there exists an $n \times n$ Hadamard matrix, or* (b) *c is even and there exists a $2n \times 2n$ Hadamard matrix. Then the lower bounds in Theorem 3 can be attained.*

Proof. Note that n must be even because each column of $\boldsymbol{X}$ is assumed to have the same number of ± 1 elements. For the values of m considered here one may compute

$$\begin{aligned}\gamma_e &= \left\lfloor \frac{mn}{2(n-1)} \right\rfloor = \left\lfloor \frac{cn}{2} + \frac{en}{2(n-1)} \right\rfloor = \begin{cases} cn/2 - 1, & e = -1,\\ cn/2, & e = 0, 1,\end{cases}\\ L_e &= M(\gamma_e + 1) - \alpha\\ &= \frac{n(n-1)(\gamma_e+1)}{2} - \frac{[c(n-1)+e]n^2}{4} = \begin{cases} n^2/4, & e = -1,\\ n(n-1)/2, & e = 0,\\ n(n-2)/4, & e = 1.\end{cases}\end{aligned}$$

For case a let $\boldsymbol{H}_l = (\mathbf{1}, \tilde{\boldsymbol{H}}_l)$, $l = 1, \ldots, c$, be $n \times n$ Hadamard matrices, where the first column of each $\boldsymbol{H}_l$ has been normalized to $\mathbf{1}$, and $\tilde{\boldsymbol{H}}_l$ denotes the last $n-1$ columns of $\boldsymbol{H}_l$. Define the $n \times c(n-1)$ design matrix to be $\boldsymbol{X}_0 = (\tilde{\boldsymbol{H}}_1, \ldots, \tilde{\boldsymbol{H}}_c)$. Since the columns of $\tilde{\boldsymbol{H}}_l$ are all orthogonal to $\mathbf{1}$, they each must have the same number of ± 1 elements. Because the rows of each $\boldsymbol{H}_l$ are orthogonal, all $L_0 = n(n-1)/2$ Hamming distances for this matrix are $d_{ij} = \gamma_0$ for $i < j$. Thus, from Lemma 2 and the proof of Theorem 3, it follows that the lower bounds in Theorem 3 are attained using $\boldsymbol{X}_0$ for $m = c(n-1)$, i.e., the case $e = 0$.

For the case of $e = -1$, define $\boldsymbol{X}_{-1}$ as all but the last column of $\boldsymbol{X}_0$. Because the column removed has the same number of ± 1 elements, for any fixed i, in comparison to $\boldsymbol{X}_0$ there are $n/2$ Hamming distances $d_{ij},\ i \neq j$, that are reduced from $\gamma_{-1} + 1 = cn/2$ to $\gamma_{-1} = cn/2 - 1$, while the other d_{ij} remain unchanged. Thus, there are a total of $L_{-1} = n^2/4$ Hamming distances $d_{ij},\ i < j$, having the value γ_{-1} with the rest having the value $\gamma_{-1} + 1$. Thus, $\boldsymbol{X}_{-1}$ attains the lower bounds in Theorem 3.

For the case of $e = 1$, define $\boldsymbol{X}_1$ as a column with equal numbers of ± 1 elements appended to $\boldsymbol{X}_0$. By a similar argument as above, there are now a total of $M - L_1 = n^2/4$ Hamming distances $d_{ij},\ i < j$, having the value $\gamma_1 + 1$, with the rest having the value γ_1. Thus, $\boldsymbol{X}_1$ attains the lower bounds in Theorem 3.

For case b choose $c/2$ $2n \times 2n$ Hadamard matrices

$$\boldsymbol{H}_l = \begin{pmatrix} \mathbf{1} & \mathbf{1} & \boldsymbol{H}_l^h \\ \mathbf{1} & -\mathbf{1} & * \end{pmatrix}, \quad l = 1, \ldots, c/2,$$

where the $n \times (2n-2)$ matrix $\boldsymbol{H}_l^h$ denotes the upper right part of H_l. Note that the rows of each $\boldsymbol{H}_l$ are orthogonal. Define $\boldsymbol{X}_0 = (\boldsymbol{H}_1^h, \ldots, \boldsymbol{H}_{c/2}^h)$. Then $\boldsymbol{X}_0$ attains the lower bounds in Theorem 3. Defining $\boldsymbol{X}_{-1}$ and $\boldsymbol{X}_1$ as in case a also gives design matrices that attain the lower bounds in Theorem 3.

Remark 2. The construction in case a is due to Tang and Wu (1997), and the construction in case b is due to Lin (1993). However, here it is shown that these constructions minimize the discrepancy as well as $E(s^2)$.

Remark 3. The constructions above *do not* guarantee that no two columns of $\boldsymbol{X}$ are multiples of each other. In many cases one may choose the Hadamard matrices so that this condition is satisfied. However, if m becomes too large, e.g., $m > n!/[2(n/2)!(n/2)!]$, then it is impossible to have no two columns of $\boldsymbol{X}$ be constant multiples of each other.

4. Conclusion

The concept of discrepancy dates back to Weyl (1916). It is used in constructing Kolmogorov-Smirnov and Cramér-von Mises statistics for testing goodness-of-fit (D'Agostino and Stephens (1986)). In this article it has been shown that

the $E(s^2)$ criterion for supersaturated designs is closely related to discrepancy. As mentioned in Remark 2, for certain choices of m and n, the $E(s^2)$-optimal designs also minimize the discrepancy $D(\boldsymbol{X}; K_d)$ defined on $\{-1, 1\}^m$. For other choices of m and n one would expect that $E(s^2)$-optimal designs would have nearly minimal discrepancy $D(\boldsymbol{X}; K_d)$, and vice versa, based on the argument in the proof of Theorem 3.

However, discrepancy is a more general, and thus more flexible criterion than $E(s^2)$. For example, $E(s^2)$ ignores possible interactions of more than one factor. However, the discrepancy $D(\boldsymbol{X}; K_d)$ includes interactions of all possible orders, and their importance may be increased or decreased by changing the value of β. The discrepancy $D(\boldsymbol{X}; K_d)$ may be extended to q-level experimental domains, $\mathcal{X}_d = \{0, \ldots, q-1\}^m$ for any positive integer q, using the same definition of K_d in (2.2). One only need ensure that $-(q-1)^{-1} \le \rho < 1$.

Acknowledgements

The authors are grateful to the Editor, an associate editor, and two referees for their valuable comments. Thanks also go to Kai-Tai Fang, Chang-Xing Ma, Rahul Mukerjee and Guo-Liang Tian for their valuable suggestions and help on this paper. This research was partially supported by Hong Kong Research Council Grant RGC/HKBU/2029/99E, Hong Kong Baptist University Faculty Research Grant FRG/99-00/II-01, and NNSF of China grant 10171051.

References

Bates, R., Buck, R., Riccomagno, E. and Wynn, H. P. (1996). Experimental design and observation for large systems. *J. Roy. Statist. Soc. Ser. B* **58**, 77-94.

Booth, K. H. V. and Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics* **4**, 489-495.

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorial. *Technometrics* **28**, 11-18.

Cheng, C.-S. (1997). $E(s^2)$-optimal supersaturated designs. *Statist. Sinica* **7**, 929-939.

D'Agostino, R. B. and Stephens, M. A., eds. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

Fang, K. T., Lin, D. K. J. and Ma, C. X. (2000). On the construction of multi-level supersaturated designs. *J. Statist. Plann. Inference* **86**, 239-252.

Fang, K. T., Lin, D. K. J., Winker, P. and Zhang, Y. (2000). Uniform design: theory and applications. *Technometrics* **42**, 237-248.

Fang, K. T., Ma, C. X. and Winker, P. (2000). Centered L_2-discrepancy of random sampling and latin hypercube design, and construction of uniform design. *Math. Comp.* **71**, 275-296.

Fang, K. T. and Mukerjee, R. (2000). A connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* **87**, 193-198.

Fang, K. T. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. Chapman and Hall, New York.

Hickernell, F. J. (1999a). Goodness-of-fit statistics, discrepancies and robust designs. *Statist. Probab. Lett.* **44**, 73-78.

Hickernell, F. J. (1999b). What affects the accuracy of quasi-Monte Carlo quadrature? In *Monte Carlo and Quasi-Monte Carlo Methods* 1998 (Edited by H. Niederreiter and J. Spanier), 16-55. Springer-Verlag, Berlin.

Li, W. W. and Wu, C. F. J. (1997). Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. *Technometrics* **39**, 171-179.

Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics* **35**, 28-31.

Lin, D. K. J. (1995). Generating systematic supersaturated designs. *Technometrics* **37**, 213-225.

Liu, M. Q. and Zhang, R. C. (2000). Construction of $E(s^2)$ optimal supersaturated designs using cyclic BIBDs. *J. Statist. Plann. Inference* **91**, 139-150.

Lu, X. and Meng, Y. (2000). A new method in the construction of two-level supersaturated designs. *J. Statist. Plann. Inference* **86**, 229-238.

Nguyen, N.-K. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics* **38**, 69-73.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.

Tang, B. and Wu, C. F. J. (1997). A method for constructing supersaturated designs and its $E(s^2)$ optimality. *Canad. J. Statist.* **25**, 191-201.

Weyl, H. (1916). Über die gleichverteilung der zahlem mod eins. *Math. Ann.* **77**, 313-352.

Wu, C. F. J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika* **80**, 661-669.

Yamada, S. and Lin, D. K. J. (1997). Supersaturated designs including an orthogonal base. *Canad. J. Statist.* **25**, 203-213.

Yamada, S. and Lin, D. K. J. (1999). Three-level supersaturated design. *Statist. Probab. Lett.* **45**, 31-39.

School of Management, Tianjin University, Tianjin 300072, China.

E-mail: mqliu@263.net

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong SAR, China.

E-mail: fred@hkbu.edu.hk

(Received November 2000; accepted November 2001)