



Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A study on design uniformity under errors in the level values

Jianfeng Yang^a, Fasheng Sun^a, Dennis K.J. Lin^{b,c}, Min-Qian Liu^{a,*}

^a Department of Statistics, School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China

^b Department of Statistics, The Pennsylvania State University, University Park, USA

^c School of Statistics, Renmin University of China, China

ARTICLE INFO

Article history:

Received 17 September 2009
Received in revised form 18 May 2010
Accepted 23 May 2010
Available online 1 June 2010

MSC:

primary 62K15
secondary 62K05

Keywords:

Discrepancy
Robust
Uniform design

ABSTRACT

Discrepancy is an important criterion for uniformity in the design of experiments. Basically, it measures the distance between two functions: the empirical distribution of the design points and the theoretical uniform distribution function. This paper studies the properties of uniformity when the factor level values are contaminated with errors. Specifically, our study focuses on the wrap-around L_2 -discrepancy. It is shown that uniform designs with errors are less uniform (in the average sense) than the original ones without errors. Related theorems are obtained. Furthermore, it can be shown that the lattice sampling outperforms Latin hypercube sampling. The latter can be viewed as a lattice sampling with error-in-level values.

© 2010 Elsevier B.V. All rights reserved.

1. Preliminary

The uniform design initially proposed by Wang and Fang (1981) is a space-filling design for computer experiments, but it can also be utilized as a fractional factorial design. In the past two decades, it has been successfully applied to a wide range of areas; for example, in industry, system engineering, pharmaceuticals and natural sciences (see Fang and Lin, 2003). Many uniform designs are available at the website <http://www.math.hkbu.edu.hk/UniformDesign>.

A U-type design $U(n; q_1, \dots, q_s)$ corresponds to an $n \times s$ matrix $U = (u_{ik})$, such that the elements u_{ik} ($i = 1, \dots, n$) in the k th column take values (equally often) from a set of q_k integers, say $\{1, \dots, q_k\}$. When some of the q_k are equal, we denote this U-type design by $U(n; q_1^{s_1} \dots q_m^{s_m})$, where $s = \sum_{k=1}^m s_k$, or $U(n; q^s)$ when all the q_k are equal. By mapping u_{ik} to $x_{ik} = (2u_{ik} - 1)/(2q_k)$ for $i = 1, \dots, n$ and $k = 1, \dots, s$, the n runs of a $U(n; q_1, \dots, q_s)$ can be transformed to n points in $C^s = [0, 1]^s$.

The uniformity of a design is usually measured by a discrepancy criterion (Hickernell, 1998a,b). There are several different discrepancies that have been defined, among which the wrap-around L_2 -discrepancy (WD_2 , for short) and centered L_2 -discrepancy (CD_2 , for short) are regarded as the most practicable ones. In this paper, the wrap-around L_2 -discrepancy is used; it can be defined as

$$WD_2(U) = \left[-\left(\frac{4}{3}\right)^s + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \prod_{k=1}^s \left(\frac{3}{2} - |x_{ik} - x_{jk}| + |x_{ik} - x_{jk}|^2 \right) \right]^{\frac{1}{2}},$$

where $x_{ik} = (2u_{ik} - 1)/(2q_k)$, $i = 1, \dots, n$, $k = 1, \dots, s$. Other criteria can be studied in a similar manner (see Section 3).

* Corresponding author. Tel.: +86 22 23504709; fax: +86 22 23506423.

E-mail addresses: mqliu@nankai.edu.cn, nkmqliu@yahoo.com.cn (M.-Q. Liu).

Table 1

A $U(6; 6^5)$ design.

Run	1	2	3	4	5
1	3	1	6	5	4
2	5	2	2	2	5
3	6	5	3	6	3
4	4	4	5	1	1
5	2	6	4	3	6
6	1	3	1	4	2

Table 2

A $U(6; 6^5)$ design under errors of $\text{Unif}(-1/3, 1/3)$.

Run	1	2	3	4	5
1	2.6644	0.5613	5.8923	4.7964	3.4486
2	4.9332	1.5265	1.7120	1.7475	4.9865
3	5.8811	4.5608	2.5803	5.8632	2.5143
4	3.7633	3.6894	4.8018	0.9712	0.5016
5	1.8786	5.8181	3.6977	2.6817	5.9172
6	0.7735	2.5395	0.6299	3.9201	1.8249

When the experiments are carried out in practice, the actually performed factor level values may be accompanied by errors; see, for example, Box (1963) and Draper and Beggs (1971). Consider the design $U(6; 6^5)$ in Table 1, obtainable from the above website for uniform designs. The WD_2^2 value of the design is 0.2030. When the factor level values are contaminated with errors, the actually performed factor levels become $u'_{ik} = u_{ik} + \varepsilon_{ik}$, with ε_{ik} being the random error. Consider the bounded support uniform distribution for the errors, namely, $\varepsilon_{ik} \sim \text{i.i.d. Unif}(-\tau, \tau)$. Then, the actually performed design with $\tau = 1/3$ may be as in Table 2, the WD_2^2 value of which is 0.2182, as compared to 0.2030 of the original WD_2^2 .

In this paper, we study the design discrepancies when the factor level values are subject to errors. We first consider the WD_2^2 values for designs with errors in all factors, and then investigate the cases in which errors only occur in some factors. Finally, we apply the results to the construction of uniform designs.

2. Main results

For any $U(n; n^s)$ design U , when the factor levels are contaminated with uniformly distributed errors, it becomes $Z = U + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = (\varepsilon_{ik})$ with $\varepsilon_{ik} \sim \text{i.i.d. Unif}(-\tau, \tau)$. In this case, the value of WD_2^2 for the resulting design Z becomes

$$WD_2^2(Z) = -\left(\frac{4}{3}\right)^s + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \prod_{k=1}^s \left(\frac{3}{2} - |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}| + |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}|^2\right),$$

where $\delta_{ik}(=\varepsilon_{ik}/n) \sim \text{i.i.d. Unif}(-a, a)$ with $a = \tau/n$. Let $B_Z(n, s, a) = E(WD_2^2(Z) - WD_2^2(U))$ represent the expected difference between the WD_2^2 values for design Z and for design U . We have the following result whose proof is given in the Appendix.

Theorem 1. For a $U(n; n^s)$ design U , with $a < 1/(2n)$, we have

- (i) $\frac{n-1}{n} \left(\frac{5}{4}\right)^s \sum_{k=1}^s \binom{s}{k} \left(\frac{8}{15}a^2\right)^k < B_Z(n, s, a) < \frac{n-1}{n} \left(\frac{3}{2}\right)^s \sum_{k=1}^s \binom{s}{k} \left(\frac{4}{9}a^2\right)^k$;
- (ii) for any fixed n and s , $B_Z(n, s, a)$ is an increasing function of a ;
- (iii) for any fixed n and a , $B_Z(n, s, a)$ is an increasing function of s .

Theorem 1 shows how the discrepancies of designs with and without errors in their level values will differ. It is shown that designs with errors are on average less uniform than those without any error. Furthermore, the larger the error, the larger the expected WD_2 value for the design will be. For example, Fig. 1 displays the result for the uniform design in Table 1. The middle curve of Fig. 1(a) shows the relationship between the values of $B_Z(n, s, a)$ and τ . Here, we take the mean value of 100 replications for each τ as $B_Z(n, s, a)$ in the simulation. The upper and lower curves are the upper and lower bounds for $B_Z(n, s, a)$, respectively. To illustrate the magnitude of $B_Z(n, s, a)$, we define the relative ratio (RR) of this expected difference to $WD_2^2(U)$ as $RR = B_Z(n, s, a)/WD_2^2(U)$. A larger RR value implies that the uniformity discrepancy changes more significantly. Fig. 1(b) shows how RR varies as τ increases. It is shown that the uniformity discrepancy is robust if the errors are relatively small (for example $\tau < 0.2$). Here, robustness refers to a situation when the expected change of discrepancies between the original design (without error) and the contaminated design (with errors) is rather insignificant.

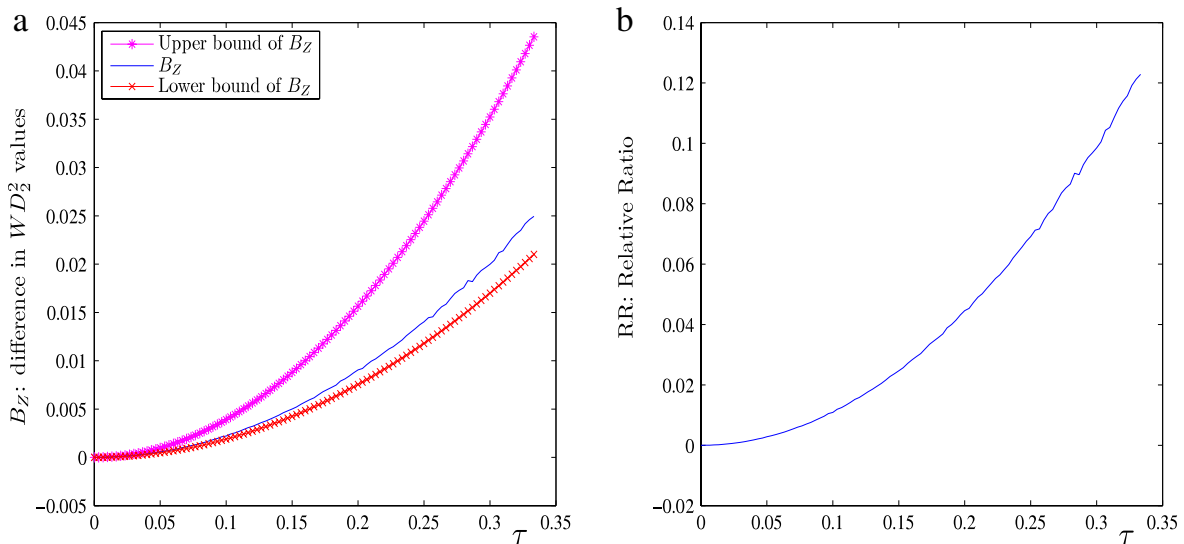


Fig. 1. The plots of (a) B_Z against τ and (b) RR against τ for the design in Table 1.

In what follows, we will generalize our main results to the case when errors only occur in some but not all factors. Without loss of generality, we assume that the first s_1 factors have errors in their levels and that the true design matrix is $Y = U + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} = (\varepsilon_{ik})$ with $\varepsilon_{ik} \sim \text{i.i.d. Unif}(-\tau, \tau)$ for $k \leq s_1$; and $\varepsilon_{ik} = 0$ for $k > s_1$. The value of WD_2^2 for design Y then becomes

$$WD_2^2(Y) = -\left(\frac{4}{3}\right)^s + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \left[\prod_{k=s_1+1}^s \left(\frac{3}{2} - |x_{ik} - x_{jk}| + |x_{ik} - x_{jk}|^2 \right) \right] \times \left[\prod_{k=1}^{s_1} \left(\frac{3}{2} - |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}| + |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}|^2 \right) \right],$$

where $\delta_{ik} = \varepsilon_{ik}/q_k$ for $k = 1, \dots, s_1$. Our finding can be stated as the following theorem.

Theorem 2. For a $U(n; n^s)$ design U , let $B_Y(n, s, a, s_1) = E(WD_2^2(Y) - WD_2^2(U))$; then

- (i) $\frac{n-1}{n} \left(\frac{5}{4}\right)^s \sum_{k=1}^{s_1} \binom{s_1}{k} \left(\frac{8}{15}a^2\right)^k < B_Y(n, s, a, s_1) < \frac{n-1}{n} \left(\frac{3}{2}\right)^s \sum_{k=1}^{s_1} \binom{s_1}{k} \left(\frac{4}{9}a^2\right)^k$;
- (ii) for any fixed n, s and s_1 , $B_Y(n, s, a, s_1)$ will increase as the value a increases;
- (iii) for any fixed n, s and a , $B_Y(n, s, a, s_1)$ will increase as s_1 , the number of factors with errors, increases.

The results in Theorems 1 and 2 indicate that, if the errors are relatively small, the traditional uniform designs are rather robust. Also, the results can be used in the construction of uniform designs. There are two conventional frameworks for construction of uniform designs on $[0, 1]^s$. The traditional one, called lattice sampling (Patterson, 1954), is to select experimental points from the centers of grids, namely, $x_{ik} = (u_{ik} - 1/2)/n$, for $1 \leq i \leq n$ and $1 \leq k \leq s$. The other one, called Latin hypercube sampling (McKay et al., 1979), is to select experimental points randomly within the grids, namely, $\tilde{x}_{ik} = (u_{ik} - \epsilon_{ik})/n$, where $\epsilon_{ik} \sim \text{i.i.d. Unif}(0, 1)$ for $1 \leq i \leq n$ and $1 \leq k \leq s$. These two frameworks have been generalized to orthogonal arrays for space fillings by Owen (1992). Fang et al. (2002) proved that lattice sampling minimizes the WD_2 discrepancy on $[0, 1]$ for one dimension ($s = 1$). Theorem 1(ii) extends their result to indicate that this is also true in expectation for higher dimensions ($s \geq 2$).

3. Conclusions

This paper discusses the design uniformity when the factor level values are contaminated with random errors. It is shown that a design with errors is less uniform than the original one. Furthermore, the fewer factors with errors and/or the smaller the errors are, the better the WD_2 uniformity is. In addition, it is shown that for U-type designs, the designs in which experimental points are selected from the centers of grids are more uniform in expectation than those in which experimental points are randomly chosen within the grids. In other words, lattice sampling outperforms Latin hypercube sampling in terms of uniformity. Note that, in this paper, we only consider the WD_2 discrepancy as the measure of uniformity. In fact, a similar study has been applied to other discrepancies, including CD_2 (see, for example, Fang and Lin (2003) and Winker and Lin (forthcoming)). The results are similar to Theorems 1 and 2, and are thus omitted here.

Though only uniform random errors are considered in this paper, it can be shown that the main results of Theorems 1 and 2 also hold for other random error structures, such as normal and beta distributions. This can be theoretically derived for the normal distribution, and via simulation for the beta distribution. For the normal distribution, because of its unbounded

support, it is more appropriate to adopt the truncated normal random errors for which the truncation points are chosen in a way to ensure that all design points will stay within the unit cube. The results are very similar, and thus are not presented here.

Acknowledgements

This work is supported by the Program for New Century Excellent Talents in University (NCET-07-0454) of China and the NNSF of China grants 10871104 and 10971107. The authors thank the Co-Editor and two referees for their valuable comments, which have significantly improved the presentation.

Appendix

To prove Theorem 1, the following lemma is needed.

Lemma 1. *If random variables $\xi, \eta \sim \text{Unif}(-\tau, \tau)$ and they are independent, then we have*

$$\begin{aligned}
 \text{(i)} \ E|\xi + c| &= \begin{cases} -c, & \text{if } c \leq -\tau, \\ \frac{\tau^2 + c^2}{2\tau}, & \text{if } -\tau < c < \tau, \\ c, & \text{if } c \geq \tau; \end{cases} \\
 \text{(ii)} \ E|\xi + c|^2 &= \frac{1}{3}\tau^2 + c^2; \\
 \text{(iii)} \ E|\xi - \eta + c| &= \begin{cases} -c, & \text{if } c \leq -2\tau, \\ \frac{1}{12\tau^2}(8\tau^3 + 6\tau c^2 + c^3), & \text{if } -2\tau < c \leq 0, \\ \frac{1}{12\tau^2}(8\tau^3 + 6\tau c^2 - c^3), & \text{if } 0 < c \leq 2\tau, \\ c, & \text{if } c > 2\tau; \end{cases} \\
 \text{(iv)} \ E|\xi - \eta + c|^2 &= \frac{2}{3}\tau^2 + c^2.
 \end{aligned}$$

Proof of Theorem 1. For ease of expression, let

$$\begin{aligned}
 T(U) &= \frac{1}{2}n^2 \left[WD_2^2(U) + \left(\frac{4}{3}\right)^s \right] - \frac{n}{2} \left(\frac{3}{2}\right)^s \\
 &= \sum_{i < j}^n \prod_{k=1}^s \left(\frac{3}{2} - |x_{ik} - x_{jk}| + |x_{ik} - x_{jk}|^2 \right).
 \end{aligned} \tag{1}$$

Then, for the true design matrix Z , we have

$$T(Z) = \sum_{i < j}^n \prod_{k=1}^s \left(\frac{3}{2} - |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}| + |x_{ik} - x_{jk} + \delta_{ik} - \delta_{jk}|^2 \right).$$

Let $c_{ijk} = x_{ik} - x_{jk}$, $d_{ijk} = \frac{3}{2} - |c_{ijk}| + |c_{ijk}|^2$ and $K = \{1, \dots, s\}$. The symbol $K_1 \subset K$ means that K_1 is a proper subset of K , $K \setminus K_1$ denotes the complementary set of K_1 in K and $|A|$ denotes the cardinality of the set A . Note that $|c_{ijk}| = |x_{ik} - x_{jk}| \geq \frac{1}{n} > 2a$; then, from Lemma 1 and the independence of δ_{ik} , we have

$$E[T(Z)] = T(U) + \sum_{i < j}^n \sum_{K_1 \subset K} \left(\frac{2}{3}a^2\right)^{|K \setminus K_1|} \prod_{k \in K_1} d_{ijk}.$$

The results (ii) and (iii) can be easily obtained from (1) and the expression of $B_Z(n, s, a)$. Now we need only to prove (i). Since $5/4 < d_{ijk} < 3/2$, we have

$$\begin{aligned}
 E[T(Z) - T(U)] &< \sum_{i < j}^n \left[\left(\frac{3}{2} + \frac{2}{3}a^2\right)^s - \left(\frac{3}{2}\right)^s \right] \\
 &= \frac{n(n-1)}{2} \left(\frac{3}{2}\right)^s \left[\left(1 + \frac{4}{9}a^2\right)^s - 1 \right] \\
 &= \frac{n(n-1)}{2} \left(\frac{3}{2}\right)^s \sum_{k=1}^s \binom{s}{k} \left(\frac{4}{9}a^2\right)^k,
 \end{aligned}$$

and

$$\begin{aligned} E[T(Z) - T(U)] &> \sum_{i < j}^n \left[\left(\frac{5}{4} + \frac{2}{3} a^2 \right)^s - \left(\frac{5}{4} \right)^s \right] \\ &= \frac{n(n-1)}{2} \left(\frac{5}{4} \right)^s \left[\left(1 + \frac{8}{15} a^2 \right)^s - 1 \right] \\ &= \frac{n(n-1)}{2} \left(\frac{5}{4} \right)^s \sum_{k=1}^s \binom{s}{k} \left(\frac{8}{15} a^2 \right)^k. \end{aligned}$$

This proves result (i) by noting expression (1). \square

References

- Box, G.E.P., 1963. The effect of error in the factor levels and experimental design. *Technometrics* 5, 247–262.
- Draper, N.R., Beggs, W.J., 1971. Errors in the factor levels and experimental design. *Ann. Math. Statist.* 41, 46–56.
- Fang, K.T., Lin, D.K.J., 2003. Uniform Experimental Design and its Applications in Industry. In: Rao, C.R., Khattree, R. (Eds.), *Handbook of Statistics in Industry*, vol. 22. North Holland, New York, pp. 131–170.
- Fang, K.T., Ma, C.X., Winker, P., 2002. Centered L_2 -discrepancy of random sampling and Latin hypercube design, and construction of uniform designs. *Math. Comp.* 71, 275–296.
- Hickernell, F.J., 1998a. A generalized discrepancy and quadrature error bound. *Math. Comp.* 67, 299–322.
- Hickernell, F.J., 1998b. Lattice rules: how well do they measure up. In: Hellekalek, P., Larcher, G. (Eds.), *Random and Quasi-Random Point Sets*. Springer, New York, pp. 106–166.
- McKay, M.D., Conover, W.J., Beckman, R.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- Owen, A.B., 1992. Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* 2, 439–452.
- Patterson, H.D., 1954. The errors of lattice sampling. *J. Roy. Statist. Soc. Ser. B* 16, 140–149.
- Wang, Y., Fang, K.T., 1981. A note on uniform distribution and experimental design. *KeXue TongBao* 26, 485–489.
- Winker, P., Lin, D.K.J., 2010. Robust uniform design with errors in the design variables. *Statist. Sinica* (forthcoming).