对称因子设计的均匀性和正交性之间的关系

李丹　罗幼喜　覃红

（华中师范大学数学与统计学学院　武汉）

摘　要：本文主要讨论了部分因子设计中的均匀性，离散偏差被用来作为均匀性的测度。对于对称部分因子设计，我们研究了正交性和均匀性之间的联系，并且发现这些准则有很好的一致性。通过对均匀性地研究，为我们合理的解释正交性原则提供了更有力的依据。另外，我们也详细证明了正交性两个的准则之间是等价的，而且还得到了离散偏差的一个新的下界。

关键词：离散偏差，正交性，均匀性，均匀设计

1 引言

均匀设计（Wang and Fang, 1981; Fang and Wang, 1994; Fang et al. 2000）是使试验点均匀的分布在试验区域，它已广泛的应用于工业领域和科学研究中。均匀性是均匀设计中的一个非常重要的概念。最近，许多研究表明均匀性在因子设计中是非常有用的。Fang et al.（2000）在中心化 L_2 -偏差（Hickernell, 1998）下给出了正交设计和均匀设计之间的联系，并且猜想任何正交设计在一定的均匀性的测度下是一个均匀设计。Ma et al.（2003）证明了如果 $q=2$ 或 q 是奇数时，对于一个全面试验 q^s，这个猜想是成立的。对于部分因子设计，Fang, Ma and Mukerjee（2002）给出了关于正交性和均匀性之间联系的结果。正交性和均匀性之间的联系为我们使用均匀设计提供了一个合理的依据。

最近，Heckernell and Liu（2002）提出了一种更为合理的均匀性的测度，即离散偏差（Discrete Discrepancy，简记为 DD）。与其它均匀性测度相比，DD 不仅大大的降低了计算费用，而且有它自身的统计合理性。Heckernell and Liu（2002）表明在 DD 度量下的均匀设计限制了效应的混杂。Liu and Heckernell（2002）使用 DD 去评判超饱和试验设计，结果表明对于 2-水平超饱和设计，DD 和 $E(s^2)$ 准则得到的最优超饱和设计是一样的。Liu（2002）用离散偏差的投影去计算对称部分因子设计的均匀性，发现均匀性和正交性之间有很好的联系。Qin and Fang（2003）给出了在 DD 度量下均匀性和其它准则（如广义最小混杂（Xu and Wu, 2001），最

小矩混杂（Xu, 2003）等）之间的联系，这些密切的联系非常有意义，它为 DD 作为因子设计的均匀性测度提供了更有力的说服力。除此之外，Qin and Fang（2003）证明了对称饱和正交设计和一类特殊的对称超饱和设计在 DD 度量下都是均匀设计。

Fang, Ma and Mukerjee（2002）首次在 2-水平或 3-水平的因子设计中建立了正交性和均匀性之间的解析联系。本文的目的在于把 Fang, Ma and Mukerjee（2002）的结果推广到更一般的对称因子设计。首先，DD 被用来作为均匀性测度，它可以被看成是 Fang, Ma and Mukerjee（2002）中讨论的均匀性测度的一般化，参看 Qin and Fang（2003）中的定理 5。其次，这篇文章中的许多结果可以化简为 2-水平或 3-水平因子设计的情况，而且结果与 Fang, Ma and Mukerjee（2002）中的结果是一致的。第三，这篇文章与 Liu and Hickernell (2002) 和 Liu（2002）不同的是，我们建立了正交性和均匀性之间的解析联系。最后，DD 度量下的均匀性不能用 Fang, Ma and Mukerjee (2002)中给出的形式来表达，因此，Fang, Ma and Mukerjee（2002）证明所使用的方法在这里不用运用，我们用到了一些新的方法。

2 几个比较准则

设 $D(n;q^s)$表示有 n 个试验且有 s 个 q 水平因素的设计的集合。

本文用 DD 来度量均匀性。在 Hilbert 空间里，我们用再生核来定义 DD（见 Hickernell and Liu, 2002），但不同的再生核会导致不同的偏差，见 Liu and Hickernell (2002), Liu (2002)和 Qin and Fang (2003)。这里我们用到了 Qin and Fang (2003)中给出的再生核，相应的离散偏差可参考他们的文章。对于给定的一个设计 d$\in D(n;q^s)$，它的离散偏差值，记为$D(d;a,b)$，可用如下公式得到：

$$[D(d;a,b)]^2 = -\left[\frac{a+(q-1)b}{q}\right]^s + \frac{a^s}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(\frac{b}{a}\right)^{\psi_{kl}}, \qquad (1)$$

这里的a和b是常数，而且$a>b>0$，ψ_{kl}是d的第k行和第l行之间的 Hamming 距离。均匀性准则是寻找一个达到最小$D(d;a,b)$值的设计。

正交设计应用于很多的领域，并形成因子设计的一个主要类。一个设计 d$\in D(n;q^s)$称为正交设计，如果任两个因子的水平组合出现相同的次数。正交设计是正交正列的一种特殊的情况。一个由 q 个字符组成的$n\times s$距阵 d 为强度为 m 的正交正列，记为 OA(n,s,q,m), 若它的任何$n\times m$子阵，所有q^m个水平组合出现的次

数都是n/q^m。关于正交正列更详细的讨论，我们可参考 Heclayat. Sloane and Stufken (1999)。

最近，一些新的准则，比如 B-准则(Fang, Lu and Winker, 2003)和 O-准则(Fang, Ma and Mukerjee, 2002)，已经被用来度量因子设计的正交性。这些准则可以看作为正交正列中强度的概念的一种推广。我们下面简短的描述一下 B-准则和 O-准则。

对于 d$\in D(n;q^s)$的某 m 列，记为$(d_{l_1}, d_{l_2}, \cdots, d_{l_m})$，定义

$$B_{l_1 \cdots l_m}(d) = \sum_{\alpha_1,\ldots,\alpha_m} (n_{\alpha_1 \ldots \alpha_m}^{(l_1 \ldots l_m)} - n/q^m)^2, \qquad (2)$$

其中$n_{\alpha_1 \ldots \alpha_m}^{l_1 \ldots l_m}$表示$(d_{l_1}, d_{l_2}, \cdots, d_{l_m})$中水平组合为$(\alpha_1 \ldots \alpha_m)$出现的次数。这里的求和针对所有$q^m$种水平组合。从定义可以看出，当$B_{l_1 \ldots l_m}(d) = 0$时，则由$(d_{l_1}, d_{l_2}, \cdots, d_{l_m})$形成的子设计是以强度为 m 的正交正列，因此，$B_{l_1 \ldots l_m}(d)$度量了$(d_{l_1}, d_{l_2}, \cdots, d_{l_m})$与强度为 m 的正交正列的接近的程度。进一步定义

$$B_m(d) = \sum_{1 \le l_1 \le \ldots \le l_m \le s} B_{l_1 \ldots l_m}(d) \Big/ \binom{s}{m} \qquad (3)$$

其中$1 \le m \le s$。显然，$B_m(d) = 0$当且仅当d为强度为 m 的正交正列。因此$B_m(d)$从总体上度量了设计d与强度为 m 的正交正列接近的程度。在 Fang, Lu and Winker（2003）中把向量（$B_1(d), B_2(d), \ldots, B_s(d)$）称为一个平衡型（Balance Pattern）。为了保证我们所选取的设计接近强度比较高的正交正列，我们需要选择d使得$B_1(d), B_2(d), \ldots, B_s(d)$序贯的小。我们称这种准则为 B-准则。

为了比较部分因子设计的正交性，Fang, Ma and Mukerjee（2002）提出了另一个向量（本文记为（$O_1(d), \ldots, O_s(d)$））来度量 d 的正交性。$O_i(d)$的详细定义可参看 Fang, Ma and Mukerjee（2002）。$\sum_{i=1}^m O_i(d)$可以度量d是否接近强度为 m 的正交列。$\sum_{i=1}^m O_i(d)$越小，则d的正交性就越好。因此，为了保证我们所选取的设计具有最好的正交性，我们应该选择合适的 d 来使$O_1(d), \ldots, O_s(d)$序贯的小，我们称

这个准则为 O-准则。在下一节，我们将证明 O-准则和 B-准则是相互等价的。

3 主要的结果

首先，我们给出下面的引理，这个引理建立了$\{B_j(d)\}$和$\{O_j(d)\}$之间的解析关系。

引理 1：设 $d\in D(n;q^s)$。那么，对任意的$1\le j\le s$和$1\le i\le s$，有

$$B_j(d)=q^{s-j}\sum_{v=1}^{j}\binom{s-v}{s-j}O_v(d)\Big/\binom{s}{j} \tag{4}$$

$$O_i(d)=\frac{1}{q^s}\binom{s}{j}\sum_{j=1}^{i}(-1)^{i-j}q^j\binom{i}{j}B_j(d) \tag{5}$$

根据引理 1，我们有如下定理：

定理 1：对于水平为 q 的对称设计 d, B-准则和 O-准则是一致的，也就是说，使$B_1(d),B_2(d),\ldots,B_s(d)$序贯的小等价于使$O_1(d),\ldots,O_s(d)$序贯的小，并且，$B_j(d)$'s 和 $O_j(d)$'s 通过（3）或（4）是线性相关的。

现在，我们考虑离散偏差度量下的均匀性和$\{O_i(d)\}$或$\{B_i(d)\}$度量下的正交性之间的关系。对于 2 水平或 3 水平因子设计，Fang, Ma and Mukerjee (2002）建立了在一些偏差度量下的均匀性和$\{O_i(d)\}$度量下的正交性之间的关系。对于 DD 度量下的均匀性，我们有更一般的结果：

定理 2：设 $d\in D(n;q^s)$,

$$[D(d;a,b)]^2=\frac{[a+(q-1)b]^s}{n^2}\sum_{v=1}^{s}\left[\frac{a-b}{a+(q-1)b}\right]^v O_v(d) \tag{6}$$

$$=\frac{b^s}{n^2}\sum_{v=1}^{s}\binom{s}{v}\left(\frac{a-b}{b}\right)^v B_v(d) \tag{7}$$

这里 a,b 都是常数，且$a>b>0$.

注 1：因为$a>b>0$，$(a-b)/(a-b+qb)$是一个小数，因此，由等式（6）知，在$[D(d;a,b)]^2$ 中$O_v(d)$的系数随着指数v增加而递减。因此，对于对称部分因子d，若有较好的正交性，即对于较小的j有较小的$O_j(d)$，那么在$[D(d;a,b)]^2$ 度量下d就有较好的均匀性。显然，由定理 1 和等式（7）可得在$B_j(d)$度量下的正交性和均匀性之间有类似的结论。

推论 1：对任一设计$d\in D(n;q^s)$，我们有下列等式成立：

(i) 当$q=2, a=5/4$以及$b=1$时，

$$[D(d;a,b)]^2=\frac{1}{n^2}\left(\frac{9}{4}\right)^s\sum_{j=1}^{s}\frac{O_j(d)}{9^j}=\frac{1}{n^2}\sum_{v=1}^{s}\binom{s}{v}\frac{B_v(d)}{4^v}\ ;$$

(ii) 当$q=2, a=3/2$以及$b=5/4$时；

$$[D(d;a,b)]^2=\frac{1}{n^2}\left(\frac{11}{4}\right)^s\sum_{j=1}^{s}\frac{O_j(d)}{11^j}$$

$$=\frac{1}{n^2}\left(\frac{5}{4}\right)^s\sum_{v=1}^{s}\binom{s}{v}\frac{B_v(d)}{5^v}\ ;$$

(iii) 当$q=3, a=3/2$以及$b=23/18$时；

$$[D(d;a,b)]^2=\frac{1}{n^2}\left(\frac{73}{18}\right)^s\sum_{j=1}^{s}\left(\frac{4}{73}\right)^jO_j(d)$$

$$=\frac{1}{n^2}\left(\frac{23}{18}\right)^s\sum_{v=1}^{s}\binom{s}{v}\left(\frac{4}{23}\right)^vB_v(d).$$

注２：由 Qin and Fang (2003)中的定理５可以清楚的知道，推论１是 Fang, Ma and Mukerjee (2002) 中结果的另一种表达形式,因此，我们的结果（定理 2）是 Fang, Ma and Mukerjee（2002）的一种推广。

4 离散偏差的一个新的下界

我们现在给出（1）式中$\left[D(d;a,b)\right]^2$ 的一个下界.

定理 3：假定$d \in D(n;q^s)$,那么$\left[D(d;a,b)\right]^2 \ge L^C(d;a,b)$， (8)

其中 $L^C(d;a,b)=\frac{b^s}{n^2}\sum_{v=1}^{s}\binom{s}{v}\left(\frac{a-b}{b}\right)^v R_{n,v,q}\left(1-R_{n,v,q}/q^v\right)$, $R_{n,v,q}$是$n\left(\bmod q^v\right)$的余数。

注 3：$L^C(d;a,b)$这个下界是根据d的列平衡（Column Balance）而得到的,它在评估一般的正交列的均匀性时非常的有用。Qin and Fang（2003）根据d的行之间的 Hamming 距离给出的$\left[D(d;a,b)\right]^2$ 的另一个下界：

$$L^S(d;a,b)=-\left[\frac{a+(q-1)b}{q}\right]^s+\frac{a^s}{n}+\frac{(n-1)\left[b(\gamma+1-\psi)+a(\psi-\gamma)\right]b^s}{nb}\left(\frac{a}{b}\right)^{\gamma},$$

这里，$\psi=s(n-q)/(q(n-1))$，γ是ψ的整数部分，显然，$L^C(d;a,b)$比$L^S(d;a,b)$更精确，后者在一些饱和或超饱和设计中能够达到，所以$L^S(d;a,b)$在评估d是否接近饱和或超饱和设计时更有用。

我们现在用一个例子来解释我们的结论。

表 1：2^5因子设计的 12 次试验

d_1					d_2					d_3				
1	1	0	1	1	1	1	0	1	1	1	0	0	1	0
0	1	1	0	0	0	1	1	0	1	1	0	1	1	0
1	0	1	1	0	1	0	1	1	0	1	1	0	1	1
0	1	0	1	0	0	1	0	1	1	1	1	1	1	1
0	0	1	0	1	0	0	1	0	1	1	0	0	0	0
0	0	0	1	1	0	0	0	1	0	1	0	0	1	1
1	0	0	0	1	1	0	0	0	1	0	1	1	0	0
1	1	0	0	0	1	1	0	0	0	0	1	1	1	1
1	1	1	0	1	1	1	1	0	0	0	1	1	0	1
0	1	1	1	1	0	1	1	1	0	0	1	0	0	1

1	0	1	1	0	1	0	1	1	1	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

例 1：取 $n=12, q=2, s=5$. 表 1 给出了 3 个非正规部分因子设计 d_1, d_2, d_3,其中 d_1 和 d_2 是从 12 阶的 Hadamard 距阵中选出的不同构的 5 列，参看 Lin and Draper(1992)。表 2 给出了一些数字结果。若从一些常用偏差度量下的均匀性和正交性的准则来看，Fang, Ma and Mukerjee（2002）推断 d_2 比 d_1 要稍微好一点，而且 d_1 和 d_2 要比 d_3 好。

容易验证

$$[D(d_2;a,b)]^2 < [D(d_1;a,b)]^2 < [D(d_3;a,b)]^2$$

因此,这也支持了 Fang, Ma and Mukerjee（2002）中的结果。

表 2：例 1 的数字结果

	d_1	d_2	d_3		d_1	d_2	d_3
$B_1(d)$	0	0	0	$O_1(d)$	0	0	0
$B_2(d)$	0	0	4.8	$O_2(d)$	0	0	6
$B_3(d)$	2	2	7.2	$O_3(d)$	5	5	0
$B_4(d)$	5	5	7.8	$O_4(d)$	2.5	2.5	1.5
$B_5(d)$	9.5	7.5	7.5	$O_5(d)$	2	0	0

$$[D(d_1;a,b)]^2 = a^5(9\beta^2+12\beta+19)(1-\beta)^3/288$$

$$[D(d_2;a,b)]^2 = 5a^5(\beta^2+4\beta+3)(1-\beta)^3/288$$

$$[D(d_3;a,b)]^2 = a^5(5\beta^3+11\beta^2+11\beta+5)(1-\beta)^2/96$$

$\beta = b/a, a > b > 0$

致谢：感谢方开泰教授、谢民育教授同作者们有益的讨论。本文得到了教育部留学回国人员启动基金和湖北省自然科学基金的支持。

参考文献：

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

[2] Fang, K.T., Lu, X., Winker, P. (2003). Lower bounds for centered and wrap-around L_2-discrepancies and construction of uniform designs by threshold accepting. *J. Complexity*, to appear

[3] Fang, K.T., Ma, C. X. and Mukerjee, R. (2002). Uniformity in fractional factorials. In: *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, eds. by K. T. Fang, F. J. Hickernell, and H. Niederreiter, Springer-Verlag, Berlin.

[4] Fang, K.T. and Wang, Y. (1994). *Number-Theoretic Methods in Statistics*. London: Chapman and Hall.

[5] Hedayat, A. S., Sloane, N. J. and Stufken, J. (1999). *Orthogonal arrays: theory and application*. Springer.

[6] Hickernell, F. J. (1998). A generalized discrepancy and quadrature error bound. *Math. Comp*. **67**, 299-322.

[7] Hickernell, F. J. and Liu, M. Q. (2002). Uniform designs limit aliasing. *Biometrika*, **89**, 893-904.

[8] Lin, D. K. J. and Draper, N. R. 91992). Projection properties of Plackett and Burman designs. *Technometrics* **34**, 423-428.

[9] Liu, M. Q. (2002). Using discrepancy to evaluate fractional factorial designs. In: *Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, eds. by K. T. Fang, F. J. Hickernell, and H. Niederreiter, Springer-Verlag, Berlin.

[10] Liu, M. Q. and Hickernell, F. J. (2002). $E\left(s^2\right)$-optimality and minimum discrepancy in 2-level supersaturated designs. *Statist. Sinica* **12**, 931-939.

[11] Ma, C. X., Fang, K. T. and Lin, D. K. J. (2003). A note on uniformity and orthogonality, *J. Statist. Planning and Inference*, **113**,323-334.

[12] MacWilliams, F. J. and Sloane. N. J. A. (1977). The theory of error-correcting codes. North-Holland. Amsterdam.

[13] Qin, H. and Fang, K. T. (2003). Discrete discrepancy in factorial designs. *Metrika*, to appear.

[14] Wang, Y. and Fang, K. T. (1981). A note on uniform distribution and experimental design. *Chin. Sci. Bull*. **26**, 485-489.

[15] Tang, B. (2001). Theory of J-characteristics for fractional factorial designs and projection justification of minimum G_2-aberration. *Biometrika* **88**, 401-407.

[16] Wang, J. C. and Wu, C. F. J. (1995). A hidden projection property of Plackett-Burman and ralated designs. *Statistica Sinica* **5**,235-250.

[17] Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for a symmetrical fractional factorial designs. *Ann. Statist*. **29**,549-560.