Statistica Sinica **14**(2004), 1069-1074

A NOTE ON REGULAR FRACTIONAL FACTORIAL DESIGNS

Kenny Q. Ye

Albert Einstein College of Medicine

Abstract: This paper shows that any two-level factorial design that has no partial aliasing must be a 2^{k-p} design or replicates of a 2^{k-p} design. This is a special case of a more general result regarding q-level factorial designs, where q is a prime power, under a certain parameterization of the factorial effects.

Key words and phrases: Indicator function, orthogonal array.

1. Regular and Non-regular Designs

Factorial designs are widely used in industrial experiments. The most popular ones are 2^{k-p} fractional factorial designs. A 2^{k-p} design is a fraction of the 2^k full factorial design, generated by p *generators* (Box, Hunter and Hunter (1978, p.383). Its aliasing structure is explicitly described by the *defining contrast group* (Wu and Hamada (2000)) generated by the p generators. Conventionally, many authors refer to these designs as *regular designs* in contrast to the designs that cannot be generated through this method. Among designs not of 2^{k-p} type are most of the Plackett-Burman designs and John's 3/4 fractions (John (1971)). Some authors refer to these designs as *irregular fractions* (Montgomery (2000)). A formal definition of *regular fractional* q^{k-p} designs, where q is a prime power, can be found in Dey and Mukerjee (1999). Such a fraction is also generated by p generators.

A very important feature of 2^{k-p} designs is that two factorial effects are either estimated independently or are fully aliased. Such a property is not found in two-level orthogonal arrays not of 2^{k-p} type, including most of the Plackett-Burman designs. For those designs, some factorial effects are partially aliased. Wu and Hamada (2000) gave another definition of *regular design* based on this property since it is essential to the application of the designs and the corresponding data analysis strategy. They call a design *regular* if for this design any two factorial effects either can be estimated independently or are fully aliased. Those that do not have this property are called *non-regular*. They choose the term "non-regular" over "irregular" because the latter sounds negative. Based on this definition, 2^{k-p} designs are regular designs and other known two-level orthogonal designs are non-regular. However, it is unclear whether these two definitions of *regular designs* are equivalent. That is, whether or not a two-level orthogonal design without partial aliasing must be a 2^{k-p} design. Theoretically, this question is rather fundamental. Practically, one might be interested in orthogonal designs without partial aliasing at run sizes other than powers of 2 if such designs exist.

For q^{k-p} regular designs in which $q > 2$, the factorial effects are not always independent depending on the parameterization system. Wu and Hamada (2000) point out that 3^{k-p} designs "can be treated as regular designs or non-regular designs ... If the orthogonal component system is used, a 3^{k-p} design is a regular design ... If the linear-quadratic system is used ... some effects in the system have absolute correlation between 0 and 1 and so the corresponding 3^{k-p} design should be treated as a non-regular design".

The main purpose of this paper is to show that the two definitions of regular designs are equivalent for two-level factorial designs. For q^{k-p} factorial designs, it is also true under orthogonal component decomposition.

2. The Main Result

Fontana, Pistone and Rogantin (2000) proved that a two-level factorial design that has no replicates must be of 2^{k-p} type if it has no partial aliasing. However, their proof does not apply directly to the designs with replicated runs. The main purpose of this paper is to show that all two-level factorial designs without partial aliasing must be a 2^{k-p} design or replicates of a 2^{k-p} design. Therefore, the two definitions mentioned above are equivalent. Furthermore, the result presented here is for general q-level factorial designs, where q is a prime power, of which the two-level designs are special cases.

Before the main theorem is presented, consider *indicator functions* that are an essential tool in the proof. They were first introduced by Fontana, Pistone and Rogantin (2000) to represent fractional factorial designs. Ye (2003) further extended them to more general case that allows replicates. The following definition is given for the general q-level designs but can be easily extended to more general mixed-level designs.

Definition 1. Let $\mathcal{D}$ be a full q^k design and let $\mathcal{A}$ be a design such that $\forall\, \mathbf{a} \in \mathcal{A}$, $\mathbf{a} \in \mathcal{D}$ but $\mathbf{a}$ might be repeated in $\mathcal{A}$. The indicator function of $\mathcal{A}$ is a function defined on $\mathcal{D}$, such that

$$F(\mathbf{x}) = \begin{cases} r_{\mathbf{x}} & \text{if } \mathbf{x} \in \mathcal{A}, \\ 0 & \text{if } \mathbf{x} \notin \mathcal{A}, \end{cases}$$

where $r_{\mathbf{x}}$ is the number of appearances of point $\mathbf{x}$ in design $\mathcal{A}$.

Traditionally, the levels of a q-level design are coded to be elements of the additive group $\mathbb{Z}_q$. Bailey (1982) used an equivalent coding system based on the multiplicative group of the roots of unity. Let the q levels of a factor be $1, e^{i\frac{2\pi}{q}}, \ldots, e^{i\frac{(q-1)2\pi}{q}}$, i.e., evenly spaced solutions of $z^q = 1$ on the unit circle in the complex plane $\mathbb{C}$. It is easy to see that when $q = 2$, they reduce to -1 and 1. Following the convention, the solutions of $z^q = 1$ are denoted by $\omega^j = e^{i\frac{2\pi j}{q}}$. The indicator function of a q-level factorial design can be represented as a polynomial on $\mathbb{C}^k$.

Define functions $Z_{\mathbf{t}}(\mathbf{z}) = \prod_{i=1}^k z_i^{t_i}$ for $\mathbf{t} \in \mathcal{T}$, where $\mathcal{T}$ is the k-fold product $\mathbb{Z}_q \times \mathbb{Z}_q \cdots \times \mathbb{Z}_q$. Note that $\mathcal{T}$ is a k-dimensional vector space on $\mathbb{Z}_q$, where vector addition and scalar multiplication in $\mathcal{T}$ are defined as usual. For vectors $\mathbf{t}_1$ and $\mathbf{t}_2$, we denote $\mathbf{t}_1 + \mathbf{t}_2 (\bmod q)$ by $\mathbf{t}_1 + \mathbf{t}_2$ whenever there is no confusion. For convenience, denote $(q-1)\mathbf{t}$ by $-\mathbf{t}$. Note that $Z_{\mathbf{0}}$ is the constant function and, for $\mathbf{t} \neq \mathbf{0}$, $\{Z_{\mathbf{t}}, Z_{2\mathbf{t}}, \ldots, Z_{(q-1)\mathbf{t}}\}$ is a set of contrasts corresponding to a factorial effect, which has $q-1$ degrees of freedom. For example, the main effect of factor z_1 corresponds to $\{z_1, z_1^2, \ldots, z_1^{q-1}\}$. Since $\{Z_{\mathbf{t}}, \mathbf{t} \in \mathcal{T}\}$ forms a basis of $\mathcal{D}$, the indicator function of a design $\mathcal{A}$, which has n runs and k factors, has a unique representation of the form

$$F_{\mathcal{A}}(\mathbf{z}) = \sum_{\mathbf{t} \in \mathcal{T}} b_{\mathbf{t}} Z_{\mathbf{t}}(\mathbf{z}). \tag{1}$$

In particular, $b_{\mathbf{0}} = n/q^k$ and

$$b_{\mathbf{t}}/b_{\mathbf{0}} = \sum_{\mathbf{z} \in \mathcal{A}} \overline{Z_{\mathbf{t}}(\mathbf{z})}/n = \sum_{\mathbf{z} \in \mathcal{A}} Z_{-\mathbf{t}}(\mathbf{z})/n. \tag{2}$$

Since indicator functions only take integer values, $F_{\mathcal{A}}(\overline{\mathbf{z}}) = F_{\mathcal{A}}(\mathbf{z})$. By comparing the coefficients, it is easy to see that $\overline{b_{\mathbf{t}}} = b_{-\mathbf{t}}$. Note that in (2) and throughout this paper, $\sum_{\mathbf{z} \in \mathcal{A}}$ denotes the summation over all design points in $\mathcal{A}$: if $\mathbf{z} \in \mathcal{D}$ repeats r times in $\mathcal{A}$, it is summed r times.

Let $\mathcal{A}$ be a regular q^{k-p} design that is generated by p generators $G_1, \ldots, G_p$. Each generator is a monomial

$$G_i(z_1, \ldots, z_k) = \prod_{j=1}^k z_j^{s_j},$$

where $0 \le s_1, \ldots, s_k < q$. Without loss of generality, all design points $\mathbf{z} \in \mathcal{A}$ satisfy $G_i(\mathbf{z}) = 1$ for $i = 1, \ldots, p$. It can be easily verified that the indicator function of $\mathcal{A}$ has the following polynomial form:

$$F_{\mathcal{A}}(\mathbf{z}) = \frac{1}{q^p} \prod_{i=1}^p \prod_{j=1}^{q-1} (G_i(\mathbf{z}) - \omega^j). \tag{3}$$

For a design point $\mathbf{z} \notin \mathcal{A}$, there must exist a generator G_{i_0} such that $G_{i_0}(\mathbf{z}) \neq 1$, i.e., $G_{i_0}(\mathbf{z}) = \omega^{j_0}$ for some $j_0 \neq 0$; hence the right side of (3) is 0. On the other hand, for a design point $\mathbf{z}_0 \in \mathcal{A}$, $G_i(\mathbf{z}_0) = 1$ for all $i = 1, \ldots, p$. Therefore the right hand of (3) is

$$\Big(\frac{1}{q}(\prod_{j=1}^{q-1}(1-\omega^j)\Big)^p,$$

which equals 1 since

$$\prod_{j=1}^{q-1}(1-\omega^j) = \lim_{z\to 1}\frac{z^q-1}{z-1} = \lim_{z\to 1}(1+z+z^2+\cdots+z^{q-1}) = q.$$

This proves (3).

Consider contrasts $Z_{\mathbf{t}_1}$ and $Z_{\mathbf{t}_2}$. They are estimated independently if $\sum_{\mathbf{z}\in\mathcal{A}} Z_{\mathbf{t}_1}(\mathbf{z})\overline{Z_{\mathbf{t}_2}(\mathbf{z})} = \sum_{\mathbf{z}\in\mathcal{A}} \overline{Z_{\mathbf{t}_1}(\mathbf{z})}Z_{\mathbf{t}_2}(\mathbf{z}) = 0$, i.e., $b_{\mathbf{t}_1-\mathbf{t}_2}/b_{\mathbf{0}} = 0$; they are fully aliased if $(1/n)\sum_{\mathbf{z}\in\mathcal{A}} Z_{\mathbf{t}_1}(\mathbf{z})\overline{Z_{\mathbf{t}_2}(\mathbf{z})} = \omega$, where ω is a root of $z^q = 1$. In the later case, $\|b_{\mathbf{t}_1-\mathbf{t}_2}/b_0\| = 1$. Since $\prod_{j=1}^{q-1}(z-\omega^j) = ((z^q-1)/(z-1)) = 1+z+z^2+\cdots+z^{q-1}$, the indicator function of a regular fractional factorial design (3) can be written as

$$F_{\mathcal{A}}(\mathbf{z}) = \frac{1}{q^p}\prod_{i=1}^{p}(1+G_i(\mathbf{z})+G_i(\mathbf{z})^2+\cdots+G_i(\mathbf{z})^{q-1}). \tag{4}$$

From the above equation, the coefficients of $F_{\mathcal{A}}$ satisfy $b_{\mathbf{t}}/b_{\mathbf{0}} = 0$ or 1.

The general definition of regular designs allows the generating relations to be $G_i(\mathbf{z}) = \omega^{j_i}$. In such cases, the indicator function of the design can be written as

$$F_{\mathcal{A}}(\mathbf{z}) = \frac{1}{q^p}\prod_{i=1}^{p}\omega^{j_i}\frac{G_i(\mathbf{z})^q-1}{G_i(\mathbf{z})-\omega^{j_i}}. \tag{5}$$

It is not hard to show that in (5), $(b_{\mathbf{t}}/b_{\mathbf{0}})^q = 1$ or $b_{\mathbf{t}}/b_{\mathbf{0}} = 0$ for all $\mathbf{t} \in \mathcal{T}$. The proof is omitted here as it is not essential for this paper.

Now, we are ready to present and prove the main theorem of the paper.

Theorem 2. *Let $\mathcal{A}$ be an $n\times k$ q-level factorial design and $F_{\mathcal{A}}(\mathbf{z}) = \sum_{\mathbf{t}\in\mathcal{T}} b_{\mathbf{t}}Z_{\mathbf{t}}(\mathbf{z})$ be its indicator function. If $\|b_{\mathbf{t}}/b_{\mathbf{0}}\| = 1$ or $b_{\mathbf{t}}/b_{\mathbf{0}} = 0$ for all $\mathbf{t} \in \mathcal{T}$, then $\mathcal{A}$ is a q^{k-p} design or replicates of a q^{k-p} design.*

Proof. Let $\mathcal{L} = \{\mathbf{t} \in \mathcal{T}, b_{\mathbf{t}} \neq 0\}$. Consider a factorial contrast $Z_{\mathbf{t}}$, $\mathbf{t} \in \mathcal{L}$. Notice that $\sum_{\mathbf{z}\in\mathcal{A}} Z_{\mathbf{t}}(\mathbf{z}) = n(b_{-\mathbf{t}}/b_{\mathbf{0}})$. Hence

$$\sum_{\mathbf{z}\in\mathcal{A}}(\frac{Z_{\mathbf{t}}(\mathbf{z})}{b_{-\mathbf{t}}/b_{\mathbf{0}}}-1) = \frac{\sum_{\mathbf{z}\in\mathcal{A}} Z_{\mathbf{t}}(\mathbf{z}) - n(b_{-\mathbf{t}}/b_{\mathbf{0}})}{(b_{-\mathbf{t}}/b_{\mathbf{0}})} = 0. \tag{6}$$

Since $Z_{\mathbf{t}}$ takes its value only among the ω^j's, it is easy to see that $\Re(Z_{\mathbf{t}}/(b_{-\mathbf{t}}/b_{\mathbf{0}}))$ ≤ 1 on $\mathcal{A}$ and the equality holds if and only if $Z_{\mathbf{t}}/(b_{-\mathbf{t}}/b_{\mathbf{0}}) \equiv 1$ on $\mathcal{A}$. Therefore, from (6), $Z_{\mathbf{t}} \equiv b_{-\mathbf{t}}/b_{\mathbf{0}}$ on $\mathcal{A}$. Consider $\mathbf{t}_1, \mathbf{t}_2 \in \mathcal{L}$. We have $Z_{-\mathbf{t}_i} \equiv b_{\mathbf{t}_i}/b_0$ on $\mathcal{A}$ and $\|b_{\mathbf{t}_i}/b_0\| = 1$ for $i = 1, 2$. Notice that

$$\frac{b_{(\mathbf{t}_1+\mathbf{t}_2)}}{b_0} = \frac{1}{n}\sum_{\mathbf{z}\in\mathcal{A}} Z_{-(\mathbf{t}_1+\mathbf{t}_2)}(\mathbf{z}) = \frac{1}{n}\sum_{\mathbf{z}\in\mathcal{A}} Z_{-\mathbf{t}_1}(\mathbf{z})Z_{-\mathbf{t}_2}(\mathbf{z}) = (\frac{b_{\mathbf{t}_1}}{b_0})(\frac{b_{\mathbf{t}_2}}{b_0}).$$

Therefore, $\mathbf{t}_1 + \mathbf{t}_2 \in \mathcal{L}$. Hence the subset $\mathcal{L}$ is closed under addition and forms a subgroup of $\mathcal{T}$. Let l be the cardinality of $\mathcal{L}$. For $\mathbf{z} \in \mathcal{A}$, consider

$$\begin{aligned} F_{\mathcal{A}}(\mathbf{z}) &= \sum_{\mathbf{t}\in\mathcal{L}} b_{\mathbf{t}} Z_{\mathbf{t}}(\mathbf{z}) = b_{\mathbf{0}} \sum_{\mathbf{t}\in\mathcal{L}} (b_{\mathbf{t}}/b_{\mathbf{0}}) Z_{\mathbf{t}}(\mathbf{z}) \\ &= b_{\mathbf{0}} \sum_{\mathbf{t}\in\mathcal{L}} Z_{-\mathbf{t}}(\mathbf{z}) Z_{\mathbf{t}}(\mathbf{z}) = b_{\mathbf{0}} l. \end{aligned} \tag{7}$$

From the above equation, $F_{\mathcal{A}} \equiv lb_{\mathbf{0}} = ln/q^k$ on $\mathcal{A}$. Because $F_{\mathcal{A}}(\mathbf{x})$ only takes positive integers on $\mathcal{A}$, both l and n must be powers of q. Let $l = q^p$. Then the subgroup $\mathcal{L}$ has p generators. When $nl/q^k = 1$, $n = q^{k-p}$ and the design is a q^{k-p} design with q^p generating words in $\mathcal{L}$. When $nl/q^k = r > 1$, $F_{\mathcal{A}}(\mathbf{z}) = r$ on $\mathcal{A}$. The design is r replicates of a q^{k-p} design defined by the defining words in $\mathcal{L}$. In this case, $n = rq^{k-p}$.

For a two-level design $\mathcal{A}$, if $\sum_{\mathbf{x}\in\mathcal{A}} X_I(\mathbf{x})X_J(\mathbf{x})/n$ equals 0 or ± 1 for all factorial contrasts X_I and X_J, then $\mathcal{A}$ must be a 2^{k-p} type design or replicates of such a design. Therefore, there is no two-level design of run size other than powers of 2 that has only full aliasing, unless the design is replicates of a 2^{k-p} fractional factorial design. For example, a 20 by 3 orthogonal design must contain partial aliasing unless it is a 2^{3-1} design replicated five times. A 20 by 4 orthogonal design must contain partial aliasing because there is no 2^{4-2} design of resolution III or higher.

3. Concluding Remarks

The main objective of this paper is to show that two definitions of regular designs, one from a construction perspective, the other from a data analysis perspective, are equivalent for two-level factorial designs. This is shown to be a special case of a general result of q-level designs, where q is a prime power. However, the result regarding q-level designs only applies to a special decomposition in which the q levels are evenly spaced on the unit circle in $\mathbb{C}$. For other decomposition systems, such as the linear-quadratic system, the two definitions are not equivalent.

Acknowledgement

This research was supported in part by National Science Foundation grant DMS-0306306. The author would like to thank an anonymous reviewer who pointed out many mistakes in the earlier versions of the paper. The author would also like to thank Professors Ching-Shui Cheng and C. F. Jeff Wu for their encouragement.

References

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters.* Wiley, New York.

Bailey, R. A. (1982). The Decomposition of Treatment Degrees of Freedom in Quantitative Factorial Experiments. *J. Roy. Statist. Soc. Ser. B* **44**, 63-70.

Fontana R., Pistone, G. and Rogantin, M. P. (2000). Classification of two-level factorial fractions. *J. Statist. Plann. Inference* **87**, 149-172.

John, P. W. M. (1971). *Statistical Design and Analysis of Experiments.* SIAM Classics in Applied Mathematics, Philadelphia, PA.

Montgomery, D. C. (2000). *Design and Analysis of Experiments.* 5th Edition. Wiley, New York.

Wu, C. F. J. and Hamada, M. (2000). *Experiments: Planning, Analysis, and Parameter Design Optimization.* Wiley, New York.

Ye, K. Q. (2003). Indicator functions and its application in two level factorial design. *Ann. Statist.* **31**, 984-994.

Albert Einstein College of Medicine, Bronx, New York 10461, U.S.A.

E-mail: kye@aecom.yu.edu

(Received January 2003; accepted January 2004)