Journal of Statistical Planning and Inference 141 (2011) 3055–3062

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Isomorphism check in fractional factorial designs via letter interaction pattern matrix

Yang Liu [a], Jian-Feng Yang [b,*], Min-Qian Liu [b]

[a] Department of Statistics, University of British Columbia, Vancouver, BC, Canada V6T 1Z2
[b] Department of Statistics, School of Mathematical Sciences and LPMC, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 29 October 2010
Received in revised form
21 March 2011
Accepted 29 March 2011
Available online 2 April 2011

Keywords:
Coset pattern matrix
Fractional factorial design
Isomorphism
Letter pattern matrix

ABSTRACT

Two fractional factorial designs are considered isomorphic if one can be obtained from the other by relabeling the factors, reordering the runs, and/or switching the levels of factors. To identify the isomorphism of two designs is known as an NP hard problem. In this paper, we propose a three-dimensional matrix named the letter interaction pattern matrix (LIPM) to characterize the information contained in the defining contrast subgroup of a regular two-level design. We first show that an LIPM could uniquely determine a design under isomorphism and then propose a set of principles to rearrange an LIPM to a standard form. In this way, we can significantly reduce the computational complexity in isomorphism check, which could only take $O(2^p)+O(3k^3)+O(2^k)$ operations to check two 2^{k-p} designs in the worst case. We also find a sufficient condition for two designs being isomorphic to each other, which is very simple and easy to use. In the end, we list some designs with the maximum numbers of clear or strongly clear two-factor interactions which were not found before.

1. Introduction

Fractional factorial designs, especially with two levels, are among the most popular experimental designs in practice. Two factorial designs are called isomorphic if one can be obtained from the other by relabeling the factors, reordering the runs and/or switching the levels of factors. Xu and Wu (2001) pointed out that two isomorphic designs have the same statistical properties in the classic ANOVA model. Thus they are considered to be equivalent in most studies in experimental designs. When constructing designs, isomorphic designs bring redundant computation work. Therefore, it is important and necessary to find a good method to check design isomorphism.

For two 2^{k-p} designs, which are expressed by their design matrices, a complete search for checking all possible reordering and relabeling needs $O(n!k!2^p)$ comparisons, where $n=2^{k-p}$ denotes the number of runs. As n or k increases, the computational complexity grows exponentially, which is known as an NP hard problem. For regular two-level fractional factorial designs, given by the defining contrast subgroup (DCS, Wu and Hamada, 2000), isomorphic designs only need to be detected through factor relabeling. The computational complexity, however, is still $O(k!)$ for a complete isomorphism check. For non-regular designs, all three kinds of transformations need to be checked. Hence, isomorphism check for non-regular designs is more complicated than that for regular designs. In all, an efficient method for checking factor relabeling is important for both regular and non-regular designs. In this paper, we only consider the case of regular two-level designs.

* Corresponding author. Tel.: +86 22 58189844; fax: +86 22 23506423.
E-mail address: jfyang@nankai.edu.cn (J.-F. Yang).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2011.03.027

Isomorphism check has received a great deal of attention in the literature, see Chen et al. (1993), Clark and Dean (2001), Ma et al. (2001), Sun et al. (2002), Block and Mee (2005), and Lin and Sitter (2008) for details. Traditionally, the word length pattern (WLP) $W=(A_1,A_2,\ldots,A_k)$ is used to characterize the aliasing pattern of a 2^{k-p} design d. However, it is easy to find non-isomorphic designs with the same word length pattern. Draper and Mitchell (1968) proposed the letter pattern matrix (LPM) $L=(l_{ij})$, where l_{ij} denotes the number of words that involve letter i and have length j, and conjectured that an LPM could uniquely determine a design under isomorphism. But a counterexample was found by Chen and Lin (1991). Then, Zhu and Zeng (2005) proposed the coset pattern matrix (CPM), which is a $2^{k-p}*k$ matrix including the word length pattern and all the $2^{k-p}-1$ cosets of the DCS. But it still fails to determine a design uniquely.

Here is an example to illustrate the definition of WLP, LPM and CPM.

Example 1. For the 2^{6-2} design d_1 defined by $I=123=1456=23456$, the WLP is the vector $W=(0,0,1,1,1,0)$ and the LPM is the 6×6 matrix:

1	2	3	4	5	6
0	0	1	1	0	0
0	0	1	0	1	0
0	0	1	0	1	0
0	0	0	1	1	0
0	0	0	1	1	0
0	0	0	1	1	0

where the ith row is the letter pattern of factor i. For example, the letter 1 appears in a word of length three (123) and a word of length four (1456), thus the (1,3) and (1,4) entries in the LPM are both 1.

There are 15 cosets of the DCS for design d_1. The coset involving a main effect is called a m.e. coset (Zhu and Zeng, 2005). The six m.e. cosets are

$$1=23=456=123{,}456,$$

$$2=13=3456=12{,}456,$$

$$3=12=2456=13{,}456,$$

$$4=156=1234=2356,$$

$$5=146=1235=2346,$$

$$6=145=1235=2345.$$

In each coset, the words are sorted according to their lengths, which follows relevant requirements in Zhu and Zeng (2005). The CPM for these m.e. cosets is as follows.

1	2	3	4	5	6
1	1	1	0	0	1
1	1	0	1	1	0
1	1	0	1	1	0
1	0	1	2	0	0
1	0	1	2	0	0
1	0	1	2	0	0

Take the first m.e. coset ($1=23=456=123456$) as an example. There is one word of length 1 (i.e., letter 1), one word of length 2 (i.e., 23), one word of length 3 (i.e., 456) and one word of length 6 (i.e., 123456), so the first three entries and the last one in the first row are 1's.

As for the computational complexity, Chen et al. (1993) checked all possible permutations of factors, which needs $\binom{k}{k-p}(k-p)!$ comparisons for two 2^{k-p} designs. Clark and Dean (2001) performed $k(k!)^2$ comparisons in the worst case and each comparison needs $O(n!)$ operations in the isomorphism check of two 2^{k-p} designs. Ma et al. (2001) requires $O(n^2k2^k)$ comparisons, each of which needs to compare 2^{k+1} centered L_2-discrepancy values. Lin and Sitter (2008) proposed a method combining word length pattern and eigenvalues, which is more efficient than the above methods in some occasions. But a major drawback of this method is that the computation of eigenvalue for an $m*n$ matrix is $O(mn^3)$, which will cause more computing burden for isomorphism check when the number of factors is relatively large.

As for the construction, a 2^{k-p} design could be obtained by selecting a subset from the column set $C=\{C_1,C_2,\ldots,C_{2^{k-p}-1}\}$, including $k-p$ independent columns and p additional columns, where C_i denotes the binary sequence of i. For example, C_1 is $(10000)'$, C_{31} is $(11111)'$ (see, for example, Chen et al., 1993; Zhang et al., 2008 for details). When k and p are relatively large, it is impractical to find non-isomorphic designs from all the $\binom{2^{k-p}-1-(k-p)}{p}$ possible designs constructed this