Statistics and Probability Letters 138 (2018) 82–89

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A note on extending the alias length pattern

Qi Zhou [a], Bing Guo [b,d,*], Shengli Zhao [c]

[a] Department of Statistics, Tianjin University of Finance and Economics , Tianjin, 300204, China
[b] School of Mathematics (Zhuhai), Sun Yat-sen University, Zhuhai, 519082, China
[c] School of Statistics, Qufu Normal University, Qufu, 273165, China
[d] LPMC and Institute of Statistics, Nankai University, Tianjin, 300071, China

ARTICLE INFO

Article history:
Received 29 September 2017
Received in revised form 30 January 2018
Accepted 26 February 2018
Available online 8 March 2018

Keywords:
Alias length pattern
Aliased effect-number pattern
Clear effect
General minimum lower order confounding
Wordlength pattern

ABSTRACT

The alias length pattern (ALP) was proposed to characterize the aliasing among two-factor interactions for regular two-level fractional factorial designs of resolution *IV*. Experimenters often pay more attention to joint estimation of main effects and some two-factor interactions. Hence, resolution *III* and *IV* designs are commonly used in practice. In this article, we first extend the ALP for regular two-level fractional factorial designs of resolution *III* and *IV* by modeling justification. Then we study the properties of the ALP and provide the relationships between the ALP and other popular criteria.

1. Introduction

Regular two-level fractional factorial designs are commonly used to estimate important effects in scientific investigations and industrial experiments. Such designs are referred to as 2^{n-m} designs, which have n two-level factors and N $(= 2^{n-m})$ runs. A 2^{n-m} design is determined by its defining relation, which consists of $2^m - 1$ defining words. The number of letters (factors) in a word is its length. The length of the shortest word in the defining relation is called the *resolution* of a design (Box and Hunter, 1961). Based on the effect hierarchy principle (Wu and Hamada, 2009), which states that lower order effects are likely more important than higher order effects and effects of the same order are equally important, main effects (ME's) and two-factor interactions (2FI's) are important and their estimation has received much attention in practice. For a design of resolution at least V, all ME's and 2FI's can be estimated if all three-factor and higher order interactions are negligible. However, such designs often require more runs than one can afford. Thus resolution *III* and *IV* designs are commonly used for estimating ME's and 2FI's.

Focusing on characterizing 2^{n-m} designs, a number of criteria have been proposed in the literature. Fries and Hunter (1980) introduced the *wordlength pattern* (WLP) of a 2^{n-m} design, obtained from its defining relation, for selecting *minimum aberration* (MA) designs. To characterize 2^{n-m} designs with more information, some criteria were proposed by considering aliasing relations, which derive from the defining relation. Wu and Chen (1992) first introduced the notion of *clear effect* (CE), and Wu and Wu (2002) proposed the CE criterion. Zhang et al. (2008) proposed the *aliased effect-number pattern* (AENP) for selecting *general minimum lower order confounding* (GMC) designs. These two criteria were devoted to summarizing the number of effects with specific order from aliasing relations. However, the number of effects considered by the CE and AENP increases geometrically as the number of factors n increases. In particular, some resolution *III* designs have no CE. Note that each aliasing relation contains 2^m effects. Hence, it is efficient to characterize 2^{n-m} designs by the distribution

* Corresponding author at: School of Mathematics (Zhuhai), Sun Yat-sen University, Zhuhai, 519082, China.
E-mail address: guobing@nankai.edu.cn (B. Guo).

https://doi.org/10.1016/j.spl.2018.02.060
0167-7152/

of aliasing relations. Based on this idea, Block and Mee (2003) introduced the *alias length pattern* (ALP) to characterize the aliasing among 2FI's of 2^{n-m} resolution *IV* designs. For both regular and nonregular designs of strength 3 (regular designs of resolution $s+1$ also have strength s), Cheng et al. (2008) proposed the *generalized alias length pattern* (GALP) with modeling justification, which can be used to assess the "goodness" of designs justified from bias caused by 2FI's. More recently, Mee (2013) extended the GALP to analyzing nonregular designs of strength 2 and 3 without modeling justification. The theory of the ALP has been summarized by Mee and Dean (Chapter 7, Dean et al., 2015).

In this article, we first propose a motivating example to illustrate that the extension of Mee (2013) cannot be directly used to study resolution *III* designs. Aiming at filling this gap, we extend the ALP for commonly used fractional factorial designs by modeling justification. The justifications of the extended ALP show that the ALP can be obtained from the alias matrix caused by the model terms that are important (ME's and 2FI's). We illustrate that the ALP presents the distribution of aliasing relations of a design, according to the number of ME's and 2FI's contained in each relation. Then we provide the relationships between the ALP and popular criteria. In particular, we prove that the 2^{n-m} designs with $N/4+1 \le n \le N-1$ and sequential maximization of the elements of the ALP have the GMC. Thus, the construction results of GMC designs, studied by Li et al. (2011), Cheng and Zhang (2010), Zhang and Cheng (2010) and Zhao et al. (2013, 2016), can be used to study 2^{n-m} designs in terms of the ALP.

This article is organized as follows. Section 2 introduces notation, existing work and a motivating example. Section 3 provides the extended ALP with modeling justification. Section 4 is devoted to studying the relationships between the ALP and some criteria. In Section 5, we provide some useful results developed from our method for practical applications.

2. Notation and motivation

We use the $2^q \times (2^q-1)$ matrix

$$H_q = \{\mathbf{1, 2, 12, 3, 13, 23, 123}, \ldots, \mathbf{12\cdots q}\}_{2^q}$$

to denote the *saturated design with the Yates order*, which is generated by the *q independent* columns with 2^q components of entries 1 or -1. We say q columns are independent if none of them can be expressed as products of other columns. Here, we specify the q independent columns of H_q as $\mathbf{1}_{2^q} = (1,-1,1,-1,\ldots,1,-1)'$, $\mathbf{2}_{2^q} = (1,1,-1,-1,\ldots,1,1,-1,-1)'$, etc. Without confusion we will omit the subscript 2^q hereafter. Furthermore, let $\mathbf{I}_{2^q}$ or $\mathbf{I}$ denote the column with all components being 1 and H_r denote the set consisting of the first 2^r-1 columns of H_q. Then, we have $H_0 = \{\mathbf{I}\}$, $H_1 = \{\mathbf{1}\}$ and $H_r = \{H_{r-1}, \mathbf{r}, \mathbf{r}H_{r-1}\}$ for $r = 2,\ldots,q$, where $\mathbf{r}H_{r-1} = \{\mathbf{rd} : \mathbf{d} \in H_{r-1}\}$ and $\mathbf{rd}$ is a component-wise product of $\mathbf{r}$ and $\mathbf{d}$.

A 2^{n-m} design d is represented by an $N \times n$ matrix $\mathbf{X_1} = \{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$. The $\mathbf{x}_j$'s are columns taken from H_q, in which q $(= n-m)$ of them are independent. Design d is determined by m independent defining words and all possible products of the m words are contained in the defining relation of design d, from which $2^{n-m}-1$ aliasing relations can be obtained. Actually, all the 2^n ME's and interactions of n factors of design d, including the grand mean I, are equally distributed into the 2^{n-m} relations (defining and aliasing). Furthermore, each $\mathbf{x}_j$ corresponds to one aliasing relation. We use the following example to illustrate this notation.

Example 1. Consider a 2^{4-1} design $d_1 = \{\mathbf{1},\mathbf{2},\mathbf{3},\mathbf{123}\}$, which is taken from $H_3 = \{\mathbf{1},\mathbf{2},\mathbf{12},\mathbf{3},\mathbf{13},\mathbf{23},\mathbf{123}\}$, and let factors: A, B, C and D be assigned to columns: $\mathbf{1}, \mathbf{2}, \mathbf{3}$ and $\mathbf{123}$ sequentially. Because of $D = ABC$, for design d_1, the defining word is $ABCD$. Then we have the defining relation: $I = ABCD$, and seven aliasing relations:

$$\begin{array}{llll} \mathbf{1}:\ A = BCD, & \mathbf{2}:\ B = ACD, & \mathbf{3}:\ C = ABD, & \mathbf{123}:\ D = ABC, \\ \mathbf{12}:\ AB = CD, & \mathbf{13}:\ AC = BD, & \mathbf{23}:\ AD = BC, & \end{array} \tag{1}$$

where $\mathbf{1} : A = BCD$ means that column $\mathbf{1}$ corresponds to the aliasing relation $A = BCD$.

Block and Mee (2003) introduced the ALP as the frequencies of the lengths of the aliasing relations containing 2FI's of resolution *IV* designs. An aliasing relation is said to have length p, if it contains p 2FI's. Let a_j $(j = 1,\ldots,l)$ denote the number of aliasing relations having length j and l be the maximum length among the aliasing relations. The vector $(a_1,\ldots,a_l)$ is called the ALP of a design. Consider strength-3 design d, Cheng et al. (2008) proposed the vector

$$\mathbf{L_1} = diag[\mathbf{X}_2'\mathbf{X}_2\mathbf{X}_2'\mathbf{X}_2/N^2] \tag{2}$$

to summarize the GALP of design d as the vector of f_j's corresponding to $\mathbf{L_1}$, where $\mathbf{X}_2$ is the corresponding matrix consisting of all 2FI's contrasts and f_j denotes the frequency of j's in $\mathbf{L_1}$. Furthermore, each element of $\mathbf{L_1}$ corresponds to one 2FI of design d. For regular designs, Cheng et al. (2008) illustrated $a_j = f_j/j$, which provided the modeling justification of the ALP. Continuing Example 1, we have that the ALP of design d_1 is $(a_1, a_2) = (0, 3)$, which means that there are three aliasing relations containing 2FI's and each of them has length 2.

Consider a nonregular design d of strength 2 and 3. Mee (2013) provided the vector

$$\mathbf{L_2} = diag[\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}/N^2] \tag{3}$$

to summarize the GALP of design d, where $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$ denotes the 2FI model matrix partitioned into n and $\frac{n(n-1)}{2}$ columns. Each element of $\mathbf{L_2}$ corresponds to one ME or 2FI of design d. The following example illustrates that the GALP from (3) cannot be directly used to analyze regular designs of resolution *III*, because it ignores the distinction between ME's and 2FI's.