Computational Statistics and Data Analysis 118 (2018) 98–111

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Robust multistratum baseline designs

Chang-Yun Lin [a,*], Po Yang [b]

[a] Department of Applied Mathematics and Institute of Statistics, National Chung Hsing University, Taichung, 40227, Taiwan
[b] Department of Statistics, University of Manitoba, Winnipeg, MB R3T 2N2, Canada

ARTICLE INFO

Article history:
Received 19 February 2017
Received in revised form 8 August 2017
Accepted 10 August 2017
Available online 18 September 2017

Keywords:
A-criterion
Coordinate-exchange
Generalized least square
Loss function
Mean squared error
Minimax
Restricted maximum likelihood
Split-plot

ABSTRACT

Baseline designs have received considerable attention recently. Most existing methods for finding best baseline designs were developed for completely randomized experiments. How to select baseline designs for experiments under multistratum structures has not been studied in the literature. The purpose of this paper is to fill this gap and extend the use of the baseline design for experiments with complex structures, such as split-plot experiments. A framework for baseline designs under multistratum structures is established and a generalized minimax A-criterion for selecting multistratum baseline designs which are efficient and model robust is proposed. The coordinate-exchange algorithm is applied and robust baseline designs under split-plot, split-split-plot, and block-split-plot structures, which can be constructed via nesting operators repeatedly, are exemplified. A real case study for industrial experiments is provided to demonstrate the application and data analysis of multistratum baseline designs.

1. Introduction

Fractional factorial designs under the orthogonal parameterization have been widely studied over the past two decades. Recently, a non-orthogonal baseline parameterization in factorial experiments has received considerable attention. Different from the orthogonal parameterization, which defines the factorial effects via orthogonal contrasts, the baseline parameterization defines the effects with reference to baseline (or control) levels of the factors. Mukerjee and Tang (2012) studied two-level fractional factorial designs under baseline parameterization and established the K-aberration criterion. Based on that criterion, Li et al. (2014) provided an efficient algorithm for searching minimum aberration baseline designs. Mukerjee and Huda (2016) proposed a minimaxity approach and applied the approximate theory to find robust efficient baseline designs. Miller and Tang (2016) explored the use of the two-level regular fractional factorial designs to generate baseline designs. Mukerjee and Tang (2016) obtained optimal two-level regular designs under baseline parameterization via coset and minimum moment aberration. Applications of the baseline parameterization in factorial experiments can be found in Yang and Speed (2002), Glonek and Solomon (2004), Banerjee and Mukerjee (2008), and Zhang and Mukerjee (2013).

The baseline designs in the above articles were mainly studied under completely randomized experiments. In practice, however, it is common that complete randomization is infeasible or uneconomic, and, hence, several stages of randomization are considered in an experiment. This results in a multistratum structure. For example, if an experiment involves factors whose levels are difficult to change, then complete randomization will increase the experimental cost due to frequently changing levels of these factors. To reduce cost, the experimenter may first conduct a randomization for the level combinations of the hard-to-change factors and then conduct a second randomization for the level combinations of the factors whose levels are easy to change. The two-stage randomization forms a split-plot structure and the design used for this experiment is

* Corresponding author.
E-mail addresses: chlin6@nchu.edu.tw (C.-Y. Lin), po.yang@umanitoba.ca (P. Yang).

http://dx.doi.org/10.1016/j.csda.2017.08.009
0167-9473/

called the split-plot design (see Fisher, 1925; Goos and Vandebroek, 2001, 2003, 2004; Goos, 2002; Mukerjee and Fang, 2002; Jones and Nachtsheim, 2009; Lin and Yang, 2015; Yang and Lin, 2017; Lin, 2018a). In split-plot experiments, the hard-to-change factors and the easy-to-change factors are called the whole-plot factors and the subplot factors, respectively, and the level combinations of the whole-plot factors and the subplot factors are called the whole plots and the subplots, respectively. Experiments with more complex structures can be found in the literature, such as split-split-plot structures (see Jones and Goos, 2009) and block-split-plot structures (see Trinca and Gilmour, 2001, 2015). The reader is referred to Cheng (2014) and Lin (2018b) for multistratum designs under orthogonal parameterization.

Fractional factorial designs under baseline parameterization for multistratum experiments have not been explored and how to select multistratum baseline designs has not been studied in the literature. The purpose of this paper is to fill this gap and extend the use of the baseline designs for experiments with complex structures. To establish a criterion for selecting multistratum baseline designs, we adopt the A-criterion in Mukerjee and Huda (2016) used for selecting baseline designs in completely randomized experiments. It is known that the A-criterion is not a model-free criterion. A model must be pre-specified for obtaining optimal A-efficient designs. If the model is misspecified, then the estimates of effects will be biased. To overcome this problem, Zhou (2001, 2008), Wilmut and Zhou (2011), Lin and Zhou (2013), and Yin and Zhou (2015) proposed several minimaxity criteria for selecting orthogonal designs which are robust under model misspecification. Mukerjee and Huda (2016) provided another version of the minimaxity criterion for obtaining robust baseline designs. These minimaxity criteria were developed based on completely randomized experiments. In this paper, we propose a generalized minimax A-criterion, which can be used for selecting baseline designs that are efficient and model robust for multistratum or completely randomized experiments. We provide three examples to demonstrate the robust baseline designs under split-plot, split-split-plot, and block-split-plot structures. These structures can be constructed via nesting operators repeatedly and are special cases of the so-called simple block structures, which are special cases of orthogonal block structures introduced in Speed and Bailey (1982).

The remainder of this paper is organized as follows. Section 2 introduces the notation and establishes a framework for baseline designs under multistratum structures. Section 3 develops the generalized minimax A-criterion for selecting robust multistratum baseline designs. A construction algorithm is given in Section 4. Section 5 provides examples for the baseline designs under the split-plot, split-split-plot, and block-split-plot structures. Section 6 studies the sensitivity of the best baseline designs to the ratios of random effects' variances. A real case study for industrial experiments is provided in Section 7 to demonstrate the application and data analysis of multistratum baseline designs. Section 8 contains the concluding remarks.

2. Baseline designs in multistratum experiments

We first introduce the background and notation of baseline parameterization for completely randomized experiments and then establish a framework and a criterion for baseline designs under multistratum structures.

2.1. Notation and background

Let D be a full factorial design with m factors $F_1, \ldots, F_m$. The levels of factor F_i, $i = 1, \ldots, m$, are coded as $0, 1, \ldots, a_i - 1$, where 0 represents the baseline level. Define the set $A_i = \{0, 1, \ldots, a_i - 1\}$. There are $N = \prod_{i=1}^{m} a_i$ treatment combinations $j_1 \cdots j_m$, where $j_i \in A_i$. The full factorial model of D under baseline parameterization can be written as

$$\tau_{j_1 \cdots j_m} = \sum_{u_1 \in \{0, j_1\}} \cdots \sum_{u_m \in \{0, j_m\}} \theta_{u_1 \cdots u_m}, \tag{1}$$

where $\tau_{j_1 \cdots j_m}$ is the treatment effect of the treatment combination $j_1 \cdots j_m$, $\theta_{u_1 \cdots u_m}$ for $u_1 \cdots u_m \neq 0 \cdots 0$ is a main or interaction effect parameter, depending on which u_i's are nonzero, and $\theta_{0 \cdots 0}$ is the baseline effect. For example, in a 2^2 factorial involving factors F_1 and F_2, each at levels 0 and 1, there are four treatment combinations, $j_1 j_2 = 00, 01, 10$, and 11, and the effects of the four treatment combinations are expressed as $\tau_{00} = \theta_{00}$, $\tau_{10} = \theta_{00} + \theta_{10}$, $\tau_{01} = \theta_{00} + \theta_{01}$, and $\tau_{11} = \theta_{00} + \theta_{10} + \theta_{01} + \theta_{11}$, where θ_{00} is the baseline effect, θ_{10} and θ_{01} are the main effects of F_1 and F_2, respectively, and θ_{11} is the interaction of F_1 and F_2. More examples can be found in Mukerjee and Tang (2012) and Mukerjee and Huda (2016). As mentioned in Mukerjee and Tang (2012), baseline parameterization can arise naturally in many situations whenever there is a control or baseline level for each factor. For instance, in toxicological study with a two-level baseline design, the baseline level 0 of each factor represents the absence of a particular toxin while the level 1 represents the presence of the toxin (Kerr, 2006); in industrial experiments for quality improvement by changing the settings of several machines, the baseline levels represent the current settings (Banerjee and Mukerjee, 2008).

Consider now a reduced model of (1) which includes the baseline effect, all main effects, and perhaps some interactions. Define the requirement set R which is constituted by the main effects and interactions in the reduced model. Denote by p the dimension of R and write θ_0 for the baseline effect $\theta_{0 \cdots 0}$. Then the reduced model with respect to R can be expressed as

$$\boldsymbol{\tau} = \theta_0 \mathbf{1}_N + \mathbf{Z}\boldsymbol{\theta}, \tag{2}$$

where $\boldsymbol{\tau}$ is the vector of N treatment effects, $\mathbf{1}_N$ is an $N \times 1$ vector of ones, $\boldsymbol{\theta}$ is the vector of the p parameters representing the factorial effects in R, and $\mathbf{Z}$ is the $N \times p$ model matrix for $\boldsymbol{\theta}$. The reader is referred to Mukerjee and Huda (2016) for details.