Choice Experiments for Estimating Main Effects and Interactions

---

A Dissertation Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

---

by
Jing Chen
January, 2011

Examining Committee Members:

Pallavi Chitturi, Advisory Chair, Statistics
Damaraju Raghavarao, Advisory Co-Chair, Statistics
Boris Iglewicz, Statistics
Milton Parnes, Statistics
Susan Mudambi, External Reader, Marketing & Supply Chain Management

ABSTRACT

Choice Experiments for Estimating Main Effects and Interactions

Jing Chen
DOCTOR OF PHILOSOPHY

Temple University, January, 2011

Professor Pallavi Chitturi, Professor Damaraju Raghavarao, Chair

Choice-based conjoint experiments are used when choice alternatives can be described in terms of attributes. The objective is to infer the value that respondents attach to attribute levels. This method involves the design of profiles on the basis of attributes specified at certain levels. Respondents are presented sets of profiles and asked to select the one they consider best. Choice sets with no dominating or dominated profiles are called Pareto optimal, and these Pareto optimal choice sets are provided to respondents. However, if choice sets have too many profiles, they may be difficult to implement. Therefore, we provide strategies for reducing the number of profiles in choice sets. We consider situations where only a subset of interactions is of interest, and obtain connected main effects plans with smaller choice sets for 2^n and 3^n designs that are capable of estimating subsets of interactions inclusive of one specific factor. We also provide plans for estimating all main effects and one two-way interaction for mixed level designs. Next, we examine the relationship between certain Pareto Optimal choice sets and g-designs. Finally, we obtain connected main effects plans with smaller choice sets for estimating different subsets of interactions, not inclusive of one specific factor.

ACKNOWLEDGEMENTS

I owe my gratitude to all those people who have made this dissertation possible and because of whom my graduate experience has been one that I will cherish forever.

My deepest gratitude is to my advisor, Dr. Pallavi Chitturi. I have been amazingly fortunate to have an advisor who gave me enormous help not only on academic research but also on teaching, career, and many other aspects to make me a better person. Her patience and support helped me overcome many crisis situations and finish this dissertation. I hope that one day I would become as successful a female professional as Dr. Chitturi is.

My co-advisor, Dr. Damaraju Raghavarao, has been always there to listen and give advice. I am deeply grateful to him for giving me the freedom to explore on my own, and at the same time the guidance to recover when my steps faltered. I want to thank him for setting high standards for his students and encouraging them to meet those standards.

I would also like to thank my committee members, Dr. Boris Iglewicz and Dr. Milton Parnes, whose insightful comments and constructive suggestions at different stages of my research were thought-provoking and they helped me focus my ideas.

Finally, I want to thank my parents. Without their unwavering faith and confidence in me, none of this would have been possible. They have been a constant source of love, concern, support and strength all these years.

To my parents, Anlan and Yibai.

TABLE OF CONTENTS

CHAPTER 1

Introduction

Valuation studies are widely used in many fields, from market research to environmental experiments. Business valuation is a set of procedures used to determine the economic value of an owner's interest in a business. It is often used to estimate the selling price of a business, resolve disputes related to estate and gift taxation, divorce litigation, allocate business purchase price among the business assets, establish a formula for estimating the value of partners' ownership interest for buy-sell agreements, and many other business and legal disputes. Contingent valuation, on the other hand, is a survey-based economic technique for the valuation of non-market resources, such as environmental preservation or the impact of contamination. While these resources do give people utility, certain aspects do not have a market price as they are not directly sold - for example, people receive benefit from a beautiful view of a mountain, but it would be tough to value using price-based models. For both studies, the major objective is to infer the valuations (explicit or implicit) that respondents attach to attributes described by the benefits and costs of a product or service.

In valuation studies, products or services are usually labeled by a combination of several attributes. Each set of multi-attribute alternatives is called a profile. Respondents are shown sets of profiles, called choice sets, to determine their preferences for these profiles. Due to the constraints of budget, production technology, market forces, etc., respondents have to trade off the attributes. By examining their preferences for profiles in each choice set, researchers may infer the value respondents attach to changes in benefits and costs relative to the value respondents attach to changes in other attributes.

1.1 Motivation

Pareto efficiency, or Pareto optimality, is an important concept in economics with broad applications in game theory, engineering and the social sciences. The term is named after Vilfredo Pareto, an Italian economist who used the concept in his studies of economic efficiency and income distribution. An outcome of a game is Pareto optimal if there is no other outcome that makes every player at least as well off and at least one player strictly better off. That is, a Pareto Optimal (PO) outcome cannot be improved upon without hurting at least one player.

In a valuation study, some profiles are better than others, or dominating, while some profiles are worse, or dominated. Choosing a dominating alternative, or not choosing a dominated one, does not involve an economic choice. If a choice set has a dominating profile, the respondent's choice is trivially made. Similarly, if a choice set has a dominated profile, it will never be selected. Therefore, we want to restrict our attention to choice sets with no dominating or dominated profiles. Such choice sets are known as Pareto Optimal subsets.

Raghavarao and Wiley (2006) discuss the feasibility of using Pareto Optimal choice sets in choice-based evaluation studies. They show how to generate PO choice sets and give general connected s^n designs to estimate main effects, two- and three-way interactions. They also discuss strategies for reducing choice set sizes by sub-setting larger subsets, sub-setting based on overlapping attributes, and/or levels, or using fractional designs. They also consider situations where a single two-way interaction is of interest for 2^n and 3^n experiments.

Obviously, PO designs enable us to infer value from the trade-offs that respondents are, or are not, willing to make. However, one disadvantage for PO subsets is that, when the number of attributes or attribute levels becomes large, usually more than six, the profiles in a single choice set become too many for respondents to make precise decisions. For example, the beneficial attributes of laptops may include processor speed, operating system, memory capacity, hard drive storage, display resolution and so on. Costs include price, weight, noise, etc. Each attribute can have a number of levels. For instance, the levels for memory capacity may be 512 megabytes, 1 gigabytes, 2 gigabytes and so on. In a typical conjoint study, respondents are presented sets of profiles (attribute level combinations) and asked to choose from, rank or rate the profiles they are shown. Each profile is similar enough that consumers will see them as close substitutes, but dissimilar enough that respondents can clearly determine a preference. Each profile is composed of a unique combination of product features. By analyzing how consumers make preferences between these products, the implicit valuation of the individual elements making up the product or service can be determined. These implicit valuations (utilities or part-worths) can be used to create market models that estimate market share, revenue and even profitability of new designs. Hence, smaller choice set sizes are necessary in such experiments. In reality, due to limits on costs and time, it might not be practical to include all the profiles in choice sets. Also, it is possible that experimenters are only interested in values related to certain factors instead of all attributes. The motivation of this dissertation is to find smaller designs based on PO choice sets which are capable of estimating all main effects and certain subsets of two- and three-way interactions.

1.2 Structure

This dissertation is organized as follows. In Chapter 2, we present an overview on choice experiments and choice-based conjoint (CBC) experiments. Section 2.1 discussed the motivation and history of conjoint analysis. The advantages and disadvantages of conjoint analysis are also described in this section. Section 2.2 discusses a special case of conjoint analysis: choice-based conjoint analysis(CBC), and steps for setting up a CBC design. Section 2.3 gives the model and notations used in this proposal. Section 2.4 introduces Pareto Optimal (PO) choice set and reviews some important results.

In Chapter 3, we describe our results on generating subsets for estimating main effects and two-way interactions inclusive of one factor, and demonstrate the approaches by examples. Section 3.1 & 3.3 contain the theorem and proof for 2^n and 3^n plans respectively. Section 3.2 & 3.4 show the corresponding examples.

In Chapter 4, we present a design that is capable of estimating main effects, two-way interactions, and one three-way interaction inclusive of one factor for 2^n plans. Section 4.1 gives the theorem and proof and we illustrate with three examples in Section 4.2. In Chapter 5, we give connected main effect plans for mixed-level or asymmetrical designs. Section 5.1 & 5.3 give the theorems and proof for $2^n \cdot 3^m$ plans and $2^n \cdot 3$ plans respectively. Section 5.2 & 5.4 give the corresponding examples.

In Chapter 6, we compare the design we provide in Section 3.1 with g-designs, which were proposed by Hedayat and Pestotan (1992). We also show certain PO designs are D-optimal. In Chapter 7, we present a design which is capable of estimating main effects and two-way interactions among three specific factors for a 2^n plan. This design is also a g-design.

In the final chapter, we summarize the result in this dissertation and discuss several possible extensions for future research.

CHAPTER 2

Literature review

Since the early 1970s, conjoint analysis has received considerable academic and industrial attention as a major technique for measuring buyer's trade-offs among multiattributed products and services. Since that time many new developments in conjoint analysis and related methods have been reported.

2.1 Conjoint analysis

Conjoint analysis is a statistical technique commonly used in behavior studies. It was introduced by Luce and Tukey (1964) and developed by Green and Wind (1975). The major goal of conjoint analysis is to determine what combination of a limited number of attributes is most influential on respondents' choice or decision making. In such designs, a product or service is described in terms of benefits (attributes for which value increases with level) or costs (attributes for which value decreases with level). For example, the beneficial attributes of laptops may include processor speed, operating system, memory capacity, hard drive storage, display resolution and so on. Costs include price, weight, noise, etc. Each attribute can then be broken down into a number of levels. For instance, the levels for memory capacity may be 512 megabytes, 1 gigabytes, 2 gigabytes and so on. In a typical conjoint study, respondents are presented sets of profiles (attribute level combinations) and asked to choose from, rank or rate the profiles they are shown. Each profile is similar enough that consumers will see them as close substitutes, but dissimilar enough that respondents can clearly determine a preference. Each profile is composed of a unique combination of product features. By analyzing how consumers make preferences between these products, the implicit valuation of the individual elements making up the product or service can be determined. These implicit valuations (utilities or part-worths) can be used to create market models that estimate market share, revenue and even profitability of new designs.

The earliest forms of conjoint analysis were what are known as Full Profile studies. Usually a small set of attributes, typically four to five, are used to create profiles that are shown to respondents, often on individual cards. Respondents then rank or rate these profiles. Using relatively simple dummy variable regression analysis, the implicit utilities for the levels can be calculated.

Two drawbacks were seen in these early designs. First, the number of attributes in use was heavily restricted. As indicated by Green (1984), industrial users of conjoint analysis have strained the methodology by requiring larger numbers of attributes and levels within attributes, thus placing a severe information overload on the respondents. When faced with such tasks, respondents resort to simplifying tactics and the resulting part-worth estimates may distort their true preference structures (Wright 1975).

Green and Srinivasan (1990) proposed three approaches to handle the problem of large numbers of attributes or attribute levels: (1) the self-explication approach; (2) hybrid conjoint analysis, and (3) Adaptive Conjoint Analysis (ACA). All three approaches were to do some form of self-explication before the conjoint tasks and some form of adaptive computer aided choice over the profiles to be shown.

The second drawback was that the task itself was unrealistic and did not link directly to behavioral theory. In real life situations the task would be some form of actual choice between alternatives rather than the more artificial ranking and rating originally used. Louviere and Woodworth (1983) pioneered an approach that used a choice task which became the basis of choice-based conjoint and discrete choice analysis.

2.2 Choice-based conjoint analysis

Choice-based conjoint (CBC) experiments are a method related to conjoint analysis and sometimes called discrete choice experiments. Louviere and Woodworth (1983) provided a theoretical foundation for choice-based conjoint analysis. They sought to integrate conjoint analysis with econometrics "discrete choice modeling". Their approach had these features:

- Sets of concepts to be shown to respondents were constructed using complex experimental designs.
- Choice sets could contain an option such as "I wouldn't choose any of these".
- Estimation of parameters was done in aggregate rather than for individual respondents, using multinomial logit analysis.

In CBC studies, instead of ranking or rating all profiles as is usually done in classic conjoint studies, respondents are asked to repeatedly choose one alternative from different sets of profiles offered to them. It may be assumed that people prefer higher levels of benefits and lower levels of costs. Therefore it is not informative to ask respondents whether they would prefer to have more of a single benefit or less of it or whether they would prefer to pay a higher price for a product or a lower price. Usually, researchers are more interested in cases in which respondents must give up an incremental level of one benefit (or incur an incremental cost) in order to acquire an incremental level of another benefit (or avoid an incremental cost). By examining alternatives in choice sets, researchers may infer the values respondents attach to changes in benefits and costs relative to the values respondents attach to changes in other attributes.

The following are general strategies for designing CBC experiments (Carter, Dubelaar and Wiley, 2000).

- Word the survey instruments simply and in a straight-forward manner.
- Keep choice tasks as realistic and natural as possible.
- Make choices credible.
- Aim to "balance" the numbers of: (1)choice sets; (2)choices; (3)attributes; (4)attribute levels, to avoid respondent overload.
- Ensure respondents understand the different product/service attributes and levels.
- Make the choice context explicit and thereby encourage realism.
- Do not set implausible attribute levels.
- Avoid including alternatives that "dominate" others because they are "better" on all benefit and cost criteria.
- Keep alternatives constant (i.e., not changing attributes or attribute levels of choices within the survey instrument).
- Include a "none of these" alternative (i.e., enabling the respondent to indicate that none of the alternatives in the specific choice set would be chosen).
- Ask respondents to indicate their most recent actual choice of the product or service being surveyed.

Most traditional conjoint analysis studies use "main effects only" assumption. However, because choice-based conjoint analysis data are analyzed by pooling or borrowing information across respondents, it is more feasible to quantify interactions. The main disadvantage is that CBC analysis is not appropriate for studies involving large number of attributes. As the number of combinations of attributes and levels increases, the number of potential profiles increases exponentially. Green and Srinivasan (1990) suggest six to ten as the maximum number of attributes to handle with full-profile concepts in traditional conjoint analysis.

Meanwhile, there has been interest in the question of whether the results of choice-based and ratings-based conjoint analysis are fundamentally different. An early study by Oliphant *et al.* (1992) compared choice-based with ratings-based conjoint analysis, concluding that there was little difference between success of the two methods in predicting holdout concepts. However, in another comparison of choice-based and ratings-based conjoint analysis, Huber *et al.* (1992) did find differences. In their study, the relative importance of attributes differed by method. One hypothesis explaining such results is that respondents may simplify choice tasks by focusing on a few key attributes or by searching for important combinations. However, it is not yet clear that the process respondents go through when choosing products is fundamentally different from what they do when they rate products. We believe that in most cases, similar results can be expected from choice-based and ratings-based conjoint analysis, even though the two approaches have different strengths and motivations.

Recent research led by Olivier Toubia et al. (2004) proposed and tested a new "polyhedral" question-design method that adapts each respondent's choice sets based on previous answers by that respondent. Polyhedral "interior-point" algorithms design questions that quickly reduce the sets of partworths that are consistent with the respondent's choices. By this approach, question design can be improved based on prior estimates of the respondent's partworths - information that is revealed by respondents' answers to prior questions.

Generally, CBC analysis provides a good way to produce relatively precise results when interactions are of concern and there are relatively few attributes. In recent years this method has become increasingly popular as a way to more directly study choice (Batsell and Louviere, 1991). Examples of areas in which CBC experiments have been used include environmental science (Adamowicz et al., 1988; Bullock et al., 1998), Geography (Oppewal et al., 1997; Waerden et al., 1993), Health (Propper, 1995; Ryan and Farrar, 2000), Marketing (Kamakura and Srivastava, 1984; Johnson and Olberts, 1991; Moore et al., 1999), Tourism (Haider and Ewing, 1990), and Transportation (Hensher et al., 1989; Fowkes and Wardman, 1988; Brandley and Gunn, 1990).

2.3 Model and notations

Among the algorithms applied in CBC designs, probabilistic choice models such as multinomial logit or probit models are most common. Responses from several respondents are aggregated and the dependent variable consists of the proportion of respondents selecting the profiles in each choice set. Analysis of these proportions is done within a logit or probit framework.

Consider a choice experiment with n attributes each at s levels denoted by 0, 1, …, $s-1$. In the experiment, respondents are shown a series of sets of profiles called choice sets. Each profile consists of a specific level on each of n attributes. Every choice set is a subset of k profiles, $k \leq s^n$. Respondents are asked to select the profile they consider best from each set. Assume $y_{x_1 x_2 \ldots x_n}$ is the proportion of respondents choosing profile $(x_1 x_2 \ldots x_n)$ from a given choice set S_l, or the transformed proportional response to that profile from that choice set, where x_i is the level of the ith attribute, $x_i = 0, 1, \ldots, s-1$, $i = 1, \ldots, n$. A regression model for this situation is set up as Eq (2.1), where μ is the general mean, $\alpha_{x_i}^{A_i}$ is the effect of attribute (or factor) A_i at x_i level, $\alpha_{x_i x_j}^{A_i A_j}$ is the effect of attributes A_i and A_j at x_i and x_j levels, respectively, …, and $e_{x_1 x_2 \ldots x_n}$ is the random error. For the purpose of this disseratation, we need only assume the expectations of the errors are zero.

$$\begin{aligned} y_{x_1 x_2 \ldots x_n} = \quad & \mu + \sum_{i=1}^{n} \alpha_{x_i}^{A_i} + \sum_{\substack{i,j=1 \\ i \neq j}}^{n} \alpha_{x_i x_j}^{A_i A_j} + \sum_{\substack{i,j,k=1 \\ i \neq j, j \neq k, i \neq k}}^{n} \alpha_{x_i x_j x_k}^{A_i A_j A_k} \\ & + \ldots + \alpha_{x_1 x_2 \ldots x_n}^{A_1 A_2 \ldots A_n} + e_{x_1 x_2 \ldots x_n} \end{aligned} \tag{2.1}$$

Without loss of generality, we assume

$\sum_{x_i} \alpha_{x_i}^{A_i} = 0$; $\sum_{x_i} \alpha_{x_i x_j}^{A_i A_j} = 0$, $\forall x_i$; $\sum_{x_j} \alpha_{x_i x_j}^{A_i A_j} = 0$, $\forall x_j$; $\ldots$; etc.
For the sake of simplicity, we use $x_1 x_2 \ldots x_n$ to denote the response variable $y_{x_1 x_2 \ldots x_n}$, and $(x_1 x_2 \ldots x_n)$ to denote a profile. However, when $x_1 x_2 \ldots x_n$ is inside a choice set, it represents a profile.

2.4 Pareto Optimal choice set

In a choice set, some profiles are better than others, or dominating, while some profiles are worse than others, or dominated. A simple example illustrating the implications of having dominated or dominating alternatives in a set is given by Raghavarao and Wiley (1998). Suppose a consumer is given the choice between a $9,000 car that uses 4 gallons of fuel per 100 miles and a $10,000 car that uses 3.5 gallons per 100 miles. If the consumer accepts the first choice, it may be inferred that he or she values $1,000 more than 0.5 additional gallons of fuel per 100 miles, assuming other things equal. Conversely, choice of the alternative would indicate that the saving of fuel was of more value than the increment in price. On the other hand, a consumer given the choice between a $9,000 car that uses 3.5 gallons per 100 miles and a $10,000 car that uses 4 gallons per 100 miles is likely to select the first set of attributes because it dominates the second set. Nothing is learned about the relative value the respondent attaches to cost versus fuel consumption. Hence, it is desirable that sets of choices presented to respondents do not contain profiles that dominate, or are dominated by, other profiles.

Choice sets with no dominating or dominated profiles are called Pareto Optimal (PO) subsets. Let S be the set of all possible profiles. Formally, subset T of S is said to be a PO subset if for every two distinct profiles $(x_1 x_2 \ldots x_m)$, $(y_1 y_2 \ldots y_m) \in T$, there exists subscripts i and j $(i \neq j)$ such that $x_i < y_i$ and $x_j > y_j$.

Wiley (1978) recognized the need of PO choice sets in choice experiments. Krieger and Green (1991) extended that work and constructed orthogonal and

PO subsets (or close to PO subsets). Huber and Hansen (1986) gave some empirical results on the comparison of PO designs and orthogonal designs, and reported that the PO designs predict better. Huber and Zwerina (1996) proposed the concept of utility balanced designs, which are similar to PO designs. They also gave methods for constructing them. Raghavarao, Wiley and Chitturi (2010) explored the design of experiment (DOE) issues that occur when constructing profiles and showed how to modify commonly used designs for solving conjoint analysis problems.

Raghavarao and Wiley (1998) consider a general setting with any number of levels for the attributes and obtained connected main effects plans. They noted that the choice sets $S_l = \{(x_1 x_2 \ldots x_n) | \sum x_i = l\}$, $l = 0, 1, \ldots, n(s-1)$, are PO. Any subset of a PO subset is itself a PO subset. And the design D of attribute profiles resulting from a single Pareto Optimal subset S_l is not a connected main effects plan. They further showed that the design based on $S_{\lfloor \frac{n}{2} \rfloor}$ and $S_{\lfloor \frac{n}{2} \rfloor + 1}$ is a connected main effects plan, where $\lfloor \frac{n}{2} \rfloor$ is the integral part of $\frac{n}{2}$.

Raghavarao and Zhang (2002) extend the above results to 2^n plans. They pointed out that for a 2^n experiment, the design based on any two PO subsets S_l and $S_{l'}$ $(0 < l < l' < n,\ l \neq l')$ is a connected main effects plan. They also examined the optimality of such designs under different situations, by using the Information Per Profile (IPP) θ as an optimality criterion. Meanwhile they showed that certain Balanced Incomplete Block Designs (BIBD) have the same IPP θ as PO subsets, but with smaller choice set sizes. Therefore under specific conditions, BIBD could be used to reduce choice set sizes. The details are given below.

Consider a BIB design with parameters $v = n$, b, k, r, λ and its complementary $v^* = n$, $b^* = b$, $k^* = n - k$, $r^* = b - r$ and $\lambda^* = b - 2r + \lambda$. This original design is actually a choice set S_k^* of b profiles, where the ith profile corresponds to the ith block, with the symbol present interpreted as the high level, and absent as the low level of that attribute. We can similarly form S_{n-k}^* from the complement BIBD. Then the $IPP\ \theta$ for the design based on S_k and

S_{n-k} is the IPP θ^* for the design based on S_k^* and S_{n-k}^*.

Raghavarao and Wiley (2006) proposed a strategy for generating PO choice sets and constructing hierarchical PO subsets to collect data for estimating main effects, two-way interactions and three-way interactions. After these subsets are chosen, the following sequential procedure will be applied. First, a lack-of-fit test will be utilized for a model including only main effects. If this model fits, inferences will be drawn only on main effects and the significance of the model parameters will be tested. If the model fails the lack-of-fit test, more data will be collected, two-way interactions will be added to the model and a lack-of-fit test will be applied for this new model. If the new model fits, inferences will be drawn on main effects and two-way interactions and the significance of the model parameters will be tested. Otherwise, the model will be updated by adding more choice sets. A lack-of-fit test is not considered at this step and inference will be drawn on parameters for main effects, two- and three-way interactions as it is likely that four or higher factor interactions are negligible. This step-by-step decision-making strategy is called sequential testing, which is of particular advantage where the objective of the research is to infer the presence of interactions.

However, some PO choice sets may be too large to be administered in a single study. For example, according to Theorem 1 in Raghavarao and Wiley (2006), at least 25 profiles are needed in order to estimate all main effects and two-way interactions for a 2^5 plan, and at least 90 profiles to estimate main effects and two-way interactions for a 3^5 plan. Moreover, if only a subset of interactions are of interest, it may be unnecessary to go over all the profiles. Raghavarao and Wiley (2006) developed strategies for breaking choice sets into smaller subsets to reduce the number of profiles and choice sets when only one two-way interaction is of interest. Their result is given below:

Theorem 2.1 *(a) The choice sets* $S_1^* = \{10\ldots00,\ 00\ldots01\}$ *and* $S_2^* = \{x_1x_2\ldots x_n | \sum_{i=1}^{n-1} x_i = 1, x_n = 1\}$ *is a connected* 2^n *main effects plan.*
(b) The choice sets S_1^* *and* $S_2^{**} = S_2^* \bigcup \{110\ldots00\}$ *is a connected main effects*

plan and is capable of estimating the two-way interaction, A_1A_2.

Theorem 2.2 (a) *The choice sets* $S_1 = \{x_1x_2\ldots x_n | \sum_{i=1}^{n} x_i = 1\}$, $S_2^* = \{x_1x_2\ldots x_n | \sum_{i=1}^{n} x_i = 2, x_i = 0 \text{ or } 2\} \bigcup \{x_1x_2\ldots x_n | \sum_{i=1}^{n-1} x_i = 1, x_n = 1\}$ *is a connected* 3^n *main effects plan.*
(b) *The choice sets* $S_3^* = \{110\ldots001, 200\ldots001, 120\ldots000\}$ *and* $S_4^* = \{220\ldots000, 210\ldots001\}$ *along with* S_1 *and* S_2^* *of* (a) *is a main effects plan that is capable of estimating two-way interaction,* A_1A_2.

Damaraju and Raghavarao (2002) also discussed a similar problem in a 2^n plan, where all n main effects and $(n-1)$ 2-way interactions with one specific factor are of interest. They concluded that when the BIBD has $b = 4(r - \lambda)$ (these designs are known to be of Family (A), Raghavarao, (1971)), the number of runs in this class of designs is same as the number of runs using a fold-over Hadamard matrix, as well as the variances of estimated main effects and two-way interactions. The limitation of this result is that this family of BIBD exist only for selected number of factors.

CHAPTER 3

Estimating main effects and two-way interactions

3.1 Theorem for 2^n plans

When the number of attributes or the number of attribute levels increases, PO choice sets may become too large to be administered. Sometimes experimenters are interested in estimating all main effects and only those two-way interactions involving one specific factor (e.g. price). In this situation we do not have to include all the profiles in our design. We present a strategy for generating PO subsets capable of estimating main effects and all two-way interactions involving one factor in a 2^n plan.

Theorem 3.1 *The choice sets* $S_1^{**} = \{100\ldots000, 000\ldots001, 010\ldots000\}$*, and* $S_2^{**} = \{x_1x_2\ldots x_n | \sum_{i=1}^{n} x_i = 2, x_1 + x_n \neq 0\}$ *is a connected main effects plan and is capable of estimating the two-way interactions,* A_1A_i*,* $i = 2, 3, \ldots n$*.*

Proof: Based on Theorem 2.1(b), the choice sets S_1^* and $S_2^* \bigcup (110\ldots00)$ is a connected main effects plan and is capable of estimating the two-way interaction, A_1A_2, such that:

$$\widehat{A_1A_2} = \tfrac{1}{2}(110\ldots000 + 000\ldots001 - 100\ldots000 - 010\ldots001).$$

Hence, in general, the choice sets S_1^* and $S_2^* \bigcup \{1100\ldots00, 1010\ldots00, \ldots, 1000\ldots10\} = S_2^* \bigcup \{x_1x_2\ldots x_n | \sum_{i=2}^{n-1} x_i = 1, x_1 = 1, x_n = 0\}$ are capable of estimating all two-way interactions involving A_1 except A_1A_n, such that:

$$\widehat{A_1A_i} \quad = \quad \tfrac{1}{2}(p_i + 000\ldots001 - 100\ldots000 - p_{i'}).$$

for all $i = 2, 3, \ldots, n-1$, where

$$\begin{aligned} p_i &= (x_1x_2\ldots x_n | \sum_{\substack{k=2 \\ k\neq i}}^{n} x_k = 0, x_1 = x_i = 1), \text{ and} \\ p_{i'} &= (x_1x_2\ldots x_n | \sum_{\substack{k=1 \\ k\neq i'}}^{n-1} x_k = 0, x_i = x_n = 1). \end{aligned}$$

In order to estimate A_1A_n, we need to set a factor other than A_1 to take the place of A_n in estimating $A_1A_i, i = 2, 3, \ldots, n-1$. Without loss of generality, we set this factor to A_2, which gives us the profile (010...000). Besides this profile, we need three others, $(100\ldots000)$, $(100\ldots001)$ and $(010\ldots001)$ to estimate A_1A_n, such that

$$\widehat{A_1A_n} \quad = \quad \tfrac{1}{2}(100\ldots001 + 010\ldots000 - 100\ldots000 - 010\ldots001).$$

It should be noticed that when two-way interactions are considered, the estimates for main effects according to Theorem 2.1 (a) are no longer unbiased. According to Raghavarao and Wiley (2006), the main effects in a 2^n design are estimated as:

$$\begin{aligned} \widehat{A}_1 &= 100\ldots001 - 000\ldots001 \\ \widehat{A}_2 &= 010\ldots001 - 000\ldots001 \\ &\ldots \\ \widehat{A}_{n-1} &= 000\ldots011 - 000\ldots001 \\ \widehat{A}_n &= 100\ldots001 - 100\ldots000. \end{aligned}$$

However, when two-way interactions are believed to exist, the expectation of $(100\ldots001 - 000\ldots001)$ is biased for the main effect of factor A_1, because:

$$\begin{aligned}
& E(100\ldots001 - 000\ldots001) \\
= \ & \mu + \frac{1}{2}(A_1 - A_2 - A_3 - \ldots - A_{n-2} - A_{n-1} + A_n \\
& -A_1A_2 - A_1A_3 - \ldots - A_1A_{n-2} - A_1A_{n-1} + A_1A_n) \\
& -[\mu + \frac{1}{2}(-A_1 - A_2 - A_3 - \ldots - A_{n-2} - A_{n-1} + A_n \\
& +A_1A_2 + A_1A_3 + \ldots + A_1A_{n-2} + A_1A_{n-1} - A_1A_n)] \\
= \ & A_1 - A_1A_2 - A_1A_3 - \ldots - A_1A_{n-2} - A_1A_{n-1} + A_1A_n
\end{aligned}$$

Hence, the estimator for the main effect of factor A_1 should be adjusted as:

$$\begin{aligned}
\widehat{A_1} &= 100\ldots001 - 000\ldots001 + \widehat{A_1A_2} + \widehat{A_1A_3} + \ldots \\
& \quad +\widehat{A_1A}_{n-2} + \widehat{A_1A}_{n-1} - \widehat{A_1A}_n,
\end{aligned}$$

where $\widehat{A_1}$ is the estimator for the main effect of A_1, $\widehat{A_1A_2}$ is the estimator for the two-way interaction between A_1 and A_2, and so on. Therefore we have,

$$\begin{aligned}
\widehat{A}_1 &= 100\ldots001 - 000\ldots001 \\
& \quad +\tfrac{1}{2}[(110\ldots000 + 000\ldots001 - 010\ldots001 - 100\ldots000) \\
& \quad +(101\ldots000 + 000\ldots001 - 001\ldots001 - 100\ldots000) \\
& \quad +\ldots \\
& \quad +(100\ldots100 + 000\ldots001 - 000\ldots101 - 100\ldots000) \\
& \quad +(100\ldots010 + 000\ldots001 - 000\ldots011 - 100\ldots000) \\
& \quad -(100\ldots001 + 010\ldots000 - 100\ldots000 - 010\ldots001)] \\
&= \tfrac{1}{2}[(n-4) \times 000\ldots001 + \sum_{i=2}^{n} p_i - 010\ldots000 \\
& \quad -(n-3) \times 100\ldots000 - \sum_{i'=3}^{n-1} p_{i'}].
\end{aligned}$$

Following the same logic, we can get the estimates for the main effects of other factors such that:

$$\widehat{A}_i = \tfrac{1}{2}(p_i + p_{i'} - 100\ldots000 - 000\ldots001), \text{ for } i = 2, 3, \ldots n-1,$$

and

$$\widehat{A}_n = \frac{1}{2}(100\ldots001 + 010\ldots001 - 100\ldots000 - 010\ldots000).$$

Therefore the choice sets

$$S_1^{**} = S_1^* \bigcup (010\ldots00)$$

and

$$S_2^{**} = \{x_1x_2\ldots x_n | \sum_{i=1}^n x_i = 2, x_1 + x_n \neq 0\}$$

provide estimates of all main effects and two-way interactions involving factor A_1. In general, the choice sets

$$S_1^{**} = S_1^* \bigcup p*,$$

where $p*$ is any profile from the choice set $\{x_1x_2\ldots x_n | \sum_{i=2}^{n-1} x_i = 1, x_1 = x_n = 0\}$, and

$$S_2^{**} = \{x_1x_2\ldots x_n | \sum_{i=1}^n x_i = 2, x_1 + x_n \neq 0\}$$

constitute a connected plan for estimating all 2^n main effects and all two-way interactions involving A_1.

3.2 Examples for 2^n plans

3.2.1 2^3 Plan

Suppose we have three factors, A, B and C, each with two levels.

(1) Consider the situation where only the three main effects are of interests. Let $S_1^* = \{100, 001\}$ and $S_2^* = \{101, 011\}$ and the main effects for factor A, B and C are:

$$\begin{aligned}\widehat{A} &= 101 - 001\\ \widehat{B} &= 011 - 001\\ \widehat{C} &= 101 - 100.\end{aligned}$$

(2) Assume we want to estimate all main effects and only those two-way interactions involving factor A.

We need S_1^* and $S_2^* \bigcup (110)$ to estimate the interaction between A and B, such as,

$$\widehat{AB} = \tfrac{1}{2}(001 + 110 - 100 - 011).$$

We need $S_1^* \bigcup (010)$ and S_2^* to estimate the interaction between A and C, such as,

$$\widehat{AC} = \tfrac{1}{2}(101 + 010 - 100 - 011).$$

In this case, the estimates for main effects in step (1) are no longer unbiased because of the existence of interactions. After adjustment, the three main effects can be estimated as:

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(110 + 101 - 010 - 001) \\
\widehat{B} &= \tfrac{1}{2}(110 + 011 - 100 - 001) \\
\widehat{C} &= \tfrac{1}{2}(101 + 011 - 100 - 010).
\end{aligned}$$

So the choice sets need to estimate all main effects and two-way interactions involving A are $S_1^{**} = S_1^* \bigcup (010) = S_1 = \{001, 010, 100\}$ and $S_2^{**} = S_2^* \bigcup (110) = \{011, 101, 110\} = S_2$. Obviously both S_1^{**} and S_2^{**} are PO sets.

3.2.2 2^4 Plan

Suppose we have four factors, A, B, C and D, each with two levels.

(1) If only main effects are of interest, let $S_1^* = \{1000, 0001\}$ and $S_2^* = \{1001, 0101, 0011\}$ and the main effects for factor A, B, C and D are:

$$\begin{aligned}
\widehat{A} &= 1001 - 0001 \\
\widehat{B} &= 0101 - 0001 \\
\widehat{C} &= 0011 - 0001 \\
\widehat{D} &= 1001 - 1000.
\end{aligned}$$

(2) Assume we want to estimate all main effects and only those two-way interactions involving factor A.

We need S_1^* and $S_2^* \bigcup (1100)$ to estimate the interaction between A and B, such as,

$$\widehat{AB} = \tfrac{1}{2}(0001 + 1100 - 1000 - 0101).$$

We need S_1^* and $S_2^* \bigcup (1010)$ to estimate the interaction between A and C, such as,

$$\widehat{AC} = \tfrac{1}{2}(0001 + 1010 - 1000 - 0011).$$

We need $S_1^* \bigcup (0100)$ and S_2^* to estimate the interaction between A and D, such as,

$$\widehat{AD} = \tfrac{1}{2}(1001 + 0100 - 1000 - 0101).$$

For the same reason, the estimates for the main effects are no longer unbiased. After adjustment, the four main effects can be estimated as:

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(1100 + 1010 + 1001 - 0011 - 1000 - 0100) \\
\widehat{B} &= \tfrac{1}{2}(1100 + 0101 - 1000 - 0001) \\
\widehat{C} &= \tfrac{1}{2}(1010 + 0011 - 1000 - 0001) \\
\widehat{D} &= \tfrac{1}{2}(1001 + 0101 - 1000 - 0100).
\end{aligned}$$

Hence the choice sets need to estimate all main effects and two-way interactions involving A are $S_1^{**} = S_1^* \bigcup (0100) = \{1000, 0100, 0001\}$ and $S_2^{**} = \{1001, 0101, 0011, 1100, 1010\}$. Obviously S_1^{**} and S_2^{**} are both PO sets.

3.2.3 2^5 Plan

Suppose we have five factors, A, B, C, D and E, each with two levels.

As mentioned previously, at least 25 profiles are needed in order to estimate all main effects and two-way interactions, based on the conclusion from

Raghavarao and Wiley (2006). However if we only want to estimate main effects and all two-way interactions inclusive of one factor, then only ten profiles are needed according to the result of this dissertation.

(1) If only main effects are of interests, let $S_1^* = \{10000, 00001\}$ and $S_2^* = \{10001, 01001, 00101, 00011\}$ and the main effects for factor A, B, C, D and E are:

$$\begin{aligned}\widehat{A} &= 10001 - 00001\\ \widehat{B} &= 01001 - 00001\\ \widehat{C} &= 00101 - 00001\\ \widehat{D} &= 00011 - 00001\\ \widehat{E} &= 10001 - 10000\end{aligned}$$

(2) Assume we want to estimate all main effects and only those two-way interactions involving factor A.

We need S_1^* and $S_2^* \bigcup (11000)$ to estimate the interaction between A and B, such as,

$$\widehat{AB} = \tfrac{1}{2}(00001 + 11000 - 10000 - 01001).$$

We need S_1^* and $S_2^* \bigcup (10100)$ to estimate the interaction between A and C, such as,

$$\widehat{AC} = \tfrac{1}{2}(00001 + 10100 - 10000 - 00101).$$

We need S_1^* and $S_2^* \bigcup (10010)$ to estimate the interaction between A and D, such as,

$$\widehat{AD} = \tfrac{1}{2}(00001 + 10010 - 10000 - 00011).$$

We need $S_1^* \bigcup (01000)$ and S_2^* to estimate the interaction between A and E, such as,

$$\widehat{AE} = \tfrac{1}{2}(10001 + 01000 - 10000 - 01001).$$

When two-way interactions exist, the adjusted main effects are:

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(11000 + 10100 + 10010 + 10001 + 00001 - 2 \times 10000 \\
&\quad -01000 - 00101 - 00011) \\
\widehat{B} &= \tfrac{1}{2}(11000 + 01001 - 10000 - 00001) \\
\widehat{C} &= \tfrac{1}{2}(10100 + 00101 - 10000 - 00001) \\
\widehat{D} &= \tfrac{1}{2}(10010 + 00011 - 10000 - 00001) \\
\widehat{E} &= \tfrac{1}{2}(10001 + 01001 - 10000 - 01000).
\end{aligned}$$

Hence the choice sets need to estimate all main effects and two-way interactions involving A are $S_1^{**} = S_1^* \bigcup (01000) = \{10000, 01000, 00001\}$ and $S_2^{**} = \{10001, 01001, 00101, 00011, 11000, 10100, 10010\}$. Obviously S_1^{**} and S_2^{**} are both PO sets.

Conclusion

If only the main effects are of interest, we need choice sets $S_1^* = \{10\ldots0, 0\ldots01\}$ and $S_2^* = \{\sum_{i=1}^{n-1} x_i = 1, x_n = 1\}$. If we want to estimate all main effects as well as all two-way interactions involving factor A_1 , we need to combine choice sets of the same level such that,

$$S_1^{**} = S_1^* \bigcup (010\ldots000),$$

and

$$S_2^{**} = \{x_1x_2\ldots x_n | \textstyle\sum_{i=1}^n x_i = 2, x_1 + x_n \neq 0\}.$$

3.3 Theorem for 3^n plans

For 3^n plans, consider only main effects and two-way interactions with factor A_1, the model we stated in section 2.3 could be simplified as:

$$y_{x_1x_2\ldots x_n} = \mu + \sum_{i=1}^{n} \alpha_{x_i}^{A_i} + \sum_{j=2}^{n} \alpha_{x_1x_j}^{A_1A_j} + e_{x_1x_2\ldots x_n}, \qquad (3.1)$$

where μ is the general mean, $\alpha_{x_i}^{A_i}$ is the effect of attribute A_i at level x_i, $\alpha_{x_1x_j}^{A_1A_j}$ is the effect of attributes A_1 and A_j at levels x_1 and x_j, respectively, and $e_{x_1x_2\ldots x_n}$

is the random error. Without loss of generality, we assume $\sum_{x_i} \alpha_{x_i}^{A_i} = 0$ and $\sum_{x_j} \alpha_{x_1 x_j}^{A_1 A_j} = 0$.

We now present a strategy for generating PO subsets capable of estimating main effects and all two-way interactions involving one factor in a 3^n plan.

Theorem 3.2 *The choice sets* $S_1^{**} = \{100\ldots000, 010\ldots000, 000\ldots001\}$, $S_2^{**} = \{x_1x_2\ldots x_n | \sum_{i=1}^{n} x_i = 2, x_i = 0 \text{ or } 1, x_1 + x_n \geq 1\} \bigcup (200\ldots000)$, $S_3^{**} = \{x_1x_2\ldots x_n | x_1 = 2, \sum_{i=2}^{n} x_i = 1\} \bigcup \{x_1x_2\ldots x_n | \sum_{i=2}^{n-1} x_i = 2, x_i = 0 \text{ or } 2, x_1 + x_n = 1\} \bigcup (100\ldots002, 010\ldots002)$, *and* $S_4^{**} = \{x_1x_2\ldots x_n | x_1 = 2, \sum_{i=2}^{n} x_i = 2, x_i = 0 \text{ or } 2\}$ *is a connected main effects plan and is capable of estimating the two-way interaction,* A_1A_i, $i = 2, 3, \ldots n$.

Proof: Suppose we want to estimate the two-way interactions between factor A_1 and A_2. By definition,

$$\begin{aligned}
A_{1_L}A_{2_L} &= \tfrac{1}{2}(\alpha_2^{A_1} - \alpha_0^{A_1})(\alpha_2^{A_2} - \alpha_0^{A_2}) \\
A_{1_L}A_{2_Q} &= \tfrac{1}{4}(\alpha_2^{A_1} - \alpha_0^{A_1})(\alpha_0^{A_2} + \alpha_2^{A_2} - 2 \times \alpha_1^{A_2}) \\
A_{1_Q}A_{2_L} &= \tfrac{1}{4}(\alpha_0^{A_1} + \alpha_2^{A_1} - 2 \times \alpha_1^{A_1})(\alpha_2^{A_2} - \alpha_0^{A_2}) \\
A_{1_Q}A_{2_Q} &= \tfrac{1}{8}(\alpha_0^{A_1} + \alpha_2^{A_1} - 2 \times \alpha_1^{A_1})(\alpha_0^{A_2} + \alpha_2^{A_2} - 2 \times \alpha_1^{A_2}).
\end{aligned}$$

So we need to find out linear combinations of profiles with expectation satisfying the above definition under the model given at the beginning of this section.

According to Raghavarao and Wiley (2006), the following nine profiles, (000. . . 001), (100. . . 000), (200. . . 000), (010. . . 001), (110. . . 000), (020. . . 001), (210. . . 000), (120. . . 000), and (220. . . 000), provide estimates for the four orthogonal contrasts of the interaction A_1A_2, such that:

$$\begin{aligned}
\widehat{A_{1_L}A_{2_L}} &= \tfrac{1}{2}(220\ldots000 + 000\ldots001 - 200\ldots000 - 020\ldots001) \\
\widehat{A_{1_L}A_{2_Q}} &= \tfrac{1}{4}(220\ldots000 + 200\ldots000 - 2 \times 210\ldots000 \\
&\quad -020\ldots001 - 000\ldots001 + 2 \times 010\ldots001) \\
\widehat{A_{1_Q}A_{2_L}} &= \tfrac{1}{4}(220\ldots000 + 020\ldots001 - 2 \times 120\ldots000 \\
&\quad -200\ldots000 - 000\ldots001 + 2 \times 100\ldots000)
\end{aligned}$$

$$\begin{aligned}\widehat{A_{1_Q}A}_{2_Q} &= \tfrac{1}{8}(220\ldots000 + 200\ldots000 - 2\times 210\ldots000 + 020\ldots001 \\ &\quad +000\ldots001 - 2\times 010\ldots001 - 2\times 100\ldots000 \\ &\quad -2\times 120\ldots000 + 4\times 110\ldots000).\end{aligned}$$

This result can be extended to all two-way interactions, A_1A_i, for $i = 2, 3, \ldots, n-1$, such that:

$$\begin{aligned}\widehat{A_{1_L}A}_{i_L} &= \tfrac{1}{2}(q_{6_i} + 000\ldots001 - 200\ldots000 - q_{3_i}) \\ \widehat{A_{1_L}A}_{i_Q} &= \tfrac{1}{4}(q_{6_i} + 200\ldots000 - 2\times q_{4_i} - q_{3_i} - 000\ldots001 + 2\times q_{1_i}) \\ A_{1_Q}A_{i_L} &= \tfrac{1}{4}(q_{6_i} + q_{3_i} - 2\times q_{5_i} - 200\ldots000 - 000\ldots001 + 2\times 100\ldots000) \\ \widehat{A_{1_Q}A}_{i_Q} &= \tfrac{1}{8}(q_{6_i} + 200\ldots000 - 2\times q_{4_i} + q_{3_i} + 000\ldots001 - 2\times q_{1_i} \\ &\quad -2\times q_{5_i} - 2\times 100\ldots000 + 4\times q_{2_i}),\end{aligned}$$

where,

$$\begin{aligned}q_{1_i} &= (x_1x_2\ldots x_n|\textstyle\sum_{\substack{j=1\\ j\neq i}}^{n-1} x_j = 0, x_i = x_n = 1) \\ q_{2_i} &= (x_1x_2\ldots x_n|x_1 = x_i = 1, \textstyle\sum_{\substack{j=2\\ j\neq i}}^{n} x_j = 0) \\ q_{3_i} &= (x_1x_2\ldots x_n|\textstyle\sum_{\substack{j=1\\ j\neq i}}^{n-1} x_j = 0, x_i = 2, x_n = 1) \\ q_{4_i} &= (x_1x_2\ldots x_n|x_1 = 2, x_i = 1, \textstyle\sum_{\substack{j=2\\ j\neq i}}^{n} x_j = 0) \\ q_{5_i} &= (x_1x_2\ldots x_n|x_1 = 1, x_i = 2, \textstyle\sum_{\substack{j=2\\ j\neq i}}^{n} x_j = 0) \\ q_{6_i} &= (x_1x_2\ldots x_n|x_1 = x_i = 2, \textstyle\sum_{\substack{j=2\\ j\neq i}}^{n} x_j = 0).\end{aligned}$$

In order to estimate A_1A_n, we need a factor other than A_1 to take the place of A_n in estimating $A_1A_i, i = 2, 3, \ldots, n-1$. Without loss of generality, we set this factor to A_2. Then the nine profiles, $(010\ldots000)$, $(100\ldots000)$, $(200\ldots000)$, $(010\ldots001)$, $(100\ldots001)$,$(010\ldots002)$,$(200\ldots001)$,$(100\ldots002)$, and $(200\ldots002)$, are capable of estimating the four orthogonal contrasts of the interaction A_1A_n, such that:

$$\begin{aligned}
\widehat{A_{1_L}A}_{n_L} &= \tfrac{1}{2}(200\ldots002 + 010\ldots000 - 200\ldots000 - 010\ldots002)\\
\widehat{A_{1_L}A}_{n_Q} &= \tfrac{1}{4}(200\ldots002 + 200\ldots000 - 2\times 200\ldots001\\
&\quad -010\ldots002 - 010\ldots000 + 2\times 010\ldots001)\\
\widehat{A_{1_Q}A}_{n_L} &= \tfrac{1}{4}(200\ldots002 + 010\ldots002 - 2\times 100\ldots002\\
&\quad -200\ldots000 - 010\ldots000 + 2\times 100\ldots000)\\
\widehat{A_{1_Q}A}_{n_Q} &= \tfrac{1}{8}(200\ldots002 + 200\ldots000 - 2\times 200\ldots001\\
&\quad +010\ldots002 + 010\ldots000 - 2\times 010\ldots001\\
&\quad -2\times 100\ldots002 - 2\times 100\ldots000 + 4\times 100\ldots001).
\end{aligned}$$

Therefore, the choice sets

$$\begin{aligned}
S_1^{**} &= \{100\ldots000, 010\ldots000, 000\ldots001\},\\
S_2^{**} &= \{x_1x_2\ldots x_n|\textstyle\sum_{i=1}^n x_i = 2, x_i = 0 \ or\ 1, x_1 + x_n \geq 1\}\bigcup(200\ldots000),\\
S_3^{**} &= \{x_1x_2\ldots x_n|x_1 = 2, \textstyle\sum_{i=2}^n x_i = 1\}\\
&\quad \bigcup\{x_1x_2\ldots x_n|\textstyle\sum_{i=2}^{n-1} x_i = 2, x_i = 0\ or\ 2, x_1 + x_n = 1\}\\
&\quad \bigcup(100\ldots002, 010\ldots002),\ \text{and}\\
S_4^{**} &= \{x_1x_2\ldots x_n|x_1 = 2, \textstyle\sum_{i=2}^n x_i = 2, x_i = 0\ or\ 2\},
\end{aligned}$$

are capable of estimating all two-way interaction involving A_1. These choice sets also provide a design for estimating all the main effects.

By definition, the linear effect and quadratic effect for factor A_i are:

$$\begin{aligned}
A_{iL} &= (\alpha_2^{A_i} - \alpha_0^{A_i})\\
A_{iQ} &= (\alpha_2^{A_i} + \alpha_0^{A_i} - 2\times\alpha_1^{A_i})
\end{aligned}$$

And based on the definition, the main effects for factor A_1 can be estimated as:

$$\begin{aligned}
\widehat{A}_{1_L} &= \tfrac{1}{3}[\textstyle\sum_{i=2}^n q_{6_i} + \sum_{i=2}^n q_{4_i} - (2n-5)\times 200\ldots000 - \sum_{i=2}^{n-1} q_{3_i}\\
&\quad -010\ldots002 - 010\ldots000 - \textstyle\sum_{i=3}^{n-1} q_{1_i} + (2n-7)\times 000\ldots001\\
&\quad +010\ldots001]
\end{aligned}$$

$$\begin{aligned}
\widehat{A}_{1_Q} &= \tfrac{1}{3}[\textstyle\sum_{i=2}^n q_{6_i} + \sum_{i=2}^n q_{4_i} - (2n-5)\times 200\ldots000 + \sum_{i=2}^{n-1} q_{3_i}\\
&\quad +010\ldots002 + 010\ldots000 + \textstyle\sum_{i=3}^{n-1} q_{1_i} - (2n-7)\times 000\ldots001\\
&\quad -010\ldots001 - 2\times\textstyle\sum_{i=2}^n q_{5_i} - 2\times\sum_{i=2}^n q_{2_i}\\
&\quad +2(2n-5)\times 100\ldots000]
\end{aligned}$$

For the factor A_i, $i = 2, 3, \ldots, n-1$, the estimates for its main effects are:

$$\begin{aligned}
\widehat{A}_{i_L} &= \tfrac{1}{3}(q_{6_i} + q_{5_i} + q_{3_i} - 200\ldots000 - 100\ldots000 - 000\ldots001) \\
\widehat{A}_{i_Q} &= \tfrac{1}{3}(q_{6_i} + q_{5_i} + q_{3_i} + 200\ldots000 + 100\ldots000 + 000\ldots001 \\
&\quad -2 \times q_{4_i} - 2 \times q_{2_i} - 2 \times q_{1_i}).
\end{aligned}$$

And for factor A_n, the estimates for its main effects are:

$$\begin{aligned}
\widehat{A}_{n_L} &= \tfrac{1}{3}(200\ldots002 + 100\ldots002 + 010\ldots002 - 200\ldots000 \\
&\quad -100\ldots000 - 010\ldots000), \\
\widehat{A}_{n_Q} &= \tfrac{1}{3}(200\ldots002 + 100\ldots002 + 010\ldots002 + 200\ldots000 \\
&\quad +100\ldots000 + 010\ldots000 - 2 \times 200\ldots001 - \\
&\quad 2 \times 100\ldots001 - 2 \times 010\ldots001).
\end{aligned}$$

3.4 Examples for 3^n plans

3.4.1 3^3 plan

Suppose we have three factors A, B and C, each at three levels.

(1) Consider the situation where only the main effects are of interest. Let $S_1^* = \{100, 010, 001\}$ and $S_2^* = \{011, 101, 002, 020, 200\}$. The linear effects and quadratic effects for factors A, B and C are:

$$\begin{aligned}
\widehat{A}_L &= 200 + 101 - 100 - 001 \\
\widehat{A}_Q &= 200 + 001 - 101 - 100 \\
\widehat{B}_L &= 020 + 011 - 010 - 001 \\
\widehat{B}_Q &= 020 + 001 - 010 - 011 \\
\widehat{C}_L &= 002 + 101 - 100 - 001 \\
\widehat{C}_Q &= 002 + 100 - 001 - 101
\end{aligned}$$

(2) Assume we want to estimate all main effects and only those two-way

interactions involving factor A.

$$
\begin{aligned}
\widehat{A_L B_L} &= \tfrac{1}{2}(220 + 001 - 200 - 021) \\
\widehat{A_L B_Q} &= \tfrac{1}{4}(220 + 200 - 2 \times 210 - 021 - 001 + 2 \times 011) \\
\widehat{A_Q B_L} &= \tfrac{1}{4}(220 + 021 - 2 \times 120 - 200 - 001 + 2 \times 100) \\
\widehat{A_Q B_Q} &= \tfrac{1}{8}(220 + 200 - 2 \times 210 + 021 + 001 - 2 \times 011 - 2 \times 120 \\
&\quad -2 \times 100 + 4 \times 110) \\
\widehat{A_L C_L} &= \tfrac{1}{2}(202 + 010 - 200 - 012) \\
\widehat{A_L C_Q} &= \tfrac{1}{4}(202 + 200 - 2 \times 201 - 012 - 010 + 2 \times 011) \\
\widehat{A_Q C_L} &= \tfrac{1}{4}(202 + 012 - 2 \times 102 - 200 - 010 + 2 \times 100) \\
\widehat{A_Q C_Q} &= \tfrac{1}{8}(202 + 200 - 2 \times 201 + 012 + 010 - 2 \times 011 - 2 \times 102 \\
&\quad -2 \times 100 + 4 \times 101).
\end{aligned}
$$

In this case the estimates for main effect given in step (1) are no longer unbiased because of the existence of interactions. After adjusting for the presence of two-way interactions involving A, the unbiased estimators of the main effects are:

$$
\begin{aligned}
\widehat{A}_L &= \tfrac{1}{3}[(220 + 202 + 210 + 201 - 200) - (021 + 012 \\
&\quad +010 + 001 - 011)] \\
\widehat{A}_Q &= \tfrac{1}{3}[(220 + 202 + 210 + 201 - 200 + 021 + 012 \\
&\quad +010 + 001 - 011) - 2 \times (120 + 102 + 110 + 101 - 100)] \\
\widehat{B}_L &= \tfrac{1}{3}(220 - 200 + 120 - 100 + 021 - 001) \\
\widehat{B}_Q &= \tfrac{1}{3}(220 + 120 + 021 + 200 + 100 + 001 - 2 \times 210 \\
&\quad -2 \times 110 - 2 \times 011) \\
\widehat{C}_L &= \tfrac{1}{3}(202 - 200 + 102 - 100 + 012 - 010) \\
\widehat{C}_Q &= \tfrac{1}{3}(202 + 102 + 012 + 200 + 100 + 010 - 2 \times 201 \\
&\quad -2 \times 101 - 2 \times 011).
\end{aligned}
$$

The choice sets used for this plan are $S_1^{**} = \{100, 010, 001\}$, $S_2^{**} = \{200, 110, 101, 011\}$, $S_3^{**} = \{201, 210, 102, 120, 021, 012\}$, and $S_4^{*} = \{220, 202\}$. Obviously all the four choice sets are PO sets.

3.4.2 3^4 Plan

Suppose we have four factors, A, B, C and D, each at three levels.

(1) Consider the situation where only the main effects are of interest. Let $S_1^* = \{1000, 0100, 0010, 0001\}$ and $S_2^* = \{0002, 0020, 0200, 2000, 0011, 0101, 1001\}$. The linear effects and quadratic effects for factors A, B, C and D are:

$$\begin{aligned}
\widehat{A}_L &= 2000 + 1001 - 1000 - 0001 \\
\widehat{A}_Q &= 2000 + 0001 - 1001 - 1000 \\
\widehat{B}_L &= 0200 + 0101 - 0100 - 0001 \\
\widehat{B}_Q &= 0200 + 0001 - 0100 - 0101 \\
\widehat{C}_L &= 0020 + 0011 - 0010 - 0001 \\
\widehat{C}_Q &= 0020 + 0001 - 0010 - 0011 \\
\widehat{D}_L &= 0002 + 1001 - 1000 - 0001 \\
\widehat{D}_Q &= 0002 + 1000 - 0001 - 1001.
\end{aligned}$$

(2) Assume we want to estimate all main effects and only those two-way interactions involving factor A.

$$\begin{aligned}
\widehat{A_L B_L} &= \tfrac{1}{2}(2200 + 0001 - 2000 - 0201) \\
\widehat{A_L B_Q} &= \tfrac{1}{4}(2200 + 2000 - 2 \times 2100 - 0201 - 0001 + 2 \times 0101) \\
\widehat{A_Q B_L} &= \tfrac{1}{4}(2200 + 0201 - 2 \times 1200 - 2000 - 0001 + 2 \times 1000) \\
\widehat{A_Q B_Q} &= \tfrac{1}{8}(2200 + 2000 - 2 \times 2100 + 0201 + 0001 - 2 \times 0101 \\
&\quad -2 \times 1200 - 2 \times 1000 + 4 \times 1100) \\
\widehat{A_L C_L} &= \tfrac{1}{2}(2020 + 0001 - 2000 - 0021) \\
\widehat{A_L C_Q} &= \tfrac{1}{4}(2020 + 2000 - 2 \times 2010 - 0021 - 0001 + 2 \times 0011) \\
\widehat{A_Q C_L} &= \tfrac{1}{4}(2020 + 0021 - 2 \times 1020 - 2000 - 0001 + 2 \times 1000) \\
\widehat{A_Q C_Q} &= \tfrac{1}{8}(2020 + 2000 - 2 \times 2010 + 0021 + 0001 - 2 \times 0011 \\
&\quad -2 \times 1020 - 2 \times 1000 + 4 \times 1010) \\
\widehat{A_L D_L} &= \tfrac{1}{2}(2002 + 0100 - 2000 - 0102) \\
\widehat{A_L D_Q} &= \tfrac{1}{4}(2002 + 2000 - 2 \times 2001 - 0102 - 0100 + 2 \times 0101) \\
\widehat{A_Q D_L} &= \tfrac{1}{4}(2002 + 0102 - 2 \times 1002 - 2000 - 0100 + 2 \times 1000) \\
\widehat{A_Q D_Q} &= \tfrac{1}{8}(2002 + 2000 - 2 \times 2001 + 0102 + 0100 - 2 \times 0101 \\
&\quad -2 \times 1002 - 2 \times 1000 + 4 \times 1001).
\end{aligned}$$

The estimates for the main effect given in step (1) are no longer unbiased. After adjusting for the presence of two-way interactions involving A, the main effects can be estimated as:

$$\begin{aligned}
\widehat{A}_L &= \tfrac{1}{3}[(2200+2020+2002+2100+2010+2001-3\times 2000)\\
&\quad -(0201+0021+0102+0100+0011-0001-0101)]\\
\widehat{A}_Q &= \tfrac{1}{3}[(2200+2020+2002+2100+2010+2001-3\times 2000)\\
&\quad +(0201+0021+0102+0100+0011-0001-0101)\\
&\quad -2\times(1200+1020+1002+1100+1010+1001-3\times 1000)]\\
\hat{B}_L &= \tfrac{1}{3}(2200-2000+1200-1000+0201-0001)\\
\hat{B}_Q &= \tfrac{1}{3}(2200+1200+0201+2000+1000+0001-2\times 2100\\
&\quad -2\times 1100-2\times 0101)\\
\widehat{C}_L &= \tfrac{1}{3}(2020-2000+1020-1000+0021-0001)\\
\widehat{C}_Q &= \tfrac{1}{3}(2020+1020+0021+2000+1000+0001-2\times 2010\\
&\quad -2\times 1010-2\times 0011)\\
\widehat{D}_L &= \tfrac{1}{3}(2002-2000+1002-1000+0102-0100)\\
\widehat{D}_Q &= \tfrac{1}{3}(2002+1002+0102+2000+1000+0100-2\times 2001\\
&\quad -2\times 1001-2\times 0101).
\end{aligned}$$

The choice sets used for this plan are $S_1^{**}=\{1000,0100,0001\}$, $S_2^{**}=\{2000, 1100,1010,1001,0101,0011\}$, $S_3^{**} = \{2001,2010,2100,1002,1020,1200,0021, 0201,0102\}$, and, $S_4^{*}=\{2200,2020,2002\}$. Obviously all the four choice sets are PO sets.

3.4.3 3^5 Plan

Suppose we have five factors, A, B, C, D and E, each at three levels.

(1) Consider the situation where only the main effects are of interest. Let $S_1^{*}=\{10000,01000,00100,00010,00001\}$ and $S_2^{*}=\{00002,00020,00200,02000, 2000,00011,00101,01001,10001\}$. The linear effects and quadratic effects for factor A, B, C and D are:

$$\begin{aligned}
\widehat{A}_L &= 20000+10001-10000-00001\\
\widehat{A}_Q &= 20000+00001-10001-10000
\end{aligned}$$

$$
\begin{aligned}
\widehat{B}_L &= 02000 + 01001 - 01000 - 00001\\
\widehat{B}_Q &= 02000 + 00001 - 01000 - 01001\\
\widehat{C}_L &= 00200 + 00101 - 00100 - 00001\\
\widehat{C}_Q &= 00200 + 00001 - 00100 - 00101\\
\widehat{D}_L &= 00020 + 00011 - 00010 - 00001\\
\widehat{D}_Q &= 00020 + 00001 - 00010 - 00011\\
\widehat{E}_L &= 00002 + 10001 - 10000 - 00001\\
\widehat{E}_Q &= 00002 + 10000 - 00001 - 10001.
\end{aligned}
$$

(2) Assume we want to estimate all main effects and only those two-way interactions involving factor A.

$$
\begin{aligned}
\widehat{A_L B_L} &= \tfrac{1}{2}(22000 + 00001 - 20000 - 02001)\\
\widehat{A_L B_Q} &= \tfrac{1}{4}(22000 + 20000 - 2 \times 21000 - 02001 - 00001 + 2 \times 01001)\\
\widehat{A_Q B_L} &= \tfrac{1}{4}(22000 + 02001 - 2 \times 12000 - 20000 - 00001 + 2 \times 10000)\\
\widehat{A_Q B_Q} &= \tfrac{1}{8}(22000 + 20000 - 2 \times 21000 + 02001 + 00001 - 2 \times 01001\\
&\quad -2 \times 12000 - 2 \times 10000 + 4 \times 11000)\\
\widehat{A_L C_L} &= \tfrac{1}{2}(20200 + 00001 - 20000 - 00201)\\
\widehat{A_L C_Q} &= \tfrac{1}{4}(20200 + 20000 - 2 \times 20100 - 00201 - 00001 + 2 \times 00101)\\
\widehat{A_Q C_L} &= \tfrac{1}{4}(20200 + 00201 - 2 \times 10200 - 20000 - 00001 + 2 \times 10000)\\
\widehat{A_Q C_Q} &= \tfrac{1}{8}(20200 + 20000 - 2 \times 20100 + 00201 + 00001 - 2 \times 00101\\
&\quad -2 \times 10200 - 2 \times 10000 + 4 \times 10100)\\
\widehat{A_L D_L} &= \tfrac{1}{2}(20020 + 00001 - 20000 - 00021)\\
\widehat{A_L D_Q} &= \tfrac{1}{4}(20020 + 20000 - 2 \times 20010 - 00021 - 00001 + 2 \times 00011)\\
\widehat{A_Q D_L} &= \tfrac{1}{4}(20020 + 00021 - 2 \times 10020 - 20000 - 00001 + 2 \times 10000)\\
\widehat{A_Q D_Q} &= \tfrac{1}{8}(20020 + 20000 - 2 \times 20010 + 00021 + 00001 - 2 \times 00011\\
&\quad -2 \times 10020 - 2 \times 10000 + 4 \times 10010)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{A_L E_L} &= \tfrac{1}{2}(20002 + 01000 - 20000 - 01002)\\
\widehat{A_L E_Q} &= \tfrac{1}{4}(20002 + 20000 - 2 \times 20001 - 01002 - 01000 + 2 \times 01001)\\
\widehat{A_Q E_L} &= \tfrac{1}{4}(20002 + 01002 - 2 \times 10002 - 20000 - 01000 + 2 \times 10000)\\
\widehat{A_Q E_Q} &= \tfrac{1}{8}(20002 + 20000 - 2 \times 20001 + 01002 + 01000 - 2 \times 01001\\
&\quad -2 \times 10002 - 2 \times 10000 + 4 \times 10001).
\end{aligned}
$$

The estimates for the main effect given in step (1) are no longer unbiased. After adjusting for the presence of two-way interactions involving A, the main effects can be estimated as:

$$\begin{aligned}
\widehat{A}_L &= \tfrac{1}{3}[(22000 + 20200 + 20020 + 20002 + 21000 + 20100 + 20010 \\
&\quad +20001 - 5 \times 20000) - (02001 + 00201 + 00021 + 01002 + \\
&\quad 01000 + 00101 + 00011 - 3 \times 00001 - 01001)] \\
\widehat{A}_Q &= \tfrac{1}{3}[(22000 + 20200 + 20020 + 20002 + 21000 + 20100 + 20010 \\
&\quad +20001 - 5 \times 20000) + (02001 + 00201 + 00021 + 01002 + \\
&\quad 01000 + 00101 + 00011 - 3 \times 00001 - 01001) \\
&\quad -2 \times (12000 + 10200 + 10020 + 10002 + 11000 + 10100 + 10010 \\
&\quad +10001 - 5 \times 10000)] \\
\widehat{B}_L &= \tfrac{1}{3}(22000 - 20000 + 12000 - 10000 + 02001 - 00001) \\
\widehat{B}_Q &= \tfrac{1}{3}(22000 + 12000 + 02000 + 20000 + 10000 + 00001 - 2 \times 21000 \\
&\quad -2 \times 11000 - 01000 - 01001) \\
\widehat{C}_L &= \tfrac{1}{3}(20200 - 20000 + 10200 - 10000 + 00201 - 00001) \\
\widehat{C}_Q &= \tfrac{1}{3}(20200 + 10200 + 00200 + 20000 + 10000 + 00001 - 2 \times 20100 \\
&\quad -2 \times 10100 - 00100 - 00101) \\
\widehat{D}_L &= \tfrac{1}{3}(20020 - 20000 + 10020 - 10000 + 00021 - 00001) \\
\widehat{D}_Q &= \tfrac{1}{3}(20020 + 10020 + 00020 + 20000 + 10000 + 00001 - 2 \times 20010 \\
&\quad -2 \times 10010 - 00010 - 00011) \\
\widehat{E}_L &= \tfrac{1}{3}(20002 - 20000 + 10002 - 10000 + 01002 - 01000) \\
\widehat{E}_Q &= \tfrac{1}{3}(20002 + 10002 + 00002 + 20000 + 10000 + 01000 - 2 \times 20001 \\
&\quad -2 \times 10001 - 00001 - 01001).
\end{aligned}$$

The choice sets used for this plan are $S_1^{**} = \{10000, 01000, 00001\}$, $S_2^{**} = \{20000, 10001, 10010, 10100, 11000, 01001, 00101, 00011\}$, $S_3^{**} = \{20001, 20010, 20100, 21000, 10002, 10020, 10200, 12000, 00021, 00201, 02001, 01002\}$, and, $S_4^{*} = \{20002, 20020, 20200, 22000\}$. Obviously all the four choice sets are PO sets.

According to Raghavarao and Wiley (2006), in order to estimate all main effects and two-way interactions, at lease 90 profiles are needed. But if we only want to estimate main effects and two-way interactions involving one factor, by following our approach, only 27 profiles are needed.

Conclusion

If only the main effects are of interest, we need choice sets $S_1^* = S_1 = \{x_1x_2\ldots x_n|\sum_{i=1}^n x_i = 1\}$ and $S_2^* = \{x_1x_2\ldots x_n|\sum_{i=1}^n x_i = 2, x_i = 0\ or\ 2\}\bigcup \{x_1x_2\ldots x_n|\sum_{i=1}^{n-1} x_i = 1,\ x_n = 1\}$. If we want to estimate all main effects as well as all two-way interactions involving factor A_1 , we need the following four choice sets,

$$\begin{aligned} S_1^{**} &= \{100\ldots000, 010\ldots000, 000\ldots001\}, \\ S_2^{**} &= \{x_1x_2\ldots x_n|\textstyle\sum_{i=1}^n x_i = 2, x_i = 0\ or\ 1, x_1 + x_n \geq 1\}\bigcup(200\ldots000), \\ S_3^{**} &= \{x_1x_2\ldots x_n|x_1 = 2, \textstyle\sum_{i=2}^n x_i = 1\}\bigcup \\ &\quad \{x_1x_2\ldots x_n|\textstyle\sum_{i=2}^{n-1} x_i = 2, x_i = 0\ or\ 2, x_1 + x_n = 1\}\bigcup \\ &\quad (100\ldots002, 010\ldots002), \end{aligned}$$

and

$$S_4^{**} = \{x_1x_2\ldots x_n|x_1 = 2, \textstyle\sum_{i=2}^n x_i = 2, x_i = 0\ or\ 2\}.$$

CHAPTER 4

Estimating a 3-way interaction in 2^n plan

In previous chapters, we obtained designs capable of estimating all main effects and all two-way interactions inclusive of one factor for both 2^n and 3^n plans. According to the sparsity-of-effects principle, a system is usually dominated by main effects and low-order interactions. Thus it is most likely that main effects and two-way interactions are the most significant responses. In other words, higher order interactions such as three-way interactions are very rare. Therefore, it is usually not necessary or realistic to investigate all main effects and higher order interactions. In this chapter we focus on the situation in which we are interested in estimating all main effects, all two-way interactions involving one factor, and only one specific three-way interaction.

4.1 Theorem

Theorem 4.1 *The choice sets* $S_1^{**} = \{ 100\ldots000,\ 010\ldots000,\ 0010\ldots001,\ 000\ldots001\}$, $S_2^{**} = \{x_1x_2\ldots x_n|\ x_1 = 1,\ \sum_{i=2}^{n} x_i = 1\} \bigcup \{x_1x_2\ldots x_n|\ x_1 = 0,\ x_2 = 1,\ \sum_{i=3}^{n} x_i = 1\} \bigcup \{x_1x_2\ldots x_n \mid x_1 = x_2 = 0,\ x_3 = 1, \sum_{i=4}^{n} x_i = 1\} \bigcup \{x_1x_2\ldots x_n|x_1 = x_2 = x_3 = 0, \sum_{i=4}^{n-1} x_i = 1, x_n = 1\}$, *and* $S_3^{**} = \{x_1x_2\ldots x_n|\ x_1 = x_2 = 1, \sum_{i=3}^{n} x_i = 1\} \bigcup \{x_1x_2\ldots x_n|x_1 = x_3 = 1, x_2 = 0, \sum_{i=4}^{n} x_i =$

$1\} \bigcup \{x_1x_2\dots x_n | x_1 = x_n = 1, x_2 = x_3 = 0, \sum_{i=4}^{n-1} x_i = 1\}$ *is a connected main effects plan and is capable of estimating all two-way interactions involving factor* A_1, A_1A_i, $i = 2, 3, \dots n$, *and one three-way interaction* $A_1A_2A_3$.

Proof of Theorem 4.1 is by verification. For a 2^n plan, if we want to estimate all three-way and lower effects, we need profiles from four consecutive choice sets, S_l, S_{l+1}, S_{l+2} and S_{l+3}, where $0 \le l \le (n-3)$ (Raghavarao and Wiley, 2006). But if we are only interested in all main effects, the two-way interactions involving one specific factor, and one three-way interaction involving the same factor, we do not need all profiles in the four choice sets.

Recall that in Chapter 2, we stated the full model for a s^n plan (equation 2.1). In this chapter, the model could be simplified as:

$$y_{x_1x_2\dots x_n} = \mu + \sum_{i=1}^{n} \alpha_{x_i}^{A_i} + \sum_{j=1}^{n} \alpha_{x_1x_j}^{A_1A_j} + \alpha_{x_1x_2x_3}^{A_1A_2A_3} + e_{x_1x_2\dots x_n}, \tag{4.1}$$

where μ is the general mean, $\alpha_{x_i}^{A_i}$ is the effect of attribute A_i at level x_i, $\alpha_{x_1x_j}^{A_1A_j}$ is the effect of attributes A_1 and A_j at levels x_1 and x_j respectively, $\alpha_{x_1x_2x_3}^{A_1A_2A_3}$ is the effect of attributes A_1, A_2 and A_3 at levels x_1, x_2 and x_3, and $e_{x_1x_2\dots x_n}$ is the random error. Without loss of generality, we assume $\sum\limits_{x_i} \alpha_{x_i}^{A_i} = 0$, $\sum\limits_{x_j} \alpha_{x_1x_j}^{A_1A_j} = 0$. For simplicity, we define the following profiles for $i = 2, 3, \dots, n$:

$$\begin{aligned}
\delta_{1i} &= (x_1x_2\dots x_n | \sum_{\substack{j=2 \\ j\neq i}}^{n} x_j = 0, x_1 = x_i = 1), \\
\delta_{2i} &= (x_1x_2\dots x_n | \sum_{\substack{j=1 \\ j\neq 2 \text{ or } i}}^{n} x_j = 0, x_2 = x_i = 1), \\
\delta_{3i} &= (x_1x_2\dots x_n | \sum_{\substack{j=1 \\ j\neq 3 \text{ or } i}}^{n} x_j = 0, x_3 = x_i = 1), \\
\delta_{ni} &= (x_1x_2\dots x_n | \sum_{\substack{j=2 \\ j\neq i}}^{n-1} x_j = 0, x_i = x_n = 1), \\
\lambda_{2i} &= (x_1x_2\dots x_n | \sum_{\substack{j=3 \\ j\neq i}}^{n} x_j = 0, x_1 = x_2 = x_i = 1), \\
\lambda_{3i} &= (x_1x_2\dots x_n | \sum_{\substack{j=2 \\ j\neq 3 \text{ or } i}}^{n} x_j = 0, x_1 = x_3 = x_i = 1),
\end{aligned}$$

$$\lambda_{ni} = (x_1 x_2 \ldots x_n | \sum_{\substack{j=2 \\ j \neq i}}^{n-1} x_j = 0, x_1 = x_n = x_i = 1).$$

By definition, the main effect for factor A_1 is

$$A_1 = \tfrac{1}{2}(\alpha_1^{A_1} - \alpha_0^{A_1}),$$

Therefore, we can verify that an unbiased estimator for A_1 is:

$$\begin{aligned} \widehat{A}_1 = {} & \tfrac{1}{(2n+3)}\{[\lambda_{23} + \delta_{12} + \delta_{13} + 2 \times \textstyle\sum_{i=4}^{n} \delta_{1i} - (2n-7) \times 100\ldots000] \\ & -[000\ldots001 + \textstyle\sum_{i=3}^{n} \delta_{2i} + \sum_{i=4}^{n-1} \delta_{3i} - (n-4) \times 010\ldots000 \\ & -(n-5) \times 0010\ldots000]\}. \end{aligned}$$

Following the same logic, we can find the unbiased estimators for the main effects of A_2 and A_3 as:

$$\begin{aligned} \widehat{A}_2 = {} & \tfrac{1}{7}(\lambda_{23} + \delta_{12} + \delta_{2n} + \delta_{23} - \delta_{13} - 100\ldots000 - 000\ldots001 \\ & -0010\ldots000), \\ \widehat{A}_3 = {} & \tfrac{1}{7}(\lambda_{23} + \delta_{13} + \delta_{3n} + \delta_{23} - \delta_{12} - 100\ldots000 - 000\ldots001 \\ & -010\ldots000). \end{aligned}$$

In general, for $k = 4, 5, \ldots, n-1$,

$$\begin{aligned} \widehat{A}_k = {} & \tfrac{1}{6}(\lambda_{2k} + \delta_{1k} + \delta_{2k} + \delta_{nk} - \delta_{12} - 100\ldots000 - 010\ldots000 \\ & -000\ldots001). \end{aligned}$$

The estimator of main effect for factor A_n is a little bit different. When estimating the main effects for other factors, we used the profile $(000\ldots001)$ in order to avoid the dominated profile $(000\ldots000)$. Hence, we need some other factor to take the role of factor A_n in estimating the main effects of factor $A_i (i = 1, 2, \ldots, n-1)$. And a proper estimator for the main effect of A_n is:

$$\begin{aligned} \widehat{A}_n = {} & \tfrac{1}{6}(\lambda_{2n} + \delta_{1n} + \delta_{2n} + \delta_{3n} - \delta_{12} - 100\ldots000 - 010\ldots000 \\ & -0010\ldots000) \end{aligned}$$

Similarly, we can verify that proper unbiased estimators for all two-way interactions involving factor A_1 are:

$$
\begin{aligned}
\widehat{A_1A_2} &= \tfrac{1}{3}(\lambda_{23} + \lambda_{2n} + 0010\ldots000 + 000\ldots001 - \delta_{13} - \delta_{1n} - \delta_{23} - \delta_{2n}),\\
\widehat{A_1A_3} &= \tfrac{1}{3}(\lambda_{23} + \lambda_{3n} + 010\ldots000 + 000\ldots001 - \delta_{12} - \delta_{1n} - \delta_{23} - \delta_{3n}),\\
\widehat{A_1A_k} &= \tfrac{1}{2}(\lambda_{2k} + \lambda_{nk} + 010\ldots000 + 000\ldots001 - \delta_{12} - \delta_{1n} - \delta_{2k} - \delta_{nk}),\\
\widehat{A_1A_n} &= \tfrac{1}{2}(\lambda_{2n} + \lambda_{3n} + 010\ldots000 + 0010\ldots000 - \delta_{12} - \delta_{13} - \delta_{2n} - \delta_{3n}).
\end{aligned}
$$

Finally, a proper unbiased estimator for $A_1A_2A_3$ is:

$$
\begin{aligned}
\widehat{A_1A_2A_3} &= \tfrac{1}{2}(\lambda_{23} + 010\ldots000 + 100\ldots000 + \delta_{3n} - \delta_{12} - \delta_{13} - \delta_{23}\\
&\quad -000\ldots001).
\end{aligned}
$$

In summary, the profiles that we used to estimate all main effects, all two-way interactions involving one factor and one three-way interaction involving the same factor, come from three consecutive Pareto Optimal choice sets:

$$
\begin{aligned}
S_1^{**} &= \{100\ldots000, 010\ldots000, 0010\ldots001, 000\ldots001\},\\
S_2^{**} &= \{x_1x_2\ldots x_n | x_1 = 1, \textstyle\sum_{i=2}^n x_i = 1\} \bigcup \{x_1x_2\ldots x_n | x_1 = 0,\\
&\quad x_2 = 1, \textstyle\sum_{i=3}^n x_i = 1\} \bigcup \{x_1x_2\ldots x_n | x_1 = x_2 = 0, x_3 = 1,\\
&\quad \textstyle\sum_{i=4}^n x_i = 1\} \bigcup \{x_1x_2\ldots x_n | x_1 = x_2 = x_3 = 0, \textstyle\sum_{i=4}^{n-1} x_i = 1, x_n = 1\},
\end{aligned}
$$

and

$$
\begin{aligned}
S_3^{**} &= \{x_1x_2\ldots x_n | x_1 = x_2 = 1, \textstyle\sum_{i=3}^n x_i = 1\} \bigcup \{x_1x_2\ldots x_n | x_1 = x_3 = 1,\\
&\quad x_2 = 0, \textstyle\sum_{i=4}^n x_i = 1, \} \bigcup \{x_1x_2\ldots x_n | x_1 = x_n = 1, x_2 = x_3 = 0,\\
&\quad \textstyle\sum_{i=4}^{n-1} x_i = 1\}.
\end{aligned}
$$

4.2 Examples

4.2.1 2^4 plan

Suppose we have four factors, A, B, C and D, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving factor A and one three-way interaction ABC.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{11}[(1110 + 1100 + 1010 + 2 \times 1001 - 1000) - (0110 + 0001 \\
&\quad +0101 + 0010)] \\
\widehat{B} &= \tfrac{1}{7}[(1110 + 1100 + 0101 + 0110) - (1010 + 1000 + 0001 + 0010)] \\
\widehat{C} &= \tfrac{1}{7}[(1110 + 1010 + 0011 + 0110) - (1100 + 1000 + 0001 + 0100)] \\
\widehat{D} &= \tfrac{1}{6}[(1101 + 1001 + 0101 + 0011) - (1100 + 1000 + 0100 + 0010)] \\
\widehat{AB} &= \tfrac{1}{3}[(1110 + 1101 + 0010 + 0001) - (1010 + 1001 + 0110 + 0101)] \\
\widehat{AC} &= \tfrac{1}{3}[(1110 + 1011 + 0100 + 0001) - (1100 + 1001 + 0110 + 0011)] \\
\widehat{AD} &= \tfrac{1}{2}[(1101 + 1011 + 0100 + 0010) - (1100 + 1010 + 0101 + 0011)] \\
\widehat{ABC} &= \tfrac{1}{2}[(1110 + 0100 + 1000 + 0011) - (1100 + 1010 + 0110 + 0001)]
\end{aligned}$$

To verify the unbiasedness of estimators, we used the condition in eq(2.1). So the choice sets need to estimate all main effects, two-way interactions involving A and one three-way interaction ABC are:

$$\begin{aligned}
S_1^* &= \{1000, 0100, 0010, 0001\} = S_1 \\
S_2^* &= \{1100, 1010, 1001, 0110, 0101, 0011\} = S_2 \\
S_3^* &= \{1110, 1101, 1011\}
\end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects, two- and three-way interactions, we need at least 15 profiles and have to include a dominating (or dominated) profile. Using the above method we only need 13 profiles and all the choice sets used are Pareto Optimal choice sets.

4.2.2 2^5 plan

Suppose we have five factors, A, B, C, D and E, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving factor A and one three-way interaction ABC.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{13}[(11100 + 11000 + 10100 + 2 \times 10010 + 2 \times 10001 - 3 \times 10000) \\
&\quad -(01100 + 00001 + 01010 + 01001 + 00110 - 01000)]
\end{aligned}$$

$$\begin{aligned}
\widehat{B} &= \tfrac{1}{7}[(11100 + 11000 + 01001 + 01100) \\
&\quad -(10100 + 10000 + 00001 + 00100)] \\
\widehat{C} &= \tfrac{1}{7}[(11100 + 10100 + 00101 + 01100) \\
&\quad -(11000 + 10000 + 00001 + 01000)] \\
\widehat{D} &= \tfrac{1}{6}[(11010 + 10010 + 01010 + 00011) \\
&\quad -(11000 + 10000 + 01000 + 00001)] \\
\widehat{E} &= \tfrac{1}{6}[(11001 + 10001 + 01001 + 00101) \\
&\quad -(11000 + 10000 + 01000 + 00100)] \\
\widehat{AB} &= \tfrac{1}{3}[(11100 + 11001 + 00100 + 00001) \\
&\quad -(10100 + 10001 + 01100 + 01001)] \\
\widehat{AC} &= \tfrac{1}{3}[(11100 + 10101 + 01000 + 00001) \\
&\quad -(11000 + 10001 + 01100 + 00101)] \\
\widehat{AD} &= \tfrac{1}{2}[(11010 + 10011 + 01000 + 00001) \\
&\quad -(11000 + 10001 + 01010 + 00011)] \\
\widehat{AE} &= \tfrac{1}{2}[(11001 + 10101 + 01000 + 00100) \\
&\quad -(11000 + 10100 + 01001 + 00101)] \\
\widehat{ABC} &= \tfrac{1}{2}[(11100 + 01000 + 10000 + 00101) \\
&\quad -(11000 + 10100 + 01100 + 00001)]
\end{aligned}$$

To verify the unbiasedness of estimators, we used the condition in eq(2.1). So the choice sets need to estimate all main effects, two-way interactions involving A and one three-way interaction ABC are:

$$\begin{aligned}
S_1^* &= \{10000, 01000, 00100, 00001\} \\
S_2^* &= \{11000, 10100, 10010, 10001, 01100, 01010, 01001, 00110, 00101, \\
&\quad 00011\} = S_2 \\
S_3^* &= \{11100, 11010, 11001, 10101, 10011\}
\end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects, two- and three-way interactions, we need at least 30 profiles. Using the above method we only need 19 profiles and all the choice sets used are Pareto Optimal choice sets.

4.2.3 2^6 plan

Suppose we have six factors, A, B, C, D, E and F, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving factor A and one three-way interaction ABC.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{15}[(111000 + 110000 + 101000 + 2 \times 100100 + 2 \times 100010 \\
&\quad +2 \times 100001 - 5 \times 100000) - (011000 + 000001 + 010100 \\
&\quad +010010 + 010001 + 001100 + 001010 - 2 \times 010000 - 001000)] \\
\widehat{B} &= \tfrac{1}{7}[(111000 + 110000 + 010001 + 011000) \\
&\quad -(101000 + 100000 + 000001 + 001000)] \\
\widehat{C} &= \tfrac{1}{7}[(111000 + 101000 + 001001 + 011000) \\
&\quad -(110000 + 100000 + 000001 + 010000)] \\
\widehat{D} &= \tfrac{1}{6}[(110100 + 100100 + 010100 + 000101) \\
&\quad -(110000 + 100000 + 010000 + 000001)] \\
\widehat{E} &= \tfrac{1}{6}[(110010 + 100010 + 010010 + 000011) \\
&\quad -(110000 + 100000 + 010000 + 000001)] \\
\widehat{F} &= \tfrac{1}{6}[(110001 + 100001 + 010001 + 001001) \\
&\quad -(110000 + 100000 + 010000 + 001000)] \\
\widehat{AB} &= \tfrac{1}{3}[(111000 + 110001 + 001000 + 000001) \\
&\quad -(101000 + 100001 + 011000 + 010001)] \\
\widehat{AC} &= \tfrac{1}{3}[(111000 + 101001 + 010000 + 000001) \\
&\quad -(110000 + 100001 + 011000 + 001001)] \\
\widehat{AD} &= \tfrac{1}{2}[(110100 + 100101 + 010000 + 000001) \\
&\quad -(110000 + 100001 + 010100 + 000101)] \\
\widehat{AE} &= \tfrac{1}{2}[(110010 + 100011 + 010000 + 000001) \\
&\quad -(110000 + 100001 + 010010 + 000011)] \\
\widehat{AF} &= \tfrac{1}{2}[(110001 + 101001 + 010000 + 001000) \\
&\quad -(110000 + 101000 + 010001 + 001001)] \\
\widehat{ABC} &= \tfrac{1}{2}[(111000 + 010000 + 100000 + 001001) \\
&\quad -(110000 + 101000 + 011000 + 000001)]
\end{aligned}$$

So the choice sets need to estimate all main effects, two-way interactions involving A and one three-way interaction ABC are:

$$\begin{aligned} S_1^* &= \{100000, 010000, 001000, 000001\} \\ S_2^* &= \{110000, 101000, 100100, 100010, 100001, 011000, 010100, \\ &\quad 010010, 010001, 001100, 001010, 001001, 000101, 000011\} \\ S_3^* &= \{111000, 110100, 110010, 110001, 101010, 101001, 100101, 100011\} \end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects, two- and three-way interactions, we need at least 56 profiles. Using the above method we only need 26 profiles and all the choice sets used are Pareto Optimal choice sets.

CHAPTER 5

Mixed-level designs

The 2^n and 3^n designs discussed in previous chapters are symmetrical designs, i.e. they have the same number of levels for each factor. In this chapter we consider mixed-level designs or asymmetrical designs. An asymmetrical design is a factorial experiment involving n factors all with possibly different numbers of levels. Consider a $2^n \cdot 3^m$ design with $n+m$ factors. Factors A_1 to A_n are two-level factors while factors A_{n+1} to A_{n+m} are all three-level factors. First, we want to estimate all the main effects for this mixed-level plan.

5.1 Theorem for a $2^n \cdot 3^m$ plan

Theorem 5.1 *The choice sets* $S_1^* = (100\ldots000) \bigcup \{x_1x_2\ldots x_{n+m} | \sum_{i=1}^{n} x_i = 0, \sum_{i=n+1}^{n+m} x_i = 1\}$ *and* $S_2^* = \{x_1x_2\ldots x_{n+m} | \sum_{i=1}^{n} x_i = 1, \sum_{i=n+1}^{n+m-1} x_i = 0,$ $x_{n+m} = 1\} \bigcup \{x_1x_2\ldots x_{n+m} | x_1 = 1, \sum_{i=2}^{n} x_i = 0, \sum_{i=n+1}^{n+m} x_i = 1\} \bigcup$ $\{x_1x_2\ldots x_{n+m} | \sum_{i=1}^{n} x_i = 0, \sum_{i=n+1}^{n+m} x_i = 2,\ x_i = 0 \text{ or } 2\}$ *is a connected main effects plan and is capable of estimating all main effects of a* $2^n \cdot 3^m$ *plan.*

Proof of Theorem 5.1 is by verification. For simplicity, we define the following profiles.

$$\begin{aligned}
\delta_{11} &= (100\ldots000),\\
\delta_{1i} &= \{x_1x_2\ldots x_{n+m}|x_1 = x_2 = \ldots = x_n = 0, \textstyle\sum_{i=n+1}^{n+m} x_i = 1,\},\\
\delta_{2i} &= \{x_1x_2\ldots x_{n+m}| \textstyle\sum_{i=1}^{n} x_i = 1, x_{n+1} = x_{n+2} = \ldots = x_{n+m-1} = 0,\\
&\quad x_{n+m} = 1\},\\
\delta_{3i} &= \{x_1x_2\ldots x_{n+m}|x_1 = 1, x_2 = x_3 = \ldots = x_n = 0, \textstyle\sum_{i=n+1}^{n+m} x_i = 1\}\\
\delta_{4i} &= \{x_1x_2\ldots x_{n+m}|x_1 = x_2 = \ldots = x_n = 0, \textstyle\sum_{i=n+1}^{n+m} x_i = 2, x_i = 0 \textit{ or } 2\}.
\end{aligned}$$

By definition, the main effect for the two-level factor A_1 is

$$A_1 = \tfrac{1}{2}(\alpha_1^{A_1} - \alpha_0^{A_1}),$$

Therefore, we can verify that an unbiased estimator for A_1 is:

$$\widehat{A}_1 = 100\ldots001 - 000\ldots001.$$

Following the same logic, we can generalize the estimator for the main effect of any two-level factors. Hence, for $i = 1, 2, \ldots n$, we have

$$\widehat{A}_i = \delta_{2i} - \delta_{1(n+m)}.$$

For three-level factors, we need to consider the linear and quadratic effects. By definition, the linear and quadratic effects for factor A_{n+1} are:

$$\begin{aligned}
A_{(n+1)L} &= (\alpha_2^{A_{n+1}} - \alpha_0^{A_{n+1}})\\
A_{(n+1)Q} &= (\alpha_2^{A_{n+1}} + \alpha_0^{A_{n+1}} - 2 \times \alpha_1^{A_{n+1}}).
\end{aligned}$$

Therefore, we can verify that proper unbiased estimators for the linear and quadratic effects of a three-level factor A_i, for $i = n+1, n+2, \ldots, n+m-1$, are:

$$\begin{aligned}
\widehat{A}_{iL} &= \delta_{4i} + \delta_{3(n+m)} - \delta_{11} - \delta_{1(n+m)},\\
\widehat{A}_{iQ} &= \delta_{4i} + \delta_{11} - \delta_{1i} - \delta_{3i}.
\end{aligned}$$

The estimation of the main effects for factor A_{n+m} is different from other three-level factors. We used the profiles $(100\ldots001)$ and $(000\ldots001)$ in estimating the linear effects for factors A_{n+1} to A_{n+m-1} in order to avoid the dominated profile $(000\ldots000)$. Therefore when come to the estimation of the main effects of this last three-way factor, we need some other factor to take the role of factor A_{n+m} in estimating all other three-way factors. Hence, two proper estimators are:

$$\begin{aligned}
\widehat{A}_{(n+m)L} &= \delta_{4(n+m)} + \delta_{3(n+1)} - \delta_{11} - \delta_{1(n+1)}, \\
\widehat{A}_{(n+m)Q} &= \delta_{4(n+m)} + \delta_{11} - \delta_{1(n+m)} - \delta_{3(n+m)}.
\end{aligned}$$

In summary, the profiles that we used to estimate all main effects come from the following two consecutive choice sets:

$$\begin{aligned}
S_1^* &= \{100\ldots000\} \bigcup \{x_1x_2\ldots x_{n+m} | \textstyle\sum_{i=1}^n x_i = 0, \sum_{i=n+1}^{n+m} x_i = 1\} \\
S_2^* &= \{x_1x_2\ldots x_{n+m} | \textstyle\sum_{i=1}^n x_i = 1, \sum_{i=n+1}^{n+m-1} x_i = 0, x_{n+m} = 1\} \\
&\quad \bigcup \{x_1x_2\ldots x_{n+m} | x_1 = 1, \textstyle\sum_{i=2}^n x_i = 0, \sum_{i=n+1}^{n+m} x_i = 1\} \\
&\quad \bigcup \{x_1x_2\ldots x_{n+m} | \textstyle\sum_{i=1}^n x_i = 1, \sum_{i=n+1}^{n+m} x_i = 2,\ x_i = 0 \textit{ or } 2,\ \}
\end{aligned}$$

Obviously, the above two choice sets are Pareto Optimal choice sets.

5.2 Examples for $2^n \cdot 3^m$ plans

5.2.1 $2^2 \cdot 3^2$ plan

Suppose we have four factors, A, B, C and D. Factors A and B have two levels while factors C and D have three levels. We are only interested in estimating all main effects.

$$\begin{aligned}
\widehat{A} &= 1001 - 0001 \\
\widehat{B} &= 0101 - 0001 \\
\widehat{C}_L &= 0020 + 1001 - 1000 - 0001 \\
\widehat{C}_Q &= 0020 + 1000 - 0010 - 1010 \\
\widehat{D}_L &= 0002 + 1010 - 1000 - 0010 \\
\widehat{D}_Q &= 0002 + 1000 - 0001 - 1001
\end{aligned}$$

The profiles used above come from two consecutive choice sets:

$$\begin{aligned}
S_1^* &= \{1000, 0010, 0001\} \\
S_2^* &= \{1010, 1001, 0101, 0020, 0002\}
\end{aligned}$$

5.2.2 $2^3 \cdot 3^2$ plan

Suppose we have five factors, A, B, C, D and E. Factors A, B and C have two levels while factors D and E have three levels. We are only interested in estimating all main effects.

$$\begin{aligned}
\widehat{A} &= 10001 - 00001 \\
\widehat{B} &= 01001 - 00001 \\
\widehat{C} &= 00101 - 00001 \\
\widehat{D}_L &= 00020 + 10001 - 10000 - 00001 \\
\widehat{D}_Q &= 00020 + 10000 - 00010 - 10010 \\
\widehat{E}_L &= 00002 + 10010 - 10000 - 00010 \\
\widehat{E}_Q &= 00002 + 10000 - 00001 - 10001
\end{aligned}$$

The profiles used above come from two consecutive choice sets:

$$\begin{aligned}
S_1^* &= \{10000, 00010, 00001\} \\
S_2^* &= \{10010, 10001, 01001, 00101, 10001, 00020, 00002\}
\end{aligned}$$

5.2.3 $2^3 \cdot 3^3$ plan

Suppose we have six factors, A, B, C, D, E and F. Factors A, B and C have two levels while factors D, E and F have three levels. We are only interested in estimating all main effects.

$$\begin{aligned}
\widehat{A} &= 100001 - 000001 \\
\widehat{B} &= 010001 - 000001 \\
\widehat{C} &= 001001 - 000001 \\
\widehat{D}_L &= 000200 + 100001 - 100000 - 000001 \\
\widehat{D}_Q &= 000200 + 100000 - 000100 - 100100 \\
\widehat{E}_L &= 000020 + 100001 - 100000 - 000001 \\
\widehat{E}_Q &= 000020 + 100000 - 000010 - 100010 \\
\widehat{F}_L &= 000002 + 100100 - 100000 - 000100 \\
\widehat{F}_Q &= 000002 + 100000 - 000001 - 100001
\end{aligned}$$

The profiles used above come from two consecutive choice sets:

$$\begin{aligned} S_1^* &= \{100000, 000100, 000010, 000001\} \\ S_2^* &= \{100100, 100010, 100001, 010001, 001001, 000200, 000020, 000002\} \end{aligned}$$

5.3 Theorem for a $2^n \cdot 3$ plan

Now we will consider the situation that all factors have two levels except one three-level factor. Our target is to estimate main effects and one two-way interaction.

Theorem 5.2 *The choice sets* $S_1^* = \{100\ldots000, 010\ldots000, 000\ldots001\}$ *and* $S_2^* = \{x_1x_2\ldots x_n | \sum_{i=1}^n x_i = 1, x_{n+1} = 1\} \bigcup (000\ldots002) \bigcup (110\ldots000)$ *is a connected main effects plan and is capable of estimating one two-way interaction* A_1A_2.

Proof of Theorem 5.2 is by verification. Suppose we have a $2^n \cdot 3$ plan and our interest is to estimate all main effects and only two-way interaction A_1A_2. In other words, all other two-way and higher interactions are regarded negligible.

For $i = 1, 2, \ldots (n+1)$, let

$$\begin{aligned} \delta_{11} &= (x_1x_2\ldots x_{n+1} | x_1 = 1, \textstyle\sum_{i=2}^{n+1} x_i = 0), \\ \delta_{12} &= (x_1x_2\ldots x_{n+1} | \textstyle\sum_{\substack{i=1 \\ i\neq 2}}^{n+1} x_i = 0, x_2 = 1), \\ \delta_{1(n+1)} &= (x_1x_2\ldots x_{n+1} | \textstyle\sum_{i=1}^{n} x_i = 0, x_{n+1} = 1), \\ \lambda_{i(n+1)} &= (x_1x_2\ldots x_{n+1} | \textstyle\sum_{\substack{j=1 \\ j\neq i}}^{n} x_j = 1, x_i = x_{n+1} = 1), \\ \lambda_{12} &= (x_1x_2\ldots x_{n+1} | x_1 = x_2 = 1, \textstyle\sum_{i=3}^{n+1} x_i = 0), \\ \lambda_{n+1} &= (x_1x_2\ldots x_{n+1} | \textstyle\sum_{i=1}^{n} x_i = 0, x_{n+1} = 2). \end{aligned}$$

By definition, the main effect for factor A_1 is

$$A_1 = \tfrac{1}{2}(\alpha_1^{A_1} - \alpha_0^{A_1}),$$

Therefore, we can verify that one proper unbiased estimator for A_1 is:

$$\widehat{A}_1 = \tfrac{1}{3}(\lambda_{1(n+1)} + \lambda_{12} - \delta_{1(n+1)} - \delta_{12}).$$

Following the same logic, we can find the estimators for the main effects of all other factors.

$$\begin{aligned} \widehat{A}_2 &= \tfrac{1}{3}(\lambda_{2(n+1)} + \lambda_{12} - \delta_{1(n+1)} - \delta_{11}) \\ \widehat{A}_j &= \lambda_{j(n+1)} - \delta_{1(n+1)}, \forall j = 3, 4, \ldots, n. \end{aligned}$$

The estimation for factor A_{n+1} is a little bit different because this is a three-level factor. We need to consider the linear and quadratic effects. By definition, the linear and quadratic effects for factor A_{n+1} are:

$$\begin{aligned} A_{(n+1)L} &= (\alpha_2^{A_{n+1}} - \alpha_0^{A_{n+1}}) \\ A_{(n+1)Q} &= (\alpha_2^{A_{n+1}} + \alpha_0^{A_{n+1}} - 2 \times \alpha_1^{A_{n+1}}). \end{aligned}$$

Therefore, we can verify that proper unbiased estimators for the linear and quadratic effects of factor A_{n+1} are:

$$\begin{aligned} \widehat{A}_{(n+1)L} &= (\lambda_{n+1} + \lambda_{1(n+1)} - \delta_{11} - \delta_{1(n+1)}) \\ \widehat{A}_{(n+1)Q} &= (\lambda_{n+1} + \delta_{11} - \delta_{1(n+1)} - \lambda_{1(n+1)}). \end{aligned}$$

Similarly, a proper unbiased estimator for A_1A_2 is:

$$\widehat{A_1A_2} = \lambda_{12} + \delta_{1(n+1)} - \lambda_{1(n+1)} - \delta_{12}.$$

In summary, the profiles that we used to estimate all main effects and one two-way interaction A_1A_2 come from the following two consecutive choice sets:

$$\begin{aligned} S_1^* &= \{100\ldots000, 010\ldots000, 000\ldots001\} \\ S_2^* &= \{x_1x_2\ldots x_n | \textstyle\sum_{i=1}^n x_i = 1, x_{n+1} = 1\} \bigcup \{\lambda_{n+1}, \lambda_{12}\} \end{aligned}$$

Obviously, the above two choice sets are Pareto Optimal choice sets.

5.4 Examples for $2^n \cdot 3$ plans

5.4.1 $2^2 \cdot 3$ plan

Suppose we have three factors, A, B and C. Factors A and B have two levels while factor C has three levels. We are only interested in estimating all main effects and one two-way interaction AB.

$$
\begin{aligned}
\widehat{A} &= \tfrac{1}{3}(101 + 110 - 001 - 010) \\
\widehat{B} &= \tfrac{1}{3}(011 + 110 - 001 - 100) \\
\widehat{C}_L &= 002 + 101 - 100 - 001 \\
\widehat{C}_Q &= 002 + 100 - 001 - 101 \\
\widehat{AB} &= 110 + 001 - 101 - 010
\end{aligned}
$$

The profiles we used above come from the following two consecutive choice sets:

$$
\begin{aligned}
S_1^* &= \{100, 010, 001\} \\
S_2^* &= \{101, 011, 110, 002\}
\end{aligned}
$$

5.4.2 $2^3 \cdot 3$ plan

Suppose we have four factors, A, B, C and D. Factors A, B and C have two levels while factor D has three levels. We are only interested in estimating all main effects and one two-way interaction AB.

$$
\begin{aligned}
\widehat{A} &= \tfrac{1}{3}(1001 + 1100 - 0001 - 0100) \\
\widehat{B} &= \tfrac{1}{3}(0101 + 1100 - 0001 - 1000) \\
\widehat{C} &= 0011 - 0001 \\
\widehat{D}_L &= 0002 + 1001 - 1000 - 0001 \\
\widehat{D}_Q &= 0002 + 1000 - 0001 - 1001 \\
\widehat{AB} &= 1100 + 0001 - 1001 - 0100
\end{aligned}
$$

The profiles we used above come from the following two consecutive choice sets:

$$
\begin{aligned}
S_1^* &= \{1000, 0100, 0001\} \\
S_2^* &= \{1001, 0101, 0011, 1100, 0002\}
\end{aligned}
$$

5.4.3 $2^4 \cdot 3$ plan

Suppose we have five factors, A, B, C, D and E. Factors A, B, C and D have two levels while factor E has three levels. We are only interested in estimating all main effects and one two-way interaction AB.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{3}(10001 + 11000 - 00001 - 01000) \\
\widehat{B} &= \tfrac{1}{3}(01001 + 11000 - 00001 - 10000) \\
\widehat{C} &= 00101 - 00001 \\
\widehat{D} &= 00011 - 00001 \\
\widehat{E}_L &= 00002 + 10001 - 10000 - 00001 \\
\widehat{E}_Q &= 00002 + 10000 - 00001 - 10001 \\
\widehat{AB} &= 11000 + 01000 - 10001 - 01000
\end{aligned}$$

The profiles we used above come from the following two consecutive choice sets:

$$\begin{aligned}
S_1^* &= \{10000, 01000, 00001\} \\
S_2^* &= \{10001, 01001, 00101, 00011, 11000, 00002\}
\end{aligned}$$

CHAPTER 6

PO choice sets and g-Designs

6.1 Introduction of g-designs

As discussed in previous chapters, the main idea of applying PO choice sets is that such a design does not include dominating or dominated profile and is capable of providing unbiased estimation while using as few profiles as possible. Other research has been done in finding designs capable of estimating main effects and certain two-way interactions in a 2^n plan.

Hedayat and Pestotan (1992) proposes the concept of g-designs, which follows the approach of Taguchi (1959, 1960) in studying the design construction for estimating the mean, the n main effects and $e \leq \binom{n}{2}$ two-way interactions in a 2^n plan. Taguchi considers a graph aided method to tackle this problem. He identifies each of the n vertices of an undirected graph $g(n, e)$ with a factor and each of the e edges with a corresponding two factor interaction. With each $g(n, e)$ he associates a 2-level fractional factorial design, which is capable of providing an unbiased estimator of the $(1 + n + e)$ parameters including the mean.

Based on Taguchi's result, Hedayat and Pestotan (1992) defined the class of g-designs as the following:

A design of the 2^n factorial will be called a g(n,e)-design if and only if (1) it is capable of providing an unbiased estimator for the parameters specified *by g, namely the mean, the main effects of the n factors, and the two-factor interactions, ..., and, (2) it is saturated, namely it contains precisely n+e+1 level combinations, which is the number of parameters in the model specified by g.*

Moreover, a $g(n, e)-design$ is non-singular because it is a saturated design. In other words, the determinant of its design matrix is nonzero.

Hedayat and Pestotan (1992) also discussed the construction of g-designs, some D-optimality results for the class of g-designs and the bounds of the determinants for the information matrices of $g(n, e)$-designs.

It is very useful to connect $g(n, e)$ graphs with $g(n, e)$-designs such that by specifying all nonisomorphic graphs on n vertices and e edges, we can find the set of all possible models for the 2^n factorial in which the parameters of interest are identified by the labeled graphs. This saves time and space in cataloging, storing and retrieving.

6.2 The concept of g-designs

Based on Hedayat and Pestotan (1992)'s definition of g-design given in section 6.1, we can show that the design given in Chapter 3, Theorem 3.1 is a $g(n, (n-1))$-design.

Theorem 3.1 *The choice sets* $S_1^{**} = \{100\ldots000, 000\ldots001, 010\ldots000\}$, *and* $S_2^{**} = \{x_1x_2\ldots x_n | \sum_{i=1}^{n} x_i = 2, x_1 + x_n \neq 0\}$*is a connected main effects plan and is capable of estimating the two-way interactions,* A_1A_i, $i = 2, 3, \ldots n$.

First, Theorem 3.1 provides a connected plan to unbiasedly estimate all n main effects and $e = (n-1)$ two-way interactions involving factor A_1. Second, there are three profiles in choice set S_1^{**} and $(2n-3)$ profiles in choice set S_2^{**}. Therefore the number of total profiles used in the plan in Theorem 3.1 is $2n$, which is exactly $n + e + 1$. Thus Theorem 3.1 gives a $g(n, (n-1)) - design$ and the design matrix is non-singular.

Although all $g(n, (n-1)) - designs$ use the same number of profiles as the design defined in Theorem 3.1, not all $g(n, (n-1)) - designs$ are PO designs since there is no restriction on including dominated or dominating profiles in a $g(n,e)-design$. Thus, Theorem 3.1 gives a design that is both PO and a g-design.

6.3 Some D-optimal PO designs

In Chapter 2, Theorem 2.1(b) given by Raghavarao and Wiley(2006) describes a design capable of estimating all main effects and one two-way interaction in a 2^n plan.

Theorem 2.1 (b) *The choice sets* S_1^* *and* $S_2^{**} = S_2^* \bigcup \{110\ldots00\}$ *is a connected main effects plan and is capable of estimating the two-way interaction,* A_1A_2.

Based on the definition of $g(n,e)$ design, such a design is a $g(n,1)$ design. The profiles used in this design are: $S_1^* = \{100\ldots000, 000\ldots001\}$ and $S_2^{**} = \{100\ldots001, 010\ldots001, \ldots, 000\ldots011, 110\ldots000\}$. The design matrix is:

$$X = \begin{pmatrix} 1 & 1 & -1 & -1 & \ldots & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 & \ldots & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & \ldots & -1 & 1 & -1 \\ 1 & -1 & 1 & -1 & \ldots & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 & \ldots & -1 & 1 & 1 \\ & \ldots & & & \ldots & & \ldots & \\ 1 & -1 & -1 & -1 & \ldots & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & \ldots & -1 & -1 & 1 \end{pmatrix}$$

And by calculation we can get that
$|det(X)| = (-1)^{n+3} \times 2^2 \times 2^n = (-1)^{n+3} \times 2^{n+2}$. Therefore the absolute value for such a design matrix X is always 2^{n+2}.

Hedayat and Pestotan (1992) defined that a square (-1,1)-matrix T of order $n+1+1$ will be called a $g(n,e)-matrix$ if and only if T is equivalent to a

(-1,1)-matrix of the form

$$[\mathbf{1}|\mathbf{X_1}|\mathbf{X_2}] \quad , \tag{6.1}$$

where
(i)X_1 has dimension $(n + e + 1) \times n$ and whose columns are labeled by the main effects in order.
(ii) X_2 has dimension $(n + e + 1) \times e$ and whose columns are labeled by the interactions.

The form (6.1) will be called the *standard form* for a $g(n, e)$-matrix and the class of all nonsingular $g(n, e)$-matrices in standard from will be denoted by $M(g, n, e)$.

Corollary 4.1 in Hedayat and Pestotan (1992) gives some D-optimality results.

(a) Any g(3,1)-matrix in the class M(g,3,1)is D-optimal. In particular, the design matrix X of the design d(g) defined in (3.2) is D-optimal in M(g,3,1) and |det(X)|=32, (b) The matrix L defined in (4.4) is a g(4,1)-matrix, where J(g)={(34)}, which is D-optimal in M(g,4,1) and |det(L)|=128.

Expressions (3.2) and (4.4) are not shown here. Yet from this corollary we can derive that the determinant or the maximum possible determinant for a $(1, -1)$ design matrix of a D-optimal design is 32 when $n = 3$ and 128 when $n = 4$. As mentioned earlier, the plan given by Theorem 2.1(b) from Raghavarao and Wiley (2006) is a $g(n, 1)$ design. When $n = 3$, the determinant of the design matrix described in Theorem 2.1(b) is $(-1)^6 \times 2^5 = 32$, which means the design is both PO and D-optimal. While $n = 4$, the determinant of the design matrix described in Theorem 2.1(b) is $(-1)^7 \times 2^6 = 64$, which means the design is only PO but not D-optimal.

CHAPTER 7

Estimating certain two-way interactions

When our interest is in estimating all main effects and certain two-way interactions, we do not need to run the full factorial design. In this chapter we present a strategy for generating PO subsets capable of estimating main effects and certain two-way interactions involving three specific factors in a 2^n plan. In past chapters we have considered subsets of two-way interactions inclusive of one specific factor. However, here we consider different subsets of two-way interactions that do not include one specific factor. In particular we consider all two-way interactions involving three specific factors.

7.1 Theorem for 2^n plan

In a 2^n plan, if we only want to estimate all main effects and all two-way interactions involving three specific factors, say A_1, A_2 and A_3, we can use the following theorem to set up the design.

Theorem 7.1 *The choice sets* $S_1^* = \{x_1x_2\ldots x_n | \sum_{i=1,i\neq j}^{n} x_i = 0, x_j = 1, j = 1 \text{ or } 2 \text{ or } 3 \text{ or } n\}$ *and* $S_2^* = \{x_1x_2\ldots x_n | \sum_{i=2}^{n} x_i = 1, x_i = 0 \text{ or } 1, x_1 = 1\} \bigcup (x_1x_2\ldots x_n | \sum_{i=1,i\neq 2 \text{ or } 3}^{n} x_i = 0, x_2 = x_3 = 1)$ *is a connected main effects* *plan and is capable of estimating all two-way interactions, A_1A_2, A_1A_3 and A_2A_3, involving the three factors A_1, A_2 and A_3.*

Proof of Theorem 7.1 is by verification. Suppose we have a 2^n plan and our interest is to estimate all main effects and three two-way interactions A_1A_2, A_1A_3, and A_2A_3 only. In other words, all other two-way and higher interactions are regarded negligible.

Under that assumption, the full model for a s^n plan (equation 2.1)could be simplified as:

$$y_{x_1x_2\ldots x_n} = \mu + \sum_{i=1}^{n} \alpha_{x_i}^{A_i} + \alpha_{x_1x_2}^{A_1A_2} + \alpha_{x_1x_3}^{A_1A_3} + \alpha_{x_2x_3}^{A_2A_3} + e_{x_1x_2\ldots x_n}, \qquad (7.1)$$

where μ is the general mean, $\alpha_{x_i}^{A_i}$ is the effect of attribute A_i at level x_i, $\alpha_{x_1x_2}^{A_1A_2}$ is the effect of attributes A_1 and A_2 at levels x_1 and x_2 respectively, and vice versa. $e_{x_1x_2\ldots x_n}$ is the random error. Without loss of generality, we assume $\sum_{x_i} \alpha_{x_i}^{A_i} = 0$, $\sum_{x_i} \alpha_{x_ix_j}^{A_iA_j} = 0$, and $\sum_{x_j} \alpha_{x_ix_j}^{A_iA_j} = 0$.

For simplicity, we define the following profiles.

$$\begin{aligned}
\eta_{1i} &= \{x_1x_2\ldots x_n | \sum_{\substack{j=1\\ i\neq j}}^{n} x_j = 0, x_i = 1\}, \\
\eta_{2i} &= \{x_1x_2\ldots x_n | \sum_{i=2}^{n} x_i = 1, x_i = 0 \; or \; 1, x_1 = 1\}, \\
\gamma_{23} &= (x_1x_2\ldots x_n | \sum_{\substack{i=1\\ i\neq 2 \text{ or } 3}}^{n} x_i = 1, x_2 = x_3 = 1).
\end{aligned}$$

By definition, the main effect for factor A_1 is

$$A_1 = \tfrac{1}{2}(\alpha_1^{A_1} - \alpha_0^{A_1}),$$

Therefore, we can verify that an unbiased estimator for A_1 is:

$$\widehat{A}_1 = \tfrac{1}{2}(\eta_{22} + \eta_{23} - \eta_{12} - \eta_{13}).$$

Following the same logic, we can find the estimators for the rest main effects as:

$$\begin{aligned}
\widehat{A}_2 &= \tfrac{1}{2}(\eta_{22} + \gamma_{23} - \eta_{11} - \eta_{13}), \\
\widehat{A}_3 &= \tfrac{1}{2}(\eta_{23} + \gamma_{23} - \eta_{11} - \eta_{12}), \\
\widehat{A}_k &= (\eta_{2k} - \eta_{11}), k = 4, 5, \ldots, n.
\end{aligned}$$

Therefore, we can verify that proper unbiased estimators for all three two-way interactions involving factors A_1, A_2 and A_3 are:

$$\begin{aligned}
\widehat{A_1A_2} &= \tfrac{1}{2}(\eta_{22} + \eta_{1n} - \eta_{2n} - \eta_{12}), \\
\widehat{A_1A_3} &= \tfrac{1}{2}(\eta_{23} + \eta_{1n} - \eta_{2n} - \eta_{13}), \\
\widehat{A_2A_3} &= \tfrac{1}{2}(\gamma_{23} + \eta_{11} + \eta_{1n} - \eta_{2n} - \eta_{12} - \eta_{13}).
\end{aligned}$$

In summary, the profiles that we used to estimate all main effects, all two-way interactions involving three specific factors, come from the following two consecutive Pareto Optimal choice sets:

$$\begin{aligned}
S_1^* &= \{x_1x_2\ldots x_n | \textstyle\sum_{i=1,i\neq j}^{n} x_i = 0, x_j = 1, j = 1 \textit{ or } 2 \textit{ or } 3 \textit{ or } n\} \\
&\quad \text{and} \\
S_2^* &= \{x_1x_2\ldots x_n | \textstyle\sum_{i=2}^{n} x_i = 1, x_i = 0 \textit{ or } 1, x_1 = 1\} \\
&\quad \bigcup (x_1x_2\ldots x_n | \textstyle\sum_{i=1,i\neq 2 \textit{ or } 3}^{n} x_i = 0, x_2 = x_3 = 1).
\end{aligned}$$

7.2 Examples

7.2.1 2^4 plan

Suppose we have four factors, A, B, C and D, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving the first three factors.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(1100 + 1010 - 0100 - 0010) \\
\widehat{B} &= \tfrac{1}{2}(1100 + 0110 - 1000 - 0010) \\
\widehat{C} &= \tfrac{1}{2}(1010 + 0110 - 1000 - 0100) \\
\widehat{D} &= 1001 - 1000 \\
\widehat{AB} &= \tfrac{1}{2}(1100 + 0001 - 1001 - 0100) \\
\widehat{AC} &= \tfrac{1}{2}(1010 + 0001 - 1001 - 0010) \\
\widehat{BC} &= \tfrac{1}{2}(0110 + 1000 + 0001 - 1001 - 0100 - 0010).
\end{aligned}$$

So the choice sets need to estimate all main effects, two-way interactions AB, AC and BC are:

$$\begin{aligned}
S_1^* &= \{1000, 0100, 0010, 0001\} = S_1 \\
S_2^* &= \{1100, 1010, 1001, 0110\}.
\end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects and two-way interactions, we need at least 14 profiles. Using the above method we only need 8 profiles and both choice sets used are Pareto Optimal choice sets.

7.2.2 2^5 plan

Suppose we have five factors, A, B, C, D, and E, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving the first three factors.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(11000 + 10100 - 01000 - 00100) \\
\widehat{B} &= \tfrac{1}{2}(11000 + 01100 - 10000 - 00100) \\
\widehat{C} &= \tfrac{1}{2}(10100 + 01100 - 10000 - 01000) \\
\widehat{D} &= 10010 - 10000 \\
\widehat{E} &= 10001 - 10000 \\
\widehat{AB} &= \tfrac{1}{2}(11000 + 00001 - 10001 - 01000) \\
\widehat{AC} &= \tfrac{1}{2}(10100 + 00001 - 10001 - 00100) \\
\widehat{BC} &= \tfrac{1}{2}(01100 + 10000 + 00001 - 10001 - 01000 - 00100).
\end{aligned}$$

So the choice sets need to estimate all main effects, two-way interactions AB, AC and BC are:

$$\begin{aligned}
S_1^* &= \{10000, 01000, 00100, 00001\} \\
S_2^* &= \{11000, 10100, 10010, 10001, 01100\}.
\end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects and two-way interactions, we need at least 25 profiles. Using the above method we only need 9 profiles and both choice sets used are Pareto Optimal choice sets.

7.2.3 2^6 plan

Suppose we have six factors, A, B, C, D, E, and F, each with two levels. We are only interested in estimating all main effects and all two-way interactions involving the first three factors.

$$\begin{aligned}
\widehat{A} &= \tfrac{1}{2}(110000 + 101000 - 010000 - 001000) \\
\widehat{B} &= \tfrac{1}{2}(110000 + 011000 - 100000 - 001000) \\
\widehat{C} &= \tfrac{1}{2}(101000 + 011000 - 100000 - 010000) \\
\widehat{D} &= 100100 - 100000 \\
\widehat{E} &= 100010 - 100000 \\
\widehat{F} &= 100001 - 100000 \\
\widehat{AB} &= \tfrac{1}{2}(110000 + 000001 - 100001 - 010000) \\
\widehat{AC} &= \tfrac{1}{2}(101000 + 000001 - 100001 - 001000) \\
\widehat{BC} &= \tfrac{1}{2}(011000 + 100000 + 000001 - 100001 - 010000 - 001000).
\end{aligned}$$

So the choice sets need to estimate all main effects, two-way interactions AB, AC and BC are:

$$\begin{aligned}
S_1^* &= \{100000, 010000, 001000, 000001\} \\
S_2^* &= \{110000, 101000, 100100, 100010, 100001, 011000\}.
\end{aligned}$$

According to Raghavarao and Wiley (2006), in order to estimate all main effects and two-way interactions, we need at least 26 profiles. Using the above method we only need 10 profiles and both choice sets used are Pareto Optimal choice sets.

7.3 Relation with g-designs

According to the definition of g-design (Hedayat and Pestotan, 1992), as mentioned in Chapter 6, Theorem 7.1 gives a $g(n, 3)$ design.

First, Theorem 7.1 provides a connected plan to unbiasedly estimate all n main effects and $e = 3$ two-way interactions involving factors A_1, A_2 and A_3. Second, there are four profiles in choice set S_1^* and n profiles in choice set S_2^*. Therefore the number of total profiles used in Theorem 7.1 is $n + 4$, which is exactly $n + e + 1$. Thus, Theorem 7.1 gives a $g(n, 3)$ design and the design matrix is non-singular.

CHAPTER 8

Conclusion

In a valuation study, some profiles are better than others, or dominating, while some profiles are worse, or dominated. If a choice set has a dominating profile, the respondent's choice is trivially made. Similarly, if a choice set has a dominated profile, it will never be selected. Including a dominating or dominated profile in a design does not provide much information to experimenters. Therefore, it is preferable to include choice sets with no dominating or dominated profiles. Such choice sets are known as Pareto Optimal (PO) choice sets. Value studies utilizing PO choice sets enable us to infer value from the trade-offs that respondents are, or are not, willing to make. However, if the number of attributes or attribute levels becomes large, the profiles in a single choice set become too numerous resulting in information overload. In this situation, respondents cannot make precise decisions. Therefore, smaller choice set sizes are necessary in such experiments. In reality, due to limits on costs and time, it might not be practical to include all the profiles in choice sets. Also, it is possible that experimenters are only interested in values related to certain factors instead of all attributes. The motivation of this dissertation is to find smaller designs based on PO choice sets which are capable of estimating all main effects and certain subsets of two- and three-way interactions.

Raghavarao and Wiley (2006) proposed a strategy for generating PO choice sets and constructing hierarchical PO subsets to collect data for estimating main effects, two-way interactions and three-way interactions. In their research, they constructed designs capable of estimating all main effects and one two-way interaction in a 2^n and 3^n plan respectively. They also developed strategies for breaking choice sets into smaller subsets to reduce the number of profiles and choice sets when only one two-way interaction is of interest.

Motivated by Raghavarao and Wiley (2006), we proposed a design which is capable of estimating main effects and two-way interactions inclusive of one factor in a 2^n plan. For example, if we need to estimate all main and two-way interactions in a 2^5 plan, we will need at least 25 profiles. However, if we only want to estimate main effects and two-way interactions inclusive of one factor, then following our approach, only ten profiles will be needed. We also proposed a design which is capable of estimating main effects and two-way interactions inclusive of one factor in a 3^n plan. For example, in a 3^5 plan, if we want to estimate all main effects and all two-way interactions, we need at least 90 profiles in three choice sets. However, if we only want to estimate main effects and two-way interactions inclusive of one factor, then following our approach, only 27 profiles will be needed. Furthermore, we presented a design that is capable of estimating main effects, two-way interactions, and one three-way interaction inclusive of one factor for 2^n plans. As shown in Section 4.2, for a 2^6 plan, in order to estimate all main effects, two- and three-way interactions. we need at least 56 profiles. However, if we only want to estimate main effects, two-way interactions, and one three-way interaction inclusive of one factor, by following our method, only 26 profiles will be needed.

Next, we considered plans in which not all attributes have the same number of levels, or mixed-level design. We gave connected main effect designs for $2^n \cdot 3^m$ plans and $2^n \cdot 3$ plans respectively. We showed by examples that we can estimate all main effects by using 12 profiles in a $2^3 \cdot 3^3$ plan. For a $2^4 \cdot 3$ plan, we can estimate all main effects along with one two-way interaction involving only two-level factors using nine profiles.

Hedayat and Pestotan (1992) proposed the concept of g-designs, which follows the approach of Taguchi (1959, 1960) in studying the design construction for estimating the mean, the n main effects and $e \leq \binom{n}{2}$ two-way interactions in a 2^n plan. They also discussed the construction of g-designs, some D-optimality results for the class of g-designs and the bounds of the determinants for the information matrices of $g(n,e)$-designs. Based on their result, we compared the design we provide in Section 3.1 with g-designs and showed certain PO designs are D-optimal.

Finally we presented a design which is capable of estimating main effects and two-way interactions among three specific factors for a 2^n plan. This design is also a g-design. For example, for a 2^6 plan, in order to estimate all main effects and two-way interactions, we need at least 14 profiles. However, if we only want to estimate main effects and two-way interactions among three factors, using our method we only need 8 profiles and both choice sets used are Pareto Optimal choice sets.

There are a few possible extensions of this work for the future. Raghavarao and Zhang (2002) showed that in a 2^n plan, the design based on any two PO subsets S_l and $S_l^{'}$ $(0 < l < l^{'} < n)$ is a connected main effects plan. They further proposed several optimal main effects designs with PO choice sets for a 2^n plan, using Information Per Profile (IPP) θ as the optimality criterion. Future research is needed to examine the optimality of the designs discussed in Chapter 3 using one or more optimality criteria. Raghavarao and Zhang (2002) also showed that certain Balanced Incomplete Block Designs (BIBD) have the same IPP θ as PO subsets, but with smaller choice set sizes. Therefore under specific conditions, BIBD could be used to reduce choice set sizes. A similar approach was discussed in Damaraju and Raghavarao (2002). They proposed a method to estimate the main effects and two-way interactions inclusive of only one factor by using a balanced incomplete block design (BIBD) and its complement (foldover) when the BIBD is of Family (A). Future research is needed to compare the advantages and disadvantages of this method with our designs.

REFERENCES

[1] Adamowicz, W., Boxall, P., Williams, M., Louviere, J., (1988): Stated Preference Approaches for Measuring Passive Use Values: Choice Experiments and Contingent Valuation, *American Journal of Agricultural Economics* 80(1), 64-75.

[2] Batsell, R.R., Louviere, J.J., (1991): Experimental Analysis of Choice, *Marketing Letters* 2(3), 199-214.

[3] Bradley, M. A., Gunn, H. F., (1990): Stated Preference Analysis of Values of Travel Time in the Netheriands, *Transportation Research Record* 1285, 78-89.

[4] Bullock, C.H., Elston, D.A., Chalmers, N.A., (1998): An Application of Economic Choice Experiments to a Traditional Land Use - Red Deer Hunting and Landscape Change the Scottish Highlands, *Journal of Environmental Management* 52, 335-351.

[5] Carter, R., Dubelaar, C., Wiley, J.B., (2000): Applying Choice Based Conjoint Measurement to Forecast Demand for A New Restaurant Category, *Journal of Food Products Marketing* 6(3), 63-78.

[6] Damaraju, L., Raghavarao, D., (2002): Fractional Plans to Estimate Main Effects and Two Factor Interactions Inclusive of a Specific Factor, *Biometrical Journal* 44(6), 780-785.

[7] Fowkes, T., Wardman, M., (1988): The Design of Stated Preference Travel Choice Experiments, with Special Reference to Inter-Personal Taste Variations, *Journal of Transport Economics and Policy* 13(1), 27-44.

[8] Green, P.E., (1984): Hybrid Models for Conjoint Analysis: An Expository Review, *Journal of Marketing Research* 21 (2), 155-169.

[9] Green, P.E., Srinivasan, V., (1990): Conjoint Analysis in Marketing: New Developements with Implications for Research and Practice, *Journal of Marketing* 54(4), 3-19.

[10] Green, P.E., Wind, Y., (1975): New Ways to Measure Consumers Judgements, *Harvard Business Review* 53, 107-117.

[11] Haider, W., Ewing, G.O., (1990): A Model of Toursit Choices of Hypothetical Caribbean Destination, *Leisure Science* 12(1), 33-47.

[12] Hedayat, A.S., Pesotan, H., (1992): Two-Level Factorial Designs for Main Effects and Selected Two-Factor Interactions, *Statistica Sinica* 2(1992), 453-464.

[13] Hensher, D.A., Barnard, P., Milthorpe, F., Smith, N., (1989): Urban Tollways and the Valuation of Travel Time Savings, *The Economic Record* 66(193), 146-156.

[14] Huber, J., Hansen, D., (1986): Testing the Impact of Dimensional Complexity and Affective Differences of Paired Concepts in Adaptive Conjoint Analysis, *Advances in Consumer Research* 14, in Melanie Wallendorf and Paul Anderson (Eds.), Provo, UT, Associations of Consumer Research, 159-163.

[15] Huber, J., Wittink, D. R., Johnson, R. M., Miller, R., (1992): Learning Effects in Preference Tasks: Choice-Based Versus Stan-

dard Conjoint, *Sawtooth Software Conference Proceedings*, 275-282.

[16] Huber, J., Zwerina, K., (1996): The Importance of Utility Balance in Efficient Choice Designs, , *Journal of Marketing Research* 33(3), 307-317.

[17] Johnson, R.M., Olberts, K.A., (1991): Using Conjoint Analysis in Pricing Studies: Is One Price Variable Enough? *American Marketing Association Advanced Research Technique Forum Conference Proceedings*, 164-173.

[18] Kamakura, W.A., Srivastava, R.K., (1984): Predicting Choice Shares Under Conditions of Brand Interdependence, *Journal of Marketing Reseach* 21, 420-434.

[19] Kreiger, A.B., Green, P.E., (1991): Designing Pareto Optimal Stimuli for Multi-Attribute Choice Experiments, *Marketing Letters* 2, 337-348.

[20] Louviere., J.J., Woodworth, G., (1983): Design and Analysis of Simulated Consumer Choice or Allocation Experiments: A Method Based on Aggregate Data, *Journal of Marketinig Research* 20, 350-367.

[21] Luce, R., Tukey, J.W., (1964): Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement, *Journal of Mathematical Psychology* 1, 1-27.

[22] Moore, W.L., Louviere, J.J., Verma, R., (1999): Using Conjoint Analysis to Help Design Product Platforms, *Journal of Product Innovation Management* 16(1), 27-39.

[23] Oliphant, K., Eagle, T.G., Louviere, J. J., Anderson, D. (1992):

Cross-Task Comparison of Ratings-Based and Choice-Based Conjoint, *Sawtooth Software Conference Proceedings*, 383-394.

[24] Oppewal, H., Timmermans, H., Louviere, J.J., (1997): Modelling the Effects of Shopping Centre Size and Store Variety on Consumer Choice Behaviour, *Environment and Planning A* 29, 1073-1090.

[25] Propper, C., (1995): The Disutility of Time Spent on the United Kingdom's National Health Service Waiting Lists, *Journal of Human Resources* 30(4), 677-700.

[26] Raghavarao, D., (1971): *Constructions and Combinatorial Problems in Design of Experiments*, Wiley, New York.

[27] Raghavarao, D., Wiley, J.B., (1998): Estimating Main Effects with Pareto Opitmal Subsets, *Australia & New Zealand Journal of Statistics* 40(4), 425-432.

[28] Raghavarao, D., Wiley, J.B., (2006): Design Strategies for Sequential Choice Experiments Involving Economic Alternatives, *Journal of Statistical Planning and Inference* 136, 3287-3306.

[29] Raghavaro, D., Wiley, J.B., Chitturi, P., (2010): *Choice-Based Conjoint Analysis: Models and Designs*, Chapman and Hall/CRC Press.

[30] Raghavarao, D., Zhang, D., (2002): 2^n Behavioral Experiments Using Pareto Optimal Choice Sets, *Statistica Sinica* 12, 1085-1092.

[31] Ryan, M., Farrar, S., (2000): Using Conjoint Analysis to Elicit Preferences for Health Care, *British Medical Journal* 320(7248), 1530-1533.

[32] Taguchi, G., (1959): Linear Graphs for Orthogonal Arrays and Their Applications to Experimental Design, with the Aid of Vari-

ous Techniques, *Reports of Statistical Application Research JUSE* 6, 1-43.

[33] Taguchi, G., (1960): Table of Orthogonal Arrays and Linear Graphs, *Reports of Statistical Application Research JUSE* 7, 1-52.

[34] Toubia, O., Hauser, J.R., Simerster, D., (2004): Polyhedral Methods for Adaptive Choice-Based Conjoint Analysis, *Journal of Marketing Research*41(1), 116-131.

[35] Waerden, P. Van Der, Oppewal, H., Timmermans, H., (1993): Adaptive Choice Behaviour of Motorists in Congested Shopping Centre Parking Lots, *Transportation* 20, 395-408.

[36] Wiley, J.B., (1978): Selecting Pareto Optimal Subsets from Multi-Attribute Alternatives, *Advances in Consumer Research* 5, 171-174.

[37] Wright, P., (1975): Consumer Choice Strategies: Simplifying vs. Optimizing, *Journal of Marketing Research* 12, 60-67.