OPTIMAL REDUCED SIZE CHOICE SETS WITH OVERLAPPING ATTRIBUTES

A Dissertation
Submitted to
the Temple University Graduate Board

---

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

---

by
Ke Huang
July, 2015

Examining Committee Members:

Pallavi Chitturi, Advisory Chair, Statistics
Robert T. Krafty, Statistics
Yuexiao Dong, Statistics
Susan Mudambi, External Reader, Marketing & Supply Chain Management

ABSTRACT

OPTIMAL REDUCED SIZE CHOICE SETS WITH OVERLAPPING ATTRIBUTES

Ke Huang

DOCTOR OF PHILOSOPHY

Temple University, July, 2015

Professor Pallavi Chitturi, Chair

Discrete choice 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. Respondents are presented sets of profiles based on attributes specified at certain levels and asked to select the profile they consider best. When the number of attributes or attribute levels becomes large, the profiles in a single choice set may be too numerous for respondents to make precise decisions. One strategy for reducing the size of choice sets is the sub-setting of attributes. However, the optimality of these reduced size choice sets has not been examined in the literature. We examine the optimality of reduced size choice sets for 2^n experiments using information per profile (IPP) as the optimality criteria. We propose a new approach for calculating the IPP of designs obtained by dividing attributes into two or more subsets with one, two, and in general, r overlapping attributes, and compare the IPP

of the reduced size designs with the original full designs. Next we examine the IPP of choice designs based on 3^n factorial experiments. We calculate the IPP of reduced size designs obtained by sub-setting attributes in 3^n plans and compare them to the original full designs.

ACKNOWLEDGEMENTS

This research would not have been possible without the support of many people and it is my pleasure to express my sincere gratitude.

My deepest gratitude is to my advisor, Dr. Pallavi Chitturi, for her motivation, keen guidance and continuous encouragement through my research in statistics during my years at Temple University. Without her dedication and enthusiasm much of the research discussed in the dissertation would not have been possible. I have been fortunate to have an advisor who gave me such strong support. I hope that one day I would become as successful a female professional as Dr. Chitturi is.

It has been an honor and a pleasure to learn and receive guidance from Dr. Damaraju Raghavarao. His brilliant ideas and broad knowledge made the research exciting, and helped me overcome some difficulties in my early part of research. Unfortunately, Dr. Raghavarao passed away on February 06, 2013. He is deeply missed by each and every one in the Department. 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 Dr. Robert Krafty and Dr. Yuexiao Dong for serving on my dissertation committee. Their insightful comments and constructive suggestions at different stages of my research were thought-provoking and helped me focus my ideas.

Finally, I want to thank my parents, Mingjin Huang and Xiuqin Zhao, for their endless love and support. Their strong belief in hard work helps me go forward. They often say that I make them proud. However, I could not be more proud of them.

Above all, I must thank my husband, Shuzhi Dong, for his love, constant support in the difficult time during research. He has been very patient with me and has encouraged me when I have been frustrated. This thesis is dedicated to him, my parents, and all that they have done for me and put up with during the past couple of years.

TABLE OF CONTENTS

ABSTRACT **iv**

ACKNOWLEDGEMENT **vi**

LIST OF TABLES **x**

1 INTRODUCTION **1**
1.1 Reducing Choice Set Sizes 4

2 LITERATURE REVIEW **7**
2.1 Conjoint analysis 8
2.2 Choice-based conjoint analysis 10
2.3 Pareto Optimal choice set 13
2.4 Models 15
2.4.1 Fit of the model 16
2.5 Sequential CBC experiments 17
2.5.1 Sequential testing strategies 18
2.6 Connected main effects PO designs and IPP 20
2.7 Smaller Choice Set Sizes 22

3 SUB-SETTING ATTRIBUTES INTO OVERLAPPING SETS FOR A 2^n PLAN **26**
3.1 Sub-setting attributes into overlapping sets with PO sets S_l and S_{l+1} 27
3.1.1 2^n plan subset into 2 sets 27
3.1.2 2^n plan subset into 2 sets for general cases 39
3.1.3 2^n plan subset into m sets 44
3.2 Comparison of IPP for 2^n plans 50
3.3 Sub-setting attributes into overlapping sets with PO sets S_l and S_{j-l} 52

4 SUB-SETTING ATTRIBUTES IN A 3^n PLAN **56**
4.1 Introduction 56
4.2 3^n plan subset into m sets 62

4.3 Comparison of IPP for 3^n plans 65

5 CONCLUSION **68**

REFERENCES **70**

LIST OF TABLES

3.1 Comparison of IPP 51

3.2 The IPP of overlapping attributes 51

4.1 The IPP of 3^n plans based on S_l and S_{l+1} 66

4.2 Comparisons of IPP for 3^n plans 66

CHAPTER 1

INTRODUCTION

Conjoint analysis and discrete choice experiments are widely used methodologies for measuring and analyzing the preferences or choices of respondents. The contribution by Luce and Tukey (1964) is viewed as the origin of conjoint analysis. Since its introduction, conjoint analysis has been researched and used in many different applications. (For reviews see Green and Srinivasan, 1990; Wittink and Cattin, 1989; Wittink et al., 1994). In a conjoint task, respondents typically sort, rank or rate a set of profiles. These profiles are experimentally designed and are described by multiple factors (attributes) and their levels. The results from conjoint analysis provide insights into how respondents evaluate certain attributes of interest.

Discrete choice experiments are a method related to conjoint analysis and are also called choice based conjoint (CBC). This method involves the design of profiles on the basis of attributes specified at certain levels. However, instead of ranking or rating all profiles, as is usually done in a conjoint analysis task, respondents are asked to repeatedly choose one alternative from different sets of profiles presented to them. Probabilistic choice models such as multinomial logit or probit models are applied to the choice data arising from such experiments. An early article describing the advantages of this approach for conjoint analysis was by Louviere and Woodworth (1983).

The advantage of discrete choice experiments over traditional conjoint analysis is that data collection involves simulated decisions (or hypothetical choices) thus providing a more realistic and simpler task for respondents than rankings or ratings. In recent years this method has become increasing popular as a way to more directly study choice (Batsell and Louviere, 1991). Consider a rail transportation example. We can consider four attributes that describe train service: fares, frequency, routes and security, and two levels: the current level (0) and an improved level (1) for each attribute i.e. current fares and a 10% reduction in fares, current frequency of service or higher frequency, current train routes or an increased number of routes, and current level of security or improved security.

If asked to choose the level of each attribute separately, respondents will naturally select the improved level for each attribute. However, due to budget constraints, it may not be feasible to expend resources on all four attributes. Rail authorities need to know which of these attributes are most important for the quality of life of citizens of the area. They can then make informed decisions about where resources should be concentrated. In such an environment, respondents are required to make tradeoffs and these tradeoffs cannot be determined by asking about each attribute separately.

For this purpose, we can create 'transportation profiles' of the characteristics of transportation and ask the respondent to make a choice. If a profile has all four attributes at the improved (higher) level (1111), it will 'dominate' other profiles that have one or more attributes at the current (lower) level and respondents will always choose the dominating profile. If we present a choice set with the profiles $\{1111, 0111, 1011, 1101\}$, the first profile dominates the others and the respondent's choice is trivially made. Similarly, if we present a choice set with the profiles $\{0011, 0111, 1011, 1101\}$, the first profile is dominated and will not be selected. For this reason, choice sets should have no dominating or dominated profiles. Such choice sets are called Pareto Optimal (PO) choice sets. Provided there are no such dominating or dominated profiles, different respondents will select different profiles and we can gain an insight into the relative importance of these factors.

Raghavarao and Wiley (1998) generalized non-dominating (PO) subsets. For an s^n experiment with n attributes each at s levels denoted by $0, 1, \ldots, s-1$, assuming $(x_1, x_2, \ldots, x_n)$ is a profile of the n attributes, 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. For example, in the rail transportation example described above, $S_2 = \{1100, 1010, 1001, 0110, 0101, 0011\}$ is a PO choice set. Since the levels in each profile of the choice set sum to two, there are no dominating or dominated profiles in the choice set S_2. They also note that any subset of a PO subset is itself a PO subset. To compare designs, Information Per Profile (IPP) can be used as an optimality criteria.

Raghavarao and Zhang (2002) provide optimal designs for 2^n experiments with n attributes each at two levels when $n = m^2$ or $n = m(m + 2)$ using IPP as the optimality criteria.

1.1 Reducing Choice Set Sizes

In choice experiments the investigator organizes profiles into systematically constructed choice sets, where the profiles are designed on the basis of all attributes. When the number of attributes or attribute levels becomes large, the profiles in a single choice set may be too numerous for respondents to make precise decisions. There is a limit to how much information respondents can process without becoming confused or overloaded. For example, in a 3^5 factorial plan with five factors at three levels each, three choice sets (S_2, S_3, and S_4) and a total of 90 profiles are needed to estimate all main effects and all two-way interactions. Furthermore, if the investigator is only interested in certain factors and interactions, it may be unnecessary to go over all profiles.

Several strategies have been developed for reducing the number of profiles and choice sets. Smaller subsets may be generated by sub-setting larger subsets. Smaller choice sets and a reduced number of alternatives may be achieved by sub-setting based on overlapping attributes and/or levels. They may also be generated using fractional designs and cyclic designs. Designs have been developed in which only a subset of two-way interactions are estimated for 2^n and 3^n experiments (Chen and Chitturi, 2012). Note that these procedures may be used to create choice sets of the same size across the choice experiment, or of varying size. Experiments with same choice sets sizes are common, but studies with varying choice sets sizes may be found (Koelemeijer and Oppewal, 1999; Oppewal et al., 1994, 2000).

In sub-setting attributes into overlapping sets, we divide the total number of attributes into subsets T_1, T_2, ..., T_m such that T_i and T_{i+1} are non-disjoint for $i = 1, 2, \ldots, (m-1)$. That is, T_i and T_{i+1} are linked by the overlap of at least one attribute. For example, in a 2^7 experiment, the seven attributes $A_1, A_2, \ldots, A_7$ can be divided into two overlapping subsets of four attributes each:

$$T_1 = A_1, A_2, A_3, A_4, \quad (1.1)$$

$$T_2 = A_4, A_5, A_6, A_7. \quad (1.2)$$

We form separate main-effects plans based on the attributes in each subset T_i. All main effect contrasts are estimable and since T_i and T_{i+1} are non-disjoint, we can test the difference of main effects of any two attributes. A limitation of this approach is that we can only estimate two-factor interactions if both attributes are in the same set T_i. Therefore, not all two-factor interactions are estimable.

In a 2^7 experiment, three choice sets S_1, S_2, and S_3 with a total of 63 profiles are needed to estimate all main effects and all two-way interactions. If we divide the seven attributes into two overlapping subsets T_1 and T_2 of four attributes each, we can use the 2^4 design twice, once with attributes A_1, A_2, A_3 and A_4, and once with attributes A_4, A_5, A_6 and A_7. In a 2^4 experiment, three choice sets S_1, S_2, and S_3, with a total of 14 profiles are needed to estimate all main effects and all two-way interactions. Thus the combined design has $2 \times 14 = 28$ profiles instead of 63. All main effects are estimable, but only those two-factor interactions are estimable for which both attributes are in the same set T_i. For example, the interaction A_1A_2 is estimable, but A_1A_7 is non-estimable. In this example, the subsets T_1 and T_2 are linked by a single attribute A_4, but subsets may be linked by the overlap of more than one attribute. Schwabe et al. (2003) and Grossmann et al. (2002) also consider the problem of sub-setting by attributes.

IPP, as an optimality criterion, has been defined for choice designs based on 2^n experiments. The IPP of designs obtained by sub-setting attributes into overlapping subsets has not been examined in the literature. In chapter 3 we develop a new approach for obtaining the IPP for smaller designs obtained by sub-setting attributes based on PO sets S_l and S_{l+1} and compare the IPP of full designs with the reduced size designs. We also discuss the calculation of IPP for sub-setting designs based on PO sets S_l and S_{n-l}. We define the IPP for choice designs based on 3^n experiments in chapter 4. Finally, we obtain the IPP of smaller designs obtained by sub-setting a 3^n plan, and compare the IPP of the full designs and smaller designs.

CHAPTER 2

LITERATURE REVIEW

As an approach to multiattribute utility measurement, conjoint analysis evolved from the theory of conjoint measurement (Luce and Tukey 1964) in mathematical psychology. Based on this theory, some psychometricians (Carroll, 1969; Kruskal, 1965; Young, 1969) developed a variety of nonmetric models for computing part-worths from respondents' preference orderings across mutliattributed stimuli, such as descriptions of products or services. In the early 1970s conjoint analysis has received considerable attention as a major set of techniques for measuring buyer's trade-offs among multiattributed products and services (Green and Rao, 1971; Johnson, 1974; Srinivasan and Shocker, 1973). With the availability of computer software for implementing the methodology conjoint analysis was widely applied in many fields of social sciences and applied sciences in the 1980s. Conjoint analysis is by far one of the most popular statistical techniques for analyzing consumer trade-offs.

2.1 Conjoint analysis

The basic idea of conjoint analysis is to handle situations in which decision makers have to deal with options that simultaneously vary across two or more attributes. In a conjoint study, decision makers solve the problem of how to trade off the possibility that option A is better than option B on attribute x while B is better than A on attribute y, and various extensions of these conflicts. Through analyzing a series of trade-offs made by participants, researchers can determine the relative importance of component attributes. Traditionally in such experiments, a product or service is described in terms of attributes. Each attribute consists of a number of levels. Participants are shown sets of combinations of levels from all attributes and asked to choose one they like best. These combinations are also called profiles. They are designed similar enough to be close substitutes and dissimilar enough so participants can clearly choose a preference.

For example, a cell phone service may have attributes such as plan, price, minutes, data and so on. Assume the attribute plan has three levels: family plan, individual plan, and mobile share plan. When a consumer chooses a service, he has to make a decision among price levels, minutes levels, and data levels such as individual plan $69.99/mo with 3 GB data and 450 minutes talk, family plan $79.99/mo with 300 MB data and 550 minutes talk for two lines, or mobile share with unlimited talk and text plan $40/mo with 1 GB data plus $45 cost for each smart phone. Based on consumers' behavior researchers can value the elements composing the product or service, and use these valuations to create models estimating market share, revenue, and even design new products.

Conjoint analysis first took the form of Full Profile studies and was used to investigate a small set of attributes (typically 4 or 5). Respondents were asked to rank or rate profiles. The utilities for the levels can be estimated by relatively simple dummy variable regression analysis. Such designs limited the effectiveness and flexibility of conjoint analysis to some extent. In order to avoid overloading information on respondents the number of applicable attributes was heavily restricted. Also ranking or rating large number of profiles was less feasible in real life situations. Considering such limitations, some approaches of data collection were introduced to solve the problem of a large number of attributes in the '80s, such as hybrid techniques (Green et al., 1981), adaptive conjoint analysis (Johnson, 1987), and self-explicated preference-data collection (Srinivasan, 1988). Including full profile techniques, these approaches became the main data collection procedures used for conjoint analysis. Extending the application of multinomial logit model to conjoint analysis Louviere and Woodworth (1983) proposed experimental designs in which only a choice task was used. This approach improved the implementation of conjoint analysis in real life and became the basis of choice-based conjoint and discrete choice analysis.

2.2 Choice-based conjoint analysis

The choice-based conjoint (CBC) model was an important technical development in conjoint analysis during the '80s. It is also referred to as discrete choice experiments. Instead of ranking various product or service profiles presented one at a time on a likelihood-of purchase scale, respondents typically see profile descriptions of two or more explicit competitors which vary on one or more attributes, and are asked to pick their most preferred profile from the set. Alternatively, respondents may allocate 100 points across the set of profiles reflecting one's relative strength of preference.

After the econometrician McFadden (1974) suggested analyzing choice data using the multinomial logit model, this method was soon recognized and adopted by a number of marketing researchers. Some of them extended and developed its applications (Gensch and Recker, 1979; Batsell and Lodish, 1981; Mahajan et al., 1982). Louviere and Woodworth (1983) were among the first to discuss experimental designs that provided a theoretical foundation for choice-based conjoint analysis. They emphasized that sets of concepts shown to respondents should be constructed using experimental designs, choice sets should contain an option such as "none of these", and employing multinomial logit model estimation of parameters should be considered in aggregate. Since the efficiency of experimental designs for multinomial logit was not as straightforward as that for traditional linear models and their designs, design efficiency became a topic of research (Kuhfeld et al., 1994; Bunch et al., 1996; Huber and

Zwerina, 1996). General rules for designing CBC experiments were proposed by Carter, Dubelaar and Wiley (2001) .

- Word 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 number of: (1)choice sets; (2)choices; (3)attributes; and (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.

In the case of valuation studies, CBC experiments typically involve economic choices in which the attributes may consist of benefits and/or costs. Researchers are interested in investigating respondents' trade-off behaviors, i.e., giving up one benefit, or incurring a cost, to secure another benefit. Benefits and costs may be defined in economic terms as follows: a benefit is something for which preference monotonically increases with increasing attribute level; a cost is something for which preference increases with decreasing attribute level. 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. For example, an investigator may want to study the relative importance consumers put on attributes like screen size, screen format, brand and price in selecting a television.

The advantage of choice-based conjoint analysis is not just in simplifying the respondent's task in the data collection procedure, but also providing a way to deal with interactions among attributes. By pooling or borrowing information across respondents, choice-based conjoint models offer the feasibility to quantify interactions, while most conjoint methods are based on the "main effects only" assumption that ignores the existence of attribute interactions.

The number of attributes is a main limitation for conjoint analysis. Green and Srinivasan (1990) recommended six to ten as the maximum number of attributes for typical data collection procedures in traditional conjoint analysis. The problem is not fully resolved under CBC analysis. When the number of attributes and levels increases, the number of potential profiles increases exponentially. However the unique ability to deal with interactions made CBC analysis increasingly popular and widely employed. In recent years, this model has emerged as a major statistical technique used in many social sciences and applied sciences including marketing (Kamakura and Srivastava, 1984; Johnson and Olberts, 1991), tourism (Haider and Ewing, 1990), geography (Oppewal et al., 1997; Waerden et al., 1993), health (Propper, 1995) and environmental science (Adamowicz et al., 1998).

2.3 Pareto Optimal choice set

Pareto optimality (or Pareto efficiency) states an allocation of resources in which it is impossible to make any one individual better off without making at least one individual worse off. This concept was pioneered by Vilfredo Pareto (an Italian economist) in his studies of economic efficiency and income distribution. If a change of allocation makes at least one individual better off without making any other individual worse off, such a case is called a Pareto improvement. In a Pareto optimal allocation, no further Pareto improvement can be made. Although the notion does not necessarily result in a socially desirable distribution of resources, Pareto optimality was efficiently employed to analyze the selection of alternatives in other similar fields.

Consider a valuation study in which we are interested in three attributes and each attribute has two levels. Profiles are combinations of 0 and 1, here 0 or 1 represents each attribute is at its low or high levels. If respondents are shown a choice set with three profiles, {111, 101, 011}, the choice is apparent since the first one dominates the other two. Similarly in choice set {110, 011, 001}, the third choice is not competitive because it is dominated. A simple example was provided by Raghavarao and Wiley (1998) to illustrate the implication of having dominating or dominated alternatives in a set. Suppose a consumer faces the choice between a \$9,000 car with 25 miles/gal, and a \$10,000 car with 32 miles/gal. If the consumer picks the first choice, it may be inferred that he values a \$1000 saving more than 7 additional miles/gal, assuming other things equal. On the other hand, choice of the other alternative might mean that the value of saving of fuel is more than the increment in price. However when given the choice between a \$9,000 car with 32 miles/gal and a \$10,000 car with 25 miles/gal, most consumers would choose the first combination. Under such a case we can not obtain relative values of cost versus fuel consumption. Therefore, researchers prefer choice sets without dominating or dominated profiles.

Wiley (1978) was first to realize the importance of Pareto optimal choice sets in choice experiments. Formally, assume S is the set of all possible profiles then subset T of S is said to be a Pareto optimal (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 exist subscripts i

and j $(i \neq j)$ such that $x_i < y_i$ and $x_j > y_j$. In other words, this definition indicates that no profile dominates another in PO sets.

Based on Wiley's work, many researchers did studies on the comparison of PO versus non-PO designs (Johnson et al., 1986; Huber and Hansen, 1986; Green et al., 1988). Although different results were drawn from their research, PO sets were still suggested as a good practice in CBC experiments. Krieger and Green (1991) extended this idea and proposed the concept of orthogonal and PO subsets. Raghavarao and Wiley (1998) generalized non-dominating (PO) subsets. For an s^n experiment with n attributes each at s levels denoted by $0, 1, \ldots, s-1$, assuming $(x_1, x_2, \ldots, x_n)$ is a profile of the n attributes, then 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. They note that any subset of a PO subset is itself a PO subset.

2.4 Models

Consider a CBC experiment with n attributes each at s levels. Respondents are shown sets of profiles which are combinations of levels from all attributes. A choice set is a subset of k profiles, $k \leq s^n$. The respondents' task is to select their preference from each choice set presented to them. Assume $y_{x_1 x_2 \ldots x_n}$ is the proportion of profile $(x_1, x_2, \ldots, x_n)$ picked by respondents from a given choice set S_l. Then $y_{x_1 x_2 \ldots x_n}$ can be modeled by the equation

$$y_{x_1x_2\ldots x_n} = \mu + \sum_{i=1}^{n} \alpha_{x_i}^{A_i} + \sum_{\substack{i,j=1\\ i\neq j}}^{n} \alpha_{x_ix_j}^{A_iA_j} + \sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n} \alpha_{x_ix_jx_k}^{A_iA_jA_k} + \cdots + \alpha_{x_1x_2\ldots x_n}^{A_1A_2\ldots A_n} + e_{x_1x_2\ldots x_n} \quad (2.1)$$

where μ is the general mean, $\alpha_{x_i}^{A_i}$ is the effect of attribute (or factor) A_i at x_i level, $\alpha_{x_ix_j}^{A_iA_j}$ is the effect of attributes A_i and A_j at x_i and x_j levels, respectively, $\ldots$, and $e_{x_1x_2\ldots x_n}$ is the random error.

The proportion, $y_{x_1x_2\ldots x_n}$, could be the outcomes of a probit process (Hausman and Wise 1978). However, most of time the $Y's$ of Eq.(2.1) are logits modeled by the multinomial logit random utility model (McFadden 1974). Generally such logits are described as the log-odd of the ratio of choice proportions for profiles in the choice sets to a common, base profile that appears in each choice set (Dhar 1997; Haaijer et al. 2001). Here the base profile represents "no choice", status quo, or delayed choice.

2.4.1 Fit of the model

Since a unique feature of the CBC model is to measure interactions among attributes, the tests of lack-of-fit and significance for models and model parameters are required in the estimation procedure. Raghavarao and Wiley (2006) provided a model for sequential testing. The hypothesis that the model fits the data could be tested by

$$SS[Y = X\beta] = Y'S^{-1}Y - b'(X'S^{-1}X)b \quad (2.2)$$

In the test, Eq.(2.1) can be rewritten as $Y = X\beta + \varepsilon$, where ε is a $P \times 1$ vector of errors with $E(\varepsilon) = 0$ and $Var(\varepsilon) = E(\varepsilon\varepsilon') = S$, P is the number of profiles tested, X is the design matrix constructed by the selected profiles, and β is the vector of unknown parameters that is to be estimated. The test statistic is distributed as a central χ^2 distribution with $(P - q)$ degrees-of-freedom, where q is the number of independent β parameters.

$$SS[C\beta = 0] = b'C'(C(X'S^{-1}X)^{-1}C')^{-1}Cb \tag{2.3}$$

Eq.(2.3) shows a general way to test the hypothesis $C\beta = 0$ given the fit of the model. This test has a central χ^2 distribution with d degrees-of-freedom, where d is the full row rank of the matrix C.

2.5 Sequential CBC experiments

Raghavarao and Wiley (1998) provided connected main effects plans for symmetric and asymmetric designs. Generalizing the results for 2^n experiments by Zhang (2001), they proposed rules of constructing choice sets to estimate main effects, two-way and three-way interactions (2006).

Theorem 2.1. ***(a)*** *We can estimate all nmain* $= n(s-1)$ *contrasts of main effects by using two consecutive choice sets* S_l *and* S_{l+1}*, where* $(s-2) \leq l \leq (n-1)(s-1)$.

(b) *We can estimate all ntwoway* $= n(n-1)(s-1)^2/2$ *contrasts of two-factor* *interactions by using three consecutive choice sets S_l, S_{l+1} and S_{l+2}, where $2(s-2) \leq l \leq (n-2)(s-1)$.*

(c) *We can estimate all nthreeway $= n(n-1)(n-2)(s-1)^3/6$ contrasts of three-factor interactions by using four consecutive choice sets S_l, S_{l+1}, S_{l+2} and S_{l+3}, where $3(s-2) \leq l \leq (n-3)(s-1)$.*

Generally, if l is assigned in the middle of its range choice sets will contain many profiles, while fewer profiles will be generated if l takes a value in the beginning or end of the range $l = 0, \ldots, n(s-1)$. In their paper Raghavarao and Wiley (2006) also provided a table to list minimum and maximum values for l in S_l for estimating all main effects, two-way, and three-way interactions, for $3 \leq n \leq 7$, $2 \leq s \leq 5$.

2.5.1 Sequential testing strategies

Based on the above rules for generating PO choice sets and testing models mentioned in section 2.4, the general sequential procedure of CBC experiments is:

1. First, collect data only considering main effects. Eq.(2.2) is used to test the lack-of-fit for a main effects model. If the model fits, conclusions will be drawn on main effects and then Eq.(2.3) is applied to test the significance of the model parameters.

2. If lack-of-fit for the main-effects model is significant, more choice sets will

be considered during the data collection so that main effects and two-way interactions could be estimated. Do a lack-of-fit test on the data to check the main effects and two-way interaction model. If the model fits, conclusions will be drawn on main effects and two-way interactions.

3. If the model still fails in the second step, we continue to update the data by adding more choice sets in data collection so that main effects, two-way and three-way interactions could be estimated. No lack-of-fit test is needed in this step. Based on data from the three steps, we draw conclusions on main effects, two-way and three-way interactions.

Raghavarao and Wiley (2006) recommended a unified level for individual contrasts testing at any of three steps, denoted by α_2. Assume α_1 is the level of the lack-of-fit test, if none of the main effects, two- and three-way interactions are significant, the probability of finding at least one significant contrast is

$$
\begin{aligned}
&(1-\alpha_1)\{1-(1-\alpha_2)^{nmain}\} + \alpha_1(1-\alpha_1)\{1-(1-\alpha_2)^{nmain+ntwoway}\} \\
&+ \alpha_1^2\{1-(1-\alpha_2)^{nmain+ntwoway+nthreeway}\}.
\end{aligned}
\tag{2.4}
$$

This probability should be under the control of the nominal investigation error rate $\alpha = 0.05$. In that paper, possible values for α_2 were given for different s and n values.

2.6 Connected main effects PO designs and IPP

Raghavarao and Wiley (1998) proposed a general setting with any number of levels for the attributes and obtained connected main effects plans. 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 $\lfloor l \rfloor$ is the integral part of l.

Raghavarao and Zhang (2002) applied the above results to 2^n plans. They proved that the design for a 2^n experiment, based on two PO subsets S_l and S_k is a connected main effects plan ($l \neq k, 0 < l, k < n$). Given the model

$$y_i = \mu + \sum_{j=1}^{n} x_{ij}\beta_j + e_i \tag{2.5}$$

where y_i is the response to the ith profile, $i = 1, \ldots, s$, μ is the general mean, β_j is the main effect of the jth attribute, x_{ij} is the level of the the jth attribute in the ith profile, and the e_i's are independently and identically distributed with mean zero and variance σ^2.

Assume $\mathbf{D} = \begin{bmatrix} 1_s & X \end{bmatrix}$ is the design matrix, where X is the $s \times n$ matrix whose $(i,j)th$ element is x_{ij} and 1_s is a s-dimensional vector of ones. The information matrix A could be computed by

$$\mathbf{A} = D'D = \begin{bmatrix} a_0 & a_1 1'_n \\ a_1 1_n & (a_0 - a_2)I_n + a_2 J_n \end{bmatrix}, \tag{2.6}$$

where I_n is the $n \times n$ identity matrix, J_n is the $n \times n$ matrix of all ones. In the case of two PO subsets S_l and S_k, a_0 the total number of profiles is defined as $a_0 = \binom{n}{l} + \binom{n}{k}$,

$$a_1 = \frac{2l-n}{n}\binom{n}{l} + \frac{2k-n}{n}\binom{n}{k}, \tag{2.7}$$

$$a_2 = \frac{(n-2l)^2-n}{n(n-1)}\binom{n}{l} + \frac{(n-2k)^2-n}{n(n-1)}\binom{n}{k}. \tag{2.8}$$

The information matrix of main effects after eliminating μ is

$$C_m = (a_0 - a_2)I_n + (a_2 - \frac{{a_1}^2}{a_0})J_n. \tag{2.9}$$

Since we have many choices for connected main effects plans in the range $0 < l, k < n (l \neq k)$, we would like to find an optimal one. Moreover, all these connected main effects plans consist of different number of profiles. Hence the concept information per profile (IPP) was developed as an optimality criterion to test designs with different number of profiles. We define IPP (θ) as the reciprocal of the average variance of estimated main effects divided by the number of profiles used in the study. Then θ is

$$\theta = \frac{n}{a_0 trace(C_m^{-1})} \tag{2.10}$$

where C_m^{-1} is given by

$$C_m^{-1} = \frac{1}{a_0 - a_2}\left[I_n - \frac{a_0a_2 - a_1^2}{a_0(a_0 - a_2) + n(a_0a_2 - a_1^2)}J_n\right] \tag{2.11}$$

for any connected main effects plan composed of PO subsets S_l and S_k.

Based on the value of IPP (θ), Raghavarao and Zhang (2002) also found the best PO sets for a 2^n experiment under different situations.

Theorem 2.2. *The design based on $S_{\lfloor \frac{n}{2} \rfloor}$ and $S_{\lfloor \frac{n}{2} \rfloor + 1}$ has maximum IPP among designs based on S_l and S_{l+1}.*

$$\theta = \frac{n}{a_0 trace(C_m^{-1})} = \frac{2(l+1)(n-l)}{n(n+1)} \tag{2.12}$$

For designs based on S_l and S_{n-l}, $l = 1, \ldots, n-1$, without loss of generality, take $l \leq \lfloor \frac{n}{2} \rfloor$

$$\theta = \frac{n}{a_0 trace(C_m^{-1})} = \frac{(1-z)(1+(n-1)z)}{1+(n-2)z} \tag{2.13}$$

where $z = \frac{a_2}{a_0} = \frac{(n-2l)^2 - n}{n(n-1)}$, as a function of l, decreases with respect to l. θ, as a function of z, decreases on $z \geq 0$ and increases on $z \leq 0$.

2.7 Smaller Choice Set Sizes

The number of attributes or attribute levels is a problem impeding the efficiency of implementation for CBC designs. Respondents are easily overloaded when facing numerous profiles. Reducing the number of profiles in choice sets is an effective way to implement CBC studies with large number attributes or attribute levels. Raghavarao and Wiley (2006) provided several strategies for reducing the sizes of choice sets. They considered the case with same size choice sets. Other researchers (including Koelemeijer and Oppewal 1999, Oppewal et al. 1994, 2000) contributed their works on studies with varying choice sets sizes.

Subsetting choice sets is the simplest way to form smaller sets from large ones. In this approach we divide each choice set into smaller sets of required size and then conduct the study. However, increasing the number of choice sets to reduce choice set size limits the application of this method, especially when the design has a large number of attributes or attribute levels. Raghavarao and Wiley (2006) listed the smallest numbers of profiles for a given number of attributes and levels required to estimate the maximum possible order of interactions.

Sub-setting attributes into overlapping sets is another approach for constructing smaller PO choice sets. By dividing all attributes into subsets $T_1, T_2, \ldots, T_j$ and linking adjacent nondisjoint subsets T_i and T_{i+1} $(i = 1, 2, \ldots, j-1)$ by the overlap of at least one attribute, we generate separate main-effect plans based on the attributes in each subset T_i. Nondisjoint subsets T_i and T_{i+1} make it possible to test the difference of main effects of any two attributes. For example, in a 2^5 experiment, five attributes could be divided into two overlapping subsets, $T_1 = \{A_1, A_2, A_3\}$ and $T_2 = \{A_3, A_4, A_5\}$. However, with this approach not all two-factor interactions could be estimated, only two-factor interactions of attributes in the same subset T_i are estimable.

Similar to the above approach, we also can reduce the size of choice sets by decreasing the number of attribute levels. Dividing all s levels $\{0, 1, \ldots, s-1\}$ into subsets $U_1, U_2, \ldots, U_f$, PO choice sets for n attributes could be constructed by using the levels of U_i for each attribute for $i = 1, 2, \ldots, f$, where U_i and U_{i+1} are nondisjoint for $i = 1, 2, \ldots, (f-1)$. For example, a 5^4 experiment could be run as two 3^4 experiments: one with levels $\{0, 1, 2\}$, the other with levels $\{2, 3, 4\}$. We should note there is no reason subsets can be linked by only a single level.

BIBDs bring unexpected benefits in smaller designs. Consider a BIBD with parameters $v = m, b, k, r, \lambda$ and its complement BIBD $v' = m$, $b' = b$, $k' = v - k$, $r' = b - r$, $\lambda' = b - 2r + \lambda$. A choice set S_k^* of b profiles could be created from the first design, 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. A similar approach could be used to form S_{m-k}^*. The IPP for designs based on S_k^* and S_{m-k}^* is the same as the IPP for designs based on S_k and S_{m-k}.

Raghavarao and Wiley (2006) also recommended a cyclic procedure for forming main-effect plans. A main-effect plan in s PO choice sets of size s might be generated using the following strategy.

1. Construct the first row as $(0, s-1, 1, s-2, \ldots)$ and cyclically permute to get the first choice set of size s with 0 on the diagonal.

2. Change 0 of the first set to 1 to get the second choice set, 2 to get the third one, $\ldots$ and $s-1$ to get the sth choice set.

Moreover, two-way interactions might be estimated for s^3 choice sets of size at most $(s-1)$ created by a cyclic construction strategy (Raghavarao and Wiley 2006).

CHAPTER 3

SUB-SETTING ATTRIBUTES INTO OVERLAPPING SETS FOR A 2^n PLAN

PO choice sets are essential and important in a CBC experiment. However, with an increasing of number of attributes or attribute levels, it is difficult to implement designs in a CBC study since they may have too many profiles. Therefore, practitioners are interested in finding effective ways to reduce the number of profiles and choice sets. Raghavarao and Wiley (2006) provided strategies to make smaller designs. In this chapter we focus on the method of sub-setting attributes into overlapping sets of attributes based on a 2^n plan.

3.1 Sub-setting attributes into overlapping sets with PO sets S_l and S_{l+1}

First we consider connected main effects plans based on two PO subsets S_l and S_{l+1}. Starting with the simplest case, we consider sub-designs with one or two overlapping attributes.

3.1.1 2^n plan subset into 2 sets

Given a 2^n experiment, assume r is the number of overlapping attributes; q is the number of non-overlapping attributes in each subset, $q = \frac{n-r}{2}$; and $j = r + q$ is the total number of attributes in each subset. In Theorem 3.1 and Theorem 3.2, we consider the simplest cases of one overlapping attribute when n is odd, and two overlapping attributes when n is even respectively. For example, a 2^7 experiment can be subset into two 2^4 designs with one overlapping attribute; while a 2^8 experiment can be subset into two 2^5 designs with two overlapping attributes.

Theorem 3.1. *Subset the attributes of a 2^n design into two sets with r overlapping attributes. When n is odd, $r = 1$, and each sub-design is based on S_l and S_{l+1}, the combined design has IPP*

$$\theta(l) = \frac{4n}{(2+7q)(1+q)(2+q)}[-l^2 + ql + (1+q)]. \tag{3.1}$$

IPP $\theta(l)$ is maximized when $l = \lfloor \frac{q+1}{2} \rfloor$.

Proof: When n is odd and $r = 1$, each 2^j experiment has information matrix

$$\mathbf{A} = D'D = \begin{bmatrix} a_0 & a_1 1_j' \\ a_1 1_j & (a_0 - a_2)I_j + a_2 J_j \end{bmatrix}, \tag{3.2}$$

where $a_0 = \binom{j}{l} + \binom{j}{k}$,

$$a_1 = \frac{2l - j}{j}\binom{j}{l} + \frac{2k - j}{j}\binom{j}{k},$$

and

$$a_2 = \frac{(j - 2l)^2 - j}{j(j-1)}\binom{j}{l} + \frac{(j - 2k)^2 - j}{j(j-1)}\binom{j}{k}.$$

The information matrix of main effects after eliminating μ is

$$C_m^{(i)} = (a_0 - a_2)I_j + (a_2 - \frac{{a_1}^2}{a_0})J_j.$$

where $i = 1, 2$ indicates the sub-design.

Altan and Raghavarao (1996) provided a way to create the C-matrix for nested designs to estimate the treatment effects after eliminating the block, row and column effects. Based on this method, we generate the information matrix of main effects after eliminating μ for a 2^n experiment with two sub-designs

$$\mathbf{C_m} = \begin{pmatrix} C_{11}^{(1)} & 0 & C_{13}^{(1)} \\ & & \\ 0 & C_{22}^{(2)} & C_{23}^{(2)} \\ & & \\ C_{31}^{(1)} & C_{32}^{(2)} & C_{33}^{(1)} + C_{33}^{(2)} \end{pmatrix}. \tag{3.3}$$

Rewriting according to the definitions of a_0, a_1, and a_2, we have

$$\mathbf{C_m} = \begin{pmatrix} (a_0 - a_2)I_q + (a_2 - \frac{{a_1}^2}{a_0})J_{qq} & 0_{qq} & (a_2 - \frac{{a_1}^2}{a_0})J_{q1} \\ & & \\ 0_{qq} & (a_0 - a_2)I_q + (a_2 - \frac{{a_1}^2}{a_0})J_{qq} & (a_2 - \frac{{a_1}^2}{a_0})J_{q1} \\ & & \\ (a_2 - \frac{{a_1}^2}{a_0})J_{1q} & (a_2 - \frac{{a_1}^2}{a_0})J_{1q} & 2[(a_0 - a_2)I_1 + (a_2 - \frac{{a_1}^2}{a_0})J_{11}] \end{pmatrix}.$$

In order to calculate C_m^{-1}, we simplify the above matrix and let

$$\begin{cases} A &= a_0 - a_2 \\ & \\ B &= a_2 - \frac{{a_1}^2}{a_0} \end{cases}$$

then

$$\mathbf{C_m} = \begin{pmatrix} I_2 \otimes (AI_q + BJ_{qq}) & J_{21} \otimes BJ_{q1} \\ & \\ J_{12} \otimes BJ_{1q} & 2[AI_1 + BJ_{11}] \end{pmatrix}.$$

Assume $\mathbf{C_m^{-1}} = \begin{pmatrix} X & Y \\ \\ Z & T \end{pmatrix}$, we have

$$X = \{I_2 \otimes (AI_q + BJ_{qq}) - [\tfrac{B^2}{2(A+B)}]J_{22} \otimes J_{qq}\}^{-1}$$

$$T = \tfrac{1}{2A}[I_1 - \tfrac{B}{A+(1+q)B}J_{11}]$$

It is easy to get $trace(T) = \frac{1}{2A}[1 - \frac{B}{A+(1+q)B}]$.

Next we try to find the trace of X. First we consider a general case of the inverse of

$$\begin{pmatrix} aI_q + bJ_{qq} & cJ_{qq} \\ \\ cJ_{qq} & aI_q + bJ_{qq} \end{pmatrix}.$$

Assume

$$\begin{pmatrix} aI_q + bJ_{qq} & cJ_{qq} \\ \\ cJ_{qq} & aI_q + bJ_{qq} \end{pmatrix} \begin{pmatrix} \alpha I_q + \beta J_{qq} & \gamma J_{qq} \\ \\ \gamma J_{qq} & \alpha I_q + \beta J_{qq} \end{pmatrix} = \begin{pmatrix} I_q & 0 \\ \\ 0 & I_q \end{pmatrix}$$

We can get

$$\begin{cases} a\alpha I_q + (b\alpha + \alpha\beta + bq\beta + cq\gamma)J_{qq} & = & I_q \\ \\ a\gamma + bq\gamma + c\alpha + aq\beta & = & 0 \\ \\ c\alpha + cq\beta + a\gamma + bq\gamma & = & 0 \\ \\ a\alpha I_q + (b\alpha + a\beta + bq\beta + cq\gamma)J_{qq} & = & I_q \end{cases}$$

$$\Longrightarrow \begin{cases} \alpha & = & \frac{1}{a} \\ \\ \beta & = & \frac{ac^2q-(a^2b+ab^2q)}{(a^2+abq)^2-(acq)^2} \\ \\ \gamma & = & \frac{a^2c}{(acq)^2-(a^2+abq)^2} \end{cases}$$

For our matrix X, we know

$$\begin{cases} a & = & A \\ \\ b & = & B - \frac{B^2}{2(A+B)} \\ \\ c & = & -\frac{B^2}{2(A+B)} \end{cases}$$

Based on the above results, we estimate parameters in the inverse of matrix X as

$$\begin{cases} \alpha &= \frac{1}{A} \\ \beta &= -\frac{B}{2A}[\frac{1}{A+(1+q)B} + \frac{1}{A+qB}] \\ \gamma &= \frac{B^2}{2A(A+qB)[A+(1+q)B]} \end{cases}$$

Thus we get

$$\begin{aligned} trace(X) &= 2q(\alpha+\beta) \\ &= \frac{q}{A}[2 - \frac{B}{A+(1+q)B} - \frac{B}{A+qB}] \end{aligned}$$

So the trace of the inverse of the C-matrix of a 2^n experiment will be

$$\begin{aligned} trace(C_m^{-1}) &= trace(X) + trace(T) \\ &= \frac{1}{2A}[(1+4q) - \frac{(1+2q)B}{A+(1+q)B} - \frac{2qB}{A+qB}] \\ &= \frac{1}{2(a_0-a_2)}[(1+4q) - \frac{(1+2q)(a_2-\frac{{a_1}^2}{a_0})}{(a_0-a_2)+(1+q)(a_2-\frac{{a_1}^2}{a_0})} - \frac{2q(a_2-\frac{{a_1}^2}{a_0})}{(a_0-a_2)+q(a_2-\frac{{a_1}^2}{a_0})}] \end{aligned}$$

where a_0, a_1, and a_2 are defined by each subset 2^j experiment.

Now we can consider the IPP of the 2^n experiment when n is odd and $r = 1$

$$\theta(l) = \frac{n}{2a_0 trace(C_m^{-1})}$$

$$= \frac{4n}{(2+7q)(1+q)(2+q)}[-l^2 + ql + (1+q)].$$

Since q has been determined at the beginning of sub-setting a higher level 2^n experiment, it can be treated as a constant. Then the IPP (θ) is a quadratic function, where $l = \lfloor \frac{q+1}{2} \rfloor$ maximizes $\theta(l)$.

Example 3.1.

Consider a 2^5 experiment with five factors A, B, C, D, E at two levels each. We can subset the attributes into two sets (A, B, C) and (D, E, C) with one overlapping attribute and use the 2^3 design twice. For set (A, B, C), we have choice sets $S_1 = \{100, 010, 001\}$ and $S_2 = \{110, 101, 011\}$. The information matrix A is

$$\mathbf{A} = D'D = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 0 & 6 & -2 & -2 \\ 0 & -2 & 6 & -2 \\ 0 & -2 & -2 & 6 \end{pmatrix},$$

and the C-matrix from this set is given by

$$\mathbf{C_m^{(1)}} = \begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}.$$

A similar procedure for set (D, E, C), gives the C-matrix

$$\mathbf{C_m^{(2)}} = \begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}.$$

Then the C-matrix for this 2^5 experiment will be

$$\mathbf{C_m} = \begin{pmatrix} C_{11}^{(1)} & 0 & C_{13}^{(1)} \\ 0 & C_{22}^{(2)} & C_{23}^{(2)} \\ C_{31}^{(1)} & C_{32}^{(2)} & C_{33}^{(1)} + C_{33}^{(2)} \end{pmatrix}$$

$$= \begin{pmatrix} 6 & -2 & 0 & 0 & -2 \\ -2 & 6 & 0 & 0 & -2 \\ 0 & 0 & 6 & -2 & -2 \\ 0 & 0 & -2 & 6 & -2 \\ -2 & -2 & -2 & -2 & 12 \end{pmatrix}.$$

Its inverse is computed as

$$\mathbf{C_m^{-1}} = \frac{1}{32}\begin{pmatrix} 7 & 3 & 1 & 1 & 2 \\ 3 & 7 & 1 & 1 & 2 \\ 1 & 1 & 7 & 3 & 2 \\ 1 & 1 & 3 & 7 & 2 \\ 2 & 2 & 2 & 2 & 4 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\ &= \tfrac{5}{12\times 1} \\ &= 0.4167. \end{aligned}$$

Theorem 3.2. *Subset the attributes of a 2^n design into two sets with r overlapping attributes. When n is even, $r = 2$, and each sub-design is based on S_l and S_{l+1}, then the combined design has IPP*

$$\theta(l) = \tfrac{6n}{(6+10q)(2+q)(3+q)}[-l^2 + (1+q)l + (2+q)]. \quad (3.4)$$

IPP $\theta(l)$ is maximized when $l = \lfloor \frac{q+2}{2} \rfloor$.

Proof: When n is even and $r = 2$, the information matrix of main effects after eliminating μ for a 2^n experiment with two sub-designs is

$$\mathbf{C_m} = \begin{pmatrix} I_2 \otimes [(a_0 - a_2)I_q + (a_2 - \frac{{a_1}^2}{a_0})J_{qq}] & J_{21} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{q2} \\ \\ J_{12} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{2q} & 2[(a_0 - a_2)I_2 + (a_2 - \frac{{a_1}^2}{a_0})J_{22}] \end{pmatrix}.$$

Assume $\mathbf{C_m^{-1}} = \begin{pmatrix} X & Y \\ \\ Z & T \end{pmatrix}$, and following a similar procedures as the above $r = 1$ case, we deduce the trace of the inverse of the C-matrix

$$\begin{aligned} trace(C_m^{-1}) &= trace(X) + trace(T) \\ &= \tfrac{1}{A}[(1+2q) - \tfrac{(1+q)B}{A+(2+q)B} - \tfrac{qB}{A+qB}] \\ &= \tfrac{1}{(a_0-a_2)}[(1+2q) - \tfrac{(1+q)(a_2 - \frac{{a_1}^2}{a_0})}{(a_0-a_2)+(2+q)(a_2-\frac{{a_1}^2}{a_0})} - \tfrac{q(a_2-\frac{{a_1}^2}{a_0})}{(a_0-a_2)+q(a_2-\frac{{a_1}^2}{a_0})}] \end{aligned}$$

The IPP of the 2^n experiment when n is even and $r = 2$ can be expressed by

$$\begin{aligned} \theta(l) &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\ &= \tfrac{6n}{(6+10q)(2+q)(3+q)}[-l^2 + (1+q)l + (2+q)] \end{aligned}$$

Here q is treated as a constant, the IPP (θ) has maximum value when $l = \lfloor \frac{q+2}{2} \rfloor$.

Example 3.2.

We can subset the attributes of a 2^6 experiment with six factors A, B, C, D, E, F into two sets (A, B, E, F) and (C, D, E, F) with two overlapping attributes and use the 2^4 design twice. For set (A, B, E, F), we get choice sets $S_2 = \{1100, 1010, 0110, 0101, 0011, 1001\}$ and $S_3 = \{0111, 1011, 1101, 1110\}$. The C-matrix of this set is

$$\mathbf{C_m^{(1)}} = \begin{pmatrix} 9.6 & -2.4 & -2.4 & -2.4 \\ -2.4 & 9.6 & -2.4 & -2.4 \\ -2.4 & -2.4 & 9.6 & -2.4 \\ -2.4 & -2.4 & -2.4 & 9.6 \end{pmatrix}.$$

The C-matrix of set (C, D, E, F) equals

$$\mathbf{C_m^{(2)}} = \begin{pmatrix} 9.6 & -2.4 & -2.4 & -2.4 \\ -2.4 & 9.6 & -2.4 & -2.4 \\ -2.4 & -2.4 & 9.6 & -2.4 \\ -2.4 & -2.4 & -2.4 & 9.6 \end{pmatrix}.$$

Then the C-matrix for a 2^6 experiment obtained by dividing attributes into two subsets is

$$\mathbf{C_m} = \begin{pmatrix} 9.6 & -2.4 & 0 & 0 & -2.4 & -2.4 \\ -2.4 & 9.6 & 0 & 0 & -2.4 & -2.4 \\ 0 & 0 & 9.6 & -2.4 & -2.4 & -2.4 \\ 0 & 0 & -2.4 & 9.6 & -2.4 & -2.4 \\ -2.4 & -2.4 & -2.4 & -2.4 & 19.2 & -4.8 \\ -2.4 & -2.4 & -2.4 & -2.4 & -4.8 & 19.2 \end{pmatrix}.$$

Its inverse is computed as

$$\mathbf{C_m^{-1}} = \frac{1}{72}\begin{pmatrix} 10 & 4 & 2 & 2 & 3 & 3 \\ 4 & 10 & 2 & 2 & 3 & 3 \\ 2 & 2 & 10 & 4 & 3 & 3 \\ 2 & 2 & 4 & 10 & 3 & 3 \\ 3 & 3 & 3 & 3 & 6 & 3 \\ 3 & 3 & 3 & 3 & 3 & 6 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\ &= \tfrac{6}{20\times\frac{13}{18}} \\ &= 0.4154. \end{aligned}$$

3.1.2 2^n plan subset into 2 sets for general cases

In the previous section we assumed r takes values 1 and 2, but there is no reason it cannot take other values. So to extend to a general case, given a 2^n experiment, we assume r is the number of overlapping attributes. If n mod 2 is nonzero then $r = n$ mod 2. If n mod 2 is zero, we define $r = r' + t$, where t is integer between 0 and $(n-2)$ and satisfies the inequality $2\binom{j+1}{\lfloor \frac{j}{2} \rfloor+1} < \binom{n+1}{\lfloor \frac{n}{2} \rfloor+1}$, and $r' = (n-t)$ mod 2. Let q be the number of non-overlapping attributes in each subset, $q = \frac{n-r}{2}$, and $j = r + q$ be the number of attributes in each subset. For example, a 2^8 experiment can be subset to two 2^5 designs with $t = 1$ or two 2^6 designs with $t = 3$.

Theorem 3.3. *When the attributes of a 2^n design are subset to two sets with r overlapping attributes, and each sub-design is based on S_l and S_{l+1}, then the combined designs has IPP*

$$\theta(l) = \frac{2n(r+1)}{[(r+1)(r+3q)+q](r+q)(r+q+1)}[-l^2 + (r+q-1)l + (r+q)]. \tag{3.5}$$

IPP $\theta(l)$ is maximized when $l = \lfloor \frac{r+q}{2} \rfloor$.

Proof: For the general case, the information matrix of main effects after eliminating μ is

$$\mathbf{C_m} = \begin{pmatrix} I_2 \otimes [(a_0 - a_2)I_q + (a_2 - \frac{{a_1}^2}{a_0})J_{qq}] & J_{21} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{qr} \\ \\ J_{12} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{rq} & 2[(a_0 - a_2)I_r + (a_2 - \frac{{a_1}^2}{a_0})J_{rr}] \end{pmatrix}. \tag{3.6}$$

Assume $\mathbf{C_m^{-1}} = \begin{pmatrix} X & Y \\ & \\ Z & T \end{pmatrix}$, the trace of inverse of the C-matrix of a given 2^n experiment is

$$
\begin{aligned}
trace(C_m^{-1}) &= trace(X) + trace(T) \\
&= \tfrac{1}{2A}[(r+4q) - \tfrac{(r+2q)B}{A+(r+q)B} - \tfrac{2qB}{A+qB}] \\
&= \tfrac{1}{2(a_0-a_2)}[(r+4q) - \tfrac{(r+2q)(a_2-\frac{a_1{}^2}{a_0})}{(a_0-a_2)+(r+q)(a_2-\frac{a_1{}^2}{a_0})} - \tfrac{2q(a_2-\frac{a_1{}^2}{a_0})}{(a_0-a_2)+q(a_2-\frac{a_1{}^2}{a_0})}]
\end{aligned}
$$

where a_0, a_1, and a_2 are defined by each subset 2^j experiment.

The IPP of the 2^n experiment is

$$
\begin{aligned}
\theta(l) &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\
&= \tfrac{2n(r+1)}{[(r+1)(r+3q)+q](r+q)(r+q+1)}[-l^2 + (r+q-1)l + (r+q)]
\end{aligned}
$$

where r and q are constants and determined at the beginning of sub-setting a higher level 2^n experiment. Therefore the IPP (θ) is maximum when $l = \lfloor \frac{r+q}{2} \rfloor$.

Example 3.3.

Based on the sub-setting rule we can subset a 2^8 experiment into either two 2^5 designs or two 2^6 designs. If we take $t = 1$, the attributes of a 2^8 experiment with factors A, B, C, D, E, F, G, H at two levels each can be subset into two sets (A, B, C, G, H) and (D, E, F, G, H) with two overlapping attributes and a 2^5 design is used twice. The C-matrix for each subset $(i = 1, 2)$ is

$$\mathbf{C_m^{(i)}} = \begin{pmatrix} 20 & -4 & -4 & -4 & -4 \\ -4 & 20 & -4 & -4 & -4 \\ -4 & -4 & 20 & -4 & -4 \\ -4 & -4 & -4 & 20 & -4 \\ -4 & -4 & -4 & -4 & 20 \end{pmatrix}.$$

Then the C-matrix for a 2^8 experiment subset into two 2^5 designs is

$$\mathbf{C_m} = \begin{pmatrix} 20 & -4 & -4 & 0 & 0 & 0 & -4 & -4 \\ -4 & 20 & -4 & 0 & 0 & 0 & -4 & -4 \\ -4 & -4 & 20 & 0 & 0 & 0 & -4 & -4 \\ 0 & 0 & 0 & 20 & -4 & -4 & -4 & -4 \\ 0 & 0 & 0 & -4 & 20 & -4 & -4 & -4 \\ 0 & 0 & 0 & -4 & -4 & 20 & -4 & -4 \\ -4 & -4 & -4 & -4 & -4 & -4 & 40 & -8 \\ -4 & -4 & -4 & -4 & -4 & -4 & -8 & 40 \end{pmatrix}.$$

Its inverse is presented as

$$\mathbf{C_m^{-1}} = \frac{1}{144}\begin{pmatrix} 10 & 4 & 4 & 2 & 2 & 2 & 3 & 3 \\ 4 & 10 & 4 & 2 & 2 & 2 & 3 & 3 \\ 4 & 4 & 10 & 2 & 2 & 2 & 3 & 3 \\ 2 & 2 & 2 & 10 & 4 & 4 & 3 & 3 \\ 2 & 2 & 2 & 4 & 10 & 4 & 3 & 3 \\ 2 & 2 & 2 & 4 & 4 & 10 & 3 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 & 6 & 3 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 6 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\ &= \tfrac{8}{40\times\frac{1}{2}} \\ &= 0.4. \end{aligned}$$

If $t = 3$ is considered, we can subset the attributes of a 2^8 experiment into sets (A, B, E, F, G, H) and (C, D, E, F, G, H) with four overlapping attributes and use the 2^6 design twice. The C-matrix for each subset $(i = 1, 2)$ is

$$\mathbf{C_m^{(i)}} = -\frac{40}{7}\begin{pmatrix} -6 & 1 & 1 & 1 & 1 & 1 \\ 1 & -6 & 1 & 1 & 1 & 1 \\ 1 & 1 & -6 & 1 & 1 & 1 \\ 1 & 1 & 1 & -6 & 1 & 1 \\ 1 & 1 & 1 & 1 & -6 & 1 \\ 1 & 1 & 1 & 1 & 1 & -6 \end{pmatrix}.$$

Then the C-matrix for a 2^8 experiment subset into two 2^6 designs is

$$\mathbf{C_m} = -\frac{40}{7}\begin{pmatrix} -6 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & -6 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & -6 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & -6 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -12 & 2 & 2 & 2 \\ 1 & 1 & 1 & 1 & 2 & -12 & 2 & 2 \\ 1 & 1 & 1 & 1 & 2 & 2 & -12 & 2 \\ 1 & 1 & 1 & 1 & 2 & 2 & 2 & -12 \end{pmatrix}.$$

Its inverse is expressed as

$$\mathbf{C_m^{-1}} = \frac{1}{400}\begin{pmatrix} 16 & 6 & 4 & 4 & 5 & 5 & 5 & 5 \\ 6 & 16 & 4 & 4 & 5 & 5 & 5 & 5 \\ 4 & 4 & 16 & 6 & 5 & 5 & 5 & 5 \\ 4 & 4 & 6 & 16 & 5 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 10 & 5 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 10 & 5 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 & 10 & 5 \\ 5 & 5 & 5 & 5 & 5 & 5 & 5 & 10 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \frac{n}{2a_0 trace(C_m^{-1})} \\ &= \frac{8}{70 \times \frac{13}{50}} \\ &= 0.4396. \end{aligned}$$

3.1.3 2^n plan subset into m sets

In this section we extend our theorem to the more general case of m subsets. Given a 2^n experiment, we want to subset the attributes into m sets. Assume r is the number of overlapping attributes. If n mod m is nonzero then $r = n$ mod m. If n mod m is zero, we define $r = r' + t$, where $0 < t < n - m$ and satisfies the inequality $m\binom{j+1}{\lfloor \frac{j}{2} \rfloor + 1} < \binom{n+1}{\lfloor \frac{n}{2} \rfloor + 1}$, and $r' = (n - t)$ mod m. Assume

q is the number of non-overlapping attributes in each subset, $q = \frac{n-r}{m}$, and $j = r + q$ is the number of attributes in each subset. For example, for a 2^9 experiment, we can subset it as three 2^5 designs with $t = 1$ or four 2^3 designs.

Theorem 3.4. *When the attributes of a 2^n design are subset to m sets with r overlapping attributes, and each sub-design is based on S_l and S_{l+1}, then the combined design has IPP*

$$\theta(l) \;=\; \frac{4n(r+1)}{\{(r+1)[2r+m(m+1)q]+m(m-1)q\}(r+q)(r+q+1)}[-l^2 + (r+q-1)l + (r+q)]. \tag{3.7}$$

IPP $\theta(l)$ is maximized when $l = \lfloor \frac{r+q}{2} \rfloor$.

Proof: For more general cases, the C-matrix of a 2^n experiment divided into m sets is expressed as

$$\mathbf{C_m} = \begin{pmatrix} I_m \otimes [(a_0 - a_2)I_q + (a_2 - \frac{{a_1}^2}{a_0})J_{qq}] & J_{m1} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{qr} \\ \\ J_{1m} \otimes (a_2 - \frac{{a_1}^2}{a_0})J_{rq} & m[(a_0 - a_2)I_r + (a_2 - \frac{{a_1}^2}{a_0})J_{rr}] \end{pmatrix}. \tag{3.8}$$

Assume $\mathbf{C_m^{-1}} = \begin{pmatrix} X & Y \\ \\ Z & T \end{pmatrix}$, and similar to the two-subset case, we deduce the trace of the inverse of the C-matrix as

$$
\begin{aligned}
trace(C_m^{-1}) &= trace(X) + trace(T) \\
&= \tfrac{1}{mA}[(m^2q+r) - \tfrac{m(m-1)qB}{A+qB} - \tfrac{(r+mq)B}{A+(r+q)B}] \\
&= \tfrac{1}{m(a_0-a_2)}[(m^2q+r) - \tfrac{m(m-1)q(a_2-\frac{a_1{}^2}{a_0})}{(a_0-a_2)+q(a_2-\frac{a_1{}^2}{a_0})} - \tfrac{(r+mq)(a_2-\frac{a_1{}^2}{a_0})}{(a_0-a_2)+(r+q)(a_2-\frac{a_1{}^2}{a_0})}].
\end{aligned}
$$

The IPP of a 2^n experiment with m subsets is

$$
\begin{aligned}
\theta(l) &= \tfrac{n}{ma_0 trace(C_m^{-1})} \\
&= \tfrac{4n(r+1)}{\{(r+1)[2r+m(m+1)q]+m(m-1)q\}(r+q)(r+q+1)}[-l^2+(r+q-1)l+(r+q)],
\end{aligned}
$$

where a_0 is calculated by each subset design, and m, r, q are constants determined at the beginning of sub-setting the 2^n experiment. As a quadratic function of l, the IPP (θ) is maximized when $l = \lfloor \frac{r+q}{2} \rfloor$.

Example 3.4.

There is no reason that a high level experiment can only be subset into two subsets. Therefore for the more general case, we can subset, for example, a 2^9 experiment into either three 2^5 designs or four 2^3 designs. If we take $t = 1$, the attributes of a 2^9 experiment with factors $A, B, C, D, E, F, G, H, I$ at two levels each can be subset into three sets (A, B, G, H, I), (C, D, G, H, I), and (E, F, G, H, I) with three overlapping attributes and a 2^5 design is used thrice. The C-matrix for each subset ($i = 1, 2, 3$) is

$$\mathbf{C_m^{(i)}} = \begin{pmatrix} 20 & -4 & -4 & -4 & -4 \\ -4 & 20 & -4 & -4 & -4 \\ -4 & -4 & 20 & -4 & -4 \\ -4 & -4 & -4 & 20 & -4 \\ -4 & -4 & -4 & -4 & 20 \end{pmatrix}.$$

Then the C-matrix for a 2^9 experiment subset into three 2^5 designs is

$$\mathbf{C_m} = \begin{pmatrix} 20 & -4 & 0 & 0 & 0 & 0 & -4 & -4 & -4 \\ -4 & 20 & 0 & 0 & 0 & 0 & -4 & -4 & -4 \\ 0 & 0 & 20 & -4 & 0 & 0 & -4 & -4 & -4 \\ 0 & 0 & -4 & 20 & 0 & 0 & -4 & -4 & -4 \\ 0 & 0 & 0 & 0 & 20 & -4 & -4 & -4 & -4 \\ 0 & 0 & 0 & 0 & -4 & 20 & -4 & -4 & -4 \\ -4 & -4 & -4 & -4 & -4 & -4 & 60 & -12 & -12 \\ -4 & -4 & -4 & -4 & -4 & -4 & -12 & 60 & -12 \\ -4 & -4 & -4 & -4 & -4 & -4 & -12 & -12 & 60 \end{pmatrix}.$$

Its inverse is presented as

$$\mathbf{C_m^{-1}} = \frac{1}{288}\begin{pmatrix} 18 & 6 & 3 & 3 & 3 & 3 & 4 & 4 & 4 \\ 6 & 18 & 3 & 3 & 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 18 & 6 & 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 6 & 18 & 3 & 3 & 4 & 4 & 4 \\ 3 & 3 & 3 & 3 & 18 & 6 & 4 & 4 & 4 \\ 3 & 3 & 3 & 3 & 6 & 18 & 4 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 & 4 & 8 & 4 & 4 \\ 4 & 4 & 4 & 4 & 4 & 4 & 4 & 8 & 4 \\ 4 & 4 & 4 & 4 & 4 & 4 & 4 & 4 & 8 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \tfrac{n}{2a_0 trace(C_m^{-1})} \\ &= \tfrac{9}{60\times\frac{132}{288}} \\ &= 0.3273. \end{aligned}$$

If we consider 2^3 experiments as sub-designs, we can subset the attributes of a 2^9 experiment into sets (A, B, I), (C, D, I), (E, F, I), and (G, H, I) with only one overlapping attributes and use a 2^3 design four times. The C-matrix for each subset $(i = 1, 2, 3, 4)$ is

$$\mathbf{C_m^{(i)}} = \begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}.$$

Then the C-matrix for a 2^9 experiment subset into four 2^3 designs is

$$\mathbf{C_m} = \begin{pmatrix} 6 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\ -2 & 6 & 0 & 0 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 6 & -2 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & -2 & 6 & 0 & 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 6 & -2 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & -2 & 6 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 6 & -2 \\ -2 & -2 & -2 & -2 & -2 & -2 & -2 & -2 & 24 \end{pmatrix}.$$

Its inverse is presented as

$$\mathbf{C_m^{-1}} = \frac{1}{64}\begin{pmatrix} 13 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\ 5 & 13 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 13 & 5 & 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 5 & 13 & 1 & 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 13 & 5 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 5 & 13 & 1 & 1 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 13 & 5 & 2 \\ 1 & 1 & 1 & 1 & 1 & 1 & 5 & 13 & 2 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 4 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \frac{n}{2a_0 trace(C_m^{-1})} \\ &= \frac{9}{24 \times \frac{108}{64}} \\ &= 0.2222. \end{aligned}$$

3.2 Comparison of IPP for 2^n plans

IPP is used as an optimality criterion to compare designs with different number of profiles. Table 3.1 shows the comparison of IPP values between the original PO designs and the designs obtained by sub-setting attributes into two subsets along with the number of profiles, a_0, in each design.

Table 3.1: Comparison of IPP

2^n design	Original Full Design		Sub-setting Design	
	a_0	IPP	a_0	IPP
2^5	20	0.6000	12	0.4167
2^6	35	0.5714	20	0.4154
2^7	70	0.5714	20	0.3652
2^8	126	0.5556	40	0.4000
2^9	252	0.5556	40	0.3600

Table 3.1 shows that for a given 2^n experiment, the IPP of designs obtained by sub-setting is smaller than that of the full design. Does this mean that the subsetting method should not be used? To answer this question we calculated the IPPs of overlapping attributes under the original design and the subsetting design. The results are listed in Table 3.2. Although designs from the subsetting method have smaller IPP overall, the overlapping attributes have higher IPP.

Table 3.2: The IPP of overlapping attributes

2^n design	Full Design	Sub-setting Design
2^5	0.6000	0.6667
2^6	0.5714	0.6000
2^7	0.5714	0.6000
2^8	0.5556	0.6000
2^9	0.5556	0.6000

The key point of the subsetting method is to subset a higher level experiment into two or more lower level experiments. In this procedure we must trade off some information to achieve the goal of reducing the number of profiles. Under such situations, we recommend that investigators assign attributes they are most interested in to be the overlapping factors. Not only can they estimate all main effects and all two factor interactions inclusive of the overlapping factors, but the overlapping attributes also have higher IPP. In reality, practitioners can choose designs with higher IPPs or with a small number of profiles based on their requirements.

3.3 Sub-setting attributes into overlapping sets with PO sets S_l and S_{j-l}

So far we have considered connected main effects plans based on two PO subsets S_l and S_{l+1}. Next we consider connected main effects plans of sub-designs based on PO sets S_l and S_{j-l} for $l = 1, \ldots, j-1$. Given a 2^n experiment, we want to subset the attributes into m sets. Assume r is the number of overlapping attributes. If n mod m is nonzero then $r = n$ mod m; if n mod m is zero, we define $r = r' + t$, where $0 < t < n - m$ and satisfies the inequality $m\binom{j+1}{\lfloor \frac{j}{2} \rfloor + 1} < \binom{n+1}{\lfloor \frac{n}{2} \rfloor + 1}$, and $r' = (n-t)$ mod m. Assume q is the number of non-overlapping attributes in each subset, $q = \frac{n-r}{m}$, and $j = r + q$ is the number of attributes in each subset.

Theorem 3.5. *When the attributes of a 2^n design are subset into m sets with r overlapping attributes, and each sub-design is based on S_l and S_{j-l}, then the combined design has IPP*

$$\theta(l) = n(1-w)[(m^2q + r) - \tfrac{m(m-1)qw}{1+(q-1)w} - \tfrac{(r+mq)w}{1+(r+q-1)w}]^{-1}, \tag{3.9}$$

where $w = \frac{a_2}{a_0} = \frac{(r+q-2l)^2-(r+q)}{(r+q)(r+q-1)}$.

Proof:

Based on PO sets S_l and S_{j-l}, each subset 2^j experiment has $a_0 = 2\binom{r+q}{l}$, $a_1 = 0$, and $a_2 = \frac{2[(r+q-2l)^2-(r+q)]}{(r+q)(r+q-1)}\binom{r+q}{l}$. The trace of the inverse of the C-matrix of the 2^n design is

$$trace(C_m^{-1}) \;=\; \frac{1}{m(a_0-a_2)}[(m^2q+r) - \frac{m(m-1)qa_2}{(a_0-a_2)+qa_2} - \frac{(r+mq)a_2}{(a_0-a_2)+(r+q)a_2}],$$

where a_0, a_1, and a_2 are defined by each subset 2^j experiment. Using the formula of IPP,

$$\theta(l) \;=\; \frac{n}{ma_0 trace(C_m^{-1})},$$

the theorem is proved.

Example 3.5.

Consider a 2^n plan based on two PO sets S_l and S_{j-l}. For example, given a 2^9 experiment, we can subset the attributes $A, B, C, D, E, F, G, H, I$ into two sets (A, B, C, G, H, I) and (D, E, F, G, H, I) with three overlapping attributes and use the 2^6 design twice. For each set (A, B, C, G, H, I) or (D, E, F, G, H, I), we get choice sets $S_2 = \{110000, 101000, 100100, 100010,$ $100001, 011000, 010100, 010010, 010001, 001100, 001010, 001001, 000110, 000101,$ $000011\}$ and $S_4 = \{111100, 111010, 111001, 110110, 110101, 110011, 101110,$

$101101, 101011, 100111, 011110, 011101, 011011, 010111, 001111\}$. The C-matrix of each subset $(i = 1, 2)$ is

$$\mathbf{C_m^{(i)}} = \begin{pmatrix} 30 & -2 & -2 & -2 & -2 & -2 \\ -2 & 30 & -2 & -2 & -2 & -2 \\ -2 & -2 & 30 & -2 & -2 & -2 \\ -2 & -2 & -2 & 30 & -2 & -2 \\ -2 & -2 & -2 & -2 & 30 & -2 \\ -2 & -2 & -2 & -2 & -2 & 30 \end{pmatrix}.$$

Then the C-matrix for a 2^9 experiment subset into two 2^6 designs based on PO sets S_l and S_{j-l} is

$$\mathbf{C_m} = \begin{pmatrix} 30 & -2 & -2 & 0 & 0 & 0 & -2 & -2 & -2 \\ -2 & 30 & -2 & 0 & 0 & 0 & -2 & -2 & -2 \\ -2 & -2 & 30 & 0 & 0 & 0 & -2 & -2 & -2 \\ 0 & 0 & 0 & 30 & -2 & -2 & -2 & -2 & -2 \\ 0 & 0 & 0 & -2 & 30 & -2 & -2 & -2 & -2 \\ 0 & 0 & 0 & -2 & -2 & 30 & -2 & -2 & -2 \\ -2 & -2 & -2 & -2 & -2 & -2 & 60 & -4 & -4 \\ -2 & -2 & -2 & -2 & -2 & -2 & -4 & 60 & -4 \\ -2 & -2 & -2 & -2 & -2 & -2 & -4 & -4 & 60 \end{pmatrix}.$$

Its inverse is expressed as

$$\mathbf{C_m^{-1}} = \frac{1}{8320}\begin{pmatrix} 283 & 23 & 23 & 3 & 3 & 3 & 13 & 13 & 13 \\ 23 & 283 & 23 & 3 & 3 & 3 & 13 & 13 & 13 \\ 23 & 23 & 283 & 3 & 3 & 3 & 13 & 13 & 13 \\ 3 & 3 & 3 & 283 & 23 & 23 & 13 & 13 & 13 \\ 3 & 3 & 3 & 23 & 283 & 23 & 13 & 13 & 13 \\ 3 & 3 & 3 & 23 & 23 & 283 & 13 & 13 & 13 \\ 13 & 13 & 13 & 13 & 13 & 13 & 143 & 13 & 13 \\ 13 & 13 & 13 & 13 & 13 & 13 & 13 & 143 & 13 \\ 13 & 13 & 13 & 13 & 13 & 13 & 13 & 13 & 143 \end{pmatrix}.$$

Then the IPP of the combined experiment is

$$\begin{aligned} \theta &= \frac{n}{2a_0 trace(C_m^{-1})} \\ &= \frac{9}{60 \times \frac{2127}{8320}} \\ &= 0.5867. \end{aligned}$$

CHAPTER 4

SUB-SETTING ATTRIBUTES IN A 3^n PLAN

4.1 Introduction

So far we have examined the IPP for full designs and sub-designs based on 2^n experiments. In this chapter we extend our method to 3^n experiments with n factors at three levels each.

Raghavarao and Wiley (1998) proposed that we can estimate all $n$main $=$ $n(s-1)$ contrasts of main effects by using two consecutive choice sets S_l and S_{l+1}, where $(s-2) \leq l \leq (n-1)(s-1)$. This theorem also indicates that any design based on two PO subsets S_l and S_{l+1} ($(s-2) \leq l \leq (n-1)(s-1)$) is a connected main-effect plan. Zhang (2001) obtained the PO design matrix for a 3^n experiment with a total number of s profiles. Given the model

$$y_i = \mu + \sum_{j=1}^{n} x_{ij}^1 \beta_j^1 + \sum_{j=1}^{n} x_{ij}^2 \beta_j^2 + e_i \tag{4.1}$$

where y_i is the response to the ith profile, $i = 1, \ldots, s$, μ is the general mean, β_j^1 and β_j^2 are the linear and quadratic effects of the jth attribute, $(x_{ij}^1, x_{ij}^2) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{6}})$ if the level of the jth attribute in ith profile is 0; $(x_{ij}^1, x_{ij}^2) = (0, \frac{-2}{\sqrt{6}})$ if the level of the jth attribute in ith profile is 1; $(x_{ij}^1, x_{ij}^2) = (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{6}})$ if the level of the jth attribute in ith profile is 2, and $e_i's$ are independently and identically distributed with mean 0 and variance σ^2.

Assume $\mathbf{D_3} = \begin{bmatrix} 1_s & X_3 \end{bmatrix}$ is the design matrix, where X_3 is the $s \times 2n$ matrix whose $(i, 2j-1)th$ element is x_{ij}^1, $(i, 2j)th$ element is x_{ij}^2, and 1_s is a s-dimensional vector of ones. The information matrix A of a 3^n experiment can be computed by

$$\mathbf{A} = D_3' D_3 = \begin{bmatrix} t & 1_n' \otimes \alpha' \\ 1_n \otimes \alpha & I_n \otimes (P_1 - P_2) + J_n \otimes P_2 \end{bmatrix}, \tag{4.2}$$

where $t = s$ is the total number of profiles, I_n is the $n \times n$ identity matrix, J_n is the $n \times n$ matrix of all ones,

$$\alpha = \begin{pmatrix} \frac{1}{\sqrt{2}}(t_0 - t_2) \\ \frac{1}{\sqrt{6}}(t_0 - 2t_1 + t_2) \end{pmatrix}, \tag{4.3}$$

$$\mathbf{P_1} = \begin{bmatrix} \frac{1}{2}(t_0 + t_2) & \frac{1}{\sqrt{12}}(t_0 - t_2) \\ \frac{1}{\sqrt{12}}(t_0 - t_2) & \frac{1}{6}(t_0 + 4t_1 + t_2) \end{bmatrix}, \tag{4.4}$$

$$\mathbf{P_2} = \begin{bmatrix} \frac{1}{2}(t_{00} - 2t_{02} + t_{22}) & \frac{1}{\sqrt{12}}(t_{00} - 2t_{01} + 2t_{12} - t_{22}) \\ \frac{1}{\sqrt{12}}(t_{00} - 2t_{01} + 2t_{12} - t_{22}) & \frac{1}{6}(t_{00} - 4t_{01} + 6t_{02} - 4t_{12} + t_{22}) \end{bmatrix}, \tag{4.5}$$

t_0, t_1, and t_2 represent the total number of profiles with levels $0, 1$, and 2 for attribute A_i, and t_{00}, t_{01}, t_{02}, t_{11}, t_{12}, and t_{22} are counted as the total number of profiles with level combinations $(0,0)$, $(0,1)$, $(0,2)$, $(1,1)$, $(1,2)$, and $(2,2)$ for attributes (A_i, A_j), $i \neq j$.

The information matrix of main effects after eliminating μ is

$$C_m = I_n \otimes (P_1 - P_2) + J_n \otimes (P_2 - \frac{\alpha\alpha'}{t}). \tag{4.6}$$

Since we can count the number of profiles in PO subset S_l for a 3^n experiment by

$$a_l^n = \sum_{i=0}^{\lfloor \frac{l}{2} \rfloor} \binom{n}{i} \binom{n-i}{l-2i},$$

in the case of two PO subsets S_l and S_{l+1} $(1 \leq l \leq 2n-2)$, the values of t can be defined as

$$t = a_l^n + a_{l+1}^n,$$

$$t_0 = a_l^{n-1} + a_{l+1}^{n-1},$$

$$t_1 = a_{l-1}^{n-1} + a_l^{n-1},$$

$$t_2 = a_{l-2}^{n-1} + a_{l-1}^{n-1},$$

$$t_{00} = a_l^{n-2} + a_{l+1}^{n-2},$$

$$t_{01} = a_{l-1}^{n-2} + a_l^{n-2},$$

$$t_{02} = a_{l-2}^{n-2} + a_{l-1}^{n-2},$$

$$t_{11} = a_{l-2}^{n-2} + a_{l-1}^{n-2},$$

$$t_{12} = a_{l-3}^{n-2} + a_{l-2}^{n-2},$$

and

$$t_{22} = a_{l-4}^{n-2} + a_{l-3}^{n-2}.$$

Then the IPP (θ) for estimating single degrees of freedom contrast is

$$\theta_3 = \frac{2n}{t \cdot trace(C_m^{-1})} \tag{4.7}$$

for any connected main-effects plan composed of PO subsets S_l and S_{l+1}.

We use R to calculate the IPP of 3^n designs for different values of l and $l+1$, the results are in Table 4.1.

Example 4.1.

For example, given a 3^3 experiment, if we take $l = 1$, the choice sets of the design are $S_1 = \{100, 010, 001\}$ and $S_2 = \{200, 020, 002, 110, 101, 011\}$. Then the C-matrix of the design is

$$\mathbf{C_m} = \begin{pmatrix} 2.1111 & 1.1547 & -0.8889 & -0.5774 & -0.8889 & -0.5774 \\ 1.1547 & 3 & -0.5774 & 0 & -0.5774 & 0 \\ -0.8889 & -0.5774 & 2.1111 & 1.1547 & -0.8889 & -0.5774 \\ -0.5774 & 0 & 1.1547 & 3 & -0.5774 & 0 \\ -0.8889 & -0.5774 & -0.8889 & -0.5774 & 2.1111 & 1.1547 \\ -0.5774 & 0 & -0.5774 & 0 & 1.1547 & 3 \end{pmatrix}.$$

Its inverse is

$$\mathbf{C_m^{-1}} = \begin{pmatrix} 1.3333 & -0.1925 & 0.8333 & 0.0962 & 0.8333 & 0.0962 \\ -0.1925 & 0.4444 & 0.0965 & -0.0556 & 0.0965 & -0.0556 \\ 0.8333 & 0.0962 & 1.3333 & -0.1925 & 0.8333 & 0.0962 \\ 0.0962 & -0.0556 & -0.1925 & 0.4444 & 0.0962 & -0.0556 \\ 0.8333 & 0.0962 & 0.8333 & 0.0962 & 1.3333 & -0.1925 \\ 0.0962 & -0.0556 & 0.0962 & -0.0556 & -0.1925 & 0.4444 \end{pmatrix}.$$

The IPP of the 3^3 experiment with $l = 1$ is

$$\begin{aligned}\theta_3 &= \tfrac{2n}{t \cdot trace(C_m^{-1})} \\ &= \tfrac{2\times 3}{9\times 1.7777\times 3} \\ &= 0.125.\end{aligned}$$

If instead $l = 3$ is considered, we have choice sets $S_3 = \{111, 102, 120, 210, 201, 012, 021\}$ and $S_4 = \{220, 202, 022, 121, 112, 211\}$. The C-matrix of the design is

$$\mathbf{C_m} = \begin{pmatrix} 3.8462 & -0.6662 & -1.6538 & 0.1998 & -1.6538 & 0.1998 \\ -0.6662 & 4.6154 & 0.1998 & 0.1154 & 0.1998 & 0.1154 \\ -1.6538 & 0.1998 & 3.8462 & -0.6662 & -1.6538 & 0.1998 \\ 0.1998 & 0.1154 & -0.6662 & 4.6154 & 0.1998 & 0.1154 \\ -1.6538 & 0.1998 & -1.6538 & 0.1998 & 3.8462 & -0.6662 \\ 0.1998 & 0.1154 & 0.1998 & 0.1154 & -0.6662 & 4.6154 \end{pmatrix}.$$

Its inverse is

$$\mathbf{C_m^{-1}} = \begin{pmatrix} 0.7614 & 0.0590 & 0.5739 & 0.0230 & 0.5739 & 0.0230 \\ 0.0590 & 0.2235 & 0.0230 & -0.0057 & 0.0230 & -0.0057 \\ 0.5739 & 0.0230 & 0.7614 & 0.0590 & 0.5739 & 0.0230 \\ 0.0230 & -0.0057 & 0.0590 & 0.2235 & 0.0230 & -0.0057 \\ 0.5739 & 0.0230 & 0.5739 & 0.0230 & 0.7614 & 0.0590 \\ 0.0230 & -0.0057 & 0.0230 & -0.0057 & 0.0590 & 0.2235 \end{pmatrix}.$$

The IPP of the 3^3 experiment with $l = 3$ is

$$\begin{aligned} \theta_3 &= \tfrac{2n}{t \cdot trace(C_m^{-1})} \\ &= \tfrac{2\times 3}{13\times 0.9849\times 3} \\ &= 0.1562. \end{aligned}$$

4.2 3^n plan subset into m sets

Next, we generate smaller designs for a 3^n plan by sub-setting attributes into overlapping sets of attributes, and calculate the IPP based on connected main effects plans with two PO subsets S_l and S_{l+1}.

Given a 3^n experiment, subset the attributes into overlapping sets and assume r is the number of overlapping attributes. If n mod m is nonzero then $r = n \bmod m$; if n mod m is zero, we define $r = r' + t$, where $0 < t < n - m$ and satisfies the inequality $m\binom{j+1}{\lfloor \frac{j}{2} \rfloor + 1} < \binom{n+1}{\lfloor \frac{n}{2} \rfloor + 1}$, and $r' = (n - t) \bmod m$. Assume

q is the number of non-overlapping attributes in each subset, $q = \frac{n-r}{m}$, and $j = r + q$ is the number of attributes in each subset.

We have the C-matrix for a 3^n design obtained by sub-setting attributes into m sets with r overlapping attributes

$$\mathbf{C_m} = \begin{pmatrix} I_q \otimes (P_1 - P_2) + J_q \otimes (P_2 - \frac{\alpha\alpha'}{t}) & 0_{2q2q} & \cdots & 0_{2q2q} & J_{qr} \otimes (P_2 - \frac{\alpha\alpha'}{t}) \\ 0_{2q2q} & I_q \otimes (P_1 - P_2) + J_q \otimes (P_2 - \frac{\alpha\alpha'}{t}) & & 0_{2q2q} & J_{qr} \otimes (P_2 - \frac{\alpha\alpha'}{t}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{2q2q} & 0_{2q2q} & \cdots & I_q \otimes (P_1 - P_2) + J_q \otimes (P_2 - \frac{\alpha\alpha'}{t}) & J_{qr} \otimes (P_2 - \frac{\alpha\alpha'}{t}) \\ J_{rq} \otimes (P_2 - \frac{\alpha\alpha'}{t}) & J_{rq} \otimes (P_2 - \frac{\alpha\alpha'}{t}) & \cdots & J_{rq} \otimes (P_2 - \frac{\alpha\alpha'}{t}) & m[I_r \otimes (P_1 - P_2) + J_r \otimes (P_2 - \frac{\alpha\alpha'}{t})] \end{pmatrix}.$$

It is simplified as

$$\mathbf{C_m} = \begin{pmatrix} I_m \otimes [I_q \otimes (P_1 - P_2) + J_q \otimes (P_2 - \frac{\alpha\alpha'}{t})] & J_{m1} \otimes [J_{qr} \otimes (P_2 - \frac{\alpha\alpha'}{t})] \\ J_{1m} \otimes [J_{rq} \otimes (P_2 - \frac{\alpha\alpha'}{t})] & m[I_r \otimes (P_1 - P_2) + J_r \otimes (P_2 - \frac{\alpha\alpha'}{t})] \end{pmatrix}.$$

Then based on the IPP definition, we get

$$\theta_3 = \frac{2n}{mt \cdot trace(C_m^{-1})}, \tag{4.8}$$

where the values of t are calculated by each sub-design 3^j experiment.

The R function is available to calculate the IPP of a smaller design obtained by sub-setting 3^n experiments.

Example 4.2.

Consider a 3^7 experiment with seven factors A, B, C, D, E, F, G at three levels each. We can subset the attributes into three sets (A, B, G), (C, D, G), and (E, F, G) with one overlapping attribute and use the 3^3 design thrice. The C-matrix for each subset $(i = 1, 2, 3)$ with $l = 3$ is

$$\mathbf{C_m^{(i)}} = \begin{pmatrix} 3.8462 & -0.6662 & -1.6538 & 0.1998 & -1.6538 & 0.1998 \\ -0.6662 & 4.6154 & 0.1998 & 0.1154 & 0.1998 & 0.1154 \\ -1.6538 & 0.1998 & 3.8462 & -0.6662 & -1.6538 & 0.1998 \\ 0.1998 & 0.1154 & -0.6662 & 4.6154 & 0.1998 & 0.1154 \\ -1.6538 & 0.1998 & -1.6538 & 0.1998 & 3.8462 & -0.6662 \\ 0.1998 & 0.1154 & 0.1998 & 0.1154 & -0.6662 & 4.6154 \end{pmatrix}.$$

Then the C-matrix for a 3^7 experiment subset into three 3^3 designs is

$$\mathbf{C_m} = \begin{pmatrix} 3.8461 & -0.6661 & -1.6538 & 0.1998 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.6538 & 0.1998 \\ -0.6661 & 4.6153 & 0.1998 & 0.1153 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1998 & 0.1153 \\ -1.6538 & 0.1998 & 3.8461 & -0.6661 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.6538 & 0.1998 \\ 0.1998 & 0.1153 & -0.6661 & 4.6153 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1998 & 0.1153 \\ 0 & 0 & 0 & 0 & 3.8461 & -0.6661 & -1.6538 & 0.1998 & 0 & 0 & 0 & 0 & -1.6538 & 0.1998 \\ 0 & 0 & 0 & 0 & -0.6661 & 4.6153 & 0.1998 & 0.1153 & 0 & 0 & 0 & 0 & 0.1998 & 0.1153 \\ 0 & 0 & 0 & 0 & -1.6538 & 0.1998 & 3.8461 & -0.6661 & 0 & 0 & 0 & 0 & -1.6538 & 0.1998 \\ 0 & 0 & 0 & 0 & 0.1998 & 0.1153 & -0.6661 & 4.6153 & 0 & 0 & 0 & 0 & 0.1998 & 0.1153 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.8461 & -0.6661 & -1.6538 & 0.1998 & -1.6538 & 0.1998 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0.6661 & 4.6153 & 0.1998 & 0.1153 & 0.1998 & 0.1153 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1.6538 & 0.1998 & 3.8461 & -0.6661 & -1.6538 & 0.1998 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1998 & 0.1153 & -0.6661 & 4.6153 & 0.1998 & 0.1153 \\ -1.6538 & 0.1998 & -1.6538 & 0.1998 & -1.6538 & 0.1998 & -1.6538 & 0.1998 & -1.6538 & 0.1998 & -1.6538 & 0.1998 & 11.5384 & -1.9985 \\ 0.1998 & 0.1153 & 0.1998 & 0.1153 & 0.1998 & 0.1153 & 0.1998 & 0.1153 & 0.1998 & 0.1153 & 0.1998 & 0.1153 & -1.9985 & 13.8461 \end{pmatrix}.$$

Its inverse is

$$\mathbf{C_m^{-1}} = \begin{pmatrix}
0.4715 & 0.0470 & 0.2840 & 0.0109 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1912 & 0.0076 \\
0.0470 & 0.2228 & 0.0109 & -0.0063 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0076 & -0.0018 \\
0.2840 & 0.0109 & 0.4715 & 0.0470 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1912 & 0.0076 \\
0.0109 & -0.0063 & 0.0470 & 0.2228 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0076 & -0.0018 \\
0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.4715 & 0.0470 & 0.2840 & 0.0109 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1912 & 0.0076 \\
0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0470 & 0.2228 & 0.0109 & -0.0063 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0076 & -0.0018 \\
0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.2840 & 0.0109 & 0.4715 & 0.0470 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1912 & 0.0076 \\
0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0109 & -0.0063 & 0.0470 & 0.2228 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0076 & -0.0018 \\
0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.4715 & 0.0470 & 0.2840 & 0.0109 & 0.1912 & 0.0076 \\
0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0470 & 0.2228 & 0.0109 & -0.0063 & 0.0076 & -0.0018 \\
0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.1448 & 0.0060 & 0.2840 & 0.0109 & 0.4715 & 0.0470 & 0.1912 & 0.0076 \\
0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0060 & 0.0003 & 0.0109 & -0.0063 & 0.0470 & 0.2228 & 0.0076 & -0.0018 \\
0.1912 & 0.0076 & 0.1912 & 0.0076 & 0.1912 & 0.0076 & 0.1912 & 0.0076 & 0.1912 & 0.0076 & 0.1912 & 0.0076 & 0.2537 & 0.0196 \\
0.0076 & -0.0018 & 0.0076 & -0.0018 & 0.0076 & -0.0018 & 0.0076 & -0.0018 & 0.0076 & -0.0018 & 0.0076 & -0.0018 & 0.0196 & 0.0744
\end{pmatrix}.$$

Then the IPP of the combined experiment with $l = 3$ is

$$\begin{aligned}
\theta &= \frac{2n}{mt \cdot trace(C_m^{-1})} \\
&= \frac{2\times 7}{39\times((0.4715+0.2228)*6+(0.2537+0.0744))} \\
&= 0.0799.
\end{aligned}$$

4.3 Comparison of IPP for 3^n plans

Table 4.1 shows the IPP values of the original full PO designs with $3 \leq n \leq 7$ and $1 \leq l \leq 2n-2$. When $l = n-1$ or $l = n$, the IPP has maximum value for a 3^n design. The same rule is applicable for the IPP calculation of the designs obtained by sub-setting. When $l = j-1$ or $l = j$ in each sub-design

3^j experiment, the IPP of combined design has maximum value.

Table 4.1: The IPP of 3^n plans based on S_l and S_{l+1}

$3^n \backslash l$	1	2	3	4	5	6	7	8	9	10	11	12
3^3	0.1250	0.1562	0.1562	0.1250								
3^4	0.0989	0.1374	0.1554	0.1554	0.1374	0.0989						
3^5	0.0789	0.1186	0.1426	0.1533	0.1533	0.1426	0.1186	0.0789				
3^6	0.0641	0.1026	0.1287	0.1444	0.1519	0.1519	0.1444	0.1287	0.1026	0.0641		
3^7	0.0529	0.0893	0.1158	0.1338	0.1453	0.1507	0.1507	0.1453	0.1338	0.1158	0.0893	0.0529

Table 4.2 lists the comparison of IPP values of the original full PO designs and the sub-setting designs, with two subsets and one or two overlapping attributes. The t and mt^* indicate the total number of profiles for a full and sub-design respectively, l and l^* display the summation of attribute levels of each profile in given choice sets of the full PO design and the sub-design.

Table 4.2: Comparisons of IPP for 3^n plans

3^n design	Original Full Design			Sub-setting Design		
	t	l	IPP	mt^*	l^*	IPP
3^5	96	4	0.1533	26	2	0.1080
3^6	267	5	0.1519	70	3	0.1162
3^7	750	6	0.1507	70	3	0.1045
3^8	2123	7	0.1499	192	4	0.1110
3^9	6046	8	0.1491	192	4	0.1016
3^{10}	17303	9	0.1486	534	5	0.1078

The IPP of designs obtained by sub-setting is comparable to that of the full design, but associated with a smaller number of profiles. As the number of attributes increases, the total number of profiles in the sub-designs become significantly smaller than the original design. These results illustrate the advantages of forming a small design as a way to deal with large number of attributes or attribute levels without losing much information. From our results, practitioners can choose the optimal design based on IPP values according to their requirements.

CHAPTER 5

CONCLUSION

Choice-based conjoint experiments are widely used methodologies for measuring consumer preferences. However, when the number of attributes or attribute levels becomes large, the profiles in a single choice set may be too numerous for respondents to make precise decisions. A popular and easy to implement strategy for reducing the size of choice sets is the sub-setting of attributes. However, the optimality of reduced size choice sets obtained by sub-setting attributes has not been examined in the literature. In this paper we examined the optimality of reduced size choice sets for 2^n experiments using information per profile (IPP) as the optimality criteria. We proposed a new approach for calculating the IPP for designs obtained by dividing attributes into m subsets with r overlapping attributes, and compared the IPP of the reduced size designs with that of the original full designs. We showed that experimenters must trade off some information in order to obtain reduced choice set sizes. We recommend that investigators assign the attributes they are most interested in to be the overlapping factors. Using this approach, they can estimate all main effects and all two factor interactions inclusive of the overlapping factors, and the overlapping attributes also have higher IPP. Next, we examined the IPP of choice designs based on 3^n factorial experiments. We calculated the IPP of reduced size designs obtained by sub-setting attributes in 3^n plans and compared them to the original full designs. Based on our results, practitioners can choose designs with higher IPP or a smaller number of profiles based on their requirements.

REFERENCES

Adamowicz, W., Boxall, P., Williams, M., and Louviere, J. (1998). Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American journal of agricultural economics*, 80(1):64–75.

Altan, S. and Raghavarao, D. (1996). Nested youden square designs. *Biometrika*, pages 242–245.

Batsell, R. R. and Lodish, L. M. (1981). A model and measurement methodology for predicting individual consumer choice. *Journal of Marketing Research*, pages 1–12.

Batsell, R. R. and Louviere, J. J. (1991). Experimental analysis of choice. *Marketing letters*, 2(3):199–214.

Bunch, D. S., Louviere, J. J., and Anderson, D. (1996). A comparison of experimental design strategies for multinomial logit models: The case of generic attributes. *University of California Davis Graduate School of Management Working Paper*, pages 11–96.

Carroll, J. D. (1969). Categorical conjoint mesaurement. *Annual Meeting of the Association of Mathematical Psychology.*

Carter, R., Dubelaar, C., and Wiley, J. B. (2001). Applying choice based conjoint measurement to forcast demand for a new restaurant category. *Journal of Food Products Marketing*, 6(3):63–78.

Chen, J. and Chitturi, P. (2012). Choice experiments for estimating main effects and interactions. *Journal of Statistical Planning and Inference*, 142(2):390–396.

Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of Consumer Research*, 24(2):215–231.

Gensch, D. H. and Recker, W. W. (1979). The multinomial, multiattribute logit choice model. *Journal of Marketing Research*, pages 124–132.

Green, P. E., Goldberg, S. M., and Montemayor, M. (1981). A hybrid utility estimation model for conjoint analysis. *The Journal of Marketing*, pages 33–41.

Green, P. E., Helsen, K., and Shandler, B. (1988). Conjoint internal validity under alternative profile presentations. *Journal of Consumer Research*, pages 392–397.

Green, P. E. and Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing research*, pages 355–363.

Green, P. E. and Srinivasan, V. (1990). Conjoint analysis in marketing: new developments with implications for research and practice. *The Journal of Marketing*, pages 3–19.

Grossmann, H., Holling, H., and Schwabe, R. (2002). Advances in optimum experimental design for conjoint analysis and discrete choice models. *Advances in econometrics*, 16:93–118.

Haaijer, R., Kamakura, W. A., and Wedel, M. (2001). The'no-choice'alternative in conjoint choice experiments. *International Journal of Market Research*, 43.

Haider, W. and Ewing, G. O. (1990). A model of tourist choices of hypothetical caribbean destinations. *Leisure Sciences*, 12(1):33–47.

Hausman, J. A. and Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica: Journal of the Econometric Society*, pages 403–426.

Huber, J. and Hansen, D. (1986). Testing the impact of dimensional complexity and affective differences of paired concepts in adaptive conjoint analysis. *Advances in consumer research*, 14(1):159–163.

Huber, J. and Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing research*, 33(3):307–317.

Johnson, E. J., Meyer, R. J., and Ghose, S. (1986). When choice models fail: Compensatory representations in efficient sets. In *ORSA/TIMS Marketing Science Conference, Dallas, TX, March*.

Johnson, R. M. (1974). Trade-off analysis of consumer values. *Journal of Marketing Research*, pages 121–127.

Johnson, R. M. (1987). Adaptive conjoint analysis. In *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing*, pages 253–265.

Johnson, R. M. and Olberts, K. A. (1991). Using conjoint analysis in pricing studies: Is one price variable enough? In *American Marketing Association Advanced Research Technique Forum Conference Proceedings*, pages 164–173.

Kamakura, W. A. and Srivastava, R. K. (1984). Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research*, pages 420–434.

Koelemeijer, K. and Oppewal, H. (1999). Assessing the effects of assortment and ambience: a choice experimental approach. *Journal of Retailing*, 75(3):319–345.

Krieger, A. M. and Green, P. E. (1991). Designing pareto optimal stimuli for multiattribute choice experiments. *Marketing Letters*, 2(4):337–348.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 251–263.

Kuhfeld, W. F., Tobias, R. D., and Garratt, M. (1994). Efficient experimental design with marketing research applications. *Journal of Marketing Research*, pages 545–557.

Louviere, J. J. and Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research*, pages 350–367.

Luce, R. D. and Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of mathematical psychology*, 1(1):1–27.

Mahajan, V., Green, P. E., and Goldberg, S. M. (1982). A conjoint model for measuring self-and cross-price/demand relationships. *Journal of Marketing Research*, pages 334–342.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers on Econometrics*, pages 105–421.

Oppewal, H., Louviere, J. J., and Timmermans, H. J. (1994). Modeling hierarchical conjoint processes with integrated choice experiments. *Journal of Marketing Research*, pages 92–105.

Oppewal, H., Louviere, J. J., and Timmermans, H. J. (2000). Modifying conjoint methods to model managers' reactions to business environmental trends: An application to modeling retailer reactions to sales trends. *Journal of Business Research*, 50(3):245–257.

Oppewal, H., Timmermans, H., and Louviere, J. J. (1997). Modelling the effects of shopping centre size and store variety on consumer choice behaviour. *Environment and Planning A*, 29(6):1073–1090.

Propper, C. (1995). The disutility of time spent on the united kingdom's national health service waiting lists. *Journal of Human Resources*, pages 677–700.

Raghavarao, D. and Wiley, J. B. (1998). Theory & methods: Estimating main effects with pareto optimal subsets. *Australian & New Zealand Journal of Statistics*, 40(4):425–432.

Raghavarao, D. and Wiley, J. B. (2006). Design strategies for sequential choice experiments involving economic alternatives. *Journal of statistical planning and inference*, 136(9):3287–3306.

Raghavarao, D. and Zhang, D. (2002). 2ˆn behavioral experiments using pareto optimal choice sets. *Statistica Sinica*, 12(4):1085–1092.

Schwabe, R., Grasshoff, U., Grossmann, H., and Holling, H. (2003). Optimal 2k paired comparison designs for partial profiles. In *PROBASTAT2002, Proceedings of the Fourth International Conference on Mathematical Statistics,* *Smolenice 2002, Tatra Mountains Mathematical Publications*, volume 26, pages 79–86.

Srinivasan, V. (1988). A conjunctive-compensatory approach to the self-explication of multiattributed preferences*. *Decision Sciences*, 19(2):295–305.

Srinivasan, V. and Shocker, A. D. (1973). Linear programming techniques for multidimensional analysis of preferences. *Psychometrika*, 38(3):337–369.

Waerden, P. V. D., Oppewal, H., and Timmermans, H. (1993). Adaptive choice behaviour of motorists in congested shopping centre parking lots. *Transportation*, 20(4):395–408.

Wiley, J. (1978). Selecting pareto optimal subsets from multi-attribute alternatives. *Advances in Consumer Research*, 5:171–174.

Wittink, D. R. and Cattin, P. (1989). Commercial use of conjoint analysis: An update. *The Journal of Marketing*, pages 91–96.

Wittink, D. R., Vriens, M., and Burhenne, W. (1994). Commercial use of conjoint analysis in europe: Results and critical reflections. *International journal of Research in Marketing*, 11(1):41–52.

Young, F. W. (1969). Polynomial conjoint analysis of similarities: Definitions for a special algorithm. *Research paper*.

Zhang, D. (2001). *Pareto optimal designs in behavioral experiments*. PhD thesis, Tempel University.