Support Points of Locally Optimal Designs for Multinomial Logistic Regression Models

by

Shuai Hao
B.Ec. (Henan University of Economics and Law) 2012

Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Chicago, 2019

Chicago, Illinois

Defense Committee:
Min Yang, Chair and Advisor
Jing Wang
Cheng Ouyang
Pingshou Zhong
Huayun Chen, Department of Epidemiology and Biostatistics

ACKNOWLEDGMENTS

First and foremost, I would like to extend my gratitude and thankfulness to my advisor Prof. Min Yang, for his patient guidance, and endless support throughout my time as a Ph.D student. Prof. Yang always care about my work and provide customized training to shape my research and writing skills. In addition, he often considerately encourage me in pursuing challenging questions and knows how to boost morale. Moreover, Prof. Yang is a statistician with big picture from whom I have broaden my views of statistics and set an example of excellence in research. I could not imagined to have a better advisor than him.

I would also like to thank those exceptional statisticians in my defense committee, Prof. Jing Wang, Prof. Cheng Ouyang, Prof. Pingshou Zhong, and Prof. Huayun Chen.

Thanks also go to all faculty members in our STAT group as well as my fellow students.

At last, my special thanks go to my amazing family members for their life long, unconditionally love and support. In particular, eternal thanks to my parents, I undoubtedly could not have done this with out them.

TABLE OF CONTENTS

CHAPTER **PAGE**

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

LIST OF ABBREVIATIONS

CRAN	Comprehensive R Archive Network
GLM	Generalized Linear Model
GET	General Equivalence Theorem
MLRM	Multinomial Logistic Regression Model
NPO	Non-proportional Odds
OWEA	Optimal Weights Exchange Algorithm
PO	Proportional Odds
PPO	Partial Proportional Odds

SUMMARY

This thesis consists of two parts. Chapter 1 $\sim$ 4 are on the topic of support points of optimal designs for multinomial logistic regression models. Chapter 5 provides details on an algorithm based framework for crossover designs and its companion R package.

We start with logistic regression models. The binary logistic regression models the log-odds between two response categories with a linear function of regression variables. Similarly, multinomial logistic regression models log-odds among more than two response categories with at least four link functions: baseline, cumulative, adjacent, and continuation ratio.

It is very common to use multinomial logistic regression models in data analysis, but research for its optimal designs are still at its infancy stage. This is because MLRM has many variants, each with its own structure of information matrix. Those are demonstrated in Chapter 1. We start with an introduction to multinomial logistic regression model (MLRM), including its mathematical representations, an unified form, as well as the information matrix for unknown parameters. Complication of information matrices are discussed. A literature review on designs for MLRM is then provided.

The study of optimal designs is in fact answering two questions: how many support points and what are their weights/repetitions. Our focus is on the first question, and we systematically characterize it through a complete class framework. Our results are presented in Chapter 2.

Our findings comprise three main theorems. In summary, (i), for proportional odds model with 3 response categories and 1 regression variable, optimal design needs at most 2 points for baseline link and at most 4 points for cumulative, adjacent, and continuation ratio link; (ii), for baseline proportional odds model with J response categories, optimal design consists at most 2 support points; (iii) for baseline proportional odds model with J response categories and p regression variables, optimal designs are from two equivalent class with 2^{p-1} support points with all of their first $(p-1)$ dimensions take extreme values. It is noticeable that those results hold for any optimal designs, regardless of optimality criterion chosen, parameters of interest, one-stage or multi-stage designs.

In Chapter 3, some numerical results are carried out to testify our theory and it turns out they are inline with our findings. A discussion is also included at the end to address our contribution and some future work. All proofs for main theorems are in Chapter 4.

In Chapter 5, we formulated an algorithm based framework for optimal/efficient designs for crossover designs with subject dropout mechanism, crossover designs with proportional carryover effects and interference model. We derived information matrices of those three models, and then numerical results are provided in comparison with designs in literature.

Finally, we provided an brief introduction of complete class framework in Appendix A, details of algorithm can be found in Appendix B, and a demonstration of R package for crossover designs is in Appendix C.

CHAPTER 1

INTRODUCTION TO MULTINOMIAL LOGISTIC REGRESSION MODEL AND OPTIMAL DESIGNS

1.1 Background on Multinomial Logistic Regression Model

Logistic regression model is a common tool for analyzing binary responses. Under many circumstances, however, binary responses may not be enough for detecting desire result (Perevozskaya et al., 2003). In a dose-response study, while binary response is used to estimate does-response curve, such response usually does not contain information about the severity of toxicity. For example, subject can suffer from five types of adverse effects ranging from self-limiting nausea to death in phase I cancer trial (Schacter et al., 1997). In general, if responses from experiment take values from a fixed set containing $J(>2)$ categories, they are called polytomous responses, which usually follow multinomial distribution. Polytomous data is often modelled by multinomial logistic regression model (MLRM) (Agresti, 2013, chap 6), a special case of generalized linear models (GLM) (McCullagh and Nelder, 1989). In particular, given a polytomous response, say Y, it is modeled by the following,

$$G(E(Y)) = \eta. \tag{1.1}$$

Here $G(\cdot)$ is called *link function* that transforms observed responses to log-odds, E is the expectation operator, and η is the *linear component*.

MLRM is a broad class of models. Despite of its simple looking like (1.1), it could be arbitrarily complex in the following perspectives.

First, unlike binary logistic regression where a single log-odds is modeled, one need to model $J-1$ log-odds simultaneously if response is polytomous and has J response categories.

Second, judged from the relation among response categories, polytomous response can be categorized into three kinds: nominal, ordinal, and hierarchical. For nominal response, categories are considered as equally important. For example, blood types, car makes, and etc.. On the contrary, there is nature order among categories of ordinal response, such as beef quality grade, peoples preference rating to a restaurant, and etc. Hierarchical response is different because some of response categories are nested in others. For example, in (McCullagh and Nelder, 1989), a study of mortality due to radiation consists of three stages. At first stage, outcomes are 'alive' and 'dead'; then at second stage, those who died are divided into 'due to cancer' and 'other cause'; at last, those who died from caner are labeled either 'other cancer' or 'leukemia'.

Third, each kind of polytomous response requires properly chosen link function. For nominal response, baseline link (1.4) is appropriate since the conclusion drawn from fitted model with baseline link is still valid if the label of categories are permuted (McCullagh and Nelder, 1989). As to ordinal response, cumulative link (1.5) (McCullagh and Nelder, 1989), or adjacent link (1.6) (Liu and Agresti, 2005; Agresti, 2013), are preferred for this case because if the order of categories are reversed, the conclusions made from fitted model with either of those link functions remain unchanged. If response is hierarchical, continuation ratio link (1.7) is recommended by (Zocchi and Atkinson, 1999).

Fourth, the complexity also lies on the linear component in (1.1). The linear components are usually summarized into three types of model assumptions: proportional odds model (*po*), non-proportional odds model (*npo*) and partial proportional odds model (*ppo*) ((Bu et al., 2019)). Here the 'odds' refers to 'log-odds', which is discussed in detail in next section. For proportional model, linear components across categories share the same set of parameters, whereas non-proportional model assumes each category has its own set of parameters that distinguish themselves across categories. The members in partial proportional model share parameters across categories while each of them possesses its own set of parameters. It is obvious that the partial proportional model is an amalgamation of po and npo models and therefore is the most general.

1.2 The Present Knowledge of Optimal Designs for Multinomial Logistic Regression Models

While MLRMs are widely applied in practice and the methodology of analyzing such models is well established, the optimal design research for MLRMs is arguably in its infancy stage with little optimality result available. The available results are scattered around and lack of systematical work.

As mentioned before there are as many as twelve types of MLRM due to the variety of link functions and model assumptions. The information matrix, which is the key to the study of optimal design, has its own structure under each model. Therefore one has to develop tools for optimal design case by case.

One major obstacle of studying optimal designs for MLRMs is that the information matrix depends on the unknown parameter $\boldsymbol{\theta}$ due to the nonlinearity. As pointed by (Ford et al., 1992), "A common approach to solve this dilemma is to use locally optimal designs, which are based on one's best guess of the unknown parameters. While a good guess is not always guaranteed, this approach remains of value to obtain benchmarks for all designs." There are other ways to address this issue, for example, by using a Bayesian approach (Chaloner and Verdinelli, 1995).

The complicated structure of the information matrix makes it is notoriously difficult to derive the corresponding optimal designs under MLRMs. There are, however, some nice attempts to attack this complex problem. (Zocchi and Atkinson, 1999) considered Bayesian D-optimal design for a multinomial logistic model based on hierarchical responses collected from an experiment on emergence of houseflies. They used Markov Chain Monte Carlo to generate a sample of parameters in order to access to the objective function. (Perevozskaya et al., 2003) explored some properties of information matrix of proportional odds model with cumulative link and locally optimal designs under multiple optimal criteria were investigated through numerical construction. (Yang et al., 2017) worked on a model with cumulative link for ordinal data and had shown the size of minimally supported design only depends on number of predictors. Locally D- and EW-D optimal designs were derived through algorithm approaches. (Bu et al., 2019) conducted an comprehensive study on all 12 variants of multinomial logistic regression models and provide general conclusions on the cardinality of minimally supported designs. Algorithm for D-optimal designs was also provided.

While these results explore some optimal designs, there lack of systematic understanding of their characterizations - arguably speaking, little is known about them. In this paper, we study the characterization of optimal designs for MLRMs through a complete class framework proposed by a series of papers (Yang and Stufken, 2009; Yang, 2010; Dette and Melas, 2011; Yang and Stufken, 2012; Dette and Schorning, 2013). The strategy is to find a subclass with simple format such that, for any design outside the complete class, say, ξ_1, there always exists a design in this subclass, say, ξ_2, and the information matrix of ξ_2 dominates that of ξ_1 in Lowner ordering. Utilizing this strategy, we obtain complete class results for a broad class of MLRMs. The results are significant for three reasons. First, it is the first time the characterizations of optimal designs under a varieties of MLRMs are derived. The results can help us understand the structure of optimal designs systematically. Second, the characterizations can significantly simplify the search of any specific optimal designs, both analytically and numerically, regardless of parameters of interest, optimality criteria, one-stage or multiple stage design. Third, a pressing research direction in big data analysis is the trade-off between computation complexity and statistical efficiency in big data analysis. The derived characterizations can guide us to develop efficient algorithms of selecting an informative subdata which can addresses the trade-off adequately (Wang et al., 2018).

The rest of this chapter is organized as follows. Section 1.3 sets up the mathematical representation of multinomial logistic regression model, its unified form, and information matrix for parameters. A brief introduction of framework of optimal designs can be found in section 1.4.

The main results are given in Chapter 2. Some applications are provided in Chapter 3 followed by a brief discussion. All proofs are included in Chapter 4.

1.3 Notations and settings

Suppose in an experiment, we observe n polytomous responses with J possible response categories from m distinct experiment settings. Particularly, at ith experiment settings, n_i responses, say y_{ij} for $j = 1, ..., n_i$, are collected, where $\sum_{i=1}^{m} n_i = n$.

Typically, y_{ij}'s from the same experiment setting are summarized into a count vector $\mathbf{Y}_i = (Y_{i1}, ..., Y_{iJ})'$, where Y_{ik} means the counts of responses obtained at ith experiment setting that belong to the kth category, or equivalently, if we code response categories as integers from 1 to J, then $Y_{ik} = \sum_{j=1}^{n_i} \mathbb{1}(y_{ij} = k)$ where $\mathbb{1}(\cdot)$ is an indicator function. Let the $\pi_{ik} = Prob(y_{ij} = k)$ for $k = 1, 2, ..., J$, $\sum_{j=1}^{J} \pi_{ij} = 1$, the distribution of $\mathbf{Y}_i$ is multinomial, $\mathbf{Y}_i \sim Multinomial(n_i, \pi_{i1}, ..., \pi_{iJ})$ with Probability Mass Function being

$$Prob[\mathbf{Y}_i = (Y_{i1}, ...Y_{iJ})] = \frac{n_i!}{Y_{i1}!, \cdots, Y_{iJ}!} \prod_{j=1}^{J} \pi_{ij}^{Y_{ij}} \tag{1.2}$$

In probability theory, multinomial distribution is generalized from binomial distribution and it belongs to the exponential family. Therefore, multinomial logistic model, a generalized version of logistic model, is appropriate to model the probabilities. For the ith experiment setting, say

$s_i = (x_{i1}, ..., x_{ip})$, one needs to model all probabilities simultaneously in the following general form,

$$G(\boldsymbol{\pi}_i) = \eta_i = \mathbf{X}_i\boldsymbol{\theta} \tag{1.3}$$

where link function $G(\cdot)$ is a map $\mathbb{R}^J \mapsto \mathbb{R}^{J-1}$, $\eta_i = (\eta_1(\boldsymbol{\pi}_i), ..., \eta_{J-1}(\boldsymbol{\pi}_i))'$ is a $(J-1) \times 1$ vector, $\boldsymbol{\pi}_i = (\pi_{i1}, ..., \pi_{iJ})'$ is a $J \times 1$ vector, $\mathbf{X}_i = (f_1(s_i), ..., f_{J-1}(s_i))'$ is a design matrix of order $(J-1) \times \nu$. Here $f_s'(\mathbf{X}_i)$ stands for its sth row, f is a function on $\mathbb{R}^p \mapsto \mathbb{R}^\nu$. $\boldsymbol{\theta}$ is a vector of unknown parameters of length ν. The linear component is $\eta = \mathbf{X}_i\boldsymbol{\theta}$.

Link function $G(\cdot)$ transforms responses to log-odds. All four links functions in previous section can be summarized as follows.

$$\begin{aligned}
&\textit{baseline} && \log\frac{\pi_{ij}}{\pi_{iJ}} && \text{for} \quad j = 1, ..., J-1 && (1.4)\\
&\textit{cumulative} && \log\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}} && \text{for} \quad j = 1, ..., J-1 && (1.5)\\
&\textit{adjacent} && \log\frac{\pi_{ij}}{\pi_{i,j+1}} && \text{for} \quad j = 1, ..., J-1 && (1.6)\\
&\textit{continuation} - \textit{ratio} && \log\frac{\pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}} && \text{for} \quad j = 1, ..., J-1 && (1.7)
\end{aligned}$$

For baseline link, the probability on denominator is the one from *reference category*, and it can be arbitrarily chosen, not necessarily being π_{iJ}. Meanwhile, there are many possible pairs to be used in getting log-odds, but most of them are redundant. For example, with J response categories and baseline link, there are $J(J-1)/2$ possible pairs, however one only needs to model $J-1$ selected pairs and the rest of them can be obtained using the existing ones. In general, for all the link functions above, $J-1$ log-odds are sufficient.

Notice that there is no standard criterion of choosing the right type of link functions. For some cases, as illustrated in (McCullagh and Nelder, 1989, chap 6), both baseline link and cumulative link yields similar parameter estimates and conclusions.

The linear component of (1.3) depends on model assumptions. There are at least three model assumptions emerge in literature (Bu et al., 2019): proportional odds (*po*), non-proportional odds (*npo*), and partial proportional odds (*ppo*). For $j = 1, ..., J-1$, let X_i^{rt} be the rtth entry of design matrix $\mathbf{X}_i$,

$$po \quad \eta_j(\boldsymbol{\pi}_i) = X_i^{j1}\theta_1 + \cdots + X_i^{j\nu}\theta_\nu, \tag{1.8}$$

$$npo \quad \eta_j(\boldsymbol{\pi}_i) = X_i^{j1}\theta_{j1} + \cdots + X_i^{j\nu}\theta_{j\nu}, \tag{1.9}$$

$$ppo \quad \eta_j(\boldsymbol{\pi}_i) = X_i^{j1}\theta_1 + \cdots + X_i^{j\tilde{\nu}}\theta_{\tilde{\nu}} + X_i^{j,\tilde{\nu}+1}\theta_{j,\tilde{\nu}+1} + \cdots + X_i^{j\nu}\theta_{j\nu}, \tag{1.10}$$

where $\tilde{\nu}$ is the number of parameters shared across categories in (1.10).

1.3.1 Unified model

In an effort to unify them, (Glonek and McCullagh, 1995) proposed a transformation that covers a wide scope of link functions between multinomial logistic model and log-linear model. It is written as

$$\mathbf{C}\log\left(\mathbf{L}\boldsymbol{\pi}_i\right) = \begin{pmatrix} \eta_i \\ \\ 0 \end{pmatrix} = \mathbf{X}_i\boldsymbol{\theta} \quad \text{for} \quad i = 1, ..., m. \tag{1.11}$$

where η_i is defined in (1.3), $\mathbf{C}$ is $J \times (2J-1)$ constant matrix, with $\mathbf{I}_{J-1}$ being identity matrix of order $J-1$ and $\mathbf{0}_{J-1}$ is a vector of $(J-1)$ 0's.

$$\mathbf{C} = \begin{pmatrix} \mathbf{I}_{J-1} & -\mathbf{I}_{J-1} & \mathbf{0}_{J-1} \\ \\ \mathbf{0}'_{J-1} & \mathbf{0}'_{J-1} & 1 \end{pmatrix} \tag{1.12}$$

$\mathbf{L}$ is a $(2J-1) \times J$ matrix varies through link functions. For baseline, cumulative, adjacent, and continuation-ratio link functions, the concrete structure of $\mathbf{L}$ matrices can be found in Chapter 4.

There is a major difference between (1.3) and (1.11). Because matrix $\mathbf{C}$ has J rows, which means (1.11) models simultaneously model J log-odds. A close look at $\mathbf{C}$ reveals that the last row is merely for imposing the constrain $\sum_{j=1}^{J} \pi_{ij} = 1$, and this is the reason that the last row of $\mathbf{L}$ is all 1's regardless of type link functions.

1.3.2 Information matrix

An important step for deriving the Fisher information matrix is to invert η_i for $\boldsymbol{\pi}_i$. Provided all the link functions we introduced, we can find the closed form of $\boldsymbol{\pi}_i$ in terms of $\mathbf{X}_i\boldsymbol{\theta}$. As an example, for baseline link, the π_{ij}'s could be calculated via

$$\pi_{ij} = \frac{exp\{\mathbf{X}_i\boldsymbol{\theta}\}}{1 + exp\{\mathbf{X}_i\boldsymbol{\theta}\}} \quad \text{for} \quad j = 1, \ldots, J-1. \tag{1.13}$$

Following (Bu et al., 2019), the information matrix for $\boldsymbol{\theta}$ in (1.11) is

$$I_i(\boldsymbol{\theta}) = (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^{'}})' diag\{\boldsymbol{\pi}_i\}^{-1} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^{'}}) \tag{1.14}$$

where $\partial \boldsymbol{\pi}_i / \partial \boldsymbol{\theta}^{'} = (\mathbf{C}\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i$ and $\mathbf{D}_i = diag\{\mathbf{L}\boldsymbol{\pi}_i\}$. Matrices $\mathbf{C}$ and $\mathbf{L}$ are defined in (1.11). $\mathbf{X}_i$ is design matrix related to design point $s_i = (s_1, ...s_p)$.

1.4 Optimal designs for multinomial logistic regression models

Let s_i be a *design point*(or *experiment setting*), which is a vector of regression variables. A collection of all possible design points is named *design space* and denoted by $\boldsymbol{\chi}$. Let d be an *exact design* with n runs and *support* $\mathcal{S}$, where $\mathcal{S} \subseteq \boldsymbol{\chi}$ is a set of m distinct design points. It can be written as

$$d = \{(s_i, n_i), s_i \in \mathcal{S}, \sum_{i=1}^{m} n_i = n, n_i \in \mathcal{Z}^+\} \tag{1.15}$$

where n_i's are repetitions associated with s_i's and are restricted to be positive integers. The set of all positive integers is $\mathcal{Z}^+$. An *optimal exact design* is therefore a collection of (s_i, n_i) that collectively optimizes an objective function involves with information matrix. Provided a design d, information matrix for unknown parameter can be represented by

$$I_d = \sum_{i=1}^{m} n_i I_i, \tag{1.16}$$

where I_i is information matrix for design point s_i. However, it is often an intractable issue to find optimal exact designs due to its restrictions on repetitions. In particular, the optimal exact design in closed form is frequently sought via combinatorial tools, but the solution only exist for certain combinations of experiment configurations, such as number of total runs, levels of regression variables and etc. Moreover, because this discrete nature on repetitions, numerical algorithms that work with derivatives are not applicable either. Consequently, optimal designs are often studied in the context of *approximate designs* by relaxing the discrete repetitions to continuous *weights*(or *proportions* in some literature). Formally, n_i are replaced by $w_i = n_i/n$ and the w_i's (weights) are assumed to be real numbers in the interval $[0, 1]$. An approximate design ξ as well as its information matrix, are written as follows,

$$\xi = \{(s_i, w_i), s_i \in S, \sum_{i=1}^{m} w_i = 1, w_i \in [0, 1]\} \tag{1.17}$$

$$I_\xi = \sum_{i=1}^{m} w_i I_i = \sum_{i=1}^{m} \frac{n_i}{n} I_i = \frac{1}{n} I_d \tag{1.18}$$

An *optimal approximate design* is hence sought for. The consequence of relaxation on repetitions is profound since there are a variety of optimization tools and numerical algorithms that are available in literature. As a trade-off, one has to take extra effort to carefully round approximate design to exact design which is either optimal or efficient prior to the implementation. Throughout this paper, the term 'optimal design' refers to optimal approximate design unless otherwise specified.

CHAPTER 2

MAIN RESULT

The strategy for developing those findings is inspired by complete class framework developed by (Yang and Stufken, 2012). A brief introduction is also provided in Appendix.

2.1 Model under Consideration

Our results are mainly on baseline proportional odds model and some special cases for models with other links. In general, a multinomial logistic model with $J(\geq 3)$ response categories and p continuous regression variables is

$$\mathbf{C}^{'} \log(\mathbf{L}\boldsymbol{\pi}_i) = \mathbf{X}_i\boldsymbol{\theta} \tag{2.1}$$

where

$$\mathbf{C}^{'} = \begin{pmatrix} \mathbf{I}_{J-1} & -\mathbf{I}_{J-1} & \mathbf{0}^{'}_{J-1} \\ \mathbf{0}_{J-1} & \mathbf{0}_{J-1} & 1 \end{pmatrix} \quad \mathbf{X}_i = \begin{pmatrix} 1 & & & x_{i1} & \cdots & x_{ip} \\ & \ddots & & x_{i1} & \cdots & x_{ip} \\ & & 1 & x_{i1} & \cdots & x_{ip} \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \tag{2.2}$$

$\mathbf{L}$ is an $(2J-1) \times J$ constant matrix and depends on choice of link functions. $\mathbf{X}_i$ is the design matrix associated to design point $s_i = (x_{i1}, ..., x_{ip})$, $\boldsymbol{\theta} = (\alpha_1, ..., \alpha_{J-1}, \beta_1, ..., \beta_p)$ is the parameter vector of which α_j's are intercepts and β_i's are coefficients of regression variables. Here we only assume $x_{ij} \in [U_j, V_j]$ for $j = 1, ..., p-1$, where U_j, V_j are real numbers, and x_{ip} is unbounded.

2.2 Information Matrix and Its Blocks

Since the analytical approach requires identification of maximal set of linear independent non-constant functions, we first introduce the general structure of information matrix.

Following (Bu et al., 2019), given design point s_i, information matrix for $\boldsymbol{\theta}$ is

$$I_i(\boldsymbol{\theta}) = (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}'})' diag\{\boldsymbol{\pi}_i\}^{-1} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}'}) \tag{2.3}$$

$$= \mathbf{X}_i' [(\mathbf{C}'\mathbf{D}_i^{-1}\mathbf{L})^{-1}]' diag\{\boldsymbol{\pi}_i\} [(\mathbf{C}'\mathbf{D}_i^{-1}\mathbf{L})^{-1}] \mathbf{X}_i \tag{2.4}$$

where $\partial \boldsymbol{\pi}_i / \partial \boldsymbol{\theta}' = (\mathbf{C}'\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i$ and $\mathbf{D}_i = diag\{\mathbf{L}\boldsymbol{\pi}_i\}$.

We let $\mathbf{U}$ be the matrix in the middle except for design matrix, then (2.3) can be written as $I_i = \mathbf{X}_i'\mathbf{U}\mathbf{X}_i$. Although the concrete expression of $\mathbf{U}$ varies case by case, according to Corollary 3.1 in (Bu et al., 2019), it has a general structure

$$\mathbf{U} = \begin{pmatrix} \mathbf{M} & \mathbf{0}_{J-1}' \\ \mathbf{0}_{J-1} & 1 \end{pmatrix} \tag{2.5}$$

where $\mathbf{M}$ is a $(J-1)\times(J-1)$ symmetric matrix. In addition, if design matrix $\mathbf{X}_i$ is partitioned as follows for blockwise matrix multiplication.

$$X_i = \begin{pmatrix} \mathbf{I}_{J-1} & \mathbf{S} \\ \mathbf{0}_{J-1} & \mathbf{0}_1 \end{pmatrix} \tag{2.6}$$

where the submatrix $\mathbf{S}$ is a $(J-1)\times p$ matrix that holds values of regression variables, and $\mathbf{0}_{J-1}$ and $\mathbf{0}_1$ are vectors of 0's with appropriate orders. As a result, we reach to the following lemma for structure of information matrix.

Lemma 1. *Given matrix partitions in* (2.5) *and* (2.6)*, information matrix at* $s_i = (x_{i1}, ..., x_{ip})$ *can be presented by blocks.*

$$\begin{aligned} I_i(\theta) &= \begin{pmatrix} \mathbf{I}_{J-1} & \mathbf{S} \\ \mathbf{0}_{J-1} & \mathbf{0}_1 \end{pmatrix}^{'} \begin{pmatrix} \mathbf{M} & \mathbf{0}_{J-1}^{'} \\ \mathbf{0}_{J-1} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{I}_{J-1} & \mathbf{S} \\ \mathbf{0}_{J-1} & \mathbf{0}_1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{M} & \mathbf{MS} \\ \mathbf{S}^{'}\mathbf{M} & \mathbf{S}^{'}\mathbf{MS} \end{pmatrix} = \begin{pmatrix} B1 & B2^{'} \\ B2 & B3 \end{pmatrix} \end{aligned} \tag{2.7}$$

We name those blocks by 'B' + numbers, where letter 'B' is short for 'Block', and $B2^{'}$ *is B2 block transposed.*

Furthermore, let $\mathbf{M} = \{M_{ij}\}$, *and define* $M_{\cdot j} = \sum_{i=1}^{J} M_{ij}$ *and* $M_{\cdot\cdot} = \sum_{i=1}^{J-1} \sum_{j=1}^{J-1} M_{ij}$,

$$B2 = \begin{pmatrix} x_1 M_{\cdot 1} & x_1 M_{\cdot 2} & \cdots & x_1 M_{\cdot p} \\ x_2 M_{\cdot 1} & x_2 M_{\cdot 2} & \cdots & x_2 M_{\cdot p} \\ \vdots & \vdots & \ddots & \vdots \\ x_p M_{\cdot 1} & x_p M_{\cdot 2} & \cdots & x_p M_{\cdot p} \end{pmatrix} \quad B3 = \begin{pmatrix} x_1^2 M_{\cdot 1} & x_1 x_2 M_{\cdot 2} & \cdots & x_1 x_p M_{\cdot p} \\ x_2 x_1 M_{\cdot 1} & x_2^2 M_{\cdot 2} & \cdots & x_2 x_p M_{\cdot p} \\ \vdots & \vdots & \ddots & \vdots \\ x_p x_1 M_{\cdot 1} & x_p x_2 M_{\cdot 2} & \cdots & x_p^2 M_{\cdot p} \end{pmatrix} \tag{2.8}$$

The proof is merely matrix multiplications, and is therefore omitted here. Lemma 1 plays an important role in following sections where those structures will be extensively exploited.

2.3 Proportional Model with 3 Response Categories and 1 Regression variable

For $J = 3$ and $p = 1$, design points $s_i = x_i$ reduces to a scalar and design matrix as well as $\boldsymbol{\theta}$ become

$$X_i = \begin{pmatrix} 1 & 0 & x_i \\ 0 & 1 & x_i \\ 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta \end{pmatrix}. \tag{2.9}$$

In the information matrix, B2 block reduces to a row vector and B3 is now a scalar. We investigated such a matrix and reach to the following theorem on complete class.

Theorem 2. *For proportional odds model* (2.1) *with* 1 *continuous regression variable* $x_i \in [U, V]$ *where* U, V *are real numbers, and* 3 *response categories, the following results on complete class hold.*

1. *For baseline link, designs with at most* 2 *support points form a complete class.*

2. *For cumulative, continuation ratio or adjacent link, designs with at most* 4 *support points form a complete class.*

The detailed proof is given in Chapter 4. Theorem 2 provides upper bounds of number of support points for MLRM with 1 continuous covariate and 3 response categories. In particular, optimal designs for such model with baseline link will have at most 2 support point. Meanwhile, the model with cumulative, continuation ratio and adjacent link will have at most 4 design points. According to (Bu et al., 2019), the minimal number of support points for this case is 2 for baseline link and 3 for the rest type of links. Combined with Theorem 2, optimal designs for baseline multinomial logistic regression model with 3 response categories are minimally supported.

2.4 Baseline Proportional Odds Model with J Categories and 1 Regression Variable

In this section, we generalized complete class result for baseline proportional odds model to the one with $J \geq 3$ response categories. The model is

$$\log(\frac{\pi_j}{\pi_J}) = \alpha_j + \beta x, \quad j = 1, ..., J-1 \tag{2.10}$$

If written in matrices like (2.1), design matrix X_i and $\boldsymbol{\theta}$ now become

$$X_i = \begin{pmatrix} 1 & & & x_i \\ & \ddots & & \vdots \\ & & 1 & x_i \\ 0 & \dots & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{J-1} \\ \beta \end{pmatrix} \tag{2.11}$$

In its information matrix, the B2 block is still a row vector and B3 is a scalar. But B1 block now is of order $J-1$ by $J-1$. We have the following complete class result for this case.

Theorem 3. *For baseline proportional odds model* (2.1) *with* $J \geq 3$ *response categories and* 1 *continuous regression variable* $x_i \in [U, V]$ *where* U, V *are real numbers, designs with at most 2 support points form a complete class.*

Theorem 3 generalizes complete class result for baseline proportional model to arbitrary number of response categories. That is, the optimal design for baseline proportional odds model consists at most 2 support points regardless of number of response categories. It broadens the scope of its applications.

2.5 Baseline Proportional Odds Model with J Categories and p Regression Variables

We now consider arbitrary J and p. The baseline proportional odds model with $J \geq 3$ response categories and $p \geq 2$ continuous regression variables can be written as

$$\log(\frac{\pi_j}{\pi_J}) = \alpha_j + \beta_1 x_1 + \cdots \beta_p x_p, \quad j = 1, \dots, J-1. \tag{2.12}$$

where Jth response category is conventionally set to be reference category, and x_i are the value of ith regression variable. Here we only assume $x_j \in [U_j, V_j]$ for $j = 1, ..., p-1$, where U_j, V_j are finite real numbers.

As introduced at the beginning of this section, the design matrices and parameter vector are exactly the same as (2.2). For example, when $J = 4, p = 2$, design matrix is

$$X = \begin{pmatrix} 1 & 0 & 0 & x_1 & x_2 \\ 0 & 1 & 0 & x_1 & x_2 \\ 0 & 0 & 1 & x_1 & x_2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{pmatrix}. \tag{2.13}$$

In general, by Lemma 1, key components in information matrix are:

The B1 block is $\mathbf{M}$ matrix of order $(J-1) \times (J-1)$ with

$$M_{ij} = \begin{cases} -\frac{e^{\alpha_i + \alpha_j + 2\sum_{t=1}^{p} \beta_t x_t}}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\sum_{t=1}^{p} \beta_t x_t}]^2}, & i \neq j \\ \frac{e^{\alpha_i + \sum_{t=1}^{p} \beta_t x_t}[1+(\sum_{s=1, s\neq j}^{J-1} e^{\alpha_s})e^{\sum_{t=1}^{p} \beta_t x_t}]}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\sum_{t=1}^{p} \beta_t x_t}]^2}, & i = j. \end{cases} \tag{2.14}$$

B2 block is a $p \times (J-1)$ matrix with the ijth entry being

$$x_i M_{\cdot j} = x_i \frac{e^{\alpha_j + \sum_{t=1}^{p} \beta_t x_t}}{[1 + (\sum_{j=1}^{J-1} e^{\alpha_j})e^{\sum_{t=1}^{p} \beta_t x_t}]^2}, \quad \text{for } j = 1, ..., J-1 \tag{2.15}$$

and B3 block is a $p \times p$ matrix with ijth entry being

$$x_i x_j M_{..} = \frac{e^{\sum_{t=1}^{p} \beta_t x_t} \sum_{s=1}^{J-1} e^{\alpha_s}}{[1 + (\sum_{j=1}^{J-1} e^{\alpha_j}) e^{\sum_{t=1}^{p} \beta_t x_t}]^2} \tag{2.16}$$

Instead of directly study the designs, we focus on the transformed design points, with support points $s_i = (x_{i1}, ..., x_{i,p-1}, c_i)$, where $c_i = \sum_{t=1}^{p} \beta_t x_t$ and $\beta_t \neq 0$ for all possible t. Note that such transformation does not change the complete class result, because of the following factorization of information matrix. For a design point x and transformed design point s,

$$I(s, \boldsymbol{\theta}) = \mathbf{X}^{'}\mathbf{U}\mathbf{X} = \mathbf{Q}^{'}\mathbf{F}^{'}\mathbf{U}\mathbf{F}\mathbf{Q} \tag{2.17}$$

where $\mathbf{X}$ is design matrix for $x = (x_1, ..., x_p)$ and $\mathbf{F}$ is design matrix for $s = (x_1, ..., c)$, and $\mathbf{FQ} = X$.

$$A(\theta) = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ 0 & \cdots & \beta_1 & \cdots & \beta_p \end{pmatrix}^{-1} \qquad F = \begin{pmatrix} 1 & & & x_1 & \cdots & c \\ & \ddots & & \vdots & \vdots & \vdots \\ & & 1 & x_1 & \cdots & c \end{pmatrix} \tag{2.18}$$

(2.19)

Let $I(s, \boldsymbol{\theta})$ stands for information matrix at $s = (x_1, ..., c)$, one can easily obtain $I(s, \boldsymbol{\theta})$ from $I(x, \boldsymbol{\theta})$ by (2.17). The structures of them are identical. Under this setting, we have the following theorem.

Theorem 4. *In the transformed design space, for an arbitrary design* $\xi = \{(s_i, w_i), i = 1, ..., m; \sum_{i=1}^m w_i = 1\}$*, there exists a design* $\tilde{\xi}$ *such that the following inequality of information matrices hold:*

$$I_\xi(\theta) \leq I_{\tilde{\xi}}(\theta) \tag{2.20}$$

where

$$\tilde{\xi} = \{(\tilde{F}_{\ell 1}, w_{\ell 1}) \quad and \quad (\tilde{F}_{\ell 2}, w_{\ell 2}), \ell = 1, ..., 2^{p-1}\} \tag{2.21}$$

and $\tilde{F}_{\ell 1} = (a_{\ell 1}, ..., a_{\ell, p-1}, \tilde{c}_1)$*,* $\tilde{F}_{\ell 2} = (a_{\ell 1}, ..., a_{\ell, p-1}, \tilde{c}_2)$*. Here* $a_{\ell, j} = U_j$ *or* V_j*, and* $(a_{\ell 1}, ..., a_{\ell, p-1})$ *are all combinations of them for* $\ell = 1, ..., 2^{p-1}$*, and* $\tilde{c}_1$ *and* $\tilde{c}_2$ *are two numbers need to be solved.*

The proof is deferred to Chapter 4 as well. Theorem 4 shows that the optimal designs for baseline proportional model with p covariates are made of two equivalent classes of design points of which the value of its first $p - 1$ covariates are easily found. The significance is not only the optimal designs for such a general model are in simple structure, also algorithms would benefit from it since it reduce the dimension of an optimization problem from p to 1.

Note that when $J = 2$, Theorem 4 reduces to theorem in (Yang et al., 2011a), where similar result for binary logistic regression is derived. Therefore, it generalizes Yang's result to baseline log-odds model.

CHAPTER 3

APPLICATIONS

All designs in this section are locally optimal. Therefore one needs to provide initial values of parameters in order to derive a design aiming at estimating them. Initial values are not randomly chosen, on the contrary, it should be determined prudently by either consulting experts opinions or look up historical experiment results. We have mentioned that locally optimal design can serve as a benchmark for other designs. However a set of badly chosen initial values which are far away from the 'truth' would result in a benchmark that has not too much practical meaning even though it is locally optimal.

We use optimal weights exchange algorithm (OWEA) in Appendix for approximate designs, detailed procedures are available in section 5.3. Although the scope of this paper is on models with continuous regression variables, we still need to discrete the design space in order to use the algorithm. A common practise is to use equally spaced grid on the design space, as we did in following examples.

3.1 Examples

Example 1: Consider the following baseline proportional odds model, with $J = 4$

$$\log(\frac{\pi_1}{\pi_4}) = \alpha_1 + \beta x \tag{3.1}$$

$$\log(\frac{\pi_2}{\pi_4}) = \alpha_2 + \beta x \tag{3.2}$$

$$\log(\frac{\pi_3}{\pi_4}) = \alpha_3 + \beta x \tag{3.3}$$

We select two sets of initial parameters $(0.5, -0.6, 0.9, 2)$ and $(1, 2, 4, -0.3)$, for the given design spaces, we use R program to find optimal approximate designs for both A- and D-optimal criteria. These approximate designs are summarized in Table I. Here the N stands for number of grid points in design space, and entries on the column 'design' are written in the format of $(point, weight)$. Finally, we count the number of support points and add them to the last column.

It is noticeable that all those designs consist of two support points, which is consistent with our findings in Theorem 3. In fact, according to (Bu et al., 2019), designs with 2 support points for this model is also minimally supported which is the minimum requirement for information matrix being non-singular and hence parameter estimation being unbiased. Therefore, optimal designs could be both optimal and minimally supported. An interesting observation is, contrary to common case where D-optimal design has equal weights, the minimally supported D-optimal design is not equally weighted. For example, for the first set of initial parameter values, D-

$(\alpha_1, \alpha_2, \alpha_3, \beta)$	criterion	design space	N	design	# of points
$(0.5, -0.6, 0.9, 2)$	D	[-3,3]	121	(-1.25,0.3106) (0.35,0.6894)	2
$(0.5, -0.6, 0.9, 2)$	A	[-3,3]	121	(-1.4,0.2198) (0.2,0.7802)	2
$(0.5, -0.6, 0.9, 2)$	D	[0,3]	61	(0.0,0.4580) (1.1,0.5420)	2
$(0.5, -0.6, 0.9, 2)$	A	[0,3]	61	(0.0,0.4509) (1.2,0.5491)	2
$(1, 2, 4, -0.3)$	D	[-10,30]	81	(6.5,0.6877) (17.0,0.3123)	2
$(1, 2, 4, -0.3)$	A	[-10,30]	81	(5.0,0.8536) (21.5,0.1464)	2
$(1, 2, 4, -0.3)$	D	[0,30]	61	(6.5,0.6877) (17.0,0.3123)	2
$(1, 2, 4, -0.3)$	A	[0,30]	61	(6.0,0.7751) (24.5,0.2249)	2

TABLE I

LOCALLY OPTIMAL DESIGNS BASELINE PROPORTIONAL ODDS MODELS

optimal design has two points $-1.25, 0.35$ with weights 0.3106 and 0.6894. Meanwhile, A-optimal design has two weights being 0.2198 and 0.7802.

Example 2: In (Perevozskaya et al., 2003), an early pioneer paper that studies designs for MLRM, they provided locally optimal designs for the following model, which is a cumulative link model with 4 response categories.

$$\log \frac{\gamma_j(x)}{1-\gamma_j(x)} = x - \alpha_j \quad \text{for } j = 1, 2, 3 \tag{3.4}$$

where $\gamma_j(x) = Prob(Y \leq j|x) = \sum_{s=1}^{j} \pi_{is}$. Here those intercept terms are unknown parameters and they set the slope to be a constant. Such a model is used for does-response study. Inspired by this paper, we reparameterize model in the fashion of proportional odds model and assume slope is also unknown. For simplicity, we only consider 3 response categories. The model is formulated as

$$\log \frac{\gamma_1(x)}{1-\gamma_1(x)} = \alpha_1 + \beta x \tag{3.5}$$

$$\log \frac{\gamma_2(x)}{1-\gamma_2(x)} = \alpha_2 + \beta x \tag{3.6}$$

Notice that there is a natural order that $\alpha_1 \geq \alpha_2$.

Similarly, two sets of initial parameter values are chosen upon which A- and D- optimal designs are derived for given design spaces. Table II summarizes key information of those designs.

Most of those designs in Table II have 3 support points, except those on the first two rows. In particular, D-optimal design on the first row have two points, $3.85, 3.90$, that can be combined as one since they are actually two adjacent grid points and 3.85 has very small weight. Such a design can be considered as the one with three support points, $0.55, 7.20$ and a where $a \in (3.85, 3.90)$. As theorem 2 shows optimal designs have at most 4 support points, the abundance of 3-point designs and absence of 4-point designs might give a hint that our current result could be improved. Finally, it is worth mentioning that, according to (Bu et al., 2019), those D-optimal designs are also minimally supported.

$(\alpha_1, \alpha_2, \beta)$	criterion	design space	N	design	# of points
$(1, 2.88, -0.5)$	D	[-10,15]	501	(0.55,0.3970) (3.85,0.0214) (3.90,0.1853) (7.20,0.3963)	4^*
$(1, 2.88, -0.5)$	A	[-10,15]	501	(-1.00,0.6539) (3.45,0.3461)	2
$(1, 2.88, -0.5)$	D	[0,15]	1501	(0.53,0.3949) (3.86,0.2066) (7.18,0.3985)	3
$(1, 2.88, -0.5)$	A	[0,15]	1501	(0.00,0.7456) (3.95,0.2209) (9.27,0.0335)	3
$(-2, 1, 0.8)$	D	[-10,10]	2001	(-2.20,0.3180) (0.62,0.3630) (3.44,0.3190)	3
$(-2, 1, 0.8)$	A	[-10,10]	2001	(-2.10,0.2800) (0.85,0.6918) (3.68,0.0282)	3
$(-2, 1, 0.8)$	D	[0,9]	901	(0.00,0.6361) (3.94,0.0454) (4.00,0.3185)	3
$(-2, 1, 0.8)$	A	[0,9]	901	(0.00,0.8630) (4.07,0.0117) (4.50,0.1253)	3

TABLE II

LOCALLY OPTIMAL DESIGNS FOR CUMULATIVE PROPORTIONAL ODDS MODEL

In practise, when there is no information on unknown parameters, an intuitive yet commonly used strategy is to implement uniform designs. Such a design puts equal weights on design points, and sometimes those points are equally spread in design space as well. However they are known as lacking efficiency. For example, (Yang et al., 2017; Bu et al., 2019) proved uniform designs are less efficient for D-optimality under some variants of multinomial logistic regression. We have the same observation here. Consider the following two uniform designs in Table III, of which puts equal weights to its support.

design space	design points	# of points	A-eff	D-eff
[-10,15]	-10,-5,0,5,10,15	6	0.5339	0.2589
[0,15]	0,2,4,8,15	5	0.6740	0.5862
[-10,10]	-10,-6,-2,0,2,6,10	7	0.4822	0.1892
[0,9]	0,1,2,3,4,5,6,7,8,9	10	0.3563	0.2278

TABLE III

UNIFORM DESIGNS AND EFFICIENCY TO OWEA DESIGNS

For simplicity, we compare optimal designs with uniform designs. Here in Table III, 'A-eff' and 'D-eff' are shorts for relative efficiency under A- and D- optimality respectively, and they are calculated by

$$eff = \frac{\Phi(\Sigma_{optimal})}{\Phi(\Sigma_{uniform})} \tag{3.7}$$

When $eff > 1$, it indicates uniform design is more efficient, and vice versa. As shown, uniform designs are not as efficient as those optimal designs. On the contrary, the difference is quit huge. For example, given design space $[-10, 15]$, there are 2 points and 4 points for A- and D- optimal designs, and they are almost 1 and 3 times more efficient than the uniform design with 5 points.

Example 3: In (Agresti, 2013, chap 6), a developmental toxicity study with pregnant mice was introduced. In this experiment, a certain chemical substance in distilled water of different concentrations (from 0 to 500 mg/kg per day) was given to pregnant mice in successive 10 days and their uterine contents were analyzed in order to examine the defects of fetuses. There are three outcomes for each fetus: nonlive, malformation, or normal. The outcome is ordinal with 'nonlive' being the most preferable. The original design has 5 levels of concentration, $0, 62.5, 125, 250, 500$, where 0 is the level of control group. Those design points spread out in the design space, $[0, 500]$. Design points and number of observations are summarized in Table IV.

dose	0	62.5	125	250	500
observations	297	242	312	299	285

TABLE IV

ORIGINAL DESIGN IN TOXICITY STUDY

A continuation-ratio proportional model is considered because the response are hierarchical. With 3 response categories, if we target at the following model,

$$\log \frac{\pi_1}{\pi_{i2} + \pi_{i,3}} = \alpha_1 + \beta x \tag{3.8}$$

$$\log \frac{\pi_2}{\pi_{i,3}} = \alpha_2 + \beta x \tag{3.9}$$

where the x means concentration. For this example, we set $(\alpha_1, \alpha_2, \beta) = (0.1, -0.5, 0.016)$, which is the initial estimate provided by (Agresti, 2013, chap 6). We use OWEA algorithm to find A- and D- optimal designs in the space $[0, 500]$. Both designs are summarized in Table V.

The numbers of support points, regardless of optimal criteria, are all equal to 2, which is less than the upper bound of 4 provided in Theorem 2. Also, the design points are not uniformly allocated.

In fact, for this toxicity study, (Agresti, 2013, chap 6) fitted a continuation-ratio non-proportional model, and give estimations for $\beta_1 = 0.0064, \beta_2 = 0.0174$.

$$\log \frac{\pi_1}{\pi_{i2} + \pi_{i,3}} = \alpha_1 + \beta_1 x \tag{3.10}$$

$$\log \frac{\pi_2}{\pi_{i,3}} = \alpha_2 + \beta_2 x \tag{3.11}$$

Although we have not derive any complete class result for such model, optimal designs can still be derived numerically. In this case, we search locally optimal designs at $(\alpha_1, \alpha_2, \beta_1, \beta_2) = (0.4, 1, 0.0064, 0.0174)$, which serves as initial 'guess' of unknown parameters.

As shown in Table V, designs have 3 levels of concentrations for D-optimal and only 2 for A-optimal. Under both criteria, the control level 0 is always included. The significance is that experimenter can reduce the levels of concentrations which saves man power and reduces the experimental cost in some sense. Lastly, we still observed that those designs have number of design points that are less than the theoretical maximum as we derived in Theorem 2.

The last column in Table V is the relative efficiency that comparing original design in (Agresti, 2013, chap 6) to optimal design in Table V. The formula is similar to (3.7). It is obvious that the original design is lacking efficiency for both A- and D- optimality. For example, D-optimal design based on 3 points for npo model is almost 2 times as efficient as

$(\alpha_1, \alpha_2, \beta_1, \beta_2)$	criterion	design space	N	design	# of points	efficiency
Locally Designs for Proportional Odds Model						
$(0.1, -0.5, 0.016)$	D	[0,500]	501	(0,0.6323) (121,0.3677)	2	0.2296
$(0.1, -0.5, 0.016)$	A	[0,500]	501	(0,0.9926) (122,0.0074)	2	0.2984
Locally Designs for Non-proportional Odds Model						
$(0.4, 1, 0.0064, 0.0174)$	D	[0,500]	501	(0,0.4653) (117,0.3741) (365,0.1606)	3	0.5108
$(0.4, 1, 0.0064, 0.0174)$	A	[0,500]	501	(0,0.9797) (105,0.0203)	2	0.2533

TABLE V

LOCALLY OPTIMAL DESIGNS FOR CONTINUATION MODELS

original design. The take away message, is optimal or efficient designs for models like (3.8) and (3.10), can be based on only limited number of design points. Those observations are inline with the spirit of theorems derived in this paper.

3.2 Discussion

Multinomial logistic regression model plays an important role in statistical analysis. However, the research on optimal designs is still at its infancy stage. Deriving optimal designs for MLRM in general is difficult. As stated in introduction, there are two major obstacles. First, the MLRM consists of at least 12 types of variants and each has its own concrete expression of both model and information matrix. So far relevant optimal designs are generated case by case. Second, the information matrix is parameter dependence, and mostly, locally optimal designs are studied.

Although there are increasing number of researches on related fields, optimal design for MLRM is still under development and almost all of existing designs emerged in literature so far are constructed in a numerical manner. While these results are helpful in some sense, they are in fact merely computational and cannot provide further insights. In recent decades, some preliminary theoretical results have been established regarding the unified model representation, information matrix and etc. There are, nevertheless, still no such studies for optimal designs from theoretical perspective.

In this paper, we accessed the optimal designs for MLRMs via an analytical approach. The main result is on the complete class of optimal designs for some prevalent models. In particular, we derived the upper bound for number of support points of optimal designs. Such results provide evidence for the claim that optimal designs for MLRM usually do not have many support points. This is important because one can expect a simple design return by numerical algorithms and if the design have more design points that exceeds the upper bound, one can instantly know there is a design which has a less support points.

Numerical examples are also explored. It is shown that number of support points of those designs are in line with our theory. In particular, some examples have number of supports that is exactly what indicated by our theorem. More interestingly, designs from some other examples have less support points than what we derived in theory. Since our theory hold regardless of initial values of parameters and optimal criteria, there might be some other cases that have exactly maximum number of support points. Moreover, and even more exciting, it might be possible to improve our result in the future.

In addition, selecting initial values for parameters is tricky. For example, (Agresti, 2013, chap 6) argues that cumulative link indicates that the cumulative probability must be stochastically ordered, otherwise, the model will be poorly fitted. The is the general guidelines for choosing initial parameters. Some bad chosen sets not only result in inadequately fitted models, but also ill-organized designs. As to our experience, some of the choice of initial parameter would result in singular information matrix, and one has to be prudent to exclude design points like this in the algorithm, since the framework of OWEA relies on non-singular information matrix.

The study of designs for multinomial logistic regression model is still under development. There are many interesting yet untouched topics in this field. For example, designs for MLRM with mixed type of regression variables, or when there are higher power terms or interactions in linear components, and etc. We hope our work can trigger more research in these topics.

CHAPTER 4

PROOFS

4.1 L Matrix

$$
\mathbf{L}_{baseline} = \begin{pmatrix}
1 & & & & 0 \\
& 1 & & & 0 \\
& & \ddots & & \vdots \\
& & & 1 & 0 \\
0 & 0 & \dots & 0 & 1 \\
0 & 0 & \dots & 0 & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \dots & 0 & 1 \\
1 & 1 & \dots & 1 & 1
\end{pmatrix}
\qquad
\mathbf{L}_{cumulative} = \begin{pmatrix}
1 & 0 & \dots & 0 & 0 \\
1 & 1 & 0 & \dots & 0 \\
\vdots & \vdots & \ddots & & \vdots \\
1 & 1 & \dots & 1 & 0 \\
0 & 1 & \dots & 1 & 1 \\
0 & 0 & 1 & \dots & 1 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \dots & 0 & 1 \\
1 & 1 & \dots & 1 & 1
\end{pmatrix}
\tag{4.1}
$$

$$
\mathbf{L}_{continuation} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & \dots & \dots & 1 \\ 0 & 0 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix} \qquad \mathbf{L}_{adjacent} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & & & \\ 0 & & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}
$$

4.2 Proof of Theorems 2

Proof. We only provide proofs for baseline link and cumulative link. For continuation ratio and adjacent link, the arguments are similar to that of cumulative link. Since information matrix at $J = 3, p = 1$ is simple, we work on them directly.

The main task to identify the complete class. For a complete class Ξ, define two designs $\xi \notin \Xi$ and $\tilde{\xi} \in \Xi$ on design space χ,

$$
\begin{aligned}
\xi &= \{(c_i, w_i), c_i \in \chi, \sum_{i=1}^{m} w_i = 1\} \\
\tilde{\xi} &= \{(\tilde{c}_i, \tilde{w}_i), \tilde{c}_i \in \chi, \sum_{i=1}^{k} \tilde{w}_i = 1\}
\end{aligned} \tag{4.2}
$$

Part I. For baseline link, information matrix at design point x is

$$I = \begin{pmatrix} \frac{e^{\alpha_1+\beta x}(1+e^{\alpha_2+\beta x})}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & -\frac{e^{\alpha_1+\alpha_2+2\beta x}}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & \frac{e^{\alpha_1+\beta x}x}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} \\ -\frac{e^{\alpha_1+\alpha_2+2\beta x}}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & \frac{e^{\alpha_2+\beta x}(1+e^{\alpha_1+\beta x})}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & \frac{e^{\alpha_2+\beta x}x}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} \\ \frac{e^{\alpha_1+\beta x}x}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & \frac{e^{\alpha_2+\beta x}x}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} & \frac{e^{\beta x}(e^{\alpha_1}+e^{\alpha_2})x^2}{(1+e^{\alpha_1+\beta x}+e^{\alpha_2+\beta x})^2} \end{pmatrix} \tag{4.3}$$

To prove the complete class result,

Step 1: (Selection) Let $c = \beta x$ (where $\beta \neq 0$), then there is a bijection between x and c, and $x = c/\beta$. Among the first two columns, select the following set as maximal linear independent nonconstant functions.

$$\Psi_1(c) = \frac{e^{\alpha_1+c}(1+e^{\alpha_2+c})}{(1+e^{\alpha_1+c}+e^{\alpha_2+c})^2} \tag{4.4}$$

$$\Psi_2(c) = -\frac{e^{\alpha_1+\alpha_2+2c}}{(1+e^{\alpha_1+c}+e^{\alpha_2+c})^2} \tag{4.5}$$

$$\Psi_3(c) = \frac{e^{\alpha_1+c}c}{\beta(1+e^{\alpha_1+c}+e^{\alpha_2+c})^2} \tag{4.6}$$

and let

$$\Psi_4(c) = \frac{c^2e^c(e^{\alpha_1}+e^{\alpha_2})}{\beta^2(1+e^{\alpha_1+c}+e^{\alpha_2+c})^2} \tag{4.7}$$

Here let $g(c) = (1+e^{\alpha_1+c}+e^{\alpha_2+c})^2$, and inequality $g(c) > 0$ holds on its domain. Such an arrangement is due to the fact that B1 block in (4.3) only has two linear independent functions in terms of c.

Step 2: (Simplification) The task is to show the following system for any two designs ξ and $\tilde{\xi}$ in (4.2),

$$\begin{aligned}
\sum_{i=1}^{m} w_i \Psi_1(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_1(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_2(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_2(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_3(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_3(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_4(c_i) &\leq \sum_{i=1}^{k} \tilde{w}_i \Psi_4(\tilde{c}_i)
\end{aligned} \tag{4.8}$$

and it is sufficient to show

$$\begin{aligned}
&\{1, \Psi_1, \Psi_2, \Psi_3\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4\} \quad \text{are Chebyshev Systems} \\
&\{1, \Psi_1, \Psi_2, \Psi_3\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, -\Psi_4\} \quad \text{are Chebyshev Systems.}
\end{aligned} \tag{4.9}$$

Due to the existence of denominators in $\Psi(c)$, the recursive construction of $F(c)$ described in Theorem 2 in (Yang and Stufken, 2012) are expected to be cumbersome and resultant $F(c)$ function can be rather complicated. Instead, we perform a series simplifications which preserve either the equality in (4.8) or the Chebyshev System in (4.9) but with more simple functions.

First, we omit the '$-$' sign in Ψ_2 and β in Ψ_3 which does not change the equality in (4.8). Then multiply all Ψ functions including the constant $\Psi_0 = 1$ by the denominator and conduct row or column operations that does not change the sign of matrix determinant. At last we get rid of positive constants like $e^{\alpha_1}, e^{\alpha_2}$ and β^2 which preserve the Chebyshev System. Eventually, a set of Ψ functions is simplified to

$$\{1, e^c, e^{2c}, ce^c, c^2e^c\} \tag{4.10}$$

To show (4.9) is equivalent to verifying either those following claims hold

$$\{1, e^c, e^{2c}, ce^c\} \quad \text{and} \quad \{1, e^c, e^{2c}, ce^c, c^2e^c\} \quad \text{are Chebyshev Systems} \tag{4.11}$$

$$\{1, e^c, e^{2c}, ce^c\} \quad \text{and} \quad \{1, e^c, e^{2c}, ce^c, -c^2e^c\} \quad \text{are Chebyshev Systems} \tag{4.12}$$

Step 3: (Calculation) The sequence of $f_{\ell,\ell}$ functions can be easily calculated according to Theorem 3 of (Yang and Stufken, 2012). Here $f_{11} = e^c, f_{22} = 2e^c, f_{33} = -e^c/2, f_{44} = 2$, and $F(c) = \prod_{i=1}^4 f_{ii}(c) = -2e^c < 0$. Then designs with at most 2 support points form a complete class is a direct consequence of case (b) of Theorem 2 in (Yang and Stufken, 2012).

Part II. For cumulative link, the information matrix at support point x is

$$I = \begin{pmatrix} \frac{e^{\alpha_1+\alpha_2+\beta x}}{(e^{\alpha_2}-e^{\alpha_1})(1+e^{\alpha_1+\beta x})^2} & -\frac{e^{\alpha_1+\alpha_2+\beta x}}{(e^{\alpha_2}-e^{\alpha_1})(1+e^{\alpha_1+\beta x})(1+e^{\alpha_2+\beta x})} & \frac{xe^{\alpha_1+\alpha_2+2\beta x}}{(1+e^{\alpha_1+\beta x})^2(1+e^{\alpha_2+\beta x})} \\ -\frac{e^{\alpha_1+\alpha_2+\beta x}}{(e^{\alpha_2}-e^{\alpha_1})(1+e^{\alpha_1+\beta x})(1+e^{\alpha_2+\beta x})} & \frac{e^{2\alpha_2+\beta x}}{(e^{\alpha_2}-e^{\alpha_1})(1+e^{\alpha_2+\beta x})^2} & \frac{xe^{\alpha_2+\beta x}}{(1+e^{\alpha_1+\beta x})(1+e^{\alpha_2+\beta x})^2} \\ \frac{xe^{\alpha_1+\alpha_2+2\beta x}}{(1+e^{\alpha_1+\beta x})^2(1+e^{\alpha_2+\beta x})} & \frac{xe^{\alpha_2+\beta x}}{(1+e^{\alpha_1+\beta x})(1+e^{\alpha_2+\beta x})^2} & \frac{x^2e^{\alpha_2+\beta x}(1+2e^{\alpha_1+\beta x}+e^{\alpha_1+\alpha_2+2\beta x})}{(1+e^{\alpha_1+\beta x})^2(1+e^{\alpha_2+\beta x})^2} \end{pmatrix} \tag{4.13}$$

Following the same steps:

Step 1: Let $c = \beta x$, we propose the assignments of functions:

$$\Psi_1 = \frac{e^{\alpha_1+\alpha_2+c}}{(e^{\alpha_2} - e^{\alpha_1})(1 + e^{\alpha_1+c})^2} \tag{4.14}$$

$$\Psi_2 = -\frac{e^{\alpha_1+\alpha_2+c}}{(e^{\alpha_2} - e^{\alpha_1})(1 + e^{\alpha_1+c})(1 + e^{\alpha_2+c})} \tag{4.15}$$

$$\Psi_3 = \frac{e^{2\alpha_2+c}}{(e^{\alpha_2} - e^{\alpha_1})(1 + e^{\alpha_2+c})^2} \tag{4.16}$$

$$\Psi_4 = \frac{ce^{\alpha_1+\alpha_2+2c}}{\beta(1 + e^{\alpha_1+c})^2(1 + e^{\alpha_2+c})} \tag{4.17}$$

$$\Psi_5 = \frac{ce^{\alpha_2+c}}{\beta(1 + e^{\alpha_1+c})(1 + e^{\alpha_2+c})^2} \tag{4.18}$$

$$\Psi_6 = \frac{c^2e^{\alpha_2+c}(1 + 2e^{\alpha_1+c} + e^{\alpha_1+\alpha_2+2c})}{\beta^2(1 + e^{\alpha_1+c})^2(1 + e^{\alpha_2+c})^2} \tag{4.19}$$

One can easily verify functions Ψ_1 to Ψ_5 form the set of maximal linear independent nonconstant functions among the first two columns $I(\boldsymbol{\theta})$.

Step 2. To show

$$\begin{aligned}
\sum_{i=1}^{m} w_i \Psi_1(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_1(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_2(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_2(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_3(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_3(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_4(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_4(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_5(c_i) &= \sum_{i=1}^{k} \tilde{w}_i \Psi_5(\tilde{c}_i) \\
\sum_{i=1}^{m} w_i \Psi_6(c_i) &\leq \sum_{i=1}^{k} \tilde{w}_i \Psi_6(\tilde{c}_i)
\end{aligned} \tag{4.20}$$

or its sufficient condition

$$\begin{aligned}
&\{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5, \Psi_6\} \quad \text{are Chebyshev Systems} \\
&\{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4, \Psi_5, -\Psi_6\} \quad \text{are Chebyshev Systems}
\end{aligned} \tag{4.21}$$

We did similar simplifications as introduced in part I and it turns out one needs to work with the following set of functions,

$$\{(1+e^{\alpha_1+c})^2(1+e^{\alpha_2+c})^2, e^c(1+e^{\alpha_2+c})^2, e^c(1+e^{\alpha_1+c})(1+e^{\alpha_2+c}), e^c(1+e^{\alpha_1+c})^2, ce^c \quad (1+e^{\alpha_1+c}),$$
$$\tag{4.22}$$
$$ze^{2c}(1+e^{\alpha_2+c}), c^2e^c(1+2e^{\alpha_1+c}+e^{\alpha_1+\alpha_2+2c})\}$$

However, the major difficulty is the resultant $F(c)$ function is still way too complicated from which one can draw conclusions regarding complete class. Instead, the best we can do so far is to investigate Chebyshev System on a augmented set of linear independent functions:

$$\{1, e^c, e^{2c}, e^{3c}, e^{4c}, ce^c, ce^{2c}, ce^{3c}, c^2e^c(1 + 2e^{\alpha_1+c} + e^{\alpha_1+\alpha_2+2c})\} \tag{4.23}$$

That is, we managed to reach to check the following,

$$\begin{aligned}&\{1, e^c, e^{2c}, e^{3c}, e^{4c}, ce^c, ce^{2c}, ce^{3c}\} \quad \text{and}\\ &\{1, e^c, e^{2c}, e^{3c}, e^{4c}, ce^c, ce^{2c}, ce^{3c}, c^2e^c(1 + 2e^{\alpha_1+c} + e^{\alpha_1+\alpha_2+2c})\} \quad \text{are Chebyshev Systems}\end{aligned} \tag{4.24}$$

or

$$\begin{aligned}&\{1, e^c, e^{2c}, e^{3c}, e^{4c}, ce^c, ce^{2c}, ce^{3c}\} \quad \text{and}\\ &\{1, e^c, e^{2c}, e^{3c}, e^{4c}, ce^c, ce^{2c}, ce^{3c}, -c^2e^c(1 + 2e^{\alpha_1+c} + e^{\alpha_1+\alpha_2+2c})\} \quad \text{are Chebyshev Systems}\end{aligned} \tag{4.25}$$

Step 3 Direct calculation shows $F(c) = \prod_{\ell=1}^{8} f_{\ell\ell} = -8e^c(3 + 2e^{\alpha_1+c} + 3e^{\alpha_1+\alpha_2+2c}) < 0$. Then according to case (b) of Theorem 2 in (Yang and Stufken, 2012), designs with at most 4 points form a complete class. □

4.3 Proof of Theorem 3

Proof. For baseline link, by Lemma 1, Information matrix at support point x is summarized blockwise.

The B1 block,

$$M_{ij} = \begin{cases} -\frac{e^{\alpha_i+\alpha_j+2\beta x}}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\beta x}]^2}, & i \neq j \\ \frac{e^{\alpha_i+\beta x}[1+(\sum_{s=1,s\neq j}^{J-1} e^{\alpha_s})e^{\beta x}]}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\beta x}]^2}, & i = j. \end{cases} \tag{4.26}$$

B2 block is a row vector and its jth entry is

$$xM_{\cdot j} = x\left\{\frac{e^{\alpha_j+\beta x}[1+(\sum_{s=1,s\neq j}^{J-1} e^{\alpha_s})e^{\beta x}]}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\beta x}]^2} - \sum_{i=1,i\neq j}^{J-1}\frac{e^{\alpha_i+\alpha_j+2\beta x}}{[1+(\sum_{s=1}^{J-1} e^{\alpha_s})e^{\beta x}]^2}\right\} \tag{4.27}$$

$$= \frac{xe^{\alpha_j+\beta x}}{[1+(\sum_{j=1}^{J-1} e^{\alpha_j})e^{\beta x}]^2}, \quad \text{for } j = 1, ..., J-1 \tag{4.28}$$

and B3 block is a scalar,

$$x^2 M_{\cdot\cdot} = \frac{x^2 e^{\beta x}\sum_{s=1}^{J-1} e^{\alpha_s}}{[1+(\sum_{j=1}^{J-1} e^{\alpha_j})e^{\beta x}]^2} \tag{4.29}$$

In order to select a maximal set of linear independent nonconstant functions, we first introduce the following lemma that summaries relevant property of information matrix.

Lemma 5. *Of information matrix for baseline proportional model with J categories, its B1 block only has two linear independent functions and its B2 block only has one linear independent function.*

The proof is evident in the calculations of B1 $\sim$ B3 blocks.

Following standard steps.

Step 1: Let $c = \beta x$ (where $\beta \neq 0$), then there is a bijection between x and c, and $x = c/\beta$.

$$\Psi_1 = \frac{e^c}{[1 + (\sum_{s=1}^{J-1} e^{\alpha_s})e^c]^2} \tag{4.30}$$

$$\Psi_2 = \frac{e^{2c}}{[1 + (\sum_{s=1}^{J-1} e^{\alpha_s})e^c]^2} \tag{4.31}$$

$$\Psi_3 = \frac{ce^c}{\beta[1 + (\sum_{j=1}^{J-1} e^{\alpha_j})e^c]^2} \tag{4.32}$$

and let

$$\Psi_4 = x^2 M_{\cdot\cdot} = \frac{c^2 e^c \sum_{s=1}^{J-1} e^{\alpha_s}}{\beta^2[1 + (\sum_{j=1}^{J-1} e^{\alpha_j})e^c]^2} \tag{4.33}$$

Let $g(c) = [1 + (\sum_{j=1}^{J-1} e^{\alpha_j})e^c]^2$, and inequality $g(c) > 0$ holds on all over its domain. Then one needs to verify either of those two claims hold.

$$\{1, \Psi_1, \Psi_2, \Psi_3\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, \Psi_4\} \quad \text{are Chebyshev Systems} \tag{4.34}$$

$$\{1, \Psi_1, \Psi_2, \Psi_3\} \quad \text{and} \quad \{1, \Psi_1, \Psi_2, \Psi_3, -\Psi_4\} \quad \text{are Chebyshev Systems} \tag{4.35}$$

After simplification, it is equivalent to work on the following set of functions,

$$\{1, e^c, e^{2c}, ce^c, c^2e^c\}. \tag{4.36}$$

That is one need to verify those following claims.

$$\{1, e^c, e^{2c}, ce^c\} \quad \text{and} \quad \{1, e^c, e^{2c}, ce^c, c^2e^c\} \quad \text{are Chebyshev Systems} \tag{4.37}$$

$$\{1, e^c, e^{2c}, ce^c\} \quad \text{and} \quad \{1, e^c, e^{2c}, ce^c, -c^2e^c\} \quad \text{are Chebyshev Systems} \tag{4.38}$$

The result in Theorem 2 applies, and designs with at most 2 support points forms a complete class. □

4.4 Proof of Theorem 4

Proof. The proof is inspired by (Yang et al., 2011a). For a given design ξ, information matrix is

$$I_\xi(\theta) = n \sum_{i=1}^{m} w_i \mathbf{F}_i^{'} \mathbf{U}_i \mathbf{F}_i \tag{4.39}$$

First of all, define following weights, $r_j = \frac{V_j - x_{ij}}{V_j - U_j}$ such that

$$r_j U_j + (1 - r_j) V_j = x_{ij} \tag{4.40}$$

$$r_j U_j^2 + (1 - r_j) V_j^2 \geq x_{ij}^2 \tag{4.41}$$

$$for \quad j = 1, ..., p - 1$$

The first equality is easy to verify and the second inequality is due to the fact that function $f(x) = x^2$ is convex. Note that this is exact the lemma appears in (Yang et al., 2011a).

For arbitrary design point, say $s_i = (x_{i1}, ..., x_{i,p-1}, c_i)$, consider the following two design points, $\tilde{s}_{i1} = (U_1, x_{i2}, ..., x_{i,p-1}, c_i)$ and $\tilde{s}_{i2} = (V_1, x_{i2}, ..., x_{i,p-1}, c_i)$, and their design matrices are $\tilde{\mathbf{F}}_{i1}$ and $\tilde{\mathbf{F}}_{i2}$. Let $\tilde{w}_{i1} = r_1 w_i$ and $\tilde{w}_{i2} = w_i - \tilde{w}_{i1}$, then $w_i \mathbf{F}_i^{'} \mathbf{U}_i \mathbf{F}_i$ and $\sum_{\ell=1}^{2} \tilde{w}_{i\ell} \tilde{\mathbf{F}}_{i\ell}^{'} \tilde{\mathbf{U}}_{i\ell} \tilde{\mathbf{F}}_{i\ell}$ are exactly the same except the first diagonal element in their B3 blocks. Here $\tilde{\mathbf{U}}$ is matrix $\mathbf{U}$ evaluated at $\tilde{s}_{i\ell}$.

This is true due to two facts. First the (4.40). Second, entries in B1, B2 as well as off-diagonal ones in B3 are linear in x_{i1}, and only the first diagonal components in B3 block is quadratic in x_{i1}. As a result,

$$w_i \mathbf{F}_i^{'} \mathbf{U}_i \mathbf{F}_i \leq \sum_{\ell=1}^{2} \tilde{w}_{i\ell} \tilde{\mathbf{F}}_{i\ell}^{'} \tilde{\mathbf{U}}_{i\ell} \tilde{\mathbf{F}}_{i\ell} \tag{4.42}$$

Repeat the procedures until $x_{i,p-1}$, and we have the following

$$w_i \mathbf{F}_i^{'} \mathbf{U}_i \mathbf{F}_i \leq \sum_{\ell=1}^{2^{p-1}} \tilde{w}_{i\ell} \tilde{\mathbf{F}}_{i\ell}^{'} \tilde{\mathbf{U}}_{i\ell} \tilde{\mathbf{F}}_{i\ell} \tag{4.43}$$

Note that the right hand side of (4.43) only depends on c_i, and they have the same set of linear independent nonconstant functions. Then following Theorem 3, there exists two points $\tilde{c}_{i1}$ and $\tilde{c}_{i2}$ such that so that

$$I_\xi(\boldsymbol{\theta}) \leq \sum_{i=1}^{m} \sum_{\ell=1}^{2^{p-1}} \tilde{w}_{i\ell} \tilde{\mathbf{F}}_{i\ell}^{'} \tilde{\mathbf{U}}_i \tilde{\mathbf{F}}_{i\ell} \leq \sum_{i=1}^{2} \tilde{w}_i \sum_{\ell=1}^{2^{p-1}} \tilde{w}_{i\ell} \tilde{\mathbf{F}}_{i\ell}^{'} \tilde{\mathbf{U}}_{i\ell} \tilde{\mathbf{F}}_{i\ell} \tag{4.44}$$

That is the complete class consists of two equivalent classes of 2^p design points in total. $\square$

CHAPTER 5

A GENERAL AND EFFICIENT ALGORITHM BASED FRAMEWORK OF OPTIMAL DESIGNS

Linear models are ubiquitous in scientific research. Among linear models, the family of crossover models have a wide spectrum of applications in multiple disciplines, such as pharmaceutical study and clinical trials (see (Bennion et al., 2002; Wang et al., 2016)), nutrition and dietetics (see (Baer et al., 2004; Harris and Raynor, 2017)), education (see (Jayaratne et al., 2013; Prunuske et al., 2016)), psychology (see (Donovan et al., 2000; Lam and Kahler, 2017)), and etc. We refer to (Stufken, 1996), (Bose and Dey, 2009) and (Jones and Kenward, 2014) for authoritative and comprehensive review of both literature and applications.

Experimental design for linear model plays a crucial role in scientific research since an optimal or highly efficient design organizes resources economically and wisely so that the subsequent analysis is reliable as well as reproducible. In a crossover design, subject receives a sequence of treatments at successive time periods, of which the sequence is determined by the design and randomly assigned to each subject. The most important advantage is its cost-effectiveness. To be specific, subject receives multiple treatments so that each single of them serves as their own control, of which the influence from the confounding factors can be reduced or eliminated. In addition, it usually requires much less subjects to achieve the same precision in parameter estimation comparing to non-crossover study because each subject generates multiple outcomes, which makes crossover design statistically efficient. This is especially important in the case where experimental subjects are scarce or expensive. While these merits make crossover design attractive, some potential issues draw extra cautions from experimenters. The most common concern lies in the fact that the effects of treatments from previous periods may be carried over time to time and have impact on the outcomes at successive periods. In the presence of such carryover effects, alternative remedies have to be considered. One of them is to insert a time gap between two consecutive time periods, which is believed sufficiently long for the carryover effects to be washed out. However, answers to the question of 'how long is a wash out period enough' subject to experts opinion, and whether the carryover effects vanish completely or still persist after wash out period remains unknown. It is therefore necessary to use a design for efficient estimations of parameters under statistical model with carryover effects. There are variants of crossover models, of which differs from the way of modeling carryover effects. Section 5.1 provides details of two of them, and we refer to (Yang and Stufken, 2008) for an extended list of models.

Theoretical results on optimal crossover design have been developed in recent four decades. Selected contributions to this field, but not limited to, are (Hedayat and Afsarinejad, 1975; Hedayat and Afsarinejad, 1978; Cheng and Wu, 1980; Laska et al., 1983; Kunert, 1984; Hedayat and Zhao, 1990; Kushner, 1997a; Kushner, 1997b; Kushner, 1998; Kunert and Stufken, 2002; Bose and Mukherjee, 2003; Hedayat and Stufken, 2003; Hedayat and Yang, 2004; Yang and Stufken, 2008) and etc. Note that The majority of researches in literature deal with the optimal design problem while assuming subject effects are fixed. (Hedayat and Yang, 2005; Hedayat et al., 2006; Hedayat and Zheng, 2010) consider the similar design problem under the assumption that subject effects are random. On other direction, for example, (Park et al., 2011; Bailey and Druilhet, 2014) provide result on crossover model with interactions. Recently, (Kempton et al., 2001; Bailey and Kunert, 2006; Bose and Stufken, 2007) develop tools for the model with carryover effects that are proportional to their direct effects. (Zheng, 2013a) adopt Bayesian design framework and establish Kiefer's type general equivalence theorem which are used later in an algorithm for approximate locally optimal design.

Despite the rich literature for crossover design, all these aforementioned work are ideal for the case where there is no subject dropout. However, on the contrary, it is quit common that subject enrolled in a crossover study would dropout at interim periods. As (Low et al., 1999) pointed out "Experience suggests that a dropout rate of between 5% and 10% is not uncommon and, in some areas, can be as high as 25%". The consequences are not trivial because the realization of an optimal crossover design is not optimal, or even worse. For instance, dropout could be so frequent that data collected would result in less efficient estimations or even render parameter inestimable. Although remedies for missing values in crossover study have been established at analysis stage, for example see (Matthews and Henderson, 2013), we believe this issue is so common that it is no doubt to be prudently considered at design stage. Along this direction, relevant work includes, (Matthews, 1988; Low et al., 1999; Godolphin, 2004; Majumdar et al., 2008; Bose and Bagchi, 2008; Zhao and Majumdar, 2012). Most of the papers focus on preserving the symmetric structure of design when subject dropout presents. In (Zheng, 2013b), Kushner's type of linear equations (see (Kushner, 1997b)) are developed as necessary and sufficient conditions for a design being universal ϕ_1 optimal where ϕ_1 is defined as a new surrogate objective function considering dropout. He extends previous work and reach to a unified result that is applicable to any configurations of experiment. Optimal or efficient designs are obtained by either solving linear equations or integer optimization problem modified from equation systems.

Although elegant results are available for crossover designs, they are not ready for use in general. For some of the combinations of periods, treatments, and total observations, optimal or efficient designs are available in literature. Under most circumstances, there is no such designs that are ready for use and practitioners have to construct them by utilizing existing theories. In addition, relevant theory may not even exist for certain experiment configurations, and this is the major obstacle to use optimal or efficient designs. Hence, a general and efficient algorithm, with easy implementation via computer program, is necessary.

In this article, we develop a general and efficient algorithm framework for crossover designs and later extend it to other models. It works for any configurations of designs as long as the information matrix satisfies certain assumption, and the resultant designs are efficient compared to designs in literature. Lower bound of efficiency is provided in the absence of optimal exact designs from literature and it turns out they are satisfactory. In addition, the computing time is fast.

The rest of this chapter is organized as follows. Section 5.1 briefly introduces necessary notations, models and assumptions, and their information matrices. Some key concepts and results in approximate design theory can be found in section 5.2. Section 5.3 gives details on

OWEA algorithm. Numerical results are presented in section 5.4, and a short discussion is in section 5.5.

5.1 Optimal Designs and Information Matrix

An exact design with n runs is typically denoted by

$$d = \{(s_i, n_i)|s_i \in \chi, \sum_{i=1}^{m} n_i = n, n_i > 0\} \tag{5.1}$$

where s_i is design point, χ is a set of all possible s_i called *design space*, m is the number of distinct design points, and n_i is positive integer that stands for the repetition of the ith design point. Given design (5.1), the corresponding information matrix for parameter vector, say θ, is

$$I(\theta) = \sum_{i=1}^{m} n_i I_i \tag{5.2}$$

where I_i is the information matrix for single design point s_i. An optimal exact design is then to allocate design points as well as their repetitions so that resultant design minimizes an objective function in terms of information matrix (5.2), usually denoted by $\Phi(I)$. Objective functions vary upon different optimal criteria while sharing some common mathematical properties, for example, convexity. Detailed description is provided later in this section.

Searching for optimal exact design is on earth an optimization problem. Unfortunately, it is intractable in general due to the fact that n_i's in (5.1) are required to be positive integers. This discrete nature disables the usage of a wide scope of numerical algorithms that require derivatives and common available combinatorial tools only work on certain configuration of experiments. Instead, optimal designs are often studied in the context of approximate design which replaces repetition n_i by weight $w_i = n_i/n$ and assumes it is continuous in $(0, 1]$. Such relaxation on n_i makes the derivation of optimal design tractable and leads to the feasible application of many numerical algorithms. A typical approximate design and its information matrix are written as follows.

$$d = \{(s_i, w_i)|s_i \in \chi, \sum_{i=1}^{m} w_i = 1, w_i \in (0, 1]\}. \tag{5.3}$$

$$I(\theta) = n \sum_{i=1}^{m} w_i I_i \tag{5.4}$$

Note that information matrix for both type of designs can be written in a form of summation. In this article, the algorithm we implemented requires the information matrix being additive with respect to design points. When this property is not satisfied, alternative way need to be sought for.

The major difference between approximate design and exact design is that approximate design does not depends on number of total runs, i.e. n, and so is its information matrix. In (5.3), m stands for number of total distinct design points, and it is free of n. Although there is a n in (5.4), it just acts as a constant and subsequent optimization only involves (s_i, w_i, m). The difference is critical because the dimension of (5.2) is fixed regardless of the value of n whereas (5.4) may have varying dimensions in both derivations or numerical searching even though n

is preset. An example is discussed in next section, and extra work has to be done in order to apply algorithms.

Notice that approximate designs may not ready to be implemented as long as it is rounded into an exact design. If $n \times w_i$ are all integers, the optimal exact design is instant. Otherwise extra cautions are necessary in rounding for the sake of obtaining an optimal or efficient exact design. In this article, we implemented a simple strategy of rounding: first multiply weights by n and round them to nearest integers; then do an one-to-one exchange between current design and design space by traversing all support points in current design until there is no more significant improvement in objective function. It turns out in later sections that we are able to obtain highly efficient exact designs. For more relevant concerns on rounding designs, we refer to (Pukelsheim and Rieder, 1992).

To evaluate the quality of rounded exact design, efficiency is introduced as the metric. There are at least three ways of evaluating efficiency. When the goal is to compare two exact designs, relative efficiency are calculated. Given two arbitrary designs ξ_1 and ξ_2, $E_1 = \Phi(\xi_1)/\Phi(\xi_2)$ is the relative efficiency of ξ_1 to ξ_2 and the value E_1 could be any positive real number. In addition, ξ_1 can be an optimal exact design available in the literature. When optimal exact designs are not available, efficiency is evaluated in other approach. Suppose ξ^* is an optimal approximate design which minimizes objective function and is then rounded to an exact design ξ. Denote their corresponding information matrices as I_{ξ^*} and I_ξ respectively. The efficiency of ξ is calculated as $E_2 = \Phi(I_{\xi^*})/\Phi(I_\xi)$ where $0 < E_2 \leq 1$. It can be interpreted as how much the optimality is preserved in after rounding or the lower bound of relative efficiency when ξ is compared with arbitrary designs. For some cases of proportional model and interference model in section 5.4, we report E_2 type of efficiency. Other than those aforementioned cases, it is possible that some of the objective functions are difficult or even unable to be accessed to. In an effort to remedy this issue, we choose a surrogate objective function for which a optimal design is sought and lower bound of efficiency can be derived.

The derivation of lower bound for efficiency depends on particular cases. The following example is how we derive lower bounds for crossover model under subject dropout which is described in next section. Under its setting, information matrix is random due to the modeling of dropout mechanism. The intuitively reasonable objective function, $E\Phi(I)$, is not easy to deal with, and $\Phi(E(I))$ is chosen as an feasible replacement. Here the operator $'E'$ means 'expectation'. Suppose we obtain an optimal approximate design ξ^* by minimizing $\Phi(E(I))$, and then rounded to exact design ξ_1. Primarily, efficiency for ξ_1 should be evaluated by $E\Phi(\xi_{opt})/E\Phi(\xi_1)$, where approximate optimal design $\xi_{opt} = \underset{\xi}{argmin} E\Phi(I_\xi)$. However, the efficiency is not accessible because ξ_{opt} is unknown. Nevertheless, one is still able to obtain the lower bound by the following lemma. This lemma is also available in (Zheng, 2013b).

Lemma 6. *For crossover design under subject dropout, define the following approximate designs*

$$\xi_{opt} = \underset{\xi}{argmin} E\Phi(I_\xi) \tag{5.5}$$

$$\xi^* = \underset{\xi}{argmin} \Phi(E(I_\xi)) \tag{5.6}$$

For any arbitrary exact design ξ_1*, the lower bound for its efficiency is* $\Phi(E(I_{\xi^*}))/E\Phi(I_{\xi_1})$.

Proof. Note that the efficiency and all three designs satisfy the following inequalities.

$$\frac{E\Phi(I_{\xi_{opt}})}{E\Phi(I_{\xi_1})} \geq \frac{\Phi(E(I_{\xi_{opt}}))}{E\Phi(I_{\xi_1})} \geq \frac{\Phi(E(I_{\xi^*}))}{E\Phi(I_{\xi_1})} \tag{5.7}$$

where $\Phi(\cdot)$ is a convex function. The first $' \geq '$ is due to Jensen's inequality with equality holds when ξ_{opt} only has one points or Φ is linear. The second $' \geq '$ is because of (5.5), and equality holds when $I_{\xi_{opt}}$ and I_{ξ^*} are the identical. □

5.1.1 Crossover Model with Subject Dropout

The ways of modeling responses from crossover study are not unique, and we stick to the one with first order carryover effects. Typically, a individual response of crossover design d with n subjects, p periods and t treatments is modeled as

$$y_{ij} = \mu + \zeta_i + \pi_j + \tau_{d(i,j)} + \gamma_{d(i,j-1)} + \epsilon_{ij}; \tag{5.8}$$

$$i = 1, ..., n; \quad j = 1, ..., p.$$

where y_{ij} is the response of jth period for ith experimental subject, and ϵ_{ij}'s are independent normally distributed random errors with mean 0 and variance $\sigma^2(>0)$. μ is general mean, ζ_i is the effect from ith experimental subject, π_j stands for the effects from jth period, $d(i,j)$ denote the treatment assignment of jth period for ith subject from design d, $\tau_d(i,j)$ is the treatment effect from $d(i,j)$, and $\gamma_{d(i,j-1)}$ is the carryover effect due to treatment $d(i,j-1)$ where $\gamma_{d(i,0)}$ is set to 0 by convention.

If the collection of responses from exact design d are gathered into a vector

$$Y_d = (y_{11}, y_{12}, ..., y_{1p}, y_{21}, ..., y_{np})^{'}$$

then model (5.8) can be written in terms of matrices as follows, which serves as the start of deriving information matrix.

$$Y_d = 1_{np}\mu + U\zeta + Z\pi + T\tau + R\gamma + \epsilon \tag{5.9}$$

where ϵ is a vector of independent errors with mean $\mathbf{0}$ and variance covariance matrix $\sigma^2 I_{np}$, and $\sigma^2 > 0$, $\pi = (\pi_1, ..., \pi_p)^{'}$, $\zeta = (\zeta_1, ..., \zeta_n)^{'}$, $\tau = (\tau_1, ..., \tau_t)^{'}$, $\gamma = (\gamma_1, ..., \gamma_t)^{'}$, $U = I_n \otimes \mathbf{1}_p = (U_1^{'}, ..., U_n^{'})^{'}$, $Z = 1_n \otimes I_p = (Z_1^{'}, ..., Z_n^{'})^{'}$, $T = (T_1^{'}, ..., T_n^{'})^{'}$ and $F = (F_1^{'}, ..., F_n^{'})^{'}$. Here U_i is $p \times n$ incidence matrices for subjects, Z_i is $p \times p$ subject to period incidence matrix, T_i and F_i are $p \times t$ period to treatment incidence matrices for ith subject depends on design d, I_n is identity matrix of dimension n and $\mathbf{1}_p$ stands for a $p \times 1$ vector of 1's.

Then the information matrix for full parameter vector $\Theta = (\zeta', \pi', \tau', \gamma')'$ is

$$I(\Theta) = \begin{pmatrix} U & Z & T & R \end{pmatrix}' \begin{pmatrix} U & Z & T & R \end{pmatrix} \tag{5.10}$$

$$= \sum_{i=1}^{n} \begin{pmatrix} U_i & Z_i & T_i & R_i \end{pmatrix}' \begin{pmatrix} U_i & Z_i & T_i & R_i \end{pmatrix} \tag{5.11}$$

$$= \sum_{i=1}^{n} I_i = \sum_{i=1}^{m} n_i I_i$$

where I_i is the information matrix for ith design point(treatment sequence).

An optimal exact design is then to choose a collection of sequences so that it optimizes an objective function related to (5.10). In this article, we obtained optimal approximate design from numerical algorithm and carefully round it to exact design which either efficient or optimal. However two issues regarding information matrix has to be settled prior to the implementation of algorithm. First, usually information matrix for direct treatment is the target, however it is not additive with respect to design point which does not fit for the prerequisite for the algorithm we introduced. Other than this, although information matrix for full parameter is additive, its dimension is changing all the time because of the inclusion of subject effects ζ iterative nature of numerical algorithm. Optimizing on an information matrix of varying dimension is problematic. Therefore, to fit the crossover design problem to our framework, we excluded subject effect from (5.10). In fact, in literature, subject effects are either treated as nuisance or assumed to be random, neither of them indicates the subject effects are the goal of subsequent analysis.

Direct calculation yields the information matrix for marginal parameter vector without subject effects, $\theta = (\pi^{'}, \tau^{'}, \gamma^{'})^{'}$,

$$I(\theta) = \begin{pmatrix} Z & T & R \end{pmatrix}^{'} pr^{\perp}(U) \begin{pmatrix} Z & T & R \end{pmatrix} \tag{5.12}$$

$$= \begin{pmatrix} Z & T & R \end{pmatrix}^{'} [I_{np} - \frac{1}{p} I_n \otimes 1_p 1_p^{'}] \begin{pmatrix} Z & T & R \end{pmatrix} \tag{5.13}$$

$$= \sum_{i=1}^{n} \begin{pmatrix} Z_i & T_i & R_i \end{pmatrix}^{'} [I_p - \frac{1}{p} 1_p 1_p^{'}] \begin{pmatrix} Z_i & T_i & R_i \end{pmatrix} \tag{5.14}$$

$$= \sum_{i=1}^{n} I_i = \sum_{i=1}^{m} n_i I_i = n \sum_{i=1}^{m} w_i I_i \tag{5.15}$$

where $pr^{\perp}(\cdot)$ is an orthogonal projection operator, and for any matrix X and $pr^{\perp}(X) = I - X(X^{'}X)^{-1}X^{'}$. Note that (5.12) is both fixed dimension and additive.

In the presence of dropout, following (Low et al., 1999; Zheng, 2013b), define the dropout mechanism of a subject as $\ell = (\ell_1, ..., \ell_p)$, where ℓ_i is the probability that the longest time of stay is i, and $\sum_{i=1}^{p} \ell_i = 1$ and assume

1. Subject's dropout is independent of the choice of design as well as the outcome of design.

2. Once a subject dropped, the chance of returning is zero.

3. Subject's dropout mechanisms are i.i.d.

With dropout mechanism, matrices Z_i, T_i, R_i in (5.12) depend on ℓ and therefore are random. Hence information matrix (5.12) becomes

$$I(\theta,\ell) = \sum_{i=1}^{n} \begin{pmatrix} Z_i & T_i & R_i \end{pmatrix}^{'} [M_i - M_i 1_p (1_p^{'} M_i 1_p)^{-1} 1_p^{'} M_i](Z_i, T_i, R_i) \tag{5.16}$$

$$= \sum_{i=1}^{n} I_i(\ell) = \sum_{i=1}^{m} n_i I_i(\ell) = n \sum_{i=1}^{m} w_i I_i(\ell) \tag{5.17}$$

where M_i is an indicator matrix depends on dropout mechanism ℓ defined by the following

$$M_i(\ell) = \begin{pmatrix} I_{(a_i \times a_i)} & O_1 \\ O_1^{'} & O \end{pmatrix}_{p \times p} \tag{5.18}$$

and a_i is the longest period of stay for subject i, O and O_1 are zero matrices with proper order.

5.1.2 Crossover Model with Proportional Carryover Effects

One of the variants of crossover model is to proportional model. It is believed the large direct effects lead to large carryover effects, that is, the carryover effects are proportional to its direct effects. Under this assumption, for design d with p periods and t treatments, individual outcomes are modeled as

$$y_{ij} = \mu + \zeta_i + \pi_j + \tau_{d(i,j)} + \lambda \tau_{d(i,j-1)} + \epsilon_{ij}; \tag{5.19}$$

$$i = 1, ..., n; \quad j = 1, ..., p.$$

where the notation are exactly the same as in (5.8), except that the additional λ is the proportion that carryover effects account for direct effects. Following similar arrangement, the proportional model can be written in matrices,

$$Y_d = \mathbf{1}_{np}\mu + U\zeta + Z\pi + T\tau + \lambda F\tau + \epsilon \tag{5.20}$$

where ϵ is a vector of independent errors with mean $\mathbf{0}$ and variance covariance matrix $I_n \otimes \Sigma$, Σ is an $p \times p$ positive definite matrix, $\pi = (\pi_1, ..., \pi_p)'$, $\zeta = (\zeta_1, ..., \zeta_n)'$, $\tau = (\tau_1, ..., \tau_t)'$, $U = I_n \otimes 1_p = (U_1', ..., U_n')'$, $Z = 1_n \otimes I_p = (Z_1', ..., Z_n')'$, $T = (T_1', ..., T_n')'$ and $F = (F_1', ..., F_n')'$. Here U_i, Z_i are incidence matrices for subjects and periods, T_i and F_i are $p \times t$ period to treatment incidence matrices for ith subject depends on design d. Therefore, following similar calculations, information matrix for partial parameters vector $\theta = (\pi', \tau', \lambda)'$ is

$$I(\theta, \lambda, \tau) = \sum_{i=1}^{n} (Z_i, T_i + \lambda F_i, F_i\tau)' [\Sigma^{-1} - \Sigma^{-1} 1_p [1_p' \Sigma^{-1} 1_p]^{-} 1_p' \Sigma^{-1}] (Z_i, T_i + \lambda F_i, F_i\tau) \tag{5.21}$$

$$= \sum_{i=1}^{n} I_i = \sum_{i=1}^{m} n_i I_i = n \sum_{i=1}^{m} w_i I_i \tag{5.22}$$

Note that proportional model is nonlinear due to the term $\lambda\tau$, and its information matrix depends on unknown parameter τ and λ.

5.1.3 Interference Model

Interference model is widely used in agricultural study to avoid the systematic bias cause by neighbor effects of block designs. Papers on this field include (Kunert and Martin, 2000; Kunert and Mersmann, 2011; Zheng, 2015) and etc.

In a study with t treatments, n total blocks of size p, an single outcome of design d, is modeled by

$$y_{ij} = \mu + \gamma_i + \tau_{d(i,j)} + \lambda_{d(i,j-1)} + \rho_{d(i,j+1)} + \epsilon_{ij} \qquad (5.23)$$

$$i = 1, ..., n; \quad j = 1, ..., p$$

where y_{ij} denotes the response from jth plot of ith block, μ is general mean, $d(i, j)$ stands for the treatment assignment of ith block and j the plot according to design d, and $\tau_{d(i,j)}$, $\lambda_{d(i,j-1)}$, and $\rho_{d(i,j+1)}$ are treatment effects from treatment itself, its left plot and right plot. By convention, we set $\lambda_{d(i,0)} = \rho_{d(i,p+1)} = 0$. Again, the responses can be gathered in a vector and modeled in terms of matrices,

$$Y_d = 1_{nk}\mu + U\gamma + T_d\tau + L_d\lambda + R_d\rho + \epsilon \qquad (5.24)$$

where $Y_d = (y_{11}, .., y_{1k}, y_{21}, ..., y_{np})$, $\gamma = (\gamma_1, ..., \gamma_n)'$, $\tau = (\tau_1, ..., \tau_t)'$, $\lambda = (\lambda_1, ..., \lambda_t)'$, $\rho = (\rho_1, ..., \rho_t)'$, $U = I_n \otimes 1_p$, $T_d = (T_1', ..., T_n')'$, $L_d = (L_1', ..., L_n')'$, $R_d = (R_1', ..., R_n')'$. Here U_i's are incidence matrices for block, T_i and L_i and R_i are $p \times t$ plot to treatment incidence matrices for ith block depends on design d. The error vector ϵ is assumed to follow $N(0, I_n \otimes \Sigma)$, where Σ is arbitrary positive definite matrix of order p. Let $\theta = (\tau', \lambda', \rho')'$ be the vector of parameters of interest. Note that block effects are excluded due to similar reason to that of subject effect for crossover model, then information matrix for θ

$$I(\theta) = \sum_{i=1}^{n} (T_i, L_i, R_i)' [\Sigma^{-1} - \Sigma^{-1} 1_p (1_p' \Sigma^{-1} 1_p)^{-1} 1_p' \Sigma^{-1}] (T_i, L_i, R_i) \tag{5.25}$$

$$= \sum_{i=1}^{n} I_i = \sum_{i=1}^{m} n_i I_i = n \sum_{i=1}^{m} w_i I_i \tag{5.26}$$

Note that when the repetition n_i is replaced by weight w_i, we switch the context from exact design to approximate design. Therefore, information matrix for all three models are free of n. In addition, they fit for the prerequisite of the algorithm introduced next.

5.2 Optimal Criteria and General Equivalence Theorem

In after-data inference, usually a differentiable function of θ, say $g(\theta)$, is of interest. The choice of g grants the flexibility on parameter vector. For example, when we target on all parameters, g is an identity map; and when comparison is the goal, one of many options of g is $g(\theta) = (\theta_1 - \theta_\nu, \theta_2 - \theta_\nu, ..., \theta_{\nu-1} - \theta_\nu))$, where ν is the length of θ. Denote $\hat{\theta}$ as the maximum likelihood estimator (MLE) of θ, then it is well known that $g(\hat{\theta})$ is also the MLE of $g(\theta)$. Under mild assumptions, delta's theorem yields the asymptotic variance covariance matrix of $g(\hat{\theta})$,

$$C_\xi(g) = \frac{\partial g}{\partial \theta'} I^{-}(\theta) (\frac{\partial g}{\partial \theta'})' \tag{5.27}$$

An optimal approximate design is the one that minimizes the objective function (denoted by Φ) in terms of (5.27). In this paper, the focus is on a unified optimality criterion introduced by (Kiefer, 1974a), namely 'Φ_p' optimality. The objective function is

$$\Phi_p(C_\xi(g)) = \left[\frac{1}{v}Tr(C_\xi(g))^p\right]^{1/p}, 0 \le p < \infty \tag{5.28}$$

Note that $\Phi_p(C_\xi(g))$ is equivalent to many prevailing optimal criteria on various values of p. When $p = 0$, $\Phi_p(C_\xi(g))$ is understood as $\lim_{p\downarrow 0}\left[\frac{1}{v}Tr(C_\xi)^p\right]^{1/p}$, which is D-optimality; when $p = 1$, we have $\Phi_p(C_\xi(g)) = Tr(C_\xi(g))$, and it is A-optimality; when $p \to \infty$, $\Phi_p(C_\xi(g))$ is equivalent to E-optimality.

For the purpose of verifying a design being optimal, the following theorem provides Kiefer's type general equivalence theorem (GET) which serves as sufficient and necessary condition.

Theorem 7. *Suppose an arbitrary design ξ with information matrix I_ξ. ξ is Φ_p optimal for $g(\theta)$ if and only if the directional derivative of Φ_p, denoted by $d_p(s, \xi)$ satisfies*

$$d_p(s, \xi) \le 0 \tag{5.29}$$

for any $s \in \chi$, with equality holds if s belongs to the support of ξ.

In addition, the $d_p(s, \xi)$ can be calculated by

$$d_p(s,\xi) = \begin{cases} Tr[(C_\xi(g))^{-1}(\frac{\partial g}{\partial \theta'})I_\xi^-(I_s - I_\xi)I_\xi^-(\frac{\partial g}{\partial \theta'})'], & p = 0 \\ (\frac{1}{v})^{1/p}Tr[(C_\xi(g))]^{1-p} \times Tr[(C_\xi(g))^{p-1}(\frac{\partial g}{\partial \theta'})I_\xi^-(I_s - I_\xi)I_\xi^-(\frac{\partial g}{\partial \theta'})'], & p > 1 \end{cases} \tag{5.30}$$

In this article, general equivalence theorem 7 is always checked prior to the claim of a design being optimal.

5.3 Algorithm

Numerical algorithms are powerful yet convenient tools for optimal designs. Existing ones are modifications of either Fedorov-Wynn algorithm (FWA, (Fedorov, 1972),(Wynn, 1970)), or multiplicative algorithm (MA, (Silvey et al., 1978)), see (Torsney, 1981; Hettich, 1983; Böhning, 1986; Harman and Pronzato, 2007; Torsney and Martín-Martín, 2009; Martín-Martín et al., 2012). (Yu, 2010) combined multiple existing designs with modifications, named 'cocktail algorithm', and achieved dramatic improvement in speed. However, all the aforementioned algorithms only focus on D-optimal designs whereas different objective of experiments requires properly chosen optimality criterion. Moreover, in the context of optimal designs for nonlinear models, multi-stage designs are necessary in some cases in order to obtain a reasonable 'guess' of unknown parameters and only (Covey-Crump and Silvey, 1970) derived two-stage designs under D- and E- optimality for a polynomial model. (Yang et al., 2013) proposed an optimal weights exchange algorithm (OWEA) for nonlinear models. It updates the support points in the same way as FWA and optimizes weights via newton's method. This is the algorithm we are going to use in this article. Besides, applications of optimization methods, which has been shown successful in many other field, emerges in finding optimal designs. One of them that worth mentioning is a meta-heuristic algorithm named particle swarm algorithm (PSO, (Kennedy and Eberhart, 1995)). Although theory on its convergence is not fully developed, it turns out work well for optimal designs, see (Chen et al., 2015; Wong et al., 2015).

The procedures of OWEA's implementation is briefly introduced in Appendix, for details, see (Yang et al., 2013). There are three tuning parameters, namely ϵ_d, ϵ_α and ϵ_0, require extra attentions. In verifying general equivalence theorem, ϵ_d needs to be non-positive for a design being optimal. As suggested by (Yang et al., 2013), a small positive number is selected due to floating errors in computer program. The other two cut-off points are simply positive numbers slightly greater than 0. Note that tuning parameters may affects the time for convergence. In all examples tested in this article, the combination of $\epsilon_d = 10^{-15}$, $\epsilon_\alpha = 10^{-6}$ and $\epsilon_0 = 10^{-10}$ work well and optimal or highly efficient designs can be found within a relatively short period of time. Besides, it is known that newton's methods is sensitive to choice of initial points and sometimes have overshooting issue. However, based on our experience, efficient designs are always found.

5.4 Examples

In this section, we implemented OWEA upon multiple models. Algorithms are programmed in R and executed on an Apple Laptop with 1.4GHz CPU and 4GB RAM. Performance metrics such as computing time as well as efficiency are summarized in tables. For convenience, hereafter, we call the design returned by OWEA algorithm *owea design*, and optimal design available from literature the *literature design*. For example, the A- optimal design obtained from OWEA is 'A-optimal owea design'.

We packed all three models as well as algorithm into an R package named **OWEA** and it is now available on CRAN. We include a demonstration of this package in Appendix. All those following numerical results can also be obtained using this package.

5.4.1 Crossover Designs with Subject Dropout

Example 1. $(t, p) = (4, 4), n = 16, \ell = (0, 0, 1/2, 1/2)$. Design d_2 derived in (Zheng, 2013b) is shown to have relatively high efficiency under a variety of optimal criteria. Under the same setting, OWEA returns the following designs under A- and D- optimality, and Table Table VI summarizes the performance of both designs. The relative efficiency is from the comparison of owea design to d_2 and lower bound of efficiency is obtained according to Lemma 6.

As shown, The A-optimal owea design has high lower bound as well as higher actual relative efficiency. Although the lower bound for D-optimal owea design is only 0.8537, its relative efficiency is as high as 0.9882. In summary, the owea designs are close to d_2 in terms of efficiency. Computing time is also reported. For this case, owea design costs 1.974 and 1.804 for A- and D- optimality respectively. Note that (Zheng, 2013b) shows the existence of universally optimal designs, which means we can obtain an exact optimal design from looking for integer solution to a linear equation system. So its computing cost is trivial.

$$A - opt = \begin{array}{cccccccccccccccc} 1&1&1&1&2&2&2&2&3&3&3&3&4&4&4&4 \\ 2&3&4&4&1&3&3&4&1&2&2&4&1&2&3&3 \\ 3&4&2&3&3&1&4&1&2&4&4&2&3&1&1&2 \\ 4&4&3&3&3&4&4&3&2&1&1&2&2&1&2&1 \end{array} \tag{5.31}$$

$$D - opt = \begin{array}{cccccccccccccccc} 1&1&1&1&2&2&2&2&3&3&3&3&4&4&4&4 \\ 2&2&3&4&1&3&4&4&1&1&2&4&1&2&3&3 \\ 3&3&2&3&3&4&1&1&4&4&4&2&2&1&1&2 \\ 4&4&2&3&3&4&3&3&2&2&4&1&2&1&1&1 \end{array} \tag{5.32}$$

Optimality	Relative Efficiency	Efficiency(Lower Bound)	Time (seconds)
A	0.9949	0.9489	1.974
D	0.9882	0.8573	1.804

TABLE VI

CROSSOVER DESIGNS WHEN T=P=4, N=16, AND DROPOUT A = (0,0,1/2,1/2)

Example 2 $(t, p) = (4, 4), n = 19, \ell = (0, 0, 1/2, 1/2)$. This is the case where Zheng's linear equation system does not have integer solution. Instead, as mentioned in (Zheng, 2013b), one can find an efficient design by solving an integer optimization problem modified from equation systems. There are many optimization solvers available online, for example, CPLEX (IBM Inc).

Hence one has to re-run the integer programming when we change the number of subjects, say from 16 to 19. In OWEA, we only need to re-do the exchange step of the algorithm. The owea designs are summarized in Table Table VII. Because there is no optimal designs available in literature, we only report lower bound of efficiency which is also obtained according to Lemma 6. Calculation shows the lower bounds for A-optimal owea design is 0.9525, which is quit high. Although the value of lower bound efficiency is only 0.8196, the actual efficiency could be higher than this. Last, computing times are 2.682 and 2.467s respectively.

$$A - opt = \begin{array}{c} 1\,1\,1\,1\,1\,2\,2\,2\,2\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4 \\ 2\,2\,2\,4\,4\,1\,1\,3\,4\,1\,2\,2\,4\,4\,1\,2\,3\,3\,3 \\ 3\,3\,3\,2\,2\,3\,4\,4\,1\,2\,4\,4\,1\,2\,3\,1\,1\,1\,2 \\ 4\,4\,4\,2\,3\,3\,3\,4\,3\,4\,1\,1\,1\,2\,2\,1\,2\,2\,1 \end{array} \tag{5.33}$$

$$D - opt = \begin{array}{c} 1\,1\,1\,1\,1\,2\,2\,2\,2\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4 \\ 2\,2\,2\,3\,3\,1\,4\,4\,4\,1\,1\,2\,4\,4\,1\,2\,3\,3\,3 \\ 3\,3\,3\,2\,4\,4\,1\,1\,3\,4\,4\,4\,2\,2\,3\,1\,1\,2\,2 \\ 4\,4\,4\,2\,2\,4\,3\,3\,3\,2\,2\,4\,1\,1\,3\,1\,1\,1\,2 \end{array} \tag{5.34}$$

Both framework of OWEA and (Zheng, 2013b) adopted a surrogate objective function and the output designs are satisfactory. In addition, both of them works for any arbitrary combination of (n, p, t) as well as dropout mechanism. For some combinations, theorems in (Zheng, 2013b) can be used to derive close form optimal design, however OWEA only generates exact design with high efficiency for crossover model with subject dropout based on our experience.

Optimality	Efficiency(Self Lower Bound)	Efficiency (Lower Bound)	Time (seconds)
A	0.9598	0.9525	2.682
D	0.9556	0.8196	2.467

TABLE VII

CROSSOVER DESIGNS WHEN T=P=4, N=19, AND DROPOUT A = (0,0,1/2,1/2)

Other than this, for arbitrary number of n, one only needs to change the value of parameters of owea instead of seeking for other tools like optimization solvers.

5.4.2 Crossover Designs with Proportional Carryover Effects

Proportional model is nonlinear. In (Zheng, 2013a), he adopted an exchangeable prior distribution of τ and optimize the prior expectation of objective function. The resultant design is symmetric and only depends on λ. (Zheng, 2013a) also provides a steepest decent algorithm for pseudo symmetric designs. The weights for symmetric blocks can be easily found and Zheng suggests to assemble symmetric blocks into highly efficient design by mimicking orthogonal arrays. Meanwhile, OWEA is able to find locally optimal designs for given values of parameters λ and τ. The locally optimal design can be regarded as a special case of Zheng's framework where the prior distribution of τ is point mass. Obviously, tools developed in (Zheng, 2013a) are applicable to more general cases. For proportional model, all the optimal designs reported are locally optimal except otherwise specified.

$\mathbf{(t, p) = (3, 3), n = 36}$. Locally optimal exact design for this case has been prove to exist. In particular, those reported owea designs (both A- and D- optimal) are equivalent to a symmetric block $\langle 123 \rangle$ defined in (Zheng, 2013a), and they are shown to have unity efficiency for A- D- T- criteria and 0.9931 under E-optimality. The computing time for this case is $0.053s$ and $0.085s$ for A- and D- optimality. We are able to recreate Zheng's design and it only takes $0.26s$ for it to identify the optimal symmetric blocks.

$$A - opt = \begin{matrix} 1\,2\,3 \\ 2\,3\,1 \\ 3\,1\,2 \end{matrix} \times 12 \tag{5.35}$$

$$D - opt = \begin{matrix} 1\,2\,3 \\ 3\,1\,2 \\ 2\,3\,1 \end{matrix} \times 12 \tag{5.36}$$

Optimality	Relative Efficiency	OWEA Time (sec)	Zheng's Time (sec)
A	1.000	0.053	0.26
D	1.000	0.085	0.26

TABLE VIII

CROSSOVER DESIGNS FOR PROPORTIONAL MODEL WHEN T=P=3, N=36

$(\mathbf{t = 4, p = 3}), \mathbf{n = 20}$. (Zheng, 2013a) shows an symmetric design with all its sequences from symmetric block that consists of treatments without repetition, for example $\langle 123 \rangle$, is A-

D- T- optimal for a mild range of λ. However it requires n being multiple of 12 to be an exact design. When there are only 20 runs allowed, universal exact optimal design does not exist. For the given configuration, owea designs are listed as follows, and the lower bound for efficiency is calculated by comparing exact design to the optimal approximate design from which it is rounded. As the result shown in Table Table IX, efficiency are all close to unity under both A- and D- optimal criteria. Note that sequences of those designs belong to the block of all distinct treatment, but they are not symmetric design. And this is why there is minor loss in efficiency.

$$A - opt = \begin{matrix} 1\,1\,1\,2\,2\,2\,2\,2\,2\,3\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4\,4 \\ 2\,3\,3\,3\,3\,4\,4\,4\,2\,2\,2\,2\,4\,4\,1\,1\,1\,2\,3\,3 \\ 3\,2\,2\,4\,4\,3\,3\,3\,4\,4\,4\,4\,1\,2\,2\,2\,2\,3\,1\,1 \end{matrix} \tag{5.37}$$

$$D - opt = \begin{matrix} 2\,2\,2\,2\,2\,2\,3\,3\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4\,4\,4 \\ 1\,3\,3\,4\,4\,4\,1\,1\,1\,2\,2\,2\,4\,1\,1\,1\,1\,2\,3\,3 \\ 4\,4\,4\,1\,1\,3\,2\,2\,2\,1\,4\,4\,1\,3\,3\,3\,3\,1\,2\,2 \end{matrix} \tag{5.38}$$

In general, if locally optimal design is the goal, one can implement a multi-stage design of which the initial stage is used for obtaining a reasonable guess of unknown parameter. Under this context, suppose the variance covariance matrix of the initial stage is C_0 with n_0 observations, and n_1 observations with covariance matrix C_1 is going to be added in the next stage, the optimal design is to minimize $\Phi_p(\frac{n_0}{n_0+n_1}C_0 + \frac{n_1}{n_0+n_1}C_1)$. This no doubt fits into the owea

Optimality	Efficiency(Lower Bound)	Time (seconds)
A	0.9973	0.299
D	0.9951	0.144

TABLE IX

CROSSOVER DESIGNS FOR PROPORTIONAL MODEL WHEN T=4,P=3, N=20

framework.

5.4.3 Designs for Interference Model

For interference model, (Zheng, 2015) developed linear equation systems as sufficient and necessary conditions for universally optimal design and provide steepest decent algorithm for identifying symmetric blocks as well as their weights. In addition, he suggest using integer optimization tools on modified linear systems for find efficient designs when optimal design is intractable. OWEA is also useful in this model. In fact, it turns out there will be less manually operations in tuning algorithms for some of the combinations of (n, t, p).

$(\mathbf{t}, \mathbf{k}) = (\mathbf{4}, \mathbf{4}), \mathbf{n} = \mathbf{10}$. As it is shown in (Kunert and Martin, 2000), the optimal weights of universal optimal design are not integers. In (Zheng, 2015)'s framework, first use steepest descend algorithm to find an approximate symmetric design, and then one needs to manually round it exact design. Some examples are provided in his paper. We listed owea designs as follows and table Table X summarizes the performance metrics. The relative efficiency of A- and D- optimal designs are 0.9880 and 0.9793 compared to Zheng's example. Computing times for owea on this case are 8.490 and 6.359 respectively. On the contrary, the steepest descend algorithm only cost 0.58, which is considerably faster.

$$A - opt = \begin{matrix} 1\,1\,1\,2\,2\,3\,3\,4\,4\,4 \\ 1\,1\,2\,2\,3\,4\,3\,2\,3\,4 \\ 3\,3\,4\,4\,1\,2\,4\,1\,1\,2 \\ 2\,3\,3\,1\,4\,2\,4\,3\,2\,1 \end{matrix} \tag{5.39}$$

$$D - opt = \begin{matrix} 1\,1\,2\,2\,2\,3\,3\,4\,4\,4 \\ 1\,4\,1\,2\,4\,1\,3\,1\,3\,4 \\ 3\,2\,3\,4\,1\,4\,2\,2\,2\,3 \\ 3\,3\,4\,4\,1\,2\,2\,3\,1\,1 \end{matrix} \tag{5.40}$$

Optimality	Relative Efficiency	Time of OWEA (sec)	Time of Zheng's Algorithm (sec)
A	0.9880	8.490	0.58
D	0.9793	6.359	0.58

TABLE X

DESIGNS FOR INTERFERENCE MODEL WHEN T=K=4, N=10

$\mathbf{(t, k) = (4, 5), n = 24}$. Universal optimal design for this case, which has been established by (Zheng, 2015) Theorem 6(ii), is a symmetric design comprises equal weights to sequences and their dual sequences, where dual sequence has the same elements as its original sequence but in reversed order. A design consists of symmetric blocks $\langle 11234 \rangle$, $\langle 12344 \rangle$, and their equivalencies are provided. Note that the owea designs are almost the same to those in (Zheng, 2015) under A- and D- optimality. Table Table XI displays the result. In comparison of efficiency, the owea designs are very close to universal optimal, with the being 0.9983 and 0.9892, where relative efficiency is by comparing with literature designs. Note that the first 12 sequences are equivalent to $\langle 11234 \rangle$ and the last 12 are equivalent to $\langle 12344 \rangle$ as well. The computing times in this case are 6.033 and 6.952. On the contrary, Zheng's steepest decent algorithm locates those two symmetric blocks in $30.12s$, which is still acceptable.

$$A - opt = \begin{matrix} 1\,1\,1\,1\,1\,1\,2\,2\,2\,2\,2\,2\,3\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4\,4 \\ 1\,1\,2\,2\,2\,4\,2\,2\,2\,3\,4\,4\,1\,2\,3\,4\,4\,4\,1\,1\,2\,2\,4\,4 \\ 4\,4\,3\,3\,3\,3\,4\,4\,4\,1\,1\,1\,4\,1\,1\,2\,2\,2\,3\,3\,1\,1\,1\,1 \\ 3\,3\,4\,4\,4\,3\,3\,3\,3\,1\,3\,3\,2\,4\,2\,1\,1\,1\,2\,2\,3\,3\,2\,2 \\ 2\,2\,4\,4\,4\,2\,1\,1\,1\,4\,3\,3\,2\,4\,4\,1\,1\,1\,2\,2\,3\,3\,3\,3 \end{matrix} \tag{5.41}$$

$$D - opt = \begin{matrix} 1\,1\,1\,1\,1\,1\,2\,2\,2\,2\,2\,2\,3\,3\,3\,3\,3\,3\,4\,4\,4\,4\,4\,4 \\ 1\,1\,1\,3\,4\,4\,1\,1\,2\,2\,4\,4\,3\,3\,3\,4\,4\,4\,1\,3\,3\,4\,4\,4 \\ 2\,2\,3\,2\,2\,2\,3\,3\,3\,3\,1\,3\,1\,1\,1\,1\,1\,1\,2\,2\,2\,2\,3\,3 \\ 4\,4\,2\,4\,3\,3\,4\,4\,1\,4\,1\,1\,4\,4\,4\,2\,2\,2\,3\,1\,1\,3\,2\,2 \\ 3\,3\,4\,4\,3\,3\,4\,4\,4\,1\,3\,1\,2\,2\,2\,2\,2\,2\,3\,1\,1\,1\,1\,1 \end{matrix} \tag{5.42}$$

Optimality	Relative Efficiency	Time of OWEA (sec)	Time of Zheng's Algorithm (sec)
A	0.9983	6.033	30.12
D	0.9892	6.952	30.12

TABLE XI

DESIGNS FOR INTERFERENCE MODELS WHEN T=4,K=5, N=24

Similar to crossover model with subject dropout, theoretical result in (Zheng, 2015) are general and powerful for interference model since it provides closed form of universally optimal design for some combinations of (t, p) and his steepest descent algorithm and the integer programming built on the linear equation system is very general. Generosity is also preserved by owea and it is faster in most cases.

5.5 Discussion

In this chapter, we provide an algorithm based framework for finding optimal/efficient crossover designs, and later extend it to interference model. Information matrix has been derived in order to apply optimal weight exchange algorithm. Result shows OWEA successfully find efficient or agree with optimal exact designs in a short period of time. The most alarming advantage of OWEA is general and convenience. It works on a myriad of models, as long as the information matrix is additive with respect to design points. From model to model, OWEA is applicable with only minor modification of information matrix. Under the same model, one only needs to change the value of case configurations, like t, p, n in those example. Note that those configurations can be arbitrary including dropout mechanism.

Optimal exact design may not tractable in some configurations of experiment. In this case, there is no unified approach of seeking for efficient exact designs. For crossover design, (Zheng, 2013b) suggests using integer optimization based on linear equation systems, whereas for proportional model or interference model, he recommend assembling symmetric blocks with references to existing symmetric structure, for example, orthogonal arrays. We reach to approximate design and then rounded it to exact design without losing too much efficiency. (Zheng, 2015) for interference model also provided the integer programming and hence it could also deal with any configuration of t,p and n. By comparison, the owea method is faster by losing very tiny efficiency sometimes. For the proportional model, the restriction to the symmetric design in (Zheng, 2013a) was due to the Bayesian framework, whereas our framework considers locally optimal designs and hence has its own flexibility.

The resultant exact designs of OWEA are shown to be either optimal or highly efficient. Although theoretical result indicates optimal designs within a subclass of pseudo symmetric designs are automatically global optimal, OWEA does not guarantee the symmetry in its output. However, based on our experience, OWEA exact designs are close to symmetric optimal designs and always have relative high efficiency or satisfactory lower bound. Moreover, OWEA works on any linear functions of unknown parameter vector. Note that all the designs in this papers focus on direct treatment effects. On other cases, for example, (Zheng, 2013a) derived corresponding results for estimating λ in proportional model, while in OWEA, one only need to change the function g.

APPENDICES

Appendix A

THE ANALYTICAL APPROACH TO COMPLETE CLASS

Deriving an optimal design for multinomial logistic regression model is to ultimately determine the support points as well as their associated weights. This is quit a complex optimization task. In general, the structure of information matrix is rather complicated and varies upon the models. Moreover, perhaps more important, it would be quit troublesome if the number of support points were large. Because there would be too many parameters to be optimized which result in either tedious or unfeasible derivations for optimal design. Nevertheless, one usually seek to take advantage of constructive information, if exists, regarding structure of optimal designs, which significantly relieves the burden from derivation.

For example, it is well known that a $(p-1)$th degree polynomial regression model needs at least p and at most $p(p+1)/2$ design points. The lower bound is to guarantee all parameters are estimable and the upper bound is the consequence of Carathéodory theorem. The significance lies in the fact that an optimal design for such a model can be based on p design points in ideal case. (Khuri et al., 2006) connected the lower bound to the findings in (de la Garza, 1954), and name it the 'de la Garza phenomenon'. In a formal representation, for a $(p-1)$th degree polynomial regression model with p parameters and $n(>p)$ runs, there exists a subclass of design, say Ξ, with p distinct design points. For any arbitrary design $\xi \notin \Xi$, there exists a

Appendix A (Continued)

design $\tilde{\xi} \in \Xi$ such that their information matrices $I_{\tilde{\xi}}$ dominates I_ξ in Lowner ordering, which is equivalent to $I_{\tilde{\xi}} - I_\xi$ is non-negative definite.

Likewise, such subclass of designs for nonlinear model is identified by an analytical approach in (Yang and Stufken, 2009; Yang, 2010; Yang et al., 2011a; Dette and Melas, 2011; Yang and Stufken, 2012). The primary goal is to identify a subclass of designs with simple structure, called *complete class* in cited papers, say Ξ, for a nonlinear model with given design space. Then for any arbitrary design $\xi \notin \Xi$, there always exist a design $\tilde{\xi} \in \Xi$ such that $I_{\tilde{\xi}}$ dominates I_ξ in Lowner ordering, that is

$$I_{\tilde{\xi}} \geq I_\xi. \tag{A.1}$$

Usually for convenience, approximate design can be represented alternatively by

$$\xi = \{(c_i, w_i), \sum_{i=1}^{m} w_i = 1, w_i \in (0, 1]\} \tag{A.2}$$

where c_i's are obtained through a bijective function of s_i's which may also involves with unknown parameter vector of length ν, say $\boldsymbol{\theta}$. Information matrix for $\boldsymbol{\theta}$ under such design ξ is

$$I_\xi(\theta) = P(\theta) \left[\sum_{i=1}^{m} w_i C(c_i, \theta) \right] P^{'}(\theta) \tag{A.3}$$

where $P(\boldsymbol{\theta})$ is a $\nu \times \nu$ nonsingular matrix that only depends on θ and $C(c_i, \boldsymbol{\theta})$ is a symmetric $\nu \times \nu$ matrix that involves both c_i and $\boldsymbol{\theta}$. Let ψ_{ij} denote the entries of ith row and jth column

Appendix A (Continued)

in C, and obviously ψ_{ij} is a function of both $\boldsymbol{\theta}$ and c_i. Throughout the rest of the paper, we omit $\boldsymbol{\theta}$ and only keep $\psi_{ij}(c)$ for purpose of simplifying notations. Matrix $C(c_i, \theta)$ therefore can be partitioned and written as

$$C(c_i, \theta) = \left(\begin{array}{c|c} C_{11}(c_i) & C_{12}(c_i) \\ \hline C_{21}(c_i) & C_{22}(c_i) \end{array}\right) \tag{A.4}$$

$$= \left(\begin{array}{ccc|ccc} \psi_{11}(c_i) & & & & & \\ \vdots & \ddots & & & & \\ \psi_{\nu_1,1}(c_i) & \dots & \psi_{\nu_1,\nu_1}(c_i) & & & \\ \hline \psi_{\nu_1+1,\nu_1+1}(c_i) & & & \psi_{\nu_1+1,\nu_1+1}(c_i) & & \\ \vdots & \ddots & & \vdots & \ddots & \\ \psi_{\nu 1}(c_i) & \dots & \psi_{\nu,\nu_1}(c_i) & \psi_{\nu,\nu_1+1}(c_i) & \dots & \psi_{pp}(c_i) \end{array}\right)$$

$$\text{for} \quad i = 1, ..., m \tag{A.5}$$

where $1 \leq \nu_1 < \nu$, $C_{12} = C_{21}^{'}$ and C_{22} is the lower principal submatrix of order $\nu_1 \times \nu_1$. If there exists a design $\tilde{\xi} = \{(\tilde{c}_i, \tilde{w}_i), \sum_{i=1}^{\tilde{m}} \tilde{w}_i = 1\}$ such that

$$\begin{aligned} \sum_{i=1}^{m} w_i C_{11}(c_i) &= \sum_{i=1}^{\tilde{m}} \tilde{w}_i C_{11}(\tilde{c}_i) \\ \sum_{i=1}^{m} w_i C_{12}(c_i) &= \sum_{i=1}^{\tilde{m}} \tilde{w}_i C_{12}(\tilde{c}_i) \\ \sum_{i=1}^{m} w_i C_{22}(c_i) &\leq \sum_{i=1}^{\tilde{m}} \tilde{w}_i C_{22}(\tilde{c}_i) \end{aligned} \tag{A.6}$$

Appendix A (Continued)

then (A.1) is just direct consequence of (A.6). The evidence that supports (A.6) involves a special set of the functions called *Chebyshev System*.

According to (Karlin and Studden, 1966; Dette and Melas, 2011; Yang and Stufken, 2012), suppose there are $k+1$ real functions $u_0, u_1, ..., u_k$ which are continuous on an interval $[A, B]$, the collection of those functions are called *Chebyshev System*(or *T-System*) on $[A, B]$ if for any set of z_i's where $A \leq z_0 < \cdots < z_k \leq B$ the following inequality of matrix determinant holds.

$$\begin{vmatrix} u_0(z_0) & u_0(z_1) & \dots & u_0(z_k) \\ u_1(z_0) & u_1(z_1) & \dots & u_1(z_k) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(z_0) & u_k(z_1) & \dots & u_k(z_k) \end{vmatrix} > 0. \tag{A.7}$$

Among the functions of first $\nu - \nu_1$ columns of $C(c_i, \theta)$, suppose $k-1$ of them form the maximal set of linear dependent functions, they are then selected and renamed as $\Psi_1, ..., \Psi_{k-1}$. Given any non-zero vector Q of length ν_1, define

$$\Psi_k = Q^{'} C_{22}(c_i) Q \tag{A.8}$$

Then (A.6) is equivalent to

$$\begin{aligned} \sum_{i=1}^{m} w_i \Psi_s(c_i) &= \sum_{i=1}^{\tilde{m}} \tilde{w}_i \Psi_s(\tilde{c}_i), \quad \text{for} \quad s = 1, ..., k-1 \\ \sum_{i=1}^{m} w_i \Psi_k(c_i) &< \sum_{i=1}^{\tilde{m}} \tilde{w}_i \Psi_k(\tilde{c}_i), \quad \text{for any nonzero } Q \end{aligned} \tag{A.9}$$

Appendix A (Continued)

(Dette and Melas, 2011) first introduces Chebyshev System into this problem and (Yang and Stufken, 2012) develops a more general tools based on such system. In particular, theorem 1 in (Yang and Stufken, 2012) connects (A.9) to chebyshev system and theorem 2 therein provides convenient tools for checking the sufficient conditions of chebyshev system. We present those theorems here in our notations.

Theorem 8. [Theorem 1 in (Yang and Stufken, 2012)]. *For* $\Psi_1, ..., \Psi_{k-1}$ *and* Ψ_k *defined in* (A.9) *with domain* $[A, B]$, *if either*

$$\{1, \Psi_1, ..., \Psi_{k-1}\} \quad \textit{and} \quad \{1, \Psi_1, ..., \Psi_{k-1}, \Psi_k\} \quad \textit{are Chebyshev Systems} \tag{A.10}$$

$$\{1, \Psi_1, ..., \Psi_{k-1}\} \quad \textit{and} \quad \{1, \Psi_1, ..., \Psi_{k-1}, -\Psi_k\} \quad \textit{are Chebyshev Systems} \tag{A.11}$$

Then the following results hold:

1. *When* $k = 2n - 1$ *and* (A.10) *holds, the designs with at most* n *support points including* B*, form a complete class.*
2. *When* $k = 2n - 1$ *and* (A.11) *holds, the designs with at most* n *support points including* A*, form a complete class.*
3. *When* $k = 2n$ *and* (A.10) *holds, the designs with at most* $n + 1$ *support points, including both* A *and* B *form, a complete class.*
4. *When* $k = 2n$ *and* (A.11) *holds, the designs with at most* n *support points form a complete class.*

Appendix A (Continued)

It is cumbersome to verify (A.10) or (A.11) by definition, and (Yang and Stufken, 2012) provides a computational approach which is feasible to be checked manually or with the aid of computer program that supports symbolic algebra, such as Mathematica. We summarize it as the following theorem.

Theorem 9. [Theorem 2 in (Yang and Stufken, 2012)]. *Given functions $\Psi_0 = 1, \Psi_1, ..., \Psi_k$, and partition in* (A.4)*, construct $f_{\ell,\ell}$ in the following recursive manner:*

$$f_{\ell,t}(c) = \begin{cases} \frac{\partial \Psi_\ell}{\partial c}, & t = 1, \ell = 1, ..., k-1 \\ \frac{\partial C_{22}}{\partial c}, & t = 1, \ell = k \\ \frac{\partial (f_{\ell,t-1}/f_{t-1,t-1})}{\partial c}, & 2 \leq t \leq k, t \leq \ell \leq k \end{cases} \tag{A.12}$$

where $\frac{\partial C_{22}}{\partial c}$ are elementwise derivatives, and $f_{k,s}$ could be matrix functions for $s = 1, ..., k$. Define $F(c) = \prod_{\ell=1}^{k} f_{\ell,\ell}$, then the following results hold:

1. *If $F(c) > 0$ then* (A.10) *holds.*
2. *If $F(c) < 0$ then* (A.11) *holds.*

To state the analytical approach in a nutshell, one has to properly choose ν_1 that stands for a partition of (A.4), then select maximal set of nonconstant linear independent functions and check theorem 9. There are two major concerns that require extra attention. First, there isn't 'the gold standard' for choice of ν_1. (Yang and Stufken, 2012) suggests to set ν_1 around $\nu/2$ as a starting point and shows such arrangement works well for the examples therein. Based on our experience, for some ν_1, the algebra of applying (A.12) would be rather tedious, or even unable

Appendix A (Continued)

to draw conclusion from. Meanwhile, the order of rows and columns of information matrix may result in different set of maximal linear independent functions and so are the $f_{\ell,\ell}$. Usually, the entries with higher order are left in the C_{22}. This maybe the best general strategy available for now and evidence can be found in (Yang and Stufken, 2012) as well.

Appendix B

OPTIMAL WEIGHT EXCHANGE ALGORITHM

Start with initial support $S^{(1)}$ with $\nu+1$ randomly picked sequences and equal weights $p^{(1)}$, at iteration t,

1. Input support $S^{(t)}$, $p^{(t)}$, and update weights using newton's method. Note that support points with zero weights will be deleted in optimizing process.

   Newton's Method:

   Input: Start with $S_1^{(t)} = S^{(t)}$, and $p_1^{(t)} = p^{(t)}$, at iteration j,

   (a) update the weights by the equation $p_{j+1}^{(t)} = p_j^{(t)} - \alpha(\frac{\partial^2\Phi}{\partial p \partial p'})^{-1}\frac{\partial\Phi}{\partial p}|_{p_j^{(t)}}$.

   (b) If $p_{j+1}^{(t)}$ has negative component, go to step (c), otherwise proceed to step (d).

   (c) Set α to $\alpha/2$ and go back to step (a). If $\alpha < \epsilon_\alpha$, remove the point with smallest weight and go back to step (a).

   (d) Check if $\frac{\partial\Phi}{\partial p}|_{p_{j+1}^{(t)}} < \epsilon_0$,if true, $\tilde{p}^{(t)} = p_{j+1}^{(t)}$ is the optimal weights otherwise go to next iteration.

   Output: Support $\tilde{S}^{(t)}$ and optimal weights $p^{(t)}$, where $\tilde{S}^{(t)} = S^{(t)}$ if no points are removed.

2. Derive $s_t^* = \underset{s\in\chi}{argmax}\ d_p(s, \xi^{(t)})$, where $\xi^{(t)} = \{(s_i, p_i)|s_i \in \tilde{S}^{(t)}, p_i \in \tilde{p}^{(t)}, \sum_i p_i = 1\}$

3. Check $d_p(s_t^*, \xi^{(t)}) < \epsilon_d$, where ϵ_d is a pre-selected small positive value. If true, $\xi^{(t)}$ is the desired design. Otherwise, let $S^{(t+1)} = \tilde{S}^{(t)} \bigcup \{s_t^*\}$, $p^{(t+1)} = \tilde{p}^{(t)} \bigcup \{0\}$, and go to step 1.

Appendix C

R PACKAGE: OWEA

C.1 Introduction

The package **OWEA** is intended to serve as a simple yet convenient tool for constructing efficient exact designs. So far it supports A- and D- optimal designs for three models, crossover model with first order carryover effects and subject dropout, crossover model with proportional carryover effects, and interference model. For given set of input, approximate design as well as exact design are returned. Meanwhile, this package also provides function for calculating efficiency both for rounding and comparisons. At last, a shiny app is provided so that one can operate on this package via a graphical interface. The app returns key information including approximate design, exact design, efficiency, and etc. Although it is convenient to use the app directly but one is recommended to call the functions in **OWEA** if more controls of input and output are needed.

The main functions of **OWEA** are summarized in the following table.

The `design` function is the internal work horse that returns approximate design as well as exact design by implementing OWEA algorithm. In particular, the exact design is rounded from approximate design. For supported models, there are parameters shared across models as well as their unique set of inputs. The details of parameters are summarized as follows.

- `model`, character, must be one of 'dropout', 'proportional', or 'interference'.

Appendix C (Continued)

FUNCTIONS	DESCRIPTION
`design`	For calculation of approximate design as well as rounding it to exact design.
`design_app`	A **Shiny** app for **OWEA**.
`eff`	For calculation of efficiency.
`effLB`	For calculation of lower bound of efficiency, only works for dropout model.
`summary`	To print out summarized result returned from `design`.

TABLE XII

FUNCTIONS IN R PACKAGE **OWEA**

- `n`, integer, total number of runs.
- `opt`, numeric, 0 = D-optimal, 1 = A-optimal.
- `t`, integer, total number of treatment levels.
- `p`, integer, total number of periods for 'dropout' and 'proportional' model; block size for 'interference' model.
- `max_iter`, integer, maximum times of iterations, default is 40.
- For `model = 'dropout'`:
  - `drop`, a numeric vector, dropout mechanism, its length must equal to `p`.
- For `model = 'proportional'`:
  - `sigma`, a matrix, assumed covariance matrix of individual error terms.
  - `tau`, a vector, initial value of initial treatment effects.
  - `lambda`, numeric, proportional coefficient in proportional model.

- For `model = 'interference'`:

  - `sigma`, a matrix. assumed covariance matrix of individual error terms.

Note that `design` function includes a feature that randomly select starting design for newton's method. In order to obtain reproducible result, we recommend to use `set.seed()` prior to invoking `design`.

C.2 Example

To install **OWEA** package and load it as well as its dependency to **R** environment

```
R> install.packages("OWEA") # install package
R> library("OWEA") # load package
R> library("gtools") # load dependency
```

Example 1. Crossover model with subject dropout at $(t, p) = (4, 4), n = 16, \ell = (0, 0, 1/2, 1/2)$. This example is from (Zheng, 2013b), where a design named d_2, is shown to have relatively high efficiency under a variety of optimal criteria. We construct D-optimal design under the same setting using `design`, and compare with Zheng's example by `eff`. In addition, the lower bound efficiency of exact design can be obtained by `effLB`.

```
R> set.seed(232) # set random seed for newton's method
R> # D-optimal Design, n = 16, t = 4, p = 4, drop mechanism (0, 0, 0.5, 0.5)
R> example1 <- design('dropout', n = 16, opt = 0, t = 4, p = 4,
+                    drop = c(0, 0, 0.5, 0.5), max_iter = 40)
R> summary(example1) # printing output
```

The output is

```
D-optimal designs for dropout model with 4 treatments 4 periods
dropout mechanism 0, 0, 0.5, 0.5 :

$exact_design
```

Appendix C (Continued)

```
Repetitions
[1,] 1 2 3 4          1
[2,] 3 2 4 4          1
[3,] 2 4 1 3          2
[4,] 3 1 4 2          2
[5,] 4 3 2 1          1
[6,] 3 4 2 1          1
[7,] 2 1 3 3          1
[8,] 4 2 1 1          1
[9,] 4 1 2 2          1
[10,] 1 3 2 2          1
[11,] 1 2 3 4          1
[12,] 1 4 3 3          1
[13,] 2 3 4 4          1
[14,] 4 3 1 1          1

$approximate_design
Weights
[1,] 1 2 3 4 0.136574074
[2,] 2 3 4 4 0.040151065
[3,] 3 2 4 4 0.047214527
[4,] 2 4 1 3 0.117909521
[5,] 3 1 4 2 0.112831469
[6,] 4 3 2 1 0.067786510
[7,] 1 3 4 2 0.023742605
[8,] 3 4 2 1 0.068787564
[9,] 2 1 3 3 0.049790692
[10,] 4 2 1 1 0.043682796
[11,] 4 1 2 2 0.053654742
[12,] 1 4 3 3 0.026048086
[13,] 4 3 1 1 0.058548636
[14,] 1 3 2 2 0.059771184
[15,] 4 1 3 3 0.007662763
[16,] 4 2 1 3 0.018664553
[17,] 3 1 4 4 0.009971946
[18,] 2 4 3 3 0.026060335
[19,] 3 4 1 1 0.011194494
[20,] 2 1 4 4 0.016088388
[21,] 1 2 3 3 0.003864050

$computing_time
```

Appendix C (Continued)

```
user   system   elapsed
2.904  0.092    3.245
```

We access to the efficiency of exact design to approximate design by

```
R> eff(example1)

$Optimal_Criterion
[1] "D-optimal"

$efficiency
[1] 0.9926396
```

Here the efficiency is the evaluated by $\Phi(EC_\xi)/\Phi(EC_d)$. We can also calculate the lower bound of efficiency using `effLB` which is according to Lemma 6.

```
R> effLB(example1)

$lower.bound
[1] 0.82123

$optimal.value
[1] 0.002452183
```

Not only the lower bound is reported, also the minimum of objective function is returned. However calculating the lower bound may last for a while when n is large.

To compare exact design to the one in (Zheng, 2013b), first construct the design as matrix. Note that to use `eff` function, the last column of the design must be the repetitions.

```
R> # Construct Zheng's Design #
R> design_compare <- cbind(t(matrix(c(2,4,3,3,1,4,2,2,2,3,1,1,3,4,1,1,3,1,
+                                    2,2,4,1,3,3,3,2,4,4,2,1,4,4,1,2,3,4,
+                                    1,2,3,4,1,3,4,2,2,4,1,3,4,3,2,1,4,3,
+                                    2,1,4,2,1,3,3,1,4,2), ncol = 16)),1)

R> # Compare Efficiency #
R> eff(example1, ex = design_compare) # relative efficiency
```

```
$Optimal_Criterion
[1] "D-optimal"

$efficiency
[1] 0.9944724
```

As shown, although the lower bound for D-optimal owea design is only 0.8213, its relative efficiency is as high as 0.9945. Therefore, the owea design is close to d_2 in terms of efficiency. Computing time is also reported. For this case, owea design costs 3.245s.

C.3 A Shiny App

If one would like to use a more straightforward approach of using **OWEA** package with only loss of some flexibility in input parameters, a shiny app is also available for a graphical user interface. To launch the shiny app, one has to install the R **shiny** package. After the installation, simply enter

```
R> design_app()
```

A pop-up window like Figure 1 the following will appear.

The top navigation bar has three tabs: 'Crossover Dropout', 'Crossover Proportional', and 'Interference'. Each tab is designed for the model that the name indicates. For example, on Figure 1, 'Crossover Dropout' tab is selected and this page is for crossover model with subject dropout. The left panel below navigation bar is where parameters are entered. Here, for crossover dropout, one need to specify 'total runs', 'periods', 'treatments' and 'dropout mechanism'. The default maximum iteration time is 40 and usually it takes less than 40 times for the algorithm to converge. The 'random seed' is used for reproducible output. When hit the 'RUN' button in blue, it will invoke two functions, `design` and `eff`. The output designs as well as other information are summarized and printed on the right panel (see Figure 2).

For 'Crossover Dropout', if the 'Lower Bound Efficiency' is checked, it would call `effLB` function and output the lower bound of efficiency on the right panel as well. But the calculation usually takes a while especially when the total number of runs is large.

There is no such options for 'Lower Bound Efficiency' when other tabs are activated. For example, when the tab 'Crossover Proportional' is clicked, the interface is now changed to (Figure 3).

The value of checking general equivalence theorem is attached to approximate design as an indicator of convergence. Usually, a small positive value that is close to 0 indicates the convergence of algorithm.

Note that except for common necessary parameter inputs, there are extra input boxes for initial value of parameters on the left panel. For proportional model, one has to provide the initial values for treatment effects as well as proportional coefficient. When all the blanks are filled, simply hit the 'RUN' button, and the output will appear on the right after a short period of time, like Figure 4. Note that the `sigma` is set as identity on the background for the purpose of simplifying input.

The 'interference' tab looks the same to 'proportional' tab except for those missing initial value inputs. In addition, `sigma` is also set to be identity on the background.

Finally, one is able to export the exact design and approximate design to a 'csv' file by clicking on 'Export' button.

Appendix C (Continued)

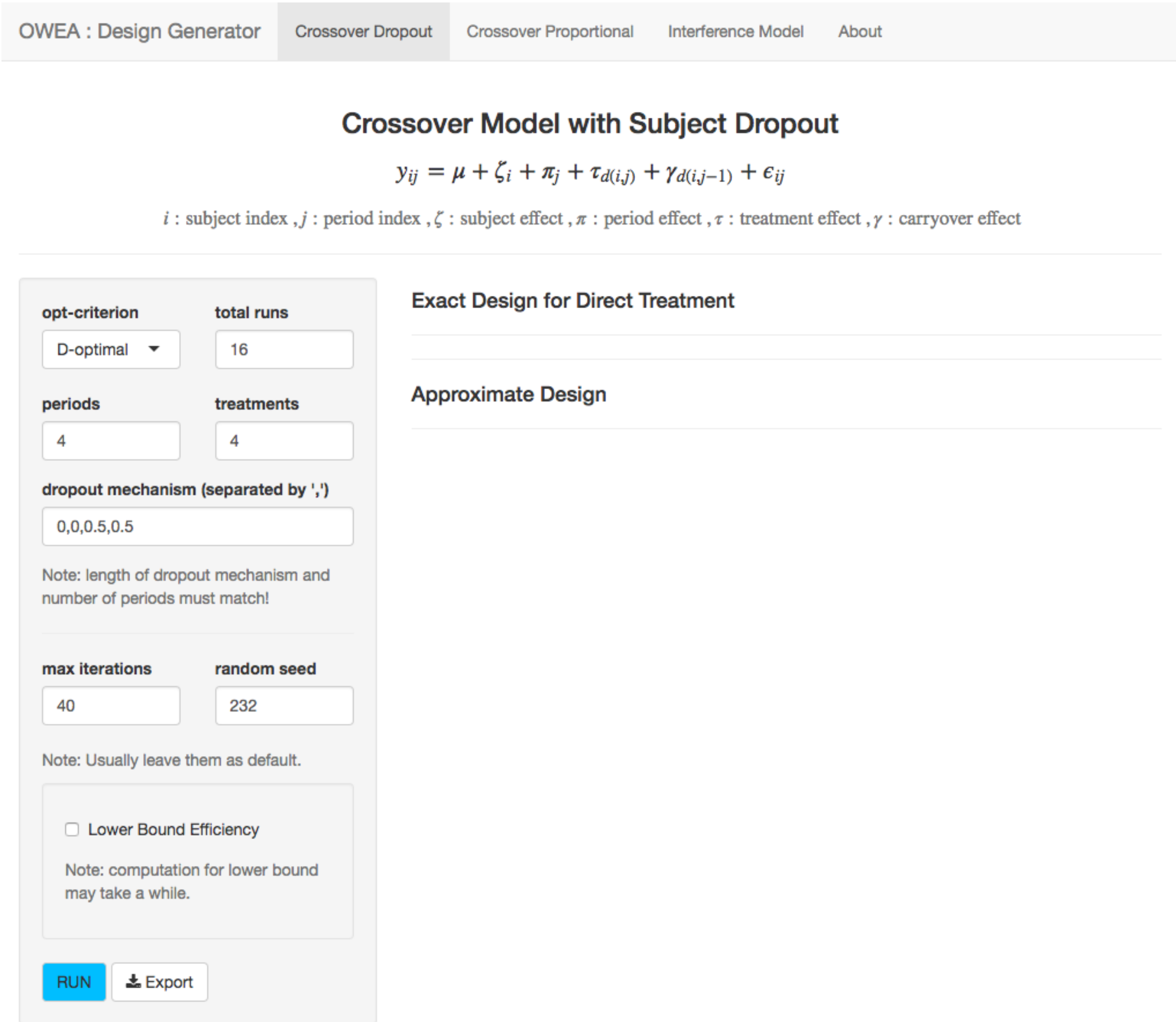

Figure 1. UI of crossover models with subject dropout

Appendix C (Continued)

OWEA : Design Generator | Crossover Dropout | Crossover Proportional | Interference Model | About

Crossover Model with Subject Dropout

$$y_{ij} = \mu + \zeta_i + \pi_j + \tau_{d(i,j)} + \gamma_{d(i,j-1)} + \epsilon_{ij}$$

i : subject index , j : period index , ζ : subject effect , π : period effect , τ : treatment effect , γ : carryover effect

opt-criterion: D-optimal

total runs: 16

periods: 4

treatments: 4

dropout mechanism (separated by ','): 0,0,0.5,0.5

Note: length of dropout mechanism and number of periods must match!

max iterations: 40

random seed: 232

Note: Usually leave them as default.

Lower Bound Efficiency

Note: computation for lower bound may take a while.

RUN   Export

Exact Design for Direct Treatment

efficiency to approximate design : 0.992639644358859
lower bound of efficiency : Not Available

	period 1	period 2	period 3	period 4	repetition
1	1	2	3	4	1
2	3	2	4	4	1
3	2	4	1	3	2
4	3	1	4	2	2
5	4	3	2	1	1
6	3	4	2	1	1
7	2	1	3	3	1
8	4	2	1	1	1
9	4	1	2	2	1
10	1	3	2	2	1
11	1	2	3	4	1
12	1	4	3	3	1
13	2	3	4	4	1
14	4	3	1	1	1

Approximate Design

verifying general equivalence theorem : 4.44089209850063e-15

	period 1	period 2	period 3	period 4	weight
1	1	2	3	4	0.136574074074074
2	2	3	4	4	0.0401510647455011
3	3	2	4	4	0.0472145266523467
4	2	4	1	3	0.117909520892882
5	3	1	4	2	0.112831468752677
6	4	3	2	1	0.0677865099259493
7	1	3	4	2	0.0237426053213972
8	3	4	2	1	0.068787564148125
9	2	1	3	3	0.0497906915269039
10	4	2	1	1	0.0436827956989287
11	4	1	2	2	0.0536547419203611
12	1	4	3	3	0.026048086205506
13	4	3	1	1	0.0585486360015873
14	1	3	2	2	0.059771184005565
15	4	1	3	3	0.00766276327198066
16	4	2	1	3	0.0186645531811926
17	3	1	4	4	0.00997194622144246
18	2	4	3	3	0.026060334528078
19	3	4	1	1	0.0111944942254094
20	2	1	4	4	0.0160883883066355
21	1	2	3	3	0.00386405039345794

Figure 2. Result page of designs for crossover model with subject dropout

Appendix C (Continued)

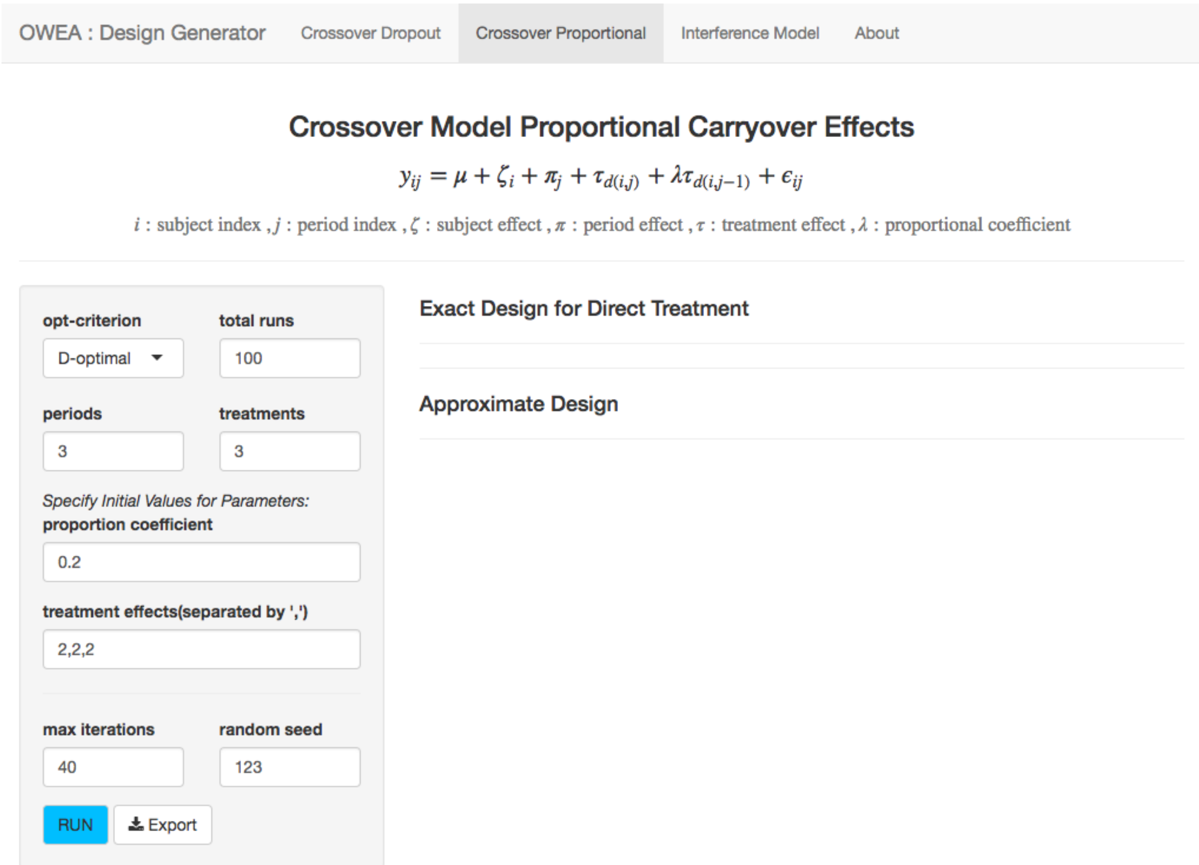

Figure 3. UI of crossover model with proportional carryover effects

Appendix C (Continued)

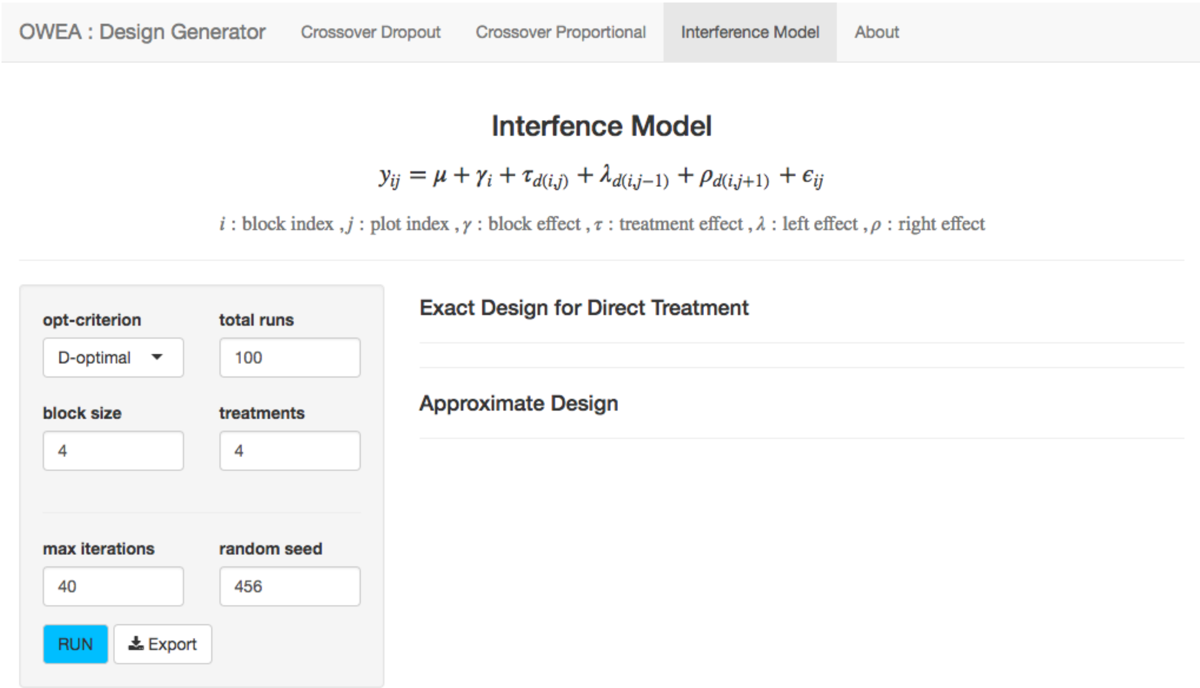

Figure 4. UI of interference model

CITED LITERATURE

Agresti, A.: Categorical Data Analysis. Wiley Series in Probability and Statistics. Wiley, 3rd edition, 2013.

Baer, D. J., Judd, J. T., Clevidence, B. A., and Tracy, R. P.: Dietary fatty acids affect plasma markers of inflammation in healthy men fed controlled diets: A randomized crossover study. American Journal of Clinical Nutrition, 79(6):969–973, jun 2004.

Bailey, R. A. and Druilhet, P.: Optimal cross-over designs for full interaction models. The Annals of Statistics, 42(6):2282–2300, dec 2014.

Bailey, R. A. and Kunert, J.: On optimal crossover designs when carryover effects are proportional to direct effects. Biometrika, 93(3):613–625, 2006.

Baltimore, D.: Expression of animal virus genomes. Bacteriological reviews, 35(3):235–241, 1971.

Bennion, N., Christensen, N. K., and McGarraugh, G.: Alternate Site Glucose Testing: A Crossover Design. Diabetes Technology & Therapeutics, 4(1):25–33, feb 2002.

Böhning, D.: A vertex-exchange-method in D-optimal design theory. Metrika, 33(1):337–347, 1986.

Bose, M. and Bagchi, S.: Optimal crossover designs under premature stopping. Utilitas Mathematica, 75, 2008.

Bose, M. and Dey, A.: Optimal Crossover Designs. WORLD SCIENTIFIC, apr 2009.

Bose, M. and Mukherjee, B.: Optimal crossover designs under a general model. Statistics and Probability Letters, 62(4):413–418, 2003.

Bose, M. and Stufken, J.: Optimal crossover designs when carryover effects are proportional to direct effects. Journal of Statistical Planning and Inference, 137(11):3291–3302, 2007.

Bu, X., Majumdar, D., and Yang, J.: D-optimal Designs for Multinomial Logistic Models. Annual of Statistics, feb 2019.

Chaloner, K. and Verdinelli, I.: Bayesian Experimental Design: A Review. Statistical Science, 10(3):273–304, aug 1995.

Chan, R. H., Wang, R. W., and Yeung, H. M.: Composition vector method for phylogenetics a review. The 9th International Symposium on Operations Research and Its Applications, pages 13–20, 2010.

Chen, R.-B., Chang, S.-P., Wang, W., Tung, H.-C., and Wong, W. K.: Minimax optimal designs via particle swarm optimization methods. Statistics and Computing, 25(5):975–988, 2015.

Cheng, C.-S. and Wu, C.-F.: Balanced Repeated Measurements Designs. The Annals of Statistics, 8(6):1272–1283, nov 1980.

Chor, B., Horn, D., Goldman, N., Levy, Y., and Massingham, T.: Genomic DNA k-mer spectra: Models and modalities. Genome Biology, 10(10), 2009.

Cortes, C. and Vapnik, V.: Support-Vector Networks. Machine Learning, 20(3):273–297, 1995.

Covey-Crump, P. A. K. and Silvey, S. D.: Optimal regression designs with previous observations. Biometrika, 57(3):551–566, 1970.

Darling, D. A.: Note on a Limit Theorem. The Annals of Probability, 3(5):876–878, oct 1975.

de la Garza, A.: Spacing of Information in Polynomial Regression. The Annals of Mathematical Statistics, 25(1):123–130, mar 1954.

Deng, M., Yu, C., Liang, Q., He, R. L., and Yau, S. S.: A novel method of characterizing genetic sequences: Genome space with biological distance and applications. PLoS ONE, 6(3), 2011.

Dette, H. and Melas, V. B.: A note on the de la Garza phenomenon for locally optimal designs. The Annals of Statistics, 39(2):1266–1281, apr 2011.

Dette, H. and Schorning, K.: Complete classes of designs for nonlinear regression models and principal representations of moment spaces. Ann. Statist., 41(3):1260–1267, 2013.

Dobson, A. J.: An introduction to generalized linear models. Chapman and Hall/CRC, 2002.

Donovan, S. J., Stewart, J. W., Nunes, E. V., Quitkin, F. M., Parides, M., Daniel, W., Susser, E., and Klein, D. F.: Divalproex treatment for youth with explosive temper and mood lability: A double-blind, placebo-controlled crossover design. American Journal of Psychiatry, 157(5):818–820, 2000.

Dudoit, S., Fridlyans, J., and Speed, T. P.: Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of the American Statistical Association, 97(457):77–87, 2002.

Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters, 27(8):861–874, 2006.

Fedorov, V.: Theory of Optimal Experiments Designs. New York : Academic Press, 1972.

Ford, I., Torsney, B., and Wu, C.-F. J.: The use of a canonical form in the construction of locally optimal designs for nonlinear problems. J. Roy. Statist. Soc. Ser. B, 54(2):569–583, 1992.

Glonek, G.: A class of regression models for multivariate categorical responses. Biometrika, 83(1):15–28, mar 1996.

Glonek, G. F. V. and McCullagh, P.: Multivariate Logistic-Models. Journal of the Royal Statistical Society Series B- Methodological, 57(3):533–546, 1995.

Godolphin, J. D.: Simple pilot procedures for the avoidance of disconnected experimental designs. Journal of the Royal Statistical Society. Series C: Applied Statistics, 53(1):133–147, jan 2004.

Grizzle, J. E., Starmer, C. F., and Koch, G. G.: Analysis of Categorical Data by Linear Models. Biometrics, 25(3):489, 1969.

Hand, D. J. and Till, R. J.: A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Machine Learning, 45(2):171–186, 2001.

Hao, B., Qi, J., and Wang, B.: Prokaryotic Phylogeny Based on Complete Genomes Without Sequence Alignment. Modern Physics Letters B, 17(03):91–94, 2003.

Harman, R. and Pronzato, L.: Improvements on removing nonoptimal support points in D-optimum design algorithms. Statistics and Probability Letters, 77(1):90–94, 2007.

Harris, J. E. and Raynor, H. A.: Crossover Designs in Nutrition and Dietetics Research. Journal of the Academy of Nutrition and Dietetics, 117(7):1023–1030, 2017.

Hastie, Trevor, Tibshirani, Robert, Friedman, J.: The Elements of Statistical Learning The Elements of Statistical LearningData Mining, Inference, and Prediction, Second Edition. Springer, 2009.

Hedayat, A. and Afsarinejad, K.: Repeated measurements designs I. A survey of statistical design and linear models (Proc. Internat. Sympos., Colorado State Univ., Ft. Collins, Colo., 1973), pages 229–242, 1975.

Hedayat, A. and Afsarinejad, K.: Repeated Measurements Designs, II. The Annals of Statistics, 6(3):619–628, may 1978.

Hedayat, A. and Zhao, W.: Optimal Two-Period Repeated Measurements Designs. The Annals of Statistics, 18(4):1805–1816, dec 1990.

Hedayat, A. S. and Stufken, J.: Optimal and efficient crossover designs under different assumptions about the carryover effects. Journal of Biopharmaceutical Statistics, 13(3):519–528, 2003.

Hedayat, A. S., Stufken, J., and Yang, M.: Optimal and Efficient Crossover Designs When Subject Effects Are Random. Journal of the American Statistical Association, 101(475):1031–1038, sep 2006.

Hedayat, A. S. and Yang, M.: Universal optimality of balanced uniform crossover designs. Annals of Statistics, 31(3):978–983, 2003.

Hedayat, A. S. and Yang, M.: Universal optimality for selected crossover designs. Journal of the American Statistical Association, 99(466):461–466, 2004.

Hedayat, A. S. and Yang, M.: Optimal and efficient crossover designs for comparing test treatments with a control treatment. The Annals of Statistics, 33(2):915–943, apr 2005.

Hedayat, A. S. and Zheng, W.: Optimal and Efficient Crossover Designs for Test-Control Study When Subject Effects Are Random. Journal of the American Statistical Association, 105(492):1581–1592, dec 2010.

Hettich, R.: A Review of Numerical Methods for Semi-Infinite Optimization. In Semi-Infinite Programming and Applications, eds. A. V. Fiacco and K. O. Kortanek, pages 158–178, Berlin, Heidelberg, 1983. Springer Berlin Heidelberg.

Huang, G., Chu, C., Huang, T., Kong, X., Zhang, Y., Zhang, N., and Cai, Y.-D.: Exploring Mouse Protein Function via Multiple Approaches. PLOS ONE, 11(11):e0166580, nov 2016.

Huang, G., Zhou, H., Li, Y., and Xu, L.: Alignment-free comparison of genome sequences by a new numerical characterization. Journal of Theoretical Biology, 281(1):107–112, 2011.

Huang, H. H., Xu, T., and Yang, J.: Comparing logistic regression, support vector machines, and permanental classification methods in predicting hypertension. In BMC Proceedings, volume 8, 2014.

Jayaratne, K., Bird, C., and McClelland, J.: Application of crossover design for conducting rigorous extension evaluations. Journal of Extension, 51(2), 2013.

John, J., Russell, K., and Whitaker, D.: Crossover: an algorithm for the construction of efficient cross-over designs. Statistics in medicine, 23(17):2645–2658, 2004.

Jones, B. and Kenward, M.: Design and Analysis of Cross-Over Trials. CRC PRESS, 3rd edition, 2014.

Karlin, S. and Studden, W. J.: Tchebycheff systems: with applications in analysis and statistics. Pure and applied mathematics. Interscience Publishers, 1966.

Kempton, R. A., Ferris, S. J., and David, O.: Optimal change-over designs when carry-over effects are proportional to direct effects of treatments. Biometrika, 88(2):391–399, 2001.

Kennedy, J. and Eberhart, R.: Particle swarm optimization. Neural Networks, 1995. Proceedings., IEEE International Conference on, 4:1942–1948 vol.4, 1995.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M.: Design Issues for Generalized Linear Models: A Review. Statistical Science, 2006.

Kiefer, J.: General Equivalence Theory for Optimum Designs (Approximate Theory). The Annals of Statistics, 2(5):849–879, sep 1974.

Kiefer, J.: General Equivalence Theory for Optimum Designs (Approximate Theory). The Annals of Statistics, 2(5):849–879, sep 1974.

Korkmaz, S., Goksuluk, D., and Zararsiz, G.: MVN: An R package for assessing multivariate normality. The R Journal, 6(2013):151–162, 2014.

Kunert, J. and Martin, R. J.: On the determination of optimal designs for an interference model. Annals of Statistics, 28(6):1728–1742, 2000.

Kunert, J. and Stufken, J.: Optimal Crossover Designs in a Model With Self and Mixed Carryover Effects. Journal of the American Statistical Association, 97(459):898–906, sep 2002.

Kunert, J. and Stufken, J.: Optimal crossover designs for two treatments in the presence of mixed and self-carryover effects. Journal of the American Statistical Association, 103(484):1641–1647, 2008.

Kunert, J.: Optimal Design and Refinement of the Linera Model with Applications to Repeated Measurements Designs. The Annals of Statistics, 11(1):247–257, mar 1983.

Kunert, J.: Optimality of Balanced Uniform Repeated Measurements Designs. The Annals of Statistics, 12(3):1006–1017, sep 1984.

Kunert, J. and Mersmann, S.: Optimal designs for an interference model. Journal of Statistical Planning and Inference, 141(4):1623–1632, 2011.

Kushner, H.: Optimality and efficiency of two-treatment repeated measurement designs. Biometrika, 84(2):455–468, jun 1997.

Kushner, H. B.: Optimal repeated measurements designs: the linear optimality equations. The Annals of Statistics, 25(6):2328–2344, dec 1997.

Kushner, H. B.: Optimal and Efficient Repeated-Measurements Designs for Uncorrelated Observations. Journal of the American Statistical Association, 93(443):1176–1187, sep 1998.

Lall, R., Campbell, M. J., Walters, S. J., and Morgan, K.: A review of ordinal regression models applied on health-related quality of life assessments, 2002.

Lam, J. A. and Kahler, C. W.: A randomized crossover trial to test the effects of positive psychology intervention delivered by text messaging. The Journal of Positive Psychology, 0(0):1–13, 2017.

Laska, E., Meisner, M., and Kushner, H. B.: Optimal Crossover Designs in the Presence of Carryover Effects. Biometrics, 39(4):1087, dec 1983.

Liu, I. and Agresti, A.: The analysis of ordered categorical data: An overview and a survey of recent developments. Test, 14(1):1–73, 2005.

Low, B. Y. J. L., Lewis, S. M., and Prescott, P.: Assessing robustness of crossover designs to subjects dropping out. Statistics and Computing, 9(3):219–227, jul 1999.

Maddouri, M. and Elloumi, M.: A data mining approach based on machine learning techniques to classify biological sequences. Knowledge-Based Systems, 15(4):217–223, may 2002.

Majumdar, D., Dean, A. M., and Lewis, S. M.: Uniformly balanced repeated measurements designs in the presence of subject dropout. Statistica Sinica, 18:235–253, 2008.

Martín-Martín, R., Rodríguez-Aragón, L. J., and Torsney, B.: Multiplicative algorithm for computing D-optimum designs for pVT measurements. Chemometrics and Intelligent Laboratory Systems, 111(1):20–27, feb 2012.

Matthews, J. N. S.: Recent Developments in Crossover Designs. International Statistical Review / Revue Internationale de Statistique, 56(2):117, aug 1988.

Matthews, J. N. and Henderson, R.: Two-period, two-treatment crossover designs subject to non-ignorable missing data. Biostatistics, 14(4):626–638, 2013.

McCullagh, P. and Nelder, J. A.: Generalized Linear Models, Second Edition. Chapman & Hall/CRC, 1989.

McFadden, M., Powers, J., Brown, W., and Walker, M.: Vehicle and Driver Attributes Affecting Distance from the Steering Wheel in Motor Vehicles. Human Factors: The Journal of the Human Factors and Ergonomics Society, 42(4):676–682, dec 2000.

Mills, E. J., Chan, A.-W., Wu, P., Vail, A., Guyatt, G. H., and Altman, D. G.: Design, analysis, and presentation of crossover trials. Trials, 10(1):27, dec 2009.

Osterstock, J. B., MacDonald, J. C., Boggess, M. M., and Brown, M. S.: Technical note: Analysis of ordinal outcomes from carcass data in beef cattle research. Journal of Animal Science, 2010.

Park, D., Bose, M., Notz, W., and Dean, A.: Efficient crossover designs in the presence of interactions between direct and carry-over treatment effects. Journal of Statistical Planning and Inference, 141(2):846–860, feb 2011.

Perevozskaya, I., Rosenberger, W. F., and Haines, L. M.: Optimal design for the proportional odds model. Canadian Journal of Statistics, 31(2):225–235, 2003.

Polychronopoulos, D., Weitschek, E., Dimitrieva, S., Bucher, P., Felici, G., and Almirantis, Y.: Classification of selectively constrained DNA elements using feature vectors and rule-based classifiers. Genomics, 104(2):79–86, 2014.

Prunuske, A. J., Henn, L., Brearley, A. M., and Prunuske, J.: A Randomized Crossover Design to Assess Learning Impact and Student Preference for Active and Passive Online Learning Modules. Medical Science Educator, 26(1):135–141, mar 2016.

Pukelsheim, F. and Rieder, S.: Efficient Rounding of Approximate Designs. Biometrika, 79(4):763, dec 1992.

Randall, J. H.: The Analysis of Sensory Data by Generalized Linear Model. Biometrical Journal, 31(7):781–793, 1989.

Schacter, L., Birkhofer, M., Carter, S., Canetta, R., Hellmann, S., Onetto, N., Weil, C. B., and Rozencweig, M. R.: Anticancer drugs. In Handbook of Phase I/II Clinical Drug Trials, eds. J. O'Grady and P. H. Joubert, pages 523–534. Boca Raton, CRC PRESS, 1st edition, 1997.

Silvey, S., Titterington, D., and Torsney, B.: An algorithm for optimal designs on a design space. Communications in Statistics - Theory and Methods, 7(14):1379–1389, jan 1978.

Sims, G. E., Jun, S.-R., Wu, G. A., and Kim, S.-H.: Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. Proceedings of the National Academy of Sciences, 106(8):2677–2682, 2009.

Speybroeck, N., Marcotty, T., Aerts, M., Dolan, T., Williams, B., Lauer, J., Molenberghs, G., Burzykowski, T., Mulumba, M., and Berkvens, D.: Titrating Theileria parva: Single stocks against combination of stocks. Experimental Parasitology, 2008.

Stufken, J.: 3 Optimal crossover designs. In Handbook of Statistics, volume 13, pages 63–90. ELSEVIER, 1996.

Stufken, J. and Yang, M.: ON LOCALLY OPTIMAL DESIGNS FOR GENERALIZED LINEAR MODELS WITH GROUP EFFECTS. Statistica Sinica, 22(4):1765–1786, 2012.

Stufken, J. and Yang, M.: Optimal Designs for Generalized Linear Models, chapter 4, pages 137–164. Wiley-Blackwell, 2012.

Torsney, B. and Martín-Martín, R.: Multiplicative algorithms for computing optimum designs, 2009.

Torsney, B.: Algorithms for a constrained optimization problem with applications in statistics and optimum design. Doctoral dissertation, University of Glasgow, 1981.

Vinga, S.: Biological sequence analysis by vector-valued functions: revisiting alignment-free methodologies for DNA and protein classification. In Advanced Computational Methods for Biocomputing and Bioimaging, pages 70–105. NOVA SCIENCE, 2007.

Vinga, S. and Almeida, J.: Alignment-free sequence comparison-a review. Bioinformatics (Oxford, England), 19(4):513–23, 2003.

Wang, H., Yang, M., and Stufken, J.: Information-based optimal subdata selection for big data linear regression. Journal of the American Statistical Association, 0(0):1–13, 2018.

Wang, T., Malone, J., Fu, H., Heilmann, C., Qu, Y., and Huster, W. J.: Crossover design and its application in late-phase diabetes studies. Journal of Diabetes, 8(5):610–618, sep 2016.

Weitschek, E., Cunial, F., and Felici, G.: LAF: Logic Alignment Free and its application to bacterial genomes classification. BioData Mining, 8(1), 2015.

Wellek, S. and Blettner, M.: On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Deutsches Arzteblatt international, 109(15):276–81, apr 2012.

Wong, W., Chen, R.-B., Huang, C.-C., and Wang, W.: A modified Particle Swarm Optimization technique for finding optimal designs for mixture models. PLoS ONE, 10(6), 2015.

Wu, F.-C.: Simultaneous Optimization of Robust Design with Quantitative and Ordinal Data. International Journal of Industrial Engineering: Theory, Applications and Practice, 15(2):231–238, 2008.

Wynn, H. P.: The Sequential Generation of \$D\$-Optimum Experimental Designs. The Annals of Mathematical Statistics, 41(5):1655–1664, 1970.

Yang, J., Tong, L., and Mandal, A.: D-optimal Designs with Ordered Categorical Data. Statistica Sinica, 27:1879–1902, 2017.

Yang, M.: On the de la Garza phenomenon. The Annals of Statistics, 38(4):2499–2524, aug 2010.

Yang, M., Biedermann, S., and Tang, E.: On optimal designs for nonlinear models: a general and efficient algorithm. J. Amer. Statist. Assoc., 108(504):1411–1420, 2013.

Yang, M. and Stufken, J.: Optimal and efficient crossover designs for comparing test treatments to a control treatment under various models. Journal of Statistical Planning and Inference, 138(1):278–285, 2008.

Yang, M. and Stufken, J.: Support points of locally optimal designs for nonlinear models with two parameters. Annals of Statistics, 37(1):518–541, 2009.

Yang, M. and Stufken, J.: Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. The Annals of Statistics, 40(3):1665–1681, jun 2012.

Yang, M., Zhang, B., and Huang, S.: Optimal designs for generalized linear models with multiple design variables. Statistica Sinica, 21(3):1415–1430, jun 2011.

Yang, M., Zhang, B., and Huang, S.: OPTIMAL DESIGNS FOR GENERALIZED LINEAR MODELS WITH MULTIPLE DESIGN VARIABLES. Statistica Sinica, 21(3):1415–1430, 2011.

Yu, Y.: Monotonic convergence of a general algorithm for computing optimal designs. The Annals of Statistics, 38(3):1593–1606, 2010.

Yu, Y.: D-optimal designs via a cocktail algorithm. Statistics and Computing, 21(4):475–481, 2011.

Zhao, S. and Majumdar, D.: On uniformly balanced crossover designs efficient under subject dropout. Journal of Statistical Theory and Practice, 6(1):178–189, 2012.

Zheng, W.: Optimal crossover designs for the proportional model. The Annals of Statistics, 41(4):2218–2235, aug 2013.

Zheng, W.: Universally optimal crossover designs under subject dropout. The Annals of Statistics, 41(1):63–90, feb 2013.

Zheng, W.: Universally optimal designs for two interference models. Ann. Statist., 43(2):501–518, 04 2015.

Zocchi, S. S. and Atkinson, a. C.: Optimum experimental designs for multinomial logistic models. Biometrics, 55(2):437–444, 1999.

VITA

Education:

- Bachelor of Economics, Henan University of Economy and Law, 2008 - 2012.
- Ph.D in Statistics, University of Illinois at Chicago, 2013 - present.

Publification:

- Hao, S., Huang, HH. and Alarcon, S., Comparisons of Classification Methods for Viral Genome and Protein Families using Alignment-free Vectorization. *Statistical Applications in Genetics and Molecular Biology.* `doi:10.1515/sagmb-2018-0004`.
- Hao, S., Han, M, and et al. Three-dimensional cephalometric analysis of the maxilla-analysis of new landmarks. *American Journal of Orthodontics and Dentofacial Orthopedics.*

Manuscript:

- Hao, S., and Yang, M. Support Points of Locally Optimal Designs for Multinomial logistic Regression Models.
- Hao, S., Yang, M., and Zheng, W. An R package for designs of crossover model with subject dropout and interference model.

Award

- Department consulting award, year of 2015 - 2016 and 2017 - 2018.