D-optimal Designs for Multinomial Logistic Models

by

Xianwei Bu
B.A. (Peking University) 1992
M.S. (Loyola University Chicago) 2011

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, 2017

Chicago, Illinois

Defense Committee:
Jie Yang, Chair and Adviser
Dibyen Majumdar
Min Yang
Jing Wang
Domenic Reda, Hines VA Hospital

To my parents,

Ms. Qin Wang and Mr. Yunan Bu

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude to my adviser Dr. Jie Yang for his tremendous academic support, wise advice, patience and encouragements. During the most difficult times, he was always able to find a good direction for me to keep exploring. His guidance helped me in all the time of research and in the writing of this thesis. My dissertation study turned out to be a pleasant academic journey instead of grueling work. Furthermore, he gave me very helpful career advice and suggestions in general, which definitely will benefit my future career life.

My sincere thanks also go to Dr. Dibyen Majumdar for his valuable suggestions and insightful comments, especially on EW optimal designs. I truly appreciate his time and efforts spent on this research project given his intensive administrative workload.

I am very grateful to the remaining members of my dissertation committee, Dr. Min Yang, Dr. Jing Wang and Dr. Domenic Reda. Dr. Min Yang is extremely knowledgeable in design theory and applications. I learned so much from his design courses and also understood some key elements about experimental design. Dr. Jing Wang has done very profound research study. She is also very nice and easy going. I had a pleasant time with her as her Teaching Assistant. Dr. Domenic Reda has rich practical experience about clinical trials and also taught "Design of Clinical Trials", which was one of the most useful classes I have ever taken. I did not realize how much this course helped me until I started my job at Abbvie.

I will forever be thankful to my former college, Loyola University Chicago. I changed my previous major to statistics at Loyola, and found a lot of interest in statistics, which became a milestone in my life. Of course, I would like to thank Professors Timothy O'Brien, Emmanuel Barron, Alan Saleski and Jerald Funk. Their courses were wonderful and provided me with a solid basis in statistics. Professor O'Brien and Professor Barron also encouraged me to pursue a PhD degree, which I had never thought of myself!

My gratitude extends to my classmates: Liyuan Jiang, Yi Huang, Xindi Wang, Shuang Zhou, Shuai Hao and Keren Li for their enthusiasm and willingness to provide me help in many different ways. I got to know many of them while attending orientation, then we took courses together, helped each other, also discussed and exchanged ideas. My student life was more interesting and exciting because of them.

I also thank Maureen Madden, Lisa Jones, Eloy Reyes and Joyce Tucker for all the administrative help. They were very helpful, friendly, considerate and cheerful, I was always happy to talk with them, no matter if it was for asking help, or just a simple greeting.

Last but not the least, I would like to thank my husband Liqun Xi and my daughter Wang Xi for their support and love. Without my family, I wouldn't have had as much motivation, strength and perseverance to finish this challenging work.

TABLE OF CONTENTS

CHAPTER **PAGE**

LIST OF TABLES

LIST OF ABBREVIATIONS

EW	E for expectation, W for the notation w_i used for information in individual observation
MW	M for median, W for the notation w_i used for information in individual observation
NPO	Non-proportional odds
PO	Proportional odds
PPO	Partial proportional odds
UIC	University of Illinois at Chicago

SUMMARY

A well-designed experiment is an efficient way of learning about the world. Because experiments cannot avoid random error even in the carefully controlled conditions of laboratories, statistical methods are essential for their efficient design and analysis.

To design experiments we need to know and apply the theory of optimal experimental design. In this thesis we aim to achieve several goals in D-optimal design about multinomial logistic models with responses having more than two categories. These are:

1° Constructing unified matrix expression for all types of multinomial logistic models. The matrix expression is in unified form, with some items expressed differently for a different model.

2° Unified form of Fisher information matrix and its determinant can be derived based on the unified matrix expression for models. Again, the results apply to all types of multinomial logistic models.

3° D-optimal designs are available for all types of multinomial logistic models, based on the unified form of Fisher information matrix. In some special cases, the analytic solutions of D-optimal design exist.

This thesis is organized as follows: Chapter 1 introduces the current multinomial logistic models, which include four different logit models for multinomial responses and three types of odds assumptions for model parameters. Basic concepts of optimal design and current study methods and results about Design of multinomial logistic models are described. Chapter 2 first summarizes all types of multinomial logistic models, then constructs unified matrix expression for all types of multinomial logistic models, following (1) and (2), but in an explicit way. The unified form of Fisher information matrix is derived and expressed in Equation 2.10, D-optimality can be invariant for a certain reparametrization based on Jacobian transformation of Fisher information matrix.

Chapter 3 starts from another form of Fisher information matrix Equation 3.7, then discusses positive definiteness of the Fisher information matrix based on that. The number of minimally supported points can be obtained as well. Chapter 4 focuses on determinant of Fisher information matrix. Fisher information matrix can be further reformulated in Equation 4.1, so the general formula for determinant of Fisher information matrix can be calculated, which gives the overall look. But in some special cases, the analytic form of Fisher information matrix's determinant can be calculated. We also tried another approach, it didn't work in many situations. However, it greatly simplified the determinant for the continuation-ratio logit model with non-proportional odds. Chapter 3 and Chapter 4 give the most important results and are a foundation in this thesis.

Chapter 5 discusses design space for multinomial logistic models. Basically the design space has no limitations for all of the models except for cumulative logit models, there the order constraints about the linear predictors exist. The solutions of multinomial response probabilities are also calculated from the model equations.

Chapter 6 introduces different D-optimal designs, including approximate design and exact design, both of them belong to local D-optimal design. We use EW D-optimal design as an efficient surrogate of Bayesian D-optimal design. Besides, minimally supported design is discussed.

Chapter 7 illustrates application of D-optimal design in two examples. One example is fitted with continuation-ratio logit model with non-proportional odds, another example applies cumulative logit model with non-proportional odds. The results show that uniform design usually is not optimal design, although it is commonly adopted in practice. The uniform design's efficiency could be improved greatly by D-optimal design. The D-optimal designs found by algorithms also confirm the corresponding theoretical results.

CHAPTER 1

INTRODUCTION

1.1 Multinomial Logistic Models

In statistics, categorical variables are those variables that fall into a particular category. Usually, the variables take on one of a number of fixed values in a set. Many categorical variables have only two categories. Such variables, for which the two categories (often given the generic labels "success" and "failure") are called binary variables.

When a categorical variable has more than two categories, we distinguish between two types of categorical scales. Variables having categories without a natural ordering are said to be measured on a nominal scale and are called nominal variables. Examples are mode of transportation to get to work (automobile, bicycle, bus, subway, walk), favorite type of music (classical, country, folk, jazz, rock), and choice of residence (apartment, condominium, house, other). For nominal variables, the order of listing the categories is irrelevant to the statistical analysis.

Many categorical variables do have ordered categories. Such variables are said to be measured on an ordinal scale and are called ordinal variables. Examples are social class (upper, middle, lower), political philosophy (very liberal, slightly liberal, moderate, slightly conservative, very conservative), patient condition (good, fair, serious, critical), and rating of a movie for Netflix (1 to 5 stars, representing hated it, didn't like it, liked it, really liked it, loved it). For ordinal variables, distances between categories are unknown. Although a person categorized as very liberal is more liberal than a person categorized as slightly liberal, no numerical value describes how much more liberal that person is.

Among the ordinal variables, there exists a special type called hierarchical variables: the responses have a clear nested or hierarchical structure, for example, in Zocchi and Atkinson's flies example (2), death before emergence and death during emergence occur in two different stages. This type of hierarchical variables have some special features, use different models compared with other ordinal variables, so we list hierarchical variables as one separate type of variables, following (2).

Generally all of the categorical response variables discussed above are assumed to follow multinomial distributions, and we often use multinomial logistic models (other models also exist, but in this thesis we focus on these types of models) to fit them. These models are actually natural extensions of ordinary logistic regression models, which are applied to binary responses and assume a binomial distribution. However, the extension could be different according to different type of response variables.

In statistical literature, four kinds of logit models have been commonly used, including the baseline-category logit model for nominal responses, for example, (2; 3), the cumulative logit model for ordinal responses (4; 5), the adjacent-categories logit model for ordinal responses (3; 6) and the continuation-ratio logit model for hierarchical responses (2; 3). The following examples are given to explain these models.

Example 1.1. ***Baseline-category logit model***

Agresti (3) introduced a famous study about food choice of alligators. The study captured 219 *alligators in four Florida lakes. The response variable is the primary food type in an alligator's stomach. This had five categories: fish, invertebrate, reptile, bird, other. These categories haven't orderings so the response is a nominal variable. The alligators are classified according to L = lake of capture (Hancock, Oklawaha, Trafford, George), G = gender (male, female), and S = size (<2.3 meters long, >2.3 meters long), These variables can serve as predictors or explanatory variables.*

Let $\mathbf{x}$ *be a fixed experimental setting for explanatory variables. For observations at that setting, the response* $\mathbf{Y}$ *could fall into one of the* J *categories as a multinomial variable with probabilities* $(\pi_1(\mathbf{x}), \cdots, \pi_J(\mathbf{x}))$*. Baseline-category logistic models pair each response category with a baseline category, such as the last one or the most common one. But any category could be treated as a baseline-category. Consider the model*

$$\log \frac{\pi_j(\mathbf{x})}{\pi_J(\mathbf{x})} = \alpha_j + \boldsymbol{\beta}_j^T \mathbf{x} \qquad j = 1, \cdots, J-1$$

The left-hand side is the log of ratio of $\pi_j(\mathbf{x})$ *over* $\pi_J(\mathbf{x})$*, it is equivalent to the logit of a conditional probability,* $logit[P(\mathbf{Y}=j|\mathbf{Y}=j \text{ or } J)]$*. This model treats the* J*th category as a baseline-category and pair each other category with the* J*th category. There are totally* $J-1$ *logits and related equations. These equations describe the effects of* $\mathbf{x}$ *on these* $J-1$ *logits. This model is a typical baseline-category logit model.*

If we look at the right-hand side, we will find j appear as subscript of $\boldsymbol{\beta}$. This indicates that the effects of predictors ($\mathbf{x}$) vary with different logit, thus the parameters for different logit equations are different. This assumption is called non-proportional odds (NPO).

Returning to this example, the prediction equation for the logit of selecting invertebrates instead of fish is

$$\log \frac{\hat{\pi_I}}{\hat{\pi_F}} = -1.55 + 1.46s - 1.66Z_H + 0.94Z_O + 1.12z_T$$

Here $s = 1$ for size < 2.3 meters and 0 otherwise, and z_H, z_O and z_T are indicator variables for Lakes Hancock, Oklawaha and Trafford. This is the only one equation, we can have other 3 logit equations, i.e., the logit of selecting reptile instead of fish, the logit of selecting bird instead of fish, the logit of selecting other instead of fish, these 4 logit equations consist of one whole baseline-category logit model.

Example 1.2. ***Cumulative logit model***

McCullagh (4) applied a cumulative logit model to an example of tonsil size study. 1398 children are classified 3 categories according to their relative tonsil size: Present but not enlarged, Enlarged and Greatly enlarged (response variables), and whether or not they were carriers of Streptococcus pyogenes (explanatory variables). The response variables have natural orders and are ordinal variables. The cumulative logit model is

$$\log \frac{\gamma_j(\mathbf{x})}{1-\gamma_j(\mathbf{x})} = \theta_j - \boldsymbol{\beta}^T \mathbf{x} \qquad 1 \leqslant j < k$$

where $\gamma_j(\mathbf{x}) = \pi_1(\mathbf{x}) + \cdots + \pi_j(\mathbf{x})$ *is the cumulative probability from the* $1st$ *category to the* jth *category, the ratio* $\gamma_j(\mathbf{x})/(1-\gamma_j(\mathbf{x}))$ *is the odds for the event* $\mathbf{Y} \leqslant j$*. Basically each logit equation calculates the log odds for the event* $\mathbf{Y} \leqslant j$*. Since there are* $k-1$ *cutoff points among* k *categories, a cumulative logit model has* $k-1$ *logit equations.*

Here we find the right-hand side has only $\boldsymbol{\beta}$*, without* j *as its subscript. Therefore, the effects of predictors are the same across different logit equations, which is often referred to as proportional odds assumption (PO).*

Example 1.3. *Adjacent-categories logit model*

The cumulative logit model is often used for ordinal responses. However, an alternative: adjacent-categories logit model (3; 6) can also be applied for ordinal responses:

$$\log[P(\mathbf{Y} = j|\mathbf{x})/P(\mathbf{Y} = j+1|\mathbf{x})] = \alpha_j - \boldsymbol{\beta}'\mathbf{x} \qquad j = 1, \cdots, c-1$$

Just like its name implies, it pairs each category with its adjacent category in logit equations. The model is a special case of the baseline-category logit model commonly used for nominal response variables (i.e., no natural ordering), with reduction in the number of parameters by utilizing the ordering to obtain a common effect. It utilizes single-category probabilities rather than cumulative probabilities, so it is more natural when one wants to describe effects in terms of odds relating to particular response categories. This model received considerable attention in the 1980s and 1990s, partly because of connections with certain ordinal loglinear models. It is *very clear that the model assumes PO since the right-hand side has common effect* $\boldsymbol{\beta}$ *related to predictors* $\mathbf{x}$ *.*

Example 1.4. ***Continuation-ratio logit model***

Price (7) did a developmental toxicity study. This study administered diEGdiME (toxic substance) in distilled water to pregnant mice. Each mouse was exposed to one of five concentration levels for 10 days early in the pregnancy. Two days later, the uterine contents of the pregnant mice were examined for defects. Each fetus has three possible outcomes (nonlive, malformation, normal). The outcomes are ordered, with nonlive the least desirable result. We fitted the continuation-ratio logit models for this study

$$\log \frac{\pi_1(x_i)}{\pi_2(x_i) + \pi_3(x_i)} = \alpha_1 + \beta_1 x_i \qquad \log \frac{\pi_2(x_i)}{\pi_3(x_i)} = \alpha_2 + \beta_2 x_i$$

We want to model (1) the probability π_1 *of a nonlive fetus, and (2) the conditional probability* $\pi_2/(\pi_2+\pi_3)$ *of a malformed fetus, given that the fetus was live. Generally the continuation-ratio logits are defined as*

$$\log \frac{\pi_j}{\pi_{j+1} + \cdots + \pi_J} \qquad j = 1, \cdots, J-1$$

Let $w_j = P(\mathbf{Y} = j | \mathbf{Y} \geqslant j)$*, The above continuation-ratio logits are just ordinary logits of these conditional probabilities:* $log[w_j/(l - w_j)]$*.*

According to (8), the continuation-ratio logit form is useful when a sequential mechanism, such as survival through various age periods, determines the response outcome.

The above examples give 4 different logit models based on different responses. We also introduced proportional odds (PO) assumptions for ordinal responses and non-proportional odds (NPO) assumptions for nominal responses. The third type–Partial proportional odds models (PPO) were proposed by Peterson (9) in order to incorporate PO and NPO components. It is given as follows:

Example 1.5. ***Partial proportional odds models (PPO)***

We assume that n *independent random observations are sampled and that the responses of these observations on an ordinal variable* Y *are classified in* $k+1$ *categories with* $Y = 0, 1, \cdots, k$*. Thus, each observation has an independent multinomial distribution. The model suggested for the cumulative probabilities is*

$$C_{ij} = Pr(Y \geqslant j|\mathbf{X}_i) = \frac{1}{1 + exp(-\alpha_j - \mathbf{X}_i'\boldsymbol{\beta} - \mathbf{T}_i'\boldsymbol{\gamma}_j)} \qquad j = 1, \cdots, k$$

where $\mathbf{X}_i$ *is a* $p \times 1$ *vector containing the values of observation* i *on the full set of* p *explanatory variables,* $\boldsymbol{\beta}$ *is a* $p \times 1$ *vector of regression coefficients associated with the* p *variables in* $\mathbf{X}_i$ *(the elements of* $\boldsymbol{\beta}$ *are denoted by* $\beta_l, l = 1, \cdots, p$*),* $\mathbf{T}_i$ *is a* $q \times 1$ *vector,* $q \leqslant p$*, containing the values of observation* i *on that subset of the* p *explanatory variables for which the proportional odds assumption either is not assumed or is to be tested,and* $\boldsymbol{\gamma}_j$ *is a* $q \times 1$ *vector of regression coefficients associated with the* q *variables in* $\mathbf{T}_i$*, so that* $\mathbf{T}_i'\boldsymbol{\gamma}_j$ *an increment associated only with the* j*th cumulative logit,* $j = 1, \cdots, k$*, and* $\boldsymbol{\gamma}_1 = 0$*. The elements of* $\boldsymbol{\gamma}_j$ *are denoted by* $\gamma_{jl}, l = 1, \cdots, q$*. If* $\boldsymbol{\gamma}_j = 0$ *for all* j*, then this model reduces to the proportional odds model.*

TABLE I: Data Structure for Multinomial Experiments

Experimental Settings	Response Observations	Total Trials
$\mathbf{x}_1(x_{11}\cdots x_{1d})$	$\mathbf{Y}_1(Y_{11}\cdots Y_{1J})$	$n_1 = Y_{11} + \cdots + Y_{1J}$
$\vdots$	$\vdots$	$\vdots$
$\mathbf{x}_m(x_{m1}\cdots x_{md})$	$\mathbf{Y}_m(Y_{m1}\cdots Y_{mJ})$	$n_m = Y_{m1} + \cdots + Y_{mJ}$

PPO models sometimes may have better performance than pure PO or NPO models (10). In fact, PPO models are available to be fitted using software including SAS (11) and R (12), as well as PO or NPO models. Actually, PPO models are more general in terms of model matrix structure, which can include PO and NPO models as special cases. Our PPO model equations are slightly different from this example in terms of parameters, but they are essentially the same.

In a multinomial response experiment, observations made on different runs are assumed to be independent; there are totally $n = n_1 + n_2 + \cdots + n_m$ runs of the process, each n_i replicates based on a given experimental setting $\mathbf{x}_i$, there are totally m experimental settings. The response $\mathbf{Y}_i$ falls into one of J categories, so observations will be held in an $J \times 1$ vector, then $\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iJ})$. A schematic of the data is given in Table I.

1.2 Optimal Designs and Efficiency

Design of experiment with categorical responses is becoming increasingly popular in a rich variety of scientific disciplines. Examples include wine bitterness study (13), trauma clinical trial (14), emergence of house flies (2), polysilicon deposition study (15), toxicity study (3), and odor removal study (16), etc.

When the response is binary, generalized linear models are widely used. For optimal designs under generalized linear models, there is a growing body of literature (see (17), (18), (19) and references therein). In this case, it is known that the minimum number of experimental settings required by a non-degenerated Fisher information matrix is equal to the number of parameters (20; 21). It is also known that the experimental units should be uniformly allocated when a minimally supported design (design with the least number of experimental settings) is adopted (21).

Multinomial categorical responses frequently occur in medical experiments or social studies. While many different multinomial logistic models are proposed to analyze these types of responses (3; 4; 5; 22), the relevant results in the design literature with multinomial logistic models are meager and restricted to special classes. Zocchi and Atkinson (2) constructed a general framework of optimal designs for multinomial logistic models (which covers all the four logit models due to the frame work built by Glonek and McCullagh (23)) but with non-proportional odds only. Perevozskaya et al. (24) discussed a special class of cumulative logit models with proportional odds. The recent study Yang et al. (16) was able to obtain comprehensive results for cumulative link models (an extension of cumulative logit models with fairly general link functions suggested by McCullagh (4)) with proportional odds.

The theoretical results and real experimental examples provided by Yang et al. (16) showed that the optimal designs for multinomial responses are very different from the ones for binary responses in at least two aspects: (1) the minimum number of experimental settings required can be strictly less than the number of parameters; (2) even for a minimally supported design, generally uniform allocation is not optimal anymore. A natural question is whether those findings are true for all multinomial logistic models, or just for cumulative link models with proportional odds. In this paper, we aim to obtain general results on optimal designs for all multinomial logistic models commonly used, which covers all the four logit models and all the three odds structure.

Now we introduce some basic concepts of optimal designs. A design point is a specification of an experimental setting or some combination of explanatory factors. $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$ is the ith experimental setting or the ith level combination of the d factors. Design points have the form $(\mathbf{x}_i), i = 1, \cdots, m$ and are usually distinct. The goal of optimal design is to reduce the cost or improve efficiency. It needs to find suitable weight or experimental runs allocated to each design point in order to optimize some objective function for a certain design criterion. For example, approximate design having m design points can be written as:

$$\xi_{approx} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ \\ w_1 & w_2 & \cdots & w_m \end{pmatrix} \tag{1.1}$$

where the weight w_i is allocated to the ith support point $\mathbf{x}_i$, and $w_1 + w_2 + \cdots + w_m = 1, 0 < w_i < 1, i = 1, \cdots, m$. If a support point has weight $1/3$, and $n = 6$ total runs are to be made, two runs should be conducted at that support point.

In practice all designs are exact, i.e. the number of experimental runs on each support point $\mathbf{x}_i$ is an integer. An exact design can be regarded as having a support point $\mathbf{x}_i$ allocated with n_i runs for all of m design points, $n_1 + n_2 + \cdots + n_m = n$ and n is the total number of runs that will be made. The exact design can be written in Equation 1.2. Although exact designs can be obtained from rounding approximate designs, it could lose efficiency when the total number of runs is small. So we use a different strategy to obtain exact design.

$$\xi_{exact} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_m \\ \\ n_1 & n_2 & \cdots & n_m \end{pmatrix} \tag{1.2}$$

Most of optimal design methods maximize or minimize a function of the Fisher information matrix about parameter estimates, which are dependent on the models. D-optimality maximizes the determinant of Fisher information matrix, which is equivalent to minimizing the generalized variance of the parameter estimates. D-optimality performs well according to other optimal criteria, so it is the most popular optimal design.

For a general non-linear model (including but not limited to multinomial logistic model), a design ξ is called D-optimal if it maximizes the objective function:

$$\phi(\xi) = |M(\xi, \boldsymbol{\theta})|$$

where $| : |$ denotes determinant. M is a Fisher information matrix containing ξ and $\boldsymbol{\theta}$ due to non-linearity in the parameters, so any optimal design will depend on the values of $\boldsymbol{\theta}$. When parameter values are assumed or given in advance, a D-optimal design is called locally D-optimal design.

In this thesis, a design ξ will be assessed via its D-efficiency, which is defined as

$$D_{eff} = \left(\frac{|M(\xi, \boldsymbol{\theta})|}{|M(\xi^*, \boldsymbol{\theta})|} \right)^{1/p}$$

where p is the number of parameters, and ξ^* is a D-optimal design (18).

If the relative performances of two designs ξ_1 and ξ_2 are of interest, their relative D-efficiency can be calculated (18), and is given by

$$D_{Reff} = \left(\frac{|M(\xi_1, \boldsymbol{\theta})|}{|M(\xi_2, \boldsymbol{\theta})|} \right)^{1/p}$$

CHAPTER 2

UNIFIED FORM OF MULTINOMIAL LOGISTIC MODELS AND ITS FISHER INFORMATION MATRIX

2.1 Unified Form of Multinomial Logistic Models

We discussed different multinomial logistic models in Section 1.1. Since PPO models are more general in terms of model structure, which can include PO and NPO models as special cases, we write the four logit models (baseline-category, cumulative, adjacent-categories, and continuation-ratio) in the same format with PPO structure as follows:

$$
\begin{aligned}
\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) &= \sum_{k=1}^{p_j} h_{jk}(\mathbf{x}_i)\beta_{jk} + \sum_{k=1}^{p_c} h_k(\mathbf{x}_i)\zeta_k\ , && \text{baseline} \\
\log\left(\frac{\pi_{i1}+\cdots+\pi_{ij}}{\pi_{i,j+1}+\cdots+\pi_{iJ}}\right) &= \sum_{k=1}^{p_j} h_{jk}(\mathbf{x}_i)\beta_{jk} + \sum_{k=1}^{p_c} h_k(\mathbf{x}_i)\zeta_k\ , && \text{cumulative} \\
\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) &= \sum_{k=1}^{p_j} h_{jk}(\mathbf{x}_i)\beta_{jk} + \sum_{k=1}^{p_c} h_k(\mathbf{x}_i)\zeta_k\ , && \text{adjacent} \\
\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}+\cdots+\pi_{iJ}}\right) &= \sum_{k=1}^{p_j} h_{jk}(\mathbf{x}_i)\beta_{jk} + \sum_{k=1}^{p_c} h_k(\mathbf{x}_i)\zeta_k\ , && \text{continuation}
\end{aligned}
$$

where π_{ij} is the probability that the response $\mathbf{Y}_i$ falls into the jth category at the ith experimental setting, $i = 1, \ldots, m$, $j = 1, \cdots, J-1$, so there are $m(J-1)$ equations for each model. $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$ is the ith experimental setting or the ith level combination of the d factors, $i = 1, \ldots, m$. Here we don't use common form of linear predictors $\alpha_j + \sum_{k=1}^{d_1} x_{ik}\beta_{jk} + \sum_{k=d_1+1}^{d} x_{ik}\zeta_k$, instead, we use function of $\mathbf{x}_i$ to make models more general.

Actually, the linear predictors of the above models can be written as vector form:

$$\eta_{ij} = \sum_{k=1}^{p_j} h_{jk}(\mathbf{x}_i)\beta_{jk} + \sum_{k=1}^{p_c} h_k(\mathbf{x}_i)\zeta_k = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$$

where $\mathbf{h}_j^T(\cdot) = (h_{j1}(\cdot), \ldots, h_{jp_j}(\cdot))$ are known functions to determine the predictors for only jth equation, $\boldsymbol{\beta}_j = (\beta_{j1}, \ldots, \beta_{jp_j})^T$ consists of p_j unknown parameters for only jth equation, $\mathbf{h}_c^T(\cdot) = (h_1(\cdot), \ldots, h_{p_c}(\cdot))$ are known functions to determine common predictors for all of equations, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{p_c})^T$ consists of p_c unknown common parameters for all of the equations.

Following (1) and (2), we rewrite these four logit models into a unified form

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta}, \qquad i = 1, \cdots, m \tag{2.1}$$

where $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})^T$,

$$\mathbf{C}^T = \begin{pmatrix} 1 & & & & -1 & & & & 0 \\ & 1 & & & & -1 & & & 0 \\ & & \ddots & & & & \ddots & & \vdots \\ & & & 1 & & & & -1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times(2J-1)} \tag{2.2}$$

is the same for all the four logit models, while $\mathbf{L}$ does take different forms as follows:

$$\mathbf{L}_{baseline} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{(2J-1)\times J} \qquad \mathbf{L}_{cumulative} = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & 1 & \cdots & 1 & 0 \\ 0 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{(2J-1)\times J}$$

$$\mathbf{L}_{continuation} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & \cdots & 1 & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{(2J-1)\times J} \qquad \mathbf{L}_{adjacent} = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 0 & 1 & & & \\ 0 & & 1 & & \\ \vdots & & & \ddots & \\ 0 & & & & 1 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{(2J-1)\times J}$$

for partial proportional odds models, the model matrix is

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{h}_1^T(\mathbf{x}_i) & \mathbf{0}^T & \cdots & \mathbf{0}^T & \mathbf{h}_c^T(\mathbf{x}_i) \\ \mathbf{0}^T & \mathbf{h}_2^T(\mathbf{x}_i) & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T & \mathbf{h}_c^T(\mathbf{x}_i) \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \mathbf{h}_{J-1}^T(\mathbf{x}_i) & \mathbf{h}_c^T(\mathbf{x}_i) \\ \mathbf{0}^T & \cdots & \cdots & \mathbf{0}^T & \mathbf{0}^T \end{pmatrix}_{J\times p} \tag{2.3}$$

and the parameter vector

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{J-1}, \boldsymbol{\zeta})^T \tag{2.4}$$

consists of $p = p_1 + \cdots + p_{J-1} + p_c$ unknown parameters in total.

Non-proportional odds models can be regarded as one degenerated case of partial proportional odds models, the model matrix is then

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{h}_1^T(\mathbf{x}_i) & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \mathbf{h}_2^T(\mathbf{x}_i) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}^T \\ \mathbf{0}^T & \cdots & \mathbf{0}^T & \mathbf{h}_{J-1}^T(\mathbf{x}_i) \\ \mathbf{0}^T & \cdots & \cdots & \mathbf{0}^T \end{pmatrix}_{J\times p} \tag{2.5}$$

and the parameter vector reduces to

$$\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{J-1})^T \tag{2.6}$$

which consists of $p = p_1 + \cdots + p_{J-1}$ unknown parameters in total.

Proportional odds models can be regarded as another degenerated case of partial proportional odds models. In this case, all of $\mathbf{h}_j^T(\mathbf{x}_i)$ functions are replaced with $1's$. The corresponding model matrix is

$$\mathbf{X}_i = \begin{pmatrix} 1 & 0 & \cdots & 0 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & \cdots & 0 & 1 & \mathbf{h}_c^T(\mathbf{x}_i) \\ 0 & 0 & \cdots & 0 & \mathbf{0}^T \end{pmatrix}_{J \times p} \tag{2.7}$$

and the parameter vector

$$\boldsymbol{\theta} = (\beta_1, \beta_2, \cdots, \beta_{J-1}, \boldsymbol{\zeta})^T \tag{2.8}$$

consists of $p = J - 1 + p_c$ unknown parameters in total. The previous $\boldsymbol{\beta}_j$ reduces to β_j serving as the cut-off point in this case. Typically, the notation α_j is used in the literature to express cut-off points. In this paper, we use β_j for consistency.

Now we give some examples for the above cases.

Example 2.1. ***Non-proportional odds models***

An example of $X_i\boldsymbol{\theta}$ for non-proportional odds models can be found in (2), which takes the form of

$$\mathbf{X}_i\boldsymbol{\theta} = \begin{pmatrix} 1 & x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_i \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \end{pmatrix}$$

$\square$

Note that different predictor functions $\mathbf{h}_j^T(\mathbf{x}_i)$ can be used for different logit equations respectively. Example 2.2 provides an example of proportional odds models.

Example 2.2. ***Proportional odds models***

An example with J=3 can be found in (24), which essentially takes the form of

$$X_i\boldsymbol{\theta} = \begin{pmatrix} 1 & 0 & x_i \\ 0 & 1 & x_i \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \zeta_1 \end{pmatrix}$$

$\square$

Example 2.3. ***Partial proportional odds models***

Suppose $d = 4$ *and* $\mathbf{x}_i = (x_{i2}, x_{i3}, x_{i4}, x_{i5})^T$*. An illustrative example of* $\mathbf{X}_i\boldsymbol{\theta}$ *with* $J = 3$ *(from (12) page 176) is*

$$\mathbf{X}_i\boldsymbol{\theta} = \begin{pmatrix} 1 & x_{i3} & x_{i5} & 0 & 0 & 0 & x_{i2} & x_{i4} \\ 0 & 0 & 0 & 1 & x_{i3} & x_{i5} & x_{i2} & x_{i4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \zeta_1 \\ \zeta_2 \end{pmatrix}$$

□

Note that in all of the above cases, $\pi_{i1} + \cdots + \pi_{iJ} = 1$ implies that $\eta_{iJ} = 0$ and the last row of $\mathbf{X}_i$ is all 0's.

2.2 Fisher Information Matrix for Multinomial Logistic Models

Following (1) and (2), the multinomial logistic models take the unified form Equation 2.1:

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta} \qquad i = 1, \cdots, m$$

for nominal, ordinal and hierarchical multinomial responses. Note that Equation 2.1 covers all of PO, NPO and PPO models. We need the following formulas of matrix differentiation (see, for example, (25) (2008, Chapter 17))

$$\begin{aligned}
\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} &= \left(\frac{\partial y_i}{\partial x_j}\right)_{ij} \\
\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}^T} &= \mathbf{A} \\
\frac{\partial \mathbf{z}}{\partial \mathbf{x}^T} &= \frac{\partial \mathbf{z}}{\partial \mathbf{y}^T} \cdot \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} \\
\frac{\partial \log \mathbf{y}}{\partial \mathbf{x}^T} &= [\mathrm{d}iag(\mathbf{y})]^{-1} \frac{\partial \mathbf{y}}{\partial \mathbf{x}^T}
\end{aligned}$$

where $\mathbf{x} = (x_i)_i,\ \mathbf{y} = (y_i)_i,\ \mathbf{z}$ and thus $\log \mathbf{y} = (\log y_i)_i$ are vectors and $\mathbf{A}$ is a constant matrix. Then

$$\begin{aligned}
\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} &= \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\eta}_i^T} \cdot \frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\theta}^T} \\
&= \left(\frac{\partial \boldsymbol{\eta}_i}{\partial \boldsymbol{\pi}_i^T}\right)^{-1} \cdot \mathbf{X}_i \\
&= \left(\frac{\partial [\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i)]}{\partial [\log(\mathbf{L}\boldsymbol{\pi}_i)]^T} \cdot \frac{\partial [\log(\mathbf{L}\boldsymbol{\pi}_i)]}{\partial [\mathbf{L}\boldsymbol{\pi}_i]^T} \cdot \frac{\partial [\mathbf{L}\boldsymbol{\pi}_i]}{\partial \boldsymbol{\pi}_i^T}\right)^{-1} \cdot \mathbf{X}_i \\
&= \left(\mathbf{C}^T [\mathrm{d}iag(\mathbf{L}\boldsymbol{\pi}_i)]^{-1} \mathbf{L}\right)^{-1} \mathbf{X}_i
\end{aligned}$$

That is,

$$\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} = (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i \tag{2.9}$$

where $\mathbf{D}_i = \mathrm{d}iag(\mathbf{L}\boldsymbol{\pi}_i)$. Note that Equation 2.9 is due to the linear form of $\mathbf{X}_i$ and $\boldsymbol{\theta}$.

Suppose for distinct $\mathbf{x}_i, i = 1, \cdots, m$, we have independent multinomial response

$$\mathbf{Y}_i = (Y_{i1}, \cdots, Y_{iJ})^T \sim \text{M}ultinomial(n_i; \pi_{i1}, \cdots, \pi_{iJ})$$

where $n_i = \sum_{j=1}^{J} Y_{ij}$. Then the log-likelihood for the multinomial model is

$$\begin{aligned} l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}) \\ &= \log \prod_{i=1}^{m} \frac{n_i!}{Y_{i1}! \cdots Y_{iJ}!} \pi_{i1}^{Y_{i1}} \cdots \pi_{iJ}^{Y_{iJ}} \\ &= constant + \sum_{i=1}^{m} \mathbf{Y}_i^T \log \boldsymbol{\pi}_i \end{aligned}$$

where $\log \boldsymbol{\pi}_i = (\log \pi_{i1}, \cdots, \log \pi_{iJ})^T$. Then the score vector

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\theta}^T} &= \sum_{i=1}^{m} \mathbf{Y}_i^T \text{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\ \frac{\partial l}{\partial \boldsymbol{\theta}} &= (\frac{\partial l}{\partial \boldsymbol{\theta}^T})^T = \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \text{d}iag(\boldsymbol{\pi}_i)^{-1} \mathbf{Y}_i \end{aligned}$$

Lemma 2.1.

$$\boldsymbol{\pi}_i{}^T diag(\boldsymbol{\pi}_i)^{-1} (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i = \mathbf{0}^T$$

Proof of Lemma 2.1: Since

$$\mathbf{C}^T = \begin{pmatrix} * & * & \cdots & 0 \\ * & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} * & * & \cdots & * \\ * & * & \cdots & * \\ & \ddots & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$$

$$\mathbf{D}_i^{-1} = \mathrm{d}iag(\mathbf{L}\boldsymbol{\pi}_i)^{-1} = \begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & \frac{1}{\mathbf{1}^T\boldsymbol{\pi}_i} \end{pmatrix} = \begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & \cdots & 0 \\ & \ddots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

then

$$\mathbf{D}_i^{-1}\mathbf{L} = \begin{pmatrix} * & \cdots & * \\ * & \cdots & * \\ & \ddots & \\ & \mathbf{1}^T & \end{pmatrix} \text{ and } \mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L} = \begin{pmatrix} * & \cdots & * \\ * & \cdots & * \\ & \ddots & \\ & \mathbf{1}^T & \end{pmatrix}$$

Rewrite $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_J)$. Then $\mathbf{1}^T\boldsymbol{a}_1 = \cdots = \mathbf{1}^T\boldsymbol{a}_{J-1} = 0$ and $\mathbf{1}^T\boldsymbol{a}_J = 1$ (just look at the last row of $\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L}$). Since $\boldsymbol{\pi}_i^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} = (1, \cdots, 1)$, then

$$\boldsymbol{\pi}_i{}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (1 \cdots 1)(\boldsymbol{a}_1, \cdots, \boldsymbol{a}_J) = (0, \cdots, 0, 1)$$

Recall that the last row of $\mathbf{X}_i$ is all 0. Then

$$\boldsymbol{\pi}_i{}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i = \mathbf{0}^T$$

□

As a direct conclusion of Lemma 2.1,

$$E(\frac{\partial l}{\partial \boldsymbol{\theta}^T}) = \sum_{i=1}^{m} n_i \boldsymbol{\pi}_i{}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i = \mathbf{0}^T$$

Then the Fisher information matrix is, see (26) (1995, Section 2.3.1)

$$\begin{aligned}
\mathbf{F} &= \mathrm{C}ov\left(\frac{\partial l}{\partial \boldsymbol{\theta}}, \frac{\partial l}{\partial \boldsymbol{\theta}}\right) = E\left(\frac{\partial l}{\partial \boldsymbol{\theta}} \cdot \frac{\partial l}{\partial \boldsymbol{\theta}^T}\right) \\
&= E\left(\sum_{i=1}^{m}(\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}\mathbf{Y}_i \cdot \sum_{j=1}^{m} \mathbf{Y}_j^T \mathrm{d}iag(\boldsymbol{\pi}_j)^{-1}\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T}\right) \\
&= E\left(\sum_{i=1}^{m}\sum_{j=1}^{m}(\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}\mathbf{Y}_i\mathbf{Y}_j^T \mathrm{d}iag(\boldsymbol{\pi}_j)^{-1}\frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T}\right)
\end{aligned}$$

Since $\mathbf{Y}_i$'s follow independent multinomial distribution, then

$$E(\mathbf{Y}_i\mathbf{Y}_i^T) = \begin{pmatrix} n_i(n_i-1)\pi_{i1}^2 + n_i\pi_{i1} & \cdots & n_i(n_i-1)\pi_{is}\pi_{it} \\ \vdots & \ddots & \vdots \\ n_i(n_i-1)\pi_{is}\pi_{it} & \cdots & n_i(n_i-1)\pi_{iJ}^2 + n_i\pi_{iJ} \end{pmatrix}$$

$$= n_i(n_i-1)\boldsymbol{\pi}_i\boldsymbol{\pi}_i^T + n_i \mathrm{d}iag(\boldsymbol{\pi}_i)$$

On the other hand, for $i \neq j$,

$$E(\mathbf{Y}_i \mathbf{Y}_j^T) = E(\mathbf{Y}_i) \cdot E(\mathbf{Y}_j^T) = n_i n_j \boldsymbol{\pi}_i \boldsymbol{\pi}_j^T$$

Then the Fisher information matrix

$$\begin{aligned}
\mathbf{F} &= \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} n_i(n_i - 1) \boldsymbol{\pi}_i \boldsymbol{\pi}_i^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\
&+ \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} n_i \mathrm{d}iag(\boldsymbol{\pi}_i) \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} \\
&+ \sum_{i \neq j} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} n_i n_j \boldsymbol{\pi}_i \boldsymbol{\pi}_j^T \mathrm{d}iag(\boldsymbol{\pi}_j)^{-1} \frac{\partial \boldsymbol{\pi}_j}{\partial \boldsymbol{\theta}^T} \\
&\triangleq (a) + (b) + (c)
\end{aligned}$$

where

$$\begin{aligned}
(b) &= \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} n_i \\
(a) + (c) &= \left[\sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \boldsymbol{\pi}_i n_i \right] \left[\sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \boldsymbol{\pi}_i n_i \right]^T \\
&- \sum_{i=1}^{m} (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} n_i \boldsymbol{\pi}_i {\boldsymbol{\pi}_i}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T}
\end{aligned}$$

Actually, let

$$\mathbf{E}_i = {\boldsymbol{\pi}_i}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} = {\boldsymbol{\pi}_i}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i$$

which is $\mathbf{0}^T$ for each i according to Lemma 2.1. Then

$$(a) + (c) = \left[\sum_{i=1}^{m} n_i \mathbf{E}_i^T\right] \left[\sum_{i=1}^{m} n_i \mathbf{E}_i^T\right]^T - \sum_{i=1}^{m} n_i \mathbf{E}_i^T \mathbf{E}_i = \mathbf{0}_{J\times J}$$

The arguments above have proved the following theorem:

Theorem 2.1. *Consider the multinomial logistic model Equation 2.1 with independent observations. The Fisher information matrix*

$$\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{F}_i \tag{2.10}$$

where

$$\mathbf{F}_i = (\frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T})^T diag(\boldsymbol{\pi}_i)^{-1} \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T},\ \frac{\partial \boldsymbol{\pi}_i}{\partial \boldsymbol{\theta}^T} = (\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \mathbf{X}_i,\ \mathbf{D}_i = diag(\mathbf{L}\boldsymbol{\pi}_i)$$

2.3 Reparametrization and D-optimaltiy

Consider a general linear predictor of generalized linear models at the ith experimental setting

$$\eta_{ij} = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \qquad j = 1, \cdots, J-1 \tag{2.11}$$

The model contains the following parameters: $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^T, \cdots, \boldsymbol{\beta}_{J-1}^T, \boldsymbol{\zeta})^T$, which can be rewritten as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$. Suppose we reparametrize the model with a different set of parameters: $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_p)^T$, such that,

$$\theta_l = h_l(\boldsymbol{\vartheta}), \qquad l = 1, \cdots, p$$

Suppose the map $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta}) = (h_1(\boldsymbol{\vartheta}), \ldots, h_p(\boldsymbol{\vartheta}))^T$ is one-to-one, h_l's are differentiable, and $|\mathbf{J}| \neq 0$, where $\mathbf{J} = (\frac{\partial h_i(\boldsymbol{\vartheta})}{\partial \vartheta_j})_{ij}$ is the $p \times p$ Jacobian matrix. Given a design $\xi = \{(\mathbf{x}_i, w_i), i = 1, \ldots, m\}$ with distinct experimental settings $\mathbf{x}_i$ and proportions $w_i \in [0, 1]$, according to (26) (1995, page 115), the Fisher information matrix $I_\xi(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}$ and the Fisher information matrix $I_\xi(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta})$ satisfy

$$I_\xi(\boldsymbol{\vartheta}) = \mathbf{J}^T I_\xi[\boldsymbol{\theta}(\boldsymbol{\vartheta})]\mathbf{J} \tag{2.12}$$

Then $|I_\xi(\boldsymbol{\vartheta})| = |\mathbf{J}|^2 \cdot |I_\xi[\boldsymbol{\theta}(\boldsymbol{\vartheta})]|$. Note that $\mathbf{J}$ contains no design points but parameters. A locally D-optimal design maximizing $|I_\xi(\boldsymbol{\vartheta})|$ also maximizes $|I_\xi[\boldsymbol{\theta}(\boldsymbol{\vartheta})]|$. That is, finding D-optimal designs for parameters $\boldsymbol{\vartheta}$ or $\boldsymbol{\theta}$ are equivalent.

In terms of Bayesian D-optimal criterion, if a prior distribution of $\boldsymbol{\vartheta}$ is available, it induces a prior distribution of $\boldsymbol{\theta}$ since $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\vartheta})$ is one-to-one.

$$\begin{aligned} E_{\boldsymbol{\vartheta}} \log |I_\xi(\boldsymbol{\vartheta})| &= E_{\boldsymbol{\vartheta}} \log \left|\mathbf{J}^T I_\xi[\boldsymbol{\theta}(\boldsymbol{\vartheta})]\mathbf{J}\right| \\ &= E_{\boldsymbol{\vartheta}} \log |\mathbf{J}|^2 + E_{\boldsymbol{\vartheta}} \log |I_\xi[\boldsymbol{\theta}(\boldsymbol{\vartheta})]| \\ &= E_{\boldsymbol{\vartheta}} \log |\mathbf{J}|^2 + E_{\boldsymbol{\theta}} \log |I_\xi(\boldsymbol{\theta})| \end{aligned}$$

Therefore, a Bayesian D-optimal design that maximizes $E_{\boldsymbol{\theta}} \log |I_\xi(\boldsymbol{\theta})|$ also maximizes $E_{\boldsymbol{\vartheta}} \log |I_\xi(\boldsymbol{\vartheta})|$.

Example 2.4. *Perevozskaya et al. (24) Consider a proportional odds model as follows:*

$$\log \frac{\gamma_j(x)}{1 - \gamma_j(x)} = \frac{x - \alpha'_j}{\beta'} \qquad j = 2, \ldots, J \tag{2.13}$$

where $\gamma_j(x) = P(Y \geqslant j|x)$. Following (24), $j = 2, \ldots, J$ are used for the response category related equations, which is actually equivalent to $j = 1, \ldots, J-1$ equations.

If we reparametrize this model into:

$$\log \frac{\gamma_j(x)}{1-\gamma_j(x)} = \alpha_j + \beta x \qquad j = 2, \ldots, J \tag{2.14}$$

Following (2) with $J = 3$,

$$\mathbf{X}_i\boldsymbol{\theta} = \begin{pmatrix} 1 & 0 & x_i \\ 0 & 1 & x_i \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_2 \\ \alpha_3 \\ \beta \end{pmatrix}$$

where $\boldsymbol{\theta} = (\alpha_2, \alpha_3, \beta)^T$. Let $\boldsymbol{\vartheta} = (\alpha_2', \alpha_3', \beta')^T$. Note that $\beta = \frac{1}{\beta'}, \alpha_2 = -\frac{\alpha_2'}{\beta'}, \alpha_3 = -\frac{\alpha_3'}{\beta'}$. Then

$$\mathbf{J} = \begin{pmatrix} -\frac{1}{\beta'} & 0 & \frac{\alpha_2'}{\beta'^2} \\ 0 & -\frac{1}{\beta'} & \frac{\alpha_3'}{\beta'^2} \\ 0 & 0 & -\frac{1}{\beta'^2} \end{pmatrix}$$

Based on Theorem 2.1, at the ith design point, the Fisher information

$$\begin{aligned} I_i(\boldsymbol{\theta}) &= \mathbf{X}_i^T[(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}]^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}\mathbf{X}_i \\ &= \begin{pmatrix} \frac{\pi_{i1}\pi_{i2,3}^2\pi_{i1,2}}{\pi_{i2}} & -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \pi_{i1}\pi_{i1,2}\pi_{i2,3}x_i \\ -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \frac{\pi_{i1,2}^2\pi_{i2,3}\pi_{i3}}{\pi_{i2}} & \pi_{i3}\pi_{i1,2}\pi_{i2,3}x_i \\ \pi_{i1}\pi_{i1,2}\pi_{i2,3}x_i & \pi_{i3}\pi_{i1,2}\pi_{i2,3}x_i & (\pi_{i1}\pi_{i2,3}^2 + \pi_{i2}(\pi_{i1}-\pi_{i3})^2 + \pi_{i1,2}^2\pi_{i3})x_i^2 \end{pmatrix} \end{aligned}$$

Here π_{ij} is the probability that the response falls into the jth catogory at the ith design point, and $\pi_{ij,k} = \pi_{ij} + \pi_{ik}$. According to Equation 2.12, $I_i(\boldsymbol{\vartheta}) = \mathbf{J}^T I_i(\boldsymbol{\theta})\mathbf{J}$ equals to

$$\begin{pmatrix} \frac{\pi_{i1}\pi_{i2,3}^2\pi_{i1,2}}{\beta'^2\pi_{i2}} & -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\beta'^2\pi_{i2}} & \frac{1}{\beta'^3}[-\frac{\alpha'_2\pi_{i1}\pi_{i2,3}^2\pi_{i1,2}}{\pi_{i2}} + \frac{\alpha'_3\pi_{i1}\pi_{i2,3}\pi_{i1,2}\pi_{i3}}{\pi_{i2}} + \pi_{i1}\pi_{i1,2}\pi_{i2,3}x_i] \\ -\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}}{\beta'^2\pi_{i2}} & \frac{\pi_{i1,2}^2\pi_{i2,3}\pi_{i3}}{\beta'^2\pi_{i2}} & \frac{1}{\beta'^3}[\frac{\alpha'_2\pi_{i1}\pi_{i2,3}\pi_{i1,2}\pi_{i3}}{\pi_{i2}} - \frac{\alpha'_3\pi_{i1,2}^2\pi_{i2,3}\pi_{i3}}{\pi_{i2}} + \pi_{i3}\pi_{i1,2}\pi_{i2,3}x_i] \\ \dots & \dots & \dots \end{pmatrix}$$

The last row is not expressed due to limited space. Since the matrix is symmetric, $I_i(\boldsymbol{\vartheta})_{31} = I_i(\boldsymbol{\vartheta})_{13}$, and $I_i(\boldsymbol{\vartheta})_{32} = I_i(\boldsymbol{\vartheta})_{23}$. The element of last row and last column is

$$\begin{aligned} I_i(\boldsymbol{\vartheta})_{33} = &\frac{\pi_{i1}\pi_{i1,2}\pi_{i2,3}\pi_{i3}(\alpha'_2 - \alpha'_3)^2}{\beta'^4\pi_{i2}} + \frac{\pi_{i1,2}\pi_{i2,3}(\alpha'_2\pi_{i1} + \alpha'_3\pi_{i3})(1-2x_i)}{\beta'^4} \\ &+ \frac{(\pi_{i1}\pi_{i2,3}^2 + \pi_{i2}(\pi_{i1} - \pi_{i3})^2 + \pi_{i1,2}^2\pi_{i3})x_i^2)}{\beta'^4} \end{aligned}$$

It can be verified that $I_i(\boldsymbol{\vartheta})$ above is equal to the corresponding one given by (24). The transformation of Equation 2.12 is confirmed in this case. Therefore, the D-optimal designs for Model (Equation 2.13) and Model (Equation 2.14) are the same.

CHAPTER 3

POSITIVE DEFINITENESS OF THE FISHER INFORMATION MATRIX

3.1 Reformulation of Fisher Information Matrix as $\mathbf{HUH}^T$

We start from the unified form Equation 2.1 and Equation 2.10. In this subsection, we do not need to specify the forms of $\mathbf{L}$, $\mathbf{X}_i$ and $\boldsymbol{\theta}$ in the model Equation 2.1. Therefore, the formulas derived are good for all of the multinomial logistic models we discussed previously. It could include 10 models $(4 \times 3 - 2)$, since there are 4 kinds of logit models combined with 3 types of odds assumptions, while the baseline logit model combined with PO or PPO doesn't follow correct logic.

According to Theorem 2.1, we need to calculate $(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1}$. Actually, if we denote

$$(\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L})^{-1} \stackrel{\triangle}{=} (\mathbf{c}_{i1}, \ldots, \mathbf{c}_{iJ})$$

where $\mathbf{c}_{ij}$ is a $J \times 1$ vector, $j = 1, \ldots, J$, we can rewrite $\mathbf{F}_i$ into a simpler form as a corollary of Theorem 2.1:

Corollary 3.1. *Under the setup of Theorem 2.1, the Fisher information at the ith design point*

$$\mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i \tag{3.1}$$

where

$$\mathbf{U}_i = \begin{pmatrix} u_{11}(\boldsymbol{\pi}_i) & \cdots & u_{1,J-1}(\boldsymbol{\pi}_i) & 0 \\ \vdots & \ddots & \vdots & \vdots \\ u_{J-1,1}(\boldsymbol{\pi}_i) & \cdots & u_{J-1,J-1}(\boldsymbol{\pi}_i) & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} \tag{3.2}$$

and

$$u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T diag(\boldsymbol{\pi}_i)^{-1}\mathbf{c}_{it} \tag{3.3}$$

for $s, t = 1, \ldots, J-1$.

Because the last row of $\mathbf{X}_i$ consists of all zeros, the entries in the last row and last column of $\mathbf{U}_i$ actually won't make any difference. In order to simplify the notations in this chapter, we rewrite

$$\begin{aligned}
\mathbf{h}_{ji} &\triangleq \mathbf{h}_j(\mathbf{x}_i) \qquad j = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
\mathbf{h}_{ci} &\triangleq \mathbf{h}_c(\mathbf{x}_i) \qquad i = 1, \ldots, m \\
u_{sti} &\triangleq u_{st}(\boldsymbol{\pi}_i) \qquad s, t = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{s\cdot i} &\triangleq \sum_{t=1}^{J-1} u_{sti} \qquad s = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{\cdot ti} &\triangleq \sum_{s=1}^{J-1} u_{sti} \qquad t = 1, \ldots, J-1; \quad i = 1, \ldots, m \\
u_{\cdot\cdot i} &\triangleq \sum_{s=1}^{J-1} \sum_{t=1}^{J-1} u_{sti} \qquad i = 1, \ldots, m
\end{aligned}$$

Based on Corollary 3.1 and when $\mathbf{X}_i$ takes partial proportional odds form Equation 2.3

$$\mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i = \begin{pmatrix} u_{11i}\mathbf{h}_{1i}\mathbf{h}_{1i}^T & \cdots & u_{1,J-1,i}\mathbf{h}_{1i}\mathbf{h}_{J-1,i}^T & u_{1\cdot i}\mathbf{h}_{1i}\mathbf{h}_{ci}^T \\ \vdots & \ddots & \vdots & \vdots \\ u_{J-1,1,i}\mathbf{h}_{J-1,i}\mathbf{h}_{1i}^T & \cdots & u_{J-1,J-1,i}\mathbf{h}_{J-1,i}\mathbf{h}_{J-1,i}^T & u_{J-1\cdot i}\mathbf{h}_{J-1,i}\mathbf{h}_{ci}^T \\ u_{\cdot 1i}\mathbf{h}_{ci}\mathbf{h}_{1i}^T & \cdots & u_{\cdot J-1,i}\mathbf{h}_{ci}\mathbf{h}_{J-1,i}^T & u_{\cdot\cdot i}\mathbf{h}_{ci}\mathbf{h}_{ci}^T \end{pmatrix}$$

Then we have

$$\mathbf{F} = \sum_{i=1}^m n_i \mathbf{F}_i \tag{3.4}$$

$$= \begin{pmatrix} \sum_{i=1}^m n_i u_{11i}\mathbf{h}_{1i}\mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{1,J-1,i}\mathbf{h}_{1i}\mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{1\cdot i}\mathbf{h}_{1i}\mathbf{h}_{ci}^T \\ \vdots & \ddots & \vdots & \vdots \\ \sum_{i=1}^m n_i u_{J-1,1,i}\mathbf{h}_{J-1,i}\mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{J-1,J-1,i}\mathbf{h}_{J-1,i}\mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{J-1\cdot i}\mathbf{h}_{J-1,i}\mathbf{h}_{ci}^T \\ \sum_{i=1}^m n_i u_{\cdot 1i}\mathbf{h}_{ci}\mathbf{h}_{1i}^T & \cdots & \sum_{i=1}^m n_i u_{\cdot J-1,i}\mathbf{h}_{ci}\mathbf{h}_{J-1,i}^T & \sum_{i=1}^m n_i u_{\cdot\cdot i}\mathbf{h}_{ci}\mathbf{h}_{ci}^T \end{pmatrix} \tag{3.5}$$

$$= \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix} \begin{pmatrix} \mathbf{U}_{11} & \cdots & \mathbf{U}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1} & \cdots & \mathbf{U}_{J-1,J-1} \end{pmatrix} \begin{pmatrix} \mathbf{H}_1^T & & & \mathbf{H}_c^T \\ & \ddots & & \vdots \\ & & \mathbf{H}_{J-1}^T & \mathbf{H}_c^T \end{pmatrix} \tag{3.6}$$

The arguments above have proved the following theorem:

Theorem 3.1. *Consider the multinomial logistic model Equation 2.1 with independent observations. The Fisher information matrix*

$$\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T \tag{3.7}$$

where

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix} \qquad \mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \cdots & \mathbf{U}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1} & \cdots & \mathbf{U}_{J-1,J-1} \end{pmatrix}$$

and

$$\mathbf{H}_j = \begin{pmatrix} \mathbf{h}_j(\mathbf{x}_1) & \cdots & \mathbf{h}_j(\mathbf{x}_m) \end{pmatrix} \qquad j = 1, \ldots, J-1$$

$$\mathbf{H}_c = \begin{pmatrix} \mathbf{h}_c(\mathbf{x}_1) & \cdots & \mathbf{h}_c(\mathbf{x}_m) \end{pmatrix}$$

$$\mathbf{U}_{st} = \begin{pmatrix} n_1 u_{st1} & & \\ & \ddots & \\ & & n_m u_{stm} \end{pmatrix} \qquad s, t = 1, \ldots, J-1$$

are block matrices.

The above $\mathbf{H}$ *is for PPO models. If models take NPO or PO form, then*

$$\mathbf{H}_{NPO} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \end{pmatrix} \qquad \mathbf{H}_{PO} = \begin{pmatrix} \mathbf{1}^T & & \\ & \ddots & \\ & & \mathbf{1}^T \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix}$$

The other expressions keep unchanged.

A singular Fisher information matrix may lead to unavailability of unbiased estimators of parameters with finite variance (27). In this section, we study when the Fisher information matrix is nonsingular, or equivalently, positive definite, under general multinomial logistic models. We start from the unified form Equation 2.1, which covers all of 10 multinomial logistic models we mentioned.

3.2 Positive Definiteness of U

Recall that the $m(J-1) \times m(J-1)$ matrix $\mathbf{U}$ in Theorem 3.1 consists of $n_i u_{st}(\boldsymbol{\pi}_i) = n_i \mathbf{c}_{is}^T \text{d}iag(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it}$ with $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$, $i = 1, \ldots, m$ and $s, t = 1, \ldots, J-1$. For typical applications, we assume $0 < \pi_{ij} < 1$ for all $i = 1, \ldots, m$, $j = 1, \ldots, J$. In order to simplify the notations, we first assume $n_i > 0$ for $i = 1, \ldots, m$.

Theorem 3.2. *Assume that $\pi_{ij} > 0$ and $n_i > 0$ for all $i = 1, \ldots, m; j = 1, \ldots, J$. Then $\mathbf{U}$ is positive definite.*

Proof of Theorem 3.2: Recall that $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\mathbf{c}_{i1}\cdots\mathbf{c}_{iJ})$ and $u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1}\mathbf{c}_{it}$, for $s, t = 1, \ldots, J-1$ and $i = 1, \ldots, m$. Denote

$$\tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{c}_{11}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m1}^T \\ \mathbf{c}_{12}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m2}^T \\ \vdots & \ddots & \vdots \\ \mathbf{c}_{1,J-1}^T & & \\ & \ddots & \\ & & \mathbf{c}_{m,J-1}^T \end{pmatrix}_{m(J-1)\times mJ}$$

and $$\tilde{\mathbf{W}} = \begin{pmatrix} n_1 \mathrm{d}iag(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m \mathrm{d}iag(\boldsymbol{\pi}_m)^{-1} \end{pmatrix}_{mJ\times mJ}$$

We claim that $\mathbf{U} = \tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T$. Actually

$$\tilde{\mathbf{C}}\tilde{\mathbf{W}} = \begin{pmatrix} n_1\mathbf{c}_{11}^T \mathrm{d}iag(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m\mathbf{c}_{m1}^T \mathrm{d}iag(\boldsymbol{\pi}_m)^{-1} \\ \vdots & \ddots & \vdots \\ n_1\mathbf{c}_{1,J-1}^T \mathrm{d}iag(\boldsymbol{\pi}_1)^{-1} & & \\ & \ddots & \\ & & n_m\mathbf{c}_{m,J-1}^T \mathrm{d}iag(\boldsymbol{\pi}_m)^{-1} \end{pmatrix}$$

and

$$\begin{aligned} \tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T &= \tilde{\mathbf{C}}\tilde{\mathbf{W}} \begin{pmatrix} \mathbf{c}_{11} & & & \cdots & \mathbf{c}_{1,J-1} & & \\ & \ddots & & \ddots & & \ddots & \\ & & \mathbf{c}_{m1} & \cdots & & & \mathbf{c}_{m,J-1} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U}_{11} & \cdots & \mathbf{U}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1} & \cdots & \mathbf{U}_{J-1,J-1} \end{pmatrix} = \mathbf{U} \end{aligned}$$

Note that $\tilde{\mathbf{W}}$ is diagonal with positive diagonal entries. Thus $\tilde{\mathbf{W}}$ is positive definite. By adjusting the rows, we can verify that $rank(\tilde{\mathbf{C}})$ is the same as $rank(\tilde{\mathbf{C}}')$, where

$$\tilde{\mathbf{C}}' = \begin{pmatrix} \mathbf{c}_{11}^T & & & \\ \vdots & & & \\ \mathbf{c}_{1,J-1}^T & & & \\ & \mathbf{c}_{21}^T & & \\ & \vdots & & \\ & \mathbf{c}_{2,J-1}^T & & \\ & & \ddots & \\ & & & \mathbf{c}_{m1}^T \\ & & & \vdots \\ & & & \mathbf{c}_{m,J-1}^T \end{pmatrix}$$

That is, $\tilde{\mathbf{C}}$ has full row rank and thus $\mathbf{U}$ is positive definite. □

Furthermore, we obtain the determinant of $\mathbf{U}$ as follows:

Theorem 3.3. *Consider the $m(J-1) \times m(J-1)$ matrix $\mathbf{U}$ in Theorem 3.1.*

$$|\mathbf{U}| = \left(\prod_{i=1}^{m} n_i\right)^{J-1} \cdot \prod_{i=1}^{m} \left(\prod_{j=1}^{J} \pi_{ij}\right)^{-1} |\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}|^{-2} \tag{3.8}$$

Theorem 3.3 clearly indicates that $\mathbf{U}$ is singular if $n_i = 0$ for some $i = 1, \ldots, m$. Since $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$ according to Theorem 3.1, $\mathbf{F}$ can still be positive definite even if $\mathbf{U}$ is singular, as long as the rank of $\mathbf{H}$ is equal to the number of its rows.

In general, given an allocation of n experimental units $(n_1, \ldots, n_m)$ with $n_i \geqslant 0$ and $\sum_{i=1}^{m} n_i = n$, we denote $k := \#\{i : n_i > 0\}$ and $\mathbf{U}_{st}^* = \mathrm{d}iag\{n_i u_{st}(\boldsymbol{\pi}_i) : n_i > 0\}$ and

$$\mathbf{U}^* = \begin{bmatrix} \mathbf{U}_{11}^* & \cdots & \mathbf{U}_{1,J-1}^* \\ \vdots & \ddots & \vdots \\ \mathbf{U}_{J-1,1}^* & \cdots & \mathbf{U}_{J-1,J-1}^* \end{bmatrix} \tag{3.9}$$

which is a $k(J-1) \times k(J-1)$ matrix. Let's remove all columns of $\mathbf{H}$ associated with $n_i = 0$ and denote the leftover as $\mathbf{H}^*$, which is a $p \times k(J-1)$ matrix. It can be verified that

Lemma 3.1. $\mathbf{H}\mathbf{U}\mathbf{H}^T = (\mathbf{H}^*)(\mathbf{U}^*)(\mathbf{H}^*)^T$.

Lemma 3.2. $|\mathbf{U}^*| = \left(\prod_{i:n_i>0} n_i\right)^{J-1} \cdot \prod_{i:n_i>0} \left(\prod_{j=1}^{J} \pi_{ij}\right)^{-1} |\mathbf{C}^T \mathbf{D}_i^{-1} \mathbf{L}|^{-2}$.

According to Lemma 3.2, $\mathbf{U}^*$ is non-singular if $\pi_{ij} > 0$ for all i satisfying $n_i > 0$ and all $j = 1, \ldots, J$. Note that $\mathbf{U}^*$ is simply $\mathbf{U}$ if $n_i > 0$ for all $i = 1, \ldots, m$. In order to check when $\mathbf{F}$ is positive definite, we still need to check if $\mathbf{U}^*$ is positive definite. The following result addresses this question as a corollary of Theorem 3.2:

Corollary 3.2. *If $\pi_{ij} > 0$ for all i satisfying $n_i > 0$ and all $j = 1, \ldots, J$, then $\mathbf{U}^*$ defined in Equation 3.9 is positive definite.*

As a direct conclusion of Lemmas 3.1 & 3.2 and Theorem 3.2, we derive a necessary and sufficient condition for $\mathbf{F}$ to be positive definite:

Theorem 3.4. *Suppose $\pi_{ij} > 0$ for all i satisfying $n_i > 0$ and all $j = 1, \ldots, J$. Then the Fisher information matrix $\mathbf{F}$ is positive definite if and only if $rank(\mathbf{H}^*) = p$, that is, $\mathbf{H}^*$ is of full row rank. Furthermore, if $n_i > 0$ for all $i = 1, \ldots, m$, then $\mathbf{F}$ is positive definite if and only if $rank(\mathbf{H}) = p$, that is, $\mathbf{H}$ is of full row rank.*

3.3 Row Rank of H Matrix

According to Theorem 3.4, the positive definiteness of the Fisher information matrix $\mathbf{F}$ depends on the row rank of $\mathbf{H}$ or $\mathbf{H}^*$.

To simplify the notations, we assume $n_i > 0, i = 1, \ldots, m$ throughout this section. In this case, $\mathbf{H} = \mathbf{H}^*$ and $\mathbf{U} = \mathbf{U}^*$. We also assume that

$$m \geqslant p_j, \quad j = 1, \ldots, J-1 \quad \text{and} \quad m \geqslant p_c \text{ if applicable} \tag{3.10}$$

due to the following lemma:

Lemma 3.3. *The rank of $\mathbf{H}$ matrix in Theorem 3.1 equals to the number of its rows p only if $rank(\mathbf{H}_j) = p_j$, $j = 1, \ldots, J-1$ and $rank(\mathbf{H}_c) = p_c$ if applicable.*

Under Assumption Equation 3.10, $rank(\mathbf{H}_j)$, $rank(\mathbf{H}_c)$ and thus $rank(\mathbf{H})$ are all equal to the numbers of rows of the corresponding matrices respectively, given that they are of full rank.

The simplest case is non-proportional odds (NPO) models, the $\mathbf{H}$ matrix in Theorem 3.1 in this case is

$$\mathbf{H}_{NPO} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \end{pmatrix}$$

which satisfies

$$rank(\mathbf{H}_{NPO}) = rank(\mathbf{H}_1) + \cdots + rank(\mathbf{H}_{J-1}) \tag{3.11}$$

In order to keep the Fisher information matrix $\mathbf{F}$ positive definite, or simply $|\mathbf{F}| > 0$, we need $\mathbf{H}_j$ is of full rank p_j for all $j = 1, \ldots, J-1$. Then a necessary condition for $|\mathbf{F}| > 0$ in this case is

$$m \geqslant \max\{p_1, \ldots, p_{J-1}\}$$

The most complicated case is partial proportional odds (PPO) models, the $\mathbf{H}$ matrix in Theorem 3.1 in this case is

$$\mathbf{H}_{PPO} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \\ \mathbf{H}_c & \cdots & \mathbf{H}_c \end{pmatrix}_{p \times m(J-1)}$$

where $\mathbf{H}_j$ is $p_j \times m$, $j = 1, \ldots, J-1$, $\mathbf{H}_c$ is $p_c \times m$, and $p = p_1 + \cdots + p_{J-1} + p_c$ in this case.

Theorem 3.5. *For partial proportional odds (PPO) models,*

$$rank(\mathbf{H}_{PPO}) = rank(\mathbf{H}_1) + \cdots + rank(\mathbf{H}_{J-1}) + r_0 \tag{3.12}$$

where $r_0 = rank(\mathbf{H}_c) - dim\left[\mathcal{M}(\mathbf{H}_c^T) \cap (\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))\right]$, $\mathcal{M}(\mathbf{H}_c^T)$ *is the column space of* $\mathbf{H}_c^T$ *or the row space of* $\mathbf{H}_c$.

Proof of Theorem 3.5: A sequence of linear subspaces are

$$\{0\} \subset \mathcal{M}(\mathbf{H}_c^T) \cap (\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T)) \subset M(\mathbf{H}_c^T)$$

with corresponding dimensions $0 \leqslant r_c - r_0 \leqslant r_c \triangleq rank(\mathbf{H}_c)$. Then there exist $\alpha_1, \cdots, \alpha_{r_c-r_0}, \alpha_{r_c-r_0+1}, \cdots, \alpha_{r_c}$ $\mathbb{R}^m$ s.t. $\{\alpha_1, \cdots, \alpha_{r_c-r_0}\}$ forms a basis of $\mathcal{M}(\mathbf{H}_c^T) \cap (\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))$ and $\{\alpha_1, \cdots, \alpha_{r_c}\}$ forms a basis of $\mathcal{M}(\mathbf{H}_c^T)$, and by simple operations $\mathbf{H}_c$ can be transformed into $\mathbf{H}_c^* = (\alpha_1, \cdots, \alpha_{r_c}, \mathbf{0}, \cdots, \mathbf{0})^T$ and $\mathbf{H}_j$ can be transformed into

$$\mathbf{H}_j^* = (\alpha_1, \cdots, \alpha_{r_c-r_0}, \alpha_{r_c-r_0+1}^{(j)}, \cdots, \alpha_{r_j}^{(j)}, \mathbf{0}, \cdots, \mathbf{0})^T$$

where $r_j = rank(\mathbf{H}_j)$, $j = 1, 2, \cdots, J-1$. Then $rank(\mathbf{H}_{PPO}) = rank(\mathbf{H}^*_{PPO})$ with

$$\mathbf{H}^*_{PPO} = \begin{pmatrix} \mathbf{H}^*_1 & & \\ & \ddots & \\ & & \mathbf{H}^*_{J-1} \\ \mathbf{H}^*_c & \cdots & \mathbf{H}^*_c \end{pmatrix}_{p\times m(J-1)}$$

Since the first $r_c - r_0$ rows of $(\mathbf{H}^*_c, \cdots, \mathbf{H}^*_c)$ can be eliminated by applying row operations of $\mathbf{H}^*_j$ onto it separately, then $rank(\mathbf{H}^*_{PPO}) = rank(\mathbf{H}^{**}_{PPO})$ where

$$\mathbf{H}^{**}_{PPO} = \begin{pmatrix} \mathbf{H}^*_1 & & \\ & \ddots & \\ & & \mathbf{H}^*_{J-1} \\ \mathbf{H}^{**}_c & \cdots & \mathbf{H}^{**}_c \end{pmatrix}_{p\times m(J-1)}$$

and $\mathbf{H}^{**}_c = (\mathbf{0}, \cdots, \mathbf{0}, \alpha_{r_c-r_0+1}, \cdots, \alpha_{r_c}, \mathbf{0}, \cdots, \mathbf{0})^T$. Therefore, $rank(\mathbf{H}_{PPO}) = rank(\mathbf{H}^{**}_{PPO}) \leqslant r_1 + \cdots + r_{J-1} + r_0$.

We claim that the nonzero rows of $\mathbf{H}^{**}_{PPO}$ are linearly independent which will lead to the final conclusion. Actually, let's denote those nonzero rows of $\mathbf{H}^{**}_{PPO}$ as $\Lambda^{(j)}_i, i = 1, 2, \cdots, r_j, j = 1, 2, \cdots, J-1$ and $\Lambda_{r_c-r_0+1}, \cdots, \Lambda_{r_c}$, where $\Lambda^{(j)}_i$ is the ith row of $(\mathbf{0}, \cdots, \mathbf{0}, \mathbf{H}^*_j, \mathbf{0}, \cdots, \mathbf{0})$, and Λ_i

is the ith row of $(\mathbf{H}_c^{**}, \cdots, \mathbf{H}_c^{**})$. Suppose there exist $a_i^{(j)} \in \mathbb{R}, i = 1, 2, \cdots, r_j, j = 1, 2, \cdots, J-1$ and $a_i \in \mathbb{R}, i = r_c - r_0 + 1, \cdots, r_c$ s.t.

$$\mathbf{0} = \sum_{j=1}^{J-1} \sum_{i=1}^{r_j} a_i^{(j)} \Lambda_i^{(j)} + \sum_{i=r_c-r_0+1}^{r_c} a_i \Lambda_i$$

then for $j = 1, \ldots, J-1$,

$$\mathbf{0} = \sum_{i=1}^{r_c-r_0} a_i^{(j)} \alpha_i + \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \alpha_i^{(j)} + \sum_{i=r_c-r_0+1}^{r_c} a_i \alpha_i$$

which implies for $j = 1, \ldots, J-1$,

$$\sum_{i=r_c-r_0+1}^{r_c} a_i \alpha_i = -\sum_{i=1}^{r_c-r_0} a_i^{(j)} \alpha_i - \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \alpha_i^{(j)} \in \mathcal{M}(\mathbf{H}_c^T) \cap \mathcal{M}(\mathbf{H}_j^T)$$

Thus, $\sum_{i=r_c-r_0+1}^{r_c} a_i \alpha_i \in \mathcal{M}(\mathbf{H}_c^T) \bigcap \left(\cap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T) \right)$. Then we must have $\sum_{i=r_c-r_0+1}^{r_c} a_i \alpha_i = \mathbf{0}$ since $\{\alpha_{r_c-r_0+1}, \ldots, \alpha_{r_c}\}$ and $\{\alpha_1, \ldots, \alpha_{r_c-r_0}\}$ are linearly independent. Therefore, $a_i = 0$ for $i = r_c - r_0 + 1, \ldots, r_c$ and thus

$$\mathbf{0} = \sum_{i=1}^{r_c-r_0} a_i^{(j)} \alpha_i + \sum_{i=r_c-r_0+1}^{r_j} a_i^{(j)} \alpha_i^{(j)}$$

It implies $a_i^{(j)} = 0$, $i = 1, \ldots, r_c - r_0, r_c - r_0 + 1, \ldots, r_j$ since $\{\alpha_1, \ldots, \alpha_{r_c-r_0}, \alpha_{r_c-r_0+1}^{(j)}, \ldots, \alpha_{r_j}^{(j)}\}$ are linear independent. $\square$

In order to apply Theorem 3.5, we need an efficient way to calculate $dim\left[\mathcal{M}(\mathbf{H}_c^T) \cap (\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T))\right]$, or in general, to calculate $dim(\bigcap_{i=1}^{n}\mathcal{M}(H_i^T))$ for general matrices. The following theorem answers to this need:

Theorem 3.6. *Suppose* $\mathbf{H}_i$ *is of* $p_i \times m$ *with rank* r_i, $i = 1, \ldots, n$. *Denote rank* $\left((\mathbf{H}_{i_1}^T, \ldots, \mathbf{H}_{i_k}^T)\right) = r_{i_1,\ldots,i_k}$ *for any* $2 \leqslant k \leqslant n$ *and* $1 \leqslant i_1 < \cdots < i_k \leqslant n$. *Then*

$$dim\left(\bigcap_{i=1}^{n}\mathcal{M}(\mathbf{H}_i^T)\right) = \sum_{i=1}^{n} r_i - \sum_{i_1<i_2} r_{i_1,i_2} + \sum_{i_1<i_2<i_3} r_{i_1,i_2,i_3} - \cdots + (-1)^{n-1} r_{1,2,\ldots,n} \tag{3.13}$$

Proof of Theorem 3.6: Recall that $dim(\mathcal{M}(\mathbf{H}_i^T)) = rank(\mathbf{H}_i^T) = r_i$ and $dim(\mathcal{M}(\mathbf{H}_{i_1}^T) + \cdots + \mathcal{M}(\mathbf{H}_{i_k}^T)) = dim(\mathcal{M}((\mathbf{H}_{i_1}^T, \cdots, \mathbf{H}_{i_k}^T))) = rank((\mathbf{H}_{i_1}^T, \cdots, \mathbf{H}_{i_k}^T)) = r_{i_1,\ldots,i_k}$, for $i_1 < \cdots < i_k$ and $k = 2, \ldots, n$, where "+" stands for the sum of two linear subspaces.

First of all, $dim(\mathcal{M}(\mathbf{H}_1^T) \cap \mathcal{M}(\mathbf{H}_2^T)) = dim(\mathcal{M}(\mathbf{H}_1^T)) + dim(\mathcal{M}(\mathbf{H}_2^T)) - dim(\mathcal{M}(\mathbf{H}_1^T) + \mathcal{M}(\mathbf{H}_2^T)) = r_1 + r_2 - r_{12}$. That is, Equation 3.13 is true for $n = 2$.

Suppose Equation 3.13 is true for $n = k$. Then for $n = k + 1$,

$$\begin{aligned}
& dim(\cap_{i=1}^{k+1}\mathcal{M}(\mathbf{H}_i^T)) = dim(\cap_{i=1}^{k}\mathcal{M}(\mathbf{H}_i^T) \cap \mathcal{M}(\mathbf{H}_{k+1}^T)) \\
= \; & dim(\cap_{i=1}^{k}\mathcal{M}(\mathbf{H}_i^T)) + dim(\mathcal{M}(\mathbf{H}_{k+1}^T)) - dim(\cap_{i=1}^{k}\mathcal{M}(\mathbf{H}_i^T) + \mathcal{M}(\mathbf{H}_{k+1}^T)) \\
= \; & \sum_{i=1}^{k} r_i - \sum_{1\leqslant i_1<i_2\leqslant k} r_{i_1 i_2} + \cdots + (-1)^{k-1} r_{12\cdots k} + r_{k+1} - \triangle
\end{aligned}$$

where

$$\begin{aligned} & \triangle = dim(\cap_{i=1}^{k}\mathcal{M}(\mathbf{H}_i^T) + \mathcal{M}(\mathbf{H}_{k+1}^T)) = dim(\cap_{i=1}^{k}\mathcal{M}((\mathbf{H}_i^T, \mathbf{H}_{k+1}^T))) \\ = & \sum_{i=1}^{k} rank((\mathbf{H}_i^T, \mathbf{H}_{k+1}^T)) - \sum_{1 \leqslant i_1 < i_2 \leqslant k} rank((\mathbf{H}_{i_1}^T, \mathbf{H}_{k+1}^T, \mathbf{H}_{i_2}^T, \mathbf{H}_{k+1}^T)) \\ & + \cdots + (-1)^{k-1} rank((\mathbf{H}_1^T, \mathbf{H}_{k+1}^T, \cdots, \mathbf{H}_k^T, \mathbf{H}_{k+1}^T)) \\ = & \sum_{i=1}^{k} r_{i,k+1} - \sum_{1 \leqslant i_1 < i_2 \leqslant k} r_{i_1,i_2,k+1} + \cdots + (-1)^{k-1} r_{1,2,\ldots,k+1} \end{aligned}$$

Therefore,

$$\begin{aligned} & dim(\cap_{i=1}^{k+1}\mathcal{M}(\mathbf{H}_i^T)) \\ = & \sum_{i=1}^{k} r_i - \sum_{1 \leqslant i_1 < i_2 \leqslant k} r_{i_1 i_2} + \cdots + (-1)^{k-1} r_{12\cdots k} + r_{k+1} \\ & - \sum_{i=1}^{k} r_{i,k+1} + \sum_{1 \leqslant i_1 < i_2 \leqslant k} r_{i_1,i_2,k+1} + \cdots + (-1)^{k} r_{1,2,\ldots,k+1} \\ = & \sum_{i=1}^{k+1} r_i - \sum_{1 \leqslant i_1 < i_2 \leqslant k+1} r_{i_1 i_2} + \cdots + (-1)^{(k+1)-1} r_{1,2,\ldots,k+1} \end{aligned}$$

That is, Equation 3.13 is true for $n = k+1$. By mathematical induction, Equation 3.13 is true for general n. □

The proportional odds (PO) models can be regarded as special cases of partial proportional odds (PPO) models. As direct conclusions of Theorem 3.5, we have

Theorem 3.7. *For proportional odds models,* $rank(\mathbf{H}_{PO}) = rank(\big(\mathbf{1}, \mathbf{H}_c^T\big)) + J - 2.$

Combining Theorem 3.4, equation Equation 3.11, Theorem 3.5 and Theorem 3.7, we obtain the theorem as follows:

Theorem 3.8. *Consider the multinomial logistic model Equation 2.1 with m distinct experimental settings $\mathbf{x}_i$ and corresponding number of observations n_i, $i = 1, \ldots, m$. Suppose $\pi_{ij} > 0$, $n_i > 0$, $i = 1, \ldots, m; j = 1, \ldots, J$. Then the Fisher information matrix $\mathbf{F}$ is positive definite if and only if*

(1) For non-proportional odds (NPO) models, $m \geqslant \max\{p_1, \ldots, p_{J-1}\}$ and $\mathbf{x}_i$'s keep $\mathbf{H}_j$ of full row rank p_j, $j = 1, \ldots, J-1$.

(2) For proportional odds (PO) models, $m \geqslant p_c + 1$ and the extended matrix $(\mathbf{1}, \mathbf{H}_c^T)$ is of full rank $p_c + 1$.

(3) For partial proportional odds (PPO) models, $m \geqslant \max\{p_1, \ldots, p_{J-1}, p_c\}$ and $\mathbf{x}_i$'s keep $\mathbf{H}_j$ of full row rank p_j, $j = 1, \ldots, J-1$; $\mathbf{H}_c$ of full row rank p_c; as well as $\mathcal{M}(\mathbf{H}_c^T) \cap \left(\cap_{j=1}^{J-1} \mathcal{M}(\mathbf{H}_j^T)\right) = \{0\}$.

(4) For a special case of PPO models with $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$, $m \geqslant p_1 + p_c$ and the extended matrix $(\mathbf{H}_1^T, \mathbf{H}_c^T)$ is of full rank $p_1 + p_c$.

Based on Theorem 3.8(3), we have the following lemma:

Lemma 3.4. *Given* (1)*rank*$(H_j) = p_j, j = 1, \cdots, J-1$, (2)*rank*$(H_c) = p_c$, $(3) M(H_C^T) \cap M_H = 0$, *where $M_H = \cap_{j=1}^{J-1} M(H_j^T)$, write $p_H = dim(M_H)$, then $m \geqslant p_H + p_c$.*

Proof of Lemma 3.4: Suppose $p_H > 0$, there exist $\alpha_1, \cdots, \alpha_{p_H} \in R^m$, which form a basis of M_H. After row transformations, H_j can be rewritten as $[\alpha_1, \cdots, \alpha_{p_H}, \alpha_{p_H+1}^{(j)}, \cdots, \alpha_{p_j}^{(j)}]^T$, write

$H_c = [\gamma_1, \cdots, \gamma_{p_c}]^T$, then (3) implies $[\alpha_1, \cdots, \alpha_{p_H}, \gamma_1, \cdots, \gamma_{p_c}] \in R^m$ are linearly independent. Thus $m \geqslant p_H + p_c$.

Then we can rewrite Theorem 3.8(3) as Theorem 3.8(3'):

(3') For partial proportional odds (PPO) models, $m \geq \max\{p_1, \ldots, p_{J-1}, p_H + p_c\}$ and $\mathbf{x}_i$'s keep $\mathbf{H}_j$ of full row rank p_j, $j = 1, \ldots, J-1$; $\mathbf{H}_c$ of full row rank p_c; and $dim\left(\cap_{j=1}^{J-1}\mathcal{M}(\mathbf{H}_j^T)\right) = p_H$.

Theorem 3.8(3') actually covers all of model situations. If it is NPO model, $p_c = 0$ and $p_H \leqslant \max\{p_1, \ldots, p_{J-1}\}$, (3') is degenerated as $m \geq \max\{p_1, \ldots, p_{J-1}\}$. If it is PO model, $p_H = p_1 = \ldots = p_{J-1} = 1$, (3') is degenerated as $m \geqslant p_c + 1$. In case of special PPO, $p_H = p_1 = \ldots = p_{J-1}$, (3') becomes $m \geq p_1 + p_c$.

Theorem 3.8 says that the number m of distinct experimental settings for a partial proportional odds model could be as low as $\max\{p_1, \ldots, p_{J-1}, p_c + p_H\}$, which is strictly less than p. The following artificial example provides such a case.

Example 3.1. *Consider an experiment with three factors* ($d = 3$)*, three response categories* ($J = 3$)*, and three distinct experimental settings* ($m = 3$)*. Then the experimental settings are* $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3})^T$, $i = 1, 2, 3$*. Consider a multinomial logistic model with partial proportional odds such that*

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \end{pmatrix}, \mathbf{H}_2^T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \mathbf{H}_c^T = \begin{pmatrix} x_{12} & x_{13} \\ x_{22} & x_{23} \\ x_{32} & x_{33} \end{pmatrix}$$

That is, $p_1 = 2, p_2 = 1, p_c = 2, p_H = 1$, $\max\{p_1, p_2, p_c + p_H\} = 3$, *and* $p = p_1 + p_2 + p_c = 5$. *In this case,*

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ x_{11} & x_{21} & x_{31} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ x_{12} & x_{22} & x_{32} & x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} & x_{13} & x_{23} & x_{33} \end{pmatrix}$$

is 5×6. *It can be verified that rank*$(\mathbf{H}) = 5$ *for general* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ *according to Theorem 3.5. That is, the minimal number of experimental settings in this case is* $m = \max\{p_1, \ldots, p_{J-1}, p_c+p_H\} = 3$.

Example 3.2. *Consider another experiment with four factors* $(d = 4)$, *three response categories* $(J = 3)$, *and four distinct experimental settings* $(m = 4)$. *Then the experimental settings are* $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})^T$, $i = 1, 2, 3, 4$. *Consider a multinomial logistic model with partial proportional odds such that*

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{pmatrix}, \ \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ 1 & x_{31} \\ 1 & x_{41} \end{pmatrix}, \ \mathbf{H}_c^T = \begin{pmatrix} x_{14} \\ x_{24} \\ x_{34} \\ x_{44} \end{pmatrix}$$

That is, $p_1 = 4, p_2 = 2, p_c = 1, p_H = 2$, $\max\{p_1, p_2, p_c + p_H\} = 4$, *and* $p = p_1 + p_2 + p_c = 7$. *In this case,*

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ x_{11} & x_{21} & x_{31} & x_{41} & 0 & 0 & 0 & 0 \\ x_{12} & x_{22} & x_{32} & x_{42} & 0 & 0 & 0 & 0 \\ x_{13} & x_{23} & x_{33} & x_{43} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & x_{11} & x_{21} & x_{31} & x_{41} \\ x_{14} & x_{24} & x_{34} & x_{44} & x_{14} & x_{24} & x_{34} & x_{44} \end{pmatrix}$$

is 7×8. *It can be verified that rank*$(\mathbf{H}) = 7$ *for general* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ *according to Theorem 3.5. That is, the minimal number of experimental settings in this case is* $m = \max\{p_1, \ldots, p_{J-1}, p_c + p_H\} = 4$.

CHAPTER 4

DETERMINANT OF FISHER INFORMATION MATRIX

4.1 Further Reformulation of Fisher Information Matrix for Multinomial Logistic Models

In order to calculate the determinant of the Fisher information matrix $\mathbf{F}$, we reformulate it into the format of $\mathbf{G}^T\mathbf{W}\mathbf{G}$ with a diagonal matrix $\mathbf{W}$ in this section.

Actually, according to Theorem 3.1, $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$. From the proof of Theorem 3.2, $\mathbf{U} = \tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T$, where $\tilde{\mathbf{W}}$ is a diagonal matrix. Therefore,

$$\mathbf{F} = \mathbf{H}\tilde{\mathbf{C}}\tilde{\mathbf{W}}\tilde{\mathbf{C}}^T\mathbf{H}^T$$

which leads to the theorem as follows:

Theorem 4.1. *Consider the multinomial logistic model Equation 2.1 with independent observations. The Fisher information matrix*

$$\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G} \tag{4.1}$$

where n is the total number of observations with n_i of them assigned to the ith experimental setting $\mathbf{x}_i$, $\mathbf{W} = diag\{w_1 diag(\boldsymbol{\pi}_1)^{-1}, \ldots, w_m diag(\boldsymbol{\pi}_m)^{-1}\}$ is an $mJ \times mJ$ matrix with $w_i = n_i/n$, $\mathbf{G}$ is an $mJ \times p$ matrix which takes the forms of

$$
\begin{aligned}
\mathbf{G}_{PPO} &= \begin{pmatrix}
\mathbf{c}_{11}\mathbf{h}_1^T(\mathbf{x}_1) & \cdots & \mathbf{c}_{1,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_1) & \sum_{j=1}^{J-1}\mathbf{c}_{1j}\cdot\mathbf{h}_c^T(\mathbf{x}_1) \\
\mathbf{c}_{21}\mathbf{h}_1^T(\mathbf{x}_2) & \cdots & \mathbf{c}_{2,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_2) & \sum_{j=1}^{J-1}\mathbf{c}_{2j}\cdot\mathbf{h}_c^T(\mathbf{x}_2) \\
\cdots & \cdots & \cdots & \cdots \\
\mathbf{c}_{m1}\mathbf{h}_1^T(\mathbf{x}_m) & \cdots & \mathbf{c}_{m,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_m) & \sum_{j=1}^{J-1}\mathbf{c}_{mj}\cdot\mathbf{h}_c^T(\mathbf{x}_m)
\end{pmatrix} \\
\mathbf{G}_{NPO} &= \begin{pmatrix}
\mathbf{c}_{11}\mathbf{h}_1^T(\mathbf{x}_1) & \cdots & \mathbf{c}_{1,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_1) \\
\mathbf{c}_{21}\mathbf{h}_1^T(\mathbf{x}_2) & \cdots & \mathbf{c}_{2,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_2) \\
\cdots & \cdots & \cdots \\
\mathbf{c}_{m1}\mathbf{h}_1^T(\mathbf{x}_m) & \cdots & \mathbf{c}_{m,J-1}\mathbf{h}_{J-1}^T(\mathbf{x}_m)
\end{pmatrix} \\
\mathbf{G}_{PO} &= \begin{pmatrix}
\mathbf{c}_{11} & \cdots & \mathbf{c}_{1,J-1} & \sum_{j=1}^{J-1}\mathbf{c}_{1j}\cdot\mathbf{h}_c^T(\mathbf{x}_1) \\
\mathbf{c}_{21} & \cdots & \mathbf{c}_{2,J-1} & \sum_{j=1}^{J-1}\mathbf{c}_{2j}\cdot\mathbf{h}_c^T(\mathbf{x}_2) \\
\cdots & \cdots & \cdots & \cdots \\
\mathbf{c}_{m1} & \cdots & \mathbf{c}_{m,J-1} & \sum_{j=1}^{J-1}\mathbf{c}_{mj}\cdot\mathbf{h}_c^T(\mathbf{x}_m)
\end{pmatrix}
\end{aligned}
$$

for PPO, NPO, PO models, respectively.

4.2 General Formula for Determinant of Fisher Information Matrix

The general determinant of Fisher Information Matrix can be obtained from $|\mathbf{G}^T\mathbf{W}\mathbf{G}|$ in Equation 4.1. Since $\mathbf{W}$ is diagonal, we obtain the following theorem as a direct conclusion of Theorem 1.1.2 of (20) or Lemma 3.1 of (21):

Theorem 4.2. *Up to the constant n^p, the determinant of Fisher information matrix is*

$$|\mathbf{G}^T\mathbf{W}\mathbf{G}| = \sum_{\alpha_1 \geq 0, \ldots, \alpha_m \geq 0 \,:\, \sum_{i=1}^m \alpha_i = p} c_{\alpha_1,\ldots,\alpha_m} \cdot w_1^{\alpha_1} \cdots w_m^{\alpha_m} \tag{4.2}$$

with

$$c_{\alpha_1,\ldots,\alpha_m} = \sum_{(i_1,\ldots,i_p)\in\Lambda(\alpha_1,\ldots,\alpha_m)} |\mathbf{G}[i_1,\ldots,i_p]|^2 \prod_{k:\alpha_k>0} \prod_{l:(k-1)J<i_l\leqslant kJ} \pi^{-1}_{k,i_l-(k-1)J} \geqslant 0 \tag{4.3}$$

where $\alpha_1, \ldots, \alpha_m$ are nonnegative integers, and

$$\Lambda(\alpha_1,\ldots,\alpha_m) = \{(i_1,\ldots,i_p)|1 \leq i_1 < \cdots < i_p \leq mJ;$$
$$\#\{i_l : (k-1)J < i_l \leqslant kJ\} = \alpha_k; k = 1,\ldots,m\}$$

and $\mathbf{G}$ stands for the submatrix consisting of the $i_1, \ldots, i_p$th rows of $\mathbf{G}$.

According to Theorem 4.2, the determinant of Fisher information matrix is an order-p homogeneous polynomial function of $w_1, \ldots, w_m$. Another important conclusion is that the coefficient $c_{\alpha_1,\ldots,\alpha_m}$ must be nonnegative according to Equation 4.3. Actually, the structure of the determinant can be significantly simplified due to the following results on the coefficient $c_{\alpha_1,\ldots,\alpha_m}$ of $w_1^{\alpha_1}\cdots w_m^{\alpha_m}$:

Lemma 4.1. *If $\max_{1\leq i\leq m} \alpha_i \geqslant J$, then $|\mathbf{G}[i_1,\ldots,i_p]| = 0$ for any $(i_1, \ldots, i_p) \in (\alpha_1, \ldots, \alpha_m)$. Therefore, $c_{\alpha_1,\ldots,\alpha_m} = 0$ in this case.*

Proof of Lemma 4.1: Actually, $\max_{1\le i\le m}\alpha_i \leqslant J$. Suppose $\max_{1\le i\le m}\alpha_i \geqslant J$, which means $\max_{1\le i\le m}\alpha_i = J$. Without any loss of generality, we assume $\alpha_1 = J$. Then $i_j = j$ for $j = 1,\ldots,J$.

According to subsection 4.3.2, we have $\mathbf{1}^T\mathbf{c}_{ij} = 0$ for $i = 1,\ldots,m$ and $j = 1,\ldots,J-1$. Then $\mathbf{1}^T(\mathbf{c}_{11} + \cdots + \mathbf{c}_{1,J-1}) = 0$ and thus $\mathbf{1}^T\mathbf{G}[i_1,\ldots,i_J] = 0$. That is, $\text{rank}(\mathbf{G}[i_1,\ldots,i_J]) \leqslant J-1$. Therefore, $\text{rank}(\mathbf{G}[i_1,\ldots,i_p]) \leqslant p-1$ and $|\mathbf{G}[i_1,\ldots,i_p]| = 0$. $\square$

Theorem 4.3. *The coefficient $c_{\alpha_1,\ldots,\alpha_m}$ as defined in Equation 4.3 is nonzero only if the restricted Fisher information matrix $\mathbf{F}_{res} = \sum_{i:\alpha_i>0}\mathbf{F}_i$ is positive definite, where $\mathbf{F}_i$ is defined in Theorem (2.1).*

Proof of Theorem 4.3: Suppose $c_{\alpha_1,\ldots,\alpha_m} \neq 0$ for some $(\alpha_1,\ldots,\alpha_m)$. Therefore, there exist $(i_1,\ldots,i_p) \in (\alpha_1,\ldots,\alpha_m)$ such that $\mathbf{G}[i_1,\ldots,i_p]$ is of full rank p. Without any loss of generality, we assume $\alpha_1 \geqslant \cdots \geqslant \alpha_k > 0 = \alpha_{k+1} = \cdots = \alpha_m$, that is, $\{i \mid \alpha_i > 0\} = \{1,\ldots k\}$. Consider the submatrix $\tilde{\mathbf{G}} := \mathbf{G}[1,\ldots,kJ]$ which is $kJ\times p$ and contains $\mathbf{G}[i_1,\ldots,i_p]$ as a submatrix. Then $\tilde{\mathbf{G}}$ is of rank p or $\tilde{\mathbf{G}}^T$ is of full row rank p. Write $\tilde{\mathbf{W}} = k^{-1}\mathrm{d}iag\{\mathrm{d}iag(\boldsymbol{\pi}_1)^{-1},\ldots,\mathrm{d}iag(\boldsymbol{\pi}_k)^{-1}\}$. Then the restricted matrix $\mathbf{F} := n\tilde{\mathbf{G}}^T\tilde{\mathbf{W}}\tilde{\mathbf{G}}$ is positive definite. On the other hand, $\mathbf{F}$ is the Fisher information matrix $n\mathbf{G}^T\mathbf{W}\mathbf{G}$ as defined in Equation 4.1 with $w_1 = \cdots = w_k = 1/k$ and $w_{k+1} = \cdots = w_m = 0$. According to Theorem 4.1 and Theorem 2.1, $\mathbf{F} = nk^{-1}\sum_{i=1}^k\mathbf{F}_i$. Therefore, $\mathbf{F}_{res} := \sum_{i=1}^k\mathbf{F}_i$ is positive definite. $\square$

Combining Theorem 3.4 and Theorem 4.3, Theorem 3.8 and Theorem 4.3, respectively, we obtain the following corollaries:

Corollary 4.1. *The coefficient $c_{\alpha_1,\ldots,\alpha_m}$ is nonzero only if $\mathbf{H}_{\alpha_1,\ldots,\alpha_m}$ is of full row rank p, where $\mathbf{H}_{\alpha_1,\ldots,\alpha_m}$ is the submatrix of $\mathbf{H}$ after removing all columns associated with $\mathbf{x}_i$ such that $\alpha_i = 0$.*

Corollary 4.2. *The coefficient $c_{\alpha_1,\ldots,\alpha_m} = 0$ if $\#\{i \mid \alpha_i > 0\} \leqslant k_{\min} - 1$, where*

(1) $k_{\min} = \max\{p_1, \ldots, p_{J-1}\}$ *for NPO models;*

(2) $k_{\min} = p_c + 1$ *for PO models;*

(3) $k_{\min} = \max\{p_1, \ldots, p_{J-1}, p_H + p_c\}$ *for PPO models;*

(4) $k_{\min} = p_1 + p_c$ *for PPO models with* $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$.

For typical applications $k_{\min} \geqslant 2$. The determinant of a PPO model could in general be more complicated than a PO model's. Here's one example which shows that $c_{\alpha_1,\ldots,\alpha_m}$ could be zero for PPO models given that $\#\{i \mid \alpha_i > 0\} = p_H + p_c - 1$.

Example 4.1. *Consider an artificial example with responses in $J = 4$ categories, $d = 5$ factors, and $m = 5$ distinct experimental settings $\mathbf{x}_i = (x_{i,1}, \ldots, x_{i,5})^T$, $i = 1, \ldots, 5$. Suppose a multinomial logistic model with*

$$\mathbf{H}_1^T = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{51} & x_{52} \end{pmatrix}, \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{11} \\ 1 & x_{21} \\ \vdots & \vdots \\ 1 & x_{51} \end{pmatrix}, \mathbf{H}_3^T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \mathbf{H}_c^T = \begin{pmatrix} x_{13} & x_{14} & x_{15} \\ x_{23} & x_{24} & x_{25} \\ \vdots & \vdots & \vdots \\ x_{53} & x_{54} & x_{55} \end{pmatrix}$$

is used. That is, $p_1 = 3$, $p_2 = 2$, $p_3 = 1$, $p_H = 1$, $p_c = 3$ and $p = 9$. The $\mathbf{G}$ matrix in this case is 20×9. $p_H + p_c = 4$ should be the minimum number of non-negative of $\alpha_1, \ldots, \alpha_m$ to make $|G[i_1, \ldots, i_p]| \neq 0$. It can be verified that $(i_1, \ldots, i_9) = (1, 2, 3, 6, 7, 8, 10, 11, 12) \in$

$\Lambda(3,3,3,0,0) = \Lambda(\alpha_1,\ldots,\alpha_5)$ *leads to Rank* $G([\, i_1,\ldots,i_9]) = 8$*, while* $(1,2,5,6,9,10,13,14,15) \in \Lambda(2,2,2,3,0)$ *leads to Rank*$(G[i_1,\ldots,i_p]) = 9$. *Therefore,* $|G[i_1,\ldots,i_p]| \neq 0$ *in general if* $(i_1,\ldots,i_p) \in \Lambda(2,2,2,3,0)$ *for such a PPO model.*

The following example shows that Lemma 4.1 and Corollary 4.2 may simplify the structure of $|\mathbf{F}|$ significantly.

Example 2.3 *(Continued)* **:** In this example, the number of factors is $d = 4$, and the experimental settings are $\mathbf{x}_i = (x_{i2}, x_{i3}, x_{i4}, x_{i5})^T$, $i = 1,\ldots,m$. Since $p_1 = p_2 = 3$, $p_c = 2$, and the number of parameters $p = p_1 + p_2 + p_c = 8$ in this case, the minimal number of experimental settings is $m = p_1 + p_c = 5$ according to Theorem 3.8. We consider the simplest case $m = 5$. That is,

$$\mathbf{H}_1^T = \mathbf{H}_2^T = \begin{pmatrix} 1 & x_{13} & x_{15} \\ 1 & x_{23} & x_{25} \\ \vdots & \vdots & \vdots \\ 1 & x_{53} & x_{55} \end{pmatrix}, \quad \mathbf{H}_c^T = \begin{pmatrix} x_{12} & x_{14} \\ x_{22} & x_{24} \\ \vdots & \vdots \\ x_{52} & x_{54} \end{pmatrix}$$

According to Theorem 4.2, the determinant of the Fisher information matrix $|\mathbf{F}|$ is an order-8 homogeneous polynomial allocating to 5 experimental settings which may contain up to $(8+5-1)!/(8!(5-1)!) = 465$ items. However, Lemma 4.1 implies $c_{\alpha_1,\ldots,\alpha_5} \neq 0$ only if $\alpha_i \in \{0,1,2\}, i = 1,\ldots,5$. On the other hand, Corollary 4.2 says $c_{\alpha_1,\ldots,\alpha_5} \neq 0$ only if $\#\{i \mid \alpha_i > 0\} \geqslant$

$p_1 + p_c = 5$, that is, $\alpha_i > 0$ for all $i = 1, \ldots, 5$. Then only items with $\alpha_i \in \{1, 2\}, i = 1, \ldots, 5$ left and the polynomial $|\mathbf{F}|$ contains only $5!/(3!2!) = 10$ items. That is,

$$|\mathbf{F}| = \prod_{i=1}^{5} w_i \cdot \sum_{1 \le i_1 < i_2 < i_3 \le 5} e_{i_1,i_2,i_3} w_{i_1} w_{i_2} w_{i_3}$$

for some coefficients e_{i_1,i_2,i_3}, where $w_i = n_i/n, i = 1, \ldots, 5$ and $w_{i_1} = n_{i_1}/n$. Actually, e_{i_1,i_2,i_3} is related to $c_{\alpha_1,\ldots,\alpha_m}$ in Equation 4.3 of Theorem 4.2. For example, $e_{1,2,3} = c_{2,2,2,1,1}$. □

4.3 Determinant of Fisher Information Matrix in Some Special Cases

The above section gives the general formula for determinant of Fisher information matrix, which could be applied to any multinomial logistic models, i.e., four types of logit models and three types of odds assumptions. In some special cases, we can get simpler formulas even analytic forms for determinant of Fisher information matrix.

4.3.1 Determinant of U matrix

Lemma 4.2. *If we just look at the determinant of* $\mathbf{U}$ *in Theorem 3.1, then*

$$|\mathbf{U}| = (\prod_{i=1}^{m} n_i)^{J-1} |\mathbf{V}|$$

where

$$
\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \cdots & \mathbf{V}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{J-1,1} & \cdots & \mathbf{V}_{J-1,J-1} \end{pmatrix}
$$

$$
= \begin{pmatrix}
u_{111} & & & \cdots & u_{1,J-1,1} & & \\
& \ddots & & & & \ddots & \\
& & u_{11m} & & & \cdots & u_{1,J-1,m} \\
\vdots & & & \ddots & & & \vdots \\
u_{J-1,1,1} & & & \cdots & u_{J-1,J-1,1} & & \\
& \ddots & & & & \ddots & \\
& & u_{J-1,1,m} & & & \cdots & u_{J-1,J-1,m}
\end{pmatrix}
$$

Note: Lemma 4.2 implies if $\mathbf{H}$ is a square matrix, then $n_1 = n_2 = \cdots = n_m = n/m$ would be D-optimal.

Kovacs et al. (28) generalized Schur's Formula (29), and their theorem is as follows:

Lemma 4.3. *Theorem 1 of (28)*

Assume that $\mathbf{M}$ *is a* $k \times k$ *block matrix, each block element* $\mathbf{A}_{ij}$ *is an* $n \times n$ *matrix.*

$$
\mathbf{M} = \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1k} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{k1} & \cdots & \mathbf{A}_{kk} \end{pmatrix}
$$

If all of $\mathbf{A}'_{ij}s$ commute pairwise, that is, $\mathbf{A}_{ij}\mathbf{A}_{lm} = \mathbf{A}_{lm}\mathbf{A}_{ij}$ for all possible pairs of indices i, j and l, m. Then

$$|\mathbf{M}| = \left| \sum_{\pi \in S_k} (sgn\pi)\mathbf{A}_{1\pi(1)}\mathbf{A}_{2\pi(2)} \cdots \mathbf{A}_{k\pi(k)} \right| \tag{4.4}$$

Here the sum is computed over all permutations π of the set $1, 2, ..., k$. A permutation is a function that reorders this set of integers. The value in the ith position after the reordering π is denoted $\pi(i)$. For example, for n = 3, the original sequence 1, 2, 3 might be reordered to π = [2, 3, 1], with $\pi(1)$ = 2, $\pi(2)$ = 3, and $\pi(3)$ = 1. The set of all such permutations (also known as the symmetric group on k elements) is denoted S_k. For each permutation π, $sgn(\pi)$ denotes the signature of π, a value that is +1 whenever the reordering given by π can be achieved by successively interchanging two entries an even number of times, and 1 whenever it can be achieved by an odd number of such interchanges.

In our case, all of $\mathbf{V}'_{ij}s$ are diagonal matrices, so they commute pairwise. Moreover, the sum of product matrices in Equation 4.4 is a diagonal matrix, in which each element is the sum of products of the corresponding elements in those matrices. If we apply the above lemma, we get

$$\begin{aligned} |\mathbf{V}| &= \left| \sum_{\pi \in S_{J-1}} (sgn\pi)\mathbf{V}_{1\pi(1)}\mathbf{V}_{2\pi(2)} \cdots \mathbf{V}_{J-1,\pi(J-1)} \right| \\ &= \prod_{i=1}^{m} \left| \sum_{\pi \in S_{J-1}} (sgn\pi)u_{1\pi(1)i}u_{2\pi(2)i} \cdots u_{J-1,\pi(J-1),i} \right| \end{aligned}$$

So the following theorem is naturally approved

Theorem 4.4.

$$|\mathbf{V}| = \prod_{i=1}^{m} |\mathbf{V}_i| \tag{4.5}$$

where

$$\mathbf{V}_i = \begin{pmatrix} u_{11}(\boldsymbol{\pi}_i) & \cdots & u_{1,J-1}(\boldsymbol{\pi}_i) \\ \vdots & \ddots & \vdots \\ u_{J-1,1}(\boldsymbol{\pi}_i) & \cdots & u_{J-1,J-1}(\boldsymbol{\pi}_i) \end{pmatrix}$$

Note that the $\mathbf{V}_i$ defined above is very similar to $\mathbf{U}_i$ we defined before, see Equation 3.2.

4.3.2 Key intermediate results for four types of logit models

In order to get $|\mathbf{V}_i|$, we need to examine its element $u_{st}(\boldsymbol{\pi}_i)$, which is related to $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}$ according to its definition Equation 3.3.

Now look at the four different kinds of multinomial logistic models. Note that $\pi_{i1}+\cdots+\pi_{iJ} = 1$, $i = 1, \ldots, m$. Then

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{baseline} = \begin{pmatrix} \frac{1}{\pi_{i1}} & 0 & \cdots & 0 & -\frac{1}{\pi_{iJ}} \\ 0 & \frac{1}{\pi_{i2}} & \ddots & \vdots & -\frac{1}{\pi_{iJ}} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{\pi_{iJ}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{cumulative} = \begin{pmatrix} \frac{1}{\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & \cdots & -\frac{1}{1-\gamma_{i1}} \\ \frac{1}{\gamma_{i2}} & \frac{1}{\gamma_{i2}} & -\frac{1}{1-\gamma_{i2}} & \cdots & -\frac{1}{1-\gamma_{i2}} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \frac{1}{\gamma_{i,J-1}} & \frac{1}{\gamma_{i,J-1}} & \cdots & \frac{1}{\gamma_{i,J-1}} & -\frac{1}{1-\gamma_{i,J-1}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{continuation} = \begin{pmatrix} \frac{1}{\pi_{i1}} & -\frac{1}{1-\gamma_{i1}} & -\frac{1}{1-\gamma_{i1}} & \cdots & -\frac{1}{1-\gamma_{i1}} \\ 0 & \frac{1}{\pi_{i2}} & -\frac{1}{1-\gamma_{i2}} & \cdots & -\frac{1}{1-\gamma_{i2}} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{1-\gamma_{i,J-1}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

$$(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})_{adjacent} = \begin{pmatrix} \frac{1}{\pi_{i1}} & -\frac{1}{\pi_{i2}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\pi_{i2}} & -\frac{1}{\pi_{i3}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\pi_{i,J-1}} & -\frac{1}{\pi_{iJ}} \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J}$$

where $\gamma_{ij} = \pi_{i1} + \cdots + \pi_{ij}$ is the cumulative categorical probability, $j = 1, \ldots, J-1$. The corresponding inverse matrices are

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{baseline} \\
&= \begin{pmatrix}
-\pi_{i1}^2 + \pi_{i1} & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{i,J-1} & \pi_{i1} \\
-\pi_{i1}\pi_{i2} & -\pi_{i2}^2 + \pi_{i2} & \cdots & -\pi_{i2}\pi_{i,J-1} & \pi_{i2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\pi_{i1}\pi_{i,J-1} & -\pi_{i2}\pi_{i,J-1} & \cdots & -\pi_{i,J-1}^2 + \pi_{i,J-1} & \pi_{i,J-1} \\
-\pi_{i1}\pi_{iJ} & -\pi_{i2}\pi_{iJ} & \cdots & -\pi_{i,J-1}\pi_{iJ} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&\triangleq \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{baseline}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{baseline} = \pi_{ij}(\mathbf{e}_j - \boldsymbol{\pi}_i)$, $j = 1, \ldots, J-1$, $(\mathbf{c}_{iJ})_{baseline} = \boldsymbol{\pi}_i$, and $\mathbf{e}_j$ is the $J \times 1$ vector with the jth coordinate 1 and all others 0. Recall that $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$.

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{cumulative} \\
&= \begin{pmatrix}
\gamma_{i1}(1-\gamma_{i1}) & 0 & \cdots & 0 & \pi_{i1} \\
-\gamma_{i1}(1-\gamma_{i1}) & \gamma_{i2}(1-\gamma_{i2}) & \ddots & \vdots & \pi_{i2} \\
0 & -\gamma_{i2}(1-\gamma_{i2}) & \ddots & 0 & \vdots \\
\vdots & \ddots & \ddots & \gamma_{i,J-1}(1-\gamma_{i,J-1}) & \pi_{i,J-1} \\
0 & \cdots & 0 & -\gamma_{i,J-1}(1-\gamma_{i,J-1}) & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&\triangleq \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{cumulative}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{cumulative} = \gamma_{ij}(1-\gamma_{ij})(\mathbf{e}_j - \mathbf{e}_{j+1})$ with $\mathbf{e}_j$ defined as above; and $(\mathbf{c}_{iJ})_{cumulative} = \boldsymbol{\pi}_i$.

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{continuation} \\
&= \begin{pmatrix}
\pi_{i1}(1-\gamma_{i1}) & 0 & \cdots & 0 & \pi_{i1} \\
-\pi_{i1}\pi_{i2} & \frac{\pi_{i2}(1-\gamma_{i2})}{1-\gamma_{i1}} & \ddots & \vdots & \pi_{i2} \\
\vdots & \vdots & \ddots & 0 & \vdots \\
-\pi_{i1}\pi_{i,J-1} & -\frac{\pi_{i2}\pi_{i,J-1}}{1-\gamma_{i1}} & \cdots & \frac{\pi_{i,J-1}(1-\gamma_{i,J-1})}{1-\gamma_{i,J-2}} & \pi_{i,J-1} \\
-\pi_{i1}\pi_{iJ} & -\frac{\pi_{i2}\pi_{iJ}}{1-\gamma_{i1}} & \cdots & -\frac{\pi_{i,J-1}\pi_{iJ}}{1-\gamma_{i,J-2}} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&= \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{continuation}
\end{aligned}
$$

where $(\mathbf{c}_{i1})_{continuation} = \pi_{i1}(1-\gamma_{i1}, -\pi_{i2}, \ldots, -\pi_{iJ})^T$, $(\mathbf{c}_{ij})_{continuation} = \frac{\pi_{ij}}{1-\gamma_{i,j-1}}(0, \ldots, 0, 1 - \gamma_{ij}, -\pi_{i,j+1}, \ldots, -\pi_{iJ})^T$ with "$1-\gamma_{ij}$" being the jth coordinate, $j = 2, \ldots, J-1$, and $(\mathbf{c}_{iJ})_{continuation} = \boldsymbol{\pi}_i$.

$$
\begin{aligned}
&(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1}_{adjacent} \\
&= \begin{pmatrix}
(1-\gamma_{i1})\pi_{i1} & (1-\gamma_{i2})\pi_{i1} & \cdots & (1-\gamma_{i,J-1})\pi_{i1} & \pi_{i1} \\
-\gamma_{i1}\pi_{i2} & (1-\gamma_{i2})\pi_{i2} & \cdots & (1-\gamma_{i,J-1})\pi_{i2} & \pi_{i2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
-\gamma_{i1}\pi_{i,J-1} & -\gamma_{i2}\pi_{i,J-1} & \cdots & (1-\gamma_{i,J-1})\pi_{i,J-1} & \pi_{i,J-1} \\
-\gamma_{i1}\pi_{iJ} & -\gamma_{i2}\pi_{iJ} & \cdots & -\gamma_{i,J-1}\pi_{iJ} & \pi_{iJ}
\end{pmatrix}_{J\times J} \\
&= \begin{pmatrix} \mathbf{c}_{i1} & \mathbf{c}_{i2} & \cdots & \mathbf{c}_{iJ} \end{pmatrix}_{adjacent}
\end{aligned}
$$

where $(\mathbf{c}_{ij})_{adjacent} = ((1-\gamma_{ij})\pi_{i1}, \ldots, (1-\gamma_{ij})\pi_{ij}, -\gamma_{ij}\pi_{i,j+1}, \ldots, -\gamma_{ij}\pi_{iJ})^T$, $j = 1, \ldots, J-1$, and $(\mathbf{c}_{iJ})_{adjacent} = \boldsymbol{\pi}_i$.

From the above results, $u_{st}(\boldsymbol{\pi}_i)$ can be calculated for 4 kinds of logit models according to Equation 3.3,

$$
u_{st}(\boldsymbol{\pi}_i)_{baseline} = \begin{cases} \pi_{is}(1-\pi_{is}), & \text{if } s = t; \\ -\pi_{is}\pi_{it} \ , & \text{if } s \neq t \end{cases} \tag{4.6}
$$

$$u_{st}(\boldsymbol{\pi}_i)_{cumulative} = \begin{cases} \gamma_{is}^2(1-\gamma_{is})^2(\pi_{is}^{-1}+\pi_{i,s+1}^{-1}), & \text{if } s=t; \\ -\gamma_{is}\gamma_{it}(1-\gamma_{is})(1-\gamma_{it})\pi_{i,s\vee t}^{-1}, & \text{if } |s-t|=1; \\ 0, & \text{if } |s-t|\geqslant 2 \end{cases} \tag{4.7}$$

$$u_{st}(\boldsymbol{\pi}_i)_{continuation} = \begin{cases} \frac{\pi_{is}(1-\gamma_{is})}{1-\gamma_{i,s-1}}, & \text{if } s=t; \\ 0, & \text{if } |s-t|\geqslant 1 \end{cases} \tag{4.8}$$

$$u_{st}(\boldsymbol{\pi}_i)_{adjacent} = \begin{cases} \gamma_{is}(1-\gamma_{is}), & \text{if } s=t; \\ \gamma_{i,s\wedge t}(1-\gamma_{i,s\vee t}), & \text{if } s\neq t \end{cases} \tag{4.9}$$

where $s, t = 1, \ldots, J-1$, $s \wedge t$ stands for $\min\{s,t\}$ and $s \vee t$ stands for $\max\{s,t\}$.

With $u_{st}(\boldsymbol{\pi}_i)$ expressed, the determinant of Fisher information matrix with non-proportional odds can be expressed in analytic form in some special cases.

4.3.3 Some preliminary results

For multinomial logit models with nonproportional odds, $\mathbf{X}_i$ is defined as in Equation 2.5 and $\boldsymbol{\theta}$ is defined as in Equation 2.6. Then $\mathbf{H}$ in Equation 3.7 is

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_1 & & \\ & \ddots & \\ & & \mathbf{H}_{J-1} \end{pmatrix}$$

Here $\mathbf{H}$ is $p \times m(J-1)$ matrix, and $p = p_1 + \cdots + p_{J-1}$, $\mathbf{U}$ is the same as defined in *Equation* 3.7, which is a $m(J-1) \times m(J-1)$ matrix. Therefore $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$ is a $p \times p$ matrix.

Lemma 4.4. $rank(\mathbf{H}) = rank(\mathbf{H}_1) + \cdots + rank(\mathbf{H}_{J-1})$

Lemma 4.5. $|\mathbf{F}| \neq 0$ *only if* $rank(\mathbf{H}) = p$

Theorem 4.5. $|\mathbf{F}| > 0$ *only if* $m \geqslant max_s p_s$

Proof of Theorem 4.5: Without loss of generality, assume $p_1 = max_s p_s$.

If $m < p_1$, then $rank(\mathbf{H}_1) < p_1$ and thus $rank(\mathbf{H}) < p$.

According to Lemma 4.5, $|\mathbf{F}| = 0$ in this case. □

Lemma 4.6. *If* $p = m(J-1)$ *and* $rank(\mathbf{H}) = p$ *then* $p_1 = \cdots = p_{J-1} = m$

Theorem 4.6. *If* $p_1 = \cdots = p_{J-1} = m$ *then*

$$|\mathbf{F}| = (\prod_{i=1}^{m} n_i)^{J-1} \prod_{j=1}^{J-1} |\mathbf{H}_j|^2 |\mathbf{V}| \tag{4.10}$$

Which implies minimally supported points D-optimal design is $n_1 = n_2 = \cdots = n_m = n/m$. Note in Theorem 4.6, we don't need $\mathbf{h}_{1i} = \cdots = \mathbf{h}_{J-1,i}$, we only need $p_1 = \cdots = p_{J-1}$. Also, all of conclusions are based on given experimental points and parameters (local optimal design).

Because of this, we conclude:

(1) The Fisher information matrix $\mathbf{F}$ is positive definite if and only if $\mathbf{H}_s = (\mathbf{h}_s(\mathbf{x}_1), \ldots, \mathbf{h}_s(\mathbf{x}_m))^T$ is of full rank p_s for each $s = 1, \ldots, J-1$.

(2) $\max\{p_s, s = 1, \ldots, J-1\}$ is also the number of support points in a minimally supported design.

(3) A uniform design is D-optimal in a minimally supported design given same dimension of

predictors for different logit equations, i.e., $m = p_1 = p_2 = \cdots = p_{J-1}$.

4.3.4 Continuation-ratio logit model with NPO

First, all of lemmas and theorems and conclusions in subsection 4.3.3 can be applied here.

Second, look at $\mathbf{V}_i$ defined in Theorem 4.4. Refer to Equation 4.8, $\mathbf{V}_i$ is a diagonal matrix, so its determinant is

$$|\mathbf{V}_i| = \prod_{s=1}^{J-1} u_{ss}(\boldsymbol{\pi}_i) = \prod_{j=1}^{J} \pi_{ij}, \quad i = 1, \ldots, m.$$

Combine this with Equation 4.5, we get

$$|\mathbf{V}_{continuation}| = \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij}$$

Plug this equation in Equation 4.10 of Theorem 4.6, we get

$$|\mathbf{F}_{continuation}| = \prod_{j=1}^{J-1} |\mathbf{H}_j|^2 \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij} (\prod_{i=1}^{m} n_i)^{J-1} \tag{4.11}$$

4.3.5 Baseline-category logit model with NPO

Consider the baseline-category logit model with nonproportional odds for nominal responses. First, all of lemmas and theorems in subsection 4.3.3 still can be applied here, and the conclusions are the same.

Look at $\mathbf{V}_i$ defined in Theorem 4.4, we first add last row and last column to it to get a larger $\mathbf{V}_i'$, then we do some column operations and row operations to change it into an upper triangular matrix, all of these operations won't change the determinant of $\mathbf{V}_i$.

$$\mathbf{V}_i' = \begin{pmatrix} \pi_{i1}(1-\pi_{i1}) & -\pi_{i1}\pi_{i2} & \cdots & -\pi_{i1}\pi_{i,J-1} & \pi_{i1} \\ -\pi_{i1}\pi_{i2} & \pi_{i2}(1-\pi_{i2}) & \cdots & -\pi_{i2}\pi_{i,J-1} & \pi_{i2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\pi_{i1}\pi_{i,J-1} & -\pi_{i2}\pi_{i,J-1} & \cdots & \pi_{i,J-1}(1-\pi_{i,J-1}) & \pi_{i,J-1} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

$$(C_j \to C_j + C_J \times \pi_{ij}, \quad j = 1, \cdots, J-1)$$

$$= \begin{pmatrix} \pi_{i1} & 0 & \cdots & 0 & \pi_{i1} \\ 0 & \pi_{i2} & \ddots & \vdots & \pi_{i2} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \pi_{i,J-1} & \pi_{i,J-1} \\ \pi_{i1} & \pi_{i2} & \cdots & \pi_{i,J-1} & 1 \end{pmatrix}$$

$$(R_J \to R_J - \sum_{j=1}^{J-1} R_j)$$

$$= \begin{pmatrix} \pi_{i1} & 0 & \cdots & 0 & \pi_{i1} \\ 0 & \pi_{i2} & \ddots & \vdots & \pi_{i2} \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & \pi_{i,J-1} & \pi_{i,J-1} \\ 0 & 0 & \cdots & 0 & 1 - \pi_{i1} - \cdots - \pi_{i,J-1} \end{pmatrix}$$

In the above operations, R_j means jth row and C_j means jth column. Therefore

$$|\mathbf{V}_i| = \prod_{j=1}^{J} \pi_{ij}, \quad i = 1, \dots, m.$$

Combine this with Equation 4.5, we get

$$|\mathbf{V}_{baseline}| = \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij}$$

Plug this equation in Equation 4.10 of Theorem 4.6, we get

$$|\mathbf{F}_{baseline}| = \prod_{j=1}^{J-1} |\mathbf{H}_j|^2 \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij} (\prod_{i=1}^{m} n_i)^{J-1} \tag{4.12}$$

It is exactly same as Equation 4.11.

4.3.6 Adjacent-categories logit model with NPO

Consider the adjacent-categories logit model with nonproportional odds. First, all of lemmas and theorems in subsection 4.3.3 still can be applied here, and the conclusions are the same.

Look at $\mathbf{V}_i$ defined in Theorem 4.4, we first add last row and last column to it to get a larger $\mathbf{V}_i'$, then we do some column operations and row operations to change it into an upper triangular matrix, all of these operations won't change the determinant of $\mathbf{V}_i$.

$$\mathbf{V}_i' = \begin{pmatrix} \gamma_{i1}(1-\gamma_{i1}) & \gamma_{i1}(1-\gamma_{i2}) & \cdots & \gamma_{i1}(1-\gamma_{i,J-1}) & \gamma_{i1} \\ \gamma_{i1}(1-\gamma_{i2}) & \gamma_{i2}(1-\gamma_{i2}) & \cdots & \gamma_{i2}(1-\gamma_{i,J-1}) & \gamma_{i2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{i1}(1-\gamma_{i,J-1}) & \gamma_{i2}(1-\gamma_{i,J-1}) & \cdots & \gamma_{i,J-1}(1-\gamma_{i,J-1}) & \gamma_{i,J-1} \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

$$(C_j \to C_j + C_J \times \gamma_{ij}, \quad j = 1, \cdots, J-1)$$

$$\Rightarrow \begin{pmatrix} \gamma_{i1} & \gamma_{i1} & \cdots & \gamma_{i1} & \gamma_{i1} \\ \gamma_{i1} & \gamma_{i2} & \cdots & \gamma_{i2} & \gamma_{i2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{i1} & \gamma_{i2} & \cdots & \gamma_{i,J-1} & \gamma_{i,J-1} \\ \gamma_{i1} & \gamma_{i2} & \cdots & \gamma_{i,J-1} & 1 \end{pmatrix}$$

$$(R_j \to R_j - R_{j-1}, \quad j = 2, \cdots, J)$$

$$\Rightarrow \begin{pmatrix} \gamma_{i1} & \gamma_{i1} & \cdots & \gamma_{i1} & \gamma_{i1} \\ 0 & \gamma_{i2}-\gamma_{i1} & \cdots & \gamma_{i2}-\gamma_{i1} & \gamma_{i2}-\gamma_{i1} \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & \gamma_{i,J-1}-\gamma_{i,J-2} & \gamma_{i,J-1}-\gamma_{i,J-2} \\ 0 & 0 & \cdots & 0 & 1-\gamma_{i,J-1} \end{pmatrix}$$

In the above operations, R_j means jth row and C_j means jth column. Therefore

$$|\mathbf{V}_i| = \prod_{j=1}^{J} \pi_{ij}, \quad i = 1, \ldots, m.$$

Combine this with Equation 4.5, we get

$$|\mathbf{V}_{adjacent}| = \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij}$$

Plug this equation in Equation 4.10 of Theorem 4.6, we get

$$|\mathbf{F}_{adjacent}| = \prod_{j=1}^{J-1} |\mathbf{H}_j|^2 \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij} (\prod_{i=1}^{m} n_i)^{J-1} \tag{4.13}$$

It is exactly same as Equation 4.11 and Equation 4.12. Note the determinants of Fisher information matrix for non-proportional odds models have the same expression for the above three logits: continuation-ratio logit, baseline-category logit and Adjacent-categories logit, only if $p_1 = p_2 = \cdots = p_{J-1} = m$.

4.3.7 Cumulative logit model with NPO

We need to simplify notation in order to express $\mathbf{V}_i$ defined in Theorem 4.4, then we do some column operations and row operations to change it into an upper triangular matrix, finally get the determinant of $\mathbf{V}_i$.

Let $f_{ij} = \gamma_{ij}(1-\gamma_{ij}), \quad j = 1, \cdots, J-1$, we have

$$\mathbf{V}_i = \begin{pmatrix} f_{i1}^2(\frac{1}{\pi_{i1}}+\frac{1}{\pi_{i2}}) & -f_{i1}f_{i2}\frac{1}{\pi_{i2}} & 0 & \cdots & 0 \\ -f_{i1}f_{i2}\frac{1}{\pi_{i2}} & f_{i2}^2(\frac{1}{\pi_{i2}}+\frac{1}{\pi_{i3}}) & -f_{i2}f_{i3}\frac{1}{\pi_{i3}} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -f_{i,J-3}f_{i,J-2}\frac{1}{\pi_{i,J-2}} & f_{i,J-2}^2(\frac{1}{\pi_{i,J-2}}+\frac{1}{\pi_{i,J-1}}) & -f_{i,J-2}f_{i,J-1}\frac{1}{\pi_{i,J-1}} \\ 0 & \cdots & 0 & -f_{i,J-2}f_{i,J-1}\frac{1}{\pi_{i,J-1}} & f_{i,J-1}^2(\frac{1}{\pi_{i,J-1}}+\frac{1}{\pi_{i,J}}) \end{pmatrix}$$

$(R_2 \to R_2 + R_1\gamma_{i1}f_{i2}/\gamma_{i2}f_{i1})$

$$\Rightarrow \begin{pmatrix} \frac{f_{i1}^2\gamma_{i2}}{\gamma_{i1}\pi_{i2}} & -f_{i1}f_{i2}\frac{1}{\pi_{i2}} & 0 & \cdots & 0 \\ 0 & \frac{f_{i2}^2\gamma_{i3}}{\gamma_{i2}\pi_{i3}} & -f_{i2}f_{i3}\frac{1}{\pi_{i3}} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -f_{i,J-3}f_{i,J-2}\frac{1}{\pi_{i,J-2}} & f_{i,J-2}^2(\frac{1}{\pi_{i,J-2}}+\frac{1}{\pi_{i,J-1}}) & -f_{i,J-2}f_{i,J-1}\frac{1}{\pi_{i,J-1}} \\ 0 & \cdots & 0 & -f_{i,J-2}f_{i,J-1}\frac{1}{\pi_{i,J-1}} & f_{i,J-1}^2(\frac{1}{\pi_{i,J-1}}+\frac{1}{\pi_{i,J}}) \end{pmatrix}$$

$(R_3 \to R_3 + R_2\gamma_{i2}f_{i3}/\gamma_{i3}f_{i2})\cdots$

$$\Rightarrow \begin{pmatrix} \frac{f_{i1}^2\gamma_{i2}}{\gamma_{i1}\pi_{i2}} & -f_{i1}f_{i2}\frac{1}{\pi_{i2}} & 0 & \cdots & 0 \\ 0 & \frac{f_{i2}^2\gamma_{i3}}{\gamma_{i2}\pi_{i3}} & -f_{i2}f_{i3}\frac{1}{\pi_{i3}} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & \frac{f_{i,J-2}^2\gamma_{i,J-1}}{\gamma_{i,J-2}\pi_{i,J-1}} & -f_{i,J-2}f_{i,J-1}\frac{1}{\pi_{i,J-1}} \\ 0 & \cdots & 0 & 0 & \frac{f_{i,J-1}^2\gamma_{i,J}}{\gamma_{i,J-1}\pi_{i,J}} \end{pmatrix}$$

Then

$$|\mathbf{V}_i| = \frac{\prod_{j=1}^{J-1} f_{ij}}{\prod_{j=1}^{J}\pi_{ij}} = \frac{\prod_{j=1}^{J-1}\gamma_{ij}(1-\gamma_{ij})}{\prod_{j=1}^{J}\pi_{ij}}, \quad i = 1, \ldots, m.$$

Combine this with Equation 4.5, we get

$$|\mathbf{V}_{cumulative}| = \frac{\prod_{i=1}^{m}\prod_{j=1}^{J-1}\gamma_{ij}(1-\gamma_{ij})}{\prod_{i=1}^{m}\prod_{j=1}^{J}\pi_{ij}}$$

Plug this equation in Equation 4.10 of Theorem 4.6, we get

$$|\mathbf{F}_{cumulative}| = \prod_{j=1}^{J-1}|\mathbf{H}_j|^2(\prod_{i=1}^{m}n_i)^{J-1}\frac{\prod_{i=1}^{m}\prod_{j=1}^{J-1}\gamma_{ij}(1-\gamma_{ij})}{\prod_{i=1}^{m}\prod_{j=1}^{J}\pi_{ij}} \tag{4.14}$$

Cumulative logit model's determinant is different from other logit model's determinants.

4.4 Alternative Approach to Explore Determinant of Fisher Information Matrix for NPO

Section 4.3 discusses determinant of Fisher information matrix in some special cases. This section considers another special case: NPO. It is our first work to explore determinant of Fisher information matrix, it just works in continuation-ratio logit models, but the determinant's formula is much simpler than the general formula in that case. For other 3 logit models, it also gives us some idea about the structure of Fisher information matrix. This part of work is not complete, but has its own meaning, so we show it in the following subsections.

4.4.1 Preliminary results for NPO

From Equation 3.5, we get the Fisher information matrix for NPO as

$$\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{F}_i = \begin{pmatrix} \mathbf{F}_{11} & \cdots & \mathbf{F}_{1,J-1} \\ \vdots & \ddots & \vdots \\ \mathbf{F}_{J-1,1} & \cdots & \mathbf{F}_{J-1,J-1} \end{pmatrix} \tag{4.15}$$

where

$$\begin{aligned} \mathbf{F}_{st} &= \sum_{i=1}^{m} n_i \mathbf{h}_s(\mathbf{x}_i) \mathbf{c}_{is}^T diag(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it} \mathbf{h}_t^T(\mathbf{x}_i) \\ &= \sum_{i=1}^{m} n_i u_{st}(\boldsymbol{\pi}_i) \mathbf{h}_s(\mathbf{x}_i) \mathbf{h}_t^T(\mathbf{x}_i) \\ &= \begin{pmatrix} \mathbf{h}_s(\mathbf{x}_1) & \cdots & \mathbf{h}_s(\mathbf{x}_m) \end{pmatrix} \begin{pmatrix} n_1 u_{st}(\boldsymbol{\pi}_1) & & 0 \\ & \ddots & \\ 0 & & n_m u_{st}(\boldsymbol{\pi}_m) \end{pmatrix} \begin{pmatrix} \mathbf{h}_t^T(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}_t^T(\mathbf{x}_m) \end{pmatrix} \\ &\triangleq \mathbf{H}_s \cdot \mathbf{U}_{st} \cdot \mathbf{H}_t^T \end{aligned}$$

and

$$u_{st}(\boldsymbol{\pi}_i) = \mathbf{c}_{is}^T diag(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it}$$

$\mathbf{F}$ is a block matrix.

4.4.2 Determinant of Fisher information matrix for continuation-ratio logit models with NPO

In this subsection, we consider the continuation-ratio logit model with nonproportional odds. We got Fisher information matrix from Equation 3.7 in the previous section. If we start from Equation 3.5, we can obtain the same results.

Recall that $\gamma_{is} = \pi_{i1} + \cdots + \pi_{is}$, $s = 1, \ldots, J-1$. We define $\gamma_{i0} = 0$ and $\gamma_{iJ} = 1$. We rewrite Equation 4.8 as follows:

$$u_{st}(\boldsymbol{\pi}_i)_{continuation} = \begin{cases} \frac{\pi_{is}(1-\gamma_{is})}{1-\gamma_{i,s-1}}, & \text{if } s = t; \\ 0, & \text{if } |s-t| \geqslant 1 \end{cases}$$

Where $s, t = 1, \ldots, J-1$. Then based on above equation and Equation 3.5 of $\mathbf{F}$ for NPO, we derive the theorem as follows:

Theorem 4.7. *For a continuation-ratio logit model with nonproportional odds, the Fisher information matrix*

$$\mathbf{F}_{continuation} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{F}_{J-1,J-1} \end{pmatrix}$$

and its determinant $|\mathbf{F}_{continuation}| = |\mathbf{F}_{11}||\mathbf{F}_{22}|\cdots|\mathbf{F}_{J-1,J-1}|$ *with*

$$\mathbf{F}_{ss} = \sum_{i=1}^{m} n_i u_{ss}(\boldsymbol{\pi}_i)\mathbf{h}_s(\mathbf{x}_i)\mathbf{h}_s^T(\mathbf{x}_i) = \mathbf{H}_s\mathbf{U}_{ss}\mathbf{H}_s^T, \quad s = 1, \ldots, J-1$$

$\mathbf{F}_{continuation}$ for NPO is a diagonal block matrix. Assume $0 < \pi_{is} < 1$ and $n_i > 0$ for $i = 1, \ldots, m$ and $s = 1, \ldots, J$. According to Equation 4.8, $n_i u_{ss}(\boldsymbol{\pi}_i) > 0$, for all $i = 1, \ldots, m$ and $s = 1, \cdots, J-1$. That is, $\mathbf{U}_{ss} = \mathrm{d}iag(n_1 u_{ss}(\boldsymbol{\pi}_1), \ldots, n_m u_{ss}(\boldsymbol{\pi}_m))$ is positive definite. Thus $\mathbf{F}_{ss}$ is positive definite if and only if the $p_s \times m$ matrix $\mathbf{H}_s = (\mathbf{h}_s(\mathbf{x}_1), \ldots, \mathbf{h}_s(\mathbf{x}_m))^T$ is of full rank p_s, which also implies $m \geq p_s$, $s = 1, \ldots, J-1$ in this case.

Therefore, in order to keep the Fisher information matrix positive definite, a minimally supported design contains at least $\max\{p_1, \ldots, p_{J-1}\}$ design points in this case.

As a direct conclusion of Theorem 4.7 and Lemma 3.1 in (21), we have the explicit form of the determinant as follows:

Theorem 4.8. *Consider an exact design $\xi = \{(\mathbf{x}_i, n_i)_{i=1,\ldots,m}\}$ with $n_i > 0, i = 1, \ldots, m$ for a continuation-ratio logit model with nonproportional odds for hierarchical responses. Suppose $0 < \pi_{is} < 1, i = 1, \ldots, m; s = 1, \ldots, J$. Then the Fisher information matrix $\mathbf{F}$ is positive definite if and only $\mathbf{H}_s = (\mathbf{h}_s(\mathbf{x}_1), \ldots, \mathbf{h}_s(\mathbf{x}_m))^T$ is of full rank p_s for each $s = 1, \ldots, J-1$. In this case, we must have $m \geq \max\{p_s, s = 1, \ldots, J-1\}$ and*

$$|\mathbf{F}| = \prod_{s=1}^{J-1} \left(\sum_{1 \leqslant i_1 < \cdots < i_{p_s} \leqslant m} |\mathbf{H}_s[i_1, \cdots i_{p_s}]|^2 n_{i_1} \cdots n_{i_{p_s}} u_{ss}(\boldsymbol{\pi}_{i_1}) \cdots u_{ss}(\boldsymbol{\pi}_{i_{p_s}}) \right) \tag{4.16}$$

where $\mathbf{H}_s[i_1, \cdots i_{p_s}]$ is the submatrix consisting of the $i_1, \ldots, i_{p_s}$th rows of $\mathbf{H}_s$.

Remark 4.1. *The Equation 4.16 shows that $|\mathbf{F}_{continuation}|$ is an order-p homogenous polynomial of $(n_1, \ldots, n_m)$. If an approximate design $\xi = \{(\mathbf{x}_i, w_i),\ i = 1, \ldots, m\}$ with $w_i = n_i/n > 0$ and $\sum_{i=1}^m w_i = 1$ is considered, then Theorem 4.8 is true with n_i replaced with w_i. In order to* *numerically find out a D-optimal design, a lift-one algorithm similar as the one in (21) can be derived accordingly.* □

Due to Equation 4.8, we can verify that

$$\prod_{s=1}^{J-1} u_{ss}(\boldsymbol{\pi}_i) = \prod_{j=1}^{J} \pi_{ij}, \quad i = 1, \ldots, m.$$

Then we can derive the corollary as follows:

Corollary 4.3. *Under the assumptions of Theorem 4.8, assume further* $\mathbf{h}_1 = \cdots = \mathbf{h}_{J-1}$ *. Then* $p_1 = \cdots = p_{J-1}$*,* $\mathbf{H}_1 = \cdots = \mathbf{H}_{J-1}$ *. Suppose further there exist* $m = p_1$ *design points* $\mathbf{x}_1, \ldots, \mathbf{x}_m$ *such that the* $p_1 \times p_1$ *matrix* $\mathbf{H}_1$ *is of full rank* p_1*. Then*

$$|\mathbf{F}_{continuation}| = |\mathbf{H}_1|^{2(J-1)} \prod_{i=1}^{m} \prod_{j=1}^{J} \pi_{ij} \left(\prod_{i=1}^{m} n_i \right)^{J-1} \tag{4.17}$$

which attains its maximum at $n_1 = \cdots = n_m = n/m$ *.*

This equation is exactly same as Equation 4.11 given same condition.

Example 4.2. *Suppose there are* $d = 2$ *design factors* x_1, x_2*,* $J = 4$ *response categories, and* $m = 3$ *support points* $\mathbf{x}_i = (x_{i1}, x_{i2})^T, i = 1, 2, 3$*. Consider an exact design* $\xi = \{(\mathbf{x}_i, n_i)_{i=1,\ldots,3}\}$ *with* $n_i > 0$ *for a continuation-ratio logit model with nonproportional odds. Assume* $\mathbf{h}_s(\mathbf{x}_i) = (1, x_{i1}, x_{i2})^T$ *is the same for* $s = 1, 2, 3$*. Then*

$$|\mathbf{F}_{continuation}| = |\mathbf{H}_1|^6 (n_1 n_2 n_3)^3 \prod_{i=1}^{3} \prod_{j=1}^{4} \pi_{ij}$$

where $|\mathbf{H}_1| = x_{31}(-x_{12}+x_{22}) + x_{21}(x_{12}-x_{32}) + x_{11}(-x_{22}+x_{32})$. □

4.4.3 Determinant of Fisher information matrix for other three logit models with NPO

In the case of cumulative logit model, we rewrite Equation 4.7

$$u_{st}(\boldsymbol{\pi}_i)_{cumulative} = \begin{cases} \gamma_{is}^2(1-\gamma_{is})^2(\pi_{is}^{-1} + \pi_{i,s+1}^{-1}), & \text{if } s = t; \\ -\gamma_{is}\gamma_{it}(1-\gamma_{is})(1-\gamma_{it})\pi_{i,s\vee t}^{-1}, & \text{if } |s-t| = 1; \\ 0, & \text{if } |s-t| \geqslant 2 \end{cases}$$

Where $s, t = 1, \ldots, J-1$ and $s \vee t$ stands for $\max\{s, t\}$. Then

$$\mathbf{F}_{cumulative} = \begin{pmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} & \ddots & \vdots \\ \mathbf{0} & \ddots & \ddots & \ddots & \mathbf{0} \\ \vdots & \ddots & \mathbf{F}_{J-2,J-3} & \mathbf{F}_{J-2,J-2} & \mathbf{F}_{J-2,J-1} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{F}_{J-1,J-2} & \mathbf{F}_{J-1,J-1} \end{pmatrix}$$

is a tri-diagonal block matrix.

In the case of nominal response model, from Equation 4.6

$$u_{st}(\boldsymbol{\pi}_i)_{baseline} = \begin{cases} \pi_{is}(1-\pi_{is}), & \text{if } s = t; \\ -\pi_{is}\pi_{it}\ , & \text{if } s \neq t \end{cases}$$

The Fisher information matrix's structure can't be simplified.

In the case of adjacent-categories logit model, from Equation 4.9

$$u_{st}(\boldsymbol{\pi}_i)_{adjacent} = \begin{cases} \gamma_{is}(1-\gamma_{is}), & \text{if } s = t; \\ \gamma_{i,s\wedge t}(1-\gamma_{i,s\vee t}), & \text{if } s \neq t \end{cases}$$

Where $s, t = 1, \ldots, J-1$, $s \wedge t$ stands for $\min\{s,t\}$ and $s \vee t$ stands for $\max\{s,t\}$. The Fisher information matrix's structure can't be simplified.

CHAPTER 5

RELATED FORMULAS FOR MULTINOMIAL LOGISTIC MODELS AND DESIGN SPACE

The multinomial logistic models Equation 2.1 include two sets of quantities related to the model parameters. One set consists of π_{ij}, $i = 1, \ldots, m;\ j = 1, \ldots, J-1$. For typical applications, we assume $0 < \pi_{ij} < 1$, $i = 1, \ldots, m;\ j = 1, \ldots, J-1$. The other set consists of $\boldsymbol{\theta}$, which typically are real vectors. A multinomial logistic model connects the two sets of quantities by a log link and the model matrix $\mathbf{X}_i$ described in Equation 2.3, which consists of both $\mathbf{h}_j^T(\mathbf{x}_i)$ and $\mathbf{h}_c^T(\mathbf{x}_i)$, $i = 1, \ldots, m$, $j = 1, \ldots, J-1$.

Definition 5.1. *The design space $\mathcal{X}$ for a multinomial logistic model Equation 2.1 is the collection of design points or level combinations of design factors $\mathbf{x} = (x_1, \ldots, x_d)^T$ such that the categorical probabilities of response $(\pi_1, \ldots, \pi_J)$ exist uniquely and satisfy $0 < \pi_j < 1$, $j = 1, \ldots, J$. That is,*

$$\mathcal{X} = \left\{\mathbf{x} = (x_1, \ldots, x_d)^T \mid 0 < \pi_j < 1, j = 1, \ldots, J \textit{ exist uniquely}\right\}$$

5.1 Baseline-Category Logit Model for Nominal Response

Recall the baseline-category logit model for nominal response (2; 3)

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j = 1, \ldots, J-1 \tag{5.1}$$

Proposition 5.1. *Fixing* $\mathbf{x}_i$*,* $\boldsymbol{\beta}_j, j = 1, \cdots, J-1$ *and* $\boldsymbol{\zeta}$ *in Equation 5.1, let* $\eta_{ij} = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$*,* $j = 1, \ldots, J-1$*. Then* $0 < \pi_{ij} < 1, j = 1, \ldots, J$ *exist uniquely if and only if* $-\infty < \eta_{ij} < \infty$*,* $j = 1, \ldots, J-1$*. In this case,*

$$\pi_{ij} = \begin{cases} \frac{e^{\eta_{ij}}}{e^{\eta_{i1}} + \cdots + e^{\eta_{i,J-1}} + 1} & 1 \leqslant j \leqslant J-1 \\ \frac{1}{e^{\eta_{i1}} + \cdots + e^{\eta_{i,J-1}} + 1} & j = J \end{cases} \tag{5.2}$$

Proof of Proposition 5.1: Write $y_j = \log \pi_{ij}$, $j = 1, \ldots, J$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ if and only if $y_j \in (-\infty, 0)$, $j = 1, \ldots, J$. In this case, Equation 5.1 implies $\eta_{ij} = y_j - y_J \in (-\infty, \infty)$, $j = 1, \ldots, J-1$.

On the other hand, for any given $\eta_{i1}, \ldots, \eta_{i,J-1} \in (-\infty, \infty)$, $y_j = \eta_{ij} + y_J$, $j = 1, \ldots, J-1$. Note that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{\eta_{i1} + y_J} + e^{\eta_{i2} + y_J} + \cdots + e^{\eta_{i,J-1} + y_J} + e^{y_J} \\ &= e^{y_J}\left(e^{\eta_{i1}} + e^{\eta_{i2}} + \cdots + e^{\eta_{i,J-1}} + 1\right) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in Equation 5.2, and thus $\pi_{ij} \in (0, 1)$ exists and is unique, $j = 1, \ldots, J$. $\square$

5.2 Cumulative Logit Model for Ordinal Response

The cumulative logit model (with partial-proportional odds) for ordinal responses (4; 5) is described in general as follows:

$$\log\left(\frac{\pi_{i1}+\cdots+\pi_{ij}}{\pi_{i,j+1}+\cdots+\pi_{iJ}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ , \quad j=1,\ldots,J-1 \tag{5.3}$$

Proposition 5.2. *Fixing $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j=1,\cdots,J-1$ and $\boldsymbol{\zeta}$ in Equation 5.3, let $\eta_{ij} = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j=1,\ldots,J-1$. Then $0<\pi_{ij}<1, j=1,\ldots,J$ exist and are unique if and only if $-\infty<\eta_{i1}<\eta_{i2}<\cdots<\eta_{i,J-1}<\infty$. In this case,*

$$\pi_{ij} = \begin{cases} \frac{\exp(\eta_{i1})}{1+\exp(\eta_{i1})} & j=1 \\ \frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} - \frac{\exp(\eta_{i,j-1})}{1+\exp(\eta_{i,j-1})} & 1<j<J \\ \frac{1}{1+\exp(\eta_{i,J-1})} & j=J \end{cases} \tag{5.4}$$

Proof of Proposition 5.2: Taking $j=1$ in Equation 5.3, we get $\log\left(\frac{\pi_{i1}}{1-\pi_{i1}}\right)=\eta_{i1}$ and thus $\pi_{i1}=\frac{\exp(\eta_{i1})}{1+\exp(\eta_{i1})}$. Then $0<\pi_{i1}<1$ if and only if $-\infty<\eta_{i1}<\infty$.

For $j=2,\cdots,J-1$,

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})} - \frac{\exp(\eta_{i,j-1})}{1+\exp(\eta_{i,j-1})}$$

which implies that $\pi_{ij} > 0$ if and only if $\eta_{ij} > \eta_{i,j-1}$. Therefore, $\pi_{iJ} = 1-(\pi_{i1}+\cdots+\pi_{i,J-1}) = 1 - \frac{\exp(\eta_{i,J-1})}{1+\exp(\eta_{i,J-1})} = \frac{1}{1+\exp(\eta_{i,J-1})}$, which indicates $0 < \pi_{iJ} < 1$ if and only if $-\infty < \eta_{i,J-1} < \infty$. Given $\pi_{i1} + \cdots + \pi_{iJ} = 1$, we have

$$-\infty < \eta_{i1} < \eta_{i2} < \cdots < \eta_{i,J-1} < \infty \Leftrightarrow \pi_{ij} \in (0,1), \quad j = 1, \ldots, J$$

□

Corollary 5.1. *For proportional odds model*

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{iJ}}\right) = \beta_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta} \ , \quad j = 1, \ldots, J-1 \tag{5.5}$$

The design space becomes

$$-\infty < \beta_1 < \beta_2 < \cdots < \beta_{J-1} < \infty \Leftrightarrow \pi_{ij} \in (0,1), \quad j = 1, \ldots, J$$

5.3 Adjacent-Categories Logit Model for Ordinal Response

The adjacent-categories logit model for ordinal responses (3; 6) takes the form as follows:

$$\log\left(\frac{\pi_{ij}}{\pi_{i,j+1}}\right) = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta} \ , \quad j = 1, \ldots, J-1 \tag{5.6}$$

Proposition 5.3. *Fixing $\mathbf{x}_i$, $\boldsymbol{\beta}_j, j = 1, \cdots, J-1$ and $\boldsymbol{\zeta}$ in Equation 5.6, let $\eta_{ij} = \mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ exist uniquely if and only if $-\infty < \eta_{ij} < \infty$, $j = 1, \ldots, J-1$. In this case,*

$$\pi_{ij} = \begin{cases} \frac{\exp(\eta_{i,J-1}+\cdots+\eta_{ij})}{\exp(\eta_{i,J-1}+\cdots+\eta_{i1})+\exp(\eta_{i,J-1}+\cdots+\eta_{i2})+\cdots+\exp(\eta_{i,J-1})+1} & j = 1, \ldots, J-1 \\ \frac{1}{\exp(\eta_{i,J-1}+\cdots+\eta_{i1})+\exp(\eta_{i,J-1}+\cdots+\eta_{i2})+\cdots+\exp(\eta_{i,J-1})+1} & j = J \end{cases} \tag{5.7}$$

Proof of Proposition 5.3: Let $y_j = \log \pi_{ij}$. Then $0 < \pi_{ij} < 1, j = 1, \ldots, J$ if and only if $y_j \in (-\infty, 0)$. In this case, Model Equation 5.6 implies $\eta_{ij} = y_j - y_{j+1} \in (-\infty, \infty)$, $j = 1, \ldots, J-1$.

On the other hand, for any given $\eta_{i1}, \ldots, \eta_{i,J-1} \in (-\infty, \infty)$, $y_j = (\eta_{i,J-1} + \cdots + \eta_{ij}) + y_J$, $j = 1, \ldots, J-1$. Note that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{y_J}\left(e^{\eta_{i,J-1}+\cdots+\eta_{i1}} + e^{\eta_{i,J-1}+\cdots+\eta_{i2}} + \cdots + e^{\eta_{i,J-1}} + 1\right) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in Equation 5.7, and thus $\pi_{ij} \in (0, 1)$ exists and is unique, $j = 1, \ldots, J$. $\square$

5.4 Continuation-Ratio Logit model for Hierarchical Response

The continuation-ratio logit model for hierarchical responses (2; 3) takes the general form as follows:

$$log\left(\frac{\pi_{ij}}{\pi_{i,j+1}+\cdots+\pi_{iJ}}\right)=\mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j+\mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}\ ,\quad j=1,\ldots,J-1 \tag{5.8}$$

Proposition 5.4. *Fixing* $\mathbf{x}_i$*,* $\boldsymbol{\beta}_j, j=1,\cdots,J-1$ *and* $\boldsymbol{\zeta}$ *in Model Equation 5.8, let* $\eta_{ij}=\mathbf{h}_j^T(\mathbf{x}_i)\boldsymbol{\beta}_j+\mathbf{h}_c^T(\mathbf{x}_i)\boldsymbol{\zeta}$*,* $j=1,\ldots,J-1$*. Then* $0<\pi_{ij}<1, j=1,\ldots,J$ *exist uniquely if and only if* $-\infty<\eta_{ij}<\infty$*,* $j=1,\ldots,J-1$*. In this case,*

$$\pi_{ij}=\begin{cases} e^{\eta_{ij}}\prod_{s=1}^{j}\left(e^{\eta_{is}}+1\right)^{-1} & j=1,\ldots,J-1\\ \prod_{s=1}^{J-1}\left(e^{\eta_{is}}+1\right)^{-1} & j=J\end{cases} \tag{5.9}$$

Proof of Proposition 5.4: Let $y_j=\log\pi_{ij}$. Then $0<\pi_{ij}<1, j=1,\ldots,J$ if and only if $y_j\in(-\infty,0)$. In this case, Equation 5.8 implies $\eta_{ij}=y_j-\log(e^{y_{j+1}}+\cdots e^{y_J})\in(-\infty,\infty)$, $j=1,\ldots,J-1$.

On the other hand, for any given $\eta_{i1},\ldots,\eta_{i,J-1}\in(-\infty,\infty)$, it can be verified by induction that

$$\begin{aligned} e^{y_{J-1}} &= e^{y_J}e^{\eta_{i,J-1}}\\ e^{y_{J-2}} &= e^{y_J}e^{\eta_{i,J-2}}\left(e^{\eta_{i,J-1}}+1\right)\\ e^{y_j} &= e^{y_J}e^{\eta_{ij}}\left(e^{\eta_{i,j+1}}+1\right)\cdots\left(e^{\eta_{i,J-1}}+1\right),\ j=J-3,J-4,\cdots,1 \end{aligned}$$

Therefore, it can be verified that

$$\begin{aligned} 1 &= \pi_{i1} + \pi_{i2} + \cdots + \pi_{i,J-1} + \pi_{iJ} \\ &= e^{y_1} + e^{y_2} + \cdots + e^{y_{J-1}} + e^{y_J} \\ &= e^{y_J}\left(e^{\eta_{i1}} + 1\right)\left(e^{\eta_{i2}} + 1\right)\cdots\left(e^{\eta_{i,J-1}} + 1\right) \end{aligned}$$

Since $\pi_{ij} = e^{y_j}$, we get solutions of π_{ij} given in Equation 5.9, and thus $\pi_{ij} \in (0,1)$ exists and is unique, $j = 1, \ldots, J$. □

5.5 Design Space

In this section, we summarize our results for different multinomial logistic models and derive the corresponding design spaces. As a direction conclusion of Propositions 5.1, 5.2, 5.3, and 5.4, we have the theorem as follows:

Theorem 5.1. *For the baseline-category logit model Equation 5.1, the adjacent-categories logit model Equation 5.6, or the continuation-ratio logit model Equation 5.8, the design space*

$$\mathcal{X} = \{\mathbf{x} = (x_1, \ldots, x_d)^T \mid \eta_j \in (-\infty, \infty),\ j = 1, \ldots, J-1\};$$

for the cumulative logit model Equation 5.3, the design space

$$\mathcal{X} = \{\mathbf{x} = (x_1, \ldots, x_d)^T \mid -\infty < \eta_1 < \eta_2 < \cdots < \eta_{J-1} < \infty\},$$

where $\eta_j = \mathbf{h}_j^T(\mathbf{x})\boldsymbol{\beta}_j + \mathbf{h}_c^T(\mathbf{x})\boldsymbol{\zeta}$, $j = 1, \ldots, J-1$.

Based on the model Equation 2.1

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta} \triangleq (\eta_{i1}, \eta_{i2}, \cdots, \eta_{i,J-1}, 0)^T$$

We may have solutions for the following models respectively.

(1) Baseline-category logit model

$$\log(\boldsymbol{\pi}_i) = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \\ & & & & -1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

(2) Adjacent-categories logit model

$$\log(\boldsymbol{\pi}_i) = \begin{pmatrix} 1 & & & & -1 \\ & 1 & & & -1 \\ & & \ddots & & \vdots \\ & & & 1 & -1 \\ & & & & -1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & & & & 0 \\ & 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \exp\left(\begin{pmatrix} 1 & 1 & \cdots & 1 & 0 \\ & 1 & \cdots & 1 & 0 \\ & & \ddots & \vdots & \vdots \\ & & & 1 & 0 \\ & & & & 1 \end{pmatrix}_{J\times J} \cdot \boldsymbol{\eta}_i\right)\right)$$

(3) Continuation-ratio logit model

$$\log(\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i - \begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & & \ddots & & \\ 1 & 1 & \cdots & 1 & \\ 1 & 1 & \cdots & 1 & 1 \end{pmatrix}_{J\times J} \cdot \log\left(\begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ & & \ddots & & \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

(4) Cumulative logit model

$$\log\left(\begin{pmatrix} 1 & & & & \\ 1 & 1 & & & \\ \vdots & & \ddots & & \\ 1 & 1 & \cdots & 1 & \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{J\times J} \cdot \boldsymbol{\pi}_i\right) = \begin{pmatrix} 1 & & & -1 & & \\ & \ddots & & & \ddots & \\ & & 1 & & & -1 \\ 0 & \cdots & 0 & 0 & \cdots & -1 \end{pmatrix}_{J\times 2(J-1)}$$

$$\cdot \log\left(\begin{pmatrix} 1 & & & 0 \\ & \ddots & & \vdots \\ & & 1 & 0 \\ 1 & & & 1 \\ & 1 & & 1 \\ & \ddots & & \vdots \\ & & 1 & 1 \end{pmatrix}_{2(J-1)\times J} \cdot \exp(\boldsymbol{\eta}_i)\right)$$

Note that the above models could be with proportional, non-proportional, or partial proportional odds.

CHAPTER 6

D-OPTIMAL DESIGN

6.1 D-optimal Approximate Design

Given experimental settings $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \ldots, m$, we are looking for D-optimal approximate allocations $\mathbf{w} = (w_1, \ldots, w_m)^T$ that maximizes $|\mathbf{G}^T\mathbf{W}\mathbf{G}|$ defined in Equation 4.1. Note that $\mathbf{x}_i \in \mathcal{X}$ implies $0 < \pi_{ij} < 1$, $j = 1, \ldots, J$.

All feasible approximate allocations form into a bounded closed convex set $S = \{(w_1, \ldots, w_m)^T \in \mathbb{R}^m \mid w_i \geqslant 0, i = 1, \ldots, m; \sum_{i=1}^m w_i = 1\}$. The objective function is $f(\mathbf{w}) = |\mathbf{G}^T\mathbf{W}\mathbf{G}|$ which is an order-p homogeneous polynomial according to Theorem 4.2. Therefore, a D-optimal approximate design that maximizes $f(\mathbf{w})$ must exist. For typical applications, we need designs coming from $S_+ = \{\mathbf{w} \in S \mid f(\mathbf{w}) > 0\}$. Due to Theorem 2.1 and log-concavity of determinant on positive semi-definite matrices, we know $f(\mathbf{w})$ is log-concave (30; 16) and S_+ is convex. A useful result as a corollary of Theorem 3.4 is as follows:

Corollary 6.1. *S_+ is nonempty if and only if $f(\mathbf{w}_u) > 0$, where $\mathbf{w}_u = (1/m, \ldots, 1/m)^T$ is the uniform allocation. In this case, $f(\mathbf{w}) > 0$ for any $\mathbf{w} = (w_1, \ldots, w_m)^T$ such that $0 < w_i < 1, i = 1, \ldots, m$.*

Proof of Corollary 6.1: We only need to verify the "only if" part. According to Theorem 3.4, if $f(\mathbf{w}) > 0$ for some $\mathbf{w} = (w_1, \ldots, w_m)^T = (n_1, \ldots, n_m)^T/n$, then the corresponding $\mathbf{H}^*$ is of full row rank. Note that $\mathbf{H}^*$ can be obtained from $\mathbf{H}$ after removing the columns of $\mathbf{H}$

corresponding to $n_i = 0$. Thus $\mathbf{H}$ is of full row rank too, which corresponds to the uniform allocation. That is, $f(\mathbf{w}_u) > 0$. In this case, any $\mathbf{w} = (w_1, \ldots, w_m)^T$ such that $0 < w_i < 1, i = 1, \ldots, m$ leads to $f(\mathbf{w}) > 0$ since it corresponds to the same $\mathbf{H}$ matrix. $\square$

In order to avoid trivial cases, we assume $f(\mathbf{w}_u) > 0$ from now on.

Following (16) (2016, Section 3) (see also (21) and (31)), we define $f_i(z) = f(w_1(1-z)/(1-w_i), \ldots, w_{i-1}(1-z)/(1-w_i), z, w_{i+1}(1-z)/(1-w_i), \ldots, w_m(1-z)/(1-w_i))$ with $0 \leqslant z < 1$ given $\mathbf{w} = (w_1, \ldots, w_m)^T \in S_+$ and $i = 1, \ldots, m$. Parallel to Theorem 3.2.6 in (16), we obtain the result as follows according to Theorem 4.2:

Theorem 6.1. *Given an approximate allocation* $\boldsymbol{w} = (w_1, \ldots, w_m)^T \in S_+$ *and* $i \in \{1, \ldots, m\}$, *for* $0 \leq z \leqslant 1$,

$$f_i(z) = (1-z)^{p-J+1} \sum_{j=0}^{J-1} b_j z^j (1-z)^{J-1-j} \tag{6.1}$$

$$f_i'(z) = (1-z)^{p-J} \sum_{j=1}^{J-1} b_j (j - pz) z^{j-1} (1-z)^{J-j-1} - p b_0 (1-z)^{p-1} \tag{6.2}$$

where $b_0 = f_i(0)$, $(b_{J-1}, \ldots, b_1)^T = \mathbf{B}_{J-1}^{-1}\mathbf{c}$, $\mathbf{B}_{J-1} = (s^{t-1})_{st}$ *is a* $(J-1) \times (J-1)$ *matrix, and* $\mathbf{c} = (c_1, \ldots, c_{J-1})^T$ *with* $c_j = (j+1)^p j^{J-1-p} f_i(1/(j+1)) - j^{J-1} f_i(0)$, $j = 1, \ldots, J-1$.

Theorem 6.1 says that $f_i(z)$ is an order-p polynomial of z. Since $f_i(1) = 0$, the solution of maximization of $f_i(z), 0 \leq z \leq 1$ could only occur at $z = 0$ or $0 < z < 1$ such that $f_i'(z) = 0$, that is,

$$\sum_{j=1}^{J-1} j b_j z^{j-1} (1-z)^{J-j-1} = p \sum_{j=0}^{J-1} b_j z^j (1-z)^{J-j-1}, \quad 0 < z < 1 \tag{6.3}$$

which is an order-$(J-1)$ polynomial of z. For $J \leq 5$, Equation 6.3 yields analytic solutions. For $J \geqslant 6$, a quasi-Newton algorithm could be applied for searching numerical solutions.

In order to find D-optimal designs numerically, a lift-one algorithm, which is essentially the same as the one in (16) (2016, Section 3), is described as follows:

Lift-one algorithm for D-optimal allocation $\mathbf{w} = (w_1, \ldots, w_m)^T$*:*

1° Start with an arbitrary allocation $\mathbf{w}_0 = (w_1, \ldots, w_m)^T$ satisfying $0 < w_i < 1$, $i = 1, \ldots, m$ and compute $f(\mathbf{w}_0)$.

2° Set up a random order of i going through $\{1, 2, \ldots, m\}$.

3° For each i, determine $f_i(z)$ according to Theorem 6.1. In this step, J determinants $f_i(0), f_i(1/2)$, $f_i(1/3)$, $\ldots$, $f_i(1/J)$ are calculated.

4° Use quasi-Newton algorithm to find z_* maximizing $f_i(z)$ with $0 \leq z \leq 1$. If $f_i(z_*) \leqslant f_i(0)$, let $z_* = 0$. Define $\mathbf{w}_*^{(i)} = (w_1(1-z_*)/(1-w_i)$, $\ldots$, $w_{i-1}(1-z_*)/(1-w_i), z_*, w_{i+1}(1-z_*)/(1-w_i), \ldots, w_m(1-z_*)/(1-w_i))^T$. Note that $f(\mathbf{w}_*^{(i)}) = f_i(z_*)$.

5° Replace $\mathbf{w}_0$ with $\mathbf{w}_*^{(i)}$, and $f(\mathbf{w}_0)$ with $f(\mathbf{w}_*^{(i)})$.

6° Repeat 2° $\sim$ 5° until convergence, that is, $f(\mathbf{w}_0) = f(\mathbf{w}_*^{(i)})$ for each i.

Lift-one algorithm is essentially of general-equivalence-theorem type (16; 31). The convergence to a global maximum is guaranteed (21).

Theorem 6.2. *When the lift-one algorithm converges, the resulting* $\mathbf{w}$ *maximizes* $f(\mathbf{w})$*.*

6.2 D-optimal Exact Design

Given distinct experimental settings $\mathbf{x}_i$, $i = 1, \ldots, m$ and the total number n of experimental units, we are looking for an integer-valued allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ that maximizes $|\sum_{i=1}^m n_i \mathbf{F}_i|$ as defined in Equation 2.10, known as a *D-optimal exact allocation.*

For simplicity in notations, we also denote the objective function $f(\mathbf{n}) = f(n_1, \ldots, n_m) = |\sum_{i=1}^m n_i \mathbf{F}_i|$ as long as there is no ambiguity. Following (31) and (16), we define $f_{ij}(z) = f(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, n_i + n_j - z, n_{j+1}, \ldots, n_m)$ with $z = 0, 1, \ldots, n_i + n_j$ for $1 \leq i < j \leqslant m$ and given $\mathbf{n} = (n_1, \ldots, n_m)^T$.

Similar to Theorem 4.2.9 in (16), we obtain the following result from Theorem 4.2, Lemma 4.1 and Corollary 4.2:

Theorem 6.3. *Suppose* $\mathbf{n} = (n_1, \ldots, n_m)^T$ *satisfies* $f(\mathbf{n}) > 0$ *and* $n_i + n_j \geq q$ *for given* $1 \leq i < j \leqslant m$*, where* $q = \min\{2J - 2, p - k_{\min} + 2, p\}$*. Then for* $z = 0, 1, \ldots, n_i + n_j$ *,*

$$f_{ij}(z) = \sum_{s=0}^{q} g_s z^s \tag{6.4}$$

where $g_0 = f_{ij}(0)$*, and* $g_1, \ldots, g_q$ *can be obtained using* $(g_1, \ldots, g_q)^T = \mathbf{B}_q^{-1}(d_1, \ldots, d_q)^T$ *with* $\mathbf{B}_q = (s^{t-1})_{st}$ *as a* $q \times q$ *matrix and* $d_s = (f_{ij}(s) - f_{ij}(0))/s$*,* $s = 1, \ldots, q$*.*

Proof of Theorem 6.3: According to Theorem 4.2,

$$f_{ij}(z) = \sum_{\alpha_i \geq 0, \alpha_j \geq 0, \alpha_i + \alpha_j \leq p} coefficient \cdot z^{\alpha_i}(n_i + n_j - z)^{\alpha_j}$$

is a polynomial with nonnegative coefficients, whose order depends on the largest possible $\alpha_i + \alpha_j$. Lemma 4.1 implies that $\max\{\alpha_i, \alpha_j\} \leqslant J - 1$ for positive coefficients and Corollary 4.2 further implies that $\alpha_i + \alpha_j \leqslant p - (k_{\min} - 2) = p - k_{\min} + 2$ for positive coefficients. Therefore, $f_{ij}(z)$ is at most an order-q polynomial of z. $\square$

For a given number n of experimental units, we assume there exists an exact allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ such that $f(\mathbf{n}) > 0$ (otherwise, the maximization problem is trivial). In this case, if $n \geqslant m$, one may choose any $\mathbf{n}$ such that $n_i > 0$, $i = 1, \ldots, m$. Then the essentially same exchange algorithm in (16) (2016, S1.5) can be applied for this case:

Exchange algorithm for D-optimal allocation $(n_1, \ldots, n_m)^T$ given $n > 0$:

1° Start with an initial allocation $\mathbf{n} = (n_1, \ldots, n_m)^T$ such that $f(\mathbf{n}) > 0$.

2° Set up a random order of (i, j) going through all pairs $\{(1,2), (1,3), \ldots, (1,m), (2,3), \ldots, (m-1,m)\}$.

3° For each (i, j), let $c = n_i + n_j$. If $c = 0$, let $\mathbf{n}_{ij}^* = \mathbf{n}$. Otherwise, there are two cases. *Case one:* $0 < c \leq q$, we calculate $f_{ij}(z)$ for $z = 0, 1, \ldots, c$ directly and find z^* which maximizes $f_{ij}(z)$. *Case two:* $c > q$, we first calculate $f_{ij}(z)$ for $z = 0, 1, \ldots, q$; secondly determine $g_0, g_1, \ldots, g_q$ in Equation 6.4 according to Theorem 6.3; thirdly calculate $f_{ij}(z)$ for $z = q+1, \ldots, c$ based on Equation 6.4; fourthly find z^* maximizing $f_{ij}(z)$ for $z = 0, \ldots, c$. For both cases, we define

$$\mathbf{n}_{ij}^* = (n_1, \ldots, n_{i-1}, z^*, n_{i+1}, \ldots, n_{j-1}, c - z^*, n_{j+1}, \ldots, n_m)^T$$

Note that $f(\mathbf{n}_{ij}^*) = f_{ij}(z^*) \geqslant f(\mathbf{n}) > 0$. If $f(\mathbf{n}_{ij}^*) > f(\mathbf{n})$, replace $\mathbf{n}$ with $\mathbf{n}_{ij}^*$, and $f(\mathbf{n})$ with $f(\mathbf{n}_{ij}^*)$.

4° Repeat 2° $\sim$ 3° until convergence, that is, $f(\mathbf{n}_{ij}^*) = f(\mathbf{n})$ in step 3° for any (i, j).

6.3 Minimally Supported Design

One important question is, how many design points we need at least to keep $\mathbf{F}$ positive definite? In other words, what is the number of distinct experimental settings for a *minimally supported design*? According to Theorem 3.4, two conditions are needed: (1) $\pi_{ij} > 0$, $j = 1, \ldots, J$ for all design point $\mathbf{x}_i$, that is, $\mathbf{x}_i \in \mathcal{X}$, $i = 1, \ldots, m$ (see Section 5); (2) $\mathbf{H}$ is of full row rank p (we assume $n_i > 0$ for $i = 1, \ldots, m$ since we are considering minimal m). Theorem 3.8 provides the lower bound of the minimal number of experimental settings needed, which is also represented by $k_{\min}$ in Corollary 4.2. More precise answers would depend on specific forms of the predictor functions $\mathbf{h}_j$ (and $\mathbf{h}_c$ if applicable).

Another important question is whether a uniform allocation $\mathbf{w}_u = (1/m, \ldots, 1/m)^T$ is D-optimal given that m is the minimal number of experimental settings. The answer is yes for $J = 2$, but no for $J \geq 3$ with cumulative link models and proportional odds (16).

Theorem 6.4. *Consider Multinomial logit model* (Equation 2.1) *with only two response categories ($J = 2$). In this case, the minimum number of support points is $m = p$. The objective function $f(\mathbf{w}) \propto w_1 \cdots w_m$ and the D-optimal allocation for a minimally supported design is* $\mathbf{w} = (1/m, \ldots, 1/m)^T$.

Proof of Theorem 6.4: In this case, the model is essentially a generalized linear model for binomial response with logit link. Theorem 4.2 says that the objective function $f(\mathbf{w}) = |\mathbf{G}^T\mathbf{W}\mathbf{G}|$ is an order-p polynomial consisting of terms $c_{\alpha_1,\ldots,\alpha_m} w_1^{\alpha_1}\cdots w_m^{\alpha_m}$. According to Lemma 4.1, $c_{\alpha_1,\ldots,\alpha_m} \neq 0$ only if $\alpha_i \in \{0,1\}, i = 1,\ldots,m$. Therefore, in order to keep $f(\mathbf{w}) > 0$, we must have $m \geq p$. In other words, a minimally supported design may contain exactly $m = p$ distinct design points or experimental settings. In this case, the objective function $f(\mathbf{w}) \propto w_1 \cdots w_m$ and the D-optimal allocation is $\mathbf{w} = (1/m, \ldots, 1/m)^T$. □

For cases with three or more response categories ($J \geq 3$), in general we have $m \leq p$ for a minimally supported design. Whether the uniform allocation is D-optimal for a minimally supported design depends on the model setup. For example, for cumulative link models with proportional odds, (16) showed that the minimal number of experimental settings is $m = p_c + 1 < p_c + J - 1 = p$, which is also verified by Corollary 4.2, and a D-optimal allocation on a minimally supported design is not uniform in general. Similar results are expected for partial proportional odds models as a more general framework than proportional odds models.

However, for certain non-proportional models for multinomial responses with three or more categories, uniform allocations could be D-optimal for minimally supported designs due to the following result as a direct conclusion of Theorem 3.1, Theorem 3.3:

Corollary 6.2. *Consider Multinomial logit model Equation 2.1 with non-proportional odds. Suppose $p_1 = \cdots = p_{J-1}$ and there exist p_1 distinct experimental settings such that rank$(\mathbf{H}_1) = \cdots$ rank$(\mathbf{H}_{J-1}) = p_1$. Then the minimal number of experimental settings is $m = p_1$ and the uniform allocation is D-optimal for a minimally supported design.*

Corollary 6.2 shows that a uniform allocation is D-optimal for npo models under the condition $p_1 = \cdots = p_{J-1}$, which confirms the D-optimal design discovered for the trauma clinical trial (Example 7.2). Nevertheless, the following lemma and example represent that even for non-proportional odds, the uniform allocations again are not D-optimal for minimally supported designs if such a condition is violated.

Lemma 6.1. *Consider the maximization problem* $f(w_1, w_2, w_3) = w_1 w_2 w_3(c_1 w_2 w_3 + c_2 w_1 w_3 + c_3 w_1 w_2)$ *given* $0 < c_1 \leq c_2 \leq c_3$ *with respect to* $0 \leq w_i \leq 1$ *and* $w_1 + w_2 + w_3 = 1$. *Then*

(i) If $c_1 = c_2 = c_3$, *then the solution is* $w_1 = w_2 = w_3 = 1/3$.

(ii) If $c_1 = c_2 < c_3$, *then* $w_1 = w_2 > w_3 > 0$. *Actually,* $w_1 = w_2 = (-2c_1 + c_3 + \Delta_1)/D_1$ *and* $w_3 = c_3/D_1$, *where* $\Delta_1 = \sqrt{4c_1^2 - c_1 c_3 + c_3^2}$ *and* $D_1 = -4c_1 + 3c_3 + 2\Delta_1$.

(iii) If $c_1 < c_2 = c_3$, *then* $w_1 > w_2 = w_3 > 0$. *Actually,* $w_1 = (-c_1 + 2c_3 + \Delta_2)/D_2$ *and* $w_2 = w_3 = 3c_3/D_2$, *where* $\Delta_2 = \sqrt{c_1^2 - c_1 c_3 + 4c_3^2}$ *and* $D_2 = -c_1 + 8c_3 + \Delta_2$.

(iv) If $c_1 < c_2 < c_3$, *then* $w_1 > w_2 > w_3 > 0$. *The procedure of obtaining analytic solutions of* w_1, w_2, w_3 *is as follows: (1) obtain* y_1 *from* (Equation 6.11)*; (2) obtain* y_2 *from* (Equation 6.9)*; (3)* $w_1 = y_1/(y_1 + y_2 + 1)$, $w_2 = y_2/(y_1 + y_2 + 1)$, $w_3 = 1/(y_1 + y_2 + 1)$.

Proof of Lemma 6.1: First of all, we only need to consider the cases of $0 < w_i < 1$, $i = 1, 2, 3$ (otherwise, $f(w_1, w_2, w_3) = 0$). It can also been verified that $0 < c_1 \leq c_2 \leq c_3$ implies that $w_1 \geq w_2 \geq w_3 > 0$ (otherwise, for example, if $w_1 < w_2$, one may replace w_1, w_2 both with $(w_1 + w_2)/2$ and strictly increase f). The same argument implies that if $c_i = c_j$, then $w_i = w_j$ in the solution.

According to Theorem 5.10 in (16), $(w_1, w_2, w_3)^T$ maximizes $f(w_1, w_2, w_3)$ if and only if

$$\frac{\partial f}{\partial w_1} = \frac{\partial f}{\partial w_2} = \frac{\partial f}{\partial w_3}$$

which is equivalent to $\partial f/\partial w_1 = \partial f/\partial w_3$ and $\partial f/\partial w_2 = \partial f/\partial w_3$ and thus equivalent to

$$c_3 w_1 w_2 (w_1 - 2w_3) + 2c_2 w_1 w_3 (w_1 - w_3) = c_1 w_2 w_3 (-2w_1 + w_3) \tag{6.5}$$

$$c_3 w_1 w_2 (w_2 - 2w_3) + 2c_1 w_2 w_3 (w_2 - w_3) = c_2 w_1 w_3 (-2w_2 + w_3) \tag{6.6}$$

Following Yang et al. (16) (2016b, Section 5.2), we denote $y_1 = w_1/w_3 > 0$ and $y_2 = w_2/w_3 > 0$. Actually, $w_1 \geq w_2 \geq w_3 > 0$ implies $y_1 \geq y_2 \geq 1$. Since $w_1 + w_2 + w_3 = 1$, it implies $w_3 = 1/(y_1 + y_2 + 1)$, $w_1 = y_1/(y_1 + y_2 + 1)$, and $w_2 = y_2/(y_1 + y_2 + 1)$. Then (Equation 6.5) and (Equation 6.6) are equivalent to

$$c_3 y_1 y_2 (y_1 - 2) + 2c_2 y_1 (y_1 - 1) = c_1 y_2 (-2y_1 + 1) \tag{6.7}$$

$$c_3 y_1 y_2 (y_2 - 2) + 2c_1 y_2 (y_2 - 1) = c_2 y_1 (-2y_2 + 1) \tag{6.8}$$

From (Equation 6.7) we get $y_2[c_3 y_1^2 - 2(c_3 - c_1)y_1 - c_1] = 2c_2 y_1 (1 - y_1)$. If $y_1 = 1$, then we must have $y_2 = 1$ and $c_3 - 2(c_3 - c_1) - c_1 = 0$, which implies $w_1 = w_2 = w_3 = 1/3$ and $c_1 = c_2 = c_3$. Actually, we can also verify that $c_1 = c_3$ implies $y_1 = 1$.

Now we assume $y_1 > 1$, which implies $c_1 < c_3$. Then

$$y_2 = \frac{2c_2(1-y_1)y_1}{c_3y_1^2 - 2(c_3-c_1)y_1 - c_1} \tag{6.9}$$

After plugging (Equation 6.9) into (Equation 6.8), we get

$$a_0 + a_1y_1 + a_2y_1^2 + a_3y_1^3 + y_1^4 = 0 \tag{6.10}$$

where $a_0 = c_1^2/c_3^2 > 0$, $a_1 = 4c_1(-2c_1+c_2+2c_3)/(3c_3^2) > 0$, $a_2 = 2(2c_1^2 - 2c_1c_2 - 7c_1c_3 - 2c_2c_3 + 2c_3^2)/(3c_3^2)$, and $a_3 = 4(2c_1+c_2-2c_3)/(3c_3)$.

Denote $h(y_1) = a_0 + a_1y_1 + a_2y_1^2 + a_3y_1^3 + y_1^4$. Note that $h(\infty) = \infty$, $h(-c_1/c_3) = -c_1^2(c_1^2 + 8c_1c_2 - 2c_1c_3 + 8c_2c_3 + c_3^2)/(3c_3^4) < 0$, $h(0) = c_1^2/c_3^2 > 0$, $h(1) = -(c_1-c_3)^2/(3c_3^2) < 0$, and $h(\infty) = \infty$. Then $h(y_1) = 0$ yields four real roots in $(\infty, -c_1/c_3)$, $(-c_1/c_3, 0)$, $(0,1)$, and $(1,\infty)$, respectively. That is, there is one and only one $y_1 \in (1,\infty)$.

According to (32) (2014, equation (12)),

$$y_1 = -\frac{a_3}{4} + \frac{\sqrt{A_1}}{2} + \frac{\sqrt{C_1}}{2}\ , \tag{6.11}$$

where

$$
\begin{aligned}
A_1 &= -\frac{2a_2}{3} + \frac{a_3^2}{4} + \frac{G_1}{3 \times 2^{1/3}} \ , \\
C_1 &= -\frac{4a_2}{3} + \frac{a_3^2}{2} - \frac{G_1}{3 \times 2^{1/3}} + \frac{-8a_1 + 4a_2a_3 - a_3^3}{4\sqrt{A_1}} \ , \\
G_1 &= \left(F_1 - \sqrt{F_1^2 - 4E_1^3}\right)^{1/3} + \left(F_1 + \sqrt{F_1^2 - 4E_1^3}\right)^{1/3} \ , \\
E_1 &= 12a_0 + a_2^2 - 3a_1a_3 \ , \\
F_1 &= 27a_1^2 - 72a_0a_2 + 2a_2^3 - 9a_1a_2a_3 + 27a_0a_3^2 \ .
\end{aligned}
$$

The calculation of G_1, A_1, C_1, and y_1 are operations among complex numbers, while y_1 at the end would be a real number.

The procedure of obtaining analytic solutions of w_1, w_2, w_3 would be, (1) obtain y_1 from (Equation 6.11); (2) obtain y_2 from (Equation 6.9); (3) $w_1 = y_1/(y_1 + y_2 + 1)$, $w_2 = y_2/(y_1 + y_2 + 1)$, $w_3 = 1/(y_1 + y_2 + 1)$.

Now we discuss some special cases.

(i) If $c_1 = c_2 < c_3$, then $w_1 = w_2$ and thus $y_1 = y_2$. Both (Equation 6.7) and (Equation 6.8) yield $y_1 = c_3^{-1}(-2c_1 + c_3 + \sqrt{4c_1^2 - c_1c_3 + c_3^2})$, which implies

$$
w_1 = w_2 = \frac{-2c_1 + c_3 + \Delta_1}{-4c_1 + 3c_3 + 2\Delta_1}, \quad w_3 = \frac{c_3}{-4c_1 + 3c_3 + 2\Delta_1}
$$

where $\Delta_1 = \sqrt{4c_1^2 - c_1c_3 + c_3^2}$. Note that $w_1 > w_3$ since $\Delta_1 > 2c_1$.

(ii) If $c_1 < c_2 = c_3$, then $w_2 = w_3$ and thus $y_2 = 1$. From (Equation 6.7) we get $y_1 = 3c_3^{-1}(-c_1 + 2c_3 + \sqrt{c_1^2 - c_1c_3 + 4c_3^2})$, which implies

$$w_1 = \frac{-c_1 + 2c_3 + \Delta_2}{-c_1 + 8c_3 + \Delta_2}, \quad w_2 = w_3 = \frac{3c_3}{-c_1 + 8c_3 + \Delta_2}$$

where $\Delta_2 = \sqrt{c_1^2 - c_1c_3 + 4c_3^2}$. Note that $w_1 > w_2$ since $\Delta_2 > c_1 + c_3$.

(iii) If $c_1 < c_2 < c_3$, then y_1, y_2 and thus w_1, w_2, w_3 can be obtained analytically. We have proven $y_1 \geq y_2 \geq 1$. Using (Equation 6.7) and (Equation 6.8), it can be verified that $y_1 \neq y_2$ unless $c_1 = c_2$; and $y_2 \neq 1$ unless $c_2 = c_3$. That is, $y_1 > y_2 > 1$ and $w_1 > w_2 > w_3$.

□

6.4 EW D-optimal Design

The D-optimal approximate designs and exact designs are known as "locally" D-optimal in the literature since the values of parameters need to be assumed in advance. It is the case for typical nonlinear models, generalized linear models, and multivariate generalized linear models as well. Bayesian criterion has been applied to address this issue (33). It maximizes $E(\log |\mathbf{F}|)$ after assigning a prior distribution on the unknown parameters. A drawback of Bayesian approach is its computational intensity since its objective function deals with multiple integrals. An alternative solution is the EW D-optimality (16; 31; 18), which maximizes $\log |E(\mathbf{F})|$ or $|E(\mathbf{F})|$ instead. (31) shows that an EW D-optimal design could be highly efficient in terms of Bayesian criterion compared with the Bayesian D-optimal one, while the computational time complexity is essentially same as the locally D-optimal one. (16) also used EW-criterion for cumulative link models with proportional odds.

According to Theorem 2.1 and Corollary 3.1, the Fisher information matrix in our case is $\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i$, where $\mathbf{U}_i$ consists of $u_{st}(\boldsymbol{\pi}_i)$, the only component involving model parameters. Therefore, in order to calculate $E(\mathbf{F})$ with respect to a prior on parameters, we only need to calculate $E(u_{st}(\boldsymbol{\pi}_i))$, $s, t = 1, \ldots, J-1$, $i = 1, \ldots, m$. Then the results and algorithms developed in the previous sections of Chapter 6 can be used for EW D-optimal designs directly.

We provide formulas in Chapter 5 for calculating π_{ij}'s given $\mathbf{X}_i$'s and the parameter values. We also provide formulas for calculating $u_{st}(\boldsymbol{\pi}_i)$'s given π_{ij}'s in Section 4.3.2.

CHAPTER 7

APPLICATIONS

7.1 Experiment on the Emergence of House Flies

Example 7.1. *Toxicological experiments involving laboratory animals often yield multinomial count data. In an experiment on the emergence of house flies (34), seven sets of 500 pupae were exposed to one of several doses of radiation. Observations from each set of pupae after a period of time included the number of flies that died before the opening of the pupae (unopened pupae)* (y_1) *, the number of flies that died before complete emergence* (y_2) *, and the number of flies that completely emerged* (y_3) *from* 500 *pupae given a dose* (x) *of gamma radiation . Given* x*, the response* (y_1, y_2, y_3) *is a trinomial random variable. In this study, the responses have a clear nested or hierarchical structure. Typical data are given in Table II.*

TABLE II: An Experiment on the Emergence of House Flies

Dose of radiation (Gy) x	Response categories y_1	y_2	y_3	Total number of pupae
80	62	5	433	500
100	94	24	382	500
120	179	60	261	500
140	335	80	85	500
160	432	46	22	500
180	487	11	2	500
200	498	2	0	500

7.1.1 Locally optimal design

We tried different models and found that continuation-ratio logit models with NPO is the best model, which is much better than other models in terms of AIC and BIC, see Table III. Actually, Atkinson (2) example adopted the non-proportional odds model with continuation-ratio logit link.

TABLE III: Model Comparison for the Flies Study (PO and NPO)

	Cumulative PO	Cumulative NPO	Continuation PO	Continuation NPO	Adjacent PO	Adjacent NPO
AIC	195.87	121.17	116.40	114.42	209.64	194.47
BIC	195.71	120.96	116.24	114.20	209.47	194.25

Following Atkinson (2), the non-proportional odds model with continuation-ratio logit link could be expressed as

$$\log\left(\frac{\pi_{i1}}{\pi_{i2}+\pi_{i3}}\right) = \beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2$$

$$\log\left(\frac{\pi_{i2}}{\pi_{i3}}\right) = \beta_{21} + \beta_{22}x_i$$

This model is equivalent to the following unified form expressions

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta}, \qquad i = 1, \cdots, 7$$

where

$$\mathbf{C}^T = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_i & x_i^2 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_i \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \end{pmatrix}$$

Fitting the model, we get the parameter estimates as follows

$$\hat{\boldsymbol{\theta}} = (-1.935, -0.02642, 0.0003174, -9.159, 0.06386)^T \tag{7.1}$$

This is parameter input for locally optimal design.

Solving the model equations, we get the probabilities as follows

$$\begin{aligned} \pi_{i1} &= \frac{exp(\beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2)}{1 + exp(\beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2)} \\ \pi_{i2} &= \frac{exp(\beta_{21} + \beta_{22}x_i)}{(1 + exp(\beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2))(1 + exp(\beta_{21} + \beta_{22}x_i))} \\ \pi_{i3} &= \frac{1}{(1 + exp(\beta_{11} + \beta_{12}x_i + \beta_{13}x_i^2))(1 + exp(\beta_{21} + \beta_{22}x_i))} \end{aligned}$$

Combining the above expressions with Corollary 3.1, we get

$$\mathbf{U}_i = \begin{pmatrix} \frac{exp(\beta_{11}+\beta_{12}x_i+\beta_{13}x_i^2)}{(1+exp(\beta_{11}+\beta_{12}x_i+\beta_{13}x_i^2))^2} & 0 & 0 \\ 0 & \frac{exp(\beta_{21}+\beta_{22}x_i)}{(1+exp(\beta_{11}+\beta_{12}x_i+\beta_{13}x_i^2))(1+exp(\beta_{21}+\beta_{22}x_i))^2} & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $\mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i$.

Since

$$\mathbf{F} = n \sum_{i=1}^{m} w_i \mathbf{F}_i$$

Given all of the above inputs, $|\mathbf{F}|$ is a polynomial of w_i. We can use Lift-one algorithm to maximize $|\mathbf{F}|$ and find the D-optimal approximate design.

The D-optimal approximate design assigns all of experimental units to the following 4 design points: dose 80, 120, 140 and 160. The D-efficiency of the original uniform allocation is $(585317/1480378)^{1/5} = 83.1\%$, see Table IV. The efficiency could be improved by $100/83.1 - 1 = 20.3\%$ if the D-optimal design is adopted.

TABLE IV: The D-optimal approximate design for the flies Data

Design Point	1	2	3	4	5	6	7	
Doses of Radiation	80	100	120	140	160	180	200	Determinant
Original sample allocation	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	585317
Optimal sample allocation	0.3116	0	0.2917	0.1071	0.2896	0	0	1480378

If we use grid search (grid=5) to find an optimal design with a finer space, then we get the following optimal design with design space $\{80, 85, 90, \cdots, 195, 200\}$.

TABLE V: The Continuous D-optimal Design for the Flies Study (Grid=5)

Doses of Radiation	80	120	125	155	160	Determinant
Optimal Sample Allocation	0.3163	0.1429	0.2003	0.1683	0.1723	1497192

The number of support points is increased to five. The efficiency of the original uniform allocation is $(585317/1497192)^{1/5} = 82.9\%$, the efficiency could be improved $100/82.9 - 1 = 20.6\%$ by the optimal design.

If we adopt an even finer design space with grid 1, then we get the following optimal design with design space $\{80, 81, 82, \cdots, 199, 200\}$.

TABLE VI: The Continuous D-optimal Design for the Flies Study (Grid=1)

Doses of Radiation	80	122	123	157	158	Determinant
Optimal Sample Allocation	0.3163	0.0786	0.2636	0.2206	0.1209	1504236

The efficiency of the original uniform allocation is $(585317/1504236)^{1/5} = 82.8\%$, the efficiency could be improved $100/82.8 - 1 = 20.8\%$ by the optimal design. Here exists two facts: (1), The optimal design's efficiency increases, but becomes slowly when the grid search is finer. (2), The number of optimal design points is still five, but with two clusters of points, each cluster consists of two points which become closer and closer to each other. It seems that the optimal design with a continuous space $[80, 200]$ would consist of three points only, which is minimally supported. We consider the following minimally supported design

TABLE VII: The Minimally Supported D-optimal Design for the Flies Study

Doses of Radiation	80	123	157	Determinant
Optimal Sample Allocation	0.3163	0.3422	0.3415	1503272

In Table VII, we accumulate the weights on point(dose) 122 and 123 to point 123, similarly accumulate the weights on point(dose) 157 and 158 to point 157. The determinant is very close to optimal design's determinant. This design's efficiency is $(1503272/1504236)^{1/5} = 99.99\%$, the minimally supported design is highly efficient.

Here the minimally supported design is not uniform. Actually we can find the analytic form of Fisher information matrix's determinant. Consider the above non-proportional odds model, where $\mathbf{h}_1(x_i) = (1, x_i, x_i^2)^T$, $\mathbf{h}_2(x_i) = (1, x_i)^T$, $J = 3$, $p_1 = 3, p_2 = 2$, and $p = 5$. According to Corollary 4.2, the minimum number of support points is $m = \max\{p_1, p_2\} = 3$, which is feasible. The objective function $f(\mathbf{w})$ is an order-5 polynomial with items $c_{\alpha_1,\alpha_2,\alpha_3} w_1^{\alpha_1} w_2^{\alpha_2} w_3^{\alpha_3}$. Lemma 4.1 implies that $\alpha_i \in \{0, 1, 2\}, i = 1, 2, 3$ in order to keep $c_{\alpha_1,\alpha_2,\alpha_3} \neq 0$. Combined with Corollary 4.2, we further know $\alpha_i \in \{1, 2\}, i = 1, 2, 3$. Therefore, the objective function

$$f(w_1, w_2, w_3) = w_1 w_2 w_3 (c_{122} w_2 w_3 + c_{212} w_1 w_3 + c_{221} w_1 w_2)$$

where

$$c_{122} = [x_2x_3(x_2 - x_3) - x_1x_3(x_1 - x_3) + x_1x_2(x_1 - x_2)]^2 (x_2 - x_3)^2 \pi_{11}(\pi_{12} + \pi_{13})\pi_{21}\pi_{22}\pi_{23}\pi_{31}\pi_{32}\pi_{33}$$

$$c_{212} = [x_2x_3(x_2 - x_3) - x_1x_3(x_1 - x_3) + x_1x_2(x_1 - x_2)]^2 (x_1 - x_3)^2 \pi_{11}\pi_{12}\pi_{13}\pi_{21}(\pi_{22} + \pi_{23})\pi_{31}\pi_{32}\pi_{33}$$

$$c_{221} = [x_2x_3(x_2 - x_3) - x_1x_3(x_1 - x_3) + x_1x_2(x_1 - x_2)]^2 (x_1 - x_2)^2 \pi_{11}\pi_{12}\pi_{13}\pi_{21}\pi_{22}\pi_{23}\pi_{31}(\pi_{32} + \pi_{33})$$

which in general does not yield a maximal occurring at $w_1 = w_2 = w_3 = 1/3$.

About D-optimal exact design, we have $\mathbf{F} = \sum_{i=1}^{m} n_i \mathbf{F}_i$, which is a polynomial about n_i, then we can use exchange algorithm to maximize $|\mathbf{F}|$ to find D-optimal exact design. The D-optimal exact designs for this model are very similar to D-optimal approximate designs due to large sample size. So we skip this part results. But D-optimal exact designs are not always equivalent to D-optimal approximate designs, especially for small sample size.

7.1.2 EW D-optimal design

Because model parameters' distribution is hard to specify, we use empirical method to obtain it. We bootstrap original sample to generate 1000 simulated samples, fit each simulated sample to get one parameter vector. The total 1000 parameter vectors can serve as parameters' empirical distribution. Then we can obtain the EW optimal design. The results are in Table VIII.

TABLE VIII: The EW D-optimal Design for the Flies Study (Bootstrap Samples)

Design Point	1	2	3	4	5	6	7	
Doses of Radiation	80	100	120	140	160	180	200	Determinant
Uniform Design	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	585185
EW Design	0.3120	0	0.2911	0.1087	0.2882	0	0	1467951

We also tried a modified EW design, that is, we take median values, entry by entry, of Fisher information matrix instead of expected **F**. We call it: MW design, indicating median of **F**. The MW optimal design's results are in Table IX.

TABLE IX: The MW D-optimal Design for the Flies Study (Bootstrap Samples)

Design Point	1	2	3	4	5	6	7	
Doses of Radiation	80	100	120	140	160	180	200	Determinant
Uniform Design	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	581275
MW Design	0.3117	0	0.2917	0.1077	0.2889	0	0	1471161

Based on the same set of bootstrap samples, we can also find Bayes optimal design, which maximizes $\phi(\mathbf{w})$.

$$\phi(\mathbf{w}) = E\hat{log}|\mathbf{F}| = \frac{1}{1000}\sum_{j=1}^{1000} log|F(\boldsymbol{\theta}_j^{(b)})| = \frac{1}{1000}\sum_{j=1}^{1000} log|\sum_{i=1}^{m} w_i|F_i(\boldsymbol{\theta}_j^{(b)})|$$

where $(\boldsymbol{\theta}_1^{(b)}, \cdots, \boldsymbol{\theta}_{1000}^{(b)})$ are the bootstrapped parameter vectors. The results are in Table X.

TABLE X: The Bayes D-optimal Design for the Flies Study (Bootstrap Samples)

Design Point	1	2	3	4	5	6	7	
Doses of Radiation	80	100	120	140	160	180	200	Determinant
Uniform Design	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	0.1429	572369
Bayesian Design	0.3159	4.4×10^{-7}	0.2692	0.1160	0.2990	1.1×10^{-6}	7.1×10^{-10}	1438052

In terms of Bayesian Criterion, the uniform design $\mathbf{w}_u$'s efficiency is $exp[(\phi(\mathbf{w}_u) - \phi(\mathbf{w}_b)/5] \times 100\% = 83.2\%$, the EW design's efficiency is $exp[(\phi(\mathbf{w}_e) - \phi(\mathbf{w}_b)/5] \times 100\% = 100.1\%$, the MW design's efficiency is $exp[(\phi(\mathbf{w}_m) - \phi(\mathbf{w}_b)/5] \times 100\% = 100.1\%$.

From the fitted model, the parameters' generalized variances are

$$\Sigma(\boldsymbol{\theta}) = \begin{pmatrix} 7.845\times 10^{-1} & -1.265\times 10^{-2} & 4.889\times 10^{-5} & 0 & 0 \\ -1.265\times 10^{-2} & 2.080\times 10^{-4} & -8.182\times 10^{-7} & 0 & 0 \\ 4.889\times 10^{-5} & -8.182\times 10^{-7} & 3.271\times 10^{-9} & 0 & 0 \\ 0 & 0 & 0 & 2.734\times 10^{-1} & -2.128\times 10^{-3} \\ 0 & 0 & 0 & -2.128\times 10^{-3} & 1.703\times 10^{-5} \end{pmatrix} \quad (7.2)$$

the correlation matrix is

$$\rho(\boldsymbol{\theta}) = \begin{pmatrix} 1 & -0.990 & 0.965 & 0 & 0 \\ -0.990 & 1 & -0.992 & 0 & 0 \\ 0.965 & -0.992 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -0.986 \\ 0 & 0 & 0 & -0.986 & 1 \end{pmatrix}$$

Since the model parameters are highly correlated, we assume the model parameters follow multivariate normal distributions with mean expressed in Equation 7.1 and variance co-variance structure as Equation 7.2. The corresponding EW D-optimal design results are very similar to bootstrap's results in terms of design points and weights, we skip the results.

In order to check robustness towards misspecified parameter values, we use 1000 bootstrap samples. For each $\boldsymbol{\theta}$, we use the lift-one algorithm to find the D-optimal allocation $\mathbf{w}_{\boldsymbol{\theta}}$ and the corresponding determinant $f(\mathbf{w}_{\boldsymbol{\theta}}) = |F(\mathbf{w}_{\boldsymbol{\theta}})|$, and then calculate the relative efficiency $D(\mathbf{w}, \boldsymbol{\theta}) = (f(\mathbf{w})/f(\mathbf{w}_{\boldsymbol{\theta}}))^{1/5}$ for $\mathbf{w} = \mathbf{w}_b$ (Bayes optimal design), $\mathbf{w}_e$ (EW optimal design), $\mathbf{w}_m$ (MW optimal design) and $\mathbf{w}_u$ (uniform optimal design), respectively. Table XI shows the summary statistics of the relative efficiencies for these optimal designs. It implies that $\mathbf{w}_b$, $\mathbf{w}_e$ and $\mathbf{w}_m$, are comparable and all of them are much better than $\mathbf{w}_u$ in terms of robustness.

TABLE XI: Summary Statistics of Relative Efficiencies in Flies Study

Design	Min	1st Quartile	Median	3rd Quartile	Max
Bayes $\mathbf{w}_b$	0.9912	0.9985	0.9989	0.9992	0.9998
EW $\mathbf{w}_e$	0.9934	0.9991	0.9996	0.9998	0.999999
MW $\mathbf{w}_m$	0.9933	0.9991	0.9996	0.9998	0.999998
Uniform $\mathbf{w}_u$	0.7762	0.8181	0.8304	0.8445	0.8861

7.2 Trauma Clinical Trial

Example 7.2. *Chuang (14) studied a group of data which have five ordered response categories ranging from "death" to "good recovery" describing the clinical outcome of trauma patients. These five categories are often called the Glasgow Outcome Scale (GOS, (35)) in the literature on critical care. There were four treatment groups labeled as Placebo, Low dose, Medium dose and High dose reported in the data. Note that the original data didn't provide the exact dosage but treat them as 1,2,3,4 instead. Table XII is regenerated from (14).*

7.2.1 Locally optimal design

In order to model ordered categorical responses, a popular choice seems to be the proportional odds model with cumulative logit link. However, Agresti (22) found some strong evidence against the assumption of proportional odds. We fit the data with non-proportional

TABLE XII: Glasgow Outcome Scales from the Trauma Clinical Trial

Treatment group	Glasgow Outcome Scale				
	Death	Vegetative state	Major disability	Minor disability	Good recovery
Placebo	59	25	46	48	32
Low	48	21	44	47	30
Medium	44	14	54	64	31
High	43	4	49	58	41

odds model with cumulative logit link and find that it is better than proportional odds model in terms of AIC and BIC (see Table XIII). We also compared it with other commonly used models including continuation-ration (for hierarchical response) and adjacent-categories logit models, the non-proportional odds model with cumulative logit link is still best. It shows us that under some circumstances, a non-proportional odds model could be better than a proportional odds model with cumulative logit link.

TABLE XIII: Model Comparison for the Trauma Data

	Cumulative PO	Cumulative NPO	Continuation PO	Continuation NPO	Adjacent PO	Adjacent NPO
AIC	107.75	99.41	108.98	101.36	107.67	101.54
BIC	104.68	94.51	105.91	96.45	104.60	96.63

From this data, the non-proportional odds model with cumulative logit link could be expressed as

$$\log\left(\frac{\pi_{i1} + \cdots + \pi_{ij}}{\pi_{i,j+1} + \cdots + \pi_{i5}}\right) = \beta_{j1} + \beta_{j2} x_i \ , \quad j = 1, 2, 3, 4$$

This model is equivalent to the following unified form expressions

$$\mathbf{C}^T \log(\mathbf{L}\boldsymbol{\pi}_i) = \boldsymbol{\eta}_i = \mathbf{X}_i\boldsymbol{\theta}, \qquad i = 1, \cdots, 4$$

where

$$\mathbf{C}^T = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_i & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_i & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_i \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \boldsymbol{\theta} = \begin{pmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \\ \beta_{31} \\ \beta_{32} \\ \beta_{41} \\ \beta_{42} \end{pmatrix}$$

Fit the model to get parameter estimators

$$\hat{\boldsymbol{\theta}} = (-0.86459, -0.11291, -0.09374, -0.26890, 0.70625, -0.18234, 1.90867, -0.11926)^T$$

This is parameter input for local optimal design.

From this model, we can solve the equation to get the probability as follows

$$\begin{aligned} \pi_{i1} &= \frac{exp(\beta_{11} + \beta_{12}x_i)}{1 + exp(\beta_{11} + \beta_{12}x_i)} \\ \pi_{ij} &= \frac{exp(\beta_{j1} + \beta_{j2}x_i)}{1 + exp(\beta_{j1} + \beta_{j2}x_i)} - \frac{exp(\beta_{j-1,1} + \beta_{j-1,2}x_i)}{1 + exp(\beta_{j-1,1} + \beta_{j-1,2}x_i)}, \qquad j = 2, 3, 4 \\ \pi_{i5} &= \frac{1}{1 + exp(\beta_{41} + \beta_{42}x_i)} \end{aligned}$$

Combine the above expression with Corollary 3.1, we can express each element of $\mathbf{U}_i$ as function of parameters and x_i.

Since

$$\mathbf{F} = n\sum_{i=1}^{m} w_i \mathbf{F}_i \qquad and \qquad \mathbf{F}_i = \mathbf{X}_i^T \mathbf{U}_i \mathbf{X}_i$$

Given all of the above inputs, $|\mathbf{F}|$ is a polynomial about w_i, then we can use Lift-one algorithm to maximize $|\mathbf{F}|$ to find D-optimal approximate design.

The D-optimal approximate design for this model is summarized as the following table, here design point corresponds to a different treatment group:

TABLE XIV: The D-optimal Approximate Design for the Trauma Data

Design Point Treatment Group	1 Placebo	2 Low	3 Medium	4 High	 Determinant
Original Sample Allocation	0.2618	0.2369	0.2581	0.2431	0.0002832108
Optimal Sample Allocation	0.5	0.0	0.0	0.5	0.002989844

Here the D-optimal exact design assigns all of experimental units to the following two design points: Placebo and High treatment. Actually, the minimally supported points are 2 according to our previous derivation: $max(p_1, \ldots, p_{J-1}, p_H + p_c)$. Since this is cumulative logit model with NPO assumption, $p_c = 0$ and $p_1 = \ldots = p_{J-1} = p_H = 2$. The optimal design found here is actually a uniform design, confirmed the conclusion of Theorem 4.6.

The efficiency of the original sample allocation is $(0.0002832108/0.002989844)^{1/8} = 74.5\%$, the efficiency could be improved $100/74.5 - 1 = 34.2\%$ by optimal design.

The D-optimal exact design has very similar results with D-optimal approximate design, so we skip its results.

7.2.2 EW D-optimal design

In trauma study, we still use bootstrap method to generate 1000 simulated samples, then obtain Bayesian, EW and MW optimal designs based on these samples. The results are in Table XV.

TABLE XV: The Bayesian, EW and MW D-optimal Designs for the Trauma Study

Design Point	1	2	3	4	
Treatment group	Placebo	Low	Medium	High	Determinant
Bayes Design	0.4997	1.0×10^{-8}	2.5×10^{-8}	0.5003	2.773737×10^{-8}
EW Design	0.5	0.0	0.0	0.5	10575
MW Design	0.5	0.0	0.0	0.5	0.007652482

Here it is very interesting that all of non-local D-optimal designs give the same sample allocation as local D-optimal designs. If we examined Theorem 4.6, it states that if it is a minimally supported design for NPO models, and $p_1 = p_2 = \cdots = p_{J-1}$, then uniform design is D-optimal design. In our case, no matter if it is a local or non-local D-optimal design, all of the above conditions are satisfied, so the results are the same, which also verified Theorem 4.6.

We also obtained EW designs assuming uniform and normal distributions for parameters, the results are the same as above, so we skip the results.

7.3 Conclusion

In this chapter, we use two real experiments to illustrate how much improvement can be made by applying D-optimal designs. Some designs probably are not applicable in real life, e.g., the trauma study optimal design just assigns two design points, but it provides us a benchmark and criterion. So we know how to make our design more efficient aiming at D-optimal design.

If model parameters are assumed to follow some unknown distributions, we bootstrap our original samples to get simulated samples. Then we can obtain Bayes, EW and MW optimal designs based on those simulated samples. These designs are robust and similar to local optimal design.

Our D-optimal designs recommended in the examples are minimally supported, actually the minimum number of experimental settings are strictly less than the number of parameters.

While in the real life, uniform design is most commonly adopted. But generally the uniform design is not D-optimal design for multinomial logistic models. Only NPO models having minimally supported points and same number of parameters for each logit equation, i.e., $m = p_1 = p_2 = \cdots = p_{J-1}$, will have uniform design as D-optimal design.

APPENDIX

NOTATIONS

b_j	Coefficients in representing $f_i(z)$, $j = 0, \ldots, J-1$
$\mathbf{B}_J$	$J \times J$ constant matrix used for deriving the coefficients of $f_i(z)$, $(s^{t-1})_{st}$
$\mathbf{C}$	$J \times (2J-1)$ constant matrix, same for all the four logit models
$\mathbf{c}$	Vector used for deriving coefficients of $f_i(z)$, $(c_1, \ldots, c_{J-1})^T$
$\mathbf{c}_{ij}$	$J \times 1$ vectors such that $(\mathbf{C}^T\mathbf{D}_i^{-1}\mathbf{L})^{-1} = (\mathbf{c}_{i1}, \ldots, \mathbf{c}_{iJ})$
c_j	$(j+1)^p j^{J-1-p} f_i(1/(j+1)) - j^{J-1} f_i(0)$, $j = 1, \ldots, J-1$
$c_{\alpha_1,\ldots,\alpha_m}$	Coefficient of $w_1^{\alpha_1} \cdots w_m^{\alpha_m}$ in the determinant of $\mathbf{G}^T\mathbf{W}\mathbf{G}$
d	Total number of design factors
d_s	$d_s = (f_{ij}(s) - f_{ij}(0))/s$, $s = 1, \ldots, q$, for coefficients in f_{ij}
$\mathbf{D}_i$	$diag(\mathbf{L}\boldsymbol{\pi}_i)$
$\mathbf{F}$	Fisher information matrix of the design, $\mathbf{F} = \sum_{i=1}^m n_i \mathbf{F}_i$
f	$f(\mathbf{w}) = f(w_1, \ldots, w_m) = \lvert\mathbf{G}^T\mathbf{W}\mathbf{G}\rvert$ which is proportional to $\lvert\mathbf{F}\rvert$; or $f(\mathbf{n}) = f(n_1, \ldots, n_m) = \lvert\sum_{i=1}^m n_i \mathbf{F}_i\rvert = \lvert\mathbf{F}\rvert$
$\mathbf{F}_i$	Fisher information matrix at the ith design point
f_i	$f_i(z) = f(w_1(1-z)/(1-w_i), \ldots, w_{i-1}(1-z)/(1-w_i), z, w_{i+1}(1-z)/(1-w_i), \ldots, w_m(1-z)/(1-w_i))$ with $0 \le z < 1$

APPENDIX (Continued)

f_{ij}	$f_{ij}(z) = f(n_1, \ldots, n_{i-1}, z, n_{i+1}, \ldots, n_{j-1}, n_i + n_j - z, n_{j+1}, \ldots, n_m)$ with $z = 0, 1, \ldots, n_i + n_j$
$\mathbf{G}$	Matrix component for Fisher information matrix such that $\mathbf{F} = n\mathbf{G}^T\mathbf{W}\mathbf{G}$, $mJ \times p$
g_s	$g_0 = f_{ij}(0)$ and $(g_1, \ldots, g_q)^T = \mathbf{B}_q^{-1}(d_1, \ldots, d_q)^T$
$\mathbf{H}$	Matrix component for Fisher information matrix such that $\mathbf{F} = \mathbf{H}\mathbf{U}\mathbf{H}^T$, consisting of $\mathbf{H}_1, \ldots, \mathbf{H}_{J-1}$ and possibly $\mathbf{H}_c$, $p \times m(J-1)$
$\mathbf{H}_c$	Matrix for the common component of $J-1$ equations, $(\mathbf{h}_c(\mathbf{x}_1), \ldots, \mathbf{h}_c(\mathbf{x}_m))$, $p_c \times m$
$\mathbf{h}_c(\mathbf{x}_i)$	Vector of common predictors for all of $J-1$ equations as known functions of the ith experimental setting, $(h_1(\mathbf{x}_i), \ldots, h_{p_c}(\mathbf{x}_i))^T$
$\mathbf{H}_j$	Matrix for the jth $J-1$ equation only, $(\mathbf{h}_j(\mathbf{x}_1), \ldots, \mathbf{h}_j(\mathbf{x}_m))$, $p_j \times m$
$\mathbf{h}_j(\mathbf{x}_i)$	Vector of predictors for the jth $J-1$ equation as known functions of the ith experimental setting, $(h_{j1}(\mathbf{x}_i), \ldots, h_{jp_j}(\mathbf{x}_i))^T$
J	Total number of response categories
$k_{\min}$	Smallest possible $\#\{i \mid \alpha_i > 0\}$ such that $c_{\alpha_1,\ldots,\alpha_m} > 0$
$\mathbf{L}$	Constant $(2J-1) \times J$ matrix, different for the four logit models
m	Total number of distinct experimental settings or design points
$\mathcal{M}(\mathbf{H})$	Column space of matrix $\mathbf{H}$, that is, the linear subspace spanned by the columns of $\mathbf{H}$

APPENDIX (Continued)

n	Total number of experimental units, $n = n_1 + \cdots + n_m$
$\mathbf{n}$	Allocation of experimental units, $(n_1, \ldots, n_m)^T$, $n_i \geq 0$, $\sum_i n_i = n$
n_i	Number of replicates at the ith experimental setting
p	Total number of parameters
p_c	Number of common parameters for all of $J-1$ equations of logit model
p_H	dimension of row space for $J-1$ intersection of H_js
p_j	Number of parameters for the jth equation only
q	$\min\{2J-2, p-k_{\min}+2, p\}$, upper bound of order of $f_{ij}(z)$
S	Collection of all feasible approximate allocations, $\{(w_1, \ldots, w_m)^T \in \mathbb{R}^m \mid w_i \geq 0, i = 1, \ldots, m; \sum_{i=1}^m w_1 = 1\}$
S_+	Collection of approximate allocations, $\{\mathbf{w} \in S \mid f(\mathbf{w}) > 0\}$
$\mathbf{U}$	Block matrix $(\mathbf{U}_{st})_{s,t=1,\ldots,J-1}$, $m(J-1) \times m(J-1)$
$\mathbf{U}_{st}$	$\mathrm{d}iag\{n_1 u_{st}(\boldsymbol{\pi}_1), \ldots, n_m u_{st}(\boldsymbol{\pi}_m)\}$, $m \times m$
$u_{st}(\boldsymbol{\pi}_i)$	$\mathbf{c}_{is}^T \mathrm{d}iag(\boldsymbol{\pi}_i)^{-1} \mathbf{c}_{it}$ for $s, t = 1, \ldots, J-1$
$\mathbf{w}$	Real-valued allocation of experimental units, $(w_1, \ldots, w_m)^T$, $w_i \geq 0$, $\sum_i w_i = 1$
$\mathbf{W}$	$\mathrm{d}iag\{w_1 \mathrm{d}iag(\boldsymbol{\pi}_1)^{-1}, \ldots, w_m \mathrm{d}iag(\boldsymbol{\pi}_m)^{-1}\}$, $mJ \times mJ$
w_i	Proportion of experimental units assigned to the ith experimental setting, n_i/n

APPENDIX (Continued)

$\mathbf{w}_u$	Uniform allocation, $(1/m, \ldots, 1/m)^T$
$\mathcal{X}$	Design space, the collection of all design points yielding strictly positive categorical probabilities of response
$\mathbf{x}_i$	The ith distinct experimental setting or design point, $(x_{i1}, \ldots, x_{id})^T$
$\mathbf{X}_i$	Model matrix at the ith design point, $J \times p$, the last row is all 0's
$\boldsymbol{\beta}_j$	Vector of parameters for the jth equation only, $(\beta_{j1}, \ldots, \beta_{jp_j})^T$
γ_{ij}	The cumulative probability from the 1th to jth catogory at the ith experimental setting, $\gamma_{ij} = \pi_{i1} + \cdots + \pi_{ij}$.
$\boldsymbol{\zeta}$	Vector of common parameters for all of the $J-1$ equations, $(\zeta_1, \ldots, \zeta_{p_c})^T$
$\boldsymbol{\eta}_i$	Vector of linear predictors at the ith experimental setting, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{iJ})^T = \mathbf{X}_i\boldsymbol{\theta}$ with $\eta_{iJ} \equiv 0$
$\boldsymbol{\theta}$	Vector of all parameters, $p \times 1$
$\boldsymbol{\pi}_i$	Vector of response category probabilities at the ith experimental setting. $\boldsymbol{\pi}_i = (\pi_{i1}, \ldots, \pi_{iJ})^T$, $\pi_{i1} + \cdots + \pi_{iJ} = 1$
π_{ij}	Probability that the response falls into the jth catogory at the ith experimental setting

CITED LITERATURE

1. Glonek, G. and McCullagh, P.: Multivariate logistic models. Journal of the Royal Statistical Society, Series B, 57(3):533–546, January 1995.

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

3. Agresti, A.: Categorical Data Analysis. Wiley Series in Probability and Statistics. Hoboken, New Jersey, John Wiley & Sons, 2013. Third Edition.

4. McCullagh, P.: Regression models for ordinal data. Journal of the Royal Statistical Society, Series B, 42(2):109–142, January 1980.

5. Christensen, R. H. B.: Analysis of ordinal data with cumulative link models – estimation with the R-package ordinal. Department of Applied Mathematics and Computer Science, Technical University of Denmark, Matematiktorvet, 2800, Kgs. Lyngby, Denmark, 2015.

6. 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.

7. Catherine J. Price, Carole A. Kimmel, J. D. G. and Marr, M. C.: The developmental toxicity of diethylene glycol dimethyl ether in mice. Fundamental and Applied Toxicology, 8(1):115–126, January 1987.

8. Tutz, G.: Sequential models in categorical regression. Computational Statistics & Data Analysis, 11(3):275–295, 1991.

9. Peterson, B. and Harrell, F.: Partial proportional odds models for ordinal response variables. Journal of the Royal Statistical Society. Series C (Applied Statistics), 31(2):205–217, January 1990.

10. R. Lall, M. J. Campbell, S. J. W. and Morgan, K.: A review of ordinal regression models applied on health-related quality of life assessments. Statistical Methods in Medical Research, 11(1):49–67, February 2002.

11. M. E. Stokes, C. S. D. and Koch., G. G.: Categorical Data Analysis Using SAS. Cary, NC: SAS Institute, 2012.

12. Yee, T. W.: Vector Generalized Linear and Additive Models: With an Implementation in R. Springer, 2015.

13. Randall, J.: The analysis of sensory data by generalised linear model. Biometrical Journal, 31:781–793, 1989.

14. Chuang-Stein, C. and Agresti, A.: Tutorial in biostatistics-a review of tests for detecting a monotone dose-response relationship with ordinal response data. Statistics in Medicine, 16(22):2599–2618, November 1997.

15. Wu, F.-C.: Simultaneous optimization of robust design with quantitative and ordinal data. International Journal of Industrial Engineering: Theory, Applications and Practice, 5:231–238, 2008.

16. J. Yang, L. T. and Mandal, A.: D-optimal designs with ordered categorical data. Technical report, MSCS, University of Illinois at Chicago, 2016.

17. Khuri, A., Mukherjee, B., Sinha, B., and Ghosh, M.: Design issues for generalized linear models: A review. Statistical Science, 21:376–399, 2006.

18. A. C. Atkinson, A. N. D. and Tobias, R. D.: Optimum Experimental Designs, with SAS. Oxford Statistical Science Series. Great Clarendon Street, Oxford OX2 6DP, Oxford, 2007.

19. Stufken, J. and Yang, M.: Optimal designs for generalized linear models. Wiley, 2012.

20. Fedorov., V. V.: Theory of Optimal Experiments. Academic Press, 1972.

21. Yang, J. and Mandal, A.: D-optimal factorial designs under generalized linear models. Communications in Statistics - Simulation and Computation, 44(9):2264–2277, October 2015.

22. Agresti, A.: Analysis of Ordinal Categorical Data. Wiley Series in Probability and Statistics. Hoboken, New Jersey, John Wiley & Sons, 2010. Second Edition.

23. Glonek, G. and McCullagh, P.: Multivariate logistic models. Journal of the Royal Statistical Society, Series B, 57:533–546, 1995.

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

25. Seber, G.: A Matrix Handbook for Statisticians. Wiley, 2008.

26. Schervish, M.: Theory of Statistics. Springer, 1995.

27. Stoica, P. and Marzetta, T. L.: Parameter estimation problems with singular information matrices. IEEE Transactions on Signal Processing, 49(1):87–90, January 2001.

28. I. Kovacs, D. S. and Williams, S.: Determinants of block matrices and schurs formula. American Mathematical Monthly, 106(5):950–952, May 1999.

29. Gantmacher, F. R.: The Theory of Matrices. Chelsea Publication, 1960.

30. Silvey, S.: Optimal Design. Chapman and Hall, 1980.

31. J. Yang, A. M. and Majumdar, D.: Optimal designs for 2^k factorial experiments with binary response. Statistica Sinica, 26(1):385–411, February 2016.

32. L. Tong, H. V. and Yang, J.: Analytic solutions for d-optimal factorial designs under generalized linear models. Electronic Journal of Statistics, 8(1):1322–1344, May 2014.

33. Chaloner, K. and Verdinelli, I.: Bayesian experimental design: a review. Statistical Science, 10(3):273–304, August 1995.

34. Itepan, N. M.: Aumento do period0 de aceitabilidade de pupas de Musca domestica L., 1758 (Diptera: Muscidae),irradiadas com raios gama, como hospedeiras de parasitbides (Hymenoptera: Pteromalidae). Centro de Energia Nuclear na AgriculturaUSP, January 1995.

35. Jennett, B. and Bond, M.: Assessment of outcome after severe brain damage. Lancet, 305(3):480–484, March 1975.

36. Golub, G. and Loan, C. V.: Matrix Computations. The Johns Hopkins University Press, 1996.

37. Martin, R.: Approximations to the determinant term in gaussian maximum likelihood estimation of some spatial models. Communications in Statististics - Theory and Methods, 22(1):189–205, January 1992.

38. Pace, R. and LeSage, J.: Chebyshev approximation of log-determinants of spatial weight matrices. Computational Statistics & Data Analysis, 45(2):179–196, March 2004.

VITA

Xianwei Bu

CONTACT

1228 Science and Engineering Offices (M/C 249)

851 S. Morgan Street, Chicago, IL 60607

xbu3@uic.edu

EDUCATION

- PH.D in Mathematics (Statistics), University of Illinois at Chicago. May 2017
- M.S. in Applied Statistics, Loyola University Chicago. May 2011
- B.S. in Atmospheric Physics, Peking University, China. July 1992

EXPERIENCE

- **Teaching Assistant**, University of Illinois at Chicago Aug 2012 - May 2016

  Chicago, IL, USA
- **Statistical Consultant**, University of Illinois at Chicago Aug 2016 - Dec 2016

  Chicago, IL, USA

- **Statistics Intern and Extern**, AbbVie Inc. July 2016 - present

  North Chicago, IL, USA

- **Intern**, BMO Financial Group May-July 2016 & May-Aug 2015

  Chicago, IL, USA

- **Math/Statistics Tutor**, Loyola University Chicago Aug 2009 - Dec 2010

  Chicago, IL, USA

- **Associate Researcher**, Second Institute of Oceanography, SOA July 1992 - July 2008

  Hangzhou, China

Presentations

Joint Statistical Meeting, Chicago, IL Aug 2016

Publications

- Discussion about Mechanism of Harmful Algal Blooms Breakout. (Original English) **Acta Oceanoligica Sinica**, 2005,24(1). Xianwei Bu, Weiyi Xu, Dedi Zhu and Gengxin Chen

- Diagnostic Calculation of the Upper-Layer Circulation in the South China Sea during the Winter of 1998. (Original English) **Acta Oceanologica Sinica**, 2004, 23 (2). Yaochu Yuan, Xianwei Bu, Guanghong Liao, Ruyun Lou, Jilan Su, Kangshan Wang

- Numerical Simulations about Dynamical Mechanism of HAB Occurrence and Spread. (Original Chinese) **Acta Oceanoligica Sinica**, 2002,24(5). Weiyi Xu, Dedi Zhu, Xianwei Bu and Gengxin Chen
- Hydrography and Current Characteristics in the South of the Yellow Sea and in the Northeast of the East China Sea in June 1999. (Original Chinese) **Acta Oceanologica Sinica**, 2002,24(suppl). Ruyun Lou, Yaochu Yuan and Xianwei Bu
- Application of P Vector Method to the Diagnostic Calculation of Circulation in the South China Sea in Summer. (Original Chinese) **Acta Oceanoligica Sinica**, 2001,23(5). Xianwei Bu, Yaochu Yuan and Yonggang Liu
- Application of P-Vector Method in the Kuroshio and the Ryukyu Current. (Original Chinese) **Acta Oceanoligica Sinica,** 2000,22(suppl). Xianwei Bu, Yaochu Yuan and Yonggang Liu
- **The Research and Assessment of Culture Zoology and Culture Capacity of Primary Fish in Xiangshan Bay.** (Original Chinese) (Monograph of project group, I am author of Chapter 5), China Ocean Press, 2002