Optimal Experimental Designs for Mixed Categorical and Continuous Responses

by

Soohyun Kim

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2017 by the
Graduate Supervisory Committee:

Ming-Hung Kao, Chair
Amylou Dueck
Rong Pan
Mark Reiser
John Stufken

ARIZONA STATE UNIVERSITY

August 2017

ABSTRACT

This study concerns optimal designs for experiments where responses consist of both binary and continuous variables. Many experiments in engineering, medical studies, and other fields have such mixed responses. Although in recent decades several statistical methods have been developed for jointly modeling both types of response variables, an effective way to design such experiments remains unclear. To address this void, some useful results are developed to guide the selection of optimal experimental designs in such studies. The results are mainly built upon a powerful tool called the complete class approach and a nonlinear optimization algorithm. The complete class approach was originally developed for a univariate response, but it is extended to the case of bivariate responses of mixed variable types. Consequently, the number of candidate designs are significantly reduced. An optimization algorithm is then applied to efficiently search the small class of candidate designs for the D- and A-optimal designs. Furthermore, the optimality of the obtained designs is verified by the general equivalence theorem. In the first part of the study, the focus is on a simple, first-order model. The study is expanded to a model with a quadratic polynomial predictor. The obtained designs can help to render a precise statistical inference in practice or serve as a benchmark for evaluating the quality of other designs.

ACKNOWLEDGMENTS

First and foremost, I thank my academic advisor, Ming-Hung (Jason) Kao for accepting me as his student. I would like to express my sincere gratitude to him for his guidance, patience, and support from the beginning to the end enabled me to develop an understanding of the subject. Without his encouragement, this dissertation would not have been possible. Also, I would like to heartily thank my committee members: Amylou Dueck, Rong Pan, Mark Reiser, and John Stufken for their encouragement and insightful comments. Thanks are also due to the School of Mathematical and Statistical Science at Arizona Statue University for my teaching assistantship from 2012 to 2016 without which I otherwise would not have been able to study statistics. I thank Dr. Kao again for his support to my research assistantship from 2016 to 2017 funded by his research grant from the National Science Foundation. I acknowledge the support provided by the Office of the Vice-President for Research and Economic Affairs, the Graduate Research Support Program, and The Graduate College at Arizona State University. I would like to thank Lauren Albin who encouraged me to develop into English writing. I would like to express my gratitude to my family: Keeran Han, Kwan Choi, Kwang-shin Choi, Hanyoon Kim, and Weonsook Koo. This dissertation would not have been possible without their love, patience, and support.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter 1

INTRODUCTION

Many experiments in engineering, medical studies, and other fields have mixed responses that contain both categorical and continuous variables. These two types of outcomes are possibly correlated to each other. For example, in the early stage of drug development studies, efficacy and toxicity of treatments of interest are simultaneously observed. Efficacy is often represented by a continuous variable, and toxicity is characterized by a categorical variable such as the occurrence of adverse events (Fedorov, Wu, and Zhang, 2012). There are similar examples in the manufacturing of solar panels and semiconductors, developmental toxicity studies, depression clinical trials, stent comparison studies, etc. (Deng and Jin, 2015; de Leon and Chough, 2013). Finding a surge in demand for using mixed response variables, researchers developed several methods for analyzing data of this sort (de Leon and Chough, 2013; Ryan, 2000).

In contrast to an animated discussion on how to analyze mixed responses, a way to design an experiment for those responses remains rather unclear. As a result, the experimenters often settle for designs that might be inefficient. What is even worse is that there is no benchmark for evaluating the efficiency of the chosen designs. When an inefficient design is chosen, the experimenter is not likely to avoid wasting limited resources. In the worst case, the experimenter gets non-informative data, so that he or she can hardly estimate the parameters of interest.

Here, we study optimal experimental designs of mixed categorical and continuous responses. We use an optimal design approach which constitutes a way to achieve 'good' designs. We will explain later what characterizes a 'good' design. Good designs are obtained by using an insightful statistical theory, or an efficient computational approach, or both. We model mixed responses by combining a generalized linear model for the categorical response and linear models for the continuous response under the different values of the categorical response variable. In the next two sections, we provide a brief discussion on the analysis and design of mixed response experiments.

1.1 Analysis of Mixed Responses

As mentioned earlier, the focus of this study is on mixed responses that contain a categorical variable as well as a continuous variable. The study of such mixed responses appeals to researchers in many fields such as developmental toxicity studies. Since the 1990s, researchers in this field have delved into a joint modeling problem using the fetal data of pregnant laboratory mice (Ryan, 2000). The outcomes contained the presence or absence of malformations, body weight and size, and sometimes organ weights.

In particular, regarding continuous and binary outcomes such as fetal weight and malformation, two major factorization approaches emerged (Catalano and Ryan, 1992; Fitzmaurice and Laird, 1995). For the continuous variable y and the binary variable z, a joint probability density function (pdf) of mixed responses was expressed as $f(y,z)=f(y)f(z|y)$ or $f(y,z)=f(z)f(y|z)$.

To formulate $f(y,z)=f(y)f(z|y)$, Catalano and Ryan (1992) introduced a latent variable y^* and made it concrete by using the form of $f(y,z)=f(y)f(z(y^*)|y)$. They assumed that, in the fetal data, the binary response of malformation was explained by the unobservable continuous latent variable. They then considered the usage of a well-established bivariate normal distribution as the joint distribution for the latent variable and the fetal weight. In addition, they used the correlation parameter to specify a correlation structure between y and y^*. They extended the model by adding more correlation parameters that account for a clustering effect in the fetus of a litter.

On the other hand, Fitzmaurice and Laird (1995) suggested a direct factorization approach using the joint pdf of $f(y,z)=f(z)f(y|z)$ where z is a binary response. They had an interest in the marginal expectation of the continuous response, $E(y)$, and considered the association between two responses a nuisance regression coefficient. This framework was inspired by the general location model in Olkin and Tate (1961). Fitzmaurice and Laird (1995) used a logistic regression model for the binary response z_i and one linear model for the continuous response $y_i|z_i$ assuming the distribution of $N(\mu+\gamma(z_i-\pi_i),\sigma^2)$ where $\mu=\beta_0+\beta_1 x_i$, β_i's are parameters, x_i is a covariate, $\pi_i=E(z_i)$, and γ is the association parameter obtained by regressing y_i on z_i.

Hirakawa (2012) also used this approach for a dose-finding study in oncology trials after changing μ to $\mu=\beta_2+(\beta_1-\beta_2)/(1+(x_i/\beta_3))^{\beta_4}$. He mentioned that a latent approach was not preferred since the simulation study showed that the existence of a latent variable undesirably pushed the level of dose toward a higher level. He considered the model to use antitumor activity as a continuous response that was in charge of the efficacy endpoint. He stated that researchers often categorized the continuous response by a threshold, which led to the loss of information.

Later, Deng and Jin (2015) suggested another approach when explaining a quality control process in the manufacturing system. The pdf in their approach was similar to the pdf in Fitzmaurice and Laird (1995), but they used two linear regression models for $y|z=0$ and for $y|z=1$ to capture an association between the continuous and categorical responses. The lapping process of a wafer, a thin slice of semiconductor material, was examined in the case study. The total thickness variation (TTV) of a rapped wafer was a continuous response and the conformity of site total indicator readings (STIR) was explained as a binary response. Their interest was on the association of mixed responses rather than on the marginal expectation of the continuous response, which was different from the case of Fitzmaurice and Laird (1995). Deng and Jin (2015) stated that they appraised a binary response under this formulation which was often informative in the manufacturing system. The focus of our study is to develop efficient designs for this recently proposed model. The details of the model will be presented in Chapter 3.

Without a consensus on how to formulate a joint pdf, various approaches are still being explored. There is a book-length collection of studies on the analysis of mixed responses (de Leon and Chough, 2013). Current models were compared, possible extensions were examined, and new modeling techniques such as a copula-based model were introduced there. Among other topics were longitudinal analysis, the Bayesian approach, and incomplete data. The benefits of joint analysis over separate univariate analysis were also measured. We refer the readers to this book for details.

1.2 Designing Mixed Response Experiments

Whereas the discussion on the analysis of mixed responses is vivid, little attention has been given to the design of mixed response experiments. While important, few studies have focused on the selection of an optimal design for such experiments.

Among a small handful of literature, Coffey and Gennings (2007) found D-optimal designs for multiple outcomes from a dose-response experiment in toxicology and other biological sciences. The D-optimality criterion will be explained in Chapter 2. They separately specified nonlinear models for five outcomes including two continuous variables, two binary variables, and one count variable. A nonlinear threshold exponential model and a logistic threshold model were used and generalized estimating equations (GEEs) were adopted. They used a model-based variance-covariance matrix instead of a working correlation matrix since the residual, which is necessary to construct a working correlation matrix, could not be obtained before experimentation. As a result, their approach did not gain one of the important benefits of the GEE, which is the consistency of parameter estimates even under misspecification of the correlation structure. This benefit is obtained by using a working correlation matrix (Fitzmaurice and Laird, 1995; Agresti, 2007; Dobson and Barnett, 2008).

On the other hand, Fedorov et al. (2012) searched for D-optimal designs for bivariate mixed responses in a dose-finding study using a latent factorization model. They considered two-stage designs and fully adaptive designs as well as locally optimal designs. The first-order exchange algorithm was used for obtaining these designs.

In their study, some parameters were not separately estimable according to the authors. In particular, for a latent variable, its mean η_2, variance σ_2^2, and a threshold parameter c_2 for discretization of a latent variable were not estimable. This might make it difficult to fully interpret the results obtained from such an approach. Notably, the association between mixed responses was hard to interpret since, in this model, the association was parametrized by a correlation parameter in the variance-covariance matrix between y and y^*, not between y and z, and then, y^* reached out to z through a cut-off point c. Additionally, the variance-covariance structure of the parameters had a complex form. For example, the information matrix, which was the inverse of the variance-covariance matrix, had a component called a_k. It was obtained by numerical integration of a function over $[-\infty, \infty]$. The function is a multiplication of several probability density functions.

In contrast, Biswas and López-Fidalgo (2013) found optimal designs for mixed responses which were modelled by a direct factorization approach. The goal of the study was to propose an optimal design for a dose-finding study in clinical trials. They also considered an experiment of mixed responses where toxicity was binary and efficacy was continuous. In particular, they used the compound optimality criterion. One component of the criterion was a traditional optimality criterion such as the D-optimality criterion and the other component was a specialized criterion for a dose-finding experiment. The latter one was devised to find designs that maximize efficacy under no toxicity. They additively combined two components using a weight λ which was a chosen value, not a design parameter.

The model contains two models which were a logistic model for the binary response and a linear model for the continuous response given the value of the binary response. A conditional variable $y_i|z_i$ had the distribution of $N(\mu + \gamma z_i, \sigma^2)$ where $\mu = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$, β_i's are parameters, x_i is a covariate and γ is the association parameter. They put a quadratic term into a predictor. They searched for optimal designs by implementing a classic first-order algorithm based on an equivalence theorem which will be explained in the next chapter. The original design space was $[0, 50]$ with the unit of mg for dose treatment. They used a linear transformation of the space and thus had the design space of $[0, 1]$.

The focus of our study is also on optimal experimental designs for mixed responses. For the underlying models, we consider a new direct-factorization approach recently proposed by Deng and Jin (2015). They involve the combination of an ordinary linear model for normal data and a generalized linear model (GLM) for non-normal data. This approach gives detailed information about the association of the two types of responses by using two conditional linear models for $y|z = 1$ and $y|z = 0$.

While results on optimal designs for each model can be found in the literature, a design problem involving both types of models together is rather complex and not much guidance is available. After we provide background knowledge about optimal designs for linear models and for GLMs, we will develop new optimal design results for mixed responses.

1.3 Outline of Dissertation

We will proceed in the following order. In Chapter 2, we first build background knowledge on optimal experimental designs and some approaches for obtaining optimal designs.

From Chapter 3 to Chapter 5, we provide three results of our studies. In Chapter 3, we identify a complete class in a systematic way and expatiate on the range of the number of support points. We note that while some of our results on the range of the number of support points can be well explained based on the previous knowledge about the design of experiments, we also discover here some rather remarkable but mathematically justifiable findings when solving this complex design problem.

The complete class results in Chapter 3 facilitate two successive numerical studies including the search for D-optimal designs and A-optimal designs. We demonstrate a basic strategy for finding optimal designs with the popular D-optimality criterion in Chapter 4 and then expand to the search for A-optimal designs in Chapter 5. The study of A-optimal designs enables us to see the effect of the variance σ^2 of continuous responses on the obtained designs, which cannot be examined in Chapter 4.

In Chapter 6, we extend our results to a model with a quadratic polynomial predictor. This setting is intended to describe a curvature trend between a covariate and a response. Such models are not uncommon in practice. Finally, we summarize the study and give concluding remarks in Chapter 7.

Chapter 2

LITERATURE REVIEW

2.1 Optimal Design Approach

The principle of an optimal design search is to hit the target right in the center by browsing all candidates of designs according to rigid mathematical standards. The object of the search is to find the best design to make an experiment as efficient as possible. It often means that, given statistical models, parameter estimation should be precise, or equivalently, the variability of estimation should be minimized. The 'goodness' of a design can then be understood in this sense when a precise parameter estimation is of interest. As Hinkelmann and Kempthorne (2005) mentioned, an optimal design approach does not provide a common answer for dealing with equipment, setting budgets, and so on. However, if experimenters want to obtain high-quality data with the analysis of it in mind, then the optimal design approach is appropriate. Using this approach, we find more statistically accurate guidelines. The data then allows for such a reliable statistical inference.

For illustration purposes, let us consider a linear regression model, $Y_i = \mathbf{f}(\boldsymbol{x}_i)^\intercal \boldsymbol{\theta} + \epsilon_i$. Y_i is the response from the ith run, $\boldsymbol{x}_i$ is a vector of independent variables, $\mathbf{f}(\boldsymbol{x}_i)$ is a model vector expressed as $(f_1(\boldsymbol{x}_i), ..., f_m(\boldsymbol{x}_i))^\intercal$, $f_j(\boldsymbol{x}_i)$ is some known function of $\boldsymbol{x}_i$, $\boldsymbol{\theta}$ is a vector of m unknown parameters, and ϵ_i is the experimental error for the ith run.

If a simple linear regression model has only one covariate x, a model vector is $\mathbf{f}(x) = [1, x]^\intercal$ and, for a quadratic model, we have $\mathbf{f}(x) = [1, x, x^2]^\intercal$. An experimental design of size N can be represented as $\{x_1,\ x_2,\ ...,\ x_N\}$; i.e. the N values of the covariate x. If the number of replicates can be expressed by a variable r_i for a trial x_i, the design of the above experiment is denoted by $\{(x_1,r_1),(x_2,r_2),...,(x_n,r_n)\}$ where $\sum_{i=1}^{n} r_i = N$ and $r_i \in \mathbb{N}$ is a positive integer. The x_i's are called support points when they are all distinct.

The variance-covariance matrix of the least squares estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is $\text{cov}(\hat{\boldsymbol{\theta}}) = \sigma^2(\mathbf{F}^\intercal\mathbf{F})^{-1}$ where $\mathbf{F}$ is a $N \times m$ full-column-rank matrix. We have $\mathbf{F}^\intercal\mathbf{F} = \sum_{i=1}^{N} \mathbf{f}(\boldsymbol{x}_i)\mathbf{f}(\boldsymbol{x}_i)^\intercal = \sum_{i=1}^{n} r_i\mathbf{f}(\boldsymbol{x}_i)\mathbf{f}(\boldsymbol{x}_i)^\intercal$. In the second member, $\mathbf{f}(\boldsymbol{x}_i)$'s represent all rows of $\mathbf{F}$ while $\mathbf{f}(\boldsymbol{x}_i)$'s of the third member are distinct rows repeated r_i times for each i in $\mathbf{F}$. We often want to minimize the determinant of the variance-covariance matrix, i.e. $|\mathbf{F}^\intercal\mathbf{F}|^{-1}$ since the square root of it is proportional to the expected volume of a confidence ellipsoid for $\boldsymbol{\theta}$. This popular criterion is called the D-optimality criterion. We mostly use a real-valued function of the variance-covariance matrix as an optimal criterion such as a determinant because it normally is difficult to rank candidate designs by a matrix $(\mathbf{F}^\intercal\mathbf{F})^{-1}$. Under this setting, the value of σ^2 is not relevant to the search for an optimal design since all candidate designs have the same value of σ^2.

Another common criterion for selecting designs is the variance of the predicted response for a given $\boldsymbol{x}$, which is $\text{var}(\hat{y}(\boldsymbol{x})) = \sigma^2\mathbf{f}(\boldsymbol{x})^\intercal(\mathbf{F}^\intercal\mathbf{F})^{-1}\mathbf{f}(\boldsymbol{x})$. After scaling the variance σ^2 and the number of trials N, the standardized variance is obtained as $d(\boldsymbol{x},\xi) = \mathbf{f}(\boldsymbol{x})^\intercal(\mathbf{F}^\intercal\mathbf{F}/N)^{-1}\mathbf{f}(\boldsymbol{x})$. We often want to find an optimal design that suppresses the worst case of the standardized variance $d(\boldsymbol{x},\xi)$. For that purpose, we use the G-optimality criterion, aiming to minimize the maximum of $d(\boldsymbol{x},\xi)$ over a design region $\mathcal{X}$.

Among candidate designs, we pick the design that achieves the minimum of $|\mathbf{F}^\intercal\mathbf{F}|^{-1}$ or that of $\max d(\boldsymbol{x},\xi)$ based on the object of an experiment. The obtained design is called a D-optimal design if the former criterion is used or a G-optimal design when the latter criterion is considered.

2.1.1 Continuous Design Theory

The development of an optimal design approach gained momentum as Kiefer (1959) introduced a continuous design theory. Assuming r_i's take any value in between 0 and N instead of being limited to integers, we avoid the complexity of discrete optimization. For example, Kiefer (1959) stated that, unless the discreteness of r_i is dismissed, D-optimal designs for a cubic regression model on an interval $[-1, 1]$ have irregular design patterns whenever N is not a multiple of 4. Such an optimization problem is in general difficult to solve. Using the continuous design theory, also known as the approximate design theory, we obtain an optimal continuous design which provides an approximate solution to a discrete design, namely an exact design. Also, if N is sufficiently large, it is known that a continuous design is close to an exact design (Berger and Wong, 2009). After obtaining an optimum continuous design, we may then use a rounding technique to get an exact design (Pukelsheim, 2006).

In the continuous design theory, designs are represented as probability measures $\xi : x_i \mapsto w_i$. We can normalize the number of replicates r_i by N. A continuous design is then expressed as $\xi = \{(x_1, r_1/N), (x_2, r_2/N), ..., (x_n, r_n/N)\}$ or $\{(x_1, w_1), (x_2, w_2), ..., (x_n, w_n)\}$, where $w_i = r_i/N$, and $\sum w_i = 1$. The size of N no longer affects the search for optimum designs. We call x a support point when the corresponding weight has $w > 0$, or a design point when the corresponding w can possibly be zero. The value of w_i is known as a 'weight' of x_i in the literature of an experimental design.

The continuous design approach is to make the information matrix $\mathbf{M}(\xi)$ as large as possible or the variance-covariance matrix $\mathbf{M}^{-1}(\xi)$ as small as possible in some sense. The information matrix is inversely proportional to the variance-covariance matrix. By the continuous design theory, the information matrix is defined as $\mathbf{M}(\xi) = \int \mathbf{f}(\boldsymbol{x})\mathbf{f}(\boldsymbol{x})^\intercal \xi(d\boldsymbol{x}) = \sum_{i=1}^{n} w_i \mathbf{f}(\boldsymbol{x}_i)\mathbf{f}(\boldsymbol{x}_i)^\intercal$ for the aforementioned linear models.

Several functions have been proposed for measuring the largeness of the information matrix or the smallness of the variance-covariance matrix. We optimize such a real-valued function of $\mathbf{M}(\xi)$ or $\mathbf{M}^{-1}(\xi)$ over all ξ that make $\mathbf{M}(\xi)$ non-singular. Some of optimal criteria that we want to minimize include:

1. G-optimality: $\Phi_G = \max_x \mathbf{f}(\boldsymbol{x})^\intercal \mathbf{M}^{-1} \mathbf{f}(\boldsymbol{x})$
2. D-optimality: $\Phi_D = |\mathbf{M}^{-1}|$
3. A-optimality: $\Phi_A = \text{trace}(\mathbf{M}^{-1})$
4. E-optimality: $\Phi_E = \lambda_{\max}(\mathbf{M}^{-1})$ where $\lambda_{\max}(\cdot)$ gives the greatest eigenvalue.

Minimizing the A-optimality criterion is equivalent to minimizing the average variance of parameter estimates. E-optimal designs minimize an upper bound of a variance of the linear combinations of parameters when the sum of the squared coefficients is 1.

There are some discussions on which optimality criterion to use when searching for optimal designs. Stufken and Yang (2012) stated that the selection of the optimality criterion depends on the purpose of experiments and personal preferences. In some cases, a design can be optimal under a broad class of optimality criteria. On the other hand, an optimal design under the certain criterion may not be optimal under another criterion. The search for optimal designs thus starts with the selection of the standard.

One significant result from the continuous design theory is the equivalence theorem established by Kiefer and Wolfowitz (1960). Over a compact design space, the theorem states that two seemingly different D-optimal and G-optimal designs are equivalent. The search for D-optimal designs is thus supplemented by that for G-optimal designs, and vice versa.

The theorem is more generally applicable under the unified optimality criterion Φ_p by Kiefer (1974). The Φ_p criterion function for a given $p \geq 0$ is expressed as:

$$\Phi_p(\mathbf{M}) = \begin{cases} |\mathbf{M}|^{-1/m} & \text{for} \quad p = 0; \\ [\frac{1}{m}\,\text{trace}((\mathbf{M}^{-1})^p)]^{1/p} & \text{for} \quad p \in (0, \infty); \\ \lambda_{\max}(\mathbf{M}^{-1}) & \text{for} \quad p = \infty, \end{cases}$$

where m is the length of a row or a column of $\mathbf{M}$. The Φ_p criterion covers D-, A-, and E-optimal criteria as when $p = 0$, $p = 1$, and $p = \infty$, respectively.

Under the Φ_p criterion, the general equivalence theorem states that, over a compact design space $\mathcal{X}$, the following three conditions are equivalent: (i) A design ξ^* is Φ_p-optimal (ii) A design ξ^* maximizes $\inf_{x \in \mathcal{X}} \phi(x, \xi)$ (iii) It holds that $\inf_{x \in \mathcal{X}} \phi(x, \xi^*) = 0$. Here, $\phi(x, \xi)$ is the directional derivative at $\mathbf{M}(\xi)$ in the direction of $\mathbf{M}(\bar{\xi})$ where $\bar{\xi}(x)$ has a unit mass at a point $x \in \mathcal{X}$. Using this theorem, we construct an optimal design or validate the optimality of some designs obtained by another approach.

Several books and papers discuss an optimal design approach and the continuous design theory (Fedorov, 1972; Silvey, 1980; Pukelsheim, 2006; Atkinson, Donev, and Tobias, 2007; Wynn, 1984; Steinberg and Hunter, 1984; Atkinson, 1996; Atkinson and Bailey, 2001).

2.1.2 Unknown Parameter Problem

A search for an optimal design of a mixed response experiment comes with the challenging issue of an unknown parameter. Khuri, Mukherjee, Sinha, and Ghosh (2006) discussed this type of problem relating to a generalized linear model (GLM). In contrast to the case of a linear model experiment, a design problem in a generalized linear model depends on unknown model parameters. Since a mixed response model contains a generalized linear model as a part of the model, we encounter the same issue described above. The same challenge also exists when finding an optimal design for a nonlinear model.

The GLM is of the form $g(E(y_i)) = \mathbf{f}(\boldsymbol{x}_i)^\intercal \boldsymbol{\beta}$, where $g(\cdot)$ is a link function that connects a linear predictor component $\eta_i = \mathbf{f}(\boldsymbol{x}_i)^\intercal \boldsymbol{\beta}$ and a response variable y_i with $\mu_i = E(y_i)$. A linear model uses two components including η_i and $E(y_i)$, but a GLM uses one more component, that is, a link function. The distribution for a response y is in an exponential family of which the pdf is $f(y_i) = \exp[(\theta y_i - b(\theta))/a(\phi) + c(y_i, \phi)]$. We then express the information matrix as $\mathbf{M}(\xi) = \sum_{i=1}^{n} w_i v_i \mathbf{f}(\boldsymbol{x}_i)\mathbf{f}(\boldsymbol{x}_i)^\intercal$ for the GLM where v_i is represented as $v_i(\theta) = [a(\phi)b''(\theta)]^{-1}(\partial\mu/\partial\eta)^2$. We can check that the components of $b''(\theta)$ and $\partial\mu_i/\partial\eta_i$ contain a parameter vector $\boldsymbol{\theta}$ by a calculation.

Consequently, v_i depends on $\boldsymbol{\theta}$ and so does the information matrix $\mathbf{M}(\xi)$. Since a parameter vector $\boldsymbol{\theta}$ is 'unknown' before executing an experiment, we have an issue called an 'unknown parameter problem' which is not the case for a linear model. Before performing an experiment, we do not have data for estimating the true value of the model parameters. Hence, when finding optimal designs for GLMs or nonlinear models, we need to handle the unknown quantities of parameters in the information matrix. To deal with this situational irony, we fix the parameter values by substituting guessed values for finding a locally optimal design. Chernoff (1953) presented the statistical validation of this approach.

Mostly, experimentation is not a one-time procedure and the previous experiment may provide reasonable initial values for parameters. Even when little information on a guessed value is available, a locally optimal design can be searched for as a benchmark. Also, in many studies of an optimal design approach, various mathematical techniques have been developed to reduce the effects of unknown parameters on an optimal design problem, for example, decomposition of the information matrix and representation of a design space.

One of the possible way of finding locally optimal designs is via the general equivalence theorem explained in the previous section. The theorem originally considers only linear models, yet, if a parameter value is fixed, the general equivalence theorem applies to nonlinear cases, too (Stufken and Yang, 2012; Pukelsheim, 2006). Another possible approach is a geometric method proposed by Elfving (1952). After generating a space by a model vector $\mathbf{f}(\boldsymbol{x})$ over all possible $\boldsymbol{x}$, the smallest ellipsoid containing the space is studied to find support points. Apart from these two approaches, Yang and Stufken (2012) recently proposed a new strategy called the complete class approach for finding a locally optimal design for nonlinear models. We apply this approach to our problem in Chapter 3. We explain the approach in detail in the next section.

2.2 Complete Class Approach

The complete class approach is a way of identifying a subclass of desirable designs. For any given design, if we identify a complete class, we know that there exists a design within the complete class that performs the same as or better than any other design. We consequently want to limit our attention to this class, when we search for optimal designs, instead of examining innumerable candidate designs. In the 1950s, there were initial attempts to conceptualize a complete class or an essentially complete class (Ehrenfeld, 1956; Kiefer, 1959). Later, some researchers proposed different ways to define and find a complete class (Pukelsheim, 1989; Cheng, 1995).

In this study, we use the complete class approach proposed by Yang and Stufken (2012). The strategy of Yang and Stufken (2012) is to identify a complete class with a simple form, so that there exists a design ξ^* in the complete class to satisfy $\mathbf{M}(\xi^*) \succeq \mathbf{M}(\xi)$ for any given design ξ under the Loewner ordering. We say that a design ξ^* is at least as good as another design ξ under the Loewner ordering if we have $\mathbf{M}(\xi^*) \succeq \mathbf{M}(\xi)$, i.e. $\mathbf{M}(\xi^*) - \mathbf{M}(\xi)$ is nonnegative definite. The Loewner ordering is one of the possible ways to compare designs. The design ξ^* is then no worse than ξ under the popularly used Φ_p criteria that include the D-, A-, E-optimality criterion; that is, $\mathbf{M}_1 \succeq \mathbf{M}_2 \Rightarrow \Phi_p(\mathbf{M}_1) \leq \Phi_p(\mathbf{M}_2)$ for $p \in [0, \infty)$. The main benefit of the complete class approach in Yang and Stufken (2012) is that we can identify an upper bound for the number of design points.

This approach evolved from several preceding studies (De la Garza, 1954; Yang and Stufken, 2009; Yang, 2010). Yang and Stufken (2009) showed an early idea using a GLM with two parameters, and Yang (2010) laid the foundation of the complete class approach based on this. As mentioned in Yang (2010), de la Garza (1954) found that, for a polynomial regression model of degree p with independently identically distributed random errors, we always have a $(p+1)$-point design whose information matrix is no worse than that of any n-point design where $n > p+1$ in the Loewner ordering. This finding implies that the upper bound for the number of design points is $p+1$ for a polynomial model. Yang (2010) analytically revived the so-called de la Garza phenomenon in nonlinear models while Dette and Melas (2011) and Dette and Schorning (2013) gave thought to this phenomenon by using the concept of the Chebyshev systems developed by Karlin and Studden (1966).

Yang and Stufken (2012) generalized the approach of Yang (2010) so that we might identify a smaller complete class than before. In Lemmas 1 and 2 and Theorem 1 of Yang and Stufken (2012), they explained how to identify a complete class and, in Theorem 2, provided a tool for identification.

A nonlinear model including a mixed response model commonly has the decomposed form of the information matrix as $\mathbf{M}(\theta) = \mathbf{B}(\theta)(\sum w_i \mathbf{C}(\theta, c_i))\mathbf{B}(\theta)^\intercal$ where a non-singular matrix $\mathbf{B}(\theta)$ depends only on θ. We also consider a represented design point of c_i instead of x_i, where c_i is obtained by x_i through a bijection. A different model requires a different transformation or decomposition. Bijections will be clearly specified in the subsequent chapters when we describe our findings. We note that the information matrix $\mathbf{M}(\theta)$ depends on the represented design point c_i only through the matrix $\mathbf{C}(\theta, c_i)$.

The complete class approach of Yang and Stufken (2012) is built upon the following fact. For a design $\xi = \{(c_i, w_i), i=1, ..., n\}$ and a design $\tilde{\xi} = \{(\tilde{c}_i, \tilde{w}_i), i=1, ..., \tilde{n}\}$ in a represented design space, it is obvious that $\sum_{i=1}^{\tilde{n}} \tilde{w}_i \mathbf{C}(\boldsymbol{\theta}, \tilde{c}_i) \geq \sum_{i=1}^{n} w_i \mathbf{C}(\boldsymbol{\theta}, c_i)$ implies $\mathbf{M}_{\tilde{\xi}}(\boldsymbol{\theta}) \succeq \mathbf{M}_{\xi}(\boldsymbol{\theta})$. After partitioning $\mathbf{C}$ as

$$\mathbf{C}(\boldsymbol{\theta}, c) = \begin{pmatrix} \mathbf{C}_{11}(c) & \mathbf{C}_{12}(c) \\ \mathbf{C}_{12}(c)^\intercal & \mathbf{C}_{22}(c) \end{pmatrix} \tag{2.1}$$

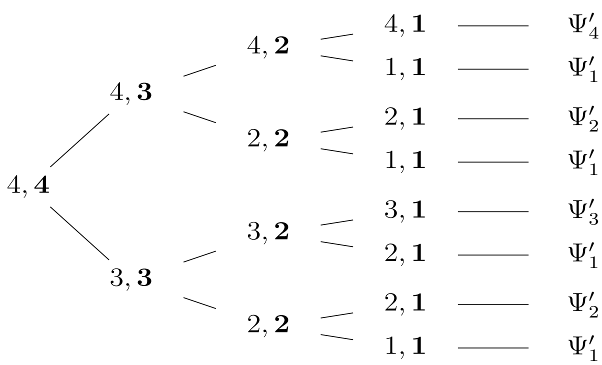
where $\mathbf{C}_{11}$ is an m_1-by-m_1 matrix, and $\mathbf{C}_{22}$ is an m_2-by-m_2 principal submatrix for some $1 \leq m_1, m_2 < m$, we have $\mathbf{M}_{\tilde{\xi}}(\theta) \succeq \mathbf{M}_{\xi}(\theta)$ if $\sum_{i=1}^{\tilde{n}} \tilde{w}_i \mathbf{C_{22}}(\boldsymbol{\theta}, \tilde{c}_i) \geq \sum_{i=1}^{n} w_i \mathbf{C_{22}}(\boldsymbol{\theta}, c_i)$, $\sum_{i=1}^{\tilde{n}} \tilde{w}_i \mathbf{C_{11}}(\boldsymbol{\theta}, \tilde{c}_i) = \sum_{i=1}^{n} w_i \mathbf{C_{11}}(\boldsymbol{\theta}, c_i)$ and $\sum_{i=1}^{\tilde{n}} \tilde{w}_i \mathbf{C_{12}}(\boldsymbol{\theta}, \tilde{c}_i) = \sum_{i=1}^{n} w_i \mathbf{C_{12}}(\boldsymbol{\theta}, c_i)$.

To identify a complete class using the tool of Theorem 2 in Yang and Stufken (2012), we extract relevant element functions from the matrix $\mathbf{C}(\boldsymbol{\theta}, c)$ and denote them Ψ_i's $(i=1, ..., k)$. The functions $\Psi_1, ..., \Psi_{k-1}$ are selected from $\mathbf{C}_{11}$ and $\mathbf{C}_{12}$, and they form a maximal set of linearly independent non-constant functions of c. We then make a sequence of Ψ_i's for $i=1, ..., k-1$ with a judiciously selected order and define $\Psi_0 = 1$ and $\Psi_k = \mathbf{C}_{22}(c)$. Note that Ψ_k can be a matrix while other Ψ_i's are scalars. A proper choice of Ψ_i's and $\mathbf{C}_{22}$ is required. With the selected $\mathbf{C}_{22}$ and Ψ-functions, we calculate $F(c)$, as suggested in Theorem 2 of Yang and Stufken (2012) and see the sign of it to check a condition of identification.

Following Yang and Stufken (2012), let us define the functions $f_{l,t}(c), 1 \leq t \leq l \leq k$ as

$$f_{l,t}(c) = \begin{cases} \Psi_l'(c) & \text{if } , t=1, l=1, ..., k-1, \\ \mathbf{C}_{22}', & \text{if } t=1, l=k, \\ \left(\frac{f_{l,t-1}(c)}{f_{t-1,t-1}(c)}\right)', & \text{if } 2 \leq t \leq k, t \leq l \leq k, \end{cases} \tag{2.2}$$

Figure 2.1: Structure of the Indices in $f_{4,4}$



$$f_{4,4}=\left[\frac{\left(\frac{(\frac{\Psi_4'}{\Psi_1'})'}{(\frac{\Psi_2'}{\Psi_1'})'}\right)'}{\left(\frac{(\frac{\Psi_3'}{\Psi_1'})'}{(\frac{\Psi_2'}{\Psi_1'})'}\right)'}\right]'$$

assuming that $\Psi_i(i=1,...,k)$ is differentiable. Then, if it holds that

$$F(c)=\prod_{l=1}^{k} f_{l,l}>0 \quad \text{or} \quad F(c)<0 \quad \text{for} \quad \text{all} \quad c\in[A,B], \tag{2.3}$$

Theorem 2 in Yang and Stufken (2012) directly affirms the existence of a complete class.

For a fixed l, we see that $f_{l,l}(c)$ in (2.2) and (2.3) is expanded as $f_{l,l}(c)=\left(\frac{f_{l,l-1}(c)}{f_{l-1,l-1}(c)}\right)'=\left(\left(\frac{f_{l,l-2}(c)}{f_{l-2,l-2}(c)}\right)'/\left(\frac{f_{l-1,l-2}(c)}{f_{l-2,l-2}(c)}\right)'\right)'=...$ until $f_{l,l}$ floats every necessary Ψ_i's and $\mathbf{C}_{22}$. For example, $f_{3,3}(c)=\left(\frac{f_{3,2}(c)}{f_{2,2}(c)}\right)'=\left(\left(\frac{f_{3,1}(c)}{f_{1,1}(c)}\right)'/\left(\frac{f_{2,1}(c)}{f_{1,1}(c)}\right)'\right)'=\left(\left(\frac{\Psi_3'}{\Psi_1'}\right)'/\left(\frac{\Psi_2'}{\Psi_1'}\right)'\right)'$. As another example, the structure of the indices in $f_{4,4}$ is shown in Figure 2.1. When $f_{k,k}$ is a matrix, $F(c)>0$ means that the matrix $f_{k,k}$ is positive definite for all $c\in[A,B]$. The differentiation of these functions can be done sequentially with a symbolic software such as Mathematica or Maple. MATLAB also supports a symbolic calculation.

Although we conveniently derive a checking condition for many nonlinear models using this tool as shown in their examples, the tool is not directly applicable to our case since mixed responses are bivariate. We therefore broaden the scope of discussion and attempt to apply their approach as follows.

We understand the strategy of the complete class approach by using the definition of the Chebyshev system (Karlin and Studden, 1966; Dette and Melas, 2011). The

$F(c)$ defined in (2.3) is such a general-purpose tool that it is applicable not only to some element functions of the information matrix relating to an optimal design problem but also to some nonlinear functions in general. This can be explained by using the Chebyshev system. The Chebyshev system is defined as a set of continuous functions $u_0, ..., u_k$ from $[A, B]$ to $\mathbb{R}$ if the inequality

$$\begin{vmatrix} u_0(x_0) & u_0(x_1) & \dots & u_0(x_k) \\ u_1(x_0) & u_1(x_1) & \dots & u_1(x_k) \\ \vdots & \vdots & \ddots & \vdots \\ u_k(x_0) & u_k(x_1) & \dots & u_k(x_k) \end{vmatrix} > 0 \tag{2.4}$$

is satisfied for all $A \leq x_0 < x_1 < ... < x_k \leq B$ (Karlin and Studden, 1966).

We assume any continuous functions $\Psi_1, ..., \Psi_k$ on $[A, B]$. We define $\Psi_0 = 1$, $\Psi_k = \mathbf{C}_{22}$ for any matrix $\mathbf{C}_{22}$, and $\Psi_k^Q = Q^T \mathbf{C}_{22} Q$ for every nonzero vector Q. If we can form pairs of Chebysyhev systems with $\{\Psi_0, \Psi_1, ..., \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, ..., \Psi_{k-1}, \Psi_k^Q\}$ or $\{\Psi_0, \Psi_1, ..., \Psi_{k-1}\}$ and $\{\Psi_0, \Psi_1, ..., \Psi_{k-1}, -\Psi_k^Q\}$ for all non-zero vectors Q, we can then find a dominant set $S^* = \{(c_i^*, w_i^*) : w_i^* > 0, i = 1, ..., n^*\}$ defined by Yang and Stufken (2012) for any given set $S = \{(c_i, w_i) : w_i > 0, i = 1, ..., N\}$ such that

$$\sum w_i^* \Psi_l(c_i^*) = \sum w_i \Psi_l(c_i), \qquad l = 0, 1, ..., k-1; \tag{2.5}$$

$$\sum w_i^* \Psi_k^Q(c_i^*) > \sum w_i \Psi_k^Q(c_i), \qquad \text{for every nonzero vector } Q, \tag{2.6}$$

where subscripts of summations are n^* and N for (c_i^*, w_i^*) and (c_i, w_i), respectively. This holds based on Lemmas 1 and 2 in Yang and Stufken (2012). In these lemmas, we find n^* of S^* from a relationship between k, n^*, and N, which will be specifically presented on the next page with a certain configuration of endpoints of c_i^*'s. The elements of S^* and S do not need to be designs as the w_i^*'s or w_i's do not need to sum to 1. If the above two equations are explained for relevant elements of the information matrix, it implies the non-inferiority of S^* to S under the Loewner ordering and we immediately identify a complete class. Yang and Stufken (2012) used this fact to render complete classes in their Theorem 1 and suggest $F(c)$ in Theorem 2. We rather use the $F(c)$ in (2.3) to verify whether the Ψ functions form Chebyshev systems in the following lemma, which is a direct consequence of (the proof of) Theorem 2 of Yang and Stufken (2012).

Lemma 2.2.1. *For any continuous function $\Psi_1, ..., \Psi_k$, if either $F(c)$ or $-F(c)$ is positive definite for all $c \in [A, B]$, then there exists a set of functions $\hat{\Psi}_1, ..., \hat{\Psi}_{k-1}$ that satisfy the following results. Here, $\hat{\Psi}_l = \Psi_l$ for some l, but $\hat{\Psi}_l = -\Psi_l$ for the other l.*

(a) If $F(c) > 0$, $\{1, \hat{\Psi}_1, ..., \hat{\Psi}_{k-1}\}$ and $\{1, \hat{\Psi}_1, ..., \hat{\Psi}_{k-1}, \Psi_k^Q\}$ form Chebyshev systems on [A, B] for all non-zero Q.

(b) If $-F(c) > 0$, $\{1, \hat{\Psi}_1, ..., \hat{\Psi}_{k-1}\}$ and $\{1, \hat{\Psi}_1, ..., \hat{\Psi}_{k-1}, -\Psi_k^Q\}$ form Chebyshev systems on [A, B] for all non-zero Q.

We note that the lemma is useful because equalities in (2.5) are true for $\hat{\Psi}_1$, ..., $\hat{\Psi}_{k-1}$ if they hold for Ψ_l's, and vice versa. Based on the above lemma, we check the sign of $F(c)$ and get one of four types of a dominant set S^* for S according to the parity of k by using the following Lemma 2 of Yang and Stufken (2012). The notations are the same as those previously defined in this section.

Lemma 2.2.2 (Lemma 2 in Yang and Stufken, 2012)**.** *Let $S = \{(c_i, w_i) : w_i > 0, A \leq c_i \leq B, i = 1, ..., N\}$. Then the following results hold:*

(a) For $k = 2n-1$, if Lemma 2.2.1 (a) holds, then there exists a dominant set S^* of size n with $c_n^* = B$ for S when $N \geq n$.

(b) For $k = 2n-1$, if Lemma 2.2.1 (b) holds, then there exists a dominant set S^* of size n with $c_1^* = A$ for S when $N \geq n$.

(c) For $k = 2n$, if Lemma 2.2.1 (a) holds, then there exists a dominant set S^* of size $n+1$ with $c_1^* = A$ and $c_{n+1}^* = B$ for S when $N \geq n$.

(d) For $k=2n$, if Lemma 2.2.1 (b) holds, then there exists a dominant set S^* of size n for S when $N \geq n+1$.

After identifying S^*, we use it to form a complete class. The emphasis is on the fact that we can reduce n^* to a certain size that depends on k. The key is to explain (2.5) and (2.6) for all element functions in the information matrix and ultimately demonstrate the Loewner ordering. In addition, we note that we do not use an assumption of symmetric designs in the complete class approach. We then distinguish a class of symmetric designs and manage it to solve a complex design problem. A symmetric design ξ_s is defined as $\xi_s = \{(\pm c_i, w_i), i=1, ..., n\}$ where there exists $-c_i$ for any c_i in a design ξ_s with the same weight w_i as a common definition. After identifying a symmetric design as a 'good' design, we will form a complete class within the collection of symmetric designs in Chapter 6.

2.3 Constrained Nonlinear Optimization

The domain of mathematical programming, i.e. mathematical optimization, has been well developed by many disciplines including applied mathematics, operations research, electrical engineering, and so on. Optimization algorithms are a powerful set of tools that can efficiently manage a subject's resources. Large scale of optimization problems can be solved reliably using various optimization algorithms if it is possible to formulate a real problem into a mathematical standard form. After submitting the form to an appropriate solver, we can get optimal solutions.

In the area of optimal design, a few researchers have used mathematical programming as an alternative to traditional algorithms such as the Fedorov-Wynn algorithm (Wynn, 1970; Fedorov, 1972) and the multiplicative algorithm (Silvey, Titterington, and Torsney, 1978). Recent works cover semi-infinite programming (Duarte and Wong, 2014), semi-definite programming (Papp, 2012), and others. In Duarte and

Wong (2014), the semi-infinite programming was available for a minimax design problem since their problem could be formulated into the semi-infinite programming. They used the solver QQNLP from the GAMS package. Papp (2012) found optimal designs for rational function regressions using semi-definite programming. He set up a two-step search procedure for design points and weights separately. After design points were determined, weights were also found. The procedure was, however, restricted to polynomial or rational function regressions which do not cover our model.

We solve the nonlinear constrained problem by using the fmincon solver in MATLAB. The Optimization Toolbox in MATLAB includes various solvers to deal with problems ranging from linear to nonlinear, from continuous to discrete, and from unconstrained to constrained. In our case, an objective function is a nonlinear smooth function. Also, there are some constraints of an optimal design approach. The constraints are the restriction of a design space and a continuous design setting for $\xi = \{(c_i, w_i), i = 1, ..., n\}$ on $[A, B]$. Taking all constraints into account, an optimal design problem is formulated into a mathematical standard form as follows.

$$\begin{aligned}
&\underset{\boldsymbol{\xi}^0}{\text{minimize}} && \Phi_p(\mathbf{M}(\boldsymbol{\xi}^0)) && \\
&\text{subject to} && \sum w_i = 1, && (i = 1, ..., n) \\
& && w_i \geq 0, && (i = 1, ..., n) \\
& && \text{and} \quad A \leq c_i \leq B && (i = 1, ..., n),
\end{aligned} \tag{2.7}$$

where $\Phi_p(\mathbf{M}(\boldsymbol{\xi}^0))$ is the objective function, and the equations in (2.7) are the set of the equality and inequality constraints. A vector $\boldsymbol{\xi}^0 = (c_1, ..., c_n, w_1, ..., w_n)$ represents decision variables consisting of design points and weights. A and B are upper and lower bound points of a design space. If n is fixed, optimization formulation can be completed. Since mathematical programming is devised for allocating given resources, the subject of resources should be decided beforehand. In our case, n should be fixed.

We still use an induced point c instead of x. In a represented design space, we identify a complete class and then without transforming it to the point of an original design space, we continue to do an algorithm search. In a represented design space $[A, B]$, if a complete class contains A, B or both, that point or those points are excluded from decision variables of $\boldsymbol{\xi}^0$. For example, if two bounds are fixed points of an at most 4-point design, decision variables or input variables are contained in $\boldsymbol{\xi}^0 = (c_2, c_3, w_1, w_2, w_3, w_4)$.

2.4 General Equivalence Theorem

The general equivalence theorem (GET) was discussed in the previous section. Using part of the theorem, we derive two types of the equations to verify the D-optimality and the A-optimality of the obtained designs. The main purpose of the verification is to prove that the design searched by our methods is optimal by the GET. If the verification is successful, the reliability of our results increases.

Let a measure $\bar{\xi}$ put unit mass at a point c and another measure ξ' be given by $\xi' = (1-\alpha)\xi + \alpha\bar{\xi}$ for $\xi = \{(c_i, w_i), i = 1, ..., n\}$. We then have

$$\mathbf{M}(\xi') = (1-\alpha)\mathbf{M}(\xi) + \alpha\mathbf{M}(\bar{\xi}).$$

Thus, the directional derivative of $\Phi_p(\mathbf{M})$ at $\mathbf{M}(\xi)$ in the direction from $\mathbf{M}(\xi)$ to $\mathbf{M}(\bar{\xi})$ is defined as

$$\begin{aligned}\phi(x, \xi) &= \lim_{\alpha \to 0+} \frac{1}{\alpha}[\Phi_p(\mathbf{M}(\xi')) - \Phi_p(\mathbf{M}(\xi))] \\ &= \lim_{\alpha \to 0+} \frac{1}{\alpha}[\Phi_p\{(1-\alpha)\mathbf{M}(\xi) + \alpha\mathbf{M}(\bar{\xi})\} - \Phi_p(\mathbf{M}(\xi))] \\ &= \lim_{\alpha \to 0+} \frac{1}{\alpha}[\Phi_p\{\mathbf{M}(\xi) + \alpha(\mathbf{M}(\bar{\xi}) - \mathbf{M}(\xi))\} - \Phi_p(\mathbf{M}(\xi))].\end{aligned}$$

Based on this result, we derive the equations for verification.

D-optimality verification For the D-optimality, $\phi(x,\xi)$ is expressed as

$$\phi(x,\xi) = \operatorname{tr}\mathbf{M}^{-1}(\xi)\mathbf{M}(\xi) - \operatorname{tr}\mathbf{M}^{-1}(\xi)\mathbf{M}(\bar{\xi}).$$

We used a represented design space of c mentioned in the previous sections. Then, noting that $\mathbf{M}(\xi) = \mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal}$ where $\mathbf{C}^{*} = \sum w_i \mathbf{C}(\theta, c_i)$, we have

$$\begin{aligned}
\phi(c,\xi) &= \operatorname{tr}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal}) - \operatorname{tr}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1}(\mathbf{B}\mathbf{C}^{*}(\bar{\xi})\mathbf{B}^{\intercal}) \\
&= \operatorname{tr}\mathbf{C}^{*-1}(\xi)\mathbf{C}^{*}(\xi) - \operatorname{tr}\mathbf{C}^{*-1}(\xi)\mathbf{C}^{*}(\bar{\xi}) \\
&= m - d(c,\xi) \geq 0 \qquad (2.8)
\end{aligned}$$

where m is the number of parameters, and $d(c,\xi) = \operatorname{tr}\ \mathbf{C}^{*-1}(\xi)\mathbf{C}^{*}(\bar{\xi})$. We use $d(c,\xi) \leq m$ for verification.

A-optimality verification From a directional derivative of Φ_A, we obtain the equation of the A-optimal verification as follows.

$$\begin{aligned}
\phi(x,\xi) &= \operatorname{tr}[-\mathbf{M}^{-\mathbf{1}}(\xi)\frac{d\mathbf{M}(\xi')}{d\alpha}\mathbf{M}^{-\mathbf{1}}(\xi)] \\
&= -\operatorname{tr}[\mathbf{M}^{-\mathbf{1}}(\xi)[\mathbf{M}(\bar{\xi}) - \mathbf{M}(\xi)]\mathbf{M}^{-\mathbf{1}}(\xi)] = \operatorname{tr}[\mathbf{M}^{-\mathbf{1}}(\xi)] - \operatorname{tr}[\mathbf{M}(\bar{\xi})\mathbf{M}^{-\mathbf{2}}(\xi)].
\end{aligned}$$

For a represented design space of c, using the same notation as in the case of the D-optimality verification, we see that

$$\begin{aligned}
\phi(c,\xi) &= \operatorname{tr}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1} - \operatorname{tr}(\mathbf{B}\mathbf{C}^{*}(\bar{\xi})\mathbf{B}^{\intercal})(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1} \\
&= \operatorname{tr}(\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1} - \operatorname{tr}\mathbf{C}^{*}(\bar{\xi})(\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal}\mathbf{B}\mathbf{C}^{*}(\xi))^{-1} \\
&= C - s(c,\xi) \geq 0 \qquad (2.9)
\end{aligned}$$

where $C = \operatorname{tr}\ (\mathbf{B}\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal})^{-1}$ is a constant for an obtained design ξ, and $s(c,\xi) = \operatorname{tr}\ \mathbf{C}^{*}(\bar{\xi})(\mathbf{C}^{*}(\xi)\mathbf{B}^{\intercal}\mathbf{B}\mathbf{C}^{*}(\xi))^{-1}$. We use $s(c,\xi) \leq C$ for A-optimality verification. The two inequality equations derived here will be extended for a mixed response model in Chapter 4, 5, and 6.

2.5 Scope of Studies and Specific Aim

In this study, we investigate locally optimal designs for an experiment where responses include both binary and continuous variables. Among possible statistical models, we adopt a direct-factorization approach to formulating the joint pdf and use one logistic model and two conditional linear models. When we are interested in an association between mixed responses, this type of modelling provides a useful analysis of mixed responses. On the other hand, Fedorov et al. (2012) found D-optimal designs for such responses based on a latent-factorization approach by assuming an unobservable continuous latent variable for the categorical variable. Also, Biswas and López-Fidalgo (2013) found optimal designs for these types of responses by using a direct factorization approach, but their model used one conditional linear model. Under the compound optimality criterion, they found 4-point designs.

We tackle our design problem using a complete class approach and a nonlinear optimization technique. We identify a complete class in an analytic way to significantly decrease the number of candidate designs. Staying within a complete class, we search for optimal designs by using a computer algorithm for nonlinear constrained optimization. The obtained designs are verified as optimal by the general equivalence theorem suggested by Kiefer (1974).

We note that the complete class approach that we consider discloses a different aspect of an optimal design approach in contrast to the GET. The GET is an important ground of the continuous design theory and is still used for constructing and validating an optimal design. However, the GET does not tell about the maximal number of support points n that optimal designs can possess. Researchers thus may start with a large n or a moderate n depending on characteristics of their algorithms and use the fine grid of a design space. Consequently, computation is expensive and the evidence of final answers is less conclusive. Hence, if a complete class is identified, we have a theoretically supported information on optimal designs and need to browse possible designs within the complete class. Then, we lighten the burden of an algorithmic search.

Since complete class results allow us to shift our focus on a small collection of designs, we still need a method for identifying an optimal design. There are some follow-up works focusing on how to derive concluding answers after identifying a complete class (Wu and Stufken, 2014; Hu, Yang, and Stufken 2015). Wu and Stufken (2014) algebraically found a Φ_p-optimal design for a generalized linear model with a quadratic polynomial predictor. Hu, Yang, and Stufken (2015) found optimal designs in various nonlinear models using Newton's algorithm along with some theoretical results. In our case, we use a nonlinear constrained optimization in mathematical programming to get numerical solutions.

The primary interest of mathematical programming is an allocation of resources so the subjects of allocation should be determined beforehand. In other words, for an optimal design problem, the number of design points should be decided. However, in most cases, this number is not determined so mathematical optimization is not more popular in optimal design studies despite its efficiency. The complete class approach helps us overcome this difficulty.

Using the GET, we will validate the optimality of the results. The GET serves as an excellent tool for optimality verification. Our study focuses only on searching locally optimal designs. A locally optimal design presumes a best-guessed parameter value to remedy an 'unknown parameter problem' regarding the information matrix which depends on unknown parameters. Despite this limitation, a locally optimal design is obtained and studied for a mixed response experiment since the obtained designs can be at least a good benchmark for evaluating other designs.

As mentioned earlier, not much work has been done in finding optimal designs for mixed response experiments. Our focus is on the search for a continuous optimal design by considering a locally optimal design approach. Our results will be provided in the next four chapters.

Chapter 3

COMPLETE CLASS RESULTS

3.1 Statistical Model and Fisher Information Matrix

We assume that the experiment is described by one independent variable x, one continuous response variable y, and one binary response variable z. Let us denote observable data as $(x_i, y_i, z_i), i=1, ..., N$, where $x_i, y_i \in \mathbb{R}$ and $z_i \in \{0, 1\}$.

A mixed response regression model is described as follows. In this model, we use the product of the marginal distribution of z and the conditional distribution of y given z for the joint probability density function (pdf) of $(y,\ z)$. The binary variable z_i is modeled by a logistic regression model with probability π_i for $z_i=1$, and the conditional distribution of y_i given $z_i=0$ or 1 is assumed to follow a normal distribution. In particular, we have the following,

$$z_i = \begin{cases} 1 & \text{with} \quad \pi_i \\ 0 & \text{with} \quad 1-\pi_i \end{cases} \quad \text{with } \pi_i = E(z_i) = \frac{\exp(\alpha_0 + \alpha_1 x_i)}{1+\exp(\alpha_0 + \alpha_1 x_i)}, \tag{3.1}$$

and

$$y_i | z_i \sim \begin{cases} N(\mu_1, \sigma^2) & \text{if} \quad z_i = 1 \\ N(\mu_2, \sigma^2) & \text{if} \quad z_i = 0 \end{cases} \quad \text{with} \quad \begin{cases} E(y_i|z_i=1) = \mu_1 = \beta_0^{(1)} + \beta_1^{(1)} x_i \\ E(y_i|z_i=0) = \mu_2 = \beta_0^{(2)} + \beta_1^{(2)} x_i \end{cases},$$

where $\alpha_0, \alpha_1, \beta_0^{(1)}, \beta_1^{(1)}, \beta_0^{(2)}, \beta_1^{(2)}$, and σ^2 are unknown parameters. The joint model describes the relationship not only between x and (y, z) but also between y and z.

The mixed response model that we consider has the joint pdf of y and z as

$$\begin{aligned} f(y_i, z_i) =& f(z_i) f(y_i|z_i) \\ &= \pi_i^{z_i}(1-\pi_i)^{1-z_i}[f(y_i|z_i=1)]^{z_i}[f(y_i|z_i=0)]^{1-z_i} \\ &= \left[\frac{\exp(\alpha_0+\alpha_1 x_i)}{1+\exp(\alpha_0+\alpha_1 x_i)}\right]^{z_i}\left[\frac{1}{1+\exp(\alpha_0+\alpha_1 x_i)}\right]^{1-z_i} \\ &\left[\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(y_i-(\beta_0^{(1)}+\beta_1^{(1)}x_i))^2}{2\sigma^2})\right]^{z_i}\left[\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(y_i-(\beta_0^{(2)}+\beta_1^{(2)}x_i))^2}{2\sigma^2})\right]^{1-z_i}. \end{aligned}$$

We denote the vector of all model parameters as $\boldsymbol{\theta_0} = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}^{(1)\intercal}, \boldsymbol{\beta}^{(2)\intercal}, \sigma^2)^\intercal = (\alpha_0, \alpha_1, \beta_0^{(1)}, \beta_1^{(1)}, \beta_0^{(2)}, \beta_1^{(2)}, \sigma^2)^\intercal$. The form of $f(y_i|z_i) = [f(y_i|z_i=1)]^{z_i}[f(y_i|z_i=0)]^{1-z_i}$ is available when z_i is binary. Then, the log-likelihood function is

$$\begin{aligned} l_N(\boldsymbol{\theta_0}) =& \log\prod_{i=1}^N f(y_i, z_i) \\ &= \log\prod_{i=1}^N f(z_i)[f(y_i|z_i=1)]^{z_i}[f(y_i|z_i=0)]^{1-z_i} \\ &= \sum_{i=1}^N \log f(z_i) + \sum_{i=1}^N z_i \log f(y_i|z_i=1) + \sum_{i=1}^N (1-z_i)\log f(y_i|z_i=0). \end{aligned}$$

If necessary, $f(y_i)$ and $f(z_i|y_i)$ can be easily derived from the models. The former one is $f(y_i) = f(y_i, z_i=0) + f(y_i, z_i=1) = (1-\pi_i)f(y_i|z_i=0) + \pi_i f(y_i|z_i=1)$. Following this, the latter one is obtained as $P(z_i=k|y_i) = f(y_i|z_i=k)P(z_i=k)/f(y_i)$ for $k=0,1$. Then, we see that

$$z_i|y_i = \begin{cases} 1 & \text{with} \quad \pi_i^0 \\ 0 & \text{with} \quad 1-\pi_i^0 \end{cases} \quad \text{with } \pi_i^0 = \frac{\exp(\alpha_0+\alpha_1 x_i)}{l+\exp(\alpha_0+\alpha_1 x_i)},$$

where $l = \exp((y_i - (\beta_0^{(2)} + \beta_1^{(2)} x_i))^2)/\exp((y_i - (\beta_0^{(1)} + \beta_1^{(1)} x_i))^2)$ as also indicated in Deng and Jin (2015). Expectations and variances of y_i and $z_i|y_i$ can also be calculated.

We have an interest in the information matrix since the inverse of an information matrix is the smallest asymptotic variance of the unbiased parameter estimates of $\boldsymbol{\theta}$ in a maximum likelihood estimation. Confining our consideration to $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\intercal, \boldsymbol{\beta}^{(1)\intercal}, \boldsymbol{\beta}^{(2)\intercal})^\intercal$, the Fisher information matrix $\mathbf{M}$ for a continuous design $\xi =$

$\{(x_i, w_i), i=1, ..., n\}$ is the following 6×6 symmetric block diagonal matrix:

$$\mathbf{M}(\xi, \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{F}^\intercal \mathbf{W} \mathbf{P}(\mathbf{I}-\mathbf{P})\mathbf{F} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F}^\intercal \mathbf{W}\mathbf{P}\mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F}^\intercal \mathbf{W}(\mathbf{I}-\mathbf{P})\mathbf{F} \end{bmatrix}, \text{ where } \mathbf{F} = \begin{bmatrix} \mathbf{f}(x_1)^\intercal \\ \mathbf{f}(x_2)^\intercal \\ \cdots \\ \mathbf{f}(x_n)^\intercal \end{bmatrix},$$

$\mathbf{f}(x_i) = [1, x_i]^\intercal$ is the model vector, $\mathbf{P} = \text{diag}(\pi_1, ..., \pi_n)$ with $\pi_i = \frac{\exp(\alpha_0 + \alpha_1 x_i)}{1+\exp(\alpha_0+\alpha_1 x_i)}$, and $\mathbf{W} = \text{diag}(w_1, ..., w_n)$. Here, σ^2 cannot be factored out as opposed to most traditional design problems under linear models.

For dealing with the information matrix conveniently, we consider a represented design point c_i and a represented design as $\xi = \{(c_i, w_i), i=1, ..., n\}$, for $c_i \in [A, B]$. Using a bijection from x_i to c_i, we define c_i as $c_i = \alpha_0 + \alpha_1 x_i$. This representation is expressed in the information matrix by a matrix $\mathbf{B}_1^{-1}(\boldsymbol{\theta}) = \left(\begin{smallmatrix} 1 & 0 \\ \alpha_0 & \alpha_1 \end{smallmatrix}\right)$ that gives $\left(\begin{smallmatrix} 1 & 0 \\ \alpha_0 & \alpha_1 \end{smallmatrix}\right)\left(\begin{smallmatrix} 1 \\ x_i \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 \\ c_i \end{smallmatrix}\right)$. Ford, Torsney, and Wu (1992) also used such a canonical form to solve an optimal design problem independently of $\boldsymbol{\theta}$, although a locally optimal design still depended on the values of θ. For a design $\xi = \{(c_i, w_i), i=1, ..., n\}$, the information matrix can be written as

$$\mathbf{M}(\xi, \boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}, \sigma)\left(\sum_{i=1}^{n} w_i \mathbf{C}(\boldsymbol{\theta}, c_i)\right)(\mathbf{B}(\boldsymbol{\theta}, \sigma))^\intercal, \tag{3.2}$$

where $\mathbf{B}(\boldsymbol{\theta}, \sigma)$ is a 6×6 nonsingular matrix that depends not only on $\boldsymbol{\theta}$ but also on σ as

$$\mathbf{B}(\boldsymbol{\theta}, \sigma) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \alpha_0 & \alpha_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma & 0 & 0 & 0 \\ 0 & 0 & \sigma\alpha_0 & \sigma\alpha_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma & 0 \\ 0 & 0 & 0 & 0 & \sigma\alpha_0 & \sigma\alpha_1 \end{pmatrix}^{-1} = \text{diag}(\mathbf{B}_1, \ \frac{1}{\sigma}\mathbf{B}_1, \ \frac{1}{\sigma}\mathbf{B}_1), \tag{3.3}$$

and $\mathbf{C}(\boldsymbol{\theta}, c)$ is a 6×6 symmetric matrix as

$$\mathbf{C}(\boldsymbol{\theta}, c) = \begin{pmatrix} \Psi_{11}(c) & \Psi_{12}(c) & 0 & 0 & 0 & 0 \\ \Psi_{12}(c) & \Psi_{22}(c) & 0 & 0 & 0 & 0 \\ 0 & 0 & \Psi_{33}(c) & \Psi_{34}(c) & 0 & 0 \\ 0 & 0 & \Psi_{34}(c) & \Psi_{44}(c) & 0 & 0 \\ 0 & 0 & 0 & 0 & \Psi_{55}(c) & \Psi_{56}(c) \\ 0 & 0 & 0 & 0 & \Psi_{56}(c) & \Psi_{66}(c) \end{pmatrix} \tag{3.4}$$

$$= \begin{pmatrix} \frac{e^c}{(1+e^c)^2} & & & & & \\ c\frac{e^c}{(1+e^c)^2} & c^2\frac{e^c}{(1+e^c)^2} & & & & \\ 0 & 0 & \frac{e^c}{1+e^c} & & & \\ 0 & 0 & c\frac{e^c}{1+e^c} & c^2\frac{e^c}{1+e^c} & & \\ 0 & 0 & 0 & 0 & \frac{1}{1+e^c} & \\ 0 & 0 & 0 & 0 & c\frac{1}{1+e^c} & c^2\frac{1}{1+e^c} \end{pmatrix}. \tag{3.5}$$

We use the notation of Ψ_{ij} for $i \leq j$ where $i, j = 1, ..., 6$ to indicate the location of the element function in $\mathbf{C}$.

3.2 Complete Class Results

We will find five complete classes using a step by step procedure with five lemmas and then identify the smallest class. The complete class that we suggest first is a collection of designs having at most five support points. The remaining complete classes are composed of at most four-point designs. The conclusion of some lemma implies that of other lemma, but here we record multiple identifications of complete classes to show the existence of many complete classes under the same model.

To apply the complete class approach introduced in Section 2.2, we first choose $\mathbf{C}_{22}$ and select a maximal set of linearly independent nonconstant functions from the matrix $\mathbf{C}$ in (3.5). After fixing $\mathbf{C}_{22}$ to Ψ_k, we make a sequence of $\Psi_1, ..., \Psi_{k-1}$ using the remaining elements. Since there are many options for choosing $\mathbf{C}_{22}$ and a sequence of Ψ functions, the process is somewhat heuristic and exhaustive.

In the case of a 1-by-1 $\mathbf{C}_{22}$, only one element is selected for $\mathbf{C}_{22}$ from the diagonal elements of $\mathbf{C}$. For an m_2-by-m_2 $\mathbf{C}_{22}$, we pick m_2 elements from the diagonal of $\mathbf{C}$, and use them as the diagonal elements of $\mathbf{C}_{22}$. We then determine what the off-diagonal elements of $\mathbf{C}_{22}$ should be. Or, in a more general way, we may consider simultaneously permuting the rows and the columns of the $\mathbf{C}$ matrix to set $\mathbf{C}_{22}$. For example, the first row and first column may be rearranged to the third row and third column, respectively. We then select the lower-right m_2-by-m_2 submatrix of the resulting matrix as $\mathbf{C}_{22}$. Different permutations may give a different $\mathbf{C}_{22}$ matrix, and thus different complete classes. It also had been observed that some of permutations will not allow us to form a complete class. We note that permuting the rows and columns within the selected $\mathbf{C}_{22}$ will not change the complete class result. A judicious selection of $\mathbf{C}_{22}$, including its size, is important to this approach.

Another important issue to consider when applying the complete class approach is the order of the Ψ functions. After selecting $\mathbf{C}_{22}$, the remaining elements in $\mathbf{C}$ will be used to generate a sequence of Ψ_i for $i=1,...,k-1$. As indicated in Yang and Stufken (2012), different orders may give different results. Specifically, Yang and Stufken (2012) indicated that the selected elements (which are represented as functions of c) should all be non-constant and linearly independent of each other. They also mentioned that the element to be represented as Ψ_1, the one to be selected as Ψ_2, and so on will have an effect on the complete class result. For a given sequence of Ψ_i's, we then calculate $f_{l,t}$ as in (2.2) and obtain $F(c)$ in (2.3).

In our problem, it is noteworthy that we have $\frac{e^c}{1+e^c} + \frac{1}{1+e^c} = 1$ by combining two functions in $\mathbf{C}$. This relationship originates from the two conditional linear models for $y|z=1$ and $y|z=0$, or more precisely, from the fact that $\Pr(z=0) + \Pr(z=1) = 1$. Because of this, Theorem 2 of Yang and Stufken (2012) is not directly applicable to our design problem since we always have $f_{l,l}=0$ for some l. If we have zero of $f_{l,l}$, it hinders a complete class approach since $F(c)$ also has zero and it is unclear whether there exists a complete class due to the lack of information. However, this issue can be taken care of as follows.

We may consider, for example, a 1-by-1 $\mathbf{C}_{22}$ with $\mathbf{C}_{22}=\frac{c^2}{e^c+1}$. We then follow the above mentioned steps to calculate $F(c)$. We have $F(c)=0$ for this case regardless how we permute the order of the Ψ functions. While this result seems to suggest considering another $\mathbf{C}_{22}$, it immediately becomes clear that useful results will never be reached as long as we keep both $e^c/(1+e^c)$ and $1/(1+e^c)$ in the sequence of Ψ functions, although they both should be included if we closely follow the procedure of Yang and Stufken (2012). With a simple modification of the approach, we have the following result.

Lemma 3.2.1. *For a mixed response model, up to a change of signs of some* Ψ_l, $l=1,...,8$, $\{\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{22}, \Psi_4=\Psi_{34}, \Psi_5=\Psi_{44}, \Psi_6=\Psi_{55}, \Psi_7=\Psi_{56}\}$ *and* $\{\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{22}, \Psi_4=\Psi_{34}, \Psi_5=\Psi_{44}, \Psi_6=\Psi_{55}, \Psi_7=\Psi_{56}, \Psi_8^Q\}$ *form Chebyshev systems for every nonzero vector* Q*. Here,* $\Psi_0=1$ *and* $\Psi_8^Q=Q^\intercal\Psi_{66}Q$*. In addition, the designs with at most 5 design points, including both* A *and* B*, form a complete class in the design space* $[A,B]$.

Proof. When we consider any of the two sets using the elements, $\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{22}, \Psi_4=\Psi_{34}, \Psi_5=\Psi_{44}, \Psi_6=\Psi_{55}, \Psi_7=\Psi_{56}$, and $\Psi_8=\Psi_{66}$, it holds that $F(c)=\frac{16}{(e^c+1)^2}>0$. Our first claim then follows from Lemma 2.2.1. Then, Lemma 2 of Yang and Stufken (2012) implies that, for any set $S=\{(c_i,w_i), i=1,...,N\}$ with $N\geq 5$, we find a set $S^*=\{(c_i^*,w_i^*), i=1,...,5\}$, including A and B as points c_i^*'s, that satisfies $\sum_{i=1}^{5} w_i^*\Psi_l(c_i^*)=\sum_{i=1}^{N} w_i\Psi_l(c_i)$, $l=0,1,...,7$, and $\sum_{i=1}^{5} w_i^*\Psi_8^Q(c_i^*)>\sum_{i=1}^{N} w_i\Psi_8^Q(c_i)$ for every nonzero vector Q. Since $\Psi_0=1$ and Ψ_{55} are parts of the Ψ_l functions, we have $\sum w_i^*\Psi_{33}(c_i^*)=\sum w_i\Psi_{33}(c_i)$ using $\Psi_{33}=1-\Psi_{55}$ for Ψ_{33} discarded from a maximal set of Ψ functions. Then, we have $\sum_{i=1}^{5} w_i^*\mathbf{C_{11}}(\boldsymbol{\theta},c_i^*)=\sum_{i=1}^{N} w_i\mathbf{C_{11}}(\boldsymbol{\theta},c_i)$. It also holds that $\sum_{i=1}^{5} w_i^*\mathbf{C_{12}}(\boldsymbol{\theta},c_i^*)=\sum_{i=1}^{N} w_i\mathbf{C_{12}}(\boldsymbol{\theta},c_i)$ and $\sum_{i=1}^{5} w_i^*\mathbf{C_{22}}(\boldsymbol{\theta},c_i^*)\geq\sum_{i=1}^{N} w_i\mathbf{C_{22}}(\boldsymbol{\theta},c_i)$. We then have $\mathbf{M}(\xi^*)\succeq\mathbf{M}(\xi)$ and the conclusion follows. □

Table 3.1: $F(c)$ Values in 1-by-1 $\mathbf{C}_{22}$ Cases

Discarded Ψ_9	$\Psi_1, \Psi_2, ..., \Psi_7$	$\Psi_8 = \mathbf{C}_{22}$	$F(c)$
$\frac{e^c}{1+e^c}$	$\frac{e^c}{(e^c+1)^2}, \frac{ce^c}{(e^c+1)^2}, \frac{c^2e^c}{(e^c+1)^2}, \frac{ce^c}{e^c+1}, \frac{c^2e^c}{e^c+1}, \frac{1}{e^c+1}, \frac{c}{e^c+1}$	$\frac{c^2}{e^c+1}$	$\frac{16}{(e^c+1)^2}(>0)$
$\frac{1}{1+e^c}$	$\frac{e^c}{(e^c+1)^2}, \frac{ce^c}{(e^c+1)^2}, \frac{c^2e^c}{(e^c+1)^2}, \frac{e^c}{e^c+1}, \frac{ce^c}{e^c+1}, \frac{c^2e^c}{e^c+1}, \frac{c}{e^c+1}$	$\frac{c^2}{e^c+1}$	$\frac{16}{(e^c+1)^2}(>0)$

Table 3.2: Other Examples in 1-by-1 $\mathbf{C}_{22}$ Cases

$\Psi_1, \Psi_2, ..., \Psi_7$	$\Psi_8 = \mathbf{C}_{22}$	$F(c)$
$\frac{e^c}{(e^c+1)^2}, \frac{ce^c}{(e^c+1)^2}, \frac{c^2e^c}{(e^c+1)^2}, \frac{1}{e^c+1}, \frac{c}{e^c+1}, \frac{c^2}{e^c+1}, \frac{ce^c}{e^c+1}$	$\frac{c^2e^c}{e^c+1}$	$\frac{16e^{2c}}{(e^c+1)^2}(>0)$
$\frac{1}{e^c+1}, \frac{c}{e^c+1}, \frac{c^2}{e^c+1}, \frac{ce^c}{e^c+1}, \frac{c^2e^c}{e^c+1}, \frac{e^c}{(e^c+1)^2}, \frac{ce^c}{(e^c+1)^2}$	$\frac{c^2e^c}{(e^c+1)^2}$	$-\frac{16e^{2c}}{(e^c+1)^2(8e^c+e^{2c}+1)}(<0)$
$\frac{1}{e^c+1}, \frac{e^c}{(e^c+1)^2}, \frac{c}{e^c+1}, \frac{ce^c}{e^c+1}, \frac{ce^c}{(e^c+1)^2}, \frac{c^2}{e^c+1}, \frac{c^2e^c}{e^c+1}$	$\frac{c^2e^c}{(e^c+1)^2}$	$-\frac{16e^{2c}}{(e^c+1)^2(8e^c+e^{2c}+1)}(<0)$
$\frac{c^2e^c}{(e^c+1)^2}, \frac{ce^c}{(e^c+1)^2}, \frac{e^c}{(e^c+1)^2}, \frac{c^2e^c}{e^c+1}, \frac{ce^c}{e^c+1}, \frac{c^2}{e^c+1}, \frac{c}{e^c+1}$	$\frac{1}{e^c+1}$	$\frac{16e^{2c}}{(e^c+1)^2((e^{2c}-1)c^2-9(e^{2c}+1)c+24(e^{2c}-1))}$

Table 3.1 shows that between $\frac{e^c}{1+e^c}$ and $\frac{1}{1+e^c}$, any functions can be discarded for the same results. Also, the complete class approach is applied to the other selection of $\mathbf{C}_{22}$ as in Table 3.2. The first row in Table 3.2 gives the same result as the previous Lemma although we choose the different $\mathbf{C}_{22}$. The second and third rows show that there exists another complete class. We then have the next lemma for which we omit the proof since it is similar to that of Lemma 3.2.1.

Lemma 3.2.2. *For a mixed response model, up to a change of signs of some* Ψ_l, $l = 1, ..., 7$, $\{\Psi_0, \Psi_1 = \Psi_{55}, \Psi_2 = \Psi_{56}, \Psi_3 = \Psi_{66}, \Psi_4 = \Psi_{12}, \Psi_5 = \Psi_{22}, \Psi_6 = \Psi_{11}, \Psi_7 = \Psi_{34}\}$ *and* $\{\Psi_0, \Psi_1 = \Psi_{55}, \Psi_2 = \Psi_{56}, \Psi_3 = \Psi_{66}, \Psi_4 = \Psi_{12}, \Psi_5 = \Psi_{22}, \Psi_6 = \Psi_{11}, \Psi_7 = \Psi_{34}, \Psi_8^Q\}$ *form Chebyshev systems for every nonzero vector* Q*. Here,* $\Psi_0 = 1$ *and* $\Psi_8^Q = Q^\intercal \Psi_{22} Q$*. In addition, the designs with at most 4 design points form a complete class in the design space* $[A, B]$*.*

Furthermore, we consider the case of a 2-by-2 $\mathbf{C}_{22}$ to search for other complete classes that have designs with the smaller number of support points. We can select two types of 2-by-2 $\mathbf{C}_{22}$ as one has nonzero off-diagonal elements such as $\begin{pmatrix} \frac{e^c}{(1+e^c)^2} & c\frac{e^c}{(1+e^c)^2} \\ c\frac{e^c}{(1+e^c)^2} & c^2\frac{e^c}{(1+e^c)^2} \end{pmatrix}$, and the other has zero off-diagonal elements such as $\mathrm{diag}(\frac{e^c}{(e^c+1)^2}, \frac{c^2e^c}{e^c+1})$.

The number of Ψ is six in the former case and it is seven in the latter. By using the principal minor test, we check the sign of a 2-by-2 $f_{k,k}$ in $F(c)$ for $k=6$, or 7. For the positive definite test, we check if the $(1,1)$ component of $f_{k,k}$ and the determinant of $f_{k,k}$ is positive. For the negative definite test, the $(1,1)$ component of $f_{k,k}$ should be negative and the determinant of $f_{k,k}$ should be positive to prove the negative definiteness of the matrix.

We examined the cases with nonzero off-diagonal elements, but $F(c)$'s were not easily tractable and we did not identify a complete class. We searched all permutations (5!) for two types of $\mathbf{C}_{22}$. Since we exclude Ψ_{33} from a maximal set, we had no $\mathbf{C}_{22}$ with Ψ_{33}.

On the other hand, when $\mathbf{C}_{22}$ is a diagonal matrix, we identify complete classes and the results are found in two following lemmas. We use a similar order of a sequence to in Lemma 3.2.1.

Lemma 3.2.3. *For a mixed response model, up to a change of signs of some Ψ_l, $l=1,...,7$, $\{\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{34}, \Psi_4=\Psi_{55}, \Psi_5=\Psi_{56}, \Psi_6=\Psi_{66}\}$ and $\{\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{34}, \Psi_4=\Psi_{55}, \Psi_5=\Psi_{56}, \Psi_6=\Psi_{66}, \Psi_7^Q\}$ form Chebyshev systems for every nonzero vector Q. Here, $\Psi_0=1$ and $\Psi_7^Q=Q^\intercal\operatorname{diag}(\Psi_{22}, \Psi_{44})Q$. In addition, the designs with at most 4 design points, including B, form a complete class in the design space $[A, B]$.*

Proof. When we consider any of the two sets using the elements, $\Psi_0, \Psi_1=\Psi_{11}, \Psi_2=\Psi_{12}, \Psi_3=\Psi_{34}, \Psi_4=\Psi_{55}, \Psi_5=\Psi_{56}, \Psi_6=\Psi_{66}, \Psi_7=\operatorname{diag}(\Psi_{22}, \Psi_{44})$, we have $f_{7,7}>0$ since the (1,1) element of $f_{7,7}$, $\frac{4e^c}{(e^c+4)^2}$), is positive, and $|f_{7,7}|=\frac{16e^{2c}\left(8e^c+e^{2c}+1\right)}{(e^c+4)^4}>0$. Also, we have $\prod_{i=1}^{6} f_{i,i}=\frac{2(e^c+4)}{(e^c+1)^2}>0$. We then verify $F(c)>0$. Our first claim then follows from Lemma 2.2.1. Consequently, Lemma 2 of Yang and Stufken (2012) implies that, for any design $\xi=\{(c_i, w_i), i=1,...,N\}$ with $N\geq 4$, we can find a design $\xi^*=\{(c_i^*, w_i^*), i=$

$1, ..., 4\}$, including B as one of the points c_i^*'s that satisfies $\sum_{i=1}^{4} w_i^* \Psi_l(c_i^*) = \sum_{i=1}^{N} w_i \Psi_l(c_i)$ for $l=0, 1, ..., 6$, and $\sum_{i=1}^{4} w_i^* \Psi_7^Q(c_i^*) > \sum_{i=1}^{N} w_i \Psi_7^Q(c_i)$ for every nonzero vector Q. In a similar way to the proof of Lemma 3.2.1, we conclude that $\mathbf{M}(\xi^*) \succeq \mathbf{M}(\xi)$ and identify the complete class. □

Lemma 3.2.4. *For a mixed response model, up to a change of signs of some* Ψ_l, $l = 1, ..., 7$, $\{\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{44}, \Psi_5 = \Psi_{55}, \Psi_6 = \Psi_{56}\}$ *and* $\{\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{44}, \Psi_5 = \Psi_{55}, \Psi_6 = \Psi_{56}, \Psi_7^Q\}$ *form Chebyshev systems for every nonzero vector Q. Here,* $\Psi_0 = 1$ *and* $\Psi_7^Q = Q^\intercal \operatorname{diag}(\Psi_{22}, \Psi_{66}) Q$. *In addition, the designs with at most 4 design points, including A, form a complete class in the design space* $[A, B]$.

Proof. When we consider any of the two sets using the elements, $\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{44}, \Psi_5 = \Psi_{55}, \Psi_6 = \Psi_{56}, \Psi_7 = \operatorname{diag}(\Psi_{22}, \Psi_{66})$, we have $f_{7,7} > 0$ since the (1,1) element of $f_{7,7}$, $\frac{2e^{2c}(2e^c+1)}{(4e^c+1)^2}$, is positive, and $|f_{7,7}| = \frac{4e^{2c}(2e^c+1)^2\left(8e^c+e^{2c}+1\right)}{(4e^c+1)^4} > 0$. Also, we have $\prod_{i=1}^{6} f_{i,i} = -\frac{4(4e^c+1)}{(e^c+1)^2(2e^c+1)} < 0$. We then verify $F(c) < 0$. Our first claim then follows from Lemma 2.2.1 and the remaining proof is similar to that of Lemma 3.2.3. □

Furthermore, we set a 3-by-3 $\mathbf{C}_{22}$ with Ψ_{22}, Ψ_{44} and Ψ_{66} and find a complete class. Here we use again the principal minor test. If a matrix is a 3-by-3 diagonal matrix such as $\mathbf{A} = \operatorname{diag}(f_1, f_2, f_3)$, we determine that $\mathbf{A}$ is positive definite if $f_1 > 0$, $f_1 f_2 > 0$, and $f_1 f_2 f_3 > 0$, i.e. $f_1 > 0$, $f_2 > 0$, and $f_3 > 0$, and that $\mathbf{A}$ is negative definite if $f_1 < 0$, $f_1 f_2 > 0$, and $f_1 f_2 f_3 < 0$, i.e. $f_1 < 0$, $f_2 < 0$, and $f_3 < 0$.

Lemma 3.2.5. *For a mixed response model, up to a change of signs of some* Ψ_l, $l = 1, ..., 6$, $\{\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{55}, \Psi_5 = \Psi_{56}\}$ *and* $\{\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{55}, \Psi_5 = \Psi_{56}, \Psi_6^Q\}$ *form Chebyshev systems for every* *nonzero vector Q. Here,* $\Psi_0 = 1$ *and* $\Psi_6^Q = Q^\intercal \operatorname{diag}(\Psi_{22}, \Psi_{44}, \Psi_{66})Q$. *In addition, the designs with at most 4 design points, including A and B, form a complete class in the design space* $[A, B]$.

Proof. When we consider any of the two sets using the elements, $\Psi_0, \Psi_1 = \Psi_{11}, \Psi_2 = \Psi_{12}, \Psi_3 = \Psi_{34}, \Psi_4 = \Psi_{55}, \Psi_5 = \Psi_{56}, \Psi_6 = \operatorname{diag}(\Psi_{22}, \Psi_{44}, \Psi_{66})$, we have $f_{6,6} > 0$ since the (1,1) element of $f_{6,6}$ is $\frac{e^c}{2} > 0$, the (2,2) element is $\frac{e^{2c}}{4} + e^{3c} > 0$, and the (3,3) element is $\frac{1}{8} e^{2c}(17e^c + 4e^{2c} + 4) > 0$. Also, we have $\prod_{i=1}^{5} f_{i,i} = \frac{4}{(e^c+1)^2} > 0$. We then verify $F(c) > 0$. Our first claim then follows from Lemma 2.2.1 and the remaining proof is similar to that of Lemma 3.2.3. □

Table 3.3: The Obtained Complete Classes for a Mixed Response Model

Complete classes	Design
Complete class 1	$\begin{pmatrix} A & c_2 & c_3 & c_4 & B \\ w_1 & w_2 & w_3 & w_4 & w_5 \end{pmatrix}$
Complete class 2	$\begin{pmatrix} c_1 & c_2 & c_3 & c_4 \\ w_1 & w_2 & w_3 & w_4 \end{pmatrix}$
Complete class 3	$\begin{pmatrix} c_1 & c_2 & c_3 & B \\ c_1 & c_2 & c_3 & c_4 \end{pmatrix}$
Complete class 4	$\begin{pmatrix} A & c_2 & c_3 & c_4 \\ w_1 & w_2 & w_3 & w_4 \end{pmatrix}$
Complete class 5	$\begin{pmatrix} A & c_2 & c_3 & B \\ w_1 & w_2 & w_3 & w_4 \end{pmatrix}$

When we denote the five complete classes that we identified in the previous lemmas as $\Xi_1, \Xi_2, \Xi_3, \Xi_4$, and Ξ_5, respectively, we see that $\Xi_5 \subset \Xi_4 \subset \Xi_2 \subset \Xi_1$ and $\Xi_5 \subset \Xi_3 \subset \Xi_2 \subset \Xi_1$ in Table 3.3. Based on our results, we use the complete class Ξ_5 as a collection of candidate designs for a search for optimal designs under a given optimality criterion in the next two chapters. We have our first main result of the study.

Theorem 3.2.6. *For any design* $\xi = \{(c_i, w_i), i = 1, ..., n\}$ *in the design space* $[A, B]$ *for a mixed response model, there exists a complete class of designs that have at most four design points including both A and B.*

3.3 Estimability and Number of Support Points

In many cases, we normally would need the number of support points of a design to be at least as large as the number of parameters of interest to make all parameters estimable. A design is sometimes called saturated when the number of support points is the same as the number of parameters in the model (Dette and Melas, 2012). However, for a mixed response model, it is not necessary to have saturated designs to make parameters estimable. In other words, we do not need at least six support points for having a linear, unbiased estimator of the six mean parameters in our model. The complete class results in the previous section provide evidence of this, and in this section, we give another explanation. At first, we remind that the Fisher information matrix is as follows (see Section 3.1).

$$\mathbf{M}(\xi,\boldsymbol{\theta})=\begin{bmatrix}\mathbf{F}^\intercal\mathbf{W}\mathbf{P}(\mathbf{I}-\mathbf{P})\mathbf{F} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F}^\intercal\mathbf{W}\mathbf{P}\mathbf{F} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F}^\intercal\mathbf{W}(\mathbf{I}-\mathbf{P})\mathbf{F}\end{bmatrix}, \text{where } \mathbf{F}=\begin{bmatrix}\mathbf{f}(x_1)^\intercal\\ \mathbf{f}(x_2)^\intercal\\ \cdots\\ \mathbf{f}(x_n)^\intercal\end{bmatrix}$$

and $\mathbf{f}(x_i)=[1,x_i]^\intercal$. Based on the following theorem, we understand that all the mean parameters can be estimated even when the number of the support points of the selected design is less than the number of parameters ($=6$).

Theorem 3.3.1. *For any design $\xi=\{(x_i,w_i),i=1,...,n,w_i>0\}$, if $0<P(z_i=1)=\pi_i<1$ for all i, and $\mathbf{F}$ is a full column rank matrix, then $\mathbf{M}\succ 0$.*

Proof. We observe that $\mathbf{WP}(\mathbf{I}-\mathbf{P})$, $\mathbf{WP}$, and $\mathbf{W}(\mathbf{I}-\mathbf{P})$ are positive definite since $w_i>0$, $\pi_i>0$ and $1-\pi_i>0$. When $\mathbf{F}$ is full column rank, all block diagonal matrices of $\mathbf{M}$ are positive definite. Therefore, $\mathbf{M}\succ 0$. □

Based on the above theorem, we know that six parameters are estimable in a mixed response model when using designs with at least two support points as long as $\mathbf{F}$ is in full column rank. Unless x_i goes to ∞ or $-\infty$, it is true that $\pi_i>0$ and $1-\pi_i>0$ since $\pi_i=\frac{\exp(\alpha_0+\alpha_1 x_i)}{1+\exp(\alpha_0+\alpha_1 x_i)}$ is not zero or one. Moreover, when we define $\mathbf{M}=\text{diag}(\mathbf{M}_1,\frac{1}{\sigma^2}\mathbf{M}_2,\frac{1}{\sigma^2}\mathbf{M}_3)$ where $\mathbf{M}_1=\mathbf{F}^\intercal\mathbf{W}\mathbf{P}(\mathbf{I}-\mathbf{P})\mathbf{F},\mathbf{M}_2=\mathbf{F}^\intercal\mathbf{W}\mathbf{P}\mathbf{F}$, and

$\mathbf{M}_3 = \mathbf{F}^\intercal \mathbf{W}(\mathbf{I} - \mathbf{P})\mathbf{F}$, we see that if $\mathbf{F}$ has full column rank, it holds that $\mathbf{M}_2 \succ \mathbf{M}_1$ and $\mathbf{M}_3 \succ \mathbf{M}_1$ since $\mathbf{M}_2 - \mathbf{M}_1 = \mathbf{F}^\intercal \mathbf{W} \mathbf{P}^2 \mathbf{F} \succ 0$ and similarly, $\mathbf{M}_3 - \mathbf{M}_1 \succ 0$.

3.4 Discussion

Our study shows that the complete class approach is helpful in tackling a locally optimal design problem for the mixed response model that we consider. We found five complete classes and, in search of an optimal design, selected one complete class that possessed designs with at most 4 design points, including A and B, in the design space $[A, B]$. Instead of a countably many number of n, we considered $n = 4$ as the number of design points which gave a small candidate set of designs. The number was clearly less than $p(p+1)/2 = 6(7)/2 = 21$ guaranteed by Carathéodory's theorem, and remarkably, it also was less than the total number of unknown parameters in the model. In addition, our case showed that a complete class approach can be successfully adapted to a nonlinear model for bivariate responses.

Our first two lemmas confirmed that, with a 1-by-1 $\mathbf{C}_{22}$, there exist complete classes which have designs with at most $n = 4$ design points, and with $n = 5$ design points including two bound points. In the cases of a 2-by-2 $\mathbf{C}_{22}$, we found two complete classes with designs that have at most $n = 4$ support points including either $c_1 = A$ or $c_4 = B$. Lastly, using a 3-by-3 $\mathbf{C}_{22}$, we found a complete class with at most $n = 4$ design points including both A and B in the design space $[A, B]$.

With a symbolic software, an exhaustive search was tried for some cases that consider every possible permutation of Ψ_i's $(i \neq k)$. The number of candidates for the pair of Chebyshev systems is (the number of selection options in $\mathbf{C}_{22}$) $\times$ (the number of permutations in $\Psi_1, ..., \Psi_{k-1}$). We found that in many cases, differently permuted sequences shared the same $F(c)$. While an exhaustive search is available, it is true that we often detected the successful results after few trials. In our case, we already knew that a simple logistic regression has a complete class from Yang and Stufken (2009), so we started to examine our case considering a possible $\mathbf{C}_{22}$ in their results. On the other hand, we observed that the smaller complete classes could be formed by larger complete classes previously identified. One of complete classes of the 2-by-2 $\mathbf{C}_{22}$ had the $\Psi_k = \text{diag}(\Psi_3, \Psi_9)$ where each diagonal element was 1-by-1 $\mathbf{C}_{22}$ in Lemmas 3.2.1 and 3.2.2. Furthermore, we found the case of a 3-by-3 $\mathbf{C}_{22}$ of which the elements were used for smaller size of $\mathbf{C}_{22}$.

For a mixed response model, we always had $f_{l,l} = 0$ for some l if we simply followed the procedure of Yang and Stufken (2012). This was because we had $\Psi_{33} + \Psi_{55} = 1$ and if both Ψ_{ij} were in a sequence of Ψ, we had $F(c) = 0$. From such a relationship, we saw that Ψ'_{33} and Ψ'_{55} were linearly dependent as $\Psi'_{33} + \Psi'_{55} = 0$, i.e. $\Psi'_{33} = -\Psi'_{55}$. Since the derivative of the $\Psi'_{33}/\Psi'_{55} = -1$ was zero, we had $f_{l,l} = 0$ for some l. As an another example, if Ψ'_4 and Ψ'_9 are linearly dependent, then, we have $f_{9,5} = 0$. In Figure 2.1, $f_{4,4}$ has Ψ'_4 on the uppermost location. Similarly, $f_{9,4}$ has the same form replacing Ψ'_4 with Ψ'_9. Then, if $\Psi'_9 = m\Psi'_4$, we obtain $f_{9,5} = f_{9,4}/f_{4,4} = 0$ during differentiation. When we use a model that we suggest for mixed responses, it should be considered that the sum of two $\Psi_{i,i}$ is one.

On the other hand, we found the range of design points as $2 \leq n \leq 4$. In many cases of the complete class results for other models, the number of design points is at most m where m is the number of parameters. Also, we normally consider that the number of support points is at least m for estimabiltiy. Combining two facts, we mostly find saturated designs with m support points under some specific criterion. However, in our case, we had at most 4-point designs, and at least 2-point designs instead of 6-point designs. In the numerical results from the next two chapters, we will see that optimal designs have $n = 2, 3$ or 4 design points. In a sense, we found 'supersaturated' optimal designs that give non-singular information matrices.

Chapter 4

NUMERICAL RESULTS 1: D-OPTIMAL DESIGNS

4.1 Optimization for the D-optimality

Based on the complete class results in the previous chapter, we will now search for D-optimal designs using a computational approach. We use a mathematical programming, specifically a nonlinear constrained algorithm of the fmincon solver in MATLAB. We first formulate the D-optimality criterion and the constraints relating to our design problem in accordance with the standard form of mathematical programming.

The D-optimality criterion was introduced in Chapter 2 as $\Phi_D = |\mathbf{M}^{-1}|$ or $\Phi_p = |\mathbf{M}^{-1}|^{1/6}$ when $p=0$. Also, we can consider $\Phi_D^0 = \log|\mathbf{M}^{-1}|$. Since three criteria share the same ordering between candidate designs, best designs are the same. Moreover, it might be desirable to use the criterion such as $\Phi_p = |\mathbf{M}^{-1}|^{1/6}$ or $\Phi_D^0 = \log|\mathbf{M}^{-1}|$ because it is a convex function having a minimum. We here adopt Φ_D^0 to have an additive form of three block matrices of the information matrix without exponent parts. Working with the decomposition of the information matrix as $\mathbf{M}(\xi, \boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}, \sigma) \left(\sum\limits_{i=1}^{n} w_i \mathbf{C}(\boldsymbol{\theta}, c_i) \right)$ $(\mathbf{B}(\boldsymbol{\theta}, \sigma))^\intercal$, we express the D-optimality criterion as

$$\Phi_D^0 = -\log|\mathbf{B}\mathbf{C}^*\mathbf{B}^\intercal| = -\log|\mathbf{B}|^2|\mathbf{C}^*| = -\log|\mathbf{C}^*| - \log|\mathbf{B}_1|^6 - \log\sigma^4 \tag{4.1}$$

where $\mathbf{C}^* = \sum_{i=1}^{n} w_i \mathbf{C}(\boldsymbol{\theta}, c_i)$ and $\mathbf{B} = \text{diag}(\mathbf{B}_1, \frac{1}{\sigma}\mathbf{B}_1, \frac{1}{\sigma}\mathbf{B}_1)$. Here, the value of $|\mathbf{B}_1|$ is a constant since the element of the matrix $\mathbf{B}_1$ is a guessed value of an unknown parameter α_0 or α_1, or a constant 0 or 1 as shown in (3.3). Also, the positive constant σ^2 does not affect the optimization procedure as long as we hold the same value of σ^2 throughout optimizing. Therefore, the D-optimality criterion Φ_D^0 is invariant with $\mathbf{B}$. We can find optimal designs by comparing values of $-\log|\mathbf{C}^*|$ instead of values of $-\log|\mathbf{M}|$ among candidate designs. Consequently, when we denote $\mathbf{C}^* = \mathrm{diag}\,(\mathbf{C}_1^*, \mathbf{C}_2^*, \mathbf{C}_3^*)$, the original minimization problem is reduced to the minimization of $-(\log|\mathbf{C}_1^*| + \log|\mathbf{C}_2^*| + \log|\mathbf{C}_3^*|)$. Within the complete class 5 identified in the previous chapter, we want to find a D-optimal design. For a design $\xi = \{(c_i, w_i), i=1,2,3,4, c_1 = A, c_4 = B\}$ in $[A, B]$, a optimization problem can be formulated as:

$$\begin{aligned} \underset{\boldsymbol{\xi}^{\mathbf{0}}}{\text{minimize}} \quad & -(\log|\mathbf{C}_1^*(\boldsymbol{\xi}^0)| + \log|\mathbf{C}_2^*(\boldsymbol{\xi}^0)| + \log|\mathbf{C}_3^*(\boldsymbol{\xi}^0)|) \\ \text{subject to} \quad & \sum w_i = 1, \quad w_i \geq 0, \text{ for } i=1,...,4 \text{ and } A \leq c_i \leq B \quad \text{for } i=2,3. \end{aligned}$$

where a vector of decision variables is $\boldsymbol{\xi}^{\mathbf{0}} = (c_2, c_3, w_1, w_2, w_3, w_4)$. We set initial values of $\boldsymbol{\xi}^{\mathbf{0}}$ as the 30th and 60th percentile points of $[A, B]$ for c_2 and c_3 and a uniformly equal weight 0.25 for w_1, w_2, w_3, and w_4. The fimincon solver requires initial values for an iterative method and the choice of initial values can thus impact outcomes of our search. Trying several initial values is recommended for the solver we use here.

The solver that we consider has five algorithms: interior point algorithm, sequential quadratic programming (SQP), sequential quadratic programming legacy (SQP-legacy), active-set algorithm, and trust-region-reflective algorithm (TRRA). To choose the most appropriate algorithm for solving our design problem, we compare the efficiency of algorithms. While we need both, the TRRA does not accommodate a bound constraint and a linear equality constraint together. The SQP-legacy algorithm is similar to the SQP method. We then exclude two algorithms from the comparison.

We set three scenarios. For each algorithm, we optimize an at most 4-point design with two fixed points in the different design spaces including $[-5, 5]$, $[-10, 10]$, and $[-100, 100]$. Computing time and the number of iteration are checked ten times with

Table 4.1: Computing Time and Number of Iterations by Algorithms (sec.)

Design space	Algorithm	Design for c				Mean of run-time	Std[a]of run-time	# of itrn[b]
$[-5, 5]$	Interior[c]	-5	-1.1067	1.1067	5	1.8207	0.0148	12
		0.1846	0.3154	0.3154	0.1846			
	SQP[d]	-5	-1.1067	1.1067	5	1.7986	0.0049	13
		0.1846	0.3154	0.3154	0.1846			
	Active-set	-5	-1.1063	1.1069	5	1.8099	0.0376	10
		0.1846	0.3154	0.3153	0.1846			
$[-10, 10]$	Interior	-10	-1.3218	1.3218	10	1.8481	0.106	13
		0.1628	0.3372	0.3372	0.1628			
	SQP	-10	-1.3218	1.3218	10	1.8221	0.0232	16
		0.1628	0.3372	0.3372	0.1628			
	Active-set	-10	-1.2966	1.4448	10	1.832	0.0062	19
		0.1444	0.4066	0.3974	0.0516			
$[-100, 100]$	interior	-100	1.5134	-1.5134	100	1.9641	0.1538	29
		0.1666	0.3334	0.3334	0.1666			
	SQP	-100	1.5134	-1.5134	100	1.8768	0.0195	26
		0.1666	0.3334	0.3334	0.1666			
	Active-set	-	-	-	-	fail	-	-
		-	-	-	-			

[a] standard deviation
[b] iteration
[c] interior point algorithm
[d] sequential quadratic programming

an initial point $\boldsymbol{\xi}_0^0 = (a, b, 0.25, 0.25, 0.25, 0.25)$ where a is the 30th quantile point and b is the 60th quantile point in a design space of $[A, B]$. Table 4.1 shows that we get the same solutions with the interior point algorithm and the SQP algorithm. Also, for each algorithm, we observe that we have the same solution over ten instances of simulation. In the case of the active-set algorithm, we do not have the answer in $[-100, 100]$ and the obtained results are different from those of the two other algorithms in $[-5, 5]$ and $[-10, 10]$. The difference does not mean a wrong answer, but we exclude the active-set algorithm because of the failure of optimization in $[-100, 100]$.

The interior-point algorithm and the SQP algorithm seem comparable to our design problem. Based on the mean of run-times in seconds, the SQP algorithm is slightly faster than the interior point algorithm. Regarding the number of iterations, the number of the interior point algorithm is smaller than the number of the SQP in $[-5, 5]$ and $[-10, 10]$ while it is larger in $[-100, 100]$. In general, they produce the same solutions and the computation time is trivial. We choose the SQP algorithm since the mean and standard deviation of the run-time is smaller than those of the interior point algorithm.

We implement the search on a computer that has a 3.5 GHz 6-core processor with 4 GB of RAM. We present the '$fval$ (function value)' in some of our results, which indicates the value of the objective function, i.e. $-(\log|\mathbf{C}_1^*| + \log|\mathbf{C}_2^*| + \log|\mathbf{C}_3^*|)$ at the obtained solution $\boldsymbol{\xi}^0$. Small values are desirable since we set our optimization problem as a minimization problem when we use the fmincon solver.

We now search for D-optimal designs using the SQP algorithm. Since the design space is arbitrary, we study several cases with different design spaces. We set a design space by two standards. The first one is the size of the design space, and the second one is its central location. For cases with variable sizes of design space, we only consider symmetric domains about zero. For cases with different central locations, we fix the length of the space to 20. Our results are summarized in Tables 4.2 and 4.3. Note that all the designs are searched and reported in terms of the represented design space of c, instead of the original design space of x.

Table 4.2: D-optimal Designs for c by the Size of Design Spaces

Design space	$fval$	Design				# of points	w_2/w_1
[-1, 1]	6.51	-1			1	2	
		0.5			0.5		1
[-2, 2]	4.78	-2	0.000		2	3	
		0.431	0.138		0.431		0.320
[-5, 5]	3.14	-5	-1.107	1.107	5	4	
		0.185	0.315	0.315	0.185		1.709
[-10, 10]	0.64	-10	-1.322	1.322	10	4	
		0.163	0.337	0.337	0.163		2.071
[-50, 50]	-5.99	-50	1.485	-1.485	50	4	
		0.166	0.334	0.334	0.166		2.003
[-100, 100]	-8.80	-100	1.513	-1.513	100	4	
		0.167	0.333	0.333	0.167		2.001

Table 4.3: D-optimal Designs for c by the Location of Design Spaces

Design space	$fval$	Design				# of points	w_2/w_1	w_3/w_1
[-25,-5]	20.82	-25	-0.696		-5	3		
		0.164	0.408		0.428		2.48	2.61
[-20, 0]	3.60	-20	-2.311		0	3		
		0.165	0.435		0.400		2.65	2.43
[-15, 5]	1.80	-15	-0.591		5	3		
		0.166	0.498		0.336		2.99	2.02
[-10, 10]	0.64	-10	-1.322	1.322	10	4		
		0.163	0.337	0.337	0.163		2.07	2.07
[-5, 15]	1.80	-5	0.591		15	3		
		0.336	0.498		0.166		1.48	0.50
[0, 20]	3.60	0	2.311		20	3		
		0.400	0.435		0.165		1.09	0.41
[5, 25]	20.82	5	6.969		25	3		
		0.428	0.408		0.164		0.95	0.38

As shown in Table 4.2, we obtain a 2-, 3-, or 4-point design according to the size of a design space. We set design spaces to $[-1,1]$, $[-2,2]$, $[-5,5]$, $[-10,10]$, $[-50,50]$, and $[-100,100]$. We note that we expect at most four design points with the complete class that we derived. When the space is $[-1,1]$, there is a 2-point design of $\{(-1,0.5),(1,0.5)\}$. The points are the two extreme points of the design space, and their weights are equal. In $[-2,2]$, the 3-point design appears as $\{(-2,0.43),(0,0.14),(2,0.43)\}$. The center and the two extreme points of the design space are included as support points, and the three corresponding weights are not all equal; the first and third weights are the same. For other design spaces, 4-point designs are obtained: the two outer points and two inner points. The inner points are symmetric about zero. The weights of the outer points are equal and so are those for the inner points. When we move the design space from $[-10,10]$ to $[-100,100]$, the inner points change from ±1.322 to ±1.513. It also can be seen that the ratio of the weights between the inner and the outer points is rather consistent for the last three designs by the values of w_2/w_1 in Table 4.2. The ratio of weights is close to 1:2:2:1.

Table 4.4: D-optimal Designs for c by the Size of Design Spaces (detail)

Design space [a]		Design				# of points	Inner points		w_1	w_2/w_1
-1	1	-1			1	2				
		0.5			0.5				0.5	1
-1.5	1.5	-1.5			1.5	2				
		0.5			0.5				0.5	1
-1.8	1.8	-1.8	0		1.8	3	0			
		0.4873	0.0255		0.4873				0.49	0.052
-2	2	-2	0		2	3	0			
		0.431	0.138		0.431				0.43	0.320
-2.5	2.5	-2.5	0		2.5	3	0			
		0.347	0.306		0.347				0.35	0.882
-2.7	2.7	-2.7	0		2.7	3	0			
		0.327	0.346		0.327				0.33	1.058
-2.9	2.9	-2.9	0		2.9	3	0			
		0.312	0.376		0.312				0.31	1.206
-2.95	2.95	-2.95	-0.184	0.184	2.95	4	-0.184	0.184		
		0.306	0.194	0.194	0.306				0.31	0.632
-3	3	-3	-0.302	0.302	3	4	-0.302	0.302		
		0.3	0.200	0.200	0.3				0.30	0.666
-3.2	3.2	-3.2	-0.544	0.544	3.2	4	-0.544	0.544		
		0.278	0.222	0.222	0.278				0.28	0.800
-3.5	3.5	-3.5	-0.739	0.739	3.5	4	-0.739	0.739		
		0.25	0.250	0.250	0.25				0.25	0.999
-4	4	-4	-0.923	0.923	4	4	-0.923	0.923		
		0.218	0.282	0.282	0.218				0.22	1.293
-5	5	-5	-1.107	1.107	5	4	-1.107	1.107		
		0.185	0.315	0.315	0.185				0.185	1.709
-6	6	-6	-1.197	1.197	6	4	-1.197	1.197		
		0.171	0.329	0.329	0.171				0.171	1.929
-7	7	-7	-1.248	1.248	7	4	-1.248	1.248		
		0.165	0.335	0.335	0.165				0.165	2.030
-8	8	-8	-1.281	1.281	8	4	-1.281	1.281		
		0.163	0.337	0.337	0.163				0.163	2.067
-9	9	-9	-1.304	1.304	9	4	-1.304	1.304		
		0.163	0.337	0.337	0.163				0.163	2.075
-10	10	-10	-1.322	1.322	10	4	-1.322	1.322		
		0.163	0.337	0.337	0.163				0.163	2.071
-50	50	-50	-1.485	1.485	50	4	-1.485	1.485		
		0.167	0.334	0.334	0.167				0.167	2.003
-100	100	-100	-1.513	1.513	100	4	-1.513	1.513		
		0.167	0.333	0.333	0.167				0.167	2.001
-150	150	-150	-1.523	1.523	150	4	-1.533	1.533		
		0.1666	0.333	0.333	0.167				0.167	2.001

[a] From now, a design space is expressed by two endpoints.

In Table 4.3, we change the central location of the design space. We set the symmetric design space $[-10, 10]$ as baseline and move it in a positive or negative direction. Except for the case of $[-10, 10]$, we obtain asymmetric 3-point designs under the D-optimality criterion. For example, in the design space of $[-20, 0]$, we obtain the design points of -20, -2.311, and 0 with weights of 0.165, 0.435, and 0.4, respectively. The second point -2.311 is not the midpoint of the two end points -20 and 0. The weights are not equal as the values of w_2/w_1 and w_3/w_1 indicate. The minimum value 0.642 of '$fval$' is achieved when a design space is symmetric as $[-10, 10]$.

We now closely observe D-optimal designs in a symmetric domain with various design spaces in Table 4.4 by considering additional symmetric design spaces from $[-1, 1]$ to $[-150, 150]$.

From Table 4.4, we see that the change in the number of support points is gradual. It also can be seen that every 2-point design has the two boundary points of the design space as support points, each with weights of 0.5. This is observed when the design space is $[-1, 1]$ or $[-1.5, 1.5]$. Every 3-point design has zero in the middle. They also are symmetric designs. When the design space is enlarged to $[-2.95, 2.95]$, a 4-point design appears. The support points are symmetric about zero. Weights are balanced as the ratio of about 3:2:2:3. By widening the design space, we observe that the ratio of weights moves gradually to 1:2:2:1.

If we look at the 'inner points' columns in Table 4.4, inner points are moving in between ± 0.184 and ± 1.533 when we have a 4-point design. In addition, the weights for the outer support points decrease as the size of the design space increases according to the 'w_1' column. It decreases from 0.5 to 0.167 when the design space changes from $[-1, 1]$ to $[-50, 50]$.

4.2 Verification of the D-optimality

Under the D-optimality criterion, we obtained designs using mathematical programming in the previous section. However, it is not guaranteed that the obtained designs are D-optimal if we completely rely on the capability of the fmincon solver for solving minimization problems. With the general equivalence theorem described in Chapter 2, we validate the D-optimality of the obtained designs. In Chapter 2, we derived the basic equation for this verification as $m - d(c,\xi) \geq 0$ where m is the number of parameters, and $d(c,\xi) = \text{tr } \mathbf{C}^{*-1}(\xi)\mathbf{C}^*(\bar{\xi})$; see (2.8). Based on this, we provide a detailed derivation of $d(c,\xi)$ for the mixed response model.

We first specify $\mathbf{C}^{*-1}(\xi)$ and $\mathbf{C}^*(\bar{\xi})$ using a model vector $\mathbf{f}(c_i)$. For a block diagonal matrix, we denote $\mathbf{C}^* = \text{diag}\{\mathbf{C}_1^*, \mathbf{C}_2^*, \mathbf{C}_3^*\} = \text{diag }\{\sum w_i \mathbf{f}(c_i)\frac{e^{c_i}}{(1+e^{c_i})^2}\mathbf{f}(c_i)^\intercal, \sum w_i \mathbf{f}(c_i)\frac{e^{c_i}}{(1+e^{c_i})}$ $\mathbf{f}(c_i)^\intercal, \sum w_i \mathbf{f}(c_i)\frac{1}{(1+e^{c_i})}\mathbf{f}(c_i)^\intercal\}$. Also, for a measure $\bar{\xi}$ with a unit mass at c, we have $\mathbf{C}^*(\bar{\xi}) = \text{diag }\{\mathbf{f}(c)\frac{e^{c}}{(1+e^{c})^2}\mathbf{f}(c)^\intercal, \mathbf{f}(c)\frac{e^{c}}{(1+e^{c})}\mathbf{f}(c)^\intercal, \mathbf{f}(c)\frac{1}{(1+e^{c})}\mathbf{f}(c)^\intercal\}$.

Then, for the D-optimality, we see that $d(c,\xi) = \text{tr }\mathbf{C}_1^{*-1}(\xi)\mathbf{C}_1^*(\bar{\xi}) + \text{tr }\mathbf{C}_2^{*-1}(\xi)\mathbf{C}_2^*(\bar{\xi}) +$ $\text{tr }\mathbf{C}_3^{*-1}(\xi)\mathbf{C}_3^*(\bar{\xi})$. From the first term, it holds that $\text{tr }\mathbf{C}_1^{*-1}(\xi)\mathbf{C}_1^*(\bar{\xi}) = \text{tr }\mathbf{C}_1^*(\bar{\xi})\mathbf{C}_1^{*-1}(\xi) =$ $\text{tr }\mathbf{f}(c)\pi(1-\pi)\mathbf{f}^\intercal(c)\mathbf{C}_1^{*-1}(\xi) = \pi(1-\pi)\,\text{tr}[\mathbf{f}^\intercal(c)\mathbf{C}_1^{*-1}(\xi)\mathbf{f}(c)] = \pi(1-\pi)\mathbf{f}^\intercal(c)\mathbf{C}_1^{*-1}(\xi)\mathbf{f}(c)$. Then, we get the following proposition.

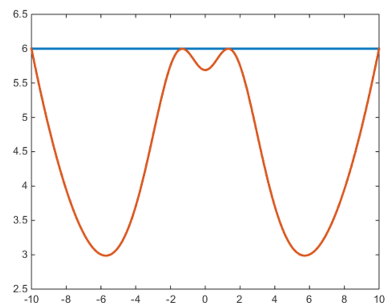
Proposition 4.2.1. *For a mixed responses model, we verify the D-optimality of an obtained design ξ if it holds that*

$$d(c,\xi) = \frac{e^c}{(1+e^c)^2}\mathbf{f}(c)^\intercal\mathbf{C}_1^{*-1}(\xi)\mathbf{f}(c)^\intercal + \frac{e^c}{(1+e^c)}\mathbf{f}(c)^\intercal\mathbf{C}_2^{*-1}(\xi)\mathbf{f}(c)^\intercal + \frac{1}{(1+e^c)}\mathbf{f}(c)^\intercal\mathbf{C}_3^{*-1}(\xi)\mathbf{f}(c)^\intercal$$

is equal to or less than 6 *based on the general equivalence theorem for all c in $[A,B]$. For a D-optimal design $\xi^* = \{(c_i^*, w_i^*), i=1,...,4, w_i^* \geq 0\}$, we have $d(c^*,\xi^*) = 6$.*

Using Proposition 4.2.1, we check the D-optimality of an obtained design ξ. $\mathbf{C}_1^*, \mathbf{C}_2^*$, and $\mathbf{C}_3^*$ are calculated by using the values of the obtained design. When we set $\mathbf{C}_k^{*-1} = ((c_{kij}^*))(ij=1,2)$ for $k=1,2,3$, we see that $\text{tr }\mathbf{C}_1^{*-1}(\xi)\mathbf{C}_1^*(\bar{\xi}) = \text{tr }\mathbf{C}_1^*(\bar{\xi})\mathbf{C}_1^{*-1}(\xi)$

Figure 4.1: D-optimality Verification of the Obtained Design in $[-10, 10]$ by the General Equivalence Theorem

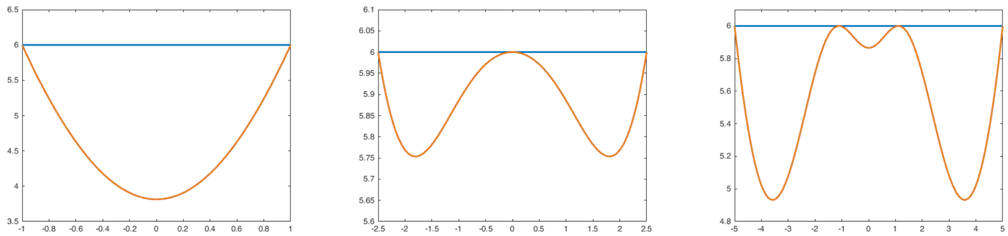


$= \pi(1-\pi)[1 \;\; c] \begin{bmatrix} c^*_{111} & c^*_{112} \\ c^*_{121} & c^*_{122} \end{bmatrix} [1 \;\; c]^\intercal = \pi(1-\pi)(c^*_{111} + (c^*_{112} + c^*_{121})c + c^*_{122}c^2)$. From here, we know that the values of c^*_{kij}'s determine the equation. For each design, we draw a plot of the GET verification.

For example, for the obtained design $\xi^* = \{(-10, 0.163), (-1.322, 0.337), (1.322, 0.337), (10, 0.163)\}$ in $[-10, 10]$ in the previous section, a function $d(c, \xi)$ is derived as $d(c, \xi) = \frac{e^c}{e^c+1}(0.1025c^2 - 0.7732c + 3.4582) + \frac{1}{e^c+1}(0.1025c^2 + 0.7732c + 3.4582) + \frac{e^c}{(e^c+1)^2}(5.0682c^2 + 8.9206)$, where $\mathbf{f}(c) = [1 \;\; c]^\intercal$. Using the derived equation, we draw the reference line $y = 6$ (the blue line in Figure 4.1) and the curve $y = d(c, \xi)$ (the orange curve in Figure 4.1). We validate that our obtained design is D-optimal after seeing that $d(c, \xi) \leq 6$ for all $c \in [-10, 10]$ in Figure 4.1. As the general equivalence theorem tells us, we observe that the tangent points are exactly the four support points we obtained.

Also, we draw the plots of the GET verification for selected 2-point, 3-point, and another 4-point design in Figure 4.2. They are the obtained designs when the design spaces are $[-1, 1]$, $[-2.5, 2.5]$, and $[-5, 5]$, respectively. In the figure, we see that all three designs are validated as D-optimal by the general equivalence theorem. Every orange curve is below the blue reference line.

Figure 4.2: D-optimality Verification of Three Different Designs



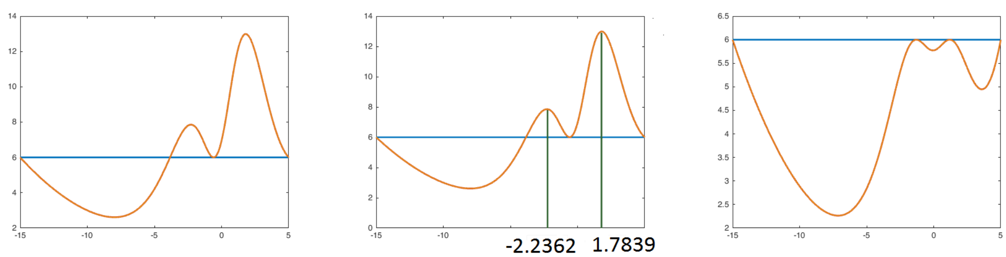
(a) design in $[-1, 1]$ (b) design in $[-2.5, 2.5]$ (c) design in $[-5, 5]$

Table 4.5: D-optimality Verification of Selected Designs by the General Equivalence Theorem

Design space		$fval$	Design				GET verification
-1	1	6.506	-1			1	sucess
			0.500			0.500	
-5	5	3.143	-5	-1.107	1.107	5	sucess
			0.185	0.315	0.315	0.185	
-10	10	0.642	-10	-1.322	1.322	10	sucess
			0.163	0.337	0.337	0.163	
-25	-5	20.823	-25	-6.969		-5	success
			0.164	0.408		0.428	
-20	0	3.604	-20	-2.311		0	success
			0.165	0.435		0.400	
-15	5	1.546	-15	-0.591		5	fail
			0.166	0.498		0.336	
-10	10	0.642	-10	-1.322	1.322	10	success
			0.163	0.337	0.337	0.163	
-5	15	1.798	-5	0.591		15	fail
			0.336	0.498		0.166	
0	20	3.604	0	2.311		20	success
			0.400	0.435		0.165	
5	25	20.823	5	6.969		25	success
			0.428	0.408		0.164	

Furthermore, we check the optimality for additional designs as shown in Table 4.5. The symmetric designs of 2 points, 3 points, and 4 points are verified to be D-optimal. In asymmetric domains, the D-optimality of the obtained designs is secured except for the two cases where the design spaces are, respectively, $[-15, 5]$ and $[-5, 15]$. The failure of the GET verification indicates that solutions from the fmincon solver are not always D-optimal.

Figure 4.3: D-optimality Verification of an Asymmetric Design in $[-15, 5]$



(a) design in $[-15, 5]$ (b) design in $[-15, 5]$ (c) optimal design in $[-15, 5]$

To remedy this failure, we change the initial points of optimization based on the obtained design. As shown in Figure 4.3, the first plot discloses the failure of the verification of the design for which we searched in $[-15, 5]$. We thus find two points, $c = -2.2362$ and 1.7839, that give the two local maximums of $d(c, \xi)$ and then use the two points as initial values for c_2 and c_3. Originally we used -9 and -3 which are the 30th and 60th quantile points in $[-15, 5]$. After we submit the new form with a changed initial value, we obtain a new design such as $\{(-15, 0.165), (-1.298, 0.383), (1.196, 0.271), (5, 0.182)\}$. This design is verified as D-optimal by the GET as shown in the third plot in Figure 4.3. We use the same procedure for the case of $[-5, 15]$. We then obtain a D-optimal design as $\{(-5, 0.182), (-1.196, 0.271), (1.298, 0.383), (15, 0.165)\}$. In asymmetric domains, D-optimal designs are not symmetric and the weights are unequal.

One way of comparing two designs is via the relative D-efficiency. The relative D-efficiency D_{rel} of ξ_1 to ξ_2 is

$$D_{rel}=\left(\frac{|\mathbf{M}(\xi_1)|}{|\mathbf{M}(\xi_2)|}\right)^{1/m}=\left(\frac{|\mathbf{C}^*(\xi_1)|}{|\mathbf{C}^*(\xi_2)|}\right)^{1/m}=\left(\frac{|\mathbf{C}_1^*(\xi_1)\mathbf{C}_2^*(\xi_1)\mathbf{C}_3^*(\xi_1)|}{|\mathbf{C}_1^*(\xi_2)\mathbf{C}_2^*(\xi_2)\mathbf{C}_3^*(\xi_2)|}\right)^{1/m}.$$

Using the values of $fval=-\log|\mathbf{C}_1^*\mathbf{C}_2^*\mathbf{C}_3^*|$, we calculate D_{rel} as

$$D_{rel}=\left(\frac{\exp(-\,\mathrm{fval}(\xi_1))}{\exp(-\,\mathrm{fval}(\xi_2))}\right)^{1/6}$$

since $m=6$.

Table 4.6: Relative D-efficiency of Two Designs

Design space		Case	Design				GET verification	$fval$	$\exp(-fval)$	D_{rel}
-5	15	1st obtained	-5	0.591		15	failure	1.798	0.166	0.882
			0.336	0.498		0.166				
		2nd obtained	-5	-1.196	1.298	15	success	1.043	0.353	
			0.182	0.271	0.383	0.165				

We compare the two designs that we obtained in the design space $[-5, 15]$ by this formula. The relative D-efficiency of the design that did not pass the GET verification to the D-optimal design is 0.822 in Table 4.6. It is clear that the first obtained design is less efficient than the second obtained design since the relative efficiency is less than 1.

In addition, by measuring the relative efficiency, we obtain a relative sample size needed for the worse design to attain the same efficiency. Let us denote the design that failed the GET verification as $\xi_f=\{(c_{fi}, w_{fi}), i=1,2,3\}$ and the optimal design as $\xi^*=\{(c_i^*, w_i^*), i=1,2,3,4\}$. Also, we define a relative sample size N_r as $N_r=N_f/N^*$ where N_f is the sample size of ξ_f and N^* is of ξ^*. Then, we want to know N_r such that $|\sum N_r w_{fi}\mathbf{C}(\xi_f)|=|\sum w_i^*\mathbf{C}(\xi^*)|$. Since it holds that $N_r^6|\sum w_{fi}\mathbf{C}(\xi_f)|=|\sum w_i^*\mathbf{C}(\xi^*)|$ and we know that $|\sum w_{fi}\mathbf{C}(\xi_f)|/|\sum w_i^*\mathbf{C}(\xi^*)|=0.882^6$ from $D_{ref}=0.882$, we have $N_r^6=1/0.882^6$, that is, $N_r=1.134$. It means we need 1.134 times the sample size when we use the failed design compared to the case when we use a D-optimal design.

4.3 Discussion

In this chapter, we found D-optimal designs for both binary and continuous responses. We consider a mixed response model with a direct factorization approach following Deng and Jin (2015). In the previous chapter, we identified a complete class with at most 4 points for a mixed response model. Within the complete class, we obtained D-optimal designs in each design space after using a constrained nonlinear optimization. The obtained designs were 2-point, 3-point, or 4-point designs depending on the size and location of design spaces. The obtained designs are the most efficient under the D-optimality criterion in each design space so that we expect D-optimal designs to minimize the determinant of the variance-covariance matrix of parameter estimates. The general equivalence theorem verifies the D-optimality of the obtained designs.

Over various symmetric design regions of $[-B, B]$, we obtain D-optimal designs including 2-point designs of $\{(-B, 0.5), (B, 0.5)\}$, 3-point designs of $\{(-B, w_1)$, $(0, w_2)$, $(B, w_1)\}$, and 4-point designs of $\{(-B, w_1), (-c_1, w_2)(c_1, w_2), (B, w_1)\}$. All support points are symmetric in terms of a value. In particular, 4-point designs have two symmetric outer points and two symmetric inner points. The outer points are the end points of the design region and the inner points are close to ± 1.4. The stable weight ratio is $1:2:2:1$ when the design space is wider than $[-7, 7]$.

We then recommend a joint experiment of mixed responses over the simple combination of two experiments, namely GLM and linear model experiments. By using optimal designs, an experiment becomes more efficient. Above all, our approach eliminated uncertainty in replicates ratios so that an effective allocation of inputs is possible, which means the saving of cost. In an arbitrary combination, it is prone to allocate resources uniformly as 1:1:1:1 due to experimenter's perception bias.

It is interesting to compare our design points with known results of D-optimal designs for a logistic regression model or a linear model. According to Stufken and Yang (2012), a D-optimal design for a logistic regression is given by $\xi = \{(-1.5434, 1/2), (1.5434, 1/2)\}$ when a design space $[-a, a]$ is wider than $[-1.5434, 1.5434]$. Otherwise, it is $\xi = \{(-a, 1/2), (a, 1/2)\}$. It is also well known that a D-optimal design for a simple linear regression model is $\{(-B, 1/2), (B, 1/2)\}$ on $[-B, B]$. In our case, 4-point designs have the form of $\{(-B, w_1), (-c_1, w_2)(c_1, w_2), (B, w_1)\}$ when a design space is approximately between $[-3, 3]$ and $[-150, 150]$. The value of c_1 is close to but not equal to 1.5434. The exact values of the inner two points are not fixed in our case. Delicate experiments such as a dose-finding study will be affected by a different range of a design space.

For a real setting, two scenarios are possible. Assume an experimenter is interested in estimating each parameter under the D-optimality criterion. At first, if an experimenter plans to use a logistic regression model, we may suggest adding a linear model to increase the utility of an experiment if it is affordable to measure a continuous response. Then, the experimenter will treat four different levels of an input variable instead of two points. Another scenario is that if an experimenter wants to use only a linear model, we may recommend measuring a binary response together since it is often done at a low cost but provides crucial information.

This study concerns only D-optimal designs. Since the complete class approach gives general results, we can use our complete class for the search of A-optimal designs. While D-optimal designs ignored the existence of a variance σ^2 of continuous responses due to the property of a determinant function, the search for A-optimal designs depends on the value of σ^2. This observation leads the study of A-optimal designs in the next chapter.

Chapter 5

NUMERICAL RESULTS 2: A-OPTIMAL DESIGNS

So far we have introduced a mixed response experiment and its related model and found D-optimal designs. After identifying a collection of candidate designs by the complete class approach, we obtained optimal designs by a constrained nonlinear optimization. Our focus was on the popular D-optimality criterion, i.e. $\Phi_D = \log|\mathbf{M}^{-1}|$.

In this chapter, we will look at what optimal designs we will obtain when we use the A-optimality criterion, $\Phi_A = \operatorname{tr}\mathbf{M}^{-1}$, which is the trace of the variance-covariance matrix. Under this criterion, we search for an A-optimal design that minimizes the average variance of the parameter estimates. Geometrically, $\operatorname{tr}\mathbf{M}^{-1}$ is equal to the square of the half-length of the diagonal of a rectangle that is circumscribed around the confidence ellipsoid of parameters (Fedorov and Leonov, 2014; Atkinson et al., 2007).

Moreover, recalling that our information matrix is $\mathbf{M} = \operatorname{diag}(\mathbf{M}_1, \frac{1}{\sigma^2}\mathbf{M}_2, \frac{1}{\sigma^2}\mathbf{M}_3)$, we see that A-optimal designs depend on the value of σ^2 since the A-optimality criterion contains inseparable σ^2 as $\Phi_A = \operatorname{tr}\mathbf{M}^{-1} = \operatorname{tr}\mathbf{M}_1^{-1} + \sigma^2 \operatorname{tr}\mathbf{M}_2^{-1} + \sigma^2 \operatorname{tr}\mathbf{M}_3^{-1}$. A parameter σ^2 is the variance of continuous responses given a binary response 0 or 1. This observation partly motivates the search for A-optimal designs. We also note that the A-optimality criterion includes other unknown parameters in addition to σ^2. In the previous case of D-optimal designs, we could not examine the effect of σ^2 since D-optimal designs are invariant to the value of σ^2 as shown in (4.1).

We will first discuss the preceding study of an A-optimal design in a treatment comparison study after mentioning the popularity of the D-optimality criterion. And then, for comparison purposes, we review a literature of A-optimal designs for a generalized linear model (GLM). We will then derive the A-optimality criterion function for a mixed response model, present the scenario that we consider for searching for A-optimal designs, and provide numerical results. There is a discussion of the results at the end. Here, we use again the previously derived complete class as the set of candidate designs.

5.1 A-optimal Designs in a Clinical Comparison Study

Researchers have recently favored the D-optimality criterion as a standard of finding optimal designs. Pukelsheim (2006) held several reasons as to why. First, the determinant of the covariance-variance matrix has often been used in multivariate analysis to measure the size of a dispersion matrix. We trace it to the generalized variance defined by Wilks (1932) who introduced the pth order determinant as the variance of a sample from a p-variate normal population analogous to a univariate case. In addition, Pukelsheim (2006) pointed out that the determinant is invariant to the representation of parameters as explained in the previous chapter.

One additional reason for the popularity of the D-optimality criterion is that, as Kiefer and Wolfowitz (1960) proved the equivalence between G-optimality and D-optimality in the context of a continuous design theory, the explanatory power of the D-optimality criterion has increased. In the field of a response surface design in engineering experiments, it is a primary concern to reduce the prediction errors of responses, which is connected to G-optimality. The G-optimality criterion is used to seek designs that minimize the maximum variance of a predictor of a response at a given x.

However, we have situations where D-optimal designs are not enough to meet the purpose of experiments, and may sometimes fail to accommodate features that are important to the scientific problem of interest. Our study of D-optimal designs in the previous chapter is one example. When we want to know the effect of the variance of the continuous response variable on design for a mixed response experiment, D-optimal designs are always the same regardless of the values of σ^2.

In a clinical study setting, it was also observed that D-optimal designs were not enough to distinguish two different situations. The first was a case where all the pairwise comparisons of the treatment effects were of interest and the second was a situation where the main focus lies only in the comparison of the treatments versus a control. When we consider the latter the objective of an experiment, we cannot fully find a 'good' design by using the D-optimality criterion. By contrast, A-optimal designs seem to better fit the corresponding objective of the study as discussed in Hedayat, Jacroux, and Majumdar (1988).

Specifically, they considered the A- and MV-optimality criteria. Denote a control as 0, each test treatment as i for $i=1,...,k$, and the effect of treatment i as t_i. The best linear unbiased estimators were notated as $\hat{t_{di}} - \hat{t_{d0}}$ for the contrasts $t_i - t_0$. A-optimal designs gave the minimum of $\sum_{i=1}^{k} \text{var}(\hat{t_{di}} - \hat{t_{d0}})$ and MV- optimal designs minimized $\max_{1 \le i \le k} \text{var}(\hat{t_{di}} - \hat{t_{d0}})$.

The reason why the authors favored these two criteria rather than the D-optimality criterion was that D-optimal designs counted the minimization of variance of the comparisons between the test treatments while this inclusion was not necessary when comparing treatments versus a control. The goal of their study was to find good designs to estimate the magnitude of contrasts between each treatment and a control of standard as precisely as possible. A-optimal designs can be used when an experimenter wants to decide which treatment is effective among new test treatments after performing an experiment.

Later, in the comments, Notz (1988) asked "what sort of criteria might be useful" and Hedayat et al. (1988) restated in the rejoinder that the A- and MV-optimality criteria have "simple and statistically meaningful interpretations" in their problem. Also, they stated that the robustness of designs over criteria might not be crucial compared to the robustness of designs over a selection of models.

In the next section, we will look at two studies of A-optimal designs for a generalized linear model. Our model is a mixed response model that contains a logistic regression model. It is therefore interesting to see what the A-optimal design of the GLM is.

5.2 A-optimal Designs for a Generalized Linear Model

Mathew and Sinha (2001) obtained A-optimal designs for a logistic regression model using two different approaches. The first one was an analytic approach focusing on a class of symmetric designs. A symmetric design here is a design with symmetric designs points. They called weights symmetric when two points had equal weights. They also defined a represented design space by a bijection which we use. They initially set a symmetric two-point design with an equal weight of 0.5 as a candidate of optimal designs. Using this design, they simplified the A-optimality criterion and derived a lower bound. After, they searched for designs that minimized the lower bound numerically. Another approach was a numerical optimization within an entire class of designs. As the authors mentioned, the first approach did not provide the best designs compared to the second case. They showed that the efficiency of the first type of designs was lower than that of the second type of designs in their table. Also, be advised that the A-optimality criterion for a logistic regression model had unknown parameters similar to our criterion. Adopting a locally optimal design approach, they used guessed values of the two parameters as $(\alpha,\beta)=(10,5),(5,5),(1,5),(10,2),(5,2),(1,2),(10,0.5),(5,0.5),(1,0.5)$ where α,β are two parameters in a simple logistic regression.

On the other hand, Yang (2008) obtained A-optimal designs for GLMs with two parameters using an algebraic method. He found that there existed A-optimal designs where two support designs were symmetric, but their weights were not equal. The author derived sufficient conditions for extending the result within the class of symmetric designs to the general result within the entire class of designs based on the min-max idea from Kunert and Stufken (2002). In the three-step approach, an A-optimal design d^* was firstly identified among a subclass of designs $\mathcal{D}_1$ that contained $d=\{(x_1,\xi_1),(x_2,\xi_2)\}$ where two points were symmetric when the represented design point is considered as $\alpha+x_1\beta=-\alpha-x_2\beta$. Then, the sufficient conditions were derived for d^* to satisfy the property of $\mathrm{tr}\,\mathbf{M}_d^{-1}\geq\mathrm{tr}\,\mathbf{M}_{d^*}^{-1}$ for any arbitrary design d. Lastly, the author verified that the models considered in the study met the sufficient conditions. Consequently, the identified d^* is A-optimal over the entire class of designs. As an example, he showed that, for a simple logistic regression, there existed A-optimal designs with two symmetric design points. The theoretical results in the study matched up with the results in Mathew and Sinha (2001).

In our case, although the mixed response model that we consider includes a simple logistic regression, we find it challenging to directly apply the methodologies in the two previous studies for finding A-optimal designs. This mainly is because our model not only contains a logistic regression model but also involves two simple linear models for the continuous response variable, given the value of the binary response variable. However, our study embraces their main ideas about the Loewner ordering by using the complete class approach of Yang and Stufken (2012). In the next section, we derive the A-optimality criterion for the mixed response model that we consider. We then explain two associated situation for the search and then obtained A-optimal designs under certain scenarios.

5.3 The A-optimality Criterion and Associated Conditions

With the same notation as in the previous chapters, the A-optimality criterion Φ_A for our mixed response model has the following form:

$$\begin{aligned}
\Phi_A &= \operatorname{tr}\mathbf{M}^{-1} = \operatorname{tr}\mathbf{M}_1^{-1} + \sigma^2 \operatorname{tr}\mathbf{M}_2^{-1} + \sigma^2 \operatorname{tr}\mathbf{M}_3^{-1} \\
&= \operatorname{tr}[\mathbf{F}^\intercal \mathbf{W}\mathbf{P}(\mathbf{I}-\mathbf{P})\mathbf{F}]^{-1} + \sigma^2 \operatorname{tr}[\mathbf{F}^\intercal \mathbf{W}\mathbf{P}\mathbf{F}]^{-1} + \sigma^2 \operatorname{tr}[\mathbf{F}^\intercal \mathbf{W}(\mathbf{I}-\mathbf{P})\mathbf{F}]^{-1} \\
&= \operatorname{tr}[\mathbf{B_1}\mathbf{C_1^*}\mathbf{B_1^\intercal}]^{-1} + \sigma^2 \operatorname{tr}[\mathbf{B_1}\mathbf{C_2^*}\mathbf{B_1}^\intercal]^{-1} + \sigma^2 \operatorname{tr}[\mathbf{B_1}\mathbf{C_3^*}\mathbf{B_1^\intercal}]^{-1} \\
&= \operatorname{tr}[\mathbf{B}_1^{-1}\mathbf{B}_1^{\intercal -1}\mathbf{C}_1^{*-1}] + \sigma^2 \operatorname{tr}[\mathbf{B}_1^{-1}\mathbf{B}_1^{\intercal -1}\mathbf{C}_2^{*-1}] + \sigma^2 \operatorname{tr}[\mathbf{B}_1^{-1}\mathbf{B}_1^{\intercal -1}\mathbf{C}_3^{*-1}].
\end{aligned}$$

Here, the first term $\operatorname{tr}[\mathbf{B}_1^{-1}\mathbf{B}_1^{\intercal -1}\mathbf{C}_1^{*-1}]$ with $\mathbf{C}_1^* = \begin{pmatrix} \sum w_i\Psi_{11}(c) & \sum w_i\Psi_{12}(c) \\ \sum w_i\Psi_{12}(c) & \sum w_i\Psi_{22}(c) \end{pmatrix}$ is:

$$\begin{aligned}
\operatorname{tr}[\mathbf{B}_1^{-1}\mathbf{B}_1^{\intercal -1}\mathbf{C}_1^{*-1}] &= \operatorname{tr}[\begin{pmatrix} 1 & 0 \\ \alpha_0 & \alpha_1 \end{pmatrix}\begin{pmatrix} 1 & \alpha_0 \\ 0 & \alpha_1 \end{pmatrix}\mathbf{C}_1^{*-1}] \qquad (5.1) \\
&= \operatorname{tr}[\begin{pmatrix} 1 & \alpha_0 \\ \alpha_0 & \alpha_0^2+\alpha_1^2 \end{pmatrix}\frac{1}{|\mathbf{C}_1^*|}\begin{pmatrix} \sum w_i\Psi_{22} & -\sum w_i\Psi_{21} \\ -\sum w_i\Psi_{12} & \sum w_i\Psi_{11} \end{pmatrix}] \\
&= \frac{1}{|\mathbf{C}_1^*|}\operatorname{tr}[\begin{pmatrix} \sum w_i\Psi_{22}-\alpha_0\sum w_i\Psi_{12} & -\sum w_i\Psi_{12}+\alpha_0\sum w_i\Psi_{11} \\ \alpha_0\sum w_i\Psi_{22}-(\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{12} & -\alpha_0\sum w_i\Psi_{12}+(\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{11} \end{pmatrix}] \\
&= \frac{1}{|\mathbf{C}_1^*|}(\sum w_i\Psi_{22} - 2\alpha_0\sum w_i\Psi_{12} + (\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{11}) \qquad (5.2) \\
&= \frac{1}{|\mathbf{C}_1^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{11} - 2\alpha_0\sum w_i\Psi_{12} + \sum w_i\Psi_{22}).
\end{aligned}$$

With similar algebra, we can show that the A-optimality criterion is

$$\begin{aligned}
\Phi_A &= \frac{1}{|\mathbf{C}_1^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{11} - 2\alpha_0\sum w_i\Psi_{12} + \sum w_i\Psi_{22}) \\
&\quad + \frac{\sigma^2}{|\mathbf{C}_2^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{33} - 2\alpha_0\sum w_i\Psi_{34} + \sum w_i\Psi_{44}) \\
&\quad + \frac{\sigma^2}{|\mathbf{C}_3^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\Psi_{55} - 2\alpha_0\sum w_i\Psi_{56} + \sum w_i\Psi_{66}) \\
&= \frac{1}{|\mathbf{C}_1^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\frac{e^{c_i}}{(1+e^{c_i})^2} - 2\alpha_0\sum w_i\frac{c_ie^{c_i}}{(1+e^{c_i})^2} + \sum w_i\frac{c_i^2e^{c_i}}{(1+e^{c_i})^2}) \qquad (5.3) \\
&\quad + \frac{\sigma^2}{|\mathbf{C}_2^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\frac{e^{c_i}}{1+e^{c_i}} - 2\alpha_0\sum w_i\frac{c_ie^{c_i}}{1+e^{c_i}} + \sum w_i\frac{c_i^2e^{c_i}}{1+e^{c_i}}) \qquad (5.4) \\
&\quad + \frac{\sigma^2}{|\mathbf{C}_3^*|}((\alpha_0^2+\alpha_1^2)\sum w_i\frac{1}{1+e^{c_i}} - 2\alpha_0\sum w_i\frac{c_i}{1+e^{c_i}} + \sum w_i\frac{c_i^2}{1+e^{c_i}}). \qquad (5.5)
\end{aligned}$$

It can be rewritten as:

$$\Phi_A = (\alpha_0^2 + \alpha_1^2)\Big(\frac{1}{|\mathbf{C}_1^*|}\sum w_i \frac{e^{c_i}}{(1+e^{c_i})^2} + \frac{\sigma^2}{|\mathbf{C}_2^*|}\sum w_i \frac{e^{c_i}}{1+e^{c_i}} + \frac{\sigma^2}{|\mathbf{C}_3^*|}\sum w_i \frac{1}{1+e^{c_i}}\Big) \tag{5.6}$$

$$- 2\alpha_0\Big(\frac{1}{|\mathbf{C}_1^*|}\sum w_i \frac{c_i e^{c_i}}{(1+e^{c_i})^2} + \frac{\sigma^2}{|\mathbf{C}_2^*|}\sum w_i \frac{c_i e^{c_i}}{1+e^{c_i}} + \frac{\sigma^2}{|\mathbf{C}_3^*|}\sum w_i \frac{c_i}{1+e^{c_i}}\Big) \tag{5.7}$$

$$+ \Big(\frac{1}{|\mathbf{C}_1^*|}\sum w_i \frac{c_i^2 e^{c_i}}{(1+e^{c_i})^2} + \frac{\sigma^2}{|\mathbf{C}_2^*|}\sum w_i \frac{c_i^2 e^{c_i}}{1+e^{c_i}} + \frac{\sigma^2}{|\mathbf{C}_3^*|}\sum w_i \frac{c_i^2}{1+e^{c_i}}\Big). \tag{5.8}$$

What follows are the issues to consider when searching for an A-optimal design.

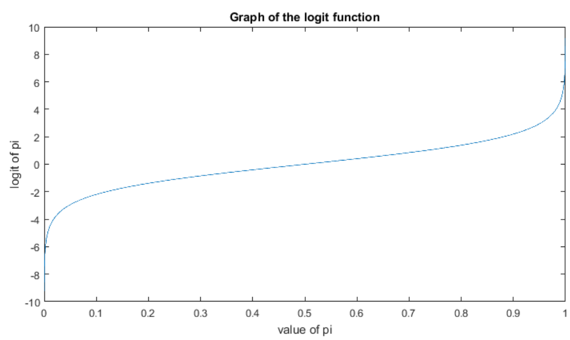
(i) Representation of design space ($x \to c$) We represent a design point x as a represented point c by defining $c = \alpha_0 + \alpha_1 x$. We used this bijection for the search for D-optimal designs in the previous chapter. The use of represented design point c normally simplified the design problem. This was because we only need to search once for the optimal combination of the c's and the corresponding weights w's (i.e., the optimal design in terms of c) from a simpler form of the criterion without unknown parameters. The values of the x and w easily derived using any given values of α_0 and α_1.

In contrast, the search for an A-optimal design does not take full advantage of the representation. This can be seen from the formula of the A-optimality criterion which is a rather complex function of c_i's, w_i's and the unknown parameters α_0, α_1, and σ^2. Specifically, we observed in (5.6), (5.7), and (5.8) that the criterion depends on these parameters through $\alpha_0^2 + \alpha_1^2$, $2\alpha_0$, and σ^2.

Additionally, we have another situation which is mathematically similar to, but statistically different from the above issue regarding the unit of a covariate. If the unit is changed, the scale of the covariate is also changed. Then, A-optimal designs are also different depending on the unit. We represent x by defining $x = m_l + m_s u$ where u is an original covariate variable, x is a standardized variable, m_l is a location parameter, and m_s is a scale parameter. This representation can be expressed by using a transformation matrix. Yet, we oscillate only between x and c without u since we can find new α_0, and α_1 for any u with given fixed values of m_l and m_s.

(ii) Size of a represented design space (c) In the previous chapter, we found that the number of support points in D-optimal designs changes from two to four when the size or location of the design space of c_i changes. However, this feature is not observed in the D- or A-optimal designs for a simple logistic regression according to other studies such as Matthew and Sinha (2001). In a simple logistic model, optimal points and weights are consistent in certain design spaces. With this in mind, we investigate A-optimal designs for a mixed response model varying design spaces.

Figure 5.1: The Graph of the Logit Function (Logit(π) vs. π)



Furthermore, we consider the following facts when determining the design spaces for numerical analysis of obtained designs. The logit function in a logistic regression model is defined as logit(π)=log[$\pi/(1-\pi)$] where π is the probability for $z=1$ with '1' corresponding to a success. Then, a logistic regression model can be expressed as $\log[\pi/(1-\pi)]=\alpha_0+\alpha_1 x$. Since $\log[\pi/(1-\pi)]=c$, we immediately derive the range of c by π. When π is in [0,1], the logit can take any number in $(-\infty,\infty)$. The logit function is increasing with π as in Figure 5.1. When we set π to certain intervals, c has a finite design space. We make the following observation.

Remark For given $\pi_{\min}$ and $\pi_{\max}$, the range of c is $(\text{logit}(\pi_{\min}),\ \text{logit}(\pi_{\max}))$. For example, assuming $\pi_{\min}=0.0001$ and $\pi_{\max}=0.9999$, we get the range of c as $(-9.21, 9.21)$. For $\pi_{\min}=0.00001$ and $\pi_{\max}=0.99999$, the range of c is $(-11.52, 11.52)$. For $\pi_{\min}=0.01$ and $\pi_{\max}=0.99$, the range of c is $(-4.60, 4.60)$.

Hence, assuming that $\pi\in[0.0001, 0.9999]$, the design space of c is about $[-10, 10]$. In Table 4.4 in the previous chapter, we observe that when design spaces are wider than $[-10, 10]$, the design points and their weights of the obtained optimal designs do not change much. Although the above remark does not provide a statistical or mathematical meaning, we acquire a spatial cognition about the design space of c for a simple logistic model or simple mixed response model.

Table 5.1: Possible Interpretation of α_1 in a Logistic Model

α_1	$\exp(\alpha_1)$	proportional increase in odds in unit change of x	$1/\exp(\alpha_1)$
-0.2	0.819	-18.13%	**1.2214**
-0.1	0.904	-9.52%	1.1052
0	1	0.00%	1
0.1	1.1052	10.52%	0.904
0.2	**1.2214**	22.14%	0.819
0.3	1.3499	34.99%	the rest is omitted
0.4	1.4918	49.18%	
0.5	1.6487	64.87%	
1	2.7183	171.83%	
2	7.3891	638.91%	
3	20.086	1908.55%	
4	54.598	5359.82%	
5	148.4132	14741.32%	
10	22026.4658	2202546.58%	
15	3269017.372	326901637.25%	

On the other hand, we also get information on a parameter α_1 as can be seen in Table 5.1. The parameter α_1 plays a vital role in a logistic regression analysis since it indicates the magnitude of an increment of the log odds of responses by one unit increase in x. The value of $\exp(\alpha_1)$ is normally interpreted as a proportional increase in odds corresponding to the one unit increase in x (Collett, 2002; Agresti, 2007; Hosmer, Lemeshow, and Sturdivant, 2013). In Table 5.1, we observe that the increasing trend of $\exp(\alpha_1)$ is exponential and if α_1 is larger than 15, the proportional increase of odds in a unit change of x starts to show extreme values.

It is possible for us to consider that α_1 rarely exceeds 15 if we assume that the proportional increase in odds is not as large as 326901637.25% in Table 5.1. We note that a unit change of x might or might not be a huge change for x. The table helps to guide our numerical study, although we also recognize that the value of α_1 depends on the scale of x.

As noted previously, we focus on the effect of σ^2 when searching for A-optimal designs. We will give particular attention to the effect of σ^2 which is not usually studied in the context of the D-optimality criterion or linear models. As a trade-off, we need to consider many combinations of guessed values of unknown parameters. When possible guessed values are far away from the real values of parameters, the obtained A-optimal designs might not be as reliable. These are the pros and cons of the delicacy of a search for A-optimal designs.

5.4 Verification of the A-optimality

After obtaining A-optimal designs, the A-optimality of the designs will be verified via the general equivalence theorem. As explained in Chapter 2, we verify the condition of $s(c,\xi)\leq C$ where $s(c,\xi)=\operatorname{tr}\mathbf{C}^*(\bar{\xi})(\mathbf{C}^*(\xi)\mathbf{B}^\intercal\mathbf{B}\mathbf{C}^*(\xi))^{-1}$ and $C=\operatorname{tr}(\mathbf{B}\mathbf{C}^*(\xi)\mathbf{B}^\intercal)^{-1}$ is a constant function calculated from the obtained design. In our current setting with $\mathbf{S}^*(\xi)=\mathbf{C}^*(\xi)\mathbf{B}^\intercal\mathbf{B}\mathbf{C}^*(\xi)$, we have:

$$\mathbf{S}^*(\xi)=\mathbf{C}^*(\xi)\mathbf{B}^\intercal\mathbf{B}\mathbf{C}^*(\xi)=\begin{bmatrix}\mathbf{C}_1^* & & \\ & \mathbf{C}_2^* & \\ & & \mathbf{C}_3^*\end{bmatrix}\begin{bmatrix}\mathbf{B}_1^\intercal & & \\ & \frac{1}{\sigma}\mathbf{B}_1^\intercal & \\ & & \frac{1}{\sigma}\mathbf{B}_1^\intercal\end{bmatrix}\begin{bmatrix}\mathbf{B}_1 & & \\ & \frac{1}{\sigma}\mathbf{B}_1 & \\ & & \frac{1}{\sigma}\mathbf{B}_1\end{bmatrix}\begin{bmatrix}\mathbf{C}_1^* & & \\ & \mathbf{C}_2^* & \\ & & \mathbf{C}_3^*\end{bmatrix}$$

$$=\begin{bmatrix}\mathbf{C}_1^*\mathbf{B}_1^\intercal\mathbf{B}_1\mathbf{C}_1^* & & \\ & \frac{1}{\sigma^2}\mathbf{C}_2^*\mathbf{B}_1^\intercal\mathbf{B}_1\mathbf{C}_2^* & \\ & & \frac{1}{\sigma^2}\mathbf{C}_3^*\mathbf{B}_1^\intercal\mathbf{B}_1\mathbf{C}_3^*\end{bmatrix}=\begin{bmatrix}\mathbf{S}_1^*(\xi) & & \\ & \mathbf{S}_2^*(\xi) & \\ & & \mathbf{S}_3^*(\xi)\end{bmatrix}.$$

We will represent $\mathbf{S}^*(\xi)=\operatorname{diag}(\mathbf{S}_1^*(\xi),\mathbf{S}_2^*(\xi),\mathbf{S}_3^*(\xi))$. The verification of the A-optimality will also depend on the guessed values of α_0 α_1, and σ^2 that we specify in the numerical study. We denote $\mathbf{S}_k^{*-1} = ((s^*_{kij}))_{ij=1,2}$ for $k=1,2,3$. Hence, $s(c,\xi)$ is expressed as follows:

$$\begin{aligned} s(c,\xi) &== \operatorname{tr} \mathbf{C}^*(\bar{\xi})(\mathbf{C}^*(\xi)\mathbf{B}^\intercal \mathbf{B}\mathbf{C}^*(\xi))^{-1} = \operatorname{tr} \mathbf{C}^*(\bar{\xi})\mathbf{S}^{*-1}(\xi) \\ &= \operatorname{tr} \mathbf{C}_1^*(\bar{\xi})\mathbf{S}_1^{*-1}(\xi) + \operatorname{tr} \mathbf{C}_2^*(\bar{\xi})\mathbf{S}_2^{*-1}(\xi) + \operatorname{tr} \mathbf{C}_3^*(\bar{\xi})\mathbf{S}_3^{*-1}(\xi), \end{aligned}$$

where $\operatorname{tr} \mathbf{C}_1^*(\bar{\xi})\mathbf{S}_1^{*-1}(\xi) = \operatorname{tr} \mathbf{f}(c)\pi(1-\pi)\mathbf{f}^\intercal(c)\mathbf{S}_1^{*-1}(\xi) = \pi(1-\pi)(s^*_{111} + (s^*_{112} + s^*_{121})c + s^*_{122}c^2)$ with a model vector $\mathbf{f}(c) = [1 \;\; c]^\intercal$. We use the following proposition to verify the A-optimality of an obtained design.

Proposition 5.4.1. *For a mixed responses model, we verify the A-optimality of an obtained design ξ if it holds that*

$$s(c,\xi) = \frac{e^c}{(1+e^c)^2}\mathbf{f}(c)^\intercal \mathbf{S}_1^{*-1}(\xi)\mathbf{f}(c)^\intercal + \frac{e^c}{(1+e^c)}\mathbf{f}(c)^\intercal \mathbf{S}_2^{*-1}(\xi)\mathbf{f}(c)^\intercal + \frac{1}{(1+e^c)}\mathbf{f}(c)^\intercal \mathbf{S}_3^{*-1}(\xi)\mathbf{f}(c)^\intercal$$

is equal to or less than $C = \operatorname{tr}\ (\mathbf{B}\mathbf{C}^*(\xi)\mathbf{B}^\intercal)^{-1}$ where $\mathbf{S}_1^*(\xi) = \mathbf{C}_1^*(\xi)\mathbf{B}_1^\intercal \mathbf{B}_1 \mathbf{C}_1^*(\xi)$, $\mathbf{S}_2^*(\xi) = \frac{1}{\sigma^2}\mathbf{C}_2^*(\xi)\mathbf{B}_1^\intercal \mathbf{B}_1 \mathbf{C}_2^*(\xi)$, and $\mathbf{S}_3^*(\xi) = \frac{1}{\sigma^2}\mathbf{C}_3^*(\xi)\mathbf{B}_1^\intercal \mathbf{B}_1 \mathbf{C}_3^*(\xi)$ based on the general equivalence theorem for all c in $[A,B]$. For an A-optimal design $\xi^* = \{(c_i^*, w_i^*), i=1,...,4, w_i^* \geq 0\}$, we have $s(c^*,\xi^*) = C$.

Using the values of an obtained design ξ, we calculate $\mathbf{C}^*$. Guessed values of α_0 and α_1 are used to obtain $\mathbf{B}$ and $\mathbf{B}_1$. Then, we find the value of each element in $\mathbf{S}_i^{*-1}(\xi^*), i=1,2,3$ and a constant C. We draw the plots of $y = s(c,\xi^*)$ and $y = C$. The A-optimality of a design ξ can then be verified by seeing the plot in a similar way as described in the D-optimality verification in Chapter 4.

5.5 Numerical Results

We make guidelines for a computational search as follows: (i) We use again the complete class consisting of designs having at most four support points with two points A and B of a design space $[A,B]$ (ii) We set initial values with two design points located at the 30th and 60th percentile in $[A, B]$ and weights being equally .25. (iii) We use the sequential quadratic programming. (iv) We check the A-optimality of all obtained designs by a general equivalence theorem at the same time. All obtained designs are in the represented design space of c.

After conducting some preliminary simulations, we observed that there are many scenarios we can consider. It was not easy to capture a general pattern from various combinations of many guessed values. We, therefore, limit our purpose of analysis to observing the effect of σ^2 on the obtained A-optimal designs. To achieve this goal, we firstly need to collect reasonable guessed values of α_0, α_1 and design spaces. Then, we will find A-optimal designs by the values of σ^2.

At first, we see the effect of design spaces for A-optimal designs for c. We set design spaces from $[-1, 1]$ to $[-50, 50]$ as in Table 5.2. We fix other parameters to $\alpha_0 = 1, \alpha_1 = 1, \sigma^2 = 1$ or $\alpha_0 = 1, \alpha_1 = 0.5, \sigma^2 = 1$. We use two different $\alpha_1 = 1$ and 0.5.

As shown in the table, we obtain 4-point optimal designs when a design space is not $[-1, 1]$. The noticeable thing is that two inner points are not symmetric both in terms of their locations and their corresponding weights. Also the two boundary outer points have unequal weights. For example, when the design space is $[-10, 10]$ with $\alpha_0 = 1, \alpha_1 = 1$, and $\sigma^2 = 1$, the two outer design points of the obtained design are -10, and 10 with the weights 0.030 and 0.022, respectively. The two inner points are -1.326, and 1.395 with the weights 0.282 and 0.667, respectively. The asymmetric pattern of two inner points with unequal weights is observed across the various design spaces that we consider. This tendency is also observed when $\alpha_1 = 0.5$ as can be seen in the lower part of Table 5.2.

Similar to D-optimal designs, two inner points do not significantly change when a design space changes. The positive inner point, c_3, is located somewhere between 0.996 and 1.395 for every 4-point A-optimal design. This second largest design point

Table 5.2: A-optimal Designs for c by Design Spaces

α_0	α_1	σ^2	Design space		Design				# of points
1	1	1	-1	1	-1	0.964		1	3
					0.309	0.000+		0.691	
			-2	2	-2	-0.849	0.996	2	4
					0.172	0.099	0.404	0.326	
			-3	3	-3	-1.080	1.285	3	4
					0.075	0.213	0.615	0.096	
			-5	5	-5	-1.179	1.381	5	4
					0.043	0.259	0.659	0.040	
			-10	10	-10	-1.326	1.395	10	4
					0.030	0.282	0.667	0.022	
			-20	20	-20	-1.479	1.387	20	4
					0.019	0.295	0.674	0.012	
			-50	50	-50	-1.605	1.382	50	4
					0.009	0.306	0.680	0.005	
1	0.5	1	-1	1	-1			1	2
					0.195			0.805	
			-2	2	-2	-0.788	1.006	2	4
					0.090	0.082	0.633	0.195	
			-3	3	-3	-0.899	1.162	3	4
					0.042	0.151	0.753	0.055	
			-5	5	-5	-0.961	1.216	5	4
					0.025	0.180	0.772	0.023	
			-10	10	-10	-1.125	1.205	10	4
					0.019	0.193	0.776	0.012	
			-20	20	-20	-1.354	1.178	20	4
					0.013	0.201	0.780	0.007	
			-50	50	-50	-1.588	1.155	50	4
					0.007	0.208	0.782	0.003	

has the largest weight among the four points. When the design space is $[-20, 20]$ or $[-50, 50]$, an A-optimal design keeps a similar pattern to that in $[-10, 10]$ in terms of the asymmetry of design points and the unequal of weights. From the observation, we reasonably set design spaces equal or less than $[-10, 10]$ for later analyses. Also, the analysis of the logit function in the previous section provides some explanation for the setting of the design space $[-10, 10]$. We confirm in this table that A-optimal designs tend to vary with the values of α_1 as well as the range of design spaces.

Table 5.3: A Comparison of A-optimal Designs for c Between a Mixed Response Model and a Logistic Model When $\alpha_0 = 1$ and $\alpha_1 = 1$

Design space		Design for a mixed response model when $\sigma^2 = 1$								Design for a logistic model			
Lower	Upper	c_1	c_2	c_3	c_4	w_1	w_2	w_3	w_4	c_1	c_2	w_1	w_2
-1	1	-1		0.964	1	0.309		0.000	0.691	-1	1	0.309	0.691
-2	2	-2	-0.849	0.996	2	0.172	0.099	0.404	0.326	-1.482	1.482	0.293	0.707
-3	3	-3	-1.080	1.285	3	0.075	0.213	0.615	0.096	-1.482	1.482	0.293	0.707
-5	5	-5	-1.179	1.381	5	0.043	0.259	0.659	0.040	-1.482	1.482	0.293	0.707
-10	10	-10	-1.326	1.395	10	0.030	0.282	0.667	0.022	-1.482	1.482	0.293	0.707
-20	20	-20	-1.479	1.387	20	0.019	0.295	0.674	0.012	-1.482	1.482	0.293	0.707
-50	50	-50	-1.605	1.382	50	0.009	0.306	0.680	0.005	-1.482	1.482	0.293	0.707

On the other hand, we obtain A-optimal designs for a logistic model as shown in Table 5.3. Based on Yang and Stufken (2009), we use a complete class with at most two-point designs, derive the A-optimality criterion, and implement the same algorithm. To compare with our results in Table 5.2, we use the same guessed values such as $\alpha_0 = 1$, and $\alpha_1 = 1$ for unknown parameters in the A-optimal criterion for a logistic model.

In Table 5.3, we find that A-optimal designs for a logistic model have symmetric points with unequal weights. When the design space is $[-1, 1]$, design points are located at -1 and 1. Their weights are 0.309 and 0.691, respectively. When the design space is wider than $[-1, 1]$, design points are constantly -1.482 and 1.482 with weights 0.293 and 0.707 regardless of a design space. In the case of a mixed response model, design points are not fixed by design spaces.

Next, we examine the effect of a change in α_0 on the obtained A-optimal designs. Fixing the represented design space to $[-10, 10]$, α_1 to 1 or 2, and σ^2 to 1, we set α_0 to 0, 0.1, 0.5, 1, 2, 5, 10, 20, and 50 which includes the values of α_0 as 1, 5, and 10 used in Mathew and Sinha (2001). A parameter α_0 is an intercept parameter in a linear predictor $\eta = \alpha_0 + \alpha_1 x$ for a logistic regression model.

As can be seen from Table 5.4, although a change of α_0 values results in a change of the obtained A-optimal design, the value change is gradual. The effect of a change in

Table 5.4: A-optimal Designs for c by α_0 Values

α_0	α_1	σ^2	Design space		Design				$\lvert c_3\rvert/\lvert c_2\rvert$	w_3/w_2
0	1	1	-10	10	-10	-1.197	1.1971	10	1.000	1.000
					0.025	0.475	0.475	0.025		
0.1					-10	-1.195	1.203	10	1.006	1.100
					0.025	0.453	0.498	0.024		
0.5					-10	-1.223	1.260	10	1.030	1.592
					0.027	0.367	0.584	0.023		
1					-10	-1.326	1.395	10	1.053	2.369
					0.030	0.282	0.667	0.022		
2					-10	-1.642	1.758	10	1.071	3.702
					0.037	0.200	0.741	0.021		
5					-10	-2.131	2.264	10	1.062	2.591
					0.058	0.253	0.654	0.035		
10					-10	-2.245	2.329	10	1.037	1.653
					0.062	0.336	0.555	0.047		
20					-10	-2.286	2.330	10	1.019	1.289
					0.061	0.387	0.499	0.053		
50					-10	-2.305	2.323	10	1.008	1.107
					0.059	0.420	0.465	0.056		
0	2	1	-10	10	-10	-1.580	1.580	10	1.000	1.000
					0.038	0.462	0.462	0.038		
0.1					-10	-1.579	1.584	10	1.003	1.050
					0.038	0.451	0.474	0.037		
0.5					-10	-1.585	1.612	10	1.017	1.271
					0.040	0.407	0.518	0.036		
1					-10	-1.624	1.676	10	1.032	1.584
					0.042	0.358	0.566	0.034		
2					-10	-1.768	1.860	10	1.052	2.181
					0.047	0.289	0.631	0.032		
5					-10	-2.125	2.244	10	1.056	2.276
					0.059	0.276	0.628	0.038		
10					-10	-2.245	2.327	10	1.036	1.627
					0.062	0.339	0.552	0.048		
20					-10	-2.286	2.330	10	1.019	1.287
					0.061	0.388	0.499	0.053		
50					-10	-2.305	2.323	10	1.008	1.107
					0.059	0.420	0.465	0.056		

the value of α_0 may not be great. We observe that, in all the cases that we studied except for when $\alpha_0 = 0$, the A-optimal designs are 4-point designs with two asymmetric inner points. Weights are not equally distributed and the second largest design point c_3 tends to get the highest weights among four weights. For example, when $\alpha_0 = 0.5, \alpha_1 = 1, \sigma^2 = 1$, the A-optimal design is $\xi = \{(-10, 0.027), (-1.223, 0.367), (1.260, 0.584), (10, 0.023)\}$. The second largest design point 1.260 has the weight of 0.584. When $\alpha_0 = 20$ where other conditions are the same, a design is $\xi = \{(-10, 0.061), (-2.286, 0.387), (2.330, 0.499), (10, 0.053)\}$, which has a similar pattern to the previous design with $\alpha_0 = 0.5$. When we have other α_0's between 0.5 and 20, we see a similar pattern. Even moving to the lower part of Table 5.4 where we use a different value of α_1, we observe that the pattern is repeated from the case of the upper part of the table.

We continue to observe that the inner design points of the obtained A-optimal designs are asymmetric, and the corresponding weights are not equal. On the other hand, when $\alpha_0 = 0$, we have symmetric optimal designs with symmetric weights.

Until now, we checked the effect of design spaces and a parameter α_0. We now move on to the effect of α_1 values on the obtained A-optimal designs. Based on the two previous results, we set α_0 as 1, and 5 and a represented design space as $[-10, 10]$. We assign α_1 to 0.1, 0.5, 1, 2, 5, 10, 20, and 50. As mentioned earlier, α_1 is the slope parameter in a logistic regression model.

In Table 5.5, we see that different values of α_1 give rise to different A-optimal designs. In most cases, the two inner points are asymmetrically located with unequal weights. For example, holding the values of $\alpha_0 = 1$ and $\sigma_2 = 1$ and a design space $[-10, 10]$, the A-optimal design points are -10, -1.225, 1.205, and 10 with weights 0.019, 0.193, 0.776, 0.012 when $\alpha_1 = 0.5$. The points -1.225 and 1.205 are not symmetric, and the weights 0.193 and 0.776 are not the same.

Table 5.5: A-optimal Designs for c by α_1 Values

α_0	α_1	σ^2	Design space		Design				$\lvert c_3\rvert/\lvert c_2\rvert$	w_3/w_2
1	0.1	1	-10	10	-10	-0.949	1.038	10	1.093	17.849
					0.005	0.053	0.940	0.003		
	0.5				-10	-1.125	1.205	10	1.071	4.022
					0.019	0.193	0.776	0.012		
	1				-10	-1.326	1.395	10	1.053	2.369
					0.030	0.282	0.667	0.022		
	2				-10	-1.624	1.676	10	1.032	1.584
					0.042	0.358	0.566	0.034		
	5				-10	-2.038	2.061	10	1.011	1.156
					0.053	0.416	0.481	0.049		
	10				-10	-2.221	2.228	10	1.004	1.047
					0.056	0.434	0.455	0.055		
	20				-10	-2.289	2.291	10	1.001	1.012
					0.057	0.440	0.446	0.057		
	50				-10	-2.311	2.311	10	1.000	1.002
					0.058	0.442	0.443	0.057		
5	0.1	1	-10	10	-10	-2.137	2.275	10	1.065	2.743
					0.058	0.243	0.666	0.034		
	0.5				-10	-2.135	2.272	10	1.064	2.703
					0.058	0.245	0.663	0.034		
	1				-10	-2.131	2.264	10	1.062	2.591
					0.058	0.253	0.654	0.035		
	2				-10	-2.125	2.244	10	1.056	2.276
					0.059	0.276	0.628	0.038		
	5				-10	-2.158	2.231	10	1.034	1.580
					0.059	0.347	0.548	0.047		
	10				-10	-2.233	2.265	10	1.015	1.210
					0.059	0.402	0.486	0.053		
	20				-10	-2.287	2.297	10	1.005	1.060
					0.058	0.430	0.456	0.056		
	50				-10	-2.310	2.312	10	1.001	1.010
					0.058	0.440	0.445	0.057		

However, the designs tend to be symmetric when α_1/α_0 increases. This tendency is obvious when checking the values of $|c_3|/|c_2|$ and w_3/w_2 that are close to 1 as α_1 increases in Table 5.5. When $\alpha_1 = 10$, design points are located at -10, -2.221, 2.228, and 10 with the weights 0.056, 0.434, 0.455, and 0.055. When $\alpha_1 = 50$, they are at -10, -2.311, 2.311, and 10 with the weights 0.058, 0.442, 0.443, and 0.057. The two inner points -2.311 and 2.311 are symmetric and their weights are almost the same for the latter case. In the previous table, we observed that we obtained symmetric designs when $\alpha_0 = 0$. In this table, we see that, as α_1/α_0 increases, the value of α_0 is relatively smaller compared to the value of α_1.

An additional noticeable thing is that A-optimal designs on the second row in Table 5.4 and on the first row 5.5 are different when we exchange the guessed values of α_0 and α_1. In the represented design space $[-10, 10]$, for the former one, we set $\alpha_0 = 0.1, \alpha_1 = 1$, and $\sigma^2 = 1$ and, for the latter one, we set $\alpha_0 = 1, \alpha_1 = 0.1$, and $\sigma^2 = 1$. When we exchange the values of α_0 and α_1, the value of $\alpha_0^2 + \alpha_1^2$ remains the same as 1.01 in (5.2), but the value of $-2\alpha_0$ changes from -0.2 with $\alpha_0 = 0.1$ to -2 with $\alpha_0 = 1$. The parameters α_0 and α_1 have different roles in a predictor $c = \alpha_0 + \alpha_1 x$ of the model and it is natural that the two obtained designs are different. The value of α_1/α_0 change from 100 to 0.1. We see that the former design is more close to a symmetric design with equal weights.

Now we will look at what optimal designs are obtained if we have different values of σ^2 in Table 5.6, Table 5.7, Figure 5.3, and Figure 5.2. In Table 5.6, we set the values of σ^2 as 0.1, 0.25, 0.5, 1, 10, and 100. We used three pairs of the guessed values for (α_0, α_1) including (1, 0.5), (5, 2), and (0, 1). There are three parts in Table 5.6 according to three pairs of the values of α_0 and α_1. The design space is $[-10, 10]$.

In the first part of Table 5.6 when $\alpha_0 = 1, \alpha_1 = 0.5$, we found that, as σ^2 increases, there are changes in two inner points c_2 and c_3 and four weights w_1, w_2, w_3, and w_4.

Table 5.6: A-optimal Designs for c by Variances in $[-10, 10]$

α_0	α_1	σ^2	Design				$var(z_1)$	$var(z_2)$	$var(z_3)$	$var(z_4)$	$\sum w_i var(z_i)$
1	0.5	0.1	-10	-1.232	1.273	10	0.00	0.17	0.17	0.00	0.17
			0.005	0.195	0.797	0.003					
		0.25	-10	-1.196	1.262	10	0.00	0.18	0.17	0.00	0.17
			0.010	0.194	0.791	0.006					
		0.5	-10	-1.165	1.241	10	0.00	0.18	0.17	0.00	0.17
			0.014	0.193	0.784	0.009					
		1	-10	-1.125	1.205	10	0.00	0.19	0.18	0.00	0.17
			0.019	0.193	0.776	0.012					
		10	-10	-0.766	0.989	10	0.00	0.22	0.20	0.00	0.19
			0.032	0.227	0.720	0.022					
		100	-10	-0.271	0.772	10	0.00	0.25	0.22	0.00	0.21
			0.037	0.331	0.601	0.032					
5	2	0.1	-10	-2.298	2.337	10	0.00	0.08	0.08	0.00	0.08
			0.013	0.294	0.685	0.008					
		0.25	-10	-2.249	2.323	10	0.00	0.09	0.08	0.00	0.08
			0.027	0.286	0.671	0.017					
		0.5	-10	-2.196	2.297	10	0.00	0.09	0.08	0.00	0.08
			0.041	0.279	0.653	0.026					
		1	-10	-2.125	2.244	10	0.00	0.10	0.09	0.00	0.08
			0.059	0.276	0.628	0.038					
		10	-10	-1.703	1.794	10	0.00	0.13	0.12	0.00	0.10
			0.124	0.300	0.485	0.091					
		100	-10	-1.085	1.191	10	0.00	0.19	0.18	0.00	0.13
			0.158	0.359	0.348	0.134					
0	1	0.1	-10	-1.286	1.286	10	0.00	0.17	0.17	0.00	0.17
			0.006	0.494	0.494	0.006					
		0.25	-10	-1.266	1.266	10	0.00	0.17	0.17	0.00	0.17
			0.012	0.488	0.488	0.012					
		0.5	-10	-1.239	1.239	10	0.00	0.17	0.17	0.00	0.17
			0.018	0.482	0.482	0.018					
		1	-10	-1.197	1.197	10	0.00	0.18	0.18	0.00	0.17
			0.025	0.475	0.475	0.025					
		10	-10	-0.907	0.907	10	0.00	0.20	0.20	0.00	0.19
			0.039	0.461	0.461	0.039					
		100	-10	-0.558	0.558	10	0.00	0.23	0.23	0.00	0.21
			0.043	0.457	0.457	0.043					

Table 5.7: The Effect of σ^2 on the Obtained A-optimal Designs

α_0	α_1	σ^2	c_1	c_4	w_1	w_4	c_2	c_3	$\lvert c_3\rvert/\lvert c_2\rvert$	$c_3 - c_2$	w_2	w_3	w_3/w_2
1	0.5	0.1	-10	10	0.005	0.003	-1.232	1.273	1.033	2.505	0.195	0.797	4.086
		0.25	-10	10	0.009	0.006	-1.196	1.262	1.055	2.457	0.194	0.791	4.082
		0.5	-10	10	0.014	0.009	-1.165	1.241	1.065	2.406	0.193	0.784	4.067
		1	-10	10	0.019	0.012	-1.125	1.205	1.071	2.330	0.193	0.776	4.022
		10	-10	10	0.032	0.022	-0.766	0.989	1.291	1.755	0.227	0.720	3.175
		100	-10	10	0.036	0.032	-0.271	0.772	2.844	1.043	0.331	0.601	1.816
5	2	0.1	-10	10	0.013	0.008	-2.298	2.337	1.017	4.635	0.294	0.685	2.328
		0.25	-10	10	0.027	0.017	-2.249	2.323	1.033	4.572	0.286	0.671	2.349
		0.5	-10	10	0.041	0.026	-2.196	2.297	1.046	4.492	0.279	0.653	2.339
		1	-10	10	0.059	0.038	-2.125	2.244	1.056	4.369	0.276	0.628	2.276
		10	-10	10	0.124	0.091	-1.703	1.794	1.054	3.497	0.300	0.485	1.618
		100	-10	10	0.158	0.134	-1.085	1.191	1.097	2.277	0.359	0.348	0.970

The case of $\alpha_0 = 5, \alpha_1 = 2$ shows similar results in Table 5.6. To see the trend, we created Table 5.7 regarding the eight values of c_1, c_2, c_3, c_4, w_1, w_2, w_3 and w_4. When σ^2 increases to 0.1, 0.25, 0.5, 1, 10, and 100, the values of the two outer points are always -10 and 10 and the weights of the two outer points increase. Looking at two inner points, as σ^2 increases, the negative value of c_2 moves toward zero and the positive value of c_3 moves toward zero. The value of $c_3 - c_2$ is decreasing and two inner points get closer. This trend is repeated when $\alpha_0 = 5, \alpha_1 = 2$ in the lower part of Table 5.7. When $\alpha_0 = 1$ and $\alpha_1 = 0.5$, the two inner points get more asymmetric and right-skewed when σ^2 increases. We observe this from the values of $|c_3|/|c_2|$ which increase to 1.033, 1.055, 1.065, 1.071, 1.291, and 2.844. However, when $\alpha_0 = 5, \alpha_1 = 2$, the asymmetry of the points are less significant than the previous case where $\alpha_0 = 1, \alpha_1 = 0.5$.

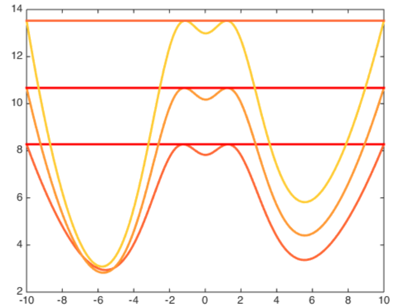
In summary, as σ^2 increases, c_1 and c_4 have the same values, w_1 and w_4 are increasing, c_2 and c_3 get closer to zero, w_2 is decreasing and increasing, and w_3 is decreasing.

When $\alpha_0 = 0, \alpha_1 = 1$ in Table 5.6, we have different designs compared to the previous two cases. In the previous table, Table 5.4, we already saw that the zero value of α_0 gives a symmetric design. Here we observe it again by using other values of σ^2.

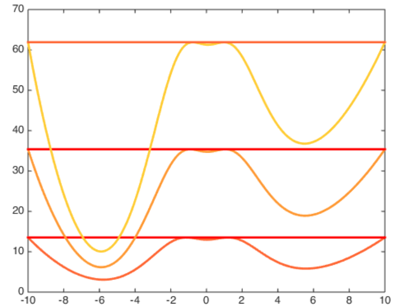
On the other hand, the variance of the binary response z_i is calculated by $\pi_i(1-\pi_i)$ using the obtained c_i and a formula $\pi_i = e^{c_i}/(1+e^{c_i})^2$. In Table 5.6, we see that if σ^2 is increasing, we have larger variances of binary responses together. We see that the variances of mixed responses are positively associated conditioning a binary response.

Using the GET, we validate the A-optimality of selected designs in Figure 5.2. We track the effect of σ^2 using these plots. From Table 5.6, we chose two cases, which are $\alpha_0 = 1, \alpha_1 = 0.5$ and $\alpha_0 = 5, \alpha_1 = 2$. Also, we set three σ^2 values. The left two plots are for $\sigma^2 = 0.1, 0.5, 1$ and the right two plots are for $\sigma^2 = 1, 5, 10$. All designs are A-

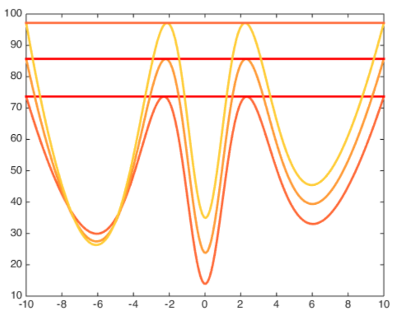
Figure 5.2: A-optimality Verification for Three σ^2 Values



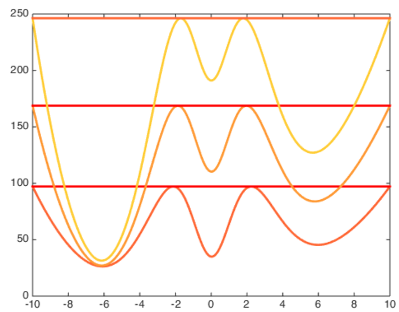
(a) $\alpha_0 = 1, \alpha_1 = 0.5, \sigma^2 = 0.1, 0.5, 1$



(b) $\alpha_0 = 1, \alpha_1 = 0.5, \sigma^2 = 1, 5, 10$



(c) $\alpha_0 = 5, \alpha_1 = 2, \sigma^2 = 0.1, 0.5, 1$



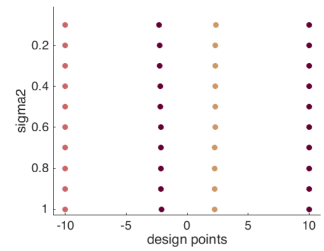
(d) $\alpha_0 = 5, \alpha_1 = 2, \sigma^2 = 1, 5, 10$

* In each plot, a red curve has the smallest value of σ^2 among three values.

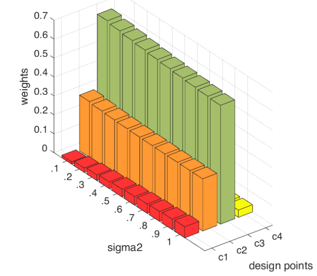
optimal. Since the GET verification depends on the guessed values of α_0, α_1, and σ^2, even the straight lines appear differently. Also, we observe the asymmetric patterns of all curves. When σ^2 increases, the GET plot is also inflated in terms of the value of the constant function C and another function $s(c, \xi)$.

Figure 5.3 summarizes the trend of obtained designs according to the value of σ^2. We set three intervals of σ^2 and generate ten equally spaced values from each interval. For each σ^2, we obtain A-optimal designs with the GET verification. The plots in the first column show the location of design points. As shown in the first three plots, the two inner points get closer as σ^2 increases.

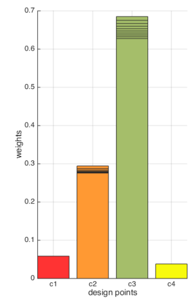
Figure 5.3: A-optimal Designs by σ^2 When $\alpha_0 = 5$ and $\alpha_1 = 2$



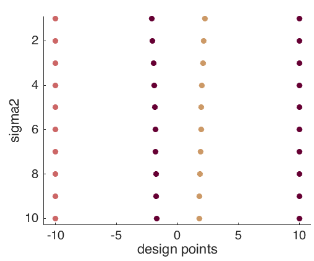
(a) points by $\sigma^2 \in [0.1, 1]$



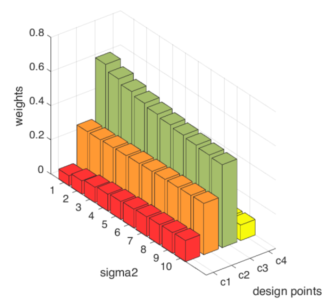
(b) weights by $\sigma^2 \in [0.1, 1]$



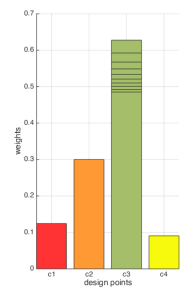
(c) summary of weights



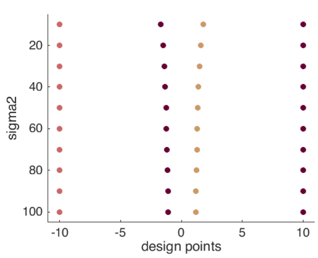
(d) points by $\sigma^2 \in [1, 10]$



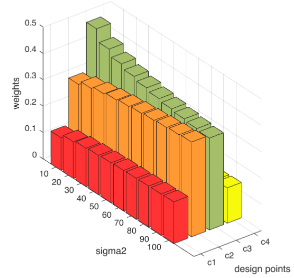
(e) weights by $\sigma^2 \in [1, 10]$



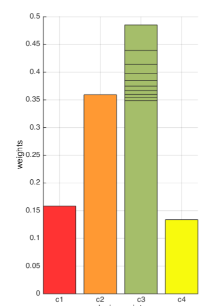
(f) summary of weights



(g) points by $\sigma^2 \in [10, 100]$



(h) weights by $\sigma^2 \in [10, 100]$



(i) summary of weights

* From each $[a, b]$, ten equal-spaced σ^2's are selected.
* (a), (d), and (g): up (small σ^2) $\rightarrow$ down (large σ^2)
* (b), (e), and (h): back (small σ^2) $\rightarrow$ front (large σ^2)

Bar charts in the second and third column explain the weights. The charts in the second column show that the weights of the boundary points increase and the weights of the second largest point c_3 decrease as σ^2 increases. When σ^2 is 100, which is the largest number in our pool of the values of σ^2, we see from the chart of the third row that the two inner points have almost equal weights and the two boundary points have more weights compared to the case when the value of σ^2 is smaller as shown in the first row.

The charts in the third column are the elevations of weights. Through the black horizontal lines in the green bar of w_3, we observe that w_3 continuously decreases while w_1, and w_4, increase. When σ^2 is large, the two boundary points weigh more. It is well known that the optimal points in a linear model experiment are usually the two boundary points that maximize the information matrix. When the variance of the linear model inflates, in other words, σ^2 increases, we assign more weights on the boundary points to account for the variance of continuous responses when we use an A-optimal design for an experiment.

Table 5.8: Design When $\alpha_0 = 0$ and $\alpha_1 = 0$ in $[-10, 10]$

α_0	α_1	σ_2	Design (initial results)*				$var(z_1)$	$var(z_2)$	$var(z_3)$	$var(z_4)$	$\sum w_i var(z_i)$
0	0	0.1	-10	0.000	0.000	10	0.00	0.25	0.25	0.00	0.25
			0.000	0.498	0.502	0.000					
		0.25	-10	0.000	0.000	10	0.00	0.25	0.25	0.00	0.25
			0.000	0.500	0.500	0.000					
		0.5	-10	-0.062	0.065	10	0.00	0.25	0.25	0.00	0.25
			0.001	0.513	0.486	0.001					
		1	-10	0.000	0.000	10	0.00	0.25	0.25	0.00	0.25
			0.000	0.500	0.500	0.000					
		10	-10	0.000	0.000	10	0.00	0.25	0.25	0.00	0.25
			0.000	0.485	0.515	0.000					
		100	-10	0.000	3.008	10	0.00	0.25	0.04	0.00	0.25
			0.000	1.000	0.000	0.000					

* All designs are one-point design with a weight 1.

In Table 5.8, we set $\alpha_0 = 0, \alpha_1 = 0$ to describe a unfortunate situation when we only have one design point $c = 0$ for all c in $[-10, 10]$. From the variances of binary responses, we check that one point design of the experiment will always produce the largest variance of binary response, 0.25, obtained as the maximum of $\pi(1-\pi)$ regardless of the value of σ^2. We check that they are not optimal. It means that in the represented design space, if an experimenter treats only one middle point of the space, there is no chance to get precise experimental data to produce an accurate estimator of parameters for a given model.

Table 5.9: A-optimal Designs for c When $\alpha_0 < 0$

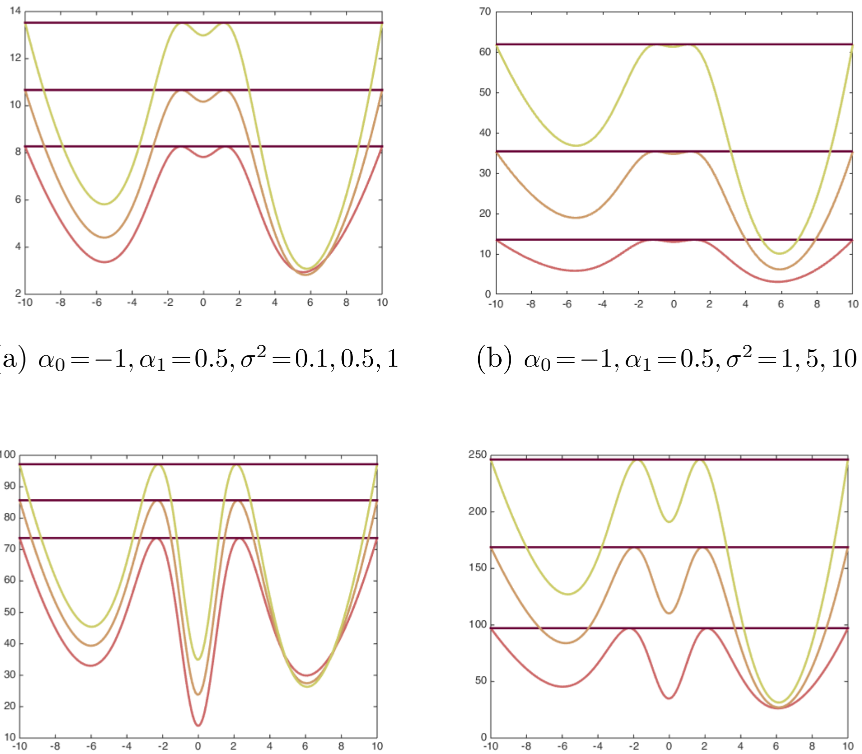
α_0	α_1	σ_2	Design space		$fval$	Design			
1	0.5	1	-10	10	13.53	-10	-1.125	1.205	10
						0.019	0.193	0.776	0.012
-1	0.5	1	-10	10	13.53	-10	-1.205	1.125	10
						0.012	0.776	0.193	0.019
5	2	1	-10	10	97.17	-10	-2.125	2.244	10
						0.059	0.276	0.628	0.038
-5	2	1	-10	10	97.16	-10	-2.244	2.125	10
						0.038	0.628	0.276	0.059

Until now, we set α_0 and α_1 to positive values. We instead discuss a design problem when α_0 is negative. The sign of α_1 does not affect an optimization since the A-optimality criterion only depends on α_1 through $\alpha_0^2+\alpha_1^2$ as shown in (5.6). On the other hand, we found that the sign change of α_0 produces reversed optimal designs as shown in Table 5.9. We verify the A-optimality of the selected designs when $\alpha_0 < 0$ in Figure 5.4.

We explain the reversion of the design points and weights in general using the A-optimality criterion. To see this, we use the fact that $\frac{e^{c_i^*}}{(1+e^{c_i^*})^2}$ is even, such as $\frac{e^{c_i^*}}{(1+e^{c_i^*})^2} = \frac{e^{-c_i^*}}{(1+e^{-c_i^*})^2}$. Also it is true that $\frac{e^{c_i^*}}{1+e^{c_i^*}} = \frac{1}{1+e^{(-c_i^*)}}$. We then see that $-2\alpha_0 \sum w_i^* \frac{c_i^* e^{c_i^*}}{1+e^{c_i^*}} = -2(-\alpha_0) \sum w_i^* \frac{(-c_i)^* \cdot 1}{1+e^{(-c_i)^*}}$ and $|\mathbf{C}_2^*(-c)| = |\mathbf{C}_3^*(c)|$.

Theorem 5.5.1. *In a mixed response model, for any A-optimal design* $\xi^* = \{(c_i^*, w_i^*), i = 1, ..., 4\}$ *under the A-optimality criterion* $\Phi_A^{\alpha_{01}}$ *with* $\alpha_0 = \alpha_{01}$*, a reflected design* $\xi_r^* = \{(-c_i, w_i), i = 1, ..., 4\}$ *is A-optimal with* $\Phi_A^{-\alpha_{01}}$ *with* $\alpha_0 = -\alpha_{01}$.

Figure 5.4: A-optimality Verification for Three σ^2 Values When $\alpha_0 < 0$



(a) $\alpha_0 = -1, \alpha_1 = 0.5, \sigma^2 = 0.1, 0.5, 1$

(b) $\alpha_0 = -1, \alpha_1 = 0.5, \sigma^2 = 1, 5, 10$

(c) $\alpha_0 = -5, \alpha_1 = 2, \sigma^2 = 0.1, 0.5, 1$

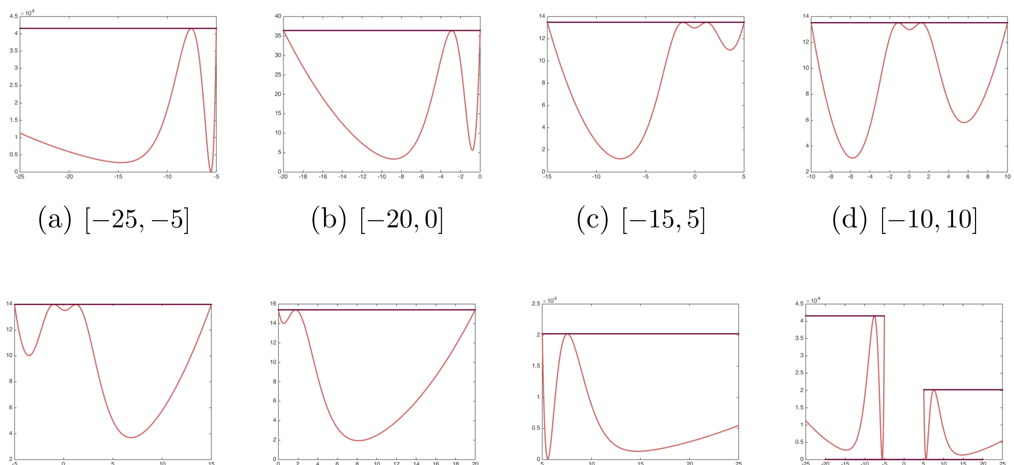
(d) $\alpha_0 = -5, \alpha_1 = 2, \sigma^2 = 1, 5, 10$

Proof. We assume that $\Phi_A^{\alpha_{01}}(\xi^*) \leq \Phi_A^{\alpha_{01}}(\xi)$ for any four-point design $\xi = \{(c_i, w_i), i = 1, 2, 3, 4\}$. Then, an A-optimal design ξ^* achieves the minimum value of the A-optimal criterion $\Phi_A^{\alpha_{01}}$. Also it can be shown using $-c^*$ as follows: $\Phi_A^{\alpha_{01}}(\xi^*(c_i^*)) = \frac{1}{|\mathbf{C}_1^*|}((\alpha_0^2 + \alpha_1^2)\sum w_i \frac{e^{c_i^*}}{(1+e^{c_i^*})^2} - 2\alpha_0 \sum w_i \frac{c_i^* e^{c_i^*}}{(1+e^{c_i^*})^2} + \sum w_i \frac{c_i^{*2} e^{c_i^*}}{(1+e^{c_i^*})^2}) + \frac{\sigma^2}{|\mathbf{C}_2^*|}((\alpha_0^2 + \alpha_1^2)\sum w_i \frac{e^{c_i^*}}{1+e^{c_i^*}} - 2\alpha_0 \sum w_i \frac{c_i^* e^{c_i^*}}{1+e^{c_i^*}} + \sum w_i \frac{c_i^{*2} e^{c_i^*}}{1+e^{c_i^*}}) + \frac{\sigma^2}{|\mathbf{C}_3^*|}((\alpha_0^2 + \alpha_1^2)\sum w_i \frac{1}{1+e^{c_i^*}} - 2\alpha_0 \sum w_i \frac{c_i^*}{1+e^{c_i^*}} + \sum w_i \frac{c_i^{*2}}{1+e^{c_i^*}}) = \frac{1}{|\mathbf{C}_1^*(-c_i^*)|}((-\alpha_0)^2 + \alpha_1^2) \sum w_i \frac{e^{-c_i^*}}{(1+e^{-c_i^*})^2} - 2(-\alpha_0) \sum w_i \frac{-c_i^* e^{-c_i^*}}{(1+e^{-c_i^*})^2} + \sum w_i \frac{(-c_i^*)^2 e^{-c_i^*}}{(1+e^{-c_i^*})^2}) + \frac{\sigma^2}{|\mathbf{C}_3^*(-c_i^*)|}(((-\alpha_0)^2 + \alpha_1^2)\sum w_i \frac{1}{1+e^{-c_i^*}} - 2(-\alpha_0)\sum w_i \frac{-c_i^*}{1+e^{-c_i^*}} + \sum w_i \frac{(-c_i^*)^2}{1+e^{-c_i^*}}) + \frac{\sigma^2}{|\mathbf{C}_2^*(-c_i^*)|}(((-\alpha_0)^2 + \alpha_1^2)\sum w_i \frac{e^{-c_i^*}}{1+e^{-c_i^*}} - 2(-\alpha_0)\sum w_i \frac{-c_i^* e^{-c_i^*}}{1+e^{-c_i^*}} + \sum w_i \frac{(-c_i^*)^2 e^{-c^*}}{1+e^{-c_i^*}} = \Phi_A^{-\alpha_{01}}(\xi^*(-c_i^*))$. The last two equations mean that an optimal design ξ_r^* achieves the minimum of the A-optimal criterion when a new α_0 has $-\alpha_{01}$. Therefore, we conclude that $\Phi_A^{-\alpha_{01}}(\xi_r^*) \leq \Phi_A^{-\alpha_{01}}(\xi_r)$ for any other design $\xi_r = \{(-c_i, w_i), i = 1, 2, 3, 4.\}$. $\square$

Table 5.10: A-optimal Designs for c in Asymmetric Design Spaces

α_0	α_1	σ^2	Design space		Design				# of points	GET verification
1	0.5	1	-25	-5	-7.5616 0.715			-5 0.285	2	success
			-20	0	-20 0.019		-2.888 0.408	0 0.573	3	success
			-15	5	-15 0.016	-1.215 0.207	1.183 0.755	5 0.022	4	success
			-10	10	-10 0.019	-1.125 0.193	1.205 0.776	10 0.012	4	success
			-5	15	-5 0.024	-1.007 0.171	1.228 0.796	15 0.009	4	success
			0	20	0 0.336	1.743 0.657		20 0.007	3	success
			5	25	5 0.313			7.561 0.687	2	success

Figure 5.5: A-optimality Verification in Asymmetric Domains When $\alpha_0 = 1, \alpha_1 = 0.5$, and $\sigma^2 = 1$



(a) $[-25, -5]$ (b) $[-20, 0]$ (c) $[-15, 5]$ (d) $[-10, 10]$

(e) $[-5, 15]$ (f) $[0, 20]$ (g) $[5, 25]$ (h) all spaces

Another interest is the existence of A-optimal designs in asymmetric domains. We set design spaces as from $[-25, -5]$ to $[5, 25]$ as shown in Table 5.10. There is a 2-, 3-, or 4-point design. Weights are not equal. It is noticeable that there exist 4-point designs in $[-15, 5]$ and $[-5, 15]$ which was not found in D-optimal designs. The general equivalence theorem confirms A-optimality in Figure 5.5.

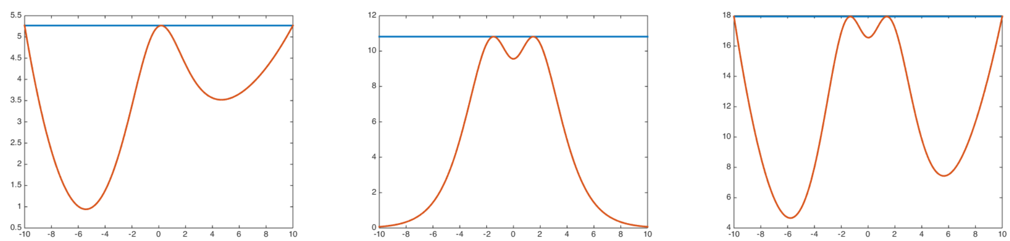
Table 5.11: The A-optimality Criteria Used for Table 5.12

Objective	A-optimility criterion
linear Model	$\mathrm{tr}\,(\mathbf{F}^{\intercal}\mathbf{W}\mathbf{F})^{-\mathbf{1}}$
two conditional LMs	$\sigma^2\,\mathrm{tr}\,(\mathbf{F}^{\intercal}\mathbf{W}\mathbf{P}\mathbf{F})^{-\mathbf{1}} + \sigma^2\,\mathrm{tr}\,(\mathbf{F}^{\intercal}\mathbf{W}(\mathbf{1}-\mathbf{P})\mathbf{F})^{-\mathbf{1}}$
logistic	$\mathrm{tr}\,(\mathbf{F}\mathbf{W}\mathbf{P}(\mathbf{1}-\mathbf{P})\mathbf{F})^{-\mathbf{1}}$
mixed response	$\mathrm{tr}\,(\mathbf{F}\mathbf{W}\mathbf{P}(\mathbf{1}-\mathbf{P})\mathbf{F})^{-\mathbf{1}} + \sigma^2\,\mathrm{tr}\,(\mathbf{F}\mathbf{W}\mathbf{P}\mathbf{F})^{-\mathbf{1}} + \sigma^2\,\mathrm{tr}\,(\mathbf{F}\mathbf{W}(\mathbf{1}-\mathbf{P})\mathbf{F})^{-\mathbf{1}}$

Table 5.12: A-optimal Designs for c under the Different A-optimality Criteria

Objective	α_0	α_1	σ^2	Design space		$fval$	Design			
linear model	none	none	none	-10	10	1.01	-10			10
							0.500			0.500
two conditional LMs	1	1	1	-10	10	5.268	-10		0.182	10
							0.055		0.887	0.057
logistic model	1	1	none	-10	10	10.815		-1.482	1.482	
								0.293	0.707	
mixed response model	1	1	1	-10	10	17.945	-10	-1.326	1.395	10
							0.030	0.282	0.667	0.022

Figure 5.6: A-optimality Verification of the Obtained Designs in Table 5.12

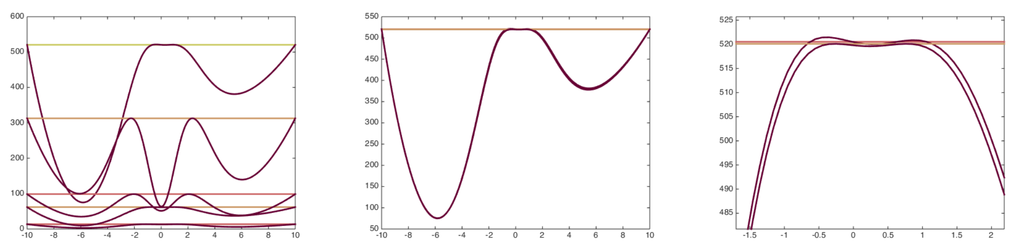


(a) Conditional linear models (b) Logistic regression model (c) Mixed response model

On the other hand, we find A-optimal designs using three different A-optimality criteria embedded in a mixed response model. As shown in Table 5.11, the first is for a linear model experiment, and the second is for two conditional linear model experiments controlled by the probability of a binary response. The third one is for a logistic regression. From three different A-optimality criterion, we obtained A-optimal designs and compared them to A-optimal designs in a mixed response model. The optimality of the designs were verified by the graph as shown in Figure 5.6.

One thing that we want to make a point of is that we used a complete class with a at most 4-point designs including two fixed points as a class of candidate designs just as we did for a mixed response model. The complete class is intended not only for the entire parameters but also a linear combination of the parameters. We thus can focus on the specific parameters of interest using the same complete class. The A-optimality criterion of a logistic regression used is not for a sole logistic regression, but for a part of our mixed response model. We thus use a complete class that is different from the complete class for a sole logistic regression model. In the case of a linear model, the model is not part of a mixed response model. However, if we assume that mixed responses are independent, we can obtain the criterion of the linear model used here after factoring out σ^2. In Table 5.12, we find A-optimal designs for each criterion. We observe that the A-optimal designs for two conditional linear models have three design points and the middle point is 0.1822 with the weight 0.8874.

Figure 5.7: A-optimality Verification of Symmetric Designs in Table 5.13



(a) five cases but the last (b) the third case (c) the third case (part)

Finally, we find a symmetric design under the A-optimality criterion to compare the previously obtained asymmetric designs. We add one more constraint as $c_2+c_3=0$ to the standard mathematical form to make c_2 and c_3 are symmetric. The results are summarized in Table 5.13. For comparison, we set several guessed values of α_0, α_1, and σ^2 in a design space $[-10, 10]$. As shown in Table 5.13, we obtained symmetric designs for every case and they are almost A-optimal since the values of $fval$ are almost the same. The ratio of weights in symmetric designs is slightly different from that of original designs, but the difference is not significant.

Table 5.13: Symmetric Designs for c in [-10, 10]

α_0	α_1	σ^2	Case	Design				$fval$	fval(sym)/fval(org)
1	0.5	1	symmetric	-10	-1.190	1.190	10	13.528	100.01%
				0.019	0.189	0.780	0.012		
			original	-10	-1.125	1.205	10	13.527	
				0.019	0.193	0.776	0.012		
1	0.5	10	symmetric	-10	-0.938	0.938	10	61.92	100.04%
				0.032	0.197	0.748	0.023		
			original	-10	-0.766	0.989	10	61.893	
				0.032	0.227	0.720	0.022		
1	0.5	100	symmetric	-10	-0.630	0.630	10	520.52	100.08%
				0.037	0.183	0.748	0.033		
			original	-10	-0.271	0.772	10	520.11	
				0.036	0.331	0.601	0.032		
1	5	1	symmetric	-10	-2.050	2.050	10	98.652	100.00%
				0.053	0.418	0.480	0.049		
			original	-10	-2.038	2.061	10	98.65	
				0.053	0.416	0.481	0.049		
10	0.5	1	symmetric	-10	-2.297	2.297	10	312.85	100.04%
				0.062	0.341	0.549	0.047		
			original	-10	-2.245	2.330	10	312.74	
				0.062	0.335	0.556	0.047		
20	10	100	symmetric	-10	-1.445	1.445	10	13001	100.02%
				0.217	0.316	0.266	0.200		
			original	-10	-1.501	1.385	10	12998	
				0.218	0.314	0.268	0.200		

The obtained designs are nearly A-optimal by the general equivalence theorem as shown in Figure 5.7. The first plot displays the plots of five symmetric designs obtained in Table 5.13. The plot of the last design is omitted because of its huge scale. The second plot shows the plot of the GET verification for the third case in Table 5.13 when $\alpha_0 = 1, \alpha_1 = 0.5$, and $\sigma^2 = 100$. In the third plot which magnifies the part of the second plot, we see that symmetric designs are slightly less efficient than original designs. So a symmetric design is almost A-optimal, but not exactly A-optimal.

5.6 Discussion

In this study, we found A-optimal designs for a mixed response model. Most designs that we obtained consisted of four points although the exact number of support points depends on the size of the represented design space and guessed values of unknown parameters α_0, α_1, and σ^2. The two outer points were on the boundary of the design space and the two inner points were asymmetrically located. The weights of the four points are unbalanced.

Since the A-optimality criterion has many unknown parameters and those parameters are not separable from the criterion, we considered different combinations of the parameter values in our numerical study. We observed some change in the obtained designs when we changed the values of α_0, and α_1. When α_0 is small compared to the value of α_1, obtained designs were close to symmetric designs.

Our focus was on how the value of σ^2 affects A-optimal designs. The parameter σ^2 is the variance of continuous responses. While the D-optimal design is invariant to the magnitude of σ^2, the A-optimal design hinges on the value of σ^2. It was observed that, when the value of σ^2 was large, the weights of the two outer design points in the obtained design increased while the weights of the two inner points decreased. It seems that when the variance of continuous responses was large, the efficient designs contained two boundary points more than two inner points. The weights of two boundary points increased as σ^2 increased.

Since the trend of σ^2 effect was relatively smooth in a fixed design space, we have useful information on experimental designs under a mixed response model. If an experimenter wants to conduct an experiment considering the change in variances of continuous responses, we recommend A-optimal designs based on our results.

Using an analysis of the logit function, we used a design space at most $[-10, 10]$. The change of the design space from $[-1, 1]$ to $[-10, 10]$ was significant in that the number of support points changed. From here, we observed again that the boundary problem is still an important issue for a design problem as it was observed in the study of D-optimal designs.

Chapter 6

EXTENSION TO A QUADRATIC MODEL

A logistic regression model is one kind of a generalized linear model (GLM). By using the logit function as the link function, we transform a discretized binary variable to a continuous variable so that a logistic model can be handled in a similar way to a linear model in some sense. For example, it is proved that the maximum likelihood estimator of the GLM is equivalent to an iterative weighted least square of the GLM similar to the case of a linear model (Charnes and Yu, 1976). Taking this point into account, we continue discussion as follows.

In classical linear models, when we detect a nonlinear relationship between an input variable and a response, we consider a sophisticated model, beyond that of a simple model, to describe that relationship. One of the common ways is to introduce a polynomial regression model as an approximation (Kutner, Nachtsheim, Neter, and Li, 2004). Moreover, according to Atkinson, Donev, and Tobias (2007), "experience indicates that in very many experiments the response can be described by polynomial models of order no greater than two (in linear models)." Hence, a second-order polynomial model is a good model to start with for investigating a curvature trend of experimental data.

Similar to the linear model, a simple logistic regression model is occasionally not enough to fit data. We found the applications of a quadratic logistic regression model to be such as a carbon disulphide study with beetles, a business management study with initial public offerings (IPOs) data, the Western Collaborative Group Study of coronary heart disease incidence and so on (Collett, 2002; Kutner et. al, 2004; Jewell, 2003). In those examples, a quadratic logistic regression model was used to capture the nonlinear trend. We can, therefore, assume that there may be situations where a quadratic model is needed when modeling binary data collected from an experiment.

For that reason, if we expand the scope of the discussion, it is worthwhile to examine an experimental design for a mixed response model with a quadratic term. A mixed response model consists of one logistic regression model and two conditional linear models. We place a quadratic term in a logistic regression model for the binary response while we set the linear models unchanged. For convenience, we call this model a quadratic mixed response model, and the model in the previous chapters is referred to as a simple mixed response model.

Meanwhile, when an experimenter wants to use a quadratic mixed response model, it presents a real challenge to designing an experiment for such an extended model. Adding one term is not merely a stretch of a simple model, but causes an almost new creation of the covariance-variance matrix of parameter estimates. Consequently, it changes the properties embedded in the covariance-variance matrix or the information matrix which we studied before and this new situation affects our strategy. We need to contemplate a more sophisticated application of a complete class approach, to write a new standard form for a nonlinear algorithm, and to derive a new inequality equation for verification via the general equivalence theorem (GET).

Among them, it is imperative that we examine the applicability of a complete class approach. So far candidate designs were fenced off by a complete class approach, and the burden of a computational search was lightened. If we have to give up such an advantage here, countably many support points can be reduced by the Carathéodory's theorem for now. Then, the upper bound of the number of design points is twenty-nine ($p(p+1)/2+1=7*(8)/2+1=29$) where $p=7$ is the number of parameters that gives the length of the row or column of the symmetric square information matrix. It is evident that computation time will increase and, in reality, twenty-nine different factor levels may not be easy or preferred to be treated when we have no way to collapse design points by another method.

Yu (2011) suggested a 'nearest neighbor exchange' in his cocktail algorithm as one of the solutions to apportionment between close points. This strategy was used to increase the efficiency of the multiplicative algorithm. The cocktail algorithm consisted of three algorithms in sequence such as vertex direction method (VDM), nearest neighbor exchanges (NNE), and multiplicative algorithm (MA). The MA was one of the traditional optimal design algorithms that updated weights based on the GET (Silvey, Titterington and Torsney, 1978). For the precision of results, more support points were recommended, and consequently, the speed of convergence could be slow. Due to the nature of the algorithm, the MA started with much more support points than the points found by Carathéodory's theorem. In this algorithm, 'n' - the number of the support points- did not quickly decrease.

The strategy of the NNE was to exclude a non-support point x_j adjacent to a support point x_i when x_j was the 'nearest neighbor' of x_i based on a certain distance standard. In numerical examples, the neighborhood structure was specified using L_1 norm as $||x_i - x_j||$ and the author stated that the choice of metric did not make much difference in his experience. The NNE was implemented before the MA and then the procedure effectively reduced computing time. During the entire iterations of the cocktail algorithm, support points x, weights w, and the number of support points n, were adjusted together.

In the cocktail algorithm, the number of design points was mostly reduced by the NNE based on a designated metric, and the basic updating rule relied on the GET. In our study, we derive a sharper bound for the number of support points based on a theoretical approach, namely the complete class approach before moving on to a computer search. In the former approach, the grid of a design space was practically considered, but in our approach we use a complete class. In his study, the GET was used as the updating rule for the MA algorithm which is the central part of the cocktail algorithm. In our study, we use the fmincon solver of mathematical programming as a general solution to finding the minimum of a nonlinear objective function given constraints that necessitate the verification procedure by the GET.

After reviewing the previous studies of quadratic models which adopted a complete class approach, we will investigate an optimal experimental design for a quadratic mixed response model. We will show the complete class results and the numerical results under the D-optimality criterion.

6.1 Previous Studies of Quadratic Models

Wu and Stufken (2014) studied locally Φ_p-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor. They considered logistic and probit models for binary responses and a log-linear model for count data. Using two groups of candidate designs, they derived a general pattern of Φ_p-optimal designs. The first group was the class of symmetric designs. This group was detected using an invariance property under the Φ_p-optimality criterion. Within a class of symmetric designs, they identified complete classes setting subdesigns for detecting Chebyshev systems.

In contrast to a simple logistic regression case in Yang and Stufken (2009), an unknown parameter problem was tricky. To confront the problem, Wu and Stufken (2014) used a vertex form of a predictor as $\eta = \theta_0 + \theta_2(x - \theta_1)^2$ where $\theta_0, \theta_1, \theta_2$ are parameters and x is a single independent variable. Under this model formulation, they represented the information matrix in terms of a represented design point d. For a quadratic logistic regression model, they obtained 3- or 4-point designs under Φ_p criterion.

Hyun (2013) also used a complete class approach to finding optimal designs for a different quadratic model. The focus of the study was to find good designs using a probit model with a quadratic term to describe a nonlinear dose-response relationship in toxicology studies. After setting a model using three parameters, the author found complete classes and then searched D- and A- optimal designs. The predictor in the model is $\eta = \theta_1 + \theta_2 x_i + \theta_3 x_i^2$ where $\theta_1, \theta_2, \theta_3$ are parameters and x_i is an independent variable. The complete classes possessed 4-point designs or 5-point designs.

On the other hand, the property of symmetric support points eased the complexity of design problems in many cases such as Wu and Stufken (2014). Mathew and Sinha (2001), and Liski, Mandal, Shah, and Sinha (2002) also used this property for various models. Mathew and Sinha (2001) considered a simple logistic regression. Liski, et al. (2002) studied optimal designs for a polynomial model with a degree $k \geq 1$. They set symmetric designs by using the operation of a reflection. For example, a symmetric design $\{(-x_2, w_2/2), (-x_1, w_1/2), (x_1, w_1/2), (x_2, w_2/2)\}$ was created from $\{(x_1, w_1), (x_2, w_2)\}$ by a reflection; $-x_1$ is a reflected point of x_1. They observed that the value of the Φ_p criterion function was invariant to reflection in their models. In Chapter 13 of Pukelsheim (2006), there was a summary of an invariance property of some optimality criteria. Generally researchers showed that symmetric designs were at least as good as any other designs for the model that they considered. We also use this property in Section 6.3.

6.2 Quadratic Mixed Response Model

To formulate a quadratic mixed response model, we use a vertex form of $\alpha_0 + \alpha_2(x_i - \alpha_1)^2$ in a logistic regression model. Then, the binary response z_i ($=1$ with a probability π_i) has $\pi_i = E(z_i) = \frac{\exp(\alpha_0 + \alpha_2(x_i - \alpha_1)^2)}{1 + \exp(\alpha_0 + \alpha_2(x_i - \alpha_1)^2)}$. We consider conditional linear models without a quadratic term. Then, the continuous response $y_i|z_i$ has $y_i|z_i = 1 \sim N(\mu_1, \sigma^2)$ and $y_i|z_i = 0 \sim N(\mu_2, \sigma^2)$ where $\mu_1 = E(y_i|z_i = 1) = \beta_0^{(1)} + \beta_1^{(1)} x_i$, and $\mu_2 =$

$E(y_i|z_i=0)=\beta_0^{(2)}+\beta_1^{(2)}x_i$. The parameter vector of interest is $\boldsymbol{\theta}=(\alpha_0,\alpha_1,\alpha_2,\beta_0^{(1)},\beta_1^{(1)}$ $\beta_0^{(2)},\beta_1^{(2)})$. The information matrix is

$$\mathbf{M}(\boldsymbol{\theta},x_i)=\begin{bmatrix}\mathbf{F_1}^\intercal\mathbf{WP}(\mathbf{I}-\mathbf{P})\mathbf{F_1} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F_2}^\intercal\mathbf{WPF_2} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \frac{1}{\sigma^2}\mathbf{F_2}^\intercal\mathbf{W}(\mathbf{I}-\mathbf{P})\mathbf{F_2}\end{bmatrix}, \text{ where } \mathbf{F_1}=\begin{bmatrix}\mathbf{f_1}(x_1)^\intercal\\ \mathbf{f_1}(x_2)^\intercal\\ \dots\\ \mathbf{f_1}(x_n)^\intercal\end{bmatrix},$$

and $\mathbf{F}_2=[\mathbf{f_2}(x_1),\mathbf{f_2}(x_2),...,\mathbf{f_2}(x_n)]^\intercal$. Also, $\mathbf{f_1}(x_i)^\intercal=[1,-2\alpha_2(x_i-\alpha_1),(x_i-\alpha_1)^2]$, $\mathbf{f_2}(x_i)^\intercal=[1 \quad x_i]$, $\mathbf{W}=\text{diag}(w_1,...,w_n)$, and $\mathbf{P}=\text{diag}(\pi_1,...,\pi_n)$ with $\pi_i=\frac{\exp(\alpha_0+\alpha_2(x_i-\alpha_1)^2)}{1+\exp(\alpha_0+\alpha_2(x_i-\alpha_1)^2)}$.

Following Wu and Stufken (2014), we define a represented point as $d_i=|\alpha_2|^{1/2}(x_i-\alpha_1)$. Then, the predictor η in the logistic regression model for the binary response is expressed as $\eta_i=\alpha_0+sign(\alpha_2)d_i^2$, and the information matrix for $\boldsymbol{\theta}$ is decomposed as $\mathbf{M}(\boldsymbol{\theta},d_i)=\mathbf{P}(\boldsymbol{\theta},\sigma^2)\left(\sum_{i=1}^n w_i\mathbf{C}(\boldsymbol{\theta},d_i)\right)\mathbf{P}(\boldsymbol{\theta},\sigma^2)^\intercal$ where $\mathbf{P}(\boldsymbol{\theta},\sigma)$ is a 7-by-7 nonsingular matrix that only depends on $\boldsymbol{\theta}$ and σ as

$$\mathbf{P}(\boldsymbol{\theta},\boldsymbol{\sigma})=\text{diag}(\mathbf{P}_1,\ \frac{1}{\sigma}\mathbf{P}_2,\ \frac{1}{\sigma}\mathbf{P}_2),$$

where $\mathbf{P}_1=\text{diag}(1,-\frac{1}{2}\,\text{sign}(\alpha_2)|\alpha_2|^{-1/2},|\alpha_2|)$, $\mathbf{P}_2=\begin{pmatrix}1 & 0\\ -\alpha_1|\alpha_2|^{1/2} & |\alpha_2|^{1/2}\end{pmatrix}^{-1}$ and $\mathbf{C}(\boldsymbol{\theta},d)$ is a 7-by-7 symmetric matrix as

$$\mathbf{C}(\boldsymbol{\theta},d)=\begin{pmatrix}\Psi_{11}(d) & \Psi_{12}(d) & \Psi_{13}(d) & & & & \\ \Psi_{12}(d) & \Psi_{22}(d) & \Psi_{23}(d) & & & & \\ \Psi_{13}(d) & \Psi_{23}(d) & \Psi_{33}(d) & & & & \\ 0 & 0 & 0 & \Psi_{44}(d) & \Psi_{45}(d) & & \\ 0 & 0 & 0 & \Psi_{45}(d) & \Psi_{55}(d) & & \\ 0 & 0 & 0 & 0 & 0 & \Psi_{66}(d) & \Psi_{67}(d)\\ 0 & 0 & 0 & 0 & 0 & \Psi_{67}(d) & \Psi_{77}(d)\end{pmatrix} \tag{6.1}$$

$$=\begin{pmatrix}\frac{e^\eta}{(1+e^\eta)^2} & & & & & & \\ d\frac{e^\eta}{(1+e^\eta)^2} & d^2\frac{e^\eta}{(1+e^\eta)^2} & & & & & \\ d^2\frac{e^\eta}{(1+e^\eta)^2} & d^3\frac{e^\eta}{(1+e^\eta)^2} & d^4\frac{e^\eta}{(1+e^\eta)^2} & & & & \\ 0 & 0 & 0 & \frac{e^\eta}{1+e^\eta} & & & \\ 0 & 0 & 0 & d\frac{e^\eta}{1+e^\eta} & d^2\frac{e^\eta}{1+e^\eta} & & \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{1+e^\eta} & \\ 0 & 0 & 0 & 0 & 0 & d\frac{1}{1+e^\eta} & d^2\frac{1}{1+e^\eta}\end{pmatrix} \tag{6.2}$$

$$=\text{diag}(\mathbf{D}_1\lambda_1,\mathbf{D}_2\lambda_2,\mathbf{D}_2\lambda_3),$$

where $\mathbf{D}_1=\begin{pmatrix}1 & d & d^2\\ d & d^2 & d^3\\ d^2 & d^3 & d^4\end{pmatrix}$, $\mathbf{D}_2=\begin{pmatrix}1 & d\\ d & d^2\end{pmatrix}$, $\lambda_1(\eta)=\frac{e^\eta}{(1+e^\eta)^2}$, $\lambda_2(\eta)=\frac{e^\eta}{1+e^\eta}$, $\lambda_3(\eta)=\frac{1}{1+e^\eta}$ and $\eta=\alpha_0+sign(\alpha_2)d^2$. We observe that it is true that $\lambda_1(-\eta)=\lambda_1(\eta)$, and $\lambda_2(-\eta)=\lambda_3(\eta)$ while it holds that $\lambda_i(\eta(-d))=\lambda_i(\eta(d))$ for $i=1,2,3$.

6.3 Class of Symmetric Designs

Based on Pukelsheim (2006), we will show that for a quadratic mixed response model, a symmetrized design is at least as good as any other design under the D-optimality criterion. In this chapter, a symmetrized, or symmetric design is defined as $\xi_s = \{(\pm d_i, w_i/2), i=1,...,m\}$ where $\sum_{i=1}^{m} w_i = 0.5$.

We create a reflected design $\xi_r = \{(-d_i, w_i), i=1,...,m\}$ from any design $\xi = \{(d_i, w_i), i=1,...m\}$ where $\sum_{i=1}^{m} w_i = 1$. We then find a transformation matrix $\mathbf{Q}$ to satisfy two conditions such that (i) $\mathbf{M}(\xi_r) = \mathbf{Q}\mathbf{M}(\xi)\mathbf{Q}^\intercal$ (ii)$-\log|\mathbf{M}(\xi)| = -\log|\mathbf{Q}\mathbf{M}(\xi)\mathbf{Q}^\intercal|$. If $|\mathbf{Q}| = \pm 1$, $\mathbf{Q}$ satisfies two conditions. We then state that the D-optimality criterion is invariant under the action of $\mathbf{Q}$ which holds that $\Phi_D(\mathbf{Q}\mathbf{M}\mathbf{Q}^\intercal) = \Phi_D(\mathbf{M})$ for all nonnegative definite matrices $\mathbf{M}$ (Pukelsheim, 2006).

In (6.2), we see that $\mathbf{C}(\xi(-d))$ has the same element functions as $\mathbf{C}(\xi(d))$ for the diagonal elements and the components of (1,3), (3,1) where (i, j) indicates the element of the ith row and jth column of a matrix. The elements of (1,2), (2,1), (2,3), (3,2), (4,5), (5,4), (6,7), (7,6) in $\mathbf{C}(\xi(-d))$ are obtained by changing the sign of the corresponding elements of $\mathbf{C}(\xi(d))$. From this observation, we know that $\mathbf{M}(\xi_s)$ has zero element in the locations of (1,2), (2,1), (2,3), (3,2), (4,5), (5,4), (6,7), and (7,6) and find a matrix $\mathbf{Q}$. We have the following result.

Lemma 6.3.1. *For an arbitrary design $\xi = \{(d_i, w_i), i=1,...,m\}$ in the design space [-D, D] for some $D > 0$, a symmetrized design $\xi_s = \{(\pm d_i, w_i/2), i=1,...,m\}$ is at least as good as the design ξ under the D-optimality criterion.*

Proof. For a matrix$\mathbf{Q} = \text{diag}(1, -1, 1, 1, -1, 1, -1)$, we have $\det(\mathbf{Q}) = -1$. Hence, it holds that $-\log|\mathbf{C}(\xi_r(d))| = -\log|\mathbf{Q}\mathbf{C}(\xi(d))\mathbf{Q}^\intercal| = -\log|\mathbf{C}(\xi(d))|$ for a reflected design $\xi_r = \{(-d_i, w_i), i=1,...,m\}$ of any design $\xi = \{(d_i, w_i), i=1,...m\}$ based on Section 13.7-8 in Pukelsheim (2006). For a symmetrized design $\xi_s = \{(\pm d_i, w_i/2), i=1,...,m\}$, using the convexity of $-\log|\cdot|$, we see that $-\log|\mathbf{M}(\xi_s)| = -\log|1/2\mathbf{M}(\xi)+1/2\mathbf{M}(\xi_r)| \leq -1/2\log|\mathbf{M}(\xi)| - 1/2\log|\mathbf{M}(\xi_r)| = -1/2\log|\mathbf{M}(\xi)| - 1/2\log|\mathbf{PQC}(\xi)\mathbf{Q}^\intercal\mathbf{P}^\intercal| = -1/2\log|\mathbf{M}(\xi))| - 1/2\log|\mathbf{P}||\mathbf{C}(\xi)||\mathbf{P}^\intercal| = -\log|\mathbf{M}(\xi)|$. $\square$

A simple mixed response model also has the property of D-invariance. In that case, we have $\mathbf{Q} = \begin{pmatrix} \mathbf{Q}^0 & 0 & 0 \\ 0 & 0 & \mathbf{Q}^0 \\ 0 & \mathbf{Q}^0 & 0 \end{pmatrix}$ with $\mathbf{Q}^0 = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. This partly explains our observation that the obtained D-optimal designs for a simple model were symmetric designs.

6.4 Complete Class Results

In this section, we identify complete classes within the class of symmetric designs based on Lemma 6.3.1. A symmetric design is now expressed as $\xi_s = \{(\pm d_i, w_i), d_i \geq 0, w_i \geq 0, i = 1, ..., m\}$ where d_i is a design point, w_i gives the weights for d_i and $-d_i$, and $\sum_{i=1}^{m} w_i = 0.5$. When we have $\pm d_i = 0$, the number of design points is odd. Also, we define a nonnegative subdesign as $\xi^+ = \{(d_i, w_i), d_i \geq 0, w_i > 0, i = 1, ..., m\}$ and a nonpositive subdesign as $\xi^- = \{(-d_i, w_i), d_i \geq 0, w_i > 0, i = 1, ..., m\}$. Here, we follow Wu and Stufken (2014) to relax the condition of $\sum w_i = 1$ to $\sum w_i = .5$ when we call ξ^+ or ξ^- a subdesign.

A reason for considering subdesigns is that, in a quadratic mixed response model, we have $f_{l,l}(d) = 0$ at $d = 0$ when we apply Theorem 2 of Yang and Stufken (2012). Consequently, a complete class cannot be formed in the entire design space $[-D, D]$. We thus separately consider two subdesigns in design spaces $[-D, 0]$ and $[0, D]$, respectively and identify the pair of two Chebyshev systems for each domain. Combining the results in two domains, we identify a complete class for the entire design space. Wu and Stufken (2004) used the same approach.

Based on Lemma 2.2.1 that we suggested in Chapter 2, we find Chebyshev systems by using $F(d)$. We first make a maximal set of necessary elements from $\mathbf{C}$. We use the notations as follows.

$$\mathbf{C}(\boldsymbol{\theta},d)=\begin{pmatrix} \lambda_1(d) & d\lambda_1(d)^* & d^2\lambda_1(d) & 0 & 0 & 0 & 0\\ d\lambda_1(d)^* & d^2\lambda_1(d) & d\lambda_1(d)^* & 0 & 0 & 0 & 0\\ d^2\lambda_1(d) & d^3\lambda_1(d)^* & d^4\lambda_1(d) & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \lambda_2(d) & d\lambda_2(d)^* & 0 & 0\\ 0 & 0 & 0 & d\lambda_2(d)^* & d^2\lambda_2(d) & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \lambda_3(d) & d\lambda_3(d)^*\\ 0 & 0 & 0 & 0 & 0 & d\lambda_3(d)^* & d^2\lambda_3(d) \end{pmatrix} \tag{6.3}$$

We do not consider even functions indicated by (*) in (6.3) as an element of a maximal set since within a class of symmetric designs ξ_s's, the corresponding elements of the information matrix vanish such as $\sum w_i d\lambda_1(d) = \sum w_i d^3\lambda_1(d) = \sum w_i d\lambda_2(d) = \sum w_i d\lambda_3(d) = 0$. We then have the equality of $\sum w_k \Psi_l(d) = \sum \tilde{w}_k \tilde{\Psi}_l(d)$ in (2.5) as "0 = 0" all the time. We delete four even functions.

Therefore, we have a set of

$$\{\lambda_1(d), d^2\lambda_1(d), d^4\lambda_1(d), d^2\lambda_2(d), \lambda_3(d), d^2\lambda_3(d)\}. \tag{6.4}$$

We already deleted $\lambda_2(d)$ due to the relationship of $\lambda_2(d) + \lambda_3(d) = \Psi_0(=1)$. The set in (6.4) corresponds to $\{\Psi_{11}, \Psi_{22}$ or $\Psi_{13}, \Psi_{33}, \Psi_{55}, \Psi_{66}, \Psi_{77}\}$ in (6.1). The index ii of Ψ_{ii} denotes a location in $\mathbf{C}$ while i of Ψ_i will express the index of a sequence for the search for Chebyshev systems. We started the search with a 3-by-3 matrix of $\mathbf{C}_{22}$ considering the results in Chapter 3, but unfortunately could not form a complete class.

Table 6.1: Candidate Sequences for the Complete Class Approach

candidates	$\Psi_1, ..., \Psi_4$	Ψ_5
Sequence 1	$\frac{1}{1+e^\eta}, \frac{e^\eta}{(1+e^\eta)^2}, \frac{d^2e^\eta}{(1+e^\eta)^2}, \frac{d^2}{1+e^\eta}$	$\mathbf{C}_{22} = \text{diag}(d^4\frac{e^\eta}{(1+e^\eta)^2}, d^2\frac{e^\eta}{1+e^\eta}),$
Sequence 2	$\frac{1}{1+e^\eta}, \frac{e^\eta}{(1+e^\eta)^2}, \frac{d^2e^\eta}{(1+e^\eta)^2}, \frac{d^2e^\eta}{1+e^\eta}$	$\mathbf{C}_{22} = \text{diag}(d^4\frac{e^\eta}{(1+e^\eta)^2}, d^2\frac{1}{1+e^\eta})$
Sequence 3	$\frac{1}{1+e^\eta}, \frac{e^\eta}{(1+e^\eta)^2}, \frac{d^4e^\eta}{(1+e^\eta)^2}, \frac{d^2e^\eta}{(1+e^\eta)^2}$	$\mathbf{C}_{22} = \text{diag}(d^2\frac{e^\eta}{1+e^\eta}, d^2\frac{1}{1+e^\eta})$
Sequence 4	$\frac{1}{1+e^\eta}, \frac{e^\eta}{(1+e^\eta)^2}, \frac{d^2e^\eta}{(1+e^\eta)^2}, \frac{d^4e^\eta}{(1+e^\eta)^2}$	$\mathbf{C}_{22} = \text{diag}(d^2\frac{e^\eta}{1+e^\eta}, d^2\frac{1}{1+e^\eta})$

We thus consider a 2-by-2 matrix $\mathbf{C}_{22}$ using two elements among $d^2\lambda_1(d), d^4\lambda_1(d), d^2\lambda_2(d)$, and $d^2\lambda_3(d)$. After selecting one type of $\mathbf{C}_{22}$, we permute the remaining elements. The number of candidate sequences is $6! \times 4!(=$ possible ways of creating

Table 6.2: Calculation of $F(d)$ for Candidate Sequences (s : sign(α_2))

candidates	$f_{1,1}$	$f_{2,2}$	$f_{3,3}$	$f_{4,4}$	$\prod f_{i,i}$	$f_{5,5}$
Seq. 1	$-\frac{2de^{\eta}s}{(1+e^{\eta})^2}$	$\frac{4de^{\eta}s}{(1+e^{\eta})^2}$	$de^{-\eta}(1+e^{\eta})^2$	$4de^{-\eta}s$	$-\frac{32d^4s^3}{(1+e^{\eta})^2}$	$\text{diag}(2de^{\eta}, -4de^{2\eta}s)$
Seq. 2	$-\frac{2de^{\eta}s}{(1+e^{\eta})^2}$	$\frac{4de^{\eta}s}{(1+e^{\eta})^2}$	$de^{-\eta}(1+e^{\eta})^2$	$-4de^{\eta}s$	$\frac{32d^4e^{2sd^2+2a}s^3}{(1+e^{\eta})^2}$	$\text{diag}(2de^{-\eta}, 4de^{-2\eta}s)$
Seq. 3	$-\frac{2de^{\eta}s}{(1+e^{\eta})^2}$	$\frac{4de^{\eta}s}{(1+e^{\eta})^2}$	$2d^3e^{-\eta}(1+e^{\eta})^2$	$-\frac{1}{d^3}$	$\frac{16d^2e^{\eta}s^2}{(1+e^{\eta})^2}$	$\text{diag}(4d^3e^{\eta}s^2, 4d^3e^{-\eta}s^2)$
Seq. 4(*)	$-\frac{2de^{\eta}s}{(1+e^{\eta})^2}$	$\frac{4de^{\eta}s}{(1+e^{\eta})^2}$	$de^{-\eta}(1+e^{\eta})^2$	$4d$	$-\frac{32d^4e^{\eta}s^2}{(1+e^{\eta})^2}$	$\text{diag}(-2de^{\eta}s^2, -2de^{-\eta}s^2)$

$\mathbf{C}_{22}$ $\times$ the number of permutation of four remaining elements). Among them, we select the four sequences of Ψ functions shown in Table 6.1 for further investigations. Other sequences of Ψ functions tend to involve rather complex $f_{l,l}$ for some l,

Furthermore, from Table 6.1, we choose Sequence 4, which is $\Psi_1 = \frac{1}{1+e^{\eta}}, \Psi_2 = \frac{e^{\eta}}{(1+e^{\eta})^2}, \Psi_3 = \frac{d^2e^{\eta}}{(1+e^{\eta})^2}, \Psi_4 = \frac{d^4e^{\eta}}{(1+e^{\eta})^2}, \Psi_5 = \text{diag}(\frac{d^2e^{\eta}}{1+e^{\eta}}, \frac{d^2}{1+e^{\eta}})$, that is to say $\Psi_1 = \lambda_3(d)$, $\Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d), \Psi_5 = \text{diag}(d^2\lambda_2(d), d^2\lambda_3(d))$. Other sequences may perhaps be used, but Sequence 4 seems to give simple results.

Table 6.3: Sign of $f_{l,l}$ in Our Choice

conditions	$f_{1,1}$	$f_{2,2}$	$f_{3,3}$	$f_{4,4}$	$\prod f_{1,1}$	$f_{5,5}$
Sequence 4	$-\frac{2de^{\eta}s}{(1+e^{\eta})^2}$	$\frac{4de^{\eta}s}{(1+e^{\eta})^2}$	$de^{-\eta}(1+e^{\eta})^2$	$4d$	$-\frac{32d^4e^{\eta}s^2}{(1+e^{\eta})^2}$	$\text{diag}(-2de^{\eta}s^2, -2de^{-\eta}s^2)$
Sequence 4	$-\text{sign}(d)\,\text{sign}(\alpha_2)$	$\text{sign}(d)\,\text{sign}(\alpha_2)$	$\text{sign}(d)$	$\text{sign}(d)$	$-$	$\text{diag}(-\text{sign}(d), -\text{sign}(d))$
$\alpha_2 > 0, d > 0$	$-$	$+$	$+$	$+$	$-$	$\text{diag}(-,-) \rightarrow f_{5,5} < 0$
$\alpha_2 > 0, d < 0$	$+$	$-$	$-$	$-$	$-$	$\text{diag}(+,+) \rightarrow f_{5,5} > 0$
$\alpha_2 < 0, d > 0$	$+$	$-$	$+$	$+$	$-$	$\text{diag}(-,-) \rightarrow f_{5,5} < 0$
$\alpha_2 < 0, d < 0$	$-$	$+$	$-$	$-$	$-$	$\text{diag}(+,+) \rightarrow f_{5,5} > 0$

Tables 6.2 and 6.3 provide $f_{l,l}$ and $F(d)$ of Sequence 4. As it turns out, we still need to consider some unknown parameters because η depends not only on the represented design point d, but also unknown parameters which was not the case of a simple mixed response model. Specifically, these unknown parameters are α_0 and α_2 of $\eta = \alpha_0 + \text{sign}(\alpha_2)d^2$. We need to use guessed values of these parameters when identifying Chebyshev systems.

Fortunately, we do not need to be concerned about the value (or more precisely, the sign) of α_0. In our case, the sign of $f_{l,l}$'s do not depend on α_0 as shown in Table

6.3. This is in contrast to Wu and Stufken (2014); some $f_{l,l}$'s in their calculation contained $(e^{\eta} - e^{-\eta})$ or a similar form, so that the value of α_0 played a role when determining the sign of such $f_{l,l}$'s. For α_2, although we still need to take the sign of α_2 into account as shown in Table 6.3, it is true that we have the same results when $\alpha_2 > 0$ and $\alpha_2 < 0$ based on Lemma 2.2.1 since the signs of $F(d)$ are the same in two cases. With Table 6.3, we have the following as a result.

Lemma 6.4.1. *(1) When $d \in [0, \infty)$, for a quadratic mixed response model, up to a change of signs of some Ψ_l, $l = 1, ..., 5$, $\{\Psi_0, \Psi_1 = \lambda_3(d), \Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d)\}$ and $\{\Psi_0, \Psi_1 = \lambda_3(d), \Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d), \Psi_5^Q\}$ form Chebyshev systems for any non-zero vector Q. Here, $\Psi_0 = 1$ and $\Psi_5^Q = Q^\intercal \operatorname{diag}(d^2\lambda_2(d), d^2\lambda_3(d))Q$. (2) When $d \in (-\infty, 0]$, for a quadratic mixed response model, up to a change of signs of some Ψ_l, $l = 1, ..., 5$, $\{\Psi_0, \Psi_1 = \lambda_3(d), \Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d)\}$ and $\{\Psi_0, \Psi_1 = \lambda_3(d), \Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d), \Psi_5^Q\}$ form Chebyshev systems for any non-zero vector Q. Here, $\Psi_0 = 1$ and $\Psi_5^Q = Q^\intercal \operatorname{diag}(d^2\lambda_2(d), d^2\lambda_3(d))Q$.*

Proof. (1) When we consider any of the two sets using the elements, $\Psi_0, \Psi_1 = \lambda_3(d), \Psi_2 = \lambda_1(d), \Psi_3 = d^2\lambda_1(d), \Psi_4 = d^4\lambda_1(d), \Psi_5 = \operatorname{diag}(d^2\lambda_2(d), d^2\lambda_3(d))$, we have $f_{5,5} < 0$ since the (1,1) component of $f_{5,5}$ is $-2de^{\eta}\operatorname{sign}(\alpha_2)^2 < 0$ and the (2,2) component of $f_{5,5}$ is $-2de^{-\eta}\operatorname{sign}(\alpha_2)^2 < 0$ in $d \in (0, \infty)$. Also, we have $\prod_{i=1}^{4} f_{i,i} = -\frac{32d^4e^{\eta}\operatorname{sign}(\alpha_2)^2}{(1+e^{\eta})^2} < 0$. We then verify $F(d) > 0$. Based on Lemma 2.2.1, we have Chebyshev systems. (2) The proof is similar to the case of (1). We omit it. □

From Lemma 6.4.1, we form a complete class in a quadratic mixed response model as follows.

Theorem 6.4.2. *For a quadratic mixed response model, in a design space $[-D, D]$, there exists a complete class of designs with at most 6 design points including $-D$*

and D. The three pairs of points located symmetrically around zero and the weights for each pair are equal.

Proof. Based on Lemma 6.4.1 and Lemma 2 of Yang and Stufken (2012), for any nonnegative subdesign $\xi^+=\{(d_i,w_i),i=1,...,m\}$ when $m\geq 3$, we can find a subdesign $\xi^{+*}=\{(d_i^*,w_i^*),i=1,2,3\}$ including D as one of points, d_i^*, that satisfies $\sum_{i=1}^{3} w_i^*\Psi_l(d_i^*)=\sum_{i=1}^{m} w_i\Psi_l(d_i)$, $l=0,1,...,4$, and $\sum_{i=1}^{3} w_i^*\Psi_5^Q(d_i^*)>\sum_{i=1}^{m} w_i\Psi_5^Q(d_i)$ for every nonzero vector Q. Then, we have a set S^{+*} of a nonnegative subdesign ξ^{+*} with at most 3 design points, including D, in $[0,D]$. Similarly, for any nonpositive subdesign $\xi^-=\{(-d_i,w_i),i=1,...,m\}$ when $m\geq 3$, we have a set S^{-*} of a nonpositive subdesign ξ^{-*} with at most 3 design points, including $-D$ in $[-D,0]$. Similar to the proof of Lemma 3.2.1, we have $\sum_{i=1}^{3} w_i^*\Psi_{44}(d_i^*)=\sum_{i=1}^{m} w_i\Psi_{44}(d_i)$ and $\sum_{i=1}^{3} w_i^*\Psi_{44}(-d_i^*)=\sum_{i=1}^{m} w_i\Psi_{44}(-d_i)$ by using $\Psi_{44}=1-\Psi_{66}$. We then consider a combined design $\xi^*=\{(-D,0.5-(w_1^*+w_2^*)),(-d_2^*,w_2^*),(-d_1^*,w_1^*),(d_1^*,w_1^*),(d_2^*,w_2^*),(D,0.5-(w_1^*+w_2^*))\}$ in $[-D,D]$. Denoting $\xi^*=\{(\pm d_i^*,w_i^*),d_i^*\geq 0,w_i^*\geq 0,i=1,2,3\}$ and omitting $\boldsymbol{\theta}$ from $\mathbf{C}$ and its partitioned matrices for convenience, we see that $\sum_{i=1}^{3} w_i^*\mathbf{C_{11}}(d_i^*)+\sum_{i=1}^{3} w_i^*\mathbf{C_{11}}(-d_i^*)=\sum_{i=1}^{m} w_i\mathbf{C_{11}}(d_i)+\sum_{i=1}^{m} w_i\mathbf{C_{11}}(-d_i)$, $\sum_{i=1}^{3} w_i^*\mathbf{C_{12}}(\boldsymbol{\theta},d_i^*)+\sum_{i=1}^{3} w_i^*\mathbf{C_{12}}(-d_i^*)=\sum_{i=1}^{m} w_i\mathbf{C_{12}}(d_i)+\sum_{i=1}^{m} w_i\,\mathbf{C_{12}}(-d_i)$, and $\sum_{i=1}^{3} w_i^*\mathbf{C_{22}}(d_i^*)+\sum_{i=1}^{3} w_i^*\mathbf{C_{22}}(-d_i^*)\geq\sum_{i=1}^{m} w_i\mathbf{C_{22}}(d_i)+\sum_{i=1}^{m} w_i\mathbf{C_{22}}(-d_i)$. We then have $\mathbf{M}(\xi^*)\succeq\mathbf{M}(\xi_s)$ and the conclusion follows. □

Using this complete class, we find D-optimal designs in the next section.

6.5 Numerical Results

Based on the results in the previous section, we set a decision vector to $\boldsymbol{\xi}^0=(d_1,d_2,w_1,w_2,w_3)$ with a fixed point $d_3=D$ in a design space $[-D,D]$. We use an objective function as $-\log|\mathbf{C}_1^*|-\log|\mathbf{C}_2^*|-\log|\mathbf{C}_3^*|$ where $\mathbf{C}_1^*=\sum w_i\mathbf{D}_1\lambda_1(\eta)$, $\mathbf{C}_2^*=\sum w_i\mathbf{D}_2\lambda_2(\eta)$, $\mathbf{C}_3^*=\sum w_i\mathbf{D}_2\lambda_3(\eta)$ similar to the case in Chapter 4. However, we need guessed values of α_0 and α_2 because of '$\eta = \alpha_0 + \text{sign}(\alpha_2)d^2$' in an objective function. Then, the formulation is written as "$\underset{\boldsymbol{\xi^0}}{\text{minimize}} \; -\log|\mathbf{C}_1^*| - \log|\mathbf{C}_2^*| - \log|\mathbf{C}_3^*|$ subject to $0 \le d_i \le D$ for $i = 1, 2$, $0 \le w_i \le 0.5$ and $\sum w_i = 0.5$ for $i = 1, 2, 3$."

Table 6.4: A Comparison of Computing Time Between Two Settings of w_3

Decision variables	Constrains	Computing time of five trials (seconds.)	Mean
w_1, w_2, w_3	$0 \le w_i \le 0.5, \sum w_i = 1$	7.22, 7.11, 7.16, 7.14, 7.15	7.16
w_1, w_2	$0 \le w_i \le 0.5$	21.53, 21.37, 21.43, 21.73, 21.80	21.57

* The design space is $[-10, 10]$

Instead of "$0.5 - w_1 - w_2$", we use w_3 and add a constraint of $\sum w_i = 0.5$ since it is more efficient as shown in Table 6.4. According to Nash (2014), mathematical programming is designed to focus on efficiently satisfying many constraints. The results in Table 6.4 are consistent with his statement. It is worthwhile to check the computation time since the time can be longer than in the previous study due to calculation of the determinant of a 3-by-3 $\mathbf{C}_1^*$.

D-optimality verification is done by the following proposition. As mentioned in Chapter 4, $\mathbf{C}_1^*, \mathbf{C}_2^*$, and $\mathbf{C}_3^*$ are determined by the values of obtained designs. The formulation is similar to Proposition 4.2.1 in a simple mixed response model, but we need guess values of α_0 and α_2 since $\eta = \alpha_0 + \text{sign}(\alpha_2)d^2$ has two unknown parameters. We previously used $\eta = c$ without any unknown parameters in a simple model.

Proposition 6.5.1. *For a quadratic mixed responses model, we verify the D-optimality of an obtained design ξ if it holds that*

$$d(d,\xi) = \frac{e^\eta}{(1+e^\eta)^2}\mathbf{f}_1(d)^\intercal \mathbf{C}_1^{*-1}(\xi^*)\mathbf{f}_1(d)^\intercal + \frac{e^\eta}{(1+e^\eta)}\mathbf{f}_2(d)^\intercal \mathbf{C}_2^{*-1}(\xi^*)\mathbf{f}_2(d)^\intercal + \frac{1}{(1+e^\eta)}\mathbf{f}_2(d)^\intercal \mathbf{C}_3^{*-1}(\xi^*)\mathbf{f}_2(d)^\intercal$$

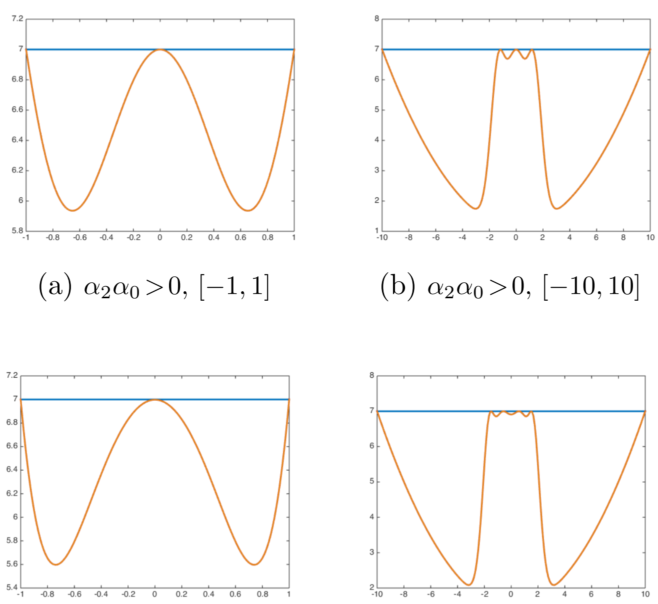
is equal to or less than 7 based on the general equivalence theorem, where $\mathbf{f}_1(d) = [1 \;\; d \;\; d^2]^\intercal, \mathbf{f}_2(d) = [1 \;\; d]^\intercal$ and $\eta = \alpha_0 + \text{sign}(\alpha_2)d^2$ for all d in $[A, B]$. For a D-optimal design $\xi^* = \{(d_i^*, w_i^*), i = 1, ..., 6, w_i \ge 0\}$, we have $d(d^*, \xi^*) = 7$.

Now we start to find D-optimal designs for a quadratic mixed model by implementing an algorithm. Firstly, we check if the guessed values of α_0 and α_2 affect

Table 6.5: D-optimal Designs for d by Values of α_0 and α_2

α_2	α_0	Design space	$fval$	Design						# of pts
1	1	$[-1,1]$	12.94	-1		0			1	3
				0.386		0.228			0.386	
		$[-10,10]$	10.34	-10	-1.180	0		1.180	10	5
				0.082	0.290	0.257		0.290	0.082	
-1	-1	$[-1,1]$	12.94	-1		0			1	3
				0.386		0.228			0.386	
		$[-10,10]$	10.34	-10	-1.180	0		1.180	10	5
				0.082	0.290	0.257		0.290	0.082	
1	-1	$[-1,1]$	9.68	-1		0			1	3
				0.401		0.197			0.401	
		$[-10,10]$	5.01	-10	-1.499	-0.578	0.578	1.499	10	6
				0.085	0.268	0.147	0.147	0.268	0.085	
-1	1	$[-1,1]$	9.68	-1		0			1	3
				0.401		0.197			0.401	
		$[-10,10]$	5.01	-10	-1.499	-0.578	0.578	1.499	10	6
				0.085	0.268	0.147	0.147	0.268	0.085	

Figure 6.1: D-optimality Verification for Designs in Table 6.5



(a) $\alpha_2\alpha_0 > 0$, $[-1,1]$

(b) $\alpha_2\alpha_0 > 0$, $[-10,10]$

(c) $\alpha_2\alpha_0 < 0$, $[-1,1]$

(d) $\alpha_2\alpha_0 < 0$, $[-10,10]$

optimization. In this study, although we are searching for D-optimal designs, we have an unknown parameter problem because the matrix $\mathbf{C}$ itself contains an unknown parameter in (6.2) regardless of the choice of a criterion where the situation is different from the case for a simple mixed response model.

In Table 6.5, we find that the results of $(\alpha_0, \alpha_2) = (1, 1)$ and $(-1, 1)$ are the same and the results of $(\alpha_0, \alpha_2) = (1, -1)$ and $(-1, 1)$ are the same. Hence, it is enough to consider only two conditions $\alpha_0\alpha_2 > 0$ and $\alpha_0\alpha_2 < 0$. It can be explained as follows.

For example, when $\alpha_0\alpha_2 \geq 0$, we set two types of η such as $|\alpha_0| + d^2$ and $-(|\alpha_0| + d^2)$ from $\eta_i = \alpha_0 + sign(\alpha_2)d_i^2$. We then see that $|\alpha_0| + d^2$ and $-(|\alpha_0| + d^2)$ share the same objective function that produces the same results. Let us denote two predictors as η_1, and η_2. Denoting an objective function that we consider as Φ_D^0, we have $\Phi_D^0(\eta_1) = -\log \sum w_i \mathbf{D}_1 \lambda_1(\eta_1) - \log \sum w_i \mathbf{D}_2 \lambda_2(\eta_1) - \log \sum w_i \mathbf{D}_2 \lambda_3(\eta_1) = -\log \sum w_i \mathbf{D}_1 \lambda_1(-\eta_1) - \log \sum w_i \mathbf{D}_2 \lambda_3(-\eta_1) - \log \sum w_i \mathbf{D}_2 \lambda_2(-\eta_1) = -\log \sum w_i \mathbf{D}_1 \lambda_1(\eta_2) - \log \sum w_i \mathbf{D}_2 \lambda_3(\eta_2) - \log \sum w_i \mathbf{D}_2 \lambda_2(\eta_2) = \Phi_D^0(\eta_2)$ using the fact that $\eta_1 = -\eta_2$, $\lambda_1(-\eta_1) = \lambda_1(\eta_1)$, and $\lambda_2(-\eta_1) = \lambda_3(\eta_1)$. Based on this result, for convenience, we continuously assume α_0 is positive and consider two cases including $\alpha_2 > 0$ and $\alpha_2 < 0$.

In addition, for α_2, we only care about the sign since sign(α_2) is a sole term for α_2. We use a vertex form of $\alpha_0 + \alpha_2(x_i - \alpha_1)^2$, α_0 is an intercept in a predictor and α_2 is the coefficient of a quadratic term. We understand that the curve opens upward if $\alpha_2 > 0$ and downward if $\alpha_2 < 0$.

Based on the above settings, we find D-optimal designs under the two cases of α_2; $\alpha_2 > 0$ and $\alpha_2 < 0$. In Table 6.6, when α_2 is positive, D-optimal designs have three or five points. On the other hand, when we have a negative value of α_2, we have 3-, 4-, or 6-point D-optimal designs as shown in Table 6.7.

For the former case, when the design space is narrower than $[-1.4, 1.4]$, two boundary points and one zero point form an optimal design while when the design space is

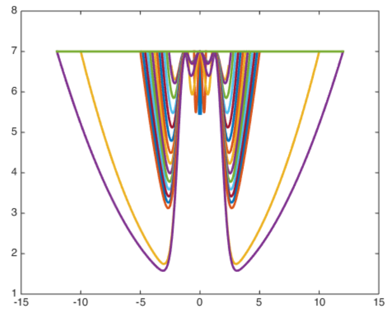
Table 6.6: D-optimal Designs for d When $\alpha_0 = 1$ and $\alpha_2 > 0$

α_2	α_0	Design space		$fval$	Design					# of pts	d_1
positive	1	-0.1	0.1	33.68	-0.1		0.000		0.1	3	-0.1
					0.400		0.200		0.400		
		-0.5	0.5	18.05	-0.5		0.000		0.5	3	-0.5
					0.398		0.205		0.397		
		-1	1	12.94	-1		0.000		1	3	-1
					0.386		0.228		0.386		
		-1.2	1.2	12.36	-1.2		0.000		-1.2	3	-1.2
					0.376		0.247		0.376		
		-1.4	1.4	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-1.6	1.6	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-1.8	1.8	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-2	2	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-2.2	2.2	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-2.4	2.4	12.29	-1.302		0.000		1.302	3	-1.302
					0.371		0.259		0.371		
		-2.6	2.6	12.29	-2.6	-1.295	0.000	1.295	2.6	5	-2.6
					0.005	0.366	0.259	0.366	0.005		
		-2.8	2.8	12.27	-2.8	-1.274	0.000	1.274	2.8	5	-2.8
					0.020	0.352	0.258	0.352	0.020		
		-3	3	12.24	-3	-1.258	0.000	1.258	3	5	-3
					0.030	0.341	0.258	0.341	0.030		
		-3.2	3.2	12.19	-3.2	-1.246	0.000	1.246	3.2	5	-3.2
					0.039	0.332	0.258	0.332	0.039		
		-3.4	3.4	12.13	-3.4	-1.236	0.000	1.236	3.4	5	-3.4
					0.046	0.326	0.258	0.326	0.046		
		-3.6	3.6	12.06	-3.6	-1.228	0.000	1.228	3.6	5	-3.6
					0.051	0.321	0.257	0.321	0.051		
		-3.8	3.8	12.00	-3.8	-1.221	0.000	1.221	3.8	5	-3.8
					0.055	0.316	0.257	0.316	0.055		
		-4	4	11.93	-4	-1.216	0.000	1.216	4	5	-4
					0.059	0.313	0.257	0.313	0.059		
		-4.2	4.2	11.86	-4.2	-1.211	0.000	1.211	4.2	5	-4.2
					0.062	0.310	0.257	0.310	0.062		
		-4.4	4.4	11.79	-4.4	-1.208	0.000	1.208	4.4	5	-4.4
					0.064	0.307	0.257	0.307	0.064		
		-4.6	4.6	11.72	-4.6	-1.204	0.000	1.204	4.6	5	-4.6
					0.066	0.305	0.257	0.305	0.066		
		-4.8	4.8	11.66	-4.8	-1.202	0.000	1.202	4.8	5	-4.8
					0.068	0.304	0.257	0.304	0.068		
		-5	5	11.59	-5	-1.199	0.000	1.199	5	5	-5
					0.069	0.302	0.257	0.302	0.069		
		-10	10	10.34	-10	-1.180	0.000	1.180	10	5	-10
					0.082	0.290	0.257	0.290	0.082		
		-12	12	9.99	-12	-1.178	0.000	1.178	12	5	-12
					0.083	0.289	0.257	0.289	0.083		

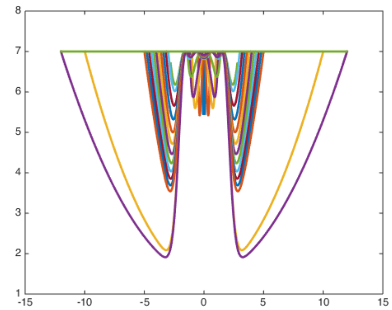
Table 6.7: D-optimal Designs for d When $\alpha_0 = 1$ and $\alpha_2 < 0$

α_2	α_0	Design space		$fval$	Design						# of pts	d_1
negative	1	-0.1	0.1	33.64	-0.1		0.000			0.1	3	-0.1
					0.400		0.200			0.400		
		-0.5	0.5	17.17	-0.5		0.000			0.5	3	-0.5
					0.402		0.197			0.402		
		-1	1	9.68	-1.0		0.000			1.0	3	-1
					0.401		0.197			0.401		
		-1.2	1.2	8.00	-1.2		0.000			1.2	3	-1.2
					0.398		0.204			0.398		
		-1.4	1.4	7.04	-1.4		0.000			1.40	3	-1.4
					0.392		0.216			0.392		
		-1.6	1.6	6.75	-1.600	-0.656			0.656	1.600	4	-1.6
					0.348	0.152			0.152	0.348		
		-1.8	1.8	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-2	2	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-2.2	2.2	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-2.4	2.4	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-2.6	2.6	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-2.8	2.8	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
		-3	3	6.74	-3	-1.618	-0.701	0.701	1.618	3	6	-3
					0.008	0.332	0.160	0.160	0.332	0.008		
		-3.2	3.2	6.72	-3.2	-1.597	-0.680	0.680	1.597	3.2	6	-3.2
					0.023	0.320	0.158	0.158	0.320	0.023		
		-3.4	3.4	6.69	-3.4	-1.581	-0.663	0.663	1.581	3.4	6	-3.4
					0.033	0.311	0.156	0.156	0.311	0.033		
		-3.6	3.6	6.64	-3.6	-1.569	-0.650	0.650	1.569	3.6	6	-3.6
					0.042	0.304	0.154	0.154	0.304	0.042		
		-3.8	3.8	6.59	-3.8	-1.559	-0.640	0.640	1.559	3.8	6	-3.8
					0.048	0.299	0.153	0.153	0.299	0.048		
		-4	4	6.53	-4	-1.550	-0.631	0.631	1.550	4	6	-4
					0.053	0.295	0.152	0.152	0.295	0.053		
		-4.2	4.2	6.47	-4.2	-1.544	-0.625	0.625	1.544	4.2	6	-4.2
					0.058	0.291	0.151	0.151	0.291	0.058		
		-4.4	4.4	6.41	-4.4	-1.538	-0.619	0.619	1.538	4.4	6	-4.4
					0.061	0.288	0.151	0.151	0.288	0.061		
		-4.6	4.6	6.35	-4.6	-1.534	-0.614	0.614	1.534	4.6	6	-4.6
					0.064	0.286	0.150	0.150	0.286	0.064		
		-4.8	4.8	6.29	-4.8	-1.530	-0.610	0.610	1.530	4.8	6	-4.8
					0.067	0.284	0.150	0.150	0.284	0.067		
		-5	5	6.22	-5	-1.526	-0.607	0.607	1.526	5	6	-5
					0.069	0.282	0.150	0.150	0.282	0.069		
		-10	10	5.01	-10	-1.499	-0.578	0.578	1.499	10	6	-10
					0.085	0.268	0.147	0.147	0.268	0.085		
		-12	12	4.66	-12	-1.497	-0.576	0.576	1.497	12	6	-12
					0.087	0.267	0.147	0.147	0.267	0.087		

Figure 6.2: D-optimality Verification in Tables 6.6 and 6.7



(a) $\alpha_2 > 0, \alpha_0 = 1$



(b) $\alpha_2 < 0, \alpha_0 = 1$

between $[-1.6, 1.6]$ and $[-2.4, 2.4]$, three points including -1.302, 0, and 1.302 form a design. When we have 5-point designs, the weights of the two outer points are small compared to the inner three points.

For the latter case, when the space is narrower than $[-1.4, 1.4]$, we have 3-point designs as $(-D, 0, D)$ in $[-D, D]$ with the weight ratio close to 2:1:2. From $[-1.8, 1.8]$ to $[-2.8, 2.8]$, we have 4-point designs with fixed points of $(-1.63, -0.71, 0.71, 1.63)$ where the weight ratio is close to 2:1:1:2. When the design space is wider than $[-3, 3]$, there are 6-point designs.

Wu and Stufken (2014) found 3-point D-optimal designs with $(-1.3089, 0, 1.3089)$ with 1/3 equal weights when $\theta_0\theta_2 > 0$ (in our case $\alpha_0\alpha_2 > 0$) in a quadratic logistic model. When $\theta_0\theta_2 < 0$, they found a subclass of designs with at most four points that are at least as good as any other designs for their model.

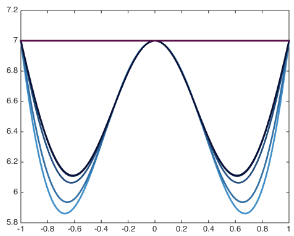
Verification is successful as shown in Figure 6.2. In summary, the sign of α_2 and the size of the design space are influential in finding optimal designs .

Next, we examine the effect of α_0 values in Table 6.8. We set three design spaces $[-1, 1]$, $[-2, 2]$, $[-5, 5]$ where we obtained different numbers of design points as shown in the previous Table 6.6 and 6.7. We set α_0 to 0.5, 1, 2.5, 5, and 10.

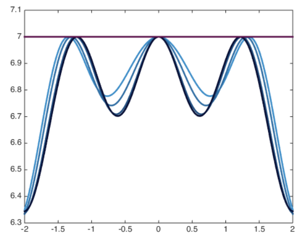
Table 6.8: D-optimal Designs for d by α_0 Values

α_2	Design space		α_0	$fval$	Design						# of pts	d_1
positive	-1	1	0.5	11.41	-1		0			1	3	-1
					0.389		0.221			0.389		
			1	12.94	-1		0			1	3	-1
					0.386		0.228			0.386		
			2.5	19.11	-1		0			1	3	-1
					0.380		0.240			0.380		
			5	31.22	-1		0			1	3	-1
					0.378		0.244			0.378		
			10	56.18	-1		0			1	3	-1
					0.378		0.245			0.378		
	-2	2	0.5	10.51	-1.346		0			1.346	3	-1.346
					0.373		0.254			0.373		
			1	12.29	-1.302		0			1.302	3	-1.302
					0.371		0.259			0.371		
			2.5	18.73	-1.238		0			1.238	3	-1.238
					0.367		0.267			0.367		
			5	30.91	-1.218		0			1.218	3	-1.218
					0.365		0.269			0.365		
			10	55.88	-1.216		0			1.216	3	-1.216
					0.365		0.270			0.365		
	-5	5	0.5	9.84	-5	-1.2447	0.000		1.2447	5	5	-5
					0.070	0.3044	0.252		0.3044	0.0695		
			1	11.59	-5	-1.1992	0.000		1.1992	5	5	-5
					0.069	0.3021	0.257		0.3021	0.0692		
			2.5	18.00	-5	-1.1321	0.000		1.1321	5	5	-5
					0.069	0.2986	0.265		0.2986	0.0688		
			5	30.16	-5	-1.1112	0.000		1.1112	5	5	-5
					0.069	0.2975	0.268		0.2975	0.0687		
			10	55.13	-5	-1.1093	0.000		1.1093	5	5	-5
					0.069	0.2974	0.268		0.2974	0.0687		
negative	-1	1	0.5	9.70	-1		0.000			1	3	-1
					0.398		0.205			0.3976		
			1	9.68	-1		0.000			1	3	-1
					0.401		0.197			0.4013		
			2.5	12.87	-1		0.000			1	3	-1
					0.408		0.184			0.4079		
			5	23.79	-1		0.000			1	3	-1
					0.410		0.179			0.4103		
			10	48.63	-1		0.000			1	3	-1
					0.411		0.179			0.4105		
	-2	2	0.5	7.76	-1.505	-0.427			0.427	1.505	4	-1.505
					0.361	0.139			0.139	0.361		
			1	6.75	-1.630	-0.712			0.712	1.630	4	-1.630
					0.338	0.162			0.162	0.338		
			2.5	4.65	-2	-1.298			1.298	1.808	4	-2
					0.298	0.202			0.202	0.2977		
			5	6.47	-2	-1.473			1.473	2	4	-2
					0.388	0.112			0.112	0.388		
			10	28.99	-2	-1.516			1.516	2	4	-2
					0.398	0.102			0.102	0.398		
	-5	5	0.5	7.17	-5	-1.391	-0.226	0.226	1.391	5	6	-5
					0.070	0.306	0.125	0.125	0.306	0.070		
			1	6.22	-5	-1.526	-0.606	0.606	1.526	5	6	-5
					0.069	0.282	0.150	0.150	0.282	0.069		
			2.5	4.34	-5	-1.926	-1.235	1.235	1.926	5	6	-5
					0.062	0.246	0.192	0.192	0.246	0.062		
			5	2.62	-5	-2.494	-1.968	1.968	2.494	5	6	-5
					0.041	0.241	0.218	0.218	0.241	0.041		
			10	0.67	-3.363	-2.979			2.979	3.363	4	-3.363
					0.264	0.237			0.237	0.264		

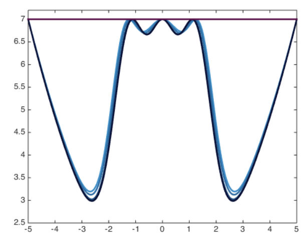
Figure 6.3: D-optimality Verification When $\alpha_0 = (0.5, 1, 2.5, 5, 10)$ in Table 6.8



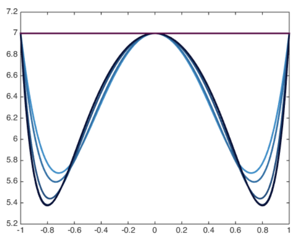
(a) $[-1, 1]$, $\alpha_2 > 0$



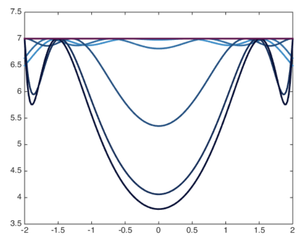
(b) $[-2, 2]$, $\alpha_2 > 0$



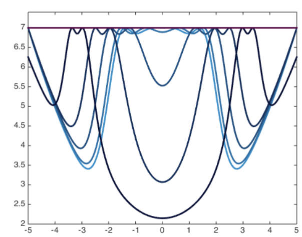
(c) $[-5, 5]$, $\alpha_2 > 0$



(d) $[-1, 1]$, $\alpha_2 < 0$



(e) $[-2, 2]$, $\alpha_2 < 0$



(f) $[-5, 5]$, $\alpha_2 < 0$

* light blue: $\alpha_0 = 0.5$, dark blue: $\alpha_0 = 10$

Table 6.9: D-optimal Designs for d When $\alpha_0 = 10$ and $\alpha_2 < 0$

α_2	α_0	Design space		$fval$	Design						# of pts	d_1
negative	10	-3	3	3.92	-3	-2.672			2.672	3	4	-3
					0.377	0.123			0.123	0.377		
		-4	4	0.67	-3.363	-2.979			2.979	3.363	4	-3.363
					0.264	0.473			0.473	0.264		
		-5	**5**	0.67	-3.363	-2.979			2.979	3.363	4	-3.363
					0.264	0.473			0.473	0.264		
		-6	6	0.64	-6	-3.350	-2.968	2.968	3.350	6	6	-6
					0.026	0.241	0.233	0.233	0.241	0.026		
		-7	7	0.51	-7	-3.337	-2.956	2.956	3.337	7	6	-7
					0.050	0.221	0.229	0.229	0.221	0.0504		
		-8	8	0.35	-8	-3.330	-2.949	2.949	3.330	8	6	-8
					0.064	0.210	0.227	0.227	0.210	0.0635		
		-9	9	0.19	-9	-3.325	-2.944	2.944	3.325	9	6	-9
					0.072	0.203	0.226	0.226	0.203	0.0716		
		-10	10	0.03	-10	-3.322	-2.941	2.941	3.322	10	6	-10
					0.077	0.198	0.225	0.225	0.198	0.077		

In general, the change of α_0 does not affect the number of support points in the same design space except for $[-5, 5]$ when $\alpha_2 < 0$. When $\alpha_2 > 0$, regardless of α_0 values, we have 3-, 3-, and 5-point designs respectively in three design spaces as can be seen in Table 6.8. On the other hand, when $\alpha_2 < 0$, the number of design points are 3, 4, and 6 in $[-1, 1], [-2, 2], [-5, 5]$ respectively except for one case in $[-5, 5]$.

To see if the case of $[-5, 5]$ when $\alpha_0 = 10$ is exceptional since we obtain a 4-point design, we create Table 6.9. As shown in the table, the change of the number of support points from four to six is gradual and the pattern is similar to the case of Table 6.7. The case is not exceptional.

The graphs in Figure 6.3 summarize the results in Table 6.8. We draw six plots when $\alpha_2 > 0$ and < 0 with three different design spaces. In each plot, five curves are created depending on the values of α_0 including 0.1, 0.5, 1, 1.5, and 2. The straight line indicates a constant function of seven, the number of parameters. The light color indicates a small value of α_0. In general, we see that the change of α_0 does not significantly break the general pattern in terms of the number of design points and the location of optimal points. When $\alpha_2 < 0$, the patterns are more deviated especially in $[-2, 2]$ and $[-5, 5]$. When $\alpha_2 < 0$, a predictor is $\eta = |\alpha_0| - d^2$ while, when $\alpha_2 > 0$, a predictor is $\eta = |\alpha_0| + d^2$ as the sum of two positive numbers. In the objective function, we have many $\exp(\eta)$ with guessed values. We conjecture that $\eta = |\alpha_0| - d^2$ causes a more complex optimization procedure than that of $\eta = |\alpha_0| + d^2$.

6.6 Discussion

In this study, we investigated D-optimal designs for a quadratic mixed response experiment. We introduced a quadratic term in a logistic regression model for mixed response experiments. Since the information matrix was more complex than in the previous study, we used a different method to apply the complete class approach and a nonlinear optimization algorithm.

We first proved that a symmetric design was at least as good as that of any other design under the D-optimality criterion when the design region was symmetric about zero. Considering these designs, we effectively moved to the complete class approach. We identified a complete class with at most 6-point designs including the two boundary points. To identify a complete class, we found Chebyshev systems for a nonnegative subdesign and a nonpositive subdesigns, respectively. We proved that two results of the search for Chebyshev systems guaranteed the Loewner ordering of the information matrices between a 'good' design and any other design. Then, we formed a complete class.

Within a complete class, we searched D-optimal designs using an algorithm we previously used. The mathematical standard form for optimization was rewritten. Obtained designs were verified as D-optimal by the general equivalence theorem. Since there existed an unknown parameter problem, we used guessed values of α_0 and α_2. When $\alpha_2 > 0$, D-optimal designs were 3- or 5-point designs with zero points. When $\alpha_2 < 0$, there were 3-, 4-, or 6-point designs depending on design spaces. The sign of α_2 mattered since the number of the obtained optimal designs was affected by the sign of α_2. The value of α_0 did not severely affect the results of optimization.

We may consider the case where the design space is not symmetric. An A-optimal study with this model can be an of interest to further studies. Also, another extended model such as the model where a linear model also has a quadratic term can be work for the future.

Chapter 7

CONCLUDING REMARKS

So far we have studied optimal designs for experiments where responses are mixed categorical and continuous responses. Specifically the study involved responses which consist of both binary and continuous variables. Although several statistical methods have been proposed for modeling these types of responses, a little attention has been paid to the design of such experiments. To fill the gap, we studied optimal designs for such experiments. Above all, we have chosen to specifically study optimal design of experiments since it connotes a clear objective of achieving the most reliable experimentation. We aimed to find the best design that attains minimization of the variance-covariance of parameter estimates.

We established a mixed response model using one simple logistic regression model and two conditional simple linear models. We derived the information matrix of parameters and then employed it to obtain optimal designs. Especially, we used the complete class approach to reduce the number of candidate designs and then implement a constrained nonlinear algorithm. The optimality of the obtained design was verified by the general equivalence theorem.

In Chapter 3, we found that at most four points were enough to construct 'good' designs that were no worse than any other designs by using the complete class approach. The four points include two endpoints of a design space. To apply the complete class approach to a mixed response model, we suggested Lemma 2.2.1 in Chapter 2 and were then able to identify the complete class. Within the class, we efficiently searched D- and A-optimal designs in Chapters 4 and 5, respectively. In the case of the D-optimal designs, the optimal designs were symmetric and the corresponding weights were the same for symmetric points in symmetric design spaces. In the case of the A-optimal designs, almost all of the two inner points of 4-point designs were not symmetric and the corresponding weights were not the same even in symmetric spaces.

The focus of the search for A-optimal designs was on the effect of σ^2 on the obtained designs. When we searched, we put guessed values of α_0, α_1 and σ^2 into the A-optimality criterion. As σ^2 increased, the weights of the two inner points moved to the two outer points. The two asymmetric inner points changed to being symmetric. The results suggested that we could manage the effect of variance in continuous responses for experiments by using A-optimal designs. This control was not available under the D-optimality criterion.

The numerical results in two chapters showed that D- or A-optimal designs had 2, 3, or 4 points. The result of the number of support points was validated by Theorem 3.3.1 in Chapter 3 regarding the positive definiteness of the information matrix. While we had six parameters of interest, two points were enough to make all parameters estimable. By combining it with the complete class result, we knew that the range of the number of support points was between two and four.

After finishing the first study with a first-order model, we considered a model with a quadratic polynomial predictor. When a curvature trend was detected, the second-order polynomial model was recommended by many authors. We added a quadratic term to a logistic regression model. We first found Chebyshev systems for two subdesigns and then formed a complete class by combining them. We also used the property of D-invariance with a symmetric design. Obtained were 3-, 4-, 5-, or 6-point designs by the condition of guessed values of unknown parameters.

The complete class approach gave a huge reduction of candidate designs. We extended an original complete class approach in Yang and Stufken (2012) to the bivariate response model. We used mathematical programming after deciding the number of support points. Computing time was trivial and the obtained designs were optimal. It was helpful to use the graphical interpretation based on the general equivalence theorem.

It has been about fifty years since prevailing models for non-normal data were unified with the name generalized linear model (GLM). For a short time, GLMs became popular and were used in many fields of natural science and social science. Khuri, Mukherjee, Sinha, and Ghosh (2006) mentioned the lack of studies of optimal designs for GLMs. Since then, a few suggestions for those types of models have emerged based on the complete class approach.

Yang and Stufken (2009) found an optimal design for a GLM with two parameters. Yang and Stufken (2012) generalized the complete class approach that enabled us to consider using a complete class approach for other nonlinear models including a mixed response model. Wu and Stufken (2014) searched for optimal designs for the GLM with a quadratic polynomial predictor. We found optimal designs for a mixed response model combining a logistic regression model and linear models.

The results of this study can serve at least as a benchmark since there are few alternatives. In conclusion, we expect that, by using optimal designs that we found, an experimenter can collect the most informative data from experiments where responses are bivariate variables that contain a binary variable and a continuous variable. Several studies are expected to be future work as discussed in each chapter.

REFERENCES

Agresti, A. (2007). *An introduction to categorical data analysis.* Hoboken.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 59-76.

Atkinson, A. C., & Bailey, R. A. (2001). One hundred years of the design of experiments on and off the pages of Biometrika. *Biometrika, 88(1)*, 53-97.

Atkinson, A., Donev, A., & Tobias, R. (2007). *Optimum experimental designs, with SAS (Vol. 34).* Oxford University Press.

Berger, M. P., & Wong, W. K. (2009). *An introduction to optimal designs for social and biomedical research (Vol. 83).* John Wiley & Sons.

Biswas, A., & López-Fidalgo, J. (2013). Compound designs for dosefinding in the presence of nondesignable covariates. *Pharmaceutical statistics, 12(2)*, 92-101.

Catalano, P. J., & Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association, 87(419)*, 651-658.

Charnes, A., Frome, E. L., & Yu, P. L. (1976). The equivalence of generalized least squares and maximum likelihood estimates in the exponential family. *Journal of the American Statistical Association, 71(353)*, 169-171.

Cheng, C. S. (1995). Complete class results for the moment matrices of designs over permutation-invariant sets. *The Annals of Statistics*, 41-54.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, 586-602.

Coffey, T., & Gennings, C. (2007). D-optimal designs for mixed discrete and continuous outcomes analyzed using nonlinear models. *Journal of agricultural, biological, and environmental statistics, 12(1)*, 78-95.

Collett, D. (2002). *Modeling Binary Data*, 2nd ed. London: Chapman & Hall.

De la Garza, A. (1954). Spacing of information in polynomial regression. *The Annals of Mathematical Statistics, 25(1)*, 123-130.

De Leon, A. R., & Chough, K. C. (Eds.). (2013). Analysis of mixed data: methods & applications. CRC Press.

Deng, X., & Jin, R. (2015). QQ Models: Joint Modeling for Quantitative and Qualitative Quality Responses in Manufacturing Systems. *Technometrics, 57(3)*, 320-331.

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

Dette, H., & Schorning, K. (2013). Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *The Annals of Statistics, 41(3)*, 1260-1267.

Dobson, A. J., & Barnett, A. (2008). *An introduction to generalized linear models.* CRC press.

Duarte, B. P., & Wong, W. K. (2014). A semi-infinite programming based algorithm for finding minimax optimal designs for nonlinear models. *Statistics and Computing, 24(6)*, 1063-1080.

Ehrenfeld, S. (1956). Complete class theorems in experimental design. *In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics.* The Regents of the University of California.

Elfving, G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics, 23(2)*, 255-262.

Fedorov, V. V. (1972). *Theory of optimal experiments* (trans. and ed. W. J. Studden and E. M. Klimko), New York: Academic Press.

Fedorov, V. V., & Leonov, S. L. (2013). *Optimal design for nonlinear response models.* CRC Press.

Fedorov, V., Wu, Y., & Zhang, R. (2012). Optimal dosefinding designs with correlated continuous and discrete responses. *Statistics in medicine, 31(3)*, 217-234.

Fitzmaurice, G. M., & Laird, N. M. (1995). Regression models for a bivariate discrete and continuous outcome with clustering. *Journal of the American statistical Association, 90(431)*, 845-852.

Ford, I., Torsney, B., & Wu, C. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 569-583.

Hedayat, A. S., Jacroux, M., & Majumdar, D. (1988). [Optimal Designs for Comparing Test Treatments with Controls]: Rejoinder. *Statistical Science, 3(4)*, 487-491.

Hedayat, A. S., Jacroux, M., & Majumdar, D. (1988). Optimal designs for comparing test treatments with controls. *Statistical Science*, 462-476.

Hinkelmann, K., & Kempthorne, O. (2005). *Design and analysis of experiments. Vol. 2, Advanced experimental design.* Wiley.

Hirakawa, A. (2012). An adaptive dosefinding approach for correlated bivariate binary and continuous outcomes in phase I oncology trials. *Statistics in medicine, 31(6)*, 516-532.

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression (Vol. 398).* John Wiley & Sons.

Hu, L., Yang, M., & Stufken, J. (2015). Saturated locally optimal designs under differentiable optimality criteria. *The Annals of Statistics, 43(1)*, 30-56.

Hyun, S. W. (2013). Optimal designs for a probit model with a quadratic term. *Statistics in Biopharmaceutical Research, 5(1)*, 18-26.

Jewell, N. P. (2003). *Statistics for epidemiology.* CRC Press.

Karlin, S., & Studden, W. J. (1966). *Tchebycheff systems: With applications in analysis and statistics.* New York: Interscience.

Khuri, A. I., Mukherjee, B., Sinha, B. K., & Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statistical Science*, 376-399.

Kiefer, J. (1959). Optimum experimental designs. *Journal of the Royal Statistical Society. Series B (Methodological)*, 272-319.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 849-879.

Kiefer, J., & Wolfowitz, J. (1960). The equivalence of two extreme problems. *Canadian Journal of Mathematics, 12(5)*, 363-365.

Kunert, J., & Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *Journal of the American Statistical Association, 97(459)*, 898-906.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied linear statistical models.* McGraw Hill.

Liski, E. P., Mandal, N. K., Shah, K. R., & Sinha, B. K. (2002). *Topics in Optimal Design, ser. Lecture Notes in Statistics.* New York: Springer, 163.

Mathew, T., & Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *Journal of Statistical Planning and Inference, 93(1)*, 295-307.

Nash, J. C. (2014). *Nonlinear parameter optimization using R tools.* John Wiley & Sons.

Notz, W. I. (1988). Optimal Designs for Comparing Test Treatments with Controls: Comment. *Statistical Science*, 3(4), 480-482.

Olkin, I., & Tate, R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables. *The Annals of Mathematical Statistics*, 448-465.

Papp, D. (2012). Optimal designs for rational function regression. *Journal of the American Statistical Association, 107(497)*, 400-411.

Pukelsheim, F. (1989). Complete class results for linear regression designs over the multi-dimensional cube. *In Contributions to Probability and Statistics* (pp. 349-356). Springer New York.

Pukelsheim, F. (2006). *Optimal design of experiments.* Society for Industrial and Applied Mathematics.

Ryan, L. M. (2000). Statistical issues in toxicology. *Journal of the American Statistical Association, 95(449)*, 304-308.

Silvey, S. D. (1980). *Optimal designs: an introduction to the theory for parameter estimation.*

Silvey, S. D., Titterington, D. M., & Torsney, B. (1978). An algorithm for optimal designs on a finite design space, *Communications in StatisticsTheory and Methods, 14*, 13791389.

Steinberg, D. M., & Hunter, W. G. (1984). Experimental design: review and comment. *Technometrics, 26(2)*, 71-97.

Stufken, J., & Yang, M. (2012). Optimal designs for generalized linear models. *Design and Analysis of Experiments, Special Designs and Applications, 3*, 137.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika, 24(3-4)*, 471-494.

Wu, H. P. (2013). Locally optimal designs for generalized linear models with a single-variable quadratic polynomial predictor (Doctoral dissertation, uga)

Wu, H. P., & Stufken, J. (2014). Locally Φ_p-optimal designs for generalized linear models with a single-variable quadratic polynomial predictor. *Biometrika, 101(2)*, 365-375.

Wynn, H. P. (1970). The sequential generation of D-optimal experimental designs. *The Annals of Mathematical Statistics, 41*, 16551664.

Wynn, H. P. (1984). Jack Kiefer's contributions to experimental design. *The Annals of Statistics, 12(2)*, 416-423.

Yang, M. (2008). A-optimal designs for generalized linear models with two parameters. *Journal of Statistical Planning and Inference, 138(3)*, 624-641.

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

Yang, M., & Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *The Annals of Statistics*, 518-541.

Yang, M., & Stufken, J. (2012). Identifying locally optimal designs for nonlinear models: A simple extension with profound consequences. *The Annals of Statistics*, 1665-1681.

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