A Comparison of Two Approaches for Constructing Exact Designs for Mixed Binary and Continuous Responses

by

Bayan Abdullah

A Thesis Presented in Partial Fulfillment
of the Requirement for the Degree
Master of Science

Approved April 2021 by the
Graduate Supervisory Committee:

Ming-Hung Kao, Chair
Dan Cheng
Yi Zheng

ARIZONA STATE UNIVERSITY

May 2021

ABSTRACT

This thesis is concerned with experimental designs for studies a controllable independent variable X, a continuous response variable Y and a binary response variable Z. It is known that judiciously selected design allows experimenters to collect informative data for making precise and valid statistical inferences with minimum cost. However, for the complex setting that this thesis consider, designs that yield a high expected estimation precision may still possess a high probability of having non-estimable parameters, especially when the sample size is small. Such an observation has been reported in some previous works on the separation issue for, e.g., the logistic regression. Therefore, when selecting a study design, it is important to consider both the expected variances of the parameter estimates, and the probability for having non-estimable parameters.

A comparison of two approaches for constructing designs for the previously mentioned setting with a mixed responses model is presented in this work. The two design approaches are the locally *A*-optimal design approach, and a penalized *A*-optimal design approach that involves the optimization of *A*-optimality criterion plus the penalty term to reduce the chance of including designs points that have a high probability to make some parameters non-estimable.

ACKNOWLEDGMENTS

I would first like to thank The Kingdom of Saudi Arabia and The Saudi Arabian Cultural Mission to the U.S. for making it possible for me to study at Arizona State University. I am so grateful to my supervisor, Professor Ming-Houg Kao, whose expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. My special and heartily thanks to my spouse, Omar Noorwali, who encouraged me throughout the time of my research.

May the Almighty God richly bless all of you.

TABLE OF CONTENTS

LIST OF TABLES

Table | Page

LIST OF FIGURES

Chapter 1

INTRODUCTION

Over the previous decade there has been an enormous progress in the development of mixed responses models for applications having multiple response of mixed variable types. This paper concentrates on generalized linear mixed responses model that allows to jointly modeling the continuous and binary responses, with a focus on studying optimal designs for it.

In studying optimal designs for mixed binary and continuous responses model, our main focus is on the model proposed by Deng and Jin (2015), but for clarity, we consider cases with one controllable continuous independent variable X. As described by Deng and Jin (2015), their model is satisfactorily devised for the prediction of the response and informative in many cases. Our goal is to provide insightful information on designs that give the 'best' set of X.

For the complex setting that we consider, designs that yield a high expected estimation precision may still possess a high probability of having non-estimable parameters, especially when the sample size is small. Such an observation has been reported in some previous works on the separation issue for, e.g., the logistic regression. Therefore, when selecting a study design, we need to consider both the expected variances of the parameter estimates, and the probability for having non-estimable parameters.

In this work, we compare two approaches for constructing designs for the previously mentioned setting with a mixed responses model. The two design approaches are the locally A-optimal design approach, and a penalized A-optimal design approach that involves the optimization of A-optimality criterion plus the penalty term to reduce the chance of including designs points that have a high probability to make some parameters non-estimable.

In the next chapter, we describe the model that we consider and some relevant model estimation. The third chapter provides a description of a design issue for mixed responses models and the optimality criteria for evaluating the quality of designs under the mixed responses model. In chapter four, a simulation study for comparing the performance of different designs is conducted. In chapter five, we adapt a design method that has recently been proposed for finding designs for generalized linear models and study its performance on finding high-quality designs for our mixed responses model.

Chapter 2

MODEL DESCRIPTION

2.1 Model

We consider a scientific procedure that involves a controllable independent variable X, a continuous response variable Y, and a binary response variable Z. For the i^{th} experimental unit with $(X,Y,Z) = (x_i, y_i, z_i)$; $i = 1, ..., N$, the joint probability function for $f(y_i, z_i)$ follows

$$f(y_i, z_i; \theta) = [\pi_i]^{z_i}[1-\pi_i]^{1-z_i}[(\frac{1}{\sigma\sqrt{2\pi}})exp(-\frac{1}{2\sigma^2}(y_i-\beta_{01}-\beta_{11}x_i)^2)]^{z_i}$$
$$[(\frac{1}{\sigma\sqrt{2\pi}})exp(-\frac{1}{2\sigma^2}(y_i-\beta_{02}-\beta_{12}x_i)^2)]^{1-z_i} \quad (2.1)$$

where $\theta = (\alpha_0, \alpha_1, \beta_{01}, \beta_{11}, \beta_{02}, \beta_{12}, \sigma^2)^T$ is a parameter vector, and $\pi_i = Pr(z_i = 1)$ is defined as below.

$$Pr(z_i = 1) = \pi_i = \frac{exp(\alpha_0 + \alpha_1 x_i)}{1 + exp(\alpha_0 + \alpha_1 x_i)}$$

with mean $E(z_i) = \pi_i$ and variance $Var(z_i) = \pi_i(1-\pi_i)$. It can also be seen that, given the value of z_i, the continuous response y_i has the following conditional distribution.

$$(y_i; z_i) \sim N([\beta_{01} + \beta_{11}x_i]z_i + [\beta_{02} + \beta_{12}x_i][1-z_i], \sigma)$$

The corresponding log-likelihood function $l(\theta)$ of θ is

$$l(\theta; y_i, z_i) = \log L(\theta; y_i, z_i) = \sum_{i=1}^{N} z_i \log(\pi_i) + (1-z_i)\log(1-\pi_i) - \log(\sigma\sqrt{2\pi})$$
$$+ z_i[-\frac{1}{2\sigma^2}(y_i - \beta_{01} - \beta_{11}x_i)^2] + (1-z_i)[-\frac{1}{2\sigma^2}(y_i - \beta_{02} - \beta_{12}x_i)^2] \quad (2.2)$$

For computing the maximum likelihood estimates (ML estimates) of θ, the following results are needed for α_0 and α_1.

- $\frac{\partial l}{\partial \alpha_0} = \frac{\partial l}{\partial \pi_i}\frac{\partial \pi_i}{\partial \alpha_0}$ where

$$\frac{\partial l}{\partial \pi_i} = \sum_{i=1}^{N}[\frac{z_i}{\pi_i} + \frac{z_i - 1}{1-\pi_i}] = \sum_{i=1}^{N}[\frac{z_i - \pi_i}{\pi_i(1-\pi_i)}];$$

$$\frac{\partial \pi_i}{\partial \alpha_0} = \frac{exp(\alpha_0 + \alpha_1 x_i)}{[1+exp(\alpha_0+\alpha_1 x_i)]^2} = \pi_i(1-\pi_i).$$

- $\frac{\partial l}{\partial \alpha_1} = \frac{\partial l}{\partial \pi_i}\frac{\partial \pi_i}{\partial \alpha_1}$ where

$$\frac{\partial \pi_i}{\partial \alpha_1} = \frac{x_i exp(\alpha_0 + \alpha_1 x_i)}{[1+exp(\alpha_0+\alpha_1 x_i)]^2} = x_i\pi_i(1-\pi_i).$$

The Gradient of $l(\theta)$, which is the vector of the first derivatives of $l(\theta)$ with respect to θ, is

$$\sum_{i=1}^{N}\begin{pmatrix} \frac{\partial l}{\partial \alpha_0} \\ \frac{\partial l}{\partial \alpha_1} \\ \frac{\partial l}{\partial \beta_{01}} \\ \frac{\partial l}{\partial \beta_{11}} \\ \frac{\partial l}{\partial \beta_{02}} \\ \frac{\partial l}{\partial \beta_{12}} \\ \frac{\partial l}{\partial \sigma^2} \end{pmatrix} = \sum_{i=1}^{N}\begin{pmatrix} z_i - \pi_i \\ x_i(z_i - \pi_i) \\ \frac{z_i}{\sigma^2}(y_i - \beta_{01} - \beta_{11}x_i) \\ \frac{z_i x_i}{\sigma^2}(y_i - \beta_{01} - \beta_{11}x_i) \\ \frac{(1-z_i)}{\sigma^2}(y_i - \beta_{02} - \beta_{12}x_i) \\ \frac{(1-z_i)x_i}{\sigma^2}(y_i - \beta_{02} - \beta_{12}x_i) \\ \frac{1}{\sigma^2} + \frac{z_i}{2\sigma^4}(y_i - \beta_{01} - \beta_{11}x_i)^2 + \frac{(1-z_i)}{2\sigma^4}(y_i - \beta_{02} - \beta_{12}x_i)^2 \end{pmatrix} \quad (2.3)$$

2.2 Information Matrix

The Fisher Information matrix of parameter vector θ is the symmetrical 7-by-7 matrix presented below.

$$I(\theta) = -E[H(\theta)] = -E[\frac{\partial^2 l}{\partial \theta^2}] = \begin{bmatrix} A & B & 0 & 0 & 0 & 0 & 0 \\ C & D & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & E & F & 0 & 0 & G \\ 0 & 0 & H & I & 0 & 0 & J \\ 0 & 0 & 0 & 0 & K & L & M \\ 0 & 0 & 0 & 0 & N & O & P \\ 0 & 0 & Q & R & S & T & U \end{bmatrix} \qquad (2.4)$$

where the Hessian matrix $H(\theta)$ is a matrix of second derivatives of $l(\theta)$ with respect to θ, and since $E(y_i) = [\beta_{01} + \beta_{11}x_i]z_i + [\beta_{02} + \beta_{12}x_i][1 - z_i]$, then the elements of the information matrix $I(\theta)$ is in Table 2.1.

As to be discussed later, the Fisher information matrix in (2.4) is also used to formulate optimality criteria for comparing the statistical efficiencies of competing designs, and to derive optimal designs for the model.

Table 2.1: Elements of the Information Matrix

Letter	Value	Letter	Value
A	$\sum_{i=1}^{N}[\pi_i(1-\pi_i)]$	B=C	$\sum_{i=1}^{N}[x_i\pi_i(1-\pi_i)]$
D	$\sum_{i=1}^{N}[x_i^2\pi_i(1-\pi_i)]$	E	$\sum_{i=1}^{N}\frac{\pi_i}{\sigma^2}$
F=H	$\sum_{i=1}^{N}\frac{x_i\pi_i}{\sigma^2}$	G	$\sum_{i=1}^{N}\frac{\pi_i(E(y_i)-\beta_{01}-\beta_{11}x_i)}{2\sigma^4}$
I	$\sum_{i=1}^{N}\frac{x_i^2\pi_i}{\sigma^2}$	J	$\sum_{i=1}^{N}\frac{x_i\pi_i(E(y_i)-\beta_{01}-\beta_{11}x_i)}{2\sigma^4}$
K	$\sum_{i=1}^{N}\frac{1-\pi_i}{\sigma^2}$	L	$\sum_{i=1}^{N}\frac{x_i(1-\pi_i)}{\sigma^2}$
M=S	$\sum_{i=1}^{N}\frac{(1-\pi_i)(E(y_i)-\beta_{02}-\beta_{12}x_i)}{\sigma^4}$	N	$\sum_{i=1}^{N}\frac{x_i(1-\pi_i)}{\sigma^2}$
O	$\sum_{i=1}^{N}\frac{x_i^2(1-\pi_i)}{\sigma^2}$	P=T	$\sum_{i=1}^{N}\frac{x_i(1-\pi_i)(E(y_i)-\beta_{02}-\beta_{12}x_i)}{\sigma^4}$
Q	$\sum_{i=1}^{N}\frac{\pi_i(E(y_i)-\beta_{01}-\beta_{11}x_i)}{\sigma^4}$	R	$\sum_{i=1}^{N}\frac{x_i\pi_i(E(y_i)-\beta_{01}-\beta_{11}x_i)}{\sigma^4}$
U	$\sum_{i=1}^{N}\frac{1}{\sigma^4}+\frac{\pi_i(E(y_i)-\beta_{01}-\beta_{11}x_i)^2}{\sigma^5}+\frac{(1-\pi_i)(E(y_i)-\beta_{02}-\beta_{12}x_i)^2}{\sigma^5}$		

2.3 Standard Errors

The asymptotic variance-covariance matrix of $\widehat{\theta}$ can be estimated by replacing θ with $\widehat{\theta}$ in the inverse of the expected information matrix:

$$cov(\widehat{\theta}) = I(\widehat{\theta})^{-1} = (-E[H(\widehat{\theta})])^{-1}$$

The above equation also gives the standard errors, since they are the square roots of the diagonal terms in the variance-covariance matrix. However, with an imprudently chosen design, some model parameters can become non-estimable or (near) non-estimable. This can result in very unstable ML estimates that have very large standard errors when a computer software package is used to find the ML estimates.

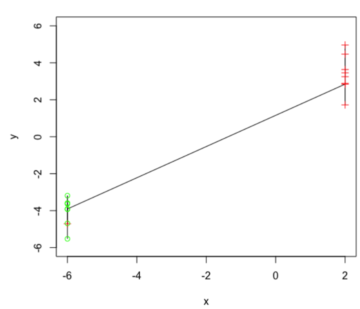
The previously described issue is linked to the so-called separation or near-separation issue in logistic regression. When this happens the estimation of one or more model parameters will fail to converge to unique parameter estimates. It occurs frequently with small sample sizes ($N \leqslant 32$), or when there is a large number of factors. Such a phenomenon appears when a hyperplane passing through the design space can completely or quasi-completely separate the design points having a response value of $Z = 0$ from the design points with a response value of $Z = 1$ (Albert and Anderson, 1984). Quasi-complete separation occurs when there are both $Z = 0$ and $Z = 1$ response on the separating hyperplane.

For instance, the data in Table 2.2 give an example of the (near-separation) issue over a given design space $\mathscr{X} = [-6, 2]$ for single independent variable X. In Figure 2.1, the y-axis represents the values of the response Y, the x-axis represents the values of the variable X, the red-point $'+'$ denotes (x_i, y_i) with $z_i = 1$, and the green-point $'+'$ denotes (x_i, y_i) with $z_i = 0$.

Table 2.2: Response Data with Quasi Separation

x	-6	-6	-6	-6	-6	-6	-6	-6	2	2	2	2	2	2	2	2
z	0	0	0	0	1	0	0	0	1	1	1	1	1	1	1	1

Figure 2.1: A Plot in the Separated Response Data Presented in Table2.2



As previously mentioned, separation forces parameters of the logistic function π_i to be (near) non-estimable and leads to huge standard errors due to an ill-conditioned Hessian matrix. Such a (nearly) non-invertible Hessian might be caused by the selection of a bad experimental design, multicollinearity or by including more explanatory variables than observations. In this work we focus on studying the selection of a good design to help avoiding this issue.

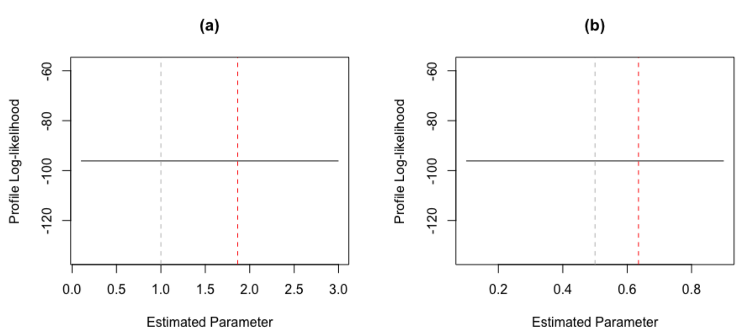
To further illustrate the effect of the separation issue on our model (2.1), we present part of the profile likelihood functions of α_0 and α_1 in Figure 2.2 given the data provided in Table 2.2. The profile likelihood functions are rather flat, making the corresponding ML estimates unstable, and they have very large (∞) standard errors.

The profile log-likelihood functions in Figure 2.2 indicate that a precise estimate of intercept and slope of model (2.1) might not be available, especially with this small sample size of $N = 16$. Clearly, this data set does not provide strong statistical information for α_0 and α_1. Also, the separation issue makes it difficult, if not impossible, to estimate the model for making valid inference. Here we would like to seek an experimental design that helps to reduce the chance for such phenomenon to occur.

Specifically, in Figure 2.2, the vertical gray dash-line corresponds to the true value of the intended parameter $\alpha_0 = 1$ and $\alpha_1 = 0.5$, while the vertical red dash-line gives the ML estimate $\widehat{\alpha_0} = 1.8653041$ and $\widehat{\alpha_1} = 0.6350151$. The maximum log-likelihood function for these estimated parameters is (-96.13022).

In such a case, unfortunately, there is no computational trick can give a stable estimation. Most textbooks advice to re-specify the model, to collect additional data, or perhaps to consider a sophisticated Bayesian approach when a prior distribution can be assumed; see also Kang et al. (2018).

Figure 2.2: A Profile Log-likelihood of (α_0, α_1),Respectively



2.4 An R Package for Model Estimation

Across many methods for estimating unknown parameters, denoted as θ, of a statistical model from data, Maximum Likelihood Estimation is both easy to compute and agrees with the initial intuition in simple examples. That is, MLE is the value of unknown parameter for which the data has the highest probability (Casella and Berger, 2002). This is accomplished by first specifying the joint density function of the observations as in Equation (2.1).

The MLE of the parameter vector θ, denoted as $\widehat{\theta_{MLE}}$, can be obtained by numerically solving some nonlinear equations involving the Gradient vector. A numerical method is needed because there is no closed form solution for $\widehat{\theta_{MLE}}$ for the model we consider. Hence statistical software package **maxLik** in R employed for solving the log-likelihood equations. The default maximisation method of **maxLik** is "NR," which is Newton-Raphson method. It includes two arguments; the first one is *logLik* which is a function that calculates the log-likelihood values as a function of the parameters; the other one is *start* which is the initial value of each parameter. As a matter of fact, Newton-type methods would be numerically more efficient in terms of number of iterations, because under appropriate conditions these approaches have a quadratic rate of convergence. However, in some applications they may require more time due to the evaluation of the Hessian and for complex models, numerical derivatives might be unreliable. In this way numerical derivatives might either slow down the estimation or even impede the convergence. Therefore, providing analytical derivatives for both the Gradient vector of $l(\theta; y_i, z_i)$ and Hessian matrix are useful otherwise, one may want to switch to a more robust estimation method that is not based on Gradients, such as the Nelder-Mead algorithm. When implementing **maxLik**, we provide analytical derivatives, as those provided in Sec. 2.1., to facilitate the search of ML estimates.

Chapter 3

EXPERIMENTAL DESIGN

Optimal designs are desirable because the use of any other design typically will require a larger N to attain the same estimation precision as the optimal design. This results in an increase of the cost and perhaps the chance of error. In many cases, an efficient way to conduct the experiment for obtaining a precise statistical inference remains rather ambiguous, and the experimenters may need to settle for inefficient designs that cause pointless expenses.

Optimal design theory spread out in its advanced structure by Kiefer (1959). Following Kiefer's work, researchers often utilize a specific statistical criterion for constructing experimental designs to fulfill the experimenter's needs. Optimal experimental design for our mixed responses model is a critical component of the research presented in this work.

The purpose of this chapter is to present an overview of optimal design theory, and to provide relevant background knowledge for our study (in the next chapter) of the performance of some experimental designs for mixed responses model in (2.1). The essential ideas of optimal design theory for standard linear models are covered in Section 3.1, with commonly used criterion functions for experimental design summarized in Section 3.2.

3.1 Optimal Design Theory

An experimental design with $m \geq 1$ support points and the weight w_i, that specifies the proportion of experimental runs of the support point x_i, is often written as

$$\xi = \begin{Bmatrix} x_1 & x_2 & \dots & x_m \\ w_1 & w_2 & \dots & w_m \end{Bmatrix}$$

There are two approaches towards experimental design optimization: exact designs and approximate designs. An exact design of experiment has a fixed sample size N and looks for minimizing some statistically meaningful function of the information matrix M_N, such as $tr(M_N^{-1})$ for the A-optimal design. In this case, Nw_i is required to be an integer, and the obtained optimal design can directly be used in experiments of sample size N. Approximate designs move away from the constraint on having a fixed N with an integer Nw_i. They allow any real value for $w_i \geq 0$ with $\sum_{i=1}^{m} w_i = 1$.

In this work, exact designs are mainly considered to avoid the need to transfer approximate designs to exact designs by, e.g., the rounding method. Another advantage of optimal exact designs is the possibility to impose additional restrictions on the design region $\{X \in \mathscr{X} : \mathscr{X} \in \mathbb{R}^{N_p}\}$. With a given model and the parameter vector of interest, a function ψ of the information matrix $M(\xi)$ is selected for evaluating competing designs to obtain an optimal one. The function ψ is called the optimality criterion. Some commonly used criteria are introduced in the next section.

3.2 Optimality Criterion

This section will cover some commonly used optimality criteria that are functional of the information matrix $M(\xi)$. One common aim in optimizing such optimality criteria is to minimize the variance of the model parameter estimates.

Table 3.1: Parameter-based Optimality Criteria

Criterion	ψ,Functional
A: $min \sum_{i=1}^{p} \lambda_i^{-1}$	$tr(M^{-1}(\xi))$
D: $min \prod_{i=1}^{p} \lambda_i^{-1}$	$\lvert M^{-1}(\xi) \rvert$

Let $\lambda_1, ..., \lambda_p$ is the eigenvalues of the information matrix. The A-optimality criterion seeks to minimize the trace of the inverse of the Fisher information matrix, which result in minimizing the average variance of the estimates of the model parameters. It also is very common to obtain the D-optimal design which maximizes the determinant of the information matrix. Recall that the determinant of a matrix $M(\xi)$ is equal to the product of all its eigenvalues, and the trace of a matrix is the sum of its eigenvalues. The two optimality criteria, presented as the smaller-the-better criteria, are summarized in Table 3.1. We note that minimizing $|M^{-1}(\xi)|$ also minimizes the generalized variance of the parameter estimates and thus gives more precise parameter estimates.

The A-optimality criterion was used by Kim (2019) to find Locally A-optimal approximate designs of model (2.1) with three different design spaces. Since the information matrices and optimal designs depend on the unknown model parameters with generalized linear mixed responses model , and one way to deal with this is to identify locally optimal designs based on the best guess of the parameters. Locally optimal designs, in fact, are important if good initial parameters are available from previous experiments, but can also function as a benchmark for designs chosen to satisfy experimental constraints. The performance of the three Locally A-optimal Designs is discussed in **Simulation Study 1** in Chapter 4.

For mixed model (2.1), it ought to be noticed that no experimental design is optimal in all aspects and minimizing (or maximizing) one particular criterion often negatively affects the experimental design of other criteria. That being said, a compound optimality criterion has discovered effective use in practice as it results in experimental designs that have many of the properties that one usually searches for. A compound design optimizes a weighted sum of two or more design criteria , and might have the form of:

$$\psi(\xi) = \kappa \quad \psi_1(\xi) \quad - \quad (1-\kappa) \quad \psi_2(\xi)$$

where ψ_1 and ψ_2 are two different functional of information matrix of a candidate design ξ, and κ is a blending coefficient that defines the weight of a candidate design's ψ_1 relative to ψ_2.

After having discussed the presence of separation in model (2.1), an optimal design-related approach that has been proposed to reduce separation probabilities is needed. In fact, *augmentation* has been proposed to resolve separation, which annexes additional trials (runs) to an initial experimental design after encountering separation or quasi-separation in the logistic regression model. Such approach presented in Park et al. (2020) uses maximum prediction variance (MPV) augmentation to eliminate separation. Augmenting design runs near the MPV region tends to produce overlapped response data. Be that as it may, the MPV-augmented designs produce much bigger A-optimal designs of equal size, which indicates that there is a trade-off between A-optimality and robustness to the separation happening. Motivated by these observations, the $A_{\{MP\}}$-criterion is considered in this paper as a compound criterion attempting to improve the separation robustness of the obtained designs.

Hence let us consider the following compound criterion that includes (a) the A-optimality criterion, and (b) a penalty term that captures the average distance of the candidate design's support points from the region of maximum prediction variance (MPV).

$$\psi_{A_{\{MP\}}} = \kappa \left[\frac{tr(M^{-1}(\xi^*))}{tr(M^{-1}(\xi))} \right] - (1-\kappa) \left[\frac{\sum_{i=1}^{N} |x_i^T \alpha|/N}{\max_{\mathscr{X}} |X^T \alpha|} \right]; \quad where \quad \kappa \in (0,1). \tag{3.1}$$

Since $tr(M^{-1}(\xi^*))$ is the minimum trace value of the information matrix obtainable across all candidate designs in $\mathscr{X}$, then $\left(\frac{tr(M^{-1}(\xi^*))}{tr(M^{-1}(\xi))} \right) \in (0,1)$.

The second term in Equation (3.1) is the normalized linear predictor penalty term, where $\alpha = (\alpha_0, \alpha_1)^T$. The numerator $\sum_{i=1}^{N} |x_i^T \alpha|/N$ is the direct measure of the average value of the linear predictor for all design points in candidate design ξ, a total measure of the distance of design points from the MPV region. A larger value of $\sum_{i=1}^{N} |x_i^T \alpha|/N$ indicates that the support points of the candidate design tend to make $Pr(Z=0)$ closer to 0 or 1. We thus would like to penalize the design points that are farther from the MPV region. Increasing the weight of the penalty term will therefore draw design points more closely to the MPV region of the design space. The denominator $\max_{\mathscr{X}} |X^T\alpha|$ denotes the maximum value of the linear predictor in $\mathscr{X}$.

Eventually, in order to minimize the first term and penalize it with the second term, we need to minimize $\psi_{A_{\{MP\}}}$. The design achieving this goal will be denoted as $\xi^*_{A_{\{MP\}}}$ and is the $A_{\{MP\}}$-optimal exact design. That is,

$$\xi^*_{A_{\{MP\}}} = \arg\min_{\xi}\{\psi_{A_{\{MP\}}}\} \tag{3.2}$$

The user-defined inputs for this criterion are:

- The number of observations in the experiment, N.
- The vector of logistic coefficients, $\alpha = (\alpha_0, \alpha_1)^T$.
- The blending coefficient that defines the weight of a candidate design's A-optimality relative to the average linear predictor magnitude penalty, $\kappa \in (0,1)$.

We note that our A_{MP}-criterion is a modification of the D_{MP}-criterion of Part et al. (2020). The D_{MP}-criterion was proposed for the logistic regression with the D-optimality criterion. We will study the performance of the $A_{\{MP\}}$-optimal exact designs of model (2.1) in **Simulation Study 2** in Chapter 5.

Chapter 4

SIMULATION STUDY 1

4.1 Locally A-Optimal Designs

Since the information matrix in (2.9) is a function of unknown parameters, a locally optimal design ξ^* that minimizes some real function ψ of $M(\xi,\theta,\sigma)$ for given values of the unknown model parameters is considered in this project. The function ψ may be the popularly used D-criterion, $\psi_0 = |M|^{-1/m}$, or A-criterion, $\psi_1 = tr(M^{-1})/m$, where m is the number of rows or columns of the information matrix $M(\xi,\theta,\sigma)$.

Based on the result of Kim (2019), an optimal design ξ^* of our setting will minimizes $\psi_1 = tr(M^{-1})$ under the A-optimality criterion. She also proved that A-optimal design can vary with values of α_0, α_1, and σ^2. With some computational and theoretical results, Kim (2019) provided some A-optimal designs for having a precise estimate of θ, excluding σ^2, as presented in Table A.1.

4.2 Data Simulation

The aim of this simulation study is to explore the performance of the three locally A-optimal Designs on the estimation of model parameters across different sample size. We will evaluate if the simulated data would allow us to successfully obtain stable ML estimates for θ, the corresponding standard errors, and study the average standard error of the parameter estimates of the design.

By giving the information in Table A.1, the selection of the range of covariate x_j is as

$$x_j = \frac{c_j - \alpha_0}{\alpha_1}$$

.

For each range of X, we consider three options for selecting value of x_i. The first option includes only the two boundary points of the range of X with equal weights for these two support points. The second option includes the boundary points and the center point and we again assign equal weights on the three support points. The third option is the corresponding A-optimal design as listed in Table A.1. For each of these designs, we then generate the binary response z_i and the continuous response y_i from the distribution specified in Chapter 2.

Moreover, the simulation for each design option will be conducted with sample size $N = 16, 30, 50$, and 100. The integer number of sample points n_j with each x_j is calculated by rounding Nw_j; where $j = 1, .., m; m$ is the selected number of distinct x-value. Hence for each design space we have three different simulated data sets. For convenience, we list the support points and the corresponding weights of the three design options for each design range in Table 4.1 to Table 4.3. Also, in this simulation study, the assumed true value for θ is as follows

$$\theta = (\alpha_0 = 1, \alpha_1 = 0.5, \beta_0^1 = 1, \beta_1^1 = -1, \beta_0^2 = -1, \beta_0^2 = 0.5, \sigma^2 = 1)^T.$$

Table 4.1: Simulation Study for Design Space [-4,0]

Design Options	x_m	Weights (w_m)
1	$\{-4, 0\}$	$\{0.5, 0.5\}$
2	$\{-4, -2, 0\}$	$\{0.333, 0.333, 0.333\}$
3	$\{-4, 0\}$	$\{0.195, 0.805\}$

.

Table 4.2: Simulation Study for Design Space [-6,2]

Design Options	x_m	Weights (w_m)
1	$\{-6,2\}$	$\{0.5,0.5\}$
2	$\{-6,-2,2\}$	$\{0.333,0.333,0.333\}$
3	$\{-6,-3.6,0,2\}$	$\{0.090,0.082,0.633,0.195\}$

Table 4.3: Simulation Study for Design Space [-12,8]

Design Options	x_m	Weights (w_m)
1	$\{-12,8\}$	$\{0.5,0.5\}$
2	$\{-12,-2,8\}$	$\{0.333,0.333,0.333\}$
3	$\{-12,-4,0.4,8\}$	$\{0.025,0.180,0.772,0.023\}$

4.3 Design Comparison

According to the simulation result, the **maxLik** function gives $\widehat{\theta_{MLE}}$ across all the sample sizes and design spaces. However, Figure 4.1, which depicts the number of times when all the SEs are finite, suggests that some of these ML estimates are not stable. Figure 4.1 indicates that the design option 3 has better performance for all of the design spaces. The design option 1 has the worst performance.

What can be conclude is that it might be hard to successfully estimate all the model parameters θ if only the boundary points are included in the design. It seems that having, additional points in the design increases the success rate in obtaining stable $\widehat{\theta_{MLE}}$. We also observe that, adequate information might be available from a given data set with $n \geqslant 30$ as an increase sample size seems to improve the parameter estimation.

Figure 4.2 presents the trace of $Var(\widehat{\theta_{MLE}})$ for the three design spaces across different sample sizes. Based on the result of Figure 4.2, boundary points help to obtain the lowest summation of $Var(\widehat{\theta_{MLE}})$ when design space is $[-4,0]$. For both design space $[-6,2]$ and $[-12,8]$, lowest summation of $Var(\widehat{\theta_{MLE}})$ is with the design option 3 . Note that low summation of $Var(\widehat{\theta_{MLE}})$ with deign option 1 when design space is $[-6,2]$ is meaningful, since deign option 1 has the lowest performance in estimating SE's. Eventually, enough sample size and adding design points assume a major part in selecting a design.

Figure 4.1: Comparison Result for Getting SE's of Three Design Space

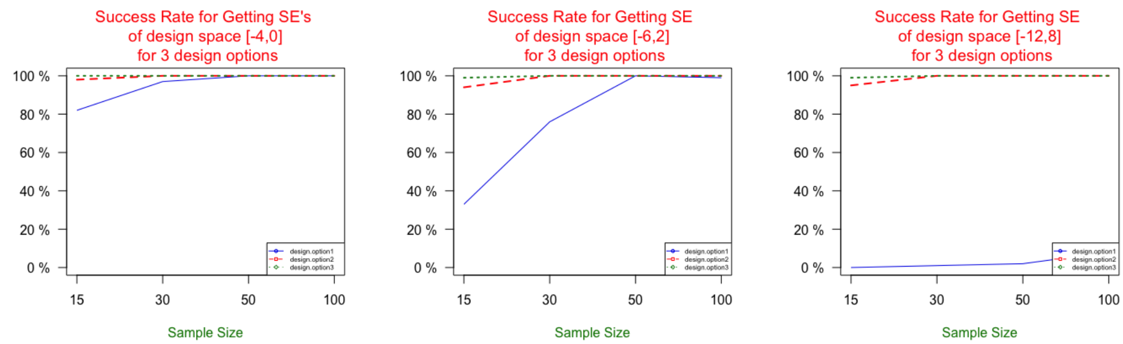
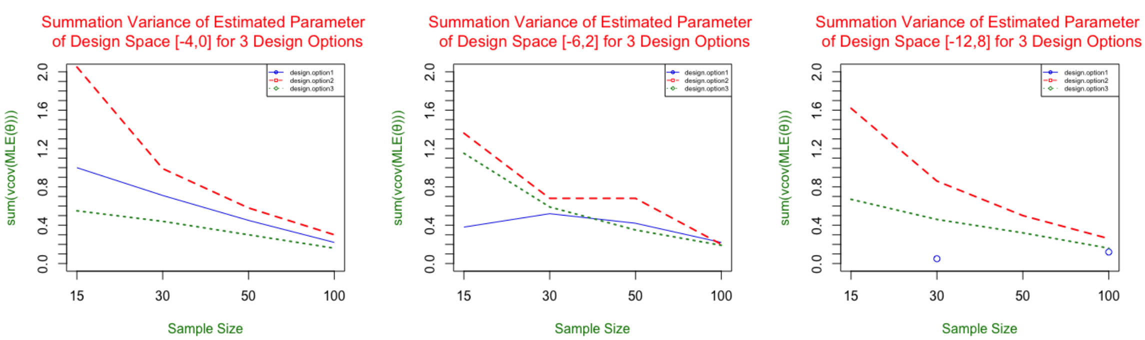


Figure 4.2: The Trace of $var(\widehat{\theta_{MLE}})$ for Three Design Spaces



Chapter 5

SIMULATION STUDY 2

5.1 Data Simulation

Based on $A_{\{MP\}}$-criterion proposed in Equation (3.1), an optimal design $\xi^*_{A_{\{MP\}}}$ of our setting will minimizes $\psi_{A_{\{MP\}}}$. The aim of this simulation study is to obtain the design that has the minimum $A_{\{MP\}}$-criteria on the estimation of model parameter across different sample sizes $N \in \{16, 30, 50, 100\}$; which allows us to evaluate the performance of $\psi_{A_{\{MP\}}}$, of successfully getting stable ML estimates for θ.

Now according to Equation 3.1, two inputs should be specified. These are the A-optimal exact design and the maximum absolute value of the linear predictor in $\mathscr{X}$. With the package **pracma** in the statistical software R, we use the R function **fmincon** to obtain A-optimal exact designs. We also have the following.

- For design space [-4,0], $\max_{\mathscr{X}} |X^T\alpha| = 1 \quad at \quad x = \{-4, 0\}$.
- For design space [-6,2], $\max_{\mathscr{X}} |X^T\alpha| = 2 \quad at \quad x = \{-6, 2\}$.
- For design space [-12,8], $\max_{\mathscr{X}} |X^T\alpha| = 5 \quad at \quad x = \{-12, 8\}$.

5.2 Designs Results

For each design space, we uses the **fmincon** function to find an A-optimal design. The the obtained A-optimal design will be used in Equation (3.1) to find an $A_{\{MP\}}$-optimal design for given κ. The **fmincon** function in R is again used for finding the latter design. Figure 5.1 presents the sum of the variances of parameter estimates, including $\hat{\sigma}^2$, of the A- and $A_{\{MP\}}$-optimal designs for the three design spaces across different sample sizes.

The blue solid line corresponds to the $A_{\{MP\}}$-optimal design and the red dash line denotes the A-optimal design provided by the **fmincon** function. Our observations are summarized below.

- For design space $[-4,0]$ with $\kappa = 0.8$, the $A_{\{MP\}}$-optimal design has slightly lower sum of variances of parameter estimates than the A-optimal design. The reduction in the total variance is mainly due to a reduction in $var(\hat{\sigma}^2)$; see also Figure 5.2. The probability of separation is reduced by 1.5% from the A-optimal design. Figure 5.3 presents our obtained $A_{\{MP\}}$-optimal designs for each N.

- For design space $[-6,2]$ with $\kappa = 0.9$, the $A_{\{MP\}}$-optimal design has a lower sum of variances of parameter estimates than the A-optimal design, especially when $N \leq 30$, and the probability of separation for both designs are similar. Figure 5.4 has the $A_{\{MP\}}$-optimal designs for each N.

- For design space $[-12,8]$ with $\kappa = 0.9$, the $A_{\{MP\}}$-optimal design has a better performance than A-optimal design and the probability of separation is reduced by 2.25%. Figure 5.5 has the $A_{\{MP\}}$-optimal designs for each N.

We also display the sum of the variances of parameter estimates with excluding $var(\hat{\sigma}^2)$ in Figure 5.2 for the three design spaces across different sample sizes.

5.3 Conclusion and Future Work

In this work, we propose the $A_{\{MP\}}$-optimality criterion for constructing designs for model (2.1) that have separation issues. For small-sample experiments with multiple factors, separation becomes a concern because it prevents the experimenter from making reasonable inferences. The results in the previous section showed that $A_{\{MP\}}$-optimal designs retain high A- values design while minimizing the probability of separation.

The development of these plans has the following restrictions. First, as is the case with every optimal design constructed for generalized linear models, $A_{\{MP\}}$-designs also suffer from the design-dependence problem. We showed that this issue may be mitigated by using compound criteria method for constructing robust designs. The advantage of the compound criteria method is that only initial parameters are required as inputs from the experimenter, which is more feasible in practice than specifying prior distributions, as required by the Bayesian method. Bayesian optimal designs, however, have the advantage of yielding more efficient designs because values of a criterion are calculated over the entire space of parameter values and take into consideration the amount of uncertainty in the initial parameter specifications. Using Bayesian methods for constructing robust $A_{\{MP\}}$- designs would be an direct extension of this work. Second, the choice of κ was explored sparingly in this project and ought to be examined officially in future work. A generalized methodology for optimization of the blending coefficient for any design situation implementing the logistic regression would be an important expansion to this project.

Figure 5.1: Comparison Result Between A-optimal Design and $A_{\{MP\}}$-optimal Design

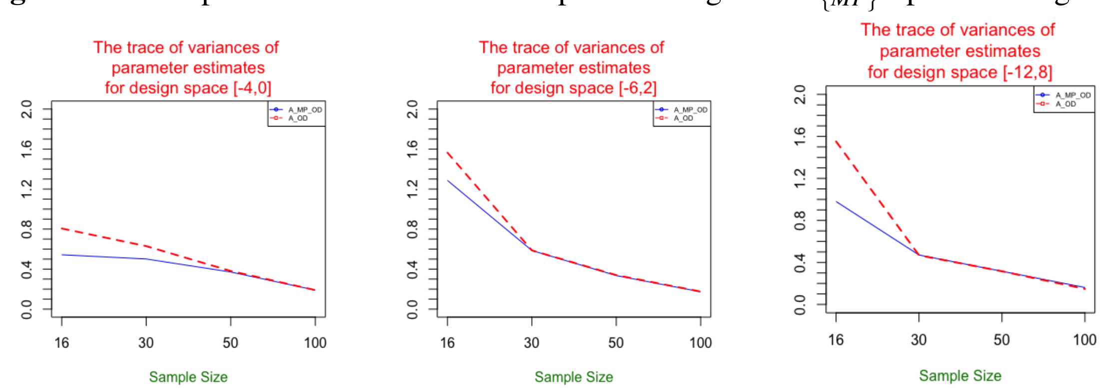


Figure 5.2: Comparison Result Between A-optimal Design and $A_{\{MP\}}$-optimal Design, Excluding σ^2

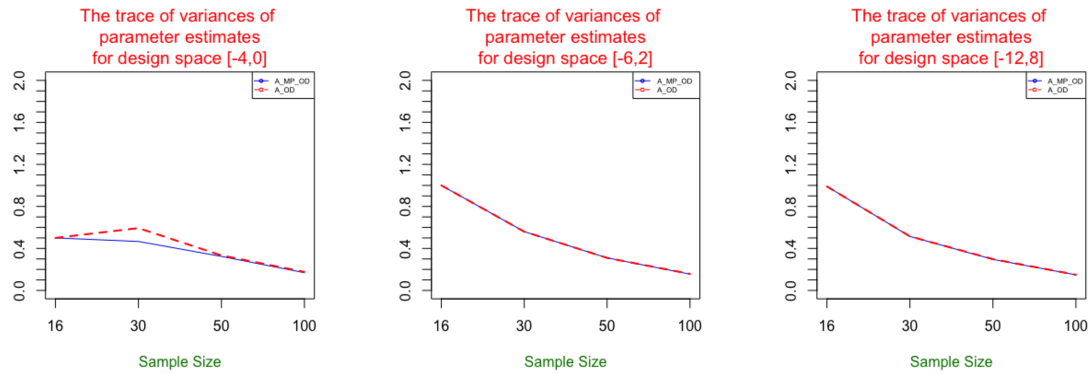


Figure 5.3: The $A_{\{MP\}}$-design Points of Design Space $[-4,0]$

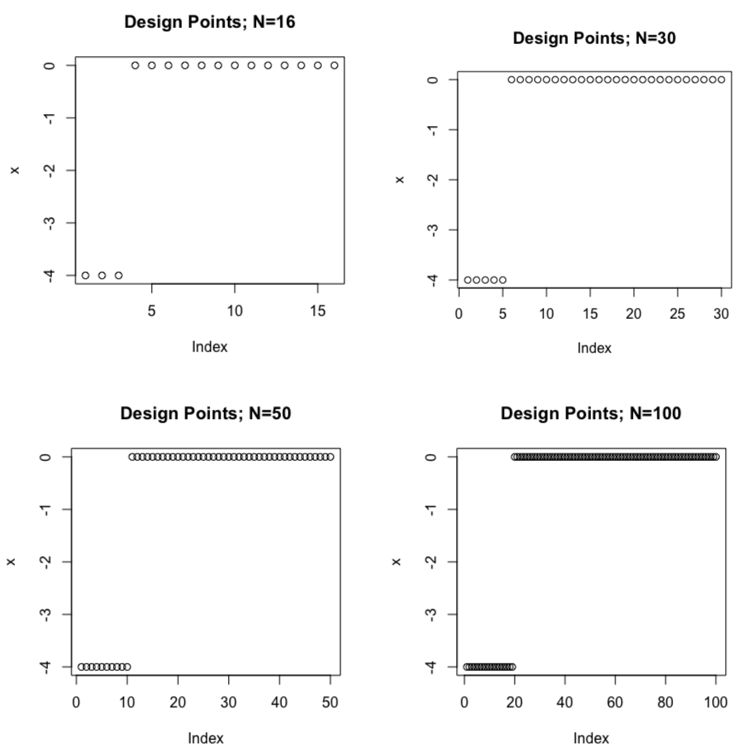


Figure 5.4: The $A_{\{MP\}}$-design Points of Design Space $[-6, 2]$

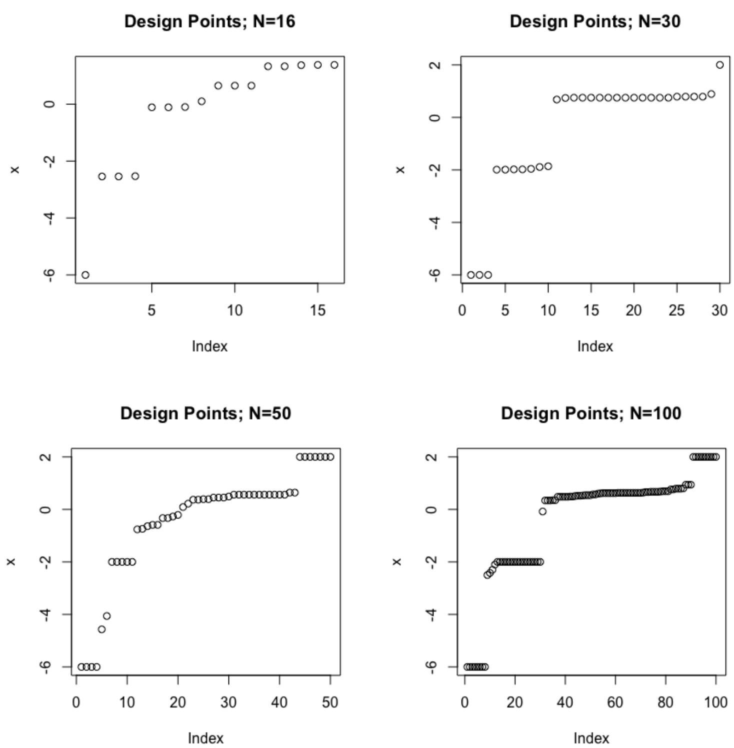
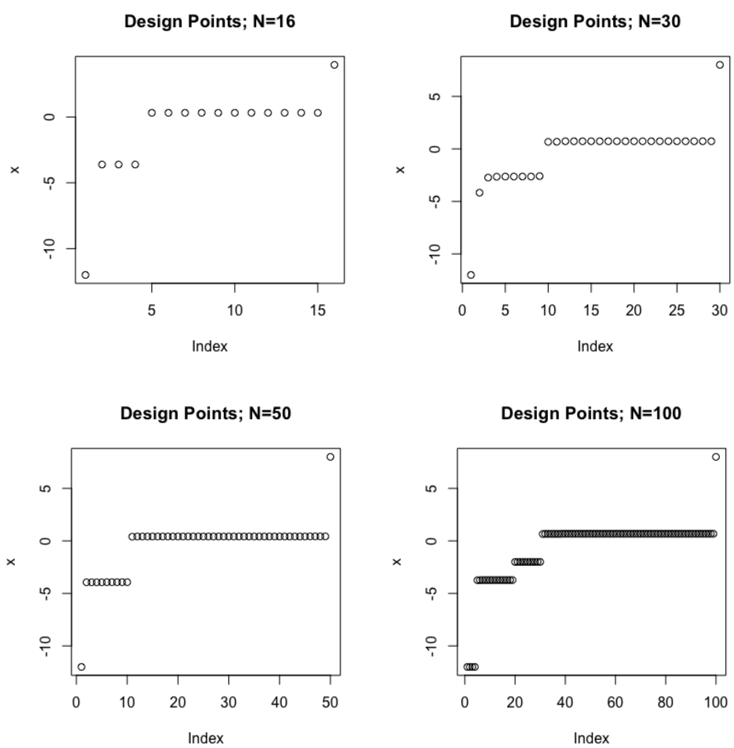


Figure 5.5: The $A_{\{MP\}}$-design Points of Design Space $[-12, 8]$



REFERENCES

Fitzmaurice, G. M. and N. M. Laird (1995). ”Regression-models for a Bivariate Discrete and Continuous Outcome with Clustering.” Journal of the American Statistical Association 90(431): 845-852.

Teixeira-Pinto, A. and S. L. T. Normand (2009). ”Correlated Bivariate Continuous and Binary Outcomes: Issues and Applications.” Statistics in Medicine 28(13): 1753-1773.

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

Kang, L., X. Kang, X. Deng and R. Jin (2018). ”A Bayesian Hierarchical Model for Quantitative and Qualitative Responses.” Journal of Quality Technology 50(3): 290-308.

Park, A. R., M. V. Mancenido and D. C. Montgomery (2020). ”Separation in D-optimal Experimental Designs for the Logistic Regression Model.” Quality and Reliability Engineering International 35(3): 776-787.

Kim, S. and M.-H. Kao (2019). ”Locally Optimal Designs for Mixed Binary and Continuous Responses.” Statistics and Probability Letters 148: 112-117.

APPENDIX A

A LOCALLY A-OPTIMAL APPROXIMATE DESIGNS BY KIM (2019)

Table A.1: Kim's *A*-optimal Designs

Type of Designs	Induced Design Space	Induced Support Points (c_m)	Weights (w_m)
A-op1	[-1,1]	(-1,1)	(0.195,0.805)
A-op2	[-2,2]	(-2,-0.788,1.006,2)	(0.090,0.082,0.633,0.195)
A-op3	[-5,5]	(-5,-0.961,1,216,5)	(0.025,0.180,0.772,0.023)