Separation in Optimal Designs for the Logistic Regression Model

by

Anson R. Park

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

Approved May 2019 by the
Graduate Supervisory Committee:

Douglas C. Montgomery, Co-chair
Michelle V. Mancenido, Co-chair
Adolfo R. Escobedo
Rong Pan

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Optimal design theory provides a general framework for the construction of experimental designs for categorical responses. For a binary response, where the possible result is one of two outcomes, the logistic regression model is widely used to relate a set of experimental factors with the probability of a positive (or negative) outcome. This research investigates and proposes alternative designs to alleviate the problem of separation in small-sample D-optimal designs for the logistic regression model. Separation causes the non-existence of maximum likelihood parameter estimates and presents a serious problem for model fitting purposes.

First, it is shown that exact, multi-factor D-optimal designs for the logistic regression model can be susceptible to separation. Several logistic regression models are specified, and exact D-optimal designs of fixed sizes are constructed for each model. Sets of simulated response data are generated to estimate the probability of separation in each design. This study proves through simulation that small-sample ($n \leq 32$) D-optimal designs are prone to separation and that separation risk is dependent on the specified model. Additionally, it is demonstrated that exact designs of equal size constructed for the same models may have significantly different chances of encountering separation.

The second portion of this research establishes an effective strategy for augmentation, where additional design runs are judiciously added to eliminate separation that has occurred in an initial design. A simulation study is used to demonstrate that augmenting runs in regions of maximum prediction variance (MPV), where the predicted probability of either response category is 50%, most reliably eliminates separation. However, it is also shown that MPV augmentation tends to yield augmented designs with lower D-efficiencies.

The final portion of this research proposes a novel compound optimality criterion,

D_{MP}, that is used to construct locally optimal and robust compromise designs. A two-phase coordinate exchange algorithm is implemented to construct exact locally D_{MP}-optimal designs. To address design dependence issues, a maximin strategy is proposed for designating a robust D_{MP}-optimal design. A case study demonstrates that the maximin D_{MP}-optimal design maintains comparable D-efficiencies to a corresponding Bayesian D-optimal design while offering significantly improved separation performance.

For Molly Eryn and Warrant boy

Life is wonderful with you two in it.

ACKNOWLEDGEMENTS

Thank you to Michelle Mancenido, Douglas Montgomery, Adolfo Escobedo, and Rong Pan. This research would not have been possible without you all and I am tremendously grateful for all of the patience, support, and inputs. To Michelle (Mickey) especially, thank you for going above and beyond to help me succeed in this program; I would be hopelessly lost without your assistance.

To my loving parents, Andrew, Wanda, Robin, and Chester, I thank you all for the money begrudgingly loaned to me as a poor graduate student. I have painstakingly worked out a sensible payment plan, and will fully settle all debts with lots of hugs and individualized smiley emojis at the end of this paragraph. You truly cannot put a price on that. ☺ ☺ ☺ ☺

I must extend a special expression of gratitude to Junki Yoshida for your gourmet teriyaki-based sauce; it has served as a delicious marinade for the thousands of chicken breasts that I grilled and consumed while in this program. For those not in the know, you may have seen him on KGW, where he performed short cooking segments, often in the guise of a strange samurai character that one might deem offensive to Asians. From the bottom of my heart, and I believe I have the requisite ethnic background to do this (25% Japanese, thanks grandma k.), you are forgiven for your shameless pandering to a predominantly white audience. *Arigatou gozaimasu* Mr. Yoshida!

There are so many additional family members, friends, training partners, and generous chipotle employees that gave me free guacamole that I should thank, but that would require a lot of extra typing and my fingers hurt. A heartfelt *mahalo nui loa* to you all.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

There are many experiments where the outcome, or response, is binary. One example is in medical trials, where a patient's status can be classified as either healthy (1) or diseased (0) (Bagley *et al.*, 2001). Binary responses are also often captured in military developmental and operational testing for the acquisition of new systems, where the outcome of an experiment can be whether a particular threat was defeated (1) or not (0). In military operational testing, measures of effectiveness are sometimes worded such that the acceptability of a new weapons system must be evaluated in terms of a binary response, such as pass or fail (Giadrosich, 1995). Efficient experimental designs for a binary response variable are therefore required across a broad spectrum of applications.

Optimal design theory provides a general framework for the construction of efficient experiments. It has been extended to include non-numeric response variables which are not normally distributed, such as the Bernoulli random variables used to model binary responses. Optimal experimental designs are constructed for an assumed model and statistical optimality criterion. When relating a set of experimental factors to a binary response, the logistic regression model is widely used. Among the proposed criteria in the literature, the D-criterion is one of the most popular, as D-optimal designs generally minimize the variance of the model parameter estimates (Montgomery, 2017). Algorithms and analytical procedures for constructing D-optimal designs for the logistic regression model, the assumed model form for binary responses in this research, have been proposed since the seminal work of Box and Lucas (1959). These procedures are implemented in statistical design packages

such as JMP® and have been found to perform better for a binary response than standard designs for normal-theory responses (Johnson and Montgomery, 2009). Optimal design criteria focusing on prediction variance, such as G-optimality, have also been used in design construction algorithms for the logistic regression model, such as the gCEA in Saleh and Pan (2016).

Efficient experimentation implies minimizing the number of experimental trials while ensuring sufficient information for model building or prediction. The paradigm of selecting minimal run sizes for binary response experiments may result in an estimation problem called *separation* (Agresti, 2013). Separation is a phenomenon that causes one or more coefficients of the logistic regression model to be inestimable through standard methods. Separated response data will produce a monotone likelihood function, which results in the non-existence of a unique maximum for at least one of the model coefficients (Albert and Anderson, 1984). While it has been recommended to use optimal designs for binary response experiments, the susceptibility of multi-factor optimal designs to separation has not been investigated.

Although the separation problem has been examined for single-factor D-optimal designs for the logistic regression model, the applicability of results from these works are limited. Fornius (2008) investigated several exact, single-factor locally D-optimal designs for a quadratic logistic regression model and noted that separation frequently occurs for small designs between 10-20 runs. Rahman (2015) proposed a compound optimality criterion to balance D-efficiency with a reduced probability of separation that was implemented for single-factor designs. However, an experimenter utilizing the D-criterion will likely have several design factors of interest, making studies of single-factor D-optimal designs less applicable in practice. The gap in this area justifies additional study of the separation problem for multi-factor D-optimal designs for the logistic regression model.

The design methodology explored in this research is *non-sequential*, where the entire experiment is planned before conducting any trials. *Sequential* design, on the other hand, involves dynamically determining the factor settings of the next experimental trial based on the results of the current trial. A non-sequential design strategy is regularly adopted in organizations which must adhere to a strict budget and rigid test plans. For example, in military operational testing, an experiment will typically not be allowed until the entire execution plan is presented and approved. This makes the use of a sequential design strategy infeasible (Giadrosich, 1995). Furthermore, a non-sequential design approach is often preferred in experiments involving animal subjects, such as the impact of the anticoagulant poison brodifacoum on cockroaches (Brooke *et al.*, 2013). In these types of studies, it is ideal for the test subjects to be sourced from the same generation and to have a homogeneous genetic profile. Sequential experimentation may not be desirable because the time required to conduct sequential trials may require different shipments of test subjects that will introduce undesired variability. Therefore, the separation problem will be approached from a non-sequential design perspective, where the focus is to create fixed designs that are robust to separation or to augment additional design runs in fixed blocks to an initial design that has encountered separation.

The main goal of this research is to provide methodologies for addressing separation in the design phase of an experiment. One aspect of this goal is to study how the separation problem manifests in multi-factor D-optimal designs for the logistic regression model as the design size varies. This will provide experimenters with sample size guidelines to minimize the risk of separation occurring in an experiment. The second aspect is to investigate an augmentation strategy, where additional runs are strategically placed in the design region once the initial design is executed, to eliminate separation as quickly as possible. The third aspect is to propose a novel

optimality criterion to generate designs for the logistic regression model that are more robust to separation.

Chapter 2 focuses on the modeling of binary response data and the problem of separation. The framework of generalized linear models (GLMs) is covered in Section 2.1, as it is necessary to develop the theoretical basis for the logistic regression model, covered in Section 2.2. Although other modeling methods for binary response data are available (Section 2.3), the logistic regression model is used solely in this research, as it is commonly employed in a variety of studies and has the clearest parameter interpretation, where the linear predictor can be viewed as a transformed odds ratio. The parameters of GLMs are estimated using the method of maximum likelihood (Section 2.4), which fails in the presence of separation (Section 2.5), as one or more of the parameter estimates will diverge to infinity. A survey of techniques for detecting separation and obtaining unique parameter estimates for the logistic regression model in the presence of separation are presented.

Chapter 3 covers optimal design theory and presents a literature survey on developments in optimal experimental designs for GLMs. Sections 3.1, 3.2, and 3.3 deal with the theoretical foundations of optimal design theory as they were originally presented by Kiefer (1959) for standard linear models, covering the difference between exact and continuous designs, alphabetic optimality criteria, and the General Equivalence Theorem (GET) (Kiefer and Wolfowitz, 1960). Section 3.4 covers the design dependence problem that manifests when constructing optimal designs for GLMs, where the information matrix, and hence the functionals of the optimality criteria, are dependent on the unknown model parameters. Several methods, such as creating locally optimal designs (Chernoff, 1953) and other robust design methods are reviewed. Techniques for constructing optimal designs for GLMs are covered in Section 3.5, with a survey of both analytical techniques and recent developments in

exact design algorithms.

The literature review in Chapters 2 and 3 identifies the following gaps:

1. There are currently no studies that examine the separation problem in exact, multi-factor optimal designs for the logistic regression model.

2. There has not been a proposed augmentation strategy to break separation in multi-factor experimental designs.

3. There is no statistical optimality criterion implemented for multi-factor designs that considers the separation problem.

Chapter 4 focuses on a study of the separation problem in multi-factor D-optimal designs for the logistic regression model. It explores how the separation problem diminishes with increasing design sizes for three logistic regression model forms with two design factors. The research in Chapter 5 contributes towards the second identified gap, as several augmentation strategies to break separation are investigated to determine an efficient methodology for eliminating separation that has occurred in an initial design. Finally, in Chapter 6, a compound optimality criterion is proposed to balance D-efficiency with separation robustness.

The major contributions of this research are as follows:

1. An exploration of the separation problem for exact, multi-factor D-optimal designs for the logistic regression model

2. An efficient strategy for augmenting design runs to eliminate separation in multi-factor experimental designs

3. A compound optimality criterion, D_{MP}, that can be implemented in a computer exchange algorithm to generate exact, multi-factor designs for the logistic regression model with reduced separation risk relative to D-optimal designs

4. A robust design methodology for exact D_{MP}-optimal designs to address the design dependence problem in GLMs

The first contribution demonstrates that small ($n \leq 32$) D-optimal designs are prone to encountering separation. Approximate sample size guidelines are provided for mitigating separation in D-optimal designs, which will allow experimenters to select a design size in accordance with their risk tolerance for encountering separation. The second contribution proposes a methodology to eliminate the separation problem as quickly as possible in multi-factor designs, enabling usable parameter estimates for inferential purposes. The last two contributions offer experimenters alternative compromise designs with high D-efficiencies and reduced separation risk. This work will be of significance to those seeking to use a non-sequential optimal design strategy in an experiment with a binary response variable, as it will provide valuable insight into mitigating, and if necessary, eliminating through augmentation, the problem of separation.

As this research serves as a first-look into addressing the separation problem in multi-factor designs, only two-factor cases are explored. The specified logistic regression models are also restricted to two-factor interaction and quadratic terms. Additionally, the large parameter standard error criterion recommended by Heinze and Schemper (2002) is solely used to detect separation, though other methods have been proposed in the literature. The exploration of the separation problem in small-sample optimal designs presented in Chapter 4 is restricted to the D-criterion, and the augmentation strategies proposed in Chapter 5 are based solely on locally and Bayesian D-optimal initial designs. Finally, in formulating a robust design strategy for the D_{MP}-criterion, only a maximin approach was explored. Alternative strategies are left for future research.

Chapter 2

MODELS

Designing efficient experiments for a binary response involves a less common application of experimental design, as the response variable is not normally distributed. For certain non-normal responses, standard design methods can still be applied through variance-stabilizing transformations on the response (Montgomery, 2017). However, fitting an appropriate generalized linear model (GLM) to the response data tends to produce better results, both in terms of the feasibility of the predicted response and the width of the confidence intervals about the mean response (Myers *et al.*, 2012; Lewis *et al.*, 2001). The logistic regression model, a type of GLM, will be the model form used for design construction methods in this research. The purpose of this chapter is to review the logistic regression model, along with certain problems that can occur when using this model in practice. This chapter proceeds as follows. An overview of GLMs is presented in Section 2.1, and the logistic regression model is covered in Section 2.2. Alternative methods of modeling binary data are covered in Section 2.3. Maximum likelihood (ML) estimation is the basis for parameter estimation in GLMs; it is discussed for the logistic regression model in Section 2.4. Finally, the problem of separation, which occurs in models with a categorical response variable, is covered in Section 2.5, along with a literature review surveying proposed methods of detecting separation and alternative model parameter estimation methods that are available with separated response data.

2.1 Generalized Linear Models

GLMs were first proposed by Nelder and Wedderburn (1972) to address the limitations of standard linear regression for normally distributed errors. A standard linear regression model relating a response variable y to a set of regressor variables, $\mathbf{x}$, based on n observations has the form shown in Equation 2.1,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.1}$$

where $\mathbf{X}$ is a matrix of covariate values, $\boldsymbol{\beta}$ is a column vector of model parameters, and $\boldsymbol{\epsilon}$ is a column vector of the error terms associated with each observation. It is assumed that ϵ_i for all $i = 1, ..., n$ are independent and identically distributed (IID) $N(0, \sigma)$ random variables (Montgomery *et al.*, 2012). With these assumptions, the method of ordinary least squares produces ML estimates (Myers and Montgomery, 1997). In cases where these assumptions are not valid, GLMs accommodate a broader range of response distributions. The structure of a GLM consists of (Myers *et al.*, 2012):

1. The *response distribution*: $y_1, ..., y_n$ are independent response observations with means $\mu_1, ..., \mu_n$, respectively. Each y_i has a distribution that is a member of the exponential family.

2. A *linear predictor* involving the regressor variables, $\eta = \mathbf{x}^T\boldsymbol{\beta}$.

3. A *link function*, $g(\cdot)$, that relates the mean of the response distribution to the linear predictor, $\eta_i = g(\mu_i)$, for all $i = 1, ..., n$. The link function is monotonic, continuous, and differentiable.

GLMs can be used to model any response variable that is a member of the exponential family of distributions. The probability density (mass) function of distributions

that are members of the exponential family has the general form shown in Equation 2.2,

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\} \tag{2.2}$$

where $a(\cdot)$, $b(\cdot)$, and $c(\cdot)$ denote functions specific to the distribution, θ is the location parameter, and ϕ is the dispersion parameter (Myers *et al.*, 2012). If the link function is chosen such that the natural location parameter is equal to the linear predictor, then $g(\mu_i)$ is referred to as a canonical link function. The use of a canonical link results in simplified mathematics when deriving ML estimates, which is covered in Section 2.5.

2.2 The Logistic Regression Model

Logistic regression is an appropriate analysis technique when a response variable, y, is categorical. See Cramer (2002) for a summary of the origins of the logistic function and the development of logistic regression in statistics. In this research, only a binary response variable will be considered, such that y results in only two dichotomous outcomes that can be coded as 0 or 1. For example, the response variable in a study of weapon accuracy, where the outcome of each trail is classified as either hit (1) or miss (0), is binary. The model relates the probability of response $y = 1$ to k regressor variables, denoted as $\mathbf{x} \in \mathbb{R}^p$, where p denotes the length of $\mathbf{x}$ when expanded to model form ($p \geq k$).

The logistic regression model has the form shown in Equation 2.3.

$$y_i \sim \text{Bernoulli}(E(y_i)), \quad \text{for } i = 1, ..., n \tag{2.3}$$

The logistic response function, $E(y_i)$, has the form shown in Equation 2.4,

$$E(y_i) = \pi_i = (1 + \exp(-\mathbf{x}_i{}^T\boldsymbol{\beta}))^{-1} \tag{2.4}$$

where $\boldsymbol{\beta}$ represents a vector of model parameters. In Equation 2.4, π_i can be interpreted as $P(y_i = 1|\mathbf{x}_i)$. The linear predictor of the logistic response function is shown in Equation 2.5.

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \eta_i = \mathbf{x}_i{}^T\boldsymbol{\beta} \tag{2.5}$$

In Equation 2.5, η_i can be interpreted as a transformed (log) odds ratio, since $\eta_i = \ln\left(\frac{P(y_i=1|\mathbf{x}_i)}{P(y_i=0|\mathbf{x}_i)}\right)$ (Montgomery *et al.*, 2012).

The logistic regression model is a type of GLM which has the following defining structure:

1. Response distribution: For binary data, it is in most cases reasonable to assume that $y_1, ..., y_n$ are independent Bernoulli random variables with means $\pi_1, ..., \pi_n$, respectively. The Bernoulli distribution is a member of the exponential family (Casella and Berger, 2002).

2. Linear predictor: $\mathbf{x}^T\boldsymbol{\beta}$ is present in the logistic response function.

3. Link function: The logit link $g(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ is used to relate the linear predictor to π_i. It is the canonical link for the assumed response distribution (Myers *et al.*, 2012).

Logistic regression is widely used in medical research (Tai and Machin, 2013), economics, business analytics (Ledolter, 2013) and many other fields. For example, Bagley *et al.* (2001) provide a survey of logistic regression methods used in testing for cancer susceptibility.

2.3 Alternative Methods of Modeling Binary Response Data

The logistic regression model uses the logistic link function, which is the canonical link for the Bernoulli distribution (see appendix A of Myers and Montgomery (1997)).

However, the choice of possible link functions extends beyond the canonical link for each response distribution. There are other link functions that are used to model binary response data in the framework of GLMs (see Piegorsch (1992)), such as the *probit* link, which takes the form shown in Equation 2.6,

$$g(E(y_i)) = \Phi^{-1}(E(y_i)) \tag{2.6}$$

where Φ represents the cumulative distribution function of the standard normal distribution. The *complimentary log-log* link function is also used, which takes the form shown in Equation 2.7.

$$g(E(y_i)) = \ln(-\ln(1 - E(y_i))) \tag{2.7}$$

The link function is a critical component of a GLM, and the choice of a particular link function is somewhat analogous to the choice of a particular transformation methodology on the response data. Although the link function is a transformation on the population mean and not on the response data, the use of an improper link function can lead to a poor model fit (Myers *et al.*, 2012).

Probit regression is a commonly used alternative to logistic regression. The use of the probit link function in Equation 2.6 leads to the probit response function shown in Equation 2.8,

$$E(y_i) = \Phi(\mathbf{x}_i{}^T\boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i{}^T\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\, dz \tag{2.8}$$

which represents the area under the standard normal curve to the left of $\mathbf{x}_i{}^T\boldsymbol{\beta}$. Generally, the shape of a fitted probit model will be similar to a logistic regression model fit from the same data, with differences becoming more pronounced in the extreme tails (Zelterman, 1999).

Beginning with the work of Bliss (1934), probit regression has been used extensively in medical research and other fields. Fisher (1935a), in an appendix to work

on fitting a dosage-mortality curve by Bliss (1935), showed how ML estimation can be used to obtain parameter estimates for the probit model. In choosing between logistic or probit regression, Berkson (1951) makes the argument that the logistic model is preferred since the score equations are easier to solve, but this problem has been mitigated by statistical software that can expediently fit both logistic and probit models. Chambers and Cox (1967) show that discriminating between a logistic and probit model is only possible for large sample sizes ($n \geq 1000$) and when the majority of the data lies at the bounds of the independent variable. The bulk of subsequent literature surveyed recommends that for most applications, the choice of using either link function comes down to personal preference, as both provide nearly identical substantive conclusions (see Long (1997), Gill (2000), Hardin and Hilbe (2007)). Research comparing various Bayesian and sample theory model selection criteria via Monte Carlo methods showed that in the case of balanced data, none of the model selection criteria could distinguish between the probit and logit regression models (Chen and Tsurumi, 2010), which illustrates that the logistic and probit models can be very similar.

The complementary log-log model, though not as commonly utilized as the logistic or probit regression models, can also be used to model a binary response. The use of the complementary log-log link function in Equation 2.7 leads to the complementary log-log response function, shown in Equation 2.9.

$$E(y_i) = 1 - \exp(-\exp(\mathbf{x}_i{}^T\boldsymbol{\beta})) \tag{2.9}$$

For the logit and probit link functions, the property $g(x) = -g(1-x)$ holds, giving the logistic and probit response functions ($E(y)$ vs. $\mathbf{x}_i{}^T\boldsymbol{\beta}$) symmetry about the point $\mathbf{x}_i{}^T\boldsymbol{\beta} = 0.5$. The complementary log-log response function does not have this symmetry. This can be seen in Figure 2.1. In cases where it would be more appropriate

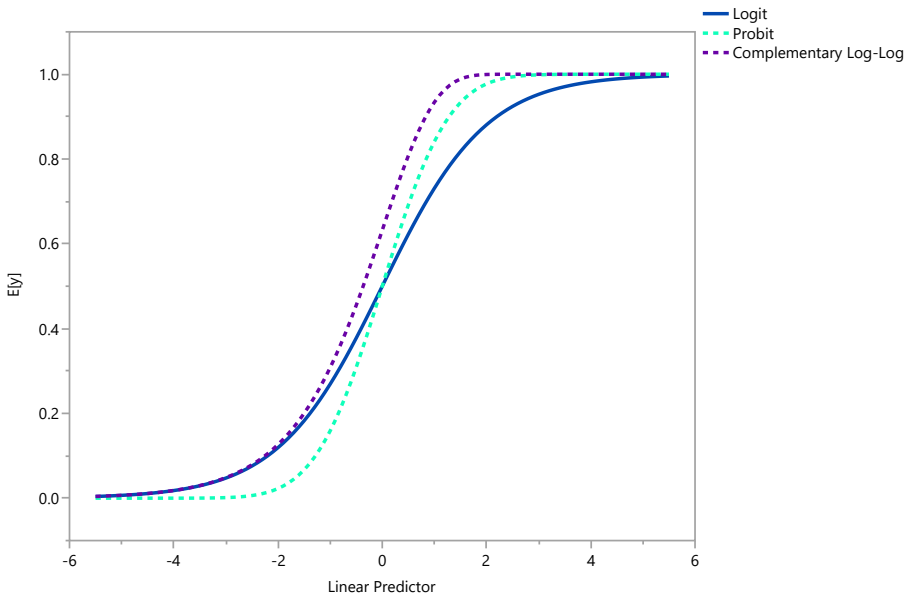


Figure 2.1: $E(y)$ vs. Linear Predictor Value for the Logit, Probit, and Complementary Log-Log Response Functions

for the response curve to increase more sharply towards 0 or 1 from $\mathbf{x}_i{}^T\boldsymbol{\beta} = 0.5$, the complementary log-log function will provide a better model fit. For recent applications of complementary log-log regression, see Boyko *et al.* (2015) and Cooper *et al.* (2015). Penman and Johnson (2009) present a simulation study that highlights some advantages of complementary log-log regression over other binomial regression methods.

The focus of this research will be on optimal experimental designs for the logistic regression model, as it is widely used and has a more straightforward log-odds interpretation than the probit and complementary log-log link functions. However, exploring optimal experimental designs for the probit or complementary log-log models may be an area for future work.

2.4 Maximum Likelihood Estimation

ML estimation is a method of estimating the parameters of a statistical model, denoted as $\boldsymbol{\beta}$. Given n observations, the parameter values that maximize the probability of realizing the observed response values are called the ML estimates of the model parameters (Casella and Berger, 2002). This is accomplished by first specifying the joint density function of the observations,

$$f(y_1, ..., y_n|\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}) \tag{2.10}$$

where the equality follows as shown in Equation 2.10 if each trial is IID. Now, the observed response values $y_1, ..., y_n$ are considered fixed, and the likelihood function, L, is equivalent in expression to the joint density function, but is viewed as a function of the parameters. Therefore, the likelihood function can be expressed as shown in Equation 2.11,

$$L(\boldsymbol{\beta}|y_1, ..., y_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}) \tag{2.11}$$

where the equality again follows with the IID observations assumption. To determine the ML estimates of the model parameters, denoted as $\hat{\boldsymbol{\beta}}_{\text{ML}}$, the set of equations $\frac{\partial L}{\partial \boldsymbol{\beta}} = \mathbf{0}$ is solved, assuming a closed-form solution to L exists (Casella and Berger, 2002). The log-likelihood transformation is often used to make the differentiation process more straightforward. Since $\ln(x)$ is monotonically increasing, solving the set of equations $\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$, where $l = \ln(L)$, will yield the same ML estimates (Allison, 2008).

When ML estimation is applied to GLMs, the log-likelihood function takes the form shown in Equation 2.12.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \tag{2.12}$$

The set of equations $\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$ are referred to as the score equations and are mostly solved through iterative methods, as $\boldsymbol{\mu}$ is usually nonlinear in $\boldsymbol{\beta}$ (Myers and Montgomery, 1997). The general form of the derivative of the log-likelihood for a GLM is shown in Equation 2.13.

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \frac{\partial l}{\partial \theta_i} \frac{\partial \theta_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \boldsymbol{\beta}} \tag{2.13}$$

However, when a canonical link function is used, $\eta_i = \theta_i$, so the score equations simplify to the form shown in Equation 2.14,

$$\sum_{i=1}^{n} (y_i - \mu_i)\, \mathbf{x}_i = \mathbf{0} \tag{2.14}$$

assuming $a(\phi)$ is constant (ϕ known). Solving the system of p equations in (2.14) will yield $\hat{\boldsymbol{\beta}}_{ML}$. If the model assumptions are satisfied and the link function is correct, it can be shown that asymptotically, $E(\hat{\boldsymbol{\beta}}_{ML}) = \boldsymbol{\beta}$ (Myers *et al.*, 2012).

The variance of the score equations, typically referred to as the information matrix, plays a crucial role in optimal design theory, which will be covered in Chapter 3. Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ denote the model matrix made up of rows ${\mathbf{x}_1}^T, ..., {\mathbf{x}_n}^T$, $\mathbf{y} \in \mathbb{R}^n$ denote a column vector of responses, $\boldsymbol{\mu} \in \mathbb{R}^n$ denote a column vector of means, and $\hat{\boldsymbol{\beta}}$ denote estimates of $\boldsymbol{\beta}$. The information matrix, denoted as $\mathbf{M}(\hat{\boldsymbol{\beta}})$, is shown in Equation 2.15 (assuming ϕ is known).

$$\mathbf{M}(\hat{\boldsymbol{\beta}}) = \mathrm{Var}\left\{\mathbf{X}^T(\mathbf{y} - \boldsymbol{\mu})\right\} = \mathbf{X}^T\mathbf{V}\mathbf{X} \tag{2.15}$$

In Equation 2.15, $\mathbf{V} \in \mathbb{R}^{n \times n}$ is a diagonal weight matrix dependent on $\boldsymbol{\mu}$, which depends on the response distribution. The asymptotic variance-covariance matrix of estimators $\hat{\boldsymbol{\beta}}$ is the inverse of the information matrix, shown in Equation 2.16 (Myers *et al.*, 2012).

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{M}^{-1}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\mathbf{V}\mathbf{X}\right)^{-1} \tag{2.16}$$

In the context of the logistic regression model, since each y_i follows a Bernoulli distribution, the corresponding probability mass function for each observation is shown

in Equation 2.17 (Casella and Berger, 2002).

$$f_i(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{2.17}$$

It cannot be assumed that the observations are identically distributed, as each y_i may have a unique π_i. However, since the observations are assumed to be independent, the likelihood function can be formulated as shown in Equation 2.18.

$$L(\boldsymbol{\beta}|y_1,...,y_n) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{2.18}$$

By utilizing the log-likelihood transformation, $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ for the logistic regression model are the solutions to the score equations shown in Equation 2.19 (Allison, 2008).

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \mathbf{x}_i y_i - \sum_{i=1}^{n} \mathbf{x}_i \pi_i = \mathbf{0} \tag{2.19}$$

For most data, no closed form solution to $\frac{\partial l}{\partial \boldsymbol{\beta}} = \mathbf{0}$ exists. Iteratively reweighed least squares with the Newton-Raphson algorithm is typically used in practice to find $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$. This method will be covered in detail in Section 2.5. For the logistic regression model, the diagonal elements of $\mathbf{V}$ are shown in Equation 2.20.

$$v_{ii} = \pi_i(1-\pi_i) = \frac{\exp(\mathbf{x}_i{}^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i{}^T\boldsymbol{\beta}))^2}, \quad \text{for } i = 1,...,n \tag{2.20}$$

2.5 The Problem of Separation

Experimental studies with categorical responses often have small samples and a large number of factors, which can sometimes result in ML estimation failing to converge to unique parameter estimates. This phenomenon, known as separation, occurs when a hyperplane passing through the design space can completely or quasi-completely separate all of the design points with a response value of $y = 0$ from all of the design points with a response value of $y = 1$ (Albert and Anderson, 1984). If there exists a column vector $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}^T\mathbf{x}_i > 0$ when $y_i = 1$ and $\boldsymbol{\beta}^T\mathbf{x}_i < 0$ when $y_i = 0$ for

all $i = 1, ..., n$, then complete separation is present. Quasi-complete separation occurs when there are both $y = 0$ and $y = 1$ responses on the separating hyperplane. If $\boldsymbol{\beta}^T \mathbf{x}_i \geq 0$ when $y_i = 1$ and $\boldsymbol{\beta}^T \mathbf{x}_i < 0$ when $y_i = 0$ for all $i = 1, \ldots, n$, then quasi-complete separation is present. The data set in Table 2.1, when used to fit a logistic regression model, illustrates a case of complete separation, where x is a single regressor variable.

Table 2.1: Response Data with Complete Separation

x	-5	-4	-3	-2	-1	1	2	3	4	5
y	0	0	0	0	0	1	1	1	1	1

For this data set, point $x = b$ along the 1-dimensional x line for any $b \in (-1, 1)$ completely or quasi-completely separates the data. This can be seen in Figure 2.2. For example, $y = 1$ when $x > 0$ and $y = 0$ when $x < 0$ holds for every observation. When

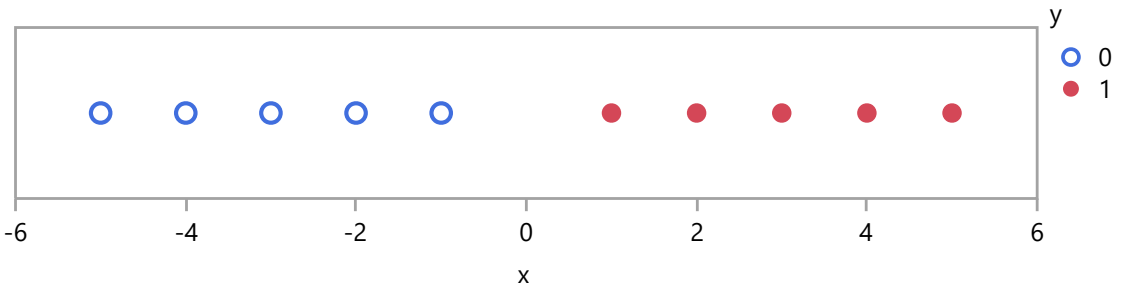


Figure 2.2: Separated Response Data Example

complete or quasi-complete separation is present, one or more of the ML estimates of the logistic regression model do not exist (Albert and Anderson, 1984). To see this, recall that the model with one regressor variable has the form shown in Equation 2.21.

$$y_i = \frac{1}{1 + \exp(-\beta_0 - \beta_1 x_i)} + \epsilon_i \tag{2.21}$$

The log-likelihood function for the data shown in Table 2.1 is shown in Equation 2.22.

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{10} y_i \left(\beta_0 + \beta_1 x_i\right) - \sum_{i=1}^{10} \ln(1 + \exp(\beta_0 + \beta_1 x_i)) \tag{2.22}$$

For simplicity, suppose that $\hat{\beta}_0^{ML} = 0$. Then, the log-likelihood function as a function of $\hat{\beta}_1$ can be expressed as shown in Equation 2.23.

$$l(\hat{\beta}_1) = 15\hat{\beta}_1 - \sum_{i=1}^{10} \ln(1 + \exp(\hat{\beta}_1 x_i)) \tag{2.23}$$

Equation 2.23 is monotonically increasing; as $\hat{\beta}_1 \to \infty$, $l \to 0$. Consequently, a unique estimate of $\hat{\beta}_1^{ML}$ does not exist, as $\hat{\beta}_1$ can be made arbitrarily large to force l arbitrarily close to zero. This is illustrated in Figure 2.3. For experimental designs constructed

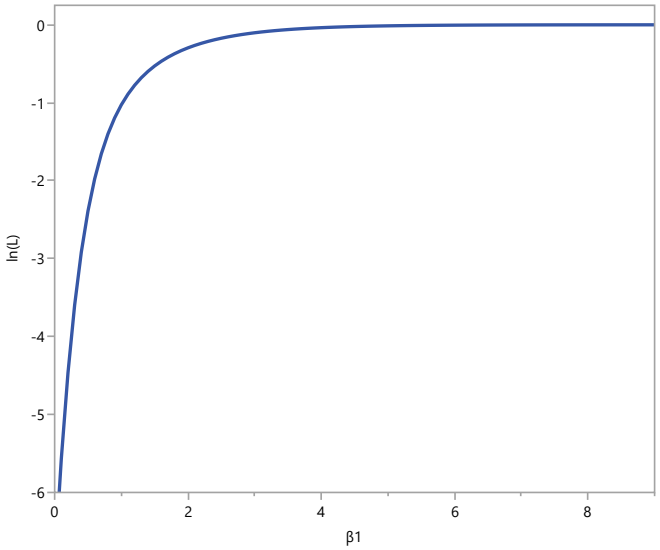


Figure 2.3: Monotone Likelihood with Separated Response Data Example

for a logistic regression model, the presence of separation will make it impossible to use standard ML estimation techniques to estimate the parameters of the underlying model. This results in the inability to use the fitted model for inferential purposes.

Significant advances have been made to detect separation and mitigate its effects in estimation. Santner and Duffy (1986) devised a mixed integer linear program capable of classifying data as either completely separated, quasi-separated or overlapped

(no separation present). A widely-used and easy detection approach is to observe the magnitude of parameter estimates and corresponding standard errors. Large parameter estimates with enormous standard errors typically indicate that some form of separation is present (Allison, 2008). Several approaches have been proposed to alleviate the problem of separation in ML estimation. It is possible to simply delete the problem variable(s) that are causing the data to be separated from the model. However, even when it is possible to determine the variable(s) that cause separation, these variables(s) are highly likely to have the greatest impact on the response. Thus, deleting these variable(s) will decrease the utility of the model. Another approach is to use exact methods for logistic regression, where instead of relying on asymptotic properties of estimators, tests of significance are based on the exact probability distribution (for details, see Agresti (2013)). This approach, however, becomes computationally inefficient with larger sample sizes and higher model complexity.

Konis (2007) presents a linear programming approach that can be used to detect separation. Consider a set of n design points, denoted as $\mathbf{x}_i$, $i = 1, ..., n$. Let E_1 denote the set of indices i such that $y_i = 0$ and E_2 denote the set of indices i such that $y_i = 1$. The distance of design point $\mathbf{x}_i$ from hyperplane H is given by $\tilde{s}_i = \mathbf{x}_i^T \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is a unit vector normal to H. If complete separation is present, there exists a $\tilde{\boldsymbol{\beta}}$ such that $\tilde{s}_i > 0$ for all $i \in E_2$ and $\tilde{s}_i < 0$ for all $i \in E_1$. In the case of quasi-complete separation, there exists a $\tilde{\boldsymbol{\beta}}$ such that $\tilde{s}_i \geq 0$ for all $i \in E_2$ and $\tilde{s}_i \leq 0$ for all $i \in E_1$, with $\tilde{s}_i = 0$ for at least one value of $i \in \{1, ..., n\}$. Since the design matrix of a candidate design is assumed to be of full rank, $\tilde{s}_i \neq 0$ holds for at least one value of $i \in \{1, ..., n\}$. With this

notation, finding $\tilde{\boldsymbol{\beta}}$ can be formulated as the optimization problem shown in (2.24).

$$\begin{aligned} \max \quad & S(\tilde{\boldsymbol{\beta}}) = \textstyle\sum_{i=1}^{n} \tilde{s}_i \\ \text{subject to} \quad & \tilde{s}_i \geq 0 \text{ for } i \in E_2 \\ & \tilde{s}_i \leq 0 \text{ for } i \in E_1 \\ & \tilde{\boldsymbol{\beta}}^T \tilde{\boldsymbol{\beta}} = 1 \end{aligned} \tag{2.24}$$

If (2.24) is feasible, the optimal separating hyperplane H^* is defined as the hyperplane with $\tilde{\boldsymbol{\beta}}^*$ given by the optimal solution to (2.24). If (2.24) is infeasible, separation is not present. In this form, (2.24) is a type of nonlinearly constrained optimization problem due to the quadratic constraint forcing $\tilde{\boldsymbol{\beta}}$ to be of unit length (Konis, 2007). The formulation of (2.24) is similar to the optimal hyperplane algorithm presented by Cortes and Vapnik (1995), where the authors prove that the determination of an optimal separating hyperplane is a type of quadratic programming problem. However, since the length of $\tilde{\boldsymbol{\beta}}$ is irrelevant in determining the presence of separation, the quadratic constraint can be removed, which yields the linear program

$$\begin{aligned} \max \quad & \mathbf{e}^T \bar{\mathbf{X}} \boldsymbol{\beta} \\ \text{subject to} \quad & \bar{\mathbf{X}} \boldsymbol{\beta} \geq \mathbf{0} \\ & \boldsymbol{\beta} \text{ free} \end{aligned} \tag{2.25}$$

where $\mathbf{e} \in \mathbb{R}^n$ denotes a column vector of ones and $\bar{\mathbf{X}} = \text{diag}(\tilde{\mathbf{y}})\mathbf{X}$, where $\tilde{\mathbf{y}} \in \mathbb{R}^n$ has elements $\tilde{y}_i = 1$ when $y_i = 1$ and $\tilde{y}_i = -1$ when $y_i = 0$ and $\mathbf{X}$ denotes the *design* matrix of a candidate design (Konis, 2007). If separation is not present, the optimal value of the objective function in (2.25) is zero. If the optimal value of the objective function in (2.25) is greater than zero, then there exists $\boldsymbol{\beta} \neq \mathbf{0}$ that is feasible, implying that separation is present. Many linear programming solvers, such as the MATLAB® `linprog` function, can be used to determine the optimal solution to (2.25).

Firth's modified score procedure, which was initially proposed to reduce the bias of ML estimates in binomial logistic regression (Firth, 1993), has been used to obtain finite, unique parameter estimates in data where separation is present (Heinze and Schemper, 2002; Heinze, 2006; Zorn, 2005). For most models and data sets, these is no closed-form solution to the likelihood score equations, and $\hat{\boldsymbol{\beta}}_{ML}$ is determined for a logistic regression model using numerical methods, such as the Netwon-Raphson algorithm. Let $\mathbf{U}(\boldsymbol{\beta})$ denote the score equations shown in Equation 2.19, and let $\mathbf{J}(\boldsymbol{\beta})$ denotes a square matrix of second derivatives as shown in Equation 2.26.

$$\mathbf{J}(\boldsymbol{\beta}) = \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i{}^T \pi_i (1 - \pi_i) \tag{2.26}$$

The Newton-Raphson algorithm in terms of these expressions is shown in Equation 2.27,

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \mathbf{J}^{-1}\left(\boldsymbol{\beta}^{(s)}\right)\mathbf{U}\left(\boldsymbol{\beta}^{(s)}\right) \tag{2.27}$$

where s denotes the iteration number (Allison, 2008). A starting set of values, $\boldsymbol{\beta}^{(0)}$, is required. The accuracy of $\boldsymbol{\beta}^{(0)}$ is not of particular importance, as $\boldsymbol{\beta}^{(0)} = \mathbf{0}$ will suffice (Allison, 2008). Iterations will continue until some convergence criterion is satisfied, where the maximum change in parameter estimates from one iteration to the next is less than some value. The convergence rate of Newton-Raphson in determining ML estimates for overlapped data is typically quite rapid; it can be shown that if the initial parameter estimates are close enough to the true parameter values, then the convergence rate is quadratic (Kuk and Cheng, 1997). In other words, the convergence rate is $\mathcal{O}(s^{-2})$, where s denotes the number of iterations. However, when the Newton-Raphson algorithm is applied to a separated data set, it will not converge; iterations will continue until the fixed iteration limit is reached. At each iteration, the parameter estimate(s) for the regressor variable (or combination of variables) causing separation will continue to increase in magnitude with each iteration, and the associated standard

errors of these parameter estimates will become very large (Allison, 2008). This is the reason that enormous parameter standard errors can serve as a reliable criterion for declaring separation in a data set; very large standard errors indicate the non-convergence of numerical algorithms, such as Newton-Raphson, caused by the non-existence of at least one ML parameter estimate with separated response data.

Firth's modified score equations have the form shown in Equation 2.28,

$$\mathbf{U}(\boldsymbol{\beta})' = \sum_{i=1}^{n} \mathbf{x}_i y_i - \sum_{i=1}^{n} \mathbf{x}_i \pi_i - \left[\sum_{i=1}^{n} h_i \mathbf{x}_i \left(\frac{1}{2} - \pi_i \right) \right] \tag{2.28}$$

where h_i is the i^{th} diagonal element of $\mathbf{H} = \mathbf{V}^{0.5}\mathbf{X}\left(\mathbf{X}^T\mathbf{V}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{0.5}$. Using the modified score equations in the implementation of the Newton-Raphson algorithm guarantees finite parameter estimates for the logistic regression model (Allison, 2008).

Heinze and Schemper (2002) presented two clinical data sets where Firth's modified score procedure was used to obtain finite parameter estimates in data sets where quasi-complete separation was present. Heinze (2006) presents a case study where ML estimation fails to produce finite estimates in a nearly separated data set. The use of Firth's procedure (also referred to as penalized ML estimation) was recommended over the use of ML and exact logistic regression, as estimation and inference was also not possible in many cases with exact logistic regression. Bayesian approaches have also been used as an alternative estimation method in the presence of separation (Gelman *et al.*, 2008). Abrahantes and Aerts (2012) proposed an approach for clustered or repeated measures data that uses penalized ML in generalized linear mixed models (GLMM). Lipsitz *et al.* (2013) presented an extension of penalized likelihood to obtain estimates for the proportional odds model, which is used to model ordinal responses.

The estimation procedures just discussed have limitations and criticisms. Rainey (2016) shows that the penalty term in Firth's score equations, which is equivalent

to the Jeffreys prior for the logistic regression model (Jeffreys, 1946), often yields point estimates that are too small. It has been noted that Firth's procedure is also susceptible to producing poor parameter estimates in a separated data set, especially when the data set is small ($n = 10,\ 20$) (Fornius, 2008).

Several authors have explored sequential design strategies to prevent separation in sensitivity tests, where the goal is to estimate the threshold level of a single design factor that will generate a success (1) versus a failure (0). Novel three-part testing strategies are proposed by Neyer (1994) and Wu and Tian (2014), where the second part of the tests are focused on generating overlapped response data to find unique ML estimates. However, these strategies are not directly applicable to this work. For single-factor experiments, detecting overlap is trivial, as it occurs when the minimum level of an observed success is smaller than the maximum level of an observed failure. With multiple design factors, determining regions of overlap is not as straightforward. Additionally, both strategies involve dynamically adjusting the locations of runs as the experiment progresses, which is not applicable to the study of separation in non-sequential designs.

Another option to mitigate separation that has not been explored thoroughly is through the selection of an adequate experimental design. The research presented in Chapters 4 through 6 will investigate the separation problem, both in the design phase of an initial experiment and for a follow-on phase to augment an initial design that encounters separation.

Chapter 3

EXPERIMENTAL DESIGNS

The field of statistical design of experiments has progressed tremendously since the pioneering work of Fisher (1935b) in agricultural studies, where the cornerstone principles of randomization, replication, and blocking were established along with the use of factorial designs. Optimal design theory, laid out in its modern form by Kiefer (1959), utilizes a specific statistical criterion for constructing experimental designs. Optimal experimental design for GLMs is a critical component of the research presented in Chapters 4 through 6.

The purpose of this chapter is to present an overview of optimal design theory and to survey developments in optimal experimental designs for GLMs and the logistic regression model, specifically algorithms for constructing exact experimental designs under different optimality criteria. The basic notation and concepts of optimal design theory for standard linear models are covered in Section 3.1, with specific optimality criteria summarized in Section 3.2. The General Equivalence Theorem, a critical tool in establishing optimality results for designs, is presented in Section 3.3. Section 3.4 summarizes the modifications of optimal design theory required for applications to GLMs. The chapter ends with Section 3.5, a literature review of optimal design construction methods available for GLMs.

3.1 Optimal Design Theory

Let ξ denote an experimental design, where weight w_i specifies the proportion of experimental runs at support point $\mathbf{x}_i$.

$$\xi = \left\{ \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_t \\ w_1 & w_2 & \ldots & w_t \end{array} \right\} \tag{3.1}$$

Equation 3.1 represents a design with t support points, where $w_i > 0$ and $\sum_{i=1}^{t} w_i = 1$. An *exact* (discrete) design of run size n requires that each $w_i = r_i/n$ be a ratio of integers, where r_i denotes the number of replicates at $\mathbf{x}_i$. As n and r_i are integers, w_i will be a ratio of integers. A *continuous* (approximate) design relaxes the integer restriction on n and r_i (Atkinson and Woods, 2015).

In practice, exact designs are favored because continuous designs often yield non-integer allocations to support points. However, it is more convenient mathematically to work with continuous designs, because closed-form, analytical solutions to exact design problems are often intractable (Berger and Wong, 2009). By relaxing the integer restrictions in continuous design problems, a design size n does not need to be specified. The focus of this research is on exact designs, where a sample size n is specified and heuristic algorithms are used to construct designs in accordance with the specific optimality criterion being used.

To create an optimal experimental design, one must specify (Atkinson and Woods, 2015):

1. A model of interest. For GLMs, one must also specify a set of parameter values $\boldsymbol{\beta}$, with a prior distribution $f(\boldsymbol{\beta})$ if applicable.

2. A design criterion, which will typically be some function of the information matrix (see Chapter 2).

3. A design region, denoted as χ, which contains feasible points in $\mathbb{R}^p$.

In optimal design theory, the empirical model must be assumed in the design stage, which may lead to a poor design if the assumed model is not correct. There has been research on optimal designs that are robust to a misspecified model (Tommasi, 2012), but this research assumes the specified model is appropriate. For this research, only optimal designs for the logistic regression model are examined.

For continuous designs, the general measure of imprecision for design ξ, denoted as $\Psi\{\mathbf{M}(\xi;\boldsymbol{\beta})\}$, is used to assess a design's optimality. An optimal design, denoted as ξ^*, minimizes $\Psi\{\mathbf{M}(\xi)\}$, where $\mathbf{M}(\xi)$ denotes the information matrix of design ξ (Atkinson and Woods, 2015). The efficiencies of several designs can be compared by assessing the objective function value for a particular optimality criterion for each design. If ξ^* is the optimal design for parameter values $\boldsymbol{\beta}$ (or prior distribution $f(\boldsymbol{\beta})$, if applicable), another design ξ can be assessed relative to the optimal design by comparing $\Psi\{\mathbf{M}(\xi)\}$ to $\Psi\{\mathbf{M}(\xi^*)\}$. When the comparison is made, the objective function is evaluated for the same $\boldsymbol{\beta}$ (or $f(\boldsymbol{\beta})$, if applicable) (Atkinson and Woods, 2015). The optimal continuous design serves as a useful benchmark when assessing the relative efficiencies of several exact designs, even when the continuous design does not have integer allocations for support points.

3.2 Criteria of Optimality

This section will cover several well-known optimality criteria, which can generally be divided into two categories:

1. Functionals of the eigenvalues of the information matrix

2. Criteria concerning the variance of predictions

The first category of optimality criteria generally serves to minimize the variance of

the model parameter estimates, which is useful when the goal of the experiment is model building or variable screening. The second category of optimality criteria is more appropriate if the experimenter desires to make predictions at any point within the design space.

3.2.1 Parameter-Based Optimality Criteria

Let $\lambda_1, ..., \lambda_p$ denote the eigenvalues of the information matrix of a candidate design, denoted as $\mathbf{M}(\xi)$. In terms of the eigenvalues of $\mathbf{M}(\xi)$, the A-, D-, and E-optimality criteria are shown in Equation 3.2.

$$\begin{aligned} A&\text{: min } \sum_{i=1}^{p} \lambda_i^{-1} \\ D&\text{: min } \prod_{i=1}^{p} \lambda_i^{-1} \\ E&\text{: min } \{\max\ (\lambda_i^{-1})\} \end{aligned} \tag{3.2}$$

A-optimal designs minimize the average variance of the parameter estimates, while D-optimal designs minimize the generalized variance of the parameter estimates (Atkinson *et al.*, 2007). E-optimal designs minimize the maximal variance among all best linear unbiased estimators of normalized linear contrasts (Filipiak and Różański, 2013). In other words, E-optimal designs maximize the minimum eigenvalue of $\mathbf{M}(\xi)$, which corresponds to minimizing the maximum variance of the parameter estimates (maximum eigenvalue of $\mathbf{M}^{-1}(\xi)$).

To illustrate how these criteria work to yield attractive parameter estimates, consider the ellipsoid of concentration on $\hat{\boldsymbol{\beta}}$ for design ξ, given in Equation 3.3.

$$(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T\ \mathbf{M}(\xi)\ (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) = R^2 \tag{3.3}$$

For a more comprehensive presentation of a concentration ellipsoid, see Nordström (1991). The confidence region for the parameter estimates takes the general form

of the concentration ellipsoid. For example, if $\hat{\boldsymbol{\beta}} \in \mathbb{R}^p$ are approximately normally distributed, then the ellipsoidal confidence region with coverage probability $1-\alpha$ is shown in Equation 3.4,

$$CI_{1-\alpha,\chi^2} = \left\{ \boldsymbol{\beta} \mid (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \, \mathbf{M}(\xi) \, (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq \chi^2_{p,\alpha} \right\} \tag{3.4}$$

where $\chi^2_{p,\alpha}$ is the $100(1-\alpha)$ percentile of the χ^2 distribution with p degrees of freedom (Fedorov and Leonov, 2013). $\mathbf{M}$ must be a positive semi-definite symmetric matrix. If $\mathbf{M}^{-1}$ exists ($|\mathbf{M}| \neq 0$), then $\mathbf{M}$ is positive definite, and the the size of the ellipsoid defined in Equation 3.4, centered at $\hat{\boldsymbol{\beta}}$, is determined by the asymptotic covariance matrix, $\mathbf{M}^{-1}(\xi)$. The lengths of the semi-axes of the ellipsoid are proportional to $\sqrt{{\lambda_i}^{-1}}$, where ${\lambda_i}^{-1}$ represents the eigenvalues of $\mathbf{M}^{-1}(\xi)$ (as $\mathbf{M}^{-1}(\xi)$ is also positive definite, ${\lambda_i}^{-1} > 0$ for all i) (Boyd and Vandenberghe, 2004).

Consider the D-criterion. Recall that the determinant of a square matrix $\mathbf{A}$ is equal to the product of all its eigenvalues, and that if λ is an eigenvalue of $\mathbf{A}$, then λ^{-1} will be an eigenvalue of $\mathbf{A}^{-1}$ (Poole, 2014). Therefore, minimizing $\prod_{i=1}^{p} \lambda_i^{-1}$ (maximizing $|\mathbf{M}(\xi)|$) generally shrinks the values of each ${\lambda_i}^{-1}$ closer to zero, which minimizes the volume of the ellipsoidal confidence region, as the volume of the ellipsoid is proportional to $|\mathbf{M}^{-1}(\xi)|^{1/2}$. This implies more precise parameter estimates. The A- and E-criteria consider other aspects of the joint confidence ellipsoid, but conceptually these criteria focus on the same goal of shrinking the joint confidence region.

Several useful extensions to D-optimality have been proposed. The D_A-criterion (Sibson, 1972) is used when only s linear combinations of $\boldsymbol{\beta}$ are of interest, where $s < p$. The D_S-criterion is appropriate when the interest is estimating a subset s of the parameters as precisely as possible (Atkinson *et al.*, 2007). Generalized D-optimality, also called S-optimality by Läuter (1976), is a linear combination of D_A criteria for multiple models. Atkinson and Cox (1974) use the generalized D-criterion

to simultaneously estimate parameters in several candidate models. A summary of the functional measures of imprecision for the optimality criteria presented in this section is shown in Table 3.1, where $\mathbf{A}$ is a $p \times s$ matrix of rank s and each w_i is a non-negative weight such that $\sum_{i=1}^{h} w_i = 1$.

Criterion	**Ψ, Functional**
A	$\text{tr}\{\mathbf{M}^{-1}(\xi)\}$
D	$\log\lvert\mathbf{M}^{-1}(\xi)\rvert$
E	$\min\ \lambda_i(\mathbf{M}(\xi)) = \lambda_{\min}$
D_A	$\log\lvert\mathbf{A}^T\mathbf{M}^{-1}(\xi)\mathbf{A}\rvert$
Generalized D	$\sum_{i=1}^{h} w_i \log\lvert\mathbf{A}_i^T\mathbf{M}^{-1}(\xi)\mathbf{A}_i\rvert$

Table 3.1: Parameter-Based Optimality Criteria

The D-criterion is the most widely used of all optimality criteria, as it has attractive mathematical properties and useful design applications. D-optimal designs for continuous factors do not depend on the scale of the design variables; linear transformations do not affect the design matrix, which is not always the case for A- and E-optimal designs (Atkinson *et al.*, 2007). Computationally, exact D-optimal designs are expedient to construct due to efficient calculation methods for updates to the inverse of a matrix determinant (Goos, 2012).

3.2.2 Prediction-Based Optimality Criteria

Suppose that the experimenter desires to make response predictions within the design region. For this application, a criterion focusing on minimizing the variance of prediction would likely be more useful than any parameter-based optimality criterion.

A G-optimal design minimizes the maximum prediction variance over the design

region. An I-optimal design (also referred to as Q-, V-, or IV-optimal) minimizes the average prediction variance over the design region (Goos, 2012). Box and Draper (1959, 1963) first presented the concept of minimizing integrated prediction variance over a design region. The functionals of these criteria are shown in Table 3.2, where χ represents the design region and $\mathbf{x}$ denotes a column vector of the design variables $x_1, ..., x_k$ expanded to model form.

Criterion	**Ψ, Functional**
G	$\max_{\mathbf{x}\epsilon\chi} \mathbf{x}^T \mathbf{M}^{-1}(\xi)\mathbf{x}$
I	$\int_\chi \mathbf{x}^T \mathbf{M}^{-1}(\xi)\mathbf{x}$

Table 3.2: Variance-Based Optimality Criteria

Implementing expedient exact design algorithms for the G-criterion is much more computationally expensive than it is for the D-criterion, as a search over χ to identify the maximum prediction variance is required for each candidate design (Rodríguez *et al.*, 2010). Several approaches, such as simulated annealing (Haines, 1987) and genetic algorithms (see Hamada *et al.* (2001), Borkowski (2003), and Heredia-Langner *et al.* (2004)) have been used to construct G-efficient designs.

3.3 The General Equivalence Theorem

For continuous designs, the General Equivalence Theorem (GET) provides necessary and sufficient conditions to declare ξ optimal over χ. As originally presented by Kiefer and Wolfowitz (1960), the GET states that for linear models, assuming that χ is compact and that Ψ is convex and differentiable, a design that is D-optimal is also G-optimal. Kiefer and Wolfowitz (1960) prove that $\max_\chi d(\mathbf{x};\xi) \geq p$, where $d(\mathbf{x};\xi) = \mathbf{x}^T \mathbf{M}^{-1}(\xi)\mathbf{x}$ is the standardized variance of prediction and p is the number of

parameters in the linear model. Therefore, a design having a maximum standardized variance of p is a sufficient condition to declare G-optimality, which implies that the design is also D-optimal.

The GET has been extended to nonlinear models, where an analogue of D- and G-optimality equivalence is presented (White, 1973). A more general presentation of the GET for linear models is given by Kiefer (1974), where the class of functions $\Phi_p(\mathbf{M}(\xi))$ is introduced, taking the form shown in Equation 3.5,

$$\Phi_p(\mathbf{M}(\xi)) = \left(v^{-1}\mathrm{tr}\left(\mathbf{M}^{-1}(\xi)\right)^p\right)^{1/p}, \quad 0 < p < \infty \tag{3.5}$$

where v is the dimension of the information matrix. Φ_0, defined as $\lim_{p\downarrow 0}\Phi_p$, corresponds to D-optimality, Φ_1 corresponds to A-optimality, and Φ_∞, defined as $\lim_{p\to\infty}\Phi_p$, corresponds to E-optimality. In its general form for continuous designs, the GET establishes the equivalence of the following conditions on the optimal design, denoted as ξ^* (Pukelsheim, 1993; Atkinson *et al.*, 2007).

1. Design ξ^* minimizes $\Psi\{\mathbf{M}(\xi)\}$
2. $\min\{\phi(\mathbf{x},\xi^*)\} \geq 0$, where $\phi(\mathbf{x},\xi^*) = \lim_{\alpha\to 0^+}\frac{1}{\alpha}(\Psi\{(1-\alpha)\mathbf{M}(\xi^*)+\alpha\mathbf{M}(\bar{\xi})\}-\Psi\{\mathbf{M}(\xi^*)\})$, the first derivative of Ψ at ξ^* in the direction of $\bar{\xi}$, which places unit mass at point $\mathbf{x}$
3. $\phi(\mathbf{x},\xi)$ achieves its minimum at the points of design ξ^*

Therefore, the GET provides the means of verifying optimality for continuous designs. At ξ^*, all gradients $\phi(\mathbf{x},\xi^*)$ are non-negative. If there exists some direction where $\phi(\mathbf{x},\xi) < 0$, the current design ξ is not optimal.

In general, the GET does not hold for exact designs (Atkinson *et al.*, 2007). However, optimal continuous designs are used as baselines to calculate design efficiencies. Two particularly convenient forms of design efficiency exist for D- and G-optimality.

Suppose that $\xi \in \chi$ denotes an arbitrary design, and ξ^* denotes the optimal continuous design in the same design region χ. The D- and G-efficiencies of ξ are given by Equations 3.6 and 3.7, respectively,

$$D_{\text{eff}} = \left(\frac{|\mathbf{M}(\xi)|}{|\mathbf{M}(\xi^*)|} \right)^{\frac{1}{p}} \tag{3.6}$$

$$G_{\text{eff}} = \left(\frac{p}{\bar{d}(\xi)} \right) \tag{3.7}$$

where $\bar{d}(\xi) = \max_{\mathbf{x} \in \chi} d(\mathbf{x}, \xi)$ (Atkinson *et al.*, 2007). It should be noted that when comparing candidate exact designs with the continuous optimal design, caution should be taken in the interpretation of efficiency metrics. Low efficiencies do not necessarily indicate that the designs perform poorly with respect to the criterion; rather, it should be a comparative metric among competing exact designs.

3.4 Optimal Designs for GLMs

Box and Lucas (1959) first addressed the problem of constructing optimal designs for non-linear models. Since then, optimal designs for GLMs has also received much attention due to the wide application of GLMs in scientific and engineering experiments. Research on optimal designs for the logistic regression model has been mostly focused on D-optimality (Khuri *et al.*, 2006).

Optimal designs for standard linear models are only functions of the design points, as the information matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ is solely based on the model matrix $\mathbf{X}$. However, as implied by Equation 2.15, this is not the case for GLMs, as the weight matrix $\mathbf{V}$ is dependent on the model parameter values. The alphabetic optimality criteria described in Section 3.2 operate in the same manner for GLMs as for linear models (typically as functionals of the information matrix), but the information matrix for GLMs is a function of both the design points and the parameter values. Therefore, constructing optimal designs for GLMs is encumbered by this dependence on the

presumably unknown parameters, a problem referred to as design dependence (Khuri *et al.*, 2006).

For the logistic regression model, the functional for D-optimality is $\Psi = \log|\mathbf{M}^{-1}(\xi)|$, with the added complexity that the information matrix is now a function of both the model matrix $\mathbf{X}$ and the weight matrix $\mathbf{V}$. For the functionals of I- and G-optimality, Myers *et al.* (1994) present a convenient form for the asymptotic average prediction variance (APV) of logit $\hat{\pi}_i$ for the single-variable case of the logistic regression model. The parameterization of the logistic regression model takes the form shown in Equation 3.8,

$$\text{logit } \pi_i = \beta_1(x_i - \mu) = z_i \tag{3.8}$$

where $\mu = -\beta_0/\beta_1$ and corresponds to the ED_{50}, which represents the value of x_i that produces $\pi_i = 0.5$, a 50 percent response rate. For the symmetric design region $[-a, a]$, the APV of logit $\hat{\pi}_i$ for an m-level design with n_i runs at the ith level is given by Equation3.9,

$$\text{APV} = \frac{n}{|\mathbf{M}(\hat{\boldsymbol{\beta}})|}\left[\sum_{i=1}^{m} n_i\pi_i(1-\pi_i){z_i}^2 + \frac{a^2}{3}\sum_{i=1}^{m} n_i\pi_i(1-\pi_i)\right] \tag{3.9}$$

where $n = \sum_{i=1}^{m} n_i$ and $\mathbf{M}(\hat{\boldsymbol{\beta}})$ represents the information matrix obtained by the fitted model logit $\hat{\pi}(z_i) = \hat{\beta}_0 + \hat{\beta}_1 z_i$. An I-optimal design will minimize APV. Assuming that the design is symmetric about ED_{50}, the maximum prediction variance (MPV) over region $[a-, a]$ takes the form shown in Equation 3.10.

$$\text{MPV} = \frac{n}{|\mathbf{M}(\hat{\boldsymbol{\beta}})|}\left[\sum_{i=1}^{m} n_i\pi_i(1-\pi_i){z_i}^2 + a^2\sum_{i=1}^{m} n_i\pi_i(1-\pi_i)\right] \tag{3.10}$$

A G-optimal design will yield the minimum value of the MPV of logit $\hat{\pi}_i$ over the region of interest. Myers *et al.* (1994) observe that the MPV of logit $\hat{\pi}$ is at the boundary of the design region. Consequently, the G-optimal design minimizes the prediction variance at the boundaries of the region of interest.

Several methods have been used to handle the design dependence problem in GLMs. One method of obtaining optimal designs for GLMs is to specify the best guess for the parameter vector $\boldsymbol{\beta}$, which can be derived from historical data, subject matter experts, or earlier experiments of a similar nature. Design points are then chosen with the assumed parameter vector. This method, proposed by Chernoff (1953), yields locally optimal designs, as these designs are optimal for the $\boldsymbol{\beta}$ that was specified. While this method makes design construction more analogous to the linear case, this method may also lead to inefficient designs if the assumed values for $\boldsymbol{\beta}$ are far from the ground truth (Chaloner and Larntz, 1989). Abdelbasit and Plackett (1983) investigated single-factor binary response experiments and demonstrated some robustness issues that can occur with poor initial guesses for $\boldsymbol{\beta}$.

An augmentation strategy, also referred to as a sequential design approach, may be used, where an initial design is used to obtain estimates of $\boldsymbol{\beta}$. Once the estimates are in place, additional runs are added to the design to maximize the optimality criterion of choice. Abdelbasit and Plackett (1983) used a two-stage procedure to reduce the effect of inaccurate initial parameter estimates, with Minkin (1987) expanding on their work by considering the quality of the estimates derived from the first stage. Some sequential design methodologies are presented in Ford *et al.* (1989). The two-stage approach has been applied to experiments for symmetric binary response models, to include the logistic and probit models (Sitter and Forbes, 1997).

One may also adopt a Bayesian approach by specifying a prior distribution for $\boldsymbol{\beta}$, denoted as $f(\boldsymbol{\beta})$, to model the uncertainty in the parameter values. This prior distribution is then incorporated into the design criterion. For example, Tsutakawa (1980) uses Bayesian techniques for the design problem of estimating an extreme percentage point of the logistic distribution. D-optimality for a GLM can be implemented by selecting design points to maximize the expectation of the logarithm of the determi-

nant of the information matrix (Johnson and Montgomery, 2009). Mathematically, this design criterion can be written as shown in Equation 3.11,

$$\Psi(\xi) = E_{\beta}(\log|\mathbf{M}(\xi)|) = \int \log|\mathbf{X}^T\mathbf{V}\mathbf{X}|f(\boldsymbol{\beta})d\boldsymbol{\beta} \tag{3.11}$$

and was initially proposed by Chaloner and Larntz (1989) to construct optimal designs for a single-factor logistic regression model. For a model with p parameters, direct use of the criterion in Equation 3.11 is computationally prohibitive, as the p-dimensional integral must be evaluated many times throughout design construction. However, Gotwalt *et al.* (2009) devised a numerical integration technique that requires only $O(p^2)$ evaluations for normal prior distributions that makes use of generalized Gauss-Laguerre quadrature. Parsa Maram and Jafari (2016) created designs for two- and three-parameter logistic regression models with a random intercept using the criterion in Equation 3.11. Chaloner and Verdinelli (1995) provide an excellent overview of the Bayesian approach to experimental design.

Pronzato and Walter (1988) present another method to create robust experimental designs for nonlinear models through maximin optimization, where the parameters are assumed to belong to some range of values without any other assumptions about their distribution. A similar approach has also been used in literature to find an analytical solution for maximin D-optimal designs for an exponential growth model, where maximin D-optimal describes a design that maximizes the minimal D-efficiency among the parameter ranges (Imhof, 2001).

Some authors have explored extensions to the GET for GLMs. Stufken and Yang (2012) derive an equivalence theorem for Φ_p-optimality, which can be applied to establish locally A-, D-, and E-optimal designs for GLMs. Recently, Li and Deng (2018) established equivalence results for I-optimality for GLMs.

3.5 Constructing Optimal Designs for GLMs

Once a GLM has been specified, an optimality criterion selected, and some method of initial parameter specification is in place, an optimal design can be constructed. However, even for a point estimate of the parameter values, which would lead to a locally optimal design, optimal design construction is a difficult problem for most practical applications, as optimal support points and their corresponding weights must be found. Analytical solutions to the optimization of the objective function are typically intractable, as the objective function is too complex, often with many variables (Stufken and Yang, 2012). A survey of the methods of design construction for GLMs, with a focus on the logistic regression model, will be presented in this section. Analytical approaches will be covered along with algorithms for exact design construction.

3.5.1 Analytical Construction Methods

The geometric approach, first proposed by Elfving (1952) for a two-variable linear model, has been used to study optimal designs for GLMs. This approach works for GLMs by transforming the weighted design problem in design space χ into an unweighted linear design problem in the induced design region Z, which depends on both χ and the parameter vector $\boldsymbol{\beta}$. The support points for an optimal design lie on the "smallest ellipsoid" centered at the origin that contains Z, where the ellipsoid can be explicitly defined based on the optimality criterion being used (Stufken and Yang, 2012). With this approach, the dependence of the optimal design on the values of $\boldsymbol{\beta}$ is replaced in the transformed problem by a design space that varies with $\boldsymbol{\beta}$. If an optimal design is found for an arbitrary Z, the design problem is implicitly solved for an arbitrary χ and $\boldsymbol{\beta}$. Ford *et al.* (1992) used the geometric approach to derive

locally D-optimal designs for several nonlinear models. Dette and Haines (1994) created E-optimal local designs for nonlinear models with two parameters. Biedermann *et al.* (2006) presented Φ_p-optimal designs for binary response experiments with two parameters.

Mathew and Sinha (2001) derived locally D-optimal, and in special cases, locally A- and E-optimal designs for a single explanatory variable logistic regression model with two parameters. The authors presented a unified approach for deriving D- and A-optimal designs for this model. Haines *et al.* (2007) created 3- and 4-support point locally D-optimal designs for the logistic regression model with two explanatory variables and no interaction term. While global D-optimality was only proven analytically for a special case of the 3-point design, an analytical proof for D-optimality for the same 3- and 4-point designs for all negative and zero intercept parameters is provided in Kabera *et al.* (2015). Yang *et al.* (2011) proposed a method for obtaining locally D-optimal designs for the logistic regression model with any number of design variables where the linear predictor takes the form $\eta(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i$ and the design region χ is unbounded for one variable, so that $\chi = [-1, 1]^{k-1} \times \mathbb{R}$. Locally D-optimal designs for the two explanatory variable logistic regression model with interaction are derived through a combination of algebraic and numerical methods by Haines *et al.* (2018). Local D-optimality for these designs is declared via numerical methods. A summary of D-optimal designs for the logistic regression model that encompasses most of the work covered in this section can be found in Atkinson and Woods (2015).

Yang and Stufken (2009) proposed an alternative strategy for deriving locally optimal designs for GLMs with two parameters and a single explanatory variable. For a particular model and design space χ, it may be possible to identify a subclass of the designs, denoted as Ξ, such that for any design $\xi \notin \Xi$, there exists a design $\tilde{\xi} \in \Xi$ such that the information matrix for $\boldsymbol{\beta}$ under $\tilde{\xi}$ dominates the information matrix

under ξ in the Loewner ordering; $\mathbf{M}_{\tilde{\xi}}(\boldsymbol{\beta}) \geq \mathbf{M}_{\xi}(\boldsymbol{\beta})$. This implies that a search for optimal designs over Ξ will produce the same result as an extensive search over χ. Ξ is referred to as a complete class, and can be particularly useful if Ξ is considerably small and robust to values of $\boldsymbol{\beta}$. Stufken and Yang (2012) present an extension of this method for GLMs with multiple covariates.

Although analytical techniques has been used extensively throughout the literature to study locally optimal designs, optimality results are difficult to generalize to a broad class of model types, parameter ranges, and optimality criteria. Furthermore, they seem to be limited to studying designs with a small number of design variables for rudimentary models, as analytical calculations become intractable for more complicated design problems. From a practical standpoint, focusing on techniques for producing efficient, but perhaps not provably optimal, designs for a fixed size n that can accommodate more complicated models and a greater number of design variables may be more useful to experimenters.

3.5.2 Exact Design Algorithms

Many iterative algorithms used to construct exact designs implement an exchange scheme. An exchange algorithm was first proposed by Fedorov (1972), and early work in this area involved point-exchange algorithms. This methodology does not guarantee optimality, as all possible design points come from a discrete candidate set, but they may still perform sufficiently well to create efficient designs. The basic concept of a point-exchange algorithm is as follows: from an initial design with n runs, all pre-specified points in a candidate set (often some fine grid of points over the design region) are individually augmented and evaluated, and the point that yields the maximum improvement with respect to the optimality criterion of choice is added to the design. The least attractive point in the $n + 1$ run design is then

deleted, and exchanges continue until some convergence criteria are satisfied. Mitchell (1974) presents the DETMAX algorithm, which is a point-exchange algorithm used to construct D-optimal designs for a standard linear model. Instead of a one-for-one exchange between design points, DETMAX allows for the replacement of multiple points at each iteration, which reduces the probability of getting trapped at a local optima. A generalized form of the DETMAX algorithm for linear models is provided by Welch (1984), which is used to implement G- and I-optimal designs. There are also algorithms, such as the k-exchange algorithm (Johnson and Nachtsheim, 1983), where k points are added and deleted simultaneously. Additional point-exchange methods are summarized in Atkinson (1988).

A coordinate-exchange algorithm for linear models was first proposed by Meyer and Nachtsheim (1995), where each coordinate (column) of a design point (row) is exchanged or iterated until no meaningful improvements are realized. In addition to the significant reduction in computing time for large problems, the coordinate-exchange method does not require the enumeration of candidate sets. Mixed design spaces, where there are combinations of categorical and numeric factors, are also handled easily. The coordinate-exchange methodology has been used in a wide application of design problems, such as generating D-optimal split-plot designs (Jones and Goos, 2007), I-optimal split-plot designs (Jones and Goos, 2012), or designing D- and I-optimal blocked experiments with prior information on the blocks (Jones and Goos, 2015).

Algorithms for exact optimal designs for the logistic regression model are less prevalent in the literature, with the majority of work focused on the D-criterion. Woods *et al.* (2006) created exact D-optimal designs for a first-order logistic regression model with four design variables, using the centroid of the parameter ranges for several uniformly distributed parameter spaces. The authors noted that the locally

D-optimal designs created using the centroid values are fairly robust across the parameter spaces when the prior ranges are not too wide. Dror and Steinberg (2006) used a point-exchange algorithm for the construction of locally D-optimal designs for a GLM, where a sequential approach can be utilized to limit the candidate set for multivariate problems, which become prohibitively large as the number of design variables increase. A clustering methodology is used to create robust designs for a range of parameter values. Both of these applications illustrate that locally D-optimal designs for a multivariate logistic regression model can be created quickly using a point-exchange algorithm with minor modifications to include the weight matrix. The numerical integration technique presented by Gotwalt *et al.* (2009) in Section 3.4 was implemented for the Bayesian criterion in Equation 3.11 by the authors in a coordinate-exchange algorithm to quickly generate 16- and 48-run Bayesian designs for the same logistic regression model with superior D-efficiencies across the same parameter ranges examined in Woods *et al.* (2006).

An exchange algorithm for generating G-optimal designs is proposed by Saleh and Pan (2016), where a combination of point-exchange and coordinate-exchange methods, along with a clustering methodology for the candidate design points is used. The algorithm can be modified to accommodate GLMs for the creation of locally G-optimal exact designs. The authors also use a methodology that is conceptually similar to the approach used by Dror and Steinberg (2006) to create robust designs.

Several gaps in the literature are noteworthy in this area. First, exact I-optimal designs for GLMs have not been well-studied. Li and Deng (2018) established the GET of I-optimality for GLMs and propose a multiplicative algorithm to construct continuous I-optimal designs. However, no methodology has been proposed to create robust exact I-optimal designs for GLMs. Second, the separation problem has not been examined in multi-factor designs for the logistic regression model. Works such

as Atkinson and Woods (2015) have mentioned the non-existence of ML estimates for small designs involving categorical (binary) data, but did not explore how the problem diminishes with increased sample sizes or manifests with different parameter values over a range of forms for the linear predictor. There are several works that have explored the separation problem from a design perspective, but only for single-factor designs. Fornius (2008) investigated the separation problem in locally D-optimal designs for a single-factor quadratic logistic regression model, and Rahman (2015) proposed a compound optimality criterion that balances D-optimality with a reduced probability of separation. There have been no studies of the separation problem in multi-factor D-optimal designs, and no studies of the separation problem in designs using other optimality criteria for the logistic regression model. It has not been established whether a certain criterion will produce designs that are distinctly less prone to separation. Additionally, for designs that encounter separation, no augmentation strategy has been proposed to break separation in multi-factor, non-sequential designs. The absence of research in these areas justifies exploring different optimality criteria to see if one yields designs with an appreciable advantage over the rest in terms of separation. Developing an augmentation strategy focused on breaking separation as part of a sequential design approach would also be a worthwhile contribution to the literature.

Chapter 4

SEPARATION IN D-OPTIMAL DESIGNS FOR THE LOGISTIC REGRESSION MODEL

4.1 Introduction

The D-criterion is often used in computer-generated experimental designs when the response of interest is binary, such as when the attribute of interest can be categorized as pass or fail. The majority of methods in the generation of D-optimal designs focus on logistic regression as the base model for relating a set of experimental factors with a binary response. Despite the advances in computational algorithms for calculating D-optimal designs for the logistic regression model, very few have acknowledged the problem of separation, a phenomenon where the responses are perfectly separable by a hyperplane in the design space. Separation causes one or more parameters of the logistic regression model to be inestimable via ML estimation. The objective of the research presented in this chapter is to investigate the tendency of computer-generated, non-sequential D-optimal designs to yield separation in small-sample experimental data. Sets of locally D-optimal and Bayesian D-optimal designs with different run (sample) sizes are generated for three specified logistic regression models. A Monte Carlo simulation methodology is then used to estimate the probability of separation for each design. Results of the simulation study confirm that separation occurs frequently in small-sample data and that separation is more likely to occur when the model has interaction and quadratic terms. Finally, this research illustrates that different designs with identical run sizes created from the same model can have significantly different chances of encountering separation.

4.2 Separation in Single-Factor D-Optimal Designs

In the literature, only single-factor designs have been studied for the phenomenon of separation. Fornius (2008) explored the separation properties of several non-sequential single-factor locally D-optimal designs for the quadratic logistic regression model, which has the linear predictor form

$$\eta(x) = \alpha + \beta(x - \mu)^2 \tag{4.1}$$

Four sets of values for the parameters α, β, and μ were used to generate two D-optimal designs for each set. The D-optimal designs had either 3 or 4 support points, a term that is used to designate the unique locations of runs within the design space; all of the trials in an experiment will be conducted at a support point. The D-optimal designs were compared to two non-optimal designs with 7 and 8 support points, respectively. 10-, 20-, 50-, and 100-run design sizes are considered for each of the four parameter sets and design types. For the smaller sample sizes, all possible permutations of the response vector are considered, while 10000 simulated response sets were used to estimate the probabilities of separation occurring for the large sample sizes (Fornius, 2008).

There was a significant problem with separation for the 10- and 20-run sizes. Even for the largest sample size, the D-optimal design for one parameter set still had a 65 percent chance of encountering separation (Fornius, 2008). It was generally noted that while the D-Optimal designs outperformed the non-optimal designs in terms of the mean squared errors of the estimators, the 7- and 8-support point designs were less prone to separation (Fornius, 2008).

Rahman (2015) proposed a compound optimal design criterion in an attempt to minimize the probability of separation in non-sequential single-factor logistic designs while considering D-efficiency. Probability-optimal (P_s) designs were created

by allocating design points in a manner that minimized the probability of separation occurring. Although the P_s designs have a much lower probability of separation than D-optimal designs of the same size, the P_s designs do not appear to be useful in practice, as they have poor D-efficiencies and appear to be runs permuted about a single point of maximum variance, where $E(y) = 0.5$.

A compound optimality criterion, DP_s, was used to balance the D-criterion with a reduced probability of separation. It is expressed as a ratio of the D-criterion to the probability of complete separation occurring in the design, weighted by an arbitrary mixing constant $\alpha \in (0, 1)$. DP_s can be represented mathematically as shown in Equation 4.2,

$$DP_s = \frac{P(S|\mathbf{x})^{1-\alpha}}{|\mathbf{X}^T\mathbf{V}\mathbf{X}|^{\alpha/p}} \tag{4.2}$$

where $P(S|\mathbf{x})$ is the probability of complete separation with design points $\mathbf{x}$, $\alpha \in [0, 1]$ is a blending constant, and p is the number of parameters in the model (Rahman, 2015). DP_s will be minimized when the probability of separation is low and the D-efficiency of the design is high. However, no definitive conclusion was reached as to the best approach for determining α and generating designs with relatively high D-efficiencies that have comparatively low chances of separation occurring.

4.3 Simulation Study

Previous studies on the probability of separation in experimental designs for the logistic regression model have focused solely on single-factor designs. A simulation study was conducted to explore the probability of separation in multi-factor locally and Bayesian D-optimal designs for different realizations of the logistic regression model. In this study, only continuous two-factor designs were considered. Furthermore, all Bayesian designs were constructed with the assumption that the prior information on the model parameters can be adequately summarized by a normal dis-

tribution. The linear predictor of the logistic regression model used in this study has the form shown in Equation 4.3,

$$\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} {x_1}^2 + \beta_{22} {x_2}^2 \tag{4.3}$$

where x_1 and x_2 represent settings of the continuous design factors.

Three representative logistic regression models were specified for the simulation study: 1) a main-effects (ME) model ($\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$) with a relatively flat response surface, 2) a ME and interaction (2FI) model ($\eta(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2$) with a steeper surface and mild curvature, and 3) a quadratic model with more complex curvature. The $\pm 2\sigma$ limits on all parameter means for each model are shown in Table 4.1. The midpoints of the ranges shown in Table 4.1 were used as point estimates for $\boldsymbol{\beta}$ to create the locally D-optimal designs, while the distributions were used to create the Bayesian D-optimal designs. The surface plots of the logistic response function for each model, generated using the midpoints of the ranges given in Table 4.1, are shown in Figure 4.1. As exact designs are generated via a heuristic exchange algorithm, the design algorithm will produce multiple distinct designs that are declared locally or Bayesian D-optimal for each model. Therefore, to comprehensively examine the probability of separation for exact n-run designs generated from a particular model, 20 designs of size n were created for each model (ten locally D-optimal, ten Bayesian D-optimal), with $n = 8$, 16, 32, and 64 runs. This corresponds to a total of 80 unique designs for each model. The nonlinear design platform of JMP Pro 13®, which implements the quadrature scheme proposed by Gotwalt *et al.* (2009), was used to create all the designs.

A Monte Carlo simulation methodology was adopted to estimate the probability of separation for each design. For each design, 2500 response sets were simulated, where the response for each run was randomly sampled from a Bernoulli process with

Table 4.1: $\pm 2\sigma$ Limits on Parameter Means

ME Model	**2FI Model**	**Quadratic Model**
$\beta_0 = 0$	$-1.5 \leq \beta_0 \leq -0.5$	$1 \leq \beta_0 \leq 3$
$1 \leq \beta_1 \leq 3$	$2 \leq \beta_1 \leq 6$	$1.5 \leq \beta_1 \leq 4.5$
$0.5 \leq \beta_2 \leq 1.5$	$0.5 \leq \beta_2 \leq 1.5$	$-3 \leq \beta_2 \leq -1$
	$-3 \leq \beta_{12} \leq -1$	$-1.5 \leq \beta_{12} \leq -0.5$
		$1.5 \leq \beta_{11} \leq 4.5$
		$-6 \leq \beta_{22} \leq -2$

$P(y_i = 1|\mathbf{x}_i) = E(y_i|\mathbf{x}_i)$, where $E(y_i|\mathbf{x}_i)$ is defined by the logistic response function of the model being used. All model parameters were assumed to be correctly specified in design creation throughout the simulation study; the parameter values were equal to the means of their given prior distributions. For each simulated response set, JMP Pro 13® was used to fit a corresponding logistic regression model, and separation was declared in the response set if the estimated variance of any parameter estimate was greater than 5000, which is a relatively conservative suggestion for declaring separation proposed by Heinze and Schemper (2002). For each design, the proportion of simulated response sets that were observed is reported as the estimated probability of separation.

A validation of the simulation methodology was conducted for all designs of size $n = 8$. A logistic regression model was fit to each of the $2^8 = 256$ possible response vectors, $\mathbf{y} \in \{0, 1\}^8$. For each response vector, separation was declared when any of the parameter estimates had a variance over 5000. The probability of separation for the design was then calculated as the summation of the probability of each unique response vector occurring that yielded separation, shown in Equation 4.4.

$$P(\text{Separation}) = \sum_{j=1}^{2^n} \left(\prod_{i=1}^{n} P(y_{ij}) \right) I_j \tag{4.4}$$

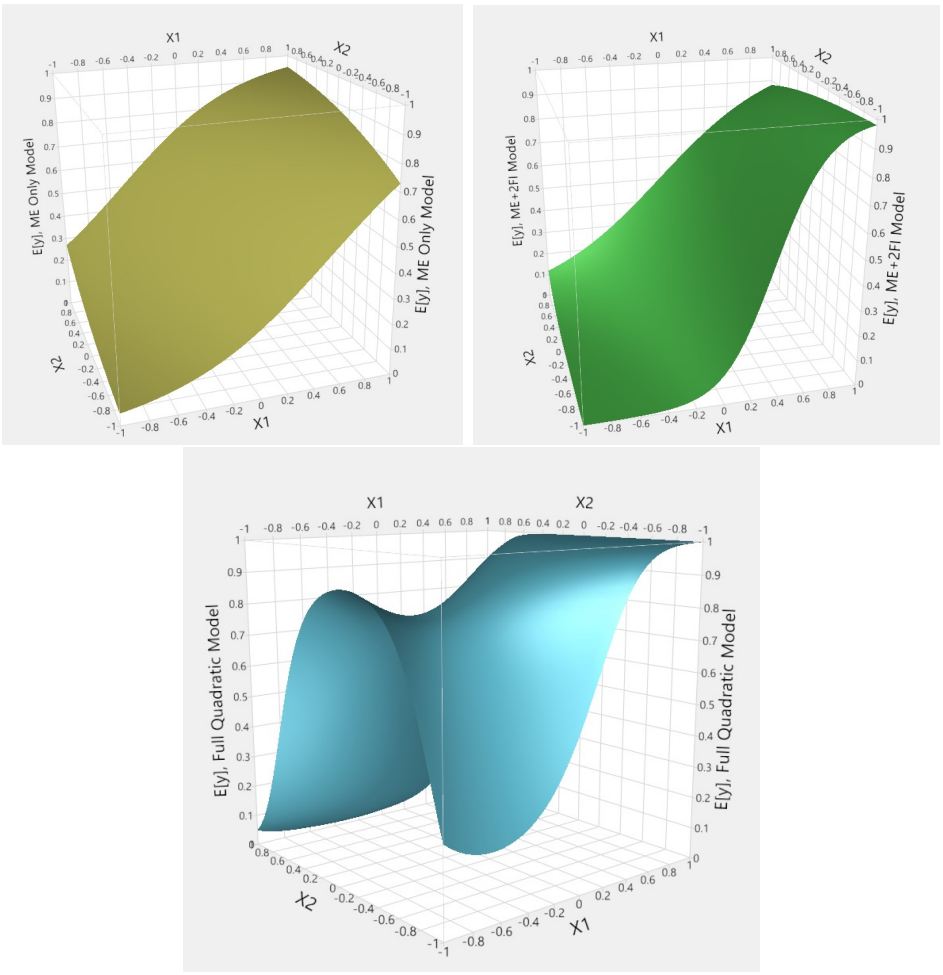


Figure 4.1: Surface Plots of the Logistic Regression Models

where

$$P(y_{ij}) = \begin{cases} \frac{1}{1+\exp(-\mathbf{x}_i^T\boldsymbol{\beta})} & \text{if} \quad y_{ij} = 1 \\ \frac{1}{1+\exp(\mathbf{x}_i^T\boldsymbol{\beta})} & \text{if} \quad y_{ij} = 0 \end{cases}$$

and

$$I_j = \begin{cases} 1 & \text{if separation is declared in response set } j \\ 0 & \text{otherwise} \end{cases}$$

As n increases, this validation methodology quickly becomes impractical, as the number of possible response vectors is 2^n. However, for designs with a small number of

runs, it is possible to estimate the separation probability of the design in this fashion. The results of the simulation study are summarized in Table 4.2. The validation results are also reported for all designs of size $n = 8$. By comparing the simulation results to the validation results, it appears that the simulation methodology produces reasonable estimates of a design's probability of separation.

Table 4.2: P(Separation) Results by Model

ME Model

	Local		**Bayesian**	
Design Runs	Average	Std Dev	Average	Std Dev
8	0.733	0.028	0.643	0.042
8 (val)	0.727	0.027	0.626	0.047
16	0.328	0.019	0.270	0.027
32	0.057	0.005	0.032	0.013
64	0.002	0.001	0.001	0.001

2FI Model

	Local		**Bayesian**	
Design Runs	Average	Std Dev	Average	Std Dev
8	0.993	0.004	0.965	0.014
8 (val)	0.993	0.003	0.961	0.014
16	0.917	0.006	0.814	0.052
32	0.619	0.007	0.468	0.117
64	0.171	0.009	0.067	0.032

Quadratic Model

	Local		**Bayesian**	
Design Runs	Average	Std Dev	Average	Std Dev
8	0.989	0.013	0.980	0.009
8 (val)	0.997	0.005	0.998	0.003
16	0.884	0.124	0.537	0.121
32	0.376	0.182	0.350	0.090
64	0.109	0.098	0.013	0.012

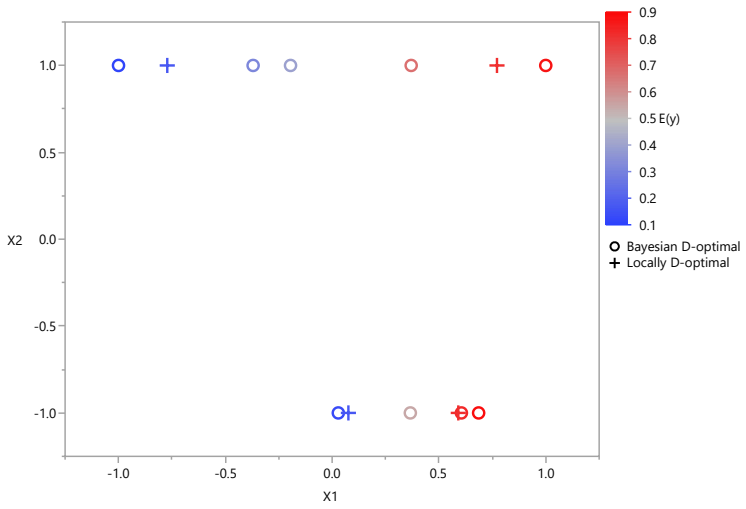
4.4 Discussion

All three representative logistic regression models examined in this study yield designs with substantial separation problems when the number of available runs is small. For 8- and 16-run designs, there were cases for the 2FI and quadratic models where the nonlinear design algorithm in JMP® created locally optimal designs that were guaranteed to either be inestimable ($\mathbf{y} = \mathbf{0}$ or $\mathbf{y} = \mathbf{1}$) or suffer from separation. A response vector that would allow for valid parameter estimates did not exist in certain small designs.

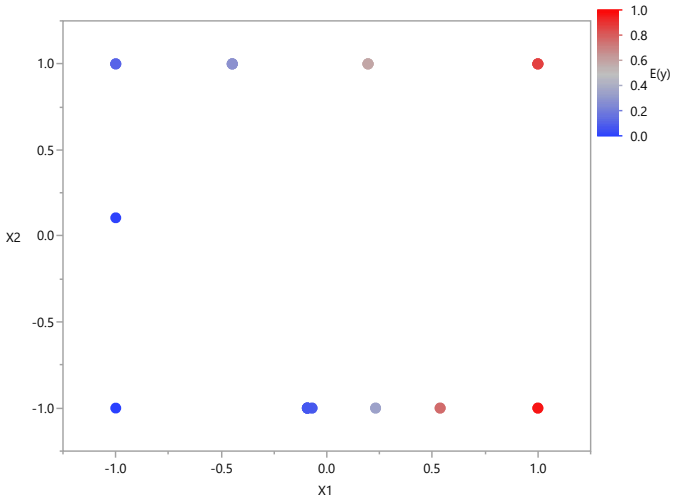
For equal design sizes, the designs constructed for the ME model, which has a comparatively flat response surface, do not have as high a probability of separation as do the designs constructed for the 2FI and the quadratic models. By the time the designs for the ME model reached a size of 32 runs, encountering separation is unlikely, as separated data occurs in only 5.7 percent of the simulated response sets. However, for designs for the 2FI and the quadratic models, this is not the case. Even with a design size of 64 runs, separation occurred over 30 percent of the time in one locally optimal design for the quadratic model.

Furthermore, the results of the simulation study indicate that the Bayesian D-optimal designs tend to perform better than the locally D-optimal designs in terms of the probability of separation. For all three models, the average probability of separation is lower for the Bayesian designs than for the locally optimal designs across all run sizes examined. Consider two 32-run designs created for the 2FI model: a locally optimal design with an estimated separation probability of 0.623, and a Bayesian design with an estimated separation probability of 0.329. The support points of each design are shown in part (a) of Figure 4.2.

The Bayesian designs have more support points than the locally optimal designs,



(a) Support Points for a Locally D-Optimal and a Bayesian D-Optimal Design



(b) Support Points for a Bayesian D-Optimal Design with Higher Variance Priors

Figure 4.2: Support Points for 2FI Model Designs

especially as the variances of the parameter prior distributions increase. As seen in part (a) of Figure 4.2, the locally optimal design has four support points, while the Bayesian design has nine. As the variance increases, the support points also tend to

cover a larger area of the design region. Chaloner and Larntz (1989) show that as the uncertainty in the prior information increases, so do the number of support points in a design. Suppose that the $\pm 2\sigma$ range on the priors for the 2FI model are doubled, as shown in (4.5).

$$\begin{aligned} -2 &\leq \beta_0 \leq 0 \\ 0 &\leq \beta_1 \leq 8 \\ 0 &\leq \beta_2 \leq 2 \\ -4 &\leq \beta_{12} \leq 0 \end{aligned} \tag{4.5}$$

The support points for a 32-run Bayesian design generated from the higher-variance prior distributions in (4.5) for the 2FI model are shown in part (b) of Figure 4.2. The design has eleven support points that cover a larger portion of the design space. Interestingly, this design has an estimated separation probability of 0.468, which is higher than the 9-support point Bayesian design generated from the original priors shown in Table 4.1.

An important observation should be made on the impact of the number of support points on the probability of separation in a design. Support points are collectively referred to as near neighbors if they occupy nearly identical spaces in the design region. In this study, as both design variables have ranges $(-1, 1)$, support points are considered near neighbors if both their x_1 and x_2 coordinates differ by less than 1×10^{-3}. If either coordinate differs by more than 1×10^{-3}, they are referred to as distinct support points. Consider the 64-run designs created for the 2FI model. The 64-run locally optimal designs have between 4 and 7 total support points, with separation probabilities ranging from 0.157 to 0.188. The design with the highest probability of separation in this group has 7 support points. However, each locally

optimal design in this group has 4 distinct support points, indicating that the number of distinct support points in a design has a larger effect on the probability of separation than the total number of support points. After grouping near neighbor support points, all of the locally optimal designs created for the 2FI model have 4 distinct support points across all run sizes. This explains the low variance of separation probabilities for the locally optimal designs from the 2FI model, seen in the standard deviations reported for these groups in Table 4.2. The 64-run Bayesian designs created for the 2FI model have between 8 and 10 distinct support points, with separation probabilities ranging from 0.029 to 0.119, which is lower than the locally optimal designs. The Bayesian design with the lowest probability of separation in this group has 10 distinct support points. By comparing the distinct support points of the locally optimal designs to the Bayesian designs across all three models, it is evident that as the number of distinct support points increase, the probability of separation tends to decrease, as the existence of a separating hyperplane becomes less likely with more distinct support points in the design region. This appears to be the reason that the Bayesian D-optimal designs are less prone to separation than the locally D-optimal designs.

As the response surfaces become more complex with the addition of interaction and quadratic terms, the variance of the estimated probabilities of separation for designs generated from the same model tends to increase. The separation probability disparity between designs is especially noticeable in the 16- and 32-run Bayesian designs from the 2FI model and all 16- and 32-run designs from the quadratic model, as seen in the standard deviations reported for these groups in Table 4.2. The large variances of the probability of separation in these groups cannot be solely attributed to differences in the number of distinct support points for each design. This can be seen in Figure 4.3, which shows the probability of separation for each of the 16-run designs

from the quadratic model grouped by the number of distinct support points. Figure

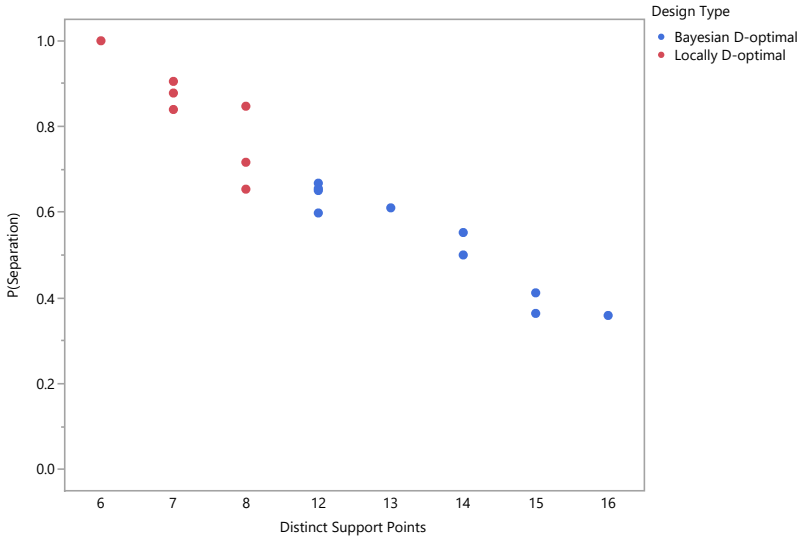
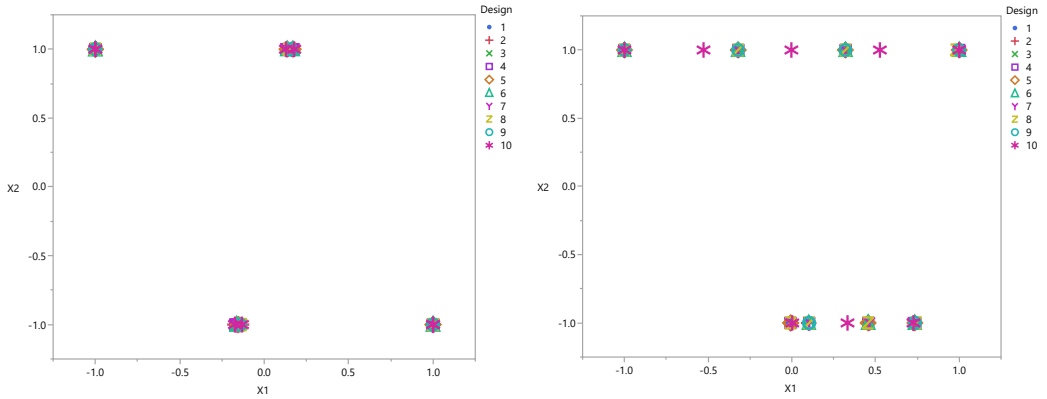


Figure 4.3: P(Separation) vs. Distinct Support Points for the 16-Run Designs from the Quadratic Model

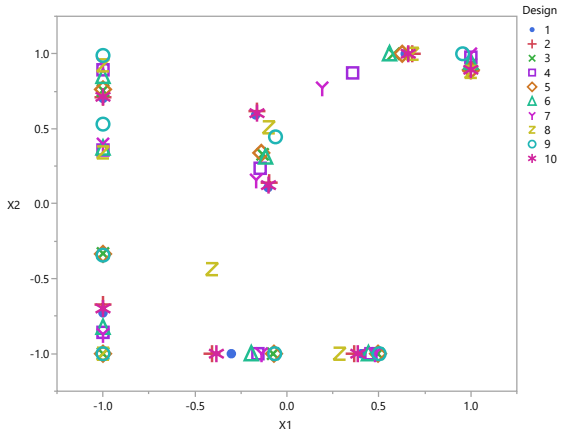
4.3 supports the conjecture that the Bayesian designs are less prone to separation due to the higher number of distinct support points, as the probability of separation is generally decreasing as the number of distinct support points increases. However, there are two 8-support point locally optimal designs with a nearly 20 percent difference in the estimated probability of separation. Differences in separation probabilities of this magnitude were observed for the 2FI and quadratic models between designs with an equal number of runs and distinct support points.

As the models become increasingly complex, the design runs collectively occupy a larger portion of the design region. This can be observed in Figure 4.4, which shows the support points for all of the 8-run Bayesian D-optimal designs across all three models. Figure 4.4 shows that the JMP® nonlinear design algorithm is always allocating runs to essentially the same four locations for the 8-run designs generated from the ME model. Since each design is nearly identical in terms of the support points,



(a) ME Model Designs

(b) 2FI Model Designs



(c) Quadratic Model Designs

Figure 4.4: Support Points of the 8-Run Bayesian D-Optimal Designs

the separation probabilities between designs are very close, as expected. However, the designs from the 2FI model have a more pronounced difference in the support point locations between designs. For the quadratic model, Figure 4.4 shows that the design algorithm is allocating runs to a wide range of support point locations throughout the design region between different designs, leading to the most disparate designs observed from the same model, even if the number of distinct support points are equivalent. This explains why designs with the same number of runs and distinct support points generated from the same model can still have significantly different

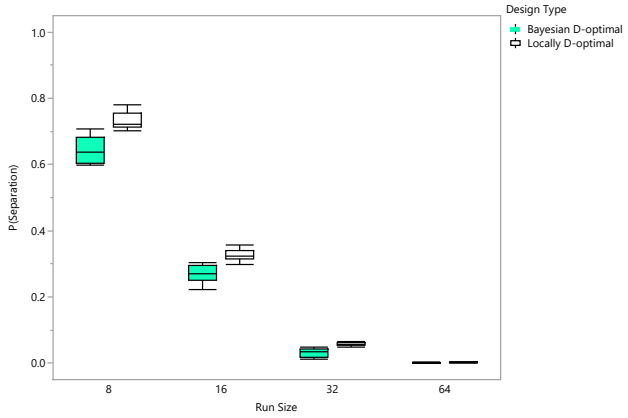
probabilities of separation.

4.5 Summary

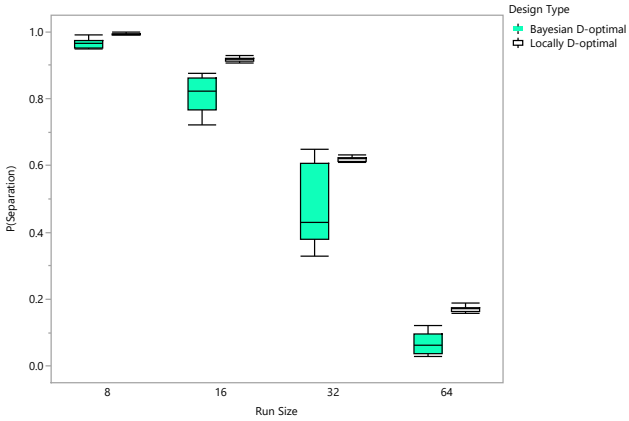
The simulation study shows that separation is a serious problem for both Bayesian and locally D-optimal two-factor designs when the run size is small. The separation problem becomes more pronounced as the models acquire active interaction and quadratic terms. For 8- and 16-run designs, encountering separation in practice would likely be a common occurrence. The problem of separation is mostly mitigated by the time the design size reaches 64 runs, although there are still 64-run locally optimal designs where separation occurred in over 30 percent of the simulated data sets.

Figure 4.5 captures the probability of separation for all of the designs produced in this study. The Bayesian designs tend to be less prone to separation than locally optimal designs of identical run size, likely due to the increased number of distinct support points. The increase in support points is especially noticeable as the models become more complex and the uncertainty in the prior information increases. However, as the example of increasing the prior variance in Section 4.4 shows, the spread of support points throughout more of the design region does not necessarily yield a design with a lower probability of separation.

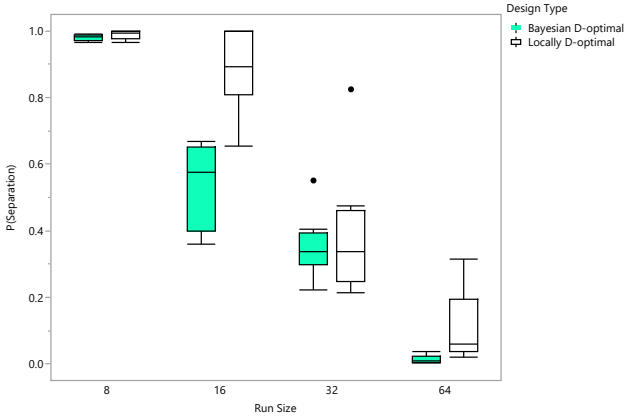
The 16- and 32-run Bayesian designs generated from the 2FI model, and both the locally optimal and Bayesian designs generated from the quadratic model illustrate that designs generated from the same model can have significantly different probabilities of separation. This is represented in Figure 4.5, where the boxes are quite elongated for the 2FI and the quadratic models, indicating a high variance in separation probabilities between designs of identical run sizes from the same model. In practice, generating several designs and examining the separation properties would be advisable, as the user should select the design with the lowest estimated probability



(a) ME Model Designs



(b) 2FI Model Designs



(c) Quadratic Model Designs

Figure 4.5: P(Separation) vs. Run Size by Model

of separation, assuming similar D-efficiencies between designs.

The simulation study presented in this chapter is a first attempt to quantify the probability of separation in non-sequential, multi-factor D-optimal designs. Future research could involve increasing the number of design factors and exploring designs generated from different prior distributions. However, an efficient design construction method for non-normal prior distributions has not been established. Future research could also involve expanding the scope of the logistic regression models that were used in this study, or designing a study that explores a wider range of prior variances. There may be model forms that yield designs with more significant separation issues.

Chapter 5

NON-SEQUENTIAL AUGMENTATION STRATEGIES TO ADDRESS SEPARATION IN LOGISTIC REGRESSION

Previous research on small sample multi-factor D-optimal designs for the logistic regression model has demonstrated that these designs are prone to encountering separation, a phenomenon where the responses are completely or quasi-completely separable by a hyperplane in the design space. Separation causes the non-existence of maximum likelihood parameter estimates and represents a serious problem for model fitting purposes. In this chapter, several non-sequential design augmentation strategies, where additional experimental trials are performed following an initial experiment that has encountered separation, are investigated. Small locally and Bayesian D-optimal initial designs are generated for several representative logistic regression models, and a Monte Carlo simulation methodology is then used to investigate the effectiveness of each augmentation strategy in eliminating separation. Results of the simulation study show that augmenting design runs (trials) in regions of maximum prediction variance (MPV) is the most effective strategy for eliminating separation. However, MPV augmentation tends to produce designs with lower D-efficiencies. The study illustrates that MPV augmentation reliably eliminates separation and can be used in practice to obtain usable parameter estimates for the logistic regression model.

5.1 Introduction

Experimental designs for categorical responses are often constructed using optimal design methods for generalized linear models (GLMs), a flexible class of models for responses belonging to the exponential family of distributions (Myers *et al.*, 2012).

Examples of responses that can be modeled using GLMs include variables that are measured in the binary (pass/fail), nominal, and ordinal scales. For *exact* (discrete) experimental designs for GLMs (see Atkinson *et al.* (2007) for the distinction between exact and continuous optimal designs), the majority of research has focused on developing efficient exchange algorithms to create designs for various optimality criteria, such as novel numerical integration techniques to expediently generate Bayesian D-optimal designs, or modifications to coordinate exchange algorithms to create robust G-optimal designs (Gotwalt *et al.*, 2009; Saleh and Pan, 2016).

Designs for binary response data, where the outcome of each experimental trial can be classified as one of two possible outcomes, typically use the logistic regression model as the base model for relating the response to a set of experimental factors. The logistic regression model, which belongs to the class of GLMs, has the form shown in Equation 2.3, where the outcome of n experimental trials (runs) are modeled as independent Bernoulli random variables. The logistic response function, $E(y)$, has the form shown in Equation 2.4, where $\mathbf{x}$ represents a vector of design variables expanded to model form and $\boldsymbol{\beta}$ represents a vector of model coefficients (parameters). The linear predictor of the logistic response function is shown in Equation 2.5.

Optimal design methods select design points such that a user-specified statistical criterion is optimized. One of the most widely used is the D-criterion, which aims to minimize the generalized variance of the parameter estimates, yielding designs tailored for model-building and screening applications. This is accomplished by constructing a design that maximizes the determinant of the Fisher information matrix, $\mathbf{M}$. For GLMs, a D-optimal design satisfies

$$\xi_D^* = \underset{\xi}{\operatorname{argmax}} |\mathbf{M}(\xi)| = \underset{\xi}{\operatorname{argmax}} |\mathbf{X}^T \mathbf{V} \mathbf{X}|, \tag{5.1}$$

where ξ is the set of all designs in the factorial space, $\mathbf{X}$ is the model matrix of a

candidate design in ξ, and $\mathbf{V}$ is a diagonal weight matrix containing the estimated variance of each observation on the diagonal Myers *et al.* (2012); Atkinson *et al.* (2007). A design given by Equation 5.1 minimizes $-\log|\mathbf{M}^{-1}|$, thus minimizing the eigenvalues of the asymptotic covariance matrix of $\boldsymbol{\beta}$ (Atkinson *et al.*, 2007). In contrast to methods for constructing optimal designs for linear models with constant error variance, optimal designs for GLMs are encumbered by their dependence on unknown model parameters, which is called the *design dependence* problem (Khuri *et al.*, 2006). Several methods have been presented to handle design dependence in GLMs. One method is to specify the best guess for the parameter vector $\boldsymbol{\beta}$, which can be derived from historical data, subject matter experts, or earlier experiments of a similar nature. Design points are then chosen with the assumed parameter vector. This method yields locally optimal designs, as these designs are only optimal for a specific estimate of $\boldsymbol{\beta}$ (Chernoff, 1953). To circumvent difficulties and issues with locally optimal designs, robust design approaches, such as Bayesian optimal designs, have been proposed. Bayesian optimal designs require the specification of a prior distribution for $\boldsymbol{\beta}$, denoted as $f(\boldsymbol{\beta})$, to model the uncertainty in the parameter values and to create designs that are more robust to parameter misspecification (Chaloner and Larntz, 1989).

Some results in the literature show that the design points of the D-optimal designs for the logistic regression model tend to be located on the boundaries of an induced subspace within the original design space; that is, depending on the specified parameters or prior distributions, the D-criterion favors locating design points on the boundaries of a region where $0.15 \leq E(y) \leq 0.85$ (Dror and Steinberg, 2005; Mancenido *et al.*, 2019). In contrast, design points of D-optimal designs for linear models with constant error variance are generally located on the boundaries of the original design space (Montgomery, 2017). Despite allocating design points on regions

where $E(y)$ should yield a sufficient proportion of both outcomes, D-optimal designs for logistic regression have been shown to be prone to the problem of *separation* in small, multi-factor experiments (Park *et al.*, 2019).

Separation is a phenomenon where responses are completely or quasi-completely separated by a hyperplane in the design space. Formally, if there exists a column vector α such that $\alpha^T \mathbf{x}_i \geq 0$ when $y_i = 1$ and $\alpha^T \mathbf{x}_i < 0$ when $y_i = 0$ holds for all observations, then some form of separation is present (Albert and Anderson, 1984). The presence of separation causes the non-existence of one or more maximum likelihood (ML) estimates for the parameters of the logistic regression model and presents a serious problem for users seeking to fit a statistical model for effect testing or inference purposes. From the perspective of analyzing experimental data, Firth likelihood (FL) estimation (Firth, 1993) was proposed to obtain finite parameter estimates for the logistic regression model with separated response data (Heinze and Schemper, 2002; Heinze, 2006). Despite the presence of recommendations for mitigating separation in the analysis of experimental data, research on addressing the problem of separation in the design phase of an experiment remains sparse.

The DP_s-criterion has been proposed to mitigate separation in the design phase for single-factor experiments for the logistic regression model (Rahman, 2015). DP_s-optimal designs use a compound design criterion that simultaneously optimizes the D-criterion and minimizes the probability of separation. These designs are more robust to separation problems for small run sizes and concurrently, produce minimum-variance estimates. However, the DP_s criterion is computationally expensive to implement for multi-factor experiments because it exhaustively enumerates all possible outcome combinations in the calculation of the probability of separation.

In this chapter, mitigating separation through a non-sequential augmentation strategy is explored, where a block of additional runs are added to the initial design to

break or remove separation. *Non-sequential augmentation* refers to situations where the entire initial experiment is planned before conducting any trials, and subsequent experimental trials (augmentation) are based only on the results of the initial experiment. On the other hand, *sequential augmentation* involves dynamically determining the factor settings of the next experimental trial based on the results observed in the current trial (Blot and Meeter, 1973). A non-sequential design strategy is typically adopted in cases where budget constraints or limited test execution windows prevent a sequential approach. For example, in the military acquisition cycle, operational testing for weapon system procurement is usually performed using a non-sequential approach to satisfy rigid test plan requirements (Giadrosich, 1995).

Although additional experimental trials are not always feasible in practice due to budget or time constraints, an augmentation strategy could still be used in settings where a fixed design size has been initially determined and could provide advantages over the use of a compound criterion. Suppose an organization has set a budget for 20 experimental runs. Instead of creating a 20-run DP_s-optimal design, an initial experiment for 15 runs can be constructed using the D-criterion. If separation occurs after executing the 15-run design, the remaining five runs can be augmented so that there is a high probability of breaking separation; if separation does not occur, the remaining five runs can be augmented using the D-criterion, which could produce a design with more precise parameter estimates than the DP_s-optimal design. Such flexibility is not offered by a one-shot experimental design strategy.

This study proposes three strategies for breaking separation in D-optimal designs for the logistic regression model using a non-sequential augmentation approach. These strategies are evaluated and compared with respect to two metrics that measure how well separation is mitigated following augmentation. The three strategies and performance metrics are discussed in Section 5.2, while the rest of the chapter is

organized as follows: Section 5.3 discusses the methodology for the Monte Carlo simulation study and demonstrates the resulting designs after augmentation; Section 5.4 summarizes and compares the performance of the augmented designs based on the estimated separation probabilities and the ML/FL estimates obtained; lastly, Section 5.5 provides recommendations for mitigating separation and potential directions for future work. While the methodology in this chapter focuses on the two-factor design case, results from this work can be extended to cases with more than two factors.

5.2 Augmentation Strategies

Several methodologies have been proposed to efficiently create *overlap*, a term used to denote the absence of complete or quasi-separation, in an initial separated data set (Neyer, 1994; Wu and Tian, 2014). The majority of work on creating overlap utilizes a sequential approach i.e., the augmented design point is dynamically determined from the results of the previous trial. For example, a three-phase procedure is presented in the literature to design optimal sensitivity experiments, where the first phase focuses on creating overlap by honing in on areas likely to produce mixed responses (Wu and Tian, 2014). The location of subsequent design points are dependent on the outcome of the previous trial, indicating a sequential approach. Proposals for sequential methods in the literature have also been limited to the single-factor case. For a non-sequential approach in multi-factor experiments, the following augmentation strategies are proposed:

1. *Maximum prediction variance* (MPV). Runs are augmented in regions of high prediction variance where $E(y_i|\boldsymbol{\beta}) \approx 0.5$.

2. *Support Point* (SP) *Replication*. Runs are augmented by replicating at the support points of the initial design. *Support points* are unique runs or factor

combinations that may or may not have been replicated in the initial design.

3. *Random.* Runs are augmented randomly throughout the design region, where the coordinates (factor levels) of each augmented design point are drawn from a Uniform distribution across the factor ranges.

The MPV strategy involves placing augmented design points in regions where $E(y_i|\boldsymbol{\beta}) \approx 0.5$ to maximize the chances of observing mixed responses to create overlap in the response data. For k total design variables and the estimates of $\boldsymbol{\beta}$ specified in design creation, $E(y|\hat{\boldsymbol{\beta}}) = 0.5$ can be expressed as a function that returns the coordinate of the kth design variable based on the coordinates of the remaining $k-1$ design variables. The factor settings of $k-1$ variables are randomly drawn from a Uniform distribution across the factor ranges. In coded units, where $x_i \in [-1, 1]$ $\forall i = 1, ..., k$, $x_i = \text{Uniform}(-1, 1)$ for $i = 1, ..., k-1$. The setting of the kth variable is determined by the function $E(y_i|x_1, ..., x_{k-1}, \hat{\boldsymbol{\beta}}) = 0.5$. The constraints $-1 \leq x_i \leq 1$ $\forall i = 1, ..., k$ are imposed to yield feasible points of MPV within the coded factorial design region. To avoid quasi-complete separation, at least one coordinate of the k design variables for each augmented point must be adjusted by a small δ. To illustrate, consider a two-factor ($k = 2$) main effects (ME) logistic regression model, with coded design variables $x_1, x_2 \in [-1, 1]$ and parameters $\beta_0 = 0$, $\beta_1 = 2$, and $\beta_2 = 1$. The logistic response function takes the form

$$E(y|\boldsymbol{\beta}) = (1 + \exp(-2x_1 - x_2))^{-1}. \tag{5.2}$$

From Equation 5.2, the region where $E(y|\boldsymbol{\beta}) = 0.5$ is a line defined by $x_2 = -2x_1$. An 8-run locally D-optimal design generated using JMP Pro 14.2® for this model consists of two replicates at each of the support points shown in Figure 5.1. Suppose that after executing the initial design depicted in Figure 5.1, the response data is

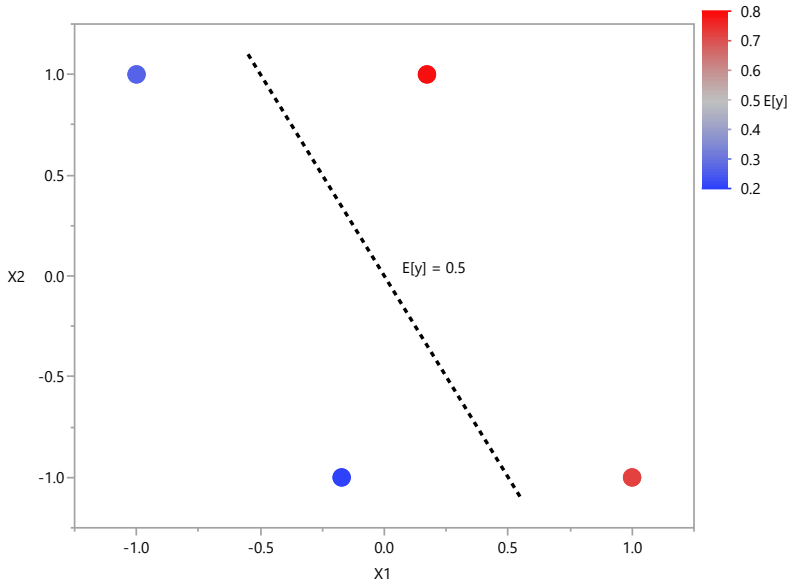


Figure 5.1: Locally D-Optimal Design with Separating Hyperplane

completely separated such that $y_i = 0$ is observed for all runs where $E(y) < 0.5$ and $y_i = 1$ is observed where $E(y) > 0.5$. In this case, any number of augmented runs that lie directly on the line $x_2 = -2x_1$ will yield, by definition, quasi-complete separation, regardless of the response values observed (Albert and Anderson, 1984). Therefore, at least one coordinate of each augmented point should be adjusted by some δ to avoid quasi-complete separation.

The proposed MPV method for breaking separation adds J augmented runs to the k-dimensional factorial coded design region, denoted as $\chi \in [-1, 1]^k$, in the following manner:

1. Define the set of candidate points, $C = \{\mathbf{x} \in \chi \mid E(y|\hat{\boldsymbol{\beta}}) = 0.5\}$. To define C, estimates for $\boldsymbol{\beta}$ are required. In practice, two methods are available when separation is present. C can be defined by the initial estimates used in design creation, such as point estimates derived from similar experimentation/the best guesses of subject matter experts, or the means of a distributional prior. FL

estimation can also be used to obtain $\hat{\boldsymbol{\beta}}$. For the simulation study presented in Section 5.3, the prior means specified in design creation are used to define C.

2. Randomly select an augmented design point $\mathbf{x}_i \in C$, for $i = 1, ..., J$, where J is the number of augmented design points.

3. Make an adjustment to at least one of the k coordinates of each $\mathbf{x}_i$ by $\pm\delta$, where $\delta \sim \text{Uniform}(a, b)$ and $a > 0$. The value of b is set to ensure that the augmented points are still in regions of high prediction variance.

4. Ensure that adjusted $\mathbf{x}_i \in \chi$ for $i = 1, ..., J$.

Strategy 2 augments runs at the support points of the initial D-optimal design. As the D-criterion will tend to locate support points in an induced sub-region of the design space with moderate response probabilities, it is reasonable to investigate whether mixed responses will be reliably observed at these support points to create overlap.

To provide baselines for the performance of MPV augmentation (strategy 1) and SP replication (strategy 2), the performance of random augmentation – where design runs are randomly placed throughout the design region (strategy 3) – is also investigated. The motivation behind this strategy is similar to the aim of space-filling computer designs, where the objective is to spread design points uniformly throughout the design region (Pronzato and Müller, 2012). This method is likely to place points in regions of moderate response probabilities, but may not have good practical properties, as each augmented run will be located in a unique region of the design space. Furthermore, the ability of random augmentation to create overlap will likely be inconsistent, as there is no system for locating the points.

Two performance metrics will be used to evaluate the three strategies: (1) the ability to reliably create overlapped response data; and (2) the standard error of estimates

that are obtained through the augmented design. The first metric is straightforward, as the primary aim of augmentation is to create overlap, which will allow for the use of ML estimation (Albert and Anderson, 1984). The second metric is a common method for evaluating the quality of an experimental design (Montgomery, 2017). The magnitude of the maximum ML estimate standard error reported by a statistical analysis package is also commonly used as a criterion for declaring separation, as enormous standard errors imply non-existence of finite ML estimates (Heinze and Schemper, 2002).

5.3 Simulation Study

A simulation study was conducted to evaluate the performance of the strategies outlined in Section 5.2. The simulation study was restricted to two-factor designs, but the methodology can be easily extended to higher-dimensional design spaces.

Three logistic regression models were used to generate the 8- and 16-run D-optimal designs used in this study: (1) a ME only model, (2) a ME with an active interaction term (2FI) model, and (3) a full quadratic model with an active interaction and second-order terms. It was assumed that the parameter prior distributions are normal; the $\pm 2\sigma$ limits on all parameter means for each model are shown in Table 4.1. Both locally and Bayesian D-optimal designs were used as the initial designs in this study. The means of the prior distributions (the midpoints of the ranges shown in Table 4.1) were used as the point estimates for $\boldsymbol{\beta}$ to generate the locally D-optimal designs, while the distributions were used to create the Bayesian D-optimal designs. Additionally, to define the candidate set for MPV augmentation, values of $\hat{\boldsymbol{\beta}}$ were assumed to be the single-point priors for the local designs and the means of the prior distributions for the Bayesian optimal designs.

To assess each augmentation strategy across designs generated from all three lo-

gistic regression models, 8-run initial designs were constructed (n = 8). For each model, ten locally D-optimal and ten Bayesian D-optimal designs were created in JMP Pro 14.2®. Due to the information-poor nature of binary response data, meaningful improvements to parameter estimates in small designs are not expected with a single, additional run. Therefore, runs were augmented to each design in a block of size J = 8, where each block was constructed using one of the three strategies. This resource allocation strategy is analogous to starting with a fractional factorial design and augmenting with a full fold-over for linear models with constant error variance, where another experiment of equal size is performed to eliminate ambiguity in a design's alias structure (Montgomery, 2017). The 16-run augmented designs (eight initial, eight augmented) are also compared to baseline 16-run D-optimal designs. This comparison indicates whether the augmented designs are more robust to separation than a D-optimal design. Additionally, the ML/FL estimates produced by the augmented designs are compared to FL estimates that are available in an equal size D-optimal design with separated response data to determine if an alternative augmentation strategy to the D-criterion is advisable.

A Monte Carlo simulation methodology was used to estimate the probability that each strategy will create overlapped response data from an initial 8-run D-optimal design. For each initial design, 2500 augmentation trials were simulated. Each trial proceeded as follows:

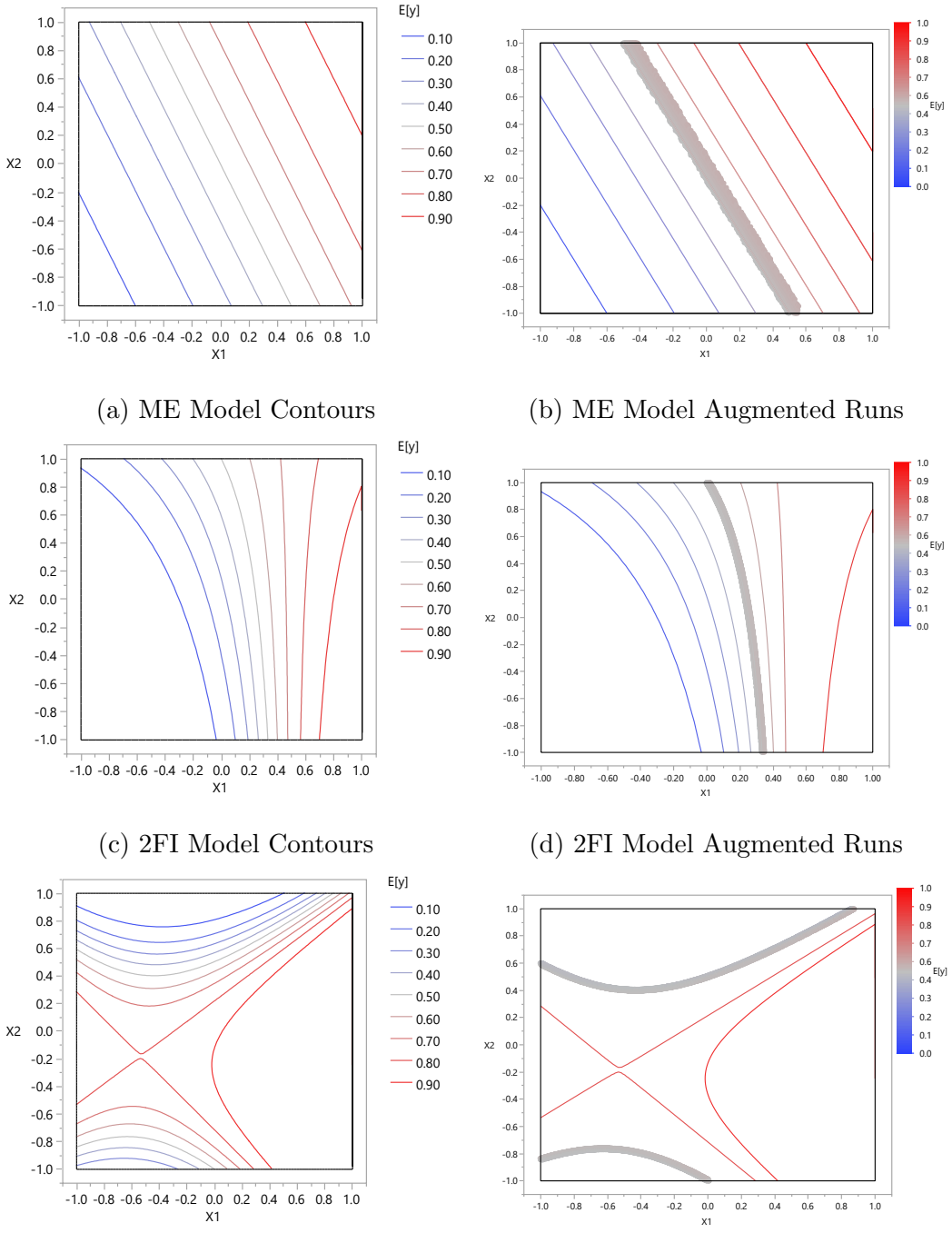
1. A separated response set for the initial design was simulated using a random Bernoulli draw with $P(y_i = 1) = E(y_i)$. A simulated response set was deemed "separated" if the estimated variance of any ML parameter estimate was greater than 5000, in accordance with previous methods in the literature (Heinze and Schemper, 2002). Response sets would be re-drawn until a separated response set was obtained.

2. The 8-run block of augmented runs was constructed using one of the augmentation strategies outlined in Section 5.2. Simulated responses for the augmented runs were randomly drawn using the same Bernoulli process.

3. From the augmented 16-run design, ML and FL estimates were determined. Separation was declared in the response data of the augmented design using the same large parameter estimate variance criterion.

Each of the 2500 augmentation trials for an initial design corresponded to a unique experimental design. The augmentation trials were performed for all three augmentation strategies outlined in Section 5.2.

The contour plots, along with 20,000 augmented runs using MPV augmentation for each model, are shown in Figure 5.2. The gray areas in the augmented run plots represent the collective plot of 20,000 design points that lie in regions of the design space where $0.46 < E(y) < 0.54$ holds. Separation results for each of the strategies are illustrated in Figure 5.3. This plot shows the estimated separation probability of the initial 16-run D-optimal designs and of the 16-run augmented designs. The separation probabilities were estimated as outlined in Park *et al.* (2019), where the reported estimate is the proportion of separated response sets that were observed out of the 2500 trials. For each simulated response set, overlap was declared if the maximum standard error of any ML parameter estimate was less than $\sqrt{5000}$ (Heinze and Schemper, 2002). Table 5.1 shows the percentage of overlapped response data for the designs constructed using the three augmentation strategies, as well as for the 16-run D-optimal designs.



(a) ME Model Contours

(b) ME Model Augmented Runs

(c) 2FI Model Contours

(d) 2FI Model Augmented Runs

(e) Quadratic Model Contours

(f) Quadratic Model Augmented Runs

Figure 5.2: Contour Plots and MPV-Augmented Runs Overlay by Model

Table 5.1: Percentage of Overlapped Response Data

	S1: MPV	**S2: Replicate SPs**	**S3: Random**	**16R *D*-Optimal**
ME	92.47%	50.90%	89.37%	70.09%
2FI	72.28%	10.59%	61.57%	13.43%
Quadratic	69.38%	10.65%	63.64%	28.95%

5.4 Discussion

Results from the simulation study show that the best strategy for breaking separation is MPV augmentation. For the three sample logistic regression models, the augmented designs created using MPV augmentation tend to have lower probabilities of separation, as shown in Figure 5.3. Statistical significance in the difference between augmentation strategies was established by fitting the log-transformed separation probabilities to a two-factor linear model with factors *augmentation strategy* (categorical, four levels: 16-run D-optimal, MPV, SP replication, and random) and *initial design type* (categorical, two levels: locally D-optimal and Bayesian D-optimal). The coefficient estimates and confidence intervals are shown in Table 5.2.

In Table 5.2, the 16-run D-optimal design level of the augmentation strategy factor is used as the base case for the effect-coded indicator variables. The results show that for all models, MPV augmentation produced designs that are most robust to separation. As the response variable of the fitted model is a negative log transformation of the separation probabilities, larger coefficient estimates for the transformed model indicate lower probabilities of separation for designs produced from the corresponding augmentation strategy. Table 5.2 also shows that the augmentation strategies produced lower probabilities of separation with initial Bayesian D-optimal designs. It

Table 5.2: Model Fit of $-\ln(P(\text{Separation}))$ to Augmentation Strategy and Design Type

ME Model	Estimate	Std Error	p-value	Lower 95%	Upper 95%
Intercept	1.69292	0.01304	<.0001	1.66694	1.71890
Design Type [Bayesian]	0.01874	0.01304	0.15490	-0.00724	0.04472
Augmentation Strategy [MPV]	0.90533	0.02259	<.0001	0.86033	0.95032
Augmentation Strategy [Replicate SPs]	-0.97667	0.02259	<.0001	-1.02167	-0.93168
Augmentation Strategy [Random]	0.54953	0.02259	<.0001	0.50454	0.59453
2FI Model	Estimate	Std Error	p-value	Lower 95%	Upper 95%
Intercept	0.62881	0.00626	<.0001	0.61635	0.64127
Design Type [Bayesian]	0.08425	0.00626	<.0001	0.07179	0.09671
Augmentation Strategy [MPV]	0.66272	0.01083	<.0001	0.64113	0.68430
Augmentation Strategy [Replicate SPs]	-0.51519	0.01083	<.0001	-0.53678	-0.49361
Augmentation Strategy [Random]	0.33438	0.01083	<.0001	0.31280	0.35597
Quadratic Model	Estimate	Std Error	p-value	Lower 95%	Upper 95%
Intercept	0.67845	0.01946	<.0001	0.63968	0.71723
Design Type [Bayesian]	0.07227	0.01946	0.00040	0.03349	0.11104
Augmentation Strategy [MPV]	0.51036	0.03371	<.0001	0.44321	0.57752
Augmentation Strategy [Replicate SPs]	-0.56252	0.03371	<.0001	-0.62968	-0.49537
Augmentation Strategy [Random]	0.34056	0.03371	<.0001	0.27340	0.40771

has been demonstrated that Bayesian D-optimal designs tend to have more support points than locally optimal designs, and as a result are more robust to separation (Park *et al.*, 2019). The results show that this initial reduction in separation probability is propagated across the augmented Bayesian designs.

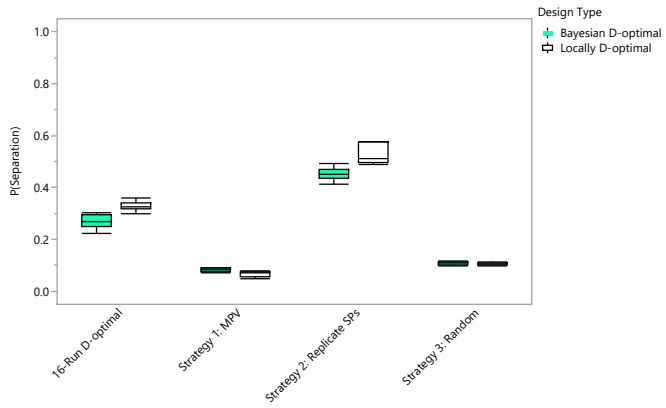
As the linear predictor of the logistic response function becomes more complex i.e., interaction or quadratic effects are added to the model, the initial D-optimal designs generated from the same model have significantly different separation probabilities, as observed in previous studies Park *et al.* (2019). This explains the increased variance in separation probabilities observed for the augmentation strategies in the quadratic

model, as the initial designs do not have uniform separation probabilities. Replicating runs at support points of the initial design is the least effective method for creating overlap, as designs augmented in this manner encountered separation more frequently than the 16-run D-optimal designs. Although MPV augmentation appears to be the best strategy for creating overlap to obtain usable ML estimates, this methodology comes with several drawbacks that will be discussed in this section.

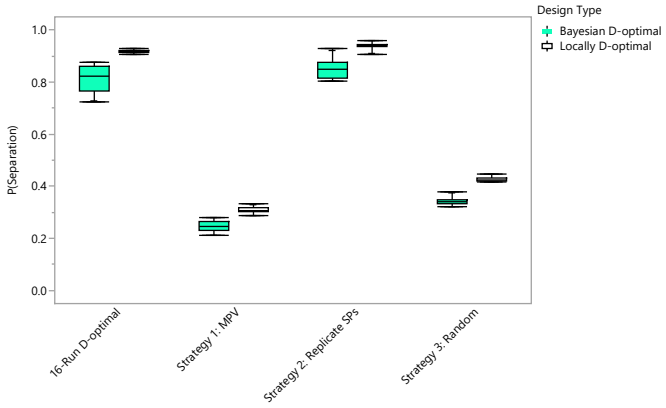
5.4.1 D-Efficiency of Augmented Designs

MPV augmentation tends to produce designs with relatively low D-efficiencies, shown in Figure 5.4. The D-efficiency of each design is calculated as shown in Equation 3.6, where $|\mathbf{M}(\xi^*)|$ is the determinant of the information matrix for a design ξ^* which achieves the maximum $|\mathbf{M}(\xi)|$ across 16-run designs and p is the number of model parameters. As a practical interpretation of the D-efficiency metric, suppose a candidate design has $D_{\text{eff}} = 0.5$. Then, two replicates of the candidate design are required to produce parameter estimates with standard errors comparable to the D-optimal design (Atkinson *et al.*, 2007). D-efficiency values close to 1 are desired for candidate designs, while values close to zero indicate poor performance with respect to the precision of parameter estimates.

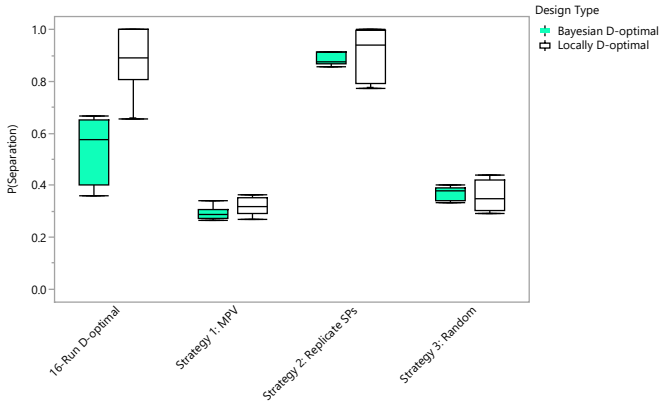
For the quadratic model, the curvature present in the response surface led to initial D-optimal designs that were noticeably more spread out over the design region, in comparison to the initial designs for the ME and 2FI models. All twenty initial 8-run D-optimal designs for the quadratic model had eight unique support points spread throughout the largest area of the design region, as opposed to four or five support points observed for the ME and 2FI models. As a result, the initial designs for the quadratic model had the largest variance in D-criterion values, which led to the relatively large variance observed in the D-efficiencies across augmented designs for



(a) ME Model Designs

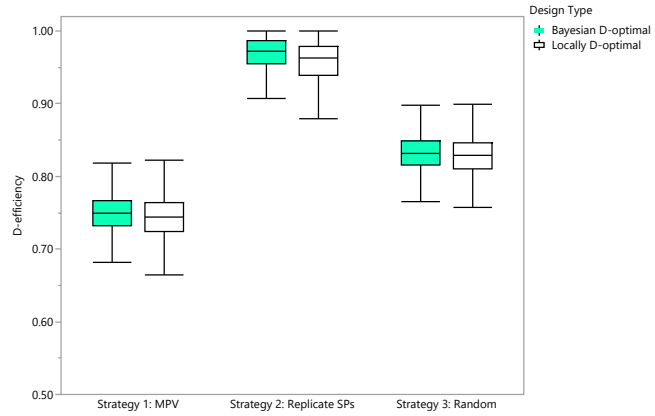


(b) 2FI Model Designs

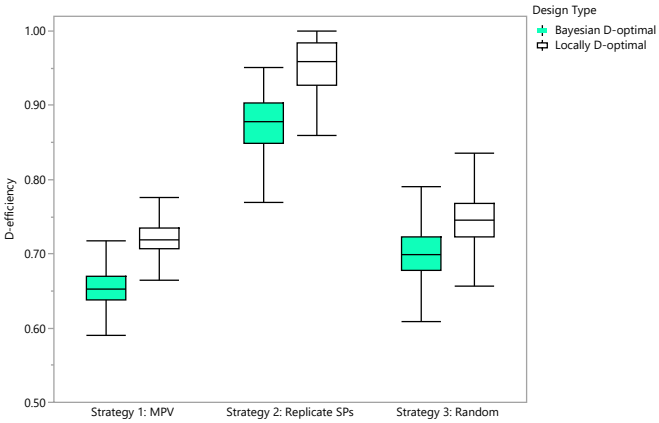


(c) Quadratic Model Designs

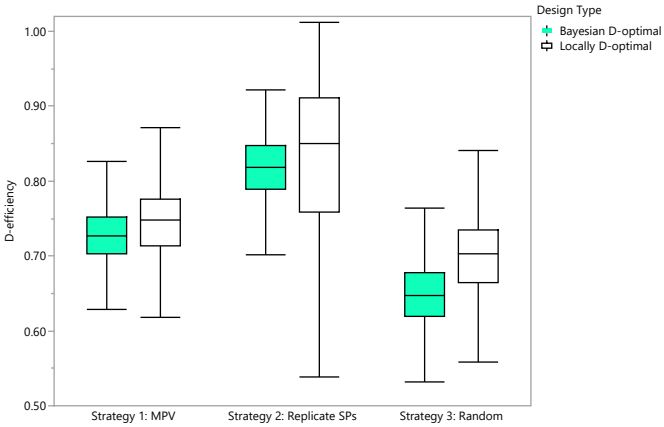
Figure 5.3: P(Separation) vs. Augmentation Strategy by Model



(a) ME Model Designs



(b) 2FI Model Designs



(c) Quadratic Model Designs

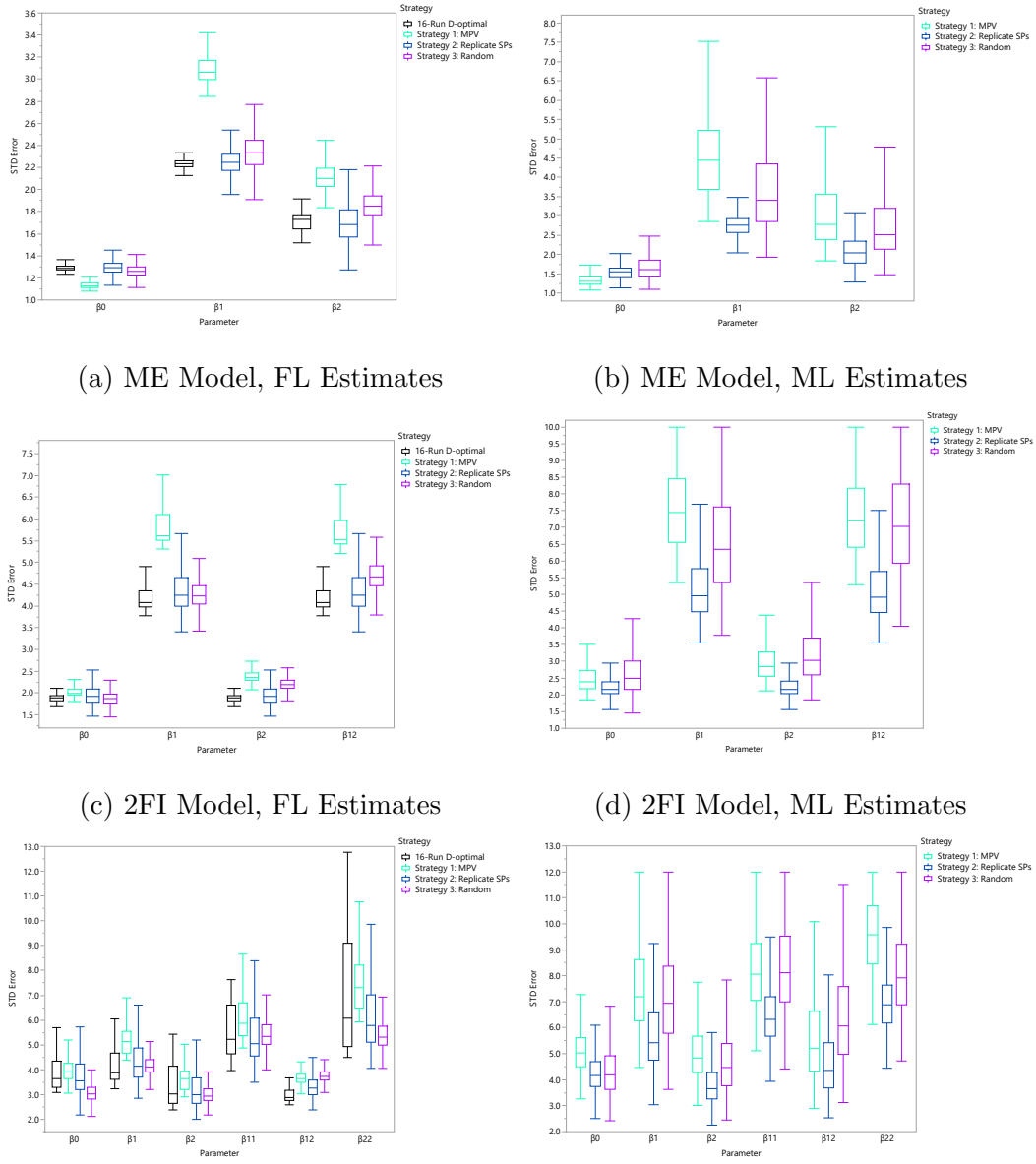
Figure 5.4: Relative *D*-Efficiency vs. Augmentation Strategy by Model

the quadratic model. Furthermore, randomly replicating at support points produced the largest observed variance in D-efficiencies for the quadratic model, as there were more support points available for replication.

The relatively low D-efficiencies of the MPV augmented designs is reflected in the standard errors of the ML and FL parameter estimates, shown in Figure 5.5. For the main effect coefficients (β_1 and β_2), MPV augmentation consistently produced estimates with the largest standard errors. The FL estimates shown for the 16-run D-optimal designs were obtained with separated response data, and as a result no valid ML estimates are available for these designs. The ML estimates shown in Figure 5.5 represent the usable ML estimates that were observed in the simulation study. With separation present in the response data, the ML estimates reported have large standard errors and are not considered usable estimates. However, even overlap present in the response data, there are still cases where the standard errors of the ML parameter estimates are too inflated to be considered usable. Therefore, the maximum standard errors observed in the FL parameter estimates for each model were used as the approximate standard error limit to screen out invalid ML estimates. Table 5.3 shows the percentage of usable ML estimates observed across all replications from each augmentation strategy for the three logistic regression models. Although MPV augmentation produced the highest percentage of overlapped response data, the percentage of usable ML estimates obtained via MPV augmentation is comparable to random augmentation.

5.4.2 ML vs. FL Estimation

As shown in Figure 5.5, the 16-run D-optimal designs generally produce FL estimates with lower standard errors than FL estimates produced by the augmentation strategies explored in this study. Additionally, the standard errors of the usable ML



(a) ME Model, FL Estimates

(b) ME Model, ML Estimates

(c) 2FI Model, FL Estimates

(d) 2FI Model, ML Estimates

(e) Quadratic Model, FL Estimates

(f) Quadratic Model, ML Estimates

Figure 5.5: Standard Errors of ML/FL Parameter Estimates, Excluding Outliers

estimates obtained with overlapped response data under any of the augmentation strategies still tend to be larger than the FL estimates obtained from the 16-run *D*-optimal designs for separated response data.

However, the point estimates of the parameters obtained through FL estimation

Table 5.3: Percentage of Usable ML Estimates Obtained in Simulation

	S1: MPV	**S2: Replicate SPs**	**S3: Random**
ME Model	73.42%	48.52%	75.29%
2FI Model	26.15%	10.14%	27.10%
Quadratic Model	18.08%	7.28%	13.57%

tend to underestimate the magnitude of the model effects more severely than ML estimation. Table 5.4 shows the average FL and ML parameter estimates obtained with 50,000 augmented designs and simulated response data (20 starting designs, 2500 augmentation trials per design), along with the standard deviations for each estimate. While the variance of the estimates is higher with ML estimation, the consistent underestimation of the magnitude of the true model parameters with FL estimation always produced uninformative Wald confidence intervals in this simulation study.

Table 5.4: Point Estimates for ME Model; $\beta_0 = 0$, $\beta_1 = 2$, $\beta_2 = 1$

	FL			**ML**		
Mean:	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
16-Run D-optimal	-0.00931	1.287853	0.622917			
Strategy 1: MPV	0.023969	0.944243	0.437915	-0.00602	2.861616	1.46304
Strategy 2: Replicate SPs	0.009731	1.017239	0.477679	0.008559	1.982252	0.972758
Strategy 3: Random	0.003441	0.959783	0.456886	0.001374	2.777335	1.387015
Std Dev:						
16-Run D-optimal	0.295307	0.274647	0.273521			
Strategy 1: MPV	0.299646	0.429178	0.322501	0.728604	1.440639	1.194952
Strategy 2: Replicate SPs	0.260802	0.372414	0.279759	0.58463	0.784901	0.638278
Strategy 3: Random	0.272463	0.363308	0.316568	0.820055	1.309719	1.162268

Wald inference is used in logistic regression to perform hypothesis tests on individual model parameters. The hypothesis test involves the null and alternative shown

in Equation 5.3.

$$H_0 : \beta_j = 0, \quad H_1 : \beta_j \neq 0. \tag{5.3}$$

Wald inference can be used to construct confidence intervals on individual model parameters; an approximate $100(1-\alpha)$ percent confidence interval is shown in Equation 5.4,

$$\hat{\beta}_j - \mathrm{Z}_{\alpha/2}\mathrm{SE}(\hat{\beta}_j) \leq \beta_j \leq \hat{\beta}_j + \mathrm{Z}_{\alpha/2}\mathrm{SE}(\hat{\beta}_j) \tag{5.4}$$

where $\mathrm{SE}(\hat{\beta}_j)$ represents the standard error of the estimate $\hat{\beta}_j$, which is given by the square root of the jth diagonal element of the covariance matrix $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \mathbf{M}^{-1}$ (Montgomery *et al.*, 2012). Constructing an approximate $100(1-\alpha)$ percent confidence interval that contains zero is equivalent to failing to reject the null hypothesis $H_0 : \beta_j = 0$ at significance level α. Failing to reject H_0 implies that there is not sufficient evidence to conclude that parameter β_j is not equal to zero, implying that factor j is not significantly impacting the response.

Due to the magnitude of the FL point estimates relative to their respective standard errors, the Wald confidence intervals constructed using the FL estimates/standard errors at significance level $\alpha = 0.1$ across all 50,000 augmentation trials of each augmentation strategy were never able to identify factor 1, the larger main effect (as $\beta_1 > \beta_2$ for the ME model), as significant; the confidence intervals for β_1 always contained zero. However, the Wald confidence intervals constructed using ML estimates/standard errors at $\alpha = 0.1$ led to the correct rejection of $H_0 : \beta_1 = 0$ a small percentage of the time, summarized in Table 5.5 for the ME model.

Only a trivial percentage of ML Wald confidence intervals correctly identified β_1 as non-zero. However, the consistent underestimation of model coefficient magnitude using the FL method is noteworthy, as it always produced overly conservative conclusions with Wald inference.

Table 5.5: Augmentation Trials where $H_0 : \beta_1 = 0$ was Rejected (ME Model)

	S1: MPV	**S3: Replicate SPs**	**S3: Random**
Count	706	242	315
Percentage	1.41%	0.48%	0.63%

5.5 Summary

This study explored three non-sequential augmentation strategies for small-sample, two-factor D-optimal designs when separation of binary response data prevents estimation of maximum likelihood (ML) estimates for the logistic regression model. The goal of augmenting initial D-optimal designs with an additional block of runs is to create overlap in the response data to eliminate quasi- or complete separation and produce usable ML estimates. The non-sequential manner of augmenting runs to an initial design is similar in spirit to full fold-overs in fractional factorial designs such that for an initial design of size n, an equal number of runs J is augmented. A simulation study was designed and executed to evaluate the three augmentation strategies by augmenting runs of size $J = 8$ to an initial design of size $n = 8$; in addition, D-optimal designs of size $n = 16$, which reflects a one-shot design strategy with no augmentation, were also constructed and evaluated.

Among the three proposed strategies, MPV augmentation was found to be the most effective at creating overlap and encountered the least frequency of separation for simulated response data. However, MPV augmentation produced designs with relatively low D-efficiencies; as a result, the ML estimates obtained from the MPV designs produced the largest standard errors. When using Firth likelihood (FL) estimation, estimates obtained from the 16-run D-optimal designs with separated

response data produced the lowest standard errors compared to estimates obtained using the three augmentation strategies.

For small-sample designs, such as the ones explored in this study, there are higher chances of encountering separation. Therefore, having overlapped response data to produce usable ML estimates may be desirable to prevent overly conservative Wald inference. It was shown in Section 5.4.2 that the magnitudes of the model coefficients were consistently underestimated with FL estimation, leading to Wald confidence intervals that always contained zero. Although the proportion of correct Wald inference did not differ significantly between FL and ML estimation in this study, the difference may be more pronounced for different model forms, prior distributions, or design sizes. This indicates that pursuing overlapped response data justifies future research.

The key assumption in the simulation study was that the parameter estimates specified in design creation matched the true parameter values ($\hat{\boldsymbol{\beta}}_{\text{initial}} = \boldsymbol{\beta}$). Response data was also simulated in accordance with this assumption. However, even with the input of subject matter experts and previous experimental results, this assumption will almost certainly not be valid. If the parameter estimates used in the initial design generation are poor, the model response surface constructed using these values will be inaccurate, and MPV augmentation may add design runs in regions of low prediction variance, where the expected value may differ significantly from 0.5. Performing this simulation study in accordance with this assumption also inflated the D-efficiencies of the D-optimal designs, as the D-optimal designs were generated using the true parameter values. Therefore, the assumption impacts the results of the simulation study in two ways. The effectiveness of MPV augmentation in creating overlap may not be achievable in practice, as determining accurate parameter estimates before initial experimentation may not be feasible. Conversely, due to the baseline 16-run D-

optimal designs being created with the true parameter values, the D-criterion values of these designs are maximized. Consequently, the difference in D-efficiency between the MPV augmented designs and the maximum efficiency achievable in practice with initial parameter uncertainty may not be as pronounced as the efficiency disparity between the MPV augmented designs and the locally D-optimal designs demonstrated in this study.

Although the performance of MPV augmentation in creating overlapped response data has been idealized due to the key assumption, the simulation study does show that placing augmented runs in regions of high prediction uncertainty is a promising strategy for creating overlap. The limitation of this augmentation strategy in practice is being able to reliably determine areas of MPV in the design region after an initial experiment. Practically, two methods are available for deducing regions of high prediction uncertainty in an initial experiment that encounters separation if the parameter estimates used in design creation are not considered reliable. First, FL estimation provides a means of updating the parameter estimates with separated response data. However, results of this simulation study indicate that the magnitude of the parameters is likely to be underestimated using FL methods. Support vector networks, specifically the optimal hyperplane algorithm described by Cortes and Vapnik (1995), may be used to both expediently identify separation and define an optimal region for locating additional design runs that is, if an optimal separating hyperplane exists in the design region, separation is present and the region spanned by the hyperplane is prime for MPV augmentation. It has been proven that generating an optimal separating hyperplane is a quadratic programming problem that can be solved efficiently (Cortes and Vapnik, 1995). Additionally, the optimal separating hyperplane represents the region that splits the design points and associated response data with maximal margin, and augmentation on or near the optimal separating hy-

perplane may replicate MPV augmentation without the requirement for an accurate determination of the response surface.

Another direction for future research in this area is the exploration of *constrained* MPV augmentation. Constraints can be formulated to restrict the eligible augmentation region to areas that will yield augmented designs of high D-efficiency. Then, the areas of MPV within the constrained regions can be excellent candidate locations for augmentation. Augmented points in these areas will serve to create overlap expediently while maintaining high D-efficiency, and the trade-off observed in this simulation study between overlap creation effectiveness and D-efficiency can be mitigated.

Chapter 6

A COMPOUND OPTIMALITY CRITERION FOR SEPARATION ROBUSTNESS

The D_{MP}-criterion is proposed to generate optimal designs for the logistic regression model with reduced separation probabilities. This compound criterion has two components: (1) the D-efficiency of the candidate design, and (2) a penalty term that captures the average distance of the candidate design's support points from the region of maximum prediction variance (MPV). A D_{MP}-optimal design maximizes the D_{MP}-criterion. The aim is to obtain compromise experimental designs with high D-efficiencies that are more robust to separation than a D-optimal design of equal size. The purpose of this chapter is to present the D_{MP}-criterion and demonstrate examples of its potential use as a means of mitigating separation in the design phase of a binary response experiment.

6.1 Introduction

The D-criterion is the most widely used of all optimality criteria, as it is computationally efficient when implemented in computer exchange algorithms and tends to produce exact designs with minimum variance parameter estimates (Goos, 2012; Montgomery, 2017). A D-optimal design minimizes $-\log|\mathbf{M}^{-1}|$, thus minimizing the standard errors of $\hat{\boldsymbol{\beta}}$ (Atkinson *et al.*, 2007). However, the simulation study in Chapter 4 indicates that small-sample ($n \leq 32$) D-optimal designs for the logistic regression model are prone to encountering separation.

The simulation study presented in Chapter 5 shows that MPV augmentation most efficiently eliminates separation that occurs in an initial D-optimal design. Augmenting design runs near the MPV region, where $E(y|\boldsymbol{\beta}) = 0.5$, tends to produce

overlapped response data more reliably then allocating additional runs in accordance with the D-criterion. However, the MPV augmented designs tend to have much lower D-efficiencies than D-optimal designs of equal size, indicating that D-optimality and separation-optimality are opposing objectives. Motivated by these observations, the D_{MP}-criterion is formulated using a compound optimal design approach to balance D-optimality with separation robustness.

The term *compound* criterion denotes an optimality criterion that is formulated as a weighted average of several design criteria, which may depend on different information matrices (Müller and Stehlík, 2010). Suppose two functionals of candidate design ξ, denoted as $\Psi_1(\xi)$ and $\Psi_2(\xi)$ exist on the space of information matrices, $\mathcal{M}$. A compound design optimizes a weighted average functional with the form shown in Equation 6.1 (Cook and Wong, 1994).

$$\Psi(\xi) = \lambda\Psi_1(\xi) + (1-\lambda)\Psi_2(\xi), \quad 0 \leq \lambda \leq 1 \tag{6.1}$$

Cook and Wong (1994) prove that this method is equivalent to a *constrained* optimal design methodology, where Ψ_2 is optimized subject to a constraint on Ψ_1, assuming that both criteria are monotonic functions of design efficiencies relative to multiple designs on each criterion and that both are defined on a space of information matrices. The constrained design problem can therefore be formulated as

$$\max\ \Psi_2(\mathbf{M}(\xi)) \text{ over } \mathbf{M}(\xi) \in \mathcal{M}^+, \text{ subject to } \Psi_1(\mathbf{M}(\xi)) \geq c, \tag{6.2}$$

where $\mathcal{M}^+$ denotes the set of nonsingular information matrices, and $c \in [-\infty, 0]$ denotes the log efficiency of design ξ relative to the optimal design ξ_1^*. For GLMs, McGree *et al.* (2008) present three possible general forms for compound criteria. For maximizing a criterion while minimizing a loss function, L, the general form is shown in Equation 6.3,

$$\Phi(\boldsymbol{\beta}, \xi) = \frac{\Psi(\boldsymbol{\beta}, \xi)}{1 + L(\boldsymbol{\beta}, \xi)} \tag{6.3}$$

as Ψ, L, and consequently Φ are dependent on the model parameters for GLMs.

Compound optimal designs are closely tied to the methodology of formulating a desirability function (DF) for the simultaneous optimization of several response variables, originally introduced by Derringer and Suich (1980). This approach allows variables on different measuring scales to be scaled to values in the range $[0, 1]$ for more balanced weighting in the DF. Meta-models, typically higher order polynomials, for each response are transformed into a range of values between zero and one, then combined into a singular objective DF capturing the total desirability of the combined responses (Dabbas *et al.*, 2003). However, the formulation of the DF assumes that all responses are equally important. In cases where the DF has a continuous first derivative, gradient-based optimization methods can be used to find an optimal design. However, in commonly used additive and multiplicative DF forms, a first derivative may not exist. Del Castillo *et al.* (1996) present modified desirability functions such that non-differentiable points are eliminated and allow the user to assign different priorities among the responses. Further modifications of this method have been presented to extend its applicability and reduce the amount of required computations (Ch'ng *et al.*, 2005).

A compound design criterion, DP_s, has been proposed by Rahman (2015) for single-factor design cases to balance D-efficiency with separation optimality. The DP_s criterion is shown in Equation 4.2. First, locally separation-optimal designs, denoted as P_s-optimal, are illustrated for single-factor main effect cases, where the model linear predictor takes the form $\eta = \beta_0 + \beta_1 x$. For designs of run size n, these designs have n support points, equally spaced by a small δ, that are centered about the point of MPV. For example, for $n = 8$, the design vector is $\mathbf{x} = [-0.003, -0.002, -0.001, 0.000, 0.001, 0.002, 0.003, 0.004]$. For this example, $\beta_0 = 0$, and the point of MPV is $x = 0$. With $\alpha = 0.5$, the average distance of the support

points of the locally optimal DP_s design from the point of MPV ($x = 0$) is reduced relative to the locally D-optimal design. The support points of the D-optimal, DP_s-optimal, and P_s optimal designs presented by Rahman (2015) for $n = 8$ are shown in Figure 6.1 (note that the design region is not on the coded scale $[-1, 1]$). The R® software package *optim*, which uses simulated annealing in conjunction with a gradient descent algorithm, was used to generate the DP_s-optimal designs (Rahman, 2015). The separation probabilities reported in the D-optimal, P_s-optimal, and DP_s-optimal designs are 0.3279, 0.0625, and 0.0884, respectively. The D-efficiencies of the the P_s-optimal and DP_s-optimal designs relative to the D-optimal design are 0.03% and 65.24%, respectively.

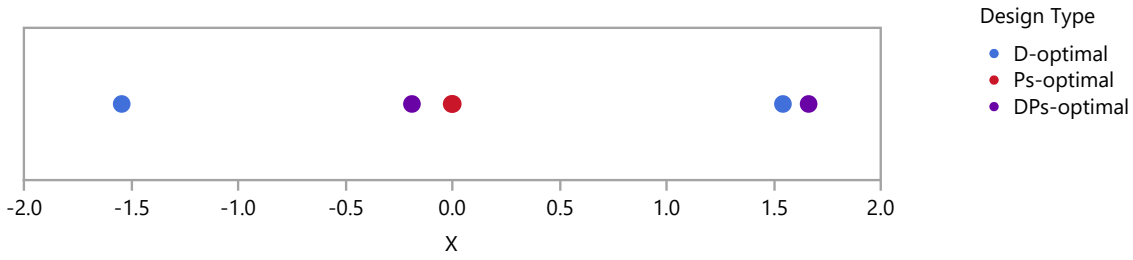


Figure 6.1: Support Points of the D-, P_s-, and DP_s-Optimal Designs

The design methodology proposed by Rahman (2015) requires an exact calculation of the probability of separation in a candidate design, which itself requires an enumeration of all 2^n possible response sets for the design and determining the presence of separation in each set. This is a straightforward procedure in one dimension, but does not extend well to multi-factor design spaces. Furthermore, the formulation of the DP_s-criterion does not scale well with different forms of the logistic regression model, as the value of $|\mathbf{M}|$ is maximized directly in the denominator. For different forms of the logistic regression model, the maximum attainable value of $|\mathbf{M}|$ is different. This means that designs with equal functional values for DP_s-optimality may

have significantly varied separation probabilities across different parameter values and forms of the linear predictor term.

A current search of the literature reveals no existing compound design criterion to mitigate separation risk for multi-factor optimal designs for the logistic regression model. Therefore, the objective of this chapter is to propose a design criterion with a tuning parameter to balance D-efficiency with reduced susceptibility to separation. The compound D_{MP}-criterion is formulated in Section 6.2. The rest of this chapter is organized as follows: Section 6.3 presents a two-phase coordinate exchange algorithm (CEA) that is used to generate exact D_{MP}-optimal designs. In Section 6.4, the performance of the D_{MP} designs with respect to D-efficiency and separation probability is evaluated in a simulation study; performance across a range of weights for the D-efficiency component will be examined in several cases with specified model forms. To address the design dependence problem in GLMs, where the optimal design is dependent on parameter estimates that must be specified in design creation (see Khuri *et al.* (2006)), a robust design methodology using a *maximin* approach is presented for the D_{MP}-criterion in Section 6.5. Section 6.6 provides recommendations for optimal weight values and potential directions for future work. While this Chapter focuses on one- and two-factor design cases, the methodology proposed can be easily scaled to three or more factors. Additionally, because the optimization proceeds through an exchange algorithm, we produce *exact*, as opposed to continuous, designs (see Atkinson *et al.* (2007) for an explanation of the distinction between exact and continuous designs).

6.2 The D_{MP}-Criterion

The functional of D_{MP}-optimality for the logistic regression model is

$$\Psi_{D_{MP}} = \lambda \left[\left(\frac{|\mathbf{M}(\xi)|}{|\mathbf{M}(\xi^*)|} \right)^{\frac{1}{p}} \right] - (1-\lambda) \left[\left(\frac{\sum_{i=1}^{n} (\mathbf{x}_i{}^T \boldsymbol{\beta})^2 / n}{\max_{\chi} (\mathbf{x}^T \boldsymbol{\beta})^2} \right)^{\frac{1}{p}} \right], \tag{6.4}$$

where ξ is an n-run candidate design in factorial space χ, $|\mathbf{M}(\xi^*)|$ is the maximum determinant value of the information matrix attainable across all candidate designs in χ, p is the number of parameters in the linear predictor of the specified logistic regression model, $\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2$ denotes the maximum squared value of the linear predictor across all possible design points in χ, and $\lambda \in [0, 1]$ is a blending coefficient. A D_{MP}-optimal design, denoted as $\xi^*_{D_{MP}}$, maximizes Equation 6.4.

$$\xi^*_{D_{MP}} = \underset{\xi}{\operatorname{argmax}} \{\Psi_{D_{MP}}\} \tag{6.5}$$

The user-defined inputs for this criterion are:

1. n - number of runs (trials) in the experiment
2. $\boldsymbol{\beta}$ - vector of model coefficients (parameters)
3. $\lambda \in [0, 1]$ - blending coefficient that defines the weight of a candidate design's D-efficiency relative to the average squared linear predictor magnitude penalty

The D_{MP}-criterion seeks to maximize the D-efficiency of the candidate design (see Equation 3.6) while penalizing the distance of each support point in the candidate design from the MPV region. Since $E(y_i|\boldsymbol{\beta}) = 0.5$ holds exclusively for points in the MPV region, this implies that all points in this region have a linear predictor value of zero.

$$E(y_i|\boldsymbol{\beta}) = (1 + \exp(-\mathbf{x}_i{}^T\boldsymbol{\beta}))^{-1} = 0.5 \iff \mathbf{x}_i{}^T\boldsymbol{\beta} = 0 \tag{6.6}$$

Therefore, the term $\sum_{i=1}^{n} (\mathbf{x}_i{}^T\boldsymbol{\beta})^2/n$ in Equation 6.4 is the *average* squared value of the linear predictor for all design points in candidate design ξ. The squared value is used

because the magnitude of the linear predictor determines the difference in expected value of a design point from 0.5. A more direct measure of the average linear predictor magnitude is $\sum_{i=1}^{n} |\mathbf{x}_i{}^T\boldsymbol{\beta}|/n$. However, the average squared term is used in Equation 6.4 instead of the average absolute value term to prevent non-differentiable points in the functional for D_{MP}-optimality. A higher value of $\sum_{i=1}^{n} (\mathbf{x}_i{}^T\boldsymbol{\beta})^2/n$ indicates that the support points of the candidate design have an average expected value closer to zero or one, implying that they are farther from the MPV region. The relationship between the value of the linear predictor at point $\mathbf{x}_i$, $E(y_i|\mathbf{x}_i)$, and $(E(y_i|\mathbf{x}_i) - 0.5)^2$ is shown in Figure 6.2.

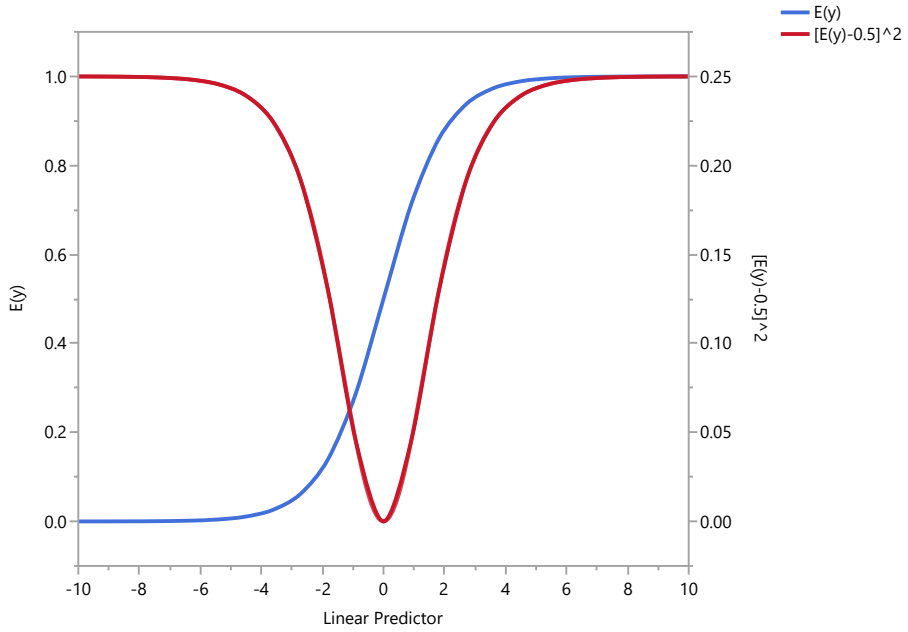


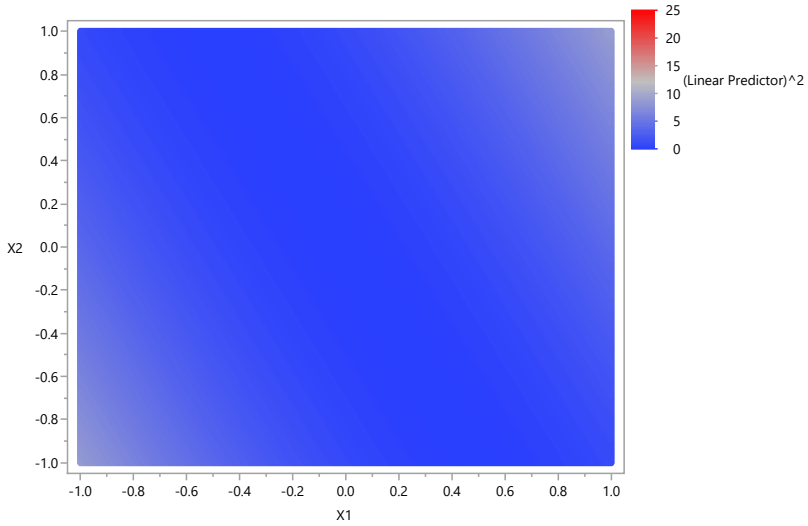
Figure 6.2: $E(y)$, $(E(y) - 0.5)^2$ vs. Linear Predictor Value for the Logistic Regression Model

The D_{MP}-criterion is formulated as shown in Equation 6.4 such that the measure of D-optimality and the linear predictor penalty are normalized, an idea that has been adopted from the DF approach to experimental design. The D_{MP}-criterion has the general form shown in Equation 6.1 for a compound criterion; the form for minimizing

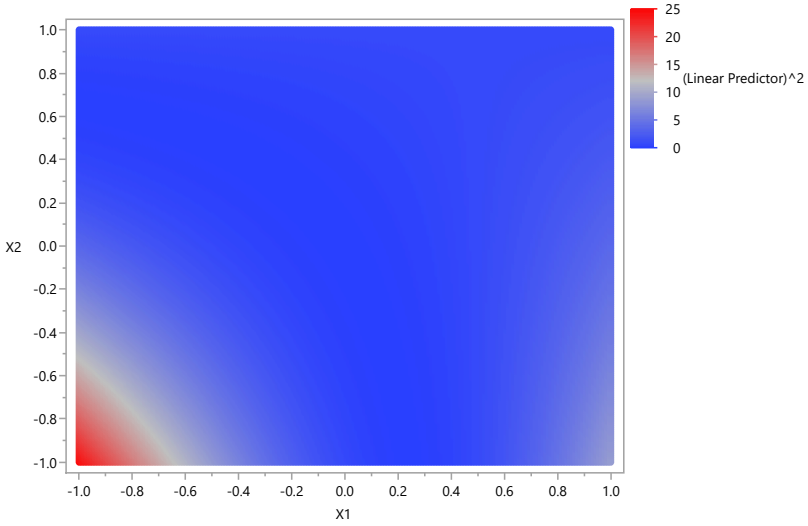
a loss function shown in Equation 6.3 is not used. The proposed formulation allows the user to scale the criterion in accordance with separation risk tolerance. For example, in an experimental situation where follow-on experimentation is feasible, the user may wish to place more weight on the D-efficiency component to maximize the precision of the parameter estimates. However, if the user is operating under tight time or resource constraints where additional experimentation is impossible, more weight can be placed on the linear predictor penalty to reduce the risk of encountering separation. This flexibility would not be available if the criterion was formulated as shown in Equation 6.3.

The first term in Equation 6.4, $\left[\left(\frac{|\mathbf{M}(\xi)|}{|\mathbf{M}(\xi^*)|}\right)^{\frac{1}{p}}\right]$, is the D-efficiency of candidate design ξ relative to the D-optimal design for parameter vector $\boldsymbol{\beta}$, denoted as ξ^*. As a D-optimal design maximizes $|\mathbf{M}(\xi)|$, the D-efficiency measure will have a range of $[0, 1]$. The second term in Equation 6.4, $\left[\left(\frac{\sum_{i=1}^{n}(\mathbf{x}_i{}^T\boldsymbol{\beta})^2/n}{\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2}\right)^{\frac{1}{p}}\right]$ is the normalized squared linear predictor penalty term adjusted for the number of parameters in the linear predictor. As $\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2$ is the maximum squared value of the linear predictor over χ, the squared linear predictor penalty term will also have a range of $[0, 1]$. The second term is taken to the $(1/p)$th power to adjust it relative to the dimensions of the specified logistic regression model in a manner similar to the D-efficiency measure. As additional model terms are added, the maximum magnitude of the linear predictor over χ tends to increase. An example is shown in Figure 6.3, where an active interaction term is added to a two-factor main effects model. The maximum squared value of the linear predictor for the main effects model is 9, while the maximum squared value of the linear predictor for the two-factor interaction (2FI) model is 25. The $(1/p)$th power adjustment therefore allows for more consistent λ recommendations across models of different size.

A limitation of the D_{MP}-criterion is that it requires two values, $|\mathbf{M}(\xi^*)|$ and



(a) ME Model, $\beta_0 = 0$, $\beta_1 = 2$, $\beta_2 = 1$



(b) 2FI Model, $\beta_0 = 0$, $\beta_1 = 2$, $\beta_2 = 1$, $\beta_{12} = -2$

Figure 6.3: Squared Linear Predictor Value over Two-Factor Design Region

$\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2$, to normalize its components. An empirical estimate for $|\mathbf{M}(\xi^*)|$ can be expediently determined through the generation of an exact locally D-optimal design for $\hat{\boldsymbol{\beta}}$, which can be done efficiently for GLMs (Dror and Steinberg, 2006). The value of $|\mathbf{M}(\xi)|$ for the exact D-optimal design can then serve as an approximation for $|\mathbf{M}(\xi^*)|$ to implement D_{MP}-optimality. The exact D-optimal design may not be optimal over all candidate designs of size n in χ, as these designs are typically generated via heuristic exchange algorithms. However, the normalization of both components of the D_{MP}-criterion does not need to be exact for the optimality function to produce effective compromise designs when implemented in an exchange algorithm; the normalization is in place to provide more consistent trade-offs between D-efficiency and separation probability across different forms of the logistic regression model. With the normalization in place, the impact of λ will be more consistent across different model forms and parameter values.

A theoretical bound on $|\mathbf{M}(\xi)|$ can be obtained using Hadamard's Inequality. For a symmetric $p \times p$ matrix $\mathbf{H}$, Hadamard's Inequality states that the determinant of $\mathbf{H}$ must be less than or equal to the product of its diagonal elements (Różański *et al.*, 2017).

$$\det(\mathbf{H}) \leq \prod_{i=1}^{p} h_{ii} \tag{6.7}$$

The information matrix for an n-run experimental design using a logistic regression base model with p parameters is $\mathbf{M} = \mathbf{X}^T\mathbf{V}\mathbf{X}$, where the maximum diagonal element of $\mathbf{V} \in \mathbb{R}^{n\times n}$ is $\max\{\text{Var}(y_i)\} = \max\{\pi_i(1-\pi_i)\} = 0.25$. As model matrix $\mathbf{X} \in [-1, 1]^{n\times p}$, the maximum diagonal element of $\mathbf{M} \in \mathbb{R}^{p\times p}$ is $0.25n$. Applying this to Equation 6.7 gives the information matrix determinant bound shown in Equation 6.8.

$$\det(\mathbf{M}) \leq \prod_{i=1}^{p} m_{ii} \leq (0.25n)^p \tag{6.8}$$

However, the theoretical bound for $|\mathbf{M}(\xi)|$ in Equation 6.8 is not tight, and imple-

menting the D_{MP}-criterion in Equation 6.4 using this bound does not provide a useful D-efficiency measure, as $|\mathbf{M}(\xi^*)|$ does not achieve a value close to this theoretical bound.

Additionally, for practical design applications that utilize the logistic regression model, determining an analytical solution for $\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2$ will be mathematically tractable. For example, for a model with only main effect terms in the linear predictor, $\max_{\chi}(\mathbf{x}^T\boldsymbol{\beta})^2$ is determined by evaluating the linear predictor at the 2^k extreme points of the design region χ, where k is the number of design factors.

6.3 Two-Phase Coordinate Exchange Algorithm for Generating D_{MP}-Optimal Designs

A two-phase CEA with a final one-pass point exchange is implemented to create exact D_{MP}-optimal designs for the logistic regression model. In Phase I, a locally D-optimal design for the specified parameter vector $\boldsymbol{\beta}$ is generated to determine an empirical estimate for $|\mathbf{M}(\xi^*)|$, which is used to assess the D-efficiency of candidate designs in Phase II. In Phase II, the user-defined weight λ is incorporated to create a D_{MP}-optimal design. As recommended by Meyer and Nachtsheim (1995), 1000 random starting designs are used in each phase.

As observed by Gotwalt *et al.* (2009), coordinate exchange often leads to designs that consist of several distinct clusters of points. To collapse the clusters into replicates, at the end of each phase, one iteration of a point (row) exchange is performed on the final design matrix given by the two-phase CEA, where the candidate set consists of the rows of the final design matrix. As a result, the efficiency of the design is marginally increased and the number of levels of the continuous factors that must be tested are reduced, which will facilitate easier experimentation in practice. The pseudo-code for the two-phase CEA is shown in Algorithm 1. The two-phase CEA

was implemented using JMP Scripting Language (JSL) in JMP Pro 14.2®; the JSL code for the two-phase CEA is available in Appendix B.

Algorithm 1 Two-Phase D_{MP}-optimal CEA

procedure 2PCEA($\boldsymbol{\beta}$, λ, n, f, iter, gap)
  $\Psi_{\text{bestD}} = M_1 \ll 0$ ▷ Phase I
  for $k = 1, 2, ...,$iter **do**
    Generate random $\xi_k \in [-1, 1]^{n \times f}$
    Expand ξ_k to $\mathbf{X}_k$
    converge = $M_2 \gg 0$
    while converge > gap **do**
      $\mathbf{X}_{\text{start}_k} = \mathbf{X}_k$
      for row $i = 1, 2, ..., n$ **do**
        for column $j = 2, ..., f + 1$ **do**
          optimize x_{ij}^k - adjust by $\pm\delta$ until $\Delta\Psi_D <$ gap
      converge = $\Psi_D(\mathbf{X}_k) - \Psi_D(\mathbf{X}_{\text{start}_k})$
    if $\Psi_D(\mathbf{X}_k) > \Psi_{\text{bestD}}$ **then**
      $\mathbf{X}_{\text{bestD}} = \mathbf{X}_k$
      $\Psi_{\text{bestD}} = \Psi_D(\mathbf{X}_{\text{bestD}})$
  $C_D = \mathbf{X}_{\text{bestD}}$
  for row $i = 1, 2, ..., n$ **do**
    Exchange row $\mathbf{x}[i]_{\text{bestD}}$ with best row in C_D

▷ Establish $\Psi_{D_{MP}}$ based on Phase I results, $|\mathbf{M}(\xi^*)| = \Psi_{\text{bestD}}$, find maximum magnitude value of linear predictor over χ (Equation 6.4)

  $\Psi_{\text{best}} = M_1 \ll 0$ ▷ Phase II
  for $k = 1, 2, ...,$iter **do**
    Generate random $\xi_k \in [-1, 1]^{n \times f}$
    Expand ξ_k to $\mathbf{X}_k$
    converge = $M_2 \gg 0$
    while converge > gap **do**
      $\mathbf{X}_{\text{start}_k} = \mathbf{X}_k$
      for row $i = 1, 2, ..., n$ **do**
        for column $j = 2, ..., f + 1$ **do**
          optimize x_{ij}^k - adjust by $\pm\delta$ until $\Delta\Psi_{D_{MP}} <$ gap
      converge = $\Psi_{D_{MP}}(\mathbf{X}_k) - \Psi_{D_{MP}}(\mathbf{X}_{\text{start}_k})$
    if $\Psi_{D_{MP}}(\mathbf{X}_k) > \Psi_{\text{best}}$ **then**
      $\mathbf{X}_{\text{best}} = \xi_k$
      $\Psi_{\text{best}} = \Psi_{D_{MP}}(\mathbf{X}_{\text{best}})$
  $C = \mathbf{X}_{\text{best}}$
  for row $i = 1, 2, ..., n$ **do**
    Exchange row $\mathbf{x}[i]_{\text{best}}$ with best row in C
  return $\mathbf{X}_{\text{best}}$

6.3.1 Validation of Two-Phase Coordinate Exchange Algorithm

To validate the performance of the two-phase CEA, it was configured to generate locally D-optimal designs ($\lambda = 1$, $|\mathbf{M}(\xi^*)| = 1$). Three tests were performed to compare the resulting designs from the two-phase CEA to the nonlinear design platform in JMP Pro 14.2® configured to generate locally optimal designs (abscissas radii set to zero, see Gotwalt *et al.* (2009) for the numerical integration method implemented in JMP Pro 14.2®). For all tests, the parameter values were selected such that the coded design region $\chi \in [-1, 1]^k$ includes areas of extreme response probability. As noted by Dror and Steinberg (2005), locally D-optimal designs for the logistic regression model will contain interior support points when the design region includes expected response values outside of the approximate range $0.15 \leq E(y) \leq 0.85$. Therefore, these validation test models will force the design algorithms to find support points that are not located at the extreme points of χ.

Validation Test 1: Single-Factor Main Effect Model

For this validation test, a single-factor main effect model with $\beta_0 = 0$ and $\beta_1 = 5$ was used. A design size of $n = 8$ was examined. Both algorithms generated a design with all runs allocated evenly between two support points. The designs generated by the two-phase CEA and JMP Pro 14.2® are shown in Figure 6.4. The designs are practically identical. The D-efficiency of the two-phase CEA design relative to the JMP 14.2 Pro® design is 99.9995%.

Validation Test 2: Two-Factor Main Effects Model

For this validation test, a two-factor main effects model with $\beta_0 = 0$, $\beta_1 = 3$, and $\beta_2 = 1$ was used. A design size of $n = 16$ was examined. Both algorithms generated a design

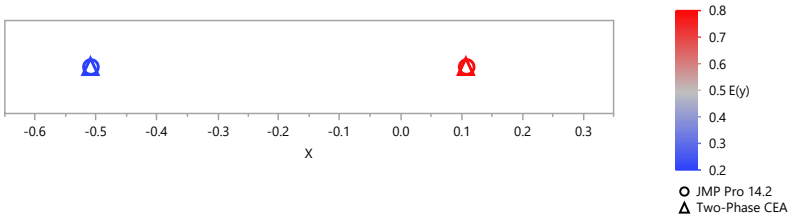


Figure 6.4: Validation Test 1 Design Comparison

with four support points. The designs generated by the two-phase CEA and JMP Pro 14.2® are shown in Figure 6.5. The designs are very similar. The D-efficiency of the two-phase CEA design relative to the JMP 14.2 Pro® design is 99.7836%.

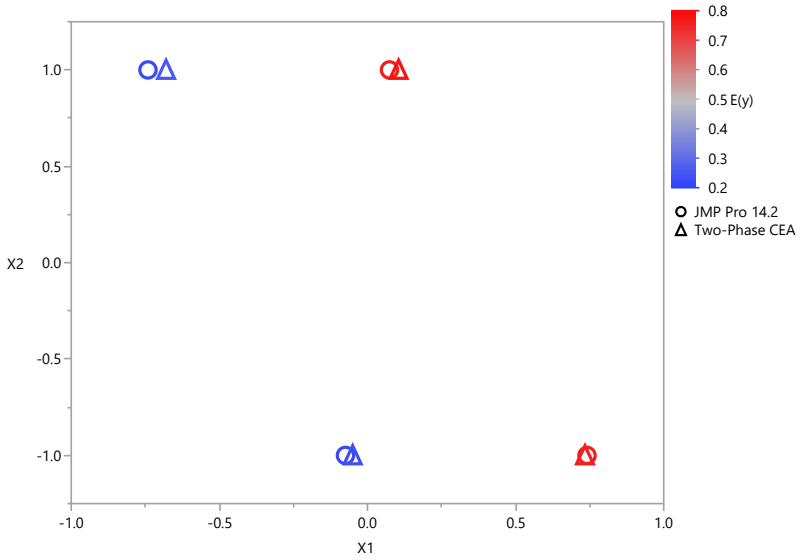


Figure 6.5: Validation Test 2 Design Comparison

Validation Test 3: Two-Factor Main Effects w/Interaction Model

For this validation test, a two-factor main effects with interaction (2FI) model with $\beta_0 = 1$, $\beta_1 = 3$, $\beta_2 = 1$, and $\beta_{12} = -1.5$ was used. A design size of $n = 16$ was examined. Both algorithms generated a design with four support points. The designs generated by the two-phase CEA and JMP Pro 14.2® are shown in Figure 6.5. The designs are

practically identical. The D-efficiency of the two-phase CEA design relative to the JMP 14.2 Pro® design is 100.0589%.

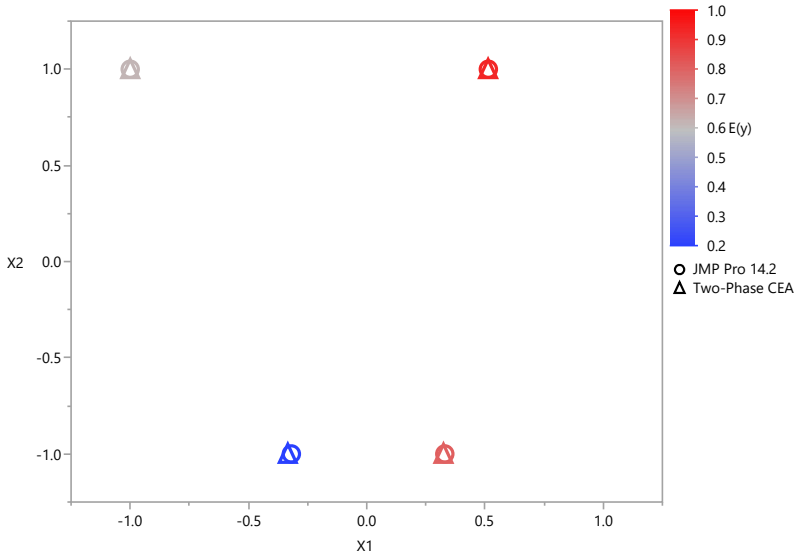


Figure 6.6: Validation Test 3 Design Comparison

6.4 Performance of Locally D_{MP}-Optimal Designs

Three test cases were performed to examine locally D_{MP}-optimal designs for different values of λ across various forms of the logistic regression model. The first case is a single-factor main effects model with n = 6, the second case is a two-factor main effects model with n = 10, and the third case is a two-factor main effects and interaction model with n = 16. As in Section 6.3.1, all test cases were chosen such that the design region contains areas of extreme response probability to force interior design points for a locally D-optimal design. The λ values examined are $\boldsymbol{\lambda}$ = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]. For each value of λ, a locally D_{MP}-optimal design of size n was created using the two-phase CEA.

Two performance metrics are used to evaluate each design: (1) D-efficiency and (2) the probability of separation. The D-efficiency of each design is calculated using

Equation 3.6, with the D-optimal design obtained in Phase I of the two-phase CEA used to determine $|\mathbf{M}(\xi^*)|$. For the first two test cases, the probability of separation for each design is calculated by enumerating all 2^n possible response vectors since the design sizes are small. Separation is declared in a response vector if the maximum parameter estimate variance is greater than 5000 (Heinze and Schemper, 2002). The total probability of separation is then calculated as shown in Equation 4.4. For the third case, the simulation methodology used in Section 4.3 is used to estimate the probability of separation for each design, where the response at each design point is simulated using a random Bernoulli draw with $P(y_i = 1) = E(y_i|\boldsymbol{\beta})$. For each design, 2500 response sets are simulated, and separation is declared in a response set if the maximum parameter estimate variance is greater than 5000 (Heinze and Schemper, 2002). The estimated probability of separation for each design is the proportion of separated response sets that are observed in the simulation. All of the D_{MP}-optimal designs described in this section are available in Appendix A.

6.4.1 Test Case 1: Single-Factor Main Effect Model

For this test case, a single-factor main effect model with $\beta_0 = 0$ and $\beta_1 = 3$ was used. A design size of $n = 6$ was examined. Since the intercept parameter $\beta_0 = 0$, the MPV point is $x = 0$. The performance of the designs are shown in Table 6.1. The probability of separation results also include the cases where the designs are inestimable ($\mathbf{y} = \mathbf{0}$ or $\mathbf{y} = \mathbf{1}$ for multi-support point designs).

At $\lambda = 0.5$, the resulting D_{MP}-optimal design has a D-efficiency of 93.68%, while the probability of separation is reduced by 0.21563. This illustrates that the D_{MP}-criterion can yield designs with high D-efficiencies that have a substantially reduced risk of encountering separation. A plot of the D-optimal and D_{MP}-optimal design with $\lambda = 0.5$ is shown in Figure 6.7. The D_{MP}-optimal design has three support

Table 6.1: Performance of the D_{MP}-Optimal Designs, Case 1

Weight (λ)	**P(Separation)**	***D*-Efficiency**
1.0	0.81065	100.00000%
0.9	0.79793	99.90679%
0.8	0.78306	99.54033%
0.7	0.76217	98.72806%
0.6	0.63295	96.99870%
0.5	0.59502	93.67517%
0.4	0.53922	86.42150%
0.3	0.45033	67.73828%
0.2	1.00000	1.82676%
0.1	1.00000	0.36970%
0.0	1.00000	0.00000%

points, while the D-optimal design has two. As observed in the simulation study conducted in Chapter 4, designs with more support points tend to be more robust to separation. As λ initially decreases, a marginal reduction of separation probability occurs in the D_{MP}-optimal designs while very high D-efficiency values are maintained. For $\lambda = 1,\ 0.9,\ 0.8$, and 0.7, two support points are observed, but the support points are drawn more closely to $x = 0$. As λ decreases further, an additional support point is observed in an area where the linear predictor value is closer to zero. As λ decreases to values below 0.3, all runs of the resulting designs are clustered closely around $x = 0$, until $\lambda = 0$ yields a D_{MP}-optimal design with a single support point at approximately $x = 0$. This occurs due to the one-pass point exchange that happens at the end of the two-phase CEA. As λ approaches zero and more weight is placed on the linear predictor penalty term, the design is converging to $x = 0$, the point of MPV. Therefore, the design run with a linear predictor value closest to zero in the Phase II

design matrix after the coordinate exchange becomes the only support point in the design following the point exchange, as this minimizes the linear predictor magnitude penalty term. The D_{MP}-optimal design for $\lambda = 0$ is a single-support point design, and as a result, the two-parameter logistic regression model of Case 1 is inestimable.

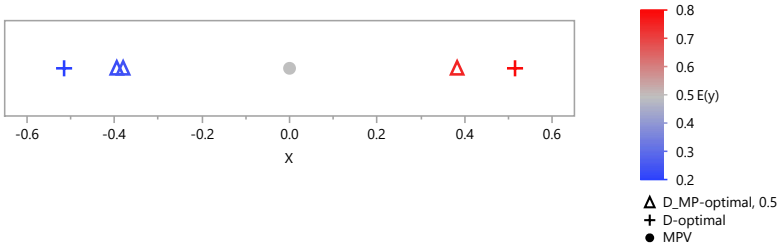


Figure 6.7: Case 1 Design Comparison; D-Optimal vs. D_{MP}-Optimal, $\lambda = 0.5$

6.4.2 Test Case 2: Two-Factor Main Effects Model

For this test case, a two-factor main effect model with $\beta_0 = 0$, $\beta_1 = 2$, and $\beta_2 = 1$ was used. A design size of $n = 10$ was examined. For this model, the MPV region is defined by the line $x_2 = -2x_1$. The performance of the designs are shown in Table 6.2. The probability of separation results also include the cases where the designs are inestimable ($\mathbf{y} = \mathbf{0}$ or $\mathbf{y} = \mathbf{1}$ for multi-support point designs).

At $\lambda = 0.5$, the D_{MP}-optimal design has a D-efficiency of 93.60%, while the probability of separation is reduced by 0.25283. This design has the best balance between D-efficiency and separation robustness of all λ values that were examined. The D_{MP}-optimal design with the lowest probability of separation observed occurs at $\lambda = 0.2$, but the D-efficiency of this design is poor. The D-optimal design has four support points, the D_{MP}-optimal design at $\lambda = 0.5$ has five, and the D_{MP}-optimal design at $\lambda = 0.3$ has seven. A plot of these three designs and the MPV region is shown in Figure 6.8. As observed in case 1, as λ initially decreases from the D-optimal case, the

Table 6.2: Performance of the D_{MP}-Optimal Designs, Case 2

Weight (λ)	**P(Separation)**	***D*-Efficiency**
1.0	0.63140	100.00000%
0.9	0.56588	99.28688%
0.8	0.57686	99.52373%
0.7	0.56134	99.20411%
0.6	0.40146	97.52968%
0.5	0.37857	93.60136%
0.4	0.37745	83.67235%
0.3	0.20645	68.06108%
0.2	1.00000	7.33787%
0.1	1.00000	0.62749%
0.0	1.00000	0.00000%

four support points of the D_{MP}-optimal design are drawn more closely to the MPV region. At $\lambda = 0.6$, the number of support points increases from four to five. As the linear predictor penalty term becomes more significant, the amount of support points tends to increase, as new points are observed closer to the MPV region. However, as λ decreases below 0.3, the support points of the designs are reduced and located very close to the MPV region. At $\lambda = 0$, the D_{MP}-optimal design collapses to a single support point that lies almost directly on the MPV region.

6.4.3 Test Case 3: Two-Factor Main Effects w/Interaction Model

For this test case, a 2FI model with $\beta_0 = -1$, $\beta_1 = 4$, $\beta_2 = 1$, and $\beta_{12} = -2$ was used. A design size of $n = 16$ was examined. For this model, the MPV region is defined by the function $x_2 = (1 - 4x_1)/(1 - 2x_1)$. The performance of the designs are shown in Table 6.3. As observed in the previous test cases, at $\lambda = 0.5$, there is a significant

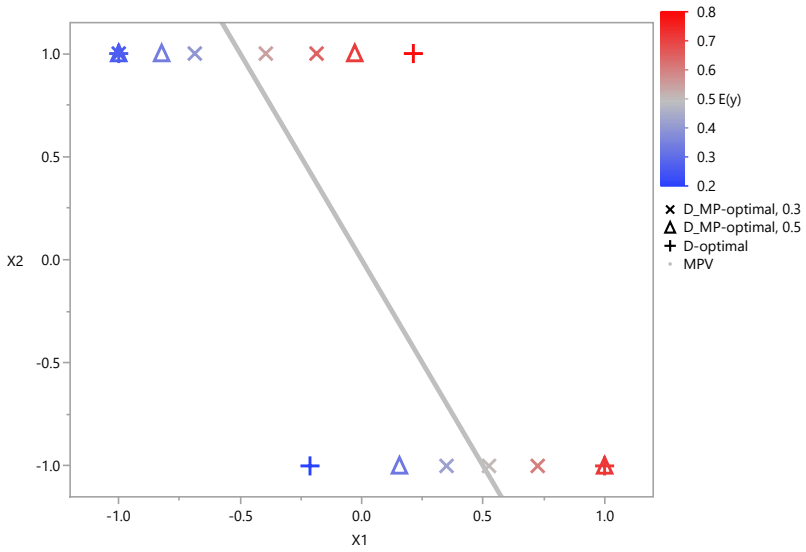


Figure 6.8: Case 2 Design Comparison; D-Optimal vs. D_{MP}-Optimal, $\lambda = 0.3, 0.5$

decrease in the probability of separation in the D_{MP}-optimal design while a very high D-efficiency is maintained.

Table 6.3: Performance of the D_{MP}-Optimal Designs, Case 3

Weight (λ)	**P(Separation)**	***D*-Efficiency**
1.0	0.91320	100.00000%
0.9	0.91680	99.97560%
0.8	0.88320	99.89897%
0.7	0.85160	99.39274%
0.6	0.80760	98.23344%
0.5	0.62520	98.34285%
0.4	0.83080	95.72332%
0.3	0.77360	90.18384%
0.2	0.86000	84.15593%
0.1	1.00000	34.01239%
0.0	1.00000	0.00000%

For this model, the locally D-optimal design has four support points. At $\lambda = 0.6$ and 0.5, the D_{MP}-optimal design has six support points. A plot of the D-optimal design, the D_{MP}-optimal design at $\lambda = 0.5$, and the MPV region is shown in Figure 6.9. However, at $\lambda = 0.4$, four support points that lie closer to the MPV region relative to the locally D-optimal design are observed, which explains the increase in the probability of separation relative to the $\lambda = 0.5$ design. Similar to the first two test cases, as λ decreases below 0.3, the support points of the designs become clustered near the MPV region. At $\lambda = 0$, the D_{MP}-optimal design collapses to a single support point that lies almost directly on the MPV region.

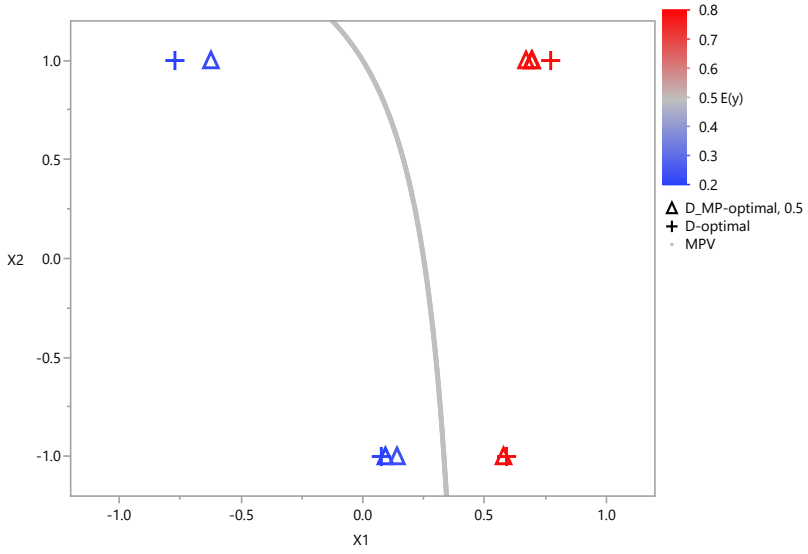


Figure 6.9: Case 3 Design Comparison; D-Optimal vs. D_{MP}-Optimal, $\lambda = 0.5$

6.5 A Methodology for Robust D_{MP}-Optimal Designs

The D_{MP}-criterion requires estimates for $\boldsymbol{\beta}$ to determine both the relative D-efficiency of the candidate design and the squared linear predictor penalty term. The problem of *design dependence* (Khuri *et al.*, 2006) that is typical for GLMs is exacerbated in implementing D_{MP}-optimality, as initial parameter inputs are required for

both components of the criterion. Consider the following example, where a two-factor main effects logistic regression model with $\boldsymbol{\beta} = [-0.06, 2.64, 1.18]$ is used. Suppose that the experimenter's initial parameter estimates are $\hat{\boldsymbol{\beta}} = [-0.71, 4.08, 0.56]$; the experimenter overestimates the impact of Factor 1 and underestimates the impact of Factor 2 on the response. For a design size of $n = 10$ and with $\lambda = 0.5$, the locally D_{MP}-optimal design for $\hat{\boldsymbol{\beta}}$ has a 48% D_{MP}-efficiency relative to the optimal design for the true parameter values, where D_{MP}-efficiency is calculated as shown in Equation 6.9.

$$D_{MP_{\text{eff}}} = \frac{\Psi_{D_{MP}}(\xi, \boldsymbol{\beta})}{\Psi_{D_{MP}}(\xi^*, \boldsymbol{\beta})} \tag{6.9}$$

The support points of both designs are shown in Figure 6.10. Consequently, robust approaches for constructing D_{MP}-optimal designs must be explored, as the criterion is sensitive to the initial parameter inputs.

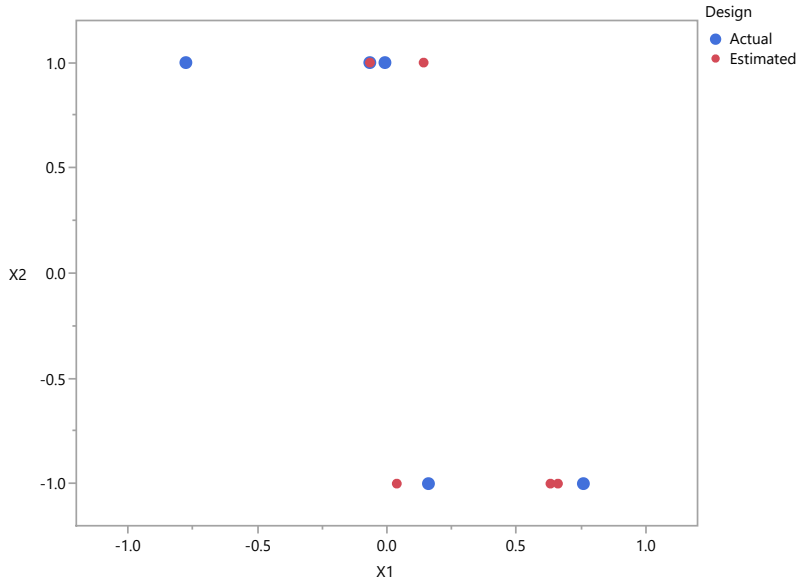


Figure 6.10: D_{MP}-Optimal Design Sensitivity Comparison, $\lambda = 0.5$

Several robust design methods for GLMs have been proposed in the literature. The standardized maximin method (Chipman and Welch, 1996) involves specifying

a collection $\mathbf{B}$ of plausible parameter values, $\mathbf{B} = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_J\}$. For each $\boldsymbol{\beta}_j$, a locally optimal design, ξ_j is determined. The relative efficiency of ξ_j is calculated over $\mathbf{B}$ for designs $j = 1, ..., J$, and the design that maximizes the minimum efficiency is designated as the robust design. Another method involves clustering the support points of a set of locally optimal designs across different plausible parameter values (Dror and Steinberg, 2006). In both of these methods, sets of locally optimal designs are generated to determine the robust design. Bayesian methods, where a prior distribution $f(\boldsymbol{\beta})$ is specified, have also been used to construct robust designs for GLMs. However, to facilitate expedient construction of exact robust designs via an exchange algorithm, certain assumptions on $f(\boldsymbol{\beta})$ must be made. For example, to construct Bayesian D-optimal designs for GLMs, Gotwalt *et al.* (2009) utilize a numerical integration approximation method to expediently evaluate the Bayesian D-optimality functional, shown in Equation 3.11. Direct use of the Bayesian optimality function in an exchange algorithm is computationally expensive and typically not practical for many design applications.

For this study, the maximin method is adopted for D_{MP}-optimality. The maximin method presents the most straightforward application of the two-phase CEA, as the current functional for D_{MP}-optimality (Equation 6.4) is formulated for point estimates of $\boldsymbol{\beta}$, yielding locally optimal designs. Additionally, the maximin approach only requires that plausible parameter ranges be specified; a prior distribution for each parameter is not required. From a practical perspective, especially when working with subject matter experts that are not well-versed in statistics and experimental design, determining parameter ranges is preferred over specifying prior distributions. The robust design for maximin D_{MP}-optimality for a fixed λ maximizes

$$\min \left\{ \frac{\Psi_{D_{MP}}(\xi, \boldsymbol{\beta})}{\Psi_{D_{MP}}(\xi[\boldsymbol{\beta}], \boldsymbol{\beta})} \middle| \boldsymbol{\beta} \in \mathbf{B} \right\}, \tag{6.10}$$

where $\xi[\boldsymbol{\beta}]$ is the locally D_{MP}-optimal design for a fixed $\boldsymbol{\beta}$ and $\mathbf{B}$ is the space of all plausible $\boldsymbol{\beta}$ values.

6.5.1 Generating a Robust D_{MP}-Optimal Design

To illustrate an example of determining a robust D_{MP}-optimal design via the maximin method, a two-factor main effects logistic regression model and a design size of $n = 10$ is used. For this example, λ is fixed at 0.5, as this value of λ was observed to yield locally optimal designs with high D-efficiencies and substantially reduced separation probabilities in the test cases performed in Section 6.4. Although this example involves arbitrarily specified run size, model form, and a fixed λ, this methodology can be extended to any n-run experiment using any form of the logistic regression model and any fixed value of $\lambda \in [0, 1]$. Suppose that the plausible ranges for each parameter are as shown in (6.11).

$$\begin{aligned} -1 &\leq \beta_0 \leq 1 \\ 1 &\leq \beta_1 \leq 5 \\ 0.5 &\leq \beta_2 \leq 1.5 \end{aligned} \tag{6.11}$$

These parameters are the coefficients in the linear predictor for the intercept, Factor 1, and Factor 2, respectively. To determine the maximin D_{MP}-optimal design, 100 parameter vectors are randomly sampled from continuous uniform distributions over the specified ranges for each parameter in (6.11); $\mathbf{B} = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{100}\}$. A locally D_{MP}-optimal design is generated for each $\boldsymbol{\beta}_j \in \mathbf{B}$ using the two-phase CEA, and the relative D_{MP}-efficiency of each design is calculated over $\mathbf{B}$ as shown in Equation 6.9. The locally optimal design with the highest worst-case relative D_{MP}-efficiency is designated as the robust design. The randomly sampled parameter vectors used for design generation are available in Appendix C.1.

The locally D_{MP}-optimal design constructed with $\boldsymbol{\beta} = [-0.083, 2.787, 1.477]$ was designated as the robust design for this example, as it had a worst-case D_{MP}-efficiency of 0.15382, which was the maximum worst-case value observed for all 100 designs over $\mathbf{B}$. The support points of the robust design and a locally D-optimal design for these parameters are shown in Figure 6.11. Relative to the D-optimal design, the robust design has a D-efficiency of 97.47%.

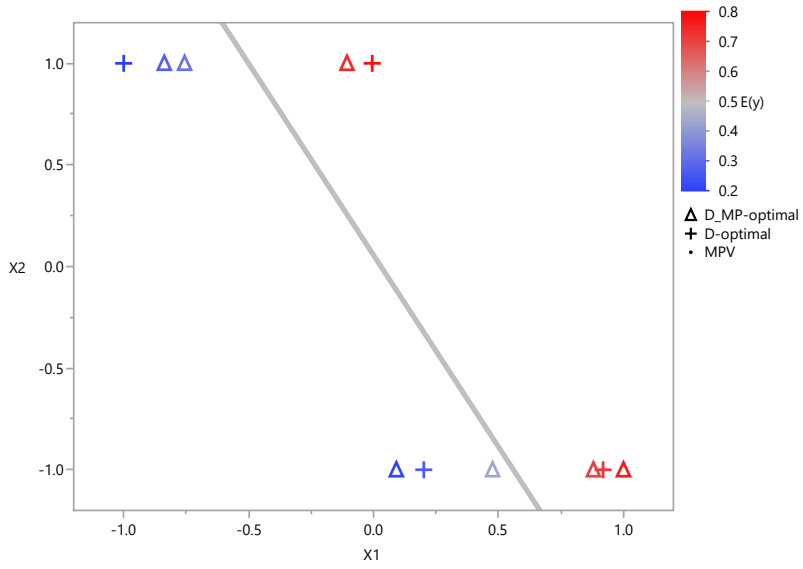


Figure 6.11: Robust D_{MP}-Optimal vs. Locally D-Optimal Design Comparison, $\boldsymbol{\beta} = [-0.083, 2.787, 1.477]$

In comparison to the other locally D_{MP}-optimal designs, the robust design performs well in terms of D-efficiency over $\mathbf{B}$, where the D-efficiencies are calculated relative to 100 locally D-optimal designs constructed for each $\boldsymbol{\beta}_j \in \mathbf{B}$. The mean, median, and standard deviation of the D-efficiencies of the robust design are 83.317%, 84.893%, and 9.470%, respectively. For the remaining 99 D_{MP}-optimal designs, the mean, median, and standard deviation are 73.569%, 76.575%, and 18.029%, respectively. Comparative box plots of the D-efficiencies for the robust design and the remaining D_{MP}-optimal designs are shown in Figure 6.12.

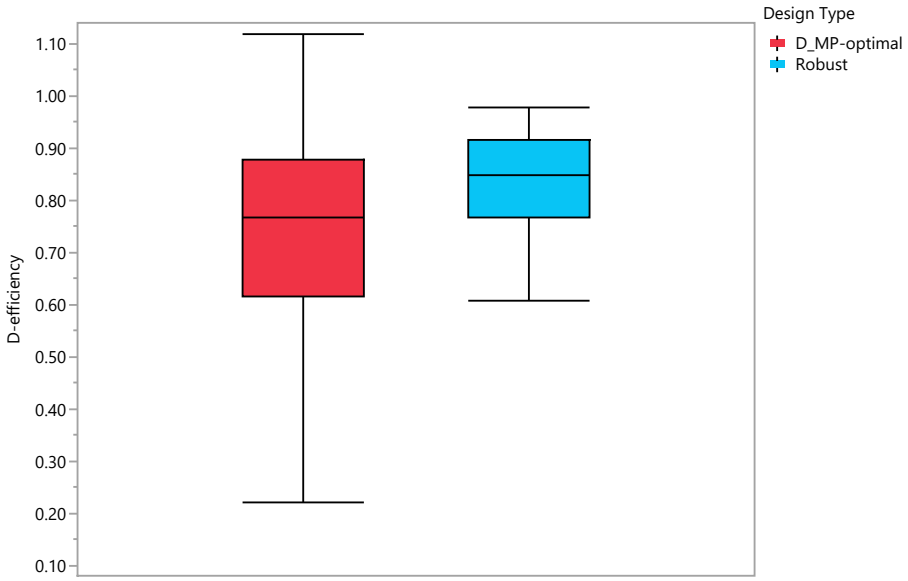


Figure 6.12: D-Efficiency over Sampled Parameter Vectors for the Locally D_{MP}-Optimal Designs

6.5.2 Performance of the Robust D_{MP}-Optimal Design

The performance of the robust D_{MP}-optimal design was evaluated against an equally sized (n = 10) Bayesian D-optimal design generated using the nonlinear design platform of JMP Pro 14.2®. The prior parameter distributions of the Bayesian D-optimal design have the ranges shown in (6.11). The support points of both designs are shown in Figure 6.13. To evaluate both designs, an additional 100 parameter vectors are randomly sampled from the same continuous uniform distributions over the ranges shown in (6.11); $\mathbf{B}_{\text{test}} = \{\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{100}\}$. The D-efficiency of both designs is evaluated for each $\boldsymbol{\beta}_j \in \mathbf{B}_{\text{test}}$ relative to a locally D-optimal design generated for $\boldsymbol{\beta}_j$. The probability of separation for both designs is also determined for each $\boldsymbol{\beta}_j \in \mathbf{B}_{\text{test}}$. For every parameter vector, each design's probability of separation is calculated by enumerating all 1024 possible response vectors, $\mathbf{y} \in \{0, 1\}^{10}$. Separation is declared in

a response vector if the maximum parameter estimate variance is greater than 5000 (Heinze and Schemper, 2002). The total probability of separation is then calculated as shown in Equation 4.4, where the probability of realizing each response vector is determined using $\boldsymbol{\beta}_j$. The parameter vectors used in the performance evaluations are available in Appendix C.2.

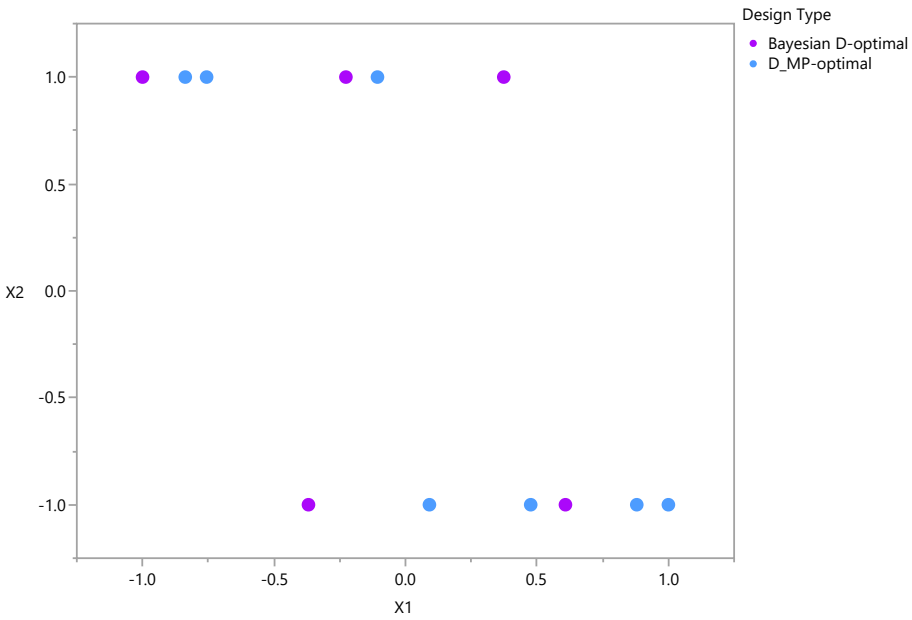


Figure 6.13: Support Point Comparison of the Robust Designs

In terms of D-efficiency over $\mathbf{B}_{\text{test}}$, the robust D_{MP}-optimal design fares well in comparison to the Bayesian D-optimal design. As expected, the performance of the Bayesian D-optimal design is better, but only marginally. The mean, median, and standard deviation of the D-efficiencies of the robust D_{MP}-optimal design are 83.141%, 84.093%, and 8.614%, respectively. For the Bayesian D-optimal design, the mean, median, and standard deviation are 84.334%, 85.644%, and 7.425%, respectively. The worst-case D-efficiency observed for the D_{MP}-optimal and Bayesian D-optimal design is 62.086% and 67.458%, respectively. Box plots comparing the robust designs are shown in Figure 6.14.

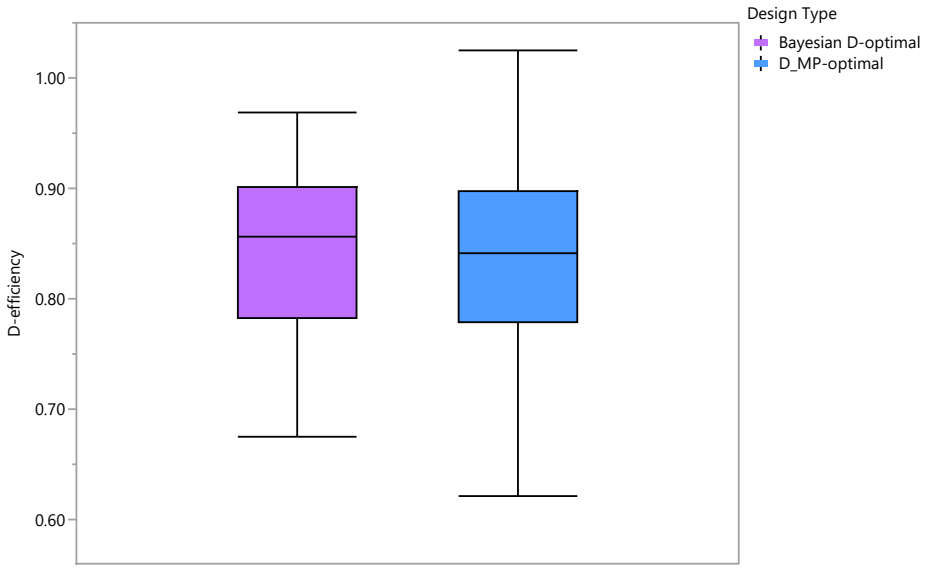


Figure 6.14: D-Efficiency over the Test Parameter Vectors for the Robust Designs

For the reduction in D-efficiency, the robust D_{MP}-optimal design offers a substantial reduction in the probability of separation over the Bayesian D-optimal design. Over $\mathbf{B}_{\text{test}}$, the mean, median, and standard deviation of the probabilities of separation for the Bayesian D-optimal design are 0.684, 0.685, and 0.140, respectively. For the robust D_{MP}-optimal design, the mean, median, and standard deviation are 0.452, 0.402, and 0.189, respectively. The worst-case probability of separation observed for the Bayesian D-optimal and D_{MP}-optimal design is 0.908 and 0.831, respectively. For all 100 parameter vectors in $\mathbf{B}_{\text{test}}$, the robust D_{MP}-optimal design outperformed the Bayesian D-optimal design in terms of separation probability. The mean, median, and standard deviation of the probability of separation reduction in the robust D_{MP}-optimal design is 0.232, 0.239, and 0.103, respectively. In other words, over $\mathbf{B}_{\text{test}}$, the robust D_{MP}-optimal design offers an average reduction in separation probability of 23.160% relative to the Bayesian D-optimal design. The Bayesian D-optimal design has five support points, while the robust D_{MP}-optimal design has seven. Therefore, a

reduction in separation probability is expected for the D_{MP}-optimal design based on the relationship between number of support points and separation probability that was observed in the Chapter 4 simulation study. The separation data for both designs and the separation probability reduction is shown in Figure 6.15. The robust design study demonstrates that the robust D_{MP}-optimal design offers a substantial reduction in separation risk with a slight decrease in D-efficiency performance relative to a Bayesian D-optimal design of equal size generated with the same logistic regression model and prior parameter distributions.

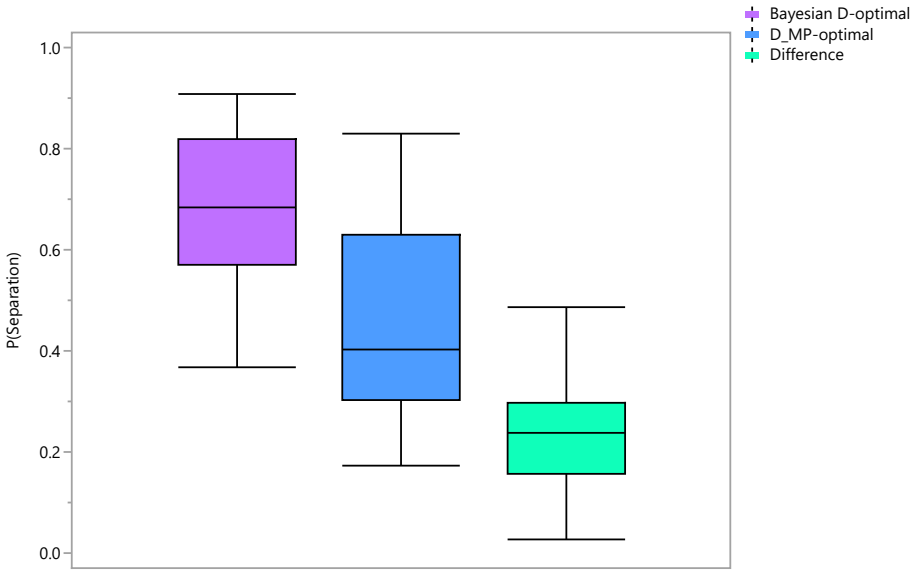


Figure 6.15: Probability of Separation over the Test Parameter Vectors for the Robust Designs and Separation Probability Reduction with the D_{MP}-Optimal Design

6.6 Summary

The simulation study in Chapter 4 demonstrated that small-sample D-optimal designs for several forms of the logistic regression model are prone to encountering separation. The design augmentation simulation study presented in Chaper 5

showed that placing additional runs in regions of MPV efficiently eliminates separation in an initial design. However, the MPV-augmented designs have relatively low D-efficiencies. Based on these observations, a compound criterion, D_{MP}, is proposed to balance D-efficiency with separation robustness, as a separation-optimal design consisting of support points solely located about the MPV region will have very poor D-efficiency. The D_{MP}-criterion has two components: D-efficiency and a normalized squared linear predictor penalty term. A candidate design point $\mathbf{x}$ will have $E(y|\mathbf{x}, \boldsymbol{\beta}) = 0.5$ if and only if $\mathbf{x}^T\boldsymbol{\beta} = 0$. Therefore, the average squared linear predictor value of a design of size n, given by $\sum_{i=1}^{n}(\mathbf{x}_i{}^T\boldsymbol{\beta})^2/n$, provides a measure of the design's average distance of the support points from the separation-optimal (MPV) region. To put this penalty term on the same scale as the D-efficiency measure, the maximum squared linear predictor value over the entire design region is used to normalize the average squared linear predictor value. Furthermore, to adjust for model size, the normalized squared linear predictor value is raised to the $(1/p)$th power, similar to how the ratio of D-criterion values used to calculate D-efficiency is adjusted. A user-defined weight value, $\lambda \in [0, 1]$, is used to adjust for separation risk tolerance. At $\lambda = 1$, the D_{MP}-criterion is equivalent to the standard D-criterion for GLMs. At $\lambda = 0$, the D_{MP}-criterion will force all design points into the MPV region. This chapter demonstrates that the D_{MP}-criterion can be incorporated in an exchange algorithm to create exact compromise optimal designs for the logistic regression model that have high D-efficiencies and a substantially reduced risk of encountering separation relative to a D-optimal design of equal run size.

A two-phase CEA is presented to implement D_{MP}-optimality in exact designs. Several logistic regression models and fixed design sizes are used to demonstrate that for $\lambda \in [0.4, 0.6]$, the locally D_{MP}-optimal designs have relatively high D-efficiencies while offering substantial reductions in separation probability over a locally D-optimal

design of equal size. It is observed that the locally D_{MP}-optimal designs tend to have more support points that the locally D-optimal designs, a characteristic that was observed to reduce a design's probability of separation in Chapter 4. To address the design dependence issue inherent in GLMs, a maximin approach is used to obtain a robust D_{MP}-optimal design that retains competitive D-efficiencies across a range of potential parameter values while offering a significant reduction in separation probability over a candidate Bayesian D-optimal design generated with the same parameter ranges (prior distributions).

In practice, exact D_{MP}-optimal designs can be expediently constructed for use in experimental situations where a binary response is of interest. If no additional experimentation will be permitted, the D_{MP}-optimal designs may provide an attractive alternative to standard D-optimal designs for the logistic regression model, as very high D-efficiencies can be maintained while reducing the risk of encountering separation. To address uncertainty in the initial parameter specifications, the maximin methodology is a practical way to generate a robust D_{MP}-optimal design over a range of potential parameter values.

Chapter 7

CONCLUSION

This research addressed the topic of separation, an estimation problem that can occur in experimental designs where the logistic regression model is used to relate a set of experimental factors with a binary response. In the presence of separated response data, maximum likelihood estimation will not provide finite estimates for the parameters of the logistic regression model. A survey of the current literature reveals that while there has been substantial research on usable estimation methods under separated data, there has not been significant work dedicated to addressing separation in the design phase of an experiment, or a detailed exploration of the separation problem in multi-factor experimental designs. In addition, there are no recommendations for alleviating the risk of separation in an initial design or follow-up phases of an experiment.

In Chapter 4, the susceptibility of two-factor locally and Bayesian D-optimal designs to separation is explored. Three forms of the logistic regression model are used as the base models to generate locally and Bayesian D-optimal designs of four distinct design sizes (n = $8, 16, 32$, and 64). A Monte Carlo simulation methodology was used to generate large sets of response data for each design. The criterion for determining the presence of separation follows the recommendation by Heinze and Schemper (2002), where the non-convergence of the Newton-Raphson algorithm in solving the score equations of the likelihood function is detected by the large magnitude ($\geq \sqrt{5000}$) of the maximum standard error of the parameter estimates. The proportion of separated response sets serves as the estimated probability of separation for each design. Based on this methodology, it was discovered that small-sample

($n \leq 32$) D-optimal designs are subject to encountering separation. Additionally, it was determined that the probability of separation is higher for constructed designs for logistic models with active interaction and quadratic terms, possibly due to the curvature of the response surface over the design region. In other words, the interaction and full quadratic models had induced regions of moderate probabilities ($0.15 \leq \pi \leq 0.85$) with nonlinear boundaries. As observed by Dror and Steinberg (2005), the support points of a locally D-optimal design tend to lie on the boundaries of the moderate response probability region, and the resulting support point patterns of the designs from the models with higher-order terms tended to be more susceptible to separation. It was also observed that the Bayesian D-optimal designs tended to be less prone to separation than the locally D-optimal designs, as the Bayesian designs tend to have more support points than their locally optimal counterparts, making the existence of a separating hyperplane in the design region less likely. Most notably, it was found that for the logistic models with active interaction and quadratic terms, designs of equal size that are generated from the same model can have significantly different separation probabilities. As the model response surface becomes more complex, designs from the same model are noticeably more disparate, as the support points of the designs collectively occupy a larger region of the design space.

The results of the D-optimal design separation study have some practical implications for experimenters in practice. First, the robust (Bayesian) designs are preferable in most design applications. In addition to accounting for the uncertainty inherent in the initial parameter specification required for design creation in GLMs, the increase in design support points is a desirable property for reducing the probability of separation. This may present an inconvenience while executing an experiment, as more distinct factor combinations will have to be planned, randomized in the run order, and executed. However, in situations where this can be accommodated, it is advisable

to do so for the reduction in separation risk. Secondly, when planning an optimal experimental design that utilizes a logistic regression model, it will be prudent to generate a candidate set of exact D-optimal designs. These designs are expedient to construct for GLMs (assuming the prior information can be adequately summarized by certain distributions), and the simulation methodology used in Chapter 4 can be used to estimate the probability of separation in each design. Assuming that each design has comparable D-efficiencies, the design with minimal separation risk should be selected as the final design.

To break separation in small-sample D-optimal designs, Chapter 5 investigates strategies for augmenting an initial design with additional runs. Two practical strategies were proposed: (1) locating augmented runs in regions of maximum prediction variance (MPV), and (2) replicating runs at the support points of the initial D-optimal design. The first strategy entails augmenting runs in regions where the predicted probability of either category of the response is 0.5. The second strategy randomly locates augmented runs at the support points of the initial design. These strategies were evaluated on the separation probability and the D-efficiency of the augmented design. To serve as a baseline for these two strategies, random augmentation was also performed, where the coordinates of each augmented design point are randomly sampled from a continuous uniform distribution over the factor range. An initial design size of $n = 8$ was examined, and design runs were augmented according to each strategy in blocks of size $n = 8$. A Monte Carlo simulation was also used to generate response data. The same base logistic regression models used in Chapter 4 were adopted for this study. For each initial design, response data was randomly sampled from a Bernoulli process until a separated response set was found. Then, a large number of augmentation trials were performed, where a block constructed based on each strategy was added to the initial design. Response data for the augmented runs

was also simulated from the same Bernoulli process. The proportion of augmentation trials that resulted in a separated response set served as the estimated probability of separation reported for each initial augmented design. Additionally, the D-efficiency of each augmented design was calculated relative to a 16-run locally D-optimal design generated for the parameter vectors of each logistic regression model.

The results of this study show that compared to initial support point replication and random augmentation, MPV augmentation most reliably eliminates separation that is encountered in an initial design. The blocks constructed with runs located in the MPV region of the design space produced the lowest proportion of separated response data. However, the MPV augmented designs tended to have lower D-efficiencies relative to the other two strategies. Initial support point replication produced designs with high D-efficiencies, but did not offer significant advantages in terms of reducing separation risk relative to the other two strategies. The parameter standard errors produced with Firth likelihood (Firth, 1993) estimation in a separated D-optimal design had lower standard errors than the maximum likelihood estimates produced through an MPV augmented design with overlapped (non-separated) response data, implying that the precision of the estimates available with an MPV-augmented design, even with overlapped data, will be low.

MPV augmentation has practical limitations. It is reliant on the initial parameter specification that is used in design creation. It is therefore not robust to parameter uncertainty. The simulation study conducted in Chapter 5 represents an ideal scenario, where it is assumed that the initial parameter estimates used in design creation match the true parameter values. Although this is not likely in practice, it does show that conceptually, locating runs in the MPV region of a design space produces designs that are much less likely to encounter separation than an equally-sized D-optimal design.

The trade-off between D-efficiency and separation robustness is notable; placing design runs in the MPV region provides a reduced risk of encountering separation at the expense of D-efficiency. Based on this observation, a compound design criterion for constructing compromise designs was presented in Chapter 6. The compound D_{MP} statistical criterion has two normalized components: (1) D-efficiency and (2) a penalty term driven by the average distance of the candidate design's support points from the MPV region. The D-efficiency of the candidate design is calculated relative to a locally D-optimal design for the specified parameters, and the average distance penalty term is captured though the average squared linear predictor value of the design runs, as a linear predictor value of zero holds if and only if the design point is located in the MPV region. To normalize the average squared linear predictor value, it is divided by the maximum squared linear predictor value over the design region, which ensures that the normalized value is in the range [0,1]. A user-defined weight, $\lambda \in [0, 1]$, is used to define the D-efficiency vs. separation risk trade-off. As the weight value initially decreases, the separation robustness of the candidate design increases relative to a locally D-optimal design of equal size from the same logistic model.

Exact D_{MP}-optimal designs are generated via a two-phase coordinate exchange algorithm (CEA). In the first phase, a locally D-optimal design for the specified model parameters is generated. In the second phase, a locally D_{MP}-optimal design is generated, where the D-criterion value of the Phase I design is used to calculate the D-efficiency of the candidate designs. The two-phase CEA was validated by configuring it to generate locally D-optimal designs and comparing the results to the designs generated by the nonlinear design platform in JMP Pro 14.2®. For several model forms and fixed run sizes, it was demonstrated that the two-phase CEA produced locally D-optimal designs that were practically identical to the designs produced by the nonlinear design platform. To assess the performance of locally D_{MP}-optimal

designs, several test cases are used to demonstrate that for weight values between 0.4 and 0.6, the locally D_{MP}-optimal designs have D-efficiencies over 90% while reducing the probability of separation by 20-30% in comparison to the locally D-optimal design. An ideal weight value is likely model-dependent, and determining an optimal weight value to generate locally D_{MP}-optimal designs is left for future work. Maximin D_{MP}-optimal designs were also proposed to address the design dependence problem in GLMs. A set of parameter vectors are randomly sampled from a continuous uniform distribution over plausible parameter ranges, and a locally D_{MP}-optimal design with fixed weight is generated for each parameter vector. The D_{MP}-optimal design with the highest worst-case D_{MP}-efficiency, calculated relative to all of the other parameter vectors in the set, is designated as the robust design. A case study shows that the robust D_{MP}-optimal design retains reasonable D-efficiencies over all parameter vectors in the set, and has a substantially reduced probability of separation relative to a corresponding Bayesian D-optimal design.

In summary, the contributions of this research are as follows:

1. An exploration of the separation problem for exact, multi-factor D-optimal designs for the logistic regression model, where a simulation methodology is used to show that small-sample ($n \leq 32$) D-optimal designs are prone to separation

2. An efficient strategy for augmenting design runs to eliminate separation in multi-factor experimental designs, where a simulation study demonstrates that MPV augmentation most reliably eliminates separation in an initial D-optimal design

3. A compound optimality criterion, D_{MP}, that is implemented in a novel two-phase CEA to generate exact, multi-factor locally D_{MP}-optimal designs for the logistic regression model with reduced separation risk relative to D-optimal

designs

4. A robust design methodology to declare minimax D_{MP}-optimal designs to address the design dependence problem in GLMs

The proposed methodologies used in this research present limitations that should be addressed in future work. The separation simulation study in Chapter 4 was restricted to D-optimal designs. Therefore, the augmentation strategies presented in Chapter 5 are limited to initial D-optimal designs. In the test cases for the locally D_{MP}-optimal designs presented in Chapter 6, a comprehensive study of the impact of the user-defined weight value was not accomplished. Therefore, the weight values that were recommended and used in the case study are limited. For the test cases that were presented, a weight value of 0.5 produced D_{MP}-optimal designs with the best balance between D-efficiency and separation robustness, but it is not clear if this weight value recommendation will extend well to higher-order models or additional design factors. The D_{MP}-criterion formulation is easily extended to more than two design factors, and the two-phase CEA can be easily configured to accommodate more than two design factors. It is hypothesized that the D_{MP}-criterion will work well for more than two design factors, assuming the model orders are restricted to active two-factor interactions and quadratic terms. How well it will perform for more complex model forms is left for future research. However, a full quadratic model is adequate as an approximating polynomial function in most response surface design applications, so the current model order scope for the D_{MP}-criterion is not overly prohibitive. All of the logistic regression models used in this study were restricted to two design factors, and it is not currently clear how separation risk for D-optimal designs and the augmentation strategy results will extend to additional design factors. Finally, the large parameter standard error criterion recommended by Heinze and Schemper

(2002) was used to declare separation in all of the simulation studies presented in this research. Alternative methods of detecting separation are currently available in the literature and should be explored in future work.

Binary responses are frequently encountered across a broad spectrum of research applications, such as biomedical development, econometrics, survey data, and military developmental and operational testing for the acquisition of new systems. Experimentation is a key component of many of these studies, and this research will be directly applicable in costly experimental work where a dichotomous non-numerical outcome is of interest. D-optimal designs are frequently used in practice, as they are expedient to construct and tend to produce precise parameter estimates. The simulation methodology used in Chapter 4 can be adopted to assess the risk of encountering separation in a small-sample D-optimal design. A candidate set of D-optimal designs can be considered, where the design with the lowest estimated separation probability can be used. Additionally, a D_{MP}-optimal design may be an attractive alternative to a standard D-optimal design if it is known that additional experimentation is infeasible. Several promising directions for this research will be proposed as areas for future study in the next chapter.

Chapter 8

AREAS AND ISSUES FOR FURTHER STUDY

Several methods for addressing the problem of separation in the design phase of an experiment were presented in this research. However, these proposed methodologies present limitations that warrant further research. In addition, from a design perspective, very little has been done on the problem of separation. In this chapter, some areas for future research are discussed.

There has been recent development in construction methods for G- and I-optimal designs for GLMs. Saleh and Pan (2016) propose a clustering approach to generating robust G-optimal designs, and Li and Deng (2018) propose a multiplicative algorithm to construct continuous I-optimal designs. Both of these optimality criteria are focused on minimizing the variance of prediction over a design region. G-optimal designs minimize the maximum prediction variance over a design region, while I-optimal designs minimize the average prediction variance over a design region (Goos, 2012). In applications where the experimenter desires to make response predictions, these designs will be more useful than the D-optimality criterion. A potential direction for future research is to investigate whether exact G- or I-optimal designs for the logistic regression model offer any advantage over D-optimal designs in terms of separation risk. A simulation study using the methodology presented in Chapter 4 can be repeated for a candidate set of exact G- or I-optimal designs of fixed sizes to assess the separation probabilities of these designs. If it is found that an alternative criterion offers a clear advantage, it may be advisable to use these designs in future applications where experimental resources are heavily constrained.

The E-criterion should also be investigated as a viable solution to mitigating

separation risk. As discussed in Section 3.2.1, an E-optimal design minimizes the maximum variance of the parameter estimates of all n-run designs over a candidate design space. For nonlinear statistical models, research on E-optimal designs is sparse in the literature. However, for linear statistical models, it has been observed that the number of support points for continuous E-optimal designs can be large relative to D-optimal designs. For example, for second-order response surface models over design space $[-1, 1]^k$, Dette and Grigoriev (2014) prove that the number of support points of an E-optimal design is given by

$$N(r_1, r_2, r_3) = \sum_{i=1}^{3} \binom{k}{r_i} 2^{k-r_i}, \tag{8.1}$$

where r_1, r_2, and r_3 are integers such that $0 \leq r_1 \leq r_2 \leq r_3 \leq k$. The authors note that the number of support points of E-optimal designs for second-order linear models is unusually large, and techniques are proposed to search for viable designs with a minimal number of support points. However, the simulation study in Chapter 4 illustrated that exact D-optimal designs with a larger number of support points tend to be less prone to separation. The large number of support points of E-optimal designs may therefore be an attractive property from the perspective of minimizing separation probability. The construction of continuous E-optimal designs becomes complicated if the multiplicity of the minimum eigenvalue of the optimal information matrix is larger than one, which is the case for second-order linear models (Dette and Grigoriev, 2014). Construction of exact E-optimal designs for the logistic regression model can be implemented via a coordinate-exchange algorithm, and investigating these designs as a potential solution to the separation problem observed in small-sample D-optimal designs is left for future work.

In terms of an augmentation strategy for follow-on experimentation, maximum prediction variance (MPV) augmentation as it is presented in Chapter 5 has limited

practical value because of the initial parameter specification required to accurately determine the MPV region. A study of the robustness of MPV augmentation to parameter value uncertainty is recommended, as it is currently not an attractive methodology if the experimentation involves a process or system that is not well understood, which may result in increased uncertainty in the parameter values. For two-factor design spaces involving the logistic regression models that were specified in this study, MPV augmentation only slightly outperformed random augmentation in eliminating separation. Determining how MPV augmentation fares relative to random augmentation in additional dimensions is left to future work. The necessity of augmentation should also be further examined to determine if eliminating separation is a worthwhile pursuit, as Firth likelihood estimation (Firth, 1993) for the logistic regression model has been shown to produce usable parameter estimates in a separated response set.

A more comprehensive study of the weight (λ) values used in the D_{MP}-criterion is recommended, as determining an optimal weight value across varying design sizes and forms of the logistic regression model has not been explored. For the weight values that were examined in the locally optimal test cases presented in Chapter 6, it was observed that D_{MP}-optimal designs generated at $\lambda = 0.5$ had the best balance between D-efficiency and reduction in separation probability. Optimizing this weight value in a variety of design applications is recommended and left for future work.

Methods to modify the D_{MP}-criterion formulation in the presence of parameter uncertainty should be explored. In its current formulation, discussed in Chapter 6, the linear predictor penalty term is entirely dependent on the initial parameter specification. As a result, the D_{MP}-criterion is extremely sensitive to parameter misspecification. An on-going work by the author is to modify the D_{MP}-criterion to make it a more practical alternative to the standard D-criterion. Bayesian methods

may potentially be used to adjust the D_{MP}-criterion to incorporate prior parameter distributions. Modifying the criterion to produce Bayesian D_{MP}-optimal designs is left for future work. Another interesting extension to this work is to investigate augmentation strategies for D_{MP}-optimal designs in the case that these designs still encounter separation. In addition to adopting the strategies that were proposed in Chapter 5 for D_{MP} designs, another potential augmentation strategy is to find new design points using the D_{MP}-criterion with an adjusted weight value. Establishing an optimal augmentation strategy for D_{MP}-optimal designs is also left for future work.

A challenging but practical extension to this research is to generalize the separation problem and the implementation of D_{MP}-optimality to nominal response variables with g possible outcomes (groups), where $g > 2$. The general case for separation is presented by Albert and Anderson (1984). Suppose that H denotes the response variable which takes values that can be coded as arbitrary integers $1, ..., g$. The conditional probabilities of group membership have the extended logistic form (Anderson, 1972):

$$P(H = s|\mathbf{x}) = \exp(\boldsymbol{\beta}_s^T \mathbf{x}) \cdot P(H = g|\mathbf{x}) \text{ for } s = 1, ..., g-1$$
$$P(H = g|\mathbf{x}) = \{1 + \textstyle\sum_{s=1}^{g-1} \exp(\boldsymbol{\beta}_s^T \mathbf{x})\}^{-1}, \tag{8.2}$$

where $\boldsymbol{\beta}_s^T = [\beta_{s0}, ..., \beta_{sp}]$, for $s = 1, ..., g-1$ and $\boldsymbol{\beta}_g^T = \mathbf{0}$. In a logistic classification model, design point $\mathbf{x}$ is classified as belonging to group s if and only if $(\boldsymbol{\beta}_s - \boldsymbol{\beta}_t)^T \mathbf{x} \geq 0$ for all $t = 1, ..., g$ except $t = s$. The model parameter vector is therefore $\boldsymbol{\beta}^T = [\boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_g^T]$. If a vector $\boldsymbol{\beta}$ exists among n multinomial sample points such that for all $i \in E_j$ and for $j, t = 1, ..., g$ $(j \neq t)$,

$$(\boldsymbol{\beta}_j - \boldsymbol{\beta}_t)^T \mathbf{x}_i \geq 0, \tag{8.3}$$

then some form of separation is present (Albert and Anderson, 1984; Konis, 2007). Therefore, to implement D_{MP}-optimality for a nominal response variable with $g > 2$

possible outcomes, the linear predictor penalty term in Equation 6.4 must be modified to include $g-1$ penalty terms with respect to $\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_{g-1}$. The generalized formulation of D_{MP}-optimal designs for a multinomial response variable is left for future work.

A critical component of this research has been detecting the presence of separation. The large parameter standard error criterion recommended by Heinze and Schemper (2002) has been used throughout this research for several reasons. First, this method is readily available in any statistical software package that implements logistic regression, even if there is not a built-in check for separation. Second, it is expedient, and an efficient method of detecting separation was required for all of the simulated response data that was generated for each design used in the simulation studies presented in Chapters 4 and 5. However, alternative methods for detecting separation have been proposed in the literature. For example, Konis (2007) proposes an algorithm for detecting separation that is based on a linear programming approach, which was discussed in Section 2.5. A valuable extension of this research would be to implement different methods of separation detection in obtaining the results presented in Chapters 4 through 6. The main challenge in these alternative approaches will be matching the efficiency of the large parameter standard error criterion, as this criterion only requires fitting a logistic regression model to the experimental data. For example, for the linear programming approach presented by Konis (2007) to roughly match the amount of time required to fit a logistic regression model using iteratively reweighted least squares, the linear program in (2.25) must be converted into a reduced linear program, and the feasibility of the dual linear program to (2.25) is assessed to declare separation in the response data. Furthermore, issues such as degeneracy and cycling, along with numerical difficulties in certain case-specific transformations of the original linear program formulation in (2.25) must be addressed (Konis, 2007).

The research presented in this dissertation serves as a first-look at addressing the

separation problem in the design phase of exact, multi-factor designs. As a result, all of the designs and strategies that were examined are restricted to two-factor design spaces. A practical addition to this work is to extend all results to higher-order logistic regression models with more than two factors. In many experimental situations, there will be more than two factors of interest, and it would be prudent to determine how these results extend to more than two dimensions.

Abdelbasit, K. M. and R. L. Plackett, "Experimental design for binary data", Journal of the American Statistical Association **78**, 381, 90–98 (1983).

Abrahantes, J. C. and M. Aerts, "A solution to separation for clustered binary data", Statistical Modelling **12**, 1, 3–27 (2012).

Agresti, A., *Categorical Data Analysis* (John Wiley & Sons, 2013).

Albert, A. and J. A. Anderson, "On the existence of maximum likelihood estimates in logistic regression models", Biometrika **71**, 1, 1–10 (1984).

Allison, P. D., "Convergence failures in logistic regression", in "SAS Global Forum", vol. 360, pp. 1–11 (2008).

Anderson, J. A., "Separate sample logistic discrimination", Biometrika **59**, 1, 19–35 (1972).

Atkinson, A., A. Donev and R. Tobias, *Optimum Experimental Designs, with SAS*, vol. 34 (Oxford University Press, 2007).

Atkinson, A. C., "Recent developments in the methods of optimum and related experimental designs", International Statistical Review/Revue Internationale de Statistique **56**, 2, 99–115 (1988).

Atkinson, A. C. and D. R. Cox, "Planning experiments for discriminating between models", Journal of the Royal Statistical Society: Series B (Methodological) **36**, 3, 321–348 (1974).

Atkinson, A. C. and D. C. Woods, "Designs for generalized linear models", in "Handbook of Design and Analysis of Experiments", edited by A. Dean, M. Morris, J. Stufken and D. Bingham, pp. 471–514 (Chapman & Hall/CRC Press, 2015).

Bagley, S. C., H. White and B. A. Golomb, "Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain", Journal of clinical epidemiology **54**, 10, 979–985 (2001).

Berger, M. P. F. and W. Wong, *An Introduction to Optimal Designs for Social and Biomedical Research*, vol. 83 (John Wiley & Sons, 2009).

Berkson, J., "Why i prefer logits to probits", Biometrics **7**, 4, 327–339 (1951).

Biedermann, S., H. Dette and W. Zhu, "Optimal designs for dose–response models with restricted design spaces", Journal of the American Statistical Association **101**, 474, 747–759 (2006).

Bliss, C. I., "The method of probits", Science **79**, 2037, 38–39 (1934).

Bliss, C. I., "The calculation of the dosage-mortality curve", Annals of Applied Biology **22**, 1, 134–167 (1935).

Blot, W. J. and D. A. Meeter, "Sequential experimental design procedures", Journal of the American Statistical Association **68**, 343, 586–593 (1973).

Borkowski, J. J., "Using a genetic algorithm to generate small exact response surface designs", Journal of Probability and Statistical Science **1**, 1, 65–88 (2003).

Box, G. E. P. and N. R. Draper, "A basis for the selection of a response surface design", Journal of the American Statistical Association **54**, 287, 622–654 (1959).

Box, G. E. P. and N. R. Draper, "The choice of a second order rotatable design", Biometrika **50**, 3/4, 335–352 (1963).

Box, G. E. P. and H. L. Lucas, "Design of experiments in non-linear situations", Biometrika **46**, 1/2, 77–90 (1959).

Boyd, S. and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

Boyko, E. J., D. W. Trone, A. V. Peterson, I. G. Jacobson, A. J. Littman, C. Maynard, A. D. Seelig, N. F. Crum-Cianflone and J. B. Bricker, "Longitudinal investigation of smoking initiation and relapse among younger and older us military personnel", American Journal of Public Health **105**, 6, 1220–1229 (2015).

Brooke, M. d. L., R. J. Cuthbert, G. Harrison, C. Gordon and M. A. Taggart, "Persistence of brodifacoum in cockroach and woodlice: implications for secondary poisoning during rodent eradications", Ecotoxicology and environmental safety **97**, 183–188 (2013).

Casella, G. and R. L. Berger, *Statistical Inference* (Duxbury Pacific Grove, 2002).

Chaloner, K. and K. Larntz, "Optimal bayesian design applied to logistic regression experiments", Journal of Statistical Planning and Inference **21**, 2, 191–208 (1989).

Chaloner, K. and I. Verdinelli, "Bayesian experimental design: A review", Statistical Science **10**, 3, 273–304 (1995).

Chambers, E. A. and D. R. Cox, "Discrimination between alternative binary response models", Biometrika **54**, 3-4, 573–578 (1967).

Chen, G. and H. Tsurumi, "Probit and logit model selection", Communications in Statistics - Theory and Methods **40**, 1, 159–175 (2010).

Chernoff, H., "Locally optimal designs for estimating parameters", The Annals of Mathematical Statistics **24**, 4, 586–602 (1953).

Chipman, H. A. and W. J. Welch, "D-optimal design for generalized linear models", Unpublished (1996).

Ch'ng, C. K., S. H. Quah and H. C. Low, "A new approach for multiple-response optimization", Quality Engineering **17**, 4, 621–626 (2005).

Cook, R. D. and W. K. Wong, "On the equivalence of constrained and compound optimal designs", Journal of the American Statistical Association **89**, 426, 687–692 (1994).

Cooper, Z., S. L. Mitchell, R. J. Gorges, R. A. Rosenthal, S. R. Lipsitz and A. S. Kelley, "Predictors of mortality up to 1 year after emergency major abdominal surgery in older adults", Journal of the American Geriatrics Society **63**, 12, 2572–2579 (2015).

Cortes, C. and V. Vapnik, "Support-vector networks", Machine Learning **20**, 3, 273–297 (1995).

Cramer, J. S., "The origins of logistic regression", Tinbergen Institute Working Paper , 2002-119/4 (2002).

Dabbas, R. M., J. W. Fowler, D. A. Rollier and D. R. McCarville, "Multiple response optimization using mixture-designed experiments and desirability functions in semiconductor scheduling", International Journal of Production Research **41**, 5, 939–961 (2003).

Del Castillo, E., D. C. Montgomery and D. R. McCarville, "Modified desirability functions for multiple response optimization", Journal of Quality Technology **28**, 3, 337–345 (1996).

Derringer, G. and R. Suich, "Simultaneous optimization of several response variables", Journal of Quality Technology **12**, 4, 214–219 (1980).

Dette, H. and Y. Grigoriev, "E-optimal designs for second-order response surface models", The Annals of Statistics **42**, 4, 1635–1656 (2014).

Dette, H. and L. M. Haines, "E-optimal designs for linear and nonlinear models with two parameters", Biometrika **81**, 4, 739–754 (1994).

Dror, H. A. and D. M. Steinberg, "Approximate local d-optimal experiment design for binary response", Technical Report RP-SO R-0501 (2005).

Dror, H. A. and D. M. Steinberg, "Robust experimental design for multivariate generalized linear models", Technometrics **48**, 4, 520–529 (2006).

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

Fedorov, V. V., *Theory of Optimal Experiments* (Elsevier, 1972).

Fedorov, V. V. and S. L. Leonov, *Optimal Design for Nonlinear Response Models* (CRC Press, 2013).

Filipiak, K. and R. Różański, "On the e-optimality of complete block designs under a mixed interference model", Journal of Statistical Planning and Inference **143**, 3, 583–592 (2013).

Firth, D., "Bias reduction of maximum likelihood estimates", Biometrika **80**, 1, 27–38 (1993).

Fisher, R. A., "The case of zero survivors in probit assays", Annals of Applied Biology **22**, 1, 164–165 (1935a).

Fisher, R. A., *The Design of Experiments* (Oliver and Boyd, 1935b).

Ford, I., D. M. Titterington and C. P. Kitsos, "Recent advances in nonlinear experimental design", Technometrics **31**, 1, 49–60 (1989).

Ford, I., B. Torsney and C. F. J. Wu, "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) **54**, 2, 569–583 (1992).

Fornius, E. F., *Optimal Design of Experiments for the Quadratic Logistic Model*, Ph.D. thesis, Stockholm University (2008).

Gelman, A., A. Jakulin, M. G. Pittau and Y. Su, "A weakly informative default prior distribution for logistic and other regression models", The Annals of Applied Statistics **2**, 4, 1360–1383 (2008).

Giadrosich, D. L., *Operations Research Analysis in Test and Evaluation* (American Institute of Aeronautics and Astronautics, 1995).

Gill, J., *Generalized Linear Models: A Unified Approach*, vol. 134 (Sage Publications, 2000).

Goos, P., *The Optimal Design of Blocked and Split-Plot Experiments*, vol. 164 (Springer Science & Business Media, 2012).

Gotwalt, C. M., B. A. Jones and D. M. Steinberg, "Fast computation of designs robust to parameter uncertainty for nonlinear settings", Technometrics **51**, 1, 88–95 (2009).

Haines, L. M., "The application of the annealing algorithm to the construction of exact optimal designs for linear–regression models", Technometrics **29**, 4, 439–447 (1987).

Haines, L. M., G. Kabera, P. Ndlovu and T. E. O'Brien, "D-optimal designs for logistic regression in two variables", in "mODa 8-Advances in Model-Oriented Design and Analysis", edited by J. López-Fidalgo, J. M. Rodríguez-Díaz and B. Torsney, pp. 91–98 (Springer, 2007).

Haines, L. M., G. M. Kabera and P. Ndlovu, "D-optimal designs for the two-variable binary logistic regression model with interaction", Journal of Statistical Planning and Inference **193**, 136–150 (2018).

Hamada, M., H. F. Martz, C. S. Reese and A. G. Wilson, "Finding near-optimal bayesian experimental designs via genetic algorithms", The American Statistician **55**, 3, 175–181 (2001).

Hardin, J. W. and J. M. Hilbe, *Generalized Linear Models and Extensions* (Stata Press, 2007).

Heinze, G., "A comparative investigation of methods for logistic regression with separated or nearly separated data", Statistics in Medicine **25**, 24, 4216–4226 (2006).

Heinze, G. and M. Schemper, "A solution to the problem of separation in logistic regression", Statistics in Medicine **21**, 16, 2409–2419 (2002).

Heredia-Langner, A., D. C. Montgomery, W. M. Carlyle and C. M. Borror, "Model-robust optimal designs: A genetic algorithm approach", Journal of Quality Technology **36**, 3, 263–279 (2004).

Imhof, L. A., "Maximin designs for exponential growth models and heteroscedastic polynomial models", Annals of Statistics **29**, 2, 561–576 (2001).

Jeffreys, H., "An invariant form for the prior probability in estimation problems", Proceedings of the Royal Society of London: Series A (Mathematical and Physical Sciences) **186**, 1007, 453–461 (1946).

Johnson, M. E. and C. J. Nachtsheim, "Some guidelines for constructing exact d-optimal designs on convex design spaces", Technometrics **25**, 3, 271–277 (1983).

Johnson, R. T. and D. C. Montgomery, "Choice of second-order response surface designs for logistic and poisson regression models", International Journal of Experimental Design and Process Optimisation **1**, 1, 2–23 (2009).

Jones, B. and P. Goos, "A candidate-set-free algorithm for generating d-optimal split-plot designs", Journal of the Royal Statistical Society: Series C (Applied Statistics) **56**, 3, 347–364 (2007).

Jones, B. and P. Goos, "I-optimal versus d-optimal split-plot response surface designs", Journal of Quality Technology **44**, 2, 85–101 (2012).

Jones, B. and P. Goos, "Optimal design of blocked experiments in the presence of supplementary information about the blocks", Journal of Quality Technology **47**, 4, 301–317 (2015).

Kabera, G. M., L. M. Haines and P. Ndlovu, "The analytic construction of d-optimal designs for the two-variable binary logistic regression model without interaction", Statistics **49**, 5, 1169–1186 (2015).

Khuri, A. I., B. Mukherjee, B. K. Sinha and M. Ghosh, "Design issues for generalized linear models: A review", Statistical Science **21**, 3, 376–399 (2006).

Kiefer, J., "Optimum experimental designs", Journal of the Royal Statistical Society: Series B (Methodological) **21**, 272–319 (1959).

Kiefer, J., "General equivalence theory for optimum designs (approximate theory)", The Annals of Statistics **2**, 5, 849–879 (1974).

Kiefer, J. and J. Wolfowitz, "The equivalence of two extremum problems", Canadian Journal of Mathematics **12**, 363–366 (1960).

Konis, K., *Linear Programming Algorithms for Detecting Separated Data in Binary Logistic Regression Models*, Ph.D. thesis, University of Oxford (2007).

Kuk, A. Y. and Y. W. Cheng, "The monte carlo newton-raphson algorithm", Journal of Statistical Computation and Simulation **59**, 3, 233–250 (1997).

Läuter, E., "Optimal multipurpose designs for regression models", Mathematische Operationsforschung und Statistik **7**, 1, 51–68 (1976).

Ledolter, J., *Data Mining and Business Analytics with R* (John Wiley & Sons, 2013).

Lewis, S. L., D. C. Montgomery and R. H. Myers, "Confidence interval coverage for designed experiments analyzed with glms", Journal of Quality Technology **33**, 3, 279–292 (2001).

Li, Y. and X. Deng, "On i-optimal designs for generalized linear models: An efficient algorithm via general equivalence theory", arXiv preprint: 1801.05861 (2018).

Lipsitz, S. R., G. M. Fitzmaurice, S. E. Regenbogen, D. Sinha, J. G. Ibrahim and A. A. Gawande, "Bias correction for the proportional odds logistic regression model with application to a study of surgical complications", Journal of the Royal Statistical Society: Series C (Applied Statistics) **62**, 2, 233–250 (2013).

Long, J. S., *Regression Models for Categorical and Limited Dependent Variables*, vol. 7 (Sage Publications, 1997).

Mancenido, M. V., R. Pan, D. C. Montgomery and C. M. Anderson-Cook, "Comparing d-optimal designs with common mixture experimental designs for logistic regression", Chemometrics and Intelligent Laboratory Systems **187**, 11–18 (2019).

Mathew, T. and B. K. Sinha, "Optimal designs for binary data under logistic regression", Journal of Statistical Planning and Inference **93**, 1/2, 295–307 (2001).

McGree, J. M., J. A. Eccleston and S. B. Duffull, "Compound optimal design criteria for nonlinear models", Journal of Biopharmaceutical Statistics **18**, 4, 646–661 (2008).

Meyer, R. K. and C. J. Nachtsheim, "The coordinate-exchange algorithm for constructing exact optimal experimental designs", Technometrics **37**, 1, 60–69 (1995).

Minkin, S., "Optimal designs for binary data", Journal of the American Statistical Association **82**, 400, 1098–1103 (1987).

Mitchell, T. J., "An algorithm for the construction of "d-optimal" experimental designs", Technometrics **16**, 2, 203–210 (1974).

Montgomery, D. C., *Design and Analysis of Experiments* (John Wiley & Sons, 2017).

Montgomery, D. C., E. A. Peck and G. G. Vining, *Introduction to Linear Regression Analysis*, vol. 821 (John Wiley & Sons, 2012).

Müller, W. G. and M. Stehlík, "Compound optimal spatial designs", Environmetrics **21**, 3/4, 354–364 (2010).

Myers, R. H. and D. C. Montgomery, "A tutorial on generalized linear models", Journal of Quality Technology **29**, 3, 274–291 (1997).

Myers, R. H., D. C. Montgomery, G. G. Vining and T. J. Robinson, *Generalized Linear Models: with Applications in Engineering and the Sciences*, vol. 791 (John Wiley & Sons, 2012).

Myers, W. R., R. H. Myers and W. H. Carter Jr, "Some alphabetic optimal designs for the logistic regression model", Journal of Statistical Planning and Inference **42**, 1/2, 57–77 (1994).

Nelder, J. A. and R. W. M. Wedderburn, "Generalized linear models", Journal of the Royal Statistical Society: Series A (General) **135**, 3, 370–384 (1972).

Neyer, B. T., "A d-optimality-based sensitivity test", Technometrics **36**, 1, 61–70 (1994).

Nordström, K., "The concentration ellipsoid of a random vector revisited", Econometric Theory **7**, 3, 397–403 (1991).

Park, A. R., M. V. Mancenido and D. C. Montgomery, "Separation in d-optimal experimental designs for the logistic regression model", Quality and Reliability Engineering International **35**, 3, 776–787 (2019).

Parsa Maram, P. and H. Jafari, "Bayesian d-optimal design for logistic regression model with exponential distribution for random intercept", Journal of Statistical Computation and Simulation **86**, 10, 1856–1868 (2016).

Penman, A. D. and W. D. Johnson, "Complementary log–log regression for the estimation of covariate-adjusted prevalence ratios in the analysis of data from cross-sectional studies", Biometrical Journal **51**, 3, 433–442 (2009).

Piegorsch, W. W., "Complementary log regression for generalized linear models", The American Statistician **46**, 2, 94–99 (1992).

Poole, D., *Linear Algebra: A Modern Introduction* (Cengage Learning, 2014).

Pronzato, L. and W. G. Müller, "Design of computer experiments: space filling and beyond", Statistics and Computing **22**, 3, 681–701 (2012).

Pronzato, L. and É. Walter, "Robust experiment design via maximin optimization", Mathematical Biosciences **89**, 2, 161–176 (1988).

Pukelsheim, F., *Optimal Design of Experiments*, vol. 50 (SIAM, 1993).

Rahman, M. L., *Bayesian Analysis of Multi-Stratum Designs and Probability-based Optimal Designs with Separation*, Ph.D. thesis, Queen Mary University of London (2015).

Rainey, C., "Dealing with separation in logistic regression models", Political Analysis **24**, 3, 339–355 (2016).

Rodríguez, M., B. Jones, C. M. Borror and D. C. Montgomery, "Generating and assessing exact g-optimal designs", Journal of Quality Technology **42**, 1, 3–20 (2010).

Różański, M., R. Wituła and E. Hetmaniok, "More subtle versions of the hadamard inequality", Linear Algebra and its Applications **532**, 500–511 (2017).

Saleh, M. and R. Pan, "A clustering-based coordinate exchange algorithm for generating g-optimal experimental designs", Journal of Statistical Computation and Simulation **86**, 8, 1582–1604 (2016).

Santner, T. J. and D. E. Duffy, "A note on a. albert and j. a. anderson's conditions for the existence of maximum likelihood estimates in logistic regression models", Biometrika **73**, 3, 755–758 (1986).

Sibson, R., "Da-optimality and duality", in "Progress in Statistics, Colloquia Mathematica Societatis János Bolyai", vol. 9, pp. 677–692 (1972).

Sitter, R. R. and B. E. Forbes, "Optimal two-stage designs for binary response experiments", Statistica Sinica **7**, 4, 941–955 (1997).

Stufken, J. and M. Yang, "Optimal designs for generalized linear models", in "Design and Analysis of Experiments, Volume 3: Special Designs and Applications", edited by K. Hinkelmann, vol. 810, pp. 137–164 (John Wiley & Sons, 2012).

Tai, B. C. and D. Machin, *Regression Methods for Medical Research* (Wiley-Blackwell, 2013).

Tommasi, C., "Optimal design robust to a misspecified model", Communications in Statistics-Simulation and Computation **41**, 7, 1220–1231 (2012).

Tsutakawa, R. K., "Selection of dose levels for estimating a percentage point of a logistic quantal response curve", Applied Statistics **29**, 1, 25–33 (1980).

Welch, W. J., "Computer-aided design of experiments for response estimation", Technometrics **26**, 3, 217–224 (1984).

White, L. V., "An extension of the general equivalence theorem to nonlinear models", Biometrika **60**, 2, 345–348 (1973).

Woods, D. C., S. M. Lewis, J. A. Eccleston and K. G. Russell, "Designs for generalized linear models with several variables and model uncertainty", Technometrics **48**, 2, 284–292 (2006).

Wu, C. F. J. and Y. Tian, "Three-phase optimal design of sensitivity experiments", Journal of Statistical Planning and Inference **149**, 1–15 (2014).

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

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

Zelterman, D., *Models for Discrete Data* (Oxford University Press, 1999).

Zorn, C., "A solution to separation in binary response models", Political Analysis **13**, 2, 157–170 (2005).

APPENDIX A

LOCALLY D_{MP}-OPTIMAL DESIGNS

A.1 Test Case 1: Single-Factor Main Effect Model

Parameters: $\beta_0 = 0$, $\beta_1 = 3$; Design Size: $n = 6$

λ	**Design**					
1.0	-0.5144575	-0.5144575	0.5144933	0.5144933	0.5144933	-0.5144575
0.9	-0.4988933	-0.4988933	0.4962152	-0.4988933	0.4962152	0.4962152
0.8	-0.4811868	0.4760174	-0.4811868	0.4760174	-0.4811868	0.4760174
0.7	0.4548911	0.4548911	-0.4515524	-0.4515524	-0.4515524	0.4548911
0.6	-0.4212407	-0.4212407	0.4247162	0.4207429	-0.4212407	0.4207429
0.5	0.3823663	0.3823663	0.3823663	-0.3797374	-0.3842654	-0.3797374
0.4	-0.3225372	-0.3225372	-0.3258209	0.3238178	0.3238178	0.3238178
0.3	0.2247608	0.2280070	-0.2270189	-0.2270189	-0.2270189	0.2247608
0.2	0.0120883	-0.0036445	-0.0036445	-0.0013941	-0.0013941	-0.0013941
0.1	-0.0017392	-0.0017392	-0.0017392	0.0004679	0.0004679	0.0004679
0.0	-0.0000188	-0.0000188	-0.0000188	-0.0000188	-0.0000188	-0.0000188

λ	***D*-crit**	***D*-eff**	***P*(Sep)**
1.0	0.20047	100.0000%	0.81065
0.9	0.20010	99.9068%	0.79793
0.8	0.19864	99.5403%	0.78306
0.7	0.19541	98.7281%	0.76217
0.6	0.18862	96.9987%	0.63295
0.5	0.17592	93.6752%	0.59502
0.4	0.14973	86.4215%	0.53922
0.3	0.09199	67.7383%	0.45033
0.2	0.00007	1.8268%	1.00000
0.1	0.00000	0.3697%	1.00000
0.0	0.00000	0.0000%	1.00000

A.2 Test Case 2: Two-Factor Main Effects Model

Parameters: $\beta_0 = 0$, $\beta_1 = 2$, $\beta_2 = 1$; Design Size: $n = 10$

λ	Factor	Design																	
1.0	X1	-1	1	0.212721	1	-1	0.212721	-0.21267	-0.21267	-1	1								
	X2	1	-1	1	-1	1	1	-1	-1	1	-1								
0.9	X1	0.012518	-0.14325	1	0.012518	1	1	-1	-1	-0.14325	-0.1432543								
	X2	1	-1	-1	1	-1	-1	1	1	-1	-1								
0.8	X1	0.214286	-1	0.214286	1	1	-1	0.214286	-1	0.034903	0.0349029								
	X2	1	1	1	-1	-1	1	1	1	-1	-1								
0.7	X1	0.192042	-1	-1	0.077423	0.077423	-1	0.192042	0.983673	0.192042	0.9836731								
	X2	1	1	1	-1	-1	1	1	-1	1	-1								
0.6	X1	-1	0.043076	0.941098	0.941098	0.043076	-1	-1	0.096176	0.045179	0.0430758								
	X2	1	1	-1	-1	1	1	1	-1	-1	1								
0.5	X1	-1	0.155535	1	1	-0.82334	-0.02837	0.155535	-0.02837	0.155535	-0.8233404								
	X2	1	-1	-1	-1	1	1	-1	1	-1	1								
0.4	X1	0.26369	0.82641	-0.0388	0.82641	-0.78014	-0.07219	0.26369	0.26369	-0.78014	-0.780138								
	X2	-1	-1	1	-1	1	1	-1	-1	1	1								
0.3	X1	0.349354	0.724621	0.52414	0.349354	-1	0.52414	-0.39449	-0.18569	-0.39449	-0.6868047								
	X2	-1	-1	-1	-1	1	-1	1	1	1	1								
0.2	X1	0.496748	0.496748	-0.48281	0.519459	0.496748	-0.48281	0.496748	0.519459	0.496748	0.4967484								
	X2	-1	-1	1	-1	-1	1	-1	-1	-1	-1								
0.1	X1	0.509913	0.509913	-0.47547	0.509913	0.509913	0.509913	0.509913	0.509913	0.510896	0.5099127								
	X2	-1	-1	1	-1	-1	-1	-1	-1	-1	-1								
0.0	X1	0.550552	0.550552	0.550552	0.550552	0.550552	0.550552	0.550552	0.550552	0.550552	0.5505524								
	X2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1								

λ	D-crit	D-eff	P(Sep)
1.0	1.9568	100.000%	0.63140
0.9	1.9153	99.287%	0.56588
0.8	1.9290	99.524%	0.57686
0.7	1.9105	99.204%	0.56134
0.6	1.8154	97.530%	0.40146
0.5	1.6047	93.601%	0.37857
0.4	1.1463	83.672%	0.37745
0.3	0.6169	68.061%	0.20645
0.2	0.0008	7.338%	1.00000
0.1	0.0000	0.627%	1.00000
0.0	0.0000	0.000%	1.00000

A.3 Test Case 3: Two-Factor Main Effects w/Interaction Model

Parameters: $\beta_0 = -1$, $\beta_1 = 4$, $\beta_2 = 1$, $\beta_{12} = -2$; Design Size: $n = 16$

λ	Factor	Design															
1.0	X1	0.591	0.772	-0.772	0.772	0.076	-0.772	0.076	0.076	0.591	0.772	-0.772	-0.772	0.076	0.591	0.772	0.591
	X2	-1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000
0.9	X1	0.581	0.067	0.581	0.581	-1.000	-0.776	-0.776	-0.776	0.779	-0.776	0.067	0.067	0.067	0.779	0.779	0.779
	X2	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
0.8	X1	0.582	0.086	0.582	-0.750	0.086	0.744	-0.750	0.744	-0.750	0.582	-0.750	0.086	0.749	0.744	0.744	0.086
	X2	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	1.000	-1.000
0.7	X1	0.122	0.603	0.765	0.765	0.603	-0.699	-0.699	0.712	0.603	0.122	0.122	0.603	0.122	-0.699	-0.699	-0.699
	X2	-1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000
0.6	X1	0.113	0.569	0.627	-0.617	0.569	0.702	0.113	-0.617	0.706	0.569	0.113	-0.617	0.569	-0.617	-0.617	0.113
	X2	-1.000	-1.000	1.000	-1.000	1.000	-1.000	-1.000	1.000	1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000
0.5	X1	0.578	0.578	0.578	-0.624	0.093	0.141	0.141	0.578	0.671	0.693	-0.624	-0.624	0.696	0.696	-0.624	0.093
	X2	-1.000	-1.000	-1.000	1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-1.000
0.4	X1	0.161	-0.537	0.589	0.589	0.589	0.589	0.663	-0.537	0.663	0.161	0.663	-0.537	0.663	-0.537	-0.537	0.161
	X2	-1.000	1.000	-1.000	-1.000	-1.000	-1.000	1.000	1.000	1.000	-1.000	1.000	1.000	1.000	1.000	1.000	-1.000
0.3	X1	0.194	0.577	-0.522	0.577	0.194	-0.522	0.194	-0.522	0.577	0.602	0.194	-0.522	0.194	0.513	0.513	0.513
	X2	-1.000	-1.000	1.000	-1.000	-1.000	1.000	-1.000	1.000	-1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	1.000
0.2	X1	0.165	0.498	0.498	0.379	0.480	-1.000	0.379	0.379	-1.000	0.480	0.379	0.379	0.165	0.480	0.165	0.165
	X2	-1.000	-1.000	-1.000	1.000	-1.000	1.000	1.000	1.000	1.000	-1.000	1.000	1.000	-1.000	-1.000	-1.000	-1.000
0.1	X1	0.230	0.230	0.428	0.095	0.487	0.095	0.428	0.095	0.095	0.340	0.095	0.095	0.095	0.340	0.310	0.415
	X2	-1.000	-1.000	-1.000	1.000	-1.000	1.000	-1.000	1.000	1.000	-1.000	1.000	1.000	1.000	-1.000	-1.000	1.000
0.0	X1	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344	0.344
	X2	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000

λ	*D*-crit	*D*-eff	*P*(Sep)
1.0	1.14318	100.0000%	0.9132
0.9	1.14206	99.9756%	0.9168
0.8	1.13856	99.8990%	0.8832
0.7	1.11566	99.3927%	0.8516
0.6	1.06451	98.2334%	0.8076
0.5	1.06926	98.3429%	0.6252
0.4	0.95981	95.7233%	0.8308
0.3	0.75619	90.1838%	0.7736
0.2	0.57339	84.1559%	0.8600
0.1	0.01530	34.0124%	1.0000
0.0	0.00000	0.0000%	1.0000

APPENDIX B

TWO-PHASE COORDINATE EXCHANGE ALGORITHM JSL CODE

Only the second phase of the CEA that generates the locally D_{MP}-optimal designs will be shown for each test case. A locally D-optimal design can be generated by setting 'lambda' = 1 and 'scale' = 1.

B.1 Single-Factor Main Effect Model Design Code

```
/*****************************/
//INITIALIZE//
/*****************************/

Names Default to Here( 1 );
Clear Log();
clear symbols();


/*****************************/
//FUNCTIONS//
/*****************************/
//Function to calculate optimality criterion value
//X is model matrix, lambda is weight of D-criterion; range [0,1]
//det_scale is max value of |X`VX| from D-opt design
//max_val is maximum value of LP^2
criterion_val = Function( {X, lambda, det_scale, max_val},

{default local},

n = nrows(X);
b0 = 0;                      //intercept
b1 = 3;                      //X coefficient
coeffs = (b0 || b1)`;
p = nrows(coeffs);

LP = J(n,1,0);               //vector to store linear predictor values
LP_sq = J(n,1,0);            //vector to store LP^2
ExpY = J(n,1,0);             //vector to store E(y) values
WeightY = J(n,1,0);          //vector to store Var(y) values

for(j = 1, j <= n, j++,
LP[j] = X[j,0]*coeffs;
LP_sq[j] = (LP[j])^2;
ExpY[j] = 1/(1+exp(-(LP[j])));
WeightY[j] = ExpY[j]*(1-ExpY[j]);
);

V = diag(WeightY);
M = X`*V*X;

Z1 = (det(M)/det_scale)^(1/p);
```

```
Z2 = (sum(LP_sq)/(n*max_val))^(1/p);
Z = (lambda)*(Z1)-(1-lambda)*Z2;

Return(Z);
);

/*****************************/
//MAIN //
/*****************************/

/*****************************/
//Inputs//
/*****************************/
iter = 1000;                          //number of random starting designs
f = 1;                                //number of factors
lambda = 0.5;                         //weight of D-criterion; range [0,1]
scale = 0.200473977115786;            //|X'VX| from D-opt; n=6, b0=0, b1=3
delta = 0.005;                        //step size of coordinate exchange
gap = 0.001;                          //size of criterion convergence
n = 6;                                //number of runs

Z_best = -1000;                       //initialize criterion value


/*****************************/
//Determine Max LP Value//
/*****************************/

b0 = 0;                               //intercept
b1 = 3;                               //X coefficient
coeffs = (b0 || b1)';

//max value of LP must occur at extreme points of x (-1 OR 1)
val1 = abs([1 1]*coeffs);
val2 = abs([1 -1]*coeffs);
max_lp = (max(val1,val2))^2;


/*****************************/
//Generate Random Initial Design//
/*****************************/
for(k = 1, k <=iter, k++,

D = J(n,f,random uniform(-1,1));
X = J(n,1,1) || D;           //creates model matrix for main effect



/*****************************/
//Coordinate Exchange//
```

```
/***************************/
delta = delta + random uniform(0,.001); //random step adjustment

X_cand1 = X;
X_cand2 = X;

//'convergence' will capture when the design is not increasing meaningfully
convergence = 1000;          //initially set at high value


while(convergence >= gap,
X_current = X;

//optimize coordinate j
for(j = 1, j <=n, j++,
//find direction of improvement, make initial replacement
store_vals = J(2,1,0);

store_vals[1,1] = X[j,2] - delta;
store_vals[2,1] = X[j,2] + delta;


X_cand1[j,2] = store_vals[1,1];
X_cand2[j,2] = store_vals[2,1];

flag = 0;

if(criterion_val(X_cand2, lambda, scale, max_lp) >= criterion_val(X_cand1,
   lambda, scale, max_lp),
//replaces best coordinate into X
X[j,2] = X_cand2[j,2];
flag = 1,
X[j,2] = X_cand1[j,2];

);

//update candidate matrices
X_cand1[j,2] = X[j,2];
X_cand2[j,2] = X[j,2];

improve = 1000;

if(flag == 1,

while(improve > 0,
Z_initial = criterion_val(X, lambda, scale, max_lp);
X[j,2] = X[j,2] + delta;
//feasibility check
if(X[j,2] > 1,
X[j,2] = 1;
```

```
flag = 2;
improve = -1,
improve = criterion_val(X, lambda, scale, max_lp)-Z_initial;
);

),

while(improve > 0,
Z_initial = criterion_val(X, lambda, scale, max_lp);

X[j,2] = X[j,2] - delta;

//feasibility check
if(X[j,2] < -1,
X[j,2] = -1;
flag = 2;
improve = -1,
improve = criterion_val(X, lambda, scale, max_lp)-Z_initial;
);

);

//set value to actual optimal coordinate
if(flag == 1,
X[j,2] = X[j,2] - delta,
flag == 0,
X[j,2] = X[j,2] + delta;
);

);

X_cand1 = X;
X_cand2 = X;

);

convergence = criterion_val(X, lambda, scale, max_lp)-criterion_val(
   X_current, lambda, scale, max_lp);
);

Z_current = criterion_val(X, lambda, scale, max_lp);

if(Z_current > Z_best,
X_best = X;
Z_best = Z_current
);

);
```

```
/***************************/
//PART 3: K-Pass Point Exchange//
/***************************/

//define candidate set
C = X_best;
X_PEA = X_best;

store_vals = J(1,n,0);

//set number of PEA cycles w/value of K
K = 1;

//K-Pass Point Exchange
for(l = 1, l<=K, l++,

for(j = 1, j <=n, j++,

//optimize row (point) j
for(i=1, i <= n, i++,
X_PEA[j,0] = C[i,0];
store_vals[i] = criterion_val(X_PEA, lambda, scale, max_lp);
);

X_best[j,0] = C[Loc Max(store_vals),0];
X_PEA[j,0] = X_best[j,0];

);
);


/***************************/
//OUTPUT//
/***************************/
as table(X_best);
```

B.2 Two-Factor Main Effects Model Design Code

```
/*****************************/
//INITIALIZE//
/*****************************/

Names Default to Here( 1 );
Clear Log();
clear symbols();


/*****************************/
//FUNCTIONS//
/*****************************/
//Function to calculate optimality criterion value
//X is model matrix, lambda is weight of D-criterion; range [0,1]
//det_scale is max value of |X`VX| from D-opt design
//max_val is maximum value of LP^2
criterion_val = Function( {X, lambda, det_scale, max_val},

{default local},

n = nrows(X);

b0 = 0;                       //intercept
b1 = 2;                       //X1 coefficient
b2 = 1;                       //X2 coefficient
coeffs = (b0 || b1 || b2)`;
p = nrows(coeffs);

LP = J(n,1,0);                //vector to store linear predictor values
LP_sq = J(n,1,0);             //vector to store LP^2
ExpY = J(n,1,0);              //vector to store E(y) values
WeightY = J(n,1,0);           //vector to store Var(y) values

for(j = 1, j <= n, j++,
LP[j] = X[j,0]*coeffs;
LP_sq[j] = (LP[j])^2;
ExpY[j] = 1/(1+exp(-(LP[j])));
WeightY[j] = ExpY[j]*(1-ExpY[j]);
);

V = diag(WeightY);
M = X`*V*X;

Z1 = (det(M)/det_scale)^(1/p);
Z2 = (sum(LP_sq)/(n*max_val))^(1/p);
Z = (lambda)*(Z1)-(1-lambda)*Z2;

Return(Z);
```

```
);

/****************************/
//MAIN //
/****************************/

/****************************/
//Inputs//
/****************************/
iter = 1000;                          //number of random starting designs
f = 2;                                //number of factors in experiment
lambda = 0.5;                         //weight of D-criterion; range [0,1]
//scale = 1;
scale = 1.95682571192056;             //|X'VX| from D-opt; n=10, b0=0, b1=2,
   b2=1
delta = 0.005;                        //step size of coordinate exchange
gap = 0.001;                          //size of criterion convergence
n = 10;                               //number of runs

Z_best = -1000;                       //initialize criterion value

/****************************/
//Determine Max LP Value//
/****************************/

b0 = 0;                               //intercept
b1 = 2;                               //X1 coefficient
b2 = 1;                               //X2 coefficient
coeffs = (b0 || b1 || b2)';

//max value of LP must occur at extreme points of region
val1 = abs([1 1 1]*coeffs);
val2 = abs([1 -1 -1]*coeffs);
val3 = abs([1 -1 1]*coeffs);
val4 = abs([1 1 -1]*coeffs);
max_lp = (max(val1,val2,val3,val4))^2;


/****************************/
//Generate Random Initial Design//
/****************************/
for(k = 1, k <=iter, k++,

D = J(n,f,random uniform(-1,1));
X = J(n,1,1) || D;                                    //creates model
   matrix for main effects



/****************************/
```

```
//Coordinate Exchange//
/***************************/
delta = delta + random uniform(0,.001); //random step adjustment

X_cand1 = X;
X_cand2 = X;

//'convergence' will capture when the design is not increasing meaningfully
convergence = 1000;          //initially set at high value


while(convergence >= gap,
X_current = X;

//point (row) j
for(j = 1, j <= n, j++,

//optimize point j, coordinate i
for(i = 1, i <= f, i++,

//find direction of improvement, make initial replacement
store_vals = J(2,1,0);

store_vals[1,1] = X[j,i+1] - delta;
store_vals[2,1] = X[j,i+1] + delta;


X_cand1[j,i+1] = store_vals[1,1];
X_cand2[j,i+1] = store_vals[2,1];

flag = 0;

if(criterion_val(X_cand2, lambda, scale, max_lp) >= criterion_val(X_cand1,
   lambda, scale, max_lp),
//replaces best coordinate into X
X[j,i+1] = X_cand2[j,i+1];
flag = 1,
X[j,i+1] = X_cand1[j,i+1];

);

//update candidate matrices
X_cand1[j,i+1] = X[j,i+1];
X_cand2[j,i+1] = X[j,i+1];

improve = 1000;

if(flag == 1,

while(improve > 0,
```

```
Z_initial = criterion_val(X, lambda, scale, max_lp);

X[j,i+1] = X[j,i+1] + delta;


//stop check for feasibility
if(X[j,i+1] > 1,
//when point becomes infeasible, exit while loop
X[j,i+1] = 1;
flag = 2;
improve = -1,
improve = criterion_val(X, lambda, scale, max_lp)-Z_initial;
);

),


while(improve > 0,
Z_initial = criterion_val(X, lambda, scale, max_lp);

X[j,i+1] = X[j,i+1] - delta;


//stop check for feasibility
if(X[j,i+1] < -1,
//when point becomes infeasible, exit while loop
X[j,i+1] = -1;
flag = 2;
improve = -1;
improve = criterion_val(X, lambda, scale, max_lp)-Z_initial;
);

);

//set value to actual optimal coordinate
if(flag == 1,
X[j,i+1] = X[j,i+1] - delta,
flag == 0,
X[j,i+1] = X[j,i+1] + delta;
);

);

X_cand1[j,i+1] = X[j,i+1];
X_cand2[j,i+1] = X[j,i+1];

)

);
```

```
convergence = criterion_val(X, lambda, scale, max_lp)-criterion_val(
    X_current, lambda, scale, max_lp);
);

Z_current = criterion_val(X, lambda, scale, max_lp);

if(Z_current > Z_best,
X_best = X;
Z_best = Z_current
);

);


/***************************/
//PART 3: K-Pass Point Exchange//
/***************************/

//define candidate set
C = X_best;
X_PEA = X_best;

store_vals = J(1,n,0);

//set number of PE cycles w/value of K
K = 1;

//K-Pass Point Exchange
for(l = 1, l<=K, l++,

for(j = 1, j <=n, j++,

//optimize row (point) j
for(i=1, i <= n, i++,
X_PEA[j,0] = C[i,0];
store_vals[i] = criterion_val(X_PEA, lambda, scale, max_lp);
);

X_best[j,0] = C[Loc Max(store_vals),0];
X_PEA[j,0] = X_best[j,0];

);
);

/***************************/
//OUTPUT//
/***************************/
as table(X_best);
```

B.3 Two-Factor Main Effects w/Interaction Model Design Code

```
/***************************/
//INITIALIZE//
/***************************/

Names Default to Here( 1 );
Clear Log();
clear symbols();


/***************************/
//FUNCTIONS//
/***************************/
//Function to calculate optimality criterion value
//X is model matrix, lambda is weight of D-criterion; range [0,1]
//det_scale is max value of |X'VX| from D-opt design
//max_val is maximum value of LP^2
criterion_val = Function( {X, lambda, det_scale, max_val},

{default local},

n = nrows(X);

b0 = -1;                                    //intercept
b1 = 4;                                     //X1 coefficient
b2 = 1;                                     //X2 coefficient
b12 = -2;                                   //Interaction Coefficient
coeffs = (b0 || b1 || b2 || b12)';

p = nrows(coeffs);

LP = J(n,1,0);                      //vector to store linear predictor
   values
LP_sq = J(n,1,0);                   //vector to store LP^2
ExpY = J(n,1,0);                    //vector to store E(y) values
WeightY = J(n,1,0);                 //vector to store Var(y) values

for(j = 1, j <= n, j++,
LP[j] = X[j,0]*coeffs;
LP_sq[j] = (LP[j])^2;
ExpY[j] = 1/(1+exp(-(LP[j])));
WeightY[j] = ExpY[j]*(1-ExpY[j]);
);

V = diag(WeightY);
M = X'*V*X;
Z1 = (det(M)/det_scale)^(1/p);

Z2 = (sum(LP_sq)/(n*max_val))^(1/p);
```

```
Z = (lambda)*(Z1)-(1-lambda)*Z2;

Return(Z);
);


//Function that expands design matrix to model matrix for 2FI model
ME2FI = function({X},

{default local},


D = J(nrow(X), 1, 1) || X;
p = ncol(D);
for(j=3, j<=p, j++,
for(i=2, i<j, i++,
D = D || E Mult(D[0,j],D[0,i]);
);
);
D;
);

/*****************************/
//MAIN //
/*****************************/

/*****************************/
//Inputs//
/*****************************/
iter = 1000;                          //number of random starting designs
f = 2;                                //number of factors in experiment
lambda = 0.5;                         //weight of D-criterion; range [0,1]
scale = 1.1431769504143;              //|X'VX| from D-opt; n=16, b0=-1, b1=4,
    b2=1, b12=-2
delta = 0.005;                        //step size of coordinate exchange
gap = 0.001;                          //size of criterion convergence
n = 16;                               //number of runs

Z_best = -1000;                       //initialize criterion value

b0 = -1;                              //intercept
b1 = 4;                               //X1 coefficient
b2 = 1;                               //X2 coefficient
b12 = -2;                             //Interaction Coefficient
coeffs = (b0 || b1 || b2 || b12)';



/*****************************/
```

```
//Determine Max LP Value//
/***************************/
res=0.01;

xv_coord = Index(-1,1,res);

//since dimensions of X1_coord/X2_coord will be equal
q = Ncol(xv_coord);

//'X_coord' will store all of the design points to be evaluated for LP
    value
X_coord = [];
X_store = J(q,2,0);

//create table of all design points
for(i =1, i <= q, i++,
for(j = 1, j <= q, j++,
X_store[j,1]=xv_coord[i];
X_store[j,2]=xv_coord[j];
);
X_coord = VConcat(X_coord,X_store);
);

X_coord = ME2FI(X_coord);

derp = Nrows(X_coord);
LP_vals = J(derp,1,0);

for(j = 1, j <= derp, j++,
LP_vals[j] = abs(X_coord[j,0]*coeffs);
);

//lp_max holds maximum value of squared linear predictor across 2D grid
lp_max = (Max(LP_vals))^2;

/***************************/
//Generate Random Initial Design//
/***************************/
for(k = 1, k <=iter, k++,

D = J(n,f,random uniform(-1,1));
X = ME2FI(D);                                                    //creates
    model matrix for 2FI


/***************************/
//Coordinate Exchange//
/***************************/
delta = delta + random uniform(0,.001); //random step adjustment
```

```
X_cand1 = X;
X_cand2 = X;

//'convergence' will capture when the design is not increasing meaningfully
convergence = 1000;          //initially set at high value


while(convergence >= gap,
X_current = X;

//point (row) j
for(j = 1, j <= n, j++,

//optimize point j, coordinate i
for(i = 1, i <= f, i++,

//find direction of improvement, make initial replacement
store_vals = J(2,1,0);

store_vals[1,1] = X[j,i+1] - delta;
store_vals[2,1] = X[j,i+1] + delta;


X_cand1[j,i+1] = store_vals[1,1];
X_cand2[j,i+1] = store_vals[2,1];

//update interaction term of candidate matrices
X_cand1[j,4] = X_cand1[j,2]*X_cand1[j,3];
X_cand2[j,4] = X_cand2[j,2]*X_cand2[j,3];

flag = 0;

if(criterion_val(X_cand2, lambda, scale, lp_max) >= criterion_val(X_cand1,
   lambda, scale, lp_max),
//replaces best coordinate into X
X[j,i+1] = X_cand2[j,i+1];
flag = 1,
X[j,i+1] = X_cand1[j,i+1];

);

//update model matrix
X[j,4] = X[j,2]*X[j,3];

//update candidate matrices
X_cand1[j,i+1] = X[j,i+1];
X_cand2[j,i+1] = X[j,i+1];

//update interaction term of candidate matrices
X_cand1[j,4] = X_cand1[j,2]*X_cand1[j,3];
```

```
X_cand2[j,4] = X_cand2[j,2]*X_cand2[j,3];

improve = 1000;

if(flag == 1,

while(improve > 0,

Z_initial = criterion_val(X, lambda, scale, lp_max);

X[j,i+1] = X[j,i+1] + delta;

//update interaction coordinate
X[j,4] = X[j,2]*X[j,3];


//stop check for feasibility
if(X[j,i+1] > 1,
X[j,i+1] = 1;
flag = 2;
//update interaction coordinate
X[j,4] = X[j,2]*X[j,3];
improve = -1,
improve = criterion_val(X, lambda, scale, lp_max)-Z_initial;
);

),

while(improve > 0,
Z_initial = criterion_val(X, lambda, scale, lp_max);

X[j,i+1] = X[j,i+1] - delta;

//update interaction coordinate
X[j,4] = X[j,2]*X[j,3];


//stop check for feasibility
if(X[j,i+1] < -1,
X[j,i+1] = -1;
flag = 2;
//update interaction coordinate
X[j,4] = X[j,2]*X[j,3];
improve = -1;
improve = criterion_val(X, lambda, scale, lp_max)-Z_initial;
);

);

//set value to actual optimal coordinate
```

```
if(flag == 1,
X[j,i+1] = X[j,i+1] - delta,
flag == 0,
X[j,i+1] = X[j,i+1] + delta;
); //end IF

//update interaction coordinate
X[j,4] = X[j,2]*X[j,3];

);

//update candidate matrices
X_cand1[j,i+1] = X[j,i+1];
X_cand2[j,i+1] = X[j,i+1];

//update interaction term of candidate matrices
X_cand1[j,4] = X_cand1[j,2]*X_cand1[j,3];
X_cand2[j,4] = X_cand2[j,2]*X_cand2[j,3];

)

);

convergence = criterion_val(X, lambda, scale, lp_max)-criterion_val(
   X_current, lambda, scale, lp_max);
);

Z_current = criterion_val(X, lambda, scale, lp_max);

if(Z_current > Z_best,
X_best = X;
Z_best = Z_current
);

);


/***************************/
//PART 3: K-Pass Point Exchange//
/***************************/

//define candidate set
C = X_best;
X_PEA = X_best;

store_vals = J(1,n,0);

//set number of PE cycles w/value of K
K = 1;
```

```
//K-Pass Point Exchange
for(l = 1, l<=K, l++,

for(j = 1, j <=n, j++,

//optimize row (point) j
for(i=1, i <= n, i++,
X_PEA[j,0] = C[i,0];
store_vals[i] = criterion_val(X_PEA, lambda, scale, lp_max);
);

X_best[j,0] = C[Loc Max(store_vals),0];
X_PEA[j,0] = X_best[j,0];

);
);


/*****************************/
//OUTPUT//
/*****************************/
as table(X_best);
```

APPENDIX C

PARAMETER SETS FOR ROBUST D_{MP}-OPTIMAL DESIGN STUDY

C.1 Parameter Set for Robust D_{MP}-Optimal Design Creation

β_0	β_1	β_2	β_0	β_1	β_2
0.235196	1.693886	1.084647	-0.44658	2.508028	0.706472
-0.4623	4.06073	1.017948	0.158406	3.902771	1.12015
-0.02395	4.892003	0.925021	0.826384	1.074975	0.580963
-0.71118	4.08233	0.556959	-0.94841	2.183921	1.300262
-0.63581	3.028928	1.140743	0.481252	1.898971	0.762102
0.195566	3.936117	1.257606	-0.30024	3.680996	0.576312
0.387921	4.97872	0.581342	-0.3372	1.018231	1.041765
-0.08288	2.786654	1.476906	-0.38912	4.126964	0.995587
0.699839	4.071851	0.939447	0.218024	3.117467	0.981682
-0.59562	1.648262	0.803959	0.289734	3.470721	0.953932
-0.11212	1.139722	0.883137	-0.70358	1.337518	0.500717
0.122523	4.728434	1.148501	0.925787	1.505506	0.943916
-0.33961	2.835342	0.833925	-0.0392	2.768391	1.490868
0.342105	2.528545	1.09869	-0.32307	4.181134	1.036226
0.498599	2.947217	1.265499	0.994863	1.575146	1.267535
0.421867	1.403616	0.622969	0.374306	3.925389	1.297518
-0.594	1.221567	1.062844	-0.8912	4.899966	0.953054
0.441108	3.055367	0.933921	0.698959	4.276264	0.593014
-0.20243	2.515835	1.403703	-0.12682	1.477163	0.559923
-0.70161	4.075445	1.267963	0.27128	1.234769	1.006315
-0.05694	2.641285	1.176301	-0.6612	4.125268	1.289849
0.688885	4.791929	0.549649	-0.69749	1.304829	0.887217
0.733205	4.158606	1.15926	-0.56547	4.658243	0.500277
-0.40437	1.15729	0.770525	0.465919	3.055999	1.295497
-0.59888	1.463734	0.883695	0.748599	2.55358	1.119004
0.772791	2.987201	1.40341	-0.8657	2.411086	1.011033
-0.02309	4.19163	0.7772	0.075563	3.185345	1.153544
0.184725	3.216093	1.230974	-0.94136	3.41653	1.113456
0.569842	1.536249	1.227263	0.059192	3.347826	0.834326
0.121757	1.053523	0.919024	0.317646	4.028766	0.733572
0.820055	1.566322	0.667019	0.27632	4.74453	0.600618
0.255708	2.921891	1.327281	-0.96402	1.901129	1.096714
0.285016	1.931018	1.303792	0.467476	4.841881	1.13905
-0.07877	1.736063	0.722935	-0.24871	4.851488	0.547996
0.835797	2.886225	0.662579	-0.48562	3.172545	0.593358
-0.6063	4.315353	0.931696	-0.41748	2.542791	1.373256
0.333882	2.638313	0.676555	-0.16789	2.916134	0.533272
-0.45054	1.55051	1.125007	0.377746	4.123676	0.912339
0.516828	3.757842	0.545553	0.354387	2.933581	0.904225
-0.15387	1.50927	1.280957	-0.81118	4.398756	1.19495
-0.57573	3.613577	0.837015	-0.80703	3.060506	1.215416
-0.38296	3.718639	0.764607	-0.68454	3.649981	1.203261
-0.52163	1.224079	1.030352	-0.43679	2.545266	1.07792
-0.86972	1.283832	0.798645	0.466768	2.592676	0.869787
-0.15065	2.262993	1.394277	0.073502	3.06434	1.27639
-0.5899	2.400975	1.152342	-0.74944	2.361922	0.876601
0.210721	2.312119	1.48715	-0.10373	3.20206	0.624355
-0.60933	3.311225	1.019628	-0.03695	2.112405	1.438344
-0.19374	2.925776	0.792661	0.826021	1.234711	0.995127
0.968026	1.711963	0.640488	0.684562	4.590801	1.288071

C.2 Parameter Set for Robust D_{MP}-Optimal Design Validation

β_0	β_1	β_2	β_0	β_1	β_2
0.162702	2.090028	0.814753	-0.46793	1.32297	1.096285
-0.76853	1.42351	1.449295	-0.23683	3.23565	1.30904
-0.45007	2.65309	0.547824	0.888944	1.307361	0.635252
0.958598	4.83382	1.464128	0.990651	4.656952	0.698461
0.610299	2.806468	1.405171	0.003209	2.240131	1.098781
0.015202	1.012337	0.957329	-0.4836	4.661544	1.298019
0.086492	4.543316	1.234511	0.817524	3.960314	0.824019
-0.23439	3.149925	1.450221	0.171624	4.99165	1.059096
-0.49976	3.799447	0.733425	0.932376	4.520087	1.238303
0.653043	3.7868	1.477642	0.669595	3.116454	0.796588
-0.15981	4.372185	1.492175	0.528075	4.017128	0.504927
0.514292	1.949356	1.119064	-0.38338	2.316363	0.750547
-0.02472	3.086975	0.877001	0.049618	2.49217	0.754143
-0.54155	4.669534	1.264742	-0.828	2.285089	0.970846
-0.69285	4.412615	1.387295	-0.00732	1.425584	0.841478
0.030813	2.117408	0.602761	0.732699	3.662541	0.867673
-0.59242	3.116412	0.950754	-0.08308	2.309687	0.595519
0.956099	2.533086	0.904305	-0.7395	3.337113	1.026998
-0.66716	1.669347	0.742461	0.084322	4.976284	0.559199
-0.14884	4.165445	0.539217	0.222802	2.352039	0.899927
-0.23049	1.927926	0.691681	0.350755	2.474787	1.159493
-0.13029	3.364065	0.831386	0.883378	1.256573	0.820379
-0.67691	1.913647	0.814841	0.827299	1.549434	1.190276
0.206803	3.938584	0.755941	0.391284	1.098513	0.551707
-0.18491	4.672441	0.844198	0.785435	2.183432	0.515926
-0.40227	2.303303	0.919475	0.032149	2.958915	1.144441
-0.06897	1.74316	0.942075	-0.26556	2.419449	1.250504
0.632274	4.58242	0.504412	0.249405	1.48252	0.810619
-0.58465	4.362132	0.849091	0.980058	1.72058	1.018685
-0.08433	4.921329	0.814473	-0.5698	3.622793	1.285999
0.085286	4.201745	0.993986	-0.83171	4.780982	1.487591
0.237145	4.664184	0.584948	0.801983	1.731309	1.078621
-0.11639	3.747349	0.579318	0.45769	3.4403	0.955467
-0.41699	4.757876	1.061617	0.896015	2.153468	1.023364
0.586799	1.937004	0.941381	-0.80327	1.174035	0.50997
0.346118	4.313012	0.636216	0.353801	4.984764	1.495879
-0.59294	3.530721	0.631435	0.819505	1.944454	0.595217
0.507564	2.272885	0.526722	0.737154	4.971926	1.188977
0.870561	2.154253	0.585291	-0.54252	3.764857	0.956185
0.10331	2.608515	0.754494	0.375539	1.826951	0.592741
-0.59137	1.187741	1.331684	0.92779	3.387664	1.434579
-0.6084	4.895589	1.056452	-0.55727	2.76129	1.099103
-0.5883	2.705976	0.683913	0.262717	1.45207	1.188725
0.931768	1.24873	1.077674	0.602306	2.423348	0.529544
-0.88782	3.24976	0.550679	-0.90423	2.592895	1.234608
0.197403	4.883651	1.178857	0.310501	4.673452	1.439146
-0.59885	2.953627	1.225704	0.63666	2.325501	0.69007
0.893192	2.883675	1.058728	0.091324	2.886465	0.67179
0.633682	4.335655	1.417743	-0.86441	3.154462	0.534525
-0.18215	2.89868	0.556881	0.179199	1.194065	1.12394