Optimal Experimental Designs for the Poisson Regression Model in Toxicity Studies

Yanping Wang

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Eric P. Smith, Co-chair
Keying Ye, Co-chair
Christine Anderson-Cook
Jeffrey B. Birch
Raymond H. Myers

July 23, 2002
Blacksburg, Virginia

Keywords: Experimental Design, Design Optimality, Poisson Regression Model

Optimal Experimental Designs for the Poisson Regression Model in Toxicity Studies

Yanping Wang

(ABSTRACT)

Optimal experimental designs for generalized linear models have received increasing attention in recent years. Yet, most of the current research focuses on binary data models especially the one-variable first-order logistic regression model. This research extends this topic to count data models. The primary goal of this research is to develop efficient and robust experimental designs for the Poisson regression model in toxicity studies.

D-optimal designs for both the one-toxicant second-order model and the two-toxicant interaction model are developed and their dependence upon the model parameters is investigated. Application of the D-optimal designs is very limited due to the fact that these optimal designs, in terms of ED levels, depend upon the unknown parameters. Thus, some practical designs like equally spaced designs and conditional D-optimal designs, which, in terms of ED levels, are independent of the parameters, are studied. It turns out that these practical designs are quite efficient when the design space is restricted.

Designs found in terms of ED levels like D-optimal designs are not robust to parameters misspecification. To deal with this problem, sequential designs are proposed for Poisson regression models. Both fully sequential designs and two-stage designs are studied and they are found to be efficient and robust to parameter misspecification. For experiments that involve two or more toxicants, restrictions on the survival proportion lead to restricted design regions dependent on the unknown parameters. It is found that sequential designs perform very well under such restrictions.

In most of this research, the log link is assumed to be the true link function for the model. However, in some applications, more than one link functions fit the data very well. To help identify the link function that generates the data, experimental designs for discrimination between two competing link functions are investigated. T-optimal designs for discrimination between the log link and other link functions such as the square root link and the identity link are developed. To relax the dependence of T-optimal designs on the model truth, sequential designs are studied, which are found to converge to T-optimal designs for large experiments.

Dedication

To my parents and my dear wife, Quan.

Acknowledgments

I would like to express my deepest appreciation to my advisors, Dr. Eric P. Smith and Dr. Keying Ye, for sharing their knowledge with me. Their mentoring, support and discipline throughout this endeavor have been invaluable. I would also like to thank Dr. Christine Anderson-Cook and Dr. Jeffrey B. Birch for their time and helpful suggestions. My special thanks go to Dr. Raymond H. Myers for his guidance and encouragement over all these years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Literature Review

1.1 Background and Motivation

Over the past few decades, a large portion of scientific research has been driven by environmental concerns and health issues. Environmental studies focus on assessing the effect of pollutants on the delicate balance of nature's ecosystems while much of the research in the health arena centers on eradication of serious illness plaguing the world's populations. The toxicity study is a common thread that runs through both environmental and health-related research.

A typical toxicity study involves application of different concentrations of the toxicants under investigation to organisms and then measuring the effect. Indicators of effect include mortality, growth effect, reproduction impairment and tumor incidence. Very often, these responses cannot be appropriately modelled as normal random variables in the analysis of effect. Two representative examples are binary responses and count responses. In such situations, classical linear models are not applicable and generalized linear models are usually used for analysis.

Generalized linear models for both binary and count data have been fully studied in the literature (McCullagh and Nelder, 1987; Myers, Montgomery and Vining, 2002). Recently, experimental design for binary data models, especially for the logistic regression model, has received increasing attention and many publications have addressed this issue from different aspects (see Section 1.3). However, very little has been done on efficient experimental designs for count data models such as the Poisson regression model. *In this research, we try to fill this gap by developing efficient and robust experimental designs for the Poisson regression model in toxicity studies.* More specifically, we are mostly interested in the negative effect of some toxicant(s) on some count biological endpoints. For example, in the cancer study, the ability of an anticancer drug to eradicate tumor cells with high proliferative potential is considered one of the most important characteristic of its effectiveness in leading

to cure or long-term remission in a patient. Therefore, in-vitro colony-formation assays have been widely used to study drug effects. In such an assay, a dose-response curve is determined from the decrease in colony formations over different concentrations of the drug. Often, the number of colonies is assumed to be a Poisson variable and the Poisson regression model is used to describe the relationship between the colony count and the concentration or dose of the drug (see Minkin, 1991; 1993). A second example occurs in environmental research, in which researchers need to determine how some chemicals in the water affect the reproduction of some fish. Poisson regression is used to model the relationship between the number of fish eggs and the concentration of the chemicals. Similar examples abound in the literature (see Oris and Bailer, 1993; Van Mullekom and Myers, 2001).

1.2 Design Optimality

Since our interest is optimal experimental design, we shall first give a brief introduction of design optimality criteria. In 1959, Kiefer and Wolfowitz formed a theoretical framework for optimal design criteria by expressing a design as a probability measure representing the allocation of observations at any particular point in the design space. Their theoretical approach to design optimality and their introduction of the D- and E-optimality criteria for the linear regression model laid the groundwork for other design criteria like A-, F-, G- and Q-optimality criteria. Each design optimality criterion addresses a specific goal in the experiment to be performed or achieves a specific property in the final fitted regression model. Many of these design criteria were originally developed for the homogeneous variance linear model, but most have been adapted for use in nonlinear and nonhomogeneous variance situations as well.

The basic idea underlying design optimality theory is that statistical inference about quantities of interest can be improved by "optimally" selecting levels of the control variables. In general, a design optimality criterion can be characterized as an "estimation criterion" or a "prediction criterion". An experimental design which is optimal with respect to an estimation criterion like D-optimality is the one that maximizes parameter information by minimizing variability of the parameter estimates. A design which is optimal with respect to a prediction criterion like Q-optimality maximizes information about a response surface by focusing on the prediction qualities of the fitted model.

Silvey (1980) and Atkinson and Donev (1992) described most commonly-used design criteria. In this research, we will mostly concentrate on D-optimal designs so D-optimality will be introduced here in detail. Some other optimality criteria like D_s-optimality and T-optimality will also receive considerable attention and they will be introduced later in the appropriate context.

The most well known and widely used design optimality criterion is D-optimality, dating to Kiefer and Wolfowitz (1959). If the model parameter vector is denoted by $\boldsymbol{\beta}$,

D-optimality suggests that the choice of design should maximizes the information on $\boldsymbol{\beta}$ by minimizing the generalized variance of its estimate. We know that a common method for estimating $\boldsymbol{\beta}$ in many nonlinear models is maximum likelihood estimation (MLE). If $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$, then the asymptotic variance matrix of $\hat{\boldsymbol{\beta}}$ is the inverse of the Fisher information matrix (see Lehmann, 1983). This information matrix forms the foundation for traditional design optimality criteria and is defined in (1.1). Obviously, if the dimension of $\boldsymbol{\beta}$ is $p \times 1$, the Fisher information matrix $I(X, \boldsymbol{\beta})$ is a $p \times p$ matrix.

$$I(X, \boldsymbol{\beta}) = -E\left[\frac{\partial^2 \log(L(X, \boldsymbol{\beta}))}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}\right], \tag{1.1}$$

where $L(X, \boldsymbol{\beta})$ is the likelihood function for the data and X is the design matrix. The D-optimality criterion, which minimizes the generalized variance of $\hat{\boldsymbol{\beta}}$, equivalently maximizes the determinant of the Fisher information matrix. The general D-optimality criterion is defined as

$$\max_{X \in \mathcal{D}} \left| \frac{I(X, \boldsymbol{\beta})}{n} \right|, \tag{1.2}$$

where $\mathcal{D}$ is the set of all possible designs and n is the sample size. Very often, the sample size n is fixed so the D-optimal design is obtained by maximizing the determinant of the Fisher information matrix. Essentially, the D-optimal design minimizes the $100(1-\alpha)\%$ confidence ellipsoid on the model parameters when the parameter estimates are asymptotically normal (Myers and Montgomery, 1995). The volume of the confidence region is relevant because it reflects how well $\boldsymbol{\beta}$ is estimated. Obviously, a smaller confidence region implies better estimation.

Silvey (1980) discussed many general properties of D-optimal designs and one of them is mentioned here. It follows from Carathéodory's Theorem (Silvey, 1980, p.72) that a continuous D-optimal design is based on a finite number of support points and, more specifically, on at most $p(p+1)/2$ points, where p is the number of model parameters. As an aside, it is interesting to observe that D-optimal designs are in fact very commonly based on exactly p points and that in such cases the weights associated with the design points are necessarily equal and thus equal to $1/p$ (Silvey, 1980, p.42), namely, $1/p$ of the total sample size is allocated to each of the p points.

Since the D-optimal design maximizes the determinant of the Fisher information matrix, it is natural to define the D-efficiency of an arbitrary design, X, as

$$\text{Eff}_D = \left(\frac{|I(X, \boldsymbol{\beta})|}{|I(X^*, \boldsymbol{\beta})|}\right)^{1/p}, \tag{1.3}$$

where X^* is the D-optimal design of the same size as X and p is the number of model

parameters. When D-optimality is employed, D-efficiency defined in (1.3) is generally used to compare different designs. Taking the ratio of the determinants in (1.3) to the $(1/p)^{th}$ power results in an efficiency measure which is proportional to experiment size, irrespective of the dimension of the model. Therefore, two replicates of a design for which $\text{Eff}_D = 0.5$ would be as efficient as one replicate of the optimal design.

For the homogeneous variance linear model, the Fisher information matrix is simply $X'X$, which depends only upon the location of the design points, not on $\boldsymbol{\beta}$. Consequently, the D-optimal design can be determined and implemented independently of $\boldsymbol{\beta}$. However, for a nonlinear model like the generalized linear model, the information matrix usually depends on the model parameters as well as the design matrix. More specifically, the information matrix, apart from a constant, for the generalized linear model is

$$I(X, \boldsymbol{\beta}) = X'WX, \tag{1.4}$$

where W is the Hessian weight matrix that depends on the unknown parameters (see McCullagh and Nelder, 1987). For a canonical generalized linear model like the logistic regression model, (1.4) reduces to

$$I(X, \boldsymbol{\beta}) = X'VX, \tag{1.5}$$

where V is a diagonal matrix and the $(i, i)^{th}$ element is the variance of the ith observation. We know that these variances depend on $\boldsymbol{\beta}$, so does the information matrix. Therefore, the D-optimal design for a generalized linear model is parameter-dependent. In order to find and implement the D-optimal design, one must know the model parameters. It is this dependence on parameters that complicates the problem of experimental designs for generalized linear models. Several methods have been proposed to circumvent this difficulty in the literature and we shall give an overview in next section.

1.3 Optimal Designs for Logistic Regression Model

As discussed earlier, optimal experimental designs for generalized linear models depend on the unknown parameters and the problem of their construction is therefore necessarily more complicated than that for classical linear models. In this section, we shall introduce some methods that have been proposed to overcome this difficulty. In the literature, the majority of current research on experimental designs for generalized linear models has been confined to the logistic regression model. Therefore, we shall go through these methods largely in the context of logistic regression model.

1.3.1 Locally Optimal Designs

A simple approach to the problem of experimental designs for generalized linear models is to adopt a best guess for the parameters and then choose the design that maximizes a selected design optimality criterion function evaluated at the guess. This approach leads to what is termed *locally optimal design*, the term introduced by Chernoff (1953). This best guess used in a locally optimal design might come from the previous experimentation, or from a pilot experiment conducted specially for this purpose, or merely a guess. We shall call this parameter guess the initial estimate or the initial guess no matter how it is obtained.

We now give a brief review of locally optimal designs for the logistic regression model. Most publications on this topic focus on the one-variable first-order logistic regression model defined as

$$y \sim Bernoulli(P), \text{ where } P = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x))}. \tag{1.6}$$

In the above model, y is the binary response (0 or 1); it has the probability P to take the value 1 and P is a function of the independent variable x. For this model, Kalish and Rosenberger (1978) developed a two-level D-optimal design. They imposed symmetry around the ED_{50} (ED is short for effective dose, ED_{100p} represents the value of x which produces the probability of p that the response is 1). The two-level D-optimal design is obtained by placing half of the experimental runs at $\text{ED}_{17.6}$ and the other half at $\text{ED}_{82.4}$. To implement the D-optimal design, one must know the values of $\text{ED}_{17.6}$ and $\text{ED}_{82.4}$, or equivalently, one should have knowledge about β_0 and β_1. In practice, the parameters are generally unknown and one needs to provide some initial guesses.

Abdelbasit and Plackett (1983) reviewed the results of two-level designs and further pursued the D-optimal design for model (1.6) by considering three-level designs. They also assumed symmetry around ED_{50}. The three-level equal sample size D-optimal design, symmetric around the ED_{50}, is formed by the doses $\text{ED}_{13.6}$, ED_{50} and $\text{ED}_{86.4}$. Minkin (1987) then followed by showing that the previously created symmetric designs (both two-level and three-level) remain optimal for a general class of designs, eliminating the restriction of symmetry.

Estimating a particular ED, especially the ED_{50}, is of great interest in many applications. Several procedures have been suggested for the construction of optimal designs for estimating ED_{50}. One procedure minimizes the asymptotic variance of the estimate of ED_{50}, which is $-\hat{\beta}_0/\hat{\beta}_1$ ($\hat{\beta}_0$ and $\hat{\beta}_1$ are estimates of β_0 and β_1, respectively). Kalish (1990) demonstrated that the optimal design based on this procedure is degenerate in that it requires all of the observations to be taken at a single point, the ED_{50}. Kalish (1990) further considered two-point designs based on this procedure. It was shown that the two-point design is not very efficient relative to the D-optimal design. However, the D-optimal design is moderately efficient for estimating the ED_{50}. A second procedure is the so-called F-optimal design, a

design that minimizes the length of the Fieller interval for the ED_{50}. Finney (1971) claimed that the Fieller interval is superior to the asymptotic confidence interval. Based upon a limited empirical study, Abdelbasit and Plackett (1983) found Finney's claim to be unfounded. However, Sitter and Wu (1993) conducted a more extensive empirical study and concluded in favor of Finney. Letsinger (1995) gave a derivation of the Fieller interval and the resulting F-optimal design for model (1.6). The F-optimal design is a symmetrical two-point design and its structure depends on the total sample size. The F-optimal design for ED_{50} has the same number of experimental runs at each point but it is not the case for other ED's (see Letsinger, 1995). Another design criterion for estimating a particular ED is to minimize the length of the likelihood-based confidence interval. Williams (1986) argued that, for small experiments (20 to 30 experimental runs), the coverage of the likelihood-based confidence interval is closer to the nominal level than that based on Fieller's Theorem. Minkin and Kundhal (1999) investigated the optimal design for model (1.6) using this procedure. Under the restriction of symmetry, they found that the optimal design for estimating the ED_{50} is also a two-point design depending on the total sample size.

For model (1.6), locally optimal designs based on prediction optimality criteria also receive some attention in the literature. Kalish and Rosenberger (1978) developed a two-level G-optimal design, which minimizes the maximum variance of the predicted probability $\hat{P}$ over the region of interest. Myers, Myers and Carter (1994) studied two-point and three-point G-optimal designs symmetric about the ED_{50} for some selected regions of interest. In addition, they developed and studied Q-optimal design for model (1.6), which minimizes the average prediction variance over the region of interest. They considered two-point and three-point Q-optimal designs with and without the constraint of symmetry about ED_{50}. They also compared various designs by calculating their G-efficiency and Q-efficiency.

Other work on locally optimal designs for the logistic regression model includes Heise and Myers (1994), who investigated locally D-optimal and Q-optimal designs for the bivariate logistic regression model; Jia (1996), who extended the topic to the two-variable logistic regression model.

Intuitively, we would expect that locally optimal designs are quite efficient if the initial guesses are close to the true parameters. However, it has been frequently demonstrated that locally optimal designs are not efficient for poor initial estimates. This is the problem of *parameters misspecification*, a term frequently used in the literature. Basically, this is the major problem that we have to face when dealing with any nonlinear experiment. Most research in this area actually focuses on designs robust to parameter misspecification.

One might argue that the locally optimal design for the generalized linear model, though simple, has limited applications due to the lack of good initial estimates. It is true that good initial estimates are seldom available in practice. However, the locally optimal design is extremely important for two reasons: (1) there do exist some situations in which practitioners have reliable knowledge of the parameters; (2) most importantly, the locally optimal design provides a useful benchmark with which we are able to calibrate any other

design. Therefore, most research in this area has focused on locally optimal designs.

1.3.2 Bayesian Optimal Designs

An alternative and, in a sense, a more realistic approach to designs for generalized linear models is to introduce a prior distribution on the parameters rather than a single guess. A Bayesian optimal design is the one that maximizes the mean of an appropriate optimality criterion function over the prior distribution. For example, if we denote the prior distribution by $\pi(\boldsymbol{\beta})$, the Bayesian D-optimality criterion is

$$\max_{X \in \mathcal{D}} \int |I(X, \boldsymbol{\beta})| \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}, \tag{1.7}$$

where $\mathcal{D}$ is the set of all possible designs. Actually, the locally optimal design can be regarded as a special type of Bayesian optimal design, for which the prior distribution degenerates to a point prior at the initial guess.

For the logistic regression model, Chaloner and Larntz (1989) gave a formal discussion on Bayesian optimal designs. The Bayesian D-optimality criterion and the Bayesian optimality criterion for estimating a particular ED are derived in their seminal paper. They developed Bayesian optimal designs using independent uniform prior distributions for the parameters. They found that, in general, the number of design points in a Bayesian optimal design increases as the uncertainty in the prior distribution increases. This is desirable since a design of more points is more robust. They also compared the efficiency of the locally optimal design and the Bayesian optimal design. The Bayesian optimal design is found to be more robust to poor initial parameter information than the locally optimal design. As an extension of their work, Letsinger (1995) studied the Bayesian D-optimal design using both normal and uniform priors for the logistic regression model. He found that the robustness of the Bayesian D-optimal design increases as the prior spreads, however, the efficiency of the Bayesian design decreases as the prior variance increases. The trade-off is between protection against poor initial parameter knowledge and high efficiency when the parameter knowledge is good.

Other work on Bayesian optimal designs includes Atkinson, et al. (1993), who studied Bayesian optimal designs for a nonlinear regression model; Dumouchel and Jones (1994), who investigated the application of Bayesian D-optimal designs to reduce the dependence on the assumed model; and Minkin (1993), who discussed Bayesian optimal designs for the Poisson regression model. Chaloner and Verdinelli (1995) gave a nice historical review of Bayesian optimal designs.

Basically, the Bayesian optimal design optimizes the weighted average of a criterion function over a prior distribution. Thus, it is more robust than the locally optimal design which maximizes the criterion at a point guess. The premise of the Bayesian design is that the

researcher's belief about the parameters can be translated into a probability density for the parameters. Often, this "translation" is criticized for its subjectivity. A second weakness of the Bayesian design is that, when the initial information is scarce and a noninformative prior is used, the efficiency is usually very low. Another disadvantage of the Bayesian approach is that it generally involves more extensive computation then the locally optimal design. This is so because of the integration part in the Bayesian optimality criterion function.

1.3.3 Minimax Approach

For the logistic regression model, Sitter (1992) proposed a minimax approach to obtain designs that are robust to poor initial parameter estimates. To use this approach, the practitioner should provide initial guesses of the unknown parameters as well as a specific region of the parameters within which he or she wishes the design to be robust. The minimax procedure chooses the design that minimizes the maximum criterion value over the preselected region. Here, the design criterion is defined such that the worst case is when the criterion function takes the maximal value. For example, we can use determinant of the inverse of the Fisher information matrix for the D-optimality.

Basically, the minimax procedure, like G-optimality, yields the design with the best "worst case" over the chosen region so that, hopefully, the worst case is not too bad. This approach, though straightforward to understand, is mathematically intractable and numerically difficult. Sitter (1992) studied minimax designs for the logistic regression model by imposing the restrictions of equal allocation and symmetry on the design. The minimax designs are robust to poor initial estimates of the parameters. The more uncertain the experimenter is in the initial parameter estimates, the more spread out is the minimax design, both in terms of coverage of design space and number of design points.

The minimax approach suffers from two major drawbacks. Firstly, as indicated earlier, it is mathematically intractable in general and some restrictions on the design are necessary to facilitate the numerical optimization. Secondly, this method, unlike the Bayesian optimal design, which optimizes the weighted average of the criterion function over the prior distribution, optimizes the worst case. So the minimax design is generally more conservative by its nature (protects against the worst case). It assumes all the points in the selected region are equally likely to occur, so points on the edges are considered as likely to occur as the middle points. Now suppose that the true parameters reside somewhere in the middle. The optimal criterion value, as one would expect, is usually found at or near the edges of the preselected region. If the region is reasonably small, this is not a problem. The efficiencies at and around the center of the region should be quite high. However, when the region is fairly large, one minimizes the maximum criterion value at a location far away from the true value so the minimax design sacrifices too much efficiency at and near the true value, in order to be efficient at a location which is quite possibly one of the most unlikely locations of the true parameters.

1.3.4 Sequential Designs

Another way of dealing with poor parameter guesses is to construct a design sequentially. In sequential experimentation, an initial experiment is chosen and implemented. Estimates of the parameters are obtained based on observations from the initial experiment. These estimates are then used to find the next design point, at which the following observation is made. This process is repeated until all the experimental resources are used up or some stopping condition is met. Compared to the locally optimal designs, a sequential design uses more reliable information about the parameters so it can improve the design efficiency in general.

The major drawback of a fully sequential design (one run at a time) is that it incurs more cost than a non-sequential design, especially when it takes a long time to obtain an observation. To reduce the time of experimentation, stage-wise designs are proposed. In a stage-wise design, a batch of experimental runs, instead of one run, are implemented at each stage. Two-stage designs and three-stage designs are typical stage-wise designs.

Box and Lucas (1959) suggested sequential experimentation. Abdelbasit and Plackett (1983) discussed two-stage and three-stage D-optimal designs for a binary data model. Their conclusions include: (1) sequential designs are robust to poor parameter guess in general and higher-stage designs are more efficient; (2) the efficiency of stage-wise designs increases with the experiment size; (3) for larger experiments, the efficiency is almost independent of the number of stages and is not much less than 1 anywhere. Wu (1985) gave sequential designs for binary data for estimating a particular ED based upon the nonparametric "up and down" method and the Robbins Monro method. Minkin (1987) proposed a two-stage D-optimal design procedure for the logistic regression model and demonstrated that the two-stage design is more robust than the locally optimal design.

Myers et al. (1996) developed a two-stage D-Q optimal design for the logistic regression model. The first stage is a locally D-optimal design and the second stage is a conditional Q-optimal design, conditioned on the estimates from the first stage. They found the two-stage design to be superior to the comparable one-stage design in all cases of parameter guesses except those extremely close to the true values. Much efficiency is gained when parameter knowledge is poor and little is lost when the guesses are correct. Letsinger (1995) argued that the performance of a two-stage design can be improved by using the Bayesian optimal design at the first stage. He studied the performance of the two-stage D-D optimal design for the logistic regression model and reached similar conclusion. Sitter and Wu (1999) also discussed a two-stage procedure for the logistic regression model. At the first stage, they proposed to use a 3 to 5 point symmetrical design to improve the robustness. They argued that any decent methodology for two-stage experimentation should allow the experimenter to evaluate the necessity of a second-stage after the completion of the first-stage analysis. Particularly, they pointed out that, in addition to the point estimates, the variance of the estimates in the first stage and the estimated coverage of the first-stage design in the design space are very useful for making decisions on the second-stage design.

One difficulty with the stage-wise design procedure is how to optimally distribute the total sample size to each stage. The optimal sample allocation depends on many factors including design of the first stage, total sample size, and the quality of the initial guesses. Further, the dependence is not clear and vary hard to quantify so it is practically impossible to analytically find the optimal sample allocation. For the logistic regression model, Myers et al. (1996) found in a limited simulation study that about 30% of the total sample is appropriate for the first stage of the two-stage D-Q design. This translates to expending 30% of the sample for the purpose of finding good initial estimates of the parameters and 70% of the sample devoted solely to reaching the final goal of the experiment through the second-stage criterion. For the same model, Letsinger (1995) did a much more extensive sample allocation study for the two-stage D-D procedure using Bayesian design with normal priors at the first stage. He recommended that one should use somewhere between 40% and 50% of the sample in the first stage for a small total sample size of 30. For a large sample size of 300, the recommendation ranges from 15% to 30% depending on the situation.

Other work on sequential designs includes McLeish (1990), who investigated sequential designs in quantal bioassay; Lin, Myers and Ye (2000), who discussed Bayesian two-stage design for mixture experiments; Schwartz et al. (2001), who studied the two-stage D-D$_s$ optimal design for a nonlinear toxicity threshold model.

1.4 Optimal Designs for Poisson Regression Model

In the last section, we reviewed current research on experimental designs for generalized linear models, most of which concentrates on the logistic regression model. As stated at the beginning, this research is mainly concerned with optimal designs for Poisson regression models in toxicity studies. In this section, we will discuss this topic. We shall first introduce models of interest and some notation. After that, current work on this topic in the literature will be reviewed and finally we will give an outline of our work in this research.

1.4.1 Models, Assumptions and Notations

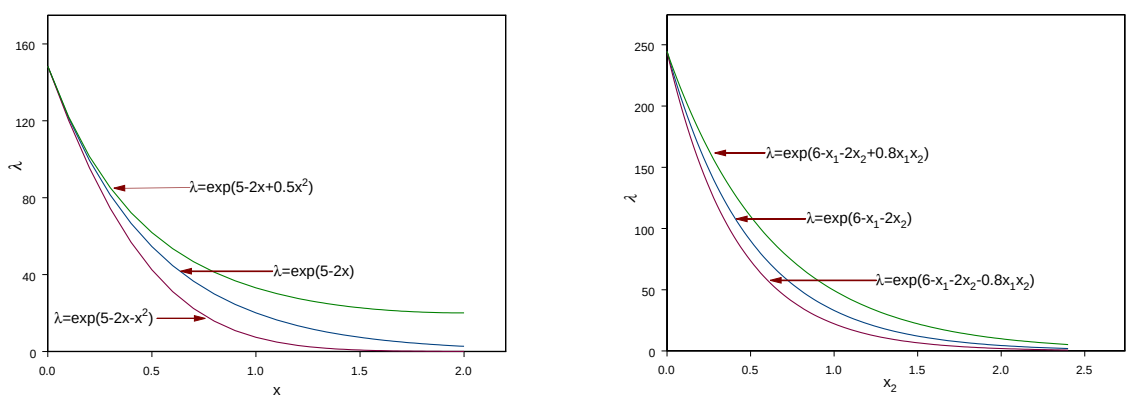
We shall first introduce Poisson regression models and some assumptions and notations that will be followed throughout this dissertation. Generally, the Poisson regression model can be written as:

$$y_{ij} \sim Poisson(\lambda_i), \text{ where } \lambda_i = exp(\mathbf{x}_i'\boldsymbol{\beta}). \tag{1.8}$$

As mentioned earlier, we are mainly interested in the negative effects of some toxicant(s) on a count biological endpoint that can be approximately modelled as a Poisson variable. Examples include the number of cancer cell colonies that survive an anticancer

treatment; the number of eggs that a fish lays after exposure to some pollutants. In such a context, we define y_{ij} in (1.8) to be the number of organisms or cells that survive the experiment for the jth replicate of the ith design point and we assume it follows a Poisson distribution with λ_i as the mean; $\mathbf{x}_i$ is the regressor vector at the ith point and $\boldsymbol{\beta}$ is the parameter vector. Equations (1.9)–(1.12) define the mean functions for the one-toxicant first-order model, one-toxicant second-order model, two-toxicant additive model and two-toxicant interaction model, respectively. Note that, in the context of toxicity studies, x's in the following equations are doses or concentrations of toxicants and thus are nonnegative.

$$\lambda_i = exp(\beta_0 + \beta_1 x_i). \tag{1.9}$$
$$\lambda_i = exp(\beta_0 + \beta_1 x_i + \beta_{11} x_i^2). \tag{1.10}$$
$$\lambda_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}). \tag{1.11}$$
$$\lambda_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i}). \tag{1.12}$$



(a) One-toxicant Model

(b) Two-toxicant Model (x_1 fixed)

Figure 1.1: Mean Function of Poisson Regression Model

For illustration, Figure 1.1 plots some mean functions of the above models. In Figure 1.1(a), one-toxicant models with different quadratic terms are compared. We can see that the a model with negative quadratic term describes a stronger effect than the first-order model does; on the other hand, a positive quadratic term means a weaker effect. In Figure 1.1(b), we plot three two-toxicant models with x_1 fixed at 0.5. One can see that a negative interaction between the two toxicants causes a larger combined effect than their additive effect; a positive interaction results in a combined effect less than the additive effect. This same pattern holds for other values of x_1.

Before we introduce some notation based on the above models, we need to make an assumption for simplicity. *Without loss of generality, we assume that all the toxicants under*

investigation have negative effects on the response and larger doses generally cause smaller mean responses. This assumption is held to be true throughout this dissertation unless otherwise stated. Though all the designs in this research are derived under this assumption, they can be easily generalized to applications that involve positive relationship between the response and regressors (see Chapter 6).

The following notation will be used throughout this dissertation unless otherwise stated:

- n denotes the total number of experimental runs and n_i is the number of runs at the ith point. The proportion of experimental runs at the ith point is denoted by p_i, i.e., $p_i = n_i/n$.
- λ_c denotes the mean response at the control point, at which there is no toxicant. Since we assume that toxicants always have some negative effect on response, λ_c is the maximum possible mean response in any experimental unit. For models (1.9)–(1.12), $\lambda_c = exp(\beta_0)$.
- λ_i denotes the mean response at the ith design point, i.e., $\lambda_i = exp(\mathbf{x}_i'\boldsymbol{\beta})$. Generally, $\lambda_i \leq \lambda_c$.
- In general, *effective dose*, or ED for short, of a toxicant is defined as the dose or the amount of the toxicant that causes a specified proportion of mortality or survival. For instance, ED_{30} of toxicant A is the dose of A that causes 70% mortality. For single-toxicant experiments, this definition is clear-cut. For multi-toxicant experiments, we need to distinguish the combined effect of all the toxicants and the contribution of each single toxicant. To this end, we define *model effective dose*, or MED for short, as all possible combinations of doses of all toxicants that cause a specified mortality or survival. For instance, in a two-toxicant experiment, MED_{30} is all possible combinations of doses of the two toxicants that cause 70% mortality. On the other hand, we define *individual effective dose*, or IED for short, of a toxicant as the dose of the toxicant that causes a specified mortality or survival. Clearly, for single-toxicant experiments, MED and IED are the same.
- At the ith design point, $q_i = \lambda_i/\lambda_c$ denotes the MED level. Note that q_i is nothing but the survival proportion at the point. For single-toxicant case, q_i is also the IED level. For multi-toxicant experiments, we use q_{1i} to denote the IED level of the first toxicant, q_{2i} to denote the IED level of the second toxicant, etc. Obviously, all these q's are between 0 and 1.

1.4.2 Optimal Designs: A Review of Current Work

We now give a review of the current work on optimal designs for the Poisson regression model.

For single-toxicant experiments, Minkin (1993) and Chiacchierini (1996) have done extensive work on locally optimal designs for the first-order model, i.e., model (1.9). For this model, Minkin (1993) studied the slope-optimal design (minimizing $var(\hat{\beta}_1)$, where $\hat{\beta}_1$ is the estimate of β_1) for estimating a particular ED. He found that the slope-optimal design is one with 22% experimental runs at the control point (i.e., $x = 0$) and the remaining 78% at $\mathrm{ED}_{7.8}$. He also examined the effect of poor initial estimates of the parameters on the optimal design and considered a Bayesian approach. Chiacchierini (1996) showed that the slope-optimal design for model (1.9) is also the F-optimal design (minimizing the width of the Fieller interval for a particular ED). In her work, Chiacchierini developed D-, Q- and F-optimal designs for model (1.9). All the optimal designs mentioned above are two-point designs and are summarized in Table 1.1, in terms of ED levels, q_i's, and proportions of experimental runs at each point, p_i's. Note that the Q-optimal design, which minimizes the average prediction variance on a specified region, depends on the region of interest and Table 1.1 gives Q-optimal designs for three different regions of interest. The design space is assumed to be $(\mathrm{ED}_0, \mathrm{ED}_{100}]$ for all the designs in Table 1.1.

Table 1.1: Some Optimal Designs for Model (1.9)

Optimality Criterion	Designs			
	Point 1		Point 2	
	p_1	q_1	p_2	q_2
D-optimal	0.50	1.0000	0.50	0.1353
F-optimal	0.22	1.0000	0.78	0.0780
Slope-optimal	0.22	1.0000	0.78	0.0780
Q-optimal				
$(\mathrm{ED}_0, \mathrm{ED}_{100}]$	0.44	1.0000	0.56	0.1870
$[\mathrm{ED}_{10}, \mathrm{ED}_{90}]$	0.40	1.0000	0.60	0.2120
$[\mathrm{ED}_{20}, \mathrm{ED}_{80}]$	0.37	1.0000	0.63	0.2500

To improve the robustness of the design to parameter misspecification, Van Mullekom and Myers (2001) studied Bayesian D-optimal designs for model (1.9). They found that only the prior distribution on β_1 has impact on the D-optimal design. They tried uniform and normal prior distributions for β_1 and derived two-point and three-point Bayesian D-optimal designs. These Bayesian optimal designs were shown to be more robust than corresponding locally optimal designs.

While some toxicity studies involve only one toxic substance, many of them are based on more complicated models which contain two or more toxicants. For example, the environmentalists may be curious about the effect of several chemicals on a specific animal population; the medical researcher may be interested in determining how a "cocktail" of drugs works to impair the reproduction of AIDS virus. Ultimately, these researchers want to know whether these substances interact and, if they do, how the interaction should be characterized.

Van Mullekom and Myers (2001) studied the D-optimal design for model (1.11). Basically, this model implies that each of the two toxicants under investigation, when applied alone, works through model (1.9); when working together, they do not interact with each other and their effects are additive. For this model, the IED level for the first toxicant is $q_1 = exp(\beta_1 x_1)$ at point (x_1, x_2); the IED level for the second toxicant is $q_2 = exp(\beta_2 x_2)$ and the MED level is $q = q_1 q_2$. Van Mullekom and Myers found that the D-optimal design for this model is a three-point design: the control point (IED_{100}, IED_{100}); and two pure component points ($\text{IED}_{13.53}$, IED_{100}) and (IED_{100}, $\text{IED}_{13.53}$), at which only one toxicant is applied, and each of these three points has 1/3 of the total experimental runs.

For model (1.12), Van Mullekom and Myers (2001) also did some work. Basically, model (1.12) implies that both toxicants, when applied alone, work through a first-order model as described in (1.9). However, when working together, their effects are not additive as they may interact with each other and the possible interaction is accounted for by the term β_{12}. For this model, the IED level is $q_{1i} = exp(\beta_1 x_{1i})$ for the first toxicant at the ith point and $q_{2i} = exp(\beta_2 x_{2i})$ for the other toxicant; the proportion of survival at this point, or MED level, is $q_i = exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i})$. In their work, Van Mullekom and Myers did not derive the D-optimal design for this model. Instead, they created a design called conditional D-optimal design, a design that is optimal only on the condition that β_{12} in model (1.12) is zero. They showed that the conditional D-optimal design is a four-point design with 1/4 of the total observations taken from each point. These four points are: the control point (IED_{100}, IED_{100}); two pure component points ($\text{IED}_{13.53}$, IED_{100}) and (IED_{100}, $\text{IED}_{13.53}$); and an interaction point ($\text{IED}_{13.53}$, $\text{IED}_{13.53}$), at which both toxicants are applied.

1.4.3 Optimal Designs: Extensions

In this research, we try to extend the current work on experimental designs for Poisson regression models in toxicity studies. Specifically, the extensions include the following points:

1. For single-toxicant experiments, we shall investigate the D-optimal design for the second-order model as defined in (1.10). Optimal designs for the first-order model have been extensively discussed in the literature. However, occasions occur that model (1.9) is not adequate to describe the dose-response relationship. Quadratic terms are found to be important in models of *Ceriodaphnia* toxicity (see Oris and Bailer, 1993). Second-order models are also needed in radiation research to quantify the relationship between the radiation dose and the count of chromosome aberrations (see Frome and DuFrain, 1986). Thus, it is of practical importance to study optimal designs for model (1.10). D-optimal and some practical designs for this mode will be investigated. D_s-optimal designs for estimating ED's are also studied.

2. For two-toxicant experiments, we will focus on the interaction model, i.e., model (1.12). The motivation is that this model is of great interest in practice as interactions between

different toxicants are expected in general. Actually, many studies are designed simply for investigation of the interactions (see Wahrendorf, Zentgraf and Brown, 1981). Therefore, model (1.12) will receive most attention throughout this research. We will first the D-optimal design for this model, which enables us to study the property of other designs like the conditional D-optimal design proposed by Van Mullekom and Myers. We will also try to generalize the D-optimal design to interaction models that involve more than two toxicants. For model (1.12), D_s-optimal designs also receive some attention.

3. As indicated earlier, robustness to parameter misspecification is a very important property of nonlinear experimental designs. To improve the robustness of locally D-optimal designs to parameter misspecification, we propose sequential designs for Poisson regression models. This includes fully sequential designs and two-stage designs. We will concentrate on sequential designs for model (1.12); however, we expect the results will be similar for other models.

4. Though the log link is usually used for Poisson regression, sometimes other link functions such as the square root link and the identity link provide a better fit. Optimal designs for discrimination between different link functions will be investigated. Firstly, locally optimal designs, which assume the true link and the parameters are known, are studied. Sequential designs are then investigated to relax this assumption.

Throughout this dissertation, the design space will be defined in terms of IED levels, i.e., q's. The advantage of using q's instead of x's is that the ranges of q's are well defined, namely, between 0 and 1. Also, we will just consider continuous designs unless otherwise stated, which means the designs will be found in terms of p_i's instead of n_i's. Algorithms to convert continuous designs to exact ones can be found in the literature (see Atkinson and Donev, 1992).

1.5 Structure of the Dissertation

This dissertation is organized as follows. Chapter 2 focuses on locally D-optimal and D_s-optimal designs for model (1.10) and model (1.12). Some practical designs for these two models like the equally spaced design and the conditional D-optimal design, which are intuitive to practitioners, also receive some attention in this chapter. The robustness of these practical designs to parameters misspecifications are investigated. In Chapter 3, we employ the general equivalence theory to verify the optimality of locally optimal designs in Chapter 2. This is necessary because numerical methods, rather than analytical approaches, are used to find the optimal designs so their optimality is in question. Taking advantage of the equivalence theory, we are able to generalize the D-optimal design to the multi-toxicant interaction model. As a solution to the problem of parameter misspecification, sequential designs for

the Poisson regression model are investigated in Chapter 4. Both fully sequential designs and two-stage designs are discussed in this chapter. Another component of Chapter 4 is sequential experimentation on restricted design space. A penalized D-efficiency is proposed for design evaluation on restricted regions. Chapter 5 is concerned with optimal designs for discrimination between competing link functions for Poisson regression models. Locally optimal designs and sequential designs are discussed. Finally, Chapter 6 lists some possible extensions for future research.

Chapter 2

Locally D-optimal Designs

2.1 Introduction

As discussed in Chapter 1, locally optimal designs for nonlinear experiments are of great importance for two reasons. Firstly, if good initial estimates are available, locally optimal designs are very efficient. Secondly and most importantly, locally optimal designs provide a useful benchmark with which we can calibrate any other design. Therefore, we start our research with locally D-optimal designs for Poisson regression models.

To implement locally optimal designs, we have to provide some initial information about the parameters. In this chapter, we assume some prior knowledge about the toxicants under investigation is available. More specifically, we assume practitioners have good estimates of ED's for the toxicants. Though the assumption of availability of good initial estimates does not always sound reasonable, there do exist situations where it is true. For instance, in the second and third phases of clinical trial experiments, knowledge about the drug is generally available from previous phase(s). Another example is that, when studying the interaction of two toxicants, researchers might know about the performance of each of the two toxicants from previous experiments.

Locally D-optimal designs for the one-variable second-order model as in (1.10) and the two-variable interaction model as in (1.12) will be investigated as they both are of practical importance. D_s-optimal designs that minimize the generalized variance of all the parameter estimates except the intercept term also receive some attention in this chapter as it is felt that D_s-optimal designs of this type are efficient for ED estimation. Notation in Chapter 1 will be followed and all the designs in this chapter will be found in terms if q_i's and p_i's instead of x_i's and n_i's.

2.2 One-toxicant Experiments: Second-order Model

As reviewed in Chapter 1, most of the current work on single-toxicant experiments focuses on the first-order model as defined in (1.9). However, as indicated in Chapter 1, the effect of some toxicants are so strong that model (1.9) is not adequate to describe the dose-response relationship. Thus, the second-order model as defined in (1.10) is necessary. In this section, we will discuss the D-optimal design for this model.

Recall that we made the assumption in Chapter 1 that larger doses generally cause smaller mean responses or, at least, increasing the dose does not increase the mean response. This means that the mean response is a decreasing function of the dose or amount of the toxicant under investigation. Equivalently, the MED level q is also a monotonically decreasing function of the dose or amount of the toxicant. For the first-order model in (1.9), this implies that $\beta_1 \leq 0$. Similarly, we shall assume $\beta_1 \leq 0$ for model (1.10). Also, β_{11} should not be positive so that model (1.10) is capable of describing a stronger effect than the first-order model as defined in (1.9). Mathematically, the signs of β_1 and β_{11} can be derived from the assumption that the MED level, or the proportion of survival, is a decreasing function of the dose or amount of the toxicant. Note that, for the second-order model, the MED level is $q = exp(\beta_1 x + \beta_{11} x^2)$, or equivalently, $\ln(q) = \beta_1 x + \beta_{11} x^2$. The assumption implies that the derivative of q or $\ln(q)$ with respect to x, the dose or amount of the toxicant, should be less than or equal to 0, i.e.:

$$\frac{\partial \ln(q)}{\partial x} = \beta_1 + 2\beta_{11} x \leq 0, \text{ where } x \geq 0.$$

Note if the above inequality holds for all nonnegative x, then we have $\beta_1 \leq 0$ and $\beta_{11} \leq 0$. Therefore, in what follows, we will focus on the second-order model with $\beta_1 \leq 0$ and $\beta_{11} \leq 0$.

2.2.1 D-optimal Designs

Clearly, the D-optimal design for the second-order model in (1.10) depends on the actual values of the parameters. The next lemma describes this dependence upon the parameters. The proof of this lemma, which involves many matrix manipulations, is given in Appendix A.1.

Lemma 2.1. *For model (1.10), the D-optimal design, in terms of ED levels, depends on the parameters only through* $r = \beta_1^2/\beta_{11}$.

Since D-optimal designs can be catalogued by r, for a given value of r, we can use some numerical method such as Nelder-Mead simplex method (Nelder and Mead, 1965) to

maximize the determinant of the Fisher information matrix and find the optimal design. We shall first discuss the design space before we shift to the optimal design.

As pointed out earlier, we will work with design space defined in terms of ED levels, i.e., q_i's. Theoretically, q_i, the proportion of survival at the ith design point, can be as small as 0. However, this rarely occurs in real experiments. In practice, a design point with very small survival proportion may not be feasible for biological, economical or physical reasons. For example, an anticancer drug may not be capable of eradicating 90% of the cancer colonies; or we have to use an excessively large dose to achieve such a goal, which might causes undesirable or even deadly side effects. Thus, it is more practical to consider designs on a restricted region than the unrestricted region. We will assume that the design space is restricted by $0 \leq c \leq q_i \leq 1$, i.e., the maximal possible impairment is $100(1-c)\%$ at any point. The lower boundary is some constant chosen by practitioners and it would typically change from case to case.

From the discussion at the beginning of this section, we know that $\beta_1 \leq 0$ and $\beta_{11} \leq 0$, which implies that $r = \beta_1^2/\beta_{11} \leq 0$. For a given value of r, we use the Nelder-Mead simplex method to numerically maximize the determinant of the information matrix and find the D-optimal design. For some representative values of r, Table 2.1 gives the D-optimal designs on different design regions. For the first design region, we set c, the lower boundary of the survival proportion, to be 0.01, which gives a practically unrestricted design space. The second and third design regions are restricted by setting c to 0.2 and 0.4, respectively. As one might expect, all the D-optimal designs are saturated three-point designs. The D-optimal design is an equal allocation design if it is saturated (Silvey, 1980). Specifically, if a k-point D-optimal design is saturated, the proportion of the sample size at each of the k points is $1/k$. For our situation, this implies p_1, p_2 and p_3 are all 1/3 and they are not listed in Table 2.1. Also, the D-optimal design always includes the control point. If we denote the control point by point 1, then $q_1 = 1$ for all the designs.

Table 2.1 shows the dependence of the D-optimal design on r. Note that the control point ($q_1 = 1$) is always in the D-optimal design. Point 3 is practically fixed at the lower boundary. Point 2, which is the middle point, depends on r and the design space. Generally, q_2 decreases as r decreases. However, the range of q_2 is not very large especially when the design region is restricted. Therefore, it seems that optimal designs for different values of r are quite close to each other on the restricted design space. This implies the possibility of finding an efficient design robust to r. We shall discuss this issue next. The optimality of the designs listed in Table 2.1 will be further verified in Chapter 3 via equivalence theory.

2.2.2 Equally Spaced Designs

In general, it is hard for practitioners to guess r and then choose the corresponding optimal design from Table 2.1. Therefore, it would be of great help if we can find some efficient designs that are robust to r. Practitioners would prefer efficient designs which are robust to

Table 2.1: D-optimal Designs for Model (1.10), $q_1 = 1$, $p_1 = p_2 = p_3 = 1/3$

r	Region 1 ($c = 0.01$)		Region 2 ($c = 0.2$)		Region 3 ($c = 0.4$)	
(β_1^2/β_{11})	q_2	q_3	q_2	q_3	q_2	q_3
0^a	0.6825	0.0729	0.7540	0.2000	0.8323	0.4000
-1	0.4959	0.0401	0.6185	0.2000	0.7223	0.4000
-2	0.4527	0.0331	0.5923	0.2000	0.7046	0.4000
-5	0.3964	0.0244	0.5626	0.2000	0.6864	0.4000
-10	0.3594	0.0190	0.5264	0.2000	0.6776	0.4000
-20	0.3304	0.0150	0.5359	0.2000	0.6737	0.4000
-50	0.3052	0.0117	0.5275	0.2000	0.6693	0.4000
$-\infty^b$	0.2856	0.0100	0.5223	0.2000	0.6668	0.4000

[a] $r = 0$ means $\beta_1 = 0$.
[b] $r = -\infty$ means $\beta_{11} = 0$.

r and easy to implement. In practice, "standard" designs with equally spaced points, in the space of ED levels, are often used due to the simplicity. We shall inspect two designs of this type: one is three-point and the other four-point.

Again, suppose the design space is defined by $0 \leq c \leq q_i \leq 1$. The 3-point design which is intuitive to practitioners consists of the two endpoints plus the middle point, namely, $q_1 = 1$, $q_2 = (1 + c)/2$ and $q_3 = c$. Furthermore, we assume the design is an equal allocation design, i.e., $p_1 = p_2 = p_3 = 1/3$. Table 2.2 gives the D-efficiency of three-point designs on different regions.

Table 2.2 indicates that three-point designs are quite robust to the values of r. For the unrestricted design space (Region 1), the lowest efficiency is about 82%. When the design space is restricted (Regions 2 and 3), the three-point design is very efficient regardless of the value of r. This should not be surprising if one looks carefully at the D-optimal design in Table 2.1. For Regions 2 and 3, we can see that the D-optimal design also consists of the two end points ($q_1 = 1$ and $q_3 = c$). Though the location of the other point changes according to the value of r, it is not far away from the middle point of the three-point design ($q_2 = (1 + c)/2$) for all possible values of r. This geometric closeness of the 3-point design to the D-optimal design explains the high D-efficiency.

We shall further study the performance of four-point equally spaced designs as the extra point allows for the lack of fit test. As before, we suppose that the design space is defined by $0 \leq c \leq q_i \leq 1$. The equally spaced four points, in the space of ED levels, are $q_1 = 1$, $q_2 = 1 - (1 - c)/3$, $q_3 = 1 - 2(1 - c)/3$ and $q_4 = c$. The allocation strategy used here is $p_1 = p_4 = 1/3$ and $p_2 = p_3 = 1/6$ as it gives larger efficiency on average than other allocations. Table 2.2 also lists D-efficiency of the four-point design, which suggests that the

Table 2.2: D-efficiency of Equally Spaced Designs for Model (1.10)

	3-point Design[a]			4-point Design [b]		
r	Region 1	Region 2	Region 3	Region 1	Region 2	Region 3
(β_1^2/β_{11})	(c=0.01)	(c=0.20)	(c=0.40)	(c=0.01)	(c=0.20)	(c=0.40)
0^c	0.8209	0.9332	0.9116	0.8334	0.9269	0.9110
-1	0.9072	0.9987	0.9965	0.8839	0.9481	0.9438
-2	0.9177	0.9998	0.9998	0.8930	0.9462	0.9433
-5	0.9239	0.9944	0.9987	0.9041	0.9430	0.9422
-10	0.9227	0.9902	0.9963	0.9095	0.9424	0.9414
-20	0.9181	0.9836	0.9944	0.9110	0.9392	0.9408
-50	0.9106	0.9796	0.9929	0.9088	0.9379	0.9404
$-\infty^d$	0.8998	0.9763	0.9916	0.9025	0.9368	0.9400

[a] $(q_1, q_2, q_3) = (1, (1+c)/2, c); p_1 = p_2 = p_3 = 1/3$.
[b] $(q_1, q_2, q_3, q_4) = (1, c + 2(1-c)/3, c + (1-c)/3, c); p_1 = 2p_2 = 2p_3 = p_4 = 1/3$.
[c] $r = 0$ means $\beta_1 = 0$.
[d] $r = -\infty$ means $\beta_{11} = 0$.

four-point design performs very well in terms of efficiency and robustness to r. In addition, testing the lack of fit is possible with a four-point design.

From the above discussion, we can see that equally spaced designs are quite efficient and very robust to the value of r. Therefore, they are highly recommended in practice since they do not require knowledge of r. Particularly, if one is concerned with model checking, the four-point equally spaced design is a good choice.

2.2.3 Parameter Misspecifications

Note that designs in previous sections are found in terms of ED levels, i.e., q_i's. To actually implement these designs, we must have some knowledge about ED's of the toxicant as we have assumed at the beginning of this chapter. Of course, instances occur when only limited knowledge of ED's is available. Without enough information about ED's, the implemented experiment might be different from the desired one. For instance, suppose that our target design $(q_1, q_2, q_3) = (1, 0.6, 0.2)$ and the corresponding ED's are $(x_1, x_2, x_3) = (0, 0.5, 1.6)$. However, due to the lack of knowledge about the toxicant, we might use $(x_1, x_2, x_3) = (0, 0.35, 1.2)$, which gives $(q_1, q_2, q_3) = (1, 0.7, 0.3)$. Therefore, we missed the target design by a little bit. This problem originates from the fact that we do not know the parameters perfectly so it is termed *parameter misspecification* in the literature.

Parameter misspecification almost always occurs in real experiments. Thus, a design with theoretically high efficiency is not necessarily a good design in practice. The question

is whether it can maintain its high efficiency under parameter misspecifications. A design that can maintain a reasonably high efficiency under parameter misspecifications is termed a *robust design*. Obviously, efficiency and robustness to parameter misspecifications are two important properties of a design.

In Table 2.2, we have demonstrated that equally spaced design are quite efficient regardless of the value of r. We shall further investigate their robustness to parameter misspecifications. Of course, we cannot exhaust all possible misspecifications. Table 2.3 lists D-efficiency of the three-point and four-point equally spaced designs when the ED levels are misspecified upward by 10%. For example, ED_{20} and ED_{60} are misspecified as ED_{30} and ED_{70}, respectively. ED_{100} cannot be misspecified since it does not require any knowledge of parameters.

Table 2.3: D-efficiency of Equally Spaced Designs for Model (1.10) under Parameter Misspecification

	3-point Design			4-point Design		
r	Region 1	Region 2	Region 3	Region 1	Region 2	Region 3
(β_1^2/β_{11})	(c=0.01)	(c=0.20)	(c=0.40)	(c=0.01)	(c=0.20)	(c=0.40)
0^a	0.9616	0.8459	0.8003	0.9391	0.8221	0.7704
-1	0.9163	0.7902	0.7210	0.8605	0.7326	0.6613
-2	0.8759	0.7585	0.6888	0.8259	0.7067	0.6377
-5	0.8048	0.7141	0.6511	0.7684	0.6731	0.6107
-10	0.7478	0.6877	0.6311	0.7212	0.6534	0.5963
-20	0.6983	0.6673	0.6187	0.6785	0.6376	0.5870
-50	0.6526	0.6529	0.6101	0.6380	0.6264	0.5806
$-\infty^b$	0.6088	0.6418	0.6038	0.5980	0.6176	0.5759

[a] $r = 0$ means $\beta_1 = 0$.
[b] $r = -\infty$ means $\beta_{11} = 0$.

From Table 2.3, we can see that the equally spaced designs lose some efficiency when parameter misspecification occurs. The loss is great when r is far from 0. An interesting point is that, for Region 1 ($c = 0.01$), parameter misspecification increases D-efficiency when r is not far from 0. For example, the efficiency of the three-point design increases from 0.8209 to 0.9616 when $r = 0$. This increase is due to the fact that the implemented design under parameter misspecification is closer to the D-optimal design. Note that the D-optimal design is $(q_1, q_2, q_3) = (1, 0.6825, 0.0729)$ and the three-point design is $(q_1, q_2, q_3) = (1, 0.5005, 0.01)$. Due to misspecification, the implemented design is actually $(q_1, q_2, q_3) = (1, 0.6005, 0.101)$, which is closer to the D-optimal design. Sometimes parameter misspecification helps by fluke and of course we cannot count on it.

Recall that a locally optimal design is constructed to be optimal only for some initial

guess of the unknown parameter. The design might maintain the high efficiency when the guess is in a close neighborhood of the true parameter; but, in general, it loses the high efficiency when the initial guess is far from the true value. Basically, this problem makes it very hard to apply locally optimal designs in practice. However, in this chapter, we assume sound information about the parameter is available so we do not consider parameter misspecification a problem here. We postpone the discussion of designs robust to parameter misspecifications to Chapter 4.

2.3 Two-toxicant Experiments: Interaction Model

As we have indicated in Chapter 1, the focus of this research is experimental designs for studies that involve more than one toxicant. The reason is twofold: (1) many of the real experiments involve more than two or more toxicants; (2) though designs for single-toxicant experiments have been extensively studied in the literature, as reviewed in Chapter 1, little has been done for designs involving two or more toxicants.

The simplest case of multi-toxicant study, of course, is the two-toxicant experiment. For two-toxicant experiments, Van Mullekom and Myers (2001) have studied the D-optimal design for the additive model as defined in (1.11), as reviewed in Chapter 1. However, model (1.12) is of more interest to researchers as interaction between two different toxicants is usually expected. Therefore, we will focus our study on D-optimal designs for this interaction model. In this section, we shall study locally D-optimal designs and some practical designs for model (1.12). In Chapter 4, we will further investigate sequential designs for this model. As we have discussed in Chapter 1, model (1.12) implies that each of the two toxicants, when applied alone, works through a first-order model as described in (1.9); however, when working together, their effects are not additive as they may interact with each other and the possible interaction is accounted for by the term β_{12}.

In this section, we will define the design space in terms of IED levels. For this model, at point (x_1, x_2), the IED level is $q_1 = exp(\beta_1 x_1)$ for the first toxicant and $q_2 = exp(\beta_2 x_2)$ for the other toxicant. The MED level, or proportion of survival at this point, is $q = exp(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2)$. As before, only continuous designs are considered.

2.3.1 D-optimal Designs

Optimal Designs

As in the last section, we start with the D-optimal design as it permits calculation of D-efficiency for any other design. The D-optimal design for model (1.12) depends on the parameters. Lemma 2.2 reveals the nature of this dependence. The proof of this lemma is outlined in Appendix A.2.

Lemma 2.2. *For model (1.12), the D-optimal design, in terms of IED levels, depends on the parameters only through* $r = \beta_{12}/\beta_1\beta_2$.

Lemma 2.2 implies that D-optimal designs for model (1.12) can be catalogued by r. For a given r, we can find the optimal design via numerical methods. The property that the optimal design depends upon a function of the unknown parameters does not make it any easier for practitioners to implement the optimal design. However, just as we have seen in the last section, this property enables us to study performance of any particular design over the whole parameter space by studying all possible values of the function.

In biomedical research, an interaction between different drugs is often classified as synergism or antagonism. Though this classification depends on the goal of the experiment, in this paper, synergism will be considered as an interaction of negative sign (i.e., negative β_{12} and thus negative r), which implies the two toxicants under investigation have a combined effect larger than their additive effect, whereas antagonism will be considered as an interaction with a positive sign. Though large synergism is possible, there is some bound for antagonism as we assume that, at any point, increasing the amounts of either or both toxicants at least does not increase the mean response. In general, larger doses should cause more impairment, which implies that the MED level, or the survival proportion, is a decreasing function of the doses of both toxicants. Based on this assumption, we can derive an upper bound for r as follows.

For model (1.12), the MED level q is

$$q = exp(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2),$$

or equivalently

$$\ln(q) = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2.$$

The assumption that the MED level is a decreasing function of the doses implies that the partial derivatives of $\ln(q)$ with respect to x_1 and x_2 should not be positive, i.e.,

$$\begin{aligned} \frac{\partial \ln(q)}{\partial x_1} &= \beta_1 + r\beta_1 \ln(q_2) \leq 0, \\ \frac{\partial \ln(q)}{\partial x_2} &= \beta_2 + r\beta_2 \ln(q_1) \leq 0. \end{aligned}$$

The above inequalities give the upper bound for r as follows

$$r \leq Ubound = \min\{-\frac{1}{\ln(q_1)}, \ -\frac{1}{\ln(q_2)}\}. \tag{2.1}$$

Though, theoretically, r can take any value that satisfies (2.1), this may not be the case in practice as in general the main effects β_1 and β_2 dominate the interaction term β_{12}.

Therefore, it is very likely that r lies between -1 and 1 as well as satisfying (2.1) in real toxicity experiments.

We now turn to the D-optimal design for model (1.12). We will consider the design space defined in terms of q_1 and q_2. As in the last section, we restrict the design region by $0 \leq c_1 \leq q_1 \leq 1$ and $0 \leq c_2 \leq q_2 \leq 1$, where the lower boundaries, c_1 and c_2, are specified by practitioners. We know that the unrestricted design space is defined by $0 \leq q_1 \leq 1$ and $0 \leq q_2 \leq 1$. When $q_1 = q_2 = 0$, the computation is generally intractable. To make computation easier, we will try to avoid literally unrestricted region of this type. Instead, throughout this dissertation, the design space restricted by $0.01 \leq q_1 \leq 1$ and $0.01 \leq q_2 \leq 1$ will be referred to as the unrestricted region or the unrestricted space since it gives a close approximation to the literally unrestricted region in practice and it is more mathematically tractable. For some typical values of r, we find the D-optimal designs on different design regions via Nelder-Mead simplex method. Table 2.4 lists these designs and some comments are in order:

1. Like most D-optimal designs, all the designs in Table 2.4 are saturated designs, which are four-point designs for model (1.12). The four points are: the control point (at which no toxicant is applied), two pure component points (at which only one toxicant is applied) and an interaction point (at which both toxicants are applied). Further, each of the four points has the same proportion of the total sample, namely, $p_1 = p_2 = p_3 = p_4 = 1/4$.

2. The control point, denoted by point 1 in Table 2.4, is located at $(q_{11}, q_{21}) = (1, 1)$. The two pure component points, denoted by points 2 and 3, are $(q_{12}, q_{22}) = (\max\{0.1353, c_1\}, 1)$ and $(q_{13}, q_{23}) = (1, \max\{0.1353, c_2\})$.

3. The location of the interaction point (point 4 in Table 2.4) depends on r. When $r = Ubound$, $(q_{14}, q_{24}) = (c_1, c_2)$. When r decreases from the upper bound, the interaction point has a tendency to move toward the control point, as illustrated in Figure 2.1.

4. The optimality of these designs will be verified via equivalence theory in Chapter 3.

Estimation

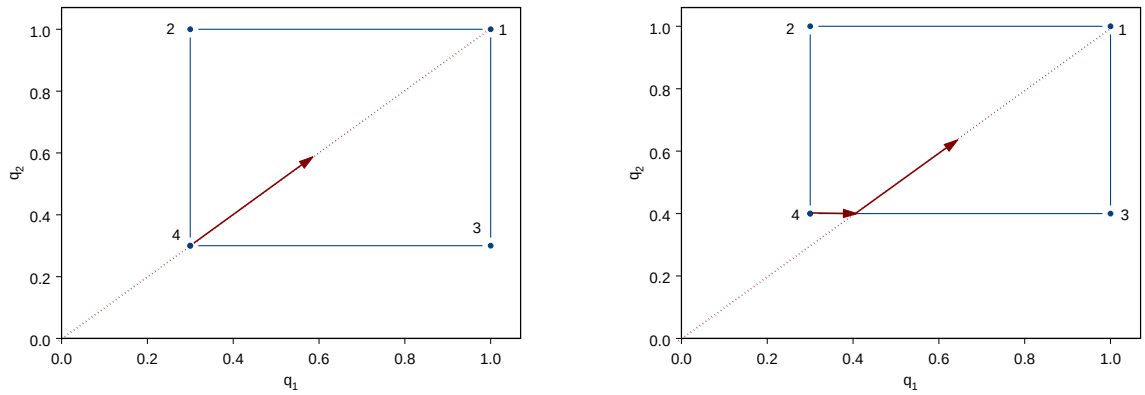
We now consider how each of the four points in the D-optimal design contributes to estimation of the parameters $\boldsymbol{\beta}$. For generalized linear models, the MLE is generally used to estimate the parameters. Therefore, we shall start with the likelihood function based on the D-optimal design. As in Table 2.4, we denote the control point, two pure component points and the interaction point by points 1, 2, 3 and 4, respectively. At the ith point, the regressors are (x_{1i}, x_{2i}) and the mean response is $\lambda_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_{12} x_{1i} x_{2i})$,

Table 2.4: D-optimal Designs for Model (1.12), $q_{11} = q_{21} = 1$, $p_1 = p_2 = p_3 = p_4 = 1/4$

r	Region 1 ($c_1 = c_2 = 0.01$)			Region 2 ($c_1 = c_2 = 0.3$)		
$(\beta_{12}/\beta_1\beta_2)$	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$
-10	0.1353	1	0.6704	0.3	1	0.6704
-5	0.1353	1	0.5823	0.3	1	0.5823
-1	0.1353	1	0.3679	0.3	1	0.3679
-0.75	0.1353	1	0.3338	0.3	1	0.3338
-0.5	0.1353	1	0.2905	0.3	1	0.3
-0.2	0.1353	1	0.2163	0.3	1	0.3
-0.1	0.1353	1	0.1812	0.3	1	0.3
0	0.1353	1	0.1353	0.3	1	0.3
0.1	0.1353	1	0.0630	0.3	1	0.3
0.2	0.1353	1	0.01	0.3	1	0.3
0.5	— [a]	—	—	0.3	1	0.3
0.75	—	—	—	0.3	1	0.3
1	—	—	—	—	—	—
Ubound[b]	0.1353	1	0.01	0.3	1	0.3

r	Region 3 ($c_1 = 0.3$, $c_2 = 0.4$)			Region 4 ($c_1 = c_2 = 0.4$)		
$(\beta_{12}/\beta_1\beta_2)$	(q_{12}, q_{22})	(q_{13}, q_{23})	(q_{14}, q_{24})	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$
-10	(0.3, 1)	(1, 0.4)	(0.6704, 0.6704)	0.4	1	0.6704
-5	(0.3, 1)	(1, 0.4)	(0.5823, 0.5823)	0.4	1	0.5823
-1	(0.3, 1)	(1, 0.4)	(0.3522, 0.4)	0.4	1	0.4
-0.75	(0.3, 1)	(1, 0.4)	(0.3056, 0.4)	0.4	1	0.4
-0.5	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
-0.2	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
-0.1	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
0	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
0.1	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
0.2	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
0.5	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
0.75	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4
1	—	—	—	0.4	1	0.4
Ubound	(0.3, 1)	(1, 0.4)	(0.3, 0.4)	0.4	1	0.4

[a] This value of r does not apply here.
[b] Ubound equals to the upper bound defined in (2.1).



(a) Symmetric Region, $c_1 = c_2 = 0.3$

(b) Asymmetric Region, $c_1 = 0.3$, $c_2 = 0.4$

Figure 2.1: Structure of D-optimal Design for Model (1.12). When $r = Ubound$, the interaction point is located at (c_1, c_2). As r decreases from the upper bound, (a) the interaction point moves toward the control point along the line $q_1 = q_2$ for symmetrical design space; (b)for asymmetrical space, the interaction point first moves on the edge of the design space and then moves along the line $q_1 = q_2$.

where $i = 1, 2, 3, 4$. Suppose that there are n_i replicates at the ith point and for the jth replicate, the response is y_{ij}; then the likelihood function of the data is given as follows:

$$L(\mathbf{y}, \boldsymbol{\beta}) = \prod_{i=1}^{4} \prod_{j}^{n_i} \frac{e^{-\lambda_i} \lambda_i^{y_{ij}}}{y_{ij}!}.$$

To facilitate the derivation, we take log of the above likelihood function, which gives the following log likelihood function:

$$l(\mathbf{y}, \boldsymbol{\beta}) = \ln(L(\mathbf{y}, \boldsymbol{\beta})) = \sum_{i=1}^{4} \sum_{j=1}^{n_i} (-\lambda_i + y_{ij} \ln(\lambda_i) - \ln(y_{ij}!)).$$

To find the MLE, we need to take derivatives of the above log likelihood function with respect to all the parameters. Setting these derivatives to zero gives the score equations. Note that, for the D-optimal design, we have $x_{11} = x_{21} = x_{22} = x_{13} = 0$. Taking advantage of this, we can simplify the score equations. The simplified score equations are as follows:

$$
\begin{aligned}
\frac{\partial l}{\partial \beta_0} &= -\sum_{i=1}^{4}\sum_{j=1}^{n_i}(\lambda_i - y_{ij}) = 0 \\
\frac{\partial l}{\partial \beta_1} &= -\sum_{j=1}^{n_2} x_{12}(\lambda_2 - y_{2j}) - \sum_{j=1}^{n_4} x_{14}(\lambda_4 - y_{4j}) = 0 \\
\frac{\partial l}{\partial \beta_2} &= -\sum_{j=1}^{n_3} x_{23}(\lambda_3 - y_{3j}) - \sum_{j=1}^{n_4} x_{24}(\lambda_4 - y_{4j}) = 0 \\
\frac{\partial l}{\partial \beta_{12}} &= -\sum_{j=1}^{n_4} x_{14}x_{24}(\lambda_4 - y_{4j}) = 0
\end{aligned}
$$

Solving the above equations gives estimates of the parameters as follows:

$$
\begin{aligned}
\hat{\beta}_0 &= \ln \bar{y}_{1.} \\
\hat{\beta}_1 &= \frac{\ln(\bar{y}_{2.}) - \ln(\bar{y}_{1.})}{x_{12}} \\
\hat{\beta}_2 &= \frac{\ln(\bar{y}_{3.}) - \ln(\bar{y}_{1.})}{x_{23}} \\
\hat{\beta}_{12} &= \frac{\ln(\bar{y}_{4.}/\bar{y}_{1.}) - (x_{14}/x_{12})\ln(\bar{y}_{2.}/\bar{y}_{1.}) - (x_{24}/x_{23})\ln(\bar{y}_{3.}/\bar{y}_{1.})}{x_{14}x_{24}}
\end{aligned}
$$

From the above equations we can see that (1) only the control point contributes to the estimation of β_0; (2) the main effect is estimated from the difference between the control point and the corresponding pure component point; (3) all four points contribute to estimation the interaction term. The same result actually applies to any four-point design that consists of the control point, two pure component points and one interaction point.

Restrictions on Design Space

In practice, certain restrictions need to be imposed on the design space for various reasons. As mentioned earlier, a design point with very small survival proportion may be unfeasible for economical, physical or biological reasons. In one-toxicant experiments, we restrict the design space by setting a lower boundary for the proportion of survival. For the one-toxicant case, this method can effectively prevent undesirably large impairment. For two-toxicant experiments, the restrictions we used are $c_1 \leq q_1$ and $c_2 \leq q_2$. This type of restriction gives a rectangular design space in terms of IED levels. It is very simple and easy to deal with. However, it cannot bound the proportion of survival as we wish to. This is so because q_1 and

q_2 are IED levels, not MED levels. To illustrate this, suppose that the two toxicants under investigation do not react with each other and we want the proportion of survival is at least 30%. By setting $0.3 \leq q_1$ and $0.3 \leq q_2$, the survival proportion at the two pure component points is 30%. However, at the interaction point, the survival proportion is only 9%, which, of course, should be avoided. If the two-toxicants have synergism interaction, the survival proportion at the interaction point is even lower.

Clearly, restrictions should be imposed on the MED level, or the survival proportion, instead of the IED levels. However, any restriction imposed on the MED level will lead to the design space dependent on the unknown parameters, which complicates the situation very much because we have to specify the design space before designing an experiment. We will discuss restrictions of this type in Chapter 4. In this chapter, we will stick to restrictions imposed on the IED levels, which gives us a design region independent of the unknown parameters.

2.3.2 Conditional D-optimal Designs

Though designs in Table 2.4 are optimal, they are not practical because of the dependence on r, which is generally unknown. Van Mullekom and Myers (2001) showed that the D-optimal design for model (1.11), in terms of IED levels, is independent of the parameters. Therefore, it is the interaction term in model (1.12) that causes the dependence and complicates the situation. To deal with the interaction in the logistic regression model, Brunden et al. (1988) proposed a so-called conditional D-optimal design. Following their idea, Van Mullekom and Myers (2001) developed the conditional D-optimal design for model (1.12). To find the conditional D-optimal design, the Fisher information matrix for model (1.12) is first derived. When the determinant of the information matrix is maximized to find the optimal design the interaction term β_{12} is assumed to be zero for convenience. Clearly, the design is optimal only on the condition that there is no interaction and that's why it is termed *conditional D-optimal design*. Obviously, the conditional D-optimal design is the D-optimal design corresponding to $r = 0$ in Table 2.4.

In developing the conditional D-optimal design, the interaction is assumed to be zero for convenience and this may not be true. One natural concern about the conditional D-optimal design is its robustness to departure from the condition that r is zero. To understand this issue, we compute the D-efficiency of the conditional D-optimal design for some representative values of r, as listed in Table 2.5. One can see that the conditional D-optimal design is very inefficient for r far away from zero; however, it performs very well when r is between -1 and 1, especially on restricted regions. We would therefore expect that the conditional optimal design to be useful in a practical sense as it is quite common that the main effects dominate the interaction term and the design region is usually restricted. Other justifications for recommending the conditional D-optimal designs include:

Table 2.5: D-efficiency of Conditional D-optimal and 5-point Designs for Model (1.12)

	Conditional D-optimal Design			5-point Design		
r ($\beta_{12}/\beta_1\beta_2$)	Region 1 (c=0.01)	Region 2 (c=0.3)	Region3 (c=0.4)	Region 1 (c=0.01)	Region 2 (c=0.3)	Region3 (c=0.4)
-10	0.0002	0.0801	0.3236	0.7331	0.8366	0.8341
-5	0.0173	0.3762	0.7087	0.8393	0.8138	0.8169
-1	0.5729	0.9717	1.0000	0.7406	0.8830	0.8863
-0.75	0.6870	0.9934	1.0000	0.7536	0.8911	0.8826
-0.5	0.8107	1.0000	1.0000	0.7930	0.8883	0.8796
-0.2	0.9511	1.0000	1.0000	0.8649	0.8811	0.8768
-0.1	0.9848	1.0000	1.0000	0.8842	0.8793	0.8760
0	1.0000	1.0000	1.0000	0.8898	0.8778	0.8752
0.1	0.9680	1.0000	1.0000	0.8561	0.8765	0.8746
0.2	0.6759	1.0000	1.0000	0.5953	0.8753	0.8740
0.5	—[a]	1.0000	1.0000	—	0.8730	0.8726
0.75	—	1.0000	1.0000	—	0.8720	0.8717
1	—	—	1.0000	—	—	0.8711
Ubound[b]	0.6278	1.0000	1.0000	0.5527	0.8719	0.8710

[a]This value of r does not apply here.
[b]Ubound equals to the upper bound defined in (2.1).

1. The conditional D-optimal design, in terms of IED levels, is independent of the parameters. Thus, it does not require any guess of the interaction term. Furthermore, it is a four-point factorial, which might appeal to practitioners. Actually, for design regions defined by $0.1353 \leq c_1 \leq q_{1i} \leq 1$ and $0.1353 \leq c_2 \leq q_{2i} \leq 1$, the conditional D-optimal design is the standard 2^2 factorial design in the space of IED levels.

2. In many situations, we know neither the magnitude nor the sign of the interaction between the two toxicants so zero interaction represents a practical solution.

For r less than -1, the efficiency of the conditional D-optimal design drops very quickly. If large synergism is possible, we can adapt the conditional D-optimal design a little bit and make it more robust. Suppose the design regions are restricted by $0 \leq c \leq q_{1i}, q_{2i} \leq 1$ and let $b = \max\{c, 0.135\}$. Table 2.4 suggests that the D-optimal design consists of the control point and two pure component point located at IED_{100b}, regardless of the value of r. As illustrated in Figure 2.1, the value of r has impact only on the location of the interaction point: the interaction point moves from (b, b) to the control point as r moves away from the upper bound defined in (2.1). The proportion of sample size at the interaction point is always 1/4. Intuitively, we can make the design more robust to r if we can spread this 1/4 of the total runs over the line $q_1 = q_2$ instead of using them up at one single point. We

investigate a five-point design of this type. The five-point design also consists of the control point and two pure component points located at IED_{100b} and each of these three points has 1/4 of the total runs. The other two points are on the line $q_1 = q_2$ with one at (b, b) and the other at $((1+b)/2, (1+b)/2)$ and each of these two points has 1/8 of the total runs. Table 2.5 also gives the D-efficiency of the five-point design. As expected, the five-point design is more robust to r than the conditional D-optimal design, especially for large synergism. Yet, the five-point design is not as efficient as the conditional optimal design when r is greater than -1. This is the price we have to pay for the robustness. Another inherent strength of the five-point design is that it allows for the lack of fit test, which we cannot do with a saturated design such as the conditional D-optimal design. Clearly, if we use more points on the line $q_1 = q_2$, the design will be more robust to negative r far from zero and less efficient for r greater than -1.

The reason that the conditional D-optimal design is inefficient for r less than -1 is that it is constructed to be optimal only for $r = 0$. Thus, it can maintain quite high efficiency only if r is close to 0. Another way to improve its robustness to r is to take the Bayesian way: instead of finding a design that maximizes the determinant of information matrix for a certain r, we could find a design that maximizes the average of the determinant over a reasonable range of r. In the Bayesian language, this actually assumes that r follows a uniform prior distribution over a certain range. Of course, generally, the prior distribution is not necessarily a uniform distribution. If we denote the prior distribution by $\pi(r)$, then the Bayesian D-optimal design is the one that maximizes the weighted average of the determinant over $\pi(r)$. Mathematically, the Bayesian criterion is:

$$\max_{X \in \mathcal{D}} \int |I(X, \boldsymbol{\beta})| \pi(r) dr, \tag{2.2}$$

where $\mathcal{D}$ is the set of all possible designs. We need some knowledge of the parameters to formulate the prior distribution $\pi(r)$. Suppose that we know the interaction is synergism but we do not much about the magnitude of the interaction; we may choose the "negative exponential distribution" as the prior distribution for r, i.e.:

$$\pi(r) = exp(r), \text{ where } r \leq 0.$$

On the unrestricted design space ($c_1 = c_2 = 0.01$), the Bayesian design using the above prior distribution turns out to be: $(q_{11}, q_{21}) = (1, 1)$, $(q_{12}, q_{22}) = (0.1353, 1)$, $(q_{13}, q_{23}) = (1, 0.1353)$, and $(q_{14}, q_{24}) = (0.2584, 0.2584)$. The only difference between the conditional D-optimal design and the Bayesian design is the location of the interaction point. Compared to the conditional D-optimal design, the Bayesian design has its interaction point closer to the control point. Since the optimal interaction point moves toward the control point as r decreases from 0, we would expect that the Bayesian design to be more efficient than the conditional D-optimal design for negative r far away from 0. Table 2.6 confirms this intuition by comparing D-efficiency of these two designs on the unrestricted space.

Table 2.6: D-efficiency of Conditional and Bayesian D-optimal Designs for Synergism on the Unrestricted Region ($c_1 = c_2 = 0.01$)

r	Designs	
$(\beta_{12}/\beta_1\beta_2)$	Conditional Design	Bayesian Design
-10	0.0002	0.0322
-5	0.0173	0.2435
-1	0.5729	0.9213
-0.75	0.6870	0.9647
-0.5	0.8107	0.9941
-0.2	0.9511	0.9911
-0.1	0.9848	0.9721
0	1.0000	0.9349

2.3.3 Parameter Misspecifications

As mentioned in the last section, robustness to parameter misspecification is an important property of a design. In this section, we will look at the robustness of the conditional D-optimal design to parameter misspecification. Remember that when constructing the conditional D-optimal design we assume that there is no interaction, so the interaction term is misspecified unless the two toxicants do not react with each other. Therefore, misspecification of the interaction term originates from the way the conditional D-optimal design is constructed.

When actually implementing the conditional D-optimal design, we need to translate the IED levels into the actual amounts or doses of the toxicants. This translation could incur misspecification of main effects since it requires the knowledge of main effects β_1 and β_2. Therefore, misspecification of main effects occurs during the course of implementing the design.

When implementing the design, we need to provide some initial guesses of the main effects. Suppose that the initial guesses are b_1 and b_2 for β_1 and β_2, respectively. We use the ratios of true parameters to their guesses to measure the extent of misspecifications. These two ratios will be denoted by m_1 and m_2, namely,

$$m_1 = \frac{\beta_1}{b_1},\ m_2 = \frac{\beta_2}{b_2}.$$

Clearly, $m_1 = m_2 = 1$ stands for the perfect guess, i.e., no misspecification. Any value other than 1 implies misspecification. Suppose that the design point is (q_1, q_2) and the implemented point, under the misspecification (m_1, m_2), is (q_1', q_2'). It can be shown that

$$q_1' = q_1^{m_1},\ q_2' = q_2^{m_2}.$$

In what follows, we will investigate D-efficiency of the conditional D-optimal design under various values of r and misspecifications of main effects. The effects of misspecification are different for unrestricted and restricted regions so we will discuss them separately.

Unrestricted Regions

As mentioned before, by unrestricted design space, we mean that the lower boundaries of both IED levels are 0.01, or Region 1 in Table 2.4. We will investigate some representative misspecifications of main effects rather than all possible misspecifications, which is impossible. Specifically, these representative misspecifications include: no misspecification ($m_1 = m_2 = 1$), mild misspecification ($m_2 = m_2 = 2$ and $m_1 = m_2 = 0.5$) and severe misspecification ($m_1 = 0.3, m_2 = 0.4$ and $m_1 = m_2 = 0.2$). For these misspecifications, Table 2.7 gives the D-efficiency of the conditional D-optimal design for different values of r:

Table 2.7: D-efficiency of Conditional D-optimal Design under Misspecification

r	Misspecification: (m_1, m_2)				
$(\beta_{12}/\beta_1\beta_2)$	(2, 2)	(1, 1)	(0.5, 0.5)	(0.3, 0.4)	(0.2, 0.2)
-1	0.0154	0.5729	0.8244	0.6084	0.2966
-0.75	0.0392	0.6870	0.8195	0.5854	0.2797
-0.5	0.0979	0.8107	0.8017	0.5544	0.2596
-0.2	0.2825	0.9511	0.7511	0.4995	0.2284
-0.1	0.3948	0.9848	0.7215	0.4736	0.2148
0	0.5412	1.0000	0.6797	0.4404	0.1982
0.1	0.7072	0.9680	0.6104	0.3904	0.1743
0.2	0.6666	0.6759	0.3954	0.2496	0.1105
Ubound[a]	0.6519	0.6278	0.3625	0.2284	0.1010

[a]Ubound equals to the upper bound defined in (2.1).

We now make two comments about the results in Table 2.7:

1. In general, misspecification of main effects causes loss of design efficiency. This can be seen by comparing the column ($m_1 = m_2 = 1$) to other columns. On average, more severe misspecification causes more loss of efficiency. For some combinations of r and misspecification of main effects such as ($r = -1, m_1 = m_2 = 2$) and ($r = 0.2, m_1 = m_2 = 0.2$), the design efficiency is extremely small, making the conditional D-optimal design practically useless.

2. In some situations, misspecification of main effects somehow counteracts misspecification of the interaction, giving a higher efficiency than under no misspecification. For instance, when $r = -1$ and $m_1 = m_2 = 0.5$, the efficiency is 0.8244, higher than the efficiency under no misspecification, which is 0.5729. For $r = -1$, upward misspecification of main effects pushes the interaction point from $(0.1353, 0.1353)$ to $(0.3678, 0.3678)$, which is much closer to the optimal interaction point located at $(0.3679, 0.3679)$, thus improving the design efficiency, by chance. However, as pointed out in the last section, we cannot count on this in practice.

3. In Table 2.7, $(m_1 - 1)(m_2 - 1) \geq 0$, which implies that β_1 and β_2 are misspecified in the same direction. Misspecifications in different directions generally cause less severe results. For instance, the efficiency corresponding to $(m_1, m_2) = (2, 0.5)$ is somewhere between the efficiencies for $(m_1, m_2) = (2, 2)$ and $(m_1, m_2) = (0.5, 0.5)$. This is so because the upward misspecification of one effect somehow counteracts the downward misspecification of the other.

Restricted Regions

We now consider the impact of parameter misspecification on the conditional D-optimal design on restricted regions. We will consider two regions: one is restricted by $c_1 = c_2 = 0.3$ and the other is restricted by $c_1 = c_2 = 0.4$. Table 2.8 lists the efficiency of the conditional D-optimal design under various misspecifications on these two regions and it can give us some insights into the impact of misspecification on the conditional D-optimal design when the design space is restricted:

1. As already shown in Table 2.5, under no misspecification of main effects, the conditional D-optimal design becomes more efficient as a stricter restriction is imposed on the design space. This can also be seen from the column $(m_1 = m_2 = 1)$ in Table 2.8. The increase of D-efficiency is due to the fact the conditional D-optimal design gets closer to the D-optimal design under stricter restrictions (see Table 2.4).

2. When main effects are upwardly misspecified, which corresponds to $m_1 < 1$ and $m_2 < 1$ in Table 2.8, the conditional D-optimal design is not efficient on average. It is less efficient on restricted regions than unrestricted regions. Further, the efficiency gets worse for stronger restriction. This can be seen by comparing the column $(m_1 = m_2 = 1)$ to columns $(m_1 = m_2 = 0.5)$, $(m_1 = 0.3, m_2 = 0.4)$ and $(m_1 = m_2 = 0.2)$. This is so because, relatively, the implemented design under upward misspecification is closer to the optimal design on regions with less strict restriction. For example, when $r = -1$ and $m_1 = m_2 = 0.5$, the interaction point on Region 1 $(c_1 = c_2 = 0.3)$ is $(0.3679, 0.3679)$ and the implemented interaction point of the conditional D-optimal design, due to misspecification, is actually $(0.5477, 0.5477)$; for Region 2 $(c_1 = c_2 = 0.4)$, the optimal interaction point is $(0.4, 0.4)$ and the implemented point is $(0.6325, 0.6325)$. We can

Table 2.8: D-efficiency of Conditional D-optimal Design under Misspecification on Restricted Regions

Region 1: $c_1 = c_2 = 0.3$					
r	Misspecification: (m_1, m_2)				
$(\beta_{12}/\beta_1\beta_2)$	(2, 2)	(1, 1)	(0.5, 0.5)	(0.3, 0.4)	(0.2, 0.2)
-1	0.3931	0.9717	0.5820	0.3508	0.1442
-0.75	0.5275	0.9934	0.5560	0.3312	0.1352
-0.5	0.6968	1.0000	0.5229	0.3078	0.1247
-0.2	0.9655	1.0000	0.4819	0.2797	0.1124
-0.1	1.0764	1.0000	0.4690	0.2710	0.1085
0	1.2000	1.0000	0.4564	0.2625	0.1048
0.1	1.3378	1.0000	0.4442	0.2542	0.1012
0.2	1.4915	1.0000	0.4323	0.2462	0.0978
0.5	2.0666	1.0000	0.3984	0.2238	0.0881
0.75	2.7120	1.0000	0.3723	0.2066	0.0807
1	—[a]	—	—	—	—
Ubound[b]	2.9603	1.0000	0.3642	0.2014	0.0785
Region 2: $c_1 = c_2 = 0.4$					
r	Misspecification: (m_1, m_2)				
$(\beta_{12}/\beta_1\beta_2)$	(2, 2)	(1, 1)	(0.5, 0.5)	(0.3, 0.4)	(0.2, 0.2)
-1	0.8524	1.0000	0.4627	0.2619	0.1018
-0.75	0.9977	1.0000	0.4448	0.2500	0.0968
-0.5	1.1678	1.0000	0.4277	0.2388	0.0921
-0.2	1.4107	1.0000	0.4079	0.2259	0.0867
-0.1	1.5024	1.0000	0.4016	0.2218	0.0849
0	1.6000	1.0000	0.3953	0.2177	0.0833
0.1	1.7040	1.0000	0.3891	0.2137	0.0816
0.2	1.8147	1.0000	0.3830	0.2098	0.0800
0.5	2.1921	1.0000	0.3654	0.1985	0.0753
0.75	2.5658	1.0000	0.3513	0.1895	0.0716
1	3.0033	1.0000	0.3377	0.1810	0.0681
Ubound	3.1811	1.0000	0.3329	0.1779	0.0668

[a]This r value does not apply here.
[b]Ubound equals to the upper bound defined in (2.1).

see that the distance between the optimal point and the implemented point is larger for stronger restriction.

3. When main effects are downwardly misspecified, D-efficiency of the conditional D-optimal design is also listed in Table 2.8. However, D-efficiency is not an appropriate measure here as we can see that the efficiency is as high as 3.2 in some case, which is higher than 1, the maximum value of D-efficiency. We know that D-efficiency assumes that the design of interest and the D-optimal design have the same design space. However, due to downward misspecification, the implemented design is actually out of the restricted region. For instance, on the region ($c_1 = c_2 = 0.3$), the implemented conditional D-optimal design under ($m_1 = m_2 = 2$) consists of $(1, 1)$, $(0.09, 1)$, $(1, 0.09)$, and $(0.09, 0.09)$. Three out of the four points are out of the restricted region. Such a design should receive a low score in evaluation; so we cannot evaluate it properly with D-efficiency.

In summary, the conditional D-optimal design, like most locally optimal designs, is not robust to parameter misspecification. On the unrestricted region, misspecification simply causes loss of efficiency. On the restricted region, misspecification could lead to experimentation outside the restricted region, which is undesirable or even dangerous in real experiments. Recall that, in this chapter, we assume availability of good initial estimates of parameters so parameter misspecification is not a major concern here. In Chapter 4, we will discuss designs robust to parameter misspecification.

2.4 D_s-optimal Designs: Addressing ED Estimation

2.4.1 Introduction

We know that D-optimal designs address estimation of all the model parameters, which is an important goal in many situations. However, as mentioned in Chapter 1, estimating a particular ED like ED_{50} is of more interest in some applications. In Chapter 1, we reviewed three different procedures proposed in the literature to construct optimal designs for estimating a particular ED for the logistic regression model. Basically, all these three procedures minimize the length of the confidence interval for the ED of interest. The difference is that different methods are employed to construct the confidence interval: the first one is based on the asymptotic variance of the ED estimate; the second one uses Filler's interval and the third one adopts the likelihood-based approach. The ideas of these procedures are straightforward; however, all these three procedures involve very lengthy derivations of the confidence interval length. Therefore, most current discussions of these procedures have been restricted to the one-variable first-order model and it is very difficult to apply them to more complicated models. So, in this section, we will discuss an easier approach to designing experiments for ED estimation.

Suppose that we want to estimate ED_{100q} as accurate as possible, where $0 < q < 1$. As before, we denote the vector of model parameters by $\boldsymbol{\beta}$. In general, ED_{100q} is a function of a subset of $\boldsymbol{\beta}$ denoted by $\boldsymbol{\beta}_{\mathrm{s}}$. For instance, for the one-toxicant second-order model in (1.10), ED_{100q} is a solution to the following equation:

$$\beta_{11}x^2 + \beta_1 x - \ln(q) = 0.$$

Obviously, it is a function of $\boldsymbol{\beta}_{\mathrm{s}} = (\beta_1, \beta_{11})'$. For convenience, we shall simply denote the ED of interest by x. We know that $x = f(\boldsymbol{\beta}_{\mathrm{s}})$ and our goal is to design an experiment so that x can be estimated as well as possible. In general, x can be estimated by $\hat{x} = f(\hat{\boldsymbol{\beta}}_{\mathrm{s}})$, where $\hat{\boldsymbol{\beta}}_{\mathrm{s}}$ is the MLE of $\boldsymbol{\beta}_{\mathrm{s}}$. Asymptotically, we have

$$var(\hat{x}) = (\partial f/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}})' var(\hat{\boldsymbol{\beta}}_{\mathrm{s}})(\partial f/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}}), \tag{2.3}$$

where $\partial f/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}} = (\partial f/\partial\beta_1, \partial f/\partial\beta_{11})'$ evaluated at $\boldsymbol{\beta}_{\mathrm{s}} = \hat{\boldsymbol{\beta}}_{\mathrm{s}}$. Intuitively, if we want to minimize the variance of $\hat{x}$, we need to minimize the variance of $\hat{\boldsymbol{\beta}}_{\mathrm{s}}$. In general, $\hat{\boldsymbol{\beta}}_{\mathrm{s}}$ is a matrix so we could minimize the generalized variance of $\hat{\boldsymbol{\beta}}_{\mathrm{s}}$ in the same spirit of D-optimality. This task lends itself to a different type of design optimality criterion known as $\mathrm{D_s}$-optimality. Basically, in the $\mathrm{D_s}$-optimality criterion, the generalized variance of the estimates of a subset of model parameters is minimized, in contrast to the D-optimality, where the generalized variance of all the parameter estimates is minimized.

Similarly, for the two-toxicant interaction model as defined in (1.12), we can also use a $\mathrm{D_s}$-optimal design to address the estimation of a particular MED. For the multi-toxicant case, we know that a particular MED is all possible combinations of doses of all the toxicants that produce a specified response. Thus, it is not a single value as in the one-toxicant case; instead it is a set or a curve. For instance, for model (1.12), MED_{100q} is actually a curve defined by

$$\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 = \ln(q).$$

Thus, to obtain a good estimate of MED_{100q}, we need to estimate every point of the curve as well as possible. Equivalently, for a given x_1, we should estimate the corresponding x_2 with the smallest variance. Obviously, for a give x_1, x_2 is a function of $\boldsymbol{\beta}_{\mathrm{s}} = (\beta_1, \beta_2, \beta_{12})'$, a subset of the model parameters. For brevity, we shall denote the function by g and generally x_2 can be estimated by $\hat{x}_2 = g(\hat{\boldsymbol{\beta}}_{\mathrm{s}})$, where $\hat{\boldsymbol{\beta}}_{\mathrm{s}}$ is the MLE of $\boldsymbol{\beta}_{\mathrm{s}}$. Similar to (2.3), we have the following asymptotic result:

$$var(\hat{x}_2) = (\partial g/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}})' var(\hat{\boldsymbol{\beta}}_{\mathrm{s}})(\partial g/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}}), \tag{2.4}$$

where $\partial g/\partial\hat{\boldsymbol{\beta}}_{\mathrm{s}} = (\partial g/\partial\beta_1, \partial g/\partial\beta_2, \partial g/\partial\beta_{12})'$ evaluated at $\boldsymbol{\beta}_{\mathrm{s}} = \hat{\boldsymbol{\beta}}_{\mathrm{s}}$. Therefore, in order to obtain good estimate of x_2, we need to estimate $\boldsymbol{\beta}_{\mathrm{s}} = (\beta_1, \beta_2, \beta_{12})'$ as well as possible.

We have argued that, for both model (1.10) and model (1.12), D_s-optimal designs that address estimation of all the parameters except the intercept term β_0 are efficient for estimation of a particular ED. In this section, we will study D_s-optimal designs of this type for the one-toxicant second-order model and the two-toxicant interaction model. We shall first give a brief introduction of D_s-optimality criterion.

Suppose that the vector of model parameters is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and we want to estimate the subset $\boldsymbol{\beta}_2$ as precisely as possible. Correspondingly, the Fisher information matrix is partitioned as:

$$I(X, \boldsymbol{\beta}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{pmatrix},$$

where I_{11} is the matrix associated with the information of the parameters comprising $\boldsymbol{\beta}_1$ and I_{22} is associated with $\boldsymbol{\beta}_2$. In order to obtain the generalized variance of $\hat{\boldsymbol{\beta}}_2$, the estimate of $\boldsymbol{\beta}_2$, we need to find the inverse of the Fisher information matrix in terms of its partitioned form, which is given by

$$I^{-1}(X, \boldsymbol{\beta}) = \begin{pmatrix} C_{11}^{-1} & I_{11}^{-1} I_{12} C_{22} \\ C_{22} I_{12}' I_{11}^{-1} & C_{22}^{-1} \end{pmatrix},$$

where $C_{11} = I_{11} - I_{12} I_{22}^{-1} I_{12}'$ and $C_{22} = I_{22} - I_{12}' I_{11}^{-1} I_{12}$. Asymptotically, we have

$$var(\hat{\boldsymbol{\beta}}_2) = C_{22}^{-1} = (I_{22} - I_{12}' I_{11}^{-1} I_{12})^{-1}.$$

Therefore, the D_s-optimal design that minimizes the generalized variance of $\hat{\boldsymbol{\beta}}_2$ is the one which minimizes $|(I_{22} - I_{12}' I_{11}^{-1} I_{12})^{-1}|$ or equivalently maximizes $|I_{22} - I_{12}' I_{11}^{-1} I_{12}|$. So the D-optimality criterion is defined as

$$\max_{X \in \mathcal{D}} |I_{22} - I_{12}' I_{11}^{-1} I_{12}|, \tag{2.5}$$

or equivalently,

$$\max_{X \in \mathcal{D}} \log |I_{22} - I_{12}' I_{11}^{-1} I_{12}|, \tag{2.6}$$

where $\mathcal{D}$ is the set of all possible designs.

2.4.2 One-toxicant Second-order Model

We now investigate the D_s-optimal design for model (1.10). More specifically, we shall study the D_s-optimal design that addresses the estimation of $\boldsymbol{\beta}_s = (\beta_1, \beta_{11})'$ since optimal designs of this type are expected to be efficient for ED estimation.

Like the D-optimal design, the D_s-optimal design for model (1.10) also depends on the model parameters. Lemma 2.3 describes this dependence, the proof of which is given in Appendix A.3.

Lemma 2.3. *For model (1.10), the D_s-optimal design for estimating $\boldsymbol{\beta}_2 = (\beta_1, \beta_{11})'$, in terms of ED levels, depends on the parameters only through $r = \beta_1^2/\beta_{11}$.*

For a given r, we can use Nelder-Mead method to optimize the D_s-optimality criterion function in (2.6) and find the optimal design. We will consider two different design regions: one is unrestricted ($q \geq 0.01$) and the other is restricted by $q \geq 0.30$. For some typical values of r, Table 2.9 lists the D_s-optimal designs on these two different design regions. Based on Table 2.9, we can make some comments about D_s-optimal designs for model (1.10):

1. Like the D-optimal design, the D_s-optimal design for model (1.10) is also a three-point design. The control point ($q_1 = 1$) is in the optimal design and point 3 is practically fixed at the lower boundary of the design space. The location of the middle point, or point 2, depends on the value of r and it moves toward the lower boundary as r decreases. However, the range of q_2 is not very large especially when the design space is restricted.

2. Unlike the D-optimal design, the D_s-optimal design for model (1.10) is not an equal allocation design. More specifically, the control point has the smallest allocation and the non-control point with lower ED level has the largest allocation. Recall that the intercept term β_0, which is estimated solely from the control point, is of less importance than the other parameters in the D_s-optimal design we considered; so less allocation is given to the control point. The value of r affects the allocation; however, the impact is negligible especially when the design space is restricted.

3. When the design space is restricted, the fact that both the locations of the design points and the allocation are robust to r implies the existence of efficient designs robust to r. For instance, when the design space is restricted by $q \geq 0.30$, we would expect that the design with $(q_1, q_2, q_3) = (1, 0.6, 0.3)$ and $(p_1, p_2, p_3) = (0.3, 0.3, 0.4)$ is quite efficient for all possible values of r.

2.4.3 Two-toxicant Interaction Model

We now turn to the two-toxicant interaction model in (1.12). For this model, we shall study the D_s-optimal design that addresses the estimation of all the parameters except the intercept

Table 2.9: D_s-optimal Designs for Model (1.10)

Design Space: $c = 0.01$						
r	Design Points			Allocations		
(β_1^2/β_{11})	q_1	q_2	q_3	p_1	p_2	p_3
0^a	1.0000	0.6081	0.0561	0.2342	0.2958	0.4701
-1	1.0000	0.4229	0.0313	0.2077	0.3134	0.4789
-2	1.0000	0.3799	0.0259	0.2001	0.3186	0.4813
-5	1.0000	0.3237	0.0192	0.1890	0.3262	0.4847
-10	1.0000	0.2868	0.0149	0.1809	0.3320	0.4872
-20	1.0000	0.2579	0.0118	0.1738	0.3370	0.4892
-50	1.0000	0.2361	0.0100	0.1685	0.3409	0.4905
$-\infty^b$	1.0000	0.2231	0.0100	0.1654	0.3445	0.4901
Design Space: $c = 0.30$						
r	Design Points			Allocations		
(β_1^2/β_{11})	q_1	q_2	q_3	p_1	p_2	p_3
0	1.0000	0.7762	0.3000	0.2819	0.3123	0.4058
-1	1.0000	0.6554	0.3000	0.2742	0.3248	0.4010
-2	1.0000	0.6337	0.3000	0.2728	0.3272	0.4000
-5	1.0000	0.6102	0.3000	0.2711	0.3300	0.3989
-10	1.0000	0.5981	0.3000	0.2702	0.3314	0.3984
-20	1.0000	0.5897	0.3000	0.2688	0.3341	0.3971
-50	1.0000	0.5831	0.3000	0.2683	0.3347	0.3970
$-\infty$	1.0000	0.5815	0.3000	0.2679	0.3352	0.3969

[a] $r = 0$ means $\beta_1 = 0$.
[b] $r = -\infty$ means $\beta_{11} = 0$.

term since optimal designs of this type are expected to be efficient for ED estimation. Like the D-optimal design, the D_s-optimal design for model (1.12) also depends on the model parameters. Parallel to Lemma 2.2, Lemma 2.4 describes this dependence, the proof of which is given in Appendix A.3.

Lemma 2.4. *For model (1.12), the D_s-optimal design for estimating $\boldsymbol{\beta}_2 = (\beta_1, \beta_2, \beta_{12})'$, in terms of IED levels, depends on the parameters only through $r = \beta_{12}/\beta_1\beta_2$.*

Nelder-Mead method is used to optimize the D_s-optimality in (2.6) and find the optimal design for a given r. We will consider two different design regions: one is unrestricted ($c_1 = c_2 = 0.01$) and the other is restricted by $c_1 = c_2 = 0.30$. For some typical values of r between -1 and the upper bound defined in (2.1), Table 2.10 lists the D_s-optimal designs on these two different design regions. Based on Table 2.10, we can summarize some features of the D_s-optimal design for model (1.12):

Table 2.10: D_s-optimal Designs for Model (1.12)

Design Space: $c_1 = c_2 = 0.01$								
	Design Points				Allocations			
r[a]	$q_{11} = q_{21}$	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$	p_1	p_2	p_3	p_4
-1	1.0000	0.0961	1.0000	0.3507	0.1200	0.2846	0.2846	0.3108
-0.5	1.0000	0.0955	1.0000	0.2754	0.1179	0.2838	0.2839	0.3144
-0.2	1.0000	0.0948	1.0000	0.2041	0.1158	0.2829	0.2829	0.3183
-0.1	1.0000	0.0944	1.0000	0.1706	0.1147	0.2825	0.2825	0.3203
0	1.0000	0.0939	1.0000	0.1271	0.1132	0.2818	0.2818	0.3232
0.1	1.0000	0.0930	1.0000	0.0587	0.1104	0.2807	0.2807	0.3282
0.2	1.0000	0.0929	1.0000	0.0100	0.1102	0.2806	0.2806	0.3287
Ub[b]	1.0000	0.0932	1.0000	0.0100	0.1112	0.2810	0.2810	0.3268
Design Space: $c_1 = c_2 = 0.30$								
	Design Points				Allocations			
r	$q_{11} = q_{21}$	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$	p_1	p_2	p_3	p_4
-1	1.0000	0.3000	1.0000	0.3573	0.1670	0.2568	0.2568	0.3194
-0.5	1.0000	0.3000	1.0000	0.3000	0.1671	0.2567	0.2567	0.3194
-0.2	1.0000	0.3000	1.0000	0.3000	0.1698	0.2586	0.2586	0.3130
-0.1	1.0000	0.3000	1.0000	0.3000	0.1709	0.2594	0.2594	0.3104
0	1.0000	0.3000	1.0000	0.3000	0.1721	0.2602	0.2602	0.3074
0.1	1.0000	0.3000	1.0000	0.3000	0.1735	0.2612	0.2612	0.3042
0.2	1.0000	0.3000	1.0000	0.3000	0.1750	0.2622	0.2622	0.3006
0.5	1.0000	0.3000	1.0000	0.3000	0.1804	0.2658	0.2658	0.2880
Ub	1.0000	0.3000	1.0000	0.3000	0.1881	0.2706	0.2706	0.2706

[a] $r = \beta_{12}/\beta_1\beta_2$.
[b] Ub equals to the upper bound defined in (2.1).

1. Like the D-optimal design, the D_s-optimal design for model (1.12) is also a saturated design. The four points are the control point, two pure component points and an interaction point. The optimal design is symmetric about the line $q_1 = q_2$ in that the control point and the interaction point are on the line $q_1 = q_2$ and the two pure component points are symmetric about the line. For the unrestricted design space, locations of the pure component points depend on r; however, the impact of r is negligible and the pure component point is practically fixed at IED_9. For the design space restricted by $c_1 = c_2 = c \geq 0.10$, the pure component points are fixed at IED_{100c}. The interaction point moves from (c_1, c_2) to the control point as r decreases from the upper bound.

2. As expected, the D_s-optimal design is not an equal allocation design. The control point

has the smallest allocation as estimation of the intercept is of less importance; and the interaction point has the largest allocation. The two pure component points have the same proportion of the total sample. Though the value of r affects the allocation of the experimental runs, the impact is negligible.

3. When the design space is restricted, we can find some efficient designs robust to the value of r. For instance, when the design space is restricted by $c_1 = c_2 = 0.30$, the design with design points located at the four corners of the design space and the allocation $(p_1, p_2, p_3, p_4) = (0.2, 0.25, 0.25, 0.3)$ is expected to be quite efficient for any value of r between -1 and the upper bound defined in (2.1).

2.5 Summary

In this chapter, we have discussed locally D-optimal and D_s-optimal designs for Poisson regression models in toxicity studies. We now conclude this chapter by summarizing some highlights:

1. To implement any design we have discussed in this chapter, practitioners should have good initial estimates of parameters or some knowledge about ED's of the toxicants. This is the core assumption of locally optimal designs.

2. Lemma 2.1 and Lemma 2.2 state that D-optimal designs depend on unknown parameters only through some function of the parameters. This property does not make it any easier for practitioners to implement the locally D-optimal design. Yet, just as we have seen in this chapter, this property enables us to study performance of any particular design over the whole parameter space by studying all possible or practical values of the function.

3. For the one-toxicant second-order model, the D-optimal design depends on β_1^2/β_{11}. In practice, equally spaced designs, three-point or four-point, are recommended for simplicity, high efficiency and good robustness. For the two-toxicant interaction model, the D-optimal design depends on $\beta_{12}/\beta_1\beta_2$. The conditional D-optimal design is a good choice in practice. To implement the conditional D-optimal design, only knowledge of main effects is needed. Information about the interaction term, which is more difficult to estimate than main effects, is not required. If model checking is needed, one can use a five-point design.

4. D-optimal designs have been the focus of this chapter. We know that D-optimality is an "estimation criterion" and, in many applications, quality of prediction is another important issue. It has been shown that continuous D-optimal designs are also G-optimal (see Silvey, 1980). We know that G-optimality is a "prediction criterion" in that it minimizes the maximum variance the of prediction over the region of interest.

Thus, D-optimal designs in this chapter have good prediction properties in the sense of G-optimality.

5. We have investigated D_s-optimal designs for estimating all the parameters except the intercept term as it is felt that designs of this type are efficient for ED estimation. Lemma 2.3 and Lemma 2.4 describe the dependence of the D_s-optimal designs upon the parameters for model (1.10) and model (1.12), respectively. For both models, the D_s-optimal design is saturated with unequal allocation. When the design space is restricted, the structure of the optimal design is quite stable regardless the value of r, which implies that efficient designs robust to r exist.

Optimal designs in this chapter are found via numerical methods. Their optimality is in question and will be verified through equivalence theory in Chapter 3. Robustness studies show that the designs we have discussed are not robust to parameter misspecifications, like most other locally optimal designs. Designs robust to parameter misspecifications will receive more attention in Chapter 4.

Chapter 3

Equivalence Theory

3.1 Introduction

In Chapter 2, the optimal designs were obtained largely through the use of numerical optimization. Since they were not found by exact mathematical optimization methods, one cannot be assured that they are theoretically the optimal designs. Equivalence theory is a tool by which we can show whether or not a particular design satisfies a given optimality criterion. Basically, equivalence theory combines techniques from measure theory and calculus to show that the property of an optimal design satisfying the criterion of interest is equivalent to some other properties of the design. Thus, if we are able to show that a design has these "other" properties, we can conclude its optimality.

Kiefer and Wolfowitz (1960) first developed equivalence theory for D-optimality and G-optimality in the context of linear models. Their results were generalized to nonlinear models by Fedorov (1972). Silvey (1980) extended equivalence theory to cover more optimality criteria for both linear and nonlinear models. These criteria include D-, G-, E-optimality and the optimality criteria known as "linear criteria functions", which includes, for example, Q-optimality as a special case.

In this chapter, we shall first give a brief introduction of the equivalence theory. Then the optimality of locally D-optimal and D_s-optimal designs discussed in Chapter 2 will be verified via equivalence theory. Finally, we will utilize equivalence theory to generalize D-optimal designs for two-toxicant interaction models to multi-toxicant experiments.

3.2 Equivalence Theory

In this section, we will give an overview of equivalence theory. This exposition on equivalence theory is mainly based on Silvey's results (Silvey, 1980).

As before, we denote the unknown parameters by $\boldsymbol{\beta}$ and the regressor vector by $\mathbf{x}$. The probability density function of the response variable y at $\mathbf{x}$ is $p(y|\mathbf{x}, \boldsymbol{\beta})$. We are given a design space from which we may choose $\mathbf{x}$ vectors at which to observe y. We denote the Fisher information matrix for a single observation y at $\mathbf{x}$ by $J(\mathbf{x}, \boldsymbol{\beta})$. Suppose that the dimension of $\boldsymbol{\beta}$ is p; then $J(\mathbf{x}, \boldsymbol{\beta})$ is a $p \times p$ matrix, whose $(i, j)^{th}$ element is $E\left(-\partial^2 \log p(y|\mathbf{x}, \boldsymbol{\beta})/\partial\beta_i\partial\beta_j\right)$. In an n-observation design, if all the n observations made at $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent of each other, then the information matrix is $I(X, \boldsymbol{\beta}) = \sum_{i=1}^{n} J(\mathbf{x}_i, \boldsymbol{\beta})$, where $X' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$.

For an n-observation design, our objective is to choose a design X from the design space that is best according to some design criterion function ϕ. In general, ϕ is a real-valued function of the Fisher information matrix, i.e., $\phi = \phi(I(X, \boldsymbol{\beta}))$. For example, ϕ is the determinant of the Fisher information matrix for D-optimality criterion. It is assumed without loss of generality that the best design according to the criterion ϕ, or the ϕ-optimal design, is the one which *maximizes* ϕ. Note that if the design criterion is one which requires the minimization of some function γ, then ϕ is taken to be the negative of γ. Thus, we have reduced the design problem to maximization of $\phi(I(X, \boldsymbol{\beta}))$, which is a numerical analysis problem. However, because of the discrete nature of the n-observation design and the consequent unpleasantness of the corresponding set $\mathcal{I}(X, \boldsymbol{\beta})$ of matrices $I(X, \boldsymbol{\beta})$ on which ϕ is defined, this numerical analysis problem can not be solved by standard optimization techniques (Silvey, 1980, p.13). This situation is analogous to the much simpler one where we wish to maximize a function defined on the integers. Because of the discrete domain, calculus techniques can not be exploited in the solution. A commonly used device for this simpler problem is to extend the definition of the function to all real numbers R, use calculus to find the number r^* that maximizes the extended function and argue that the maximum of the function over the integers will occur at an integer adjacent to r^*. This idea has been adapted for the n-observation design problem and leads to what Kiefer has termed *approximate theory*.

We shall briefly outline the idea of approximate theory for n-observation designs. Generally, an n-observation design consists of m distinct design points, denoted by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, with each $\mathbf{x}_i$ replicated n_i times, where $i = 1, 2, \ldots, m$. Such an n-observation design can be characterized by the probability distribution on $\mathcal{X}$, which is the set of all $\mathbf{x}$ vectors in the design space. This probability distribution, denoted by η_n, assigns probability $p_i = n_i/n$ to the point $\mathbf{x}_i$ for $i = 1, 2, \ldots, m$. This is a discrete probability, and the n-observation design is considered a discrete design. Suppose $\mathbf{x}$ is a random vector with probability distribution η_n and we denote $E(J(\mathbf{x}, \boldsymbol{\beta}))$ by $M(\eta_n, \boldsymbol{\beta})$; then, for an n-observation design, we have

$$M(\eta_n, \boldsymbol{\beta}) = E(J(\mathbf{x}, \boldsymbol{\beta})) = \sum_{i=1}^{m} p_i J(\mathbf{x}_i, \boldsymbol{\beta}) = I(X, \boldsymbol{\beta})/n. \tag{3.1}$$

Note that (3.1) implies that $I(X, \boldsymbol{\beta}) = nM(\eta_n, \boldsymbol{\beta})$. According to Silvey, for most of the commonly used criterion functions such as D-optimality criterion, the following relationship holds for any real positive constant a:

$$\phi(aM(\eta_n, \boldsymbol{\beta})) = k\phi(M(\eta_n, \boldsymbol{\beta})),$$

where k is a positive constant. Therefore, for an n-observation design, we have

$$\phi(I(X, \boldsymbol{\beta})) = c\phi(M(\eta_n, \boldsymbol{\beta})),$$

where c is some positive constant. Recall that the n-observation design problem is to find a design X to maximize $\phi(I(X, \boldsymbol{\beta}))$, which is hard to deal with because of the discreteness. According to the above equation, this problem is equivalent to finding η_n^*, a probability measure on the set $\mathcal{X}$ corresponding to an n-observation design, that maximizes $\phi(M(\eta_n, \boldsymbol{\beta}))$. Unfortunately, this reinterpretation of the problem in itself does not make the problem any easier to solve because the set $\mathcal{M}(\eta_n, \boldsymbol{\beta})$ of matrices $M(\eta_n, \boldsymbol{\beta})$ on which ϕ is defined is also discrete and does not lend itself to the standard optimization techniques. However, this problem can be remedied by extending the definition of ϕ.

We now extend the definition of $\mathcal{M}(\eta_n, \boldsymbol{\beta})$ to the set H of *all* probability distributions on $\mathcal{X}$. The extended set is defined as follows:

$$\mathcal{M}(\eta, \boldsymbol{\beta}) = \{M(\eta_n, \boldsymbol{\beta}) : \eta \in H\}.$$

Now consider the problem of finding η^* to maximize $\phi(M(\eta, \boldsymbol{\beta}))$ over H. This is much more tractable since $\mathcal{M}(\eta, \boldsymbol{\beta})$ is a set with much nicer properties than $\mathcal{M}(\eta_n, \boldsymbol{\beta})$ and we have the opportunity of exploiting calculus techniques in solving it. If we can find a η^* to solve this problem then hopefully an n-observation design X^*, whose associated probability distribution approximates η^*, will be close to optimal for the n-observation design problem. In the literature, η^* corresponds to the continuous optimal design and X^* corresponds to the discrete optimal design. All the designs discussed in Chapter 2 are actually continuous designs. Algorithms converting continuous designs to discrete designs can be found in Atkinson and Donev (1992).

We now shift to the mathematical details which lead to the equivalence theory. We first give the definition of a directional derivative called the *Fréchet derivative*, which plays a pivotal role in the equivalence theory. The *Fréchet derivative* is defined as follows:

Definition 3.1. *For any $M_1, M_2 \in \mathcal{M}$, the Fréchet derivative is defined as:*

$$F_\phi(M_1, M_2) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}[\phi\{(1-\varepsilon)M_1 + \varepsilon M_2\} - \phi(M_1)].$$

In the above definition, M_1 and M_2 are short for $M(\eta_1, \boldsymbol{\beta})$ and $M(\eta_2, \boldsymbol{\beta})$, where $\eta_1, \eta_2 \in H$. We will follow this convention throughout this chapter unless otherwise stated.

After giving the definition of the *Fréchet derivative*, we are now ready for the equivalence theory. Silvey (1980) proves the following theorem, which is the cornerstone of Silvey's equivalence theory. This theorem is also referred to as *general equivalence theory* in the literature.

Theorem 3.1. (Silvey, 1980) *If $\boldsymbol{\beta}$ is fixed and ϕ is concave on $\mathcal{M}$, the following three statements are equivalent to each other:*

1. *η^* maximizes $\phi(M(\eta, \boldsymbol{\beta}))$ over H;*
2. *$F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) \leq 0$ for all $\mathbf{x} \in \mathcal{X}$;*
3. *$F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ achieves its maximum at the points of the design.*

Theorem 3.1 is the central result on which the optimum design of experiments depends. It provides a method for checking the optimality of a design. Basically, in order to verify the optimality of a particular design η^* , one can evaluate the *Fréchet derivative* of that design at all the points in the design space. In turn, the *Fréchet derivative* must be less than or equal to zero at all points in the design space and equal to zero at the design points. In next section, we will use this method to verify the optimality of the locally D-optimal $\mathrm{D_s}$-optimal designs we have discussed in Chapter 2.

3.3 Optimality of Locally D-optimal Designs

In this section, we will use the general equivalence theorem to verify optimality of the locally D-optimal designs listed in Chapter 2. In general, when deriving the D-optimal design, we simply define the design criterion function as the determinant of the information matrix, i.e., $\phi(M(\eta, \boldsymbol{\beta})) = |M(\eta, \boldsymbol{\beta})|$. Recall that Theorem 3.1 requires that the design criterion function ϕ should be concave on $\mathcal{M}$. To ensure the concavity, an equivalent design criterion, $\phi(M(\eta, \boldsymbol{\beta})) = \log(|M(\eta, \boldsymbol{\beta})|)$, is generally adopted. Taking the logarithm of the determinant leads to maximization of a concave function, so that any maximum found will certainly be global rather than local. We shall use the logarithm of the determinant as the criterion function throughout this chapter.

To verify the optimality of a particular design η^*, we need to demonstrate that the *Fréchet derivative* $F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ is bounded by 0. From Definition 3.1, we can see that computation of $F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ involves taking the limit, which is awkward to derive in general. Fortunately, Silvey (1980) gives a computation formula of the *Fréchet derivative* as follows:

$$F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) = trace[J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})] - p, \tag{3.2}$$

where p is the number of parameters. Therefore, if we can demonstrate that trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta^*, \boldsymbol{\beta})$ is less than or equal to p over the design space, then we can conclude the design η^* is optimal. In turn, at the design points, the trace should take the value of p. Next, we will try to evaluate the trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})$ for the designs claimed to be optimal in Chapter 2.

3.3.1 One-toxicant Second-order Model

In general, $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on the parameter $\boldsymbol{\beta}$. For the one-toxicant second-order model and the D-optimality criterion, the following lemma reveals the dependence.

Lemma 3.1. *For model (1.10) and the D-optimality criterion, $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_1^2/\beta_{11}$.*

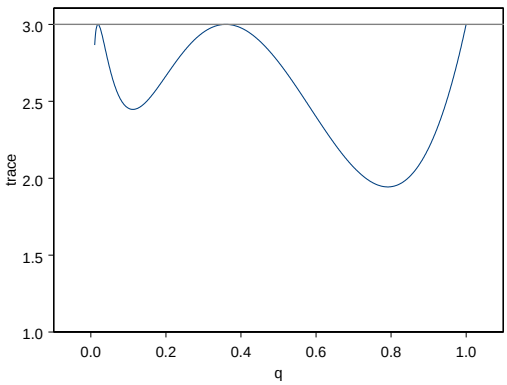
The proof of Lemma 3.1 is given in Appendix A.4. With Lemma 3.1, our work of verifying optimality for all possible sets of parameters is reduced to all possible or reasonable values of r. It has the same function as Lemma 2.1 in finding the D-optimal design. Lemma 3.1 enables us to verify optimality of the designs catalogued by r in Section 2.2.

We have evaluated the trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})$ over the design space for all the designs listed in Table 2.1. For all these designs, we have found that, over the design space, the trace achieves the maximum of 3, the number of model parameters, at the design points. Thus, they are optimal designs as claimed. Figure 3.1 plots the trace function over the design space for four cases. Some conclusions can be drawn from the plots:

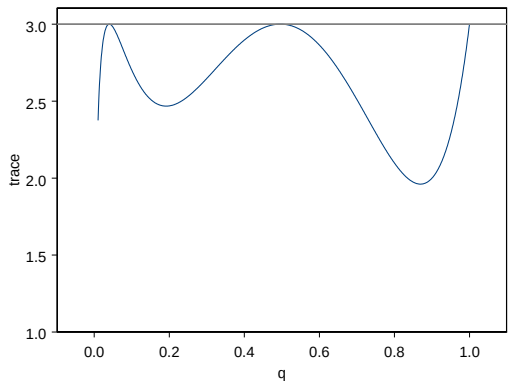
1. In all the plots, the trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})$ is less than or equal to 3, the number of parameters, which further implies that $F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) \leq 0$ over the design space. By Theorem 3.1, we can conclude that η^* maximizes $\phi(M(\eta, \boldsymbol{\beta}))$. In other words, the design is D-optimal. The plots also show that the trace achieves its maximum at the design points, demonstrating equivalence of statement (2) and statement (3) in Theorem 3.1.

2. In Figure 3.1(a) and 3.1(b), the design space is unrestricted. For Figure 3.1(a), $r = -10$ and the trace function achieves its maximum at $q = 1, q = 0.3594$ and $q = 0.0190$, corresponding to the three design points listed in Table 2.1. For Figure 3.1(b), we have the same observation. When comparing these two plots, we can find that the two non-control points move closer to the control point when r increases from -10 to -1, which is in accordance with the conclusion in Section 2.2.

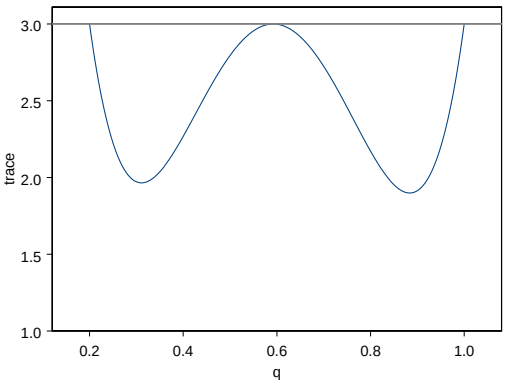
3. Figure 3.1(c) and 3.1(d) plot the trace function on restricted regions. As we can see that the function has the two end points as its maximum points, in agreement with the fact that the D-optimal design on restricted regions always includes the two end points.



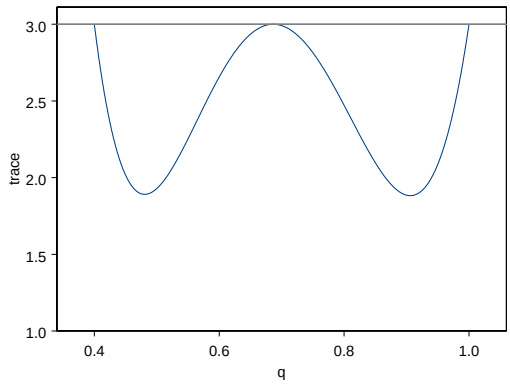
(a) Unrestricted Region: $r = -10$



(b) Unrestricted Region: $r = -1$



(c) Restricted Region ($c = 0.2$): $r = -2$



(d) Restricted Region ($c = 0.4$): $r = -5$

Figure 3.1: D-optimality of Designs for One-toxicant Second-order Model. The trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta^*, \boldsymbol{\beta})$ (1) is less than or equal to 3, the number of parameters, over the design space; (2) achieves its maximum at the design points, confirming D-optimality of the designs.

3.3.2 Two-toxicant Interaction Model

We now discuss D-optimality of the designs for the two-toxicant interaction model. Parallel to Lemma 3.1, the following lemma describes how the *Fréchet derivative* of D-optimality

criterion depends on $\boldsymbol{\beta}$ for this model.

Lemma 3.2. *For model (1.12) and the D-optimality criterion, $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_{12}/\beta_1\beta_2$.*

The proof of the above lemma is given in Appendix A.5. For the two-toxicant interaction model, Lemma 3.2 enables us to verify optimality of the designs catalogued by r in Section 2.3.

We have evaluated the trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})$ over the design space for all the designs listed in Table 2.4. For all these designs, we have found that, over the design space, the trace achieves the maximum of 4, the number of model parameters, at the design points. Thus, they are D-optimal designs. Figure 3.2 plots this function over the design space for four cases. Some comments are in order:

1. In all the plots, the trace function is less than or equal to 4, the number of parameters, over the design space, implying that $F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) \leq 0$ over the design space. By Theorem 3.1, all these designs are D-optimal. Also, the trace achieves its maximum at the design points, demonstrating equivalence of statement (2) and statement (3) in Theorem 3.1.

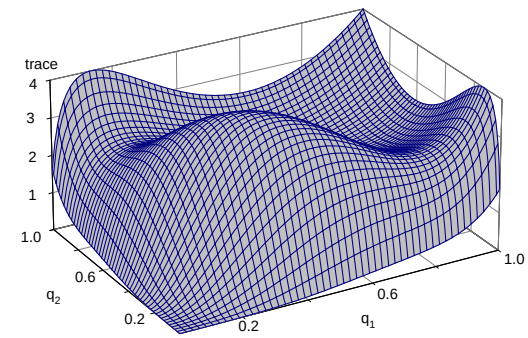
2. In Figure 3.2(a) and 3.2(b), the design space is unrestricted. In both plots, the maximum of the trace function is obtained at the control point, two pure component points located at $\text{IED}_{13.53}$, and an interaction point. For $r = -1$, the interaction point is $(0.3679, 0.3679)$ in both the optimal design listed in Table 2.4 and the Figure 3.2(a). When r increases to 0 in Figure 3.2(b), the interaction point moves to $(0.1353, 0.1353)$, corresponding to the interaction point in the conditional D-optimal design. The movement of the interaction point according to r is in agreement with the conclusion in Section 2.3.

3. Figure 3.2(c) and 3.2(d) plot the trace function over restricted regions. In both cases, the trace function achieves the maximum at the four corners of the restricted design space. Recall that, in Table 2.4, the corresponding D-optimal designs have the same structure.

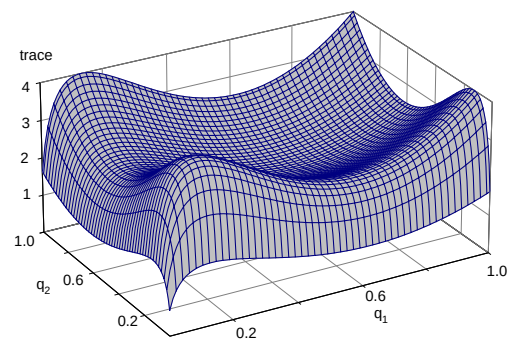
In summary, the locally D-optimal designs discussed in Section 2.2 and Section 2.3 are really optimal as claimed.

3.4 Optimality of Locally $\mathrm{D_s}$-optimal Designs

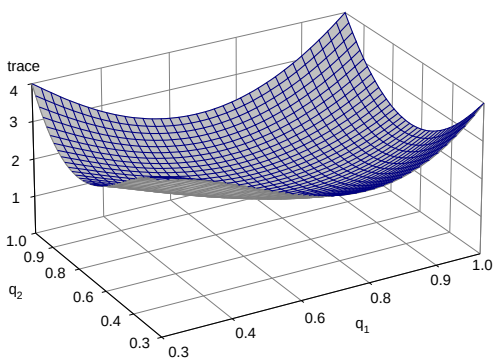
In this section, we will adopt the general equivalence theory to verify optimality of the $\mathrm{D_s}$-optimal designs discussed in Chapter 2. To this end, we need to demonstrate that the



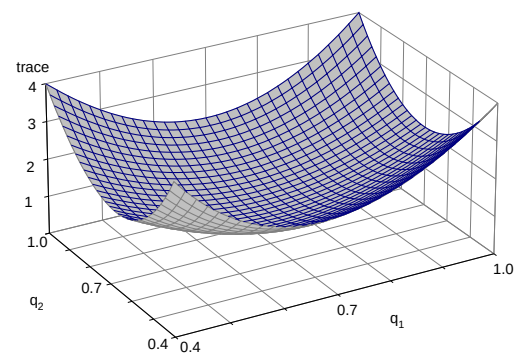
(a) Unrestricted Region: $r = -1$



(b) Unrestricted Region: $r = 0$



(c) Restricted Region ($c_1 = c_2 = 0.3$): $r = -0.5$



(d) Restricted Region ($c_1 = c_2 = 0.4$): $r = 0.5$

Figure 3.2: D-optimality of Designs for Two-toxicant Interaction Model. The trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta^*, \boldsymbol{\beta})$ (1) is less than or equal to 4, the number of model parameters, over the design space; (2) achieves its maximum at the design points, confirming D-optimality of the designs.

Fréchet derivative of the locally optimal design is less than or equal to 0 over the design space. Suppose that the vector of the model parameters is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ and we want to estimate the subset $\boldsymbol{\beta}_2$ as precisely as possible. Correspondingly, for a design η, the Fisher information matrix is partitioned as:

$$M(\eta) = \begin{pmatrix} M_{11}(\eta) & M_{12}(\eta) \\ M_{12}'(\eta) & M_{22}(\eta) \end{pmatrix}.$$

In the same fashion, a design point $\mathbf{x}$ is partitioned as $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$. For a generalized linear model, it can be shown that the *Fréchet derivative* of the $\mathrm{D_s}$-optimality criterion for a design η can be computed as follows (see Atkinson and Donev, 1992):

$$F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) = v(\mathbf{x})(\mathbf{x}' I^{-1}(\eta)\mathbf{x} - \mathbf{x}_1' I_{11}^{-1}(\eta)\mathbf{x}_1) - s, \tag{3.3}$$

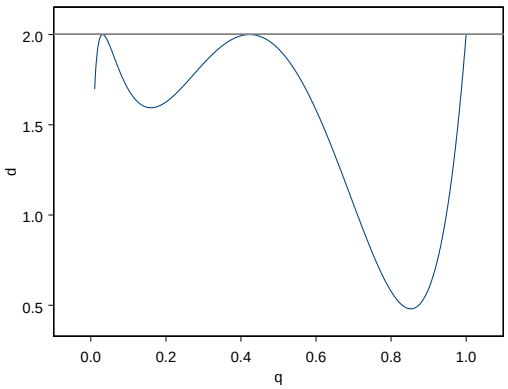
where η is any design, $\mathbf{x}$ is any point in the design space, $v(\mathbf{x})$ is the variance function of the response at $\mathbf{x}$ and s is the dimension of $\boldsymbol{\beta}_2$. Thus, if we can demonstrate that, for a design η, $d(\mathbf{x}, \eta) = v(\mathbf{x})(\mathbf{x}' I^{-1}(\eta)\mathbf{x} - \mathbf{x}_1' I_{11}^{-1}(\eta)\mathbf{x}_1)$ is less than or equal to s at any point in the design space, then we can conclude that η is $\mathrm{D_s}$-optimal by equivalence theory. For the locally $\mathrm{D_s}$-optimal designs in Chapter 2, we will evaluate and plot the $d(\mathbf{x}, \eta)$ over the design space to demonstrate that they are optimal designs as claimed.

3.4.1 One-toxicant Second-order Model

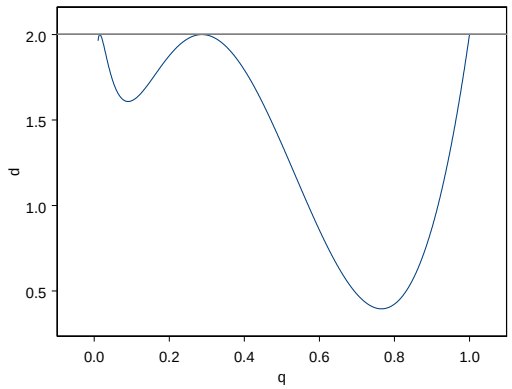
Parallel to Lemma 3.1, for model (1.10) and $\mathrm{D_s}$-optimality criterion, the next lemma describes the dependence of the *Fréchet derivative* on the model parameters. Its proof is outlined in Appendix A.6.

Lemma 3.3. *For model (1.10) and the D_s-optimality criterion, $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_1^2/\beta_{11}$.*

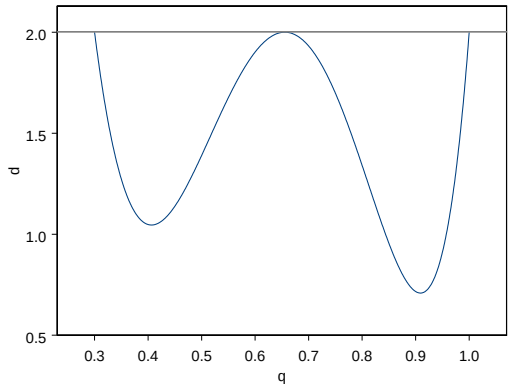
The above lemma enables us to evaluate the *Fréchet derivative* of the designs listed in Table 2.9, which are catalogued by r. For all the designs in Table 2.9, we have found that the *Fréchet derivative* is less than or equal to 0 so they are $\mathrm{D_s}$-optimal. Figure 3.3 plots the function $d(\mathbf{x}, \eta)$ over the design space for four cases. In all the plots, we can see that $d(\mathbf{x}, \eta)$ is less than or equal to 2, the dimension of $\boldsymbol{\beta}_2$. Further, $d(\mathbf{x}, \eta)$ achieves the maximum of 2 at all the design points. Thus, the *Fréchet derivative* is less than or equal to 0 over the design space and achieves the maximum of 0 at all the design points, which implies that all the designs are $\mathrm{D_s}$-optimal according to the equivalence theory.



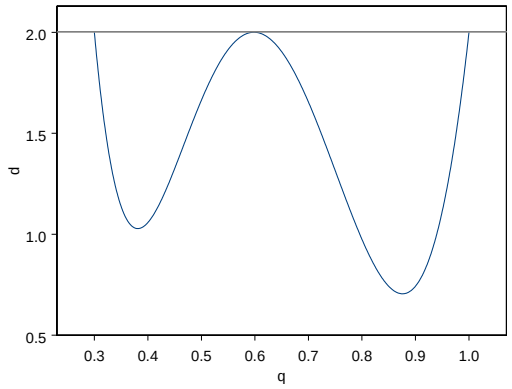
(a) Unrestricted Region: $r = -1$



(b) Unrestricted Region: $r = -10$



(c) Restricted Region ($c = 0.3$): $r = -1$



(d) Restricted Region ($c = 0.3$): $r = -10$

Figure 3.3: D_s-optimality of Designs for One-toxicant Second-order Model. The function $d(\mathbf{x}, \eta)$ (1) is less than or equal to 2, the dimension of $\boldsymbol{\beta}_2 = (\beta_1, \beta_{11})'$, over the design space; (2) achieves its maximum at the design points, confirming D_s-optimality of the designs.

3.4.2 Two-toxicant Interaction Model

For model (1.12) and the D_s-optimality criterion, Lemma 3.4 states the dependence of the *Fréchet derivative* on the parameters. Its proof is outlined in Appendix A.6.

Lemma 3.4. *For model (1.12) and the D_s-optimality criterion, $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_{12}/\beta_1\beta_2$.*

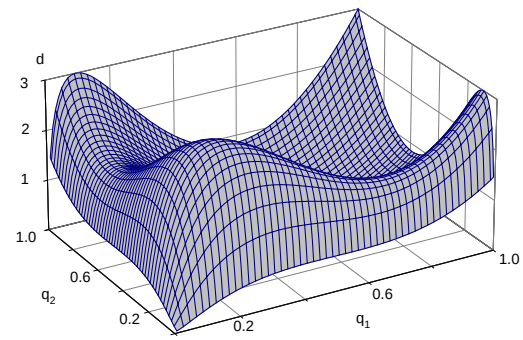
Based on the above lemma, we can ignore values of the parameters and verify optimality of designs catalogued by r. We have found that, for all the designs in Table 2.10, the *Fréchet derivative* is less than or equal to 0 and takes 0 only at all the design points so, by the general equivalence theory, they are optimal as claimed. Figure 3.4 plots the function $d(\mathbf{x}, \eta)$ over the design space for four cases. In all the plots, we can see that $d(\mathbf{x}, \eta)$ is less than or equal to 3, the dimension of $\boldsymbol{\beta}_2 = (\beta_1, \beta_2, \beta_{12})'$. Further, $d(\mathbf{x}, \eta)$ achieves the maximum of 3 only at all the design points. Thus, the *Fréchet derivative* is less than or equal to 0 over the design space and achieves the maximum of 0 only at all the design points, which implies that all the designs are D_s-optimal.

In summary, the locally D_s-optimal designs discussed in Section 2.4 are really optimal as claimed.

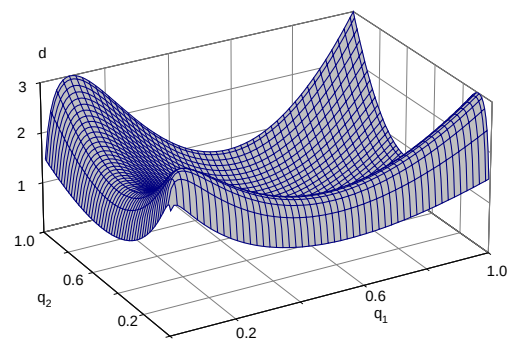
3.5 D-optimal Designs for Multi-toxicant Interaction Model

Up to this point, we have just considered one-toxicant and two-toxicant experiments. However, it is common that real toxicity experiments involve more than two toxicants. In this section, we try to generalize the discovery concerning D-optimal designs in Chapter 2 to multi-toxicant experiments. Particularly, we will focus on the multi-toxicant model that only involves main effects and two-way interactions, which, for brevity, will be referred to as the multi-toxicant interaction model.

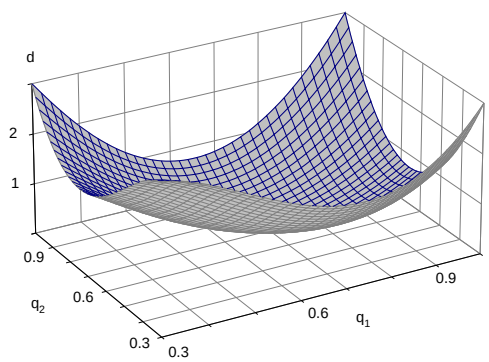
In Chapter 2, numerical methods such as Nelder-Mead simplex method were employed to maximize the determinant of the Fisher information matrix and find the optimal designs. The difficulty to extend this procedure to multi-toxicant models is that most numerical methods give unstable solutions to the problem as the number of variables goes up. Take the three-toxicant interaction model for example; we have seven parameters to estimate so we need at least a seven-point design. At each design point we need to determine three IED's and the proportion of total sample at this point. For a seven-point design, we have 21 variables. Even if we assume that the design is equal-allocation and the control point is in the design, the total number of variables is still 18, which is too large for most numerical methods. The problem becomes more complicated for experiments involving more toxicants.



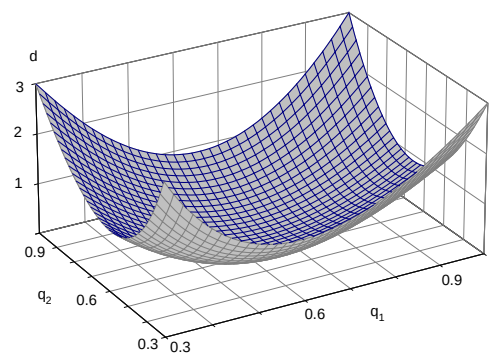
(a) Unrestricted Region: $r = -0.2$



(b) Unrestricted Region: $r = 0.1$



(c) Restricted Region ($c_1 = c_2 = 0.3$): $r = -0.5$



(d) Restricted Region ($c_1 = c_2 = 0.3$): $r = 0.5$

Figure 3.4: D_s-optimality of Designs for Two-toxicant Interaction Model. The function $d(\mathbf{x}, \eta)$ (1) is less than or equal to 3, the dimension of $\boldsymbol{\beta}_2 = (\beta_1, \beta_2, \beta_3)'$, over the design space; (2) achieves its maximum at the design points, confirming D_s-optimality of the designs..

Actually, even the determinant of the information matrix itself is hard to extract for multi-toxicant model because of the high dimension.

For multi-toxicant interaction model, fortunately, the equivalence theory provides an alternative solution to the problem. From the structure of D-optimal design for the two-toxicant interaction model we may conjecture the structure of D-optimal design for the multi-toxicant interaction model. Then we can employ the general equivalence theory to verify whether the conjecture is right or not. We shall illustrate this idea with the three-toxicant interaction model.

3.5.1 Three-toxicant Interaction Model

we know that the three-toxicant interaction model is the simplest multi-toxicant model. The mean function for this model is:

$$\lambda_i = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_{12} x_{1i} x_{2i} + \beta_{13} x_{1i} x_{3i} + \beta_{23} x_{2i} x_{3i}). \tag{3.4}$$

The above model says that each of the three toxicants, when applied alone, works through a first-order model as described in (1.9). However, when working together, they may interact with each other. The three two-way interactions are accounted for by β_{12}, β_{12} and β_{23}. The three-way interaction is assumed to be negligible.

We now try to conjecture the D-optimal design for model (3.4) on the unrestricted design space. Since most D-optimal designs are saturated designs, we would naturally conjecture that the D-optimal design for model (3.4) is a seven-point design. Recall that, for the two-toxicant interaction model, the D-optimal design consists of the control point, two pure component points located at $\text{IED}_{13.53}$ and an interaction point, whose location depends on $r = \beta_{12}/\beta_1\beta_2$. Recall that the intercept term β_0 is estimated solely from the control point; two main effects β_1 and β_2 are estimated from the difference between the control point and the corresponding pure component point and all the four points contribute to estimation of β_{12}, the interaction between the two toxicants. We would imagine that the D-optimal design for model (3.4) takes similar structure: the control point estimating β_0, three pure component points located at $\text{IED}_{13.53}$ estimating the three main effects β_1, β_2 and β_3, and three binary blend points estimating the three two-way interactions β_{12}, β_{13} and β_{23}. More specifically, we would expect that locations of the three binary blend points depend on the parameters only through $r_{12} = \beta_{12}/\beta_1\beta_2$, $r_{13} = \beta_{13}/\beta_1\beta_3$ and $r_{23} = \beta_{23}/\beta_2\beta_3$, respectively. Figure 3.5 geometrically illustrates the imaginary structure of the design. In Figure 3.5, point 1 is the control point; points 2, 3 and 4 are the pure component points located at $\text{IED}_{13.53}$ and points 5, 6 and 7 are the binary blend points. We would expect that the location of point 5 only depends on $r_{12} = \beta_{12}/\beta_1\beta_2$; point 6 only depends on $r_{13} = \beta_{13}/\beta_1\beta_3$ and point 7 only depends on $r_{23} = \beta_{23}/\beta_2\beta_3$. Further, we expect that the way a binary blend point

depends on the corresponding ratio is the same as how the interaction point depends on the ratio in a two-toxicant interaction model. For example, if $r_{13} = \beta_{13}/\beta_1\beta_3 = -1$, we would expect the location of point 6 is $(q_{16}, q_{26}, q_{36}) = (0.3679, 1, 0.3679)$ because the location of the interaction point in the two-toxicant interaction model is $(q_1, q_2) = (0.3679, 0.3679)$ for $r = -1$. Basically, this means that, in Figure 2.1, points 1, 2, 4 and 6 form the optimal design for model $\lambda = exp(\beta_0 + \beta_1 x_1 + \beta_3 x_3 + \beta_{13} x_1 x_3)$. Similarly, points 1, 3, 4 and 7 form the D-optimal design for the interaction model that only involves the second and third toxicants. To further illustrate the conjecture, Table 3.1 lists the D-optimal designs for two cases on the unrestricted design space. Note the locations of the binary blend points are the same as the corresponding interaction point listed in Table 2.4.

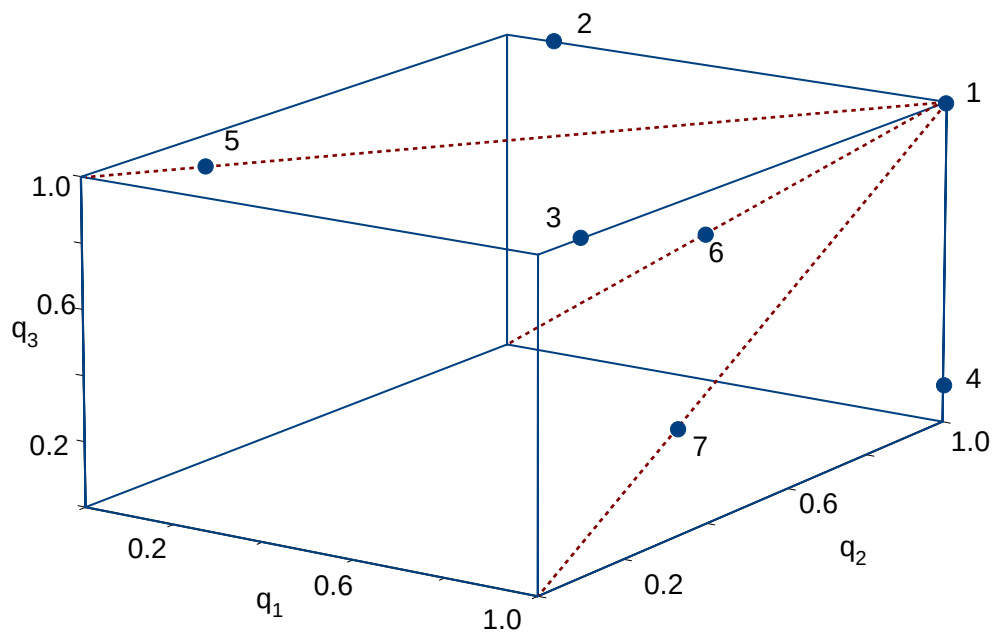


Figure 3.5: Structure of D-optimal Design for Three-toxicant Interaction Model

Of course, we need to verify the above conjectures via equivalence theory. We have evaluated the trace of $[J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})]$ for the two designs listed in Table 3.1 over the design space. We have found that this trace function achieves the maximum of 7 at the design points, confirming that the above conjecture is right. We also use this procedure to investigate optimal designs for some other three-toxicant interaction models and some four-toxicant interaction models. They all confirm that the above conjecture is right.

Table 3.1: D-optimal Designs for Three-toxicant Interaction Model, $p_1 = \cdots = p_7 = 1/7$.

	$r_{12} = -1, r_{13} = -0.1, r_{23} = -1$			$r_{12} = 0.2, r_{13} = -0.5, r_{23} = 0$		
point	q_1	q_2	q_3	q_1	q_2	q_3
1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	0.1353	1.0000	1.0000	0.1353	1.0000	1.0000
3	1.000	0.1353	1.0000	1.0000	0.1353	1.0000
4	1.0000	1.0000	0.1353	1.0000	1.0000	0.1353
5	0.3679	0.3679	1.0000	0.0100	0.0100	1.0000
6	0.1812	1.0000	0.1812	0.2905	1.0000	0.2905
1	1.0000	0.3679	0.3679	1.0000	0.1353	0.1353

3.5.2 Multi-toxicant Interaction Model

The above results can be easily generalized to the k-toxicant case, where k is any integer larger than 3. Again, we only consider the k-toxicant interaction model, which only includes the k main effects and two-way interactions. For such a model, the number of the parameters is $p = 2^k - 1$.

For the k-toxicant interaction model, we would expect that the D-optimal design is a saturated design including $p = 2^k - 1$ design points and the design is an equal allocation design, i.e., each point receiving $1/p$ of the total experimental runs. Further, the design includes the control point for estimation of the intercept, k pure component points located at $\text{IED}_{13.53}$ for estimation of the k main effects, and $\binom{k}{2} = k(k-1)/2$ binary blend points for estimation of all the two-way interactions. The location of a binary blend point depends on the corresponding ratio of the interaction term to the product of main effects, in the same way as the interaction point depends on r for the two-toxicant interaction model. In other words, for any two of these k toxicants, the control point and their pure component and interaction points form the D-optimal design for the interaction model of these two toxicants.

3.6 Summary

In this chapter, with equivalence theory, we have verified optimality of the locally D-optimal and D_s-optimal designs found in Chapter 2 and discussed D-optimal designs for the multi-toxicant interaction model. We now conclude this chapter by giving a summary:

1. In this chapter, we have employed Silvey's general equivalence theory to verify optimality of the locally D-optimal and D_s-optimal designs discussed in Chapter 2. By

evaluating and plotting the trace of $J(\mathbf{x}, \boldsymbol{\beta})M^{-1}(\eta, \boldsymbol{\beta})$ over the design space, we have geometrically demonstrated that the *Fréchet derivative* is less than or equal to zero for all the locally D-optimal designs in Chapter 2. According to the general equivalence theory, all these designs are D-optimal. Similarly, plots of the function $d(\mathbf{x}, \eta)$ confirm $\mathrm{D_s}$-optimality of the designs discussed in Section 2.4.

2. For the multi-toxicant model, we have trouble finding the optimal design by numerically maximizing the determinant of the Fisher information matrix as discussed in Chapter 2. The reason is twofold: (1) it is very hard to extract the determinant of the Fisher information matrix because of high dimension of the matrix; (2) generally, numerical methods do not work well when the number of variables is large. For the multi-toxicant interaction model, we go around this difficulty with a two-step procedure: firstly, we conjecture the structure of the D-optimal design based on the structure of D-optimal design for the two-toxicant interaction model; and then we use the general equivalence theory to verify our conjecture. It has been found that D-optimal design for the multi-toxicant interaction model has similar structure as D-optimal design for the two-toxicant interaction model.

Chapter 2 and Chapter 3 have focused on locally D-optimal designs. As we have demonstrated in Chapter 2, locally D-optimal designs does not perform well when the initial guesses are poor. In practice, we need efficient designs that are robust to initial poor information. In next chapter, we will discuss this issue and investigate how sequential designs improve the robustness to parameter misspecification.

Chapter 4

Sequential Designs

4.1 Introduction

As in Chapter 3, we denote the information matrix arising from a single observation made at $\mathbf{x}$ by $J(\mathbf{x}, \boldsymbol{\beta})$. Then the information matrix from n independent observations made at $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is $I(X, \boldsymbol{\beta}) = \sum_{i=1}^{n} J(\mathbf{x}_i, \boldsymbol{\beta})$, where $X' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. Our objective is to choose X to maximize the determinant of $I(X, \boldsymbol{\beta})$ for the true parameter $\boldsymbol{\beta}$. The maximization usually depends on $\boldsymbol{\beta}$ and since $\boldsymbol{\beta}$ is generally unknown we cannot achieve this objective in practice.

Alternatively, if we have good initial estimates of $\boldsymbol{\beta}$ then locally D-optimal designs discussed in Chapter 2 can be used. However, sound information about the parameters is not always available. For instance, generally we do not know much about $\boldsymbol{\beta}$ when we are testing new toxicants. Unfortunately, locally optimal designs do not perform well if poor initial estimates of $\boldsymbol{\beta}$ are used. Or in other words, locally optimal designs are not robust to parameter misspecification. Basically, this is the major difficulty we have to face when dealing with nonlinear experiments.

In Chapter 1, we have already indicated several possible ways to circumvent this difficulty and in this chapter we will investigate one of them, namely, sequential designs for the Poisson regression model. Both fully sequential designs (one run at a time) and two-stage designs will be discussed for improvement of design robustness. For brevity, we only consider the two-toxicant interaction model in this chapter; however, the methodology can be applied to any other model.

4.2 Sequential Designs

In a sequential design, we start with some initial experiment and use the observations to estimate the unknown parameters $\boldsymbol{\beta}$. We then choose the next design point so that it gives the largest increase of information at this estimated value. After the observation is taken at that point, we update our estimates and find the next design point based on the updated estimates. This process is repeated until all the experimental resources are run out or some stopping condition is met.

In a locally D-optimal design, placement of all the design points depends solely upon some initial guess of the parameters. If this initial guess is unreliable, then locally D-optimal design will generally be far away from the D-optimal design. In a sequential design, after the initial experiment, we make better use of the experimental resources in the sense that the placement of a design point is made based on our up-to-date estimates other than some unreliable guess. Generally, we would expect the up-to-date estimates will get closer to the true parameters as more observations are made and eventually will converge to the true parameters after enough runs. Suppose that we have n experimental runs in total and after the first r runs our estimates converge to the true parameters, then we actually make nearly optimal use of the remaining $n - r$ runs as their placement will be close to the optimal placement. This better use of part of the experimental resources improves the efficiency of the whole experiment.

4.2.1 Algorithm

We now formally describe the procedure of sequential experimentation. In a sequential design, one can start with an arbitrary design such as the Bayesian design or the locally optimal design. In this chapter, we shall use the conditional D-optimal design for the two-toxicant interaction model, with one replicate at each of the four design points. Then the design is implemented based on some initial guess and observations are made.

After the initial experiment, steps to proceed are as follows:

1. After taking k observations, we estimate the parameters based on the k observations by $\hat{\boldsymbol{\beta}}_k$.
2. The $(k+1)^{th}$ point, $\mathbf{x}_{k+1}$, is found by maximizing $|I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)|$, i.e.,

$$\mathbf{x}_{k+1} = \operatorname{Arg}\max_{\mathbf{x}} |I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)|.$$

3. The $(k+1)^{th}$ observation is taken at $\mathbf{x}_{k+1}$.
4. Steps 1, 2 and 3 are repeated until all the n experimental runs have been used up or sufficient accuracy has been achieved.

Clearly, step 2 is the most important step in implementing the above algorithm. In this step, firstly we need to derive $|I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)|$ and then maximize it. It can be shown (see the Appendix A.2) that

$$|I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)| = \left(\frac{\lambda_c}{\hat{\beta}_{1k}\hat{\beta}_{2k}}\right)^4 |A_1 + A_2|, \tag{4.1}$$

where

$$A_1 = \begin{pmatrix} \sum_{i=1}^k d_i & \sum_{i=1}^k d_i a_i & \sum_{i=1}^k d_i b_i & \sum_{i=1}^k d_i a_i b_i \\ \sum_{i=1}^k d_i a_i & \sum_{i=1}^k d_i a_i^2 & \sum_{i=1}^k d_i a_i b_i & \sum_{i=1}^k d_i a_i^2 b_i \\ \sum_{i=1}^k d_i b_i & \sum_{i=1}^k d_i a_i b_i & \sum_{i=1}^k d_i b_i^2 & \sum_{i=1}^k d_i a_i b_i^2 \\ \sum_{i=1}^k d_i a_i b_i & \sum_{i=1}^k d_i a_i^2 b_i & \sum_{i=1}^k d_i a_i b_i^2 & \sum_{i=1}^k d_i a_i^2 b_i^2 \end{pmatrix},$$

and

$$A_2 = \begin{pmatrix} d & da & db & dab \\ da & da^2 & dab & da^2 b \\ db & dab & db^2 & dab^2 \\ dab & da^2 b & dab^2 & da^2 b^2 \end{pmatrix}.$$

In matrix A_1, $a_i = \hat{\beta}_{1k} x_{1i}, b_i = \hat{\beta}_{2k} x_{2i}$ and $d_i = exp(a_i + b_i + a_i b_i \hat{\beta}_{12k}/\hat{\beta}_{1k}\hat{\beta}_{2k})$; a, b and d in matrix A_2 are defined similarly. Note that, at each step, $\left(\frac{\lambda_c}{\hat{\beta}_{1k}\hat{\beta}_{2k}}\right)^4$ is a constant; thus, maximization of $|I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)|$ is equivalent to maximization of $|A_1 + A_2|$.

Numerical methods will be used to maximize $|A_1 + A_2|$ and find the design points. As before, for design space with linear restrictions (restrictions imposed on IED levels), we use Nelder-Mead simplex method for its simplicity. For design space with nonlinear restrictions (restrictions imposed on MED levels), Quasi-Newton method is used for its capability of nonlinear optimization.

Recall that our goal is to maximize the determinant of the Fisher information matrix. However, if we construct a design sequentially as above, the determinant we actually maximize in step (2) is not that of the Fisher information matrix. This is so because the observations are not independent of each other after the rth run. For example, the distribution of y_{r+1} depends on $\mathbf{x}_{r+1}$, which is determined by $y_1, \ldots, y_r$. Thus the $(i, j)^{th}$ component of $J(\mathbf{x}_{r+1}, \boldsymbol{\beta})$ is $E\left(-\partial^2 \log p(y_{r+1}|y_1, \ldots, y_r, \boldsymbol{\beta})/\partial\beta_i\partial\beta_j\right)$ and to get Fisher information matrix for the $(r+1)^{th}$ observation we should take the expectation of this over $y_1, \ldots, y_r$, which is, in general, extremely difficult. However, as far as design construction is concerned we are completely at liberty to proceed as proposed above (Ford and Silvey, 1980; Silvey, 1980). All that matters is whether or not the method leads to good designs. But it is necessary to bear this point in mind when we consider the problem of making repeated-sampling inferences based on the data obtained from a sequentially constructed design.

4.2.2 An example

To illustrate the algorithm discussed above, a simulated example of sequential design is given in this section.

In the example, we assume that the total of experimental runs is 48 and the true model is

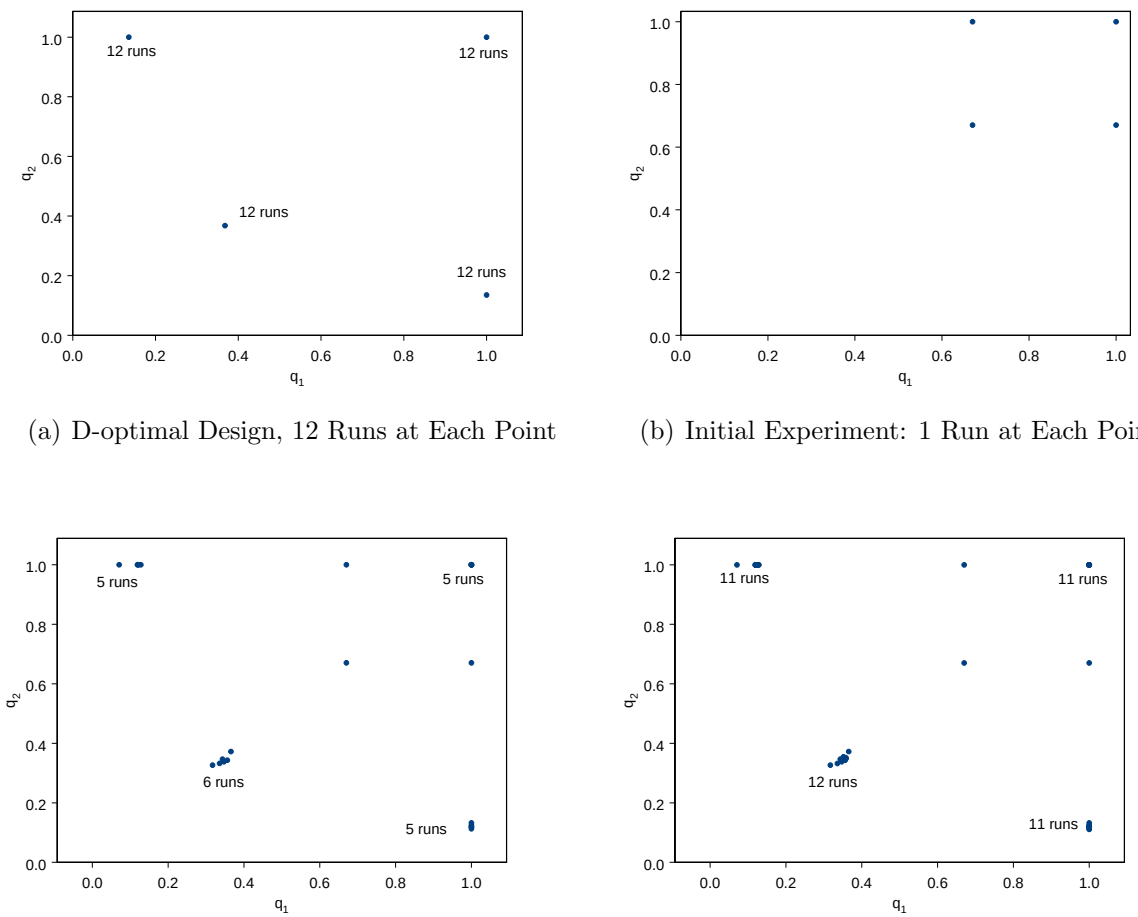
$$y_i \sim Poisson(\lambda_i) \text{ where } \lambda_i = exp(6 - x_{1i} - 0.4x_{2i} - 0.4x_{1i}x_{2i}).$$

We start with the conditional D-optimal design based on some initial parameter guess, $\mathbf{b} = (b_0, b_1, b_2, b_{12})'$. For the conditional D-optimal design, we have $b_{12} = 0$, i.e., no interaction. We shall use ratios of true parameters to the initial estimates to measure the extent of misspecifications of main effects, namely, $m_1 = \beta_1/b_1$ and $m_2 = \beta_2/b_2$ will be used to measure how severely β_1 and β_2 are misspecified. For simplicity, we assume both m_1 and m_2 are 0.2 in the example, i.e., both main effects are upwardly misspecified five times, which is quite severe. We use 4 runs in the initial experiment, one at each point. Then we simulate 4 observations based on the true model and the initial experiment. Using these 4 observations, we obtain estimates of the parameters as $\hat{\boldsymbol{\beta}}_4 = (6.009\ \text{-}1.139\ \text{-}0.392\ \text{-}0.262)'$. Maximizing $|I(X_4, \hat{\boldsymbol{\beta}}_4) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_4)|$ gives the location of the fifth run at $\mathbf{q}_5 = (q_{15}, q_{25}) = (0.2800, 0.2828)$. Based on $\hat{\boldsymbol{\beta}}_4$, the implemented fifth run is actually at $\mathbf{q}_5 = (q_{15}, q_{25}) = (0.3269, 0.2756)$. The fifth observation is then simulated and the estimates are updated so the sixth run can be located. This process is repeated until all 48 runs are exhausted. Table 4.1 lists some of the 48 runs and the up-to-date estimates.

Table 4.1: An Example of Sequential Design. The number of total runs is 48 and 4 runs are used in the initial experiment. The true model is $\lambda = 6 - x_1 - 0.4x_2 - 0.4x_1x_2$ and parameter misspecifications are specified by $m_1 = m_2 = 0.2$.

Run Number	q_1	q_2	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_{12}$
4	0.6703	0.6703	6.009	-1.139	-0.392	-0.262
5	0.3269	0.2756	5.996	-1.132	-0.353	-0.625
8	0.3968	0.4461	5.978	-0.848	-0.378	-0.532
9	0.0802	1.0000	5.985	-0.892	-0.379	-0.514
18	1.0000	0.1279	5.995	-0.939	-0.397	-0.443
19	0.3542	0.3756	5.995	-0.939	-0.397	-0.384
28	1.0000	1.0000	6.001	-0.960	-0.396	-0.437
29	0.1210	1.0000	6.011	-0.956	-0.396	-0.438
38	1.0000	0.1311	6.006	-0.957	-0.399	-0.436
39	0.3661	0.3818	6.005	-0.956	-0.399	-0.432
47	0.3702	0.3806	6.002	-0.977	-0.403	-0.423
48	1.0000	1.0000	6.001	-0.977	-0.403	-0.423

From Chapter 2, we know that the D-optimal design for the assumed true model is a saturated design with equal allocation. Thus, for our example, the D-optimal design allocates 12 runs to the control point, 12 runs to each of the two pure component points located at $\text{IED}_{13.53}$ and the remaining 12 runs to the interaction point, which is located at $(0.3679, 0.3679)$, as illustrated in Figure 4.1(a).



(a) D-optimal Design, 12 Runs at Each Point

(b) Initial Experiment: 1 Run at Each Point

(c) Sequential Design After 24 Runs

(d) Sequential Design After 48 Runs

Figure 4.1: An Example of Sequential Design. The conditional D-optimal design is used in the initial experiment. The number of total runs is 48 and 4 runs are used in the initial experiment. The true model is assumed to be $\lambda = exp(6 - x_1 - 0.4x_2 - 0.4x_1x_2)$ and parameter misspecification is specified by $m_1 = m_2 = 0.2$.

From Table 4.1, we can see that the design points in the sequential design get closer to those in the D-optimal design as more observations are made. Figure 4.1 illustrates the process of sequentially constructing the design. Figure 4.1(b) depicts the initial experiment, namely, the conditional D-optimal design based on $\mathbf{b}$. Due to parameter misspecification,

the conditional D-optimal design is far away from the D-optimal design and if we use up all 48 runs based on this design we would expect that the experiment is very inefficient. Figure 4.1(c) shows the distribution of experimental points after 24 runs and we can see four clusters are formed around the optimal design points. After 48 runs, the design is closer to the D-optimal design, as we can see in Figure 4.1(d). Thus, we would expect that the sequential design is very efficient since it is geometrically close to the D-optimal design. In next section, we will formally discuss evaluation of sequential designs.

4.2.3 Evaluation of Sequential Designs

The example in the last section shows that the sequential design is geometrically close to the D-optimal design. In this section, we will evaluate sequential designs and see if the method proposed in Section 4.2.1 works or not in general.

Recall that our goal is to maximize the determinant of the Fisher information matrix $I(\boldsymbol{\beta})$. So one logical way to evaluate a sequential design is to compare the determinant of the Fisher information matrix of the sequential design to that of the D-optimal design. Basically, we can define the efficiency of a sequential design as follows:

$$\text{Eff}_{seq} = \left(\frac{|I(X, \boldsymbol{\beta})| \text{ of the sequential design}}{|I(X, \boldsymbol{\beta})| \text{ of the D-optimal design}} \right)^{1/p}, \tag{4.2}$$

where p is the number of parameters.

It is straightforward to derive the Fisher information matrix $I(\boldsymbol{\beta})$ for the D-optimal design. However, it is much more complicated for a sequential design as the observations are not independent.

Suppose that we have n runs in total and we use r of them in the initial experiment. If we use $\mathbf{y}_{i-1} = (y_1, \ldots, y_{i-1})'$ to denote the up-to-date observation vector, the likelihood function of the n observations is

$$\begin{aligned} &L(\boldsymbol{\beta}; \mathbf{y}_n) = \\ &\quad L_1(\boldsymbol{\beta}; y_1) \cdots L_r(\boldsymbol{\beta}; y_r) L_{r+1}(\boldsymbol{\beta}; y_{r+1}|\mathbf{y}_r) \cdots L_n(\boldsymbol{\beta}; y_n|\mathbf{y}_{n-1}). \end{aligned} \tag{4.3}$$

Taking log of the likelihood function in (4.3) yields the log likelihood function in (4.4):

$$\ln(L) = \ln(L_1) + \cdots + \ln(L_r) + \ln(L_{r+1}) + \cdots + \ln(L_n). \tag{4.4}$$

When constructing sequential designs as discussed in Section 4.2.1, what we try to maximize is not the determinant of the Fisher information matrix but that of an observed information matrix as in (4.5), which basically is the negative of a partial expectation of the second derivative of (4.4):

$$I(\boldsymbol{\beta}|\mathbf{y}_{n-1}) = \\ J_1(\boldsymbol{\beta}) + \cdots + J_r(\boldsymbol{\beta}) + J_{r+1}(\boldsymbol{\beta}|\mathbf{y}_r) + \cdots + J_n(\boldsymbol{\beta}|\mathbf{y}_{n-1}). \tag{4.5}$$

In (4.5), $J_i(\boldsymbol{\beta}|\mathbf{y}_{i-1}) = -E_{y_i}\left(\partial^2 \ln L_i(\boldsymbol{\beta}; y_i|\mathbf{y}_{i-1})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\right)$ and E_{y_i} stands for taking the expectation over y_i, where $i = r+1, \ldots, n$.

To get the Fisher information matrix, we need to take the expectation of the observed information matrix in (4.5) with respect of $y_1, \ldots, y_{n-1}$, which is

$$\begin{aligned} I(\boldsymbol{\beta}) &= E_{\mathbf{y}_{n-1}}(I(\boldsymbol{\beta}|\mathbf{y}_{n-1})) = J_1(\boldsymbol{\beta}) + \cdots + J_r(\boldsymbol{\beta}) \\ &\quad + E_{\mathbf{y}_r}(J_{r+1}(\boldsymbol{\beta}|\mathbf{y}_r)) + \cdots + E_{\mathbf{y}_{n-1}}(J_n(\boldsymbol{\beta}|\mathbf{y}_{n-1})). \end{aligned} \tag{4.6}$$

The expectations in (4.6) are extremely hard, if possible, to compute. Alternatively, we could use simulation to estimate these expectations. Basically, for a given model, we can simulate many observation vectors, $\mathbf{y}_{n-1} = (y_1, \ldots, y_{n-1})'$, for each of which we can compute the observed information matrix as defined in (4.5). Given a sample of observed information matrices, we can average out the observation vector $\mathbf{y}_{n-1}$ and get an estimate of the Fisher information matrix. The idea here can be expressed in (4.7):

$$I(\boldsymbol{\beta}) = E_{\mathbf{y}_{n-1}}(I(\boldsymbol{\beta}|\mathbf{y}_{n-1})) \approx \frac{\sum_{k=1}^{K} I(\boldsymbol{\beta}|\mathbf{y}_{n-1}^k)}{K}. \tag{4.7}$$

We will make use of (4.7) to estimate the Fisher information matrix for sequential designs, based on which we can obtain the efficiency of a sequential design as defined in (4.2). Before carrying out the simulation, we need to clarify the following issues:

1. the true model and total runs,
2. simulation size (number of iterations, i.e., K in (4.7)),
3. parameter misspecification.

In what follows, we shall discuss the above issues.

True Model and Experiment Size

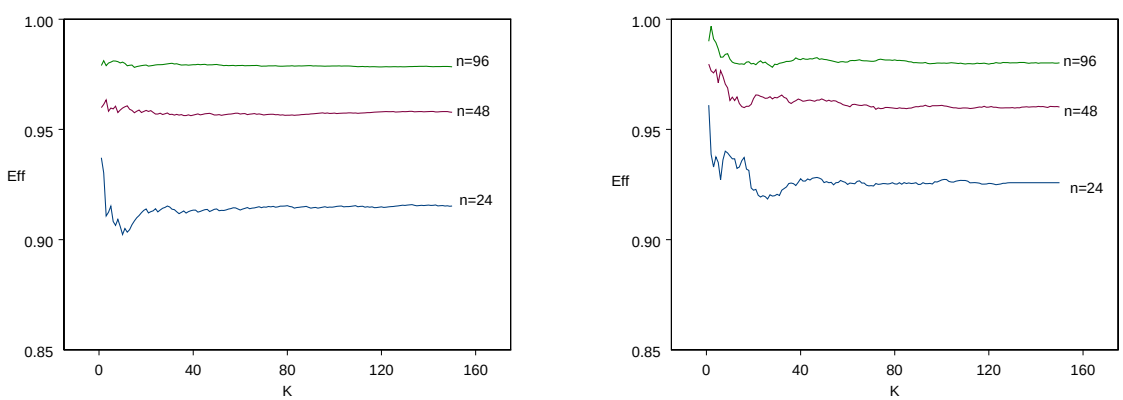
We do not intend to evaluate sequential designs for all models, which is impossible. Instead, we will demonstrate that sequential designs are preferable for some representative

two-toxicant interaction models. Particularly, we will try to cover models with different r values. We will consider four models with $r = -1, -0.5, 0$ and 0.2, respectively (see Table 4.2).

Intuitively, we would expect that sequential designs work better for larger experiments. In the simulation, we will investigate $n = 24, 48$ and 96 for each model considered as they can represent small, moderate and large experiment sizes. We purposely choose these multiples of 4 so that exact designs can possibly achieve optimality.

Simulation Size: K

Obviously, the larger K is, the better the approximation in (4.7) is. However, for large K, the simulation is very time-consuming as each iteration includes many complicated tasks such as model fitting and numerical optimization of the determinant of the information matrix. Thus, a reasonable K needs to be determined. A plot of the D-efficiency versus K, like Figure 4.2, can give us some idea of a reasonable K.



(a) $r = -0.5$ (b) $r = 0.2$

Figure 4.2: D-efficiency versus K. In (a), the true model is assume to be $\lambda = exp(6 - x_1 - 0.5x_2 - 0.25x_1x_2)$ and parameter misspecifications are specified by $m_1 = m_2 = 0.2$; in (b) the true model is assumed to be $\lambda = exp(5.5 - 0.5x_1 - 2x_2 + 0.2x_1x_2)$ and parameter misspecifications are given by $m_1 = 0.3, m_2 = 0.4$.

Figure 4.2 indicates that the estimate of the Fisher information matrix based on (4.7) converges after about 50 iterations for an experiment as small as 24 runs. For larger experiments, convergence takes place after fewer runs. Similar pattern have been found in plots for some other representative models with different parameter misspecifications. Therefore, large K such as 500 or 1000 is not necessary. In subsequent simulations, $K = 100$ is used unless otherwise stated.

Parameter Misspecification

Again, we cannot evaluate sequential designs for all possible situations of parameter misspecification. We will consider some representative cases that range from mild misspecification to severe misspecification. As before, we use $m_1 = \beta_1/b_1$ and $m_2 = \beta_2/b_2$ to measure the extent of misspecifications, where β_1 and β_2 are true parameters; b_1 and b_2 are corresponding initial guesses. We will consider four cases of parameter misspecifications: (1) $m_1 = m_2 = 2$; (2) $m_1 = m_2 = 1$; (3) $m_1 = 0.3$, $m_2 = 0.4$ and (4) $m_1 = m_2 = 0.2$. These four cases represent different severity of misspecifications.

Simulation Results

Table 4.2 gives the simulated D-efficiency for four different models with four different parameter misspecifications. Based on Table 4.2, we can make two comments about sequential designs as follows:

1. Sequential designs are very efficient and robust to parameter misspecification. As shown in Table 4.2, the D-efficiency is as high as 0.88 even for the worst case. In general, it works much better than the conditional D-optimal design under parameter misspecifications (see Section 2.3).

2. Though sequential designs work better for larger experiments, they are quite efficient even for an experiment as small as 24 runs. For larger experiments, the portion of experimental runs based on misspecified parameters is smaller so the design efficiency is larger. Table 4.2 shows that the design efficiency is at least 0.97 for experiments of 96 runs.

For the same parameter misspecification, the performance of the sequential design is model-dependent. Thus, there is no point comparing D-efficiency of sequential designs for different models in Table 4.2. To illustrate this point, we consider the first two models, for which $r = -1$ and -0.5, respectively. Though sequential designs are efficient for both models, there is slight difference between sequential designs for these two models in terms of D-efficiency. For example, when there is no misspecification of main effects ($m_1 = m_2 = 1$), sequential designs work a little better for model 2 than for model 1. Apart from randomness, this slight difference mainly results from the initial experiment. For the initial experiment, we use the conditional D-optimal design, which assumes that $r = 0$. For model 1, the interaction point of the D-optimal design is $(0.3679, 0.3679)$, and for model 2, this point is located at $(0.2905, 0.2905)$. When $m_1 = m_2 = 1$, the implemented interaction point of the initial experiment is $(0.1353, 0.1353)$, which is closer to the optimal interaction point for model 2 than for model 1. This slight difference explains the slightly higher D-efficiency of sequential designs for model 2. However, sequential designs perform a little better for model

Table 4.2: D-efficiency of Sequential Designs

$\lambda = exp(6.0 - 2x_1 - 0.2x_2 - 0.4x_1x_2)$, $r = -1$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.9161	0.9391	0.9537	0.9234
48	0.9583	0.9702	0.9771	0.9640
96	0.9790	0.9850	0.9883	0.9811
$\lambda = exp(6.0 - 0.4x_1 - x_2 - 0.2x_1x_2)$, $r = -0.5$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.9106	0.9624	0.9420	0.9193
48	0.9545	0.9813	0.9708	0.9592
96	0.9774	0.9908	0.9854	0.9791
$\lambda = exp(5.0 - x_1 - 1.5x_2 - 0x_1x_2)$, $r = 0$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.8870	0.9665	0.9035	0.8809
48	0.9406	0.9771	0.9444	0.9402
96	0.9666	0.9902	0.9699	0.9674
$\lambda = exp(5.0 - 2x_1 - 3x_2 + 1.2x_1x_2)$, $r = 0.2$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.9142	0.9350	0.8869	0.8793
48	0.9434	0.9614	0.9369	0.9286
96	0.9700	0.9767	0.9628	0.9604

1 when there are parameter misspecifications of main effects. This can also be explained by looking at the conditional D-optimal design in the initial experiment. When the main effects are downwardly misspecified ($m_1 = m_2 = 2$), the mean response at the interaction point of the initial experiment is close to zero so the simulated responses at that point are all zeros. Zero responses at the interaction point give a biased estimate of the interaction term β_{12} and thus a biased estimate of r. It happens that the bias is smaller for model 1 than for model 2 so sequential designs are slightly better for model 1. When $m_1 = 0.3$ and $m_2 = 0.4$, the interaction point of the initial experiment is located at $(0.5488, 0.4493)$ and it is closer to the interaction point of the D-optimal design for model 1 than for model 2. Thus, the D-efficiency is slightly smaller for model 2. For $m_1 = m_2 = 0.2$, we can explain the difference similarly.

4.2.4 Restricted Design Space

We have discussed sequential designs on the unrestricted design space. As discussed earlier, certain restrictions need to be imposed on the design space in many applications. In Section 2.3, we discussed a simple way to restrict the design space, namely, we set some lower

boundaries for the IED levels, q_1 and q_2. However, we also pointed out that restrictions of this type are not strong enough in practice as moderate IED levels could give very small MED level. In this section, we will investigate stronger restrictions and investigate how sequential designs perform under such restrictions.

In this section, we will restrict the design region by forcing the MED level, instead of IED levels, to be greater than some lower boundary. Restrictions of this type can avoid undesirable small proportions of survival, which is necessary in a lot of applications. However, under such restrictions, the design space depends on the unknown parameters $\boldsymbol{\beta}$. In what follows, we will discuss this dependence.

Suppose the lower boundary of MED levels is some desirable constant c, i.e.:

$$exp(\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2) \geq c.$$

Recall that $q_1 = exp(\beta_1 x_1)$, $q_2 = exp(\beta_2 x_2)$ and $r = \beta_{12}/\beta_1\beta_2$; the above inequality implies that

$$\ln(q_1) + \ln(q_2) + r \ln(q_1) \ln(q_2) \geq \ln(c). \tag{4.8}$$

Obviously, the design space defined by (4.8) depends on r. For illustration, the restricting curve in (4.8) is plotted for $c = 0.3$ and different values of r in Figure (4.3). In the figure, the areas above the restricting curves are the restricted design regions. We can see that larger r gives larger design space.

We have assumed that, at any point, increasing the dose or amount of either or both toxicants at least does not increase the mean response. This assumption implies that the MED level q is not greater than the IED levels q_1 and q_2 at any point. Therefore, if we restrict the MED level by (4.8), we will have $q_1 \geq c$ and $q_2 \geq c$.

Usually, we need to specify the design space before we design an experiment. Unfortunately, the above discussion reveals that the restrictions imposed on the design space depend on the unknown parameters. Thus, misspecification of parameters has impact on both the design and the design space. This implies that parameter misspecification becomes a more serious issue. In what follows, we will discuss how to use sequential designs to deal with this issue.

Locally D-optimal Designs

First of all, we shall study the locally D-optimal design as it provides a benchmark with which to calibrate any other design. Obviously, the locally D-optimal design depends on r in the sense that both the design and the design space depend on r. Lemma 2.2 still applies here. Specifically, for a given r, we firstly specify the design space by using (4.8). Then, as

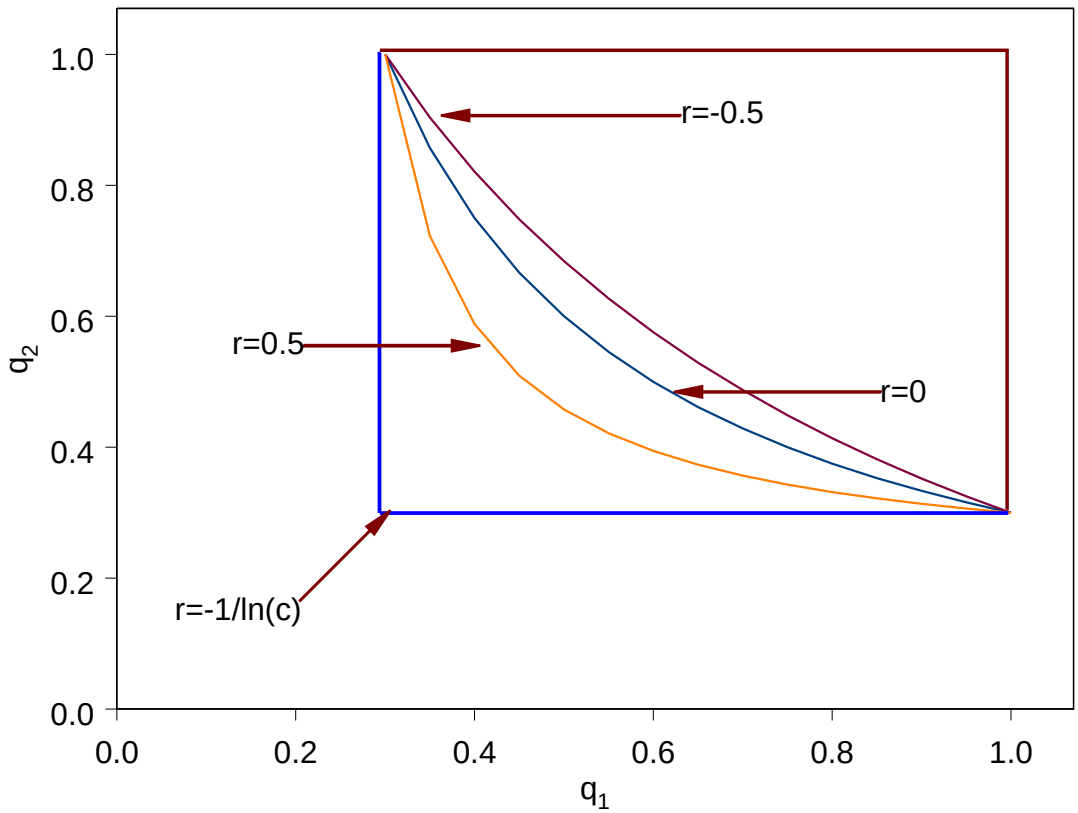


Figure 4.3: Restricted Design Regions for Different r, $c = 0.3$

before, we use some numerical methods to find the design under the nonlinear restriction. The numerical method adopted here is Quasi-Newton method for its capability of nonlinear optimization.

Without loss of generality, we shall assume that $c = 0.3$ in this section. The upper boundary defined in (2.1) also applies here. Actually, when r reaches the upper boundary, $-1/\ln(c)$, the design space is just the rectangular region defined by $q_1 \geq c$ and $q_2 \geq c$. Theoretically, r can take any value less than $-1/\ln(c)$. However, we will mainly focus on $-1 \leq r \leq -1/\ln(c)$ as this range is very common in practice. For some typical values of r between -1 and $-1/\ln(0.3)$, Table 4.3 lists the corresponding D-optimal designs on the restricted region defined by $\ln(q_1) + \ln(q_2) + r\ln(q_1)\ln(q_2) \geq \ln(0.3)$.

Like D-optimal designs on the unrestricted design space discussed in Chapter 2, the designs in Table 4.3 are also four-point designs with equal allocation, i.e., $p_1 = p_2 = p_3 = p_4 = 1/4$. These four points are the control point; two pure component points located

Table 4.3: D-optimal Designs on Restricted Region, $q \geq 0.3$, $p_1 = p_2 = p_3 = p_4 = 1/4$

r	Designs			
$\beta_{12}/\beta_1\beta_2$	$q_{11} = q_{21}$	$q_{12} = q_{23}$	$q_{13} = q_{22}$	$q_{14} = q_{24}$
-1	1	0.3	1	0.6160
-0.75	1	0.3	1	0.6029
-0.5	1	0.3	1	0.5878
-0.25	1	0.3	1	0.5698
-0.2	1	0.3	1	0.5658
-0.1	1	0.3	1	0.5572
0	1	0.3	1	0.5477
0.1	1	0.3	1	0.5373
0.2	1	0.3	1	0.5255
0.25	1	0.3	1	0.5191
0.5	1	0.3	1	0.4780
0.75	1	0.3	1	0.3993
Ubound	1	0.3	1	0.3000

at $(0.3, 1)$ and $(1, 0.3)$; and an interaction point whose location depends on r. D-optimal designs for other values of c have the same structure. More specifically, for any $c \geq 0.1353$, the two pure component points are located at $(c, 1)$ and $(1, c)$; and the interaction point is nothing but the intersection of the line $q_1 = q_2$ and the restricting curve $\ln(q_1) + \ln(q_2) + r\ln(q_1)\ln(q_2) \geq \ln(c)$. It can be shown that this intersection point is located at $(exp(-1/r + \sqrt{1 + r\ln(c)}/r), exp(-1/r + \sqrt{1 + r\ln(c)}/r))$.

Conditional D-optimal Designs

Both the design and design space being dependent on the unknown ratio r makes the locally D-optimal design in Table 4.3 impractical. As before, the conditional D-optimal design, which assumes that $r = 0$, might be a practical choice. However, we would not expect that the conditional D-optimal design performs well due to three facts: (1) misspecification of r gives wrong location of the interaction point; (2) misspecification of main effects is still a problem as before; (3) misspecification of r results in misspecification of design space, which means some points might situate out of the desirable restricted design space. Therefore, the implemented conditional D-optimal design may be far away from the D-optimal design so we need a practical design better than the conditional D-optimal design. Next, we will discuss the sequential design under such restrictions and see if it provides a better solution.

Sequential Designs

In this part, we will investigate how sequential designs perform on the restricted design space. After each experimental run in a sequential design, we use the up-to-date estimates to redefine the design space as well as to locate the next design point. As the up-to-date estimates get close the true parameters, both the design space and the design point should be close to their counterparts under no misspecification. Intuitively, we would expect that the sequential design is close to the optimal design in general, especially for experiments of large size.

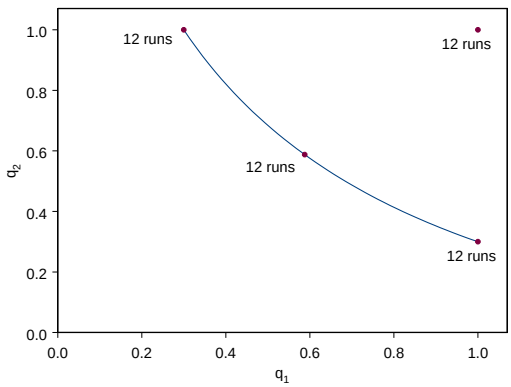
The algorithm described in Section 4.2.1 can be used here with one adjustment: in step (2), the maximization of $|I(X_k, \hat{\boldsymbol{\beta}}_k) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_k)|$ should be subject to the nonlinear constraint defined in (4.8) with $\hat{r}_k$ substituted for r, where $\hat{r}_k$ is the estimate of r after making k observations.

To illustrate the procedure, a simulated example is given next. In the example, we assume that the true model is $\lambda = exp(5.5 - 0.5x_1 - 2x_2 - 0.5x_1x_2)$ and the design space is restricted by $\ln(q_1)+\ln(q_2)+r\ln(q_1)\ln(q_2) \geq \ln(0.3)$. The initial parameter misspecifications are specified by $m_1 = m_2 = 2$ and the total of experimental runs is assumed to be 48.

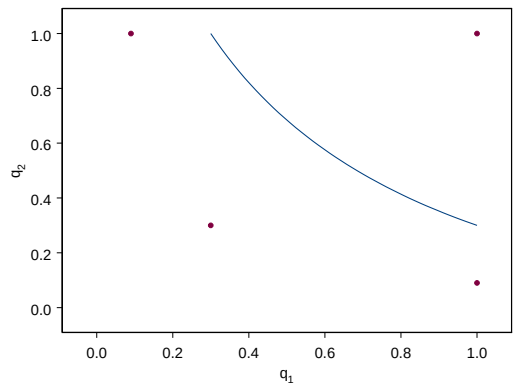
For the assumed true model, $r = -0.5$. From Table 4.3, we can see that the D-optimal design consists of four points, which are the control point, pure component points $(0.3, 1)$ and $(1, 0.3)$, and the interaction point $(0.5878, 0.5878)$, as illustrated in Figure 4.4(a). Note that the interaction point is the intersection of the line $q_1 = q_2$ and the restricting curve $\ln(q_1) + \ln(q_2) + r\ln(q_1)\ln(q_2) \geq \ln(0.3)$.

We start with the conditional D-optimal design with 4 runs as illustrated in Figure 4.4(b). Due to misspecification of main effects ($m_1 = m_2 = 2$), the four points are the control point, two pure component points $(0.09, 1)$ and $(1, 0.09)$, and the interaction point $(0.3, 0.3)$. We can see that three out of the four runs are out of the restricted region due to misspecification of main effects. Clearly, the conditional D-optimal design is not a good choice for this situation. We then simulate 4 responses from the true model for this initial experiment. Based on these responses, we can obtain the estimates of parameters as $\hat{\boldsymbol{\beta}}_4 = (5.4424, -0.5544, -1.7236, -0.6019)'$ and $\hat{r}_4 = -0.6299$. We then use this estimate of r to redefine the design space by $\ln(q_1) + \ln(q_2) - 0.6299\ln(q_1)\ln(q_2) \geq \ln(0.3)$ and locate the fifth run by maximizing $|I(X_k, \hat{\boldsymbol{\beta}}_4) + J(\mathbf{x}, \hat{\boldsymbol{\beta}}_4)|$ over the redefined design space. The fifth run turns out to be $(q_{15}, q_{25}) = (1, 1)$, i.e., the control point. Implementing the fifth run based on $\hat{\boldsymbol{\beta}}_4$ gives one more response and then we can update the estimates. The above process is repeated until 48 runs are used up.

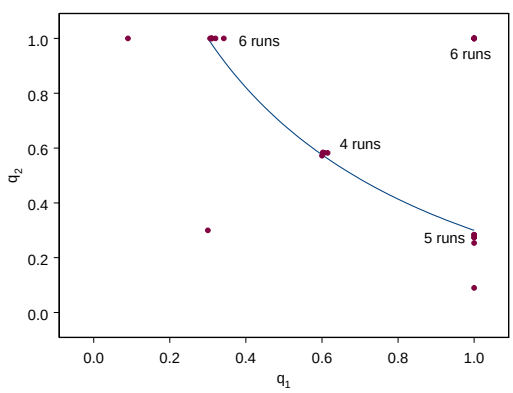
Figure 4.4 pictures the process of this simulated sequential experiment. Figure 4.4(a) plots the structure of the D-optimal design for the true model. The curve in the plot is the restricting curve. One can see that three out of the four points of the optimal design are located on the curve, or the lower boundary of the design space. Any efficient design should take similar structure. Figure 4.4(b) is the initial experiment, i.e., the conditional D-optimal



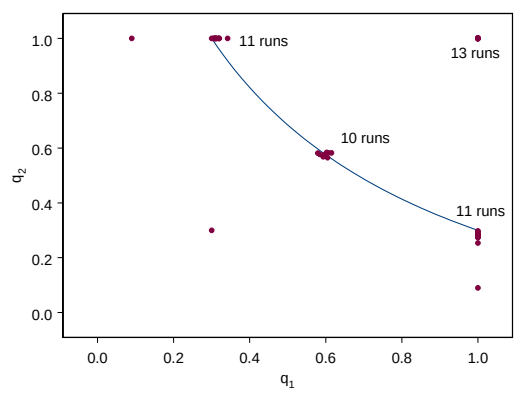
(a) D-optimal Design, 12 Runs at Each Point



(b) Initial Experiment: 1 Run at Each Point



(c) Sequential Design After 24 Runs



(d) Sequential Design After 48 Runs

Figure 4.4: An Example of Sequential Design on Restricted Region. The conditional D-optimal design is used in the initial experiment. The total runs are 48 and 4 runs are used in the initial experiment. The true model is assumed to be $\lambda = exp(5.5 - 0.5x_1 - 2x_2 - 0.5x_1x_2)$ and parameter misspecifications are specified by $m_1 = m_2 = 2$)

design with one run at each of the four points. We can see that three out of the four runs are out of the restricted design space. Figure 4.4(c) depicts the structure of the sequential experiment after 24 runs. One can see that all the experimental runs are clustered around the optimal points except the initial four runs. After 48 runs, the sequential experiment forms four clusters, corresponding to the four points in the D-optimal design. Each of the four clusters has about 1/4 of the total runs. This is also close to the D-optimal design, which is an equal-allocation design.

From the geometric closeness, we would expect that the sequential design performs well relative to the D-optimal design on the restricted design space. Of course, this is just a particular example of sequential designs and we will discuss evaluation of restricted sequential designs in general next.

Evaluation

In the last part, we gave a simple example of sequential designs. Figure 4.4 geometrically illustrates that it is quite close to the D-optimal design. In this part, we will discuss the evaluation of sequential designs on restricted regions and see how this method performs in general.

Up to this point, D-efficiency has been used to evaluate a design of interest. For instance, we used the D-efficiency to evaluate the conditional D-optimal design in Chapter 2; we also computed D-efficiency defined in (4.2) to see how sequential designs perform on unrestricted regions. Basically, D-efficiency compares a particular design with the D-optimal design in terms of the determinant of the Fisher information matrix. As discussed in Chapter 1, D-efficiency has a practical meaning (it gives the relative sample size of a particular design to achieve the same precision as the D-optimal design) and it is a very useful measure.

A basic assumption of D-efficiency is that the design of interest and the D-optimal design have the same design space. However, for restricted design regions, we have seen that some points of the conditional D-optimal design or the sequential design are out of the restricted region due to parameter misspecification, as illustrated in Figure 4.4. Since design regions are not the same, it does not make sense to compare two designs in terms of D-efficiency. Besides, D-efficiency, which addresses parameter estimation, is not the only concern here. Recall that we impose some constraints on the design space because of physical, biological or economical reasons. Experimentation out of a particular region is not desirable or even not possible. Therefore, for experimentation on restricted regions, a design should first follow the restrictions and then, under this condition, high efficiency is pursued. A design out of the region is not desirable even it is efficient. Take the conditional D-optimal design in Figure 4.4(b) for example. If we use up all the 48 runs based on this design, it can be computed that the D-efficiency of this design is 1.53. It is more efficient than the D-optimal design because it is out of the region in which the D-optimal design is restricted to reside. Recall that, due to parameter misspecification, the four points of the conditional

D-optimal design are $(1, 1)$, $(0.09, 1)$, $(1, 0.09)$, and $(0.3, 0.3)$. Remember that in this example we restrict the proportion of survival should at least be 0.3 at any point; however, the actual MED levels at point 2, 3 and 4 are 0.09, 0.09, and 0.04, respectively. Such low proportions of survival are not desirable or even unachievable and should be avoided.

Clearly, D-efficiency is not an appropriate tool for evaluation of a design on the restricted region as it does not penalize designs out of the specified space. We need some new measure. Basically, any desirable measure should praise designs close to the D-optimal design and penalize those far away. The failure of D-efficiency here is due to the fact that it gives a high score to the design that is out of the restricted region and far away from the D-optimal design.

If we still want to use D-efficiency as an evaluation measure, adjustment must be made. Basically, certain penalty must be imposed on the design points out of the restricted region. Clearly, when penalizing the outside points, we should follow the rule that a point that is far away from the restricted region should receive more severe penalty than a point close to the restricted region. Of course, there are many different types of penalty mechanisms and the choice is really case-dependent. For example, if the consequence of an outside experimental run is serious, a severe penalty rule should be enforced; on the other hand, if experimentation outside the restricted region is tolerable, though undesirable, a minor penalty should be given. In what follows, we will discuss a simple type of penalty mechanism and the resulting penalized D-efficiency.

We know that an outside point wrongly inflating D-efficiency should be penalized when computing D-efficiency. One way to correct the wrong inflation is that we can map the outside point to an inside point and use this "mirror" point, instead of the outside point itself, to compute D-efficiency. We will first introduce a simple mapping and discuss its justification afterwards.

Note that we can classify all the design points into four categories: control point ($q_1 = q_2 = 1$, or equivalently $x_1 = x_2 = 0$), pure component point I ($q_1 < 1$ and $q_2 = 1$, or equivalently $x_1 > 0$ and $x_2 = 0$), pure component point II ($q_1 = 1$ and $q_2 < 1$, or equivalently $x_1 = 0$ and $x_2 > 0$) and interaction point ($q_1 < 1$ and $q_2 < 1$, or equivalently $x_1 > 0$ and $x_2 > 0$). Suppose that an outside point is (x_1', x_2') and the point of the same category in the D-optimal design is (x_1, x_2). Its mirror point is given as (x_1'', x_2''), where $x_1'' = 2x_1 - x_1'$ and $x_2'' = 2x_2 - x_2'$. Equivalently, if we denote the outside point and its corresponding optimal point by (q_1', q_2') and (q_1, q_2), respectively, then its mirror point is (q_1'', q_2''), where $q_1''/q_1 = q_1/q_1'$ and $q_2''/q_2 = q_2/q_2'$. For completeness, we define the mirror point of an inside point or a point on the boundary as itself. Suppose that the original design matrix is X and after mapping each point to its mirror point as discussed above, the design matrix is mapped to its mirror matrix $\mathcal{X}$. We then derive the Fisher information matrix for the design X based on its mirror matrix $\mathcal{X}$. The penalized D-efficiency for a particular sequential design on the restricted region is defined in terms of its mirror design matrix as follows:

$$\text{Eff}_{penalized} = \left(\frac{|I(\mathcal{X}, \boldsymbol{\beta})| \text{ of the sequential design}}{|I(X, \boldsymbol{\beta})| \text{ of the D-optimal design}} \right)^{1/p}, \tag{4.9}$$

where p is the number of parameters.

We now consider the rationale behind the above mapping and the penalized D-efficiency. For any design, the control point is not affected by misspecification and it's always inside the restricted region. The two pure component points and the interaction point are possibly out of the desirable region due to misspecification. Note that the pure component points and the interaction point of the D-optimal design are located on the restricting curve as defined in (4.8). An inside pure component or interaction point contributes less to D-efficiency of the design than its optimal counterpart and, generally, its contribution becomes smaller as its distance from its corresponding optimal point increases. On the other hand, an outside point generally contribute more to D-efficiency than its corresponding optimal point but it is not desirable as it does not follow the restriction. In general, the further an outside point is from the its optimal counterpart, the more it will cost to run an experiment at this point. Thus a further point should be punished more severely. The above mapping, which maps an outside point to its inside mirror point, is able to provide such a mechanism of penalty. It is easy to see that an outside point close to the boundary will be mapped to an inside point that is also close to the boundary. Therefore, the penalty is minor since such an inside point is close to the optimal point in terms of D-efficiency. A further outside point will be mapped to an inside point further from the boundary and thus be severely penalized as a further inside point contributes less to D-efficiency.

We shall use the penalized D-efficiency in (4.9) to evaluate sequential designs on the restricted region. As discussed in Section 4.2.3, the Fisher information matrix of a sequential design is intractable and simulation will be used to estimate the Fisher information matrix. We will evaluate three true models with $r = -0.5, 0$ and 0.5, respectively. Table 4.4 lists the penalized D-efficiency of sequential designs under three different misspecifications: no misspecification ($m_1 = m_2 = 1$), mild misspecification ($m_1 = m_2 = 2$) and severe misspecification ($m_1 = m_2 = 0.2$).

We shall give some comments on the penalized D-efficiency and sequential designs based on the simulation results in Table 4.4:

1. Firstly, all the penalized efficiencies are between 0 and 1 as desired. The best case occurs when there is no misspecification ($r = 1, m_1 = m_2 = 0$), which, of course, deserves the highest score. The worst two cases are ($r = -0.5, m_1 = m_2 = 2$) and ($r = 0.5, m_1 = m_2 = 0.2$). This can be explained by inspecting the initial experiments for these two cases. Recall that the initial experiment is the conditional D-optimal design, which assumes that $r = 0$. Figure 4.3 shows that the lower boundary of the design space for $r = 0$ is outside of the boundary for $r = -0.5$. Thus the design space is misspecified. Downward misspecification of main effect ($m_1 = m_2 = 2$) makes the

Table 4.4: Penalized D-efficiency of Sequential Design and Conditional D-optimal Design on Restricted Region, $m_1 = m_2 = m$.

$\lambda = exp(6.0 - 0.5x_1 - 2.0x_2 - 0.5x_1x_2)$, $r = -0.5$						
	Sequential Design			Conditional Optimal Design		
n	$m = 2.0$	$m = 1.0$	$m = 0.2$	$m = 2.0$	$m = 1.0$	$m = 0.2$
24	0.8539	0.9477	0.8651			
48	0.9304	0.9626	0.9242	0.0000	0.9061	0.0932
96	0.9623	0.9703	0.9530			
$\lambda = exp(6.0 - x_1 - 2x_2 - 0x_1x_2)$, $r = 0$						
	Sequential Design			Conditional Optimal Design		
n	$m = 2.0$	$m = 1.0$	$m = 0.2$	$m = 2.0$	$m = 1.0$	$m = 0.2$
24	0.8631	0.9574	0.8523			
48	0.9486	0.9678	0.9177	0.0000	1.0000	0.0824
96	0.9712	0.9733	0.9503			
$\lambda = exp(6.5 - 2x_1 - 1.5x_2 + 1.5x_1x_2)$, $r = 0.5$						
	Sequential Design			Conditional Optimal Design		
n	$m = 2.0$	$m = 1.0$	$m = 0.2$	$m = 2.0$	$m = 1.0$	$m = 0.2$
24	0.8662	0.9330	0.8406			
48	0.9394	0.9546	0.9145	0.0000	0.8534	0.0673
96	0.9646	0.9676	0.9483			

situation even worse in that the three runs on the boundary for $r = 0$ are now outside of the already-misspecified design space due to misspecification of main effects. Therefore, these runs in the initial experiment, perhaps some runs after the initial experiment, are severely punished for being outside of the restricted region and far away from the boundary. This severe penalty causes low design efficiency. Similarly, we can explain the low efficiency for $(r = 0.5, m_1 = m_2 = 0.2)$, where the initial experimental runs are inside the restricted region but too far away from the boundary. This justifies, at least partially, the penalized D-efficiency as a rational measure.

2. Table 4.4 indicates that, in terms of the penalized D-efficiency, the sequential design works quite well on restricted regions. For experiments of 48 runs, the efficiency is always above 90% and it gets better for larger experiments.

3. As mentioned before, there is no point comparing design efficiency of sequential designs across different models because, for the same misspecification, performance of the sequential design is model-dependent. Take the first two models in Table 4.4 for example. For $m_1 = m_2 = 2$ and $m_1 = m_2 = 1$, sequential designs for the second model $(r = 0)$ are more efficient because there is no misspecification of design space under this model. However, for $m_1 = m_2 = 0.2$, the first model $(r = -0.5)$ has larger efficiency because upward misspecification of main effects somehow counteracts the effect

of misspecification of design space.

For comparison, Table 4.4 also gives the penalized D-efficiency of the conditional D-optimal design. Note that efficiency of the conditional D-optimal design is independent of the sample size n. Under no misspecification ($r = 0, m_1 = m_2 = 1$), the conditional D-optimal design, which happens to be the D-optimal design, beats the sequential design. However, when misspecification occurs, sequential designs are much more efficient. Particulary, for $m_1 = m_2 = 2$, three of the four points for the conditional D-optimal design are way out of the restricted region. The penalty is so severe that all the three points are mapped to the control point. Thus, the mapped design degenerates to a one-point design and the corresponding information matrix is singular, which explains the zero efficiency.

Recall that the above mapping scheme maps an outside point (x_1', x_2') to an inside point (x_1'', x_2''), where $x_1'' = x_1 + (x_1 - x_1')$, $x_2'' = x_2 + (x_2 - x_2')$ and (x_1, x_2) is the corresponding optimal point. This mapping scheme can be adjusted according to the nature of the restriction. If experimentation out of the restricted region is extremely costly and should be avoided whenever possible, an outside point should be more severely penalized. To this end, we can adjust the mapping scheme so that $x_1'' = x_1 + \delta(x_1 - x_1')$ and $x_2'' = x_2 + \delta(x_2 - x_2')$, where δ is some constant greater than 1. Obviously, larger δ implies more severe penalty. Choice of δ is arbitrary and depends on the cost of an outside experimental run. Note that the mapping scheme we discussed earlier simply chooses 1 for δ. On the other hand, if an outside run, though undesirable, does not incur serious consequence, we should give less severe penalty to an outside point so δ should be between 0 and 1. If we denote the outside point and its corresponding optimal point by (q_1', q_2') and (q_1, q_2), respectively, based on the adjusted mapping scheme, its mirror point is (q_1'', q_2''), where $q_1''/q_1 = (q_1/q_1')^\delta$ and $q_2''/q_2 = (q_2/q_2')^\delta$.

Up to this point, we have demonstrated that sequential designs are very efficient and robust to parameter misspecifications on both unrestricted and restricted regions. Next, we will discuss statistical inferences based on sequential designs.

4.2.5 Inference Based on Sequential Designs

In constructing sequential designs, we actually work with the observed information matrix rather than the Fisher information matrix (see Section 4.2.1). As we have stated earlier, this raises no question of principle; all that matters is whether the method is effective or not in practice. We have demonstrated that sequential experimentation is indeed effective. However, having constructed a design sequentially, we will usually wish to make inferences about model parameters $\boldsymbol{\beta}$, and questions arise about how the sequential construction affects inferences. In this section, we will discuss this issue.

Suppose that we construct a sequential design $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The likelihood function arising from this sequentially constructed design is given in (4.3). If the design had not

been constructed sequentially and we had merely taken independent observations at the predetermined design points $\mathbf{x}_1, \ldots, \mathbf{x}_n$, we can easily see that the likelihood function will stay the same. Therefore, if either a Bayesian or likelihood-based approach to inferences is adopted, we can just ignore the fact that the design is sequential and make inferences as if all the observations are independent of each other.

However, if a repeated-sampling approach is adopted, the situation is not so clear-cut. The repeated-sampling properties of sequential designs are quite different from those of predetermined designs; the design achieved varies from occasion to occasion in the former, but not in the latter. Still, when making inferences from a sequential design, one is tempted to ignore the sequential construction. Standard methods and software for analyzing predetermined designs are often used to make inferences from sequentially constructed designs. But is it a right or reasonable thing to do?

Suppose, in particular, that we wish to construct a confidence region for the model parameters $\boldsymbol{\beta}$ based on a sequentially constructed design $X' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$. As before, we denote the observed information matrix for the kth observation by $J(\mathbf{x}_k, \boldsymbol{\beta})$, the $(i,j)^{th}$ component of which is $E\left(-\partial^2 \log p(y_k|y_1, \ldots, y_{k-1}, \boldsymbol{\beta})/\partial\beta_i\partial\beta_j\right)$. The observed information matrix of all the n observations is $I(\boldsymbol{\beta}|\mathbf{y}_{n-1}) = \sum_{k=1}^{n} J(\mathbf{x}_k, \boldsymbol{\beta})$. For a pre-determined design with independent observations, the $(i,j)^{th}$ component of $J(\mathbf{x}_k, \boldsymbol{\beta})$ is simply $E\left(-\partial^2 \log p(y_k)/\partial\beta_i\partial\beta_j\right)$ and $I(\boldsymbol{\beta}|\mathbf{y}_{n-1}) = \sum_{k=1}^{n} J(\mathbf{x}_k, \boldsymbol{\beta})$ is the Fisher information matrix. Given sufficient regularity, standard large-sample theory applies and, for large n, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is approximately normally distributed with mean $\boldsymbol{\beta}$ and variance matrix $I^{-1}(\boldsymbol{\beta}|\mathbf{y}_{n-1})$. From this, confidence regions for $\boldsymbol{\beta}$ can be constructed as in (4.10). Note that the degree of freedom for χ^2 distribution in (4.10) is the dimension of $\boldsymbol{\beta}$.

$$\{\boldsymbol{\beta} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' I(\boldsymbol{\beta}|\mathbf{y}_{n-1})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \leq \chi^2_\alpha\}. \tag{4.10}$$

Of course, $\boldsymbol{\beta}$ is unknown in practice. so we replace $I(\boldsymbol{\beta}|\mathbf{y}_{n-1})$ with $I(\hat{\boldsymbol{\beta}}|\mathbf{y}_{n-1})$. We know (4.10) is the right way to construct the confidence region for independent observations because for independent observations $I(\boldsymbol{\beta}|\mathbf{y}_{n-1})$ is the Fisher information matrix. However, if the design $X' = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ is achieved by sequential construction, we know that $I(\boldsymbol{\beta}|\mathbf{y}_{n-1})$ is not the Fisher information matrix. But is it still an adequate approximation to the variance matrix of $\hat{\boldsymbol{\beta}}$? Should we use instead the inverse of the Fisher information matrix?

Unfortunately, there is no asymptotic theory in the literature to support the use of (4.10) for repeated-sampling inferences based on sequentially constructed designs. Empirical investigation of questions like those posed above therefore seems to be necessary. Ford and Silvey (1980) studied the sequential design for estimating a nonlinear function of parameters for a linear regression model. They looked at some of the above questions using a simulation study. For the one-variable logistic regression, Minkin (1987) empirically investigated the actual coverage of the confidence region defined in (4.10) for the two-stage design. Both studies reach the same conclusion that (4.10) works well for sequential designs. The extent

to which similar conclusions apply to other examples such as our situation is, of course, open to question. In the following, we will do some simulations to investigate if we can use (4.10) for sequential designs we have discussed.

For a given model and parameter misspecification, we can simulate a sequential experiment and estimate $\boldsymbol{\beta}$. Then $I(\hat{\boldsymbol{\beta}}|\mathbf{y}_{n-1})$ can be obtained and a $100(1-\alpha)\%$ confidence region defined in (4.10) is constructed. We then check if this region contains the true parameters. Suppose that we repeat this process K times and C out the K confidence regions cover the true parameters; then the ration, C/K, is an estimate of actual coverage if K is large enough. If this ratio is close to the nominal level, $100(1-\alpha)\%$, we can conclude that (4.10) works well for this situation.

We consider five models with different values of r. For each model and given misspecification of main effects, we simulate 500 sequential experiments, each with 48 runs. For each of these 500 experiments, we use the parameter estimates and observed information matrix to construct the confidence region based on (4.10). We then count how many of the 500 confidence regions cover the true parameters and then compute the actual coverage. On the other hand, based on the discussion in Section 4.2.3, for each model, we can use the average of the 500 observed information matrices to approximate the expected information matrix. We then use the parameter estimates and the approximated expected information matrix to construct confidence regions. These two types of confidence regions will be compared in terms of their actual coverage. Two nominal confidence levels will be considered: one is 90% and the other is 95%. Table 4.5 gives the simulation results.

Table 4.5: Coverage of Confidence Regions Based on Expected and Observed Information Matrices for Sequential Designs

True Model			Actual Coverage			
			90% Confidence Region		95% Confidence Region	
$\boldsymbol{\beta}'$	m_1	m_2	Expected[a]	Observed[b]	Expected	Observed
$(5.5, -2.0, -0.2, -0.4)$	2.0	2.0	0.904	0.908	0.940	0.940
$(5.0, -1.0, -1.0, -0.5)$	2.0	0.5	0.900	0.904	0.944	0.946
$(4.5, -0.2, -1.0, -0.2)$	0.5	0.5	0.901	0.903	0.949	0.951
$(5.0, -1.2, -2.5, -0.0)$	0.2	1.0	0.896	0.908	0.952	0.964
$(4.0, -3.0, -2.0, \quad 1.2)$	2.0	1.0	0.904	0.910	0.942	0.952

[a]actual coverage based on expected information matrix.
[b]actual coverage based on observed information matrix.

From Table 4.5, we can see that the confidence regions based on the observed information matrix have slightly higher coverage than those based on the expected information matrix. However, they are actually close to each other and both are close to nominal confidence levels. Thus, there is no compelling reason for using the expected information matrix to construct confidence regions. In other words, we can just ignore the fact that the design is

sequentially constructed and simply use the observed information matrix to make inferences. This simulation result is in accordance with the conclusion in Ford and Silvey (1980) and Minkin (1987).

4.3 Two-stage Designs

4.3.1 Two-stage Procedure

In the last section, we have seen that sequential designs are very efficient and robust to parameter misspecification. However, sequential experimentation has a major drawback: a sequential experiment is generally more costly than a non-sequential one. In general, sequential experimentation costs more time, especially when it takes a long time to obtain an observation. However, we have also seen that non-sequential designs in Chapter 2 can be very inefficient under parameter misspecification.

One way out of this dilemma is that we could carry out the experiment sequentially but in less stages. The sequential design in the last section is implemented in approximately n stages and at each stage there is only one experimental run. This type of sequential designs (one run at a time) is referred to as fully sequential designs or just sequential designs for brevity. We could reduce the number of stages by increasing the number of runs at each stage. Thus, the whole experiment, which has less stages, takes less time than a fully sequential experiment. Meanwhile, we can still update our knowledge about the unknown parameters after each stage and make better use of experimental resources at the next stage. Sequential designs of this type (a batch of runs at a time) are referred to as stage-wise designs.

In this section, we will discuss a special case of stage-wise designs, namely, two-stage designs. In a two-stage design, as the name indicates, we implement the whole experiment in just two stages. Design of the first stage is primarily used to obtain "good" parameter for use in the second stage. Usually, little knowledge is available at the first stage and we design and implement the experiment based on some initial guess. After the experiment of the first stage is carried out and the responses are observed, we can estimate the parameters and we then design the experiment for the second stage based on the estimates. We shall discuss the procedure of two-stage experimentation by starting with the likelihood function.

Suppose that we have n experimental runs in total and we use n_1 experimental runs at the first stage and the remaining $n_2 = n - n_1$ at the second stage. The likelihood function for the observations from the first stage is:

$$L_1(\boldsymbol{\beta}; \mathbf{y}_1) = \prod_{i=1}^{n_1} p(\mathbf{y}_1|\boldsymbol{\beta}) = \prod_{i=1}^{n_1} \frac{exp(-\lambda_{1i})\lambda_{1i}^{y_{1i}}}{y_{1i}!}, \tag{4.11}$$

where $\mathbf{y}_1$ is the observation vector from the first stage and $\lambda_{1i} = exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} +$

$\beta_{12}x_{1i}x_{2i}$) for $i = 1, 2, \ldots, n_1$.

Based on (4.11), we can derive the Fisher information matrix for the first stage, namely, $I_1(\boldsymbol{\beta}) = -E_{\mathbf{y}_1}\left(\partial^2 \ln L_1(\boldsymbol{\beta}; \mathbf{y}_1)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\right)$. Then we can maximize the determinant of this information matrix, based on some initial guesses of the parameters, to find the design for the first stage. In this section, as before, we assume little is known about the interaction term and the conditional D-optimal design will be used for the first stage. After the experiment of the first stage is implemented and observations are made, we can obtain estimates of the parameters.

The joint likelihood function of both stages is the product of the likelihood function for the first stage and the conditional likelihood function for the second stage given the first stage, as given in (4.12). Note that, in (4.12), $\mathbf{y}_1$ and $\mathbf{y}_2$ are the observation vectors of the first stage and second stage, respectively.

$$L_{1,2}(\boldsymbol{\beta}; \mathbf{y}_1, \mathbf{y}_2) = L_1(\boldsymbol{\beta}; \mathbf{y}_1) L_{2|1}(\boldsymbol{\beta}; \mathbf{y}_2|\mathbf{y}_1), \tag{4.12}$$

where

$$L_{2|1}(\boldsymbol{\beta}; \mathbf{y}_2|\mathbf{y}_1) = \prod_{i=1}^{n_2} \frac{exp(-\lambda_{2j})\lambda_{2j}^{y_{2j}}}{y_{2j}!} \tag{4.13}$$

is the conditional likelihood function in which $\lambda_{2j} = exp(\beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \beta_{12} x_{1j} x_{2j})$. The second stage is conditioned on the first stage in the sense that the parameter estimates from the first stage are used to compute the design levels in the second stage, i.e., the design matrix X_2 depends on the first stage.

The information matrix is the expectation of the second derivative of the log likelihood function for the data. Thus, taking log of the joint likelihood in (4.12) yields $\ln(L_{1,2}) = \ln(L_1) + \ln(L_{2|1})$, which further implies:

$$I_{1,2}(\boldsymbol{\beta}) = I_1(\boldsymbol{\beta}) + E_{\mathbf{y}_1}(I_{2|1}(\boldsymbol{\beta}|\mathbf{y}_1)), \tag{4.14}$$

where $I_{2|1}(\boldsymbol{\beta}) = -E_{\mathbf{y}_2}\left(\partial^2 \ln L_{2|1}(\boldsymbol{\beta}; \mathbf{y}_2|\mathbf{y}_1)/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'\right)$ is the observed information matrix for the second stage.

At the second stage, we should find a design that maximizes the determinant of the joint information matrix defined in (4.14) with $\boldsymbol{\beta}$ replaced by estimates from the first stage. However, the calculation of the expectation, $E_{\mathbf{y}_1}(I_{2|1}(\boldsymbol{\beta}|\mathbf{y}_1))$, is extremely difficult. Instead, we can find the design for the second stage by maximizing the determinant of the observed information matrix defined as follows:

$$I_{1,2}(\boldsymbol{\beta}|\mathbf{y}_1) = I_1(\boldsymbol{\beta}) + I_{2|1}(\boldsymbol{\beta}|\mathbf{y}_1). \tag{4.15}$$

Clearly, $I_{1,2}(\boldsymbol{\beta}|\mathbf{y}_1)$ defined in (4.15) is not the Fisher information matrix. As stated earlier in last section, this is of no account as far as design construction is concerned; all that matters is whether this method results in good designs.

Suppose the conditional D-optimal design is adopted in the first stage and a four-point design is used in the second stage. It can be shown (see Appendix A.2) that the determinant of the observed joint information matrix based on (4.15) is:

$$|I_{1,2}(\boldsymbol{\beta}|\mathbf{y}_1)| = \left(\frac{\lambda_c N}{\beta_1\beta_2}\right)^4 |A_1 + A_2|, \tag{4.16}$$

where A_1 is

$$\begin{pmatrix} \sum_{i=1}^{n_1} d_i & \sum_{i=1}^{n_1} d_i a_i & \sum_{i=1}^{n_1} d_i b_i & \sum_{i=1}^{n_1} d_i a_i b_i \\ \sum_{i=1}^{n_1} d_i a_i & \sum_{i=1}^{n_1} d_i a_i^2 & \sum_{i=1}^{n_1} d_i a_i b_i & \sum_{i=1}^{n_1} d_i a_i^2 b_i \\ \sum_{i=1}^{n_1} d_i b_i & \sum_{i=1}^{n_1} d_i a_i b_i & \sum_{i=1}^{n_1} d_i b_i^2 & \sum_{i=1}^{n_1} d_i a_i b_i^2 \\ \sum_{i=1}^{n_1} d_i a_i b_i & \sum_{i=1}^{n_1} d_i a_i^2 b_i & \sum_{i=1}^{n_1} d_i a_i b_i^2 & \sum_{i=1}^{n_1} d_i a_i^2 b_i^2 \end{pmatrix},$$

and A_2 is

$$\begin{pmatrix} \sum_{j=1}^{n_2} d_j & \sum_{j=1}^{n_2} d_j a_j & \sum_{j=1}^{n_2} d_j b_j & \sum_{j=1}^{n_2} d_j a_j b_j \\ \sum_{j=1}^{n_2} d_j a_j & \sum_{j=1}^{n_2} d_j a_j^2 & \sum_{j=1}^{n_2} d_j a_j b_j & \sum_{j=1}^{n_2} d_j a_j^2 b_j \\ \sum_{j=1}^{n_2} d_j b_j & \sum_{j=1}^{n_2} d_j a_j b_j & \sum_{j=1}^{n_2} d_j b_j^2 & \sum_{j=1}^{n_2} d_j a_j b_j^2 \\ \sum_{j=1}^{n_2} d_j a_j b_j & \sum_{j=1}^{n_2} d_j a_j^2 b_j & \sum_{j=1}^{n_2} d_j a_j b_j^2 & \sum_{j=1}^{n_2} d_j a_j^2 b_j^2 \end{pmatrix}.$$

In both matrices $a_k = \ln(q_{1k}), b_k = \ln(q_{2k})$ and $d_k = p_k q_{1k} q_{2k} exp(r a_k b_k)$.

From the above discussion, the procedure of the two-stage design can be outlined as follows:

1. Decide n_1 and n_2.
2. Design the experiment for the first stage. This is done by maximizing $|A_1|$ in which $\boldsymbol{\beta}$ is replaced by some initial guess. We will use the conditional D-optimal design for the first stage in this section.

3. Carry out the experiment, observe the responses and obtain estimates of the parameters, $\hat{\boldsymbol{\beta}}$.

4. Find the design for the second stage. This is done by maximizing $|A_1 + A_2|$, where $\boldsymbol{\beta}$ is replaced by $\hat{\boldsymbol{\beta}}$.

5. Carry out the experiment for the second stage, observe the responses and update estimates of the parameters.

4.3.2 An Example of Two-stage Designs

In this section, we shall give a simple example of two-stage designs to illustrate the two-stage procedure discussed in the last section.

In the example, we assume that the number of total experimental runs is 48 and the true model is $\lambda = exp(6 - x_1 - 2x_2 - 0.5x_1x_2)$. We know that for this model the D-optimal design is a saturated and equal-allocation design as listed in Table 4.6. Suppose that we equally allocate the total experimental runs to the two stages, i.e., $n_1 = n_2 = 24$. For the first stage, we use the conditional D-optimal design based on some initial parameter guess, $\mathbf{b} = (b_0, b_1, b_2, b_{12})'$. Of course, in the conditional D-optimal design, we have $b_{12} = 0$, i.e., no interaction. In this particular example, we assume that we have perfect knowledge about the main effects, i.e., $b_1 = \beta_1 = -1$ and $b_2 = \beta_2 = -2$. The design for the first stage, in terms of IED levels and proportions of total runs, is also listed in Table 4.6. We then simulate 24 observations for the first stage design from the true model. Based on these 24 observations, we obtain the estimates of parameters as $\hat{\boldsymbol{\beta}} = (6.0088, -1.0303, -1.9539, -0.6331)'$. Given the estimates of $\boldsymbol{\beta}$, we can then find the design for the second stage, which is also listed in Table 4.6. Then the design for the second stage is implemented based on $\hat{\boldsymbol{\beta}}$. The implemented points for the second-stage design are $(1, 1), (0.1441, 1), (1, 0.1286)$ and $(0.2553, 0.2361)$. Note that they are not the same as but quite close to the optimal points. Finally, the observations are made and the estimates are updated.

Table 4.6: An Example of Two-stage Design. There are 48 runs in total and 24 runs in each of the two stages. The true model is assumed to be $\lambda = exp(6 - x_1 - 2x_2 - 0.5x_1x_2)$ and parameter misspecifications are specified by $m_1 = m_2 = 1$.

	D-optimal Design		First-stage Design		Second-stage Design	
Point	p_i	(q_{1i}, q_{2i})	p_i	(q_{1i}, q_{2i})	p_i	(q_{1i}, q_{2i})
1	0.25	(1, 1)	0.125	(1, 1)	0.116	(1, 1)
2	0.25	(0.1353, 1)	0.125	(0.1353, 1)	0.115	(0.1359, 1)
3	0.25	(1, 0.1353)	0.125	(1, 0.1353)	0.115	(1, 0.1348)
4	0.25	(0.2313, 0.2313)	0.125	(0.1353, 0.1353)	0.154	(0.2450, 0.2441)

We use this example to illustrate how a two-stage design works under parameter misspecifications. To make the explanation easier, we have purposely assumed there is no misspecification of main effects and only the interaction is misspecified at the first stage. Due to misspecification, we can see that the interaction point in the first stage is $(0.1353, 0.1353)$. Note that the optimal interaction point is $(0.2313, 0.2313)$. Thus, in the first stage, the MED level at the interaction point is smaller than the optimal value. Therefore, this design fails to provide as much information about the interaction as possible. The second-stage design, which maximizes the determinant of joint information matrix based on more accurate information, tries to make up the misspecification in the first stage by extracting more information about the interaction. This is achieved in two ways: (1) the second-stage design puts slightly more experimental resources at the interaction point than any other point; (2) the second stage design pushes the MED level at the interaction point $(0.2450, 0.2441)$ higher than the corresponding MED level in the D-optimal design, in an effort to correct the unduly small MED level at the first stage. In a word, the two-stage design improves the design efficiency by making better use of the experiment at the second stage. One might notice that the designs in Table 4.6 are continuous. In practice, we need to convert them to exact design according to some algorithms. Atkinson and Donev (1992) discussed such algorithms in their book.

4.3.3 Evaluation of Two-stage Designs

This section discusses evaluation of two-stage designs. Similar to (4.2), D-efficiency of a two-stage design is defined as:

$$\text{Eff}_{2stage} = \left(\frac{|I(X, \boldsymbol{\beta})| \text{ of the two-stage design}}{|I(X, \boldsymbol{\beta})| \text{ of the D-optimal design}} \right)^{1/p}, \tag{4.17}$$

where p is the number of parameters.

It is straightforward to get the Fisher information matrix $I(\boldsymbol{\beta})$ for the D-optimal design. However, as mentioned earlier, it is more complicated for a two-stage design because it is constructed sequentially and the observations from the first and second stages are not independent. Following the same idea in Section 4.2.3, we will use simulation to approximate the Fisher information matrix of a two-stage design. Basically, for a given model, we can simulate many first-stage observation vectors $\mathbf{y}_1$, for each of which we can compute the observed information matrix defined in (4.15). Given a sample of observed information matrices, we can average out the observation vector $\mathbf{y}_1$ and get an estimate of the Fisher information matrix. The idea here can be expressed by (4.18):

$$I(\boldsymbol{\beta}) = E_{\mathbf{y}_1}(I(\boldsymbol{\beta}|\mathbf{y}_1)) \approx \frac{\sum_{k=1}^{K} I(\boldsymbol{\beta}|\mathbf{y}_1^k)}{K}. \tag{4.18}$$

Given an estimate of Fisher information matrix based on (4.18), we can compute its D-efficiency as defined in (4.17). As in Section 4.2.3, we need to clarify some issues concerning the simulation:

1. true model and total runs,
2. parameter misspecification,
3. simulation size (number of iterations, i.e., K in (4.18)),
4. strategy to allocate total runs to each stage.

In what follows, we will discuss the above issues. Though most of them are similar to those in Section (4.2.3), the last issue, allocation of total runs to the two stages, is a special question for two-stage designs.

True Model, Experiment Size and Parameter Misspecification

Clearly, we cannot evaluate two-stage designs for all true models, neither can we exhaust all possible experiment sizes or parameter misspecifications. Following Section 4.2.3, we will try to cover some representative situations. For comparison, we will consider the same situations as those listed in Table 4.2.

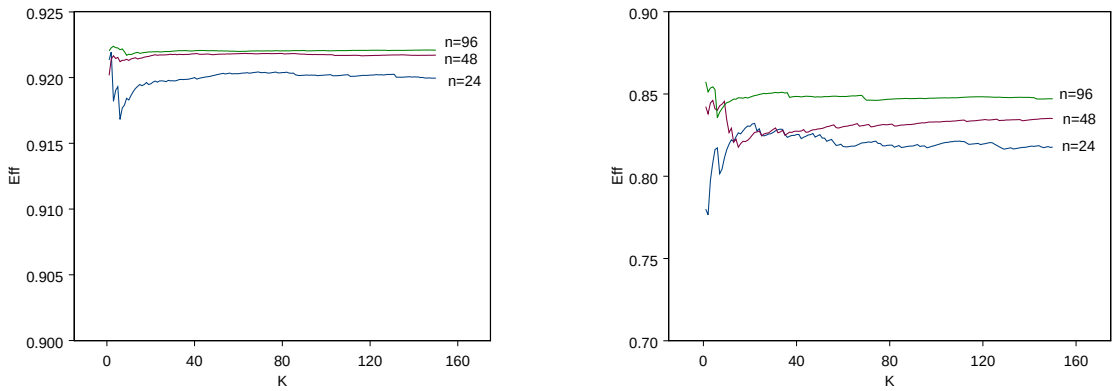
Simulation Size: K

Obviously, the larger K is, the better the approximation in (4.18) is. As mentioned earlier, the simulation is very time-consuming as each iteration includes many complicated tasks such as model fitting and numerical optimization of the determinant of the information matrix. Thus, a reasonable K needs to be determined. As in Section 4.2.3, a plot of the D-efficiency versus K, like Figure 4.5, can give us some idea of a reasonable K.

Figure 4.5 indicates that the simulated D-efficiency converges after about 70 iterations. Actually, the variation of D-efficiency is quite small after the first 40 iterations, implying that large K is not necessary. Similar pattern have been found in plots for some other representative models with different parameter misspecifications. In subsequent simulations, $K = 100$ is used unless otherwise stated.

Sample Allocation

Suppose that the total sample size is n. Of all the n experimental runs, n_1 are implemented at the first stage and n_2 at the second stage. Our question here is how to optimally allocate the n runs to the two stages. In other words, which (n_1, n_2) gives the best design efficiency?



(a) $r = -0.5$ (b) $r = 0.2$

Figure 4.5: D-efficiency versus K. In (a), the true model is $\lambda = exp(6.0 - x_1 - 2x_2 - x_1x_2)$ and parameter misspecifications are specified by $m_1 = m_2 = 0.5$; in (b), the true model is $\lambda = exp(6.0 - 2x_1 - 0.5x_2 + 0.2x_1x_2)$ and parameter misspecifications are specified by $m_1 = m_2 = 0.5$.

The optimal allocation is a very complicated question as it depends on many factors. Clearly, parameter misspecification plays an important role in the allocation. Intuitively, we can allocate more experimental resources to the first stage without damage to the design efficiency when we have sound information about parameters. The extreme case is that, if there is no misspecification, we should adopt a one-stage design since the D-optimal design is a one-stage design. Estimates from the first stage also affect allocation. Better estimates gives a better design for the second stage, thus making more efficient use of the experimental runs at the second stage. The quality of the first-stage estimates depends on many things such as the true model, the design of the first stage, and the experiment condition. Total sample size is another factor that has an impact on allocation. Though different allocation strategies give different efficiencies for an experiment of 50 runs, we would not expect any significant difference between them for a 1000-run experiment. There are some other factors that might affect how to optimally allocate the total experimental resources. The way the optimal allocation depends on these factors is not clear and very hard, if possible, to quantify. Therefore, it is practically impossible to analytically work out the optimal allocation. Many authors (Minkin, 1987; Sitter and Wu, 1999) just ignored this issue when they discussed two-stage designs. Myers et al. (1996) empirically investigated this issue for the one-variable logistic regression model. Basically, they evaluated some typical allocation strategies by simulation and made some rough recommendation for the particular case they studied. Letsinger (1995) did a more extensive simulation for the same model. We will also do some empirical studies for our situation. Basically, given the true model, the experiment size and the parameter misspecification, we will simulate the D-efficiency for some reasonable allocations and then pick the one that gives the highest efficiency. The setup of simulation

is outlined as follows:

1. We will consider four models: $\boldsymbol{\beta} = (6.0, -1.0, -0.5, -0.5)'$, $\boldsymbol{\beta} = (6.0, -3.0, -4.0, -6.0)'$, $\boldsymbol{\beta} = (5.0, -2.0, -1.0, -0.0)'$ and $\boldsymbol{\beta} = (5.0, -1.0, -1.0, 0.2)'$. Note that $r = -1, -0.5, 0$ and 0.2 for these four models, respectively.

2. For each model, we choose four misspecifications of main effects to represent different severity of misspecification: no misspecification ($m_1 = m_2 = 1$), mild misspecification ($m_1 = m_2 = 2$ and $m_1 = 0.3, m_2 = 0.4$), and severe misspecification ($m_1 = m_2 = 0.2$).

3. For each combination of the true model and the parameter misspecification of main effects, three sample sizes, $n = 24, 48$ and 96, are considered as they can represent small, moderate and large experiments.

4. For $n = 24$, we evaluate 4 different allocations for each combination of the true model and the parameter misspecification: $n_1 = 8, 12, 16$ and 20; for $n = 48$, we also consider four allocations: $n_1 = 8, 16, 24$ and 32; for $n = 96$, we choose $n_1 = 16, 24, 32, 48$ and 64.

For each sample size n and allocation n_1, we simulate the D-efficiency for each combination of the true model and the parameter misspecification. For simplicity, we take the average of these simulated efficiencies over all the true models and misspecifications. We use the average efficiency to evaluate the performance of this particular allocation n_1. Figure 4.6 plots the average D-efficiency versus different allocations for each experiment size.

Figure 4.6 can give us some insights into the optimal allocation:

1. For the situations we have considered, the optimal allocations are $n_1 = 12, 16$ and 24 for $n = 24, 48$ and 96, respectively. Note that the optimal proportions of sample size at the first stage are not the same for experiments of different sizes. For our situation, the optimal proportions are $1/2, 1/3$ and $1/4$, respectively. The proportion at the first stage has a tendency to decrease as the experiment size gets large. This result is in agreement with Letsinger's (1995) discovery.

2. The performance of different allocations varies less for larger experiments. This can be seen by comparing the plots for different n in Figure 4.6.

3. The average D-efficiency is higher for larger experiments, which is quite intuitive.

We know that, to obtain more reliable estimates for the second-stage design, we need to have more experimental runs at the first stage; however, to reduce the loss of efficiency due to misspecification at the first stage, we need to have more runs at the second stage. Thus, optimal allocation is a balance point of these two forces. Unfortunately, this balance point

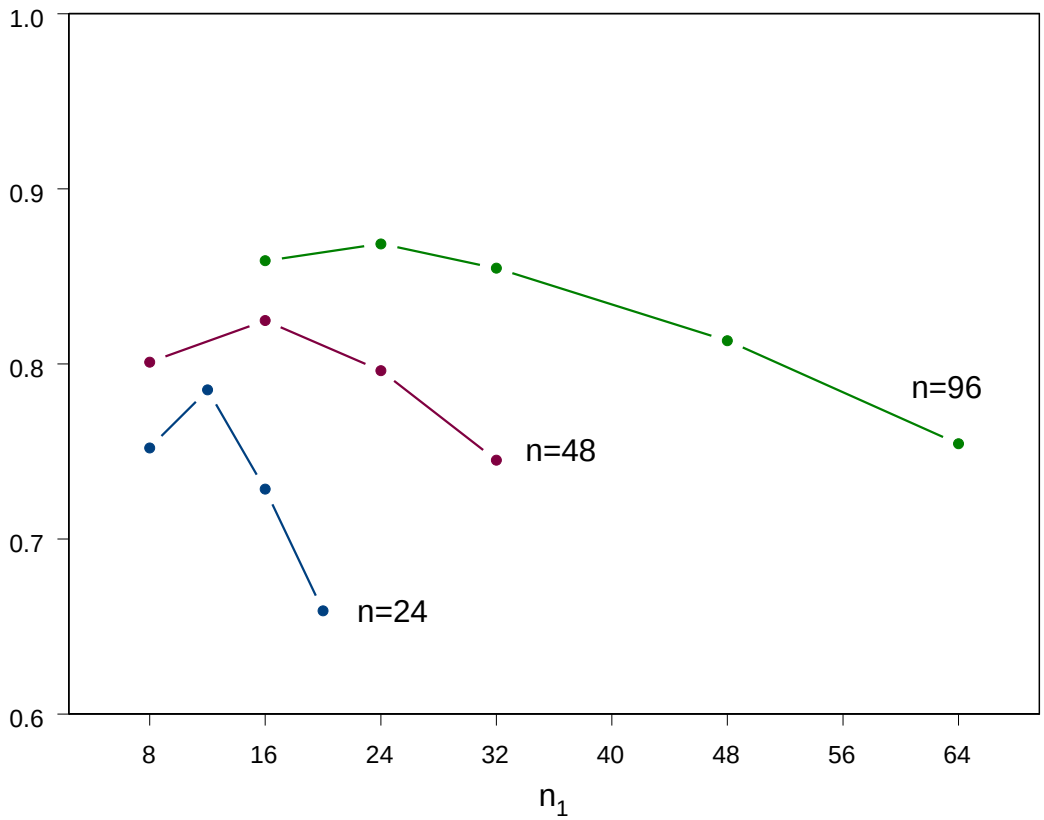


Figure 4.6: Average D-efficiency versus n_1 for Different n

is case-dependent and there is no general rule. Based on the situation we have considered, the first-stage experiment should have no less than 3 replications at each point; on the other hand, the proportion of experimental runs at the first stage should not exceed 1/2 under any circumstance; further, this proportion should be smaller for larger experiments.

Simulation Results

As indicated earlier, for comparison reason, we shall evaluate two-stage designs for the same true models, experiment sizes and misspecifications as listed in Table 4.2. Based on the above discussion about sample allocation, we use 1/2 of the total runs at the first stage for $n = 24$; 1/3 for $n = 48$ and 1/4 for $n = 96$. Table 4.7 lists the simulated D-efficiency for each situation.

We now conclude this section by making some comments about two-stage designs

Table 4.7: D-efficiency of Two-stage Designs

$\lambda = exp(6.0 - 2x_1 - 0.2x_2 - 0.4x_1x_2)$, $r = -1$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.7618	0.7940	0.8672	0.7905
48	0.8412	0.8373	0.9144	0.8707
96	0.8807	0.8659	0.9373	0.9037
$\lambda = exp(6.0 - 0.4x_1 - x_2 - 0.2x_1x_2)$, $r = -0.5$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.7344	0.9214	0.8411	0.7439
48	0.8110	0.9510	0.8968	0.8398
96	0.8491	0.9635	0.9242	0.8795
$\lambda = exp(5.0 - x_1 - 1.5x_2 - 0x_1x_2)$, $r = 0$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.6925	0.9795	0.7300	0.6600
48	0.7535	0.9889	0.8092	0.7133
96	0.8057	0.9912	0.8487	0.7668
$\lambda = exp(5.0 - 2x_1 - 3x_2 + 1.2x_1x_2)$, $r = 0.2$				
n	$m_1 = 2$, $m_2 = 2$	$m_1 = 1$, $m_2 = 1$	$m_1 = 0.3$, $m_2 = 0.4$	$m_1 = 0.2$, $m_2 = 0.2$
24	0.8235	0.8904	0.7366	0.6711
48	0.8751	0.9156	0.8097	0.7072
96	0.9036	0.9402	0.8265	0.7601

based on the simulation results in Table 4.7:

1. In general, two-stage designs are quite robust to parameter misspecifications. Except for a few cases, the D-efficiency is above 80% in Table 4.7.

2. Like sequential designs, two-stage designs are more efficient for larger experiments. This can be seen from both Figure 4.6 and Table 4.7.

3. When parameters are severely misspecified, the efficiency of two-stage designs drops below 80% even for experiments of 96 runs. In Table 4.7, this is indicated by efficiencies for the last two models when $m_1 = m_2 = 0.2$. Due to severe misspecification, the estimates from the first stage are not reliable so the second-staged design is also based on misspecified parameters, which causes loss of design efficiency.

4. We shall compare the two-stage design to the sequential design and the conditional D-optimal design in next section.

4.4 Summary

In this chapter, we have investigated sequential experimental designs for the two-toxicant interaction model. We now give an summary of this chapter:

1. Firstly, we studied sequential designs on the unrestricted design space for the two-toxicant interaction model. We outlined the algorithm of sequential designs and illustrated this algorithm with an example. Due to the dependence between the observations, it is very difficult to analytically extract the Fisher information matrix for sequential designs. We discussed how to use simulation to approximate the Fisher information matrix, which further enabled us to compute D-efficiency of sequential designs. We evaluated the performance of sequential designs by simulation for some representative models and sample sizes. It has been demonstrated that sequential designs are very efficient and robust to parameter misspecifications.

2. We discussed restrictions imposed on the design space, which is necessary in many applications. In Chapter 2, we imposed some lower boundaries on IED levels to prevent undesirably large proportion of mortality. It turns out that restrictions of this type, though simple to deal with, are not strong enough at the interaction point. In this chapter, we proposed a new type of restrictions, namely, restrictions imposed on the MED level instead of on IED levels. Bounding the MED level can effectively prevent extremely low survival proportion; however, restrictions of this type depend on parameters through r. Thus, misspecifications of parameters have impact on both the design and the design space so the locally optimal design is not a good choice due to its poor robustness to misspecifications. We proposed sequential designs for this situation. At each step, we redefine the design space and find the next design point using the up-to-date estimates. To evaluate designs on restricted regions, we proposed penalized D-efficiency since the original D-efficiency is not an appropriate measure on the restricted design space. Based on this penalized D-efficiency, we have demonstrated that the sequential design works well under restrictions.

3. We discussed statistical inferences based on sequential designs. We pointed out that, if either a Bayesian or likelihood-based approach to inferences is adopted, we can just ignore the fact that the design is sequentially constructed and make inferences as if all the observations are independent of each other. For the repeated-sampling approach, we did a small simulation to demonstrate that the actual coverage of the confidence region based on the observed information matrix is close to that based on the expected information matrix and both are close to the nominal level. Therefore, when constructing the confidence region, we can just pretend that the observations are independent and treat the observed information matrix as the Fisher information matrix.

4. To reduce the stages of sequential designs, we proposed two-stage designs for the Poisson regression model. The algorithm for two-stage designs is similar to that for sequential designs. A particular problem with two-stage designs is how to optimally allocate the total sample to the two stages. For some representative sample sizes, we did a simulation to evaluate some typical allocation strategies and found optimal allocations. We then simulated D-efficiencies of two-stage designs based on these optimal allocations. We have found that two-stages are quite efficient unless parameters are severely misspecified.

Table 4.8: D-efficiency of Sequential, Two-stage and Conditional D-optimal Designs

$\lambda = exp(6.0 - 2x_1 - 0.2x_2 - 0.4x_1x_2)$, $r = -1$								
n	$m_1 = 2.0$, $m_2 = 2.0$		$m_1 = 1.0$, $m_2 = 1.0$		$m_1 = 0.3$, $m_2 = 0.4$		$m_1 = m_2 = 0.2$	
24	0.9161[a]	0.7618[b]	0.9391	0.7940	0.9537	0.8672	0.9234	0.7905
48	0.9583	0.8412	0.9702	0.8373	0.9771	0.9144	0.9640	0.8707
96	0.9790	0.8807	0.9850	0.8659	0.9883	0.9373	0.9811	0.9037
c[c]	0.0154		0.5729		0.6084		0.2966	
$\lambda = exp(6.0 - 0.4x_1 - x_2 - 0.2x_1x_2)$, $r = -0.5$								
n	$m_1 = 2.0$, $m_2 = 2.0$		$m_1 = 1.0$, $m_2 = 1.0$		$m_1 = 0.3$, $m_2 = 0.4$		$m_1 = 0.2$, $m_2 = 0.2$	
24	0.9106	0.7344	0.9624	0.9214	0.9420	0.8411	0.9193	0.7439
48	0.9545	0.8110	0.9813	0.9510	0.9708	0.8968	0.9592	0.8398
96	0.9774	0.8491	0.9908	0.9635	0.9854	0.9242	0.9791	0.8795
c	0.0979		0.8107		0.5544		0.2596	
$\lambda = exp(5.0 - x_1 - 1.5x_2 - 0x_1x_2)$, $r = 0$								
n	$m_1 = 2.0$, $m_2 = 2.0$		$m_1 = 1.0$, $m_2 = 1.0$		$m_1 = 0.3$, $m_2 = 0.4$		$m_1 = 0.2$, $m_2 = 0.2$	
24	0.8870	0.6925	0.9665	0.9795	0.9035	0.7300	0.8809	0.6600
48	0.9406	0.7535	0.9771	0.9889	0.9444	0.8092	0.9402	0.7133
96	0.9666	0.8057	0.9902	0.9912	0.9699	0.8487	0.9674	0.7668
c	0.5412		1.0000		0.4404		0.1982	
$\lambda = exp(5.0 - 2x_1 - 3x_2 + 1.2x_1x_2)$, $r = 0.2$								
n	$m_1 = 2.0$, $m_2 = 2.0$		$m_1 = 1.0$, $m_2 = 1.0$		$m_1 = 0.3$, $m_2 = 0.4$		$m_1 = 0.2$, $m_2 = 0.2$	
24	0.9142	0.8235	0.9350	0.8904	0.8869	0.7366	0.8793	0.6711
48	0.9434	0.8751	0.9614	0.9156	0.9369	0.8097	0.9286	0.7072
96	0.9700	0.9036	0.9767	0.9402	0.9628	0.8265	0.9604	0.7601
c	0.6519		0.6278		0.2284		0.1010	

[a]The first column of each misspecification is D-efficiency for sequential designs.
[b]The second column of each misspecification is D-efficiency for two-stage designs with optimal allocations.
[c]The last row of each model is D-efficiency for the conditional D-optimal Design.

5. We now compare the performance of sequential designs, two-stage designs and conditional D-optimal designs. Table 4.8 lists the D-efficiencies of these designs for different

models under various parameter misspecifications. When there is no misspecification ($r = 0, m_1 = m_2 = 1$), we know that the conditional D-optimal design is the D-optimal design so its D-efficiency is 1; the efficiency of the two-stage design is slightly less than 1 but a little higher than the efficiency of the sequential design. The efficiency loss of the sequential design is due to randomness of the estimates at each stage. When misspecification occurs, D-efficiency of the sequential design is higher than that of the two-stage design, which in turn is higher than the conditional D-optimal design. Thus, from the viewpoint of D-efficiency, the sequential design is recommended whenever possible. However, as we indicated earlier, sequentially implementing a design is more costly in general, especially when it takes a long time to obtain an observation. In this sense, locally optimal designs like the conditional D-optimal design are preferable. In practice, we have to balance the design efficiency and the cost incurred by sequential construction of a design. If design efficiency is of ultimate importance and we do not have much information about the parameters, we should go for sequential designs. If we have reliable information about the parameters, one-stage design like locally D-optimal designs is recommended. The two-stage design is a good choice to balance the design efficiency and the cost of carrying out the experiment sequentially.

In this chapter, we have only considered sequential design for the two-toxicant interaction model. However, the methodology we have discussed can be applied to other models and it is anticipated that the results will be similar.

Chapter 5

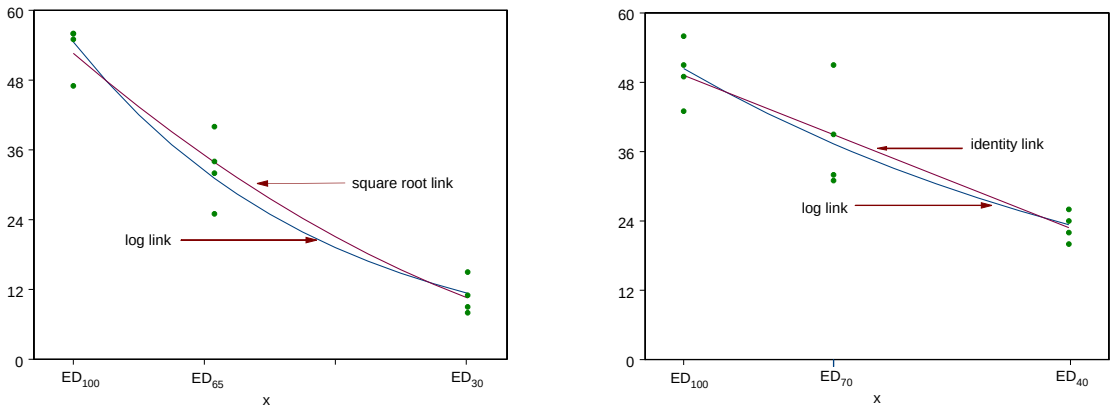
Optimal Designs for Link Discrimination

5.1 Introduction

For binary data, the logit link function and the probit link function are both widely used for analysis, leading to *logit analysis* and *probit analysis*, respectively. McCullagh and Nelder (1989) compared these two functions and pointed out that they are almost linearly related over the interval $0.1 \leq \pi \leq 0.9$, where π is the probability of "success". For this reason, it is usually difficult to discriminate between these two link functions. For count data, it also happens that two or more competing models with different link functions may be suitable to describe the data, especially when the design space is restricted. We shall give two hypothetical examples of one-toxicant experiments for illustration.

In both examples, it is believed that the first-order model is adequate to describe the dose-response relationship. Three-point equally spaced designs are adopted. In the first example, the design space is restricted by $q \geq 0.3$ and the equally spaced design is formed by $q_1 = 1, q_2 = 0.65$ and $q_3 = 0.3$, with 4 replicates at each of the three points. Similarly, in the second example, the design space is defined by $q \geq 0.4$ and the design points are located at $q_1 = 1, q_2 = 0.7$ and $q_3 = 0.4$, with 4 runs at each point. We assume that the log function is the true link for both examples. More specifically, in both examples, we assume that the mean function of the true model is $\lambda = exp(4-x)$. Responses are simulated based on the true model for both experimental designs. After the data are obtained, we usually try different models to fit the data. For each example, we found that, at least, two models fit the simulated data quite well and compete with each other. For the first example, both the log link and the square root link provide very good fits; the two fitting curves are $\lambda = exp(4.0 - 1.3x)$ and $\lambda = (7.25 - 3.33x)^2$ and their deviance values are 7.78 and 7.46, respectively. For the second example, the log link and the identity link compete with

each other; the fitting curves are $\lambda = exp(3.92 - 0.84x)$ and $\lambda = 49.21 - 28.78x$ and the corresponding deviance values are 9.18 and 9.12, respectively. Figure 5.1 plots the raw data and fitting curves of the competing models for both examples.



(a) Log Link and Square Root Link

(b) Log Link and Identity Link

Figure 5.1: Examples of Competing Link Functions. In both examples, the true model is $\lambda = exp(4 - x)$. In (a), two fitting curves, $\lambda = exp(4.0 - 1.3x)$ and $\lambda = (7.25 - 3.33x)^2$, compete with each other; in (b), two fitting curves, $\lambda = exp(3.92 - 0.84x)$ and $\lambda = 49.21 - 28.78x$, compete with each other.

The above examples illustrate the situations in which two distinct models, or two different link functions, compete with each other in fitting the data, making model selection very hard. Actually, if the model selection is purely based on some goodness-of-fit statistics like the deviance, the true model that generates the data will be missed in both examples. We can alleviate such a problem by carefully designing the experiment. In general, unplanned experimental designs usually give rise to situations when a number of models may provide reasonable fits to the data, hence demanding model selection procedures to be applied. This is done so as to make a sensible choice of model. It is possible, though, that even the most efficient model selection procedure may be unable to correctly identify the true model, due to the amount of confounding introduced by the underlying design, as illustrated in the above examples. More carefully designed experiments, however, would stand more chances of generating data leading to an adequate fit for the most appropriate model, thus enabling the experimenter to discard all but one model. For instance, in Figure 5.1, though the two competing curves are very close to each other at both end points, discernible differences exist at the middle point; intuitively, if we do not use equal allocation designs and have more experimental runs at the middle point, we would stand more chances of discriminating between the competing models.

In this chapter, we shall discuss experimental designs for discrimination of models for the Poisson data. For simplicity, we mostly focus on the one-toxicant first-order model;

however, all the methods discussed in this chapter can be applied to more complicated models.

5.2 T-optimality

To discriminate between competing models, Atkinson and Fedorov (1975) introduced T-optimality design criterion in the context of optimal design theory. For classical linear models, the T-optimal design assumes that the true model is known and maximizes the residual sum of squares arising from the fit of the competing model, which, equivalently, maximizes the non-centrality parameter of the χ^2 distribution of the residual sum of squares for the competing model. Thus, the T-optimal design provides the most powerful lack-of-fit test for the competing model and it is in this sense that the T-optimal design is optimal in discriminating the true model from the competing model.

Ponce de Leon and Atkinson (1992) extended the T-optimality criterion to include generalized linear models belonging to the same subclass. The extension is based on the analogy between the residual sum of squares in classical linear regression theory and the deviance in generalized linear model theory. Just as the former plays an important role in the concept of T-optimality for linear models, the latter provides the basis on which the criterion function is defined in the context of generalized linear models.

Since we are dealing with the Poisson regression model, a special case of generalized linear models, we shall give an introduction of T-optimality criterion for discrimination between two competing generalized linear models. This exposition is mainly based on Ponce de Leon and Atkinson's work.

In Atkinson and Ponce de Leon's approach to designing experiments to discriminate between two generalized linear models, the true model and its parameters are assumed to be known. We will discuss how to relax this assumption later. The two competing models are assumed to belong to the same subclass of the generalized linear model. Their link functions and/or linear predictor structures may or may not be the same, so the models may or may not be nested. There is no restriction on the kind of linear structure nor on the number of linear predictor parameters.

Let $\boldsymbol{\mu}_i$ and $\boldsymbol{\beta}_i$ denote the vector of means and the vector of model parameters corresponding to the ith model, where $i = 1, 2$. More specifically, $\boldsymbol{\mu}_i' = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in})$ for an n-observation experiment. Now assume that the first model is the true model and that its parameters $\boldsymbol{\beta}_1$ are known. When the second model is fitted to the data that are generated by the first model, we estimate the parameters $\boldsymbol{\beta}_2$ by minimizing the deviance function:

$$\min_{\beta_2 \in B_2} D(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2), \tag{5.1}$$

where B_2 is the parameter space for the second model and the deviance function $D(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ depends on the generalized linear model subclass regarded (see McCullagh and Nelder, 1989).

Suppose that the estimates are $\hat{\boldsymbol{\beta}}_2$ and the corresponding estimated mean vector is $\hat{\boldsymbol{\mu}}_2 = \boldsymbol{\mu}_2(\hat{\boldsymbol{\beta}}_2)$; then the deviance function from fitting the second model is $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$. Under the above assumptions, Atkinson and Ponce de Leon (1992) introduced the generalized T-optimality criterion function which consists of maximizing the deviance arising from the fit of model 2 when data are generated by model 1, i.e.,

$$\max_{X \in \mathcal{D}} D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2), \tag{5.2}$$

where $\mathcal{D}$ is the set of all possible designs. We know that the deviance function $D(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$ is simply the sum of deviance at each point. Suppose that the deviance at the design point $\mathbf{x}_j$ is $d(\mathbf{x}_j, \boldsymbol{\beta}_2)$; then the above T-optimality criterion can be equivalently written as

$$\max_{\eta \in H} \min_{\beta_2 \in B_2} \sum_j p_j d(\mathbf{x}_j, \boldsymbol{\beta}_2), \tag{5.3}$$

where, following the notation in Chapter 3, η is a design measure; H is the set of all design measures on the design space $\mathcal{X}$ and p_j is the proportion of sample size at the jth point.

It is important to recall that, for the moment, the true model and its parameters are assumed to be known. Therefore, the above criterion function depend solely on the design measure η. Our goal is to find a design measure η^* that satisfies this criterion. In (5.3), the deviance function depends on the generalized linear model subclass considered. Particularly, for the Poisson distribution, the deviance function at the point $\mathbf{x}$ is given as

$$d(\mathbf{x}, \boldsymbol{\beta}_2) = \mu_1(\mathbf{x}) \ln\left(\frac{\mu_1(\mathbf{x})}{\mu_2(\mathbf{x}, \boldsymbol{\beta}_2)}\right) - (\mu_1(\mathbf{x}) - \mu_2(\mathbf{x}, \boldsymbol{\beta}_2)). \tag{5.4}$$

Ponce de Leon and Atkinson (1992) argued that the general equivalence theory discussed in Chapter 3 also applies to T-optimality. According to the equivalence theory, a design η^* is T-optimal if and only if the deviance function $d(\mathbf{x}, \hat{\boldsymbol{\beta}}_2^*)$ is bounded by the value of the criterion function at the maximum, where $\mathbf{x}$ is any point in the design space and $\hat{\boldsymbol{\beta}}_2^*$ are the estimates of $\boldsymbol{\beta}$ based on the design η^*. Additionally, they pointed out that the upper limit is achieved at all the points of the T-optimal design. This property enables us to check the T-optimality of any design. Specifically, if we are able to demonstrate that $d(\mathbf{x}, \hat{\boldsymbol{\beta}}_2^*)$ achieves its maximum at all the design points, we can conclude that η^* is the T-optimal design. In practice, the deviance function is computed over a grid in the design space and then plotted to demonstrate the optimality of a particular design, as what we did in Chapter 3.

Recall that, according to the Carathéodory's Theorem (Silvey, 1980, p.72), a continuous D-optimal design is based on a finite number of support points and, more specifically, on at most $p(p+1)/2$ points, where p is the number of model parameters. However, as far as the discrimination problem is concerned, there is no result in the literature so far concerning limits for the number of support points of the T-optimal design, a lack of which complicates considerably the process of searching for a T-optimal design.

5.3 Locally T-optimal Designs

Recall that the T-optimality criterion depends upon knowledge of the true model and of its associated parameter values. Accordingly, the resulting design is only locally optimal. However, as indicated in previous chapters, though the locally optimal design has limited applications in practice, it provides a useful benchmark for evaluation of any other design. Thus, in this section, we will discuss locally T-optimal designs for discrimination between two competing models for Poisson data.

Basically, to find the locally T-optimal design, we need to specify the true model and the structure of the competing model. Then we can use numerical methods to find the design satisfying the criterion in (5.3). For this T-optimality criterion function, the numerical optimization involves both maximization and minimization so it is more complicated than the task of finding the locally D-optimal design discussed in Chapter 2. For simplicity, we only consider one-toxicant models. *Particularly, we assume that the first-order model is adequate to describe the dose-response relationship so the T-optimal design is actually the optimal design for discrimination between two competing link functions.*

Though the log link is widely used to model Poisson data, other link functions such as the identity link and the square root link also find many applications (see Frome and DuFrain, 1986; Myers, Montgomery and Vining, 2002). In many situations, the log link and the other two link functions provide very close fits especially when the design space is restricted, as illustrated in Figure 5.1. Their closeness makes model selection very difficult and a carefully designed experiment for link discrimination is necessary. Thus, we will concentrate on designs for discrimination between the log link and the other two link functions. We will consider three different restricted design regions defined by $q \geq c$ with $c = 0.2, 0.3$ and 0.4, respectively. The unrestricted design space is not considered here because it is easy to discriminate the log link from the other two functions on the unrestricted design space. This is so because the log function has a heavier tail than the other two functions on the unrestricted space.

As indicated earlier, the true model (both the link function and its parameters) and the competing link function need to be specified. For instance, we can assume that the log link generates the data and the square root is the rival link function. We need to specify the parameters for the log link and estimate the parameters for the square root link by

minimizing the deviance arising from the fitting, as given in (5.1). Then the optimal design is found by maximizing the minimum deviance over the design space. Similarly, we also consider three other cases: the square root link is the true link and the log is the rival link; the log function is the true link and the identity is the rival link and vice versa. Newton-Raphson method is adopted for optimizing the T-optimality criterion function defined in (5.3). As mentioned earlier, there is no result in the literature concerning the number of support points for T-optimal designs so we start with a two-point design and increase the number of support points by 1 each time until the optimality is verified via equivalence theory. Table 5.1 lists the optimal designs for these four cases on different design regions. We can see some features of T-optimal designs based on Table 5.1:

1. All the optimal designs in Table 5.1 are three-point designs. Particularly, the two boundaries of the design space, ED_{100} and ED_{100c}, are always in the optimal designs. The location of the middle point is case-dependent.

2. For Poisson data, the T-optimality criterion function in (5.3) is asymmetric with respect to the model truth, which results from the asymmetry of the deviance function in (5.4). As a result, for the same two competing link functions, the optimal designs are not the same or symmetric with respect to the model truth.

3. For all the optimal designs, about half of the total sample size is allocated to the middle point. As illustrated in Figure 5.1, though the log function is close to the other two functions at the two end points, discernible difference exists in the middle. Therefore, putting more experimental runs at the middle points helps the discrimination. When the log function is the true link, more runs are put at the control point than at the other end point. Conversely, the other end point ($q = c$) has more runs than the control point when log is not the true link.

4. The restriction on the design space has impact on the locations of the non-control points. More specifically, the non-control points move closer to the control point as the restriction becomes stronger. The restriction also affects distribution of the sample size but the effect is actually negligible.

As indicated earlier, all the designs in Table 5.1 have been verified to be optimal via equivalence theory. We now give a brief discussion on how the general equivalence theory works here.

In Section 5.1, we discussed Ponce de Leon and Atkinson's method for checking T-optimality of a design. Basically, when optimum of the T-optimality criterion function in (5.3) is achieved, the parameter estimates for the rival model, $\hat{\boldsymbol{\beta}}_2^*$, can be obtained. Based on the estimates, the deviance function $d(\mathbf{x}, \hat{\boldsymbol{\beta}}_2^*)$ can be computed as defined in (5.4). We then evaluate and plot the deviance function over the design space and if we are able to demonstrate that this deviance function achieves its maximum at all the designs points,

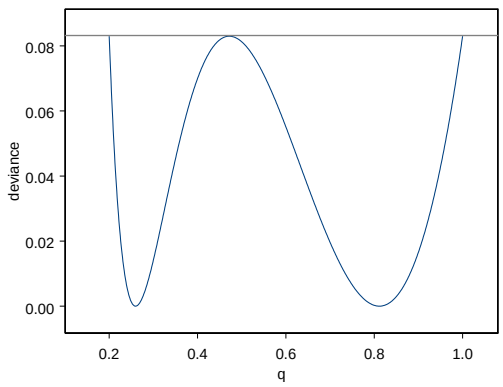
according to Ponce de Leon and Atkinson's (1992) result, we can conclude that the design is T-optimal. By using this method, we have verified T-optimality of all the designs in Table 5.1. For illustration, Figure 5.2 plots the deviance function for four situations. In all the plots, one can see that the deviance function achieves its maximum at all the design points, confirming T-optimality of all the designs.

We have studied locally T-optimal designs for discrimination between the log link and the square root link or the identity link. To develop these designs, we assume the true link and its parameters are known. This assumption restricts application of locally T-optimal designs. To relax this assumption, two approaches have been discussed in the literature: one is Bayesian approach and the other is sequential experimentation. For the Bayesian design, one need to specify the prior distributions for both the true link function and the model parameters. Ponce de Leon and Atkinson (1991) discussed Bayesian T-optimal designs for linear regression models. The idea of Bayesian design is very simple but it is computationally difficult to apply it to generalized linear models due to the integration part in the design criterion. For classical linear models, Atkinson and Fedorov (1975) discussed sequential experimentation which converges to the T-optimal design. We will extend their idea to our situation in next section.

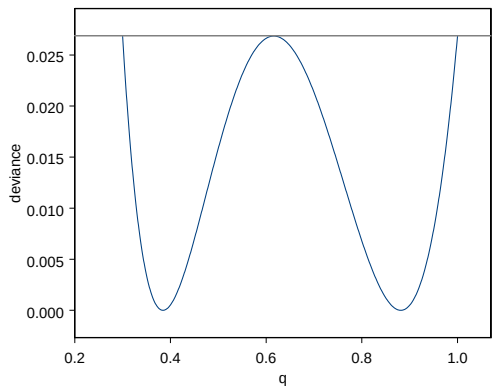
5.4 Sequential Designs

5.4.1 Algorithm

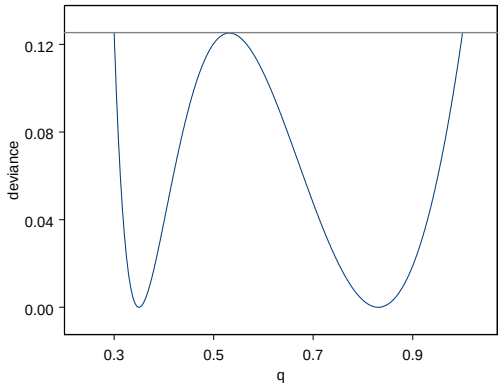
Obviously, for experiments that are both affordable and can be performed in a short period of time, sequential designs could be used to relax the assumption of locally T-optimal that the link function and the parameters are known. In a sequential design, we can update our information about the true model after each observation is made. After enough observations are made, our updated information will converge to the model truth so we can make nearly optimal use of the remaining experimental resources. For discrimination between two linear regression models, the procedure of the sequential design proposed in Atkinson and Fedorov (1975) consists of, at a given stage, taking the next observation at the point which maximizes the squared difference between the predicted responses from the two models. Asymptotically, this criterion leads to the T-optimal design. Atkinson and Fedorov's idea can be extended to include generalized linear models, according to the analogy that exists between residual sum of squares in linear regression and the deviance for generalized linear models. Müller and Ponce de Leon (1996) followed this idea and investigated sequential designs for discriminating between two rival binary data models, logit and probit models. In this section, we will discuss sequential designs for discrimination between two rival link functions of Poisson data models. As in the last section, we only consider first-order models. We assume that one of the two competing link functions is the true link. However, we do not know which one is true and its parameters are also unknown. Based on the above assumptions, we give the algorithm of



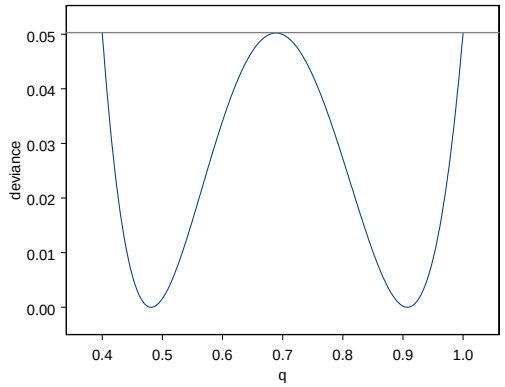
(a) Log (true link) and Square Root



(b) Square Root (true link) and Log



(c) Log (true link) and Identity



(d) Identity (true link) and Log

Figure 5.2: Optimality of Locally T-optimal Designs. The lower boundaries for ED levels are 0.2, 0.3, 0.3 and 0.4 in (a), (b), (c) and (d), respectively. The true link and the rival link are specified in each figure, for instance, in (a), the true link function is the log function and the rival link is the square root, etc. In all the figures, the deviance functions achieve the maximum at all the design points, confirming T-optimality of the designs.

Table 5.1: T-optimal Designs for One-toxicant First-order Model

Log (true link) and Square Root (rival link)						
c^a	q_1	q_2	q_3	p_1	p_2	p_3
0.2	1.0000	0.4719	0.2000	0.2644	0.5070	0.2286
0.3	1.0000	0.5645	0.3000	0.2611	0.5038	0.2351
0.4	1.0000	0.6436	0.4000	0.2587	0.5022	0.2392
Square Root (true link) and Log (rival link)						
c	q_1	q_2	q_3	p_1	p_2	p_3
0.2	1.0000	0.5496	0.2000	0.1796	0.4747	0.3457
0.3	1.0000	0.6161	0.3000	0.1949	0.4854	0.3197
0.4	1.0000	0.6776	0.4000	0.2067	0.4914	0.3019
Log (true link) and Identity (rival link)						
c	q_1	q_2	q_3	p_1	p_2	p_3
0.2	1.0000	0.4235	0.2000	0.3182	0.5340	0.1478
0.3	1.0000	0.5313	0.3000	0.3039	0.5190	0.1771
0.4	1.0000	0.6214	0.4000	0.2927	0.5108	0.1964
Identity (true link) and Log (rival link)						
c	q_1	q_2	q_3	p_1	p_2	p_3
0.2	1.0000	0.5748	0.2000	0.1543	0.4701	0.3757
0.3	1.0000	0.6330	0.3000	0.1739	0.4824	0.3437
0.4	1.0000	0.6888	0.4000	0.1897	0.4895	0.3208

[a]The design space is restricted by $c \leq q$.

sequential designs to discriminate between two link functions for the Poisson data:

1. Start with an initial design, which could be an arbitrarily chosen design. Based on the structure of the T-optimal design, we will use an equally spaced three-point design as the initial design with one run at each point. Implement the initial design and take the observations.

2. After k observations are made, find the maximum likelihood estimates $\hat{\boldsymbol{\beta}}_{1k}$ and $\hat{\boldsymbol{\beta}}_{2k}$ for model 1 and model 2, respectively. Meanwhile, obtain the deviances D_{1k} and D_{2k} from fitting the two competing models.

3. Denote $\mu_1(x, \hat{\boldsymbol{\beta}}_{1k})$ and $\mu_2(x, \hat{\boldsymbol{\beta}}_{2k})$ by $\hat{\mu}_{1k}$ and $\hat{\mu}_{2k}$, respectively. The $(k+1)^{th}$ design point x_{k+1} is found as follows:

$$x_{k+1} = \begin{cases} \text{Arg} \max\limits_{x} \{\hat{\mu}_{1k} \ln\left(\frac{\hat{\mu}_{1k}}{\hat{\mu}_{2k}}\right) - (\hat{\mu}_{1k} - \hat{\mu}_{2k})\}, & \text{if } D_{1k} < D_{2k}; \\ \text{Arg} \max\limits_{x} \{\hat{\mu}_{2k} \ln\left(\frac{\hat{\mu}_{2k}}{\hat{\mu}_{1k}}\right) - (\hat{\mu}_{2k} - \hat{\mu}_{1k})\}, & \text{otherwise.} \end{cases}$$

4. Take the $(k+1)^{th}$ observation at x_{k+1}. Based on the $k+1$ observations, find new maximum likelihood estimates $\hat{\boldsymbol{\beta}}_{1(k+1)}$ and $\hat{\boldsymbol{\beta}}_{2(k+1)}$ corresponding to model 1 and model 2, respectively.

5. Repeat steps 2, 3 and 4 until all the experimental resources are run out or there is no change in the structure of the design.

Clearly, the most important step in the above procedure is Step 3, in which the next design point is found based on all the available observations. Basically, the $(k+1)^{th}$ point is found to be the one that maximizes the deviance between the two models evaluated at the parameter estimates from the first k observations. The rationale behind this is the additivity of the deviance at each data point. Recall that the deviance based on all the data is simply the sum of the deviance at each point. Thus, by maximizing the deviance at each data point, we approximately maximize the deviance based on all the observations. For discriminating between two linear regression models, Atkinson and Fedorov's approach simply maximizes the squared difference between the predicted responses from the two models to find the next design point. This makes sense for their situation because the T-optimality criterion function is symmetric with respect to model truth for linear regression models. However, for generalized linear models, we have demonstrated in last section that, for the same two competing models, the T-optimal design depends on which model is the true model. Thus, in Step 3, we "estimate" the model truth using the deviance functions based on all the available observations. Basically, if the deviance for the first model is smaller, then we assume that the first model is the true model when finding the next observation; otherwise, the second model is assumed to be the true model. For the first few runs, the "estimate" of the model truth may not be accurate; but after obtaining enough observations, we will have more accurate information about the model truth. After enough observations are made, the "estimate" will converge to the model truth. Provided that one of the two models is true, either $\hat{\mu}_{1k}$ or $\hat{\mu}_{2k}$ will converge to the true model as $k \to \infty$. The sequential design strategy will then converge to the T-optimal design.

To illustrate the above procedure of sequential design, we give a simple simulated examples. In the example, the true model is assumed to be $\lambda = (7-2x)^2$ and the rival link is the log function. The design space is restricted by $q \geq 0.2$. For this case, we know that the T-optimal design has three points located at ED_{100}, ED_{55} and ED_{20} with approximately 18%, 47% and 35% of the total sample allocated to these three points, respectively. For the sequential design, we start with a three-point equally spaced design with one run at each point. We assume that the middle point ED_{55} is misspecified as ED_{60} in the initial experiment. Thus, the three initial designs points are ED_{100}, ED_{60} and ED_{20}. Based on the initial experiment, we simulate three observations from the true model. With these simulated observations, we can estimate the parameters for both models and obtain the deviances from fitting these two models. Then we can find the next design point by maximizing the estimated deviance. This process is repeated until we have 100 runs in total. Figure 5.3 illustrates this process. In Figure 5.3, the T-optimal design, initial design, sequential designs after 20,

50 and 100 runs are plotted. In the figure, dots stand for locations of design points and numbers represent the number of experimental runs. We can see that, after 20 runs, design points of the sequential design form three cluster corresponding the three design points in the T-optimal design. Particularly, two of these three clusters are just the two end points of the T-optimal design and the other cluster center around the middle point of the optimal design. However, allocation of the sample size is not close to the optimal allocation. After 50 runs, both the location of design points and the allocation are close to those of the optimal design. The sequential design gets closer to the T-optimal design after 100 runs.

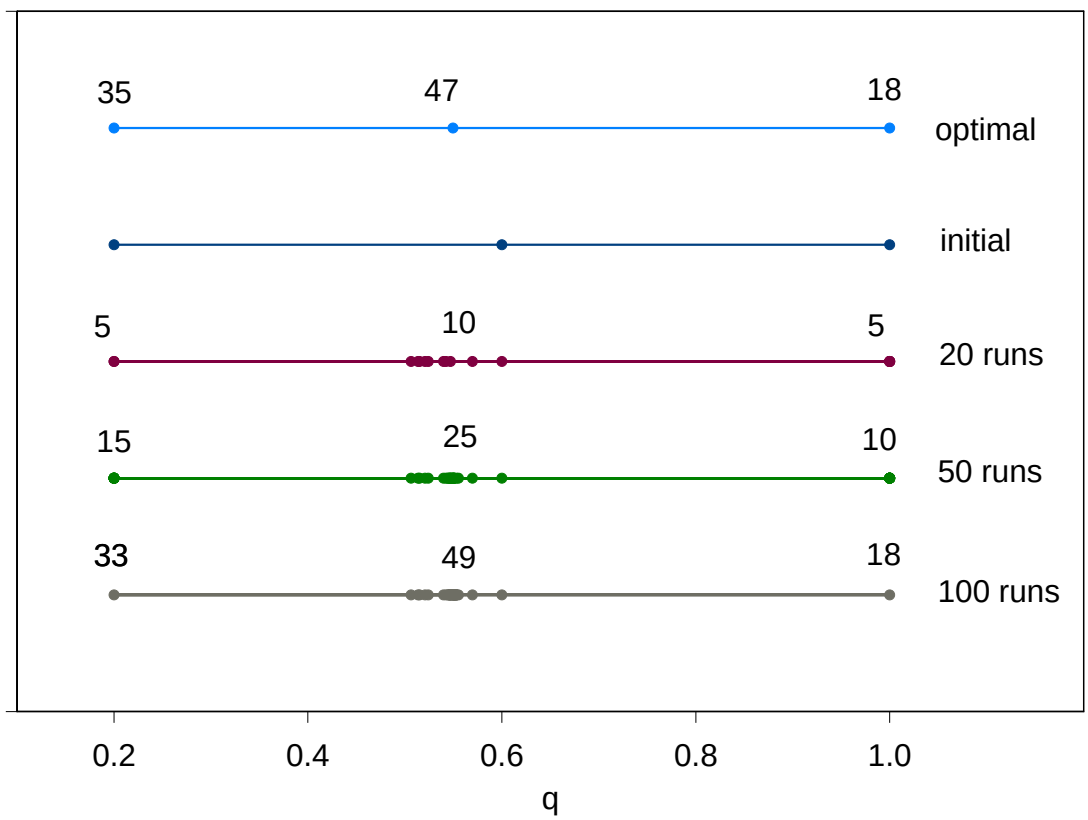


Figure 5.3: An Example of Sequential Design for Discrimination of Two Link Functions. The true model is assumed to be $\lambda = (7 - 2x)^2$ and the rival link is the log link. The design space is defined by $0.2 \leq q$. The top line is the T-optimal design and the other lines illustrate the process of sequential experimentation. Numbers in the plot stand for the number of replicates at the corresponding point or cluster.

The above example illustrates that sequential experimentation works well for discrimination between two competing link functions. We shall formally discuss evaluation of

sequential designs next.

5.4.2 Evaluation

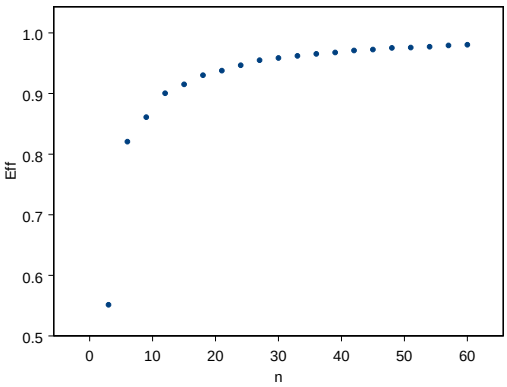
For two competing linear regression models, the efficiency of sequential designs is measured by the ratio of non-centrality parameter to that of the T-optimal design, where the non-centrality parameter is the residual sum of squares in the absence of error (see Atkinson and Donev, 1992). Similarly, we can define the efficiency of a sequential design for generalized linear models as the ratio of its deviance to that of the T-optimal design. As before, we denote the mean vectors for the two competing models by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, respectively. Without loss of generality, we assume that the first model is the true model. Suppose that the deviance function arising from fitting the second model is $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$; then the efficiency of a sequential design can be defined as in (5.5). Clearly, a design with T-efficiency close to 1 provides a as nearly powerful lack-of-fit test as the T-optimal design.

$$\text{Eff}_T = \frac{D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2) \text{ of the sequential design}}{D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2) \text{ of the T-optimal design}}. \tag{5.5}$$

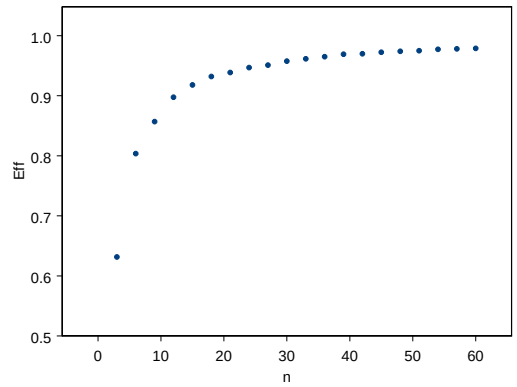
For the T-optimal design, it is easy to extract $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$. However, it is not an easy task for a sequential design because the sequential design is random and so is the corresponding $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$. As in Chapter 4, we need to use simulation to approximate the mean of $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$ for the sequential design. Basically, we can simulate many sequential designs for discrimination between two competing models and for each of these sequential designs we compute the deviance function. This will give us a sample of the random variable $D(\boldsymbol{\mu}_1, \hat{\boldsymbol{\mu}}_2)$ and, if the sample is large enough, the sample mean provides a good approximation to the population mean.

We shall evaluate sequential designs using simulation as discussed above. We will consider four situations: (1) the true model is assumed to be $\lambda = exp(3.9 - 0.7x)$ and the rival model is also a first-order model with the square root link; (2) the true model is $\lambda = (7 - 2x)^2$ and the rival link is assumed to be the log function; (3) the true model is $\lambda = 3.9 - 0.7x$ and the rival link is the identity link; (4) the true model is $\lambda = 45 - 15x$ and the rival link is assumed to be the log function. In all the situations, we assume the design space is restricted by $q \geq 0.2$ and the three-point equally spaced design with one replicate at each point is used for initial experiments. As we have see in Chapter 4, sequential designs are very robust to initial parameter misspecifications so we do not systematically consider parameter misspecifications here. However, we do assume some arbitrary misspecifications in all the initial designs. Plots of efficiency versus simulation size like Figure 4.2 are used to determine a reasonable simulation size. We have found that the simulated efficiency converges after about 40 iterations. In all the subsequent simulations, 100 iterations are used.

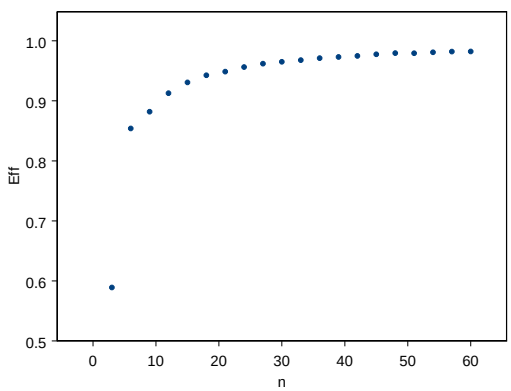
Figure 5.4 plots the simulated efficiency versus experiment size for the four situations



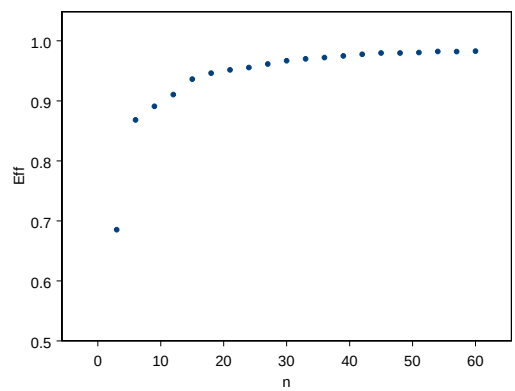
(a) Log (true link) and Square Root



(b) Square Root (true link) and Log



(c) Log (true link) and Identity



(d) Identity (true link) and Log

Figure 5.4: Efficiency of Sequential Designs versus Experiment Size for Discrimination between Two Rival Link Functions. The design space is restricted by $q \geq 0.2$ in all the plots. The true link and the rival link are specified in each plot; for instance, the true link is the log function and the rival link function is the square root in (a). The efficiency of all the sequential design converges to 1 for large experiments.

discussed above. From the plots we can see that all the sequential designs are very efficient for large experiments. More specifically, the T-efficiency is greater than 95% for experiments of 30 runs in all the plots and it gets higher for larger experiments. The fact that T-efficiency converges to 1 for large sample size implies that the sequential design converges to the T-optimal design after enough runs.

5.5 Summary

In this chapter, we have discussed T-optimal designs for discrimination between two competing link functions for Poisson data models. Though the log link is widely used for Poisson data models, the square root link and the identity link also provide very good fits in many applications. Thus, we have focused on discrimination between log link and the other two competing link functions. For brevity, we just considered one-toxicant first-order models. However, all the methodology discussed in this chapter can be applied to more complicated models like the two-toxicant interaction model.

We first studied locally T-optimal designs as they provide a basis to evaluate other designs. All the locally T-optimal designs we have considered are three-point designs with two points located at the two boundaries of the design space. Also, all the optimal designs allocate about half of the sample size to the middle point because the middle point is where the log function deviates from the other two link functions. General equivalence theory is adopted to verify optimality of all the locally optimal designs.

To relax the assumption of the locally T-optimal design that the true model is known, we extend Atkinson and Fedorov's idea of sequential experimentation to Poisson data models. In a sequential design, we update our estimates of the model, both the link and the parameters, after each observation is made and the next design point is found to maximize the deviance function evaluated at the updated estimates. Efficiency study indicates that the sequential design converges to the T-optimal design after enough runs.

Chapter 6

Future Research

The primary goal of this research was to develop efficient and robust experimental designs for Poisson regression models in toxicity studies. Throughout this research, we have made an important assumption that increasing any design variable causes a decrease in the mean response, which is generally true in toxicity studies. However, in industrial applications, positive relationship between a design variable and the expected response is also very common. In Van Mullekom and Myers (2001), a design variable that has negative effect on the response is termed *toxicant* and a design variable with positive effect is named *stimulant*. Accordingly, models are classified into three categories: toxicant model, stimulant model and toxicant-stimulant model which involves at least one toxicant and one stimulant. Obviously, all the models considered in this research belong to the category of toxicant model. However, Van Mullekom and Myers (2001) pointed out that the same optimal experiment designs, in terms of ED levels, for toxicant models can be applied to both stimulant models and toxicant-stimulant models with generalizing the concept ED. For a toxicant, the ED level at a point is the ratio of the mean response at this point to the maximum mean response. For a stimulant, we also define the ED level as the ratio of the mean response at a particular point to the maximum mean response. The difference is that the maximum mean response occurs at the control point for a toxicant while it occurs at the upper limit of the design variable for a stimulant. With this generalization of ED, all the designs we have discussed apply to models of any type. Thus, results of this research are not restricted to toxicity studies.

The work described in this research has potential extensions in several directions and they are listed as follows:

1. The D-optimality criterion, which is a "estimation criterion", is the focus in this research. In many applications, quality of prediction is another important issue and Q-optimality is commonly chosen as a "prediction criterion" so the Q-optimal design for the Poisson regression model makes another topic of practical importance. In this

work, D_s-optimal designs are studied as it is felt that high quality estimates of ED's may result from high quality estimates of specific model parameters. However, the relationship between ED estimation and the corresponding D_s-optimal design is not so clear-cut. The optimality criterion that directly minimizes the length of the confidence interval of a particular ED, like the F-optimality criterion, is more straightforward. Comparison between D_s-optimal designs and other optimality criteria that address ED estimation is also an interesting topic.

2. For the logistic regression model, Chaloner and Larntz (1989) found that the number of support points of the Bayesian optimal design increases as the uncertainty in the prior distribution increases. This is a desirable property since designs of more points are more robust to parameter misspecifications. However, for the Poisson regression model, Van Mullekom and Myers (2001) found that the Bayesian design is a saturated design regardless of uncertainty in the priors. The reason why the number of support points of the Bayesian optimal design fails to vary according to priors need to be further examined for the Poisson regression model.

3. In Chapter 5, we considered experimental designs for discrimination between two rival link functions. This work could be extended at least in two directions. Firstly, both the T-optimal design and the sequential procedure could be extended to discriminate between more than two rival link functions. Secondly, using a generalized link function and designing optimal experiments for estimation of the link parameter is another plausible method to deal with link selection. For the Poisson regression model, the following generalized link could be used:

$$\mathbf{x}'\boldsymbol{\beta} = \frac{\lambda^{\alpha} - 1}{\alpha}.$$

When α takes 0, 0.5 and 1, the above generalized link function corresponds to the log link, square root link and identity link, respectively. The D_s-optimal design for estimating α and the D-optimal design for estimating both α and $\boldsymbol{\beta}$ can be developed.

Bibliography

[1] Abdlebasit, K.M. and Plackett, R.L. (1983), "Experimental Design for Binary Data", *Journal of American Statistical Association*, **78**, 90-98.

[2] Atkinson, A.C., Chaloner, K., Herzberg, A.M. and Juritz, J. (1993), "Optimum Experimental Designs for Properties of a Compartmental Model", *Biometrics*, **49**, 325-337.

[3] Atkinson, A.C. and Donev, A. N. (1992), *Optimum Experimental Designs*, Oxford Univ. Press, Oxford.

[4] Atkinson, A.C. and Fedorov, V.V. (1975), "The Design of Experiments for Discriminating between Two Rival Models", *Biometrika*, **62**, 57-70.

[5] Atkinson, A.C. and Haines, L.M. (1996), "Designs for Nonlinear and Generalized Linear Models," In: S. Ghosh and C.R. Rao, eds., *Handbook of Statistics*, **13**, Elsevier Science B.V, 437-475.

[6] Box, G.E.P. and Lucas, H.L. (1959), "Design of Experiments in Nonlinear Situations", *Biometrics*, **46**, 77-90.

[7] Brunden, M.N., Vidmar, T. J., and McKean, J.W. (1988), *Drug Interaction and Lethality Analysis*, CRC Press, Boca Raton, FL.

[8] Chernoff, H. (1953), "Locally Optimal Designs for Estimating Parameters", *Annals of Mathematical Statistics*, **24**, 586-602.

[9] Chiachierrini, L.M., (1996), *Experimental Design Issues in Impaired Reproduction Studies*, PhD Thesis, Virginia Tech.

[10] Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian Design Applied to Logistic Regression Experiments," *Journal of Statistical Planning and Inference*, **21**, 191-208.

[11] Chaloner, K. and Verdinelli, I. (1995), "Bayesian Experimental Design: A Review," *Statistical Science*, **10**, 273-304.

[12] Dumouchel, W. and Jones, B. (1994), "A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model," *Technometrics*, **36**, 37-47.

[13] Fedorov, V.V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

[14] Fieller, E.C. (1944), "A Fundamental Formula in the Statistics of Biological Assay and Some Applications," *Quarterly Journal of Pharmacy and Pharmacology*, **17**, 117-123.

[15] Finney, D.J. (1971), *Probit Analysis*, Cambridge University Press, Cambridge.

[16] Ford, I. and Silvey, S.D. (1980), "A Sequentially Constructed Design for Estimating a Nonlinear Parametric Function," *Biometrika*, **67**, 381-388.

[17] Frome, E.L., DuFrain, R.J. and Brown, C.C. (1986), "Maximum Likelihood Estimation for Cytogenetic Dose-Response Curves," *Biometrics*, **42**, 73-84.

[18] Heise, M.A. and Myers, R.H. (1996), "Optimal Designs for Bivariate Logistic Regression," *Biometrics*, **52**, 613-624.

[19] Jia, Y. (1996), *Optimal Experimental Designs for Two-Variable Logistic Regression Models*, PhD Thesis, Virginia Tech.

[20] Kalish, L.A. and Rosenberger J.L. (1978), "Optimal Design for the Estimations of the Logistic Function", Technical Report, **33**, Department of Statistics, Pennsylvania State University.

[21] Kalish, L.A. (1990), "Efficient Design for Estimation of Median Lethal Dose and Quantal Dose-Response Curves," *Biometrics*, **46**, 737-748.

[22] Kiefer, J. and Wolfowitz, J. (1959), "Optimum Designs in Regression Problems", *Annals of Mathematical Statistics*, **30**, 271-294.

[23] Kiefer, J. and Wolfowitz, J. (1960), "The Equivalence of Two Extremum Problems", *Canadian Journal of Mathematics*, **12**, 363-366.

[24] Lehmann, E.L. (1983), *Theory of Point Estimation*, John Wiley & Sons, New York.

[25] Lestinger, W.C. (1995), *Optimal One and Two-Stage Designs for the Logistic Regression Model*, PhD Thesis, Virginia Tech.

[26] Lin, H., Myers, R.H. and Ye, K. (2000), "Bayesian Two-Stage Optimal Design for Mixture Models," *Journal of Statistical Computation and Simulation*, **66**, 209-231.

[27] McCullagh, P. and Nelder, J.A. (1987), *Generalized Linear Models*, Chapman and Hall, London.

[28] McLeish, D.L. and Tosh, D. (1990), "Sequential Designs in Bioassay," *Biometrics*, **46**, 103-116.

[29] Minkin, S. (1987), "Optimal Designs for Binary Data", *Journal of American Statistical Association*, **82**, 1098-1103.

[30] Minkin, S. (1991), "A Statistical Model for In-Vitro Assessment of Patient Sensitivity to Cytotoxic Drugs," *Biometrics*, **47**, 1581-1591.

[31] Minkin, S. (1993), "Experimental Design for Clonogenic Assays in Chemotherapy," *Journal of the American Statistical Association*, **88**, 410-420.

[32] Minkin, S. and Kiran K. (1999), "Likelihood-Based Experimental Designs for Estimation of ED_{50}," *Biometrics*, **55**, 1030-1037.

[33] Müller, W.G. and Ponce de Leon, A.C. (1996), "Discrimination between Two Binary Data Models: Sequentially Designed Experiments," *Journal of Statistical Computation and Simulation*, **55**, 87-100.

[34] Myers, R.H. and Montgomery, D.C. (1995), *Response Surface Methodology*, John Wiley & Sons, New York.

[35] Myers, R.H., Montgomery, D.C., and Vining, G.G. (2002), *Generalized Linear Models*, John Wiley & Sons, New York.

[36] Myers, W.R., Myers, R.H. and Carter, W.H. Jr. (1994), "Some Alphabetic Optimal Designs for the Logistic Regression Model", *Journal of Statistical Planning and Inference*, **42**, 57-77.

[37] Myers, W.R., Myers, R.H., Carter, W.H. Jr., and White, K.L. (1996), "Two-stage Designs for the Logistic Regression Model in a Single Agent Bioassay", *Journal of Biopharmaceutial Statistics*, **6**, 283-301.

[38] Nelder, J.A. and Mead, R. (1965), "A Simplex Method for Function Minimization," *Computer Journal*, **7**, 308-313.

[39] Oris, J.T. and Bailer, A.J. (1993), "Statistical Analysis of the Ceriodaphnia Toxicity Test: Sample Size Determination for Reproductive Effects," *Environmental Toxicology and Chemistry*, **12**, 85-90.

[40] Ponce de Leon, A.C. and Atkinson, A.C. (1991), "Optimum Experimental Design for Discriminating between Two Rival Models in the Presence of Prior Information", *Biometrika*, **78**, 601-608.

[41] Schwartz, P.F., Gennings, C., Teuschler, L.K. and Fariss, M.W. (2001), "Optimizing the Precision of Toxicity Threshold Estimation Using a Two-Stage Experimental Design," *Journal of Agricultural, Biological, and Environmental Statistics*, **6**, 409-428.

[42] Silvey, D. (1980), *Optimal Design*. Chapman and Hall, London.

[43] Sitter, R.R. (1992), "Robust Designs for Binary Data", *Biometrics*, **48**, 1145-1155.

[44] Sitter, R.R. and Wu, C.F.J. (1993), "On the Accuracy of Fieller Intervals for Binary Data," *Journal of the American Statistical Association*, **88**, 1021-1025.

[45] Sitter, R.R. and Wu, C.F.J. (1999), "Two-stage Design of Quantal Response Studies", *Biometrics*, **55**, 396-402.

[46] Van Mullekom and Myers (2001), "Optimal Experimental Designs for Poisson Impaired Reproduction", Technical Report, **01-1**, Department of Statistics, Virginia Tech.

[47] Wahrendorf, J. and Zentgraf, R. (1981), "Optimal Designs for the Analysis of Interactive Effects of Two Carcinogens or Other Toxicants," *Biometrics*, **37**, 45-54.

[48] Williams, D.A. (1986), "Interval Estimation of the Median Lethal Dose," *Biometrics*, **42**, 641-645.

[49] Wu, C.F.J. (1985), "Efficient Sequential Designs with Binary Data", *Journal of the American Statistical Association*, **80**, 974-984.

Appendix A

Proofs and Derivations

The following notation will be used throughout this appendix:

- λ_c denotes the mean response at the control point;
- λ_i denotes the response at the ith design point;
- n denotes the sample size, or total experimental runs;
- p_i denotes the proportion of sample size at the ith design point;
- $q_i = \lambda_i/\lambda_c$ denotes the survival proportion, or MED level, at the ith point;
- $q_{1i} = exp(\beta_1 x_{1i})$ denotes the IED level of the first toxicant at the ith point;
- $q_{2i} = exp(\beta_2 x_{2i})$ denotes the IED level of the second toxicant at the ith point.

A.1 Proof of Lemma 2.1

Proof. First, we assume that the D-optimal design for model (1.10) is a saturated design like most D-optimal designs. This assumption is verified to be true via equivalence theory in Chapter 3. For a saturated design for model (1.10), the information matrix is given as follows:

$$I(X, \boldsymbol{\beta}) = \begin{pmatrix} \sum_{i=1}^{3} np_i\lambda_i & \sum_{i=1}^{3} np_i\lambda_i x_i & \sum_{i=1}^{3} np_i\lambda_i x_i^2 \\ \sum_{i=1}^{3} np_i\lambda_i x_i & \sum_{i=1}^{3} np_i\lambda_i x_i^2 & \sum_{i=1}^{3} np_i\lambda_i x_i^3 \\ \sum_{i=1}^{3} np_i\lambda_i x_i^2 & \sum_{i=1}^{3} np_i\lambda_i x_i^3 & \sum_{i=1}^{3} np_i\lambda_i x_i^4 \end{pmatrix}$$

It can be shown that:

$$
\begin{aligned}
|I(X,\boldsymbol{\beta})| &= \left(-\frac{n\lambda_c}{4\beta_{11}}\right)^3 \begin{vmatrix} \sum_{i=1}^{3} p_i q_i & \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 \\ \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 \\ \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 & \sum_{i=1}^{3} p_i q_i Q_i^4 \end{vmatrix} \\
&\propto p_1 p_2 p_3 q_1 q_2 q_3 \left((Q_1 - Q_2)(Q_1 - Q_3)(Q_2 - Q_3)\right)^2
\end{aligned}
$$

where $Q_i = -\sqrt{-r} + \sqrt{-r - 4\ln(q_i)}$ and $r = \beta_1^2/\beta_{11}$ for $i = 1, 2, 3$.

We know that the D-optimal design maximizes $|I(X,\boldsymbol{\beta})|$. The above equation indicates that maximization of $|I(X,\boldsymbol{\beta})|$ with respect to p_i and q_i only depends on $r = \beta_1^2/\beta_{11}$. In addition, it can be shown that maximization of the determinant requires that $p_1 = p_2 = p_3 = 1/3$. Thus, the D-optimal design for model (1.10), in terms of ED levels, depends on the parameters only through $r = \beta_1^2/\beta_{11}$. Also, the D-optimal design is an equal-allocation design. □

A.2 Proof of Lemma 2.2

Proof. As before, we assume that the D-optimal design for model (1.12) is a saturated design like most D-optimal designs, which is verified to be true via equivalence theory in Chapter 3. For a saturated design for model (1.12), the information matrix is given as follows:

$$
I(X,\boldsymbol{\beta}) = \begin{pmatrix} \sum_{i=1}^{4} np_i\lambda_i & \sum_{i=1}^{4} np_i\lambda_i x_{1i} & \sum_{i=1}^{4} np_i\lambda_i x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i} \\ \sum_{i=1}^{4} np_i\lambda_i x_{1i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}^2 & \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}^2 x_{2i} \\ \sum_{i=1}^{4} np_i\lambda_i x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{2i}^2 & \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i}^2 \\ \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}^2 x_{2i} & \sum_{i=1}^{4} np_i\lambda_i x_{1i}x_{2i}^2 & \sum_{i=1}^{4} np_i\lambda_i x_{1i}^2 x_{2i}^2 \end{pmatrix}.
$$

It can be shown, by expanding the determinant, that:

$$
|I(X,\boldsymbol{\beta})| = \left(\frac{n\lambda_c}{\beta_1\beta_2}\right)^4 |A|
$$

where

$$A = \begin{pmatrix} \sum_{i=1}^{4} d_i & \sum_{i=1}^{4} d_i a_i & \sum_{i=1}^{4} d_i b_i & \sum_{i=1}^{4} d_i a_i b_i \\ \sum_{i=1}^{4} d_i a_i & \sum_{i=1}^{4} d_i a_i^2 & \sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i a_i^2 b_i \\ \sum_{i=1}^{4} d_i b_i & \sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i b_i^2 & \sum_{i=1}^{4} d_i a_i b_i^2 \\ \sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i a_i^2 b_i & \sum_{i=1}^{4} d_i a_i b_i^2 & \sum_{i=1}^{4} d_i a_i^2 b_i^2 \end{pmatrix}$$

In the above matrix, $r = \beta_{12}/\beta_1\beta_2, d_i = p_i q_{1i} q_{2i} exp(r \ln(q_{1i}) \ln(q_{2i})), a_i = \ln(q_{1i}), b_i = \ln(q_{2i})$, where $i = 1, 2, 3, 4$.

Expending the determinant of A and collecting terms give that:

$$|I(X, \boldsymbol{\beta})| = \left(\frac{n\lambda_c}{\beta_1\beta_2}\right)^4 |A| \propto p_1 p_2 p_3 p_4 f(r, q_{1i}, q_{2i})$$

where $f(r, q_{1i}, q_{2i}) = \left(\prod_{i=1}^{4} exp(a_i + b_i + r a_i b_i)\right)(a_1a_3b_1b_2 - a_2a_3b_1b_2 - a_1a_4b_1b_2 + a_2a_4b_1b_2 - a_1a_2b_1b_3 + a_2a_3b_1b_3 + a_1a_4b_1b_3 - a_3a_4b_1b_3 + a_1a_2b_2b_3 - a_1a_3b_2b_3 - a_2a_4b_2b_3 + a_3a_4b_2b_3 + a_1a_2b_1b_4 - a_1a_3b_1b_4 - a_2a_4b_1b_4 + a_3a_4b_1b_4 - a_1a_2b_2b_4 + a_2a_3b_2b_4 + a_1a_4b_2b_4 - a_3a_4b_2b_4 + a_1a_3b_3b_4 - a_2a_3b_3b_4 - a_1a_4b_3b_4 + a_2a_4b_3b_4)^2$, $a_i = \ln(q_{1i})$ and $b_i = \ln(q_{2i})$ for $i = 1, 2, 3, 4$.

We know that D-optimal design maximizes $|I(X, \boldsymbol{\beta})|$. From the above equation, we can see that the maximization of $|I(X, \boldsymbol{\beta})|$ with respect to q_{1i} and q_{2i} depends on the parameters only through $r = \beta_{12}/\beta_1\beta_2$. Further, it can be shown that the product $p_1p_2p_3p_4$ achieves its maximum when $p_1 = p_2 = p_3 = p_4 = 1/4$, which implies that the D-optimal design is an equal-allocation design. □

A.3 Proof of Lemma 2.3 and Lemma 2.4

Proof. The D_s-optimal design of interest focus on estimation of all the parameters except the intercept term; accordingly, the vector of model parameters is partitioned as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, where $\boldsymbol{\beta}_1$ is simply the intercept term β_0. Similarly, the Fisher information matrix is partitioned as

$$I(X, \boldsymbol{\beta}) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12}' & I_{22} \end{pmatrix}$$

where $I_{11} = I[1,1]$ is just a scalar corresponding β_0.

We know that the $\mathrm{D_s}$-optimality criterion function is $\phi(I) = |I_{22} - I'_{12}I_{11}^{-1}I_{12}|$. Since $I_{11} = I[1,1]$ is a scalar, we have

$$\begin{aligned}
\phi(I) &= |I_{22} - I'_{12}I_{11}^{-1}I_{12}| \\
&= \left| I_{22} - \frac{I'_{12}I_{12}}{I_{11}} \right| \\
&= \frac{|I_{11}I_{22} - I'_{12}I_{12}|}{|I_{11}|} \\
&= \frac{|I|}{|I_{11}|}
\end{aligned}$$

From the proof of Lemma 2.1, we know that, for model (1.10),

$$|I_{11}| = n\lambda_c \sum_{i=1}^{3} p_i q_i$$

and

$$\begin{aligned}
|I| &= \left(-\frac{n\lambda_c}{4\beta_{11}}\right)^3 |A| \\
&= \left(-\frac{n\lambda_c}{4\beta_{11}}\right)^3 \begin{vmatrix} \sum_{i=1}^{3} p_i q_i & \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 \\ \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 \\ \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 & \sum_{i=1}^{3} p_i q_i Q_i^4 \end{vmatrix}
\end{aligned}$$

where $Q_i = -\sqrt{-r} + \sqrt{-r - 4\ln(q_i)}$ and $r = \beta_1^2/\beta_{11}$ for $i = 1, 2, 3$. Thus, we have

$$\begin{aligned}
\phi(I) &= \frac{|I|}{|I_{11}|} \\
&= \frac{(n\lambda_c)^2}{(-4\beta_{11})^3} |A| / \sum_{i=1}^{3} p_i q_i \\
&\propto |A| / \sum_{i=1}^{3} p_i q_i
\end{aligned}$$

Obviously, the D_s-optimal design which maximizes the above criterion function $\phi(I)$ only depends on $r = \beta_1^2/\beta_{11}$ as stated in Lemma 2.3.

Similarly, for model (1.12), we have

$$\begin{aligned}
\phi(I) &= \frac{|I|}{|I_{11}|} \\
&= \left(-\frac{(n\lambda_c)^3}{(\beta_1\beta_2)^4}\right)^2 |B| \Big/ \sum_{i=1}^{4} d_i \\
&\propto |B| \Big/ \sum_{i=1}^{4} d_i
\end{aligned}$$

where

$$B = \begin{pmatrix}
\sum_{i=1}^{4} d_i & \sum_{i=1}^{4} d_i a_i & \sum_{i=1}^{4} d_i b_i & \sum_{i=1}^{4} d_i a_i b_i \\
\sum_{i=1}^{4} d_i a_i & \sum_{i=1}^{4} d_i a_i^2 & \sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i a_i^2 b_i \\
\sum_{i=1}^{4} d_i b_i & \sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i b_i^2 & \sum_{i=1}^{4} d_i a_i b_i^2 \\
\sum_{i=1}^{4} d_i a_i b_i & \sum_{i=1}^{4} d_i a_i^2 b_i & \sum_{i=1}^{4} d_i a_i b_i^2 & \sum_{i=1}^{4} d_i a_i^2 b_i^2
\end{pmatrix}$$

In the above matrix, $r = \beta_{12}/\beta_1\beta_2, d_i = p_i q_{1i} q_{2i} exp(r \ln(q_{1i}) \ln(q_{2i})), a_i = \ln(q_{1i}), b_i = \ln(q_{2i})$, where $i = 1, 2, 3, 4$.

Clearly, the D_s-optimal design which maximizes the above criterion function $\phi(I)$ only depends on $r = \beta_{12}/\beta_1\beta_2$ as stated in Lemma 2.4. □

A.4 Proof of Lemma 3.1

Proof. For model (1.10), by definition,

$$M(\eta^*, \boldsymbol{\beta}) = \begin{pmatrix}
\sum_{i=1}^{3} p_i\lambda_i & \sum_{i=1}^{3} p_i\lambda_i x_i & \sum_{i=1}^{3} p_i\lambda_i x_i^2 \\
\sum_{i=1}^{3} p_i\lambda_i x_i & \sum_{i=1}^{3} p_i\lambda_i x_i^2 & \sum_{i=1}^{3} p_i\lambda_i x_i^3 \\
\sum_{i=1}^{3} p_i\lambda_i x_i^2 & \sum_{i=1}^{3} p_i\lambda_i x_i^3 & \sum_{i=1}^{3} p_i\lambda_i x_i^4
\end{pmatrix}$$

and

$$J(\mathbf{x}, \boldsymbol{\beta}) = \begin{pmatrix} \lambda & \lambda x & \lambda x^2 \\ \lambda x & \lambda x^2 & \lambda x^3 \\ \lambda x^2 & \lambda x^3 & \lambda x^4 \end{pmatrix}.$$

For $\phi(M(\eta, \boldsymbol{\beta})) = \log(|M(\eta, \boldsymbol{\beta})|$, by definition,

$$\begin{aligned} F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}[\phi\{(1-\varepsilon)M(\eta^*, \boldsymbol{\beta}) + \varepsilon J(\mathbf{x}, \boldsymbol{\beta})\} - \phi(M(\eta^*, \boldsymbol{\beta}))] \\ &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left[\log\left(\frac{|M(\eta^*, \boldsymbol{\beta}) + \varepsilon(J(\mathbf{x}, \boldsymbol{\beta}) - M(\eta^*, \boldsymbol{\beta}))|}{|M(\eta^*, \boldsymbol{\beta})|}\right)\right]. \end{aligned}$$

It can be shown that:

$$|M(\eta^*, \boldsymbol{\beta}) + \varepsilon(J(\mathbf{x}, \boldsymbol{\beta}) - M(\eta^*, \boldsymbol{\beta}))| = \left(-\frac{\lambda_c}{4\beta_{11}}\right)^3 |A_1|$$

and

$$|M(\eta^*, \boldsymbol{\beta})| = \left(-\frac{\lambda_c}{4\beta_{11}}\right)^3 |A_2|.$$

To give A_1 and A_2, we need to define the following quantities:

- $r = \beta_1^2/\beta_{11}, \gamma = 1 - \varepsilon$;
- corresponding to $M(\eta^*, \boldsymbol{\beta})$, $Q_i = -\sqrt{-r} + \sqrt{-r - 4\ln(q_i)}$ for $i = 1, 2, 3$;
- corresponding to $J(\mathbf{x}, \boldsymbol{\beta})$, $Q = -\sqrt{-r} + \sqrt{-r - 4\ln(q)}$.

Given the above notation, we have

$$A_1 = \begin{pmatrix} \gamma \sum_{i=1}^{3} p_i q_i + \varepsilon q & \gamma \sum_{i=1}^{3} p_i q_i Q_i + \varepsilon q Q & \gamma \sum_{i=1}^{3} p_i q_i Q_i^2 + \varepsilon q Q^2 \\ \gamma \sum_{i=1}^{3} p_i q_i Q_i + \varepsilon q Q & \gamma \sum_{i=1}^{3} p_i q_i Q_i^2 + \varepsilon q Q^2 & \gamma \sum_{i=1}^{3} p_i q_i Q_i^3 + \varepsilon q Q^3 \\ \gamma \sum_{i=1}^{3} p_i q_i Q_i^2 + \varepsilon q Q^2 & \gamma \sum_{i=1}^{3} p_i q_i Q_i^3 + \varepsilon q Q^3 & \gamma \sum_{i=1}^{3} p_i q_i Q_i^4 + \varepsilon q Q^4 \end{pmatrix}$$

and

$$A_2 = \begin{pmatrix} \sum_{i=1}^{3} p_i q_i & \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 \\ \sum_{i=1}^{3} p_i q_i Q_i & \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 \\ \sum_{i=1}^{3} p_i q_i Q_i^2 & \sum_{i=1}^{3} p_i q_i Q_i^3 & \sum_{i=1}^{3} p_i q_i Q_i^4 \end{pmatrix}.$$

Since A_1 and A_2, in terms of q, only depend on r, we can conclude that, for model (1.10), the *Fréchet derivative* $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_1^2/\beta_{11}$.

□

A.5 Proof of Lemma 3.2

Proof. For model (1.12), by definition,

$$M(\eta^*, \boldsymbol{\beta}) = \begin{pmatrix} \sum_{i=1}^{4} p_i\lambda_i & \sum_{i=1}^{4} p_i\lambda_i x_{1i} & \sum_{i=1}^{4} p_i\lambda_i x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i} \\ \sum_{i=1}^{4} p_i\lambda_i x_{1i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}^2 & \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}^2x_{2i} \\ \sum_{i=1}^{4} p_i\lambda_i x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{2i}^2 & \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i}^2 \\ \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}^2x_{2i} & \sum_{i=1}^{4} p_i\lambda_i x_{1i}x_{2i}^2 & \sum_{i=1}^{4} p_i\lambda_i x_{1i}^2x_{2i}^2 \end{pmatrix}$$

and

$$J(\mathbf{x}, \boldsymbol{\beta}) = \begin{pmatrix} \lambda & \lambda x_1 & \lambda x_2 & \lambda x_1x_2 \\ \lambda x_1 & \lambda x_1^2 & \lambda x_1x_2 & \lambda x_1^2x_2 \\ \lambda x_2 & \lambda x_1x_2 & \lambda x_2^2 & \lambda x_1x_2^2 \\ \lambda x_1x_2 & \lambda x_1^2x_2 & \lambda x_1x_2^2 & \lambda x_1^2x_2^2 \end{pmatrix}.$$

For $\phi(M(\eta, \boldsymbol{\beta})) = \log(|M(\eta, \boldsymbol{\beta})|$, by definition,

$$\begin{aligned} F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}[\phi\{(1-\varepsilon)M(\eta^*, \boldsymbol{\beta}) + \varepsilon J(\mathbf{x}, \boldsymbol{\beta})\} - \phi(M(\eta^*, \boldsymbol{\beta}))] \\ &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left[\log\left(\frac{|M(\eta^*, \boldsymbol{\beta}) + \varepsilon(J(\mathbf{x}, \boldsymbol{\beta}) - M(\eta^*, \boldsymbol{\beta}))|}{|M(\eta^*, \boldsymbol{\beta})|}\right)\right]. \end{aligned}$$

It can be shown that:

$$|M(\eta^*, \boldsymbol{\beta}) + \varepsilon(J(\mathbf{x}, \boldsymbol{\beta}) - M(\eta^*, \boldsymbol{\beta}))| = \left(\frac{\lambda_c}{\beta_1\beta_2}\right)^4 |A_1|$$

and

$$|M(\eta^*, \boldsymbol{\beta})| = \left(\frac{\lambda_c}{\beta_1\beta_2}\right)^4 |A_2|.$$

To give A_1 and A_2, we need to define the following quantities:

- $r = \beta_{12}/\beta_1\beta_2, \gamma = 1 - \varepsilon$;
- corresponding to $M(\eta^*, \boldsymbol{\beta})$, $d_i = p_i q_{1i} q_{2i} exp(r \ln(q_{1i}) \ln(q_{2i})), a_i = \ln(q_{1i})$, and $b_i = \ln(q_{2i})$, where $i = 1, 2, 3, 4$;
- corresponding to $J(\mathbf{x}, \boldsymbol{\beta})$, $d = pq_1q_2exp(r \ln(q_1) \ln(q_2)), a = \ln(q_1)$ and $b = \ln(q_2)$.

Given the above notation, we have

$$A_1 = \begin{pmatrix} \gamma \sum_{i=1}^4 d_i + \varepsilon d & \gamma \sum_{i=1}^4 d_i a_i + \varepsilon da & \gamma \sum_{i=1}^4 d_i b_i + \varepsilon db & \gamma \sum_{i=1}^4 d_i a_i b_i + \varepsilon dab \\ \gamma \sum_{i=1}^4 d_i a_i + \varepsilon da & \gamma \sum_{i=1}^4 d_i a_i^2 + \varepsilon da^2 & \gamma \sum_{i=1}^4 d_i a_i b_i + \varepsilon dab & \gamma \sum_{i=1}^4 d_i a_i^2 b_i + \varepsilon da^2 b \\ \gamma \sum_{i=1}^4 d_i b_i + \varepsilon db & \gamma \sum_{i=1}^4 d_i a_i b_i & \gamma \sum_{i=1}^4 d_i b_i^2 + \varepsilon db^2 & \gamma \sum_{i=1}^4 d_i a_i b_i^2 + \varepsilon dab^2 \\ \gamma \sum_{i=1}^4 d_i a_i b_i + \varepsilon dab & \gamma \sum_{i=1}^4 d_i a_i^2 b_i & \gamma \sum_{i=1}^4 d_i a_i b_i^2 + \varepsilon dab^2 & \gamma \sum_{i=1}^4 d_i a_i^2 b_i^2 + \varepsilon da^2 b^2 \end{pmatrix}$$

and

$$A_2 = \begin{pmatrix} \sum_{i=1}^4 d_i & \sum_{i=1}^4 d_i a_i & \sum_{i=1}^4 d_i b_i & \sum_{i=1}^4 d_i a_i b_i \\ \sum_{i=1}^4 d_i a_i & \sum_{i=1}^4 d_i a_i^2 & \sum_{i=1}^4 d_i a_i b_i & \sum_{i=1}^4 d_i a_i^2 b_i \\ \sum_{i=1}^4 d_i b_i & \sum_{i=1}^4 d_i a_i b_i & \sum_{i=1}^4 d_i b_i^2 & \sum_{i=1}^4 d_i a_i b_i^2 \\ \sum_{i=1}^4 d_i a_i b_i & \sum_{i=1}^4 d_i a_i^2 b_i & \sum_{i=1}^4 d_i a_i b_i^2 & \sum_{i=1}^4 d_i a_i^2 b_i^2 \end{pmatrix}.$$

Since A_1 and A_2, in terms of q_1 and q_2, only depend on r, we can conclude that, for model (1.12), the *Fréchet derivative* $F_\phi(M(\eta, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta}))$ depends on $\boldsymbol{\beta}$ only through $r = \beta_{12}/\beta_1\beta_2$. □

A.6 Proof of Lemma 3.3 and Lemma 3.4

Proof. From the discussion in Appendix A.3, we know that the D_s-optimality criterion function is $\phi(M) = |M|/M[1,1]$, or equivalently,

$$\phi(M) = \log\left(\frac{|M|}{M[1,1]}\right).$$

For the above criterion function, by definition, the *Fréchet derivative* of the design η^* is

$$\begin{aligned} F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}[\phi\{(1-\varepsilon)M(\eta^*, \boldsymbol{\beta}) + \varepsilon J(\mathbf{x}, \boldsymbol{\beta})\} - \phi(M(\eta^*, \boldsymbol{\beta}))] \\ &= \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left[\log\left(\frac{|S|}{S[1,1]}\right) - \log\left(\frac{M}{M[1,1]}\right)\right] \end{aligned}$$

where $S = (1-\varepsilon)M(\eta^*, \boldsymbol{\beta}) + \varepsilon J(\mathbf{x}, \boldsymbol{\beta})$.

For model (1.10), M and J are defined in Appendix A.4. It can be shown that

$$\frac{|S|}{S[1,1]} = \frac{(\lambda_c)^2|A_1|}{(-4\beta_{11})^3 q}$$

and

$$\frac{M}{M[1,1]} = \frac{(\lambda_c)^2|A_2|}{(-4\beta_{11})^3 \sum_{i=1}^{3} p_i q_i}$$

where q is the ED level at any design point in the design space, p_i's and q_i's correspond to the design η^*, A_1 and A_2 are defined in Appendix A.4 and both matrices depend on the parameters only through $r = \beta_1^2/\beta_{11}$. Further, the *Fréchet derivative* can be written as

$$F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left[\log\left(|A_1| \sum_{i=1}^{3} p_i q_i\right) - \log\left(q|A_2|\right)\right].$$

Clearly, the above *Fréchet derivative* depends on the parameters only through $r = \beta_1^2/\beta_{11}$ as stated in Lemma 3.3.

Similarly, for model (1.12), we have

$$F_\phi(M(\eta^*, \boldsymbol{\beta}), J(\mathbf{x}, \boldsymbol{\beta})) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\left[\log\left(|A_1| \sum_{i=1}^{4} p_i d_i\right) - \log\left(d|A_2|\right)\right].$$

where d, d_i, A_1 and A_2 are defined in Appendix A.5. Since A_1 and A_2 depend on the parameters only through $r = \beta_{12}/\beta_1\beta_2$, we can see that the above *Fréchet derivative* depends only upon $r = \beta_1^2/\beta_{11}$ as stated in Lemma 3.4. $\square$

Vita

Yanping Wang, son of Jicheng Wang and Shuying Li, was born on December 26, 1974 in Anhui, China. In 1993, he graduated from Fuyang High School, Anhui, China. He attended the University of Science and Technology of China and received a Bachelor of Science degree in Information Management and Decision Science in 1998. Upon graduation, he was awarded the Excellent Graduate Award for the academic excellence. He began graduate school in August 1998 at Virginia Polytechnic Institute and State University in Blacksburg, Virginia. He completed the requirements for a Master of Science degree in Statistics in December 1999, and was awarded the Boyd Harshbarger Award for the academic achievement among the first year graduate students in Statistics.

Yanping Wang is a member of the honor society Phi Kappa Phi. He is also a member of the American Statistical Association.