Experimental Design Issues in Impaired Reproduction Applications

by

Lisa Marie Chiacchierini

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

APPROVED:

Raymond H. Myers, Chairman

Eric P. Smith

Jesse C. Arnold

Marvin Lentner

Robert V. Foutz

December 4, 1996
Blacksburg, VA

Keywords: Impaired reproduction, Response surface design, Poisson regression models

Experimental Design Issues in Impaired Reproduction Applications

by

Lisa Marie Chiacchierini

Raymond H. Myers, Chairman

Department of Statistics

ABSTRACT

Within the realms of biological and medical research, toxicity studies which measure impaired reproduction are becoming more and more common, yet methods for efficiently designing experiments for these studies have received little attention. In this research, response surface design criteria are applied to four models for impaired reproduction data. The important role of control observations in impairment studies is discussed, and for one model, a normal error linear model, a design criterion is introduced for allocating a portion of the sample to the control. Special attention is focused on issues surrounding optimal design of experiments for two of the models, a Poisson exponential model and a Poisson linear model. As most of the optimal designs for these models are obtained via numerical methods rather than directly from criteria, equivalence theory is used to prove analytically that the numerically obtained designs are truly optimal. A further complication associated with designing experiments for Poisson regression is the need to know parameter values in order to implement the optimal designs. Thus, two stage design of experiments is investigated as one solution to this problem. Finally, since researchers frequently do not know the appropriate model for their data a priori, the optimal designs for these two different models are compared, and designs which are robust to model misspecification are highlighted.

Acknowledgments

I would like to extend my sincerest gratitude to Dr. Raymond H. Myers for being an excellent advisor and a good friend. His guidance and encouragement throughout the duration of this research have been invaluable. I would also like to thank Drs. Eric P. Smith, Jesse C. Arnold, Robert V. Foutz and Marvin Lentner for their time and their helpful suggestions.

I am especially grateful for the many close friends I have made at Virginia Tech, who, without exception, have provided me with moral support and added much laughter and enjoyment to my years in graduate school. In particular, I thank Matthew Rotelli for his friendship and encouragement, and his willingness to lend an ear or a hand whenever it was needed. I also would like to express my deepest appreciation to my parents Richard and Kathleen Chiacchierini, my brother, Paul, and my other family members for their love and support.

Finally, I would like to thank the Graduate School and Statistics Department at Virginia Tech and the Air Force Office of Scientific Research for their financial support during this research.

Table of Contents

Page

Page

Page

List of Tables

Table Page

List of Figures

Chapter 1
Introduction

§1.1 Motivation

Over the past few decades, much of the spotlight in scientific research has come to focus on environmental and health issues. In fact, much of the biological research being done today is directed toward two issues of particular concern. In the environmental realm, a good deal of research concentrates on determining how and to what extent pollutants from the modern industrialized world affect the environment and the various forms of life supported by it. Much of the research in the health arena centers on finding cures and treatments to combat the growing number of serious health conditions plaguing the world's populations. At the heart of both of these types of research are toxicity studies, studies which assess the impact of single toxicants and mixtures of toxicants on the environment and the organisms which live there. Due to the serious nature of these studies and the possible repercussions for human life, proper design of experiments and statistical analysis are crucial.

One primary tool of toxicity studies is inverse regression. In most cases, the researcher is interested in determining the effective concentration (EC), the concentration of a single toxicant or the combination of concentrations in a mixture of toxicants, which will produce a particular final response. For example, the goal of one study might be to determine the dose of a chemical which will be lethal to 50% of a population of fruitflys. Another researcher may wish to determine the dose combination of several drugs in a treatment which will cause 80% of the patients in the study to respond positively to the treatment. In both of these examples and in many other situations in which an EC is sought, the data to be analyzed is dichotomous, and the response modeled is the

probability of an observation having one of the two possible values. Logistic regression is an appropriate statistical technique for these cases (Finney, 1978). There are, however, studies involving non-dichotomous data in which ECs are of interest.

Some of the toxicity studies being performed today are performed with the purpose of learning how toxicants impair the reproductive abilities of organisms or cells, and many of these studies involve count data (Oris and Bailer, 1993, Minkin, 1993). The EC in this type of situation is the concentration of a toxicant which causes a particular amount of impairment in reproduction, or a particular percentage decrease in the cell or organism count due to the presence of toxicants. An example of such a study in the statistics department at Virginia Tech was recently funded by the United States Air Force. On occasion, Air Force jets accidentally discharge fuel into aquatic ecosystems. It is therefore a concern of the Air Force to investigate how the toxicity of the chemicals in the jet fuel is affecting aquatic organisms. The particular focus of this study was to determine how various combinations of levels of the toxicants impair population growth through decreased organism reproduction.

Although studies similar in nature to the one conducted by the Air Force are prevalent in biological research, little has been accomplished in terms of efficient design of experiments for this type of study. In response surface methodology, optimal design of experiments offers techniques for designing experiments to most efficiently address specific goals of the experiment. For example, if inverse regression is the aim of a particular study, an optimality criterion exists for selecting a design which best estimates a specific EC. Response surface design is especially useful when the number of available experimental units is limited, as it frequently is in medical research.

In this dissertation, four different models for impaired reproduction data are considered, a linear model with normal errors, a Poisson exponential model, a Poisson linear model and a model known as a Poisson power model. The majority of the work, however, focuses on just two of these, the Poisson exponential and the Poisson linear model. Optimal experimental designs are found for the models based on three design

optimality criteria. One of these criteria addresses estimation of model parameters, one concentrates on prediction with the final fitted model, and one addresses EC estimation. These three criteria will be developed further in Chapter 2. The optimal designs, primarily obtained through numerical methods, are given for the linear model with normal errors in Chapter 3, for the Poisson exponential model in Chapter 4, for the Poisson linear model in Chapter 5, and for the Poisson power model in Chapter 8. Several issues and complications related to designing experiments for impaired reproduction data are identified and discussed as well, including the use of equivalence theory to verify the optimality of some of the designs and robustness issues concerning the designs. These topics appear in Chapters 6 and 7, respectively. Ideas for expanding the work in this dissertation are presented in Chapter 9.

§1.2 The Models

A standard experiment in this type of research consists of a number of experimental units, some of which are treated with particular concentrations of the toxicants and some of which are controls. Controls are experimental units which are treated with no toxicants, and the importance of control observations in impaired reproduction studies will be discussed in the next section. At a specific time after the toxicants have been administered, the organisms or cells in each experimental unit are counted. These counts measure reproduction, and four ways of modeling this data are considered here.

§1.2.1 The Linear Model with Normal Errors

One approach for modeling impaired reproduction data involves creating a "percent impairment" response from the counts as follows:

$$\%\ \text{impairment}_{ij} = \frac{y_{ij}}{\bar{y}_c} \tag{1.2.1}$$

where y_{ij} is the final organism/cell count for experimental unit ij (the experimental unit which is the jth replicate at the ith design point) and $\bar{y}_c$ is the average of the final organism/cell counts at the controls. Frequently, this response is modeled by

$$1 - \frac{y_{ij}}{\bar{y}_c} = \mathbf{x}_i'\beta + \varepsilon_{ij} \tag{1.2.2}$$

where $\mathbf{x}_i$ is a k x 1 vector of dose levels and combinations of dose levels of the toxicants, β is a p x 1 vector of unknown parameters, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. For simplicity, the following equivalent model will actually be used in this research:

$$\frac{y_{ij}}{\bar{y}_c} = \mathbf{x}_i'\beta + \varepsilon_{ij} \tag{1.2.3}$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$. In this model, the "percent of the maximum possible reproduction which occurs" is being modeled rather than the "percent impairment." The models in (1.2.2) and (1.2.3) are equivalent since, for example, modeling 40% impairment is the same as modeling 60% of the maximum possible reproduction actually occurring.

While modeling the "percent impairment" response as a normal random variable with homogeneous variance is done in practice, it is probably not the best way to model impaired reproduction data. It is unlikely that this response actually has a normal distribution since there is a chance that is has nonhomogeneous variance, particularly near the 0 and 1 extremes. In fact, this model may only be appropriate over a small range of the data.

§1.2.2 The Poisson Exponential Model and the Poisson Linear Model

An alternative modeling approach is to model the actual cell or organism counts as Poisson random variables. The remaining three models used in this research are

Poisson regression models. Two of these are created using generalized linear models (GLM) theory (Nelder and McCullagh, 1989).

In GLM theory, models may be formed for any error distribution from an exponential family using link functions to determine the nature of the models. The name generalized linear models comes from the fact that the models all contain the linear function $\mathbf{x}'\beta$, often called the "linear predictor." The density function for a random variable from an exponential family can be written in the form

$$\exp\{r(\phi)[y_i\theta_i - g(\theta_i)] + h(y_i, \phi)\} \tag{1.2.4}$$

where ϕ is a scale parameter and θ_i is a natural location parameter which is related to the mean of the distribution. In GLM, a link function is a function, s, of the mean of the distribution which is modeled by the linear predictor, $\mathbf{x}_i'\beta$. Thus, building a model in GLM begins with the equation

$$s(\mu_i) = \mathbf{x}_i'\beta \tag{1.2.5}$$

where s is some function which is monotonic in $\mathbf{x}_i'\beta$. The actual model for the mean of the distribution is

$$\mu_i = s^{-1}(\mathbf{x}_i'\beta). \tag{1.2.6}$$

Numerous links can be used with the same probability distribution to create many different models. However, each distribution from an exponential family has a "natural" or canonical link function. Recall that in (1.2.4), θ_i, the natural location parameter, is a function of the mean dictated by the density function. The canonical link function for a distribution is

$$s(\mu_i) = \theta_i = \mathbf{x}_i'\beta. \tag{1.2.7}$$

Use of the canonical link has some nice features including a simple form of the Fisher Information Matrix and sufficient statistics for the model parameters which are linear functions of the observations.

A general nonlinear model for the data in an impaired reproduction study can be written as

$$y_{ij} = f(\mathbf{x}_i, \beta) + \varepsilon_{ij}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. The function, $f(\mathbf{x}_i, \beta)$, used to model the mean of the data, λ_i, is determined by a link function. The two generalized linear models for the mean studied in this research are the model derived from the log link function and the model derived from the identity link function.

The joint likelihood function for N Poisson random variables is

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{\,y_i}}{y_i!} \\ &= \prod_{i=1}^{N} e^{-\lambda_i + y_i \ln \lambda_i - \ln y_i!}. \end{aligned} \tag{1.2.8}$$

From the exponential form in (1.2.8), it is clear that the canonical link for the Poisson distribution is the log link. The log link leads to an exponential model for the mean as follows

$$\begin{aligned} &\ln \lambda_i = \mathbf{x}_i'\beta \\ \Rightarrow \quad &\lambda_i = f(\mathbf{x}_i, \beta) = e^{\mathbf{x}_i'\beta} \end{aligned} \tag{1.2.9}$$

for $i = 1, 2, \ldots, n_i$. The exponential model used in this dissertation is the single regressor exponential model,

$$y_{ij} = e^{\beta_0 + \beta_1 x_i} + \varepsilon_{ij} \tag{1.2.10}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$.

Now consider the situation in which the identity link function is used rather than the log link. A linear model is formulated since $\lambda_i = \mathbf{x}_i'\beta$ is already in terms of the mean of the Poisson distribution. The single regressor linear model is

$$y_{ij} = \mathbf{x}_i'\beta + \varepsilon_{ij} \tag{1.2.11}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$.

§1.2.3 The Poisson Power Model

The Poisson power model is not a generalized linear model as it does not contain the linear predictor, $\mathbf{x}_i'\beta$. This single regressor model is

$$y_{ij} = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_{ij} \qquad (1.2.12)$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. This particular model finds use in some impairment applications (See Smith). However, sometimes this model is used to model the "percent impairment" response rather than the organism or cell counts, and, in that case, it is assumed to have homogeneous variance.

§1.3 Control Observations

The concept of impairment in reproduction implies deviation from the reproduction that would have been expected had toxicants not been present in the environment. Therefore, modeling impairment requires knowledge of a population's size in the presence of no toxicants. Consequently, it is necessary that at least one observation in impaired reproduction studies be taken at a control, an experimental unit treated with no toxicants. Without controls, it would be impossible to gauge the amount of impairment which different levels of the chemicals produce. Because control observations play a role of primary importance in studies of this nature, it is necessary to have a method for determining how many control observations are taken. For the Poisson models, the number of observations to be allocated to the control is taken into account through the design optimality criteria. For the linear model with normal errors, however, a separate criterion is needed to allocate observations to the control. Reasons for this will be discussed in Chapter 3.

§1.4 Notes on the Poisson Models

Theoretically, the support for a Poisson random variable is the integers in $[0, \infty)$. For any practical Poisson design problem, however, it must be assumed that there is some finite maximum value that the mean of the random variable in the experiment can take on, as in the impaired reproduction studies. In impaired reproduction studies, the

presence of the control places an upper limit on the mean of the cell counts. Consequently, λ_c, the mean reproduction at the control, is the maximum possible reproduction that can occur in any experimental unit. This implies that, for each of the Poisson models, the expected response at any non-control design point is a fraction of the expected response at the control, or

$$E(y_{ij}) = \lambda_i = q_i\lambda_c$$

where $0 < q_i < 1$. The quantity q_i is the "percent of the maximum reproduction that occurs at design point i" or 1 - (% impairment at design point i). With the models outlined and discussed, a presentation of the design optimality criteria and their development for both linear and nonlinear models is appropriate.

Chapter 2

Design Optimality

§2.1 Introduction

During the 1970's and 1980's, design optimality enjoyed an increase in popularity as the idea of designing experiments for regression situations began to appeal to researchers in a variety of fields (Myers and Montgomery, 1995). The concepts in optimal design, however, have not been as quick to gain a foothold in biological studies as they have been in other areas, such as engineering. This is perhaps partially because the models in biological studies are often nonlinear or have nonhomogeneous variance. The alphabetic optimality criteria used to create optimal designs are more difficult to apply in nonlinear and nonnormal error situations, primarily because some knowledge of the model parameters is often necessary. This is not the case in linear regression with normal errors.

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. Most of the optimality criteria are identified by letters and thus are called alphabetic optimality criteria. Kiefer and Wolfowitz (1959) did some of the earliest work in design optimality for regression models. 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 familiar design criteria like A, F, G, and Q optimality. Many of these design criteria were originally developed for the homogeneous variance linear model, but most have since been adapted for use in nonlinear and nonhomogeneous variance situations as well. In this research, we will concentrate on D, F, and Q, as well

as slope optimality, which is a special case of F for one of the impairment models. The design criteria will be developed for the general model

$$y_{ij} = f(\mathbf{x}_i, \beta) + \varepsilon_{ij} \tag{2.1.1}$$

where i = 1, 2,...,d and d is the number of distinct design points, j = 1, 2,...,n_i and n_i is the number of replicates at ith design point, y_{ij} is the jth response at the ith design point, $f(\mathbf{x}_i,\beta)$ is a function of known form (linear or nonlinear in β), β is a p x 1 vector of unknown parameters, $\mathbf{x}_i$ is a k x 1 vector of regressors, and ε_{ij} is an unknown error with an assumed probability distribution.

When designing experiments for regression models, practitioners can define two regions for the k design variables. One region is the region of operability. This is the portion of $\Re^k$ that is determined by the combinations of levels of the design variables that are physically possible for the experimenter to use. This region is also known as the design space. The other region is the region of interest. This is the portion of $\Re^k$ in which the experimenter is primarily concerned with the response. Usually, it is the region over which the experimenter hopes to predict well with the final fitted model. In many cases the regions of operability and interest are the same. In cases, however, the region of interest is only a part of the region of operability. For D and F-optimality, typically only one of these regions is needed. For Q-optimality, however, both regions must be specified.

§2.2 D-Optimality

The D-optimality criterion, probably the most popular of the design criteria, seeks a design which will estimate β as well as possible. The design which best estimates β is the design which minimizes the generalized variance of the estimate of β in the fitted model. A common method for estimating β in many nonlinear models is maximum likelihood. If $\mathbf{b}$ is the MLE of β, then the asymptotic variance-covariance matrix of $\mathbf{b}$ is

the inverse of the Fisher Information Matrix (Lehmann, 1983). The Fisher Information Matrix is given by

$$\mathbf{I} = -\mathrm{E}\left[\frac{\partial^2 \ln L(\beta)}{\partial\beta'\partial\beta}\right] \tag{2.2.1}$$

where $L(\beta)$ is the joint likelihood function for the data. This is a p x p matrix of second partial derivatives. For nonlinear and nonhomogeneous variance models, these derivatives are likely to depend on β. Since nearly all the design criteria are based upon the information matrix, knowledge of the parameters is frequently necessary to apply the design criteria in nonlinear and nonhomogeneous variance situations. The general D-optimality criterion is

$$\max_{\mathcal{D}} \left|\frac{\mathbf{I}(\mathbf{x}_i,\beta)}{N}\right| \tag{2.2.2}$$

where $\mathcal{D}$ indicates all possible designs over the region of operability and N is the number of available experimental units (Box and Lucas, 1959). Essentially, this criterion minimizes the determinant of $N\mathbf{I}^{-1}$, which in turn, minimizes a $(1-\alpha)100\%$ confidence ellipsoid on the model parameters when the parameter estimates are asymptotically normal (Myers and Montgomery, 1995).

§2.3 Q-Optimality

While D-optimality addresses parameter estimation, the Q-optimality criterion seeks a design which will make prediction with the fitted model as accurate as possible. It accomplishes this by finding a design which minimizes the average prediction variance over the region of interest. In linear normal error situations, the prediction variance is straightforward. In nonlinear or nonhomogeneous variance situations, prediction variance is usually approximated by using a first order Taylor series expansion about the true parameter values. This approximation method, which is equivalent to the Delta Method, produces the following expression for the prediction variance at a point $\mathbf{x}_0$:

$$\text{var}(\hat{y}(\mathbf{x}_0)) = \text{var}(f(\mathbf{x}_0,\hat{\beta})) \approx \mathbf{w}(\mathbf{x}_0)'\mathbf{I}^{-1}\mathbf{w}(\mathbf{x}_0) \tag{2.3.1}$$

where

$$\mathbf{w}(\mathbf{x}_0)' = \left[\frac{\partial f(\mathbf{x}_0,\beta)}{\partial\beta_1},\cdots,\frac{\partial f(\mathbf{x}_0,\beta)}{\partial\beta_p}\right] \tag{2.3.2}$$

The actual Q-optimality criterion is

$$\begin{aligned}&\min_{\mathcal{D}} \frac{N}{K}\int_R \text{var}(f(\mathbf{x}_0,\hat{\beta}))d\mathbf{x}_0 \\ &= \min_{\mathcal{D}} \frac{N}{K}\int_R \mathbf{w}(\mathbf{x}_0)'\mathbf{I}^{-1}\mathbf{w}(\mathbf{x}_0)d\mathbf{x}_0\end{aligned} \tag{2.3.3}$$

where $\mathcal{D}$ is the design space or region of operability, R is the region of interest (the region over which accurate prediction is desired) and K is the volume of the region of interest.

§2.4 F-Optimality and Slope-Optimality

F-optimality addresses estimation of the effective concentration (EC) which yields a particular response. For example, suppose a cancer researcher wishes to design an experiment to best determine what dose of a new drug will allow reproduction of malignent cells to attain only 10% of their maximum possible reproduction over a specified period of time. If the effective concentration can be expressed as a ratio of random variables, a fiducial interval called a Fieller interval can be constructed for the ratio (Fieller, 1944). A design which minimizes the width of this Fieller interval would most accurately estimate the EC (Finney, 1971, 1978).

Unlike many of the other design criteria, F-optimality was originally developed for the logistic regression model and has been applied extensively in logistic design situations (Sitter and Wu, 1993, Letsinger, 1995). The F-optimality criterion depends largely on the parameterization of a particular model, so no general form for the criterion will be given here. This criterion will be developed, however, for the normal error linear

model in Chapter 3, the Poisson exponential model in Chapter 4 and the Poisson linear model in Chapter 5.

The slope optimality criterion is a design criterion which is of interest in some regression problems with only one regressor variable. The criterion is given by

$$\min_{\mathcal{D}} \ \text{var}(b_1)$$

where b_1 is the MLE of β_1. This criterion is of interest mainly in cases in which the slope parameter has a special interpretation. One such case is a Poisson exponential model in which impaired reproduction is the dependent variable. This case will be discussed at length in Chapter 4.

§2.5 Application to the Linear Model with Normal Errors

Recall the linear model

$$y_{ij} = \mathbf{x}_i'\beta + \varepsilon_{ij} \tag{2.5.1}$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$. For this model, $\mathbf{w}(\mathbf{x}_0)' = \mathbf{x}_0'$ and $\mathbf{I} = \mathbf{X}'\mathbf{X}$ where $\mathbf{X}$ is the standard model matrix for normal linear regression. Thus, the D-optimality criterion becomes

$$\max_{\mathcal{D}} \left| \frac{\mathbf{X}'\mathbf{X}}{N} \right|, \tag{2.5.2}$$

and the Q-optimality criterion is

$$\begin{aligned} &\min_{\mathcal{D}} \ \frac{N}{K} \int_R \text{var}(\hat{y}(\mathbf{x}_0))d\mathbf{x}_0 \\ &= \min_{\mathcal{D}} \ \frac{N}{K} \int_R \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 d\mathbf{x}_0. \end{aligned} \tag{2.5.3}$$

Another useful form of the Q-optimality criterion is

$$\min_{\mathcal{D}} \ \text{tr}[N(\mathbf{X}'\mathbf{X})^{-1}]\left[\frac{1}{K}\int_R \mathbf{x}_0\mathbf{x}_0' d\mathbf{x}_0\right]. \tag{2.5.4}$$

One should note that these expressions do *not* depend on the model parameters.

§2.6 Application to Nonlinear and Nonhomogeneous Variance Models

Application of the criteria to nonlinear and nonhomogeneous variance cases is more complex. Among specific nonlinear models, the logistic model for binary data has received the most attention in terms of optimal design of experiments. Within the context of GLM models, the logistic model is created by using a logit link with the Bernoulli distribution. Some optimal design has been considered for binomial data with other links, but the logistic model has been the most common.

The majority of the published work has dealt with a single regressor logistic model. The D-optimal design for this model was determined by Kalish and Rosenberger (1978). Myers, Myers, and Carter (1994) derived Q-optimal designs for the same logistic model, and F-optimality was addressed for quantal data using models formed from several different link functions by Sitter and Wu (1993). Abdelbasit and Plackett (1983), Minkin (1987), Letsinger (1995) and Myers, Myers, Carter, and White (1996) developed multi-stage design procedures for the one variable logistic model.

Recall that three nonhomogeneous variance models are of interest in this research, the Poisson exponential model, the Poisson linear model and the Poisson power model. Two of these, the exponential model and the power model, are nonlinear as well. To apply the D, Q, and F-optimality criteria to these models, the information matrix for each model is needed.

Consider first the exponential model

$$y_{ij} = e^{\beta_0+\beta_1 x_i} + \varepsilon_{ij} \qquad (2.6.1)$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. The Fisher Information Matrix is obtained by taking second partial derivatives of the log of the likelihood function for the data. The joint likelihood for N Poisson random variables is

$$L(\beta) = \prod_{i=1}^{N} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} = \prod_{i=1}^{N} e^{-\lambda_i + y_i \ln \lambda_i - \ln y_i!} \tag{2.6.2}$$

so the log likelihood is given by

$$\ln L(\beta) = -\sum_{i=1}^{N} \lambda_i + \sum_{i=1}^{N} y_i \ln \lambda_i - \sum_{i=1}^{N} \ln(y_i!). \tag{2.6.3}$$

After substituting the exponential model for the mean, the log-likelihood becomes

$$\ln L(\beta) = -\sum_{i=1}^{N} e^{\beta_0 + \beta_1 x_i} + \sum_{i=1}^{N} y_i(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{N} \ln(y_i!), \tag{2.6.4}$$

and the information matrix for this model is

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \lambda_i & \sum_{i=1}^{N} x_i \lambda_i \\ \sum_{i=1}^{N} x_i \lambda_i & \sum_{i=1}^{N} x_i^2 \lambda_i \end{bmatrix}. \tag{2.6.5}$$

Note that the elements of the information matrix contain λ_i, the expected response at design point i. Thus, the expressions for the design criteria will all contain λ_i as well. Consequently, the design criteria will yield the optimal designs in terms of the *expected responses* at the optimal design levels, the $e^{\beta_0+\beta_1 x_i}$'s, *not* the design levels themselves, the x_i's. Unfortunately, to find the x's from the λ's, knowledge of the model parameters is needed. At the very least, initial guesses at the values of the model parameters are necessary, and this complication is what makes optimal design of experiments for nonlinear and nonhomogeneous variance models complex. However, there are techniques for addressing this problem, and these will be discussed in Chapter 7.

Although optimal design for the Poisson exponential model has not received much attention in design literature, F-optimality for this model was recently addressed by Minkin(1993). Minkin, in finding a slope-optimal design for the one regressor log-linear model, actually found the F-optimal design for the Poisson exponential model as well. Details pertaining to this are shown in Chapter 4. In addition, he examined the effects of

poor initial parameter estimates on the optimal design and investigated a Bayesian solution to the problem of having to guess the model parameters.

Consider now the Poisson linear model,

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, \tag{2.6.6}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. The log-likelihood for this model is

$$\ln L(\beta) = -\sum_{i=1}^{N} \beta_0 + \beta_1 x_i + \sum_{i=1}^{N} y_i \ln(\beta_0 + \beta_1 x_i) - \sum_{i=1}^{N} \ln(y_i!), \tag{2.6.7}$$

and the information matrix is

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i}{\lambda_i} \\ \sum_{i=1}^{N} \frac{x_i}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^2}{\lambda_i} \end{bmatrix}. \tag{2.6.8}$$

Like the information matrix for the exponential model, this matrix contains model parameters for which values will be needed to find the optimal design levels.

Finally, consider the power model

$$y_{ij} = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_{ij} \tag{2.6.9}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. The log-likelihood function for this model is

$$\ln L(\beta) = -\sum_{i=1}^{N} \beta_0 + \beta_1 x_i^{\beta_2} + \sum_{i=1}^{N} y_i \ln(\beta_0 + \beta_1 x_i^{\beta_2}) - \sum_{i=1}^{N} \ln(y_i!), \tag{2.6.10}$$

and the information matrix is

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^{\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{\beta_2} \ln(x_i)}{\lambda_i} \\ \sum_{i=1}^{N} \frac{x_i^{\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^{2\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{2\beta_2} \ln(x_i)}{\lambda_i} \\ \sum_{i=1}^{N} \frac{\beta_1 x_i^{\beta_2} \ln(x_i)}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{2\beta_2} \ln(x_i)}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1^2 x_i^{2\beta_2} (\ln(x_i))^2}{\lambda_i} \end{bmatrix}. \tag{2.6.11}$$

Once again, parameter guesses are needed to determine the optimal design levels. For all three of the Poisson models, the optimal designs are obtained when the appropriate

information matrices are substituted into the general design criteria described earlier in this chapter.

In this dissertation, D, Q, and F-optimal designs are eventually presented for three of the four impairment models, the linear model with normal errors, the Poisson exponential model and the Poisson linear model. Only D-optimal designs are obtained for the Poisson power model. The next chapter is dedicated to the designs for the linear model with normal errors. Designs for the Poisson models appear in later chapters.

Chapter 3

Single Toxicant and Mixture Problems Using the Linear Model with Normal Errors

§3.1 Introduction

The model under consideration in this chapter is

$$\frac{y_{ij}}{\bar{y}_c} = \mathbf{x}_i'\beta + \varepsilon_{ij} \qquad (3.1.1)$$

where y_{ij} is the final organism count for the jth replicate at the ith design point, $\bar{y}_c$ is the average of the final organism counts at the control observations, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. This is the only non-Poisson model discussed in this dissertation, and the response represents the "percent of the maximum possible reproduction which occurs" in an experimental unit. D, Q, and F-optimal designs for this model are presented in the following sections with a special section dedicated to allocating observations to the control for this model. While the importance of control observations to an impaired reproduction study was indicated in Chapter 1, the question of how many observations to allocate to the control remains. In this chapter, a design criterion is introduced to determine the appropriate number of control observations for the model in (3.1.1).

§3.2 Optimal Designs Based on the D and Q Criteria

In this chapter, the linear model of interest is considered to be first order or first order with interaction. In response surface applications, for a linear model with normal errors, the levels of the k regressors are typically coded such that for any point in the region of operability,

$$x_{coded} = \frac{x_i - \frac{1}{2}(\max\{x_i\} + \min\{x_i\})}{\frac{1}{2}(\max\{x_i\} - \min\{x_i\})} \qquad (3.2.1)$$

(Myers and Montgomery, 1995). Thus, the smallest value of each regressor receives a coded value of -1, and the largest value receives a coded value of +1. For the remainder of *this* chapter, x will refer to the *coded* value of the regressor, or the coded dose. The D and Q-optimal designs for a first order linear model with normal errors are well-known. In fact, one design satisfies both design criteria. This optimal design is a 2^k factorial with all design points at the -1 and +1 extremes (the corners of the region of operability) and with columns of the **X** matrix mutually orthogonal. This design is optimal regardless of the values of the model parameters. Unlike the D and Q-optimal design, however, the F-optimal design for this model does depend on the model parameters.

§3.3 Designs for Estimation of Effective Concentration (EC)

Let the EC_{100p} be the point in the design space, or dose space, at which (100 x p)% of the maximum possible reproduction occurs. Now consider an experimenter who wishes to estimate the EC_{100p} for some p of interest. Recall that F-optimality addresses this concern. In this section, the F-optimality criterion is applied to three linear models with normal errors.

§3.3.1 The Simple Linear Regression Model

Consider the simple linear regression model

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij} \qquad (3.3.1)$$

where the ε_{ij}'s are $N(0,\sigma^2)$. In this case, the F-optimal design is the design which minimizes the width of a 100(1-α)% Fieller interval on the x which causes reproduction to be (100 x p)% of the reproduction at the control. Suppose the fitted model is

$$y_i = b_0 + b_1 x_i . \qquad (3.3.2)$$

Let k be the response of interest (the expected value of the response at the EC_{100p}, or p), and let x_k be the EC_{100p}. Then

$$E(k - b_0 - b_1 x_k) = 0$$

$$\text{and } \hat{Var}(k - b_0 - b_1 x_k) = s^2(var(b_0) + x^2 var(b_1) + 2cov(b_0, b_1))$$

where $var(b_0)$, $var(b_1)$ and $cov(b_0, b_1)$ are apart from s^2, the model MSE. The Fieller interval is derived from

$$-t_{\frac{\alpha}{2}} \le \frac{k - b_0 - b_1 x_k}{s\sqrt{v_{00} + x_k^2 v_{11} + 2x_k v_{12}}} \le t_{\frac{\alpha}{2}} \qquad (3.3.3)$$

where $t_{\alpha/2}$ is the value from the t-table with the appropriate degrees of freedom and with right tail probability $\alpha/2$, $v_{ii} = var(b_i)$ and $v_{ij} = cov(b_i, b_j)$. Squaring both sides of this expression and rearranging terms yields the quadratic equation

$$(b_1^2 - t^2 s^2 v_{11})x_k^2 - 2(t^2 s^2 v_{01} + b_1(k - b_0))x_k + (k - b_0)^2 - t^2 s^2 v_{00} = 0 . \quad (3.3.4)$$

This quadratic equation in x_k has roots (after simplification)

$$\frac{t^2 s^2 v_{01} + b_1(k - b_0)}{b_1^2 - t^2 s^2 v_{11}}$$

$$\pm \frac{ts\sqrt{t^2 s^2 (v_{01}^2 - v_{00} v_{11}) + 2b_1 v_{01}(k - b_0) + b_1^2 v_{00} + v_{11}(k - b_0)^2}}{b_1^2 - t^2 s^2 v_{11}} . \qquad (3.3.5)$$

These two roots define the endpoints of the Fieller interval. Minimization of the width of the Fieller interval is equivalent to minimization of the squared half-width of the interval. From (3.3.5) the half-width is

$$\frac{ts\sqrt{t^2 s^2 (v_{01}^2 - v_{00} v_{11}) + 2b_1 v_{01}(k - b_0) + b_1^2 v_{00} + v_{11}(k - b_0)^2}}{b_1^2 - t^2 s^2 v_{11}} . \qquad (3.3.6)$$

Dividing both the numerator and denominator by b_1^2 and letting $g = t^2 s^2 v_{11}/b_1^2$ yields

$$\frac{\frac{ts}{b_1}\sqrt{\frac{-t^2 s^2}{b_1^2}(v_{01}^2 - v_{00} v_{11}) + 2v_{01}\frac{(k - b_0)}{b_1} + v_{00} + v_{11}\frac{(k - b_0)^2}{b_1^2}}}{1 - g} . \qquad (3.3.7)$$

Now, $(k-b_0)/b_1$ is $\hat{x}_k$, the point estimate of the EC_{100p}, and for simple linear regression,

$$v_{00} = \frac{1}{N} + \frac{\bar{x}^2}{S_{xx}},\quad v_{11} = \frac{1}{S_{xx}},\quad v_{01} = \frac{-\bar{x}}{S_{xx}},\quad \text{and}\quad (v_{01}{}^2 - v_{00}v_{11}) = \frac{-1}{NS_{xx}} = \frac{-1}{|\mathbf{X}'\mathbf{X}|}$$

where N is the number of available experimental units. With the appropriate substitutions and after extensive simplification, the half-width becomes

$$\frac{\frac{ts}{b_1}\sqrt{\frac{-t^2s^2}{b_1{}^2|\mathbf{X}'\mathbf{X}|} + \frac{1}{N} + \frac{(\bar{x}-\hat{x}_k)^2}{S_{xx}}}}{1-g}. \tag{3.3.8}$$

Squaring the half-width and obtaining a common denominator gives

$$\left(\frac{1}{(1-g)^2}\right)\left(\frac{t^2s^2}{b_1{}^2}\right)\left(\frac{1}{N} + \frac{Nb_1{}^2(\bar{x}-\hat{x}_k)^2 - t^2s^2}{b_1{}^2|\mathbf{X}'\mathbf{X}|}\right). \tag{3.3.9}$$

Now, in most practical designs, observations are allocated to the coded design levels such that $\bar{x} = 0$. Thus, the final expression for the squared half-width of the Fieller interval is

$$\left(\frac{1}{(1-g)^2}\right)\left(\frac{t^2s^2}{b_1{}^2}\right)\left(\frac{1}{N} + \frac{N(k-b_0)^2 - t^2s^2}{b_1{}^2|\mathbf{X}'\mathbf{X}|}\right). \tag{3.3.10}$$

All terms are constant except g and the determinant of $\mathbf{X}'\mathbf{X}$. These two terms depend on design. Clearly, minimization of the last piece of (3.3.10) requires that $|\mathbf{X}'\mathbf{X}|$ be as large as possible. To minimize the first piece, recall that

$$g = \frac{t^2s^2v_{11}}{b_1{}^2} = \frac{t^2s^2}{b_1{}^2S_{xx}} = \frac{Nt^2s^2}{b_1{}^2|\mathbf{X}'\mathbf{X}|}. \tag{3.3.11}$$

Now, $0 < g < 1$ because if $g > 1$,

$$\frac{b_1{}^2}{s^2v_{11}} < t_{\frac{\alpha}{2}}{}^2.$$

Thus, when $g > 1$, β_1 is not statistically significant, and the Fieller interval is considered to be infinite. Since $0 < g < 1$, the first piece of (3.3.10) is minimized when g as close to 0 as possible. For g to be close to 0, $|\mathbf{X}'\mathbf{X}|$ must be as large as possible. Therefore, maximizing $|\mathbf{X}'\mathbf{X}|$ will minimize the squared half-width of the Fieller interval for the one

regressor normal error linear model. Recall that for fixed N, the D-optimality criterion also requires the determinant of **X'X** to be maximized. Therefore, an appropriate design for predicting the EC_{100p} well in simple linear regression is the D and Q-optimal design, the one which will also minimize the variance of the parameter estimates.

As mentioned in section 3.2, the D-optimal design for the model in (3.3.1) is an orthogonal design with all levels at the -1 and +1 extremes of the design region. In reality, a practitioner would probably add center runs to the design. Center runs are runs at the point coded to 0 on each variable which allow a test for lack-of-fit.

§3.3.2 The Two Regressor First Order Linear Model - No Interaction

In this case, the model of interest is

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_{ij} \qquad (3.3.12)$$

where $\varepsilon_{ij} \sim N(0,\sigma^2)$. When the model involves two regressors, the EC_{100p} is a line or curve defined by all the combinations of x_1 and x_2 which give the desired response. Given a particular x_1, minimizing the variance of the estimate of the EC_{100p} simply becomes a matter of minimizing the variance of $\hat{x}_2$. Thus, if x_1 is fixed, the F-optimality criterion seeks a design which minimizes the width of the Fieller interval on x_2 given x_1. In reality, however, accurate estimation of the EC_{100p} at a single x_1 value would probably not be of much interest. Typically, either the whole EC_{100p} curve or portions of the curve would be of interest. Therefore, it is desirable to have a criterion which minimizes the width of the Fieller interval on x_2 given x_1 for a range of x_1 values. One way to make the optimal design independent of the choice of x_1 is to average over x_1. The squared width of the Fieller interval on x_2 given x_1 can be integrated over x_1 on the region of interest before it is minimized. This is the approach taken in this research.

Again let the expected response at the EC_{100p} be k. Given x_1, let x_{2k} be the value of x_2 for which the $E(y) = k$. Figure 3.1 illustrates this situation for the case in which the

EC_{100p} is a straight line. Letting $v_{ii} = var(b_i)$ and $v_{ij} = cov(b_i,b_j)$, the 100(1-α)% Fieller interval for x_{2k} given x_1 is derived from

$$-t_{\frac{\alpha}{2}} \le \frac{k - b_0 - b_1x_1 - b_2x_{2k} - 0}{s\sqrt{v_{00} + x_1^2v_{11} + x_{2k}^2v_{22} + 2x_1v_{01} + 2x_{2k}v_{02} + 2x_1x_{2k}v_{12}}} \le t_{\frac{\alpha}{2}}. \quad (3.3.13)$$

Treating x_1 as fixed, squaring both sides and collecting terms yields the following quadratic equation in x_{2k}:

$$\begin{aligned}(b_2^2 - t^2s^2v_{22})x_{2k}^2 - 2(b_2(k - b_0 - b_1x_1) + t^2s^2v_{02} + t^2s^2x_1v_{12})x_{2k} \\ + (k - b_0 - b_1x_1)^2 - t^2s^2(v_{00} + x_1^2v_{11} + 2v_{01}x_1) = 0.\end{aligned} \quad (3.3.14)$$

The roots of this equation, the endpoints of the Fieller interval, are

$$\frac{b_2(k - b_0 - b_1x_1) + t^2s^2(v_{02} + x_1v_{12}) \pm ts\sqrt{R}}{b_2^2 - t^2s^2v_{22}} \quad (3.3.15)$$

where

$$\begin{aligned}R = {} & x_1^2(t^2s^2(v_{12}^2 - v_{11}v_{22}) + b_1^2v_{22} - 2b_1b_2v_{12} + b_2^2v_{11}) \\ & + 2x_1(t^2s^2(v_{02}v_{12} - v_{01}v_{22}) + (k - b_0)(b_2v_{12} - b_1v_{22}) + b_2^2v_{01} - b_1b_2v_{02}) \\ & + t^2s^2(v_{02}^2 - v_{00}v_{22}) + b_2^2v_{00} + 2b_2(k - b_0)v_{02} + v_{22}(k - b_0)^2.\end{aligned} \quad (3.3.16)$$

The squared half-width of the Fieller interval is given by

$$\frac{t^2s^2R}{(b_2^2 - t^2s^2v_{22})^2}. \quad (3.3.17)$$

If the numerator and denominator are each divided by b_2^2 and $g = t^2s^2v_{22}/b_2^2$, the following expression for the squared Fieller half-width (SFW) is obtained:

$$\begin{aligned}SFW = {} & \frac{1}{(1-g)^2}\left(\frac{t^2s^2}{b_2^2}\right)\left[x_1^2\left[\frac{g}{v_{22}}(v_{12}^2 - v_{11}v_{22}) + \left(\frac{b_1}{b_2}\right)^2v_{22} - 2\left(\frac{b_1}{b_2}\right)v_{12} + v_{11}\right]\right. \\ & + 2x_1\left[\frac{g}{v_{22}}(v_{02}v_{12} - v_{01}v_{22}) + \left(\frac{k - b_0}{b_2}\right)\left(v_{12} - \frac{b_1v_{22}}{b_2}\right) + v_{01} - \left(\frac{b_1}{b_2}\right)v_{02}\right] \\ & \left. + \left[\frac{g}{v_{22}}(v_{02}^2 - v_{00}v_{22}) + v_{00} + 2\left(\frac{k - b_0}{b_2}\right)v_{02} + \left(\frac{k - b_0}{b_2}\right)^2v_{22}\right]\right].\end{aligned} \quad (3.3.18)$$

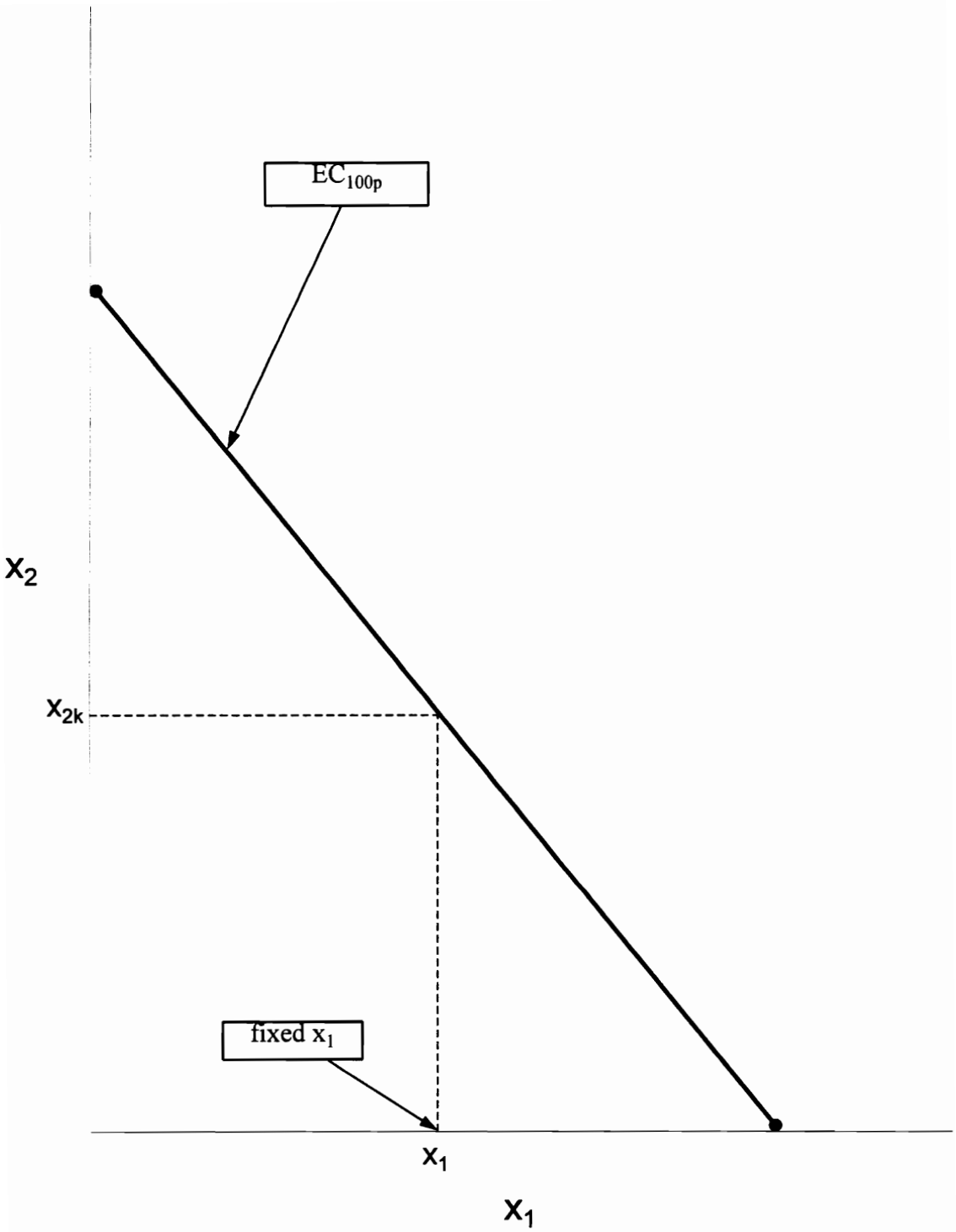

Figure 3.1 Plot of x_{2k} for a given value of x_1.

Notice that this expression is a function of α, the MSE, the expected response at the EC_{100p}, parameter estimates and variances and covariances determined by the design.

Integration of this squared half-width gives the final expression for the criterion. As mentioned in section 3.2, when designing experiments for linear regression with normal errors, the values in the region of interest on x_1 are typically coded from -1 to 1. Thus the integral of the squared half-width of the Fieller interval on x_2 given x_1 is

$$\int_{-1}^{1} SFW dx_1$$

$$= \frac{1}{(1-g)^2}\left(\frac{t^2s^2}{b_2^{\,2}}\right)\left[\frac{2}{3}\left[\frac{t^2s^2}{b_2^{\,2}}(v_{12}^{\,2} - v_{11}v_{22}) + \left(\frac{b_1}{b_2}\right)^2 v_{22} - 2\left(\frac{b_1}{b_2}\right)v_{12} + v_{11}\right]\right. \qquad (3.3.19)$$

$$\left. + 2\left[\frac{t^2s^2}{b_2^{\,2}}(v_{02}^{\,2} - v_{00}v_{22}) + v_{00} + 2\left(\frac{k-b_0}{b_2}\right)v_{02} + \left(\frac{k-b_0}{b_2}\right)^2 v_{22}\right]\right].$$

The design which minimizes the integrated squared half-width (ISFW) will be call the F*-optimal design.

A computer search over a set of candidate designs was needed to find the F*-optimal design. Some reasonable restrictions were placed on the designs in the candidate list to keep the number of eligible designs manageable. Thus, the candidate list contained only designs in which the columns for the regressors are orthogonal to the column for the intercept in the **X** matrix and designs with only two levels, symmetric about 0, of each regressor. The first constraint ensures that $v_{01} = v_{02} = 0$, which, for the model in (3.3.12), implies that $v_{00} = 1/N$ where N is the available number of experimental units for the design. The second property causes $\bar{x}_1$ and $\bar{x}_2$ to both be 0.

The restrictions on the designs in the candidate set allow for some simplification of the expression for the integrated squared half-width. Letting $d = t^2s^2/b_2^{\,2}$ so that $g = dv_{22}$, the integrated squared half-width becomes

$$\frac{2d}{(1-dv_{22})^2}\left[\frac{1}{3}\left[d(v_{12}^{\ 2}-v_{11}v_{22})+\left(\frac{b_1}{b_2}\right)^2 v_{22}-2\left(\frac{b_1}{b_2}\right)v_{12}+v_{11}\right]\right.$$
$$\left.+\left[\left(\left(\frac{k-b_0}{b_2}\right)^2-dv_{00}\right)v_{22}+v_{00}\right]\right]. \qquad (3.3.20)$$

As in the previous section, if $g \geq 1$, then β_2 is not significantly different from 0, and the Fieller interval has infinite width. Thus, $0 < g < 1$.

The variables in (3.3.20) which are determined by design are v_{11}, v_{12}, and v_{22}. In particular, these quantities are influenced by the collinearity between x_1 and x_2 and the levels of the regressors. Orthogonal designs, the type chosen by criteria such as D and Q-optimality, are known to have no collinearity between x_1 and x_2. However, the equation in (3.3.20) gives no reason to assume the same will be true of F*-optimality. Indeed, it is expected that the optimal design will contain some collinearity. Thus, designs with varying amounts of collinearity between x_1 and x_2 were in the set of candidate designs. The collinearity between x_1 and x_2 for each design was expressed in terms of a correlation type measure, ρ^*, defined by

$$\rho^* = \frac{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2}} = \frac{\sum_{i=1}^{n}x_{1i}x_{2i}}{\sqrt{\sum_{i=1}^{n}x_{1i}^{\ 2}\sum_{i=1}^{n}x_{2i}^{\ 2}}}. \qquad (3.3.21)$$

The set of candidate designs consisted of designs with $\rho^* = \pm 0.25, \pm 0.5, \pm 0.75, \pm 0.875, \pm 0.9375$. For each degree of collinearity, a grid search was performed to determine the optimal levels for the regressors. Again, all of the candidate designs were two level designs with the levels symmetric about 0, and many different combinations of possible levels of the regressors were examined. Since the coded range on each regressor is -1 to 1, the grid was constructed to divide the range from -1 to 1 into increments of 0.1. Values were set for d, b_1/b_2 and $(k - b_0)/b_2$. The ratio, b_1/b_2, was assigned six different

values: ± 0.1, ± 1 and ± 10. The ratio, $(k - b_0)/b_2$, was assigned three values: 0.1, 1, and 10. The parameter, d, ranged in value from 0 to n-1 since restricting $g < 1$ ensures $d < n$.

The computer search revealed that the F*-optimal design depends heavily on the values of the parameter estimates. For many combinations of the parameter estimates, the design with $\rho^* = 0$ and v_{11} and v_{22} as small as possible (the D and Q-optimal design) had the smallest ISFW. For some combinations, however, it did not. In particular, when d was near 0, $(k - b_0)/b_2$ was near 0 and b_1/b_2 was near -1, the F*-optimal designs turned out to be designs with large collinearity between x_1 and x_2. These types of designs are considered highly undesirable when judged by other popular design criteria.

Obviously, there are no general characteristics of an F*-optimal design. Only prior knowledge of the parameter estimates would make any one design more attractive than the others. A brief investigation, however, revealed that the ISFW for the D-optimal design was relatively close to that for the F*-optimal design for nearly all of the cases. The ISFWs for the D-optimal design were compared to the ISFWs for the F*-optimal designs by means of the following ratio:

$$D - F^* \text{ Eff} = \frac{\text{ISFW}_D}{\text{ISFW}_{F^*}}. \qquad (3.3.22)$$

These efficiencies are largest when $(k - b_0)/b_2 = 0$. Table 3.1 shows the efficiencies for this case. A D-F* Efficiency near of 1 indicates that the D-optimal design and the F*-optimal design are the same. A D-F* Efficiency greater than 1 indicates that the D-optimal design is not the best. It seems that when $(k - b_0)/b_2$ is near 0, the D-F* Efficiency is bounded on the upper end by 1.25. As $(k - b_0)/b_2$ increases, the efficiencies approach 1. The implication is that not much is lost in terms of F*-optimality if the D-optimal design is used, and, of course, use of the D-optimal design provides the additional advantages of better estimation of model coefficients and better prediction.

Table 3.1 D-F* Efficiencies for $(k - b_0)/b_1 = 0$ and Various Values of d, b_1/b_2 and Sample Size, N.

	d = 0			d = 0.0005			d = 0.05		
b_1/b_2	N = 16	N = 64	N = 128	N = 16	N = 64	N = 128	N = 16	N = 64	N = 128
0	1	1	1	1	1	1	1	1	1
±0.5	1.060	1.060	1.060	1.060	1.060	1.060	1.059	1.059	1.059
±1.0	1.200	1.200	1.200	1.201	1.201	1.201	1.193	1.199	1.200
±1.5	1.184	1.184	1.184	1.184	1.184	1.184	1.181	1.183	1.184
±2.0	1.141	1.141	1.141	1.141	1.141	1.141	1.139	1.140	1.140
±2.5	1.107	1.107	1.107	1.107	1.107	1.107	1.106	1.107	1.107
±3.0	1.083	1.083	1.083	1.083	1.083	1.083	1.082	1.083	1.083

	d = 3			d = 6			d = 9	d = 12	d = 15
b_1/b_2	N = 16	N = 64	N = 128	N = 16	N = 64	N = 128	N = 128	N = 128	N = 128
0	1	1	1	1	1	1	1	1	1
±0.5	1.037	1.050	1.054	1.032	1.044	1.050	1.047	1.044	1.042
±1.0	1.096	1.144	1.163	1.073	1.120	1.144	1.130	1.120	1.112
±1.5	1.105	1.152	1.166	1.076	1.131	1.152	1.140	1.131	1.123
±2.0	1.089	1.123	1.131	1.062	1.109	1.123	1.115	1.109	1.103
±2.5	1.070	1.095	1.101	1.049	1.086	1.095	1.090	1.086	1.081
±3.0	1.055	1.074	1.078	1.038	1.067	1.074	1.071	1.067	1.064

§3.3.3 The Two Regressor First Order Linear Model with Interaction

In this section, a design is sought which minimizes the variance of the EC_{100p} for a two regressor model in which the interaction between the independent variables is significant. The model for this case is

$$y_{ij} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_I x_{1i} x_{2i} + \varepsilon_i \tag{3.3.23}$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$. As in the previous sections, k is the expected response at the EC_{100p} for some p of interest, and x_{2k} is the EC_{100p}. It seems reasonable that the strategy used in the last section would work in this case as well. However, basing the F*-optimality criterion on the integrated squared half-width of the Fieller interval for x_{2k} given x_1 does not work well with the interaction model. In fact, the integral does not exist for some combinations of the parameter values. Perhaps a criterion which minimizes the variance of the *point* estimate, $\hat{x}_{2k}$ given x_1, rather than the width of an interval on x_{2k} given x_1 would work better.

Suppose the fitted model is

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + b_I x_{1i} x_{2i}. \tag{3.3.24}$$

Using this model, the estimate of the EC_{100p} given x_1 is

$$\hat{x}_{2k} = \frac{k - b_0 - b_1 x_1}{b_2 + b_I x_1}. \tag{3.3.25}$$

Use of the Delta Method yields

$$\begin{aligned} \operatorname{var}(\hat{x}_{2k}) \approx & \frac{v_{00} + x_1^2 v_{11} + 2x_1 v_{01}}{(\beta_2 + \beta_I x_1)^2} \\ & + \frac{2(k - \beta_0 - \beta_1 x_1)[v_{02} + x_1(v_{0I} + v_{12}) + x_1^2 v_{II}]}{(\beta_2 + \beta_I x_1)^3} \\ & + \frac{(k - \beta_0 - \beta_1 x_1)^2 [v_{22} + x_1^2 v_{II} + 2x_1 v_{2I}]}{(\beta_2 + \beta_I x_1)^4} \end{aligned} \tag{3.3.26}$$

Unfortunately, even though this expression is simpler than the expression for the squared Fieller half-width, the integral of this expression does not exist for certain parameter

values either. Since it is not possible to average the variances of $\hat{x}_{2k}$ over x_1 using integration, they will be averaged numerically. From the model in (3.3.23), given x_1,

$$\frac{k-\beta_0-\beta_1 x_1}{\beta_2+\beta_1 x_1} = x_{2k}. \tag{3.3.27}$$

Therefore, the expression for the approximate variance simplifies to

$$\begin{aligned} \operatorname{var}(\hat{x}_{2k}) \approx \left(\frac{1}{(\beta_2+\beta_1 x_1)^2}\right)[&(v_{00}+x_1{}^2 v_{11}+2x_1 v_{01}) \\ &+2x_{2k}(v_{02}+x_1(v_{01}+v_{12})+x_1{}^2 v_{11}) \\ &+x_{2k}{}^2(v_{22}+x_1{}^2 v_{11}+2x_1 v_{21})]. \end{aligned} \tag{3.3.28}$$

Clearly, the variance of the estimate of x_{2k} given x_1 depends not only on the particular value of x_1, but on the true value of x_{2k} as well. Thus, a computer search will have to be employed to average the above variance over values of both x_1 and x_2.

As in the previous section, the computer search requires a set of "candidate designs" for which the various variances and covariances in (3.3.28) differ. The candidate list will again contain only designs in which the columns for the regressors are orthogonal to the column for the intercept in the **X** matrix. As before, this ensures that $v_{01} = v_{02} = 0$. However, in the interaction case, an additional consequence of this constraint is that $v_{11} = v_{21} = 0$. Unfortunately, $v_{01} \neq 0$ as a result of this assumption. As it turns out, though, v_{01} is completely determined by the collinearity between x_1 and x_2, or v_{12}. Thus, as v_{12} is varied, v_{01} is simultaneously varied. This implies that the variances and covariances of the parameter estimates (and hence the F*-optimal design), in the interaction case, are completely determined by the levels of x_1 and x_2 and the collinearity between x_1 and x_2 just as in the no interaction case.

The set of candidate designs searched for the interaction case was the same set described in the previous section. To find a design which minimizes $\operatorname{var}(\hat{x}_{2k})$ averaged over values of x_1 and x_2, a grid of possible x_1, x_2 combinations was used. The grid divided the region from -1 to 1 on each variable into increments of 0.05. For each design in the candidate set, the variance of $\hat{x}_{2k}$ was computed at each point on the grid. Then

these values were averaged over the grid. The design which gave the smallest average variance over the grid was considered the F*-optimal design. The results of this analysis indicate that the design with $\rho^* = 0$, and x_1 and x_2 both at ± 1 levels (the D and Q-optimal design for the normal error linear model) is the best design *on the average*. However, the study also shows that for certain combinations of x_1 and x_{2k}, there are other designs which give a much smaller $\text{var}(\hat{x}_{2k})$. Some of these have much higher collinearity than the D-optimal design and have levels different from -1 and 1. If a practitioner has a prior idea of the parameter values and a particular x_1 of interest, there may be a better design than the D-optimal to use for his or her situation. In general, however, the D-optimal design would be the best to use for minimizing the variance of the estimate of the EC_{100p} for a two regressor first order linear model with interaction.

§3.4 Allocation of Observations to the Control

Recall that the normal error linear model may only be a reasonable model over a small range of the impairment data, probably away from the extremes of the response. For this reason, the control would probably not be in the experimenter's region of operability when using this model, and therefore, would probably not be selected as a design point by the alphabetic optimality criteria. Nevertheless, the observations at the control are necessary for calculating the "percent of maximum reproduction" or "percent impairment," and observations should be allocated to the control in some optimal way. A separate design criterion is presented in this section for allocating a portion of the sample to the control. One should recall that in the previous section, the F-optimal and F*-optimal designs were frequently the D and Q-optimal design. Thus, the criterion for allocation of experimental units to the control is derived with the assumption that the design which will acutally be analyzed (the design apart from the controls) is the D and Q optimal design.

The criterion used in this section to determine the optimal number of control observations is minimization, over all possible designs on the region, of the prediction variance at a particular EC of interest. The model of interest is

$$\frac{y_{ij}}{\bar{y}_c} = \mathbf{x}_i'\beta + \varepsilon_{ij} \tag{3.4.1}$$

where the ε_{ij}'s are assumed to be i.i.d. $N(0,\sigma^2)$, y_{ij} is the organism count for the jth replicate at the ith design point, and $\bar{y}_c$ is the average of the organism counts at the control. Let the EC of interest again be referred to as the EC_{100p}, or the point in the design space at which (100 x p)% of the reproduction at the control is achieved.

The prediction variance for the model in (3.4.1) is

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) = \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \mathrm{Var}\left(\frac{y_{ij}}{\bar{y}_c}\right). \tag{3.4.2}$$

Clearly this expression contains the variance of a ratio of random variables. The approximate variance of a ratio of random variables can be derived using the Delta Method. The approximation is

$$\mathrm{Var}\left(\frac{y_{ij}}{\bar{y}_c}\right) \approx \left(\frac{E(y_{ij})}{E(\bar{y}_c)}\right)^2 \left[\frac{\mathrm{var}\ y_{ij}}{E(y_{ij})^2} + \frac{\mathrm{var}\ \bar{y}_c}{E(\bar{y}_c)^2} - \frac{2\,\mathrm{cov}(y_{ij},\bar{y}_c)}{E(y_{ij})E(\bar{y}_c)}\right]. \tag{3.4.3}$$

This approximation does not depend on the distribution of the individual random variables in the ratio. As the observations are i.i.d., the numerator and denominator of the ratio are independent, and the approximate variance in (3.4.3) simplifies to

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) \approx \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \left[\frac{\mu_{y_{ij}}^2}{\mu_{\bar{y}_c}^2}\left(\frac{\mathrm{var}\ y_{ij}}{\mu_{y_{ij}}^2} + \frac{\mathrm{var}\ \bar{y}_c}{\mu_{\bar{y}_c}^2}\right)\right]. \tag{3.4.4}$$

Since the model in (3.4.1) assumes homogeneous variance, var y_{ij} = var y_c = var y = σ^2. Therefore, var $\bar{y}_c = \sigma^2/n_c$ where n_c is the number of observations at the control settings, and

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) \approx \mathbf{x}_i' (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \left[1 + \frac{\mu_{y_{ij}}^2}{n_c \mu_{\bar{y}_c}^2}\right] \frac{\sigma^2}{\mu_{\bar{y}_c}^2}. \tag{3.4.5}$$

This variance approximation is known to work well when $\mu_{\bar{y}_c}$ is large and σ^2 is small.

§3.4.1 Simple Linear Regression

Consider the simple linear regression model

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}. \tag{3.4.6}$$

For this model, the approximate prediction variance at $x = x_i$ is

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) \approx \left(\frac{1}{n_d} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n_d} x_i^2 - n_d \bar{x}^2}\right)\left(1 + \frac{\mu_{y_{ij}}^2}{n_c \mu_{\bar{y}_c}^2}\right)\left(\frac{\sigma^2}{\mu_{\bar{y}_c}^2}\right) \tag{3.4.7}$$

where n_d is the number of observations in the design and n_c is the number of observations at the control. Recall that the design criterion seeks a design which minimizes the prediction variance at a particular EC_{100p}. Thus, p can be substituted for $\mu_{y_i} / \mu_{\bar{y}_c}$. The expression for the approximate prediction variance at the EC_{100p} becomes

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) \approx \left(\frac{1}{n_d} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n_d} x_i^2 - n_d \bar{x}^2}\right)\left(1 + \frac{p^2}{n_c}\right)\left(\frac{\sigma^2}{\mu_{\bar{y}_c}^2}\right). \tag{3.4.8}$$

Minimization of this quantity is simplified greatly if two further assumptions are made. These assumptions are actually constraints on the placement of the design points in the design to be analyzed (the non-controls). The first constraint requires the coded levels of the independent variable to be symmetric about $x = 0$, and the second requires the same number of replicates to be placed at each level. The D and Q optimal design incorporates both constraints. The additional assumptions ensure that $\bar{x} = 0$ and that the design is orthogonal. Let N be the total number of available experimental units and p_c be

the proportion of the available experimental units that should be allocated to the control settings. With the same number of replicates at each design point, the expression for the prediction variance at $x = x_i$ becomes

$$\mathrm{Var}\left(\frac{\hat{y}_{ij}}{\bar{y}_c}\right) \approx \left(\frac{k + x_i^2}{rN(1-p_c)}\right)\left(1 + \frac{p^2}{Np_c}\right)\left(\frac{\sigma^2}{\mu_{\bar{y}_c}^2}\right) \tag{3.4.9}$$

where r is a constant determined by the number of levels of the design variable to be used. Minimizing this expression with respect to p_c yields the solution

$$p_c = \frac{-p^2 + \sqrt{p^4 + p^2 N}}{N}. \tag{3.4.10}$$

The constant, r, and x_i drop out of the expression. This implies that knowledge of the actual EC_{100p} is not necessary to determine the number of controls which would minimize the prediction variance at that location.

For many applications, including impaired reproduction studies, $p \leq 1$. Figure 3.2 shows the optimal proportion of controls by sample size given values of p for this case. Sometimes, however, $p \geq 1$. Figure 3.3 shows the optimal proportion of controls for this situation. It is clear that more controls are needed as p gets large. For situations in which toxicants impair reproduction, the necessary proportion of controls rarely gets higher than 20%. In the situations where chemicals enhance reproduction, a larger proportion of controls is necessary, in some cases nearly 50%.

§3.4.2 Multiple Linear Regression

The one variable case works out very nicely. Most often, though, the model in (3.4.1) will contain more than one independent variable. Let b = the number of independent variables in the model. Then, the model of interest is

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_b x_{bi} + \beta_{12} x_{1i} x_{2i} + \cdots + \beta_{(b-1)b} x_{(b-1)i} x_{bi} + \varepsilon_i \tag{3.4.11}$$

where $\varepsilon_i \sim N(0,1)$. The model is still first order, but now it may contain two-factor

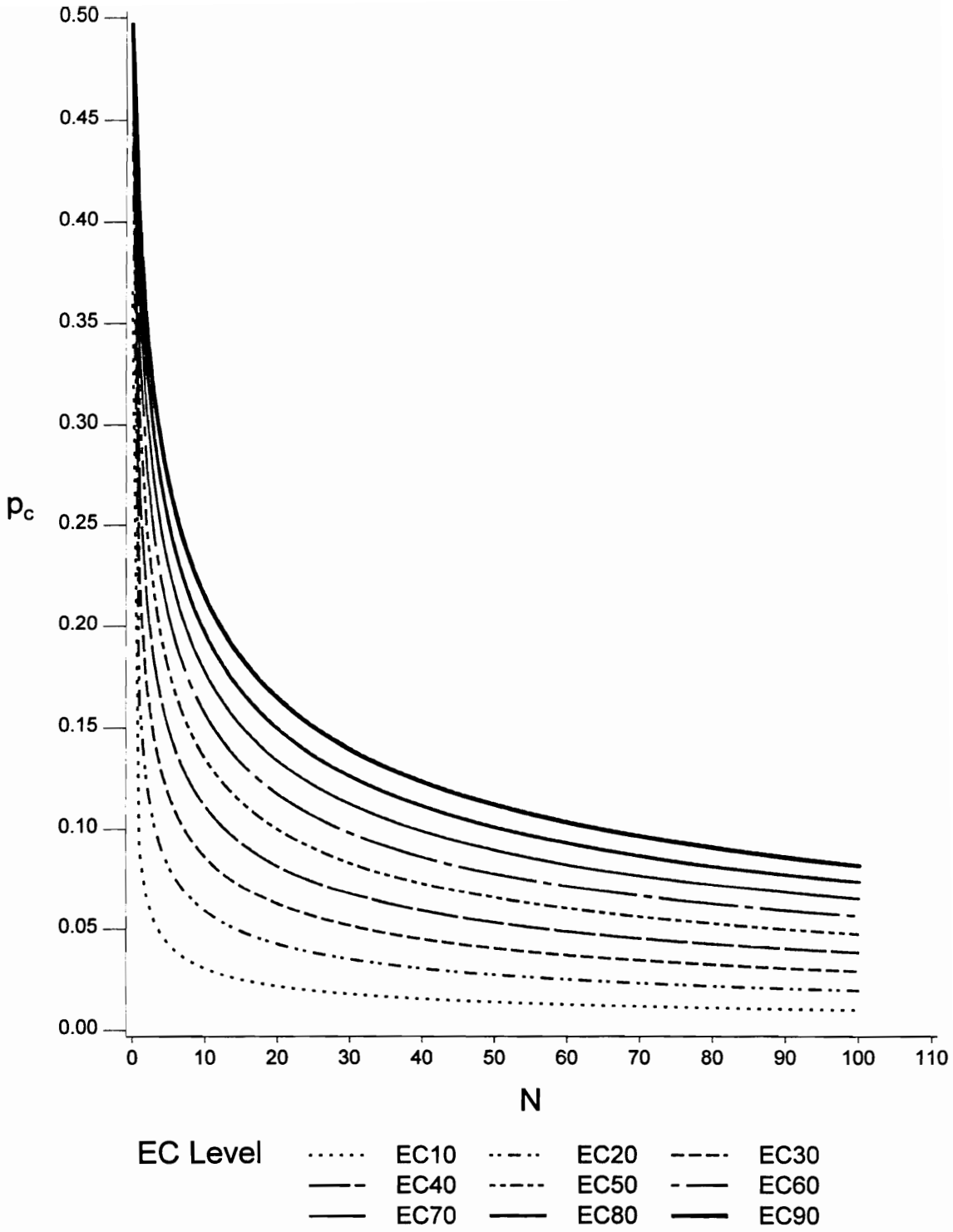

Figure 3.2 Plot of the optimal proportion of controls by total sample size for various ECs of interest between the EC_{10} and EC_{90}.

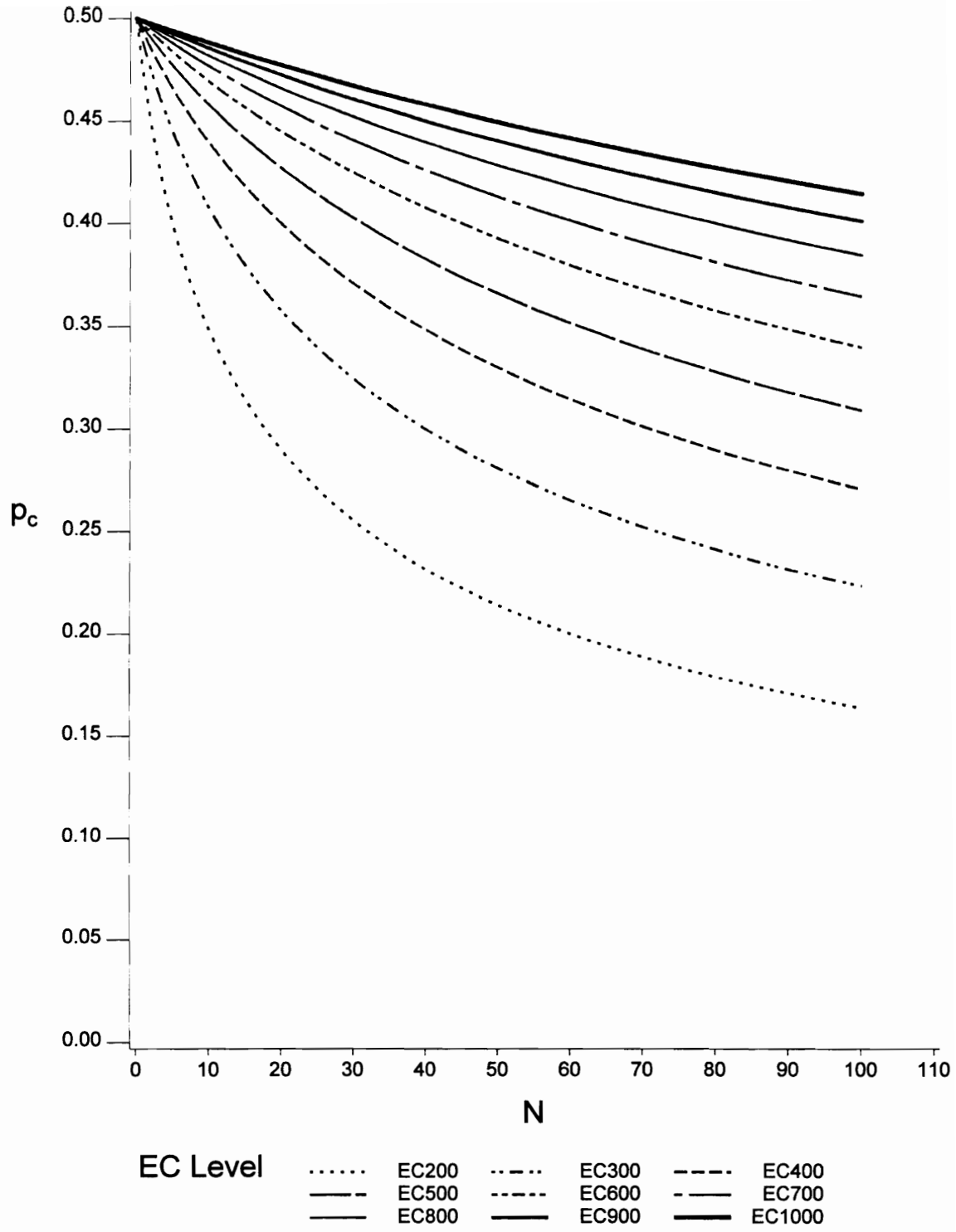

Figure 3.3 Plot of the optimal proportion of controls by total sample size for various ECs of interest between the EC_{200} and EC_{1000}.

interactions as does the model in (3.4.11). Nevertheless, when the two previously mentioned design assumptions (levels symmetric about 0, same number of replicates at each design point) hold, some nice properties in the $\mathbf{X'X}$ matrix arise. In particular, $\sum_{i=1}^{n_d} x_{mi}^2 = r_1$ for m = 1,...,b and $\sum_{i=1}^{n_d} x_{mi}^2 x_{ji}^2 = r_2$ for m = 1,...,b, and j = 1,...,b, i < j. The constants r_1 and r_2 are determined by the number of levels of the regressors. Because of these features, minimization of the prediction variance in the multiple variable case gives the same results as in the single variable case. Figures 3.2 and 3.3, then, describe the relationship between total sample size and optimal proportion of controls for any number of independent variables when a first order or first order with interaction model is used. One should keep in mind, however, that homogeneous variance is assumed, and that two design constraints are imposed.

§3.5 Conclusions

Two important conclusions can be drawn from the work on the linear model with normal errors. First, when the response is "percent of the reproduction at the control" or "percent impairment," there is a reasonable criterion for finding the appropriate number of observations to take at the control so long as the control points are not also used as design points. Secondly, and perhaps more importantly, the D-optimal design for linear regression with normal errors is also a design which will provide reasonably accurate prediction of the EC_{100p} for models with one and two variables (with and without interaction).

An additional note on the normal error linear model warrants discussion. Frequently, the actual organism or cell counts measured in toxicity studies are modeled as Poisson data using a log-linear model, a model in which the log transform of the response is modeled by a linear model. If this log-linear model is assumed to have normal errors and homogeneous variance, the F-optimality results given this research also apply to this

model. The optimal control results do not apply since the ratio response is used to form the criterion. (The same is true of models for count data in which a square root transformation of the counts is modeled using a linear model with homogeneous variance.)

To this point, the focus has been on the model in (3.1.1), that is, the homogeneous variance model. As previously mentioned, this may not be a sensible assumption for the impaired reproduction data. In Chapter 5, however, optimal designs for a heterogeneous variance model, a Poisson linear model, are discussed. First, however, optimal designs for a nonlinear Poisson model, the exponential model, are considered.

Chapter 4

Optimal Designs for the Poisson Exponential Model

§4.1 Introduction

The model under consideration in this chapter is the Poisson exponential model,

$$y_{ij} = e^{\beta_0+\beta_1 x_i} + \varepsilon_{ij}, \tag{4.1.1}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. D and Q-optimal, as well as slope and F-optimal designs for this model will be found and discussed in the following sections. Slope optimality is developed for this model alone because in this specific case, it is equivalent to F-optimality. Recall that the Fisher Information matrix plays a key role in all of the design optimality criteria for nonlinear and non-normal error models. In Chapter 2, the information matrix for the model in (4.1.1) was derived. The information matrix for an experiment with N experimental runs is

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \lambda_i & \sum_{i=1}^{N} x_i\lambda_i \\ \sum_{i=1}^{N} x_i\lambda_i & \sum_{i=1}^{N} x_i^2\lambda_i \end{bmatrix}. \tag{4.1.2}$$

§4.2 Characterization of Optimal Designs

All of the designs for the three Poisson regression models have some common characteristics. First, the total number of available experimental units, N, will be assumed to be constant. Secondly, as mentioned in Chapter 2, because the chemicals used in these studies impair reproduction, maximum reproduction occurs at the control.

If λ_c is the expected reproduction at the control, then any experimental unit treated with a non-control level of the toxicant experiences impairment and yields a mean reproduction of $p\lambda_c$, for some $0 < p < 1$. The design point which has expected response $p\lambda_c$ is called the EC_{100p}, and is the point at which (1-p)100% impairment in reproduction occurs. Due to the presence of the Poisson mean, λ_i, in the information matrices for the Poisson models, the design criteria produce optimal designs in terms of optimal λ's, the expected responses at the optimal design levels, and knowledge of the model parameters is needed to obtain the actual optimal design levels, the x's which produce these expected responses. To illustrate, suppose the optimal design has 50% of the observations at the EC_{15} and 50% of the observations at the EC_{50}. Guesses of the model parameters are needed to determine what x_i's will yield responses of approximately $0.15\lambda_c$ and $0.50\lambda_c$, respectively. Of course, not all guesses are close to the true model parameters, and Chapter 7 provides some methods for addressing this problem. Thirdly, sometimes the regions of operability and interest must be restricted to a portion of the entire dose space in practical problems. For the Poisson regression models, when design space is restricted, the regions of operability and interest given in terms of EC's, or x's. Finally, for all three Poisson models, optimal sample allocation at the control is incorporated in the alphabetic optimality design criteria.

§4.3 D-Optimal Designs

Recall that the D-optimal design is the design which minimizes the generalized variance of the estimates of the model parameters. In particular, it is the design which maximizes $|\mathbf{I}|/N$. For the exponential model,

$$|\mathbf{I}| = \sum \lambda_i \sum x_i^2 \lambda_i - \left(\sum x_i \lambda_i\right)^2 . \qquad (4.3.1)$$

Consider an optimal two level design. This design has $n_1 = p_1N$ experimental units assigned to a level x_1 in the design space and $n_2 = (1-p_1)N$ experimental units assigned to a level x_2. For a two level design,

$$\begin{aligned} |\mathbf{I}| &= (n_1\lambda_1 + n_2\lambda_2)(n_1x_1^{\,2}\lambda_1 + n_2x_2^{\,2}\lambda_2) - (n_1x_1\lambda_1 + n_2x_2\lambda_2)^2 \\ &= n_1n_2\lambda_1\lambda_2(x_1 - x_2)^2. \end{aligned} \tag{4.3.2}$$

However,

$$\begin{aligned} x_i &= \frac{\ln\lambda_i - \beta_0}{\beta_1} \\ \Rightarrow \quad \frac{|\mathbf{I}|}{N} &= \left(\frac{1}{N\beta_1^{\,2}}\right) n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2. \end{aligned} \tag{4.3.3}$$

The D-optimal design is the design which maximizes this determinant apart from the constants $1/\beta_1^{\,2}$ and N. It is of interest to note the tradeoff between the need to spread the points out and the need to place them close together. By inspection, n_1n_2 is maximized when $n_1 = n_2$. Thus, the optimal design will have half the available experimental units at each of the two levels. Similarly, $\lambda_1\lambda_2$ is maximized when $\lambda_1 = \lambda_2 = \lambda_c$ since λ_c is the largest possible mean. However, if $\lambda_1 = \lambda_2$, then $\ln(\lambda_1/\lambda_2) = 0$, and the determinant is clearly not maximized. Assume $\lambda_2 \geq \lambda_1$. Then, $\lambda_1 = k\lambda_2$ where $0 < k \leq 1$. In Appendix A, it is shown that if $\lambda_2 = \lambda_c$ and $k = e^{-2} = 0.1353$, the determinant in (4.3.3) is maximized. Thus, the two level D-optimal design has half of the observations taken at the control and half of the observations taken at the x which yields a response of $0.1353\lambda_c$, or the $EC_{13.5}$.

In general, practitioners would probably not want to use a two level design as it allows for no test of lack-of-fit for the model. Suppose a researcher is interested in a three level design. The control was automatically chosen as one of the design points in the optimal two level design. In toxicity studies, however, the control is almost always run whether or not it is chosen as a design point for analysis. Therefore, the three level designs in this work will be required to contain the control as one design point (whether or not it is in the region of operability) and will be allowed to have two design levels *in*

addition to the control. The determinant of the information matrix for a three level design is

$$|\mathbf{I}| = \left(\frac{1}{\beta_1^{\,2}}\right)\left(n_1 n_2 \lambda_1 \lambda_2 \left(\ln\frac{\lambda_1}{\lambda_2}\right)^2 + n_1 n_3 \lambda_1 \lambda_3 \left(\ln\frac{\lambda_1}{\lambda_3}\right)^2 + n_2 n_3 \lambda_2 \lambda_3 \left(\ln\frac{\lambda_2}{\lambda_3}\right)^2\right)$$
$$= \left(\frac{1}{\beta_1^{\,2}}\right)\left(n_1 n_2 \lambda_1 \lambda_2 \left(\ln\frac{\lambda_1}{\lambda_2}\right)^2 + n_1 n_c \lambda_1 \lambda_c \left(\ln\frac{\lambda_1}{\lambda_c}\right)^2 + n_2 n_c \lambda_2 \lambda_c \left(\ln\frac{\lambda_2}{\lambda_c}\right)^2\right) \quad (4.3.4)$$

since $\lambda_3 = \lambda_c$ by assumption. Maximizing (4.3.4) requires a numerical solution, so the Nelder-Mead numerical minimization algorithm (Nelder and Mead, 1965) was used. To simplify the procedure slightly for the Nelder-Mead algorithm, an additional constraint was imposed on the design. It was assumed that the same number of observations would be taken at each of the non-control levels, i.e. $n_1=n_2$. Since the third level is fixed at the control, Nelder-Mead searched for the optimal first and second levels (x_1 and x_2), the optimal proportion of observations at each of these levels ($p_1=p_2$), and the optimal proportion of observations at the control (p_c).

Optimal three level designs were sought over several regions of operability. The designs are displayed in Table 4.1. As seen in Cases A-D, the original attempt to find three level designs resulted in two level designs. Since the goal was to find designs with three distinct levels, for Cases E-H, Nelder-Mead forced the non-control design levels to be distinct by forcing the means at those levels to be symmetric about the mean in the center of the region of interest/operability.

The designs in Table 4.1 are compared to the optimal design by means of D-Efficiencies. The D-Efficiency of a design is calculated by

$$\text{D} - \text{Efficiency} = \left(\frac{|\mathbf{I}| \text{ for the design of interest}}{|\mathbf{I}| \text{ for optimal design}}\right)^{\frac{1}{2}} \quad (4.3.5)$$

when both the design of interest and the optimal design contain the same number of experimental units. The first column of D-Efficiencies contains efficiencies which compare all of the designs to the overall two level D-optimal design on the entire dose

Table 4.1. Three Level D-Optimal Designs for the Poisson Exponential Model and the D-Efficiencies for Each.

Case	Region of Interest/Operability	Levels Symmetric about Center	p_c	$p_1=p_2$	x_1	x_2	D-Efficiency	D-Efficiency (restricted)
A	$(EC_0, EC_{100}]$	no	0.5	0.25	$EC_{13.5}$	$EC_{13.5}$	1	1
B	$[EC_{20}, EC_{80}]$	no	0.5	0.25	EC_{20}	EC_{20}	0.9783	1
C	$[EC_{10}, EC_{90}]$	no	0.5	0.25	$EC_{13.5}$	$EC_{13.5}$	1	1
D	$[EC_{10}, EC_{50}]$	no	0.5	0.25	$EC_{13.5}$	$EC_{13.5}$	1	1
E	$(EC_0, EC_{100}]$	yes	0.154	0.423	$EC_{9.4}$	$EC_{90.6}$	0.9139	0.9139
F	$[EC_{20}, EC_{80}]$	yes	0.31	0.345	EC_{20}	EC_{80}	0.8437	0.8623
G	$[EC_{10}, EC_{90}]$	yes	0.164	0.418	EC_{10}	EC_{90}	0.9135	0.9135
H	$[EC_{10}, EC_{50}]$	yes	0.486	0.257	$EC_{19.2}$	$EC_{40.8}$	0.8984	0.8984

space. Most of the designs in the table have high D-Efficiencies in this column. The highest D-Efficiencies are achieved by the designs which have the $EC_{13.5}$ in the region of operability/interest. The column labeled *D-Efficiency (restricted)* contains efficiencies which compare only designs on the same regions of operability. These are the more sensible efficiencies because if a researcher can only operate from the EC_{20} to the EC_{80}, then the optimal design on this region is *the* optimal design for that researcher. The design that is optimal on a larger region is not even a practical consideration. Although only a few restricted regions were used to construct the table, the algorithm developed for this section can be used to find the D-optimal design for any portion of $(0, \lambda_c]$.

§4.4 Q-Optimal Designs

The Q-optimality criterion addresses prediction. For nonlinear regression, the criterion is

$$\min_{\mathcal{D}} \frac{N}{K}\int_R \text{var}(\hat{f}(\mathbf{x}_0,\hat{\beta}))\partial\mathbf{x}_0 \qquad (4.4.1)$$

where $\hat{f}(\mathbf{x}_0,\hat{\beta}) = \hat{y}(\mathbf{x}_0) = e^{\hat{\beta}_0+\hat{\beta}_1 x_0}$, R is the region of interest and K is the volume of that region. In Chapter 2, a general expression for the approximate prediction variance at a point, $\mathbf{x}_0$, based on a nonlinear (or nonhomogeneous variance) model was given. The expression is

$$\text{var}(f(\mathbf{x}_0,\hat{\beta})) \approx \mathbf{w}(\mathbf{x}_0)'\mathbf{I}^{-1}\mathbf{w}(\mathbf{x}_0)$$

where

$$\mathbf{w}(\mathbf{x}_0)' = \left[\frac{\partial f(\mathbf{x}_0,\beta)}{\partial\beta_1},\cdots,\frac{\partial f(\mathbf{x}_0,\beta)}{\partial\beta_p}\right].$$

For the exponential model and a design with a general number of levels,

$$\mathbf{I}^{-1} = \frac{1}{\sum\lambda_i\sum x_i^2\lambda_i - \left(\sum x_i\lambda_i\right)^2}\begin{pmatrix}\sum x_i^2\lambda_i & -\sum x_i\lambda_i \\ -\sum x_i\lambda_i & \sum\lambda_i\end{pmatrix} \qquad (4.4.2)$$

and

$$\mathbf{w}(\mathbf{x}_0)' = \left[e^{\beta_0+\beta_1 x_0} \quad x_0 e^{\beta_0+\beta_1 x_0}\right] = \left[\lambda_0 \quad x_0\lambda_0\right]. \tag{4.4.3}$$

Thus,

$$\mathrm{var}(\hat{f}(x_0,\hat{\beta})) \approx \frac{\lambda_0^2\left(\sum x_i^2\lambda_i - 2x_0\sum x_i\lambda_i + x_0^2\sum\lambda_i\right)}{\sum\lambda_i\sum x_i^2\lambda_i - \left(\sum x_i\lambda_i\right)^2} \tag{4.4.4}$$

where λ_0 is the mean at the point of interest x_0. The prediction variance for a two level design is

$$\frac{\lambda_0^2\left(n_1\lambda_1\left(\ln\frac{\lambda_1}{\lambda_0}\right)^2 + n_2\lambda_2\left(\ln\frac{\lambda_2}{\lambda_0}\right)^2\right)}{n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}. \tag{4.4.5}$$

Recall that the Q criterion requires integration of this expression over the region of interest. After substituting $e^{\beta_0+\beta_1 x_i}$ for λ_i, integration can be accomplished using integration by parts. As mentioned in section 4.2, the integration limits are expressed in terms of EC's. Let x_U be the upper bound on the region of interest, the point at which the expected response is λ_U, and let x_L be the lower bound on the region of interest, the point at which the expected response is λ_L. Then, for two levels, after simplification,

$$Q = \int_{x_L}^{x_U} \mathrm{var}(\hat{f}(\mathbf{x}_0,\hat{\beta})) = \frac{NA}{4n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_U}{\lambda_L}\right)\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2} \tag{4.4.6}$$

where

$$\begin{aligned} A = {} & n_1\lambda_1\lambda_U^2\left[2\left(\ln\frac{\lambda_1}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_U}\right) + 1\right] - n_1\lambda_1\lambda_L^2\left[2\left(\ln\frac{\lambda_1}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_L}\right) + 1\right] \\ & + n_2\lambda_2\lambda_U^2\left[2\left(\ln\frac{\lambda_2}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_U}\right) + 1\right] - n_2\lambda_2\lambda_L^2\left[2\left(\ln\frac{\lambda_2}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_L}\right) + 1\right]. \end{aligned} \tag{4.4.7}$$

Note that the denominator in (4.4.6) is the volume of the region of interest multiplied by the determinant of the information matrix apart from $1/\beta_1^2$. The determinant of the information matrix becomes a part of the denominator because the Q criterion contains the inverse of the information matrix. Thus, making the determinant large is a goal of both D and Q-optimality, and the connection between the two types of optimality can be seen. Once again, the Nelder-Mead algorithm was used to search for the optimal designs. The optimal two level design with regions of operability and interest both $(0, \lambda_c]$ is a design with 56.1% of the experimental units assigned to the $EC_{18.7}$ and 43.9% of the units assigned to the control.

Three level Q-optimal designs were also found. As in the previous section, one of the three levels is forced to be at the control, and $n_1=n_2$ by assumption. For three levels, the prediction variance is

$$\frac{\lambda_0^2\left(n_1\lambda_1\left(\ln\frac{\lambda_1}{\lambda_0}\right)^2+n_2\lambda_2\left(\ln\frac{\lambda_2}{\lambda_0}\right)^2+n_c\lambda_c\left(\ln\frac{\lambda_c}{\lambda_0}\right)^2\right)}{n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2+n_1n_c\lambda_1\lambda_c\left(\ln\frac{\lambda_1}{\lambda_c}\right)^2+n_2n_c\lambda_2\lambda_c\left(\ln\frac{\lambda_2}{\lambda_c}\right)^2}. \tag{4.4.8}$$

The final Q-optimality criterion for three levels is

$$Q=\frac{NA}{4D\left(\ln\frac{\lambda_U}{\lambda_L}\right)} \tag{4.4.9}$$

where

$$\begin{aligned}A=\;&n_1\lambda_1\lambda_U^2\left[2\left(\ln\frac{\lambda_1}{\lambda_U}\right)^2+2\left(\ln\frac{\lambda_1}{\lambda_U}\right)+1\right]-n_1\lambda_1\lambda_L^2\left[2\left(\ln\frac{\lambda_1}{\lambda_L}\right)^2+2\left(\ln\frac{\lambda_1}{\lambda_L}\right)+1\right]\\&+n_2\lambda_2\lambda_U^2\left[2\left(\ln\frac{\lambda_2}{\lambda_U}\right)^2+2\left(\ln\frac{\lambda_2}{\lambda_U}\right)+1\right]-n_2\lambda_2\lambda_L^2\left[2\left(\ln\frac{\lambda_2}{\lambda_L}\right)^2+2\left(\ln\frac{\lambda_2}{\lambda_L}\right)+1\right]\\&+n_c\lambda_c\lambda_U^2\left[2\left(\ln\frac{\lambda_c}{\lambda_U}\right)^2+2\left(\ln\frac{\lambda_c}{\lambda_U}\right)+1\right]-n_c\lambda_c\lambda_L^2\left[2\left(\ln\frac{\lambda_c}{\lambda_L}\right)^2+2\left(\ln\frac{\lambda_c}{\lambda_L}\right)+1\right]\end{aligned} \tag{4.4.10}$$

and

$$D = n_1 n_2 \lambda_1 \lambda_2 \left(\ln \frac{\lambda_1}{\lambda_2} \right)^2 + n_1 n_c \lambda_1 \lambda_c \left(\ln \frac{\lambda_1}{\lambda_c} \right)^2 + n_2 n_c \lambda_2 \lambda_c \left(\ln \frac{\lambda_2}{\lambda_c} \right)^2 . \quad (4.4.11)$$

As in the two level case, this quantity, D, is the determinant of the information matrix apart from $1/\beta_1^2$. Q-optimal designs for various regions of operability and interest are displayed in Table 4.2. As in the previous section, the optimal three level designs reverted to two level designs, so in Cases E-H and f-h, the non-control levels were forced to be distinct.

Q-Efficiencies for the designs are also given in Table 4.2. These efficiencies compare the designs by means of the following formula:

$$Q - \text{Efficiency} = \frac{Q \text{ for the optimal design}}{Q \text{ for the design of interest}} . \quad (4.4.12)$$

There are two columns for Q-Efficiency in the table. The efficiencies in the *Q-Efficiency (whole region)* column compare all the designs to Case A, *the* optimal design on the whole region (the two level design). The column for *Q-Efficiency (restricted)* contains efficiencies which compare only designs with the same operability and interest regions. Given a particular region of interest and operability, the optimal design for this column is considered to be the design with no symmetry imposed, and these are probably the only fair comparisons. In the second column of efficiencies, nearly all of the restricted Q-optimal designs have efficiencies over 90%. Even when compared across regions in the first column of efficiencies, many of the Q-optimal designs are highly efficient. In general, once restrictions on regions are imposed, it is difficult to interpret Q-Efficiencies. Although only a small number of possible situations are represented in Table 4.2, the algorithm developed for this section can be used to find optimal designs over any combination of regions of operability and interest on $(0, \lambda_c]$.

Table 4.2. Three Level Q-optimal Designs for the Poisson Exponential Model and the Q-Efficiencies for Each.

Case	Region of Operability	Region of Interest	Symmetric Levels	p_c	$p_1=p_2$	x_1	x_2	Q-Efficiency (whole region)	Q-Efficiency (restricted)
A	$(EC_0, EC_{100}]$	$(EC_0, EC_{100}]$	no	0.438	0.281	$EC_{18.7}$	$EC_{18.7}$	1	1
B	$(EC_0, EC_{100}]$	$[EC_{20}, EC_{80}]$	no	0.366	0.317	EC_{25}	EC_{25}	0.9700	1
C	$(EC_0, EC_{100}]$	$[EC_{10}, EC_{90}]$	no	0.404	0.298	$EC_{21.2}$	$EC_{21.2}$	0.9941	1
D	$(EC_0, EC_{100}]$	$[EC_{10}, EC_{50}]$	no	0.178	0.411	$EC_{21.5}$	$EC_{21.5}$	0.6942	1
E	$(EC_0, EC_{100}]$	$(EC_0, EC_{100}]$	yes	0.064	0.468	$EC_{12.6}$	$EC_{87.4}$	0.9267	0.9267
F	$(EC_0, EC_{100}]$	$[EC_{20}, EC_{80}]$	yes	0.028	0.486	EC_{17}	EC_{83}	0.8946	0.9445
G	$(EC_0, EC_{100}]$	$[EC_{10}, EC_{90}]$	yes	0.038	0.481	$EC_{14.2}$	$EC_{85.8}$	0.9206	0.9330
H	$(EC_0, EC_{100}]$	$[EC_{10}, EC_{50}]$	yes	0	0.50	$EC_{15.6}$	$EC_{84.4}$	0.8995	0.7736
b	$[EC_{20}, EC_{80}]$	$[EC_{20}, EC_{80}]$	no	0.366	0.317	EC_{25}	EC_{25}	0.9700	1
c	$[EC_{10}, EC_{90}]$	$[EC_{10}, EC_{90}]$	no	0.404	0.298	$EC_{21.2}$	$EC_{21.2}$	0.9941	1
d	$[EC_{10}, EC_{50}]$	$[EC_{10}, EC_{50}]$	no	0.178	0.411	$EC_{21.5}$	$EC_{21.5}$	0.6942	1
f	$[EC_{20}, EC_{80}]$	$[EC_{20}, EC_{80}]$	yes	0.054	0.473	EC_{20}	EC_{80}	0.8628	0.9380
g	$[EC_{10}, EC_{90}]$	$[EC_{10}, EC_{90}]$	yes	0.038	0.481	$EC_{14.2}$	$EC_{85.8}$	0.9206	0.9330
h	$[EC_{10}, EC_{50}]$	$[EC_{10}, EC_{50}]$	yes	0.126	0.437	$EC_{19.1}$	$EC_{40.9}$	0.6257	0.9226

§4.5 Designing Experiments to Estimate the EC_{100p}

F-optimality is the criterion which addresses estimation of the EC_{100p} for some p of interest. The F-optimality criterion is developed in this section for the one regressor exponential Poisson model. For this model, however, slope optimality, a criterion which addresses minimization of the variance of the slope, has a special interpretation which makes it similar to F-optimality. Thus, a discussion of slope optimality is presented first.

Consider again the one regressor exponential model,

$$E(y_i) = e^{\beta_0+\beta_1 x_i}.$$

When maximum reproduction occurs at the control ($x_i = 0$), as in the Air Force study, the maximum reproduction is $\lambda_c = e^{\beta_0}$. Suppose it is of interest to estimate the EC_{100p} as accurately as possible. The expected response at the EC_{100p} is $p\lambda_c$. Thus,

$$\begin{aligned} p\lambda_c &= e^{\beta_0+\beta_1(EC_{100p})} \\ \Rightarrow\ pe^{\beta_0} &= e^{\beta_0}e^{\beta_1(EC_{100p})} \\ \Rightarrow\ p &= e^{\beta_1(EC_{100p})} \\ \Rightarrow\ \frac{\ln p}{\beta_1} &= EC_{100p} \quad \text{and} \quad \hat{EC}_{100p} = \frac{\ln p}{\hat{\beta}_1}. \end{aligned}$$

Minkin observed that since p is known, estimating the EC_{100p} well simply involves minimizing the $\text{var}(\hat{\beta}_1)$. Minimizing $\text{var}(\hat{\beta}_1)$ produces the slope-optimal design. The same result can be obtained from the Delta Method. The expression for the approximate variance as derived by the Delta Method is

$$\begin{aligned} \text{var}(\hat{EC}_{100p}) &= \text{var}\left(\frac{\ln p}{\beta_1}\right) \\ &\approx \left(\frac{\ln p}{\beta_1}\right)^2\left(\frac{0}{(\ln p)^2} + \frac{\text{var}(\hat{\beta}_1)}{\beta_1^{\,2}} - \frac{2(0)}{\beta_1 \ln p}\right) \\ &= \left(\frac{(\ln p)^2}{\beta_1^{\,4}}\right)\text{var}(\hat{\beta}_1). \end{aligned}$$

Thus, it is clear that to estimate the EC_{100p} well, one simply needs to estimate β_1 well. Minkin's approach focuses on minimizing the asymptotic variance of the point estimate of the EC_{100p}. What about the interval or F-optimality approach?

For the single regressor exponential model, the Fieller interval reduces to a simple confidence interval on β_1. Recall that b_1, the maximum likelihood estimate for β_1, is asymptotically normal. Thus, asymptotically,

$$\frac{b_1 - \beta_1}{\sqrt{\text{var}(b_1)}} \sim N(0,1). \tag{4.5.1}$$

Letting $\sigma_b = \sqrt{\text{var}(b_1)}$ and $z_{\alpha/2}$ be the standard normal random variable which has upper tail probability $\alpha/2$, a $100(1-\alpha)\%$ asymptotic confidence interval on the EC_{100p} can be derived as follows:

$$\begin{aligned} &\Pr\left(-z_{\alpha/2} \le \frac{b_1 - \beta_1}{\sigma_b} \le z_{\alpha/2}\right) = 1-\alpha \\ \Rightarrow &\Pr\left(z_{\alpha/2} \ge \frac{\beta_1 - b_1}{\sigma_b} \ge -z_{\alpha/2}\right) = 1-\alpha \\ \Rightarrow &\Pr\left(\frac{\ln p}{b_1 - z_{\alpha/2}\sigma_b} \le \frac{\ln p}{\beta_1} \le \frac{\ln p}{b_1 + z_{\alpha/2}\sigma_b}\right) = 1-\alpha \end{aligned} \tag{4.5.2}$$

(since the fact that $0 \le p \le 1$ implies $\ln p \le 0$).

The squared half-width of the interval is

$$\frac{(\ln p)^2 z^2 \text{ var}(b_1)}{(b_1^2 - z^2 \text{ var}(b_1))^2}. \tag{4.5.3}$$

The F-optimal design minimizes this expression. Since the toxicants in this research are assumed to impair reproduction, it is expected that b_1 is a negative number. Thus, for β_1 to be significant, it is necessary that

$$\frac{b_1}{\sqrt{\text{var}(b_1)}} \le -z_{\frac{\alpha}{2}}. \tag{4.5.4}$$

This implies that $b_1^2 - z^2 \text{ var}(b_1) \ge 0$. To minimize the expression in (4.5.3), the denominator must be made as large as possible. Thus, a design must be selected which

makes the var(b_1) as small as possible. Minimizing var(b_1) also minimizes the numerator of (4.5.3). Therefore, the design which makes the var(b_1) as small as possible will minimize the width of the confidence interval on the EC_{100p}.

Clearly, the single task of minimizing the variance of b_1 will satisfy both the slope and F-optimality criteria and, consequently, will provide for the most accurate estimation of the EC_{100p}. From the inverse of the Fisher information matrix,

$$\text{var}(b_1) = \frac{\sum \lambda_i}{\sum \lambda_i \sum x_i^2 \lambda_i - \left(\sum x_i \lambda_i\right)^2}. \qquad (4.5.5)$$

Minkin(1993) showed analytically that the design which minimizes this expression is a two level design with 78% of the observations taken at the $EC_{7.8}$ and 22% of the observations taken at the control. Minkin also examined the effects of poor initial estimates of β_1 on the optimal design, considered a Bayesian approach to obtaining guesses for β_1, and studied the slope-optimal design in the presence of overdispersion.

Three level designs for accurately estimating the EC_{100p} were examined in the this dissertation. The investigation was conducted in a fashion similar to that used for the D-optimal designs, and the results of the investigation appear in Table 4.3. The F-Efficiency is computed as

$$\frac{\text{var}(b_1) \text{ for optimal design}}{\text{var}(b_1) \text{ for the design of interest}}. \qquad (4.5.6)$$

This efficiency measure provides efficiencies which are somewhat lower than those provided by the D and Q-Efficiency measures. The highest F-Efficiency for any restricted F-optimal design is only about 80%. The efficiencies in Table 4.3 seem low compared to the D and Q efficiencies seen earlier. This is due to the nature of the F criterion and due to the regions chosen. Recall that the optimal designs on the entire dose space for all three criteria have two points, a control and a non-control. In the D and Q cases, the restricted regions chosen typically either included the optimal non-control point or came close to including it. Thus, the efficiencies of the designs on restricted regions

Table 4.3. Three Level F-Optimal/Slope-Optimal Designs for the Exponential Poisson Model and the F-Efficiencies for Each.

Case	Region of Interest/Operability	Levels Symmetric about Center	p_c	$p_1=p_2$	x_1	x_2	F-Efficiency	F-Efficiency (restricted)
A	$(EC_0, EC_{100}]$	no	0.218	0.391	$EC_{7.8}$	$EC_{7.8}$	1	1
B	$[EC_{20}, EC_{80}]$	no	0.31	0.345	EC_{20}	EC_{20}	0.7974	1
C	$[EC_{10}, EC_{90}]$	no	0.24	0.380	EC_{10}	EC_{10}	0.9867	1
D	$[EC_{10}, EC_{50}]$	no	0.24	0.380	EC_{10}	EC_{10}	0.9867	1
E	$(EC_0, EC_{100}]$	yes	0	0.5	$EC_{8.3}$	$EC_{91.7}$	0.7080	0.7080
F	$[EC_{20}, EC_{80}]$	yes	0.114	0.443	EC_{20}	EC_{80}	0.5124	0.6426
G	$[EC_{10}, EC_{90}]$	yes	0	0.5	EC_{10}	EC_{90}	0.7004	0.7099
H	$[EC_{10}, EC_{50}]$	yes	0.28	0.36	$EC_{11.7}$	$EC_{48.3}$	0.6162	0.6245

tend to be high. In the F-optimal case, the optimal non-control point was the $EC_{7.8}$. None of the restricted regions included this point. Thus, the F-efficiencies tend to appear low.

Finally, because F-optimality is a very important criterion to biological researchers, and biological researchers typically prefer designs with several levels, some four leveldesigns were sought based on the F-optimality criterion. For a four level design, the criterion is simply an extension of the criteria for the two and three level designs. The best four level designs are given in Table 4.4. In all of these designs, one level was fixed to be the control, and the sample allocation at the three non-control levels was always required to be the same. In some cases (iii, iv, vi, and ix), equal sample allocation was required at all four levels. Note that the third column in the table, labeled "number of equally-spaced levels" always contains a 3 or a 4. The presence of a 3 indicates that the three non-controls were required to be equidistant from each other, while the lower-most non-control level could be any distance from the control. The presence of a 4 in this column indicates that all four levels were required to be equally spaced. Like the optimal three level designs, the optimal four level designs reverted to two levels. Consequently, in cases ii, iv, vii, and x, the three non-control design levels were forced to be distinct. All of the efficiencies in Table 4.4 are restricted in the sense that they compare the design of interest to the optimal design on the same restricted region.

§4.6 Conclusions

It is possible, and in fact, likely in some cases, that a biological researcher will use none of the optimal designs. Many use "traditional designs" which are preferred because they are convenient for the experimenter and not because they are statistically appealing. For example, the design in a single toxicant study may consist of a number of equally spaced levels with the experimental units divided evenly among them. Designs in mixture studies frequently consist of a few set combinations of concentrations with

Table 4.4. Four Level F-Optimal Designs for the Poisson Exponential Model and the F-Efficiencies for Each.

Case	Region	No. Equally-Spaced Levels	p_1, p_2, p_3	p_c	x_1	x_2	x_3	F-Efficiency
i	$(EC_0, EC_{100}]$	3	0.2607	0.2179	$EC_{7.8}$	$EC_{7.8}$	$EC_{7.8}$	1
ii	$(EC_0, EC_{100}]$	4	0.2573	0.2281	$EC_{7.2}$	$EC_{38.1}$	$EC_{69.1}$	0.5161
iii	$(EC_0, EC_{100}]$	3	0.25	0.25	$EC_{8.3}$	$EC_{8.3}$	$EC_{8.3}$	0.9953
iv	$(EC_0, EC_{100}]$	4	0.25	0.25	$EC_{7.3}$	$EC_{38.2}$	$EC_{69.1}$	0.5154
v	$[EC_{20}, EC_{80}]$	3	0.2303	0.3091	EC_{20}	EC_{20}	EC_{20}	1
vi	$[EC_{20}, EC_{80}]$	3	0.25	0.25	EC_{20}	EC_{20}	EC_{20}	0.9818
vii	$[EC_{20}, EC_{80}]$	3	0.2586	0.2242	EC_{20}	EC_{50}	EC_{80}	0.5176
viii	$[EC_{10}, EC_{50}]$	3	0.2532	0.2404	EC_{10}	EC_{10}	EC_{10}	1
ix	$[EC_{10}, EC_{50}]$	3	0.25	0.25	EC_{10}	EC_{10}	EC_{10}	0.9995
x	$[EC_{10}, EC_{50}]$	3	0.2343	0.2971	EC_{10}	EC_{30}	EC_{50}	0.6120

design points at serial dilutions of these mixtures. (This was the initial type of design used in the Air Force study.) The computer algorithms developed for this research can be adapted to find optimal designs in a host of experimental situations. Even when these optimal designs are not used in the actual experiment, efficiencies of the designs which are used can be computed based on these results. Since much of biological research and nonlinear design rely on information from previous experiments, the most efficient designs from a statistical standpoint will be the most useful in both the current and future research.

The optimal designs in this chapter will be discussed further in future chapters. In particular, they will be compared with designs for the Poisson linear model, the optimality of some of the designs which were numerically found will be proven with equivalence theory, and the issue of having to know the parameters to locate the actual optimal dose

Chapter 5

Optimal Designs for the Poisson Linear Model

§5.1 Introduction

In the previous section, Poisson counts were modeled by an exponential model which was created using a log link function. In this section, D, Q, and F-optimal designs are sought for a different Poisson model, the linear model

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij} \tag{5.1.1}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. This is the model formulated using the identity link with the Poisson distribution. Unlike the linear model with normal errors in Chapter 3, this is a heterogeneous variance model. From Chapter 2, the information matrix for this model and an experiment with N experimental runs is

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i}{\lambda_i} \\ \sum_{i=1}^{N} \frac{x_i}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^2}{\lambda_i} \end{bmatrix}. \tag{5.1.2}$$

§5.2 D-Optimal Designs

Recall that the goal in finding a D-optimal design is to maximize $|\mathbf{I}|/N$. In general, for the Poisson error linear model,

$$\frac{|\mathbf{I}|}{N} = \frac{1}{N}\left[\sum \frac{1}{\lambda_i} \sum \frac{x_i^2}{\lambda_i} - \left(\sum \frac{x_i}{\lambda_i}\right)^2\right] \tag{5.2.1}$$

For a two level design, this simplifies to

$$\frac{|\mathbf{I}|}{N} = \left(\frac{1}{N\beta_1^2}\right)\left(\frac{n_1 n_2(\lambda_1 - \lambda_2)^2}{\lambda_1\lambda_2}\right). \tag{5.2.2}$$

The D-optimal design is the design which maximizes this apart from $1/\beta_1^2$ and N. By inspection, n_1n_2 is maximized when $n_1 = n_2$. The Nelder-Mead algorithm was used to find the optimal two level design apart from this product. The optimal design contains two levels spaced very far apart. Recall that the expected response at the control, λ_c, is the maximum possible reproduction for the types of experiments considered in this research. Thus, the possible values for expected responses are in $[0, \lambda_c]$, and the two design levels on the entire region which result in maximum spacing are the EC_0 and the EC_{100}. However, both λ_1 and λ_2 are in the denominator of $|\mathbf{I}|$. Thus, the EC_0 is not a feasible design point. The two level D-optimal design for the model in (5.1.1), then, is a design with half of the observations taken at the EC_{100}, the control, and half of the observations taken at the x_i which yields an expected response as close to 0 as possible. The practitioner's particular situation or experiment will dictate which point has the smallest possible non-zero mean. For the remainder of this section, it will be assumed that the x which yields an expected response as close to 0 as possible is the $EC_{0.01}$, or the x which yields an expected response of $0.0001\lambda_c$.

Once again, since practitioners typically prefer to use designs with more than two levels, optimal three level designs were explored. As usual, one of the three levels was forced to be at the control, and the sample allocation at the two other levels was forced to be the same ($n_1 = n_2$). For three levels,

$$|\mathbf{I}| = \left(\frac{1}{\beta_1^2}\right)\left(\frac{n_1 n_2(\lambda_1 - \lambda_2)^2}{\lambda_1\lambda_2} + \frac{n_1 n_c(\lambda_1 - \lambda_c)^2}{\lambda_1\lambda_c} + \frac{n_2 n_c(\lambda_2 - \lambda_c)^2}{\lambda_2\lambda_c}\right) \tag{5.2.3}$$

The optimal three level designs for various situations are given in Table 5.1. All of these designs were obtained using the Nelder-Mead algorithm. As in the previous section, the optimal three level designs reverted to two level designs. Thus, a column on the table

Table 5.1. Three Level D-Optimal Designs for the Poisson Linear Model and the D-Efficiencies for Each.

Case	Region of Interest/Operability	Levels Symmetric about Center	p_c	$p_1=p_2$	x_1	x_2	D-Efficiency	D-Efficiency (restricted)
A	$[EC_{0.01}, EC_{100}]$	no	0	0.5	$EC_{0.01}$	EC_{100}	1	1
B	$[EC_{20}, EC_{80}]$	no	0.5	0.25	EC_{20}	EC_{20}	0.0179	1
C	$[EC_{10}, EC_{90}]$	no	0.5	0.25	EC_{10}	EC_{10}	0.0285	1
D	$[EC_{10}, EC_{50}]$	no	0.5	0.25	EC_{10}	EC_{10}	0.0285	1
E	$[EC_{0.01}, EC_{100}]$	yes	0	0.5	$EC_{0.01}$	EC_{100}	1	1
F	$[EC_{20}, EC_{80}]$	yes	0.236	0.382	EC_{20}	EC_{80}	0.0158	0.8813
G	$[EC_{10}, EC_{90}]$	yes	0.11	0.445	EC_{10}	EC_{90}	0.0269	0.9442
H	$[EC_{10}, EC_{50}]$	yes	0.386	0.307	EC_{10}	EC_{50}	0.0230	0.8076

indicates whether or not the two non-control design levels were forced to be distinct, creating a true three level design. Recall that D-Efficiencies for designs with the same number of experimental units are computed as follows:

$$\text{D} - \text{Efficiency} = \left(\frac{|\mathbf{I}| \text{ for the design of interest}}{|\mathbf{I}| \text{ for optimal design}} \right)^{\frac{1}{2}}. \tag{5.2.4}$$

In the first D-Efficiency column, the column comparing all designs to the overall optimal design, the efficiencies are very low. This is because the overall optimal design has an expected response at non-control design level as close to 0 as possible. As the expected response approaches 0, the determinant of $\mathbf{I}$ becomes infinite, and, thus, the expression in (5.2.4) becomes close to 0. The D-Efficiencies for the designs on restricted regions are much better, and one should bear in mind that these are really the only sensible efficiencies.

§5.3 Q-Optimal Designs

The Q-optimality criterion, once again, is

$$\min_{\mathcal{D}} \frac{\text{N}}{\text{K}} \int_{\text{R}} \text{var}(\hat{\text{f}}(\mathbf{x}_0, \hat{\beta})) \partial \mathbf{x}_0 \tag{5.3.1}$$

where R is the region of interest and K is the volume of that region. From Chapter 2, the general expression for the approximate prediction variance at a point, $\mathbf{x}_0$, is

$$\text{var}(\text{f}(\mathbf{x}_0, \hat{\beta})) \approx \mathbf{w}(\mathbf{x}_0)' \mathbf{I}^{-1} \mathbf{w}(\mathbf{x}_0) \tag{5.3.2}$$

where

$$\mathbf{w}(\mathbf{x}_0)' = \left[\frac{\partial \text{f}(\mathbf{x}_0, \beta)}{\partial \beta_1}, \cdots, \frac{\partial \text{f}(\mathbf{x}_0, \beta)}{\partial \beta_p} \right]. \tag{5.3.3}$$

For the Poisson linear model, $\hat{\text{f}}(\mathbf{x}_0, \hat{\beta}) = \hat{\text{y}}(\mathbf{x}_0) = \hat{\beta}_0 + \hat{\beta}_1 \text{x}_0$, $\mathbf{w}(\mathbf{x}_0)' = \mathbf{x}_0'$ and

$$\mathbf{I}^{-1} = \frac{1}{\sum\frac{1}{\lambda_i}\sum\frac{x_i^2}{\lambda_i} - \left(\sum\frac{x_i}{\lambda_i}\right)^2}\begin{pmatrix} \sum\frac{x_i^2}{\lambda_i} & -\sum\frac{x_i}{\lambda_i} \\ -\sum\frac{x_i}{\lambda_i} & \sum\frac{1}{\lambda_i} \end{pmatrix}. \tag{5.3.4}$$

Thus,

$$\mathrm{var}(f(\mathbf{x}_0,\hat{\beta})) \approx \frac{\sum\frac{x_i^2}{\lambda_i} - 2x_0\sum\frac{x_i}{\lambda_i} + x_0^2\sum\frac{1}{\lambda_i}}{\sum\frac{1}{\lambda_i}\sum\frac{x_i^2}{\lambda_i} - \left(\sum\frac{x_i}{\lambda_i}\right)^2}. \tag{5.3.5}$$

Let x_U be the upper boundary on the region of interest and x_L be the lower boundary on the region of interest. Then the final integrated expression for Q is

$$Q = \left(\frac{N}{3(x_U - x_L)}\right)\left(\frac{3(x_U - x_L)\sum\frac{x_i^2}{\lambda_i} - 3(x_U^2 - x_L^2)\sum\frac{x_i}{\lambda_i} + (x_U^3 - x_L^3)\sum\frac{1}{\lambda_i}}{\sum\frac{1}{\lambda_i}\sum\frac{x_i^2}{\lambda_i} - \left(\sum\frac{x_i}{\lambda_i}\right)^2}\right).$$

The simplified expression for a two level design is

$$Q = \left(\frac{N}{3}\right)\left(\frac{3\lambda_1\lambda_2[(n_1\lambda_1 + n_2\lambda_2) - (\lambda_U + \lambda_L)N]}{n_1n_2(\lambda_1 - \lambda_2)^2} + \frac{(n_1\lambda_2 + n_2\lambda_1)(\lambda_U^2 + \lambda_U\lambda_L + \lambda_L^2)}{n_1n_2(\lambda_1 - \lambda_2)^2}\right). \tag{5.3.6}$$

All of the Q-optimal designs were obtained using the Nelder-Mead algorithm. The overall optimal two level design has approximately 1% of the observations taken at the $EC_{0.01}$ and approximately 99% of the observations taken at the control. Again, the optimal design requires the two design points to be placed as far apart as possible. Thus, one point is again required to have an expected response near 0. As with the D-optimal design, the $EC_{0.01}$ was assumed to be this point, and 1% was the optimal sample allocation there. However, if a different point with an expected response near 0 was

chosen (i.e. the $EC_{0.1}$ or the $EC_{0.001}$), a different optimal sample allocation would have been selected for that point.

Three level Q-optimal designs were also sought in a fashion similar to that in Chapter 4. The Q expression for three levels, after simplification, is

$$Q = \frac{NH}{3D} \tag{5.3.7}$$

where

$$\begin{aligned} H = {} & 3\lambda_1\lambda_2\lambda_c[(n_1\lambda_1 + n_2\lambda_2 + n_c\lambda_c) - 3N(\lambda_U + \lambda_L)] \\ & + (n_1\lambda_2\lambda_c + n_2\lambda_1\lambda_c + n_c\lambda_1\lambda_2)(\lambda_U^2 + \lambda_U\lambda_L + \lambda_L^2) \end{aligned} \tag{5.3.8}$$

and

$$D = n_1n_2\lambda_c(\lambda_1 - \lambda_2)^2 + n_1n_c\lambda_2(\lambda_1 - \lambda_c)^2 + n_2n_c\lambda_1(\lambda_2 - \lambda_c)^2. \tag{5.3.9}$$

Q-optimal three level designs for a variety of cases are given in Table 5.2. The Q-Efficiencies are fairly low for the same reasons the D-Efficiencies were low. The Q expression can be made very small as the lower optimal level is made closer to 0. Thus, Q values for other designs will compare poorly to this value. As with the D-optimal designs, however, the restricted Q-Efficiencies are the only sensible efficiencies, and these are quite high.

§5.4 Designing Experiments to Estimate the EC_{100p}

The F-optimality criterion, recall, seeks a design which minimizes the width of a Fieller interval around the EC_{100p} for some p of interest. As we saw in section 4.5, in the log link case, F-optimality ultimately required minimization of the variance of b_1, the maximum likelihood estimate of β_1. In the identity link case, F-optimality is more complicated. Recall that the model which results from use of the identity link is

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij} \tag{5.4.1}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. At the control, $x_i = 0$ and $E(y_{ij}) = \lambda_c$. Thus, $\beta_0 = \lambda_c$, and the expression for the estimate of the EC_{100p} is derived as follows:

Table 5.2. Three Level Q-Optimal Designs for the Poisson Linear Model and the Q-Efficiencies for Each.

Case	Region of Operability	Region of Interest	Symmetric Levels	p_c	$p_1=p_2$	x_1	x_2	Q-Efficiency (whole region)	Q-Efficiency (restricted)
A	$[EC_{0.01}, EC_{100}]$	$[EC_{0.01}, EC_{100}]$	no	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3173	1
B	$[EC_{0.01}, EC_{100}]$	$[EC_{20}, EC_{80}]$	no	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3778	1
C	$[EC_{0.01}, EC_{100}]$	$[EC_{10}, EC_{90}]$	no	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3488	1
D	$[EC_{0.01}, EC_{100}]$	$[EC_{10}, EC_{50}]$	no	0.956	0.022	$EC_{0.01}$	EC_{100}	1	1
E	$[EC_{0.01}, EC_{100}]$	$[EC_{0.01}, EC_{100}]$	yes	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3173	1
F	$[EC_{0.01}, EC_{100}]$	$[EC_{20}, EC_{80}]$	yes	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3778	1
G	$[EC_{0.01}, EC_{100}]$	$[EC_{10}, EC_{90}]$	yes	0.980	0.010	$EC_{0.01}$	EC_{100}	0.3488	1
H	$[EC_{0.01}, EC_{100}]$	$[EC_{10}, EC_{50}]$	yes	0.956	0.022	$EC_{0.01}$	EC_{100}	1	1
b	$[EC_{20}, EC_{80}]$	$[EC_{20}, EC_{80}]$	no	0.594	0.203	EC_{20}	EC_{20}	0.2031	1
c	$[EC_{10}, EC_{90}]$	$[EC_{10}, EC_{90}]$	no	0.508	0.137	EC_{10}	EC_{10}	0.2161	1
d	$[EC_{10}, EC_{50}]$	$[EC_{10}, EC_{50}]$	no	0.726	0.246	EC_{10}	EC_{10}	0.4216	1
f	$[EC_{20}, EC_{80}]$	$[EC_{20}, EC_{80}]$	yes	0.314	0.343	EC_{20}	EC_{80}	0.1922	0.9460
g	$[EC_{10}, EC_{90}]$	$[EC_{10}, EC_{90}]$	yes	0.49	0.255	EC_{10}	EC_{90}	0.2115	0.9788
h	$[EC_{10}, EC_{50}]$	$[EC_{10}, EC_{50}]$	yes	0.296	0.352	EC_{10}	EC_{50}	0.3727	0.8838

$$\lambda_i = \beta_0 + \beta_1 x_i$$

$$\Rightarrow \quad p\lambda_c = \lambda_c + \beta_1 EC_{100p}$$

$$\Rightarrow \quad EC_{100p} = \frac{\lambda_c(p-1)}{\beta_1} \tag{5.4.2}$$

$$\Rightarrow \quad \hat{EC}_{100p} = \frac{\hat{\lambda}_c(p-1)}{b_1}.$$

Now, letting $x_p = EC_{100p}$,

$$E(\hat{\lambda}_c(p-1) - \hat{\beta}_1 x_p) = (p-1)\lambda_c - \beta_1 x_p = 0 \tag{5.4.3}$$

and

$$\begin{aligned} Var(\hat{\lambda}_c(p-1) - \hat{\beta}_1 x_p) &= (p-1)^2\, var(\hat{\lambda}_c) + x_p{}^2\, var(\hat{\beta}_1) - 2x_p(p-1)\, cov(\hat{\lambda}_c, \hat{\beta}_1) \\ &= (p-1)^2\, var(b_0) + x_p{}^2\, var(b_1) - 2x_p(p-1)\, cov(b_0, b_1) \end{aligned} \tag{5.4.4}$$

where b_i is the MLE of β_i. Asymptotically, then,

$$\frac{\hat{\lambda}_c(p-1) - \hat{\beta}_1 x_p - 0}{\sqrt{(p-1)^2\, var(\hat{\lambda}_c) + x_p{}^2\, var(\hat{\beta}_1) - 2x_p(p-1)\, cov(\hat{\lambda}_c, \hat{\beta}_1)}} \sim N(0,1). \tag{5.4.5}$$

Let $v_{ii} = var(b_i)$ and $v_{ij} = cov(b_i, b_j)$. The $(1-\alpha)100\%$ Fieller interval on the estimate of the EC_{100p} is derived from

$$-z_{\frac{\alpha}{2}} \le \frac{(p-1)b_0 - b_1 x_p}{\sqrt{(p-1)^2 v_{00} + x_p{}^2 v_{11} - 2x_p(p-1)v_{01}}} \le z_{\frac{\alpha}{2}}. \tag{5.4.6}$$

where $z_{(\alpha/2)}$ is the standard normal random variable with right tail probability $\alpha/2$. (For the remainder of this section, the subscript $\alpha/2$ will be dropped from the z.) Squaring this equation and simplifying yields the following quadratic equation

$$(b_1{}^2 - z^2 v_{11})x_p{}^2 - 2(p-1)(b_0 b_1 - z^2 v_{01})x_p + (p-1)^2(b_0{}^2 - z^2 v_{00}) = 0. \tag{5.4.7}$$

The roots of this equation are the endpoints of the Fieller interval. After some simplification, the roots are given by

$$\frac{(p-1)(b_0b_1 - z^2v_{01}) \pm (p-1)\sqrt{(b_0b_1 - z^2v_{01})^2 - (b_1^2 - z^2v_{11})(b_0^2 - z^2v_{00})}}{b_1^2 - z^2v_{11}}. \quad (5.4.8)$$

Thus, the squared half-width of the Fieller interval is given by

$$\frac{(p-1)^2(z^2(z^2(v_{01}^2 - v_{00}v_{11}) + b_0^2v_{11} + b_1^2v_{00} - 2b_0b_1v_{01}))}{(b_1^2 - z^2v_{11})^2}$$

$$= \frac{(p-1)^2\left(\frac{z^2}{b_1^2}\right)\left(\frac{-z^2}{b_1^2|\mathbf{I}|} + \left(\frac{b_0}{b_1}\right)^2 v_{11} + v_{00} - \left(\frac{b_0}{b_1}\right)v_{01}\right)}{\left(1 - \frac{z^2v_{11}}{b_1^2}\right)^2}. \quad (5.4.9)$$

After substitution of variances and covariances from the inverse of the information matrix and after extensive simplification, the squared Fieller width for a two level design is

$$\frac{(p-1)^2z^2}{\beta_1^2(1-g)^2}\left[\frac{\lambda_1\lambda_2(n_1\lambda_1 + n_2\lambda_2 - z^2)}{n_1n_2(\lambda_1 - \lambda_2)^2}\right] \quad (5.4.10)$$

where $g = z^2v_{11}/b_1^2$, λ_1 and λ_2 are the expected responses at the two F-optimal design points and n_1 and n_2 are the number of observations at those points, respectively. Now, $0 \le g < 1$ for β_1 to be significant. Since λ_c is the maximum expected response, all design points will have an expected response of $q\lambda_c$ where $0 < q \le 1$. Thus, $\lambda_1 = q_1\lambda_c$ and $\lambda_2 = q_2\lambda_c$. Also, if N is the total number of observations and p_1 is the proportion of those points allocated to the first design point, $n_1 = Np_1$ and $n_2 = N(1-p_1)$. With these substitutions, the final expression for the squared half-width for a two level design is

$$\frac{z^2\lambda_c(p-1)^2}{N\beta_1^2}\left[\frac{q_1q_2\left(p_1q_1 + (1-p_1)q_2 - \frac{z^2}{N\lambda_c}\right)}{p_1(1-p_1)(q_1-q_2)^2\left(1 - \frac{z^2}{N\lambda_c}\left(\frac{p_1q_2 + (1-p_1)q_1}{p_1(1-p_1)(q_1-q_2)^2}\right)\right)}\right]. \quad (5.4.11)$$

The criterion value depends on design and on the ratio $z^2/N\lambda_c$. One should note that it does *not* depend upon p, which specifies the EC_{100p} of interest. In most applications, the

values of N and λ_c are at least moderately large, and even for moderate values of N and λ_c, $z^2/N\lambda_c$ is close to 0.

As in the D-optimality and Q-optimality cases, F-optimality favors a design with the two design levels placed as far apart as possible. Thus, the control was one design level and the $EC_{0.01}$ was again used as the design level as far away from the control as possible. Table 5.3 shows the optimal two level designs for various values of the ratio, $z^2/N\lambda_c$. In the table, $z^2/N\lambda_c$ is always less than 1 because optimal designs do not exist for situations in which this ratio is greater than or equal to 1. Furthermore, depending upon how close the expected response at the upper optimal design point can get to 0, optimal designs may not exist for some values of the ratio which are less than 1. Nevertheless, as stated before, even for moderate values of N and λ_c, this ratio will be small enough that optimal designs do exist.

Note that 0 is listed in Table 5.3 as a possible value for the ratio, $z^2/N\lambda_c$. Although, in reality, this ratio can never be exactly 0, it turns out that the expression for

Table 5.3. Two Level F-Optimal Designs for the Poisson Linear Model on the Region $[EC_{0.01}, EC_{100}]$.

$\frac{z^2}{N\lambda_c}$	p_1	x_1	x_2
0	0.9901	$EC_{0.01}$	EC_{100}
0.0001	0.9859	$EC_{0.01}$	EC_{100}
0.001	0.9668	$EC_{0.01}$	EC_{100}
0.01	0.8995	$EC_{0.01}$	EC_{100}
0.1	0.6836	$EC_{0.01}$	EC_{100}
0.2	0.5526	$EC_{0.01}$	EC_{100}
0.5	0.2927	$EC_{0.01}$	EC_{100}
0.9	0.0511	$EC_{0.01}$	EC_{100}
0.95	0.0252	$EC_{0.01}$	EC_{100}

the asymptotic variance of the $\hat{EC}_{100p}$ is identical to the expression for the F-optimality criterion when $z^2/N\lambda_c$ is considered to be 0. Thus, the design in the table which corresponds to $z^2/N\lambda_c = 0$ is the design which minimizes the asymptotic variance of the *point* estimate of the EC_{100p}.

Three level F-optimal designs were found for the Poisson linear model as well. For each of the three level designs, one level was fixed to be the control and the sample size at the two non-control design points was required to be the same. Thus, $\lambda_1 = q_1\lambda_c$, $\lambda_2 = q_2\lambda_c$, and $\lambda_3 = \lambda_c$. Also, if p_1 is the proportion of observations allocated to the non-control design points x_1 and x_2, then $n_1 = n_2 = Np_1$ and $n_3 = (1\text{-}2p_1)N$. The final expression for the squared half-width for this type of three level design is

$$\frac{(p-1)^2 z^2 \lambda_c}{\beta_1^2 N(1-g)^2}\left[\frac{q_1 q_2\left(p_1(q_1+q_2)+(1-2p_1)-\dfrac{z^2}{N\lambda_c}\right)}{p_1^2(q_1-q_2)^2+p_1(1-2p_1)(q_2(q_1-1)^2+q_1(q_2-1)^2)}\right] \qquad (5.4.12)$$

where

$$g=\frac{z^2}{N\lambda_c}\left[\frac{p_1(q_1+q_2)+(1-2p_1)q_1q_2}{p_1^2(q_1-q_2)^2+p_1(1-2p_1)(q_2(q_1-1)^2+q_1(q_2-1)^2)}\right]. \qquad (5.4.13)$$

The optimal three level designs for various situations are given in Tables 5.4(a) and 5.4(b). The columns for the F-efficiencies compare designs by

$$\text{F}-\text{Efficiency}=\frac{\text{squared Fieller half - width for optimal design}}{\text{squared Fieller half - width for design of interest}}. \qquad (5.4.14)$$

Once again, most of the designs compare very poorly to the overall optimal design but fare much better when compared with other designs on exactly the same region.

A few four level F-optimal designs were also sought. All of these were obtained under the assumption that $z^2/N\lambda_c = 0$. This is because $z^2/N\lambda_c$ is near 0 in most experimental situations, and it is evident in Tables 5.3 and 5.4 that the designs for values of this ratio close to 0 are very similar to the design for $z^2/N\lambda_c = 0$. The F-optimality criterion for a four level design is simply an extension of the criterion for three levels.

Table 5.4(a). Three Level F-Optimal Designs for the Poisson Linear Model and the F-Efficiencies for Each.

Case	Region of Interest/ Operability	$\frac{z^2}{N\lambda_c}$	Levels Symmetric about Center	p_c	$p_1=p_2$	x_1	x_2	F-Efficiency	F-Efficiency (restricted)
A	$[EC_{0.01}, EC_{100}]$	0.1	no	0.3164	0.3418	$EC_{0.01}$	$EC_{0.01}$	1	1
B	$[EC_{20}, EC_{80}]$	0.1	no	0.5	0.25	EC_{20}	EC_{20}	0.0001	1
C	$[EC_{10}, EC_{90}]$	0.1	no	0.4348	0.2826	EC_{10}	EC_{10}	0.0005	1
D	$[EC_{10}, EC_{50}]$	0.1	no	0.4348	0.2826	EC_{10}	EC_{10}	0.0005	1
E	$[EC_{0.01}, EC_{100}]$	0.1	yes	0	0.5	$EC_{0.01}$	EC_{100}	0.8558	0.8558
F	$[EC_{20}, EC_{80}]$	0.1	yes	0.2628	0.3686	EC_{20}	EC_{80}	0.00009	0.6424
G	$[EC_{10}, EC_{90}]$	0.1	yes	0.0074	0.4963	EC_{10}	EC_{90}	0.0004	0.8570
H	$[EC_{10}, EC_{50}]$	0.1	yes	0.3416	0.3292	EC_{10}	EC_{50}	0.0003	0.5537
A	$[EC_{0.01}, EC_{100}]$	0.01	no	0.1006	0.4497	$EC_{0.01}$	$EC_{0.01}$	1	1
B	$[EC_{20}, EC_{80}]$	0.01	no	0.3364	0.3318	EC_{20}	EC_{20}	0.0002	1
C	$[EC_{10}, EC_{90}]$	0.01	no	0.2684	0.3658	EC_{10}	EC_{10}	0.0005	1
D	$[EC_{10}, EC_{50}]$	0.01	no	0.2684	0.3658	EC_{10}	EC_{10}	0.0005	1
E	$[EC_{0.01}, EC_{100}]$	0.01	yes	0	0.5	$EC_{0.01}$	EC_{100}	0.6055	0.6055
F	$[EC_{20}, EC_{80}]$	0.01	yes	0.0126	0.4937	EC_{20}	EC_{80}	0.0001	0.7312
G	$[EC_{10}, EC_{90}]$	0.01	yes	0	0.5	EC_{10}	EC_{90}	0.0004	0.7826
H	$[EC_{10}, EC_{50}]$	0.01	yes	0.1394	0.4303	EC_{10}	EC_{50}	0.0003	0.5959

Table 5.4(b). Three Level F-Optimal Designs for the Poisson Linear Model and the F-Efficiencies for Each.

Case	Region of Interest/ Operability	$\frac{z^2}{N\lambda_c}$	Levels Symmetric about Center	p_c	$p_1=p_2$	x_1	x_2	F-Efficiency	F-Efficiency (restricted)
A	$[EC_{0.01}, EC_{100}]$	0.001	no	0.0332	0.4834	$EC_{0.01}$	$EC_{0.01}$	1	1
B	$[EC_{20}, EC_{80}]$	0.001	no	0.312	0.344	EC_{20}	EC_{20}	0.0002	1
C	$[EC_{10}, EC_{90}]$	0.001	no	0.2432	0.3784	EC_{10}	EC_{10}	0.0005	1
D	$[EC_{10}, EC_{50}]$	0.001	no	0.2432	0.3784	EC_{10}	EC_{10}	0.0005	1
E	$[EC_{0.01}, EC_{100}]$	0.001	yes	0	0.5	$EC_{0.01}$	EC_{100}	0.5337	0.5337
F	$[EC_{20}, EC_{80}]$	0.001	yes	0	0.5	EC_{20}	EC_{80}	0.0001	0.7360
G	$[EC_{10}, EC_{90}]$	0.001	yes	0	0.5	EC_{10}	EC_{90}	0.0004	0.7629
H	$[EC_{10}, EC_{50}]$	0.001	yes	0.1096	0.4452	EC_{10}	EC_{50}	0.0003	0.5982
A	$[EC_{0.01}, EC_{100}]$	0	no	0.0098	0.4951	$EC_{0.01}$	$EC_{0.01}$	1	1
B	$[EC_{20}, EC_{80}]$	0	no	0.309	0.3455	EC_{20}	EC_{20}	0.0002	1
C	$[EC_{10}, EC_{90}]$	0	no	0.2402	0.3799	EC_{10}	EC_{10}	0.0005	1
D	$[EC_{10}, EC_{50}]$	0	no	0.2402	0.3799	EC_{10}	EC_{10}	0.0005	1
E	$[EC_{0.01}, EC_{100}]$	0	yes	0	0.5	$EC_{0.01}$	EC_{100}	0.5100	0.5100
F	$[EC_{20}, EC_{80}]$	0	yes	0	0.5	EC_{20}	EC_{80}	0.0001	0.7363
G	$[EC_{10}, EC_{90}]$	0	yes	0	0.5	EC_{10}	EC_{90}	0.0004	0.7605
H	$[EC_{10}, EC_{50}]$	0	yes	0.106	0.447	EC_{10}	EC_{50}	0.0003	0.5834

The four level designs are given in Table 5.5, and the columns have the same interpretations as in Chapter 4.

§5.5 Conclusions

The remarks made in the conclusion of the previous chapter pertain to the designs in this chapter as well. The algorithms developed for this model may be used to find optimal designs in many situations, and even when not used in the actual experiment, these designs may at least be used to determine the efficiencies of the designs that are used. Now that designs for two GLM Poisson models have been obtained, it will be interesting to compare the two sets of designs, and study the effects of parameter misspecification on them. Before working with the designs further, however, it would be nice to at least verify that the numerically obtained optimal designs in Chapters 4 and 5 are indeed optimal. The tool which will be used to do this is equivalence theory, the topic of the next chapter.

Table 5.5. Four Level F-Optimal Designs for the Poisson Linear Model and the F-Efficiency for Each.

Case	Region	No. Equally-Spaced Levels	p_1, p_2, p_3	p_c	x_1	x_2	x_3	F-Efficiency
i	$[EC_{0.01}, EC_{100}]$	3	0.3333	0	$EC_{0.01}$	$EC_{27.3}$	$EC_{54.6}$	0.3398
ii	$[EC_{0.01}, EC_{100}]$	4	0.3333	0	$EC_{0.01}$	$EC_{33.3}$	$EC_{66.7}$	0.3397
iii	$[EC_{0.01}, EC_{100}]$	3	0.25	0.25	$EC_{0.01}$	$EC_{14.9}$	$EC_{29.9}$	0.2550
iv	$[EC_{0.01}, EC_{100}]$	4	0.25	0.25	$EC_{0.01}$	$EC_{33.3}$	$EC_{66.7}$	0.2550
v	$[EC_{20}, EC_{80}]$	3	0.2303	0.3091	EC_{20}	EC_{20}	EC_{20}	1
vi	$[EC_{20}, EC_{80}]$	3	0.25	0.25	EC_{20}	EC_{20}	EC_{20}	0.9817
vii	$[EC_{20}, EC_{80}]$	3	0.3103	0.0691	EC_{20}	EC_{50}	EC_{80}	0.4963
viii	$[EC_{10}, EC_{50}]$	3	0.2532	0.2404	EC_{10}	EC_{10}	EC_{10}	1
ix	$[EC_{10}, EC_{50}]$	3	0.25	0.25	EC_{10}	EC_{10}	EC_{10}	0.9995
x	$[EC_{10}, EC_{50}]$	3	0.2797	0.1609	EC_{10}	EC_{30}	EC_{50}	0.4334

Chapter 6

Equivalence Theory in Design Optimality

§6.1 Introduction

In Chapters 4 and 5, D, Q, and F-optimal designs were obtained for two separate models: an exponential Poisson model and a linear Poisson model. The designs were obtained largely through the use of numerical minimization. Since they were not obtained directly from criteria using mathematical minimization and maximization techniques, one cannot be assured that these are theoretically the optimal designs. Equivalence theory is a tool by which other properties of the optimal designs which *can* be shown using measure theory and calculus are shown to be equivalent to the property of the design satisfying the criterion of interest.

Traditional equivalence theory, originally developed by Kiefer and Wolfowitz (1960), is used to show the D-optimality of a design. Kiefer and Wolfowitz introduced equivalence theory for linear models, and Federov (1972) generalized their results to include nonlinear models. A type of equivalence theory pertaining to other optimality criteria as they apply to both linear and nonlinear models was developed by Silvey (1980). Silvey's theory includes D-optimality as well as a variety of optimality criteria whose criteria functions can be categorized as "linear criteria functions." The Q-optimality criterion function and the F-optimality criterion function for the exponential model qualify. The F-optimality criterion function for the linear model does not. Thus, in this section, equivalence theory is used to verify the optimality of D and Q-optimal designs for both models and F-optimal design for the exponential model only.

Recall that for both the exponential and the linear model, the optimal designs on the various regions specified were *two* level designs. Although optimal three level

designs were sought in each case, the two level designs were actually preferred by all of the optimality criteria. These two level optimal designs are the designs to which equivalence theory is applied in this section.

§6.2 Equivalence Theory

The terminology of Silvey will be used throughout this section. Consider the model

$$y = f(\mathbf{x}, \beta) + \varepsilon \tag{6.2.1}$$

where β is a p x 1 vector of fixed, unknown model parameters and $f(\mathbf{x}, \beta)$ is a known linear or nonlinear function. Now consider the problem of designing an experiment to estimate β using some design criterion function, ϕ. It is assumed without loss of generality that the best design according to the criterion, ϕ, or the ϕ-optimal design is the design which *maximizes* ϕ. (If the design criterion is one which requires the minimization of some function, g, then ϕ is taken to be the negative of g.) Thus, the overall goal of optimal design is to find a design with N observations which maximizes ϕ.

An N-observation design consists of m distinct design points, $\mathbf{x}_1, \mathbf{x}_2,...,\mathbf{x}_m$, with each $\mathbf{x}_i$ replicated r_i times, i = 1,...,m. If $\mathcal{X}$ is the set of all vectors, $\mathbf{x}$, in the design space, then, according to Silvey, any N-observation design may be characterized by the probability distribution, η_N, on $\mathcal{X}$ which assigns probability r_i/N to the point $\mathbf{x}_i$, i = 1,...,m. This is a discrete probability distribution, and an N-observation design is considered a discrete design.

Now, recall that the Fisher Information Matrix is given by

$$\mathbf{I}(\mathbf{x},\beta) = -E\left[\frac{\partial^2 \ln L(\beta)}{\partial\beta'\partial\beta}\right] \tag{6.2.2}$$

where $L(\beta)$ is the likelihood function for the data. Let $\mathbf{J}(\mathbf{x}, \beta)$ be the information matrix for a *single observation* taken at the point $\mathbf{x}$, a random vector with probability distribution η_N, and let

$$\mathbf{M}(\eta_N, \beta) = E(\mathbf{J}(\mathbf{x}, \beta)) = \frac{1}{N}\mathbf{I}(\mathbf{x},\beta)\,. \qquad (6.2.3)$$

Note that this implies that $\mathbf{I}(\mathbf{x}, \beta) = N\mathbf{M}(\eta_N, \beta)$. If designs are viewed as design measures, η_N, then finding an optimal design according to the criterion function ϕ can be viewed as of finding the design, η_N^*, which maximizes $\phi(\mathbf{M}(\eta_N, \beta))$.

From this point of view, ϕ is defined on the set of η_N which is a "discrete domain" (Silvey, p.13). Thus, calculus techniques cannot be employed in finding the optimal designs. However, Silvey states that if H is defined to be the set of *all* η on $\mathcal{X}$, then calculus techniques can be used to find the ϕ-optimal design, η^*, the design which maximizes $\phi(\mathbf{M}(\eta, \beta))$ over all H. The original design problem, then, becomes a problem of finding an N-observation design with a probability distribution close to η^*. Both Kiefer and Silvey refer to this as the approximate theory problem. According to Silvey, there is always a discrete design with at most (1/2)(k)(k+1) + 1 points which maximizes $\phi(\mathbf{M}(\eta, \beta))$ over H, and thus, the "ultimate aim [of optimal design] is to approximate this discrete η^* by a design measure corresponding to an N-observation design" (Silvey, p.16).

A directional derivative called the Frechet directional derivative plays a pivotal role in Silvey's design theory. The Frechet directional derivative of ϕ at $\mathbf{M}_1$ in the direction of $\mathbf{M}_2$ is defined as

$$F_\phi(\mathbf{M}_1,\mathbf{M}_2) = \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon}\left[\phi\{(1-\varepsilon)\mathbf{M}_1 + \varepsilon\mathbf{M}_2\} - \phi(\mathbf{M}_1)\right] \qquad (6.2.4)$$

(Silvey, p. 18). This definition is obtained from the traditional definition of a directional derivative by taking the derivative of ϕ at $\mathbf{M}_1$ in the direction of $\mathbf{M}_2 - \mathbf{M}_1$. Letting $\mathcal{M} = \{\, \mathbf{M}(\eta, \beta) : \eta \in H\}$, and using the Frechet derivative, Silvey proves the following:

Theorem 1: If β is fixed and if ϕ is increasing and concave on $\mathcal{M}$ and differentiable at $\mathbf{M}(\eta^*, \beta)$, then η^* is ϕ-optimal if and only if $F_\phi[\mathbf{M}(\eta^*, \beta), \mathbf{J}(\mathbf{x}, \beta)] \le 0$ for all $\mathbf{x} \in \mathcal{X}$ (Silvey, p. 54). (6.2.5)

This theorem is the cornerstone of Silvey's equivalence theory. When various criterion functions are substituted for ϕ, the optimality of particular designs from Chapters 4 and 5 can be shown.

§6.3 Equivalence Theory for D-Optimal Designs

Recall that a D-optimal design is a design for which the determinant of the information matrix is maximized. Thus, for D-optimality, $\phi(\mathbf{M}(\eta, \beta)) = |\mathbf{M}(\eta, \beta)|$, although Silvey actually uses the log of the determinant. Silvey shows that for this definition of the criterion function, ϕ,

$$F_{\phi}(\mathbf{M}(\eta, \beta), \mathbf{J}(\mathbf{x}, \beta)) = \text{tr}(\mathbf{J}(\mathbf{x}, \beta)\mathbf{M}(\eta, \beta)^{-1}) - p \qquad (6.3.1)$$

where p is the number of unknown parameters in the model in (6.2.1). Using the theorem in (6.2.5), the design η^* is optimal if and only if $F_{\phi}[\mathbf{M}(\eta^*, \beta), \mathbf{J}(\mathbf{x}, \beta)] \leq 0$ for all $\mathbf{x} \in \mathcal{X}$. Now,

$$\begin{aligned} & F_{\phi}[\mathbf{M}(\eta^*, \beta), \mathbf{J}(\mathbf{x}, \beta)] \leq 0 \\ \Rightarrow\ & \text{tr}(\mathbf{J}(\mathbf{x}, \beta)\, \mathbf{M}(\eta^*, \beta)^{-1}) - p \leq 0 \\ \Rightarrow\ & \text{tr}(\mathbf{J}(\mathbf{x}, \beta)\, \mathbf{M}(\eta^*, \beta)^{-1}) \leq p. \end{aligned} \qquad (6.3.2)$$

Note that if $\mathbf{J}(\mathbf{x}, \beta)$ can be written as $\mathbf{v}(\mathbf{x})\mathbf{v}(\mathbf{x})'$,

$$\begin{aligned} \text{tr}(\mathbf{J}(\mathbf{x}, \beta)\, \mathbf{M}(\eta^*, \beta)^{-1}) &= \text{tr}(\mathbf{v}(\mathbf{x})\mathbf{v}(\mathbf{x})'N\mathbf{I}(\mathbf{x}, \beta)^{-1}) \\ &= N\text{tr}(\mathbf{v}(\mathbf{x})'\mathbf{I}(\mathbf{x}, \beta)^{-1}\mathbf{v}(\mathbf{x})) \\ &= N\, \mathbf{v}(\mathbf{x})'\mathbf{I}(\mathbf{x}, \beta)^{-1}\mathbf{v}(\mathbf{x}). \end{aligned} \qquad (6.3.3)$$

For both the exponential Poisson model and the linear Poisson model, p = 2. Thus, the two level D-optimal designs obtained in Chapters 4 and 5 are indeed D-optimal if, when these designs are substituted for η^*, they satisfy

$$N\, \mathbf{v}(\mathbf{x})'\mathbf{I}(\mathbf{x}, \beta)^{-1}\mathbf{v}(\mathbf{x}) \leq 2 \quad \text{for all } \mathbf{x} \in \mathcal{X}. \qquad (6.3.4)$$

This result is consistent with Federov's results for nonlinear models.

§6.3.1 Federov's Equivalence and Its Application to the Normal Error Linear Model

Federov defined a generalized scaled prediction variance to be

$$G(\eta, \mathbf{x}) = N \, tr(\mathbf{I}(\eta, \beta)^{-1}\mathbf{J}(\mathbf{x}, \beta)). \tag{6.3.5}$$

His equivalence theorem for D-optimal designs, states that the following three properties are equivalent:

$$\begin{aligned} &\text{(i)} \ |\mathbf{I}(\eta,\beta)| = \max_{\eta\in H} |\mathbf{I}(\eta,\beta)| \\ &\text{(ii)} \ \max_{x\in\mathcal{I}} G(\eta^*,\mathbf{x}) = \min_{\eta\in H} \max_{x\in\mathcal{I}} G(\eta,\mathbf{x}) \\ &\text{(iii)} \ \max_{x\in\mathcal{I}} G(\eta^*,\mathbf{x}) = p. \end{aligned} \tag{6.3.6}$$

Federov proved that any one of the above three statements implies the other two. The third property is identical to the last line of (6.3.2).

The application of this theory to the normal error linear model will help to illustrate how it works. Consider the standard textbook model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{6.3.7}$$

where ε_i are i.i.d. $N(0, \sigma^2)$. For this model ,

$$\mathbf{I}(\eta, \beta) = \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X} = \frac{1}{\sigma^2}\begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 \end{bmatrix}, \tag{6.3.8}$$

so

$$\mathbf{J}(x_i, \beta) = \frac{1}{\sigma^2}\begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} = \frac{1}{\sigma^2}\mathbf{x}_i\mathbf{x}_i' \tag{6.3.9}$$

where $\mathbf{x}_i' = [1 \ \ x_i]$. Then,

$$\begin{aligned} G(\eta, \mathbf{x}) &= N \, tr(\mathbf{I}(\eta, \beta)^{-1}\mathbf{J}(\mathbf{x}, \beta)) \\ &= N \, tr((\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i\mathbf{x}_i') \\ &= N \, \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \end{aligned} \tag{6.3.10}$$

which is the scaled prediction variance for the normal error linear model. Thus, Federov's properties become

$$\text{(i) } |\mathbf{I}(\eta,\beta)| = \underset{\eta\in H}{\text{Max}}\left|\frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right|$$

$$\text{(ii) } \underset{x\in\mathcal{I}}{\text{Max}}\ G(\eta^*,\mathbf{x}) = \underset{\eta\in H}{\text{Min}}\ \underset{x\in\mathcal{I}}{\text{Max}}\ N\mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i \qquad (6.3.11)$$

$$\text{(iii) } \underset{x\in\mathcal{I}}{\text{Max}}\, G(\eta^*,\mathbf{x}) = p.$$

The second condition is the criterion for G-optimality for the normal linear model. G-optimality requires minimization of the maximum prediction variance over a region. Thus, for a normal linear model, the D and G optimal designs are the same, and the design which satisfies these criteria also has maximum scaled prediction variance on the region equal to p, the number of model parameters.

Recall that for the normal error linear model, the regressors can be coded such that each point in the design space has a coded value between -1 and +1, inclusively. For the particular linear model in (6.3.7), it is known that the D-optimal design is a design with N/2 observations at the design point coded to -1 and N/2 observations at the design point coded to +1 (See Myers and Montgomery, 1995). For this design,

$$\mathbf{X}'\mathbf{X} = N\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}. \qquad (6.3.12)$$

Therefore, the prediction variance at a point, x_0, is

$$N\, x_0'(\mathbf{X}'\mathbf{X})^{-1}x_0 = 1 + x_0^2. \qquad (6.3.13)$$

Now, each point in the design space, $\mathcal{D}$, is coded such that $-1 \le x \le 1$ for all $x \in \mathcal{D}$. Thus, $0 \le x^2 \le 1$ for all $x \in \mathcal{D}$, and $1 \le 1 + x^2 \le 2$ for all $x \in \mathcal{D}$. The maximum prediction variance on design region is indeed equal to 2, and attains 2 at the point -1 and +1, the design points. Therefore, the design which is known to be D-optimal also satisfies the condition (iii) in (6.3.11) which means it is G-optimal as well, and the equivalence theory is illustrated.

§6.3.2 The Exponential Model

Recall the exponential model

$$y_{ij} = e^{\beta_0+\beta_1 x_i} + \varepsilon_{ij} \tag{6.3.14}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$. The information matrix for this model and a design with N experimental runs is

$$\mathbf{I}(\mathbf{x},\beta) = \begin{bmatrix} \sum_{i=1}^{N}\lambda_i & \sum_{i=1}^{N}\lambda_i x_i \\ \sum_{i=1}^{N}\lambda_i x_i & \sum_{i=1}^{N}\lambda_i x_i^2 \end{bmatrix} \tag{6.3.15}$$

where λ_i is the expected response at design level i. The information matrix for a single observation at x_j is

$$\begin{aligned} \mathbf{J}(x_j,\beta) &= \begin{bmatrix} \lambda_j & \lambda_j x_j \\ \lambda_j x_j & \lambda_j x_j^2 \end{bmatrix} \\ &= \mathbf{v}(x_j)\mathbf{v}(x_j)' \end{aligned} \tag{6.3.16}$$

where

$$\mathbf{v}(x_j)' = \sqrt{\lambda_j}\begin{pmatrix} 1 & x_j \end{pmatrix}. \tag{6.3.17}$$

Therefore, for a two level design,

$$\begin{aligned} N\mathbf{v}(x_j)\mathbf{I}(\mathbf{x},\beta)^{-1}\mathbf{v}(x_j)' &= \frac{N\lambda_j\left(n_1\lambda_1(x_1-x_j)^2 + n_2\lambda_2(x_2-x_j)^2\right)}{|\mathbf{I}|} \\ &= \frac{N\lambda_j\left(n_1\lambda_1\left(\ln\frac{\lambda_1}{\lambda_j}\right)^2 + n_2\lambda_2\left(\ln\frac{\lambda_2}{\lambda_j}\right)^2\right)}{n_1 n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2} \end{aligned} \tag{6.3.18}$$

where n_i is the number of observations at design level i. Therefore, the D-optimal two level design must satisfy the inequality

$$\frac{N\lambda_j\left(n_1\lambda_1\left(\ln\frac{\lambda_1}{\lambda_j}\right)^2+n_2\lambda_2\left(\ln\frac{\lambda_2}{\lambda_j}\right)^2\right)}{n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}\leq 2. \qquad (6.3.19)$$

According to (6.3.6), when the two level D-optimal design on a region is substituted into this expression, the left-hand side should have a maximum value of 2, and it should reach that maximum value at the design points.

Recall from Chapter 4 that the two level D-optimal design on the entire region $(EC_0, EC_{100}]$, obtained numerically, had $n_1 = 0.5$ N, $n_2 = 0.5$ N, $\lambda_1 = 0.135\ \lambda_c$ and $\lambda_2 = \lambda_c$. Substitution of this design into the left-hand side of (6.3.19) and the additional substitution of $\lambda_j = q_j\lambda_c$, $\lambda_1 = q_1\lambda_c$, and $\lambda_2 = q_2\lambda_c$ gives

$$\frac{2q_j\left(0.135\left(\ln\frac{0.135}{q_j}\right)^2+\left(\ln\frac{1}{q_j}\right)^2\right)}{4}\leq 2. \qquad (6.3.20)$$

If this is indeed the D-optimal two level design, then this equation is satisfied for all q_j in the design region, that is, all $0 < q_j \leq 1$.

Let the function on the left-hand side of this equation be called *f*. Maximization of *f* with respect to q_j using calculus techniques involves first equating the first derivative of *f* with respect to q_j equal to 0. The first derivative after simplification is

$$q_j(2 + 2\ln q_j + 4.195(\ln q_j)^2). \qquad (6.3.21)$$

When this derivative is equated to 0, two solutions are found, $q_j = e^{-2}$ and $q_j = \exp\{-4/(1+e^{-2})\}$. Substitution of each into the expression for the second derivative of *f* with respect to q_j reveals that $q_j = e^{-2}$ is a maximum while $q_j = \exp\{-4/(1+e^{-2})\}$ is a minimum. Substitution of e^{-2} for q_j in the function *f* yields a function value of 2. This satisfies equation (6.3.19). However, there may be a point on the boundary of the region with a value greater than 2. The lower boundary $q_j = 0$ makes *f* nonexistent. Nevertheless, as values closer and closer to 0 are substituted into *f*, *f* gets closer and closer

to 0. When the upper boundary, $q_j = 1$, is substituted into f, it yields a function value of 2. This is demonstrated in Figure 6.1.

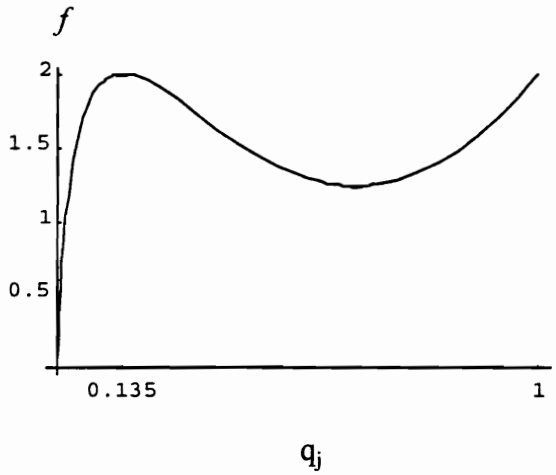

Figure 6.1 Plot of f on the region $(EC_0, EC_{100}]$.

Thus, the maximum value of f on the region is 2 for this two level design and that value is attained by the function at the two design levels. Having met Silvey's conditions, this design is the true two level D-optimal design.

Now consider the restricted region $[EC_{20}, EC_{100}]$. Table 4.1 gives the two level D-optimal design that was numerically obtained on this region. (Recall that when searching for the D-optimal three level design on the region $[EC_{20}, EC_{80}]$, the optimal design turned out to be a two-level design.) The D-optimal two level design obtained for this region has 1/2 the observations at the EC_{20} and 1/2 of the observations at the EC_{100}, the control. Substituting this design into f and equating the first derivative of f to 0 on this new region yields one solution which is a minimum of f. Therefore, the maximum must be sought at the boundaries of the region. At the lower boundary, the EC_{20}, $f = 2$. At the upper boundary, the control, $f = 2$ as well. Figure 6.2 illustrates this. Since f attains a maximum value of 2 on the region, and it attains this value at the design points, this design is truly the D-optimal design on the restricted region $[EC_{20}, EC_{100}]$.

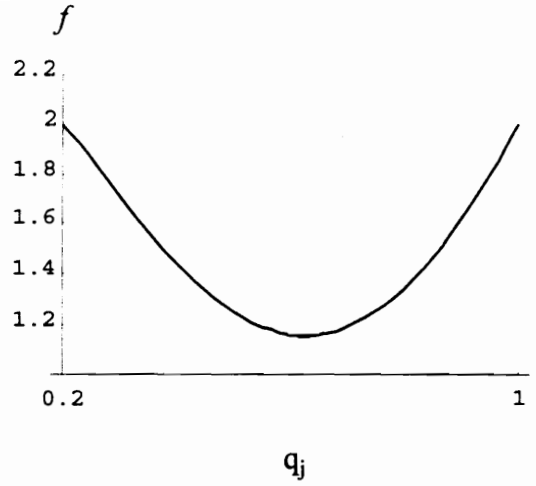

Figure 6.2 Plot of f on the region $[EC_{20}, EC_{100}]$.

Using equivalence theory, it has been shown that the truly D-optimal design on this restricted region has two levels. However, in previous sections, three level designs were obtained because practitioners would probably not want to use two level designs. For the sake of comparison, it is interesting to see how a design on the region $[EC_{20}, EC_{100}]$ which is forced to have three levels does not satisfy Silvey's theory and is not, therefore, truly D-optimal on that region. Consider the three level design given in Table 4.1 which has 34.5% of the observations at each of the EC_{20} and the EC_{80} and has 31% of the observations at the control. Substitution of this design into f gives

$$q_j (2 + 3.59 \ln q_j + 6.80 (\ln q_j)^2). \qquad (6.3.22)$$

When the first derivative of this function is equated with 0, only one solution in the region arises, and this solution is a minimum. The maximum value, then, must be on a boundary. Substituting $q_j = 1$ into f gives a value of 2, but substituting $q_j = 0.2$ into f gives a value of 2.7693 which is clearly greater than 2. Figure 6.3 shows a plot of f for the best three level design. Since the maximum value of f on the region is greater than 2, this design does not satisfy Silvey's theorem, and is thus *not* the D-optimal design on the region.

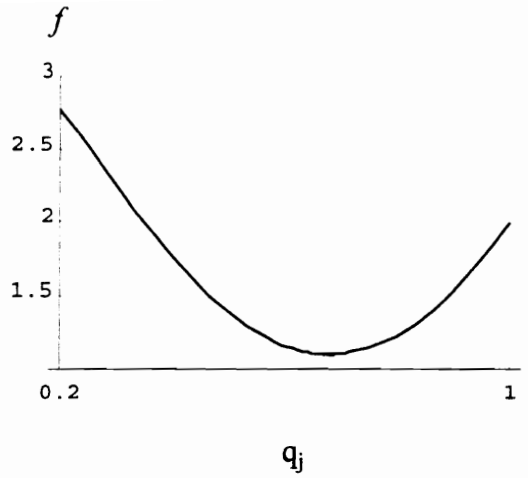

Figure 6.3 Plot of f on the region $[EC_{20}, EC_{100}]$ for the best three level design.

§6.3.3 The Linear Model

The linear Poisson model considered in Chapter 5 was

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij} \tag{6.3.23}$$

where $y_{ij} \sim \text{Poisson}(\lambda_i)$, and the information matrix for this model and a design with N experimental runs is

$$\mathbf{I}(\mathbf{x},\beta) = \begin{bmatrix} \sum \frac{1}{\lambda_i} & \sum \frac{x_i}{\lambda_i} \\ \sum \frac{x_i}{\lambda_i} & \sum \frac{x_i^2}{\lambda_i} \end{bmatrix}. \tag{6.3.24}$$

The information matrix for a single observation at the point x_j is

$$\begin{aligned} \mathbf{J}(\mathbf{x},\beta) &= \begin{bmatrix} \frac{1}{\lambda_j} & \frac{x_j}{\lambda_j} \\ \frac{x_j}{\lambda_j} & \frac{x_j^2}{\lambda_j} \end{bmatrix} \\ &= \mathbf{v}(x_j)\mathbf{v}(x_j)' \end{aligned} \tag{6.3.25}$$

where

$$\mathbf{v}(x_j)' = \frac{1}{\sqrt{\lambda_j}}[1 \quad x_j].$$

The expression for the Frechet directional derivative is

$$\begin{aligned} N\mathbf{v}(x_j)\mathbf{I}(\mathbf{x},\beta)^{-1}\mathbf{v}(x_j)' &= \frac{\frac{N}{\lambda_j}\left(\frac{n_1}{\lambda_1}(x_j - x_1)^2 + \frac{n_2}{\lambda_2}(x_j - x_2)^2\right)}{|\mathbf{I}|} \\ &= \frac{\frac{N}{\lambda_j}\left(\frac{n_1}{\lambda_1}(\lambda_j - \lambda_1)^2 + \frac{n_2}{\lambda_2}(\lambda_j - \lambda_2)^2\right)}{\frac{n_1 n_2}{\lambda_1\lambda_2}(\lambda_1 - \lambda_2)^2}. \end{aligned} \tag{6.3.26}$$

Thus, for the model in (6.3.23), a two level D-optimal design must satisfy

$$\frac{\frac{N}{\lambda_j}\left(\frac{n_1}{\lambda_1}(\lambda_j - \lambda_1)^2 + \frac{n_2}{\lambda_2}(\lambda_j - \lambda_2)^2\right)}{\frac{n_1 n_2}{\lambda_1\lambda_2}(\lambda_1 - \lambda_2)^2} \le 2. \tag{6.3.27}$$

Again, let the left hand side of this equation be *f*.

From Chapter 5, the two level D-optimal design on the whole region $[EC_{0.01}, EC_{100}]$ had $n_1 = N/2$, $n_2 = N/2$, $\lambda_1 = 0.0001\ \lambda_c$ and $\lambda_2 = \lambda_c$. Substitution of this design into *f* and substitution of $\lambda_j = q_j\lambda_c$ gives

$$-0.00080016 + \frac{0.00020006}{q_j} + 2.0006\, q_j. \tag{6.3.28}$$

Figure 6.4(a) gives a plot of this function over the region $[EC_{0.01}, EC_{100}]$. Since Figure 6.4(a) does not show clearly the behavior of the function around the lower endpoint, Figure 6.4(b) shows *f* in a smaller neighborhood of the lower boundary. From the plots, it can be seen that there is a minimum of the function in the region, and the function attains a maximum value of 2 at the boundaries, the two design points. Calculus methods verify that this is indeed the case. Thus, this design is the truly D-optimal design for the linear model in this region.

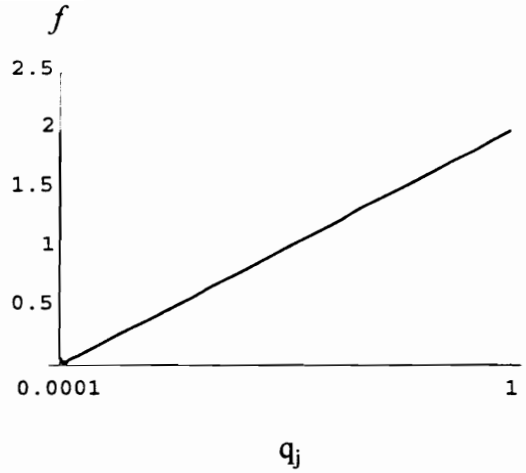

Figure 6.4(a) Plot of f on the region $[EC_{0.01}, EC_{100}]$.

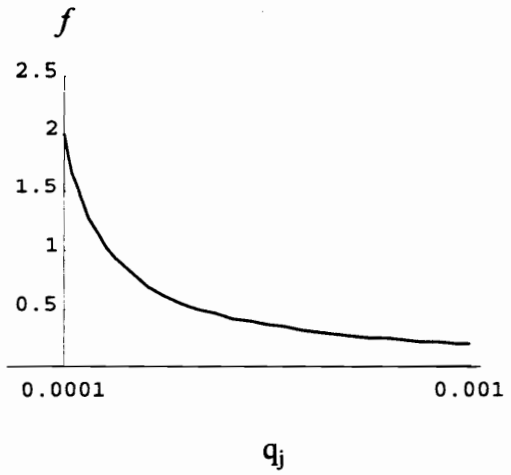

Figure 6.4(b) Plot of f on the region $[EC_{0.01}, EC_{0.1}]$.

§6.4 Equivalence Theory for Linear Criteria Functions

Consider the criterion function given by

$$\phi(\mathbf{M}(\eta,\beta)) = -\mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}) \tag{6.4.1}$$

where **A** is a p x s matrix of rank s. One can easily see that if **A** equals the vector, **a**′, the result is the variance of **a**′β. Silvey categorizes this function as a linear criteria function. Many design criteria including the Q-optimality criterion and the F-optimality criterion for the exponential model can be written in this form. Silvey shows that for the function in (6.4.1),

$$F_\phi[\mathbf{M}(\eta,\beta),\mathbf{J}(\mathbf{x},\beta)] = \mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{J}(\mathbf{x},\beta)\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}) - \mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}). \tag{6.4.2}$$

When $\mathbf{J}(\mathbf{x}, \beta)$ can be written as $\mathbf{v}(\mathbf{x})\mathbf{v}(\mathbf{x})'$,

$$F_\phi[\mathbf{M}(\eta,\beta),\mathbf{J}(\mathbf{x},\beta)] = \mathbf{v}(\mathbf{x})'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{v}(\mathbf{x}) - \mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}). \tag{6.4.3}$$

Theorem 1 states that the design η^* is optimal if and only if $F_\phi[\mathbf{M}(\eta^*, \beta), \mathbf{J}(\mathbf{x}, \beta)] \le 0$ for all $\mathbf{x} \in \mathcal{X}$. Therefore, the optimal design η^* based on the criterion, ϕ, satisfies

$$\begin{aligned} & F_\phi[\mathbf{M}(\eta^*,\beta),\mathbf{J}(\mathbf{x},\beta)] \le 0 \\ \Rightarrow\ & \mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{A}\mathbf{A}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x}) \le \mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{A}) \end{aligned} \tag{6.4.4}$$

for all $\mathbf{x} \in \mathcal{X}$. The result in (6.4.4) will be important in dealing with Q and F-optimal designs.

§6.5 Equivalence Theory for Q-Optimal Designs

Recall that the Q-optimal design on a specified region of operability is the design which minimizes the average prediction variance over a particular, possibly different, region of interest. The general form of the Q-optimality criterion is

$$Q = \int_R \mathbf{c}(\mathbf{x})'\mathbf{M}(\eta,\beta)^{-1}\mathbf{c}(\mathbf{x})d\mathbf{x} \tag{6.5.1}$$

where $\mathbf{c}(\mathbf{x})$ is the p x 1 vector which makes the argument of the integral the prediction variance at an arbitrary point $\mathbf{x}$ given a particular model. The goal in finding the Q-optimal design is to find the design which minimizes Q or maximizes -Q. Thus,

$$\begin{aligned}\phi(\mathbf{M}(\eta,\beta)) &= -Q \\ &= -\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})'\mathbf{M}(\eta,\beta)^{-1}\mathbf{c}(\mathbf{x})d\mathbf{x} \\ &= -\,\mathrm{tr}\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})'\mathbf{M}(\eta,\beta)^{-1}\mathbf{c}(\mathbf{x})d\mathbf{x} \\ &= -\,\mathrm{tr}[\mathbf{M}(\eta,\beta)^{-1}\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})\mathbf{c}(\mathbf{x})'d\mathbf{x}].\end{aligned} \tag{6.5.2}$$

Silvey asserted that $\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})\mathbf{c}(\mathbf{x})'d\mathbf{x} = \mathbf{B}$ where $\mathbf{B}$ is a non-negative definite matrix, and that if $\mathbf{B}$ is of rank s, it can be written as $\mathbf{AA}'$. Therefore,

$$\begin{aligned}\phi(\mathbf{M}(\eta,\beta)) &= -Q \\ &= -\,\mathrm{tr}[\mathbf{M}(\eta,\beta)^{-1}\mathbf{AA}'] \\ &= -\,\mathrm{tr}[\mathbf{A}'\mathbf{M}(\eta,\beta)^{-1}\mathbf{A}].\end{aligned} \tag{6.5.3}$$

This is identical to the function in (6.4.1), so the theory in section 6.4 applies to Q-optimality. The Q-optimal design must η^* satisfy

$$\begin{aligned}&\mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{AA}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x}) \le \mathrm{tr}[\mathbf{A}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{A}] \\ \Rightarrow\ &\mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{AA}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x}) \le \mathrm{tr}[\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{AA}'] \\ \Rightarrow\ &\mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})\mathbf{c}(\mathbf{x})'d\mathbf{x}\,\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x}) \\ &\qquad\qquad \le \mathrm{tr}[\mathbf{M}(\eta^*,\beta)^{-1}\int_{\mathbf{x}\in\mathcal{I}} \mathbf{c}(\mathbf{x})\mathbf{c}(\mathbf{x})'d\mathbf{x}].\end{aligned} \tag{6.5.4}$$

for all $\mathbf{x} \in \mathcal{I}$. The right-hand side of this equation is Q. Let the left-hand side be called g.

§6.5.1 Application to the Normal Error Linear Model

Since the Q-optimal design for the normal error linear model is known, it may be used to once again illustrate how the equivalence theory works for a design that we know a priori is Q-optimal. For the normal error linear model, the prediction variance at a point $\mathbf{x}$, apart from σ^2, is given by $\mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}$. Therefore, the Q criterion is to minimize

$$\frac{N}{K}\int_R \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}d\mathbf{x} \tag{6.5.5}$$

where K is the volume of the region of interest. Also, for the linear normal error model,

$$\mathbf{M}(\eta^*,\beta)^{-1} = N(\mathbf{X}'\mathbf{X})^{-1} \tag{6.5.6}$$

which means

$$\mathbf{c}(\mathbf{x}) = \frac{1}{\sqrt{K}}\mathbf{x}. \tag{6.5.7}$$

Thus, the right hand side of (6.5.4) for this model is

$$\mathrm{tr}[N(\mathbf{X}'\mathbf{X})^{-1}]\left[\frac{1}{K}\int_R \mathbf{x}\mathbf{x}'d\mathbf{x}\right] \tag{6.5.8}$$

where

$$\left[\frac{1}{K}\int_R \mathbf{x}\mathbf{x}'d\mathbf{x}\right] \tag{6.5.9}$$

is called the region moment matrix. The expression in (6.5.8) is the alternate form of the Q-optimality criterion given in Chapter 2.

The Q-optimal design for the normal error linear regression model is the same as the D-optimal design, i.e. N/2 observations at -1 and N/2 observations at +1. From section 6.3.1, $N(\mathbf{X}'\mathbf{X})^{-1}$ for this design is the identity matrix, $\mathbf{I}_0$. The volume of the region of interest, K, is 1- (-1) = 2, so the region moment matrix is

$$\left[\frac{1}{K}\int_R \mathbf{x}\mathbf{x}'d\mathbf{x}\right]$$

$$= \frac{1}{2}\int_{-1}^{1}\begin{bmatrix}1 & x\\ x & x^2\end{bmatrix}dx \qquad (6.5.10)$$

$$= \begin{bmatrix}1 & 0\\ 0 & 1/3\end{bmatrix}.$$

Therefore, the right hand side of (6.5.4) for this design is $\mathrm{tr}[N(\mathbf{X}'\mathbf{X})^{-1}]\left[\frac{1}{K}\int_R \mathbf{x}\mathbf{x}'d\mathbf{x}\right] =$ $1 + 1/3 = 4/3$. The left hand side is

$$\mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\int_{\mathbf{x}\in\mathcal{X}}\mathbf{c}(\mathbf{x})\mathbf{c}(\mathbf{x})'d\mathbf{x}\,\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x})$$

$$= [1 \quad x]\mathbf{I}_0\begin{bmatrix}1 & 0\\ 0 & 1/3\end{bmatrix}\mathbf{I}_0\begin{bmatrix}1\\ x\end{bmatrix} \qquad (6.5.11)$$

$$= 1+\frac{x^2}{3}.$$

Since $-1 \le x \le 1$ for all $x \in \mathcal{D}$, the design space, $0 \le x^2 \le 1$ for all $x \in \mathcal{D}$. This implies that $1 \le 1 + x^2/3 \le 4/3$ for all $x \in \mathcal{D}$. Thus, the maximum value the left hand side of (6.5.4) can achieve is 4/3, and it attains 4/3 at the point -1 and +1, the design points. The design known to be Q-optimal, then, satisfies Theorem 1, and Silvey's equivalence theory for Q-optimality is demonstrated with this very standard textbook situation.

§6.5.2 The Exponential Model

Recall from Chapter 4 that for the one regressor exponential model, the prediction variance at a point x_0 is

$$\mathrm{var}(f(x, \hat{\beta})) \approx \mathbf{w}(x)'\mathbf{I}^{-1}\mathbf{w}(x) \qquad (6.5.12)$$

where

$$\mathbf{w}(x)' = [e^{\beta_0+\beta_1 x} \quad xe^{\beta_0+\beta_1 x}] = [\lambda \quad x\lambda] = \lambda[1 \quad x]. \qquad (6.5.13)$$

Recall also that $\mathbf{I}(\mathbf{x}, \beta) = N\mathbf{M}(\eta, \beta)$. Thus,

$$\begin{aligned} \text{var}(f(x, \hat{\beta})) &\approx \mathbf{w}(x)'\mathbf{I}^{-1}\mathbf{w}(x) \\ &= N^{-1}\lambda^2[1 \quad x]\mathbf{M}(\eta,\beta)^{-1}\begin{bmatrix}1\\ x\end{bmatrix}. \end{aligned} \tag{6.5.14}$$

Now,

$$\begin{aligned} Q &= \frac{N}{K}\int_R \text{var}(\hat{f}(x, \hat{\beta}))\partial x \\ &= \int_R \mathbf{c}(x)'\mathbf{M}(\eta,\beta)^{-1}\mathbf{c}(x)dx \end{aligned} \tag{6.5.15}$$

where

$$\mathbf{c}(x)' = K^{-1/2}\lambda[1 \quad x] \tag{6.5.16}$$

and K is the volume of the region of interest. Now,

$$\mathbf{c}(x)\mathbf{c}(x)' = K^{-1}\lambda^2\begin{bmatrix}1 & x\\ x & x^2\end{bmatrix} = K^{-1}e^{2(\beta_0+\beta_1 x)}\begin{bmatrix}1 & x\\ x & x^2\end{bmatrix} \tag{6.5.17}$$

and

$$\int_R \mathbf{c}(x)\mathbf{c}(x)'dx = \frac{\lambda^2}{2\beta_1 K}\left[\begin{matrix}1 & \left(x-\frac{1}{2\beta_1}\right)\\ \left(x-\frac{1}{2\beta_1}\right) & \left(x^2-\frac{x}{\beta_1}+\frac{1}{2\beta_1^2}\right)\end{matrix}\right]_R. \tag{6.5.18}$$

The Q-optimal design satisfies the final expression in (6.5.4). Now, in Chapter 4, it was determined that for a two level design and region of interest $[x_L, x_U]$, the right hand side of (6.5.4) is

$$Q = \int_{x_L}^{x_U} \text{var}(\hat{f}(\mathbf{x}, \hat{\beta})) = \frac{NA}{4n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_U}{\lambda_L}\right)\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2} \tag{6.5.19}$$

where

$$A = n_1\lambda_1\lambda_U{}^2\left[2\left(\ln\frac{\lambda_1}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_U}\right) + 1\right] - n_1\lambda_1\lambda_L{}^2\left[2\left(\ln\frac{\lambda_1}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_L}\right) + 1\right]$$
$$+ n_2\lambda_2\lambda_U{}^2\left[2\left(\ln\frac{\lambda_2}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_U}\right) + 1\right] - n_2\lambda_2\lambda_L{}^2\left[2\left(\ln\frac{\lambda_2}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_L}\right) + 1\right]. \quad (6.5.20)$$

From above,

$$\mathbf{v}(x_j)' = \sqrt{\lambda_j}\,(1 \quad x_j)$$

for the exponential model. Thus, the left hand side of (6.5.4) after much simplification is

$$g = \frac{N\lambda_j(\lambda_U{}^2 f(\lambda_U) - \lambda_L{}^2 f(\lambda_L))}{4n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_U}{\lambda_L}\right)\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2} \quad (6.5.21)$$

where

$$f(\lambda_m) = n_1{}^2\lambda_1{}^2\left(\ln\frac{\lambda_1}{\lambda_j}\right)^2\left[2\left(\ln\frac{\lambda_1}{\lambda_m}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_m}\right) + 1\right]$$
$$+ 2n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_j}\right)\left(\ln\frac{\lambda_2}{\lambda_j}\right)\left[2\left(\ln\frac{\lambda_1}{\lambda_m}\right)\left(\ln\frac{\lambda_2}{\lambda_m}\right) + \left(\ln\frac{\lambda_1}{\lambda_m}\right) + \left(\ln\frac{\lambda_2}{\lambda_m}\right) + 1\right] \quad (6.5.22)$$
$$+ n_2{}^2\lambda_2{}^2\left(\ln\frac{\lambda_2}{\lambda_j}\right)^2\left[2\left(\ln\frac{\lambda_2}{\lambda_m}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_m}\right) + 1\right].$$

For the Q-optimal design characterized by n_1, n_2, λ_1 and λ_2, the left hand side will be maximized when it equals the right hand side and when λ_j is either λ_1 or λ_2.

Consider the numerically obtained Q-optimal design on the whole region (EC_0, EC_{100}]. This design had $n_1 = 0.561$ N, $n_2 = 0.439$ N, $\lambda_1 = 0.187\ \lambda_c$ and $\lambda_2 = \lambda_c$. For this design the right-hand side of (6.5.4), Q, is 0.6558. The left-hand side, after substituting $q_j\lambda_c$ for λ_j, is

$$g = q_j\,(0.655128 + 0.104576 \ln q_j + 1.07771\,(\ln q_j)^2). \quad (6.5.23)$$

Figure 6.5 shows a plot of g for this design over the region.

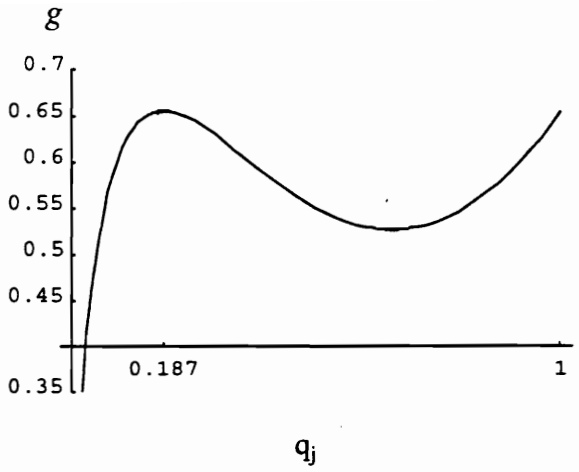

Figure 6.5 Plot of g on the region $(EC_0, EC_{100}]$.

Calculus techniques verify that one maximum and one minimum exist in the region. The maximum value of g in the region is 0.6558 and it occurs at the $EC_{18.7}$, a design point. Examination of the boundaries reveals that only the upper boundary could be a maximum, and the value of the function at the right boundary is 0.6558. Recall that the upper boundary is the point at which the response is λ_c. Thus, this design satisfies Theorem 1 and is the true Q-optimal design on the entire region.

Now, consider the restricted case in which both the region of operability and the region of interest are $[EC_{20}, EC_{100}]$. The optimal design on this region was found numerically to have $n_1 = 0.634$ N, $n_2 = 0.366$ N, $\lambda_1 = 0.25\ \lambda_c$ and $\lambda_2 = \lambda_c$. For this design, Q = 2.2952, and Figure 6.6 shows a plot of g. Calculus techniques verify that the function attains this maximum value at two different locations in the region. One location is the point with expected response $0.25\ \lambda_c$, the EC_{25}, and the other location is the point with expected response λ_c, the control. Since these are the two design points, this is the Q-optimal design on the restricted region.

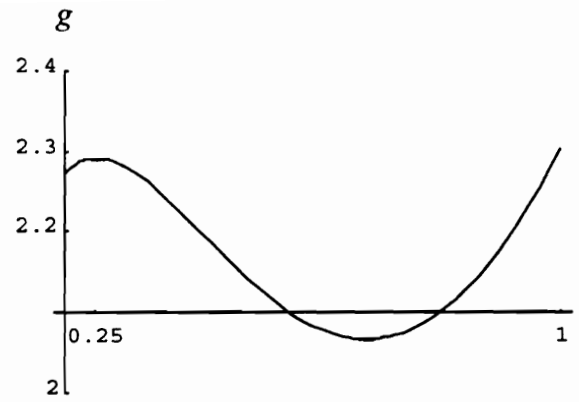

Figure 6.6 Plot of g on the region $[EC_{20}, EC_{100}]$.

§6.5.3 The Linear Model

For the one regressor linear Poisson model,

$$\text{var}(f(x, \hat{\beta})) \approx \mathbf{w}(x)'\mathbf{I}^{-1}\mathbf{w}(x)$$

where

$$\mathbf{w}(x)' = [1 \quad x].$$

Thus,

$$\begin{aligned} Q &= \frac{N}{K}\int_R \text{var}(\hat{f}(x, \hat{\beta}))\partial x \\ &= \int_R \mathbf{c}(x)'\mathbf{M}(\eta,\beta)^{-1}\mathbf{c}(x)dx \end{aligned} \tag{6.5.24}$$

where

$$\mathbf{c}(x)' = K^{-1/2}[1 \quad x]. \tag{6.5.25}$$

Now,

$$\mathbf{c}(x)\mathbf{c}(x)' = K^{-1}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} = K^{-1}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} \tag{6.5.26}$$

and

$$\int_R \mathbf{c}(x)\mathbf{c}(x)'dx = \frac{1}{K}\begin{bmatrix} x & \frac{x^2}{2} \\ \frac{x^2}{2} & \frac{x^3}{3} \end{bmatrix}_R . \tag{6.5.27}$$

Again, the Q-optimal design satisfies the final expression in (6.5.4). From Chapter 5, the right hand side of expression (6.5.4) for a two level design on region of interest $[x_L, x_U]$ is

$$\begin{aligned} Q = \left(\frac{N}{3}\right)\Bigg(&\frac{3\lambda_1\lambda_2[(n_1\lambda_1 + n_2\lambda_2) - (\lambda_U + \lambda_L)N]}{n_1 n_2(\lambda_1 - \lambda_2)^2} \\ &+ \frac{(n_1\lambda_2 + n_2\lambda_1)(\lambda_U{}^2 + \lambda_U\lambda_L + \lambda_L{}^2)}{n_1 n_2(\lambda_1 - \lambda_2)^2}\Bigg). \end{aligned} \tag{6.5.28}$$

Since for the linear model,

$$\mathbf{v}(x_j)' = \frac{1}{\sqrt{\lambda_j}}[1 \quad x_j],$$

the left hand side of (6.5.4) after simplification is

$$\begin{aligned} \frac{N}{3\lambda_j}\Bigg[&(\lambda_U{}^2 + \lambda_U\lambda_L + \lambda_L{}^2)\left(\frac{n_1(\lambda_1 - \lambda_j)}{\lambda_1} + \frac{n_2(\lambda_2 - \lambda_j)}{\lambda_2}\right)^2 \\ &- 3(\lambda_U + \lambda_L)\left(\frac{n_1{}^2\lambda_1{}^2}{\lambda_1} + n_1 n_2\lambda_1\lambda_2\left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2}\right) + \frac{n_2{}^2\lambda_2{}^2}{\lambda_2}\right) \\ &+ 3(n_1(\lambda_1 - \lambda_j) + n_2(\lambda_2 - \lambda_j))^2\Bigg] \end{aligned} \tag{6.5.29}$$

Again, for the Q-optimal design characterized by n_1, n_2, λ_1 and λ_2, (6.5.29) will be maximized when it equals (6.5.28) and when λ_j is either λ_1 or λ_2.

Consider the numerically obtained Q-optimal design for the linear model on the region $[EC_{0.01}, EC_{100}]$. This is the design characterized by $n_1 = 0.01$ N, $n_2 = 0.99$ N, $\lambda_1 = 0.0001\ \lambda_c$ and $\lambda_2 = \lambda_c$. For this design, Q, the right-hand side of the inequality is 1.0201. Figure 6.7(a) describes the behavior of g.

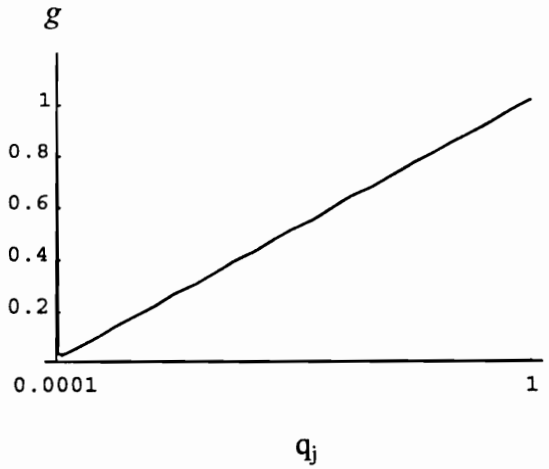

Figure 6.7(a) Plot of g on the region $[EC_{0.01}, EC_{100}]$.

Figure 6.7(b) more clearly illustrates the behavior of g in the neighborhood of the lower boundary.

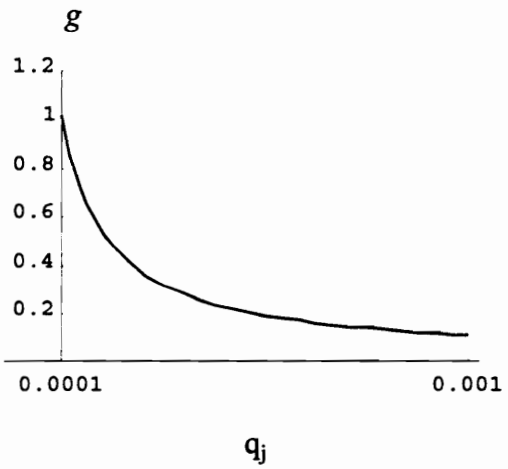

Figure 6.7(b) Plot of g on the region $[EC_{0.01}, EC_{0.1}]$.

Again, the use of derivatives shows that this function has one minimum in the region indicating that the maximum must be on a boundary. Both the upper and lower boundaries yield a function value of 1.0201. Thus, this design satisfies Silvey's theorem and is Q-optimal.

§6.6 Equivalence Theory for F-Optimal Designs

As stated previously, F-optimal designs can only be verified by Silvey's theory if the F-optimality criterion can be written as a linear criteria function. Though this is not the case for the F-optimality criterion for linear model, it is true of the exponential model. Thus, the F-optimal designs can be verified for the exponential model using the equivalence theory.

Recall that Silvey's theory can be applied to criteria functions of the form

$$\phi(\mathbf{M}(\eta,\beta)) = -\mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)\mathbf{A}) \tag{6.6.1}$$

where **A** is a p x s matrix of rank s. It was shown in Chapter 4 that for the exponential model, the F-optimality criterion is

$$\min_{\mathcal{D}} \ \mathrm{var}(b_1)$$

where $\mathcal{D}$ represents all possible designs on the region. Now,

$$\begin{aligned} \mathrm{var}(b_1) &= [0 \quad 1]\mathbf{I}^{-1}\begin{bmatrix}0\\1\end{bmatrix} \\ &= [0 \quad 1]\mathbf{N}^{-1}\mathbf{M}^{-1}\begin{bmatrix}0\\1\end{bmatrix} \\ &= \mathbf{a}'\mathbf{M}^{-1}\mathbf{a} \end{aligned} \tag{6.6.2}$$

where $\mathbf{a}' = [0 \quad N^{-1/2}]$. Since the goal is to minimize this quantity and since the trace of a scalar equals that scalar, the F-optimality criterion function for the exponential model is of the form

$$\phi(\mathbf{M}(\eta,\beta)) = -\mathrm{tr}(\mathbf{A}'\mathbf{M}(\eta,\beta)\mathbf{A})$$

where **A** = **a**, a p x 1 matrix of rank 1. Therefore, the theory in section 6.4 holds for this criterion function and the F-optimal design η^* satisfies

$$\mathbf{v}(\mathbf{x})'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{a}\mathbf{a}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{v}(\mathbf{x}) \ \le \ \mathrm{tr}(\mathbf{a}'\mathbf{M}(\eta^*,\beta)^{-1}\mathbf{a}) \tag{6.6.3}$$

for all $\mathbf{x} \in \mathcal{X}$. The right hand side of (6.6.3) is simply the $\mathrm{var}(b_1)$ which for a two level design is

$$\text{var}(b_1) = \frac{\beta_1^{\,2}(n_1\lambda_1 + n_2\lambda_2)}{n_1 n_2 \lambda_1 \lambda_2 \left(\ln\frac{\lambda_1}{\lambda_2}\right)^2}. \tag{6.6.4}$$

The left hand side, after simplification is

$$\frac{\lambda_j\, \beta_1^{\,2} \left(n_1\lambda_1 \ln\frac{\lambda_j}{\lambda_1} + n_2\lambda_2 \ln\frac{\lambda_j}{\lambda_2}\right)^2}{n_1^{\,2} n_2^{\,2} \lambda_1^{\,2} \lambda_2^{\,2} \left(\ln\frac{\lambda_1}{\lambda_2}\right)^4}. \tag{6.6.5}$$

Let this quantity be called *h*. By the theorem in (6.6.3), for the F-optimal design characterized by n_1, n_2, λ_1 and λ_2, *h* will be maximized when it equals $\text{var}(b_1)$ and when λ_j is either λ_1 or λ_2.

For the exponential model, the F-optimal design over the entire region $(EC_0, EC_{100}]$ was found via numerical methods to have $n_1 = 0.78$ N, $n_2 = 0.22$ N, $\lambda_1 = 0.078\ \lambda_c$ and $\lambda_2 = \lambda_c$. For this design, $\text{var}(b_1) = 3.224$ and

$$h = q_j\,(3.19279 + 11.525 \ln q_j + 10.4004 (\ln q_j)^2). \tag{6.6.6}$$

Figure 6.8 shows a plot of *h* over the region for this design.

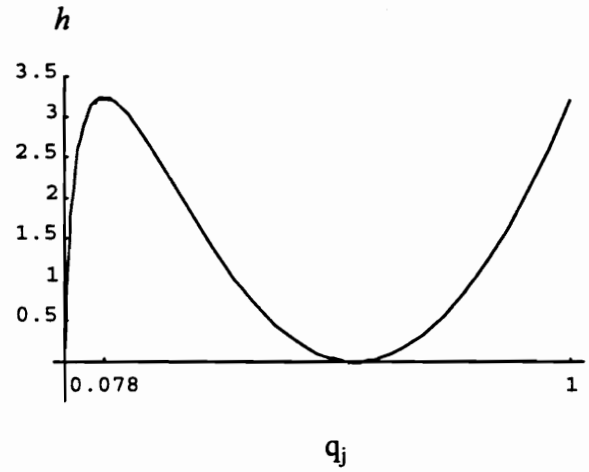

Figure 6.8 Plot of *h* on the region $(EC_0, EC_{100}]$.

The plot and calculus techniques reveal that the function, *h*, has a maximum value of 3.224 over the region. The maximum occurs at the point where the expected response is 0.078 λ_c, the $EC_{7.8}$, and at the point where the expected response is λ_c, the control. In other words, the maximum of *h* equals the var(b_1) for this design, and it is attained at the design points. Therefore, this is truly the F-optimal design. This result is not surprising since Minkin (1993) showed this to be the optimal design analytically from the criterion.

On the restricted region [EC_{20}, EC_{100}], the optimal design obtained numerically has n_1 = 0.69 N, n_2 = 0.31 N, λ_1 = 0.2 λ_c and λ_2 = λ_c. For this design, var(b_1) = 4.043. Figure 6.9 shows a plot of *h* for this design.

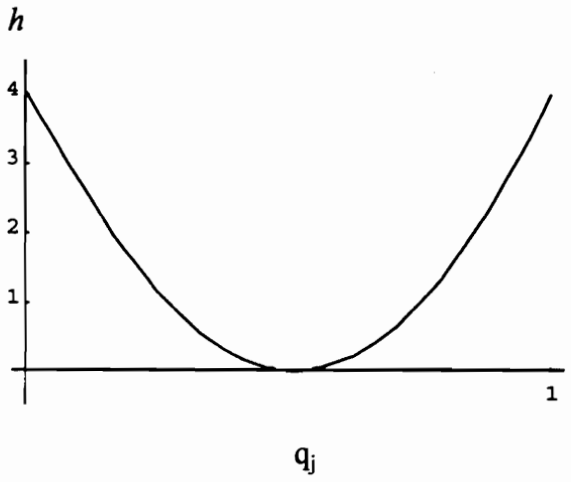

Figure 6.9. Plot of *h* on the region [EC_{20}, EC_{100}].

This figure and the use of derivatives show that this is the F-optimal design on this restricted region.

As shown in the cases above, the truly optimal designs on the region [EC_{20}, EC_{100}] are two level designs. Consequently, two level designs were continually obtained when three level designs were originally sought on the region [EC_{20}, EC_{80}]. As shown in the D-optimality section, forcing a third distinct level to be present in the design causes the design to no longer be optimal. Although only designs on two different

regions were used for each model in the above examples, the results of this section indicate that equivalence theory would most likely show the two level designs found numerically on the other regions in Chapters 4 and 5 to be the truly optimal designs as well.

Chapter 7

Robustness Issues

§7.1 Introduction

There are several ways in which designs may be robust to difficulties that arise when designing experiments for impaired reproduction data. Two robustness issues are discussed in this chapter, and robust designs presented for each issue. The first issue concerns robustness of the optimal designs to the chosen link function, or to the selected model. The second addresses robustness of the optimal designs to misspecification of the model parameters.

§7.2 Robustness to Link Function

A comparison of the designs chosen by the various criteria in Chapters 4 and 5 reveals that the optimal designs differ considerably for the two different models. This leads to an important question. What if a practitioner does not know which model is appropriate for his or her data? What design should he or she use? Do any designs exist which are robust to the assumed model?

By comparing the tables for the two models one can see that there are some D and F-optimal designs on restricted regions which are exactly the same for both models. One should keep in mind that the F-optimal designs for the linear model depend on the ratio $z^2/N\lambda_c$, and the F-optimal designs for the linear model which are used in this section were obtained from the criterion value when this ratio is 0. The designs which are the same or very similar for the two models are highlighted in Table 7.1. Unfortunately, these are two level designs, and the difficulties with two level designs have already been mentioned.

Table 7.1. Two Level Designs on Restricted Regions which are Robust to Link Function

	Region of Operability	Type of Optimality	p_1	p_2	x_1	x_2	Cross-Efficiency
Exponential Model	$[EC_{20},EC_{80}]$	D	0.5	0.5	EC_{20}	EC_{100}	1
Linear Model			0.5	0.5	EC_{20}	EC_{100}	1
Exponential Model	$[EC_{10},EC_{90}]$ or	D	0.5	0.5	$EC_{13.5}$	EC_{100}	0.8272
Linear Model	$[EC_{10},EC_{50}]$		0.5	0.5	EC_{10}	EC_{100}	0.9896
Exponential Model	$[EC_{20},EC_{80}]$	F	0.69	0.31	EC_{20}	EC_{100}	1
Linear Model			0.69	0.31	EC_{20}	EC_{100}	1
Exponential Model	$[EC_{10},EC_{90}]$ or	F	0.76	0.24	EC_{10}	EC_{100}	1
Linear Model	$[EC_{10},EC_{50}]$		0.76	0.24	EC_{10}	EC_{100}	1

However, there are some situations in which the three and four level designs are similar for the two models. Table 7.2 shows some of the three level designs (all are D optimal), and Table 7.3 shows some four level designs which are robust. The four level designs are all F-optimal because F-optimality was the only criterion used to obtain four level designs for the exponential and linear models. It is clear that some designs do exist which, when used, would not cause a severe loss in efficiency even if they are based on a link function which is different from that which should have been used. Also, based on these results, it is not unrealistic to assume that robust designs such as these exist on other restricted design regions as well.

§7.3 Robustness to Parameter Misspecification

In the previous section, some designs were highlighted which are robust to misspecification of the link function. In this section, a different type of robustness is addressed, robustness to misspecification of model parameters. Recall that when designing experiments for Poisson regression, knowledge or some guess of the model parameters is needed. This is due to the fact that the optimal designs are in terms of EC_{100p}'s. The expected responses at the optimal design points are known, but the actual dose levels of the toxicants which produce these responses are not. The only way to know which doses yield particular responses is to know the model parameters. Clearly, the model parameters are not known, and misspecification of the parameters will result in a loss of efficiency of the design. One should keep in mind that, in this section, it is assumed that the form of the model is correctly specified.

Often "guesses" for the values of the model parameters are obtained using literature on certain toxicants or by using results from a previous experiment. The quality of these guesses may vary considerably from application to application. One approach for combating misspecification of these initial guesses is Bayesian design (Chaloner and Larntz, 1989). In Bayesian design, prior distributions are assumed for the model

Table 7.2. Three Level Designs on Restricted Regions which are Robust to Link Function

	Region of Operability	Type of Optimality	$p_1 = p_2$	p_3	x_1	x_2	x_3	Cross-Efficiency
Exponential Model	$[EC_{20}, EC_{80}]$	D	0.345	0.31	EC_{20}	EC_{80}	EC_{100}	0.9952
Linear Model			0.382	0.236	EC_{20}	EC_{80}	EC_{100}	0.9487
Exponential Model	$[EC_{10}, EC_{90}]$	D	0.418	0.164	EC_{10}	EC_{90}	EC_{100}	0.9981
Linear Model			0.445	0.11	EC_{10}	EC_{90}	EC_{100}	0.9979
Exponential Model	$[EC_{10}, EC_{50}]$	D	0.257	0.486	EC_{10}	EC_{50}	EC_{100}	0.9866
Linear Model			0.307	0.386	EC_{10}	EC_{50}	EC_{100}	0.9939

Table 7.3. Four Level Designs on Restricted Regions which are Robust to Link Function

	Region of Operability	Type of Optimality	$p_1 = p_2 = p_3$	p_4	x_1	x_2	x_3	x_4	Cross-Efficiency
Exponential Model	$[EC_{20}, EC_{80}]$	F	0.2586	0.2242	EC_{20}	EC_{50}	EC_{80}	EC_{100}	0.9546
Linear Model			0.3103	0.0691	EC_{20}	EC_{50}	EC_{80}	EC_{100}	0.9252
Exponential Model	$[EC_{10}, EC_{50}]$	F	0.2343	0.2971	EC_{10}	EC_{30}	EC_{50}	EC_{100}	0.9481
Linear Model			0.2797	0.1609	EC_{10}	EC_{30}	EC_{50}	EC_{100}	0.9092

parameters rather than initial point estimates. While Bayesian design has received much attention in recent statistical literature, it is not the approach taken in this work. The method for addressing parameter misspecification used in this dissertation is two-stage design. In this chapter, a brief investigation of the effect of parameter misspecification is conducted using both the exponential and the linear models, the two-stage design theory is presented and the benefits of the two-stage procedure when misspecification is present are revealed.

§7.3.1 Two Stage Design

As the name implies, in two stage design, the experiment is run in two different stages. Consequently, two design criteria are used, one to design the experiment for the first stage and one to design the experiment for the second stage. The first stage is primarily used to obtain "good" parameter estimates for use in the second stage design. Although any of the design criteria could be used in either stage, D-optimality is usually a reasonable choice for the first stage. The second stage design criterion may then be any which is of particular interest to the experimenter. The design type is typically named by both criteria. For example, if D-optimality is the criterion used in both stages, the resulting design is called D-D optimal.

Two stage design was investigated for binary data by Abdelbasit and Plackett (1983), Minkin (1987), Letsinger (1995) and Myers, Myers, Carter, and White (1996). Abdelbasit and Plackett used *the* optimal design in each stage for the criterion being used in that stage. For example, for a D-D optimal two stage design, they used the same design in both stages, namely the two level D-optimal design. The only difference between the two stages was that the parameter estimates from the first stage were used to locate the two optimal levels in the second stage whereas the optimal levels for the first stage were based on a "guess." Minkin chose to design the experiment so that the second stage of the design augmented the first stage. In other words, not only did his second

stage use parameter estimates from the first stage, but it took into account the misspecification in the first stage and, in a sense, tried to correct for it. This was also the approach taken by Letsinger and Myers, Myers, Carter and White. In this section, Minkin's type of two stage design is used, and D-D optimal and D-F optimal designs are found for both the exponential and linear Poisson regression models.

Recall that both the D and F optimality criteria depend upon the information matrix for the data given the particular model. Minkin's method of two stage design utilizes the total, or joint, information matrix for the data when determining the optimal second stage design. This is the information matrix derived from the joint likelihood of the data. The likelihood of the data for the first stage is given by

$$\mathbf{L}_1(\beta,\mathbf{y}_1|\mathbf{x}_1)=\prod_{i=1}^{m_1}\frac{e^{-\lambda_{1i}}\lambda_{1i}{}^{y_{1i}}}{y_{1i}} \tag{7.3.1}$$

where

$$\lambda_{1i}=e^{\beta_0+\beta_1 x_{1i}} \tag{7.3.2}$$

is the Poisson mean at the ith design point in the first stage for the exponential model,

$$\lambda_{1i}=\beta_0+\beta_1 x_{1i} \tag{7.3.3}$$

is the Poisson mean at the ith design point in the first stage for the linear model, m_1 is the number of design points in the first stage, x_{1i} is the m_1 x 1 vector of design points in the first stage and y_{1i} is the m_1 x 1 vector of responses in the first stage. Recall that the responses are the counts at the various design points. The joint likelihood of the data is the product of the likelihood for the first stage and the conditional likelihood for the second stage given the first stage, or

$$\mathbf{L}_{1,2}(\beta,\mathbf{y}_1,\mathbf{y}_2,\mathbf{x}_2|\mathbf{x}_1)=\mathbf{L}_1(\beta,\mathbf{y}_1|\mathbf{x}_1)\mathbf{L}_{2|1}(\beta,\mathbf{y}_2|\mathbf{x}_2) \tag{7.3.4}$$

where

$$\mathbf{L}_{2|1}(\beta,\mathbf{y}_2|\mathbf{x}_2)=\prod_{i=1}^{m_2}\frac{e^{-\lambda_{2j}}\lambda_{2j}{}^{y_{2j}}}{y_{2j}} \tag{7.3.5}$$

is the conditional likelihood in which

$$\lambda_{2j} = e^{\beta_0+\beta_1 x_{2j}} \tag{7.3.6}$$

is the Poisson mean at the ith design point in the second stage for the exponential model,

$$\lambda_{2j} = \beta_0 + \beta_1 x_{2j} \tag{7.3.7}$$

is the Poisson mean at the ith design point in the second stage for the linear model, m_2 is the number of design points in the second stage, x_{2i} is the m_2 x 1 vector of design points in the second stage and y_{2i} is the m_2 x 1 vector of responses in the second stage. The second stage is conditioned on the first in the sense that the parameter estimates from the first are used to compute the design levels in the second. Since all of the optimal designs which were verified using equivalence theory have two levels, the first and second stage designs will each have two levels. This allows each stage to be as close to optimal as possible.

The information matrices are obtained by taking derivatives of the log of the likelihood function for the data. Thus, taking the log of the joint likelihood in (7.3.4) yields

$$\begin{aligned} \ln(\mathbf{L}_{1,2}) &= \ln(\mathbf{L}_1) + \ln(\mathbf{L}_{2|1}) \\ &= -\sum_{i=1}^{m_1}\lambda_{1i} + \sum_{i=1}^{m_1} y_{1i}\ln\lambda_{1i} - \sum_{i=1}^{m_1}\ln(y_{1i}!) \\ &\quad - \sum_{j=1}^{m_2}\lambda_{2j} + \sum_{j=1}^{m_2} y_{2j}\ln\lambda_{2j} - \sum_{j=1}^{m_2}\ln(y_{2j}!). \end{aligned} \tag{7.3.8}$$

This implies that the joint information matrix for the two stages, $\mathbf{I}_{1,2}$, is the sum of the information matrices derived from equations (7.3.1) and (7.3.5). The joint information matrix for the exponential model is

$$\mathbf{I}_{1,2} = \begin{bmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{bmatrix} \tag{7.3.9}$$

where

$$I_{11} = \sum_{i=1}^{k_1} n_{1i}\lambda_{1i} + \sum_{j=1}^{k_2} n_{2j}\lambda_{2j}, \tag{7.3.10}$$

$$I_{12} = \sum_{i=1}^{k_1} n_{1i}\lambda_{1i}x_{1i} + \sum_{j=1}^{k_2} n_{2j}\lambda_{2j}x_{2j}, \qquad (7.3.11)$$

$$I_{22} = \sum_{i=1}^{k_1} n_{1i}\lambda_{1i}x_{1i}^{\ 2} + \sum_{j=1}^{k_2} n_{2j}\lambda_{2j}x_{2j}^{\ 2}, \qquad (7.3.12)$$

k_1 is the number of distinct design points in the first stage, n_{1i} is the number of replicates at the ith distinct design point in the first stage, k_2 is the number of distinct design points in the second stage and n_{2j} is the number of replicates at the jth distinct design point in the second stage. The joint information matrix for the linear model is analogous to this. It should be noted that this is the information matrix for a particular two stage design, not the two stage procedure in general. To obtain the joint information matrix, the expectation was taken only over the y's. The expectation over $\mathbf{x}_2$, the second stage design, has not been taken.

When the criterion of interest in the second stage is the D-optimality criterion the determinant of $\mathbf{I}_{1,2}$ is maximized, or equivalently, the determinant of $\mathbf{I}_{1,2}^{-1}$ is minimized. When the second stage criterion is that for F-optimality, elements of $\mathbf{I}_{1,2}^{-1}$, the asymptotic variance covariance matrix of $\hat{\beta}$, are used. Only the information matrix for the first stage, the matrix derived from the likelihood in (7.3.1) is used in determining the design for the first stage. Since this was the matrix used in the one stage procedures, the optimal first stage designs are those that were determined for the exponential and linear models in Chapters 4 and 5, respectively. The optimal second stage design is found via the Nelder-Mead algorithm and uses $\mathbf{I}_{1,2}$ where the first stage design levels are fixed. In other words, the experimenter runs the first stage experiment using "guesses" for the parameter values in order to determine the doses, the x_{1i}'s, corresponding to the optimal design levels. After the first stage is run, the experimenter uses the parameter estimates from the first stage to obtain improved estimates of the λ_{1i}'s. Given this information about the first stage, the experimenter fixes this information in the joint information matrix $\mathbf{I}_{1,2}$, and uses an algorithm to determine the optimal second stage design.

§7.3.2 Evaluation of the Two Stage Procedure

There is obviously a need to evaluate the two stage design procedure compared with the optimal one stage design and compared with the misspecified one stage design. In a manner similar to that used in previous sections, the designs in this section are evaluated using efficiencies. The D-D optimal two stage designs are be evaluated in terms of D-efficiencies, and the D-F optimal designs are evaluated in terms of F-efficiency.

The D-Efficiency is found in the usual way whereas the F-Efficiency is given by

$$\text{F} - \text{Efficiency} = \frac{\text{asymptotic variance of } \hat{EC}_{100p} \text{ for one stage design}}{\text{asymptotic variance of } \hat{EC}_{100p} \text{ for two stage procedure}}. \quad (7.3.13)$$

Both efficiencies require elements of the matrix, $\mathbf{I}_{1,2}^{-1}$. Computation of these efficiencies require special considerations since the second stage design is a random variable. Thus, to assess the general performance of the two stage procedure, the following expression for the variance of a random variable, W, which is always observed in the presence of another random variable, U, is needed:

$$\text{Var}(W) = E_U[\text{Var}(W|U)] + \text{Var}_U[E(W|U)]. \quad (7.3.14)$$

In this case,

$$\text{Var}(\hat{\beta}) = E_{\mathbf{d}_2}[\text{Var}(\hat{\beta}|\mathbf{d}_2)] + \text{Var}_{\mathbf{d}_2}[E(\hat{\beta}|\mathbf{d}_2)] \quad (7.3.15)$$

where $\mathbf{d}_2 = [\mathbf{n}_2, \mathbf{x}_2]$, the vector of design levels and sample allocations for the particular second stage design. Since $\hat{\beta}$ is the maximum likelihood estimate of β, the conditional expectation, $E(\hat{\beta}|\mathbf{d}_2)$, is β. Thus, asymptotically, $\text{Var}_{\mathbf{d}_2}[E(\hat{\beta}|\mathbf{d}_2)] = 0$, and $\text{Var}(\hat{\beta}) = E_{\mathbf{d}_2}[\text{Var}(\hat{\beta}|\mathbf{d}_2)]$. Let V_{ij} be the element in the ith row and jth column of the unconditional variance covariance matrix for $\hat{\beta}$. Let v_{ij} be the element in the ith row and jth column of the variance covariance matrix of $\hat{\beta}$ given a particular second stage design. Then $V_{ij} = E_{\mathbf{d}_2}[v_{ij}]$, and since the density function for $\mathbf{d}_2$ is unknown, this expectation

must be estimated by the average of the conditional variances and covariances taken over many two stage experiments. This necessitates simulation.

Computer simulations provided estimates of the unconditional asymptotic variances and covariances for the two stage procedure in general. One hundred simulations were run to obtain each efficiency. The total number of observations in each simulation was 60, and the control mean was assumed to be 100. An important question in two stage design is how many observations to allocate to the first stage and how many to the second. The answer to this question was not extensively sought in this research. Rather, two different situations were investigated. In one case, 1/2 the observations were used in the first stage and 1/2 were used in the second. In the second case, 1/3 of the observations were allocated to the first stage and 2/3 were used in the second. D-D optimal and D-F optimal designs were found for each of these cases and for each of the exponential and linear Poisson models. Several different misspecifications in the first stage were assumed.

The results of the simulations appear in Tables 7.4-7.7. Tables 7.4 and 7.5 display the efficiencies of the D-D optimal and D-F optimal designs, respectively, for the exponential model. Tables 7.6 and 7.7 display the efficiencies of the D-D optimal and D-F optimal designs, respectively, for the linear model. In each case, Table (a) contains the results when 1/2 of the total number of experimental units were allocated to the first stage and 1/2 were allocated to the second, and Table (b) contains the results when 1/3 of the sample was used in the first stage and 2/3 was used in the second. The M_D-Efficiency in Tables 7.4 and 7.6 is the efficiency of the misspecified one stage design with respect to the *D-optimal* one stage design. The TS_D-Efficiency is the efficiency comparing the two stage procedure given the misspecified first stage with the one stage *D-optimal* design. The M_F-Efficiency in Tables 7.5 and 7.7 is the efficiency of the misspecified one stage design with respect to the *F-optimal* one stage design. The TS_F-Efficiency is the efficiency comparing the two stage procedure given the misspecified first stage with the one stage *F-optimal* design.

Table 7.4(a) D-D Optimal Two Stage Design Efficiencies for the Exponential Model: 1/2 Observations in First Stage, 1/2 Observations in Second stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_D-Eff	TS_D-Eff
0.25	$EC_{0.01}$	0.25	EC_{100}	0.1252	0.5547
0.25	EC_5	0.25	EC_{100}	0.9104	0.9586
0.25	EC_{20}	0.25	EC_{100}	0.9783	0.9899
0.25	EC_{40}	0.25	EC_{100}	0.7876	0.9245
0.25	EC_{50}	0.25	EC_{100}	0.6662	0.9008

Table 7.4(b) D-D Optimal Two Stage Design Efficiencies for the Exponential Model: 1/3 Observations in First Stage, 2/3 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_D-Eff	TS_D-Eff
0.1667	$EC_{0.01}$	0.1667	EC_{100}	0.1252	0.6077
0.1667	EC_5	0.1667	EC_{100}	0.9104	0.9724
0.1667	EC_{20}	0.1667	EC_{100}	0.9783	0.9929
0.1667	EC_{40}	0.1667	EC_{100}	0.7876	0.9513
0.1667	EC_{50}	0.1667	EC_{100}	0.6662	0.9357

Table 7.5(a) D-F Optimal Two Stage Design Efficiencies for the Exponential Model: 1/2 Observations in First Stage, 1/2 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_F-Eff	TS_F-Eff
0.25	$EC_{0.01}$	0.25	EC_{100}	0.0213	0.3002
0.25	EC_5	0.25	EC_{100}	0.9585	0.9788
0.25	EC_{20}	0.25	EC_{100}	0.7623	0.9263
0.25	EC_{40}	0.25	EC_{100}	0.3493	0.7849
0.25	EC_{50}	0.25	EC_{100}	0.2179	0.7403

Table 7.5(b) D-F Optimal Two Stage Design Efficiencies for the Exponential Model: 1/3 Observations in First Stage, 2/3 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_F-Eff	TS_F-Eff
0.1667	$EC_{0.01}$	0.1667	EC_{100}	0.0213	0.3009
0.1667	EC_5	0.1667	EC_{100}	0.9585	0.9897
0.1667	EC_{20}	0.1667	EC_{100}	0.7623	0.9497
0.1667	EC_{40}	0.1667	EC_{100}	0.3493	0.8574
0.1667	EC_{50}	0.1667	EC_{100}	0.2179	0.8307

Table 7.6(a) D-D Optimal Two Stage Design Efficiencies for the Linear Model: 1/2 Observations in First Stage, 1/2 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_D-Eff	TS_D-Eff
0.25	EC_{10}	0.25	EC_{100}	0.6699	0.8984
0.25	EC_{20}	0.25	EC_{100}	0.4211	0.8585
0.25	EC_{30}	0.25	EC_{100}	0.3008	1.0685
0.25	EC_{40}	0.25	EC_{100}	0.2233	0.8308
0.25	EC_{50}	0.25	EC_{100}	0.1664	0.9365

Table 7.6(b) D-D Optimal Two Stage Design Efficiencies for the Linear Model: 1/3 Observations in First Stage, 2/3 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_D-Eff	TS_D-Eff
0.1667	EC_{10}	0.1667	EC_{100}	0.6699	0.9250
0.1667	EC_{20}	0.1667	EC_{100}	0.4211	0.9507
0.1667	EC_{30}	0.1667	EC_{100}	0.3008	1.0136
0.1667	EC_{40}	0.1667	EC_{100}	0.2233	1.0264
0.1667	EC_{50}	0.1667	EC_{100}	0.1664	0.8511

Table 7.7(a) D-F Optimal Two Stage Design Efficiencies for the Linear Model: 1/2 Observations in First Stage, 1/2 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_F-Eff	TS_F-Eff
0.25	EC_{10}	0.25	EC_{100}	0.3794	0.8048
0.25	EC_{20}	0.25	EC_{100}	0.1145	0.6645
0.25	EC_{30}	0.25	EC_{100}	0.0473	0.6785
0.25	EC_{40}	0.25	EC_{100}	0.0219	0.5969
0.25	EC_{50}	0.25	EC_{100}	0.0105	0.6236

Table 7.7(b) D-F Optimal Two Stage Design Efficiencies for the Linear Model: 1/3 Observations in First Stage, 2/3 Observations in Second Stage

Misspecified First Stage				Efficiencies	
p_{11}	x_{11}	p_{12}	x_{12}	M_F-Eff	TS_F-Eff
0.1667	EC_{10}	0.1667	EC_{100}	0.3794	0.8943
0.1667	EC_{20}	0.1667	EC_{100}	0.1145	0.7846
0.1667	EC_{30}	0.1667	EC_{100}	0.0473	0.7561
0.1667	EC_{40}	0.1667	EC_{100}	0.0219	0.8374
0.1667	EC_{50}	0.1667	EC_{100}	0.0105	0.7042

The investigation into the efficiency of the two stage procedure did not cover numerous restricted regions. For the exponential model, only the optimal designs on the whole dose space (EC_0, EC_{100}] were considered, and for the linear model, only optimal designs on the restricted region [EC_5, EC_{100}] were considered. Recall that all of the D and F-optimal designs found for the Poisson exponential and linear models are two level designs. In fact, the D-optimal design on the region (EC_0, EC_{100}] for the exponential model has 1/2 the observations at the $EC_{13.5}$ and 1/2 the observations at the EC_{100}, the control, and the F-optimal design for the same model on the same region has 78% of the observations at the $EC_{7.8}$ and 22% of the observations at the control. The D-optimal design on the region [EC_5, EC_{100}] for the linear model has 1/2 the observations at the EC_5 and 1/2 the observations at the EC_{100}, the control, and the F-optimal design for the linear model on the same region has 82% of the observations at the $EC_{7.8}$ and 18% of the observations at the control. Fortunately, in all of the optimal two level designs in Chapters 4 and 5, the control was selected as one of the two design levels. That leaves only one design level, x_{11}, to be misspecified. This allows the misspecified one stage designs in Tables 7.4-7.7 to enjoy better efficiencies than might be expected and also allows the two stage procedures to be very efficient.

The highest efficiencies of the one stage misspecified designs are in Table 7.4. In this table, even a misspecification as high as the EC_{40} results in a reasonably efficient design. In general, the two-stage procedure performs better when 1/3 of the observations are used in the first stage and 2/3 are used in the second. Also, the D-D optimal designs enjoy higher efficiencies for the most part. This is probably because the D-optimality criterion was used in both stages of the design. The D-F efficiencies are compared to the F-optimal design, yet the F-optimality criterion was only used in one stage. It would be of interest in future studies to examine the efficiencies of F-F optimal designs. These may yield better results than the D-F designs.

The linear model efficiencies do not show the nice trends that are displayed in the exponential model tables. The efficiencies for the linear model seem to be more

inconsistent and do not necessarily decrease as the misspecification in the first stage increases. There are two possible reasons for this. One reason is that the D and F-optimality criteria for the linear model prefer designs in which the design levels are as far apart as possible. Thus, a two stage design which has points around the EC_5, the EC_{50} and the EC_{100} has more mutual distance among the design levels than a design with points around the EC_5, EC_{30} and EC_{100}. Thus, a design with very high misspecification may actually prove more efficient once it is complemented by the second stage. Another reason the linear model is more inconsistent is because the optimal designs for this model require the non-control design point to be the point with the smallest possible non-zero reproduction. For the simulations in this research this was the EC_1. However, the entire dose space was not being considered in this two stage study. Only the region [EC_5, EC_{100}] was considered. Nevertheless, even though the practitioner would only allow the *expected response* at the non-control design point to be as small as $0.05\lambda_c$, the parameter estimates from the first stage could have mapped this response back to a dose smaller than the EC_5. Such a design would be more efficient than the best one-stage design on the region [EC_5, EC_{100}]. This is more likely to happen as the misspecification in the first stage becomes more serious because the first stage design is getting further and further from the design which minimizes the variance on the parameter estimates, the D-optimal design. This is also the reason that some of the efficiencies for the linear model are greater than 1.

One should note that the tables for the linear model contain only misspecified designs in which the non-control dose, x_{11}, is *larger* than the EC_5 (the optimal non-control design level on the region of interest). Recall that the D-optimality criterion for the linear model prefers a design in which the design levels have the maximum possible spread between them. Now, if the researcher misspecifies the non-control dose such that he or she actually uses a dose *smaller* than the EC_5, the misspecified one-stage design will be *more* efficient than the best one-stage design on the region [EC_5, EC_{100}]. Thus, a two stage design will be much less efficient than the one-stage misspecified design

because it will pull the non-control design point in to a less optimal level, a level closer to the EC_5, in the second stage. In cases such as these, two-stage design does not improve design efficiency. Thus, these situations were left out of the simulation study.

The main conclusion to be drawn from the two stage simulations is that the two stage designs are highly efficient in most cases. However, there is an additional advantage to using the two stage procedure. Since the first stage contained a misspecified non-control design point, the second stage which is correcting for this error will require a different non-control design point to compensate for the misspecification. Thus, the two stage procedure automatically produces an design which is more than two levels. It has already been noted that such designs are more robust to *model* misspecification, are more robust to overdispersion and are more popular among biologists.

Chapter 8

D-Optimal Designs for the Poisson Power Model

§8.1 Introduction

Recall the Poisson power model

$$y_{ij} = \beta_0 + \beta_1 x_i^{\beta_2} + \varepsilon_{ij} \tag{8.1.1}$$

where $y_{ij} \sim$ Poisson(λ_i). Because the model has three parameters, the experimental designs for this model will have to have at least three design levels. Therefore, three and four level optimal designs are found in this chapter based on the power model. Only D-optimal designs are found in this dissertation because the expressions for the F-optimal and Q-optimal designs became unmanageable even with only three levels. Nevertheless, since the model is used in practice, it is of interest to see some optimal designs for a non-GLM Poisson model however limited the scope of the study.

§8.2 D-Optimal Designs

From Chapter 2,

$$\mathbf{I} = \begin{bmatrix} \sum_{i=1}^{N} \frac{1}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^{\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{\beta_2} \ln(x_i)}{\lambda_i} \\ \sum_{i=1}^{N} \frac{x_i^{\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{x_i^{2\beta_2}}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{2\beta_2} \ln(x_i)}{\lambda_i} \\ \sum_{i=1}^{N} \frac{\beta_1 x_i^{\beta_2} \ln(x_i)}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1 x_i^{2\beta_2} \ln(x_i)}{\lambda_i} & \sum_{i=1}^{N} \frac{\beta_1^2 x_i^{2\beta_2} (\ln(x_i))^2}{\lambda_i} \end{bmatrix}. \tag{8.2.1}$$

The determinant of **I** for a three level design with n_1 observations taken at design point x_1, n_2 observations taken at design point x_2 and n_3 observations taken at design point x_3 is, after much simplification, given by

$$|\mathbf{I}| = \frac{n_1 n_2 n_3 \lambda_c}{\beta_1^2 \beta_2^2 q_1 q_2 q_3}[(q_2 - q_1 q_2 - q_3 + q_1 q_3)\ln(q_1 - 1) + (q_1 q_2 - q_1 + q_3 - q_2 q_3)\ln(q_2 - 1) + (q_1 - q_2 - q_1 q_3 + q_2 q_3)\ln(q_3 - 1)]^2 \quad (8.2.2)$$

where $\lambda_1 = q_1\lambda_c$, $\lambda_2 = q_2\lambda_c$, and $\lambda_3 = q_3\lambda_c$ are the means at design points x_1, x_2 and x_3, respectively. The goal is to maximize this determinant. The quantity $n_1 n_2 n_3$ will be maximized when $n_1 = n_2 = n_3$. Therefore, one-third of the observations should be at each of the three design levels in the optimal designs. Note that the control, the EC_{100}, cannot be selected as one of the three design points because q_i at the control is 1, and ln(0) does not exist. Furthermore, the EC_0 cannot be chosen as a design point because q_i at this point is 0, and that would appear in the denominator of the determinant.

The optimal three level designs for the power model for several regions of operability are given in Table 8.1. The optimal design on the entire region (EC_0, EC_{100}) has one-third of the observations at the design level which produces the smallest non-zero reproduction possible, one-third of the observations at the design level which produces the largest possible reproduction smaller than that at the control and one-third of the observations at an EC near the center of the region. The center level depends on what the upper and lower levels can be, and this of course will depend on the individual experimental situation. The optimal designs for this model tend to be designs which provide the maximum possible distance among the three design points.

Optimal four level designs for the power model were also considered. The expression for the determinant for a four level design is analogous to the expression for the three level design. As in previous chapters, due to limitations of the Nelder-Mead algorithm, restrictions were placed on the designs. Since the control is not a possible design point, a level could not be fixed there. Thus, the restrictions were that each design level had to be equidistant from the level on either side of it, and equal sample allocation

Table 8.1 Three Level D-Optimal Designs for the Poisson Power Model and the D-Efficiency for Each.

Region of Operability	p_1	p_2	p_3	x_1	x_2	x_3	D-Efficiency
$[EC_1, EC_{99}]$	1/3	1/3	1/3	EC_1	$EC_{46.9}$	EC_{99}	1
$[EC_{20}, EC_{80}]$	1/3	1/3	1/3	EC_{20}	$EC_{48.5}$	EC_{80}	0.0461
$[EC_{10}, EC_{90}]$	1/3	1/3	1/3	EC_{10}	$EC_{47.3}$	EC_{90}	0.1566
$[EC_{10}, EC_{50}]$	1/3	1/3	1/3	EC_{10}	$EC_{27.3}$	EC_{50}	0.0216

of 1/4 at each design point was assumed. The optimal four level designs over several regions are given in Table 8.2. In each case, the optimal design levels divide the region into three equal partitions.

The efficiencies of the designs in both tables are quite poor, but these again may not be fair comparisons. These efficiencies compare the designs to the overall optimal design on the entire region (in the both cases, this is considered to be the first design in the table, the best design on the region $[EC_1, EC_{99}]$). As stated in previous chapters, if the experimenter can only operate on a limited region of the dose space, then the best design on that region is the optimal design, and designs on regions not possible for the experimenter should not be considered.

§8.3 Conclusions

Of the two GLM Poisson models, the linear model more closely resembles the Poisson power model. Like the criteria for the linear model, the D-optimality criterion for this model prefers designs with maximum spacing among the design levels. Any attempt, either through constraints or restricted regions, to pull the levels in or make them closer together results in a rather serious loss in efficiency. The designs on restricted regions with the fewest constraints, however, should technically be considered the optimal designs on those regions. If limitations on the experimenter required the region of operability to be restricted, then the best design on that restricted region is the best that the experimenter can do.

Table 8.2 Four Level D-Optimal Designs for the Poisson Power Model and the D-Efficiency for Each.

Region of Operability	$p_1 = p_2 = p_3 = p_4$	x_1	x_2	x_3	x_4	D-Efficiency
$[EC_1, EC_{99}]$	1/4	EC_1	$EC_{33.7}$	$EC_{66.3}$	EC_{99}	1
$[EC_{20}, EC_{80}]$	1/4	EC_{20}	EC_{40}	EC_{60}	EC_{80}	0.0454
$[EC_{10}, EC_{90}]$	1/4	EC_{10}	$EC_{36.7}$	$EC_{63.3}$	EC_{90}	0.1550
$[EC_{10}, EC_{50}]$	1/4	EC_{10}	$EC_{23.3}$	$EC_{36.7}$	EC_{50}	0.0212

Chapter 9
Future Research and Other Considerations

§9.1 Asymptotics

Any designs which are based upon the Fisher Information matrix are only asymptotically the best designs. A small simulation study was conducted to investigate how well some of the one stage designs in this dissertation perform for smaller sample sizes. In the study, sample sizes of 16, 30 and 50 were used, and it was found that even for a sample size of 16, the simulated variances are very close to the elements of the inverse of the information matrix.

An additional concern present in the two stage design procedure is that the variances in the two stage procedure and the variances in the one stage procedure do not converge to the asymptotic variances at the same rate. Letsinger (1995) conducted a study to investigate this question for the designs developed from the logistic regression model. His results show that variances for the two stage procedure converge more quickly than those for the one stage. Thus, in general, the efficiencies based on the asymptotic variances are probably underestimates of the true efficiencies.

A study similar to the one conducted by Letsinger could be done for the models in this dissertation as well, but that will be left to future research. Furthermore, although Letsinger showed that the variances for the two stage procedure converge more quickly than the variance for the one stage, the results are not quite as drastic if the sample size is 60 or greater. A sample size of 60 was used in all two stage simulations in this paper, and the results of the one stage asymptotic investigation conducted in this research indicate that a sample size as small as 50 still produces variances close to the asymptotic values.

§9.2 Other Models

Although four models were considered in this dissertation, this research is certainly not exhaustive. Many other Poisson models and impairment models exist. A logical extension of this research is to apply some of the optimal design techniques to other models of interest to researchers. A study of robustness to model among several models would be interesting.

Transforming the response is another popular approach to modeling Poisson data. A transformation to make the data as close to normal as possible is the usual technique. Naturally, optimal design of experiments for normal error linear models is a well developed topic. However, data transformation carries its own set of difficulties. Nevertheless, an interesting extension of this work would be to compare designing for transformed models to designing for actual Poisson regression models, in terms of performance of the final models.

§9.3 Sample Allocation in Two Stage Design

As previously mentioned in Chapter 7, only two possibilities were considered in this research for sample allocation to the two stages in the two stage design procedure. No study was conducted to try to allocate observations to the two stages in any optimal way. Since sample allocation is important, particularly in fields in which the available number of experimental units is limited, this would be a worthwhile investigation.

§9.4 Bayesian Design

In Chapter 7, Bayesian design was mentioned as a way to deal with not knowing the model parameter before experimentation. Bayesian methods have been successfully

applied to designing for logistic regression (Letsinger, 1995), and may be applied to the designs in this dissertation as well.

§9.5 Models with More Than One Regressor

The only model considered in this research which was permitted to have more than one regressor was the normal error linear model with "percent impairment" as the response variable. In the Air Force study, one major goal was to investigate interaction among toxicants. However, when more independent variables become involved in the expressions for the design criteria, the expressions become unmanageable very quickly. Some two variable design has been considered for the logistic model (Jia, 1996), and similar techniques could be carried over to Poisson regression models.

Bibliography

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

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

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

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.

Finney, D.J. (1971), *Probit Analysis* (3rd edition), Cambridge University Press.

Finney, D.J. (1978), *Statistical Method in Biological Assay* (3rd edition), London: C. Griffin.

Jia, Y. (1996), *Optimal Experimental Designs for Two-Variable Logistic Regression Models*, Ph.D. Thesis, Virginia Tech

Kalish, L.A. and Rosenberger, J.L. (1978), "Optimal Designs for the Estimation of the Logistic Function," Technical Report 33, Department of Statistics, Pennsylvania State University.

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

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

Letsinger, W. (1995), "Optimal One and Two-Stage Designs for the Logistic Regression Model," Ph.D. Thesis, Virginia Tech

McCullagh, P. and Nelder, J.A. (1989), *Generalized Linear Models* (2nd edition), New York: Chapman and Hall.

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

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

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

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.

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 Biopharmaceutical Statistics,* **6** (to appear).

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

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.

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.

Smith, E.P., personal communication

Appendix

APPENDIX A

When $\lambda_2 \geq \lambda_1$ and $\lambda_1 = k\lambda_2$ for some $0 < k \leq 1$,

$$|\mathbf{I}| = \left(\frac{1}{\beta_1^2}\right) n^2 k \lambda_2^2 (\ln k)^2. \tag{A1}$$

From this expression, it seems that, for a given k, maximizing λ_2 will maximize the determinant of **I**. The largest λ_2 can be is λ_c, the mean at the control. Given that $\lambda_2 = \lambda_c$, the optimal value for k can be found using calculus. The first derivative of the determinant of **I** with respect to k is given by

$$\frac{\partial |\mathbf{I}|}{\partial k} = \left(\frac{1}{\beta_1^2}\right)[n^2 \lambda_2^2 (\ln k)^2 + 2n^2 \lambda_2^2 k (\ln k)(1/k)]. \tag{A2}$$

Setting this derivative equal to 0 is equivalent to setting

$$(\ln k)^2 + 2 \ln k = 0. \tag{A3}$$

Solving this expression for k gives

$$\ln k(\ln k + 2) = 0. \tag{A4}$$

Thus, either $\ln k = 0$ or $\ln k = -2$. The first solution makes the determinant of **I** equal to 0, and hence is a minimum. The second solution,

$$\ln k = -2$$

$$\Rightarrow \quad k = e^{-2}$$

maximizes the determinant.

APPENDIX B

The approximate prediction variance at a point x_0 for a two level design is

$$\frac{\lambda_0^2\left(n_1\lambda_1\left(\ln\frac{\lambda_1}{\lambda_0}\right)^2+n_2\lambda_2\left(\ln\frac{\lambda_2}{\lambda_0}\right)^2\right)}{n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2} \tag{B1}$$

$$= \frac{\lambda_0^2(n_1\lambda_1(\ln\lambda_1-\ln\lambda_0)^2+n_2\lambda_2(\ln\lambda_2-\ln\lambda_0)^2)}{n_1n_2\lambda_1\lambda_2(\ln\lambda_1-\ln\lambda_2)^2}$$

where $\lambda_0 = e^{\beta_0+\beta_1x_0}$ and $\ln\lambda_0 = \beta_0 + \beta_1x_0$. Substituting these quantities and letting $n_1n_2\lambda_1\lambda_2\left(\ln\frac{\lambda_1}{\lambda_2}\right)^2 = D$ yields the following expression for the prediction variance:

$$\frac{e^{2(\beta_0+\beta_1x_0)}}{D}[c_1+c_2x_0+c_3x_0^2] \tag{B2}$$

where

$$c_1 = n_1\lambda_1(\beta_0 - \ln\lambda_1)^2 + n_2\lambda_2(\beta_0 - \ln\lambda_2)^2, \tag{B3}$$

$$c_2 = 2\beta_1[n_1\lambda_1(\beta_0 - \ln\lambda_1) + n_2\lambda_2(\beta_0 - \ln\lambda_2)], \tag{B4}$$

and $$c_3 = \beta_1^2(n_1\lambda_1 + n_2\lambda_2). \tag{B5}$$

Now,

$$\int_{R_L}^{R_U}\frac{c_1e^{2(\beta_0+\beta_1x_0)}}{D} = \left(\frac{c_1}{2\beta_1D}\right)e^{2(\beta_0+\beta_1x_0)}\Bigg|_{R_L}^{R_U}. \tag{B6}$$

Using integration by parts,

$$\int_{R_L}^{R_U}\frac{c_2e^{2(\beta_0+\beta_1x_0)}}{D} = \left(\frac{c_2}{2\beta_1D}\right)e^{2(\beta_0+\beta_1x_0)}\left(x_0-\frac{1}{2\beta_1}\right)\Bigg|_{R_L}^{R_U} \tag{B7}$$

and

$$\int_{R_L}^{R_U}\frac{c_3e^{2(\beta_0+\beta_1x_0)}}{D} = \left(\frac{c_3}{2\beta_1D}\right)e^{2(\beta_0+\beta_1x_0)}\left(x_0^2-\frac{x_0}{\beta_1}+\frac{1}{2\beta_1^2}\right)\Bigg|_{R_L}^{R_U}. \tag{B8}$$

Now

$$R_U = \frac{\ln\lambda_U - \beta_0}{\beta_1}$$
$$\text{and } R_L = \frac{\ln\lambda_L - \beta_0}{\beta_1}. \tag{B9}$$

Evaluating the integrals in B6, B7, and B8 at these points and adding the integrals gives

$$\left(\frac{1}{2\beta_1 D}\right)\left(\lambda_U{}^2 - \lambda_L{}^2\right)\left(n_1\lambda_1(\beta_0 - \ln\lambda_1)^2 + n_2\lambda_2(\beta_0 - \ln\lambda_2)^2\right)$$
$$+\left(\frac{1}{2\beta_1 D}\right)\left(n_1\lambda_1(\beta_0 - \ln\lambda_1) + n_2\lambda_2(\beta_0 - \ln\lambda_2)\right)\left(\lambda_U{}^2(2\beta_1\left(\frac{\ln\lambda_U - \beta_0}{\beta_1}\right) - 1) - \right.$$
$$\left.\lambda_L{}^2(2\beta_1\left(\frac{\ln\lambda_L - \beta_0}{\beta_1}\right) - 1)\right) \tag{B10}$$
$$+\left(\frac{1}{4\beta_1 D}\right)\left(n_1\lambda_1 + n_2\lambda_2\right)\left(\lambda_U{}^2(2\beta_1{}^2\left(\frac{\ln\lambda_U - \beta_0}{\beta_1}\right)^2 - 2\beta_1\left(\frac{\ln\lambda_U - \beta_0}{\beta_1}\right) + 1) - \right.$$
$$\left.\lambda_L{}^2(2\beta_1{}^2\left(\frac{\ln\lambda_L - \beta_0}{\beta_1}\right)^2 - 2\beta_1\left(\frac{\ln\lambda_L - \beta_0}{\beta_1}\right) + 1)\right).$$

After much simplification, the β's cancel out, and the expression for the integral becomes A/D where

$$A = n_1\lambda_1\lambda_U{}^2\left[2\left(\ln\frac{\lambda_1}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_U}\right) + 1\right] - n_1\lambda_1\lambda_L{}^2\left[2\left(\ln\frac{\lambda_1}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_1}{\lambda_L}\right) + 1\right]$$
$$+ n_2\lambda_2\lambda_U{}^2\left[2\left(\ln\frac{\lambda_2}{\lambda_U}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_U}\right) + 1\right] - n_2\lambda_2\lambda_L{}^2\left[2\left(\ln\frac{\lambda_2}{\lambda_L}\right)^2 + 2\left(\ln\frac{\lambda_2}{\lambda_L}\right) + 1\right].$$

Vita

Lisa Marie Chiacchierini, daughter of Richard P. and Kathleen O. Chiacchierini, was born on November 23, 1970 in Radford, Virginia. She grew up Gaithersburg, Maryland, and graduated as salutatorian from Seneca Valley High School in 1988. In 1992, she graduated Summa Cum Laude with a Bachelor of Science degree in Mathematics from Villanova University in Villanova, Pennsylvania. Upon graduation, she was awarded the Emil Amelotti Award for the highest grade point average among Mathematics majors. In August 1992, she entered the graduate program in Statistics at the Virginia Polytechnic Institute and State University in Blacksburg, Virginia on a Cunningham Fellowship. She completed the requirements for a Master of Science degree in Statistics in December 1993, and was awarded the Klaus Hinkelmann Award for the Outstanding Graduate Assistant in 1996. She received her Doctor of Philosophy degree in Statistics in December 1996.

Lisa M. Chiacchierini