Optimal One and Two-stage Designs for the Logistic Regression Model

by

William C. Letsinger II

Dissertation submitted to the faculty of
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

Jesse C. Arnold

Klaus H. Hinkelmann

Marion R. Reynolds, Jr.

Jeffrey B. Birch

February 7, 1995
Blacksburg, Virginia

Optimal One and Two-stage Designs for the Logistic Regression Model

by

William C. Letsinger II

Raymond H. Myers, Chair
Department of Statistics
Virginia Polytechnic Institute and State University

ABSTRACT

Binary reponse data is often modeled using the logistic regression model, a well known nonlinear model. Designing an optimal experiment for this nonlinear situation poses some problems not encountered with a linear model. The application of several optimality design criteria to the logistic regression model is explored, and many resulting optimal designs are given. The implementation of these optimal designs requires the parameters of the model to be known. However, the model parameters are not known. If they were, there would be no need to design an experiment. Consequently the parameters must be estimated prior to implementing a design.

Standard one-stage optimal designs are quite sensitive to parameter misspecification and are therefore unsatisfactory in practice. A two-stage Bayesian design procedure is developed which effectively deals with poor parameter knowledge while maintaining high efficiency. The first stage makes use of Bayesian design as well as Bayesian estimation in order to cope with parameter misspecification. Using the parameter estimates from the first stage, the second stage conditionally optimizes a chosen design optimality criterion. Asymptotically, the two-stage design procedure is considerably more efficient than the one-stage design when the parameters are misspecified and only slightly less efficient when the parameters are known. The superiority of the two-stage procedure over the one-stage is even more evident for small samples.

DEDICATION

My late father, Warren C. Letsinger, instilled in me a thirst for knowledge that has never wavered. It is to him that I would like to dedicate this work.

ACKNOWLEDGEMENTS

I would like to thank my family, especially my mother, Sara H. Letsinger, and my wife, Jennifer D. Letsinger for their love, support, and encouragement. The support of my family and friends has enabled me to complete this endeavor.

In addition I wish to thank Dr. Raymond H. Myers for being my advisor and sharing his knowledge with me. His guidance, support, and friendship has been invaluable.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction and Literature Review

§1.1 Introduction

With the current emphasis in the sciences and industry on the need for developing new products as well as improving quality of existing products on ever shrinking development/research budgets, response surface methodology (RSM) is enjoying increasingly widespread use. The increased demand being placed upon RSM over the past number of years has stimulated a considerable amount of research aimed at broadening and improving the scope and applications which cover areas including design of experiments and linear and nonlinear modeling.

The aspect of RSM of direct interest in this dissertation is design of experiments. The majority of research done in this area is based on the assumption of a linear model, although some important work has also been done with a nonlinear model. Regardless of the nature of the model, it is safe to assume that cost is an increasingly important factor in most experiments today. Consequently, it is desirable to find designs which yield a maximal amount of pertinent information about the system under study while using a minimal number of observations. In addition, it is also desirable for a design to be attractive to practitioners in terms of usefulness and applicability as well as being theoretically attractive to the statistician. One objective of this research is to find designs which have a combination of these attributes. The assumed model is the logistic regression model, a widely known nonlinear model for binary response data.

§1.2 Linear Model and Design Optimality

In the quest for optimal designs, many optimal design criteria have been created, along with optimal design theory. In 1959, Kiefer and Wolfowitz formed a theoretical framework for optimal design criteria by expressing a design as a probability measure representing the allocation of observations at any particular point in the design space. They formed the D and E optimality criteria, which led to the term "alphabetic optimality." In addition, many other design performance criteria have been created such as A, F, G, and Q. With the exception of F optimality, these criteria were initially developed for a linear model. This is not to imply that these criteria could not be used in a nonlinear situation. However, the linear model does lend itself more nicely to design optimization than does the nonlinear model as will become apparent shortly.

Consider the linear model in standard form,

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

where $\mathbf{y}$ is an n dimensional vector of responses, $\mathbf{X}$ is an $n \times (h+1)$ matrix formed by an $n \times h$ model matrix of h regressor variables augmented by a column of ones, $\boldsymbol{\beta}$ is a h+1 dimensional vector of unknown parameters, and $\boldsymbol{\epsilon}$ is an n dimensional vector of unknown errors. Note that a regressor variable might be defined as some function of the other regressor variables such as interactions (products) or quadratic terms.

Under the assumption that $\boldsymbol{\epsilon}$ is $N(\mathbf{0},\sigma^2\mathbf{I})$, the method of least squares is used to estimate the vector of parameters, $\boldsymbol{\beta}$. Under this assumption, the least squares estimator of $\boldsymbol{\beta}$ is equivalent to the maximum likelihood estimator, $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ which is the minimum variance unbiased estimator of $\boldsymbol{\beta}$. The variance of $\mathbf{b}$, the inverse of the Fisher information matrix, is given by $(\mathbf{X}'\mathbf{X})^{-1}\sigma^2$. This variance covariance matrix of $\mathbf{b}$ as a whole or in part drives many of the design criteria which have been developed. The matrix $(\mathbf{X}'\mathbf{X})^{-1}$ does not depend upon the unknown parameter vector $\boldsymbol{\beta}$, which is precisely why the linear model lends itself so nicely to design optimization. In the next section, however, we will see that the same does not hold true for a nonlinear model.

The most widely known design criterion is D optimality, dating to Kiefer and Wolfowitz (1959). This criterion suggests one should make $(\mathbf{X'X})^{-1}$ small or equivalently, make $(\mathbf{X'X})$ large in some sense, thus attempting to minimize the variance of $\mathbf{b}$. The determinant of the variance covariance matrix of $\mathbf{b}$ is termed the generalized variance of $\mathbf{b}$ and was used as scalar measure of dispersion by Wilks (1932). Thus, the D optimal design is the design which maximizes the determinant of $(\mathbf{X'X})$. The D optimality criterion is given by

$$\underset{\mathfrak{D}}{\text{Max}} \left| \frac{\mathbf{X'X}}{N} \right|$$

where $\mathfrak{D}$ is the set of possible designs and N is the sample size. The matrix and determinant depend only upon the location of the design points, not on β. Consequently, the D optimal design can be determined and implemented independently of β.

Another well known design criterion is Q optimality. The goal of the Q criterion is to find a design which offers good prediction throughout the region in which the experimenter wants to predict. Consider the prediction variance at a point $\mathbf{x}_0$,

$$\text{var}(\hat{y}(\mathbf{x}_0)) = \text{var}(\mathbf{x}_0'\mathbf{b}) = \mathbf{x}_0'(\mathbf{X'X})^{-1}\mathbf{x}_0\sigma^2$$

where $\mathbf{x}_0' = (1, x_{01}, x_{02}, \ldots , x_{0h})$ is a location in the region of interest written in model space identical to rows of the $\mathbf{X}$ matrix. One way to address the quality of prediction in a region is to average the prediction variances over that region. The Q optimal design is the design which minimizes the average prediction variance (APV) over all possible locations within the region of interest. The Q optimality criterion is given by

$$\underset{\mathfrak{D}}{\text{Min}} \ \frac{N}{K} \int_R \mathbf{x}_0'(\mathbf{X'X})^{-1}\mathbf{x}_0 \, d\mathbf{x}_0$$

where R is the region of interest and K is the volume of that region. The idea of average prediction variance originated with Box and Draper (1959) in an effort to find a design criterion which protects against model misspecification.

E optimality is another criterion which addresses prediction variance. The E optimal design is the design which minimizes the maximum eigenvalue of $(\mathbf{X}'\mathbf{X})^{-1}$ which has been shown for the linear model situation to be equivalent to minimizing maximum prediction variances on spheres of radius r. The E optimality criterion is given by

$$\underset{\mathcal{D}}{\text{Min}} \ \underset{i}{\text{Max}} \ \lambda_i.$$

Many such design criteria have been developed and all tend to fall within one of two categories, those concerned with parameter estimation and those concerned with prediction. Regardless of the category, in the linear model situation, the D, Q, E, optimal designs are found and implemented independently of the model parameters and are also independent of the scale of the design variables. More is said later about scale invariance. In addition, these three criteria yield optimal designs which are quite similar in most circumstances. It will be shown that the same cannot be said of the same criteria in the context of the logistic model. For more detailed instruction on the design criteria discussed and their use in RSM see Myers, Khuri, and Carter(1989).

§1.3 Nonlinear Model (Logistic) and Design Optimality

The use of nonlinear models is becoming increasingly prevalent as the applications of statistics broadens. In many instances the normality and constant variance assumptions used in most linear models contexts are not valid. Situations often arise in which the errors are more likely distributed Poisson, binomial, exponential, gamma, or some other member of the GLM family of models. When one of these situations arise, nonlinear models typically exist which are better suited than the linear model. Consider the nonlinear model,

$$y_{ij} = f(\mathbf{x}_i, \boldsymbol{\beta}) + \epsilon_{ij} \qquad (1.3.1)$$

where i=1,...,k, j=1,...,n_i, y_{ij} is the j^{th} response for the i^{th} design point, f is a known function which is nonlinear in β, where β is a $p \times 1$ vector of unkown model parameters, $\mathbf{x}$ is an $h \times 1$ vector of regressor variables, and ϵ_{ij} is an unknown error.

An appropriate method for estimating β is maximum likelihood estimation. The asymptotic variance covariance matrix of $\mathbf{b}$, the MLE of β, is the inverse of the Fisher information matrix (Lehmann 1983). This information matrix forms the foundation for traditional design performance criteria and is defined by

$$\mathbf{I}_{ij}(\mathbf{x},\beta) = \left[-E\left(\frac{\partial^2 L(\beta)}{\partial \beta_i \partial \beta_j}\right) \right] ; \quad (i=1,..., p \text{ and } j=1,..., p)$$

where $L(\beta)$ is the log-likelihood function from model 1.3.1.

Relatively little research has been done in the area of design optimization in the case of nonlinear models, but there is good reason for this. Box and Lucas (1959) demonstrated that the D optimal design for a nonlinear model depends on the model parameters, β, through the partial derivatives in the Fisher information matrix. In order to implement the D optimal design one must know the model parameters.

The particular nonlinear model under study in this dissertation, as previously mentioned, is the logistic regression model. This model is well suited for binary data which often arise in areas including biological assays of drugs, material fatigue to failure studies, and stress and fracture studies. Research questions of interest may center around the amount of drug needed to produce a 50% chance of response, the level of poison (pesticide/herbicide) required to yield an 80% death rate, the amount a piece of composite material can be stressed before the chance that it fractures exceeds 20%, or the minimum temperature a component can experience and still give chance of failure not to exceed 25%. The logistic function and model was first introduced by Berkson (1944) in the context of biological drug dose/response experiments, the subject matter context in which it has remained until recently. The logistic distribution has a symmetric probability density function and apart from constants is given by

$$f(\mathbf{x}) = \frac{\exp^{-\boldsymbol{\beta}'\mathbf{x}}}{(1 + \exp^{-\boldsymbol{\beta}'\mathbf{x}})2} \tag{1.3.2}$$

The single variable logistic regression model is derived from the corresponding cumulative distribution function and is commonly seen in the following form,

$$y_{ij} = P_i + \epsilon_{ij}$$

or

$$= \frac{1}{1 + \exp^{-\boldsymbol{\beta}'\mathbf{x}_i}} + \epsilon_{ij} \tag{1.3.3}$$

where, $j=1,\ldots,n_i$, $y_{ij}= 0$ or 1 is the j^{th} response at the i^{th} design point, P_i is the unknown probability of success, $\boldsymbol{\beta}$ is an $m \times 1$ vector of unknown parameters, and $\mathbf{x}_i$ is a $(h+1) \times 1$ vector of h regressor variables augmented by a 1, and $y_{ij} \sim$ Bernoulli(P_i). In this research h=1 and m=2; i.e. one regressor and two parameters. In this form the model is written as

$$y_{ij} = \frac{1}{1 + \exp^{-(\beta_0+\beta_1 x_i)}} + \epsilon_{ij}$$

or

$$= \frac{1}{1 + \exp^{-\beta(x_i-\mu)}} + \epsilon_{ij}$$

where $\beta_1=\beta$, the scale parameter and $-\frac{\beta_0}{\beta_1} = \mu$, the location parameter. The parameter μ represents the value of x which yields a probability of success of .50, often termed the effective dose 50 (ED_{50}).

Suppose one is interested primarily in the estimation of μ. A seemingly reasonable design criterion would involve the width of an asymptotic confidence interval for μ. Finney (1971, 1978) suggested the use of a Fieller interval criterion (F optimality) instead. The F criterion is based on a fiducial interval about μ, where μ is estimated by the ratio of two random variables, (Fieller, 1944). The fiducial interval used in F optimality is given by Fieller's theorem and is therefore termed a Fieller interval. The F optimal design is the design which minimizes the

length of a Fieller interval about the parameter of interest. Although some view the Fieller interval as a fiducial interval, it is best viewed as an improved confidence interval.

Abdelbasit and Plackett (1983) studied the use of the Fieller interval as a design criterion for the probit model. Based upon a limited empirical study they found Finney's claim of the Fieller interval being superior to the asymptotic confidence interval to be unfounded. Sitter and Wu (1993) conducted a more extensive empirical study of coverage and tail probabilities and concluded in favor of Finney. Several F optimal as well as D and A optimal designs are given for the logistic model as well as other nonlinear models in another paper by Sitter and Wu (1994). The Fieller interval and the F optimality criterion are derived in detail in chapter 2.

Many such optimal design criteria for nonlinear regression models have either been developed or adapted from design criteria used in linear model situations. For a binary response model with a symmetric tolerance distribution, White (1975) derived the D optimal design. The D optimality expression becomes

$$\underset{\mathfrak{D}}{\text{Max}} \left| \frac{\mathbf{I}(\mathbf{x},\boldsymbol{\beta})}{\mathrm{N}} \right| .$$

Kalish and Rosenberger (1978) derived the D optimal and G optimal designs for the logistic model. Myers, Myers, and Carter (1994) derived the Q and G optimal designs for the logistic model in which the prediction variance of logit($\hat{\mathrm{P}}$) is addressed rather than the prediction variance of $\hat{\mathrm{P}}$. The logit transformation, assuming the logistic model in (1.3.3), is defined by

$$\text{logit}(\mathrm{P}) = \log(\frac{\mathrm{P}}{1\text{-}\mathrm{P}}) = \boldsymbol{\beta}'\mathbf{x}$$

and ranges from minus to plus infinity unlike P which is bounded by (0,1). The Q optimality expression based upon the prediction variance of Logit($\widehat{\mathrm{P}}$) is

$$\underset{\mathcal{D}}{\text{Min}} \quad \frac{N}{K} \int_R \mathbf{x}_0'(\mathbf{I}^{-1})\mathbf{x}_0 \, d\mathbf{x}_0.$$

The logit transformation linearizes the logistic model making the prediction variance expression parallel that of the linear model. Consequently closed formed expressions can be obtained for the integral. If however one attempts to address the prediction variance of $\hat{P}$ instead of logit($\hat{P}$), the situation is much more difficult.

Obtaining an expression for the prediction variance of $\hat{P}$ can in itself be a formidable task. This prediction variance can however be approximated by the delta method which makes use of a first order Taylor series expansion about the true parameters. Using this approximation, the Q expression is given by

$$\underset{\mathcal{D}}{\text{Min}} \quad \frac{N}{K} \int_R P_0^2(1-P_0)^2\mathbf{x}_0'(\mathbf{I}^{-1})\mathbf{x}_0 \, d\mathbf{x}_0 \qquad (1.3.4)$$

which is, apart from the squared Bernoulli variance at x_0, identical to the expression for the prediction variance of Logit($\hat{P}$). The derivation of the approximation is given in appendix A. The squared Bernoulli variance simply forces the prediction variance to zero as P goes to zero or one. Due to the nature of P, this integral has no apparent closed form solution, making it far less attractive to work with than the prediction variance of the logit.

For the one variable logistic model, the slope of the logistic curve at a regressor location in logit space, i.e. $z_0=-\beta(x_0-\mu)$, is given by the Bernoulli variance at that point z_0. This result is shown in appendix B. The second Q expression given by 1.3.4 can now be viewed in a more intuitive manner. The approximate prediction variance of $\hat{P}$ goes to zero as the slope of the logistic or response curve goes to zero. The term Bernoulli variance should be used synonymously with the term slope.

The E optimal design criterion minimizes the maximum eigenvalue of the inverse of the Fisher information matrix and is given by

$$\underset{\mathfrak{D}}{\text{Min}} \ \underset{i}{\text{Max}} \ \lambda_i,$$

where λ is an eigenvalue from $I(\mathbf{x},\beta)$. The existence of an equivalency in terms of prediction variance is addressed in chapter 3. Kalish (1990) gives several optimal designs using criteria that deal with the estimation of μ and also for the estimation of β for the logistic model. Through simulation, Kalish made comparisons among these designs for the logistic model as well as two other models. For additional references and information on design criteria and optimal design of experiments see Silvey(1980), Atkinson (1982), Box and Draper (1987).

Although several design criteria have been developed for various nonlinear models, the problem still exists concerning the need to know the parameters when trying to implement an optimal design. The optimal designs for all of the criteria given thus far can be expressed in terms of probabilities of response. However, the associated effective dose or level of the regressor variable is needed in order to implement the design. For instance, the D optimal design for the one variable logistic model is a two point design, with design points at the $ED_{17.6}$ and $ED_{82.4}$. This information is useless in practice unless one knows where these ED's occur in terms of the regressor variable. The parameters of the model are needed to transform the optimal probabilities of response into the optimal locations of the design points in terms of the explanatory variable.

In practice, parameters are not known and must be estimated or "guessed" prior to implementing a design. This motivates the notion of robustness to parameter misspecification. Consequently, to answer the call of practicality, work has been done in order to find optimal designs which are robust to poor parameter guesses.

A minimax procedure was proposed by Sitter (1992) which directly addresses the problem of poor parameter guesses. The procedure requires the experimenter to not only make *a priori* point estimates of the parameters but also supply intervals in which he or she believes the true parameters to reside. The procedure

finds the design with the best "worst case" within these chosen intervals with respect to some criterion. Most of the typical optimality criteria can be used with this minimax procedure. Sitter gives many minimax designs for the D, F, and asymptotic confidence interval criteria. The minimax procedure using the F criterion is discussed in greater detail in Chapter 4.

Also examined in chapter 4 is a Bayesian procedure for incorporating prior information about the parameters. The use of Bayesian methods in creating optimal designs is not new. Covey-Crump and Silvey (1970) introduced the notion of a Bayesian D optimality criterion in the linear model case. Several Bayesian optimal design methods have been developed and investigated for the linear model, see Brooks (1972), Bernardo (1979), Näther and Pilz (1980), and Chaloner (1984). An extensive bibliography is provided by Chaloner.

In 1989, Chaloner and Larntz developed a Bayesian design procedure for the logistic regression model. Their procedure, which can be used with most design criteria, can accommodate extremely poor parameter information, as does the minimax procedure, but without the conservative nature of a minimax criterion. The Bayesian aspect is only incorporated in the determination of the design and not in any inferences about the parameters.

Another way of dealing with poor parameter guesses is to proceed in stages. Box and Lucas (1959) suggested sequential experimentation. Wu (1985) gave sequential designs for binary data for estimating particular ED's based upon the nonparametric "up and down" method and the Robbins Monro method. Two-stage designs for the one variable logistic model were discussed by Abdelbasit and Plackett (1983) and Minkin (1987) using the D optimality criterion.

Myers (1991) developed a two-stage D-Q design for the logistic model. The first stage is a D optimal design which is relatively robustness to parameter misspecification. The second stage is a conditional Q optimal design, conditioned on the first stage. Myers found the two-stage design to be superior to the comparable one-stage design in all cases of parameter guesses except those extremely close to the true values. Much efficiency is gained when parameter knowledge is poor, but little is lost when the guesses are correct.

An important focus of this dissertation is the development and investigation of a practical two-stage design procedure utilizing Bayesian design and estimation in the first stage. However two other main points of interest are investigated first. Much attention is given to the study of various design criteria as they apply to the logistic model and to one-stage designs and their robustness properties.

The D, Q, F, E, and asymptotic confidence interval criteria are investigated as they apply to the logistic model. Unlike the other criteria, the F criterion recommended by Finney and Sitter and Wu is not intuitive and therefore needs development. Consequently, the F criterion is derived. Next the minimax procedure proposed by Sitter is studied and critiqued with respect to the F criterion. In addition, the Bayesian design procedure proposed by Chaloner and Larntz is discussed. Then an extensive effort is made to find simplified forms of the design criterion expressions in an attempt to investigate and relate the criteria. The purpose of this is to illustrate why the design criteria yield drastically different designs for the logistic model, unlike for the linear model. In addition, attention is given to the notion of scale invariance.

Many one-stage D, Q, E, and F optimal designs are given. Additionally, the Bayesian design procedure is applied to the D optimality criterion, and numerous one-stage Bayesian D optimal designs are provided. The robustness of these designs to parameter misspecification is studied through the use of asymptotic efficiencies. Furthermore, comparisons are made between Bayesian and non-Bayesian one-stage designs.

The robustness study of the one-stage designs motivates the need for better designs. A two-stage optimal design procedure is developed to fill this need. The first stage of the two-stage procedure makes use of the Bayesian design procedure. In addition, Bayesian estimation methods are used to estimate the model parameters at the first stage. The second stage then conditionally optimizes the second stage design criterion.

The two-stage design procedure is compared to one-stage designs using asymptotic efficiencies over many different levels of parameter misspecification. Additionally, the use of Bayesian design and estimation is studied through

comparisions among two-stage designs. A small sample study is then provided which displays the finite sample performance of the two-stage design procedure compared to the one-stage design. In addition, the small sample study investigates the appropriateness of the use of asymptotics for both one and two-stage designs.

CHAPTER 2

THE FIELLER INTERVAL AND F CRITERION FOR THE LOGISTIC MODEL

The Fieller interval for an ED is an alternative to the standard confidence interval. The derivation of the Fieller interval is quite intuitive while the resulting form of the interval is not. Both the Fieller interval and standard confidence interval are derived in order to illustrate their similarities and differences. The F optimality design criterion is then defined.

§2.1 Parameter of Interest

The one variable two parameter logistic regression model can be parameterized in several ways, one being:

$$y_{ij}=\frac{1}{1+\exp^{(\xi-\beta(x_i-\mu))}} + \epsilon_{ij} \qquad (2.1.1)$$

where $y_{ij} \sim$ Bernoulli(P_i), β is the "slope" parameter, ξ is a known adjustment parameter equal to -Logit(p) such that $\mu=ED_{100p}$. This parameterization will be

referred to as the centered model. Although μ can be equated to any ED, it is usually equated to the ED of primary interest or the ED at the center of the region of interest.

If μ is equated to the ED of interest, then F optimality is directly concerned with the estimation of μ. However, since μ=ED_{100p}, F optimality can be viewed equivalently as addressing inverse prediction given p. Although μ need not be equated to the ED of interest in order to address that particular ED in terms of an F optimal design, in this dissertation μ is always equated to the ED of interest. However, keep in mind that this is merely cosmetic. The Fieller interval about the ED_{100p} and F optimal design for predicting the ED_{100p} will be the same whether or not the ED_{100p} is equated to μ.

§2.2 The Fieller Interval and its Relationship to the Asymptotic Confidence Interval

Fieller's thereom and the Fieller interval can be easily derived if another more convenient parameterization of the logistic model is considered,

$$y_{ij} = \frac{1}{1 + \exp^{(\xi + \alpha - \beta x_i)}} + \epsilon_{ij} \qquad (2.2.1)$$

where β is the same as in 2.1.1 and $\mu = \frac{\alpha}{\beta}$. Both the centered and noncentered parameterizations will be used throughout the dissertation. The likelihood function for model 2.2.1 is given by

$$L(\alpha,\beta;\, r_i) = \prod_{i=1}^{k} \binom{n_i}{r_i} P_i^{r_i} (1-P_i)^{n_i - r_i}, \qquad (2.2.2)$$

where $$P_i = \frac{1}{1 + \exp^{(\xi + \alpha - \beta x_i)}},$$

n_i= the number of observations at the i^{th} design point location x_i, r_i=the number of "successes" out of n_i, and r_i~Binomial(n_i,P_i). Taking the natural logarithm of 2.2.2 yields

$$\text{Log L} = \sum_{i=1}^{k} (\log\binom{n_i}{r_i} + r_i \log(P_i) + (n_i - r_i)\log(1-P_i)).$$

Using this log likelihood equation, the Fisher information matrix is defined by

$$I(\alpha,\beta) = \begin{pmatrix} -E(\frac{\partial^2 \text{logL}}{\partial \alpha^2}) & -E(\frac{\partial^2 \text{logL}}{\partial \alpha \partial \beta}) \\ -E(\frac{\partial^2 \text{logL}}{\partial \alpha \partial \beta}) & -E(\frac{\partial^2 \text{logL}}{\partial \beta^2}) \end{pmatrix} \qquad (2.2.3)$$

where the elements of the Fisher information matrix for this situation are

$$I_{11} = -E(\frac{\partial^2 \text{logL}}{\partial \alpha^2}) = \sum_{i=1}^{k} n_i P_i (1-P_i) ,$$

$$I_{12} = -E(\frac{\partial^2 \text{logL}}{\partial \alpha \partial \beta}) = -\sum_{i=1}^{k} x_i n_i P_i (1-P_i) ,$$

and

$$I_{22} = -E(\frac{\partial^2 \text{logL}}{\partial \beta^2}) = \sum_{i=1}^{k} x_i^2 n_i P_i (1-P_i).$$

The inverse of the Fisher information matrix is then given by

$$I^{-1}(\alpha,\beta) = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{pmatrix} \qquad (2.2.4)$$

where $V_{11} = \frac{I_{22}}{D}$, $V_{12} = -\frac{I_{12}}{D}$, and $V_{22} = \frac{I_{11}}{D}$ and D is the determinant of $I(\alpha,\beta)$.

Notice that both 2.2.3 and 2.2.4 are functions of the model parameters, α and β, through P, and any design criterion based upon these matrices in part or in whole will also depend upon the parameters.

The inverse of the Fisher information matrix is the asymptotic variance covariance matrix for a and b, where a and b are the maximum likelihood estimates of α and β respectively. This matrix can therefore be used to form an asymptotic confidence interval for $\mu=\frac{\alpha}{\beta}$ and subsequently a Fieller interval for μ. Since a and b are mle's, they are distributed asymptotically normal with means α and β, respectively, and variance structure as described by the inverse of the Fisher information matrix. Hence, the linear combination $(a-\mu b)$ is also asymptotically normally distributed with mean 0 and variance V^2, where

$$V^2=V_{11}-2\mu V_{12}+\mu^2 V_{22}.$$

A $(1-\gamma)100\%$ asymptotic confidence for μ can then be formed by creating the probability statement

$$\Pr\left[-zV \leq (a - \mu b) \leq zV\right] = 1-\gamma \qquad (2.2.5)$$

where z is the $100(1-\frac{\gamma}{2})$th percentile of the standard normal distribution. Solving 2.2.5 for μ, we obtain

$$\Pr\left[\frac{a}{b} - z\frac{V}{b} \leq \mu \leq \frac{a}{b} + z\frac{V}{b}\right] = 1-\gamma.$$

The asymptotic confidence interval for μ is

$$\frac{a}{b} \pm z\frac{V}{b} \qquad (2.2.6)$$

or equivalently,

$$\frac{a}{b} \pm z\left(AV(\frac{a}{b})\right)^{\frac{1}{2}}$$

where $AV(\frac{a}{b}) = \frac{V^2}{b^2}$ is the asymptotic variance of the estimate of μ.

In order for the asymptotic interval to be applied, one must substitute into $AV(\frac{a}{b})$ the mle's of α and β. The asymptotic confidence interval assumes $AV(\frac{a}{b})$ and thus V are known even though they are functions of μ and $P_i(i=1...k)$, which in turn are functions of the unknown parameters α and β. This is precisely where the Fieller interval differs from the CI. The Fieller interval does not assume that V is known but rather treats it as a function of μ. However both intervals make use of $I^{-1}(\alpha,\beta)$ when the mle's of α and β are inserted. Consequently, the Fieller interval is also an asymptotic interval. The Fieller interval takes into account the imprecision in V due to μ being an unknown but does not address the imprecision in V due to P which also contains μ.

§2.3 Derivation of the Fieller Interval

To derive a Fieller interval, one simply solves the inequality in (2.2.5) for μ realizing that all three components are functions of μ. For simplicity, the inequality is squared and put into equation form so that

$$(a - \mu b)^2 = z^2 (V_{11}-2\mu V_{12}+\mu^2 V_{22}). \tag{2.3.1}$$

The roots of this quadratic equation in μ form Fieller's theorem and a $(1-\gamma)100\%$ Fieller interval for μ. With some rearranging, equation (2.3.1) can be written as a quadratic equation in μ in standard form,

$$\mu^2(b^2-z^2V_{22}) + \mu(2z^2V_{12} - 2(\tfrac{a}{b})b^2) + (b^2(\tfrac{a}{b})^2 - z^2V_{11}) = 0.$$

The quadratic formula is employed to solve this equation. In an unsimplified form the the solutions to equation (2.3.1) are

$$\frac{\left(2(\frac{a}{b})b^2-2z^2V_{12}\right) \pm \left\{(2z^2V_{12}-2(\frac{a}{b})b^2)^2 - 4(b^2-z^2V_{22})(b^2(\frac{a}{b})^2-z^2V_{11})\right\}^{\frac{1}{2}}}{2(b^2-z^2V_{22})}.$$

Letting $\mu=\frac{a}{b}$ and $g=\frac{z^2V_{22}}{b^2}$ and after some simplification, we have

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{V_{12}}{V_{22}}\right) \pm \frac{z}{b^2}\left\{z^2(V_{12})^2-2V_{12}\left(\frac{a}{b}\right)b^2+b^2V_{11}+V_{22}b^2\left(\frac{a}{b}\right)^2-z^2V_{22}V_{11}\right\}^{\frac{1}{2}}\right).$$

Finally, the roots to the quadratic equation are

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{V_{12}}{V_{22}}\right) \pm \frac{z}{b}\left\{V_{11} - 2V_{12}\left(\frac{a}{b}\right) + V_{22}\left(\frac{a}{b}\right)^2 + g\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}^{\frac{1}{2}}\right). \qquad (2.3.2)$$

This is Fieller's theorem. The two values given by 2.3.2 form the boundaries of the $(1-\gamma)100\%$ Fieller interval for μ. The interval is defined to be infinite if $g \geq 1$ or equivalently $\frac{|b|}{\sqrt{V_{22}}} \leq z$, in which case β is not significantly different from zero at the γ level.

Substituting the asymptotic variance of $\frac{a}{b}$, denoted $AV(\frac{a}{b})$, into 2.3.2, we can gain some insight into the relationship between the two intervals,

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{V_{12}}{V_{22}}\right) \pm z\left\{AV\left(\frac{a}{b}\right) + \frac{g}{b^2}\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}^{\frac{1}{2}}\right). \qquad (2.3.3)$$

Asymptotically, the Fieller interval and the asymptotic confidence interval are equivalent. The $(1-\gamma)100\%$ Fieller interval for μ converges to the corresponding $(1-\gamma)100\%$ confidence interval for μ as n, the total sample size, increases, and g goes to 0, see Sitter and Wu (1993, 1994) and Abdelbasit and Plackett (1983). However, as can be seen in (2.3.3), the Fieller interval is not necessarily symmetric about $\hat{\mu}$. Additionally, Sitter and Wu (1993) showed that the Fieller

interval is always longer than the asymptotic confidence interval. Finney suggests that the Fieller interval is only worth the extra effort over calculating the asymptotic CI when g>.05, or equivalently when $z < |b|/\sqrt{V_{22}} < 4.47z$. For g<.05, the difference between the two intervals is considered by Finney to be negligible due to an extremely significant β. In the limited study by Abdelbasit and Plackett, g was less than .05, and consequently, they noticed little difference between the two intervals.

The larger size of the Fieller interval over the asymptotic CI is intuitive since it is accounting for some of the imprecision in V due to μ being unknown. For small n, the difference can be quite dramatic. The significance of b relative to $(V_{22})^{\frac{1}{2}}$, through the quantity g, plays the main role in determining the additional length as well as the asymmetry of the interval. This is why Finney weighs the usefulness of the Fieller interval in terms of g. Typically, the more observations one has, assuming a reasonable design, the better the estimate of β, the smaller g becomes, and hence the closer the Fieller interval resembles the asymptotic CI.

The Fieller interval and resulting F optimal design depend upon n and γ. The effect of n is quite dramatic, and thus many values of n are investigated. For instance, the 95% F interval is nearly three times longer than the asymptotic CI for $\mu=ED_{50}$ when using the F optimal two point design with $n_1=n_2=5$. The estimate of the slope is relatively poor with such a small number of observations. For the same situation, however, the two intervals are very similar in length when $n_1=n_2=55$; the F interval is only about 4% longer. The convergence rate is slower for smaller γ as well as when μ represents ED's other than the ED_{50}.

Sitter and Wu (1993) showed for a total sample size as low as 40, the coverage probabilities are reasonable for both intervals, although the Fieller interval is better. The main discrepancy resides in the individual tail areas. The F interval has tail areas closer to nominal, $\frac{\gamma}{2}$, than does the asymptotic confidence interval for small to moderate sample sizes. The asymmetries of the F interval allow it to maintain the close to nominal tail area.

§2.4 The F Optimality Design Criterion

The F optimal design is the design, δ, from the set of all possible designs, $\mathfrak{D}$, which minimizes the length of a (1-γ)100% Fieller interval for μ. The half length multiplied by β is an equivalent measure which is scale invariant. More is said about scale invariance in the next chapter. The F optimality design criterion reduces to

$$\underset{\mathfrak{D}}{\text{Min}} \ \frac{1}{(1\text{-}g)}\left\{V_{11} - 2\mu V_{12} + \mu^2 V_{22} + g\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}^{\frac{1}{2}}. \qquad (2.4.1)$$

This definition implies that the parameters need to be known, as they appear throughout the expression directly and indirectly. Conseqently initial guesses of the parameters must be provided in order to implement the F optimal design.

CHAPTER 3

VARIOUS DESIGN OPTIMALITY CRITERIA FOR THE LOGISTIC REGRESSION MODEL : SOME INSIGHTS AND RELATIONSHIPS

The various design criteria discussed thus far create very different designs for the logistic regression model. This is in sharp contrast to the linear model situation where relatively similar designs are created. Even though these design criteria address different aspects of design optimality and produce drastically different designs for the logistic model, similarities exist among them. These similarities and differences are studied in this chapter through expressions or components including Bernoulli variances (slopes) at the design points, spread of the design points, and distance from the points to μ. Each criterion is expressed in a scale-free form and when possible in terms of these components. Much insight can be gained by observing which component(s) the criterion exploits. The D, asymptotic CI, F, Q, and E optimality criteria are examined in this manner. However, first the notion of scale invariance is addressed.

§3.1 Scale Invariance

Two distinct areas of scale invariance warrant discussion in the context of design optimality criteria and the logistic regression model. Let the property of absolute scale invariance imply that the value of a design criterion expression is invariant to location shifts and scale changes of the design variable. Consequently, the resulting optimal design (in terms of P_i's) is also invariant to such changes in location and scale. For the logistic model the location shifts and scale changes are specified by changes in μ and β, respectively. Next consider the property of relative scale invariance which implies that the value of a criterion expression is not invariant but proportional for changing μ and β. Such a situation also yields optimal designs that are invariant to changes in the model parameters.

For the linear model, the D and E criteria possess the property of relative scale invariance, while the Q criterion has the property of absolute scale invariance. The A criterion which has not been previously mentioned represents the trace of $(\mathbf{X}'\mathbf{X})^{-1}$. The trace is not scale invariant and is therefore rarely used as a design criterion.

In the logistic model situation, uncentered or centered, the design criteria D, F, Asy.CI, and Q are absolutely scale invariant to μ (shifts) and are relatively scale invariant to β (changes in scale). Absolute scale invariance to β is achieved by multiplying the expressions by a power of β. The need for relative invariance is obvious when choosing an optimal design, and the need for absolute invariance is clear when a criterion value is required over a range of parameter values as is the case in the minimax and Bayesian design procedures to be discussed in chapter 4.

With respect to scale invariance, the E criterion behaves differently than the others discussed. This criterion depends upon the eigenvalues of the inverted information matrix, and although the product of the eigenvalues have the invariance attribute, individually the eigenvalues do not. Using the uncentered model, the E criterion value depends on both μ and β. Therefore the centered model must be used, in which case the criterion displays absolute invariance to μ. However the criterion is still dependent upon β. This dependency is a result of the individual variance components being scale dependent. Scaling the inverted information matrix from the centered model with β corrects the problem.

For the logistic model the scale of the regressor variable x and the model parameter μ is given by $\frac{1}{\beta}$. Therefore the scaling problem is easily solved by multplying the expressions by the appropriate power of β. This scaling is illustrated in the following development.

Using the centered model (2.1.1), the Fisher information matrix is given by

$$I(\mu,\beta) = \begin{pmatrix} \beta^2 \sum_{i=1}^{k} n_i P_i(1-P_i) & -\beta \sum_{i=1}^{k} (x_i-\mu) n_i P_i(1-P_i) \\ -\beta \sum_{i=1}^{k} (x_i-\mu) n_i P_i(1-P_i) & \sum_{i=1}^{k} (x_i-\mu)^2 n_i P_i(1-P_i) \end{pmatrix}$$

and the inverse of the Fisher information matrix is then given by

$$I^{-1}(\mu,\beta)=\begin{pmatrix} V_{c11} & V_{c12} \\ V_{c12} & V_{c22} \end{pmatrix}$$

where V_{c11}, V_{c22}, and V_{c12} represent the asymptotic variances of the mle's of μ and β and their covariance, respectively. The scaled version of the inverted information matrix which yields a scale invariant E criterion is as follows:

$$I^{*-1}(\mu,\beta)=\begin{pmatrix} \beta^2 V_{c11} & V_{c12} \\ V_{c12} & \frac{1}{\beta^2} V_{c22} \end{pmatrix} \tag{3.1.1}$$

which implies that the scaled verion of the information matrix is

$$I^{*}(\mu,\beta) = \begin{pmatrix} \sum_{i=1}^{k} n_i P_i(1-P_i) & -\beta \sum_{i=1}^{k} (x_i-\mu) n_i P_i(1-P_i) \\ -\beta \sum_{i=1}^{k} (x_i-\mu) n_i P_i(1-P_i) & \beta^2 \sum_{i=1}^{k} (x_i-\mu)^2 n_i P_i(1-P_i) \end{pmatrix}.$$

It can be easily shown that the relationship among the determinants for the scaled and unscaled versions of the centered model information matrix and the uncentered model information matrix is

$$|I(\mu,\beta)| = |I^{*}(\mu,\beta)| = \beta^2 |I(\alpha,\beta)|$$

or

$$D_c = D_c^{*} = \beta^2 D.$$

Using the centered model with the scaled information matrix yields absolute scale invariance in the D, F, Asy.CI, Q, and E criterions. However, the centered model does not always lend itself to be easily applied to all criteria. The F and Q

criteria come more naturally from the uncentered model although they can be artificially derived from the centered model through the following equivalencies:

$$V_{c11} = \frac{1}{\beta^2}(V_{11}-2\mu V_{12}+\mu^2 V_{22}),$$

$$V_{c12} = \frac{1}{\beta}(V_{12}-\mu V_{22}),$$

and $$V_{c22} = V_{22}.$$

Both models will therefore be used in the following section with distinctions being made through the use of the c notation.

§3.2 Two Point D Optimality Design Criterion

The two point D optimal design for the logistic model where $\mu=ED_{50}$ was given by Kalish and Rosenberger (1978). They derived a very informative expression for the determinant of the Fisher information matrix for this model. In the derivation to follow, however, a more general model, one which includes ξ, will be used.

The determinant of the scaled information matrix is given by,

$$D_c = |I^*(\mu,\beta)|$$

$$= \Big(\sum_{i=1}^{2} n_i P_i(1-P_i)\Big)\Big(\beta^2 \sum_{i=1}^{2} (x_i-\mu)^2 n_i P_i(1-P_i)\Big) - \Big(-\beta \sum_{i=1}^{2} (x_i-\mu) n_i P_i(1-P_i)\Big)^2.$$

After expanding and simplifying, the determinant becomes

$$\beta^2\Big(n_1n_2(x_1-\mu)^2P_1(1-P_1)P_2(1-P_2) + n_1n_2(x_2-\mu)^2P_1(1-P_1)P_2(1-P_2)$$

$$- 2n_1n_2(x_1-\mu)(x_2-\mu)P_1(1-P_1)P_2(1-P_2)\Big).$$

Further simplification results in the final form,

$$\beta^2 n_1 n_2 P_1(1-P_1)P_2(1-P_2)(x_1-x_2)^2. \qquad (3.2.1)$$

The D optimal design is the design which maximizes 3.2.1. This expression has the property of absolute scale invariance.

An intuitive view of D optimality is brought forth by expression 3.2.1. The first part of the expression is comprised of the binomial variances of the random variables r_1 and r_2 at design point locations x_1 and x_2, respectively. To maximize this part, one needs to maximize the product of the two Bernoulli variances, implying that each individual variance should be maximized. The Bernoulli variance or slope is maximized when P=.50, or equivalently $x=ED_{50}$. Consequently, the first part of the expression exhibits a need to place the two design points at the ED_{50}. However, the second part of the expression counteracts the first part. The quadratic term $(x_1-x_2)^2$ suggests a need to spread the point out. The compromise is the D optimal design. In addition, if the distance between design points is held constant, the Bernoulli variances are maximized when they are equal. This occurs when the design points are placed symmetrically about the ED_{50}. Also, equal sample allocation is optimal since the product n_1n_2 is maximized when $n_1=n_2$.

Another interesting point concerning the D criterion is the comparison between 3.2.1 and its counterpart for the linear model. Apart from the two Bernoulli variances and β^2, expression 3.2.1 is identical to the D criterion expression one would obtain for the linear model restricted to one design variable and two design points. This result follows from the fact that the information matrix for the logistic model differs from the corresponding linear model information matrix only by the inclusion of the Bernoulli variances of the design points in each of the four terms. Consequently, for the linear model, the D optimal design consists of design points spread as far as is allowed by the design region. The inclusion of the Bernoulli variances in the logistic model situation prevents the design points from being spread too far or, in other words, into locations of near constant success or failure.

The D optimal design can easily be found numerically from expression 3.2.1 by making substitutions for x_1 and x_2 and minimizing with respect P_1 and P_2. Solving for the design point location in the assumed model,

$$x_i = \frac{\xi+\alpha+\text{Log}(\frac{P_i}{1-P_i})}{\beta}. \tag{3.2.2}$$

Substituting expression 3.2.2 for x_1 and x_2 in equation 3.2.1, we obtain after some simplification another form of the D optimality criterion expression:

$$n_1n_2P_1(1-P_1)P_2(1-P_2)\left\{\text{Log}(\frac{P_1}{1-P_1})-\text{Log}(\frac{P_2}{1-P_2})\right\}^2. \tag{3.2.3}$$

Maximizing expression 3.2.3 numerically with respect to P_1, P_2, and n_1 yields the values .176, .824, and $\frac{n}{2}$. Consequently, the D optimal design is equal sample allocation at $ED_{17.6}$ and $ED_{82.4}$. Notice that the absolute scale invariance property of the D criterion is apparent from expression 3.2.3.

§3.3 Two Point Asymptotic Confidence Interval Criterion for μ

The expression for the squared half length of an asymptotic confidence interval for μ for a two point design is derived for the logistic regression model. The $(1-\gamma)100\%$ asymptotic CI as given in Chapter 2 is

$$\frac{a}{b} \pm \frac{z}{\beta}(V_{11}-2\mu V_{12}+\mu^2V_{22})^{\frac{1}{2}}$$

where a and b are the mle's of α and β, respectively, z is the $100(1-\gamma)^{\text{th}}$ percentile from the standard normal distribution, μ and β are the unknown parameters, and V_{11}, V_{12}, and V_{22} are the asymptotic variances and covariance of a and b taken from the inverse of the Fisher information matrix given by 2.2.4. An equivalent confidence interval as derived from the centered model is given by

$$\frac{a}{b} \pm z(V_{c11})^{\frac{1}{2}}$$

Consequently, the absolute scale invariant squared half length of the asymptotic confidence interval apart from constants is given by,

$$\beta^2 V_{c11} = (V_{11}-2\mu V_{12}+\mu^2 V_{22})$$

which is equal to

$$\left(\frac{\sum_{i=1}^{2} x_i^2 n_i P_i(1-P_i)}{D}\right) - 2\mu\left(\frac{\sum_{i=1}^{2} x_i n_i P_i(1-P_i)}{D}\right) + \mu^2\left(\frac{\sum_{i=1}^{2} n_i P_i(1-P_i)}{D}\right)$$

where $D=|I(\alpha,\beta)|$. After expanding, collecting terms, and simplifing we obtain

$$\frac{n_1 P_1(1-P_1)}{D}(x_1-\mu)^2 + \frac{n_2 P_2(1-P_2)}{D}(x_2-\mu)^2.$$

Substituting the expression for D from 3.2.1, apart from β^2 since $D=|I(\alpha,\beta)|$, the final form becomes

$$\frac{(x_1-\mu)^2}{n_2 P_2(1-P_2)(x_1-x_2)^2} + \frac{(x_2-\mu)^2}{n_1 P_1(1-P_1)(x_1-x_2)^2}. \qquad (3.3.1)$$

The asymptotic CI optimal design is the design which minimizes expression 3.3.1. The expression consists of three main components: the squared distances of the design points from μ, the Bernoulli variances, and the squared distance between the two design points. Separately, the goals would be to place the design points at μ, maximize the Bernoulli variances (points at ED_{50}), and maximize the distance between the points. However, these goals are in conflict, and the design which minimizes expression 3.3.1 is not obvious. Numerical minimization of expression 3.3.1 yields a single point design located at μ for reasonable ED's of interest, ranging from ED_{04} to ED_{96}. For ED's of interest outside this range, numerical minimization results in a two point design at $ED_{8.3}$ and $ED_{91.7}$. This two point design is also the slope optimal design; the design which minimizes the variance of the estimate of β. The single point design does not allow for the estimation of the logistic model, but rather only μ.

§3.4 Two Point F Optimality Design Criterion

The $(1-\gamma)100\%$ Fieller interval for μ for the logistic model was derived in Chapter 2. In this section, an expression for the squared half length of the Fieller interval using the logistic regression model is derived for the special case of a two point design. Apart from constants, the absolute scale invariant squared half length from 2.3.2 is

$$\frac{1}{(1-g)^2}\left\{\left[V_{11} - 2\mu V_{12} + \mu^2 V_{22}\right] + g\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}.$$

The expression in brackets is the expression which is minimized for the asymptotic CI criterion. Consequently, result 3.3.1 is substituted and we obtain

$$\frac{1}{(1-g)^2}\left\{\frac{(x_1-\mu)^2}{n_2P_2(1-P_2)(x_1-x_2)^2} + \frac{(x_2-\mu)^2}{n_1P_1(1-P_1)(x_1-x_2)^2} + g\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}.$$

Replacing g by $\frac{z^2V_{22}}{b^2}$ in one part of the expression yields

$$\frac{1}{(1-g)^2}\left\{\frac{(x_1-\mu)^2}{n_2P_2(1-P_2)(x_1-x_2)^2} + \frac{(x_2-\mu)^2}{n_1P_1(1-P_1)(x_1-x_2)^2} + \frac{z^2}{b^2}\left((V_{12})^2 - V_{11}V_{22}\right)\right\}.$$

The determinant of $I(\alpha,\beta)^{-1}=V_{11}V_{22}-(V_{12})^2=\frac{1}{|I(\alpha,\beta)|}=\frac{1}{D}$; therefore, the expression can be written as

$$\frac{1}{(1-g)^2}\left\{\frac{(x_1-\mu)^2}{n_2P_2(1-P_2)(x_1-x_2)^2} + \frac{(x_2-\mu)^2}{n_1P_1(1-P_1)(x_1-x_2)^2} - \frac{z^2}{b^2D}\right\}.$$

Substituting the expression for D obtained from 3.2.1, apart from β^2, and letting $Q_i=(1-P_i)$, the final form becomes

$$\frac{1}{(1-g)^2}\left\{\frac{(x_1-\mu)^2}{n_2P_2Q_2(x_1-x_2)^2} + \frac{(x_2-\mu)^2}{n_1P_1Q_1(x_1-x_2)^2} - \frac{z^2}{b^2n_1n_2P_1Q_1P_2Q_2(x_1-x_2)^2}\right\}. \qquad (3.4.1)$$

The design which minimizes expression 3.4.1 is the F optimal design. In general, for a given ED of interest, the F optimal design varies with sample size and level of confidence. However if the ratio z^2/n is held constant, where n is the total sample size and z is a standard normal deviate, the F interval and corresponding F optimal design is invariant to changes in n and z. This result is shown in appendix C. In a later section, an equivalence is formed between the F criterion which depends upon z^2/n and ξ and the Q criterion which depends upon region.

Expression 3.4.1 contains both the asymptotic CI and D optimality expressions. With respect to minimization, these two expressions as they appear in 3.4.1 are in agreement: a single point design at μ for reasonable ED's of interest. However, the two expressions are contrasted by g, which measures the significance of the slope β. Numerical minimization of expression 3.4.1 indicates that as the ratio z^2/n increases, g approaches 1, and the design converges to the slope or β optimal design which is $ED_{8.3}$ and $ED_{91.7}$. As g nears 1, it and the minimization of the slope variance begins to dominates expression 3.4.1. Conversely, numerical minimization indicates that as the ratio z^2/n decreases, g approaches 0, and the design converges to the μ optimal design, all points at μ for reasonable ED's of interest.

The scenarios of g near 0 or 1 represent extremes. Consequently, the F optimal design is a compromise between the bounds formed by the μ optimal and slope optimal designs. The F optimal design depends upon g, yet g is proportional to the ratio z^2/n. Therefore the F optimal design is a function of ξ, the ED of interest, and z^2/n.

§3.5 Two Point Q Optimality Design Criterion

The Q optimality criterion expression derived by Myers, Myers, and Carter (1994) for the two parameter one variable logistic model is derived in this section in a form that allows the reader to gain some insight into how it compares with the other criteria. The Q optimal design is based upon a region of interest instead of a parameter of interest. Let R_L and R_U represent the boundaries of this region

of interest over the logistic curve in the natural units of x. If a region is specified in logit space, Z, let $R_j=\frac{Z_j+\alpha+\xi}{\beta}$. The absolute scale invariant expression for the average prediction variance of logit($\hat{P}$) over the region of interest is

$$\frac{1}{K}\int_{R_L}^{R_U}(V_{11} - 2x_0V_{12} + x_0^2V_{22})\,dx_0$$

where $K=R_U-R_L$, the volume of the region. After integrating, the expression becomes

$$\frac{1}{(R_U-R_L)}\left((R_U-R_L)V_{11} - (R_U^2-R_L^2)V_{12} + (R_U^3 - R_L^3)\frac{V_{22}}{3}\right). \qquad (3.5.1)$$

For a two point design, 3.5.1 can be written as

$$\frac{1}{D}\Big(n_1P_1(1-P_1)\big(x_1^2-x_1(R_U+R_L)+\tfrac{1}{3}(R_U^2+R_UR_L+R_L^2)\big) +$$

$$n_2P_2(1-P_2)\big(x_2^2-x_2(R_U+R_L)+\tfrac{1}{3}(R_U^2+R_UR_L+R_L^2)\big)\Big)$$

which simplifies to give the form,

$$\left\{\frac{(x_1-\overline{R})^2}{n_2P_2Q_2(x_1-x_2)^2} + \frac{(x_2-\overline{R})^2}{n_1P_1Q_1(x_1-x_2)^2} + \frac{(R_U-R_L)^2}{12n_2P_2Q_2(x_1-x_2)^2} + \frac{(R_U-R_L)^2}{12n_1P_1Q_1(x_1-x_2)^2}\right\} \qquad (3.5.2)$$

where $Q_i=1-P_i$ and $\overline{R}=\frac{1}{2}(R_U+R_L)$. Substituting V_{22} into expression 3.5.2 yields

$$\left\{\frac{(x_1-\overline{R})^2}{n_2P_2Q_2(x_1-x_2)^2} + \frac{(x_2-\overline{R})^2}{n_1P_1Q_1(x_1-x_2)^2} + \frac{(R_U-R_L)^2}{12}V_{22}\right\}. \qquad (3.5.3)$$

The design which minimizes expression 3.5.3 is the Q optimal design. The expression is composed of three main components. From section 3.3 it is apparent that a single point design at the center of the region of interest minimizes the first two components. However the third component, variance of b, which is weighted by the variance of a uniform random variable, is of course minimized by the slope optimal design, $ED_{8.3}$ and $ED_{91.7}$. The variance from a uniform comes from integrating region moments uniformly over the region of interest. For a single point region at μ, the right hand term of expression 3.5.3 is zero and the Q optimal design is the asymptotic confidence interval (μ optimal) design. If on the other hand, the region of interest is infinite in length, the right hand term being weighted by the squared width of the region dominates the first two components which are unaffected by region width. The resulting Q optimal design is the asymptotic confidence interval (β optimal) design. Both of these results are easily shown numerically. The Q optimal design is a compromise between the single point design at $\overline{R}$ and the slope optimal design based upon the region of interest both in terms of location, denoted by $\overline{R}$, and width.

§3.6 An Emperical Q, F Relationship

Both the Q and F optimality expressions represent compomises between the slope optimal and μ optimal designs. For a given ED of interest the degree of compromise for the F expression is driven by the ratio n/z^2 whereas the size of the region of interest drives the compromise for the Q criterion. Since the F and Q expressions have this relationship, situations should exist in which the the two criteria yield similar designs. In this section we give equations which approximate the region width and location which yields a Q optimal design that is equivalent to an F optimal design with a particular value of n/z^2 and ED of interest. The main insight to be gained from such equations is to help understand how an F optimal design relates to a region of interest as well as an ED of interest.

The equations only offer approximations of the relationships. Closed form expressions for the equivalences could not be found due to the complexities of the logistic model and the criteria. With no other foreseable option, a trial and error method was employed. Region width and location (in logit units) was adjusted

until the Q optimal two point design (P_1, P_2, and Pn, where Pn is the proportion of the total sample taken at P_1) equaled the F optimal design for a particular n/z^2 and ED of interest. This procedure was employed for over 30 different ξ and n/z^2 combinations. At that point, trial and error as well as Proc NLIN (SAS, Statistical Analysis System) was used to find equations which fit the points.

The fitted equation for region width is given by

$$R_W = \frac{5.3813 + .6900\xi^2}{(n/z^2)^{.456} - 1.234} \tag{3.6.1}$$

where R_W is the region of interest corresponding to ξ, the ED of interest, both of which are expressed in logit units. The equation for the region location is then denoted by region center and is

$$R_C = \frac{(n/z^2)\xi}{\frac{5.862}{z^2}\sqrt[3]{|\xi|} - (n/z^2)}$$

$$= \frac{n\xi}{5.862\sqrt[3]{|\xi|} - n} \tag{3.6.2}$$

where R_C is the corresponding region center again expressed in terms of logit units.

The approximations work quite well for $n/z^2 > 6.0$. This requirement encompasses most reasonable situations encountered. Some examples are given below to illustrate the degree of approximation of 3.6.1 and 3.6.2.

n=200, z=1.5		P_1	P_2	Pn
$\mu=ED_{40}$	F optimal design	.3141	.6852	.733
$\Rightarrow R_W=.8447$	Q optimal design	.3142	.6884	.736
$R_C=-.4145$	$\Rightarrow$ equivalent region of interest ED_{30} to ED_{50}			

n=90, z=2.1		P_1	P_2	Pn
$\mu=ED_{15}$	F optimal design	.1405	.8594	.827
$\Rightarrow R_W=2.740$	Q optimal design	.1409	.8594	.826
$R_C=-1.882$	$\Rightarrow$ equivalent region of interest ED_4 to ED_{38}			

n=60, z=2.25		P_1	P_2	Pn
$\mu=ED_{50}$	F optimal design	.2300	.7701	.500
$\Rightarrow R_W=2.892$	Q optimal design	.2305	.7693	.500
$R_C=0.000$	$\Rightarrow$ equivalent region of interest ED_{19} to ED_{81}			

n=600, z=2.6		P_1	P_2	Pn
$\mu=ED_8$	F optimal design	.1041	.8958	.896
$\Rightarrow R_W=1.461$	Q optimal design	.1053	.8945	.896
$R_C=-2.472$	$\Rightarrow$ equivalent region of interest ED_4 to ED_{15}			

§3.7 E Optimality Design Criterion

For reasons of scaling, as mentioned earlier, the E optimality criterion can only be derived using matrix 3.1.1 or the scaled information matrix itself from the centered model. The scaled, inverted information matrix from the centered model, matrix 3.1.1, is used for the derivation of the E optimality expression.

The characteristic equation for matrix 3.1.1 is given by

$$\lambda^2 - \lambda(\beta^2 V_{c11} + \frac{V_{c22}}{\beta^2}) + V_{c11}V_{c22} - V_{c12}^2 = 0. \qquad (3.7.1)$$

The eigenvalues of matrix 3.1.1 are then given by the roots of equation 3.7.1 which are

$$\lambda_1 = \frac{1}{2}\left[(\beta^2 V_{c11} + \frac{V_{c22}}{\beta^2}) + \left((\beta^2 V_{c11} - \frac{V_{c22}}{\beta^2})^2 + 4(V_{c12})^2\right)^{\frac{1}{2}}\right]$$

and $$\lambda_2 = \frac{1}{2}\left[(\beta^2 V_{c11} + \frac{V_{c22}}{\beta^2}) - \left((\beta^2 V_{c11} - \frac{V_{c22}}{\beta^2})^2 + 4(V_{c12})^2\right)^{\frac{1}{2}}\right].$$

Clearly λ_1 is the largest eigenvalue of matrix 3.1.1. Therefore the E optimal design is the design which minimizes λ_1. The E optimal design is determined by

$$\underset{\mathfrak{D}}{\text{Min}} \left[(\beta^2 V_{c11} + \frac{V_{c22}}{\beta^2}) + \left((\beta^2 V_{c11} - \frac{V_{c22}}{\beta^2})^2 + 4(V_{c12})^2\right)^{\frac{1}{2}}\right] \qquad (3.7.2)$$

which is the absolute scale invariant E optimality expression.

No useful or intuitive simplification into elementary components could be found in expression 3.6.2. However, a nice interpretation exists if one considers the prediction variance of Logit($\hat{p}$). Minimizing the maximum eigenvalue of 3.1.1 can be shown to be equivalent to minimizing the maximum prediction variance of Logit($\hat{p}$) on spheres of varying radii centered about μ. This result is shown in the appendix D. Here we only have one explanatory variable resulting in only two prediction variances per sphere. Consequently, the E optimal design for the logistic model minimizes the maximum prediction variance of Logit($\hat{p}$) at either location μ+d or μ-d for all d>0. This result implies that the E optimal design should depend upon the ED associated with μ.

CHAPTER 4

DESIGN CRITERIA SUPPLEMENTS

The design criteria discussed in the previous chapter yield optimal designs which typically consist of only two design points. In chapter 6, these designs are shown to be quite sensitive to poor parameter knowledge. For this reason, these criteria and the designs they produce are not very useful in practice by themselves. Consequently, procedures have been developed to enhance the performance of these criteria. Two of these procedures are discussed here: the minimax procedure by Sitter (1992) and the Bayesian design procedure by Chaloner and Larntz (1989).

§4.1 Minimax Procedure and Critique

The minimax procedure chooses a design based upon the knowledge the experimenter has about the parameters. The experimenter not only makes direct guesses, μ_0 and β_0, of the parameters μ and β, but also creates an interval rectangle in which he or she believes the true parameters reside: $\beta_L<\beta<\beta_U$ and $\mu_L<\mu<\mu_U$. This region is assumed to be symmetric about the initial parameter guesses μ_0 and β_0. The minimax procedure then chooses, for a fixed sample size and criterion, the design, symmetric about μ_0, which minimizes the maximum criterion value over the rectangle. Most any reasonable criterion could be used, such as D, Q, E, F, or Asymptotic CI optimality. In general, Sitter defines the minimax design as follows:

$$R(\delta^*) = \min_{\delta \in \mathfrak{D}} \max_{(\mu,\beta) \in \mathfrak{P}} R(\delta, \mu, \beta) \qquad (4.1.1)$$

where R represents any reasonable criterion defined such that the worst case scenario is when the criterion is maximum. $\mathfrak{D}$ represents the set of designs for a fixed sample size which are symmetric about μ_0, $\mathfrak{P}=\{(\mu,\beta)\colon \beta_L<\beta<\beta_U,\ \mu_L<\mu<\mu_U\}$, and $\delta^*$is the minimax design.

Sitter discovered that the set of designs $\mathfrak{D}$ was too large to find δ^* in a practical manner. The Nelder-Mead optimization algorithm (1965) was being used and could not find the optimal number of points in addition to the optimal location and sample allocation to those points. The dimensionality of the problem was too high. Consequently, $\mathfrak{D}$ was restricted to "kk" designs which allocate the sample equally to design point locations which are equally spaced.

Many minimax designs are given by Sitter for the D, F, and asymptotic CI criteria for a large variety of *a priori* parameter knowledge. The tables of designs are indexed by $\frac{\beta_U}{\beta_L}$ and $\mu_\Delta=\beta_L(\mu_U-\mu_0)=\beta_L(\mu_0-\mu_L)$ since the minimax design, which is specified by optimal values of k=number of design points and d=scale-free distance between design points, can be determined independently of μ_0, β_0, and β_L. This result is shown in the appendix E.

An F optimality minimax design given by Sitter is used to critique the procedure. The particular minimax design comes from $\mu_\Delta=4.0$, $\frac{\beta_U}{\beta_L}=2.0$, which yields d=4.3, k=5. Arbititrary values for the center of the rectangle are choosen to be $\mu_0=40$ and $\beta_0=.90$, resulting in an interval rectangle bounded by μ:(34, 46) and β:(.6, 1.2). The longest Fieller interval occurs when (μ, β) equal (46, .6) and (34, .6), on the edge of the rectangle.

The procedure assumes each possible parameter pair, (μ,β), is equally likely to occur, so that pairs on the edge of the rectangle are considered as likely to occur as μ_0,β_0. The maximum criterion value as one would expect is found at or near the edges of the rectangle. When the ranges of the parameter are reasonably small, this is not a problem. The efficiencies at and around the center of the rectangle are quite high. However when the ranges are fairly large as in the example choosen, one minimizes the maximum criterion value at a location far away from the best single guess of μ_0,β_0. The minimax designs sacrifice too much efficiency at and near μ_0,β_0, in order to be able to be efficient at a location which

is quite possibly one of the most unlikely locations of the true parameter values within the interval rectangle. At the location with the longest F interval, the F efficiency is 60% while the efficiency at the center is only 50%, where F efficiency is defined by

$$\text{F\_EFF} = \frac{\text{F interval length for optimal design given } \mu,\beta}{\text{F interval length for minimax design centered at } \mu_0 \text{ given } \mu,\beta}.$$

Another problem exists with the minimax F optimal designs given by Sitter. The Fieller interval length which was used is not invariant to scale. The scale-free length should be used so that the location with the worst F efficiency would drive the procedure. Intuitively, it is easy to see that the range in x space where .10<P<.90, decreases as β increases. Therefore, the interesting range of the response curve is narrower when the slope is 1.20 than when it is .6, and one would expect an equally efficient interval about μ in the first case to be considerably more narrow than in the latter. In fact, an equally efficient scale dependent F interval on μ when β=1.2 is exactly half as long as when β=.6. The length of an Fieller interval is of order $\frac{1}{\beta}$. The minimax procedure should minimize the maximum scale invariant F interval length as given in 2.4.1 assuming F is the criterion being used.

Even when using the scale invariant F criterion, the minimax design still offers an unsatisfactory solution to the problem of robustness. The motivation and idea behind the minimax procedure is of great practical interest even though the procedure itself is not. The minimax procedure is simply too conservative.

§4.2 Bayesian Design Procedure

In another effort aimed at creating optimal robust designs, a Bayesian procedure was developed by Chaloner and Larntz (1989) which utilizes the experimenter's apriori belief about the parameters in the form of a density. The procedure finds optimal Bayesian designs based upon the experimenter's belief

about the parameters. The prior distributions were not used for making inferences due to the complexity of the problem. As with the minimax procedure, most reasonable design criteria can be used with the Bayesian procedure.

In general the Bayesian design procedure is given by the following expression

$$\underset{\mathfrak{D}}{\text{Max}} \int_{\boldsymbol{\theta}} \text{R}(\delta,\boldsymbol{\theta})\ \pi(\boldsymbol{\theta})\ \text{d}\boldsymbol{\theta} \tag{4.2.1}$$

where $\boldsymbol{\theta}$ is a vector of model parameters, $\pi(\boldsymbol{\theta})$ is the prior density of $\boldsymbol{\theta}$, $\text{R}(\delta,\boldsymbol{\theta})$ is any design optimality criterion expression of choice, and δ is any design from the set of possible designs $\mathfrak{D}$. Expression 4.2.1 comes from Bayesian decision theory where $-\text{R}(\delta,\boldsymbol{\theta})$ represents the Bayes risk. The Bayesian optimal design minimizes the expected risk. Although expression 4.2.1 is analogous to Bayesian estimation, it is quite different. Unlike Bayesian estimation, $-\text{R}(\delta,\boldsymbol{\theta})$ does not represent an expected loss function. The decision, δ, to be made is therefore not an estimator but rather a design.

For the one variable logistic model with two parameters the experimenter would need to select a criterion and a prior $\pi(\alpha,\beta)$ or $\pi(\mu,\beta)$ which reflects his or her knowledge about the parameters. The expression for the Bayesian design procedure in this situation is given by

$$\underset{\mathfrak{D}}{\text{Max}} \int_{(\mu,\beta)} \text{R}(\delta,\mu,\beta)\ \pi(\mu,\beta)\ \text{d}(\mu,\beta). \tag{4.2.2}$$

Chaloner and Larntz give several examples of these one-stage designs for the logistic model for a number of different priors and two different criteria. In general, they found that as the variance of the prior increases, so does the dispersion of the optimal Bayesian design and the number of required design levels. The additional design points and width allow the Bayesian design to cope with poor parameter knowledge. Essentially, the Bayesian procedure would be as robust as the experimenter believes he or she needs depending on the prior knowledge of the parameters.

This procedure addresses the same problem as the minimax procedure without necessarily being so conservative. The Bayesian design procedure is incorporated throughout this dissertation and is shown to be a useful addition to the standard design criteria. The typical point guesses normally associated with logistic model designs are represented by the product of two single point densities. These standard single point guesses will be referred to as a point prior. Any other degree of parameter knowledge is represented by the class of normal or uniform densities, although many other densities could be used. The one-stage designs given in the next chapter make use of the Bayesian procedure as it is seen in 4.2.2, yielding one-stage optimal Bayesian designs.

Although the Bayesian design procedure improves the usefulness of design criteria in a one-stage design, the robustness and efficiency is shown in chapter 6 to still be relatively poor. It is in the two-stage designs of chapters 7, 8, and 9 when the true usefulness of the prior is realized. These designs make use of the prior twice: first as is done in the one-stage designs and secondly to yield Bayesian estimates of the parameters from the first stage optimal Bayesian design.

CHAPTER 5

ONE-STAGE OPTIMAL DESIGNS FOR THE LOGISTIC MODEL

Several design optimality criteria have been presented for the logistic regression model in the one variable case. In this chapter we present the optimal designs these criteria produce. Extensive lists of F, E, and Q optimal designs are given for a wide variety of situations in which a point prior is used. The μ and slope optimal designs do not require extensive lists, for they vary one dimensionally if at all when using a point prior. The μ optimal design is achieved by placing all n observations at the ED of interest for ED's ranging from ED_{04} to ED_{96}, and the slope optimal design is achieved by placing $\frac{n}{2}$ observations at $ED_{8.3}$ and $\frac{n}{2}$ observations at $ED_{91.7}$. The effect of different priors is demonstrated using the D criterion. Many Bayesian D optimal designs are given. All designs are found numerically using the Nelder-Mead algorithm (Nelder and Mead, 1965).

§5.1 F Optimal Designs, Point Prior

The F optimal design consists of two points and depends upon ξ, n, and z. Three different one-stage design scenarios are considered using an F criterion. The first situation results from finding the F optimal design with respect to the number of design points and the location and sample allocation to those design points. This first case yields the unconstrained F optimal designs.

Secondly, consider a two point design in which the sample allocation is held equal. This situation yields constrained F optimal designs. Then consider designs with more than two levels. Many four and five level designs are given which are not F optimal, but are highly F efficient.

Using the Nelder-Mead algorithm and no restrictions other than a fixed sample size and ED of interest, we found that the F optimal design with a point prior is in fact a two level design. The required sample allocation, however, is not necessarily equal nor are the points necessarily symmetric. Although, if μ=ED_{50}, the sample allocation is equal and the points symmetric.

Sample size has a profound effect on the F optimal design. In fact, the optimal design changes complexion with sample size. For sample sizes of 500 or less and $\mu \neq ED_{50}$ the complexion of the design is a majority of the sample at the design point closest to μ and the rest at the symmetric counterpart. For very large samples, the optimal design complexion changes to equal allocation and symmetry about the ED of interest.

Several F optimal designs are given in table 5.1.1 using a 95% Fieller interval. In this dissertation, a 95% F interval is used for all designs using the F criterion. The Fieller interval depends on the ratio, n/z^2, and so the squared reciprical of z has the same effect as n. Consequently, studying the effect of n is equivalent to studying z. Additionally, only designs for ED_{100p}, $p \leq .50$ are given explicitly since the logistic function is symmetric.

Table 5.1.1. F optimal designs, point prior, μ=ED_{100p}

p	n	n_1	n_2	P_1	P_2
.50	10	5	5	.10	.90
.50	20	10	10	.17	.83
.50	40	20	20	.22	.78
.50	100	50	50	.28	.72
.50	150	75	75	.30	.70
.50	1000	500	500	.375	.625

Table 5.1.1. F optimal designs, point prior, μ=ED_{100p}, continued

p	n	n_1	n_2	P_1	P_2
.40	20	12	08	.16	.84
.40	40	25	15	.21	.79
.40	100	66	34	.26	.74
.40	150	102	48	.28	.72
.40	1000	500	500	.28	.52
.30	20	13	07	.14	.86
.30	40	28	12	.19	.81
.30	100	77	23	.23	.77
.30	150	121	29	.24	.76
.30	1000	501	499	.19	.41
.20	20	14	06	.12	.88
.20	40	30	10	.15	.85
.20	100	83	17	.17	.83
.20	150	129	21	.18	.82
.20	1000	944	56	.19	.81
.10	20	13	07	.10	.90
.10	40	30	10	.11	.89
.10	100	84	16	.12	.88
.10	150	130	20	.12	.88
.10	1000	943	57	.11	.89

Four plots have been created to give a graphical representation of the F optimal designs from table 5.1.1 for a continous array of sample sizes ranging from 15 to 100. The plots rely on the knowledge that P_2=1-P_1. The optimal value of P_1 and the ratio n_1/n is plotted for each n. The plots, which are in figures 5.1.1-5.1.4, show that the F optimal designs are affected relatively little by changes in the ED of interest. Once one moves away from μ=ED_{50} to a region from p=.10 to .30, the F optimal designs are quite similar for the sample range, n=15 to 100. This point relates back to section 3.6, where it was stated that F optimality addresses a region of interest about the ED of interest.

Now consider the situation in which the sample allocation of the two point designs is held equal. The Nelder-Mead algorithm was employed to find the optimal locations of the two design points under such a constraint. Many of these constrained F optimal designs are given in table 5.1.2 for a few ED of interests as well as a number of sample sizes. When μ=ED_{50}, the constrained F optimal designs are equivalent to the F optimal designs. The designs, though different from those of table 5.1.1, exhibit the same complexion change, but at a much smaller sample size. The dramatic change of the optimal design appears at first to be rather abrupt. However, further investigation revealed that as sample size approaches the point of the change, two different locations of near optimality arise. At a particular infinitesimal sample size increment the optimum moves from one location to the other. At one location, the design is given by points spread symmetrically far apart about the ED_{50}, and at the other location the design is given by both points symmetrically close μ.

The reason for such a change in complexion is linked back to the discussion of the F criterion in section 3.4. Two forces are at odds, that being the efficient estimation of μ and β. The primary goal is of course the efficient estimation of μ which pulls the F optimal design towards allocation of all points at μ. The secondary goal concerns g, which measures the significance of the slope. However, in order to achieve the primary goal, the secondary goal must be satisfactorily met first, thus the trade-off. The change in the design complexion comes about when the secondary goal is met and the emphasis moves to the primary goal.

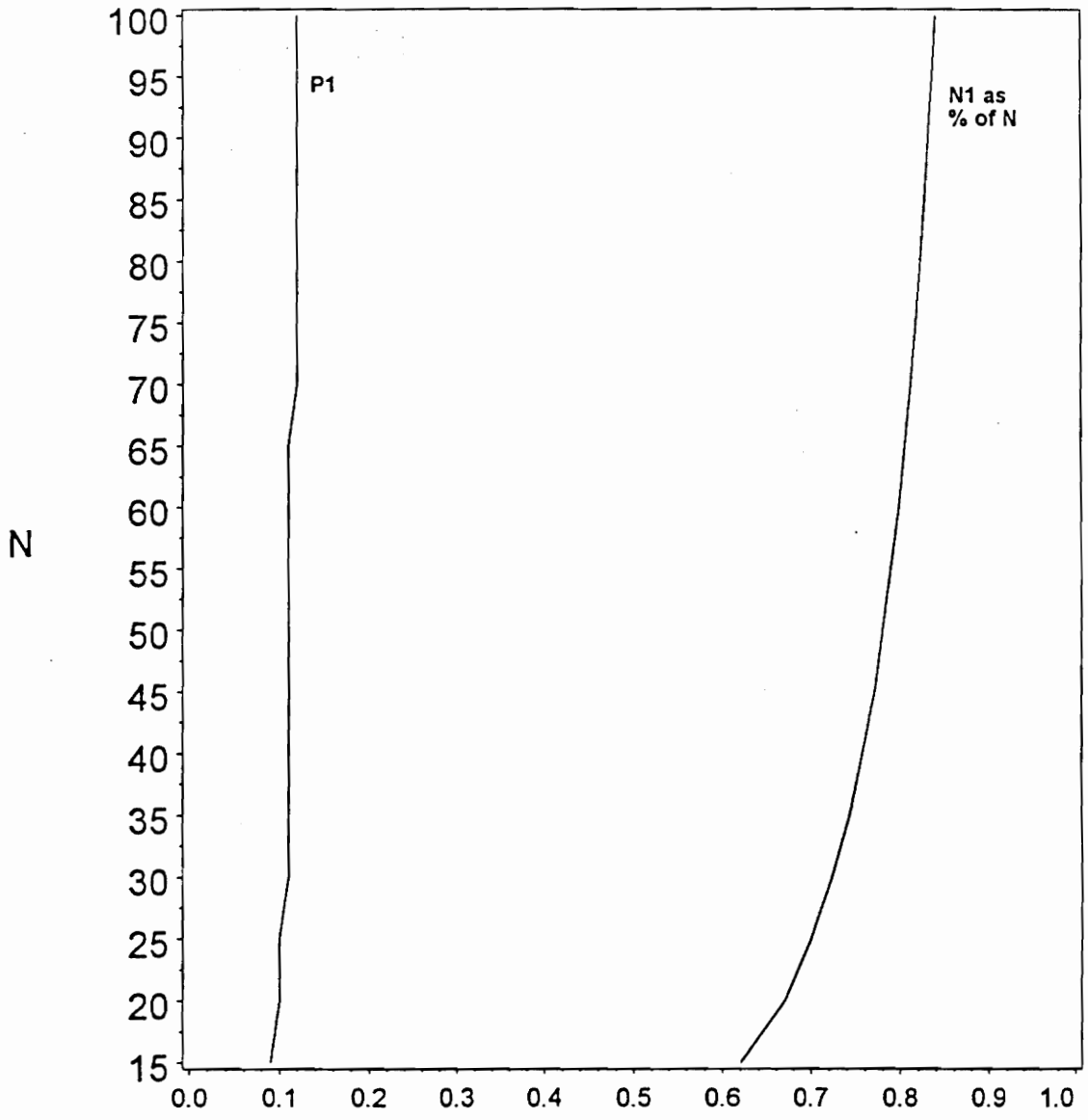

Figure 5.1.1 F optimal designs as a function of N, when μ=ED_{10}

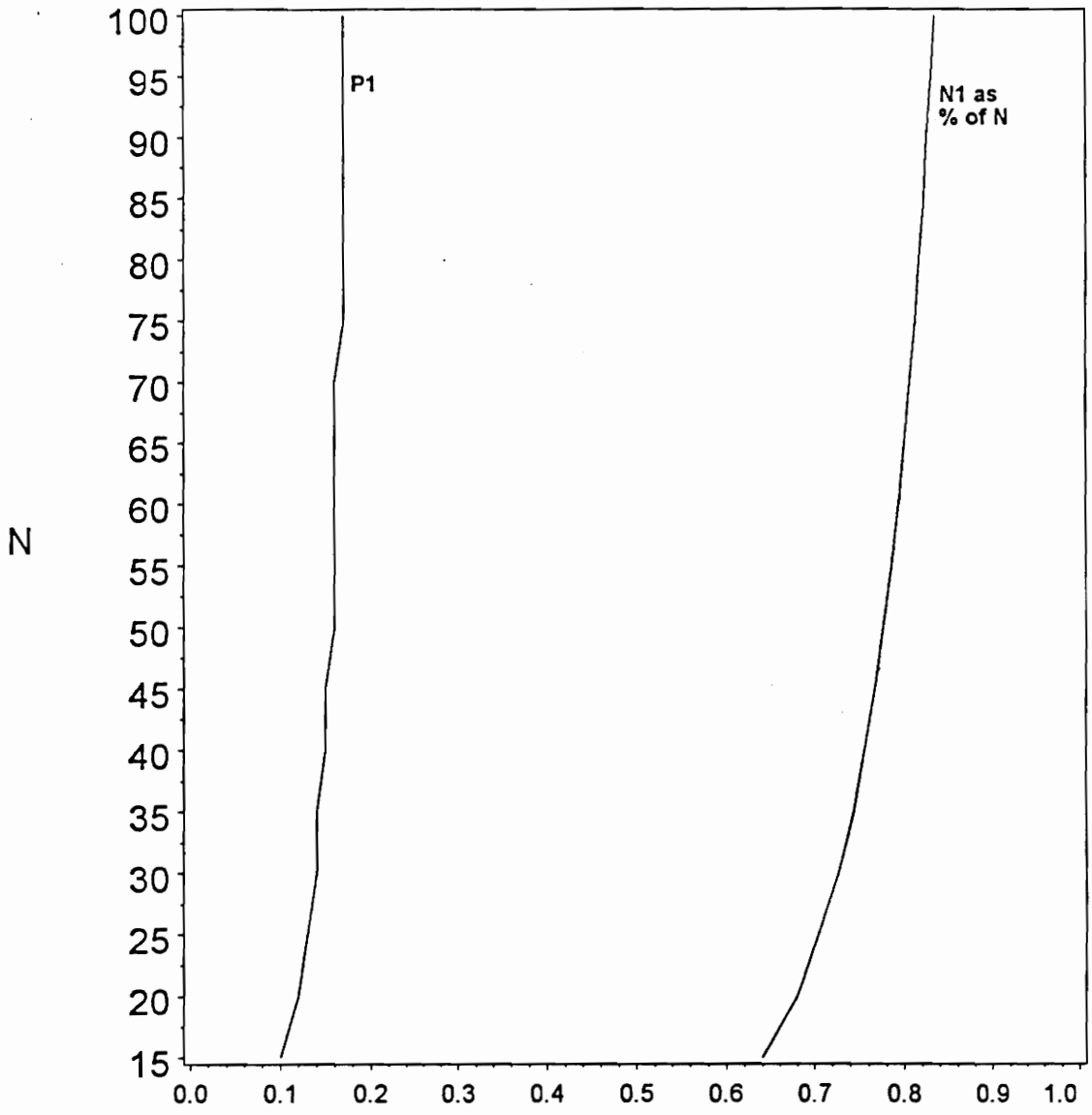

Figure 5.1.2 F optimal designs as a function of N, when μ=ED_{20}

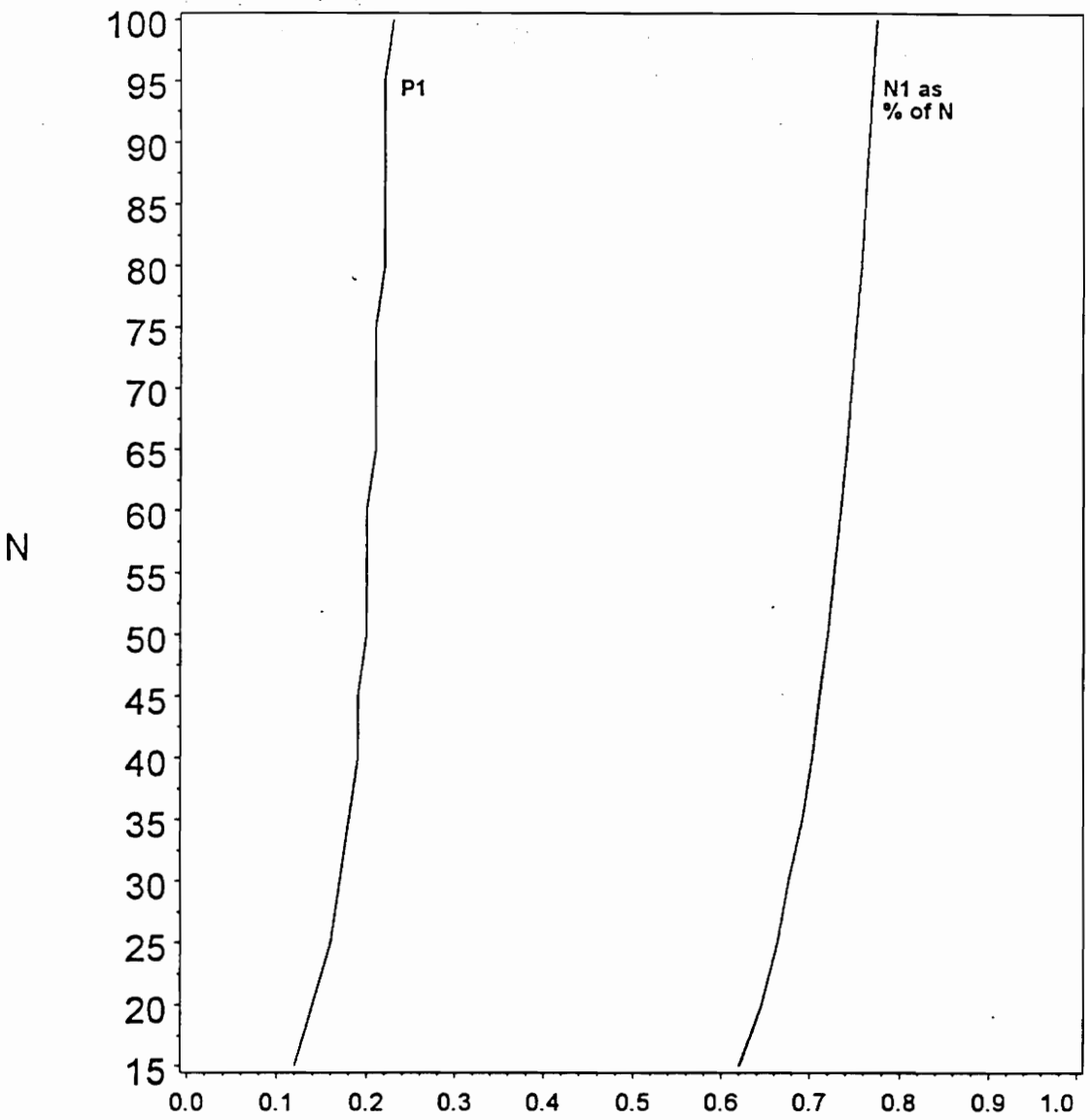

Figure 5.1.3 F optimal designs as a function of N, when μ=ED_{30}

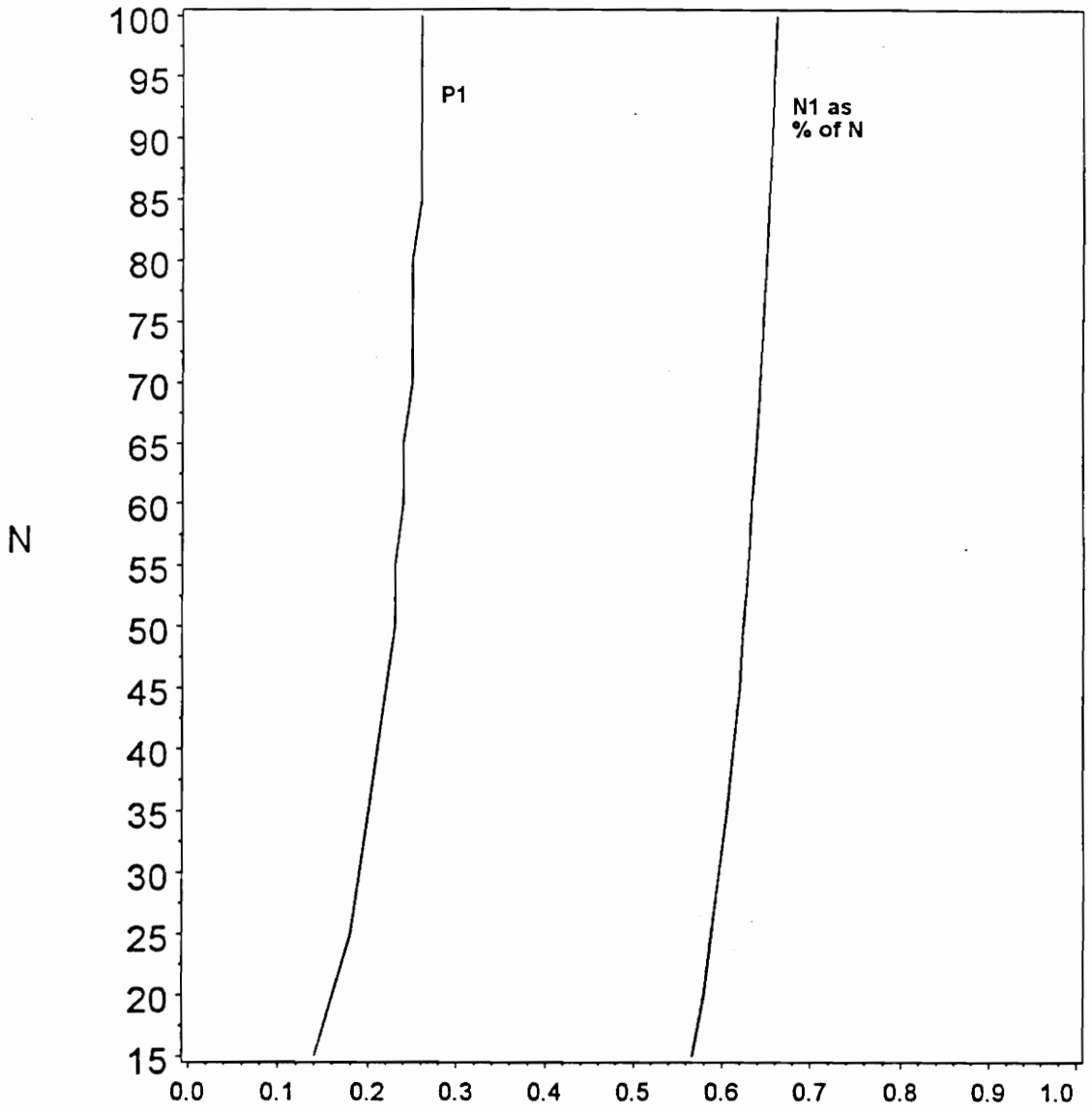

Figure 5.1.4 F optimal designs as a function of N, when μ=ED_{40}

Table 5.1.2 Contrained F optimal designs, point prior, μ=ED_{100p}

p	n	n_1	n_2	P_1	P_2
.33	10	5	5	.10	.92
.33	20	10	10	.165	.89
.33	40	20	20	.16	.77
.33	100	50	50	.14	.53
.33	200	100	100	.16	.50
.20	10	5	5	.09	.92
.20	20	10	10	.14	.92
.20	40	20	20	.18	.93
.20	100	50	50	.07	.37
.20	200	100	100	.08	.32

Plots are given in figures 5.1.5 and 5.1.6 which accompany table 5.1.2. These figures offer a graphical representation of the constrained F optimal designs. The optimal P_1 and P_2 values are plotted as a function of sample size. Figure 5.1.6 shows the change in complexion of the optimal design to be quite dramatic. Again, this abrupt change is a bit misleading. As the sample size approaches the location of the change, a near optimal design arises which is drastically different from the optimal design. At the point of change, the roles of these two designs reverses.

The F optimal design and constrained F optimal design behave much in the same way. Together they represent and exhibit the characteristics of the F optimality design criterion. The value of such two level designs in practice is easily questioned, but the usefulness of them in gaining knowledge of a design optimality criterion is invaluable.

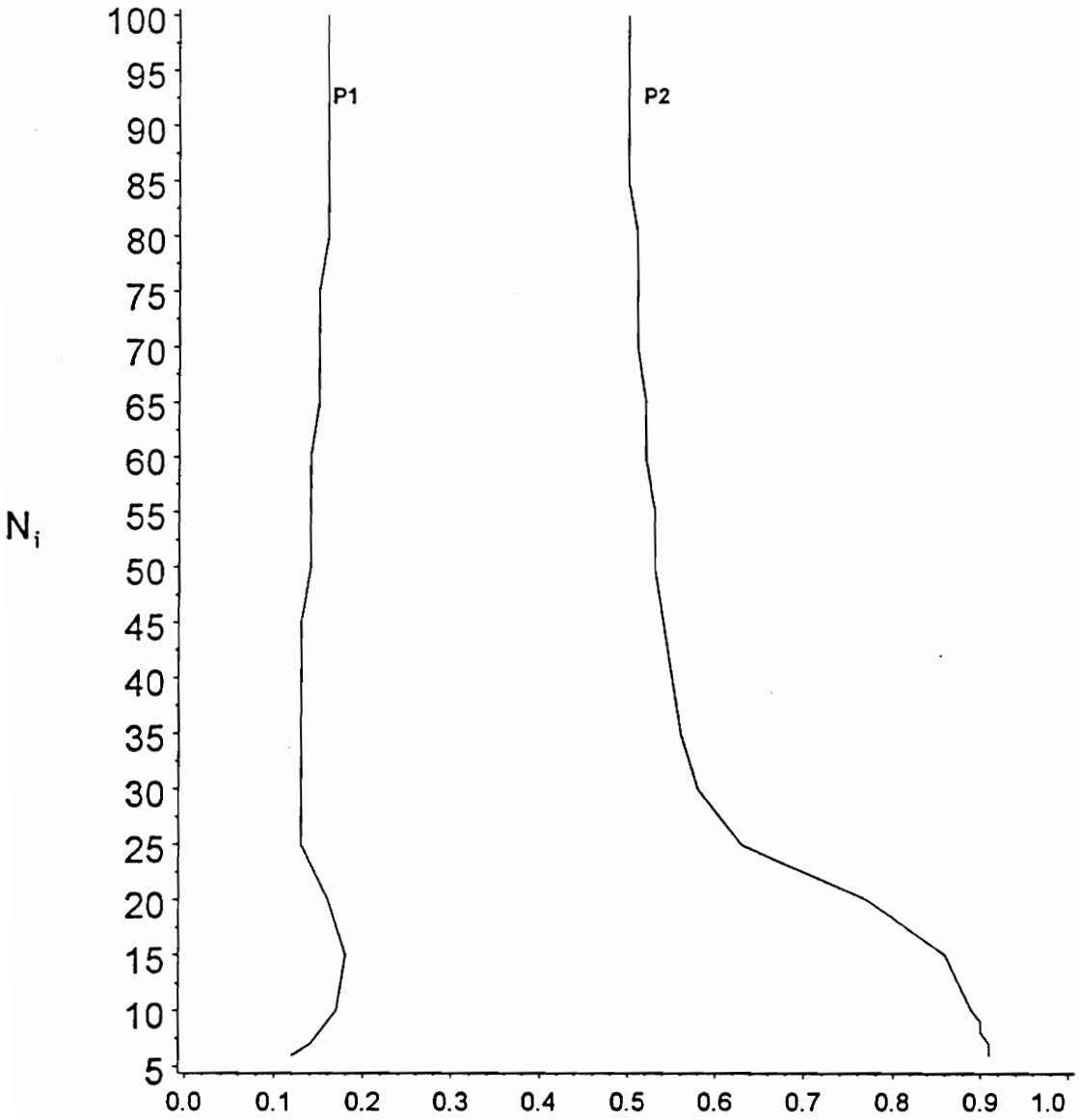

Figure 5.1.5 Constrained F optimal designs as a function of N_i when μ=$ED_{33.3}$

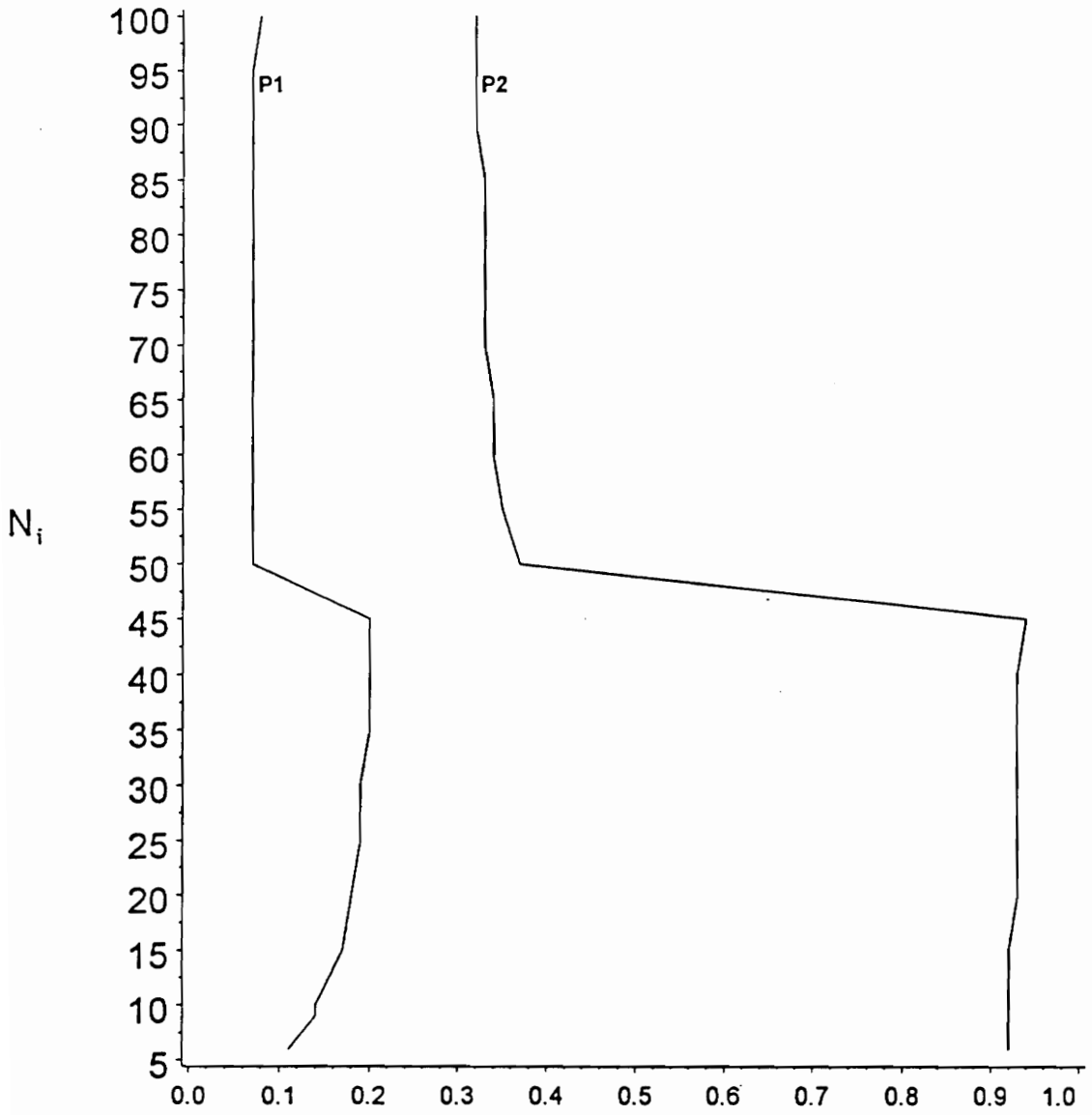

Figure 5.1.6 Constrained F optimal designs as a function of N_i when μ=ED_{20}

Many two level designs have thus far been given. These designs, however, are typically not considered very practical. The two level designs are too dependent on the assumptions of correct model and good parameter knowledge. What is needed in practice are designs which are nearly optimal when the model and parameter guesses are correct but are robust to poor guesses. Additionally, these designs should allow for a check of model adequacy.

This need for more practical one-stage designs leads to the use of more design levels. Attempts to construct F optimal four and five level designs using constraints to force more than two levels failed using the Nelder-Mead algorithm, due to dimensionality and the number of constraints needed. Other methods, however, were employed with success. Two reasonable levels were chosen about the ED of interest, and the algorithm found the best two additional levels to complement the first. This method led to an intuitive feel for creating highly efficient four and five level designs which outperformed those created with the intervention of the algorithm. Many four and five level designs using point priors were found which are nearly optimal and are given in tables 5.1.3 and 5.1.4.

An important point to notice is that the complexion of these designs resembles that of the optimal two level design: place a majority of the sample close to the ED of interest for ED's not close to the ED_{50} and the rest near the symmetric counterpart. The four and five point designs simply allow that majority to be spread out to essentially cover a region about the ED of interest. The remote point far from the region of interest is a necessity for slope estimate stability for the four and five level designs as it is for the optimal two point design.

The designs given in tables 5.1.3 and 5.1.4 in no way represent the only "good" one-stage four or five point designs with respect to the F optimality criterion. However, the designs which are given are nearly optimal and offer a nice alternative to the optimal two point designs. The F efficiencies are defined by

$$\text{F\_EFF} = \frac{\text{F interval length for F optimal design}}{\text{F interval length for given design}} .$$

Table 5.1.3. Four point nearly F optimal designs, point prior, μ=ED_{100p}

p	n	design	F_EFF
.10	50	n_i: 22, 10, 8, 10 P_i: .10, .20, .30, .90	.91
.30	50	n_i: 8, 10, 22, 10 P_i: .10, .20, .30, .80	.95
.50	50	n_i: 12, 13, 13, 12 P_i: .10, .30, .70, .90	.95
.10	100	n_i: 44, 20, 16, 20 P_i: .10, .20, .30, .90	.94
.30	100	n_i: 16, 20, 44, 20 P_i: .10, .20, .30, .80	.98
.50	100	n_i: 25, 25, 25, 25 P_i: .20, .40, .60, .80	.99
.10	150	n_i: 74, 31, 20, 25 P_i: .10, .20, .30, .90	.95
.30	150	n_i: 30, 68, 24, 30 P_i: .20, .30, .40, .80	.96
.50	150	n_i: 37, 38, 38, 37 P_i: .15, .35, .65, .85	.95

Table 5.1.4. Five point nearly F optimal designs, point prior, μ=ED_{100p}

p	n	design	F_eff
.10	50	n_i: 19, 14, 5, 6, 6 P_i: .10, .20, .30, .80, .90	.91
.30	50	n_i: 10, 13, 15, 6, 6 P_i: .10, .20, .30, .80, .90	.97
.50	50	n_i: 10, 10, 10, 10, 10 P_i: .10, .30, .50, .70, .90	.96
.10	100	n_i: 40, 30, 10, 10, 10 P_i: .10, .20, .30, .80, .90	.93
.30	100	n_i: 15, 25, 35, 15, 10 P_i: .10, .20, .30, .40, .80	.94
.50	100	n_i: 15, 20, 30, 20, 15 P_i: .10, .30, .50, .70, .90	.97
.10	150	n_i: 60, 45, 15, 15, 15 P_i: .10, .20, .40, .80, .90	.91
.30	150	n_i: 37, 53, 30, 15, 15 P_i: .20, .30, .40, .80, .90	.96
.50	150	n_i: 20, 30, 50, 30, 20 P_i: .10, .30, .50, .70, .90	.97

§5.2 E Optimal Designs, Point Prior

The E optimal design is a two point design which depends only on ξ. The designs are symmetric about the ED_{50} but with unequal sample allocation. E optimal designs for a variety of ED's of interests were found using the Nelder-Mead algorithm and are given in table 5.2.1.

Recall that these designs minimize the maximum prediction variance of Logit($\hat{p}$) at either location μ+d or μ-d for all d>0. As the ED of interest nears zero or one, the placement of the E optimal design moves towards that of the slope optimal design. This result is shown numerically in table 5.2.1. The E optimal designs in table 5.2.1 are indexed by μ and give the two optimal design points as well of the proportion of the sample, denoted by $\frac{n_1}{n}$, required at P_1.

Table 5.2.1. E optimal designs, point prior, $\mu=ED_{100p}$

p	$\frac{n_1}{n}$	P_1	P_2
1×10^{-12}	.54	.0832	.9168
1×10^{-6}	.59	.0832	.9168
.0001	.63	.0833	.9167
.01	.74	.09	.91
.02	.77	.09	.91
.04	.80	.10	.90
.06	.81	.11	.89
.08	.81	.12	.88
.10	.81	.12	.88
.20	.76	.16	.84
.30	.68	.19	.81
.40	.60	.21	.79
.50	.50	.21	.79

§5.3 Q Optimal Designs, Point Prior

The Q optimal design is a two point design with points placed symmetrically about the ED_{50}. As previously mentioned, the Q optimal design is region dependent. In fact, the sample allocation and location of the symmetric Q optimal design points vary with the size and location of the region of interest. Since we have a region of interest instead of a particular ED of interest, μ is equated to the most neutral ED, the ED_{50}.

Several Q optimal designs are given in tables 5.3.1 and 5.3.2. Designs for symmetric regions of interest about the ED_{50} are provided in table 5.3.1. Asymmetric regions are then studied in table 5.3.2. The symmetric regions have equal sample allocation while the asymmetric regions do not. For the asymmetric regions, the designs place one design point in the region and one at the symmetric counterpart. As the region narrows, the concentration of the sample increases at the point in the region. As the region size goes to zero, all of the sample will converge to a single design point at the center of that region. In contrast, as the region size goes to infinity, the Q optimal design converges to the slope optimal design. Tables 5.3.1 and 5.3.2 are indexed by region of interest and provide the two optimal design points and the proportion of the sample to be placed at the first design point.

Table 5.3.1 Q optimal designs, point prior, μ=ED_{50}, symmetric region

region of interest	$\frac{n_1}{n}$	P_1	P_2
$ED_{00.0}$ to ED_{100}	.50	.0832	.9168
$ED_{00.1}$ to $ED_{99.9}$	.50	.11	.89
ED_{01} to ED_{99}	.50	.14	.86
ED_{10} to ED_{90}	.50	.19	.81
ED_{15} to ED_{85}	.50	.21	.79
ED_{25} to ED_{75}	.50	.26	.74
ED_{35} to ED_{65}	.50	.31	.69
ED_{45} to ED_{55}	.50	.38	.62
ED_{50} to ED_{50}	.50	.50	.50

Table 5.3.2 Q optimal designs, point prior, μ=ED_{50}, asymmetric regions

region of interest	$\frac{n_1}{n}$	P_1	P_2
ED_{01} to ED_{10}	.82	.09	.91
ED_{01} to ED_{20}	.81	.10	.90
ED_{01} to ED_{30}	.80	.11	.89
ED_{01} to ED_{40}	.78	.11	.89
ED_{01} to ED_{50}	.77	.12	.88
ED_{01} to ED_{60}	.75	.13	.87
ED_{01} to ED_{70}	.72	.13	.87
ED_{01} to ED_{80}	.69	.14	.86
ED_{01} to ED_{90}	.64	.14	.86
ED_{10} to ED_{10}	1.0	.10	---
ED_{10} to ED_{20}	.94	.15	.85
ED_{10} to ED_{30}	.89	.17	.83
ED_{10} to ED_{40}	.84	.19	.81
ED_{10} to ED_{50}	.79	.20	.80
ED_{10} to ED_{60}	.73	.21	.79
ED_{10} to ED_{70}	.67	.21	.79
ED_{10} to ED_{80}	.59	.21	.79
ED_{20} to ED_{20}	1.0	.20	---
ED_{20} to ED_{30}	.93	.23	.77
ED_{20} to ED_{40}	.85	.25	.75
ED_{20} to ED_{50}	.77	.26	.74
ED_{20} to ED_{60}	.68	.26	.74
ED_{20} to ED_{70}	.59	.26	.74
ED_{30} to ED_{30}	1.0	.30	---
ED_{30} to ED_{40}	.88	.31	.69
ED_{30} to ED_{50}	.74	.31	.69
ED_{30} to ED_{60}	.61	.30	.70

§5.4 D Optimal Designs, General Prior

The D optimal design with a point prior gives equal allocation to $ED_{17.6}$ and $ED_{82.4}$. If the design is forced to have a third design point, the equal allocation D optimal three point design is given by the $ED_{13.6}$, ED_{50}, and $ED_{86.4}$. The family of D optimal designs is invariant to the choice of sample size and ξ, rendering it a good candidate for studying the effect of different priors. Many Bayesian D optimal designs are given in this section.

The designs are found using the Nelder-Mead algorithm and a numerical expectation. The Bayesian D optimal expression in general is given by

$$\underset{\mathfrak{D}}{\text{Max}} \int_{(\mu,\beta)} |I(\mu,\beta)|\ \pi(\mu,\beta)\ d(\mu,\beta). \tag{5.4.1}$$

In this dissertation, due to computing limitations, the class of designs $\mathfrak{D}$ is limited to kk designs which are centered at the mean of the prior for μ. Both joint normal with zero correlation and joint uniform priors are investigated. The expectations are approximated by a 20 point segmentation of the density in the uniform case and a 20 point segmentation from within ± 4 standard deviations in the normal case. In the normal case a 15 point segmentation was initially explored within ± 3 standard deviations. Neglible differences were found between the 20 point and 15 point approximations. Consequently, the 20 point segmentations are assumed to offer a good approximation to the expectation and were therefore used to find the designs given in table 5.4.1.

The tables are indexed similar to the minimax designs by Sitter. The knowledge of the slope is indicated by the ratio of the experimenter's upper and lower bounds on β, (β_U/β_L), denoted β_R. The location information is indicated by μ_D and is formed by a scaled difference given by $\beta_L(\mu_U-\mu_L)/2$. Values of β_R and μ_D each specify a prior. The product of these two priors yields the joint prior, $\pi(\mu,\beta)$. For the uniform prior, the bounds given in each parameter are taken as the endpoints of the joint uniform density, and in the case of the normal, the bounds are assumed to be ± 2 standard deviations from the mean of the bounds.

The Bayesian optimal designs are specified by k, the optimal number of design points, and w, the optimal scale invariant width between the outermost design points. The actual width of a design in natural units for a particular situation is simply w divided by the average of the slope boundaries, or $2w/(\beta_U+\beta_L)$. For a point prior, the optimal design has two points and a scale invariant width of 3.0875. This translates into the known $ED_{17.6}$ and $ED_{82.4}$.

Table 5.4.1 Bayesian D optimal designs, normal prior, characterized by values of k and w

	β_R					
	1.0	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**
	2	2	2	2	2	2
0.0	3.088	3.081	3.063	3.042	3.020	2.998
	2	2	2	2	2	2
0.5	3.103	3.106	3.105	3.102	3.103	3.105
	2	2	2	2	2	2
1.0	3.144	3.167	3.192	3.219	3.245	3.272
μ_D	2	2	2	2	3	3
1.5	3.194	3.234	3.278	3.320	4.538	4.663
	2	2	2	3	3	3
2.0	3.241	3.291	3.345	4.642	4.802	4.945
	2	2	3	3	3	3
2.5	3.284	3.339	4.650	4.842	5.009	5.820
	2	3	3	3	4	4
3.0	3.319	4.589	4.803	4.997	5.828	6.041
	2	3	3	4	4	6
3.5	3.348	4.711	4.928	5.761	6.002	6.903

If an experimenter is interested in a one-stage D optimal design for the logistic model, table 5.4.1 should be very useful. The Bayesian D optimal design adapts to the prior information available on the model parameters. As the prior widens, table 5.4.1 indicates that the Bayesian D optimal design, normal prior, also widens with additional design points. In addition notice that the effect of β_R depends upon the value μ_D.

The following example illustrates how table 5.4.1 is intended to be used. Suppose an experimenter consults a statistician about designing a one-stage experiment using the logistic model. The statistician requests the experimenter to think of the 95% normal theory confidence interval concept and express his or her beliefs about the parameters in that manner. Once the boundaries are chosen, the Bayesian D optimal design can be extracted from table 5.4.1. Suppose the boundaries are β:(.1, .3) and μ:(210, 280) yielding β_R=3.0 and μ_D=3.5 which in turn yields k=4 and w=6.002. The equal allocation design is centered at 245.0 and has four equally spaced levels in natural units with a total width of the design in natural units of 30.01. The four design levels are 230.00, 240.00, 250.00, and 260.00. If the centers of the boundaries are indeed the correct parameter values, then the design points are placed at EDs 4.7, 26.9, 73.1, and 95.3.

D optimal designs using a uniform prior are given in table 5.4.2. These designs are as expected more disperse than those using a normal prior. Each parameter pair within the choosen boundaries are considered to be equally likely to be the true situation. Consequently, the optimal designs of table 5.4.2 have more design points spread further apart than those of table 5.4.1 in order to cope with the increased likelihood of true parameters being near the boundaries.

For reasons of practicality and limitations on CPU time the search for these designs encompassed kk designs with k ranging from 2 to 12 and a 20 point expectation approximation. For the ranges given in table 5.4.1, a twelve point optimal design never arose. However, using a uniform prior, the need for a twelve or more point kk design does arise. Even so, only Bayesian, uniform prior, D optimal designs with the $k \leq 12$ restriction are tabled. Nevertheless table 5.4.2 shows as does table 5.4.1 that the impact of poor knowledge of β on Bayesian optimal design requirements depends upon the knowledge of μ.

Table 5.4.2 Bayesian D optimal designs, uniform prior, characterized by values of k and w, $k \leq 12$

μ_D \ β_R	1.0	1.5	2.0	2.5	3.0	3.5
0.0	2 3.088	2 3.077	2 3.050	2 3.013	2 2.969	2 2.926
0.5	2 3.111	2 3.117	2 3.119	2 3.120	2 3.125	2 3.136
1.0	2 3.174	2 3.220	2 3.278	2 3.347	3 4.631	3 4.853
1.5	2 3.259	2 3.338	3 4.672	3 4.991	4 5.897	4 6.313
2.0	2 3.342	3 4.730	3 5.098	4 6.125	5 6.991	6 7.803
2.5	3 4.675	3 5.045	4 6.088	5 7.000	10 8.397	12 9.191
3.0	3 4.922	4 5.894	6 7.033	12 8.313	12 8.950	12 9.569
3.5	4 5.675	6 6.725	12 7.944	12 8.538	12 9.131	12 9.725

The designs of table 5.4.2 are considerably wider and have more points than those of table 5.4.1. The uniform prior considers parameter situations near and far from the middle to have equal chance of occurrance, causing the design to spread out relatively quickly. In contrast, the normal prior "downweights" parameter values further away from the middle, resulting in a much slower spread of the Bayesian D optimal design. In either case however, the effect of the prior is clear. The implementation of the Bayesian design procedure in the D optimal criterion forces the D optimal design to adapt to the information available. Consequently the Bayesian designs are quite practical in a one-stage design setting. The less precise the information, generally, the wider the design becomes and the more design points are needed.

CHAPTER 6

ROBUSTNESS TO PARAMETER MISSPECIFICATION

Many optimal and near optimal designs were given in chapter 5. In order, for any of the point prior optimal designs to be implemented, the parameter values must be specified. For these designs to actually be optimal when implemented, the true parameter values must be known. If the model parameters are misspecified, the implemented design will be less than optimal. This chapter briefly examines the effect of parameter misspecification on the asymptotic efficiency of the point prior designs as well as general prior Bayesian designs and motivates the need for designs which are more robust than the ones currently given. The robustness properties of several F optimal designs using a point prior along with several Bayesian D optimal design are investigated.

Robustness tables containing asymptotic efficiency values are used to perform this study. The experimenter is assumed to make guesses of the parameters μ and β, through a prior which may be a point prior. The mean of the prior yields the guessed values μ_0 and β_0. These values along with the true parameter values specify a degree of parameter misspecification and an entry in the table through the indexes $\beta_k=\frac{\beta_0}{\beta}$ and $\mu_k=\beta(\mu_0-\mu)$. When $\mu=ED_{50}$, the F and D efficiencies are symmetric about $\mu_k=0$, and so only the overestimation of μ is tabled.

Table 6.0.1 is provided to give the reader an idea of how the indexes of the table correspond to the actual locations of the design points in terms of probabilities of response. The table contains the actual probabilities associated with the design points given a particular level of parameter misspecification. The table uses the D optimal design, which is given by the $ED_{17.6}$ and $ED_{82.4}$, to illustrate the effect of design implementation when the parameters are misspecified. If parameter knowledge is perfect, β_k=1 and μ_k=0, we indeed have two design points which correspond to the ideal probabilities .176 and .824, respectively. If, however, the parameter knowledge is not perfect and the parameters are consequently misspecified, the two design point locations will not correspond to the ideal probabilities but rather something else.

Table 6.0.1 Design point locations (ED's) as a function of parameter guesses

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	.071	.127	**.176**	.216	.249	.276
		.929	.873	**.824**	.784	.751	.724
	0.5	.112	.193	.260	.313	.354	.386
		.956	.919	.885	.856	.832	.812
	1.0	.172	.283	.367	.429	.474	.509
		.923	.949	.927	.908	.891	.877
	1.5	.255	.394	.489	.553	.598	.631
		.983	.969	.955	.942	.931	.922
	2.0	.361	.518	.612	.672	.710	.738
		.990	.981	.972	.964	.957	.951
	2.5	.482	.639	.722	.771	.802	.823
		.994	.988	.983	.978	.973	.970

An example of parameter guesses for a one variable logistic model in an actual study is given by Dr. Kimber White at Medical College of Virginia in 1991, see Myers (unpublished dissertation, 1991). The study was conducted to investigate the anesthesizing effect of a drug called Brevital on mice. Dr. White's guesses correspond to locations in table 6.0.1 of β_k=.58 and μ_k=-2.20, which corresponds to μ_k=2.20. The guesses which are located at the edge of the table indicate to some degree that the table is in fact representative of the parameter knowledge experimenters have in practice.

§6.1 F Optimal Designs, Point Prior

Robustness tables are given for the n=40 and n=100, μ=ED_{50} F optimal designs as well as the n=100, μ=ED_{20} F optimal design in tables 6.1.1, 6.1.2, and 6.1.3, respectively. The F efficiency values used in these tables is given by

$$\text{F\_EFF}= \frac{(\text{length of F interval for the given design} \mid \mu,\beta)}{(\text{length of F interval for the given design} \mid \mu_0, \beta_0)}.$$

Table 6.1.1 F efficiencies of the n=40, μ=ED_{50} F optimal design, point prior

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	.82	.96	1.00	.97	.89	.74
	0.5	.82	.95	.96	.88	.72	.48
	1.0	.80	.90	.80	.58	.33	.08
	1.5	.76	.70	.43	.18	.03	.00
	2.0	.55	.23	.01	.00	.00	.00
	2.5	.00	.00	.00	.00	.00	.00

Table 6.1.2 F efficiencies of the n=100, μ=ED_{50} F optimal design, point prior

		β_k 0.6	0.8	1.0	1.2	1.4	1.6
	0.0	.89	.98	1.00	.98	.94	.87
	0.5	.90	.95	.92	.83	.71	.57
μ_k	**1.0**	.89	.83	.68	.53	.39	.26
	1.5	.79	.59	.41	.27	.16	.07
	2.0	.54	.32	.19	.09	.01	.00
	2.5	.26	.12	.04	.00	.00	.00

Tables 6.1.1 and 6.1.2 display a few trends which need to be noted. As expected, since the F optimal designs depend upon sample size, the designs become more robust as sample size increases. In addition, if μ=ED_{50}, it is better to underestimate the slope than to overestimate it. This fact becomes more evident as the prior estimate of μ degrades. The zero efficiencies in tables 6.1.1 and 6.1.2 occur when $g \geq 1$ and the Fieller interval is defined to be infinite. This results from a design which does not allow for a reasonable estimate of the slope.

Table 6.1.3 F efficiencies of the n=100, μ=ED_{20} F optimal design, point prior

		β_k 0.6	0.8	1.0	1.2	1.4	1.6
	- 2.5	.54	.62	.50	.26	.06	.00
	- 2.0	.59	.68	.66	.50	.29	.11
	- 1.5	.63	.75	.77	.71	.56	.38
	- 1.0	.69	.81	.86	.85	.79	.68
	- 0.5	.76	.90	.95	.96	.94	.90
μ_k	**0.0**	.85	.98	1.00	.99	.96	.92
	0.5	.75	.90	.89	.85	.79	.72
	1.0	.18	.51	.57	.55	.50	.44
	1.5	.14	.16	.25	.26	.24	.21
	2.0	.00	.00	.04	.07	.07	.05
	2.5	.00	.00	.00	.00	.00	.00

Recall that the F optimal designs, $\mu \neq ED_{50}$, tend to place a majority of the sample at the design point closest to μ and the rest at the symmetric counterpart. Therefore intuitively, the location of the design point with the majority of the sample is more critical than the location of the other design point. Consequently the robustness of these designs is more sensitive to the location of the design point near μ. Table 6.1.3 relays this charateristic. The majority of the sample is placed at what is thought to be the ED_{20}. The overestimation of μ would quickly place this design point in a location of zero response as is illustrated by the poor efficiencies in the lower part of table 6.1.3. On the other hand, slight to moderate underestimation allows this design point to reside in the transitional part of the curve accounting for the better efficiences in the upper part.

Robustness tables for the other F optimal designs show similar trends to those of tables 6.1.1-6.1.3 and are therefore not given. The main point which needs to be taken from this very brief examination is that these one stage F optimal designs are not particularly robust to parameter misspecification. The optimal designs having only two points makes them quite sensitive to incorrect initial specification of the parameters. One would therefore think the four and five level nearly F optimal designs given in tables 5.1.3 and 5.1.4 would be considerably more robust. This however is not the case. The designs resemble the optimal designs so closely that they offer little or no improvement of robustness to parameter misspecification. Consequently, additional design points alone do not insure a robust experimental plan.

§6.2 D Optimal Designs, General Prior

Robustness tables are given for D optimal designs utilizing point priors as well as other priors. Tables 6.2.1 and 6.2.2 contain D efficiencies for the point prior, two and three point D optimal designs, respectively. In addition tables 6.2.3, 6.2.4, and 6.2.5 examine three Bayesian D optimal designs with nontrivial normal priors, and in table 6.2.6 a Bayesian nontrivial uniform prior D optimal design is studied. The D efficiency used in the tables is defined by

$$D\_EFF=\frac{(\ |I(\mu,\beta)|\ \text{ for given design } |\ \mu_0,\beta_0)}{(\ |I(\mu,\beta)|\ \text{for given design } |\ \mu,\beta)}. \qquad (6.2.1)$$

Table 6.2.1 D efficiencies of the D optimal design, point prior

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	.57	.91	1.00	.95	.85	.74
	0.5	.55	.86	.93	.87	.77	.67
μ_k	**1.0**	.50	.73	.75	.68	.59	.50
	1.5	.41	.54	.52	.45	.37	.31
	2.0	.31	.35	.31	.25	.21	.17
	2.5	.20	.20	.16	.13	.10	.08

Table 6.2.2 D efficiencies of the three point D optimal design, point prior

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	.68	.93	1.00	.96	.87	.77
	0.5	.69	.92	.96	.90	.81	.70
μ_k	**1.0**	.71	.85	.83	.74	.64	.55
	1.5	.69	.73	.64	.53	.44	.36
	2.0	.61	.54	.43	.33	.26	.20
	2.5	.48	.35	.25	.18	.13	.10

As can be seen from tables 6.2.1 and 6.2.2, neither design can be said to be particularly robust to parameter misspecification. Even so, the constraint to the use of three points results in a somewhat more robust design. Although the three point design is more robust, it is not necessarily the better design. Unlike the two point design the three point design represents a situation where a constraint was used to force a third level. Consequently, the determinant of the information matrix from the two point design is larger than that from the three point design when the parameters are known. The ratio of the determinant from the three point design to the determinant from the two point design yields another D

efficiency which does not measure robustness but rather compares two competing designs. In this situation the three point design is 86% as efficient as the two point design when the parameters are known.

Tables 6.2.3, 6.2.4, and 6.2.5 compare three different Bayesian D optimal designs to the D optimal point prior design using D efficiency as defined by expression 6.2.1.

Table 6.2.3 D efficiencies, β_R=1.5, μ_D=1.5, normal prior D optimal design

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	.51	.87	1.00	.97	.89	.78
	0.5	.50	.83	.93	.90	.81	.71
	1.0	.45	.71	.76	.70	.62	.53
	1.5	.38	.53	.53	.47	.40	.33
	2.0	.29	.35	.32	.27	.22	.18
	2.5	.20	.21	.17	.14	.11	.09

Table 6.2.4 D efficiencies, β_R=2.0, μ_D=2.5, normal prior D optimal design

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	.48	.80	1.00	1.07	1.06	1.00
	0.5	.50	.81	.98	1.03	1.01	.94
	1.0	.54	.81	.92	.91	.85	.76
	1.5	.57	.77	.78	.71	.62	.53
	2.0	.57	.66	.59	.49	.40	.32
	2.5	.52	.50	.39	.29	.22	.17

Table 6.2.5 D efficiencies, β_R=3.5, μ_D=3.0, normal prior D optimal design

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	.52	.79	1.00	1.13	1.17	1.15
	0.5	.50	.77	.98	1.09	1.12	1.09
μ_k	**1.0**	.47	.73	.91	.98	.98	.93
	1.5	.44	.69	.81	.82	.77	.69
	2.0	.43	.64	.68	.63	.54	.46
	2.5	.44	.57	.52	.42	.33	.26

As seen from tables 6.2.3, 6.2.4, and 6.2.5 the robustness of the Bayesian D optimal designs increases as the prior spreads. Unfortunately however, the efficiency of the Bayesian designs in comparison to the D optimal design decreases as the prior variance increases. The trade-off is between protection against poor initial parameter knowledge and high efficiency in comparison to the best when the parameter knowledge is good. A desirable design in practice would be as robust as is deemed necessary yet be nearly optimal in the presence of good parameter guesses. The designs studied in tables 6.2.3, 6.2.4, and 6.2.5 are respectivelly 99.6%, 80%, and 73% as efficient as the D optimal design when parameters are known.

Table 6.2.6 D efficiencies, β_R=2.0, μ_D=3.0, uniform prior D optimal design

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	.40	.67	1.00	1.33	1.60	1.79
	0.5	.39	.67	1.01	1.33	1.58	1.75
μ_k	**1.0**	.38	.68	1.02	1.31	1.51	1.62
	1.5	.37	.70	1.02	1.25	1.37	1.39
	2.0	.37	.72	.99	1.13	1.15	1.08
	2.5	.39	.71	.90	.94	.87	.76

The uniform prior Bayesian D optimal design studied in table 6.2.6 is quite robust to the mispecification of μ if β is not underspecified. The underspecification of β causes the already wide uniform Bayesian design to become excessively broad. As is expected the efficiency relative to the D optimal design when the parameters are known is low, at 44%. The one-stage designs which are most robust are the same designs which are least efficient.

Tables 6.2.1-6.2.6 contain some interesting robustness patterns. In order to understand how these patterns come about, one needs only to look at the location of the design points in terms of ED's. The movement of these design points as the parameter guesses change paints the robustness picture.

This robustness study, though brief, shows the advantage in terms of robustness of the Bayesian designs with a nontrivial prior over those with the trivial point prior. In fact, using the Bayesian designs, the robustness properties can be controlled, to some degree, in order to suit the needs of the experimenter. However, it is also clear that when parameter knowledge is good this advantage is soon outweighed by the inefficiency of these designs in comparison to the D optimal design. A one-stage design simply cannot address both issues effectively. The trade-off between robustness and efficiency is often unsatisfactory. A reasonable solution is to avoid the trade-off and address the two goals separately.

CHAPTER 7

TWO-STAGE DESIGNS

In this chapter, a two-stage experimental design procedure which utilizes Bayesian techniques is developed for the logistic regression model. This procedure offers a practical alternative to the single stage designs of chapter 5. The shortcomings illustrated in chapter 6 of the one-stage designs in terms of the trade-off between efficiency and robustness are dealt with effectively. The two-stage procedure will allow the two goals to be addressed separately instead of simultaneously. In addition, a section is devoted to the evaluation of the procedure as is a section concerning the sample distribution between the stages. This two-stage procedure in part originated with some other two-stage work by Minkin (1987) and Myers, Myers, and Carter (1994). Myers et.al. found and studied two-stage D-Q (non-Bayesian) optimal designs which were found to be considerably more robust and nearly as efficient as the one-stage designs. The main extension from their work and Minkin's work is in the use of Bayesian design and Bayesian estimation.

§7.1 Two-stage Design Procedure

The two-stage procedure uses two design optimality criteria, one in the first stage and one in the second. The first stage design should be quite robust to poor parameter guesses while not necessarily being very efficient if parameter knowledge is good. In order to cope with varying degrees of parameter knowledge, the first stage design will incorporate the Bayesian design framework of section 5.2.

Many reasonable design criteria could be used in the first stage. However, since the two-stage design procedure is under study and not the optimality criterion, the D criterion is used. The invariance of the D criterion to sample size and ξ makes it a good choice. For these reasons, the second stage makes use of the D criterion as well, even though any reasonable criterion could be used. Following the convention set by Myers, Myers, and Carter (1994), a two-stage Bayesian D-D optimal design strategy is studied.

The purpose of the first stage is to provide good point estimates of the parameters to the second stage criterion. Given the first stage design, the second stage design maximizes the determinant of the total information matrix. The first stage addresses the problem of robustness with respect to parameter misspecification while the second stage addresses efficiency when the parameters are assumed known. The second stage complements the first and corrects for any estimated misplacement of points in an effort to best achieve the overall goal of high efficiency.

In order for the first stage to provide good estimates of the parameters to the second stage the optimal Bayesian designs from section 5.4 are used. However, these designs do not ensure good estimates. At times, absurd maximum likelihood estimates of the parameters are obtained due to the unstable nature of the slope estimate when the design is poor. A poor design would likely result from misspecified parameters, even when using a Bayesian design. To improve this situation, the prior described in section 5.4 is used a second time to create Bayesian parameter estimates. The prior contains the experimenters *Apriori* beliefs about μ and β. The information is combined with the likelihood forming a posterior likelihood. Ideally, one would want to find the mean of this posterior. However, estimating the posterior mean numerically would require an excessively large amount of CPU time. Consequently, the posterior mode or maximum posterior likelihood estimate, which is much easier to obtain is used.

The posterior mode is found using the Nelder-Mead algorithm (Nelder and Mead, 1965), solving an adjusted set of nonlinear equations, adjusted from the set of equations obtained for maximum likelihood estimation. The concern which arises when using the posterior mode is the possibility of multiple modes. To be

confident that the algorithm is converging to the global mode if more than one exists, the algorithm was begun at various starting points for several different situations. Each case resulted in the same mode, giving satisfactory confidence that we indeed have the global mode.

The set of nonlinear equations one solves to obtain the maximum likelihood estimates of μ and β is given by

$$\sum_{i=1}^{k}(r_i\text{-}n_iP_i)(x_i\text{-}\mu) = 0 \qquad (7.1.1)$$

and

$$-\sum_{i=1}^{k}(r_i\text{-}n_iP_i)\beta = 0$$

where k is the number of design points, r_i is the number of successes at the i^{th} design point out of n_i, and P is the centered logistic model. When a nontrivial prior is used and a posterior is found, the set of equations no longer represents the sum of residuals and weighted residuals. The likelihood function apart from constants for the centered logistic regression model is given by

$$L(\mu,\beta;\, \mathbf{r}) \propto \prod_{i=1}^{k} P_i^{r_i}(1\text{-}P_i)^{n_i\text{-}r_i}.$$

The expression for a joint independent normal prior apart from constants is then

$$\pi(\mu,\beta) \;\propto\; \mathbb{EXP}[-\tfrac{1}{2}(\tfrac{\beta\text{-}\mu_\beta}{\sigma_\beta})^2]\;\mathbb{EXP}[-\tfrac{1}{2}(\tfrac{\mu\text{-}\mu_\mu}{\sigma_\mu})^2]$$

where β and μ are random variables and μ_β, σ_β and μ_μ, σ_μ are their respective means and standard deviations. Consequently, the posterior likelihood, again apart from constants, is

$$PL \;\propto\; \mathbb{EXP}[-\tfrac{1}{2}(\tfrac{\beta\text{-}\mu_\beta}{\sigma_\beta})^2]\;\mathbb{EXP}[-\tfrac{1}{2}(\tfrac{\mu\text{-}\mu_\mu}{\sigma_\mu})^2]\;\prod_{i=1}^{k} P_i^{r_i}(1\text{-}P_i)^{n_i\text{-}r_i}.$$

Taking the log, the posterior likelihood can be written in the following form

$$\text{Log(PL)} \propto -\frac{1}{2}\left(\frac{\beta-\mu_\beta}{\sigma_\beta}\right)^2 - \frac{1}{2}\left(\frac{\mu-\mu_\mu}{\sigma_\mu}\right)^2 + \sum_{i=1}^{k}\left(r_i\text{Log}(P_i) + n_i\text{Log}(1-P_i) - r_i\text{Log}(1-P_i)\right). \tag{7.1.2}$$

Differentiating expression 7.1.2 with respect to β yields

$$\frac{\partial\text{Log(PL)}}{\partial\beta} = \left(\frac{\mu_\beta-\beta}{\sigma_\beta^2}\right) + \sum_{i=1}^{k}\left(\frac{r_i}{P_i}\frac{\partial P_i}{\partial\beta} - \frac{n_i}{(1-P_i)}\frac{\partial P_i}{\partial\beta} + \frac{r_i}{(1-P_i)}\frac{\partial P_i}{\partial\beta}\right) \tag{7.1.3}$$

where $$\frac{\partial P_i}{\partial\beta} = P_i(1-P_i)(x_i-\mu).$$

Simplifying expression 7.1.3 we obtain

$$\frac{\partial\text{Log(PL)}}{\partial\beta} = \left(\frac{\mu_\beta-\beta}{\sigma_\beta^2}\right) + \sum_{i=1}^{k}(r_i-n_iP_i)(x_i-\mu). \tag{7.1.4}$$

Differentiating expression 7.1.2 with respect to μ yields

$$\frac{\partial\text{Log(PL)}}{\partial\mu} = \left(\frac{\mu_\mu-\mu}{\sigma_\mu^2}\right) + \sum_{i=1}^{k}\left(\frac{r_i}{P_i}\frac{\partial P_i}{\partial\mu} - \frac{1}{(1-P_i)}\frac{\partial P_i}{\partial\mu} + \frac{r_i}{(1-P_i)}\frac{\partial P_i}{\partial\mu}\right) \tag{7.1.5}$$

where $$\frac{\partial P_i}{\partial\mu} = -\beta(P_i(1-P_i)).$$

After some simplification expression 7.1.5 becomes

$$\frac{\partial\text{Log(PL)}}{\partial\mu} = \left(\frac{\mu_\mu-\mu}{\sigma_\mu^2}\right) - \sum_{i=1}^{k}\beta(r_i-n_iP_i). \tag{7.1.6}$$

Equating expressions 7.1.4 and 7.1.6 to zero produces the set of nonlinear equation which when solved yield the posterior mode. The pair of equations,

$$\left(\frac{\mu_\beta - \beta}{\sigma_\beta^2}\right) + \sum_{i=1}^{k} (r_i - n_i P_i)(x_i - \mu) = 0 \qquad (7.1.7)$$

and

$$\left(\frac{\mu_\mu - \mu}{\sigma_\mu^2}\right) - \sum_{i=1}^{k} (r_i - n_i P_i)\beta = 0,$$

is merely an adjusted version of the set given by 7.1.1. The adjustment terms keep the estimation of μ and β from straying too far away from their respective prior means in units of prior variance.

If a joint uniform distribution is used as the prior, the posterior mode is simply given by the maximum likelihood estimates subject to the constraints that $\beta_L \leq \widehat{\beta} \leq \beta_U$ and $\mu_L \leq \widehat{\mu} \leq \mu_U$. The estimates become the nearest endpoint of the prior if the maximum likelihood estimates should fall outside the uniform prior.

The first stage information matrix, apart from notation, is identical to the one given in section 3.1 for the single stage experiment using the centered model. To derive the second stage information matrix, we must begin with the derivation of the conditional likelihood. Abdelbasit and Plackett (1983) created two-stage designs based upon the first stage information matrix. The conditional likelihood and total information matrix were not used for finding the optimal second stage design. They used the same design (D optimal) in terms of probabilities of response for both stages. The second stage was not allowed to complement the first. Minkin, however, found the two-stage D optimal design by maximizing the determinant of the total information matrix. Nevertheless, few details were given as to the derivation of the joint information matrix and conditional likelihood. In addition, little investigation or exploration of the two-stage design was performed.

The likelihood for the first stage is the same as the likelihood for the single stage experiment except the notation has been changed slightly to allow for differentiation between the two stages.

$$L_1(\mu,\beta;\, \mathbf{r}_1|\, \mathbf{x}_1)=\prod_{i=1}^{k_1}\binom{n_{1i}}{r_{1i}} P_{1i}^{\,r_{1i}}(1-P_{1i})^{n_{1i}-r_{1i}} \qquad (7.1.8)$$

where $$P_{1i} = \frac{1}{1 + \exp^{(\xi - \beta(x_{1i}-\mu))}}$$

is the probability of "success" at the i^{th} design point in the first stage, $\mathbf{r}_1$ is a $(k_1 \times 1)$ vector of responses in the first stage in which each element is the number of "successes" out of n_{1i}, the sample size at the i^{th} design point, k_1 is the number of design points in the first stage, and $\mathbf{x}_1$ is a $(k_1 \times 1)$ vector of design point locations. We know that the total or joint likelihood can be expressed as a product of the first stage likelihood and the second stage likelihood conditioned on the first.

$$L_{1,2}(\mu,\beta;\, \mathbf{r}_1, \mathbf{r}_2|\, \mathbf{x}_1, \mathbf{x}_2) = L_1(\mu,\beta;\, \mathbf{r}_1|\, \mathbf{x}_1)\; L_{2|1}(\mu,\beta;\, \mathbf{r}_2|\, \mathbf{r}_1, \mathbf{x}_1, \mathbf{x}_2) \qquad (7.1.9)$$

where the conditional likelihood for the second stage given the first stage is given by

$$L_{2|1}(\mu,\beta;\, \mathbf{r}_2|\, \mathbf{r}_1, \mathbf{x}_1, \mathbf{x}_2) = \prod_{j=1}^{k_2}\binom{n_{2j}}{r_{2j}} P_{2j}^{\,r_{2j}}(1-P_{2j})^{n_{2j}-r_{2j}} \qquad (7.1.10)$$

and $$P_{2j} = \frac{1}{1 + \exp^{(\xi - \beta(x_{2j}-\mu))}}$$

is again the probability of "success" at the j^{th} design point in the second stage, $\mathbf{r}_2$ is a $(k_2 \times 1)$ vector of responses in the second stage in which each element is the number of "successes" out of n_{2j}, the sample size at the j^{th} design point, and k_2 is the number of design points in the second stage.

In this dissertation, the number of design points in the second stage, k_2, is two. In order for the entire two-stage design to be highly efficient, it must resemble the optimal two level designs. Even though the first stage may have many levels or points in order to address its goal of robustness, the second stage needs just two levels to complete a two-stage design which has the complexion of an optimal two level design.

Taking the natural logarithm of equation 7.1.9 results in

$$\text{Log } L_{1,2} = \text{Log } L_1 + \text{Log } L_{2|1}.$$

After expanding,

$$\text{Log } L_{1,2} = \sum_{i=1}^{k_1} \left(\log\binom{n_{1i}}{r_{1i}} + r_{1i}\log(P_{1i}) + (n_{1i}\text{-}r_{1i})\log(1\text{-}P_{1i})\right)$$

$$+ \sum_{j=1}^{k_2} \left(\log\binom{n_{2j}}{r_{2j}} + r_{2j}\log(P_{2j}) + (n_{2j}\text{-}r_{2j})\log(1\text{-}P_{2j})\right). \quad (7.1.11)$$

Equation 7.1.11 can then be used directly to find the Fisher information matrix. However, the total or joint information matrix is merely the sum of the individual information matrices formed from equations 7.1.8 and 7.1.10, the two likelihoods. The elements of the information matrix will parallel those given in section 3.1. The total information matrix for the two-stage procedure is then

$$I_{1,2}(\mu,\beta) = \begin{pmatrix} I_{11} & I_{12} \\ I_{12} & I_{22} \end{pmatrix}$$

where, $$I_{11} = \beta^2 \sum_{i=1}^{k_1} n_{1i}P_{1i}(1\text{-}P_{1i}) + \beta^2 \sum_{j=1}^{k_2} n_{2j}P_{2j}(1\text{-}P_{2j}),$$

$$I_{12} = -\beta \sum_{i=1}^{k_1} (x_{1i} - \mu) n_{1i} P_{1i}(1-P_{1i}) - \beta \sum_{j=1}^{k_2} (x_{2j} - \mu) n_{2j} P_{2j}(1-P_{2j}),$$

and $$I_{22} = \sum_{i=1}^{k_1} (x_{1i} - \mu)^2 n_{1i} P_{1i}(1-P_{1i}) + \sum_{j=1}^{k_2} (x_{2j} - \mu) n_{2j} P_{2j}(1-P_{2j}).$$

As before, the inverse of the information matrix is given by,

$$I_{1,2}^{-1}(\mu,\beta) = \begin{pmatrix} V_{11} & V_{12} \\ V_{12} & V_{22} \end{pmatrix} \tag{7.1.12}$$

where $V_{11} = \frac{I_{22}}{D}$, $V_{12} = -\frac{I_{12}}{D}$, $V_{22} = \frac{I_{11}}{D}$, and D is the determinant of $I_{1,2}(\mu,\beta)$.

The first stage of the D-D design is found by using the components from the first stage inverted information matrix and is given by

$$\underset{\mathfrak{D}}{\text{Min}} \int_{(\mu,\beta)} |I_1^{-1}(\mu,\beta)| \, \pi(\mu,\beta) \, d(\mu,\beta). \tag{7.1.13}$$

The second stage design is found using the elements of $I_{1,2}^{-1}(\mu,\beta)$, in which the first stage Bayesian estimates, $(\hat{\mu},\hat{\beta})$ are used throughout. Within this matrix it is assumed that the first component in the expressions I_{11}, I_{12}, and I_{22} are fixed with respect to first stage design. The first component represents the information obtained about μ and β in the first stage, and the second component represents the information due to the second stage experiment given the first. The second stage design is determined with respect to (x_{2j}'s) and (n_{2j}'s) by

$$\underset{\mathfrak{D}}{\text{Min}} \; |I_{1,2}^{-1}(\mu,\beta)|. \tag{7.1.14}$$

These two factors, the second stage design points and corresponding sample distribution, are in essence determined so that the information from the second stage experiment best complements the information which has already been collected on μ and β from the first stage when the end goal is to minimize the two-stage generalized asymptotic variance. A computer program is provided in appendix F which produces the second stage design of a ?-D optimal two-stage design.

Although the two-stage design procedure has been developed in the context of the D criterion, the methodology given is easily extended. Many other possibilities of two-stage design procedures exist simply by using different combinations of the D, E, F, and Q criteria.

§7.2 Evaluation of the Two-stage Design Procedure

In this section a method is developed for evaluating the two-stage design procedure. The two-stage designs are to be compared against one-stage designs in addition to other two-stage designs. Also of interest is an investigation of the use of various priors. D efficiency expressions are given to perform such comparisons.

Given $\pi(\mu,\beta)$, the one-stage design is deterministic. The two-stage procedure, however, results in a random design when given $\pi(\mu,\beta)$. Although the first stage of the two-stage procedure behaves exactly like a one-stage procedure, the second stage does not. The second stage design is random due to the second stage initial parameter estimates being random. In a one-stage procedure, a fixed amount of design dependent information is gained about μ and β for a given prior situation whereas in a two-stage procedure, the amount of design dependent information obtained about the parameters is random for a given prior situation.

The D efficiency for the comparison of a two-stage procedure to a one-stage design each with the same prior is defined to be

$$\text{D\_EFF}=\frac{(\ |\mathrm{I}^{-1}(\mu,\beta)|\ \text{ for one-stage design}\mid \pi(\mu,\beta))}{(\ |\mathrm{I}_{1,2}^{-1}(\mu,\beta)|\ \text{for two-stage procedure}\mid \pi(\mu,\beta))}. \qquad (7.2.1)$$

Unlike the numerator, the denominator of 7.2.1 needs some further development. Let $\nu_{11}(\theta)$, $\nu_{12}(\theta)$, and $\nu_{22}(\theta)$ denote the asymptotic variances and covariance of the coefficients μ and β for the two-stage procedure where $\theta=\{\mu,\beta\}$, and $\mathbf{t_2}=\{\mathbf{x_2},\mathbf{n_2}\}$ denotes the second stage design with density $f(\mathbf{t_2})$. We then make use of an equality (see Rao 1973) for the variance of a random variable Y which is observed only in the presence of a random variable W:

$$\mathrm{Var}(Y) = E_W[\mathrm{Var}(Y|W)] + \mathrm{Var}_W[E(Y|W)].$$

In the context needed here, the variance expression becomes

$$\mathrm{Var}(\hat{\theta}) = E_{\mathbf{t_2}}[\mathrm{Var}(\hat{\theta}|\mathbf{t_2})] + \mathrm{Var}_{\mathbf{t_2}}[E(\hat{\theta}|\mathbf{t_2})].$$

The conditional expectation, $E(\hat{\theta}|\mathbf{t_2})$ equals θ, a constant, asymptotically, since $\hat{\theta}$ is the MLE of θ. Consequently, the variance expression simplifies to give the following form

$$\mathrm{Var}(\hat{\theta}) = E_{\mathbf{t_2}}[\mathrm{Var}(\hat{\theta}|\mathbf{t_2})]$$

This in turn implies

$$\nu_{ij}(\hat{\theta}) = E_{\mathbf{t_2}}\left(V_{ij}(\hat{\theta}|\mathbf{t_2})\right) \qquad (7.2.2)$$

where $V_{ij}(\hat{\theta}|\mathbf{t_2})=V_{ij}$ denotes the elements from $I^{-1}_{1,2}(\mu,\beta)$ for a particular two-stage design. Expression 7.2.2 can then be written equivalently as

$$\nu_{ij}(\hat{\theta}) = \int_{\mathbf{t_2}} V_{ij}(\hat{\theta}|\mathbf{t_2})\, f(\mathbf{t_2})\, d\mathbf{t_2}.$$

The expectation over $\mathbf{t_2}$ suggests that for evaluation purposes, since the joint density of $\{\mathbf{x_2},\mathbf{n_2}\}$ is unknown, one should average the estimated variance of coefficients, from many two-stage experiments. These experiments will be identical with the exception of the random nature of the first stage parameter estimates and the second stage design. This average is an estimate of the asymptotic variance of a coefficient using the two-stage procedure.

These variances would then be used to estimate the determinant of the asymptotic generalized variance from the two-stage procedure by

$$|\hat{\Gamma}_{1,2}^{-1}(\mu,\beta)| = \hat{\nu}_{11}\,\hat{\nu}_{22} - \hat{\nu}_{12}^2$$

where $\Gamma_{1,2}^{-1}(\mu,\beta)$ is the inverted total information matrix for the two-stage procedure, not to be confused with $I_{1,2}^{-1}(\mu,\beta)$, the inverted total information matrix for a particular two-stage design.

Various other efficiencies can be defined using expression 7.2.2 when the two-stage procedure is involved. The comparison of a two-stage procedure using one prior and a one-stage design using another prior uses the D efficiency defined by

$$\text{D\_EFF}=\frac{(\ |I^{-1}(\mu,\beta)|\ \text{for one-stage design}\ |\ \pi_1(\mu,\beta))}{(\ |\Gamma_{1,2}^{-1}(\mu,\beta)|\ \text{given}\ \pi_2(\mu,\beta))} \qquad (7.2.3)$$

where π_1 and π_2 have the same mean and π_1 could be a point prior. This efficiency, for example, could study the effect of the prior or perhaps make comparisons between the two-stage procedure and the D optimal design, a two point one-stage design. Recall that for a one-stage design, a point prior can be used since the prior is only used for finding the Bayesian design. Although the two-stage design procedure could use a point prior, it would be counterproductive since it is used in the estimation of the parameters as well as design.

The effect of the prior can be studied with approaches other than that given in expression 7.2.3. One would be to compare two two-stage procedures which have different priors with the same mean. The D efficiency is defined as

$$D\_EFF=\frac{(\ |\Gamma_{1,2}^{-1}(\mu,\beta)|\ \text{given}\ \pi_1(\mu,\beta))}{(\ |\Gamma_{1,2}^{-1}(\mu,\beta)|\ \text{given}\ \pi_2(\mu,\beta))}. \tag{7.2.4}$$

Also the effect of using the Bayesian estimation in addition to Bayesian design might be of interest. This effect is addressed by

$$D\_EFF=\frac{(\ |\Gamma_{1,2}^{-1}(\mu,\beta)|\ \text{given}\ \pi(\mu,\beta)\ \text{using MLE's})}{(\ |\Gamma_{1,2}^{-1}(\mu,\beta)|\ \text{given}\ \pi(\mu,\beta)\ \text{using BE's})} \tag{7.2.5}$$

where MLE and BE denote maximum likelihood and Bayesian estimates respectfully.

The method for evaluation of the two-stage procedure has been illustrated using the D optimality criterion. However, expression 7.2.2 can acommodate any criterion making use of the variance components. Consequently, all given efficiency expressions can be easily converted to the criterion of choice.

§7.3 Optimal Sample Distribution

The optimal distribution of the sample to the first and second stages of the two stage procedure is an important point of consideration. Myers (1991) found in a limited simulation study that about 30% of the total sample is appropriate for the first stage of the D-Q procedure. This translates to expending 30% of the sample for the purpose of finding good initial estimates of the parameters and 70% of the sample devoted solely to reaching the final goal of the experiment through the second stage criterion. A much more extensive sample allocation study is illustrated in this section for the normal prior situation.

The optimal sample allocation depends on several factors including the first stage design (prior), total sample size, and the severity of the parameter misspecification. A cross-section of three first stage Bayesian D optimal designs from table 5.4.1 were used in the study: k=2 and w=3.234, k=3 and w=4.650, and k=4 and w=6.041. In addition, three different levels of parameter misspecification are used: μ_k=0.0 and β_k=1.0, μ_k=2.0 and β_k=0.6, and μ_k=2.0 and β_k=1.4. For each of the resulting nine combinations, the sample sizes ranging from 30 to 300 are studied.

A two dimensional grid was then created for each of the nine situations. Each are indexed by sample size and percent of the sample used in the first stage. In order to overcome CPU time contraints, a second order reponse surface model of the design criterion values over each grid is estimated using 35-40 100 run simulations for each of the nine combinations. The fitted models are then maximized with respect to percent in the first stage. The result is a set of lines which estimates the optimal sample allocation as a function of sample size.

Preliminary simulations showed that a second order model is not appropriate for the entire grid. However, such a model is a reasonable approximation for a subset of the grid about the line of maximum. The design region is taken to be this subset and a grid of approximately 30 locations in the region make up the design points. A grid design instead of a more conventional second-order design was used in order to help minimize the effects of bias due to a misspecified model. Ten points were then replicated to allow for a test of model adequacy as well as offer an estimate of simulation error through the coefficient of variation.

Using the D optimality criteria the second-order model is given by

$$|\mathrm{I}(\mu,\beta)| = \beta_0 + \beta_1 \mathrm{N} + \beta_2 \mathrm{Pn} + \beta_3 \mathrm{NPn} + \beta_4 \mathrm{N}^2 + \beta_5 \mathrm{Pn}^2 \qquad (7.3.1)$$

where N is the total sample size and Pn is the percent of the sample in the first stage. Differentiating 7.3.1 with respect to Pn yields

$$\frac{\partial|\mathrm{I}(\mu,\beta)|}{\partial \mathrm{Pn}} = \beta_2 + \beta_3 \mathrm{N} + 2\beta_5 \mathrm{Pn}. \qquad (7.3.2)$$

Setting 7.3.2 equal to zero and solving for Pn results in the following linear equation,

$$Pn = \frac{-\beta_2 - \beta_3 N}{2\beta_5}$$

which gives the optimal first stage sample proportion as a function of sample size assuming $\beta_5<0$ and the model is correct at the conditional locations of maximum determinant.

Table 7.3.1 Estimated optimal sample allocation between stages, N=30-300

β_k	μ_k	β_R=3.5 μ_D=3.0 (k=4 w=6.041)	R^2	CV(%)
1.0	0.0	$\hat{P}n = 55.07 - .11N$	.94	.97
0.6	2.0	$\hat{P}n = 46.47 - .08N$	.96	1.60
1.4	2.0	$\hat{P}n = 49.93 - .08N$	.95	1.19
β_k	μ_k	β_R=2.0 μ_D=2.5 (k=3 w=4.650)	R^2	CV(%)
1.0	0.0	$\hat{P}n = 50.82 - .12N$	.96	.39
0.6	2.0	$\hat{P}n = 43.12 - .06N$	.99	.76
1.4	2.0	$\hat{P}n = 53.86 - .09N$	.96	1.08

Table 7.3.1 continued

β_k	μ_k	β_R=1.5 μ_D=1.5 (k=2 w=3.234)	R^2	CV(%)
1.0	0.0	Pn = 100.0 - 0.0N	1.00	----
0.6	2.0	$\hat{P}n$ = 24.77 - 0.0N	.97	1.22
1.4	2.0	$\hat{P}n$ = 47.89 - .06N	.96	2.14

Table 7.3.1 gives the estimated optimal sample allocation as a function of N for three different priors and three different degrees of parameter misspecification. The R^2 values given are a measure of fit for the fitted second order regression model, and the CV calculated using pure (replication) error is a measure of simulation error.

When the first stage design resembles the D optimal design and parameter knowledge is good, the optimal two-stage design has all of the sample in the first stage, thus reverting back to a one-stage design. If either or both the first stage design diverges from the optimal two-point design to a three or four point design or parameter knowledge is poor, more of the sample is put in the second stage. In addition, sample proportion decreases as sample size increases. A trade-off exists between trying to maintain the complexion of the optimal two-point design for the entire two-stage design and coping with poor parameter knowledge in the first stage. The first stage requires sufficient sample size to estimate the parameters well, unless the first stage design is so good that creating a second stage based upon estimates is counterproductive. When the initial design deviates sufficiently from the optimal two point design, the optimal allocation line is quite consistent for a wide variety of parameter knowledge.

Two of the three prior situations have relative agreement across the range of parameter knowledge. The exception to this is when the first stage design resembles the optimal design. While they are in disagreement, little efficiency is lost when parameter knowledge is good and one fails to place the entire sample in

the first stage. The same can be said of the situation when the optimal value of Pn is a constant 24.77. Consequently, the recommendation for the last scenario of table 7.3.1 is to use the predicted line of maximum for the β_k=1.4 μ_k=2.0 case.

The response surface surrounding the location of the lines is reasonably flat, making the linear equations of table 7.3.1 quite useful as a guide to approximate the optimal first stage sample size for a nearly endless number of possible situations. Because these surfaces are reasonably flat and consistent over different priors, generalizations are easily made. For a small total sample size of 30, one should use somewhere between 40 and 50% of the sample in the first stage. This generalization certainly holds for priors yielding two, three, and four point first stage designs and most likely holds for even wider priors. For a large sample size of 300, the recommendation ranges from 15 to 30% depending upon the situation.

CHAPTER 8

ASYMPTOTIC TWO-STAGE DESIGN PERFORMANCE

The performance of the two-stage design procedure is studied using tables similar to those of chapter 6. Comparisons are made between various one and two-stage designs. In addition the effect of various priors is investigated through comparisons among different two-stage designs. The effectiveness of using Bayesian estimation in the first stage is also explored. The three priors used in section 7.3 cover a wide range of parameter knowledge and therefore are used to study two-stage performance. A sample size of 300 is used in the study with the allocation between stages coming from section 7.3. The D efficiencies as defined in chapter 7 are used.

These D efficiencies are based upon asymptotic variances and are therefore asymptotic themselves. With a sample size of 300, the asymptotic efficiencies should be representative of the actual efficiencies if the true variances were known. However, for a smaller sample size, the asymptotic efficiencies are likely to be misleading since the asymptotic variances severely underestimate the actual variances. Small sample properties are investigated in chapter 9.

§8.1 One-stage to Two-stage Comparison

Using the D efficiency defined by expression 7.2.3, the following tables compare various two-stage D-D optimal designs with the one-stage D optimal design. The prior for the one-stage design consists of a single point which coincides with the two-stage prior mean. Efficiencies exceeding one imply that the two-stage D-D optimal design is better than the one-stage D optimal design for that given situation. Each efficiency value is based upon 100 simulation runs.

Table 8.1.1 Two-stage D-D, β_R=1.5 μ_D=1.5 joint normal prior, Pn=30% vs one-stage D optimal, N=300

		β_k					
		0.6	0.8	1.0	1.2	1.4	1.6
	0.0	1.18	1.01	0.98	1.00	1.05	1.11
	0.5	1.20	1.05	1.03	1.07	1.14	1.19
μ_k	**1.0**	1.26	1.18	1.21	1.29	1.41	1.54
	1.5	1.44	1.48	1.63	1.87	2.11	2.26
	2.0	1.77	2.14	2.65	3.16	3.38	3.64
	2.5	2.65	3.65	5.00	5.69	6.30	6.38

standard error = .006

Table 8.1.2 Two-stage D-D, β_R=2.0 μ_D=2.5 joint normal prior, Pn=20% vs one-stage D optimal, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.11	0.92	0.92	0.96	1.04	1.14
	0.5	1.16	0.99	0.99	1.05	1.13	1.21
μ_k	**1.0**	1.30	1.18	1.21	1.32	1.46	1.62
	1.5	1.61	1.56	1.69	1.93	2.19	2.35
	2.0	2.10	2.34	2.81	3.36	3.62	4.12
	2.5	3.20	3.95	5.19	6.15	7.20	7.63

standard error = .028

Table 8.1.3 Two-stage D-D, β_R=3.5 μ_D=3.0 joint normal prior, Pn=20% vs one-stage D optimal, N=300

		β_k					
		0.6	0.8	1.0	1.2	1.4	1.6
μ_k	0.0	1.26	0.91	0.87	0.92	1.02	1.16
	0.5	1.27	0.97	0.94	1.01	1.12	1.25
	1.0	1.38	1.11	1.16	1.26	1.42	1.68
	1.5	1.66	1.50	1.62	1.84	2.22	2.55
	2.0	2.19	2.29	2.67	3.28	3.76	4.41
	2.5	3.35	3.95	4.94	6.00	7.40	8.63

standard error = .020

Table 8.1.1 shows that this particular four point two-stage design, two levels in each stage, is considerably more efficient than the D optimal two point design except when parameter knowledge is near perfect. The use of the two-stage design becomes more advantageous as parameter knowledge worsens. The same can be seen from tables 8.1.2 and 8.1.3. From 8.1.2, the five point two-stage design succumbs to the one-stage D optimal design in a small region about the perfect guesses but outperforms it as parameter knowledge diminishes. The six point design in table 8.1.3 exemplifies the pattern to an even greater degree.

In fact, at the most extreme situation of parameter misspecification in table 8.1.3, the determinant of the total information matrix from the two-stage D-D design is 8.63 times larger than the determinant from the D optimal design. However, when the guesses are correct the two-stage determinant is 13% smaller than that from the one-stage design. Consequently, much efficiency is gained when the parameters are misspecified, but little is lost if they are not.

The next set of tables, 8.1.4, 8.1.5, and 8.1.6 compare the same three two-stage D-D optimal designs to the three point one-stage D optimal design. The D efficiency defined by expression 7.2.3 is used once again. D efficiencies greater than one imply the two-stage design is better than the three point design.

Table 8.1.4 Two-stage D-D, β_R=1.5 μ_D=1.5 joint normal prior, Pn=30% vs three point one-stage D optimal, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	1.14	1.15	1.14	1.14	1.19	1.24
	0.5	1.10	1.14	1.16	1.19	1.26	1.31
	1.0	1.03	1.16	1.26	1.38	1.51	1.64
	1.5	0.98	1.27	1.55	1.83	2.05	2.26
	2.0	1.04	1.60	2.22	2.82	3.23	3.44
	2.5	1.26	2.35	3.81	4.93	5.73	5.67

standard error = .004

Table 8.1.5 Two-stage D-D, β_R=2.0 μ_D=2.5 joint normal prior, Pn=20% vs three point one-stage D optimal, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	1.07	1.05	1.07	1.10	1.17	1.27
	0.5	1.07	1.08	1.11	1.17	1.24	1.33
	1.0	1.07	1.16	1.26	1.41	1.56	1.72
	1.5	1.10	1.33	1.60	1.89	2.13	2.35
	2.0	1.23	1.74	2.35	3.00	3.45	3.89
	2.5	1.52	2.54	3.95	5.33	6.55	6.78

standard error = .045

Table 8.1.6 Two-stage D-D, β_R=3.5 μ_D=3.0 joint normal prior, Pn=20% vs three point one-stage D optimal, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.22	1.04	1.01	1.05	1.16	1.30
	0.5	1.17	1.05	1.05	1.13	1.23	1.38
μ_k	**1.0**	1.13	1.09	1.21	1.34	1.53	1.79
	1.5	1.13	1.29	1.53	1.80	2.16	2.55
	2.0	1.28	1.70	2.24	2.93	3.59	4.17
	2.5	1.60	2.55	3.76	5.20	6.73	7.67

standard error = .012

All three two-stage designs being studied are more efficient than the three point one-stage D optimal design almost uniformly over the parameter knowledge ranges given. The only exception occurs at a single table location with the four point two-stage design. Consequently, the specific four, five, and six point two-stage designs shown here are as a whole more efficient than the best three point one-stage design even when the parameter guesses are perfect. These results should obviously apply to all two-stage D-D optimal designs using first stage Bayesian normal prior designs which fall into the range used in the tables.

The next three tables compare the two-stage D-D optimal design to the corresponding one-stage D optimal design using the same prior. The one-stage design uses the prior only to specify the design while the two-stage uses the prior a second time for first stage parameter estimation. The D efficiencies used in tables 8.1.7, 8.1.8, and 8.1.9 are defined by expression 7.2.1. D efficiencies above one indicate that the two-stage design is better than the one-stage design.

Table 8.1.7 Two-stage D-D, Pn=30% vs. one-stage D optimal, both using β_R=1.5 μ_D=1.5 joint normal prior, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.31	1.06	0.98	0.98	1.01	1.05
	0.5	1.35	1.08	1.03	1.04	1.09	1.13
μ_k	**1.0**	1.40	1.21	1.21	1.26	1.34	1.45
	1.5	1.55	1.51	1.60	1.79	1.95	2.12
	2.0	1.90	2.14	2.56	2.93	3.23	3.44
	2.5	2.65	3.65	4.71	5.29	6.30	5.67

standard error = .006

Table 8.1.8 Two-stage D-D, Pn=20% vs. one-stage D optimal, both using β_R=2.0 μ_D=2.5 joint normal prior, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.66	1.32	1.15	1.06	1.04	1.05
	0.5	1.60	1.31	1.16	1.10	1.09	1.08
μ_k	**1.0**	1.51	1.32	1.25	1.23	1.26	1.33
	1.5	1.43	1.38	1.40	1.53	1.62	1.70
	2.0	1.44	1.55	1.85	2.15	2.38	2.69
	2.5	1.52	1.98	2.68	3.48	4.24	4.36

standard error = .031

Table 8.1.9 Two-stage D-D, Pn=20% vs. one-stage D optimal, both using β_R=3.5 μ_D=3.0 joint normal prior, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.95	1.46	1.21	1.06	1.02	1.01
	0.5	1.94	1.51	1.23	1.11	1.05	1.05
μ_k	**1.0**	2.03	1.53	1.32	1.19	1.17	1.24
	1.5	2.13	1.62	1.42	1.38	1.44	1.55
	2.0	2.19	1.74	1.66	1.78	1.98	2.21
	2.5	2.09	1.88	2.03	2.52	2.96	3.63

standard error = .013

The one-stage Bayesian designs were found in chapter 6 to be more robust than the standard one-stage designs. However, they were quite inefficient. Tables 8.1.6 to 8.1.9 show that the addition of the second stage in the two-stage designs offer a distinct advantage over the corresponding one-stage Bayesian design. In fact, the two-stage designs are uniformly more efficient with exception to a small region of parameter specification in table 8.1.7.

The two-stage designs are nearly as efficient as the best design when parameter knowledge is good while being considerably more efficient than the so called robust one-stage designs when parameter knowledge is poor. The two-stage D-D optimal design using Bayesian design and estimation in the first stage has effectively separated and addressed the goals of robustness and efficiency.

§8.2 Two-Stage to Two-Stage Comparison

A comparison among the three two-stage designs is examined in the next three tables. The tables show how the three designs relate to each other and also show the trend of increasing robustness to parameter misspecification as the prior widens. The D efficiency given by expression 7.2.4 is used in tables 8.2.1, 8.2.2, and 8.2.3. Efficiencies exceeding one indicate the first two-stage design is more efficient than the second.

Table 8.2.1 Two-stage D-D, joint normal priors, β_R=2.0 μ_D=2.5, Pn=20% vs β_R=1.5 μ_D=1.5, Pn=30%, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	0.94	0.91	0.94	0.96	0.99	1.02
	0.5	0.97	0.94	0.96	0.98	0.99	1.01
μ_k	**1.0**	1.03	1.00	1.00	1.02	1.04	1.05
	1.5	1.12	1.05	1.04	1.04	1.04	1.04
	2.0	1.18	1.09	1.06	1.06	1.07	1.13
	2.5	1.21	1.08	1.04	1.08	1.14	1.20

Table 8.2.2 Two-stage D-D, joint normal priors, β_R=3.5 μ_D=3.0, Pn=20% vs β_R=1.5 μ_D=1.5, Pn=30%, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
	0.0	1.07	0.90	0.89	0.92	0.98	1.05
	0.5	1.06	0.92	0.91	0.95	0.98	1.05
μ_k	**1.0**	1.10	0.94	0.96	0.98	1.01	1.09
	1.5	1.15	1.01	0.99	0.99	1.05	1.13
	2.0	1.24	1.07	1.01	1.04	1.11	1.21
	2.5	1.26	1.08	0.99	1.05	1.17	1.35

Table 8.2.3 Two-stage D-D, joint normal priors, β_R=3.5 μ_D=? vs β_R=2.0 μ_D=2.5, Pn=20%, N=300

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	
μ_k	**0.0**	1.14	0.99	0.95	0.96	0.99	1.02
	0.5	1.09	0.98	0.95	0.97	0.99	1.04
	1.0	1.06	0.94	0.96	0.96	0.98	1.04
	1.5	1.03	0.96	0.95	0.95	1.01	1.08
	2.0	1.05	0.98	0.95	0.98	1.04	1.07
	2.5	1.05	1.00	0.95	0.98	1.03	1.13

Tables 8.2.1, 8.2.2, and 8.2.3 show that relatively little difference exists among the three two-stage designs. The wider priors do cause the design to be slightly more robust at the extremes of the table, but they also cause the designs to be less efficient in the center. These results also show that the choice of prior is not critical to the performance of the design. The properties of the design changes slowly with the prior. Nevertheless, a recommendation can be made on the basis of these tables for those who cannot determine a reasonable prior for their situation but rather only a single point guess. In most situations, the five and six point two-stage designs would be more desirable than the four point designs. However, it should be noted that more gain in robustness is made going from the four point to five point than from the five to the six. Consequently, the five point two-stage D-D optimal designs offer an attractive compromise between being highly efficient and robust.

In the next set of tables, the usefulness of the Bayesian estimation is studied. We have seen the advantage of the two-stage Bayesian designs which use Bayesian estimation in the first stage over the one-stage designs. Tables 8.2.4, 8.2.5, and 8.2.6 compare two-stage Bayesian D optimal designs using Bayesian estimation with the same two-stage Bayesian design using maximum likelihood estimation. The D efficiency to be used is defined by expression 7.2.5. D efficiencies exceeding one imply that first stage Bayesian estimation is better than maximum likelihood estimation.

Table 8.2.4 Two-stage D-D, joint normal prior, β_R=1.5 μ_D=1.5, Pn=30% N=300, First stage Bayesian estimation vs. maximum likelihood

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	0.92	1.03	1.07	1.06	1.02	0.98
	0.5	0.96	1.02	1.09	1.08	1.02	0.98
	1.0	1.00	1.12	1.12	1.09	1.06	1.03
	1.5	1.04	1.29	1.23	1.20	1.13	1.03
	2.0	1.10	1.56	1.52	1.46	1.34	1.68
	2.5	1.18	1.66	1.86	1.85	1.50	1.55

Table 8.2.5 Two-stage D-D, joint normal prior, β_R=2.0 μ_D=2.5, Pn=20% N=300, First stage Bayesian estimation vs. maximum likelihood

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	1.40	1.22	1.15	1.12	1.09	1.11
	0.5	1.33	1.27	1.21	1.12	1.07	1.13
	1.0	1.18	1.19	1.12	1.17	1.18	1.13
	1.5	1.14	1.18	1.16	1.14	1.21	1.18
	2.0	1.07	1.24	1.25	1.31	1.33	1.27
	2.5	1.03	1.52	2.02	1.95	1.76	1.79

Table 8.2.6 Two-stage D-D, joint normal prior, β_R=3.5 μ_D=3
N=300, First stage Bayesian estimation vs. max

		β_k					
		0.6	**0.8**	**1.0**	**1.2**	**1.4**	**1.6**
μ_k	**0.0**	1.31	1.14	1.19	1.06	1.06	1.10
	0.5	1.32	1.17	1.14	1.10	1.06	1.01
	1.0	1.19	1.23	1.18	1.12	1.05	1.02
	1.5	1.48	1.27	1.17	1.15	1.14	1.08
	2.0	1.42	1.16	1.15	1.30	1.20	1.29
	2.5	1.46	1.32	1.16	1.32	1.42	1.68

From tables 8.2.4, 8.2.5, and 8.2.6 it is clear that the Bayesian estimation method is an important addition to the two-stage procedure in terms of increasing the expected efficiency. The patterns within these tables, however, is not as uniform as was seen in earlier tables. This result is due to an increase in the variability of the determinant of the two-stage information matrix when Bayesian estimation is not used.

The efficiencies used in this chapter represent asymptotic results for a sample size of 300. Indeed, these results indicate that the two-stage D-D optimal design using Bayesian design and estimation in the first stage is asymptotically superior to the one-stage D optimal designs. However, the question remains as to the actual performance of the two-stage designs for small samples as well as a sample size of 300.

CHAPTER 9

SMALL SAMPLE ONE AND TWO-STAGE DESIGN PERFORMANCE

In chapter 8, the performance of the two-stage design procedure is evaluated using asymptotic efficiencies. This chapter addresses the question of whether these asymptotic efficiencies are indeed appropriate, even for a sample size as large as 300. Furthermore, the sample size at which the asymptotic properties begin to fail is investigated. Simulated variances of parameter estimates and efficiencies are used. Using simulation, the five point two-stage D-D optimal design specified by β_R=2.0 μ_D=2.5, normal prior, is compared to the one-stage D optimal design, $ED_{17.6}$ and $ED_{82.4}$. The comparison is made for three different levels of parameter misspecification over a wide range of sample sizes. Both tables and plots are used to illustrate the study.

§9.1 Correct Parameter Specification

The asymptotic variances of parameter estimates given by the inverted information matrix are based upon an approximation to order $\frac{1}{n}$. Using methods provided by Shenton and Bowman (1977), Kalish (1990) gives first and second order variance expressions for a one-stage equal allocation two point symmetric design. The second order expression (order $\frac{1}{n^2}$) for the variance of the slope estimate is

$$\mathrm{Var}(\widehat{\beta}) \simeq \frac{\beta^2}{nPQ\lambda^2} + \frac{\beta^2(4P^2 - 3PQ + 4Q^2)}{(nPQ)^2\lambda^2} \qquad (9.1.1)$$

where Q=1-P, n is the total sample size, and λ=Logit(P)=$\mathrm{Log}\left[\frac{P}{1-P}\right]$. The design has equal allocation to ED_{100P} and ED_{100Q}. The first and second terms of expression 9.1.1 represent the first and second order components in terms of the

inverse of n of the variance approximation respectfully. An absolutely scale invariant expression of the slope variance is obtained by dividing expression 9.1.1 by β^2. The second order variance expression for $\widehat{\mu}$ is given by

$$\mathrm{Var}(\widehat{\mu}) \simeq \frac{1}{\beta^2 \mathrm{nPQ}} + \frac{(3\mathrm{P}^2 - \mathrm{PQ} + 3\mathrm{Q}^2)\lambda^2 - 6(\mathrm{P\text{-}Q})\lambda + 3}{\beta^2(\mathrm{nPQ})^2\lambda^2}. \qquad (9.1.2)$$

As one might expect, the absolutely scale invariant expression of the variance of $\widehat{\mu}$ is obtained by multiplying expression 9.1.2 by β^2.

In the case of the two point one-stage D optimal design and correct parameter specification, P=.176, Q=.824, and λ=-1.544. From 9.1.1 and 9.1.2, these values for P, Q, and λ result in the following second order absolutely scale invariant variance approximations:

$$\mathrm{Var}(\widehat{\beta}) \simeq \frac{2.893}{\mathrm{n}} + \frac{46.725}{\mathrm{n}^2} \qquad (9.1.3)$$

and

$$\mathrm{Var}(\widehat{\mu}) \simeq \frac{6.895}{\mathrm{n}} + \frac{34.475}{\mathrm{n}^2}. \qquad (9.1.4)$$

Expressions 9.1.3 and 9.1.4 are then used to create tables and plots of the first and second order asymptotic variances as a function of sample size. The first order variance comes from the first term of each of the two expressions 9.1.3 and 9.1.4.

Table 9.1.1 gives the first and second order asymptotic approximations for the variance of $\widehat{\beta}$ from the one-stage design for a variety of sample sizes when β_{k}=1.0 and μ_{k}=0.0 (no parameter misspecification). In addition two simulated variance estimates of $\widehat{\beta}$ are given, one for the one-stage D optimal design and the other for the five point two-stage β_{R}=2.5 and μ_{D}=2.0 D-D optimal design. Standard error of estimates are also offered and are denoted in small print. Simulation results for the each of the one-stage variances were obtained from 3 sets of 20,000 runs each while the two-stage variances were based on at least 3 sets of 1000.

Table 9.1.1 Scale invariant variance of $\widehat{\beta}$ as a function of n, β_k=1.0 and μ_k=0.0

			simulated	
	1st order	2nd order	one-stage	two-stage
n	A.Var($\widehat{\beta}$)	A.Var($\widehat{\beta}$)	Var($\widehat{\beta}$)	Var($\widehat{\beta}$)
300	.0096	.0102	.0101(.0001)	.0104(.0006)
200	.0145	.0157	.0156(.0001)	.0167(.0002)
150	.0193	.0214	.0215(.0001)	.0220(.0005)
100	.0283	.0336	.0408(.0022)	.0365(.0008)
80	.0362	.0435	.0945(.0058)	.0480(.0013)
60	.0482	.0612	.4985(.0197)	.1249(.0441)

Table 9.1.1 indicates that the use of the asymptotic (1st order) variance of $\widehat{\beta}$ is appropriate for a sample size of 300 when the parameters are correctly specified. While the second order asymptotic expression follows the actual variances more closely than the first order expression, its addition adds relatively little to the asymptotic variance expression until the sample size is 100 or below. However, for small sample sizes, even with a substantial contribution from the second order term, the asymptotic variance expression is a gross underestimate of the true variance and is therefore inappropriate.

The estimated variance in the case of the one-stage D optimal design is well represented by the asymptotic variances until sample size falls below 100. When sample size is 60, the one-stage variance is many times greater than that given by the asymptotic expression. The two-stage design, however, does not substantially depart from the asymptotic values until $n \leq 80$, and when n=60, the two-stage variance, though still several times greater than the asymptotic value, is considerably less than that obtained from the one-stage design. One should keep in mind that under no parameter misspecification the first order asymptotic variance serves as a lower bound for both the one and two-stage designs.

Figure 9.1.1 offers a graphical representation of table 9.1.1. Again, the variances are presented in a scale invariant form. From the plot and table it becomes very clear that the variance of $\widehat{\beta}$ for the one-stage design departs from the asymptotic variance at a higher sample size and to a much greater degree than does the variance from the two-stage.

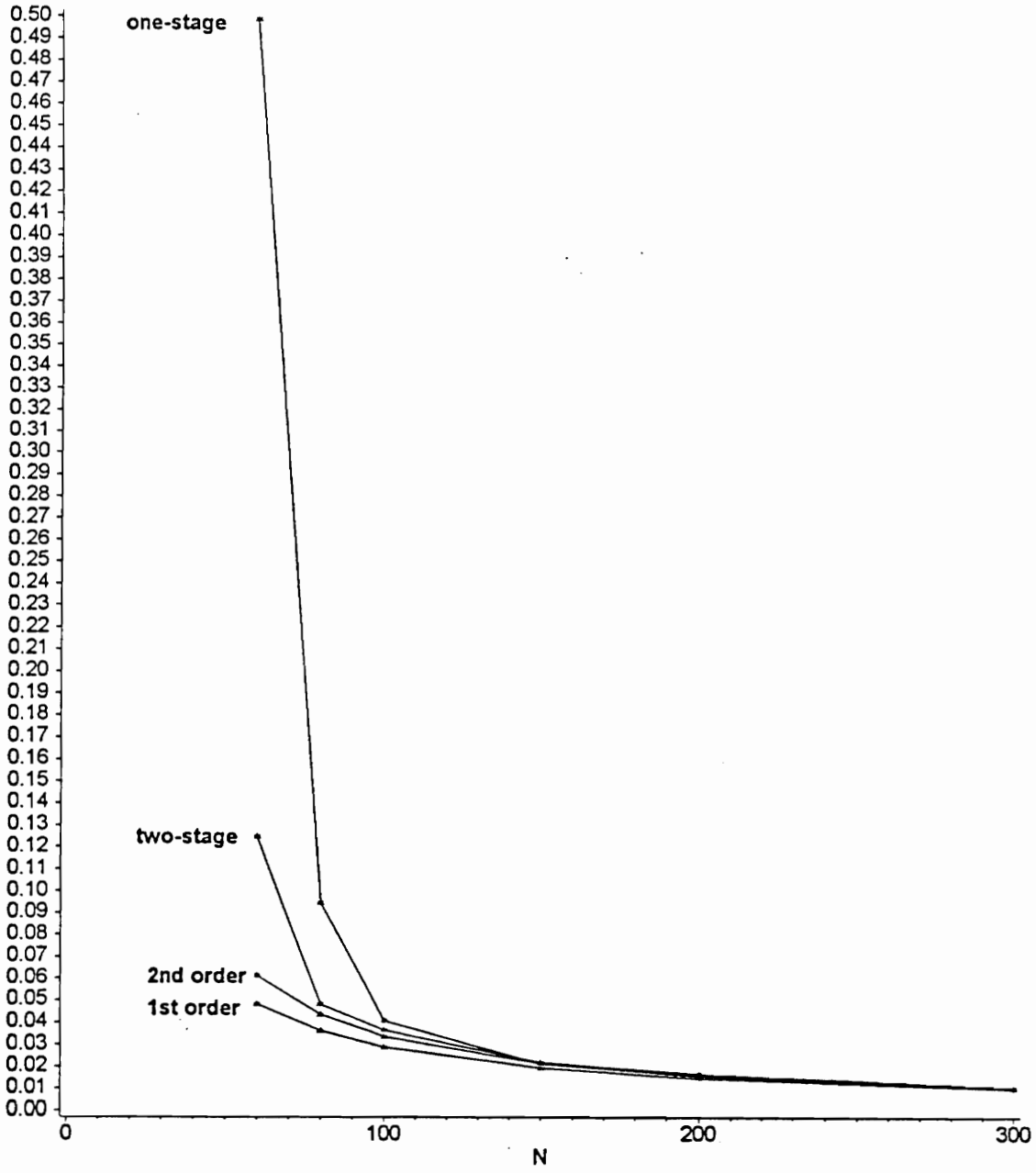

Figure 9.1.1 Small sample and asymptotic estimates of variance of $\widehat{\beta}$ as a function of design and N when β_k=1.0 and μ_k=0.0.

In the same format as table 9.1.1, table 9.1.2 contains the first and second order asymptotic variance approximations of $\hat{\mu}$, again assuming β_k=1.0 and μ_k=0.0 and the one-stage design. Likewise, the simulated variances of $\hat{\mu}$ obtained from the one-stage D optimal design and the two-stage, β_R=2.5 and μ_D=2.0, D-D optimal design are also included.

Table 9.1.2 Scale invariant variance of $\hat{\mu}$ as a function of n, β_k=1.0 and μ_k=0.0

			simulated	
	1st order	2nd order	one-stage	two-stage
n	A.Var($\hat{\mu}$)	A.Var($\hat{\mu}$)	Var($\hat{\mu}$)	Var($\hat{\mu}$)
300	.0230	.0234	.0234(.0003)	.0233(.0003)
200	.0345	.0354	.0352(.0003)	.0350(.0003)
150	.0460	.0475	.0473(.0001)	.0479(.0006)
100	.0690	.0724	.0721(.0003)	.0707(.0017)
80	.0862	.0916	.0916(.0007)	.0868(.0026)
60	.1149	.1245	.1314(.0006)	.1214(.0022)

While the appropriateness of the asymptotic variance of $\hat{\beta}$ fell into question for small sample sizes in table 9.1.1, the same does not hold true for the asymptotic variance of $\hat{\mu}$, according to table 9.2.1. For the sample size of 300, the first order asymptotic expression is quite representative of the actual variance. In fact, the first order approximation is appropriate for sample sizes as low as 60.

When the parameters are correctly specified, it appears that any problem with the use of asymptotic approximations in small sample situations arises from the estimation of β and not μ. The estimation of μ remains relatively stable for small sample sizes for either the one or two-stage design. This stability allows the asymptotic variance to closely approximate the actual variance. This result is in sharp contrast to the estimation of β which becomes extremely unstable for small sample sizes.

Figure 9.1.2 graphically describes the information displayed in table 9.1.2. Little departure from the asymptotic approximation is apparent for either design. Notice that the scales of figures 9.1.1 and 9.1.2 are vastly different.

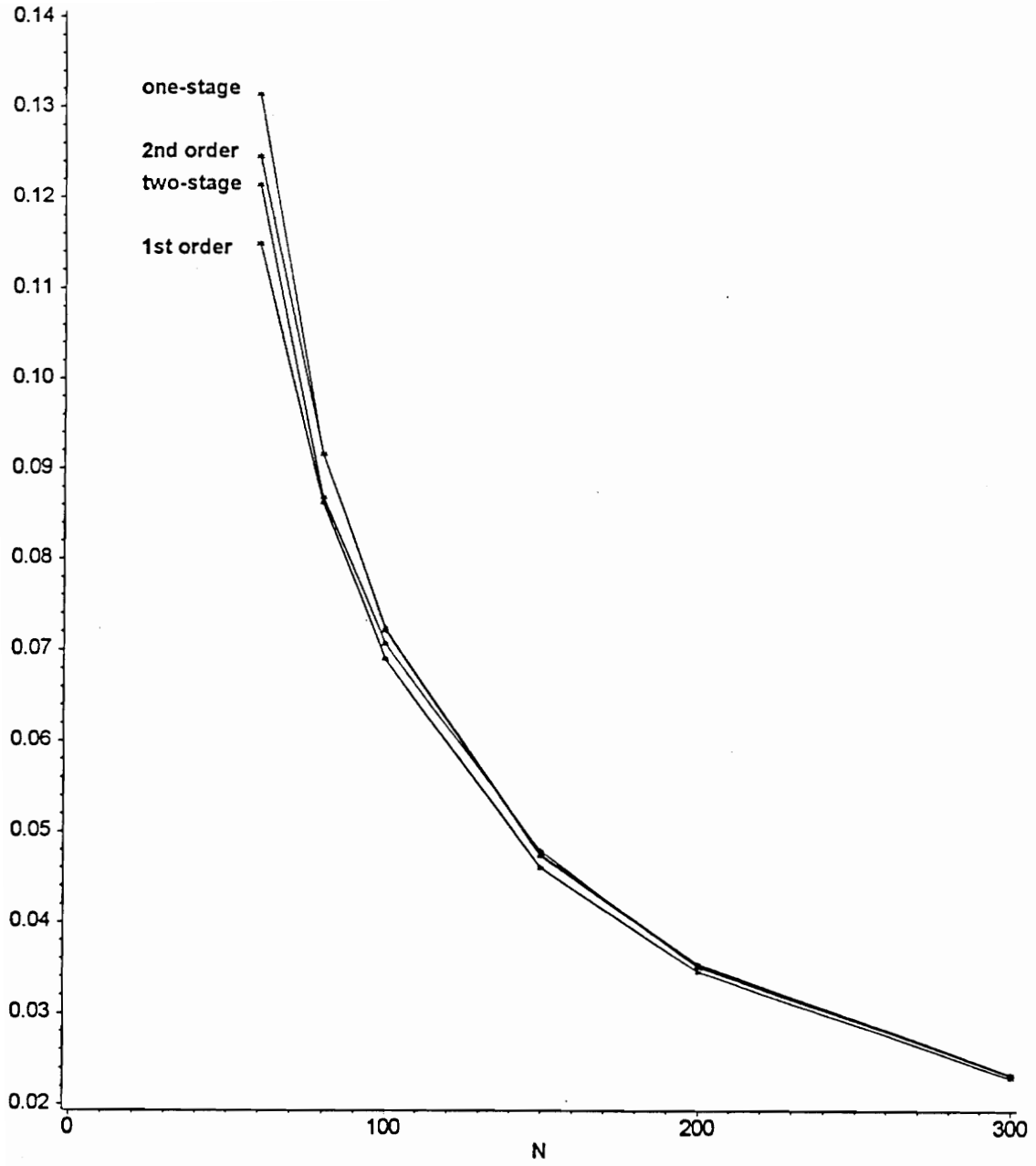

Figure 9.1.2 Small sample and asymptotic estimates of variance of $\widehat{\mu}$ as a function of design and N when β_k=1.0 and μ_k=0.0.

§9.2 Parameter Misspecification

Now suppose the model parameters are no longer correctly specified. Table 9.2.1 contains the first order asymptotic variance approximations of $\hat{\beta}$ when the parameter misspecification is of degree β_k=0.8 and μ_k=1.5. The simple asymptotic expressions given by 9.1.3 and 9.1.4 no longer hold. Consequently, no second order approximation is given, and the scale invariant first order approximation comes from expression 3.1.1. The asymptotic variances are for the parameter misspecified one-stage design. In addition, estimated actual variances of $\hat{\beta}$ are presented for the D optimal one-stage and two-stage, β_R=2.5 and μ_D=2.0, D-D optimal design under parameter misspecification.

The maximum likelihood estimates of β and μ are biased for finite sample sizes. When the parameters are correctly specified, the bias was found to be negligible for nearly all situations listed in tables 9.1.1 and 9.1.2. Consequently, the bias was not tabled in those instances. However, when the parameters are misspecified the bias is often not negligible and can therefore play an important role in determining the appropriateness of the asymptotic variance. The simulated scale invariant bias defined by $\hat{E}\left(\frac{\hat{\beta}-\beta}{\beta}\right)$ is given in table 9.2.1.

Table 9.2.1 Scale invariant variance of $\hat{\beta}$ as a function of n, β_k=0.8 and μ_k=1.5

		one-stage		two-stage	
n	**A. Var($\hat{\beta}$)**	**Var($\hat{\beta}$)**	**Bias($\hat{\beta}$)**	**Var($\hat{\beta}$)**	**Bias($\hat{\beta}$)**
300	.0167	.3697(.0090)	.0844	.0097(.0001)	.0158
200	.0249	1.649(.0205)	.2927	.0150(.0001)	.0194
150	.0333	3.391(.0349)	.3281	.0209(.0004)	.0253
100	.0500	6.682(.0424)	1.281	.0375(.0028)	.0455
80	.0625	8.343(.0210)	1.740	1.142(1.360)	.0798
60	.0833	9.943(.0427)	2.369	4.565(1.592)	.1759

Table 9.2.1 indicates that the asymptotic variance of $\hat{\beta}$ severely underestimates the true variance for the one-stage design for a sample size as large as 300. Although, the asymptotic variance in table 9.2.1 no longer represents a lower bound for the two-stage design, it is still serves as a useful basis of comparison.

The bias of $\hat{\beta}$ offers further evidence in favor of the two-stage design. Both the variance and bias of the two-stage design are uniformly less than the variance and bias for the one-stage design over the sample range 60-300. For small sample sizes, the values are substantially smaller for the two-stage.

For the same degree of parameter misspecification as table 9.2.1, β_k=0.8 and μ_k=1.5, table 9.2.2 investigates the variance of $\hat{\mu}$. The simulated scale invariant bias of $\hat{\mu}$ defined by $\hat{E}(\beta(\hat{\mu}-\mu))$ is also provided.

Table 9.2.2 Scale invariant variance of $\hat{\mu}$ as a function of n, β_k=0.8 and μ_k=1.5

		one-stage		two-stage	
n	**A. Var($\hat{\mu}$)**	**Var($\hat{\mu}$)**	**Bias($\hat{\mu}$)**	**Var($\hat{\mu}$)**	**Bias($\hat{\mu}$)**
300	.0248	.0260(.0003)	-.0116	.0284(.0007)	-.0004
200	.0372	.0416(.0002)	-.0193	.0434(.0010)	.0050
150	.0496	.0578(.0002)	-.0382	.0613(.0020)	-.0040
100	.0744	.0888(.0004)	-.0711	.0913(.0013)	-.0070
80	.0930	.1096(.0004)	.0926	.1142(.0026)	-.0053
60	.1240	.1422(.0002)	.1238	.1745(.0033)	-.0065

The actual variance estimate of $\hat{\mu}$ is well represented by the asymptotic variance for the one and two-stage designs for all sample sizes tabled except for 60. Interestingly, the simulated variance of $\hat{\mu}$ from the one-stage procedure is uniformly less than that obtained from the two-stage procedure. However, the two-stage procedure remains relatively unbiased over the entire range of tabled sample sizes while the one-stage procedure yields an increasingly biased location estimate as sample size decreases.

Figures 9.2.1 and 9.2.2 graphically demonstrate some of the information contained in tables 9.2.1 and 9.2.2. Clearly for the parameter misspecification, β_k=0.8 and μ_k=1.5, the small sample properties of this two-stage design are considerably more attractive than those of the one-stage design when considering both variance and bias of the parameter estimates.

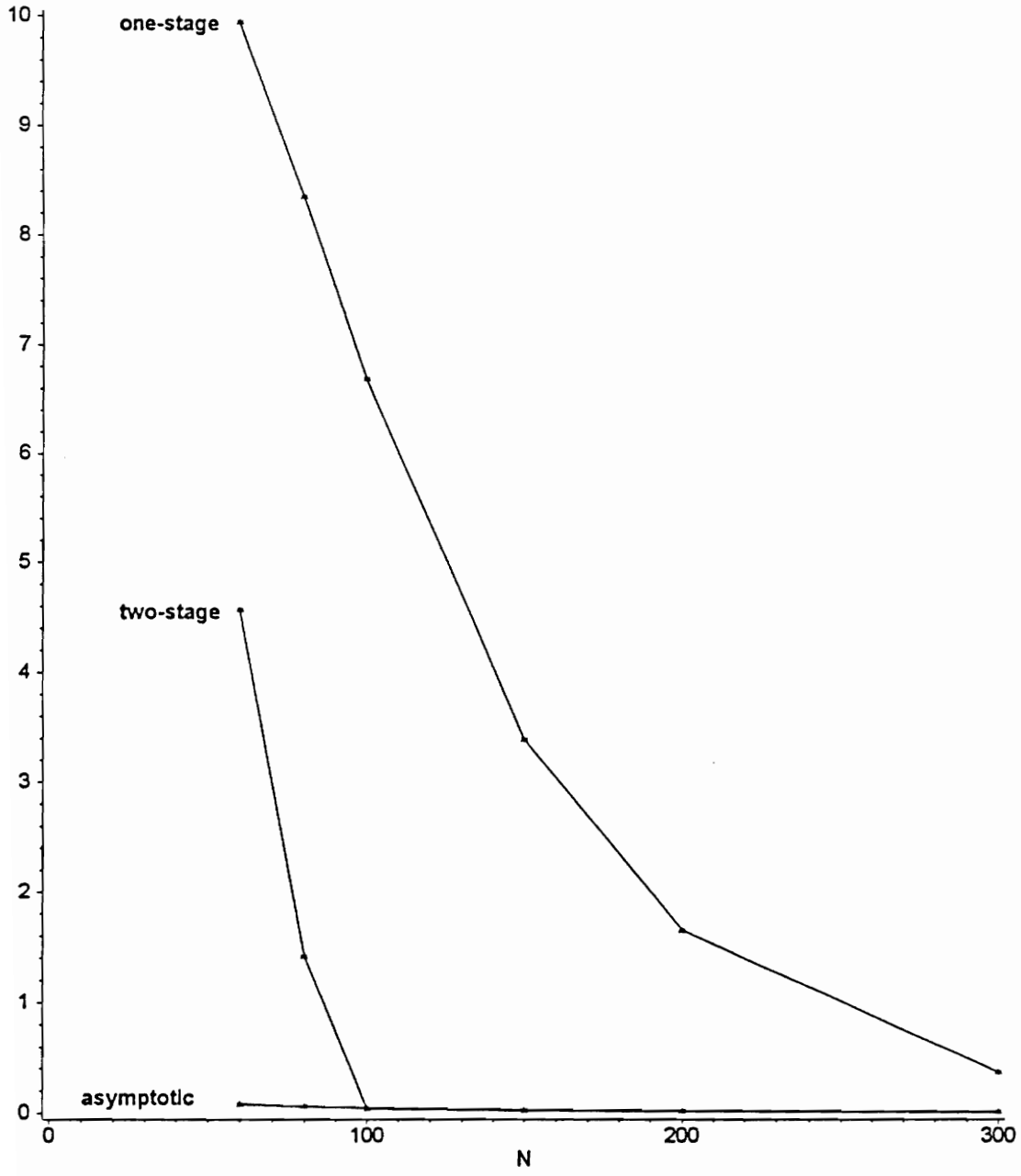

Figure 9.2.1 Small sample and asymptotic estimates of variance of $\widehat{\beta}$ as a function of design and N when β_k=0.8 and μ_k=1.5.

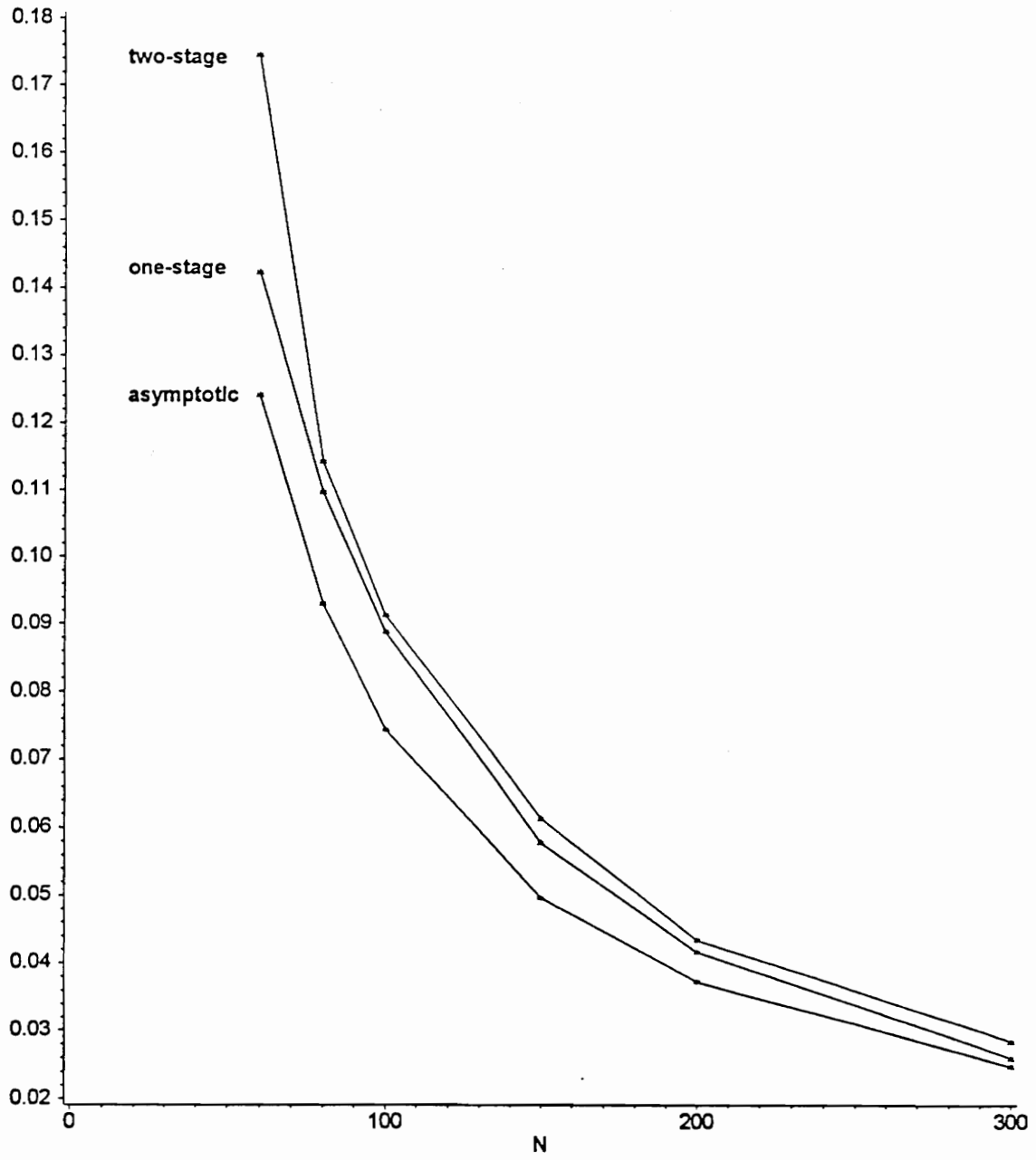

Figure 9.2.2 Small sample and asymptotic estimates of variance of $\hat{\mu}$ as a function of design and N when β_k=0.8 and μ_k=1.5.

The third and final level of parameter misspecification used in this small sample study is given by β_k=1.4 and μ_k=1.0. As before both tables and plots are used to display the results. Similar to tables 9.2.1 and 9.2.2, tables 9.2.3 and 9.2.4 contain variance and bias estimates for $\widehat{\beta}$ and $\widehat{\mu}$, respectively. In addition, figures 9.2.3 and 9.2.4 relay the same information graphically.

Table 9.2.3 Scale invariant variance of $\widehat{\beta}$ as a function of n, β_k=1.4 and μ_k=1.0

		one-stage		two-stage	
n	A. Var($\widehat{\beta}$)	Var($\widehat{\beta}$)	Bias($\widehat{\beta}$)	Var($\widehat{\beta}$)	Bias($\widehat{\beta}$)
300	.0196	.0211(.0001)	.0134	.0138(.0003)	.0209
200	.0294	.0328(.0002)	.0212	.0211(.0002)	.0257
150	.0392	.0767(.0031)	.0316	.0286(.0004)	.0336
100	.0588	.4465(.0302)	.0743	.0485(.0016)	.0592
80	.0736	1.240(.0157)	.1514	.0579(.0006)	.0577
60	.0981	3.763(.6680)	.3846	.0885(.0035)	.0803

Table 9.2.4 Scale invariant variance of $\widehat{\mu}$ as a function of n, β_k=1.4 and μ_k=1.0

		one-stage		two-stage	
n	A. Var($\widehat{\mu}$)	Var($\widehat{\mu}$)	Bias($\widehat{\mu}$)	Var($\widehat{\mu}$)	Bias($\widehat{\mu}$)
300	.0244	.0264(.0002)	-.0130	.0208(.0009)	.0016
200	.0366	.0410(.0001)	-.0170	.0304(.0010)	-.0018
150	.0489	.0575(.0003)	-.0230	.0424(.0012)	-.0010
100	.0733	.1003(.0002)	-.0375	.0613(.0012)	-.0110
80	.0916	.1520(.0095)	-.0504	.0752(.0008)	-.0067
60	.1221	>>>	>>>	.1044(.0011)	-.0180

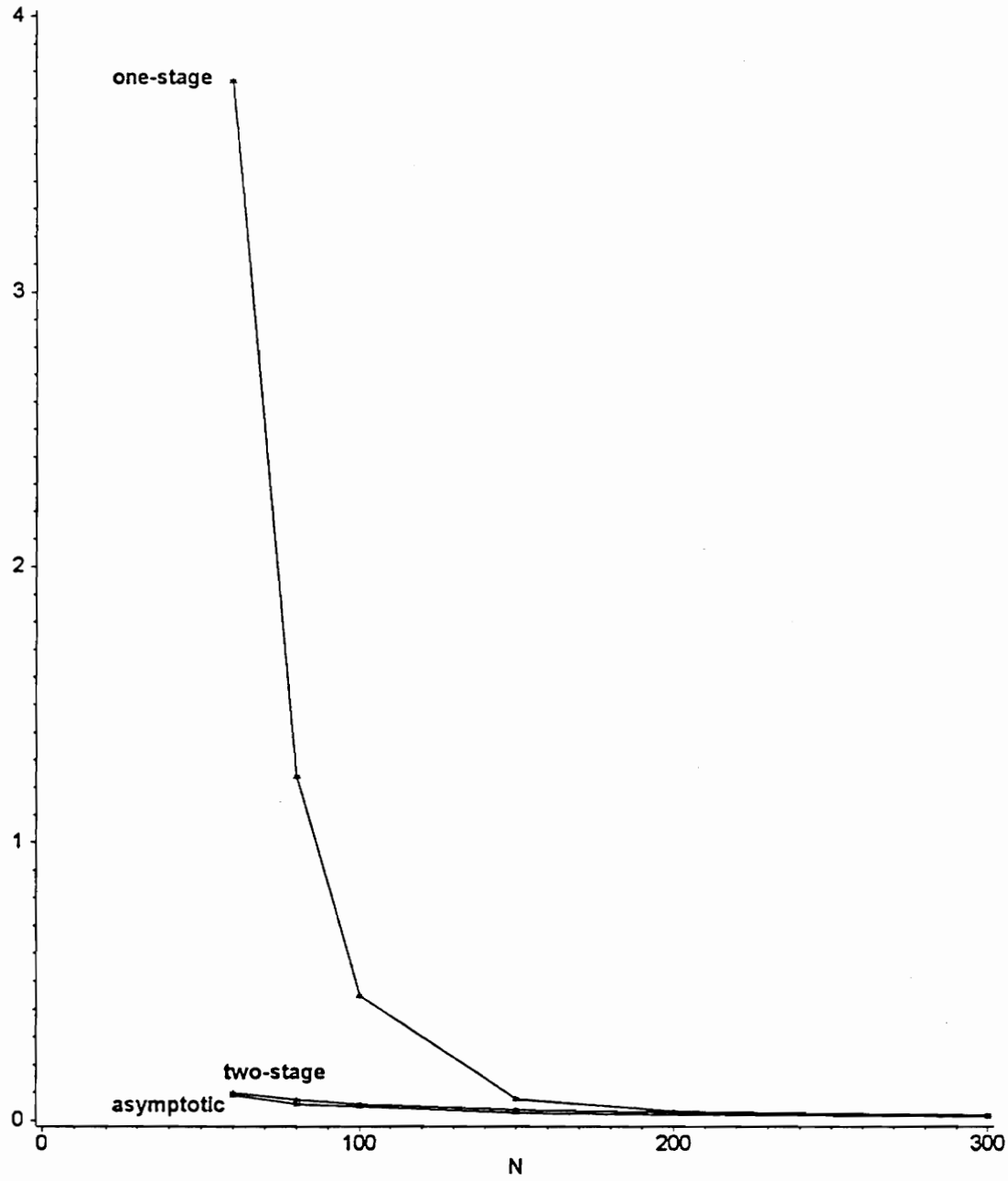

Figure 9.2.3 Small sample and asymptotic estimates of variance of $\widehat{\beta}$ as a function of design and N when β_k=1.4 and μ_k=1.0.

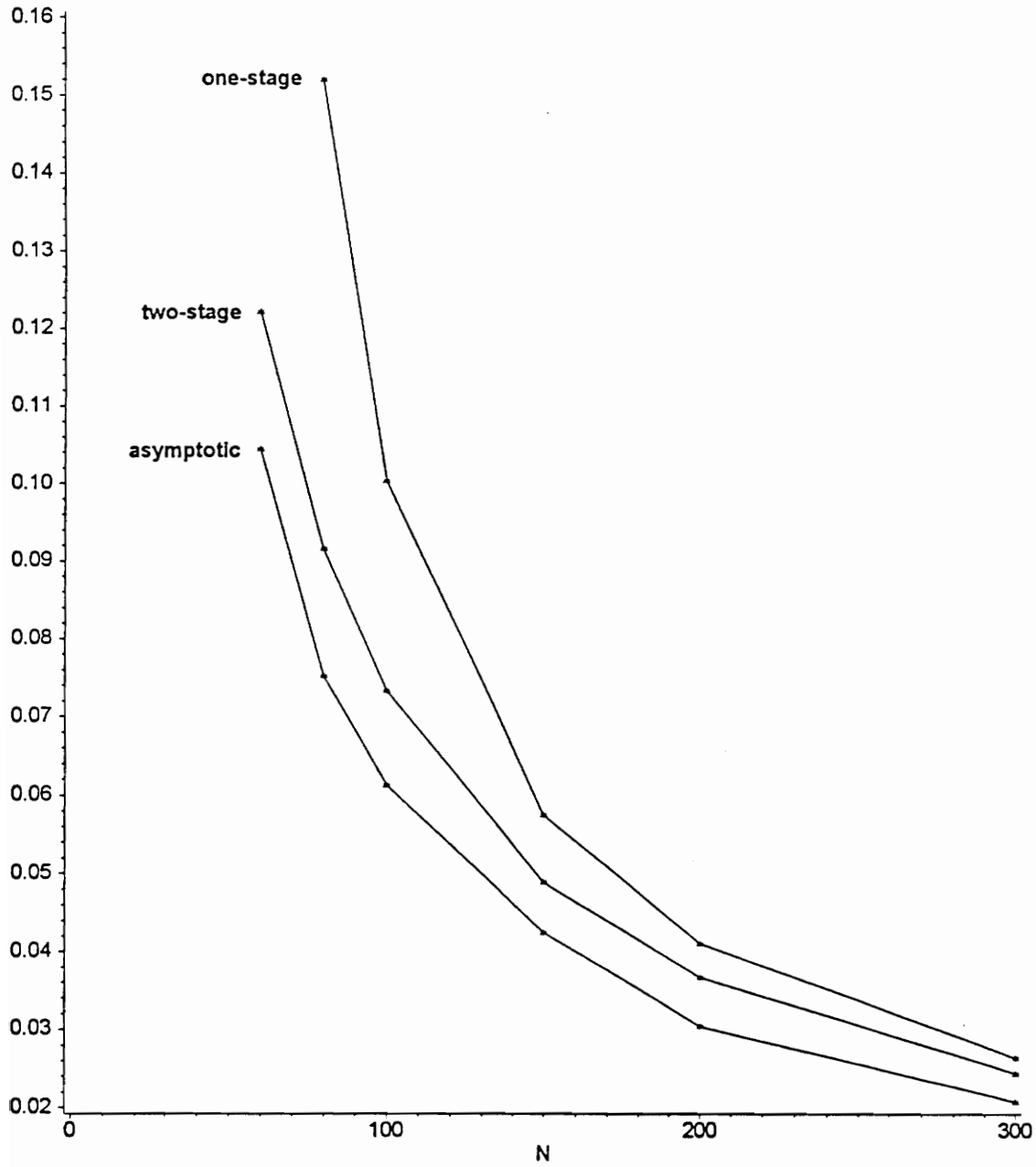

Figure 9.2.4 Small sample and asymptotic estimates of variance of $\hat{\mu}$ as a function of design and N when β_k=1.4 and μ_k=1.0.

In table 9.2.4 the >>> signs under the one-stage column indicate arbitrarily large numbers. This situation occurs when the estimate of β is approximately zero resulting in an arbitrary estimate of μ. The problem arises from extremely poor estimation of β. Tables 9.2.3 and 9.2.4 along with the corresponding figures illustrate the same message as the previous tables in chapter nine. With respect to variance and bias of parameter estimates, the two-stage is superior to the one-stage design for small samples.

Now consider the ratio of generalized variances as a method of studying small sample properties of the one and two-stage designs. Table 9.2.5 contains D efficiencies as defined by

$$\text{D\_EFF} = \frac{|\,\text{variance matrix, one-stage design}\,|}{|\,\text{variance matrix, two-stage design}\,|} \tag{9.2.1}$$

where the individual variances are estimated through simulation. The same one and two-stage designs are used in table 9.2.5 as was used previously in this chapter. D efficiencies exceeding one indicate that the two-stage design is more efficient than the one-stage design with respect to generalized variance.

Table 9.2.5 Small sample D efficiencies, one-stage vs two-stage

	D efficiency		
n	β_k=1.0, μ_k=0.0	β_k=0.8, μ_k=1.5	β_k=1.4, μ_k=1.0
asy.	0.92	1.56	1.46
300	0.94	33.59	1.56
200	0.94	92.53	1.67
150	0.97	125.61	3.23
100	1.16	140.22	14.87
80	2.03	90.32	41.66
60	6.25	17.89	>>>

The superiority of the two-stage D-D optimal design over the one-stage D optimal design as seen in table 9.2.5 is quite impressive. Notice that even in the situation of perfect parameter knowledge, the two-stage is superior to the one-stage for small samples. Recall that the one-stage design is only asymptotically optimal. The asymptotic efficiencies represent a minimum, indicating that the performance of the two-stage relative to the one-stage improves as sample size decreases. The performance of the one-stage design degrades more quickly and to a greater degree than the two-stage design.

Table 9.2.5 indicates that a sample size of 300 is not necessarily large enough for the asymptotic efficiency to be representative of the actual efficiency. This problem originates with the variance of $\widehat{\beta}$ from the one-stage design, not the two-stage. The one-stage variances become grossly misrepresented for sample sizes as low as 300. Consequently, the asymptotic D efficiencies used throughout chapter 8 are not necessarily representative of the actual efficiencies especially when the parameters are moderately misspecified. However, these asymptotic efficiencies do represent a baseline in favor of the two-stage procedure. Therefore, the efficiencies are still quite useful even though they underestimate the true capability of the two-stage procedure.

Certainly, in the design of experiments for the logistic model and nonlinear models in general, the use of asymptotic properties and results are commonplace. This chapter casts some doubt on the appropriateness of such asymptotic results for creating optimal designs. This is particularly true for the optimal one-stage designs. The two-stage procedure however does not reflect the same amount of pessimism. Consequently, the appropriatness of the asymptotic results lies in large part with the design.

A five point nearly D optimal one-stage design was briefly examined as to the appropriatness of the asymptotic approximations. Although, this design showed a marked improvement over the one-stage D optimal, it fell quite short of the two-stage design. It therefore arises once again that the addition of design points alone does not ensure good design. Rather it is the additional design points coupled with the complementary nature of the two-stage procedure which makes

the two-stage design so superior to the one-stage design. In the same light, it is assumed that the conclusions of this chapter apply not only to the particular two-stage design used in the small sample study but rather to the D-D optimal two-stage procedure in general.

CHAPTER 10

SUMMARY AND FUTURE RESEARCH

The primary goal of this research was to develop and study a useful, practical, and highly efficient two-stage design procedure for fitting the one variable logistic regression model. Secondary goals included the study of various design optimality criteria and the one-stage designs which they yield. These one and two-stage designs were primarily studied asymptotically. However, some small samples results were given.

In chapter 3, several design optimality criteria were investigated. Simplistic forms were given where possible in an effort to better understand the nature of the criteria. In addition, an emperical relationship between the Q and F criteria was provided. This relationship yields insight into the equivalency of a region of interest and an ED of interest coupled with a particular level of confidence and sample size.

Chapter 5 contains numerous one-stage (asymptotically) optimal and near optimal designs. Many Bayesian D optimal designs are also provided. The robustness to parameter misspecification of these one-stage designs was then studied in chapter 6. The optimal one-stage designs, while being very efficient

when parameter knowledge is good, are quite inefficient otherwise. The Bayesian designs on the other hand are very appealing in that the robustness properties can be customized to some degree through the choice of the prior. However, not surprisingly, the goals of robustness to parameter misspecification and being highly efficient under good parameter knowledge are in conflict.

A two-stage design procedure was developed in chapter 7 which effectively separates the two contradictory goals of robustness and efficiency. The first stage design of the two-stage procedure addresses the issue of robustness to parameter misspecification, and therefore makes use of Bayesian design as well as Bayesian estimation. The second stage design uses the parameter estimates obtained from the first stage in an effort to best achieve the overall goal of the experiment. Consequently, the second stage addresses efficiency.

The properties of a particular two-stage procedure, the D-D optimal design procedure, D optimality in both stages, was evaluated in chapter 8. The procedure was evaluated asymptotically for a sample size of 300. The large sample size was used solely in hopes that the asymptotic approximations would be representative of the actual. As measured by an asymptotic D efficiency, the D-D optimal designs were found to be vastly superior to the one-stage optimal designs in all situations except correct or nearly correct parameter specification.

The asymptotic D efficiencies of chapter 8 underestimate the actual efficiencies, particularly under moderate to severe parameter misspecification. Nevertheless, they are quite useful, in that they provide a minimum efficiency in favor of the two-stage procedure. The two-stage procedure performs its worst in comparison to the one-stage optimal design for infinite sample size. **The advantage of the two-stage procedure is much more than can be realized from asymptotic D efficiencies.** For small samples the two-stage procedure is even more efficient than the one-stage in the situation of perfect parameter knowledge.

This dissertation has provided a two-stage design procedure which effectively addresses the desires and concerns of designing an experiment for the logistic regression model. Although, only D-D optimal two-stage designs were studied, the two-stage procedure is easily generalized for the use of most design optimality

criteria. The reasonable assumption is then made that the superiority of the two-stage procedure over the one-stage design is not unique to D optimality. The nature of the two-stage procedure coupled with Bayesian design and estimation should allow it to perform quite well with the use of any criterion.

Several ideas for areas of extension and interest has arisen from this research, and they are listed below.

1) Extend the application of the two-stage procedure to more than one variable. The difficulty of a single variable has been illustrated in this dissertation. The task of multiple variables and their possible interactions significantly complicates the situation.

2) Consider using Bayesian estimation in both stages instead of only the first. This would require one to redefine and derive new optimal design theory, theory based upon Bayesian estimation instead of likelihood estimation.

3) Multiple stage designs containing three or more stages is a natural extension of the two-stage design. The idea of multiple stages may allow for additional design adaption in the event of a misspecified model in addition to the misspecification of parameters.

4) Certainly, other nonlinear models should be considered besides the logistic model as well as error structures other than Bernoulli. For instance, the two-stage design concept could be applied to a nonlinear model with Poisson or exponential errors.

5) The small sample exploration for the logistic model could be expanded. The need exists for good small sample variance approximations. In addition, the small sample study should accompany (4) above in the expansion of the two-stage concept to other error structures and other nonlinear models.

Bibliography

Bibliography

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

Atkinson, A. C. (1982), "Developments in the Design of Experiments," *International Statistical Review*, **50**, 161-177.

Berkson, J. (1944), "Application of the Logistic Function to Bio-assay," *Journal of the American Statistical Association*, **39**, 357-365.

Bernardo, J. M. (1979), "Expected Information as Expected Utility," *Annals of Statistics*, **7**, 686-690.

Box, G. E. P. and Draper, N. R. (1959), "A Basis for the Selection of a Response Surface Design," *Journal of the American Statistical Association*, **54**, 622-654.

---------- (1987), *Emperical Model-Building and Response Surfaces*, New York: John Wiley & Sons.

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

Box, G. E. P. and Wilson, K. B. (1951), "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society*, Ser. B, 13, 1-45.

Brooks, R. J. (1972), "A Decision Theory Approach to Optimal Regression Design," *Biometrika,* **59**, 563-571.

Chaloner, K. (1984), "Optimal Bayesian Experimental Design for Linear Models," *Annals of Statistics,* **12**, 283-300.

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

Covey-Crump, P. A. K. and Silvey, S. D. (1970), "Optimal Regression with Previous Observations," *Biometrika*, **57**, 551-566.

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 Ed.), Cambridge University Press.

Hill, W. J. and Hunter, W. G. (1966), "A Review of Response Surface Methodology: A Literature Review," *Technometrics*, **8**, 571-590.

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

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.

Khuri, A. I. and Cornell, J. A. (1987), *Repsonse Surfaces: Design and Analyses*, New York: Marcel Dekker.

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.

Mead, R. and Pike, D. J. (1975), "A Review of Response Surface Methodology from a Biometric Viewpoint," *Biometrics*, **31**, 803-851.

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

Myers, R. H. (1976), *Response Surface Methodology*, Blacksburg, Virginia: Author.

Myers, R. H., Khuri, A. I., and Carter, W. H., Jr. (1989), "Response Surface Methodology: 1966-1988," *Technometrics*, **31**, 137-157.

Myers, W. R. (1991), "Optimal Experimental Design for Fitting the Logistic Regression Model," unpublished Ph.D. thesis, Virginia Commeonwealth University, Richmond, Virginia, Department of Biostatistics.

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.

Näther, W. and Pilz, J. (1980), "Estimation and Experimental Design in a Linear Model using Prior Information," *Zastosowania Matematyki*, **16**, 565-577.

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

Shenton, L. R. and Bowman, K. O. (1977). *Maximum Likelihood Estimation in Small Samples*, London: Griffin.

Silvey, S. D. (1980), *Optimal Designs*, London: Chaptman & Hall.

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

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.

Sitter, R. R. and Wu, C. F. J. (in press 1994), "Optimal Designs in Binary Response Experiments: Fieller, D, and A Criteria," *Scandinavian Journal of Statistics*, to appear.

White, L. V. (1975), "The Optimal Design of Experiments for Estimation in Nonlinear Models," unpublished Ph.D thesis, Imperial College, London, Department of Mathematics.

Wilks, S. S. (1932), "Certain Generalizations in the Analysis of Variance," *Biometrika*, **24**, 471-494.

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

APPENDIX

Appendix A

Using the delta method approximation, the prediction variance of $\hat{P}(x_0)$ is the product of the squared bernoulli variance at x_0 and the prediction variance of $\text{logit}(\hat{P}(x_0))$.

Consider the following fitted logistic model and corresponding predicted response at a point x_0,

$$\hat{P}(x_0) = \frac{1}{1 + e^{-(\hat{\beta}_0+\hat{\beta}_1 x_0)}}.$$

By the delta method, the variance of $\hat{P}(x_0)$ is approximated using a first order Taylor series expansion of $\hat{P}(x_0)$ about β_0 and β_1. The approximation is given by

$$\text{Var}(\hat{P}(x_0)) \simeq \text{Var}(\hat{\beta}_0)\left[\left(\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_0}\right)^2\right]_{\beta_0,\beta_1} + \text{Var}(\hat{\beta}_1)\left[\left(\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_1}\right)^2\right]_{\beta_0,\beta_1}$$

$$+ 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)\left[\left(\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_0}\right)\left(\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_1}\right)\right]_{\beta_0,\beta_1}. \qquad \text{(A1)}$$

The first differential needed for expression A1 is given by

$$\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_0} = \left[\frac{1}{1 + \text{EXP}(-(\hat{\beta}_0+\hat{\beta}_1 x_0))}\right]^2 \left[\text{EXP}(-(\hat{\beta}_0+\hat{\beta}_1 x_0))\right]$$

which can also be written as

$$= \hat{P}(x_0)\left(1-\hat{P}(x_0)\right)$$

or

$$= \hat{P}_0(1-\hat{P}_0).$$

Similarly, the last differential needed for expression A1 is given by

$$\frac{\partial \hat{P}(x_0)}{\partial \hat{\beta}_1} = \left[\frac{1}{1 + \mathbb{EXP}(-(\hat{\beta}_0+\hat{\beta}_1 x_0))}\right]^2 \left[\mathbb{EXP}(-(\hat{\beta}_0+\hat{\beta}_1 x_0))\right] x_0$$

which likewise can be rewritten as

$$= x_0\hat{P}(x_0)\left(1-\hat{P}(x_0)\right).$$

or

$$= x_0\hat{P}_0(1-\hat{P}_0).$$

The assumption is made that the parameter estimates are in fact the maximum likelihood estimates. The variance components of the parameter estimates used in expression A1 are then taken to be the asymptotic variances from the inverted information matrix.

Substituting in expression A1, the prediction variance becomes

$$\mathrm{Var}(\hat{P}(x_0)) \simeq \hat{P}_0^2(1-\hat{P}_0)^2\mathrm{Var}(\hat{\beta}_0) + x_0^2\hat{P}_0^2(1-\hat{P}_0)^2\mathrm{Var}(\hat{\beta}_1)$$

$$+ 2x_0\hat{P}_0^2(1-\hat{P}_0)^2\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1). \qquad \text{(A2)}$$

A2 can then be simplified to yield

$$\mathrm{Var}(\hat{P}(x_0)) \simeq \hat{P}_0^2(1-\hat{P}_0)^2\left[\mathrm{Var}(\hat{\beta}_0) + x_0^2\mathrm{Var}(\hat{\beta}_1) + 2x_0\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)\right]$$

which in turn is equivalent to

$$= P_0^2(1-P_0)^2\mathbf{x}_0'(\mathbf{I}^{-1})\mathbf{x}_0 \qquad \text{(A3)}$$

where $\mathbf{x}_0=(1, x_0)$ and $I^{-1}=\begin{pmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) \end{pmatrix}$.

Expression A3, the prediction variance of $\hat{P}(x_0)$, represents the product of the squared bernoulli variance at x_0 and the prediction variance of $\mathrm{logit}(\hat{P}(x_0))$.

Appendix B

The slope of the one variable logistic regression model at a regressor location in logit space is given by the bernoulli variance at that location. This result is shown using the following form of the one variable logistic regression model,

$$P_0 = \frac{1}{1 + e^{-\beta(x_0-\mu)}}, \tag{B1}$$

and the logit function given by

$$\text{logit}(P) = \beta(x-\mu).$$

Let z=logit(P) and the model B1 can rewritten as

$$P_0 = \frac{1}{1 + e^{-z_0}}. \tag{B2}$$

The derivative of B2 with respect to z is then

$$\frac{\partial P_0}{\partial z_0} = \frac{e^{-z_0}}{(1 + e^{-z_0})^2}, \tag{B3}$$

which represents the slope of the tangent line on the logistic response curve at the location z_0 in regressor space. Expression B3 is then rewritten as follows

$$\left(\frac{1}{1 + e^{-z_0}}\right)\left(\frac{e^{-z_0}}{1 + e^{-z_0}}\right),$$

which is equivalent to $P_0(1-P_0)$, the bernoulli variance at the regressor location z_0. Consequently the slope of the logistic curve is given by the bernoulli variance.

Appendix C

The dependence of the Fieller interval on the ratio $\frac{z^2}{n}$ and not on n and z separately holds for any design. However, the property will be proved for a two point design. The Fieller interval for μ in general is given by

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{V_{12}}{V_{22}}\right) \pm \frac{z}{b}\left\{V_{11} - 2V_{12}\left(\frac{a}{b}\right) + V_{22}\left(\frac{a}{b}\right)^2 + g\left(\frac{(V_{12})^2}{V_{22}} - V_{11}\right)\right\}^{\frac{1}{2}}\right)$$

where $g=\frac{z^2V_{22}}{b^2}$.

The goal is to show that the interval relies only on n and z through the ratio $\frac{z^2}{n}$. First consider g, which can be written for a two point design using results from sections 3.2 and 3.3 as

$$g=\frac{z^2\left(n_1P_1(1-P_1)+n_2P_2(1-P_2)\right)}{b^2n_1n_2P_1(1-P_1)P_2(1-P_2)(x_1-x_2)^2}. \qquad \text{(C1)}$$

Now let $n_1=nh_1$ and $n_2=nh_2$ where n is the total sample size and h_1 and h_2 are the proportions of n used at the first and second design points, respectfully. Consequently equation C1 can be written in the following form

$$\frac{z^2\left(nh_1P_1(1-P_1)+nh_2P_2(1-P_2)\right)}{b^2nh_1nh_2P_1(1-P_1)P_2(1-P_2)(x_1-x_2)^2}. \qquad \text{(C2)}$$

Simplifying equation C2, we obtain

$$g=\frac{\left(\frac{z^2}{n}\right)\left(h_1P_1(1-P_1)+h_2P_2(1-P_2)\right)}{b^2h_1h_2P_1(1-P_1)P_2(1-P_2)(x_1-x_2)^2} \qquad \text{(C3)}$$

which is a function of n and z only in the form of the ratio $\left(\frac{z^2}{n}\right)$.

Equation C3 shows that the variance component V_{22} depends on n only in the form of a reciprical product. Let $V_{22}=\frac{1}{n}V'_{22}$, where V'_{22} is not a function of n. The remaining two variance components can be expressed in a similar manner. Therefore equation C1 can be expressed as,

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{\frac{1}{n}V'_{12}}{\frac{1}{n}V'_{22}}\right) \pm \frac{z}{b}\left\{\frac{1}{n}V'_{11} - 2\frac{1}{n}V'_{12}\left(\frac{a}{b}\right) + \frac{1}{n}V'_{22}\left(\frac{a}{b}\right)^2 + g\left(\frac{\left(\frac{1}{n}V'_{12}\right)^2}{\frac{1}{n}V'_{22}} - \frac{1}{n}V'_{11}\right)\right\}^{\frac{1}{2}}\right)$$

where V'_{11}, V'_{12}, and V'_{22} are no longer functions of n. Simplification yields the following form of the Fieller interval which is a function of z and n only through the ratio of the square of z to n.

$$\left(\frac{1}{1-g}\right)\left(\left(\left(\frac{a}{b}\right)-g\frac{V'_{12}}{V'_{22}}\right) \pm \frac{1}{b}\left\{\frac{z^2}{n}\left(V'_{11} - 2V'_{12}\left(\frac{a}{b}\right) + V'_{22}\left(\frac{a}{b}\right)^2 + g\left(\frac{\left(V'_{12}\right)^2}{V'_{22}} - V'_{11}\right)\right)\right\}^{\frac{1}{2}}\right)$$

Therefore the fieller interval and corresponding F optimal designs depend on n and z only through the ratio of the square of z to n. The proof can be extended to an arbitrary design.

Appendix D

The E optimal design which minimizes the maximum eigenvalue of $I^*(\mu,\beta)$ is equivalent to the design which minimizes the maximum prediction variance on spheres of varying radii centered about μ. Recall that $I^*(\mu,\beta)$ is the scaled information matrix for the centered one variable logistic model.

Consider the prediction variance of Logit($\hat{p_0}$) given by

$$\text{Var}(\text{Logit}(\hat{p}_0))=\text{Var}(\beta(x_0-\mu)) \tag{D1}$$

where the model is assumed to be centered. Prediction variance is invariant to changes in location and scale. Therefore expression D1 is equivalent to

$$\text{Var}(\alpha-\beta x_0) \tag{D2}$$

where the model is not centered. The prediction variance is then given by

$$(V_{11}-2x_0V_{12}+x_0^2V_{22})$$

where the variance components come from the uncentered model. Using expressions from 3.2.1, the prediction variance can then be equivalently expressed as a function of the variances as they appear from the centered model, scaled information matrix case. Consequently,

$$\text{Var}(\beta(x_0-\mu)) = (\beta^2V_{c11}-2\beta(x_0-\mu)V_{c12}+(x_0-\mu)^2V_{c22}) \tag{D3}$$

in which the c notation denotes the use of the centered model. Expression D3 is then equivalent to $\mathbf{x}_0'(I^{*-1})\mathbf{x}_0$ where $\mathbf{x}_0=(1,\ x_0-\mu)$ and I^{*-1} is given by 3.1.1.

Now consider the following expression

$$\underset{\mathfrak{D}}{\text{Min}}\left[\underset{||\mathbf{x}_0||=r^2}{\text{Max}}\ \mathbf{x}_0'(I^{*-1})\mathbf{x}_0\right]$$

where r is an arbitrary real number and $||\mathbf{x}_0||$ is dot product. In order to find the maximum prediction at a distance $\rho^2=(r^2\text{-}1)$, Lagrangian multipliers are used. Let the expression to maximized be

$$L = \mathbf{x}_0'(I^{*\text{-}1})\mathbf{x}_0 - \lambda(\mathbf{x}_0'\mathbf{x}_0\text{-}r^2) \tag{D4}$$

where λ is a Lagrangian multiplier. Differentiating L with respect to $\mathbf{x}_0$ yields

$$\frac{\partial L}{\partial \mathbf{x}_0} = 2(I^{*\text{-}1})\mathbf{x}_0 - 2\lambda\mathbf{x}_0. \tag{D5}$$

Setting expression D5 equal to zero, we obtain after some simplification

$$\left((I^{*\text{-}1}) - \lambda I\right)\mathbf{x}_0 = 0 \tag{D6}$$

where I is the identity matrix. By definition the location at which L is maximized is an eigenvector from $I^{*\text{-}1}$. The Lagrangian multiplier is in fact the eigenvalue associated with that eigenvector.

In order to determine which of the two eigenvalues and eigenvectors maximizes L, consider expression D6 again. Premultiply the expression by $\mathbf{x}_0$ to obtain

$$\mathbf{x}_0'(I^{*\text{-}1})\mathbf{x}_0 - \lambda\mathbf{x}_0'\mathbf{x}_0 = 0 \tag{D7}$$

where by definition λ is an eigenvalue of $I^{*\text{-}1}$. Solving expression D7 for λ and substituting r^2 for $\mathbf{x}_0'\mathbf{x}_0$ yields

$$\lambda = \frac{\mathbf{x}_0'(I^{*\text{-}1})\mathbf{x}_0}{r^2}.$$

Rewriting once again as

$$\lambda\, r^2 = \mathbf{x}_0'(I^{*\text{-}1})\mathbf{x}_0$$

it becomes clear that the prediction variance is maximized for an arbitrary r when λ is equated to the largest eigenvalue of $I^{*\text{-}1}$. Consequently, minimizing the

maximum eigenvalue of I^{*-1} is equivalent to minimizing the maximum prediction variance distances d from μ. In this situation, with only one regressor, the E optimal design minimizes the maximum prediction variance which occurs either at μ+d or μ-d for arbitrary d.

Appendix E

Optimal values of both k and d can be determined independently of μ_0, β_0, and β_L when given $\frac{\beta_U}{\beta_L}$ and $\mu_\Delta=\beta_L(\mu_U-\mu_0)=\beta_L(\mu_0-\mu_L)$. This fact is easily shown by rewriting the logistic model,

$$P = \frac{1}{1+\exp^{-\beta(x-\mu)}} \tag{E1}$$

so that it is a function only of the ratio $\frac{\beta_U}{\beta_L}$ and scaled difference μ_Δ.

The parameter β is subject to the bounds (β_L,β_U) and can therefore be expressed over the range as $\beta = q\beta_L$, where $q=1,\ldots,\frac{\beta_U}{\beta_L}$. Similarly, the parameter μ is subject to the bounds (μ_L,μ_U) which can be expressed as a function of β_L, μ_0, and μ_Δ by

$$(\mu_0-\frac{\mu_\Delta}{\beta_L},\ \mu_0+\frac{\mu_\Delta}{\beta_L}).$$

Consequently, the parameter μ can be expressed over its range as

$$\mu = \mu_0 + \frac{l\mu_\Delta}{\beta_L}$$

where $l = -1,\ldots,1$. Substituting for μ and β into E1, we obtain

$$P = \frac{1}{1+\exp^{-q\beta_L(\mu_0+\frac{id}{\beta_L}-\mu_0-\frac{l\mu_\Delta}{\beta_L})}} \tag{E2}$$

where $i=\pm\frac{1}{2},\pm1,\pm2,\ldots$ until there are k total design points if k is even or $i=0,\pm1,\pm2,\ldots$ until there are k total design points if k is odd. The scale-free distance between each design point is given by d while the actual distance is given by $\frac{d}{\beta_L}$.

The model expression of E2 is then simplified to yield

$$= \frac{1}{1+\exp^{-q(id-l\mu_{\Delta})}}$$

The model is no longer a function of μ_0, β_0, β_L, and μ_L. Consequently, the model and resulting optimal designs do not depend upon μ_0, β_0, β_L, and μ_L but rather only on the ratio $\frac{\beta_U}{\beta_L}$ and scaled difference $\mu_{\Delta}=\beta_L(\mu_U-\mu_0)=\beta_L(\mu_0-\mu_L)$.

APPENDIX F

This computer program which is written in SAS Proc IML determines the optimal second stage of a two-stage design. Given a first stage design and first stage parameter estimates, the program finds the optimal second stage design using the D optimality criterion. The second stage consists of two design point locations and the sample allocation to those design points.

The first stage design, parameter estimates, and second stage sample size must be provided by the user. The program can accomodate a first stage design of eight design points or less. However, the program is easily extended to more than eight. The first stage design is entered in section "main". The design point locations in natural units and corresponding sample distribution to those design points are denoted by X1i and N1i, respectfully. The first stage estimates of β and μ are denoted by "MU_HAT" and "BETA_H" respectfully. The second stage sample size is denoted by "N2ND".

```
PROC IML;

START FUNCTION;
************************
* computes value of estimated *
* joint determinant           *
************************;
P21 = PARMS[1,1];
P22 = PARMS[1,2];
N21 = PARMS[3,1];

N22 = N2ND -N21;
X21 = LOG(P21/(1-P21))/B + U;
X22 = LOG(P22/(1-P22))/B + U;
```

```
**********************************
* components of the joint information matrix *
**********************************;
WT = W1 + (B**2)*(N21*P21*(1-P21) + N22*P22*(1-P22));
XWT = XW1 - B*((X21-U)*P21*(1-P21) + (X22-U)*P22*(1-P22));
X2WT = X2W1 + ((X21-U)**2)*N21*P21*(1-P21)
            + ((X22-U)**2)*N22*P22*(1-P22);

**************************
* negative joint determinant value *
**************************;
FN_VALUE = -(WT*X2WT - XWT*XWT);

FINISH FUNCTION;
*////////////////////////////////////////////////*



*////////////////////////////////////////////////*;
START SIMPLEX;
**************************
* Nelder-Mead Simplex Algorithm*
* for function minimization      *
**************************;

_NITER_=900;_EPS_=1.0E-7;_K_=NROW(IN_PARMS);_KK_=_K_+1;
_P_=J(_K_,_KK_,0);_Y_=J(1,_KK_,0);
COUNT=0;_DABIT_=2.04607E-20;_BIGNUM_=1.0E38;_KONVGE_=5;
_PBAR_=J(_K_,1,0);_PSTAR_=_PBAR_;_P2STAR=_PBAR_;
_RCOEFF_=1.0;_ECOEFF_=2.0;_CCOEFF_=0.5;

** construct initial simplex **;

_P_[,_KK_]=IN_PARMS;PARMS=IN_PARMS;RUN FUNCTION;
_A_=FN_VALUE;_Y_[_KK_]=_A_;COUNT=COUNT+1;
```

```
DO _I_=1 TO _K_;
_P_[,_I_]=IN_PARMS;_P_[_I_,_I_]=_P_[_I_,_I_]+IN_STEPS[_I_];
_TEMP_=_P_[,_I_];PARMS=_TEMP_;RUN FUNCTION;
_A_=FN_VALUE;_Y_[,_I_]=_A_;COUNT=COUNT+1;
END;

** simple is now constructed **;

HILO;
_YLO_=MIN(_Y_);_YNEWLO_=MAX(_Y_);

DO _I_=1 TO _KK_;
IF _Y_[,_I_]=_YLO_ THEN _ILO_=_I_;
IF _Y_[,_I_]=_YNEWLO_ THEN _IHI_=_I_;
END;

** perform convergence check on function **
** the ratio of the largest to smallest vertex function test **;
_DCHK_=(_YNEWLO_+DABIT_)/(_YLO_+_DABIT_)-1;
IF ABS(_DCHK_)<_EPS_ THEN GOTO BEST;
_KONVGE_=_KONVGE_-1;
IF _KONVGE_=0 THEN DO;_KONVGE_=5;
   DO _I_=1 TO _K_;
      _COORD1_=_P_[_I_,1];_COORD2_=_COORD1_;
      DO _J_=2 TO _KK_;
      IF _P_[_I_,_J_]<_COORD1_ THEN _COORD1_=_P_[_I_,_J_];
      IF _P_[_I_,_J_]>_COORD2_ THEN _COORD2_=_P_[_I_,_J_];
      END;
       _DCHK_=(COORD2_+_DABIT_)/(_COORD1_+_DABIT_)-1;
       IF ABS(_DCHK_)<=_EPS_ THEN GO TO BEST;
   END;
END;
IF COUNT>_NITER_ THEN GOTO BEST;
```

```
** calculate _PBAR_, the centroid of the simplex vertices **
** except when _Y_ = _YNEWLO_ **;

DO _I_=1 TO _K_;_Z_=0;
    DO _J_=1 TO _KK_;_Z_=_Z_+_P_[_I_,_J_];END;
    _Z_=_Z_-_P_[_I_,_IHI_];_PBAR_[_I_]=_Z_/_K_;
END;
_PSTAR_=(1+_RCOEFF_)*_PBAR_-_RCOEFF_*_P_[,_IHI_];

** reflection through the centroid **;

PARMS=_PSTAR_;RUN FUNCTION;_YSTAR_=FN_VALUE;
COUNT=COUNT+1;
IF _YSTAR_>=_YLO_ THEN GOTO NOEXT;
IF COUNT>=_NITER THEN GOTO RETAIN;

** successful reflection, so extension **;

_P2STAR_=_ECOEFF_*_PSTAR_+(1-_ECOEFF_)*_PBAR_;
PARMS=_P2STAR_;RUN FUNCTION;_Y2STAR_=FN_VALUE;
COUNT=COUNT+1;

**retain extension or contraction **;

IF _Y2STAR_>=_YSTAR_ THEN GOTO RETAIN;

EXCON:
   _P_[,_IHI_]=_P2STAR_;
   _Y_[_IHI_]=_Y2STAR_;
   GOTO HILO;

** no extension **;

NOEXT;
_L_=0;
```

```
DO _I_=1 TO _KK_;
IF _Y_[_I_]>_YSTAR_ THEN _L_=_L_+1;
END;
IF _L_>1 THEN GOTO RETAIN;

** contraction on the reflection side of the centroid **;

IF _L_=1 THEN DO;
_P_[,_IHI_]=_PSTAR_;
_Y_[_IHI_]=_YSTAR_;
END;

** contraction on the _Y_[_IHI_] side of the centroid **;

IF COUNT>=_NITER_ THEN GOTO BEST;
_P2STAR_=_CCOEFF_*_P_[,_IHI_]+(-_CCOEFF_+1)*_PBAR_;
PARMS= _P2STAR_;RUN FUNCTION;_Y2STAR_=FN_VALUE;
COUNT=COUNT+1;
IF _Y2STAR_<_Y_[_IHI_] THEN GOTO EXTCON;

** CONTRACT THE WHOLE SIMPLEX **;

DO _J_=1 TO _KK_;
   DO _I_=1 TO _KK_;
   _P_[_I_,_J_]=0.5*(_P_[_I_,_J_]+_P_[_I_,_ILO_]);
   END;_XMIN_=_P_[,_J_];
   PARMS=_XMIN_;RUN FUNCTION;_A_=FN_VALUE;
   _Y_[,_J_]=_A_;
END;
COUNT=COUNT+_KK_;
IF COUNT<_NITER_GOTO HILO;ELSE GOTO BEST;

RETAIN;
_P_[,_IHI_]=_PSTAR_;_Y_[_IHI_]=_YSTAR_;GOTO HILO;
```

```
BEST;
DO _J_=1 TO _KK_;_XMIN_=_P_[,_J_];
PARMS=_XMIN_;RUN FUNCTION;_A_FN_VALUE;_Y_[_J_]=_A_;
END;
_YNEWLO_=_BIGNUM_;
DO _J_ TO _KK_;
    IF _Y_[_J_]<_YNEWLO_ THEN DO;
       _YNEWLO_=_Y_[_J_];_IBEST_=_J_;
   END;
END;
_Y_[_IBEST_]=_BIGNUM_;_YSEC_=_BIGNUM_;
DO _J_=1 TO _KK_;
    IF _Y_[_J_]<_YSEC_ THEN DO;
       _YSEC_=_Y_[_J_];_ISEC_=_J_;
    END;
END;
_XMIN_=_P_[,_IBEST_];_XSEC_=_P_[,_ISEC_];
PARMS=_XMIN_;FN_VALUE=_YNEWLO_;

FREE _NITER_ _EPS_ _K_ _KK_ _P_ _Y_ _DABIT_ _BIGNUM_
_KONVGE_;
FREE _PBAR_ _PSTAR_ _P2STAR_ _RCOEFF_ _ECOEFF_
_CCOEFF_;
FREE _A_ _I_ _TEMP_ _YLO_ _YNEWLO_ _ILO_ _IHI_;
FREE _DCHK_ _COORD1_ _COORD2_ _Z_ _YSTAR_ _L_ _J_;
FREE _XMIN_ _IBEST_ _YSEC_ _XSEC_;

FINISH  SIMPLEX;
*//////////////////////////////////////////*
```

```
*////////////////////////////////////*;

                    **********************
                  **          MAIN        **
                    **********************;


** enter first stage parameter estimates **;
MU_HAT = ?????? ;
BETA_H = ?????? ;

** enter first stage design point locations (natural units) **
** and corresponding sample allocations **;
N11 = ??? ; X11 = ????? ;
N12 = ??? ; X12 = ????? ;
N13 = ??? ; X13 = ????? ;
N14 = ??? ; X14 = ????? ;
N15 = ??? ; X15 = ????? ;
N16 = ??? ; X16 = ????? ;
N17 = ??? ; X17 = ????? ;
N18 = ??? ; X18 = ????? ;

** enter the second stage sample size **;
N2ND = ??? ? ;


*////////////////////////////*;

U = MU_HAT;
B = BETA_H;

**    estimated probability of reponse at design point i **;
P11 = 1/(1+EXP(-B*(X11-U)));
P12 = 1/(1+EXP(-B*(X12-U)));
P13 = 1/(1+EXP(-B*(X13-U)));
```

```
P14 = 1/(1+EXP(-B*(X14-U)));
P15 = 1/(1+EXP(-B*(X15-U)));
P16 = 1/(1+EXP(-B*(X16-U)));
P17 = 1/(1+EXP(-B*(X17-U)));
P18 = 1/(1+EXP(-B*(X18-U)));

** first stage information matrix components **;
W1 = (B**2)*(N11*P11*(1-P11)+N12*P12*(1-P12)+N13*P13*(1-P13)+
             N14*P14*(1-P14)+N15*P15*(1-P15)+N16*P16*(1-P16)+
             N17*P17*(1-P17)+N18*P18*(1-P18);

XW1 = -B*((X11-U)*N11*P11*(1-P11)+(X12-U)*N12*P12*(1-P12)+
          (X13-U)*N13*P13*(1-P13)+(X14-U)*N14*P14*(1-P14)+
          (X15-U)*N15*P15*(1-P15)+(X16-U)*N16*P16*(1-P16);
          (X17-U)*N17*P17*(1-P17)+(X18-U)*N18*P18*(1-P18);

X2W1 = ((X11-U)**2)*N11*P11*(1-P11)+((X12-U)**2)*N12*P12*(1-P12)+
       ((X13-U)**2)*N13*P13*(1-P13)+((X14-U)**2)*N14*P14*(1-P14)+
       ((X15-U)**2)*N15*P15*(1-P15)+((X16-U)**2)*N16*P16*(1-P16)+
       ((X17-U)**2)*N17*P17*(1-P17)+((X18-U)**2)*N18*P18*(1-P18);


** initial starting values of the two second stage design points **
** in terms of probabilities of response **;
P21 = .20  ;
P22 = .80  ;
N21 = N2ND/2;

IN_PARMS = P21//P22//N21;
IN_STEPS = {.10, .10, 1};

CALL SIMPLEX;
```

```
** determinant of joint information matrix of two-stage design **;
DET = -FN_VALUE;

** prints second stage design along with determinant of information matrix **;
PRINT P21 P22 X21  X22 N21 N22 DET;

*////////////////////////////////////////////*;
```

VITA

William C. Letsinger, II, the son of Warren C. (deceased) and Sara H. Letsinger, was born on June 6, 1968 in Kokomo, Indiana. In 1986, he graduated from Northwestern High School, Kokomo, Indiana. He attended the University of West Florida in Pensacola, Florida and received a Bachelor of Arts degree in Mathematics in 1990, graduating cum laude. In that same year he received the UWF Mathematics Award and was inducted into the national mathematics honor society, Pi Mu Epsilon. He began graduate school in August 1990 at Virginia Polytechnic Institute and State University. In 1991 he received the Boyd Harshbarger Award for academic acheivement among the first year graduate students in statistics, was inducted in the national statistics honor society, Mu Sigma Rho, and received a Master of Science degree in Statistics.