INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.
Harper, William Victor

CONSIDERATION OF OPTIMAL DESIGN FOR BINARY RESPONSE EXPERIMENTS

The Ohio State University

Ph.D. 1984

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106
CONSIDERATION OF OPTIMAL DESIGN FOR BINARY RESPONSE EXPERIMENTS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

by

William Victor Harper, B.S., M.S.

The Ohio State University
1984

Reading Committee:
Walter C. Giffin, Ph.D.
Clark A. Mount-Campbell, Ph.D.
John B. Neuhardt, Ph.D.

Approved By
John B. Neuhardt, Ph.D.
Department of Industrial and Systems Engineering
This dissertation is dedicated to the memory of my dog, Sam, who passed away on June 12, 1984.
ACKNOWLEDGMENTS

I would like to acknowledge the good-natured support received from my advisor, Dr. John B. Neuhardt. Without his encouragement and understanding, this dissertation would not have been completed. I would also like to acknowledge the financial support of the Battelle Memorial Institute. In addition to providing computer resources, the one-year scholarship proved invaluable in the completion of this degree.
**VITA**

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>Born - Blue Island, Illinois</td>
</tr>
<tr>
<td>1967-1971</td>
<td>United States Air Force</td>
</tr>
<tr>
<td>1973</td>
<td>B.S., Computer and Information Science, Engineering College, summa cum laude, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>1976</td>
<td>M.S., Statistics, The Ohio State University, Columbus, Ohio</td>
</tr>
<tr>
<td>1976-1978</td>
<td>Statistician, Ross Labs, Columbus, Ohio</td>
</tr>
<tr>
<td>1978-Present</td>
<td>Statistician, Battelle Memorial Institute, Columbus, Ohio</td>
</tr>
</tbody>
</table>

**PUBLICATIONS**


Harper, W.V., Sensitivity/Uncertainty Analysis Techniques for Non-
stochastic Computer Codes, ONWI-444, May 1983, Office of Nuclear
Waste Isolation, Battelle Memorial Institute, Columbus, Ohio.

Harper, W.V., "Closed Form Solutions for the Propagation of Uncer-
tainty", 1983 American Statistical Association Meeting in Toronto,
Canada, Abstract in 1983 Program & Abstracts Joint Statistical Meet-
ings, pp. 93, American Statistical Association, Washington, D.C.

Harper, W.V., "Model Validation Via Microcomputer", Computer Applica-
tions in Health Physics, Proceedings of the 17th Midyear Topical
Symposium of the Health Physics Society, pp. 4.41-4.48, 1984, Colum-
bia Chapter of the Health Physics Society, Richland, Washington.

Borehole Scenario Comparing Latin Hypercube Sampling and
Deterministic Sensitivity Approaches, BMI/OWNI-516, October 1983,
Office of Nuclear Waste Isolation, Battelle Memorial Institute,
Columbus, Ohio.

Harper, W.V. and L. Kroitoru, Measures of Central Tendency for
Permeability Data, O/TM-30, May 1983, Office of Nuclear Waste
Isolation, Battelle Memorial Institute, Columbus, Ohio.

Harper, W.V., G.E. Raines, and E.M. Oblow, "Evaluation of Sensitiv-
ity/Uncertainty Analysis Techniques for Salt Repository Performance
Storage Program Information Meeting, DOE/NWTS-30, December, 1982,
pp. 210-213, United States Department of Energy, Washington, D.C.

Harper, W.V. and D. A. Waite, Long-Term Performance Objectives and
Measures for Repositories, ONWI-398, January 1983, Office of Nuclear
Waste Isolation, Battelle Memorial Institute, Columbus, Ohio.

Waite, D.A. and W.V. Harper, "ALARA Beyond Dollars Per Person-Rem",
Proceedings of the Fourth DOE Environmental Protection Information
Meeting, CONF-821215, August 1983, pp. 367-385, United States
Department of Energy, Washington, D.C.

of Existing Risk Assessment Results for High Level Nuclear Waste
Repositories", Proceedings of the International Radiation Protection
Association 6th International Congress Radiation - Risk - Protection,
Volume 3, West Berlin, Germany, May 7-12, 1984, pp 1366-1369.

FIELDS OF STUDY

Major Fields: Operations Research and Applied Statistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>iii</td>
</tr>
<tr>
<td>VITA</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
</tbody>
</table>

## CHAPTER

1. REVIEW OF LITERATURE.................................................. 5
   - Introduction.................................................. 5
   - Nonparametric Approaches.................................... 10
   - Bayesian Methods............................................ 22
   - Classical Parameter Analysis................................ 24
   - Problem of Interest.......................................... 28

2. OPTIMAL DESIGN AND LOGISTIC PROPERTIES................................ 29
   - Introduction.................................................. 29
   - Optimality Criteria - Survey of Methods.................... 29
   - Extensions of the Kiefer-Wolfowitz General Equivalence Theorem | 32   |
   - Properties of the Logistic Distribution.................... 37
   - Use of Logistic Properties in Optimal Design Theory...... 42
   - Experimental Design Problem to be Studied in Detail....... 45

3. SEQUENTIAL DESIGN ALGORITHMS FOR THE LOGISTIC......................... 48
   - Introduction.................................................. 48
   - Fixed Optimal Experimental Design for the Logistic Model. | 50   |
# TABLE OF CONTENTS

(Continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal Two-Point Design When</td>
<td>54</td>
</tr>
<tr>
<td>( x_1 \neq x^* )</td>
<td></td>
</tr>
<tr>
<td>Construction of D-Optimal Designs Subject</td>
<td>61</td>
</tr>
<tr>
<td>to Constraints on the Design Space</td>
<td></td>
</tr>
<tr>
<td>Algorithms to be Compared</td>
<td>67</td>
</tr>
<tr>
<td>4. PILOT STUDY</td>
<td>71</td>
</tr>
<tr>
<td>Introduction</td>
<td>71</td>
</tr>
<tr>
<td>Pilot Study Design and Results</td>
<td>72</td>
</tr>
<tr>
<td>5. DETERMINATION OF THE BEST SEQUENTIAL ALGORITHM AND RECOMMENDATION</td>
<td>81</td>
</tr>
<tr>
<td>FOR ITS USE</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>81</td>
</tr>
<tr>
<td>Reparameterization</td>
<td>83</td>
</tr>
<tr>
<td>Determination of Key Factors</td>
<td>85</td>
</tr>
<tr>
<td>Selection of Best Sequential Algorithm</td>
<td>88</td>
</tr>
<tr>
<td>Comparison of ALG7 and MAR</td>
<td>92</td>
</tr>
<tr>
<td>Robustness of ALG7</td>
<td>95</td>
</tr>
<tr>
<td>Least Squares Prediction of ALG7 Performance</td>
<td>98</td>
</tr>
<tr>
<td>6. SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH</td>
<td>103</td>
</tr>
<tr>
<td>Summary</td>
<td>103</td>
</tr>
<tr>
<td>Areas of Future Research</td>
<td>107</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>111</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Categorization of Methods</td>
<td>9</td>
</tr>
<tr>
<td>2.</td>
<td>Algorithm Means and Analysis of Variance Results for $a_0 = -6$, $b_0 = 1$</td>
<td>76</td>
</tr>
<tr>
<td>3.</td>
<td>Algorithm Means and Analysis of Variance Results for $a_0 = -2.25$, $b_0 = 0.375$</td>
<td>77</td>
</tr>
<tr>
<td>4.</td>
<td>Scheffé Multiple Comparison at 0.05 Significance Level for $b = 0.5$</td>
<td>90</td>
</tr>
<tr>
<td>5.</td>
<td>Scheffé Multiple Comparison at 0.05 Significance Level for $b = 1$</td>
<td>90</td>
</tr>
<tr>
<td>6.</td>
<td>$\theta$ Values, $\mu = 6$</td>
<td>91</td>
</tr>
<tr>
<td>7.</td>
<td>ALG7 and MAR Means by UDIFF and $b_0/b$ for $b = 0.5$</td>
<td>93</td>
</tr>
<tr>
<td>8.</td>
<td>ALG7 and MAR Means by UDIFF and $b_0/b$ for $b = 1.0$</td>
<td>94</td>
</tr>
<tr>
<td>9.</td>
<td>ALG7, MAR Means Averaged Over $b_0/b$</td>
<td>95</td>
</tr>
<tr>
<td>10.</td>
<td>ALG7 and MAR Mean, Minimum, Maximum, and Standard Deviation by $\mu$ for $b = 0.5$ and $b = 1$</td>
<td>96</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Typical Binary Response Curve</td>
<td>6</td>
</tr>
<tr>
<td>2.</td>
<td>Ford's Three Optimal Regions</td>
<td>54</td>
</tr>
<tr>
<td>3.</td>
<td>Plot of Typical $</td>
<td>M</td>
</tr>
<tr>
<td>4.</td>
<td>Plot of $\frac{\partial \ln</td>
<td>M</td>
</tr>
<tr>
<td>5.</td>
<td>Pictorial Summary of Algorithms</td>
<td>70</td>
</tr>
<tr>
<td>6.</td>
<td>Plot of $</td>
<td>M_A</td>
</tr>
<tr>
<td>7.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\mu = 4, \beta = 0.5$, $TSS = 22.8$</td>
<td>86</td>
</tr>
<tr>
<td>8.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\mu = 6, \beta = 0.5$, $TSS = 17.2$</td>
<td>86</td>
</tr>
<tr>
<td>9.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\mu = 8, \beta = 0.5$, $TSS = 16.3$</td>
<td>86</td>
</tr>
<tr>
<td>10.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\mu = 4, \beta = 1$, $TSS = 31.6$</td>
<td>87</td>
</tr>
<tr>
<td>11.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\mu = 6, \beta = 1$, $TSS = 25.1$</td>
<td>87</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

(Continued)

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\beta = 8$, $\beta = 1$, TSS = 30.5.</td>
<td>87</td>
</tr>
<tr>
<td>13.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\beta = 0.5$, TSS = 55.6.</td>
<td>89</td>
</tr>
<tr>
<td>14.</td>
<td>Total ANOVA Sum of Squares (TSS) Indicating Proportion Variance Explained by Main Effects and Interactions for $\beta = 1$, TSS = 96.2.</td>
<td>90</td>
</tr>
<tr>
<td>15.</td>
<td>Plot of Range of MAR for $\beta = 0.5$.</td>
<td>97</td>
</tr>
<tr>
<td>16.</td>
<td>Plot of Range of MAR for $\beta = 1$.</td>
<td>97</td>
</tr>
<tr>
<td>17.</td>
<td>Plot of Range of ALG7 for $\beta = 0.5$.</td>
<td>97</td>
</tr>
<tr>
<td>18.</td>
<td>Plot of Range of ALG7 for $\beta = 1$.</td>
<td>97</td>
</tr>
<tr>
<td>19.</td>
<td>Quadratic Regression Prediction of ALG7 as a Function of $\mu_0$ and $\beta_0/\beta$ for $\mu = 4$ and $\beta = 0.5$ ($R^2 = 0.78$, Root M.S.E. = 0.03).</td>
<td>99</td>
</tr>
<tr>
<td>20.</td>
<td>Quadratic Regression Prediction of ALG7 as a Function of $\mu_0$ and $\beta_0/\beta$ for $\mu = 6$ and $\beta = 0.5$ ($R^2 = 0.67$, Root M.S.E. = 0.03).</td>
<td>99</td>
</tr>
<tr>
<td>21.</td>
<td>Quadratic Regression Prediction of ALG7 as a Function of $\mu_0$ and $\beta_0/\beta$ for $\mu = 8$ and $\beta = 0.5$ ($R^2 = 0.82$, Root M.S.E. = 0.03).</td>
<td>100</td>
</tr>
<tr>
<td>22.</td>
<td>Quadratic Regression Prediction of ALG7 as a Function of $\mu_0$ and $\beta_0/\beta$ for $\mu = 4$ and $\beta = 1$ ($R^2 = 0.71$, Root M.S.E. = 0.05).</td>
<td>100</td>
</tr>
<tr>
<td>23.</td>
<td>Quadratic Regression Prediction of ALG7 as a Function of $\mu_0$ and $\beta_0/\beta$ for $\mu = 6$ and $\beta = 1$ ($R^2 = 0.87$, Root M.S.E. = 0.04).</td>
<td>101</td>
</tr>
<tr>
<td>FIGURE</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>24. Quadratic Regression Prediction of ALG7 as a Function of ( \mu ) and ( \beta_0/\beta ) for ( \mu = 8 ) and ( \beta = 1 ) (( R^2 = 0.86 ), Root M.S.E. = 0.04)</td>
<td>101</td>
<td></td>
</tr>
</tbody>
</table>
INTRODUCTION

The following chapters study experimental design for a binary response that is a function of the continuous variable time. This relationship is common in many areas in which an item degrades with time. If the independent response variable may be categorized into two groups such as "good" or "bad", then the relationship may be analyzed using techniques such as are presented in this study.

The overall objective of this research is to develop and compare sequential (or adaptive) algorithms to existing fixed designs (entire experiment preplanned) for binary data. The criterion used for these comparisons is an extension of D-optimality to nonlinear models such as the nonlinear logistic model used to study a binary response in this work. D-optimality, as detailed later, is a generalized variance criterion. Unlike linear models, the D-optimality of nonlinear designs depends on the parameter values of the underlying model. The existing D-optimal fixed design algorithms are thus only locally optimal in a small area around the true unknown parameter values. The purpose of performing an experiment is to estimate these unknown parameters; therefore, if one already knew the true parameter values there would be no need to conduct the experiment. When the initial estimates of these parameters are not close to the true values, the fixed D-optimal design scheme results
in a poor design. Rather than preplanning the entire experiment, sequential designs allow the experimenter to update his initial estimates and use the current estimates to plan the remainder of the experiment. When the initial estimates are poor, one would expect the sequential design to perform better than the fixed design. If the initial estimates are good (i.e., close to true values), both the sequential and fixed D-optimal designs should perform well. Four new sequential algorithms are developed for the nonlinear binary response problem detailed in this paper. These are compared to existing fixed designs, and recommendations are given as to their suggested use.

The first chapter introduces a typical binary response curve. The published literature is surveyed and categorized based on five criteria. As in linear models, most of the literature is concerned with the analysis of experimental data rather than the construction of the experimental design. The main problem of interest is briefly given at the end of Chapter 1 and covered in more detail in subsequent chapters. Briefly, this problem involves product degradation (though the methods presented are not limited to this) over time for a given batch of product. The binary response is a function of time and as the experiment progresses, one cannot go backwards in time to collect data. This "forward only" nature of this problem is a constraint on the sequential algorithms that will be developed in later chapters. Chapter 2 introduces optimal experiment design theory for linear models with a primary focus on D- and G-optimality. The Kiefer-Wolfowitz General Equivalence Theorem is
presented for design measures with three separate extensions of it by Federov, White, and Ford, respectively. A detailed study of the logistic distribution used to model our binary response as a function of time shows that the three General Equivalence Theorem extensions are identical for our logistic model. The third chapter develops the sequential design algorithms to be compared to existing fixed designs in subsequent chapters. The fixed two-point D-optimal design algorithm in the literature is given. Theory is developed for a two-point optimal design subject to the constraint of an arbitrary first point. The theory is used in the first two sequential algorithms presented. The last two sequential design algorithms are based on a fixed design algorithm developed by Federov for linear models.

The next two chapters examine the sensitivity of the algorithms to the variables used in the experimental designs constructed for algorithm comparison. Chapter 4 provides the results of a pilot study that is used to screen the independent variables used in the comparative experimental design. Both chapters quantify the performance of each algorithm using a ratio on the interval [0, 1] to indicate the degree of D-optimality achieved. Chapter 5 has a different parameterization than Chapter 4 to allow a broader comparison of algorithms. Due to the large number of simulation runs used, statistical significance was not always a good method of portraying the importance of the independent variables. Other methods including graphical approaches were used to identify important factors when all the factors were found to be statistically significant as a
result of the large sample size. Chapter 4 found two sequential algorithms that performed well over the range of experimental variables used in the study. These same two algorithms also were the top two sequential algorithms in Chapter 5; however, the better of these two was chosen for further comparison against the selected fixed design algorithm. Detailed comparison of the sequential and fixed algorithm shows that the sequential algorithm is more robust than the fixed design and results in a better design except when the initial logistic parameters estimates are close to the true unknown parameter values. To supplement these results, predictive contour plots are developed that provide the expected performance of the sequential algorithm over the continuum of the key independent variables studied. The last chapter summarizes this research and provides areas for future research.

The knowledge gained from this research indicates that sequential algorithms are useful in the determination of a binary response curve. They are much more robust than fixed designs for the nonlinear binary model studied here. How much better they can do depends on the amount and quality of available information. The work of Chapters 4 and 5 determined the conditions under which an adaptive scheme can be recommended, and verified that the adaptive algorithms are better than the fixed design algorithms when the initial estimates differ moderately from the actual values.
CHAPTER 1 REVIEW OF LITERATURE

INTRODUCTION

Experiments are usually designed to study the relationship between response (dependent) and explanatory (independent) variables. These variables may be quantitative or qualitative in nature. Our concern is focused on those situations where the dependent variable, y, is binary and the independent variables are quantitative and continuous. We will study y as a function of one independent variable x. While it is obvious the response is binary in experiments that result in one of two possible outcomes, there are many investigations, although containing continuous responses, in which errors may imply the actual responses cannot be accurately determined. Such continuous responses can be made binary and thus studied by the same mathematical methods. Mathematically, this relationship between a binary dependent variable and a continuous independent variable is as follows:

\[ y_x = \begin{cases} 
1 & \text{with probability } F(x) \\
0 & \text{with probability } 1-F(x) 
\end{cases} \]

Letting \( \theta_x = E(y_x) \) (implying \( 0 \leq \theta_x \leq 1 \)), a typical response curve is given in Figure 1 where \( \theta_x \) is assumed to be a nondecreasing function of x. Such a curve could represent, say, the probability
FIGURE 1. TYPICAL BINARY RESPONSE CURVE

of insect death as a function of dosage of poison, or perhaps the probability of failure of a machine component versus stress (temperature, humidity, etc.). In such cases, experience has shown that a nondecreasing function is realistic.

Depending on one's experimental purposes, different aspects of this curve may be of interest. One may want to estimate the entire curve; whereas, another may only care about selected points along the curve. This chapter will briefly survey methods to do either.

The relationship of this binary response variable with a continuous independent variable can be studied using either sequential or fixed (entire experiment preplanned) experimental designs. Also nonparametric, parametric, or Bayesian methods may be utilized in the estimation procedure. Regardless of the method employed, let $x_1, x_2, \ldots, x_k$, be the ordered sequence of the levels of the independent variable sampled. For an observation at $x_i$, the probability of a response ($P(y_x = 1)$) is $F(x_i)$ and the probability of nonresponse ($P(y_x = 0)$) is $1 - F(x_i)$. Points can be defined at a level, $L_p$, for which the probability of a response is $p$, i.e.,
E(y|L_p) = p. It is obvious that 0 ≤ p ≤ 1. At any point x along this curve a random sample of n independent observations of y|x can be seen to be binomally distributed with probability of "success" \( \theta_x \). The entire curve, \( \theta_x \) versus x, is assumed to be a nondecreasing function of x. What is assumed about the curve determines whether one is applying parametric or nonparametric assumptions, and if classical or Bayesian approaches are to be followed. Suppose at \( x_j \), \( n_j \) observations are taken. Let \( r_j \) be the number of positive responses (\( y = 1 \)) at \( x_j \). Then \( \hat{p}_j = \frac{r_j}{n_j} \) is an estimate of \( \theta_{x_j} \). In many applications of binary responses one is interested in \( \theta_{x} = .50 \). Areas in which this has been shown to apply are biostatistics (Finney, 1971) and psychophysics (Cornsweet, 1962).

The nondecreasing nature of the curve and the fact that \( 0 < \theta_x < 1 \) allows the relationship of \( \theta_x \) versus x to be thought of as a statistical cumulative distribution function. While it is not necessary to view the curve in this manner, it has proved to be a useful way of defining the modeled functional relationship between \( \theta_x \) and x. This will be seen throughout this paper. The two most common relationships assumed are:

\[
\begin{align*}
\theta_x &= \frac{e^{x\beta}}{1+e^{x\beta}} \quad \text{logistic} \\
\theta_x &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \, dz \quad \text{normal}
\end{align*}
\]

where \( \beta \) is \( tx1 \) vector of parameters

\( x \) is \( 1xt \) vector of independent variables.
In this paper, \( t \) will equal 2 and \( \beta = (\alpha, \beta) \). Theoretical justification for the logistic may be found in Berkson (1944, 1951), and in Finney (1971) for the normal.

The methods covered or referenced in this chapter fall into different categories. For our purposes, we have established the following categorical structure for Table 1:

1. \( L_{50} \) versus Global (the problem)
2. Classical versus Bayesian (the analysis procedure)
3. Nonparametric versus Parametric (the assumptions)
4. Analysis versus Design (major emphasis)
5. Sequential versus Fixed (the design approach).

While these terms will become more clear as the chapter progresses, it is worthwhile to provide brief definitions now. Most methods are designed to estimate a specific percentile, usually \( L_{50} \), or the entire \( \theta_X \) versus \( x \) curve. If a particular method is used to estimate \( L_{50} \) or more generally \( L_p \), it will be placed in the \( L_{50} \) category. If it can be used to estimate an arbitrary \( L_p \), then \( (L_p) \) will follow the method name. Estimation of the entire \( \theta_X \) versus \( x \) curve places a method in the global category. Classical methods as used here imply no prior distributions are developed for the parameter estimates \( (\hat{\alpha}, \hat{\beta}, \hat{\theta}_X) \); whereas, Bayesian methods incorporate prior knowledge formally into prior probability distributions for parameters of interest. Nonparametric methods do not make specific assumptions about the functional relationship of \( \theta_X \) versus \( x \) as do parametric methods which assume specific models. Some methods are more concerned with analysis of experimental data rather than the
<table>
<thead>
<tr>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-MSA(L_p)</td>
</tr>
<tr>
<td>UP-DOWN</td>
</tr>
<tr>
<td>TUP-D(L_p)</td>
</tr>
<tr>
<td>DUP-D(L_p)</td>
</tr>
<tr>
<td>PEST</td>
</tr>
<tr>
<td>S-K</td>
</tr>
<tr>
<td>R-M</td>
</tr>
<tr>
<td>D-B</td>
</tr>
<tr>
<td>HARPER</td>
</tr>
<tr>
<td>BETA(L_p)</td>
</tr>
<tr>
<td>OWEN</td>
</tr>
<tr>
<td>Probit</td>
</tr>
<tr>
<td>Logit</td>
</tr>
<tr>
<td>BETA (Winkler, 1972)</td>
</tr>
<tr>
<td>D-B (Dragstedt and Lang, 1928; Behrens 1929)</td>
</tr>
<tr>
<td>D-OPT (White, 1973)</td>
</tr>
<tr>
<td>DIRICH (Ramsey, 1972)</td>
</tr>
<tr>
<td>DUP-D (Brownlee, 1953)</td>
</tr>
<tr>
<td>FREE (Freeman, 1970)</td>
</tr>
<tr>
<td>HARPER (Harper, 1980)</td>
</tr>
<tr>
<td>LOGIT (Berkson, 1944)</td>
</tr>
<tr>
<td>OWEN (Owen, 1975)</td>
</tr>
<tr>
<td>PEST (Pollack, 1968, 1970)</td>
</tr>
<tr>
<td>PROBIT (Finney, 1971)</td>
</tr>
<tr>
<td>R-M (Reed and Muench, 1938)</td>
</tr>
<tr>
<td>R-MSA (Robbins and Monro, 1951)</td>
</tr>
<tr>
<td>S-K (Spearman, 1908; Karber, 1931)</td>
</tr>
<tr>
<td>TSUT (Tsutakawa, 1972, 1980)</td>
</tr>
<tr>
<td>TUP-D (Wetherill, 1975)</td>
</tr>
<tr>
<td>UP-DOWN (Dixon and Mood, 1948)</td>
</tr>
<tr>
<td>WASAN (Wasan, 1969)</td>
</tr>
</tbody>
</table>
experimental design used to gather data, i.e., selection of the x's. The former methods will be characterized as the "analysis" classification and the latter the "design" classification. Some approaches utilize sequential methods in which observations are gathered at more than one time with design decisions usually made after each collection and are thus categorized as sequential; whereas, others pre-plan the entire experiment and are categorized as fixed.

Unfortunately, such a classification is not as simple to use as it sounds. Some approaches can be applied to both parts of the 5 dichotomies presented above. Therefore, some methods can be seen in more than one category of Table 1. The methods given in this table will be either described in more detail in Chapter 1, or referenced.

**Nonparametric Approaches**

Nonparametric statistical methods do not assume any unique functional relationship between the expected response and the independent variable, i.e., no functional relation between $\theta_x$ and $x$ is assumed. Even though nonparametric approaches do not make as restrictive a set of assumptions as classical parametric analyses do, there are nonetheless assumptions usually underlying them.

The most often cited nonparametric method for fixed design is the Spearman-Karber method. This method is easy to apply and has been shown to have both good small-sample (Bross, 1950) and asymptotic (Miller, 1973) properties. Spearman (1908) and Karber (1931) both proposed this method which is used to estimate $L_{50}$ when the
response curve is assumed to be symmetric. The estimate of \( L_{50} \), \( \hat{L}_{50} \), is obtained as follows:

\[
L_{50} = \frac{k+1}{\sum_{j=1}^{k} (\hat{p}_j - \hat{p}_{j-1}) \{1/2(x_j + x_{j-1})\}}
\]

where \( \hat{p}_j = \frac{r_j}{n_j} \) (\( j = 1, 2, \ldots, k \)); \( \hat{p}_0 = 0 \) and \( \hat{p}_{k+1} = 1 \).

It does not assume equal spacing nor equal number of observations per each level of \( x \) observed, though usually an equal number of observations are taken at equally spaced levels of \( x \).

Brown (1961) discusses properties of the Spearman-Karber estimate. Since it is a nonparametric estimator, he is concerned with its efficiency over a wide class of distributions. Brown found high efficiencies when the underlying distribution was either normal or logistic, but asymptotically zero efficiency when the Cauchy distribution (statistical distribution with infinite variance) was the underlying distribution. In another article, Brown (1966) presents an easy to follow guide for designing an experiment when the Spearman-Karber procedure is to be used.

There are two other nonparametric methods that have been competitors to the Spearman-Karber method; however, their popularity does seem to be diminishing in light of recent research. They are the Reed-Muench and Dragstedt-Behrens estimates. The Dragstedt-Behrens method was proposed independently by Dragstedt and Lang in 1928 and by Behrens in 1929. The Reed-Muench method (Reed and
Muench, 1938) is a variant of the Dragstedt-Behrens method. To see the relationship between these three nonparametric methods, the comparison below (based on Miller, 1973) will assume equally spaced intervals for the independent variable and an equal number of observations per each data collect point.

More specifically, let there be $k$ levels of the independent variable $x$, i.e., $x_1, x_2, \ldots, x_k$. Each level is $d$ units apart with $n$ observations to be taken at each level of $x$. Let $P(y=1|x=x_i) = p_i$. The true unknown $p_i$ is estimated by

$$\hat{p}_i = \frac{r_i}{n}$$

where $r_i$ is the sum of positive responses (i.e., $y_i = 1$) at level $x_i$. The unknown underlying curve is $F$, and $p_i = F(x_i)$ are thus points along this curve. With these guidelines, the Spearman-Karber estimator is defined by

$$\hat{u}_{sk} = x_k + \frac{1}{2}d - d \sum_{i=1}^{k} \hat{p}_i.$$

The Dragstedt-Behrens is based on the following:

$$p_j = \frac{\sum_{i=1}^{j} r_i}{\sum_{i=1}^{j} r_i + \sum_{i=j}^{k} (n-r_i)}$$

Then the estimator is defined as
\[ u_{DB} = \begin{cases} x_k - d(k-j) & \text{if there is a } \hat{p}_j = \frac{1}{2} \\ x_k - d(k-j) + df & \text{else} \end{cases} \]

where

\[ j = \max \{ j \mid 1 \leq j \leq k, \sum_{i=1}^{j} r_i \leq \sum_{i=j}^{k} (n-r_i) \} \]

and

\[ f = \frac{\frac{1}{2} - \hat{p}_j}{\hat{p}_{j+1} - \hat{p}_j} \]

The Reed-Muench estimator is defined as

\[ \hat{u}_{RM} = \begin{cases} x_k - d(k-j) & \text{when } \sum_{i=1}^{j} r_i = \sum_{i=k}^{k} (n-r_i) \\ x_k + d(1-\hat{p}_j) - d \sum_{i=1}^{k} \hat{p}_i + x_j(\hat{p}_{j+1} - \hat{p}_j) & \\ 1 + \hat{p}_{j+1} - \hat{p}_j & \text{otherwise} \end{cases} \]

where \( \hat{j} \) is as defined previously.

Thus both \( \hat{p}_{j+1} \) and \( \hat{p}_j \) should be close to 1/2 when \( k \) is large, and \( \hat{u}_{RM} \) is then very close to \( \hat{u}_{SK} \). Also note that when there exists a sample point \( x_j \) such that \( \sum_{i=1}^{j} r_i = \sum_{i=1}^{k} (n-r_i) \), then the Dragstedt-Behrens and Reed-Muench are equivalent. As Miller (1973) states, \( \hat{u}_{DB} \) and \( \hat{u}_{RM} \) will always be within the same adjacent pair of \( x_i \).
Their basic difference is how they interpolate to find their estimator.

Various authors (e.g., Bross, 1950; Finney, 1971) have made small sample comparisons between these three competing estimators. Their results show the superiority of the Spearman-Karber method. As Finney (1971) states, there appears to be no reason to continue use of the other two methods. Miller (1973), not denying the small sample results, shows all three to be asymptotically equivalent.

Recently there has been considerable interest in robust nonparametric estimation. In this area, the goal is to obtain estimates that are not heavily influenced by one or two observations. Miller and Halpern (1980) have studied robust nonparametric estimates for \( L_{50} \). Their work compares asymptotic efficiencies for the Spearman-Karber, trimmed mean, and Tukey biweight estimators. They examined the efficiencies for seven different distributions. The Spearman-Karber was better for the logistic and normal distributions; however, it had zero efficiency for the Cauchy and slash (unit normal divided by a uniform (0,1)) distributions. Overall, the trimmed mean performed the most consistently, and is their recommended estimator when little is known about the underlying distribution.

Thus far we have reviewed only nonparametric fixed design estimators for \( L_{50} \). We now begin a review of nonparametric sequential estimators of \( L_{50} \). For some, we will also examine their properties for \( L_p \), a general percentile along the binary response curve.
There are two major approaches to sequential nonparametric estimation of this type. They are the up-and-down (or staircase) method and the Robins-Monro stochastic approximation method. There are many variants of these two general approaches, and a few of the more important will be reviewed.

The up-and-down method was created by Dixon and Mood (1948). While their initial work assumed the response curve could be modeled by a normal c.d.f., this need not be assumed, as Wetherill (1963) shows. After an initial level of $x$, and an interval, $d$, are chosen, the next level of $x$ to be observed is chosen as follows:

$$x_{i+1} = \begin{cases} x_i + d & \text{if } y_i = 0 \\ x_i - d & \text{if } y_i = 1 \end{cases}$$

(10)

where $y_i$ is the binary outcome for the $i^{th}$ observation. The estimator of the median for $N$ experiments is

$$\hat{L}_{.50} = \frac{1}{N} \sum_{i=2}^{N+1} x_i$$

(11)

Brownlee et al. (1953) studied the small sample properties of this estimate and found it was more efficient than probit analysis (discussed later), and that it performed well for sample sizes as small as 5 to 10. Methods based on the up-and-down procedure have been heavily used in the psychophysical areas, but have not seen heavy usage elsewhere. The major reason for this is the same for most
sequential statistical procedures, i.e., the inconvenience of obtaining only one response at a time. If considerable time is required to determine the outcome of one experiment, then not many observations can be rapidly collected via sequential designs. Even if the outcome time is short, the necessary set-up time for the next experiment may cause one to avoid sequential methods. In psychophysics where often one patient is being tested, such problems do not usually occur. As long as the next design point may be rapidly chosen, it should take about the same time to run a sequential experiment for \( N \) observations as an \( N \) observation fixed design since the subject can perform only one experiment at a time. The real advantage in sequential (or adaptive as it is often called in psychophysics) approaches then is that usually fewer observations need to be collected when \( L_{50} \) is being estimated sequentially as opposed to fixed designs.

There are many variants of the standard up-and-down procedure. Cornsweet (1962) advocates the use of simultaneously running two up-and-down procedures (a double staircase in his terminology). In this procedure two starting points are chosen and the up-and-down procedure is performed from each of these two points. A random choice is made as to which staircase will be used next, i.e., the next observation to be obtained is collected from either the first or second up-and-down procedure on a random basis. Such a procedure has been shown to have value when human subjects are being tested in the psychophysics field. It is important there
not to let the human subjects learn the pattern of observations being presented.

Brownlee et al. (1953) propose a delayed up-and-down method. To avoid large biases caused by starting the sequence far from \( L_{.50} \), they propose the following estimate:

\[
L_{.5i} = \frac{N+1}{\sum_{i=i^*} x_i/(N-i^*+2)}
\]

where \( N \) experiments are run and \( i^* \) is the first observation where both 0's and 1's have now been observed. Thus this delayed method censors some of the original observations that would normally be used in the up-and-down method.

The up-and-down procedure is easy to use and efficient for \( L_{.50} \); however, different points on the response curve are often of interest. Wetherill (1963) introduced a method which has been the basis for a series of techniques known as transformed up-and-down methods. In these methods, the level of the independent variable does not change with each observation. One simple example would have the following set of rules for going up or down from a stimulus level just arrived at.

**Go down if response pattern 11 observed**

\((y=1 \text{ followed by } y=1);\)

**Go up if response pattern 10, or 0 observed**

\((y=1 \text{ followed by } y=0 \text{ or an initial } y=0).\)
These schemes will eventually center around the point where \( P(\text{Up}) = P(\text{Down}) \). Here \( P(\text{Down}) = 0.5 = P(y=1) P(y=1) \) and \( P(\text{Up}) = 0.5 = P(y=1) P(y=0) + P(y=0) \). It can be seen the \( P(y=1) = 0.707 \), and thus this scheme estimates \( L_{0.707} \) in which the levels are equally spaced. More complicated schemes are given for different percentiles. While this extension of Dixon and Mood's up-and-down method allows percentiles other than \( L_{0.50} \) to be estimated, it is difficult to establish rules to estimate an arbitrary point. Also it is now necessary to keep track of all outcomes since arriving at a new design level.

Wasan (1969) presents a modification of the up-and-down procedure based on Derman's (1957) work that allows any \( L_p \) to be estimated. In addition, a decision is made after each observation; therefore, it is not necessary to perform the record keeping needed for Wetherill's transformed up-and-down technique. The decision procedure for choosing the next level of the independent variable \( x \) is as follows:

\[
(13) \quad x_{i+1} = \begin{cases} 
  x_i - d & \text{with probability } \frac{1}{2\alpha} \text{ if } y_i=1 \\
  x_i + d & \text{with probability } 1 - \frac{1}{2\alpha} \text{ if } y_i=1 \\
  x_i + d & \text{with probability } 1 \text{ if } y_i=0 
\end{cases}
\]

where \( \frac{1}{2} \leq \alpha < 1 \), and \( L_\alpha \) is the desired point.

If \( 0 < \alpha < \frac{1}{2} \), the necessary modifications are obvious. Harper (1980) has briefly studied this procedure for \( \alpha = 0.75 \). Using methods based on Brownlee et al. (1953), he found increased bias and error variance for these estimates of \( L_{0.75} \) compared to \( L_{0.50} \).
The up-and-down approaches described above all use fixed step sizes for the independent variable. Taylor and Creelman (1967) proposed a method called Parameter Estimation by Sequential Testing (PEST) that incorporates rules for changing the step size. Their method allows any arbitrary $L_p$ to be estimated. PEST is a complicated procedure utilizing detailed rules for what levels to try next, when to stop the procedure, and how to estimate $L_p$ once the procedure is stopped. Pollack (1968, 1970) studies PEST and makes recommendations for simplifying the procedure. Rendleman et al. (1970) examine Pollack's 1968 version of PEST and compare it to the up-and-down method when $L_{.50}$ is of interest. For this version of PEST, they found it to be inferior to the up-and-down method. While for large samples, they found roughly comparable variability in $L_{.50}$ estimates by the two methods, they found the up-and-down method to have less variability than PEST for small sample sizes.

The Robbins-Monro stochastic approximation method (Robbins and Monro, 1951) is the other major nonparametric sequential approach. The usual form of the Robbins-Monro procedure for binary data is based on $n_i$ subjects tested at $x_i$. The next level of $x$ to be chosen is as follows:

$$x_{i+1} = x_i - \frac{c_i}{n_i} \left( \frac{r_i}{n_i} - p \right)$$

(14)
where

\[ p = \text{desired percentile} \]
\[ c = \text{constant whose optimal value depends on underlying model.} \]

It is easy to see that the step size changes with each observation. The current estimate at each step of \( L_p \) is \( x_{i+1} \). So when the procedure is stopped, the estimate of \( L_p \) is the level of \( x \) to be chosen next.

Wetherill (1975) gives formulas for the asymptotic variance of \( L_p \). Davis (1971) derives optimal values of \( c \) for several common distributions. Cochran and Davis (1964) found the Robbins-Monro method to be somewhat more accurate than Dixon and Mood's up-and-down procedure for estimating \( L_{.50} \). However, as they state, the differences in performance are not sensational. Wetherill (1963) concludes that the estimation procedure is robust with respect to \( c \) varying from the optimal value.

The Robbins-Monro procedure performs well for \( L_{.50} \) and is in close agreement to asymptotic theory for this case. However, Wetherill (1975) reports large biases and variances greatly in excess of asymptotic theory for values away from \( L_{.50} \). When the sequence overshoots the desired point, say \( L_{.75} \), the method requires many observations to downward compensate. As mentioned previously, the Wasan modified up-and-down procedure encountered similar problems away from \( L_{.50} \).

Variations on the above Robbins-Monro procedure have been suggested. Two of these are given in Davis (1971). The first,
attributed to Cochran, replaces the step multiplier $c_i$ by $c$ until both 0 and 1 responses have been detected. Suppose this occurs at observation $i^*$. Then use \[ \frac{c}{1-i^* + 2} \] for $i > i^*$ for the step multiplier. Davis calls this the delayed Robbins-Monro process. The second adaptation was developed by Kesten (1958). Instead of step multipliers $c_i$, he proposed the use of \[ \frac{c}{\delta(i)} \] where $\delta(i)$ is as follows:

$\delta(1) = 1$

$\delta(2) = 1$

(15) $\delta(i) = \left\{ \begin{array}{ll} \delta(i-1) & \text{if } (x_i - x_{i-1}) (x_{i-1} - x_{i-2}) > 0 \\ \delta(i-1) + 1 & \text{otherwise} \end{array} \right.$

For $i > 2$.

Thus the step multiplier is decreased only when the two previous steps were in opposite directions. The idea behind both of these adaptations is to speed convergence when the points being sampled are not close to the desired $L_p$.

With all the sequential methods described above, there are no clear cut stopping rules. In Monte Carlo comparisons of these methods, a total sample size can be fixed in advance, or sample size can be treated as a random variable. In Davis (1971), a fixed sample size is assumed and the following methods compared on the basis of mean squared error for estimation of $L_{.50}$:

1. Robbins-Monro stochastic approximation
2. Delayed Robbins-Monro
3. Kesten version of Robbins-Monro
4. Up-and-down method
5. Delayed up-and-down method
6. Non-sequential Spearman-Karber

For these the following factors are varied:

1. Starting x value (center for Spearman-Karber)
2. Interval between adjacent x values if appropriate
3. Arrangement of observations where appropriate
4. Underlying distribution

While no design is best overall, Davis recommends both the delayed Robbins-Monro and the delayed up-and-down procedure. They perform well over the spectrum of conditions. Under most conditions the Spearman-Karber method performed worse than the five sequential methods.

**Bayesian Methods**

Bayesian approaches to estimation of the binary response have been focused on the analysis of the collected data. Recall that at any point, x, the number of positive responses (y_x = 1) is described as a binomial random variable with probability of a positive response or success (P(y_x) = 1) = \theta_x. The well known relationship between the beta and binomial distributions can be easily used for a Bayesian analysis at a selected x value. More recent work in the Bayesian estimation utilizes data for the entire response curve. Both approaches are briefly described below.

The use of Bayesian methods requires the development of prior distributions. These may be derived from past historical data, but often involve some subjective judgment. Often no
historical data exist for the same setup, and one must utilize his best knowledge to arrive at these priors. When interest is focused on the estimation of $\theta_i$ for a given $x_i$, a beta distribution is commonly used as the prior distribution. This beta distribution is of the following form:

\begin{equation}
    h(\theta) = \begin{cases} 
        \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1}(1-\theta)^{b-1} & \text{for } 0 < \theta < 1 \\
        0 & \text{elsewhere}
    \end{cases}
\end{equation}

The $r_i$ successes observed at $x_i$ will be a binomial random variable.

\begin{equation}
    f(r_i | \theta) = \binom{n_i}{r_i} \theta^{r_i}(1-\theta)^{n_i-r_i} \quad 0 < \theta < 1
\end{equation}

The posterior distribution for $\theta$ will also be beta distributed as indicated below.

\begin{equation}
    g(\theta | r_i) = \frac{\Gamma(n_i+a+b)}{\Gamma(a+r_i)\Gamma(n_i-r_i+b)} \theta^{r_i+a-1}(1-\theta)^{n_i-r_i+b-1} \quad 0 < \theta < 1
\end{equation}

The updated posterior distribution of $\theta_i$ will usually have a smaller variance than the prior distribution as a result of the $n_i$ observations collected. Since the beta is a conjugate prior for the binomial, it is also ideal for sequential estimation where the posterior becomes the prior for the next step.
Mazzuchi and Singpurwala (1981) considered the following situation for the entire response curve:

$$0 \equiv \theta_0 < \theta_1 < \theta_2 < \ldots < \theta_k < \theta_{k+1} \equiv 1$$

They assign (based on work by Ramsey, 1972) a Dirichlet for a prior distribution for the successive differences $\theta_1, \theta_2 - \theta_1, \ldots, \theta_k - \theta_{k-1}$. They were not able to obtain posterior distributions for the $\theta_i$, $i=1, ...k$. Ramsey (1972) has been able to obtain estimates for $(\theta_1, \theta_2, \ldots, \theta_k)$ by using the modal value of the resulting joint posterior distributions. Mazzuchi (1982) has extended this work and can obtain all the moments of the marginal posterior distributions of the $\theta_i, i=1, \ldots, k$. He uses these moments to approximate the posterior distributions. However, Disch (1981) obtains posterior distributions for the $\theta_i, i=1, \ldots, k$, and for any arbitrary $L_p$. The c.d.f. of $L_p$ is a linear combination of incomplete beta functions. Since, in practice, this c.d.f. is usually numerically unmanageable, Disch provides two methods of approximation.

Bayesian analysis has also been used for binary data by Freeman (1970), Lindley (1964), Tsutakawa (1972, 1980), Benardo (1977), Owen (1975), Ferguson (1973, 1974, 1979), and Bhattacharya (1981).

**Classical Parameter Analysis**

The classical fixed design approach to binary response data usually assumes the $\theta_x$ versus $x$ curve may be represented by a cumulative distribution which is normal or logistic. Finney, in his classic book *Probit Analysis* (Finney, 1971) assumes that a normal
c.d.f. fits the binary response curve. Berkson (1944, 1949, 1951) advocated the assumption of an underlying logistic c.d.f. to fit this stimulus-response curve. D. R. Cox, in his book *Analysis of Binary Data* (Cox, 1970), uses the logistic distribution. Most analyses and research documented in the statistical literature now use the logistic distribution (Aranda-Ordaz, 1981). As many parties have observed, the normal and logistic distributions result in nearly identical results. Mathematically, as will be seen below, the logistic is easier to work with than the normal. It has also found favor in the epidemiological fields because of the natural ways in which the odds ratio (explained below) occurs when the logistic distribution is utilized (Kleinbaum, 1982).

The probability of $y = 1$ is a function of the independent variable $x$. Since our response curve is assumed here to be cumulative distribution function, $F(x)$, we have the following relationship:

$$p_i = F(x_i) = \int_{-\infty}^{x_i} f(x_i) \, dx$$  \hspace{1cm} (19)$$

For the normal c.d.f.,

$$p_i = \int_{-\infty}^{x_i} \frac{1}{\sigma \sqrt{2\pi}} \exp \left( - \frac{(x_i - \mu)^2}{2\sigma^2} \right) \, dx$$

$$= \int_{-\infty}^{\lambda_i} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2} \mu^2) \, d\mu = \Phi(\lambda_i)$$
where $\lambda_i = (x_i - \mu)/\sigma$.

For the logistic c.d.f.,

$$(21) \quad p_i = \frac{e^{\alpha + \beta x_i}}{1 + e^{\alpha + \beta x_i}} = \frac{e^\lambda_i}{1 + e^\lambda_i}$$

where $\lambda_i = \alpha + \beta x_i$. Using the log odds ratio below, we obtain a relationship that is much easier to work with.

$$(22) \quad \ln \frac{p_i}{1 - p_i} = \lambda_i = \alpha + \beta x_i$$

No such simplifying transformation is available for the normal distribution. Instead

$$(23) \quad \lambda_i = \varphi^{-1}(p_i) \text{ in this case.}$$

Historically, the estimates for the slope and intercept of the following lines $\lambda_i = (x_i - \mu)/\sigma$ when a normal is assumed or $\lambda_i = \alpha + \beta x_i$ for the logistic were obtained by an eye-ball fit of a straight line to the available data. These initial estimates would often be iteratively updated by using tables of working probits (for the normal distribution) or working logits (for the logistic distribution) (see Finney, 1971, for details). With the advent of computers, these methods were commonly abandoned in favor of an iterative weighted regression approach. The weights to be used are
functions of the unknown parameters, and thus are updated in each iteration as new estimates of the parameters are derived.

Another current approach to estimation of these parameters uses non-linear optimization routines to find the maximum likelihood estimates directly from the log of the likelihood function. Algorithms such as the Fletcher-Powell (1963) or Nelder-Mead (1965) optimization routines are often employed in the search for the maximum of the log likelihood function.

In many binary response situations \( L_{50} \) is the percentile of interest. Normally a fixed design approach is used to estimate the response curve, and from that, \( L_{50} \) can be estimated. Of course, once the underlying c.d.f. has been estimated, any percentile, i.e., an arbitrary \( L_p \), can be easily estimated. Brand et al. (1973) provide asymptotic confidence bands for an arbitrary logistic \( L_p \). Finney (1971) concentrates on confidence bands for \( L_{50} \).

Most literature in this area is concerned with analysis rather than experimental design. In comparison to the work on the continuous response problem, very little research has been done in the design of experiments for binary response problems. The binary response model has much in common with nonlinear models, as we will see in a later chapter.

Finney (1978) provides fixed experimental designs geared at minimizing the fiducial confidence bands for \( L_{50} \) for both the logistic and normal distributions. He examines the situation where 2, 3, or 4 distinct levels of \( x \) are used. The optimal selection of the \( x \) locations for these three cases depends upon the total sample
size to be used for the experiment. \( L_{50} \) is often used in biostatistics for the comparison of two treatments. In bioassay, it is assumed that the two transformed response curves are parallel (implying the underlying c.d.f.'s have equivalent standard deviations), and then the difference between the \( L_{50} \) estimates for each treatment is used to measure the relative effectiveness of the treatment.

**Problem of Interest**

For this research we will develop and assess methods that primarily fall into the category global-classical-parametric-design-sequential. The algorithms developed will be compared against published work in the global-classical-parametric-design-fixed category, i.e., the main thrust of this work is to develop and compare adaptive (sequential) algorithms against fixed designs that are based on the same underlying assumptions. There will be one independent variable, time. One interesting aspect of this problem is that one can only move forward in time. We assume that the logistic distribution defines the functional relationship between \( \theta_X \) and \( x \). We desire estimates of \( \alpha, \beta \) that describe this logistic distribution to be optimal in a global sense. D-optimality, to be detailed in Chapter 2, will be our criterion for comparing various designs to accomplish this goal. In addition, Chapter 2 will detail particular aspects of the logistic distribution for this problem.
CHAPTER 2 OPTIMAL DESIGN AND LOGISTIC PROPERTIES

Introduction

Chapter 1 surveyed the standard methods found in the literature for the analysis and/or design of binary experiments. This chapter will detail various aspects of the problem briefly mentioned at the end of Chapter 1, and include a more detailed explanation of this problem. In the following sections, various optimal methods will be introduced, and equivalence theorems relating some of them will be presented. Next, properties of the logistic distribution will be covered and these will be shown to tie in nicely with the equivalence theorems presented.

Optimality Criteria - Survey of Methods

Optimality theory for the design of experiments has generally been focused on linear models of the following form:

(24) \[ y = X \beta + \epsilon \]

where

- \( y \) is an \( n \times 1 \) vector of observations
- \( X \) is an \( n \times p \) matrix containing \( n \) rows with row \( i \) containing the independent variable settings corresponding to observation \( y_i \)
$\beta$ is a $p \times 1$ vector of unknown parameters
$\xi$ is an $n \times 1$ vector of independent, identically distributed random variables with expected value zero and a common variance $\sigma^2$.

Some of the most important results in this area have been based on continuous normalized designs in which the design is viewed as a probability measure, $\xi$, on the potential experimental region, $X$, of $X$. This design measure satisfies the following (St. John and Draper, 1975):

\[
\xi(x) \geq 0 \quad , \quad x \in X
\]

\[
\int_{X} \xi(dx) = 1
\]

There need not be an integer valued number of observations at $x_i$. For exact or discrete n-point designs actually used in practice, $\xi(x)$ must have values that are multiples of $1/n$. Thus, we have the following situation:

\[
\sum_{i=1}^{n} \xi(x_i) = 1 \quad \text{with} \quad \xi(x_i) = n_i / n
\]

where $n_i$ observations are observed at $x_i$.

For the models above, $E(y_i | x_i) = f'(x_i)\beta$, where $f'(x_i)$ (1 x p vector) depends on the form of the response function assumed. The Fisher information matrix is defined (Federov; 1972) as
For discrete designs, \( n \mathbf{M}(\xi) = \mathbf{X}'\mathbf{X} \). We now consider some of the more widely used measures of optimality for experimental designs. \( \xi^* \) is D-optimal if and only if \( \mathbf{M}(\xi^*) \) is nonsingular and \( |\mathbf{M}(\xi^*)| = \sup_{\xi \in \mathcal{X}} |\mathbf{M}(\xi)| \). This functional form is the inverse of the generalized variance criterion. In the literature (Whittle, 1973; Silvey, 1980), this is sometimes replaced by \( \sup_{\xi \in \mathcal{X}} \log |\mathbf{M}(\xi)| \). It is obvious that \( \xi^* \) is the same in both cases.

Another optimality measure for experimental designs is G-optimality. G-optimality is concerned with the accuracy of the estimated response surface rather than the parameters \( \mathbf{\beta} \) as is D-optimality. It is often referred to as the mini-max criteria for reasons to be seen below. Let

\[
(28) \quad d(x, \xi) = f(x)'M_1(\xi)f(x)
\]

where \( \mathbf{M}(\xi) \) is again assumed to be nonsingular. A design measure \( \xi^* \) is G-optimal if it satisfies the following

\[
(29) \quad \inf_{\xi} \sup_{x \in \mathcal{X}} d(x, \xi) = \sup_{x \in \mathcal{X}} d(x, \xi^*)
\]

Trace-optimality or A-optimality is an average variance criterion in which the functional form used is the trace of \( \mathbf{M}^{-1}(\xi) \). This criterion has been proposed by Elfving (1952) and Chernoff (1953). In an exact design, the trace of \( \mathbf{M}^{-1}(\xi) \) is proportional to the sum of the parameter variances since \( \mathbf{M}^{-1}(\xi) = (\mathbf{X}'\mathbf{X})^{-1} \) in this instance, and the \( \text{cov}(\mathbf{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \).
E-optimality, proposed by Ehrenfeld (1955), involves maximization of the minimum eigenvalue of $M(\xi)$. $D_s$-optimality is based on D-optimality criteria when only a subset ($s < p$) of the $\mathbf{R}$ parameters are of interest. Other optimality criteria have also been proposed (Federov, 1972; Silvey, 1980) for experimental design.

Kiefer (1958, 1959, 1961a, 1961b, 1962a, 1962b) and Kiefer and Wolfowitz (1959, 1960) developed the foundation of the theory of optimal experimental designs based on the design measure $\xi$. Their work resulted in the General Equivalence Theorem (G.E.T.) that proved the equivalence of different optimality criteria for design measures. The optimal designs that are arrived at using the design measures can then be used as the basis for discrete designs used in practical situations. The G.E.T. for linear models such as we have discussed thus far in this chapter states that for design measures $\xi$, a design $\xi^*$ is D-optimal if and only if it is G-optimal. The G.E.T. also shows that for $p$ parameters in $\mathbf{R}$

$$\max_{x \in \mathbf{X}} d(x, \xi^*) = p$$

This last condition is often used to verify that a particular measure is indeed optimal.

Extensions of the Kiefer-Wolfowitz General Equivalence Theorem

For linear models with a homogeneous variance, $\sigma^2$, the above theory has proved invaluable in the attainment of optimal experimental designs. However, many situations do not satisfy the
above stated assumptions. Therefore, extensions to the Kiefer-Wolfowitz G.E.T. were sought. Federov (1972) in his classic book, Theory of Optimal Experiments, provides one of the more important extensions. Federov generalizes $M(\xi)$ as follows:

\begin{equation}
M(\xi) = \int \lambda(\mathbf{x}) f(\mathbf{x}) f'(\mathbf{x}) \xi(d\mathbf{x})
\end{equation}

He calls $\lambda(\mathbf{x})$ an efficiency function and assumes it is a known function. Under this generalization, he develops an extension to the G.E.T. in which he proves the following three statements to be equivalent:

1. The design measure $\xi^*$ maximizes $|M(\xi)|$.

2. $\xi^*$ minimizes $\max_{\mathbf{x} \in \mathcal{X}} \lambda(\mathbf{x}) d(\mathbf{x}, \xi)$.

3. $\max_{\mathbf{x} \in \mathcal{X}} \lambda(\mathbf{x}) d(\mathbf{x}, \xi^*) = \rho$.

For models where $E(y|\mathbf{x}) = n(\mathbf{x}, \mathbf{\theta})$ is a nonlinear function of $\mathbf{\theta}$, the Kiefer-Wolfowitz G.E.T. does not directly apply. Parameter estimates are not easily obtainable as they are for linear models, and the optimal design is now a function of the unknown parameters $\mathbf{\theta}$. Therefore, we cannot a priori construct globally optimal designs; however, locally optimal designs may be devised that are optimal in a small neighborhood in the range of parameter values.
Chernoff (1953) introduced the concept of asymptotically optimal designs. Since the true parameter covariance matrix is usually intractable for nonlinear models, Chernoff suggested the use of the Fisher Information matrix. When maximum likelihood estimates, \( \hat{\theta} \), are obtained for \( \theta \), the inverse of the Fisher Information matrix is the asymptotic covariance matrix for \( \hat{\theta} \). In practice for parameter estimation using nonlinear models, maximum likelihood estimates and the associated Fisher Information matrix are the most commonly used approach. Thus, it is reasonable to consider using the same tools to obtain asymptotically optimal designs for these nonlinear models. Chernoff (1953) and others (Box and Lucas (1959); Box and Hunter (1965)) have obtained locally optimal designs by linearizing the model around preliminary values \( \theta_0 \) using Taylor series approximations. The optimality theory mentioned previously can be utilized on the linearized model.

While the use of methods based on the Kiefer-Wolfowitz G.E.T. for linearized versions of nonlinear models is a practical approach to the development of optimal experimental designs, others have developed general equivalence theorems for nonlinear models. The most well known of these is White (1973). In this work, White extends the G.E.T. to nonlinear models by generalizing \( M(\xi) \) and \( d(x, \xi) \). Letting \( M(\xi, \theta) \) be the Fisher Information matrix for \( \theta \), a design measure, \( \xi^* \), is called \( D(\theta) \)-optimal if

\[
(32) \quad |M(\xi^*, \theta)| = \max_{\xi} |M(\xi, \theta)|
\]
for \( \beta \) having its true value. This is the same criterion as Chernoff (1953) presented. Thus

\[
M(\xi, \beta) = \int X I(x, \beta) \xi(dx)
\]

where

\[
I(x, \beta) = \left( \begin{array}{c}
-E(\beta_1^2 \log f/\beta_1^2) & \cdots & -E(\beta_1 \beta_p) \\
& \cdots & \\
-E(\beta_p \beta_1) & \cdots & -E(\beta_p^2)
\end{array} \right)
\]

and \( f(y; x, \beta) \) is the conditional probability density function for the random variable \( y \). White introduces an analogue of G-optimality for nonlinear models. She defines this nonlinear analogue as

\[
d(x, \xi, \beta) = \text{tr} \ I(x, \beta) M^{-1}(\xi, \beta)
\]

where \( \text{tr} \) denotes the trace of the matrix. Under this definition, \( \xi^* \) is G(\( \beta \))-optimal if

\[
\sup_{\beta} d(x, \xi^*, \beta) = \min_{\xi} \sup_{\beta} d(x, \xi, \beta)
\]

for \( \beta \) taking its true value. White's nonlinear general equivalence theorem states that D(\( \beta \))-optimality and G(\( \beta \))-optimality are equivalent for a design measure \( \xi \), and that \( \sup_{\xi} d(x, \xi, \beta) = p \). More will be given on this general equivalence theorem in later sections.

Ford (1976) provides another general equivalence theorem for nonlinear models. He utilizes the same \( M(\xi, \beta) \) that White used.
For G-optimality, he uses the first order approximation suggested by Chernoff (1953). That is

\[
(36) \quad \text{var} \eta(x, \hat{\beta}) \approx \eta_{\hat{\beta}}(x, \hat{\beta}) M(\xi, \hat{\beta})^{-1} \eta_{\hat{\beta}}(x, \hat{\beta})
\]

where

\[
\eta_{\hat{\beta}}(x, \hat{\beta}) = \left( \frac{\partial \eta(x, \hat{\beta})}{\partial \beta_1}, \ldots, \frac{\partial \eta(x, \hat{\beta})}{\partial \beta_p} \right)
\]

His general equivalence theorem may be summarized as follows:

An optimal design measure \( \xi^* \) can be equivalently characterized by:

1. \( \xi^* \) maximizes \( \log |M(\xi, \hat{\beta})| \)
2. \( \xi^* \) minimizes \( \frac{1}{\sigma^2} \max_{x \in X} \text{var} \eta(x, \hat{\beta}) - p \)
3. \( \frac{1}{\sigma^2} \max_{x \in X} \text{var} \eta(x, \hat{\beta}) = p \)

where \( \sigma^2 = \text{var} (y) \).

When \( \eta(x, \hat{\beta}) = E(y|x) \) and \( I(x, \hat{\beta}) = \frac{1}{\text{var}(y|x)} \eta_{\hat{\beta}}(x, \hat{\beta}) \)
\( \eta_{\hat{\beta}}(x, \hat{\beta}) \), the \( \sigma^2 \) in 2 and 3 above may be replaced by \( \text{var}(y|x) \). In this case, the function given in 2 and 3 (ignoring \( p \)) is approximately equal to \( \frac{\text{var} \eta(x, \hat{\beta})}{\text{var}(y|x)} \).

In the next sections we will examine the particular properties of the form of \( \eta(x, \hat{\beta}) \) related to our binary problem of interest. These properties will be used, where appropriate, on the material we have presented in this section. In particular, we will
examine the extensions of the Kiefer-Wolfowitz G.E.T. given by Federov, White, and Ford for our particular problem.

Properties of the Logistic Distribution

The problem we will study in detail will be that of a binary random variable \( y \) as a function of one independent variable \( x \). We assume that this relationship follows a logistic distribution. In this section, properties of the logistic distribution are presented. More specific properties of the logistic distribution that relate to the optimality criteria presented earlier will be given later.

The logistic binary random variable \( y \) has the following probability density function for a single independent variable \( x \):

\[
\begin{align*}
 f(y) &= \left[ \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \right]^y \left[ \frac{1}{1 + \exp(\alpha + \beta x)} \right]^{1-y} \\
&\text{for } y = 0, 1
\end{align*}
\]

From this we see that \( P(y_i = 1) = \theta_i = E(y_i) = \exp(\alpha + \beta x_i)/(1 + \exp(\alpha + \beta x_i)) \) and \( P(y_i = 0) = 1 - \theta_i = 1/(1 + \exp(\alpha + \beta x_i)) \). Thus

\[
\begin{align*}
y_i &= \begin{cases} 
0 \text{ with probability } 1 - \theta_i \\
1 \text{ with probability } \theta_i
\end{cases} \\
\Rightarrow 0 \leq E(y_i) = \theta_i \leq 1
\end{align*}
\]
By performing what is usually called the logit transform (or log odds ratio), the following is obtained:

\[ \lambda_i = \ln\left(\frac{\theta_i}{1 - \theta_i}\right) = \alpha + \beta x_i \]

The relationship between \( y \) and \( x \) is nonlinear in terms of the parameters; however, the above log odds ratio results in a model linear in \( \alpha \) and \( \beta \). This is an example of a generalized linear model (McCullough and Nelder; 1983). This linearity is between \( \lambda \) and \( x \) and not between the original dependent variable \( y \) and \( x \). The log odds ratio, \( \ln\left(\frac{\theta}{1-\theta}\right) \), is called the link function that results in a linear relationship between a function of \( E(y|x) = \theta_x \) and \( x \). In practice, one must obtain estimates of \( \alpha \) and \( \beta \) from the available data. Typically, maximum likelihood estimation is used to obtain the maximum likelihood estimates \( \hat{\alpha}, \hat{\beta} \). For \( n \) independent observations, the likelihood function \( L(\alpha, \beta) \) is defined as follows:

\[ L(\alpha, \beta) = \prod_{i=1}^{n} \left[ \exp(\alpha + \beta x_i) \right]^{y_i} \left[ \frac{1}{1 + \exp(\alpha + \beta x_i)} \right]^{1-y_i} \]

\[ \log L(\alpha, \beta) = \sum_{i=1}^{n} \log\left[ \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right] \]

\[ + \sum (1-y_i) \left[ \log \frac{1}{1 + \exp(\alpha + \beta x_i)} \right] = \sum_{i=1}^{n} y_i (\alpha + \beta x_i) - \sum \log \left( 1 + \exp(\alpha + \beta x_i) \right) \]
Maximum likelihood estimates $\hat{\alpha}, \hat{\beta}$ must satisfy the following set of equations:

\[
\begin{align*}
\frac{\partial \log L(\alpha, \beta)}{\partial \alpha} \bigg|_{\alpha = \hat{\alpha}} &= 0 \\
\frac{\partial \log L(\alpha, \beta)}{\partial \beta} \bigg|_{\beta = \hat{\beta}} &= 0.
\end{align*}
\]  

(42)

Pursuing this, we find

\[
\begin{align*}
\frac{\partial \log L(\alpha, \beta)}{\partial \alpha} &= E y_i - E \frac{\exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} \\
\frac{\partial \log L(\alpha, \beta)}{\partial \beta} &= E x_i y_i - E \frac{x_i \exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)}.
\end{align*}
\]  

(43)

It is not possible to set the above two equations equal to zero and find simple closed formed solutions for $\hat{\alpha}, \hat{\beta}$. Nonlinear optimization routines are used to search for the maximum of the log likelihood to find $\hat{\alpha}, \hat{\beta}$. The asymptotic covariance matrix for these estimates is the inverse of the Fisher Information matrix $M(\alpha, \beta)$, where $M(\alpha, \beta)$ is defined as:

\[
M(\alpha, \beta) = E \begin{bmatrix}
-\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha^2} & -\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} \\
-\frac{\partial^2 \log L(\alpha, \beta)}{\partial \alpha \partial \beta} & -\frac{\partial^2 \log L(\alpha, \beta)}{\partial \beta^2}
\end{bmatrix}.
\]  

(44)

Carrying out these calculations results in
(45) \[
M(\alpha, \beta) = \begin{bmatrix}
\sum \frac{\exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} & \sum \frac{x_i \exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} \\
\sum \frac{x_i \exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} & \sum \frac{x_i^2 \exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)}
\end{bmatrix}
\]

Since we have shown \( E(y_i) = \theta_i = \frac{\exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} \) and \( (1 - \theta_i) = \frac{1}{1 + \exp (\alpha + \beta x_i)} \), we may rewrite \( M(\alpha, \beta) \) in the following less cumbersome form:

\[
M(\alpha, \beta) = \begin{bmatrix}
\sum \theta_i (1 - \theta_i) & \sum x_i \theta_i (1 - \theta_i) \\
\sum x_i \theta_i (1 - \theta_i) & \sum x_i^2 \theta_i (1 - \theta_i)
\end{bmatrix}
\]

or \( M(\alpha, \beta) = \begin{bmatrix}
\sum f(x_i, \alpha, \beta) & \sum x_i f(x_i, \alpha, \beta) \\
\sum x_i f(x_i, \alpha, \beta) & \sum x_i^2 f(x_i, \alpha, \beta)
\end{bmatrix}\)

where \( f(x_i, \alpha, \beta) = \frac{\exp (\alpha + \beta x_i)}{1 + \exp (\alpha + \beta x_i)} \). This latter form is similar to weighted linear models where each observation has a weight associated with it; however, here the weight is a function of the unknown parameters \( \alpha, \beta \) that we wish to estimate.

The curve \( E(y_x) = \theta_x \) versus \( x \) has the shape depicted at the beginning of Chapter 1. That is, it is a nondecreasing function of \( x \), and \( \theta_x \) is bounded by 0 and 1. At any point \( x_i \) along this curve the random variable \( y_i \) may be seen to be a binomial random variable with probability success, \( P(y_i = 1) \), equal to \( \theta_i \). The variance for an individual observation at \( x_i \) will be \( \theta_i (1 - \theta_i) \). Thus, we see in \( M(\alpha, \beta) \) that each observation is weighted by its variance \( \theta_i (1 - \theta_i) \).
$M(\alpha, \beta)$ given above is the Fisher Information matrix for an exact or discrete design based on $n$ observations. We now structure this formulation to fit the design measure approach described earlier in this chapter. First, we assume there are $k$ distinct $x_j$, $j = 1, 2, \ldots, k$ at which we consider taking observations. Then we define

(47)

$$M(\xi, \alpha, \beta) = \begin{pmatrix}
\sum_{j=1}^{k} p_j \theta_j (1-\theta_j) & \sum_{j=1}^{k} p_j x_j \theta_j (1-\theta_j) \\
\sum_{j=1}^{k} p_j x_j \theta_j (1-\theta_j) & \sum_{j=1}^{k} p_j x_j^2 \theta_j (1-\theta_j)
\end{pmatrix}$$

Where $\sum_{j=1}^{k} p_j = 1$. Later we will show how this form can be used in some of the extensions of the Kiefer-Wolfowitz General Equivalence Theorem. Now let us examine some additional properties of $M(\xi, \alpha, \beta)$. First, the determinant of $M(\xi, \alpha, \beta)$ may be succinctly expressed.

(48)

$$|M(\xi, \alpha, \beta)| = \left[ \sum_{j=1}^{k} p_j \theta_j (1-\theta_j) \right] \left[ \sum_{j=1}^{k} p_j x_j^2 \theta_j (1-\theta_j) \right] - \left[ \sum_{j=1}^{k} p_j x_j \theta_j (1-\theta_j) \right]^2$$

By multiplying the above factors and collecting like terms, this reduces to

(49)

$$|M(\xi, \alpha, \beta)| = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} p_i p_j c_{ij}$$

where $c_{ij} = (x_i - x_j)^2 \theta_i \theta_j (1-\theta_i)(1-\theta_j)$. 
If one were interested in D-optimality in which \(|M(\xi, \alpha, \beta)|\) was to be maximized, and if the true parameter values \(\alpha, \beta\) were known, then all the quantities \(a_{ij}\) would be known. Therefore, the decision variables would be the \(p_i, i=1, 2, \cdots, k\). Methods for selection of \(p_i, i=1, 2, \cdots, k\) will be addressed later.

Use of Logistic Properties in Optimal Design Theory

For the logistic distribution, \(E(y_x) = \theta_x\) is the function we would like to estimate. Letting

\[
(50) \quad \eta(x, \beta) = \eta(x, \alpha, \beta) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)},
\]

we see that \(E(y_x) = \eta(x, \beta)\). We now study how the Ford, White, and Federov extensions to the Kiefer-Wolfowitz G.E.T. can be applied. For D-optimality, all three of these extensions involve maximization of \(|M(\xi, \alpha, \beta)|\) or \(\log |M(\xi, \alpha, \beta)|\). For G-optimality or analogues to G-optimality, each has a different approach. We will examine how the logistic fits each of these.

For Ford's work, we must examine

\[
(51) \quad \eta_\beta(x, \beta) = \left[ \frac{\partial \eta(x, \beta)}{\partial \alpha}, \frac{\partial \eta(x, \beta)}{\partial \beta} \right]'.
\]

Carrying out these partial derivatives, we find

\[
(52) \quad \eta_\beta(x, \beta) = \left[ \frac{\exp(\alpha + \beta x)}{(1 + \exp(\alpha + \beta))^2}, \frac{x \exp(\alpha + \beta x)}{(1 + \exp(\alpha + \beta))^2} \right]'.
\]
Now we must determine \( I(x, \beta) \). Based on the work presented so far, we see that \( I(x, \beta) \) is the Fisher Information matrix for one point \( x \).

Then

\[
I(x, \beta) = \begin{pmatrix}
\theta_x (1-\theta_x) & x \theta_x (1-\theta_x) \\
 x \theta_x (1-\theta_x) & x^2 \theta_x (1-\theta_x)
\end{pmatrix}
\]

\[= \theta_x (1-\theta_x) \begin{pmatrix}
1 & x \\
x & x^2
\end{pmatrix} = \text{var}(y|x) \begin{pmatrix}
1 & x \\
x & x^2
\end{pmatrix}
\]

We need to show that \( I(x, \beta) = \frac{1}{\text{var}(y|x)} \eta_\beta(x, \beta) \eta_\beta(x, \beta)' \).

Rewriting \( \eta_\beta(x, \beta) \) as \( \text{var}(y|x)(1,x)' \), we immediately see that

\[
\eta_\beta(x, \beta) \eta_\beta(x, \beta)' = \text{var}(y|x)^2 \begin{pmatrix}
1 & x \\
x & x^2
\end{pmatrix},
\]

and thus \( I(x, \beta) = \frac{1}{\text{var}(y|x)} \eta_\beta(x, \beta) \eta_\beta(x, \beta)' \). This last issue was necessary so that we could use Ford's more general version of G-optimality. Now

\[
\frac{1}{\text{var}(y|x)} \eta_\beta(x, \beta)' M^{-1}(\xi, \beta) \eta_\beta(x, \beta) \text{ is approximately equal to }
\]

\[
\frac{1}{\text{var}(y|x)} \text{var}(\hat{x}, \beta). \]

Letting \( M^{-1}(\xi, \beta) = \begin{bmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{bmatrix} \), we see that

\[
\frac{1}{\text{var}(y|x)} \eta_\beta(x, \beta)' M^{-1}(\xi, \beta) \eta_\beta(x, \beta) =
\]

\[
\frac{1}{\text{var}(y|x)} (\text{var}(y|x))^2 (1,x) \begin{bmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} =
\]

\[
\text{var}(y|x) (M_{11} + 2xM_{12} + x^2M_{22}).
\]
White's nonlinear general equivalence theorem uses the

$$\text{tr}(I(x, \theta) M^{-1}(\xi, \theta))$$

as the G-analogue where tr denotes the trace:

\begin{equation}
\text{tr}(I(x, \theta) M^{-1}(\xi, \theta)) = \text{var}(y|x) \begin{pmatrix}
1 & x \\
x & x^2
\end{pmatrix} \begin{pmatrix}
M_{11} & M_{12} \\
M_{12} & M_{22}
\end{pmatrix}
\end{equation}

$$= \text{var}(y|x)(M_{11} + 2xM_{12} + x^2M_{22})$$.

Thus, for the logistic distribution, the G-optimality analogue of Ford and White are the same function. As we stated earlier, both use the same D-optimality criteria; therefore, for the logistic distribution we use, both Ford and White's general equivalence theorems are identical.

Re-examining Federov's extension to the Kiefer-Wolfowitz General Equivalence Theorem, we see that careful selection of \(\lambda(x)\) will allow the logistic distribution to fit his general equivalence theorem. In his book, *Theory of Optimal Experiments*, he assumes that \(\lambda(x)\) is proportional to \(\frac{1}{\text{var}(y|x)}\); however, this is not needed in any of the proofs he provides. Indeed, if we set \(\lambda(x) = \text{var}(y|x)\), then the necessary conditions are satisfied and the logistic model satisfy his G.E.T. extension. It should also be pointed out that Federov (page 188) uses an analogue of G-optimality for the general nonlinear models which, for the logistic, is equivalent to Ford and White's G-optimality analogue.
At this point, it has been shown that not only does the logistic distribution satisfy the general equivalence theorems of Ford, White, and Federov, but that all three are identical in this case. Therefore, we can proceed comfortably to focus on either D-optimality or G-optimality realizing that the results we obtain will be optimal for both criteria when design measures $\xi$ are considered.

Experimental Design Problem to be Studied in Detail

The main focus of the following chapters will be on the experimental design problem we now discuss. This problem involves a binary random variable $y$ that is assumed to follow a logistic distribution. It is a function of one independent variable $x$. There are $k$ distinct values of this independent variable, $x_i$, $i=1, 2, \ldots, k$, at which we may collect observations. The independent variable will be time. This creates an interesting situation since once the experiment begins, the problem is a "forward only" one, i.e., one cannot go backward in time to collect data at a point in time already passed.

A specific example of the above would involve the degradation of a canned milk product. This product has a finite shelf life due primarily to physical stability cosmetic defects. A selected can of the product at given point in time will either have or not have this cosmetic defect. The probability of the defect increases with the age of the product. Batches of this product undergo frequent compositional changes for economic or regulatory reasons or in hopes of increasing the shelf life. While numerous pilot plant
batches are made to test these new formulations, the first production batch is very closely monitored. It is desired that the degradation curve of this product be estimated from the first batch. While past batches provide good information on the functional relationship of degradation over time, the parameters of this function need to be estimated accurately based on this first batch. A specified number of cans, say N, of this batch will be set aside for estimation of these parameters. The goal is to maximize the information available from these N experiments subject to the constraints of the problem. The canned samples may be examined at either the time of manufacture or at the end of each subsequent month over the duration of the study. Thus a finite number of sampling intervals are available. Since time is the independent variable, we have "forward only" sampling capabilities. Since the response variable is binary with a time increasing probability of having the defect, this relationship follows the plot given early in Chapter 1.

Our main focus will be on D-optimality (and, thus, G-optimality by the general equivalence theorems). In Chapter 3, published results for optimal designs for the logistic distribution are presented. These are based on perfect knowledge of the parameters $\alpha, \beta$ and fixed designs. We will present new algorithms and make comparisons with these published results when perfect knowledge of $\alpha, \beta$ is not available; rather, we have initial estimates $\alpha_0, \beta_0$. Also, our focus will be on sequential designs where we update our estimates of $\alpha, \beta$ as new data are available. We will also examine situations where constraints, such as a minimum number of samples,
\( \epsilon_i \), must be taken at design point \( x_i \). Such constrained situations are not amenable to present algorithms. The objectives of this work are the following:

1. Develop and compare new algorithms based on D-optimality for this problem.
2. Determine the sensitivity of the algorithm responses to the variables used in the experimental designs of later chapters.
3. Make recommendations for the use of the algorithms.

These algorithms to be compared will be discussed in the next chapter. Additional properties of the logistic distribution will be derived, where appropriate, for these methods. Some approaches will be examined analytically; whereas, others will require computer simulations for comparative studies.
CHAPTER 3  SEQUENTIAL DESIGN ALGORITHMS FOR THE LOGISTIC

Introduction

Chapter 1 provided a review of the literature in the design and analysis of binary response experiments. As was seen there, many approaches exist. The choice as to which technique is most appropriate depends on the aim of the experimenter. The experimenter must decide on his/her goals, i.e., what aspects of the response curve on which to focus. In addition, the assumptions that can be made about the relationship of the dependent binary variable, \( y \), and the independent variable, \( x \), need to be made clear. The number of observations that can be used for the experiment, any constraints on the allocation of observations, and the cost of different experimental plans must be determined prior to the start of data collection. Under cost considerations, one must consider the potential cost differential of observations having different levels of the independent variable, and the cost differences of performing a fixed (entire experiment pre-planned) versus sequential design. Set-up costs are thus incorporated into such pre-experiment study.

Chapter 2 introduced the concept of optimal experimental designs, primarily D-optimal and G-optimal plans. The well-known Kiefer-Wolfowitz General Equivalence Theorem and several extensions
proving the equivalence of various forms of D- and G-optimal designs for design measures, $\xi$, for both linear and nonlinear models were presented. Properties of the logistic model for binary response experiments were investigated. These properties were then related to the Federov, White, and Ford extensions of the General Equivalence Theorem. While the D-optimality criteria were obviously the same for these 3 extensions, the G-optimality criteria, in general, were different. However, for the logistic model that we are interested in, it was shown that these different G-optimality conditions resulted in the optimization of the same function. Thus, the Federov, White, and Ford theorems are the same for our particular problem.

In this chapter, we present a variety of approaches to the optimal design of experiments for the logistic model. Theory developed by White, Ford, and Silvey for the optimal design of fixed experiments will be shown. The properties of this approach will be presented. Next we present various sequential design algorithms and their associated properties. As explained in Chapter 2, optimal nonlinear designs depend on our knowledge of the true unknown parameters $\alpha$ and $\beta$. The main purpose of this work is to compare new proposed sequential algorithms to the fixed optimal design theory of White, Ford, and Silvey. When our initial knowledge of these parameters is poor, we expect the sequential methods to be superior; whereas, when our initial knowledge is good, both approaches should perform well.
Fixed Optimal Experimental Design for the Logistic Model

Analytic solutions for the construction of optimal experimental designs are possible only in relatively simple problems. Each distinct problem requires special consideration. Most solutions presented in the literature arise from special considerations such as symmetry of the allocation of design points (Federov, 1972). Then the optimality of the proposed design can be verified from the general equivalence theorems. For the logistic model, White (1973, 1975), Ford (1976), and Silvey (1980) present a method for obtaining a 2 point (2 distinct levels of the independent variable \( x \)) D-optimal design for the logistic model. This optimal design is based on knowing exactly the parameters of the underlying logistic distribution, \( \alpha \) and \( \beta \).

As proposed in their work, they consider the binary response model where

\[
P(y=1|x, \alpha, \beta) = F(\alpha + \beta x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}
\]

where \( F(\alpha + \beta x) \) is the cumulative distribution function for the logistic distribution. They assume \( f(\alpha + \beta x) \), is centered about the origin, and consider a D-optimal 2-point design. They show that the optimal design places equal weight at

\[
\frac{\pm \alpha^* - \alpha}{\beta}
\]

where \( \alpha^* = 1.5434 \), if these points lie in the range of \( x = [-1, 1] \).

From this one can show that \( \lambda_x = \alpha + \beta x \) must equal \( \pm 1.5434 \) for this
to hold. Looking at this from the dependent variable point of view where $E(y|x) = \theta_x$, and recalling that $\lambda = \log \frac{\theta_x}{1-\theta_x}$, we see that for $\lambda = \pm 1.5434$,

$$\frac{\theta_x}{1-\theta_x} = e^\lambda \quad \text{or}$$  

$$\theta_x = \frac{e^\lambda}{1+e^\lambda}, \text{ so that}$$  

$$\theta_{x_1^*} = \frac{e^{-1.5434}}{1 + e^{-1.5434}} = .176$$  

$$\theta_{x_2^*} = \frac{e^{1.5434}}{1 + e^{1.5434}} = 1-\theta_{x_1^*} = .824$$

where $\theta_{x_1^*}$ and $\theta_{x_2^*}$ denote the points on the dependent variable axis corresponding to the optimal $x_1^*$ and $x_2^*$ locations. These points corresponding to $\theta_{x_1^*}$ and $\theta_{x_2^*}$ are independent of the values of $\alpha$ and $\beta$. Of course, this assumes the $x_1^*$ and $x_2^*$ are both in the range of $x$.

If this is not true and $x$ is symmetric about $\chi$, the mean of the logistic, then the optimal $x$ values are at the endpoints of $x$.

Ford generalizes this to include the case where the distribution is not symmetric about the origin implying that $x$ is not symmetric about the logistic mean $-\frac{\alpha}{\beta}$, Ford also shows that the optimal design will fall in one of three regions depending on the value of $\alpha$ and $\beta$. For this work, he assumes that $x \in [-1,1]$. He considers only the first quadrant of the parameter space where $\alpha, \beta > 0$. 
Theorem 1. Let $x_1^*, x_2^*$ be the 2 point D-optimal design points in the first quadrant of the design space $(\alpha, \beta > 0)$. Then the 2 point D-optimal design points are $x_1^*, x_2^*$ in the third quadrant, and $-x_1^*, -x_2^*$ in the second and fourth quadrant.

Proof. The points $x_1^*, x_2^*$ maximize the determinant of the Fisher information matrix below for $\alpha, \beta > 0$.

\[
|M(\xi)| = \frac{1}{4} \theta_1 \theta_2 (1-\theta_1)(1-\theta_2) (x_1-x_2)^2
\]

\[
= \frac{1}{4} V(x_1, \alpha, \beta) V(x_2, \alpha, \beta) (x_1-x_2)^2
\]

where $V(x, \alpha, \beta) = \frac{\exp(\alpha+\beta x)}{(1+\exp(\alpha+\beta x))^2} = \frac{e^{\lambda}}{(1+e^{\lambda})^2}$

is the variance of the $E(y|x)$. For the third quadrant where $\alpha, \beta < 0$, consider $V(x, -\alpha, -\beta)$.

\[
V(x, -\alpha, -\beta) = \frac{e^{-\lambda}}{(1+e^{-\lambda})^2} = \frac{e^{\lambda}}{(1+e^{\lambda})^2} = V(x, \alpha, \beta).
\]

Therefore, $x_1^*, x_2^*$ are the optimal design points for $\alpha, \beta < 0$. In the fourth quadrant $(\alpha > 0, \beta < 0)$, $V(-x, \alpha, -\beta) = V(x, \alpha, \beta)$. Also in the second quadrant $(\alpha < 0, \beta > 0)$ $V(-x, -\alpha, \beta) = V(x, \alpha, \beta)$.

Thus $-x_1^*, -x_2^*$ are the optimal design points for the second and fourth quadrants. This completes the proof.

The work that follows is an updated version of Ford's solution finding the three areas in which $x_1^*, x_2^*$ may lie for $\alpha, \beta > 0$. 
1. Solve the equations below for $x_1, x_2$:

(64) \[ \theta_{x_1} = .176 = \frac{\exp (\alpha + \beta x_1)}{1 + \exp (\alpha + \beta x_1)} \]

(65) \[ \theta_{x_2} = .824 = \frac{\exp (\alpha + \beta x_2)}{1 + \exp (\alpha + \beta x_2)} \]

(66) implying \[ \alpha + \beta x_1 = -1.5434 \]
\[ \alpha + \beta x_2 = 1.5434 \]

(67) or \[ x_1 = \frac{-1.5434 - \alpha}{\beta} \]
\[ x_2 = \frac{1.5434 - \alpha}{\beta} \]

2. If $x_1, x_2 \in [-1, 1]$, $x_1^\# = x_1, x_2^\# = x_2$. If not, set $x_1^\# = -1$, and solve

(68) \[ \exp (\alpha + \beta x_2) = \frac{2 + (x_2 + 1) \beta}{-2 + (x_2 + 1) \beta} \]

3. If $x_2$ from step 2 lies in $[-1, 1]$, then $x_2^\# = x_2$.

Otherwise $x_2^\# = 1$.

Ford illustrates these results in the following diagram.
Area A corresponds to $x_1^*, x_2^* \in [-1, 1]$; area B corresponds to $x_1^* \in [-1, 1]$ and $x_2^* = 1$; and area C corresponds to $x_1^* = -1$ and $x_2^* = 1$.

If one knows $\alpha, \beta$ for the logistic distribution, then the above procedure provides a method for obtaining D-optimal 2 point fixed designs. In most practical situations, perfect knowledge of $\alpha, \beta$ will not be available. If one knew that $\alpha, \beta$ were in area C, then exact knowledge of $\alpha, \beta$ is not necessary since the optimal design will be at -1 and 1.

**Optimal Two-Point Design when $x_1 \neq x_1^*$**

The previous section presented the results of Ford (1976), White (1973, 1975), and Silvey (1980) for obtaining a D-optimal 2 point design with the logistic, assuming known values of $\alpha, \beta$. It is interesting to investigate how well a two-point design might perform if $x_1 \neq x_1^*$ was already selected. This could occur when one was
constrained in his selection of \( x_1 \) or chose \( x_1 \) based on estimates \( \alpha_0, \beta_0 \) that did not equal the true parameter values \( \alpha, \beta \). The theory developed in this section will be utilized as a basis of two of the sequential algorithms presented later.

Since the D-optimal two-point design for the logistic was based on the independent variable confined to the \(-1 \) to \(+1 \) range, we need to convert our problem of interest to this range. Our problem with time as the independent variable can be placed on any set of time units such as days, weeks, or months. We will let the possible levels of time be \( 0, 1, 2 \ldots k-1 \). For example, \( k-1 \) may be \( 12 \) and the time units may be months. Thus the levels of \( x \) that we can select observations at could be at time \( 0 \), end of 1 month, end of two months, \( \ldots \), end of 12 months or 1 year. Thus we will observe a process over the span of a 1-year time period in this instance. We assume that this interval is created so that it is symmetric about the mean of the logistic distribution, \( \frac{-\alpha}{\beta} \). Let

\[
\tilde{x}_i = \frac{x_i + \frac{\alpha}{\beta}}{-\frac{\alpha}{\beta}} \quad \text{for } i = 0, 1, \ldots, k-1.
\]

(69) \[ \tilde{x}_i \]

(70) so that \( \tilde{x}_0 = -1 \)

\( \tilde{x}_k = +1 \).

Thus the transformed variable \( \tilde{x} \in [-1, 1] \). It is obvious that \( \theta_{\tilde{x}_1} = \theta_{\tilde{x}_1} \) and therefore
Substituting $\tilde{x}_i = \frac{-\beta x_i}{\alpha} - 1$, and solving for $\tilde{\alpha}, \tilde{\beta}$ in terms of $\alpha, \beta$, we find the following:

(72)\[ \tilde{\alpha} = 0 \]
(72)\[ \tilde{\beta} = -\alpha . \]

Let us investigate how we might find the optimal $\tilde{x}_2^*$ subject to the constraint that $\tilde{x}_1$ is already selected. Recall that

(73)\[ |M| = p_1 p_2 \theta_1 \theta_2 (1-\theta_1)(1-\theta_2)(\tilde{x}_1-\tilde{x}_2)^2 \]
(73)\[ = k \theta_2 (1-\theta_2) (\tilde{x}_1-\tilde{x}_2)^2 \]

where $k$ is known since $p_1, p_2,$ and $\theta_1$ are known at this point. This may also be written as

(74)\[ |M| = k \frac{\exp(\alpha+\beta x_2)}{(1+\exp(\alpha+\beta x_2))^2} (x_1-x_2)^2 . \]

Converting this to the $-1$ to $+1$ range we have

(75)\[ |M| = \frac{k \exp(\tilde{\beta} \tilde{x}_2)}{(1+\exp(\tilde{\beta} \tilde{x}_2))^2} (\tilde{x}_2-\tilde{x}_1)^2 . \]

Now
\[
\ln |M| = \ln k + \beta x_2 - 2 \ln (1 + \exp(\beta x_2)) + 2 \ln(x_2 - x_1) .
\]

Taking the partial of this with respect to \(x_2\) and setting it equal to zero results in the following:

\[
\frac{\partial \ln |M|}{\partial x_2} = 0 = \beta - \frac{2 \beta \exp(\beta x_2)}{1 + \exp(\beta x_2)} + \frac{2}{x_2 - x_1}
\]

\[
- \beta = - 2 \beta \theta_2 + \frac{2}{x_2 - x_1}
\]

\[
\frac{\beta}{2} = \beta \theta_2 - \frac{1}{x_2 - x_1}
\]

\[
\frac{1}{2} = \theta_2 - \frac{1}{\beta(x_2 - x_1)}
\]

\[
\theta_2 = \frac{1}{2} + \frac{1}{\beta(x_2 - x_1)}
\]

The extreme points of \(|M|\) are the same for \(\ln |M|\). To find these, one may solve the last equation above for \(x_2\). Since

\[
\theta_2 = \exp(\beta x_2) / (1 + \exp(\beta x_2)),
\]

an iterative search scheme is necessary. A typical plot of \(|M|\) as a function of \(\tilde{x}_2\) given \(\tilde{x}_1\) is shown in Figure 3.
From the plot we see there are two local maximums. Before we elaborate on a method of determining which is the global maximum, let us first examine $\frac{\partial^2 \ln |M|}{\partial x_2^2}$.

\[(82) \quad \ln |M| = \ln k + 2 \ln |x_2 - x_1| + \ln \theta_2 + \ln (1-\theta_2)\]

\[(83) \quad \Rightarrow \quad \frac{\partial \ln |M|}{\partial x_2} = \frac{2}{(x_2 - x_1)} + \bar{\beta}(1-2\theta_2)\]

\[(84) \quad \Rightarrow \quad \frac{\partial^2 \ln |M|}{\partial x_2^2} = \frac{-2}{(x_1 - x_2)^2} - 2\bar{\beta}(\beta \theta_2 (1-\theta_2)) < 0\]

Thus any solutions to

\[(85) \quad \theta_2 = \frac{1}{2} + \frac{1}{\beta(x_2 - x_1)}\]

are at least local maximums. Let $\bar{x}_2$ be a solution to the above equation. Since $\bar{x} \in [-1,1]$, we must follow the following rules:

- If $\bar{x}_2 < -1$, set $\bar{x}_2 = -1$.
- If $\bar{x}_2 > 1$, set $\bar{x}_2 = 1$. 

![Figure 3. Plot of typical $|M|$ versus $\bar{x}_2$ given $\bar{x}_1$.](image-url)
It is worth examining $\frac{\partial \ln |M|}{\partial \tilde{x}_2}$ versus $\tilde{x}_2$. A typical graph of this relationship is shown below.

It is seen that $\frac{\partial \ln |M|}{\partial \tilde{x}_2}$ asymptotically levels off as $\tilde{x}_2 \to -\infty$ and $\tilde{x}_2 \to +\infty$. Taking limits we find the following:

\[
\lim_{\tilde{x}_2 \to -\infty} \frac{\partial \ln |M|}{\partial \tilde{x}_2} = \lim_{\tilde{x}_2 \to +\infty} \frac{\partial \ln |M|}{\partial \tilde{x}_2} = 0 + \beta(l-2\theta_2)
\]

In a similar fashion,
\[
\lim_{\tilde{x}_2 \to \infty} \frac{\partial \ln |M|}{\partial \tilde{x}_2} = -\bar{\beta}.
\]

From the analysis above, we have seen that two extreme points exist for this problem. How does one determine which of the two is the global maximum? Since there are only two points to consider one could easily calculate \(|M|\) for each and choose the one giving the larger value for \(|M|\). However, this is not necessary for we can show which will be the global maximum. As seen from the above work, one value will be less than \(\tilde{x}_1\) and the other greater than \(\tilde{x}_1\). If \(\tilde{x}_1 = 0\), then both values will be symmetric about \(\tilde{x}_1\) and either will provide an optimal \(|M|\). We now examine the case where \(\tilde{x}_1 \neq 0\).

**Theorem 2.** For \(\tilde{x} \in [-1, 1]\) and \(\tilde{x}_1 \neq 0\), the global maximum of the determinant of the Fisher Information matrix \(|M|\) for a 2 point design given \(\tilde{x}_1\) is found by selecting the root of \(\theta_2 = \frac{1}{2} + \frac{1}{B(\tilde{x}_2 - \tilde{x}_1)}\) farthest from \(\tilde{x}_1\).

**Proof.** For the two roots of \(\theta_2 = \frac{1}{2} + \frac{1}{B(\tilde{x}_2 - \tilde{x}_1)}\), denote the \(\tilde{x}_2\) value closer to \(\tilde{x}_1\) by \(\tilde{x}_{2c}\), and the one farther by \(\tilde{x}_{2F}\). Recalling that \(|M| = k\theta_2(1-\theta_2)(\tilde{x}_1-\tilde{x}_2)^2\), we immediately see that \((\tilde{x}_1-\tilde{x}_{2c})^2 < (\tilde{x}_1-\tilde{x}_{2F})^2\). Now, suppose \(|\tilde{x}_1-\tilde{x}_{2c}| = \Delta = |\tilde{x}_1-\tilde{x}_F|\) where \(\tilde{x}_F\) is on the opposite side of \(\tilde{x}_1\) from \(\tilde{x}_{2c}\). Since in this case \(\tilde{x}_F\) will be closer to 0 than \(\tilde{x}_{2c}\), \(\theta_\bar{\gamma}(1-\theta_\bar{\gamma}) > \theta_\bar{\gamma}(1-\theta_\bar{\gamma})\). Therefore, \(k\theta_\bar{\gamma}(1-\theta_\bar{\gamma})(\tilde{x}_1-\tilde{x}_F)^2 > k\theta_\bar{\gamma}(1-\theta_\bar{\gamma})(\tilde{x}_1-\tilde{x}_{2c})^2\). Hence \(\tilde{x}_{2c}\) cannot be the global maximum. Rather \(\tilde{x}_{2F}\) must be the global maximum. This completes the proof.
In this section, we have examined the problem of finding an optimal $\tilde{x}_2^*$ given that $\tilde{x}_1$ was already chosen. If $\tilde{x}_1 = \tilde{x}_1^*$, then $\tilde{x}_2^*$ will be the same as found by Ford, White, and Silvey. If $\tilde{x}_1 \neq \tilde{x}_1^*$, then the method presented will result in a maximum $|M|$ given that $\tilde{x}_1$ is already chosen. This theory may be used to provide an upper bound on $|M|$ in later sections as well as providing the basis for two algorithms presented.

**Construction of D-Optimal Designs Subject to Constraints on the Design Space**

Thus far in this chapter, we focused on two point D-optimal designs. In general, these two point designs will be superior to $k$ point ($k > 2$) designs with respect to D-optimality. However, there are many instances in which one may desire to collect data at more than two points in the domain of the independent variable. There may also be economic, legal, psychological, regulatory, or information needs that cannot be satisfied by a two point design. For example, often the information needs required for quality assurance require data at least quarterly for a food product with a shelf life over six months. Constraints may have to be enforced such as at least a selected percentage of the total sample size $N$ must be allocated at various levels of the independent variable $x$. At present, methods for solving such problems from a D-optimal perspective have not been found in the published literature.

Federov (1972), page 102, presents an iterative algorithm for the construction of D-optimal fixed designs for linear models,
and proves it converges. Working from this algorithm we develop two new algorithms for the sequential construction of D-optimal designs for our logistic model with time as the independent variable.

Recall that this "forward only" problem will not allow us to go backward in time. As presented, Federov's algorithm is not suitable for this issue. In the first algorithm, this "forward only" constraint is the only constraint we handle. The second algorithm is a generalization of the first in which constraints on the minimum number of observations to be collected at selected design points are allowed.

Before these two modifications are presented, Federov's algorithm is summarized. Let $\xi_0$ be the initial design such that $M(\xi_0)$ is nonsingular and $x_1, x_2, \ldots, x_k$ are the initial design points where $k$ is at least as large as the number of parameters, $m$, to be estimated. Associated with $x_i$ in $\xi_0$ is $p_i$, $i = 1, 2, \ldots, k$ where $\sum p_i = 1$. On this foundation, the algorithm follows where $j$ is initially 1.

1. Find $x_0$ for which $\lambda(x_0)d(x_0, \xi_{j-1}) = \max \lambda(x)d(x, \xi_{j-1})$

   This step selects the point with the largest var($y|x$).

2. Construct design $\xi_j = (1-\gamma_0)\xi_{j-1} + \gamma_0 x_0$

   where $\gamma_0 = \delta_0/[(\delta_0 + (m-1))m$,

   $\delta_0 = \lambda(x_0)d(x_0, \xi_{j-1})-m$.

   This step creates a new design $\xi_j$ that is a weighted average of the previous design $\xi_{j-1}$ and the point $x_0$ selected in Step 1. The weight $\gamma_0$ results in the maximal increase in $|M(\xi_j)|$. 
3. Compute $M(\xi_j) = \sum_{i} \lambda(x_i)f(x_i)f'(x_i)$ and its inverse $D(\xi_j)$.

This step creates a new Fisher Information matrix $M(\xi_j)$ and its inverse for the new design $\xi_j$.

4. Repeat steps 1-3 until a selected convergence criterion is met.

Examples of convergence criteria are the following:

$$\gamma_0 \leq \delta_1$$

or

$$\frac{|M(\xi_j)| - |M(\xi_{j-1})|}{|M(\xi_j)|} < \delta_2$$

Several items are worth noting here. First the summation in step 3 is over all design points that are presented in design $\xi_j$.

Federov's algorithm conceivably adds a new design point $x_0$ on each iteration, but never completely removes an existing design point.

Also the determination of $x_0$ is a nontrivial matter in general.

Selection of the initial design measure $\xi_0$ can greatly impact the speed of convergence to the D-optimal design.

For our problem of interest, we have only $k$ design points from which to build our design. Thus step 1 of Federov's algorithm can be easily accomplished with just a simple search of the $k$ points. We have shown in Chapter 2 that $\lambda(x)d(x,\xi)$ is the equivalent G-optimality criterion for the 3 extensions to the G.F.T. presented there. Thus we will be increasing $p_i$ corresponding to the
point \( x_1 \) with the maximum variance analogue. However, since we have
time as our independent variable, once the experiment begins we
cannot go backwards in time. This is accounted for in our
modifications of the above algorithm.

The first extension focus on the changes that must be
incorporated in a sequential forward only problem. At time \( t \), no
changes can be made in \( p_i \) for \( i < t \). We cannot increase or decrease
the number of observations collected. Thus there are constraints
that must be enforced. Let \( \epsilon_i \) be the percentage of the total sample
size already allocated at time \( i, i < t \). Then the constraints to be
enforced are \( p_i = \epsilon_i, i < t \). We now modify Federov's algorithm to
handle this situation. For the present, we will assume an initial
design, \( \xi_0 \), exists.

1. Find \( x_1^* \), \( \lambda(x_1^*)d(x_1^*, \xi_{j-1}) = \max_{i \geq t} \lambda(x_i)d(x_i, \xi_{j-1}) \)
2. Calculate \( y_0 \) as in Federov's algorithm
3. Let \( y_0^* = y_0 \left( \sum_{i \geq t} p_i \right) \)
4. Set
   \[ p_i = \epsilon_i, i < t \]
   \[ p_i = p_i(1-y_0), i \geq t \text{ and } i \neq i^* \]
   \[ p_i = y_0^* + (1-y_0)p_i, i = i^* \]
5. Compute \( M(\xi_j) \) and \( D(\xi_j) \).
6. Repeat steps 1-5 until a desired convergence
criterion is hit.
7. Collect \( p_i^* \) observations at \( x_i^* \).
8. Increment \( t \).
9. If \( t \leq k \), go to 1; otherwise, stop.
As seen above, this modified algorithm must be run until its convergence criterion is met for each time interval.

This can be generalized to account for the possibility of another form of constraint. For many reasons (e.g., quality control) an individual may desire a minimum percentage of observations be collected at any of the design points. Thus constraints such as \( p_i \geq m_i \), where \( m_i \) is the desired minimum percentage at \( x_i \), must be handled. The second extension of Federov's algorithm that follows accomplishes this. In this algorithm, \( \varepsilon_i \geq m_i \), \( i < t \) since the percentage of observations already collected at \( x_i \) must be at least the minimums that were prespecified. Assuming an initial design \( \xi_0 \) was given and that the experiment is at time \( t \), the following must be changed in our above modified Federov algorithm to handle both type of constraints.

3. Let \( \gamma_0 = \gamma_0(1 - \sum_{i<t} \varepsilon_i - \sum_{i\geq t} m_i) \)

4. Set

\[ P_i = \varepsilon_i, \quad i < t \]
\[ P_i = m_i + (p_i - m_i)(1 - \gamma_0), \quad i \geq t \text{ and } i \neq i^* \]
\[ P_i = m_i + (p_i - m_i)(1 - \gamma_0) + \gamma_0, \quad i = i^* \]

In this manner, both the equality constraints \( p_i = \varepsilon_i \), \( i < t \), and the inequality constraints \( p_i \geq m_i \), \( i \geq t \) are satisfied. One can show in a straightforward manner that the constraint \( \sum_i P_i = 1 \) is assured, and that the percentage allocated to \( x_i^* \) is increased as a result of the above. The convergence of this algorithm will now be addressed.

**Theorem 3.** At each point in time, \( t \), the second modified Federov algorithm converges.
Proof. Federov (1972, pp. 100-101) shows that for any \( \gamma \), \( |M(\xi_j)| \) is an increasing function of \( \lambda(x_0) d(x_0, \xi_{j-1}) \). While the quantity \( \gamma_0 \) results in the maximal increase, it ignores the constraints imposed on our problem. The use of \( \gamma_0^{*} \) will still result in a monotonically increasing \( |M(\xi_j)| \) for the current parameter estimates \( \hat{\alpha}, \hat{\beta} \). The quantity \( |M(\xi_j)| \) is bounded above by \( |M(\xi^{*}, \alpha, \beta)| \) which is the D-optimal design using the true parameters \( \alpha, \beta \). Since any bounded monotonic nondecreasing sequence converges, the sequence \( |M(\xi_j)| \) will converge. This completes the proof.

While it was seen above that the algorithm converges at each time step, there are two issues that must be addressed. The first concerns the use of \( \hat{\alpha}, \hat{\beta} \) as estimates of \( \alpha \) and \( \beta \). Since we do not know the true parameter values \( \alpha, \beta \) while using this algorithm it is seen that while convergence is guaranteed at each point in time, we do not know what the sequence \( |M(\xi_j)| \) converges to. It is obvious that the closer \( \hat{\alpha}, \hat{\beta} \) are to \( \alpha, \beta \), then the closer the sequence \( |M(\xi_j)| \) will be to \( |M(\xi^{*}, \alpha, \beta)| \). The second issue concerns the final Fisher Information matrix obtained at the end of the entire experiment, i.e., after all the data have been collected and the final estimates of \( \alpha, \beta \) are obtained. The closeness of the determinant of this Fisher Information matrix to \( |M(\xi^{*}, \alpha, \beta)| \) depends on the estimates of \( \alpha \) and \( \beta \) at each time step. While the maximum likelihood estimates \( \hat{\alpha}, \hat{\beta} \) are consistent and we will have a fairly large total sample size, \( N \), the early time period estimates have much fewer cumulative samples. As a result, the early sample allocation
might be expected to vary considerably from the optimal allocation. This will impact the final design.

Several items relative to the above will be briefly discussed. So far we have not discussed stopping rules for the above algorithms. Federov (1972), page 104, provides conditions that can be used. For our purposes we could stop when \( \gamma^*_o N \leq 1 \), where \( N \) is the total sample to be used in the experiment. Under these conditions, at most 1 observation is being allocated to \( x_0 \). Another interesting point concerns the selection of the \( m_i, i \geq t \). One could conceivably change these during the course of the experiment for \( i \geq t \). There may be financial or other reasons to do this. For example, sufficient information may have been gathered and the experimenter may wish to terminate the experiment early. It should be kept in mind that previous allocations were based on the earlier values of \( m_i \), and may no longer be optimal; however, the selection of the \( m_i \) is under the control of the experimenter.

Algorithms to be Compared

In this section various algorithms are presented that will be compared in later chapters. All of these are based on material presented earlier in this chapter. Where possible, analytical comparisons of these methods will be made. When necessary, Monte Carlo simulation is used. The specific conditions under which these comparisons are made are given in Chapter 4.
The Ford-Silvey-White theory results in an optimal design as a function of true $\alpha$, $\beta$ when the independent variable $x$ can take on any value. However, in practice, perfect knowledge is seldom available. Indeed if perfect knowledge were available, then there would be no reason to perform the experiment. Also we consider only discrete levels of $x$. To provide an upper bound we will compute $|M_{\text{opt}}|$ using $\alpha$, $\beta$ assuming $x$ can be sampled anywhere. For comparison purposes with other algorithms, two algorithms below based on the Ford-Silvey-White approach are computed using the initial estimates $\alpha_0$, $\beta_0$ of the true values $\alpha$ and $\beta$.

**MA** The first two point algorithm based on Ford-Silvey-White theory using $\alpha_0$, $\beta_0$, allows any value in the range of interest of $x$ to be selected.

**MAR** There exists an imposed grid on the independent variable time. This grid allows observation to be taken only at the $k$ discrete grid points. Therefore, this second algorithm based on Ford-Silvey-White theory using $\alpha_0$, $\beta_0$ rounds the selected two points from algorithm MA to the nearest two grid points. This second version is, in the author's opinion, more realistic in terms of our stated problem of interest.

**ALG3** Equal $n_i$ at all $k$ locations. This is the standard procedure in which an equal number of observations are collected at each design point. This is included to provide a comparison between this common approach and the other algorithms considered here.
**ALG4** Equal $n_i$ until at least 50 percent of the observations are allocated, then collect the remainder at $x_2^*$. The point $x_2^*$ is estimated using our optimal $x_2|x_1$ theory where $x_1$ will be equal to the average $x$ value of the observations already allocated.

**ALG5** Small $n_i = .5 \frac{(N/k)}{}$ until $x_1^*$, allocate (.5N-$\Sigma n_i$) at $x_1^*$, estimate $\alpha, \beta$ and place remainder of observations at $x_2^*$. The point $x_1^*$ is obtained using the Ford-Silvey White method based on $\alpha_0$, $\beta_0$. The parameter estimates $\hat{\alpha}, \hat{\beta}$ are estimated after $x_1^*$ is hit, and then $x_2^*$ is selected using our optimal $x_2|x_1$ theory where $x_1$ will be given the average $x$ value of the first 50 percent of the observations used.

**ALG6** First modified Federov scheme where we force the forward only nature of our problem into the algorithm. The initial design $\xi_0$ sets $p_i = N/k$ for $i = 1, 2, \ldots, k$.

**ALG7** Second modified Federov scheme where constraints such as a minimum percentage of observations may have to be collected at selected design points. For this study $\varepsilon_i = .5 \frac{(N/k)}{}$. The initial design $\xi_0$ sets $p_i = N/k$ for $i = 1, 2, \ldots, k$.

A comparison between the new sequential algorithms (ALG4, ALG5, ALG6, and ALG7) and the fixed design algorithms (MA, MAR, and ALG3) will indicate the conditions under which these new approaches are better or worse. Chapter 4 will detail the experimental conditions and the criterion that will be used for these comparisons, and give the results of an initial pilot study.
Below is a pictorial summary of the algorithms that may aid the reader in the chapters ahead. These pictures are samples of what may occur.

---

**FIGURE 5. PICTORIAL SUMMARY OF ALGORITHMS**
CHAPTER 4 PILOT STUDY

Introduction

The previous chapters have provided the theoretical basis for what follows in the remainder of this report. This included the General Equivalence Theorem and its extensions, and important properties of the logistic distribution. The extensions of the Kiefer-Wolfowitz G.E.T. by Federov, White, and Ford were shown to be equivalent for the logistic model. Chapter 3 briefly introduced the algorithms that will be compared in the remainder of this study. Several sequential algorithms will be compared to fixed designs on the basis of D-optimality.

Two phases of comparison of the algorithms were performed. The first phase was a pilot study that covered a wide range of experimental values and variables. The experimental design and a brief summary of the corresponding results will be discussed in this chapter. The secondary phase involved a more in-depth analysis of a restricted number of variables and variable ranges. The second phase also examined the problem from a different perspective described later. The experimental design for this latter phase will be detailed in the next chapter, along with the results of the analysis.
The objectives of this work are to allow algorithm comparisons on the basis of D-optimality and to determine the key independent variables that influence algorithm performance. The pilot study will primarily serve as a screening mechanism on these independent variables. It will result in a reduced set of variables that will be examined in detail in Chapter 5. The follow-on analysis of Chapter 5 will be used for detailed comparison of the best overall sequential algorithm with a fixed algorithm based on the Ford-Silvey-White theory. The knowledge gained will result in suggested uses of a sequential algorithm for our binary response problem when D-optimality is the measure of interest.

Pilot Study Design and Results

Before a detailed description is provided concerning the variables involved with the pilot study experimental design, it is important to explain the point of view underlying this phase of the study. In the opinion of the author this point of view may be very typical of a "real" problem setting. In this case, the individual wishing to study a binary response as a function of time has initial estimates for the parameters of the binary response curve. For the logistic distribution, this means he has initial estimates $\alpha_0$ and $\beta_0$ of the true parameters $\alpha$ and $\beta$. The experiment will be designed using these initial estimates. Of course, the sequential designs will update the initial estimates with the collected experimental data, and modify their designs accordingly. The pilot study point
of view examines the impact of different states of nature (i.e., $\alpha$ and $\beta$) given these initial estimates of $\alpha_o$ and $\beta_o$.

The variables upon which the pilot study is based are as follows:

- $(\alpha, \beta)$ true parameter values for the logistic distribution
- $(\alpha_o, \beta_o)$ initial estimates of $(\alpha, \beta)$
- $AR = \alpha/\alpha_o$
- $BR = \beta/\beta_o$
- $N$ total sample size for entire experiment
- $K$ number of grid points or time intervals.

The time span covered by the experiment was arbitrarily chosen to be the interval $[0, 12]$. The results are not dependent on this interval since the interval can be scaled to any desired time period. The $[0,12]$ interval was chosen because of the association with the canned milk product example of Chapter 2. Furthermore, the choice of $AR = \alpha/\alpha_o$ and $BR = \beta/\beta_o$ allows one to determine the extent of the deviations of the state of nature from the initial estimates on a ratio basis. This selection was not without its problems as will be seen later in this chapter.

The pilot study experimental design examined the following levels for the variables mentioned above.

- $(\alpha_o, \beta_o) = (-6, 1), (-2.25, 0.375)$
- $AR = \alpha/\alpha_o = 0.5, 1, 1.5$
- $BR = \beta/\beta_o = 0.5, 1, 1.5$
The design was a full factorial involving 720 factor combinations. Ten iterations of each of the 720 combinations were run for the sequential algorithms using specially written FORTRAN-77 software on large Control Data Corporation machines. For each algorithm, the determinant of the Fisher Information Matrix, |M|, was calculated. The 2 point D-optimal design Fisher Information Matrix determinant using α, β for the appropriate combination of variables above was used to normalize the results to an [0, 1] scale to allow for easier comparison and interpretation. Thus, the performance measure used in the analysis is the ratio |M_A|/|M_{opt}| where |M_A| is the actual determinant of the Fisher Information Matrix for a given iteration of a particular algorithm, and |M_{opt}| is the optimal 2 point design using the true values of α and β. This ratio is then bounded by 0 and 1. |M_A| is calculated using the true values of α and β for a given realization of a design. Thus, for the nonsequential algorithms for which the sample allocations are predetermined and not affected by the results of the collected data, |M_A| will be invariant. Therefore, it is not necessary to perform multiple iterations of these fixed design algorithms which must be done, however, for the sequential algorithms. These sequential algorithms modify the experimental design based on collected data. Also, |M_{opt}| was not restricted to the [0, 12] range in the optimal selection of the two design points.
Due to the large sample size, statistical significance was found for most main effects and interactions even when the absolute difference between the simulation means was too small to have any practical significance. For example, differences in some mean combinations were found to be statistically significant when the means differed by less than 0.01. In the author's judgment such small differences are not practically significant. Therefore, statistical significance was not the key tool used in the pilot study. Numerous analyses were performed, but only an overall summary will be given in this chapter. The pilot study had a primary role of screening variables so that a more in-depth analysis on fewer variables could be analyzed in the secondary phase in Chapter 5. Overall, the effect of the variables K and N was minimal. Increasing the total sample size N from 260 to 1040 did slightly increase the means portrayed in the following tables, but did not alter the algorithm ordering. This increase in sample size raised the mean values for the sequential algorithms an average of approximately 0.03. The levels of K, the number of time grid points at which observations could be sampled, resulted in a mean difference of less than 0.01. Thus, these two variables were eliminated and not considered in the secondary phase of the experiment. Tables 2 and 3 show mean results that have been averaged over both K and N.

The initial estimates \((\alpha_0, \beta_0)\) chosen result in an estimate of \(6\) for the logistic mean \(\mu_0\). Since the logistic distribution is symmetric, the mean and median (50th percentile of the distribution) are the same. The two sets of values were chosen
TABLE 2. ALGORITHM MEANS AND ANALYSIS OF VARIANCE
RESULTS FOR $\alpha_0 = -6$, $\beta_0 = 1$

<table>
<thead>
<tr>
<th>$AR = \alpha/\alpha_0$</th>
<th>0.5</th>
<th></th>
<th>1.0</th>
<th></th>
<th>1.5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>.84</td>
<td>7</td>
<td>A</td>
<td>.31</td>
<td>6</td>
<td>all \approx 0</td>
</tr>
<tr>
<td>A</td>
<td>.81</td>
<td>4</td>
<td>A</td>
<td>.29</td>
<td>7</td>
<td>all \approx 0</td>
</tr>
<tr>
<td>AB</td>
<td>.78</td>
<td>5</td>
<td>B</td>
<td>.18</td>
<td>4</td>
<td>all \approx 0</td>
</tr>
<tr>
<td>B</td>
<td>.72</td>
<td>6</td>
<td>C</td>
<td>.02</td>
<td>5</td>
<td>all \approx 0</td>
</tr>
<tr>
<td>MA</td>
<td>.56</td>
<td></td>
<td>MA</td>
<td>.02</td>
<td></td>
<td>all \approx 0</td>
</tr>
<tr>
<td>MAR</td>
<td>.70</td>
<td></td>
<td>MAR</td>
<td>.03</td>
<td></td>
<td>all \approx 0</td>
</tr>
<tr>
<td>ALG3</td>
<td>.82</td>
<td></td>
<td>ALG3</td>
<td>.19</td>
<td></td>
<td>all \approx 0</td>
</tr>
<tr>
<td>\mu</td>
<td>6</td>
<td></td>
<td>\mu</td>
<td>12</td>
<td></td>
<td>all \approx 0</td>
</tr>
<tr>
<td>BR = \beta/\beta_0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>.58</td>
<td>4</td>
<td>A</td>
<td>.71</td>
<td>6</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>.53</td>
<td>7</td>
<td>A</td>
<td>.68</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>C</td>
<td>.43</td>
<td>5</td>
<td>AB</td>
<td>.60</td>
<td>7</td>
<td>B</td>
</tr>
<tr>
<td>D</td>
<td>.09</td>
<td>6</td>
<td>B</td>
<td>.54</td>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>MA</td>
<td>.08</td>
<td></td>
<td>MA</td>
<td>1.0</td>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MAR</td>
<td>.09</td>
<td></td>
<td>MAR</td>
<td>.93</td>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>ALG3</td>
<td>.31</td>
<td></td>
<td>ALG3</td>
<td>.38</td>
<td></td>
<td>ALG3</td>
</tr>
<tr>
<td>\mu</td>
<td>3</td>
<td></td>
<td>\mu</td>
<td>6</td>
<td></td>
<td>\mu</td>
</tr>
<tr>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>.32</td>
<td>7</td>
<td>A</td>
<td>.45</td>
<td>5</td>
<td>A</td>
</tr>
<tr>
<td>B</td>
<td>.21</td>
<td>5</td>
<td>A</td>
<td>.43</td>
<td>4</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>.19</td>
<td>4</td>
<td>B</td>
<td>.39</td>
<td>7</td>
<td>C</td>
</tr>
<tr>
<td>D</td>
<td>.00</td>
<td>6</td>
<td>C</td>
<td>.30</td>
<td>6</td>
<td>D</td>
</tr>
<tr>
<td>MA</td>
<td>.00</td>
<td></td>
<td>MA</td>
<td>.12</td>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MAR</td>
<td>.00</td>
<td></td>
<td>MAR</td>
<td>.11</td>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>ALG3</td>
<td>.14</td>
<td></td>
<td>ALG3</td>
<td>.18</td>
<td></td>
<td>ALG3</td>
</tr>
<tr>
<td>\mu</td>
<td>2</td>
<td></td>
<td>\mu</td>
<td>4</td>
<td></td>
<td>\mu</td>
</tr>
</tbody>
</table>

\( \mu \) above is the logistic mean (\( \mu = -\alpha/\beta \)); Scheffé is the Scheffé multiple comparison test. (Algorithms not having common letters are significantly different at the .05 significance level.)
**TABLE 3. ALGORITHM MEANS AND ANALYSIS OF VARIANCE RESULTS FOR \( \alpha_0 = -2.25, \beta_0 = 0.375 \)**

\[
AR = \frac{\alpha}{\alpha_0}
\]

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th></th>
<th>1.0</th>
<th></th>
<th>1.5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
</tr>
<tr>
<td>A</td>
<td>.58</td>
<td>7</td>
<td>A</td>
<td>.45</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>.58</td>
<td>4</td>
<td>A</td>
<td>.40</td>
<td>7</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>.54</td>
<td>6</td>
<td>B</td>
<td>.32</td>
<td>6</td>
<td>B</td>
</tr>
<tr>
<td>B</td>
<td>.34</td>
<td>5</td>
<td>C</td>
<td>.23</td>
<td>5</td>
<td>B</td>
</tr>
<tr>
<td>MA</td>
<td>.56</td>
<td></td>
<td>MA</td>
<td>.33</td>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MAR</td>
<td>.54</td>
<td></td>
<td>MAR</td>
<td>.32</td>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>ALG3</td>
<td>.43</td>
<td></td>
<td>ALG3</td>
<td>.27</td>
<td></td>
<td>ALG3</td>
</tr>
<tr>
<td>( \mu )</td>
<td>6</td>
<td></td>
<td>( \mu )</td>
<td>12</td>
<td></td>
<td>( \mu )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th></th>
<th>1.0</th>
<th></th>
<th>1.5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
</tr>
<tr>
<td>A</td>
<td>.70</td>
<td>7</td>
<td>A</td>
<td>1.00</td>
<td>6</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>.69</td>
<td>6</td>
<td>AB</td>
<td>.90</td>
<td>7</td>
<td>AB</td>
</tr>
<tr>
<td>B</td>
<td>.51</td>
<td>5</td>
<td>B</td>
<td>.87</td>
<td>4</td>
<td>AB</td>
</tr>
<tr>
<td>B</td>
<td>.51</td>
<td>4</td>
<td>C</td>
<td>.58</td>
<td>5</td>
<td>C</td>
</tr>
<tr>
<td>MA</td>
<td>.69</td>
<td></td>
<td>MA</td>
<td>1.00</td>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MAR</td>
<td>.69</td>
<td></td>
<td>MAR</td>
<td>1.00</td>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>ALG3</td>
<td>.62</td>
<td></td>
<td>ALG3</td>
<td>.82</td>
<td></td>
<td>ALG3</td>
</tr>
<tr>
<td>( \mu )</td>
<td>3</td>
<td></td>
<td>( \mu )</td>
<td>6</td>
<td></td>
<td>( \mu )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.5</th>
<th></th>
<th>1.0</th>
<th></th>
<th>1.5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
<td>Scheffé Mean</td>
<td>Alg</td>
</tr>
<tr>
<td>A</td>
<td>.56</td>
<td>4</td>
<td>A</td>
<td>.80</td>
<td>4</td>
<td>A</td>
</tr>
<tr>
<td>A</td>
<td>.55</td>
<td>7</td>
<td>A</td>
<td>.80</td>
<td>7</td>
<td>AB</td>
</tr>
<tr>
<td>B</td>
<td>.37</td>
<td>5</td>
<td>B</td>
<td>.61</td>
<td>5</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>.27</td>
<td>6</td>
<td>B</td>
<td>.59</td>
<td>6</td>
<td>C</td>
</tr>
<tr>
<td>MA</td>
<td>.27</td>
<td></td>
<td>MA</td>
<td>.58</td>
<td></td>
<td>MA</td>
</tr>
<tr>
<td>MAR</td>
<td>.27</td>
<td></td>
<td>MAR</td>
<td>.59</td>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>ALG3</td>
<td>.41</td>
<td></td>
<td>ALG3</td>
<td>.67</td>
<td></td>
<td>ALG3</td>
</tr>
<tr>
<td>( \mu )</td>
<td>2</td>
<td></td>
<td>( \mu )</td>
<td>4</td>
<td></td>
<td>( \mu )</td>
</tr>
</tbody>
</table>

\( \mu \) above is the logistic mean (\( \mu = -\alpha/\beta \)); Scheffé is the Scheffé multiple comparison test. (Algorithms not having common letters are significantly different at the .05 significance level.)
to cover a wide range of situations. The \((\alpha_0 = -6, \beta_0 = 1)\) combination corresponds to a binary response having a probability of being defective of 0.002 at time 0 and by symmetry 0.998 at time 12. The second choice \((\alpha_0 = -2.25, \beta_0 = 0.375)\) corresponds to a probability of being defective of 0.095 at time 0 to 0.905 at time 12. Defective as used above implies the presence of a particular binary attribute.

Tables 2 and 3 provide the mean results for the sequential algorithms. These means are the results of 40 iterations that have been averaged over the 2 levels of both K and N. These numbers cannot be compared across all cells. This is a result of the ratios that are used as the dependent variable in the study. The denominators, \(|M_{opt}|\), used are not the same in the cells. The value of \(|M_{opt}|\) is dependent on the state of nature, i.e., the true values of \(\alpha\) and \(\beta\). Figure 5 below is a plot of \(|M_{opt}|\) versus \(|M_\mu|\) for algorithm 3 when both \(\mu_0\) and \(\mu\) equal 6. As seen in Figure 6, \(|M_{opt}|\) varies considerably even though the true mean \(\mu\) is the same for all three cases. This is the major problem of comparison with the point of view taken in this pilot study; however, this is a common approach in which various states of nature are contrasted with an individual's initial estimates.
Within each cell, one way analysis of variance tests were performed to compare the sequential algorithms. The Scheffe multiple comparison test was used (at the 0.05 significance level) to allow pairwise comparison of the means. This multiple comparison procedure protects the experiment-wise type I errors at the sacrifice of the power of some of the other popular multiple comparison tests. These tests were performed using SAS™ (Statistical Analysis System) on a Digital VAX 11/780. Algorithms sharing any common letter are not significantly different. In addition to providing the means for the four sequential algorithms, the corresponding values for the following fixed algorithms are given where \((\alpha_0, \beta_0)\) is used in their design.
MA - 2 point Ford-Silvey-White algorithm where any points in [0, 12] may be selected
MAR - the 2 points selected in MA are rounded to the nearest grid interval
ALG3 - Equal $n_i = N/K$ allocation at each of the $K$ time-grid points (Algorithm 3).

Also, the true mean $\mu = -a/b$ is given for each cell. The numbers 4, 5, 6, and 7 represent algorithms 4, 5, 6, and 7, respectively.

As can be seen from Tables 2 and 3, no algorithm dominates completely; however, algorithms 4 and 7 appear to perform better over the entire ranges of $AR = \alpha/\alpha_0$ and $BR = \beta/\beta_0$. At this point of our study, we are content to eliminate the variables $K$ and $N$ from further consideration and not eliminate any algorithms. Looking at the means $\mu$ given in each of the cells of these two tables, it becomes apparent that the choice of the parameters $AR = \alpha/\alpha_0$ and $BR = \beta/\beta_0$ with their range of 0.5 to 1.5 centers the logistic distribution on values ranging asymmetrically around $\mu_0 = 6$. In fact, the range of $\mu$ values is from 2 to 18. In 2 of the 9 cells in each table, $\mu$ is either at the boundary ($\mu = 12$) of the experimental region of interest [0, 12] or outside ($\mu = 18$) this region. In the second phase of analysis, steps will be taken to avoid extreme cases such as this.
CHAPTER 5 DETERMINATION OF THE BEST SEQUENTIAL ALGORITHM
AND RECOMMENDATION FOR ITS USE

Introduction

Our primary objective in this research is to study the "usefulness" of adaptive (sequential) design algorithms in estimating a typical binary response curve. The measure that quantifies "usefulness" is an extension of D-optimality to nonlinear models such as represented by the binary logistic model. This extension of D-optimality is a nonlinear analog of the linear model generalized variance. More specifically, it is the asymptotic generalized variance for the maximum likelihood estimates \( \hat{\alpha}, \hat{\beta} \) based on the Fisher Information matrix. Chapter 2 covered the concepts of optimal designs for linear models and the extensions to nonlinear situations. Properties of the logistic model were developed, and it was shown that the extensions of the Kiefer-Wolfowitz General Equivalence Theorem were identical for this logistic model even though the extensions in general were different. Chapter 3 built on the results of the second chapter and derived methods that could serve as the foundation of sequential design algorithms. This resulted in four sequential algorithms (ALG4, ALG5, ALG6, and ALG7). The thrust of Chapter 4 and this chapter is to evaluate and compare these.
sequential procedures to fixed design approaches. The questions that must be addressed include when should adaptive algorithms be used, how robust are fixed and sequential design algorithms, and the sensitivity of sequential algorithms to various independent variables. Chapter 4 examined the independent variables that impact algorithm performance in a pilot study. This chapter continues the experimental comparison of algorithms.

The pilot study in Chapter 4 played an important role. It examined the experimental variables over wide ranges and provided insight into potential difficulties of comparing the algorithms for the parameterization used. The pilot study screened from further consideration the number of time-grid points at which data might be gathered (K), and the total sample size for the experiment (N). This chapter presents a secondary phase of the experiment examining algorithm performance. This analysis selects the best overall sequential algorithm and compares it in detail to the 2 point D-optimal fixed design based on Ford, Silvey, and White theory rounded to the nearest grid point (MAR). In addition to examining mean performance, the robustness of both algorithms is compared. Least squares regression analysis of the performance of the sequential algorithm is graphically presented in countour plots to show its predicted behavior over the entire range of the key variables. A reparameterization of the experimental design of Chapter 4 was necessary to accomplish the above and is presented early in this chapter. The major factors of this new experimental design are identified and graphically presented. Finally, this
chapter ends with recommendations for the use of the chosen
sequential algorithm.

Reparameterization

In this second phase, an in-depth study of a smaller set
of experimental variables is performed for the algorithm comparison. This more detailed phase looks at a narrower range of values for the variables, but examines numerous levels within these ranges. In addition, the problem is structured so that wider comparison can be made. To allow this, we view the problem from a different point of view. In the pilot phase, the individual had initial parameter estimates and addressed the issue of how good are the algorithms for different states of nature given these initial estimates. For the second phase, we examine how well the algorithms perform for different initial estimates given a state of nature. While similar, this second approach allows a much wider comparison of results. We will now be able to produce tables in which all numbers may be compared due to having the same denominator in the ratio performance measure $|M_A|/|M_{\text{opt}}|$. Recall that $|M_{\text{opt}}|$ is the determinant of the Fisher Information matrix for the 2 point D-optimal design using the true values of $\alpha$ and $\beta$. The determinant $|M_A|$ is the determinant of the Fisher Information matrix for a given realization of one of the algorithms being compared.

Because of the problems with the parameters $AR = \alpha/\alpha_0$ and $BR = \beta/\beta_0$ used in the pilot study, a new parameterization was sought for the second phase. The idea for the reparameterization chosen is
based on Abdelbasit and Plackett (1983). In their work, they study the efficiency of the 2 point D-optimal Ford-Silvey-White algorithm when initial estimates are used. Their parameterization relating the initial estimates to the true parameter values is the following:

1. \( B(p_0 - \mu) \)
2. \( B_0/B \).

The \( B_0/B \) is the inverse of \( BR \) used in the pilot study. Abdelbasit and Plackett provide a compact formulation relating the efficiency of the 2 point design based on initial estimates to the true 2 point D-optimal design using the parameterization \( B(p_0 - \mu) \) and \( B_0/B \). For our sequential algorithms, no concise relationship such as they present could be found for measuring the efficiency (the ratio \( |M_A|/|M_{opt}| \)) of these algorithms to the 2 point D-optimal solution when the true parameters are unknown. It was decided to use \( \mu_0 - \mu \) instead of \( B(p_0 - \mu) \) because it is easier to understand. In the following UDIFF is defined to be \( \mu_0 - \mu \).

The following selection of levels for variables was used for phase two. For these choices, the number of time-grid points (K) was set at 13 and the total sample size (N) was 260.

- UDIFF: -3, -2, -1, 0, 1, 2, 3
- \( B_0/B \): .6, .8, 1, 1.2, 1.4
- \( \mu \): 4, 6, 8
- \( B \): 0.5, 1.0

For the 210 combinations above, 5 replications were performed for the sequential algorithms.
**Determination of Key Factors**

For each pair of $\mu$ and $\beta$ values, the same denominator $|M_{opt}|$ was used to create the ratios $|M_{A}|/|M_{opt}|$. Therefore, for each such combination all the corresponding simulation results may be compared. Three way analysis of variance tests were performed for all 6 of these combinations. The 3 factors for these 6 ANOVA's were UDIFF, $\beta_{o}/\beta$, and ALG (for the 4 sequential algorithms). Below, we will summarize the major findings of these tests. Additional analysis approaches and results will be given later.

All main effects and interactions in the analysis of variance models that follow were found to be significant at the 0.0001 significance level used in the SAS ANOVA procedure. The large sample size (700) was a major factor in this. Once again, we must resort to additional methods to identify important factors. Figures 7 through 12 show the partitioning of the total sum of squares (TSS) for each of the six cases. This provides an important aid in the determination of key factors. The effects accounting for at least 10 percent of the TSS and the error sum of squares proportions are labelled in these boxes. The unmarked cell of each box represents the sum of all effects that each account for less than 10 percent of the TSS.
FIGURE 7. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 4, B = 0.5, \text{TSS} = 22.8 \)

FIGURE 8. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 6, B = 0.5, \text{TSS} = 17.2 \)

FIGURE 9. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 8, B = 0.5, \text{TSS} = 16.3 \)
FIGURE 10. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 4, \beta = 1, \text{TSS} = 31.6 \)

<table>
<thead>
<tr>
<th>Fraction TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDIFF</td>
</tr>
<tr>
<td>ALG*UDIFF</td>
</tr>
<tr>
<td>ALG</td>
</tr>
<tr>
<td>3 Way</td>
</tr>
<tr>
<td>((B_\theta/\beta)*UDIFF)</td>
</tr>
<tr>
<td>((B_\theta/\beta)*ALG)</td>
</tr>
<tr>
<td>(B_\theta/\beta)</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>

FIGURE 11. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 6, \beta = 1, \text{TSS} = 25.1 \)

<table>
<thead>
<tr>
<th>Fraction TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDIFF</td>
</tr>
<tr>
<td>ALG*UDIFF</td>
</tr>
<tr>
<td>ALG</td>
</tr>
<tr>
<td>3 Way</td>
</tr>
<tr>
<td>((B_\theta/\beta)*UDIFF)</td>
</tr>
<tr>
<td>((B_\theta/\beta)*ALG)</td>
</tr>
<tr>
<td>(B_\theta/\beta)</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>

FIGURE 12. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR \( \mu = 8, \beta = 1, \text{TSS} = 30.5 \)

<table>
<thead>
<tr>
<th>Fraction TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG</td>
</tr>
<tr>
<td>UDIFF</td>
</tr>
<tr>
<td>ALG*UDIFF</td>
</tr>
<tr>
<td>3 Way</td>
</tr>
<tr>
<td>((B_\theta/\beta)*UDIFF)</td>
</tr>
<tr>
<td>((B_\theta/\beta)*ALG)</td>
</tr>
<tr>
<td>(B_\theta/\beta)</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>
From Figures 7 through 12, we see that the main effects due to UDIFF ($\mu_0 - \mu$) and ALG (Algorithm) and their interaction ALG*UDIFF consistently total a large fraction of the TSS. The factor $\beta_0/\beta$ and its interactions account for a smaller portion of the TSS. Therefore, it seems worthwhile to average over the $\beta_0/\beta$ factor when presenting summary means. This is done in some of the analysis that follows. Later regression analysis lends further credence to this averaging over $\beta_0/\beta$.

Selection of Best Sequential Algorithm

As seen above, much broader comparisons are possible in this second phase than in the pilot study. Indeed, an even broader comparison will now be made. This broader comparison will also provide a good opportunity to compare the four sequential algorithms and select the one that is best overall. For the true logistic mean values $\mu$ chosen (4, 6, and 8), the 2 point D-optimal designs selected design points within the (0, 12) boundary. Recall from earlier chapters that the following:

$$|M_{opt}| = \theta_1^* (1 - \theta_1^*) \theta_2^* (1 - \theta_2^*) (x_1^* - x_2^*)^2$$

where

$$\theta_1^* = .176$$
$$\theta_2^* = .824$$

and $x_1^*$, $x_2^*$ are the design points at the optimal $\theta_1^*$, $\theta_2^*$ values, respectively.
For the same $\beta$ values, $(x_1^* - x_2^*)^2$ will be a constant. Therefore the same $|M_{opt}|$ denominator is used in phase two for the performance ratio when $\beta$ is the same. Thus all the data having $\beta = 0.5$ (and similarly for $\beta = 1$) can be compared in an analysis of variance test. The additional term $\mu$ can be added as one of the effects.

Below is a summary of these two analysis of variance tests. Figures 13 and 14 provide a breakdown of the TSS for $\beta = 0.5$ and $\beta = 1$, respectively. Tables 4 and 5 give the results of a Scheffe multiple comparison test on the algorithm main effects for $\beta = 0.5$, and $\beta = 1$, respectively.

\[
\begin{array}{|l|c|}
\hline
\text{Fraction} & \text{TSS} \\
\hline
\text{ALG} & .10 \\
\text{UDIFF} & .15 \\
\text{ALG} & .13 \\
\mu & .08 \\
(\beta_0/\beta) & .06 \\
(\beta_0/\beta) & .04 \\
\mu & .04 \\
\mu & .04 \\
\mu & .03 \\
\mu & .02 \\
\mu & .02 \\
\mu & .01 \\
\mu & .00 \\
\mu & .00 \\
\mu & .24 \\
\text{Error} & 1.00 \\
\hline
\end{array}
\]

**FIGURE 13. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR $\beta = 0.5$, TSS = 56.6**
Table 4. Scheffe Multiple Comparison at 0.05 Significance Level for $\beta = 0.5$

<table>
<thead>
<tr>
<th>Letter</th>
<th>Fraction</th>
<th>ALG6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.78</td>
<td>ALG7</td>
</tr>
<tr>
<td>B</td>
<td>.75</td>
<td>ALG4</td>
</tr>
<tr>
<td>C</td>
<td>.69</td>
<td>ALG5</td>
</tr>
<tr>
<td>D</td>
<td>.65</td>
<td>ALG6</td>
</tr>
</tbody>
</table>

|        | .68      | MA   |
|        | .68      | MAR  |
|        | .74      | ALG3 |

<table>
<thead>
<tr>
<th>Fraction TSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDIFF</td>
</tr>
<tr>
<td>ALG*UDIFF</td>
</tr>
<tr>
<td>$\mu$*ALG</td>
</tr>
<tr>
<td>$\mu$</td>
</tr>
<tr>
<td>($B_0/\beta$)<em>ALG</em>UDIFF</td>
</tr>
<tr>
<td>($B_0/\beta$)*UDIFF</td>
</tr>
<tr>
<td>4 Way</td>
</tr>
<tr>
<td>ALG</td>
</tr>
<tr>
<td>($B_0/\beta$)*ALG</td>
</tr>
<tr>
<td>$B_0/\beta$</td>
</tr>
<tr>
<td>$\mu$<em>ALG</em>UDIFF</td>
</tr>
<tr>
<td>$\mu$*($B_0/\beta$)*UDIFF</td>
</tr>
<tr>
<td>$\mu$*($B_0/\beta$)*ALG</td>
</tr>
<tr>
<td>$\mu$*UDIFF</td>
</tr>
<tr>
<td>$\mu$*($B_0/\beta$)</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

FIGURE 14. TOTAL ANOVA SUM OF SQUARES (TSS) INDICATING PROPORTION VARIANCE EXPLAINED BY MAIN EFFECTS AND INTERACTIONS FOR $\beta = 1$, TSS = 96.2

Table 5. Scheffe Multiple Comparison at 0.05 Significance Level for $\beta = 1$

<table>
<thead>
<tr>
<th>Letter</th>
<th>Fraction</th>
<th>ALG6</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>.48</td>
<td>ALG7</td>
</tr>
<tr>
<td>B</td>
<td>.45</td>
<td>ALG4</td>
</tr>
<tr>
<td>C</td>
<td>.42</td>
<td>ALG6</td>
</tr>
<tr>
<td>D</td>
<td>.40</td>
<td>ALG5</td>
</tr>
</tbody>
</table>

|        | .41      | MA   |
|        | .37      | MAR  |
|        | .35      | ALG3 |
For both $\beta = 0.5$ and $\beta = 1$ the main effect $UDIFF = \mu_0 - \mu$ and the interaction $ALG*UDIFF$ are the 2 largest contributors to the total sum of squares (TSS) that is explained by the model. When $\beta = 0.5$ the ALG main effect was a relatively large portion of the TSS; however, it was a very small part of the TSS when $\beta = 1$. It is not surprising that $UDIFF$ and $ALG*UDIFF$ account for 47 percent of the TSS when $\beta = 1$ as compared with 28 percent of the TSS when $\beta = 0.5$ when the table below is examined for $\mu = 6$. The increased steepness of the $\beta = 1$ curve leaves little information at many of the potential design points.

Table 6. $\theta$ Values, $\mu = 6$

<table>
<thead>
<tr>
<th>Time</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.05</td>
<td>.00</td>
</tr>
<tr>
<td>1</td>
<td>.08</td>
<td>.01</td>
</tr>
<tr>
<td>2</td>
<td>.12</td>
<td>.02</td>
</tr>
<tr>
<td>3</td>
<td>.18</td>
<td>.05</td>
</tr>
<tr>
<td>4</td>
<td>.27</td>
<td>.12</td>
</tr>
<tr>
<td>5</td>
<td>.38</td>
<td>.27</td>
</tr>
<tr>
<td>6</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>7</td>
<td>.62</td>
<td>.73</td>
</tr>
<tr>
<td>8</td>
<td>.73</td>
<td>.88</td>
</tr>
<tr>
<td>9</td>
<td>.82</td>
<td>.95</td>
</tr>
<tr>
<td>10</td>
<td>.88</td>
<td>.98</td>
</tr>
<tr>
<td>11</td>
<td>.92</td>
<td>.99</td>
</tr>
<tr>
<td>12</td>
<td>.95</td>
<td>1.00</td>
</tr>
</tbody>
</table>

From the $\beta = 1$ portion of Table 6, one can see how quickly $\theta$ changes around the median (Time = 6). Thus differences in $UDIFF = \mu_0 - \mu$ have a much larger effect on the ratio $|M_A|/|M_{opt}|$. 
The Scheffé multiple comparison test for both the $\beta = 0.5$ and $\beta = 1$ cases shows that overall algorithm 7 performed significantly better than the other sequential algorithms. Therefore we will focus primarily on this algorithm in the remainder of this study. We will compare it with MAR the 2 point D-optimal design approach using $\alpha_0$, $\beta_0$ that has been rounded to the nearest of the 13 grid points in time.

**Comparison of ALG7 and MAR**

This section provides a detailed comparison of the best sequential algorithm ALG7 with the fixed algorithm MAR based on the D-optimal work of Ford, Silvey, and White. The fixed algorithm MAR selects its 2-point design using the initial estimates $\alpha_0, \beta_0$ and then rounds the chosen points to the nearest grid points. The sequential algorithm ALG7, detailed in Chapter 3, updates the estimates of $\alpha, \beta$ as new data is collected. Recall that ALG7 incorporates the constraint that each time-grid point will have at least $0.5(N/K)$ samples allocated to it. Therefore, only 50 percent of the total sample size $N$ may be freely allocated during the course of the experiment. Thus ALG7 can never achieve 100 percent D-optimality. In addition the forward only nature of the sequential algorithms further limits the allocation of samples. The comparison of ALG7 with MAR that follows will lead to recommendations for the use of both algorithms.

In Tables 7 and 8, algorithm 7 (ALG7) and MAR values are shown by $U\text{DIFF} = \mu_0 - \mu$ and $\beta_0/\beta$ for both $\beta = 0.5$ and $\beta = 1.0$. The
Table 7. ALG7 and MAR Means by UDIFF and $\beta_0/\beta$ for $\beta = 0.5$

<table>
<thead>
<tr>
<th>UDIFF = $\mu_0 - \mu$</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.67</td>
<td>0.68</td>
<td>0.73</td>
<td>0.73</td>
<td>0.78</td>
</tr>
<tr>
<td>MAR</td>
<td>0.35</td>
<td>0.38</td>
<td>0.37</td>
<td>0.37</td>
<td>0.25</td>
</tr>
<tr>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.69</td>
<td>0.75</td>
<td>0.77</td>
<td>0.80</td>
<td>0.79</td>
</tr>
<tr>
<td>MAR</td>
<td>0.59</td>
<td>0.65</td>
<td>0.67</td>
<td>0.67</td>
<td>0.52</td>
</tr>
<tr>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.74</td>
<td>&lt;0.81</td>
<td>&lt;0.83</td>
<td>&lt;0.82</td>
<td>&lt;0.81</td>
</tr>
<tr>
<td>MAR</td>
<td>0.70</td>
<td>&lt;0.87</td>
<td>&lt;0.93</td>
<td>&lt;0.93</td>
<td>&lt;0.70</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.77</td>
<td>&lt;0.81</td>
<td>&lt;0.80</td>
<td>&lt;0.81</td>
<td>&lt;0.78</td>
</tr>
<tr>
<td>MAR</td>
<td>0.75</td>
<td>&lt;0.88</td>
<td>&lt;1.00</td>
<td>&lt;1.00</td>
<td>&lt;0.77</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.78</td>
<td>&lt;0.82</td>
<td>&lt;0.80</td>
<td>&lt;0.79</td>
<td>&lt;0.78</td>
</tr>
<tr>
<td>MAR</td>
<td>0.61</td>
<td>&lt;0.88</td>
<td>&lt;0.93</td>
<td>&lt;0.93</td>
<td>&lt;0.70</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.83</td>
<td>0.80</td>
<td>0.78</td>
<td>0.79</td>
<td>0.81</td>
</tr>
<tr>
<td>MAR</td>
<td>0.72</td>
<td>0.78</td>
<td>0.75</td>
<td>0.75</td>
<td>0.52</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>0.79</td>
<td>0.78</td>
<td>0.77</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>MAR</td>
<td>0.69</td>
<td>0.65</td>
<td>0.51</td>
<td>0.51</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Table 8. ALG7 and MAR Means by UDIFF and $\beta_0/\beta$ for $\beta = 1.0$

<table>
<thead>
<tr>
<th>UDIFF = $\mu_0 - \mu$</th>
<th>.6</th>
<th>.8</th>
<th>1.0</th>
<th>1.2</th>
<th>1.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>ALG7</td>
<td>.32</td>
<td>.31</td>
<td>.34</td>
<td>.33</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.09</td>
<td>.08</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>-2</td>
<td>ALG7</td>
<td>.36</td>
<td>.38</td>
<td>.42</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.27</td>
<td>.35</td>
<td>.35</td>
<td>.18</td>
</tr>
<tr>
<td>-1</td>
<td>ALG7</td>
<td>.38</td>
<td>.46</td>
<td>.51</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.33</td>
<td>.71</td>
<td>.71</td>
<td>.52</td>
</tr>
<tr>
<td>0</td>
<td>ALG7</td>
<td>.52</td>
<td>.57</td>
<td>.59</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.37</td>
<td>.88</td>
<td>.88</td>
<td>.77</td>
</tr>
<tr>
<td>1</td>
<td>ALG7</td>
<td>.58</td>
<td>.57</td>
<td>.53</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.33</td>
<td>.71</td>
<td>.71</td>
<td>.52</td>
</tr>
<tr>
<td>2</td>
<td>ALG7</td>
<td>.54</td>
<td>.51</td>
<td>.48</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.30</td>
<td>.35</td>
<td>.35</td>
<td>.18</td>
</tr>
<tr>
<td>3</td>
<td>ALG7</td>
<td>.52</td>
<td>.52</td>
<td>.49</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
<td>.19</td>
<td>.12</td>
<td>.12</td>
<td>.04</td>
</tr>
</tbody>
</table>

circled results indicate those cells where MAR > ALG7. As these results indicate, algorithm 7 performs very well in comparison to MAR. It is only when the initial estimates ($\alpha_0, \beta_0$) are close to the true values ($\alpha, \beta$) (implying UDIFF close to 1 and $\beta_0/\beta$ close to 1) that MAR is superior to algorithm 7. Overall, both procedures result in higher $|M_\alpha|/|M_{opt}$ ratios for $\beta = 0.5$ than $\beta = 1.0$. This is not surprising based on Table 6 shown earlier.

Tables 7 and 8 presented results averaged over the three $\mu$ values 4, 6, and 8. Another similar look at the data is presented in Table 9. These results are averaged over $\beta_0/\beta$ but are now sorted by UDIFF and by $\mu$ for both $\beta = 0.5$ and $\beta = 0.1$. Again the circled
Table 9. ALG7, MAR Means Averaged Over $\beta_0/\beta$

<table>
<thead>
<tr>
<th>UDIFF=$\mu_0-\mu$</th>
<th>$\beta = 0.5$</th>
<th>$\beta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>-3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.78</td>
<td>.73</td>
</tr>
<tr>
<td>MAR</td>
<td>.10</td>
<td>.47</td>
</tr>
<tr>
<td>-2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.79</td>
<td>.78</td>
</tr>
<tr>
<td>MAR</td>
<td>.52</td>
<td>.68</td>
</tr>
<tr>
<td>-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.82</td>
<td>.82</td>
</tr>
<tr>
<td>MAR</td>
<td>.88</td>
<td>.80</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.79</td>
<td>.80</td>
</tr>
<tr>
<td>MAR</td>
<td>.91</td>
<td>.85</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.78</td>
<td>.83</td>
</tr>
<tr>
<td>MAR</td>
<td>.80</td>
<td>.80</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.77</td>
<td>.84</td>
</tr>
<tr>
<td>MAR</td>
<td>.65</td>
<td>.68</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG7</td>
<td>.75</td>
<td>.81</td>
</tr>
<tr>
<td>MAR</td>
<td>.46</td>
<td>.54</td>
</tr>
</tbody>
</table>

Entries indicate those cells where MAR > ALG7. As can be seen, algorithm 7, the modified Federov algorithm, performs better than the 2 point D-optimal algorithms based on initial estimates $\alpha_0$, $\beta_0$ whenever $|UDIFF| > 1$. This section provides the comparisons necessary to allow recommendations to be made for the suggested use of these two algorithms.

Robustness of ALG7

In addition to comparing the mean performance of the modified "forward only" modified Federov algorithm (ALG7) to the 2 point
D-optional Ford-Silvey-White algorithm based on initial estimates of $\alpha_0$ and $\beta_0$ (MAR), it is worthwhile to compare the variability of the performance of these two algorithms over the range of input assumptions used by $UDIFF = \mu_0 - \mu$ and $\beta_0 / \beta$. Across the 35 different combinations of $UDIFF$ and $\beta_0 / \beta$ (7 levels of $UDIFF$ times 5 levels of $\beta_0 / \beta$) the mean, minimum, maximum, and standard deviations are shown in Table 10 below for MAR and ALG7 for the 3 $\mu$ and 2 $\beta$ levels used in this analysis. The table also summarizes the results over the 3 $\mu$ levels. Recall a summary over $\beta$ is not possible since different $|M_{\text{opt}}|$ denominators are used in the performance measure $|M_A| / |M_{\text{opt}}|$ for $\beta = 0.5$ and $\beta = 1.0$. As Table 10 indicates ALG7 exhibits considerably less variability than MAR. This can be seen graphically in the plots seen in Figures 15 through 18. Thus ALG7 is considerably more robust than MAR. Also note that averaged over all the $\beta = 0.5$ and $\beta = 1.0$ data ALG7's mean performance is better by 0.10, and 0.11, respectively.

<table>
<thead>
<tr>
<th>$\beta = 0.5$</th>
<th>$\beta = 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>Mean</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>4</td>
<td>ALG7</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>6</td>
<td>ALG7</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>8</td>
<td>ALG7</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
</tr>
<tr>
<td>All</td>
<td>ALG7</td>
</tr>
<tr>
<td></td>
<td>MAR</td>
</tr>
</tbody>
</table>
FIGURE 15. PLOT OF RANGE OF MAR FOR $\beta = 0.5$.

FIGURE 16. PLOT OF RANGE OF MAR FOR $\beta = 1$.

FIGURE 17. PLOT OF RANGE OF ALG7 FOR $\beta = 0.5$.

FIGURE 18. PLOT OF RANGE OF ALG7 FOR $\beta = 1$. 
Least Squares Prediction of ALG7 Performance

While the previous tables and figures have highlighted various summary aspects of ALG7, we would like to estimate its performance over the entire design space. The analysis to this point has examined the performance of ALG7 only at discrete points in the design space of the independent variables. It is important to predict this algorithm's D-optimal efficiency over the continuum of the design space so that one can estimate its behavior at any desired values. This section provides contour maps that will serve as a guide to accomplish this goal. Based on the 35 mean values corresponding to the levels of UDIFF and $B_0/B$, a least squares quadratic regression curve was fit for each of the 6 combinations of $\mu$ and $\beta$. These results are depicted in the contour maps in Figures 19 through 24 along with summary statistics. The quadratic model equation thus has UDIFF and $B_0/B$ as the independent variables used to predict ALG7 performance. The most striking aspect of these predicted contours for ALG7 is that the impact of UDIFF on the prediction is much greater than the ratio $B_0/B$. This agrees with the earlier analysis of variance results. These contour surfaces now allow one to predict the performance of ALG7 for any combination within the range of the two independent variables used.
FIGURE 19. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_0$ AND $B_0/B$ FOR $\mu = 4$ AND $B = 0.5$ ($R^2 = 0.78$, ROOT M.S.E. = 0.03)

FIGURE 20. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_0$ AND $B_0/B$ FOR $\mu = 6$, AND $B = 0.5$ ($R^2 = 0.67$, ROOT M.S.E. = 0.03)
FIGURE 21. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_O$ AND $B_O/B$ FOR $\mu = 8$ AND $B = 0.5$ ($R^2 = 0.82$, ROOT M.S.E. = 0.03)

FIGURE 22. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_O$ AND $B_O/B$ FOR $\mu = 4$ AND $B = 1$ ($R^2 = 0.71$, ROOT M.S.E. = 0.05)
FIGURE 23. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_0$ AND $B_o/B$ FOR $\mu = 6$, AND $B = 1$ ($R^2 = 0.87$, ROOT M.S.E. = 0.04)

FIGURE 24. QUADRATIC REGRESSION PREDICTION OF ALG7 AS A FUNCTION OF $\mu_0$ AND $B_o/B$ FOR $\mu = 8$, AND $B = 1$ ($R^2 = 0.86$, ROOT M.S.E. = 0.04)
Suggested Use of ALG7, MAR

This chapter has summarized the results of the second phase of analysis comparing various fixed and sequential algorithms on the basis of D-optimality when time was the independent variable for the binary logistic response. Of the sequential algorithms, ALG7 provided the best results over the experimental region of interest. It was compared to the fixed algorithm MAR based on previous D-optimal work for the logistic distribution. ALG7 is a robust sequential algorithm that performs respectably over the experimental range of interest. Only when the initial estimates $\alpha_0$ and $\beta_0$ result in $|UDIFF| < 1$ and $0.8 < \beta_0/\beta < 1.2$ is MAR superior. If one feels confident that one's initial estimates satisfy these constraints, then MAR is preferred over ALG7. Otherwise, ALG7 is the better choice.

The knowledge gained by this research indicates that sequential algorithms can result in better D-optimal designs than fixed algorithms when the initial estimates of $\alpha$ and $\beta$ differ moderately from their true unknown values. What is meant by moderately is quantified above. The key independent variables that impact algorithm performance have been identified. In addition, predictive contour surfaces have been developed to estimate the behavior of the best sequential algorithm over a continuum. In the next chapter we briefly summarize this entire study, and indicate areas of future research.
CHAPTER 6 SUMMARY AND RECOMMENDATIONS FOR FUTURE RESEARCH

Summary

The previous chapters have presented the problem of experimental design for a binary response with an expected value that is an increasing function of time. Such a problem is a common one in many areas in which an item degrades in some respect with time. As long as the dependent variable of the relationship may be categorized into two divisions such as "good-bad", "pass-fail", "operational-nonoperational", then the process exhibits a binary response which may be analyzed using techniques like those presented in this study.

The first chapter introduced the binary response curve as a function of a continuous independent variable. A review of the published literature was performed and the documented methods categorized based on five classification variables. The algorithms developed in this study fit the global-classical-parametric-design-sequential cell of Table 1. Some of the major approaches to the binary response problem were detailed in Chapter 1. Chapter 2 examined experimental design optimality criteria, in particular D and G optimality. The General Equivalence Theorem and several extensions were given. The Federov extension and the nonlinear general
equivalence theorems of White and Ford were shown to be equivalent for the logistic distribution. Major properties of the logistic distribution relating to maximum likelihood estimation and the Fisher Information matrix are presented. An example, based on a canned milk product, of the binary response problem to be examined in this study is given at the end of Chapter 2, followed by a listing of the study objectives below:

1. Develop and compare new algorithms based on D-optimality for this problem.

2. Determine the sensitivity of the algorithm responses to the variables used in the experimental designs of later chapters.

3. Make recommendations for the use of the algorithms.

Chapter 3 examined experimental design for the logistic distribution. Included in this chapter are the 2 point D-optimal algorithm of Ford, Silvey, and White, and new theoretical developments for optimal 2 point designs for which the first point is already given. This latter item serves as the basis for two of the four sequential algorithms proposed at the end of Chapter 3. The remaining two sequential algorithms are based on "forward only" sequential extensions of a fixed design algorithm by Federov. The second of the extensions of Federov's work allows the incorporation of constraints requiring at least a minimum number of observations be collected at time points pre-specified by the individual designing the experiment.
Chapters 4 and 5 presented the results of comparing the algorithms mentioned in Chapter 3 using simulated results. The new algorithms presented (ALG4, ALG5, ALG6, and ALG7) are sequential and vary the experimental design as the current estimates of the true unknown logistic parameters $\alpha, \beta$ are updated each time new data are collected. These algorithms are compared with fixed design algorithms that do not vary the experimental design during the course of the experiment. The pilot study experimental design had the following design variables for each algorithm:

- $(\alpha_0, \beta_0)$: initial estimates of $(\alpha, \beta)$
- $AR = \alpha/\alpha_0$
- $BR = \beta/\beta_0$
- $K$: number of grid points in $[0, 12]$ interval
- $N$: total sample size.

The performance measure studied for both the pilot and secondary study is the ratio $|M_A|/|M_{\text{opt}}|$ where $|M_A|$ is the determinant of the Fisher Information matrix for a given iteration of an algorithm, and $|M_{\text{opt}}|$ is the determinant of the D-optimal 2 point design using the true parameter values $\alpha$ and $\beta$. Because of the structure of the pilot study, experimental design algorithm comparisons were made within the cells of Tables 2 and 3. However, the pilot study did allow the removal of design variables $K$ and $N$. Of the four sequential algorithms, ALG4 and ALG7 performed the best.
over the range of design variables used. Considerable insight into the problem resulted from the pilot study and led to a secondary analysis that allowed more extensive algorithm comparisons.

Chapter 5 presented a second phase of analysis based on a different experimental design structure. In this phase two analyses, an in-depth study of a smaller set of experimental variables was performed. Due to a different way of viewing the problem, algorithm comparisons could now be carried out over a wider range of the design variables. This different viewpoint is based on a parameterization that resulted in the following design variables for each algorithm:

\[ \text{UDIFF} = \mu_0 - \mu \]
\[ \frac{\beta_0}{\beta} \]
\[ \mu \]
\[ \beta \]

Several methods were used to analyze the simulated results. The design variables accounting for the larger portions of the total sum of squares for the analysis of variance tests performed were given. Of the four sequential algorithms, ALG7 was identified as the best overall. Detailed comparisons were made between ALG7 and the fixed design MAR. Unless the initial parameter estimates are close to the true unknown parameter values, ALG7 is recommended over MAR. The modified Federov ALG7 also exhibited considerably less variability than MAR for the conditions tested.
Contour plots were developed predicting the behavior of ALG7 as a function of \( \mu_0 \) and \( \beta_0 / \beta \) for the values of \( \mu \) and \( \beta \) studied. These plots further emphasize earlier results indicating that knowledge of \( \mu \) is more important than the ratio of \( \beta_0 / \beta \) in predicting the algorithm's performance. Finally, suggested use of both ALG7 and MAR is given.

**Areas for Future Research**

In the use of algorithm 7 (ALG7) the minimum percentage of observations, \( m_i \), to be collected at each design point was selected so that \( n_i \geq 0.5 \, (N/K) \), i.e., at least 50 percent of the equal allocation (ALG3 where \( n_i = N/K \) for all \( i \)) number of observations would be collected at each possible design point. Based on this, it can be seen that ALG7 is bounded in design by ALG3 and ALG6 (modified Federov with no constraint on the minimum number of samples collected at any design point). This could be treated parametrically where \( \rho = \frac{m_i}{N/K} \in [0,1] \). For ALG7 studied, \( \rho = 0.5 \). When \( \rho = 0 \) this algorithm becomes ALG6, and when \( \rho = 1 \) it becomes ALG3. Further research could study the effect of different values of \( \rho \) on the performance measure \( |M_A|/|M_{opt}| \) used in this study. Another approach would involve dynamic modification of \( \rho \) using dynamic programming techniques.

Another application of ALG7 involves the selection of different initial constraints at the potential design points. There exists nothing in our extension of Federov's work that forces this
constraint to be the same over all points. Therefore, the experimenter could choose those constraints most meaningful to him. Indeed the experimenter may change the forward constraints during the running of the experiment.

This study focused on the "forward only" binary response problem. If the independent variable was not time or subsequent batches of the same product could be assumed to be identical, then ALG7 could also be applied with a minor modification. In its D-optimal search for the $p_i$ values after each collection of observations, it would have to scan all design points instead of only the forward ones. Another trivial extension would impose a maximum number of observations that could be selected at any design point. This extension would involve the removal of any design point at its maximum from the D-optimal search of $p_i$ values. Thus, it can be seen that ALG7 has considerable flexibility and can easily be adapted for many needs.

Instead of using a modified Federov approach to select the $p_i$ that determine the sample allocation, recall from Chapter 2 equation (49). All the constraints imposed in this study have been linear. Therefore, quadratic programming could be used with equation (49) as the objective function to be maximized. The speed of convergence of the modified Federov and quadratic programming approaches could be compared, and the quicker of the two used to select the $p_i$ at each point in time.

The present work is univariate in both the dependent and independent variables. Much of the present work could be extended
to allow multiple independent variables. The algorithmic design construction in ALG7 could be expanded for a grid that encompasses the design points to be considered. If a continuum instead of a grid is desired, then a nonlinear optimization routine must be used to identify the point at which future data collections should occur. In this case, the bookkeeping would also become significantly more difficult since no longer would there be a fixed number of possible design points. Extensions to multivariate dependent variables may be possible. Federov (1972) would be a good source of initial information.

The problem of studying product degradation over time may be viewed from a wider economic sense. The cost of experimentation could be incorporated into the problem as in previous work with linear models by Mount-Campbell and Neuhardt (1982). There is risk associated with the first batch of any new product released to the market. Based on the current parameter estimates \( \alpha, \beta \) management could make a decision to sample more observations at the current time period to help determine if the product is acceptable or perhaps should be recalled from warehouses before more of it is released to retail stores. Typically, food retail stores have a quick turnover, so removal of warehouse stock will greatly diminish the chance of the public receiving older cans of product that may be aging faster than the company desires. Such a recall for cosmetic defects can be quietly undertaken without the need for public awareness in the matter. The cost associated with the risk of poor publicity on a given product may be balanced with the cost of
information loss of a non-optimal design to arrive at a design that is optimal in a larger economic sense. Whenever possible, global performance measures such as this should be investigated.
LIST OF REFERENCES


Karber, G. "Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche", Arch. Exp. Path Pharmak, 1931, 162, pp. 480-487.


SAS/GRAPH Enhancements and Updates for 82.2 Release, SAS Institute Inc., 1982, Cary, N.C.


