One-Stage and Bayesian Two-Stage Optimal Designs for Mixture Models

By

Hefang Lin

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

DOCTOR OF PHILOSOPHY
IN
STATISTICS

APPROVED:

Raymond H. Myers (Co-chair)

Keying Ye (Co-chair)

Christine M. Anderson-Cook

Robert V. Foutz

Marion R. Reynolds

December 9, 1999
Blacksburg, Virginia

One-Stage and Bayesian Two-Stage Optimal Designs

for Mixture Models

by

Hefang Lin[1], Raymond H. Myers, Keying Ye

In this research, Bayesian two-stage D-D optimal designs for mixture experiments with or without process variables under model uncertainty are developed. A Bayesian optimality criterion is used in the first stage to minimize the determinant of the posterior variances of the parameters. The second stage design is then generated according to an optimality procedure that collaborates with the improved model from first stage data. Our results show that the Bayesian two-stage D-D optimal design is more efficient than both the Bayesian one-stage D-optimal design and the non-Bayesian one-stage D-optimal design in most cases. We also use simulations to investigate the ratio between the sample sizes for two stages and to observe least sample size for the first stage. On the other hand, we discuss D-optimal second or higher order designs, and show that D_s-optimal designs are a reasonable alternative to D-optimal designs.

Keywords: Bayesian, Two-Stage, Mixture Experiments, Process Variables, Optimality

[1] If anyone want to use Bayesian two-stage D-D optimal design, please fell free to contact Hefang lin for the SAS program. Her email is helin@vt.edu and home address is 10570 Cross Fox Lane, Columbia MD 21044.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to my advisors, Dr. Raymond H. Myers and Dr. Keying Ye for sharing their knowledge with me. I have been fortunate to have both of them as my advisors, because they helped to make my dreams come true. I am deeply indebted to their mentoring, encouragement, support, and discipline over these four and half years. I will always treasure their friendship. I would also like to thank my colleagues for many fruitful discussions I had with them.

I thank my parents for their love and sacrifice over the years of my study abroad. Finally and most importantly, I thank my husband Honghui Chen, and my daughter Jessica Chen, whose constant love, understanding, support made everything possible.

Contents

	Page
Abstract	i
Acknowledgements	ii
List of Tables	vi
List Figures	iv
Chapter 1 Introduction and Motivation	1
1.1 Introduction	1
1.2 D-Optimality Criterion, Bayesian Designs, and Two-Stage Designs	2
1.3 D_s-Optimality Criterion	5
1.4 Outline of the Dissertation	5
Chapter 2 Literature Review	7
2.1 Mixture Experiments	7
2.2 Mixture Experimental Model and Design D-optimality	9
2.3 Bayesian One-Stage D-optimal Design	11
2.4 Two-Stage Designs	13
2.4.1 Non-Bayesian D-D Optimal Design	14
2.4.2 Bayesian Two-Stage D-D Optimal Design	15
Chapter 3 Bayesian One-Stage D-Optimal Design for Mixture Models	20
3.1 The Selection of Design Optimality	20
3.2 Existing Bayesian D-optimal Designs	21

3.3 Modified Bayesian D-optimal Designs 22

3.4 Example of Bayesian D-optimal Designs 23

3.5 Comparisons and Results 27

Chapter 4 Bayesian Two-Stage D-D Optimal Design for Mixture Models 34

4.1 Selection and Analysis for the First Stage D-optimal Design 34

4.2 Uncertainty in τ^2 37

4.3 Selection of the Second Stage Bayesian D-optimal Designs 38

4.4 Examples of Bayesian D-D Optimal Designs 40

4.4.1 A Linear Primary and Quadratic Potential Model for Three Components 40

4.4.2 A Quadratic Primary and Cubic Potential Model for Four Components 44

4.5 Two Classes of Potential Terms 48

4.5.1 The Posterior Density of M_i Given y_1 49

4.5.2 Uncertainty in τ_1^2 and τ_2^2 50

4.5.3 An Example of A Linear Primary and Both Quadratic and Cubic Potential Model for Three Components 52

Chapter 5 Evaluation of Bayesian Two-Stage D-D Optimal Designs 56

5.1 Method of Evaluating Bayesian Two-Stage D-D Optimal Designs 56

5.2 A Example of Results Three Components 58

5.3 The Results of Evaluation for Four Components 61

5.3.1 A Linear Primary and Quadratic Potential Model 61

5.3.2 A Quadratic Primary and Cubic Potential Model 64

5.4 A Example of the Results of Evaluation for Three Components with Two Classes of Potential Terms 66

5.5 Conclusion 69

Chapter 6 Distribution of the Sample Size between Two Stages 70

6.1 A Reasonable Ratio of the Total Size between Two Stages 71

6.2 The Choice of Sufficient Data for the First Stage Design 73

Chapter 7 Application with Restrictions on Component Proportions 76

7.1 Case where Feasible Region is a Simplex 76

7.2 Case in Which the Feasible Region is not a Simplex 78

Chapter 8 Bayesian Two-Stage Designs for Mixture Experiments with Process Variables 81

8.1 Introduction 81

8.2 Examples of Bayesian Two-Stage Designs 83

8.2.1 Three Components with One Process Variable 84

8.2.2 Three Components with Two Process Variables 87

8.3 Evaluation of the designs 91

8.4 Results of Bayesian Two-Stage D-D Designs 92

8.4.1 Three Components with One Process Variable 92

8.4.2 Three Components with One Process Variable for Two Classes of Potential Terms 94

8.4.3 Three Components with Two Process Variables 96

8.5 Conclusion 98

Chapter 9 D_s-Optimal Designs for Mixture Models 100

9.1 Criteria for D_s-Optimal designs 101

9.2 D_s-Optimal Designs for Three Components with Second Order Model 101

9.3 D_s-Optimal Designs for Four Components with Second Order Model 105

9.4 D_s-Optimal Designs for Four Components with Special Cubic Model 107

9.5 Comparing Designs between D_s and D through Plots along Prediction Rays 111

Chapter 10 Discussions and Future Research 122

Bibliography 124

Appendix

A. Calculation of $p(y|M_i)$ and $p(M_i|y)$ for section 4.1 129

B. Calculation of $p(y|M_i)$ and $p(M_i|y)$ for section 4.5.1 132

Vita 135

List of Tables

Table		Page
3.1	The determinant of $D=(X'X+K/\tau^2)$	27
4.1	Posterior probabilities of candidate models for three components with linear primary and second-order potential terms model	43
4.2	Posterior probabilities of candidate models for four components with second-order primary and special cubic potential terms model	47
4.3	Posterior probabilities of candidate models for three components with two classes of potential terms	55
5.1	Values of D* and A* for three components with only primary terms in the true model	60
5.2	Values of D* and A* for three components with primary terms and one potential term in the true model	61
5.3	Values of D* and A* for three components with primary terms and two potential terms in the true model	61
5.4	Values of D* and A* for three components with primary terms and three potential terms in the true model	61
5.5	Values of D* and A* (in parentheses) for four components with linear primary and second-order potential terms model	63
5.6	Values of D* and A* (in parentheses) for four components with second order primary and special cubic potential terms model	65
5.7	Values of D* and A* (in parentheses) for three components with two classes of potential terms	68

6.1	Simulation study for data allotment with sample size 24	72
6.2	Simulation study for data allotment with sample size 30	72
6.3	Simulation study for sufficient data in the first stage (n=16)	74
6.3	Simulation study for sufficient data in the first stage (n=12)	74
8.1	Posterior probabilities of candidate models for three components with one process variable	86
8.2	The values of D* for three components with one process variable	94
8.3	The values of D* for three components with one process variable when full model contains two classes of potential terms	96
8.4	The values of D* for three components with two process variables	98
9.1	The determinant of D=|X'X| for three components with second-order model between D_s and D-optimal designs	104
9.2	The determinant of $D_s=\|M_{22}-M_{12}'M_{11}^{-1}M_{12}\|$ for three components with second-order model between D_s and D-optimal designs	104
9.3	The determinant of D=|X'X| for four components with second-order model between D_s and D-optimal designs	107
9.4	The determinant of $D_s=\|M_{22}-M_{12}'M_{11}^{-1}M_{12}\|$ for four components with second-order model between D_s and D-optimal designs	107
9.5	The determinant of D=|X'X| for four components with special cubic model between D_s and D-optimal designs	110
9.6	The determinant of $D_s=\|M_{22}-M_{12}'M_{11}^{-1}M_{12}\|$ for four components with special cubic model between D_s and D-optimal designs	110
9.7	D-optimal design for four components with second-order model (n=15)	113

9.8 D_s-optimal design for four components with second-order model (n=15) 113

9.9 D-optimal design for six components with second-order model (n=38) 116

9.10 D_s-optimal design for six components with second-order model (n=38) 117

List of Figures

Figure		Page
2.1	Constraints factor space for three components	8
3.1	The sequence of designs with n=8 for a linear primary and a quadratic potential model in an unconstrained region (our designs)	25
3.2	The sequence of designs with n=8 for a linear primary and a quadratic potential model for three components (Andere-Rendon et al.'s designs)	26
3.3	The sequence of designs with n=12 for a linear primary and a quadratic potential model for four components	31
3.4	The sequence of designs with n=9 for a linear primary and both a quadratic and a special potential model for three components	32
3.5	The sequence of designs with n=16 for a quadratic primary and a special cubic potential model for four components	33
4.1	Bayesian first stage D-D optimal design for three components with a quadratic full model	41
4.2	The posterior density of τ^2 given the true model contains only primary terms	42
4.3	Bayesian second stage D-D optimal design for three components given the true model contains only primary terms	43
4.4	The combined two-stage D-D optimal design for three components given the true model contains only primary terms	44
4.5	Bayesian first stage D-D optimal design for four components with a special cubic full model	45

4.6	The posterior density of τ^2 given the true model contains one of the potential terms	46
4.7	Bayesian second stage D-D optimal design for four components given the true model contains a cubic term $x_1x_2x_3$ primary term	47
4.8	The combined two-stage D-D optimal design for four components given the true model contains a cubic term $x_1x_2x_3$ primary term	48
4.9	Bayesian first stage D-D optimal design for three components with two classes of potential terms	53
4.10	The posterior density of τ_1^2 and τ_2^2 given the true model contains one from each of classes of potential terms	54
4.11	Bayesian second stage D-D optimal design for three components given the true model contains second order x_1x_3 and a cubic term $x_1x_2x_3$ potential term	55
4.12	The combined two-stage D-D optimal design for three components given the true model contains second order x_1x_3 and a cubic term $x_1x_2x_3$ potential term	55
7.1	Example where the feasible region is a simplex	76
7.2	Example where the feasible region is an inverted simplex region	77
7.3	The first stage of Bayesian two-stage design for three constraints components	79
7.4	A second stage of Bayesian two stage design for three constraints components	79

7.5	The first stage of Bayesian two-stage design for four constraints components	80
7.6	A second stage of Bayesian two stage design for four constraints components	80
8.1	Bayesian first stage design for three components with one process variable (n_1=12)	85
8.2	The posterior density of τ^2 given the true model contains only the primary terms	86
8.3	A Bayesian second stage design for three components with one process variable (n_2=12)	87
8.4	A Bayesian first stage design for three components with two process variables (n_1=17)	89
8.5	A Bayesian second stage design for three components with two process variables (n_2=17)	91
9.1	D_s- and D-optimal designs for three components with a quadratic model	103
9.2	D_s- and D-optimal designs for four components with a quadratic model	106
9.3	D_s- and D-optimal designs for four components with a cubic model	109
9.4	Plot of prediction variance traces for D and D_s-optimal designs: x_1, magnesium; x_2, sodium nitrate; x_3, strontium nitrate; x_4, binder	114
9.5.1	The prediction variance along X_1	118
9.5.2	The prediction variance along X_2	
9.5.3	The prediction variance along X_3	119
9.5.4	The prediction variance along X_4	119

9.5.5 The prediction variance along X_5 120

9.5.6 The prediction variance along X_6 120

Chapter 1

Introduction and Motivation

1.1 Introduction

To develop, improve, and optimize processes, scientists use response surface methodology (RSM), a collection of statistical and mathematical techniques. RSM has important applications in the design, development, and formulation of new products as well as in the improvement of existing product designs. The direct interest in RSM for this research is in regard to the mixture experiment. A mixture experiment is a special type of response surface experiment with two primary features: factors that are the ingredients or components of a mixture, and a response that depends on the proportions of ingredients (Myers and Montgomery, 1995). For example, suppose the flavor of a fruit punch consisting of juices from apples, pineapples, and oranges is studied. Here, the flavor of the punch depends on the proportions of the juices in the blend, not on the amount of punch. In some mixture experiments, the response depends not only on the proportion of the components in the blend, but also on the processing conditions. Process

variables are factors that do not constitute any portion of the mixture, but whose levels, when changed, could affect the response (Cornell, 1990).

Regardless of the nature of a model, one can safely assume that cost is an important factor in most experiments. Therefore, one should try to find the most efficient design to estimate parameters in the right model. In other words, one should try to have designs that are not only theoretically efficient but attractive to practitioners in their practicality. The primary objective of this dissertation is to find designs that have a combination of these attributes. The particular model in this research is the mixture experiment model with or without process variables.

1.2 D-Optimality Criterion, Bayesian Designs, and Two-Stage Designs

Out of the variety of **alphabetic optimality criteria**, D-optimality, dating back to Kiefer and Wolfowitz (1959), is the most widely known design criterion. This criterion suggests one should reduce uncertainty in model coefficient estimates. The D-optimal design for a normal linear model is defined as the one that maximizes the determinant of (X'X/n), where X is the model matrix and n is the sample size. A design in which X contains too many regressors (overspecification) could well estimate unimportant parameters (effects), thus wasting some observations. On the other hand, a design in which X contains too few regressors (underspecification) will not allow all necessary parameters to be estimated, and an inadequate model will result. Overspecification or underspecification of regressors is called model misspecification. Hence, D-optimal design will not give an explicit provision for diagnosing inadequacies in the model. Unfortunately,

experimenters do not always have adequate information about X. Therefore, the D-optimality criterion is often difficult to utilize.

Since the D-optimal design is greatly dependent on the assumed fitting model, a Bayesian approach is used to reduce this dependency. DuMouchel and Jones (1994) proposed a Bayesian approach to obtain a D-optimal design that is less dependent on the assumed model, and Andere-Rendon, Montgomery, and Rollier (1997) used a Bayesian modification to study the performance of Bayesian D-optimal design for mixture experiments. The important factor of these designs is to use prior distributions of the coefficients in a model to control the uncertainty of their corresponding variables. Compared to the standard D-optimal designs, Bayesian D-optimal designs for mixture experiment models produce significantly smaller bias errors and improve the coverage of the factor space. In addition, they still allow the fitting of a larger number of high order terms and maintain good variance properties.

On the other hand, many authors have discussed Bayesian designs in general situations. Covey-Crump and Silvey (1970) introduced the notion of a Bayesian optimality criterion for the linear model case. Brooks (1972), Bernardo (1979), Nather and Pilz (1980), and Chaloner (1984) developed and investigated several Bayesian optimal designs for the linear model. Chaloner and Larntz (1989) developed a Bayesian design procedure for the logistic regression model.

Another way of dealing with model misspecification is to run the experiment in stages. Box and Lucas (1959) suggested sequential experimentation. Abdelbasit and Plackett (1983) and Minkin (1987) discussed two-stage designs for one variable logistic regression using the D-optimality criterion. They found that non-Bayesian two-stage

D-D optimal designs for the logistic model were more robust to poor parameter guesses than the single stage equivalent. Letsinger (1995) combined the Bayesian design with a two-stage approach to generate an efficient D-optimal design for the logistic model. This two-stage approach results in good efficiency and robustness properties. Myers, Myers, Carter, and White (1996) developed a two-stage D-Q design for the logistic model. Both studies found that the two-stage design is superior to the comparable one-stage design in most cases where parameter guesses are made. Neff (1996) developed Bayesian two-stage designs under model uncertainty for normal linear models. She found when the right model is a subset of an assumed full model, designs obtained by using a Bayesian two-stage procedure are more efficient than traditional designs.

Numerous design procedures have been developed to address the problem of model misspecification for a normal linear model and a logistic model. Very little work has been done to address the problem of model misspecification for mixture experimental models with or without process variables. To generate efficient designs with reduced dependence on regressor specification for mixture experimental models with or without process variables, Bayesian two-stage D-D optimal designs are developed. In these designs, a Bayesian D-optimality criterion is used in the first stage to minimize the determinant of the posterior variances of the parameters. The second stage design is then generated according to a D-optimality procedure that uses the improved model from the first stage data.

1.3 D_s–Optimality Criterion

D_s–optimality is an extension to accommodate the situation in which a subset of the parameter vector (β_s) is of interest, which is defined by minimizing the determinant of the variance-covariance matrix corresponding to the estimator of subset β_s. Kiefer (1961) first suggested D_s–optimality. Hill and Hunter (1974) discussed an application of the D_s–criterion for a nonlinear regression problem. Wardrop and Myers (1990) found that in many response surface studies there is more to be gained by good estimation of the second-order coefficients. Huffman and Myers (1998) discussed D_s-optimal designs for exponential models. Neither of them studied the performance of D_s-optimal design for a mixture experimental model. Therefore, this dissertation discusses the D_s-optimal designs for mixture models and makes comparisons between D_s-optimal designs and D-optimal designs.

1.4 Outline of the Dissertation

In Chapter 2, a literature review of one-stage D and two-stage D-D optimal designs, and mixture experiments is provided. Chapter 3 presents the development of Bayesian one-stage D-optimal designs for mixture experimental models. In Chapter 4, Bayesian two-stage D-D optimal designs for mixture experimental models are developed. In Chapter 5, evaluations of Bayesian two-stage D-D optimal designs are made. In Chapter 6 distribution of the sample sizes between the two stages is considered. In Chapter 7, the application of Bayesian two-stage D-D optimal design to the mixture experiment with constraint region is discussed. In Chapter 8, Bayesian two-stage D-D optimal designs for mixture models with process variables are generated. In Chapter 9, D_s-optimal designs

for mixture models are generated, and comparisons between D_s-optimal design and D-optimal design are made. Finally, in Chapter 10, the work is summarized, and future research goals are proposed.

Chapter 2

Literature Review

2.1 Mixture Experiments

Suppose a product is formed by q ingredients or components, and let x_i be the proportion of the i^{th} ingredient or component, then x_i must satisfy (1) $0 \le x_i \le 1$, and (2) $\sum x_i = x_1 + x_2 + \cdots + x_q = 1$. Because of the latter constraint, the levels of the factor x_i are not independent and mixture experiments are different from the usual response surface experiments. The design space for three components is shown in Figure 2.1, which is a regular-sided hyperplane with only two dimensions rather than a cubic with three dimensions. Here, the vertices represent pure blends, the points on the edges represent binary blends, and the center point represents a mixture with equal proportions of the three components. In general, the design space with q components is a simplex, which is a regular-sided hyperplane with q vertices in (q-1) dimensions. If there are more constraints on the proportions, then the factor space is a general convex (q-1) dimensional polyhedron.

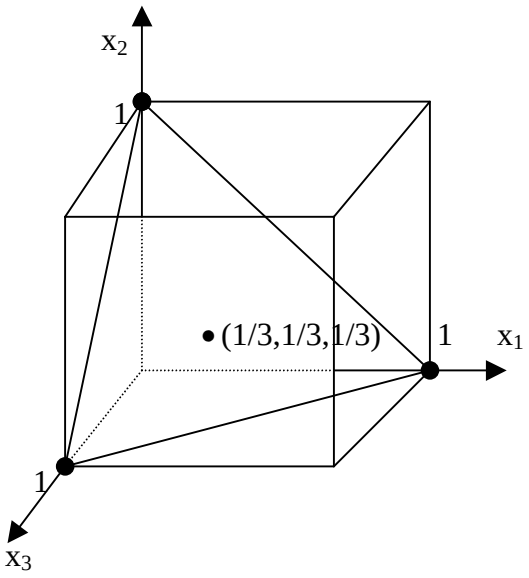


Figure 2.1. Constraints factor space for three components

As in standard non-mixture experiments, the choice of a design for a mixture experiment is also based on the particular assumed fitting model. Typically, mixture experimental models are of the Scheffe type models (Scheffe 1958, 1959, 1965), which are called canonical models, where the first-order, second-order and special cubic of a third-order models are written as

$$y = \beta_1 x_1 + \ldots + \beta_q x_q + \varepsilon = \sum_{i=1}^{q} \beta_i x_i + \varepsilon,$$

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \beta_{ij} x_i x_j + \varepsilon,$$

and

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \beta_{ij} x_i x_j + \sum_{i<j<k}^{q} \beta_{ijk} x_i x_j x_k + \varepsilon,$$

respectively.

The canonical models can be reparameterized to slack models that contain a constant term, or intercept β_0 by simply deleting one of the $\beta_i x_i$ terms, say $\beta_q x_q$. Thus, instead of the linear mixture model

$$y = \beta_1 x_1 + \ldots + \beta_q x_q + \varepsilon,$$

we can write

$$y = \beta_0 + \beta^*_1 x_1 + \ldots + \beta^*_{q-1} x_{q-1} + \varepsilon.$$

In addition, rather than the following quadratic model

$$y = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \beta_{ij} x_i x_j + \varepsilon,$$

we can write

$$y = \beta_0 + \sum_{i=1}^{q-1} \beta^*_i x_i + \sum_{i<j}^{q} \beta_{ij} x_i x_j + \varepsilon$$

and so on.

In a mixture experiment, the coefficients of a canonical model allow for ease in interpreting the results of the experiments, while the coefficients of a slack model do not provide a convenient interpretation. Therefore, a canonical model is preferred to a slack model (Cornell, 1990).

2.2 Mixture Experimental Model and Design D-Optimality

A mixture experimental model can also be written in standard form,

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

In this model,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad \mathrm{X} = \begin{bmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1k} \\ x_{21} & \cdot & \cdot & \cdot & x_{2k} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{nk} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix},$$

where n is the number of experimental runs and k is the number of parameters in the assumed model. The error $\boldsymbol{\varepsilon}$ is assumed to have a normal distribution with mean zero and common variance σ^2. Under the assumption that $\boldsymbol{\varepsilon}$ is N(0, $\sigma^2 I$), the least squares estimator of β is equal to the maximum likelihood estimator, $\mathbf{b} = (X'X)^{-1}X'y$, which is the uniformly minimum variance unbiased estimator of β. The variance of **b** is $\sigma^2(X'X)^{-1}$, which is the inverse of the Fisher information matrix. Many of the design criteria have been developed regarding the variance covariance matrix of **b**. In this research, D-optimality is considered. This criterion suggests that one should make $(X'X)^{-1}$ small or equivalently, make $(X'X)$ large, thus attempting to minimize the variance of **b**. The generalized variance of **b** is the determinant of the variance covariance matrix of **b**, and Wilk (1932) used it as a scalar measure of dispersion. Thus, a D-optimal design maximizes the determinant of $(X'X/n)$. The D-optimality criterion is given by

$$\underset{D}{\mathrm{Max}} \left| \frac{X'X}{n} \right|,$$

where D is the set of possible designs and n is the total number of runs in the experiment. where k is the number of the parameters in the full model. Notice that the determinant does not depend on the unknown parameter vector β, only on the location of design points, assumed model and sample size.

Once the model form, design space, and sample size have been chosen, it is relatively easy to select the corresponding design by using "proc optex," a procedure in SAS. The D-optimal design for the mixture experiment consists of the extreme vertices for a first-order model. For a second–order model the D-optimal design for the mixture experiment contains the extreme vertices and the edge midpoints. Finally, the D-optimal design with a special cubic model for three components consists of the three vertices, the edge midpoints, and the overall centroid (Andere-Rendon et al., 1997).

2.3 Bayesian One-Stage D-optimal Design

Andere-Rendon et al. (1997) used a Bayesian modification to reduce the dependency of D-optimal designs on the assumed models for mixture experiments. Compared to standard D-optimal designs, their designs produce significantly smaller bias errors and improved coverage of the factor space. In addition, they allow fitting of a large number of higher order terms and maintain good variance properties. This section presents the method of Bayesian D-optimal design.

The regressors believed to be important in modeling the response are called primary terms, while the uncertain regressors are treated as potential terms. For example, the experimenter may know from past experience that linear terms must be included in the model but is uncertain whether quadratic terms are needed. In this case, we may treat linear terms as primary terms and quadratic terms as potential terms. For this case, the model can be written as $\mathbf{y} = X_{pri}\ \beta_{\mathbf{pri}} + X_{pot}\beta_{\mathbf{pot}} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ is the vector of $(n \times 1)$ responses distributed as $N(X\beta, \sigma^2 I)$, $X = (X_{pri} \mid X_{pot})$ is a matrix of $(n \times (p+r))$, $\beta_{\mathbf{pri}}$ is a vector of $(p \times 1)$, and $\beta_{\mathbf{pot}}$ is a vector of $(r \times 1)$.

To minimize any correlation between primary and potential terms, DuMouchel and Jones (1994) recommended a scaling convention. The idea of scaling convention is to make each potential term and all primary terms orthogonalized. That is, $\sum X_1 X_2 = 0$, for each primary term X_1 and each potential term X_2. Here, the summation is over the set of candidate points for the design, which consists of extreme vertices, the midpoint of edges, constraint plane centroids, the axial check blends, the overall centroid, as well as a lattice grid where each proportion is a multiple of 0.05.

The method of the scaling is as follows. Let $X = [X_{pri}|X_{pot}]$ be the set of candidate points in the model space, partitioned by the two sets of terms. The potential terms are first regressed on the primary terms, and $\alpha = (X_{pri}'X_{pri})^{-1}X_{pri}'X_{pot}$ is the alias matrix measuring the confounding of the primary and potential terms. Define $R = X_{pot} - X_{pri}\ \alpha$ and Z=R/(max{R}-min{R}), where the max{R} and min{R} are taken separately for each column of R. The definition of X then becomes (X_{pri} | Z) or (X_{pri} | R) instead of (X_{pri} | X_{pot}). Note that α is the least squares regression coefficient of X_{pot} on X_{pri} and R contains the residuals from this regression. The purpose of this scaling is to make the primary and potential terms nearly unrelated.

The assumptions for the model are as follows. The errors are assumed to be independently and identically distributed (iid) as N (0, σ^2). The error variance σ^2 is a known constant. Since primary terms are usually very important factors to the practitioners, these terms are assumed to have a normal distribution with an arbitrary prior mean and prior variance, tending to infinite. On the other hand, the potential terms are unknown to the experimenter, and are assumed to have a normal distribution with a zero prior mean and a finite variance N (0, $\tau^2\sigma^2 I$). A large value of τ^2 means that some

of the potential terms are very likely for the right model. A small value of τ^2 means that all of the potential terms are unlikely to appear in the right model. The value of τ^2 is the ratio of variance of potential terms and experimental errors. The parameter τ^2 will determine the choice of design, since it reflects the degree of uncertainty associated with potential terms, relative to σ^2.

Given the above prior distributions, the posterior distributions for the model parameters are normal (DuMouchel and Jones (1994), and Pilz (1991)),

$$p(\beta \mid y, \sigma^2) \sim N[(X'X + \frac{K}{\tau^2})^{-1} X'y, \sigma^2 (X'X + \frac{K}{\tau^2})^{-1}].$$

Here, K is a (p+r) × (p+r) diagonal matrix, with the p diagonal elements associated with the primary terms equal to 0 and the remaining r diagonal elements equal to 1 (p and r are the numbers of primary and potential terms, respectively). The Bayesian information matrix is $[(1/n)(X'X + K/\tau^2)]$, where n is the total number of observations in Bayesian D-optimal design.

DuMouchel and Jones' (1994) D-optimal design was to use a Bayesian D-optimal approach, which is formed by a subset of candidate points by numerically maximizing the determinant of $(X`X+K/\tau^2)$. Notice that when τ is close to zero, the D-optimal design for the model of primary terms is obtained, and when τ is large and $n \geq p+r$, the D-optimal design for the full model is generated.

2.4 Two-Stage Designs

A design in which the experiment is run sequentially in two different stages is called a two-stage design. The first stage is primarily used to obtain "good" parameter estimates in order to choose efficiently in the second stage design. Usually, D-optimality is a

reasonable choice for the first stage. The second stage design criterion can be chosen by the experimenter according to a special goal. The design type is typically named by both criteria. For example, if D-optimality is the criterion used in both stages, the resulting design is called two-stage D-D optimal design.

Two-stage optimal designs have been investigated for binary data by Abdelbasit and Plackett (1983), Minkin (1987), Letsinger and Myers (1995), and Myers et al. (1996). Also, two-stage designs have been investigated by Neff (1996) for a normal linear model.

In this section, non-Bayesian D-D optimal design as well as Bayesian D-D optimal design are presented for a linear model.

2.4.1 Non-Bayesian D-D Optimal Design

Suppose that $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, and let $p(\mathbf{y}|\beta)$ be the probability distribution of the response. The likelihood function for the first stage is

$$L_1(y_1, \beta | x_1) = \prod_{i=1}^{n_1} p(y_i | \beta),$$

and the Fisher information is

$$I_1(\beta) = - E\left[\frac{\partial}{\partial \beta'}\left[\frac{\partial \ln L_1(\beta, y_1)}{\partial \beta}\right]\right].$$

Under the assumption that $\mathbf{y_1}$ has a normal distribution, the Fisher information for the first stage design is

$$I_1(\beta) = \frac{1}{\sigma^2} X_1' X_1,$$

where X_1 is the $n_1 \times k$ model matrix, k is the number of parameters in the assumed fitting model, and n_1 is the number of runs for the first stage. The D-optimal design for the first stage is selected by maximizing the determinant of ($X_1' X_1$) in the model space.

Let n_2 be the sample size for the second stage. The likelihood for the parameter with two-stage responses is

$$L_{1,2}(y_1,y_2, \beta |X_1,X_2) = L_1(y_1, \beta |X_1)L_{2|1}(y_2, \beta |y_1,X_1,X_2), \quad (2.1)$$

where $L_{2|1}(y_2, \beta |y_1,X_1,X_2) = \prod_{i=1}^{n_2} p(y_i \mid \beta)$.

Taking the log of both sides of (2.1), we have

$$\log L_{1,2} = \log L_1 + \log L_{2|1},$$

and the Fisher information

$$I_{1,2}(\beta) = -E\left[\frac{\partial}{\partial \beta'}\left[\frac{\partial \ln L_{1,2}(\beta, y)}{\partial \beta}\right]\right].$$

The combined Fisher information is

$$I_{1,2}(\beta) = I_1(\beta) + I_{2|1}(\beta) = \frac{1}{\sigma^2}[X_1' X_1 + X_2' X_2].$$

It is clear that the second stage design is chosen by maximizing the determinant of ($X_1' X_1 + X_2' X_2$) in the model space. Notice that the non-Bayesian D-D optimal design is equivalent to the one-stage non-Bayesian D-optimal design for normal linear model.

2.4.2 Bayesian Two-Stage D-D design

Letsinger (1995) developed a two-stage procedure for designing D-optimal designs for the logistic regression model. Bayesian D-optimality was used to select the first stage design but non-Bayesian D-optimality was applied to the second stage. Neff (1996)

developed numerous Bayesian two-stage design procedures for normal linear models. The following describes one of the two-stage Bayesian D-D procedures for the normal linear model in which one experiences model uncertainty.

The Bayesian D-optimality criterion which maximizes $|X_1'X_1+K/\tau^2|$ is used to select the first stage design, according to the method which is presented in Section 2.3. The parameter $\tau = 5$ is recommended in both the first and the second stage in Neff 's thesis (1996) because of the ability to produce designs which is most robust to model misspecification.

Box and Meyer (1993) developed a method of calculating the weights using posterior probabilities of the candidate models given the first stage data. Label these models in order as M_0, M_1, ..., M_m, where M_0 denotes the model containing only the primary terms and model M_i contains the parameters $\beta_i = \begin{bmatrix} \beta_{pri} \\ \beta_{pot(i)} \end{bmatrix}$, where β_{pri} contains the p primary parameters and $\beta_{pot(i)}$ contains a subset of q_i of the q potential parameters. The density of the predictive $\mathbf{y_1}$, given model M_i, $p(y_1| M_i)$, is written as

$$p(y_1 \mid M_i) = \int_{B_i} p(y_1 \mid M_i, \beta_i) p(\beta_i \mid M_i) d\beta_i . \qquad (2.2)$$

Here, B_i is the set containing all possible values β_i, $p(y_1| M_i, \beta_i)$ is the sampling distribution of y_1 given the model M_i, and $p(\beta_i|M_i)$ is the prior probability density of β_i. Given the first stage data, y_1, the posterior probability of the model M_i becomes

$$p(M_i \mid y_1) = \frac{p(M_i) p(y_1 \mid M_i)}{\sum_{j=0}^{m} p(M_j) p(y_1 \mid M_j)},$$

where $p(M_i)$ is the prior probability of M_i and $p(M_i|y_1)$ quantifies the likelihood that M_i is the model that best describes the relationship between the response and the regressors, upon observation of the first stage data.

Suppose that the prior probability of each potential term to be active in the true model is π. The prior probability of M_i, containing q_i active potential terms, can be expressed by

$$p(M_i) = \pi^{q_i}(1-\pi)^{q-q_i}, \quad I = 0, 1, \ldots, m.$$

The value of π should be chosen to represent the proportion of potential terms believed to be active or important relative to the primary terms. Since the experimenter may expect that only a few potential terms are important, an appropriate value for π would be in the range of 0 to 0.5. Box and Meyer (1993) found that sensible results can be achieved with a value of π=0.25. The model posterior probabilities would be conservatively evaluated by using values of π such as 0.2, 0.3,and 0.5. In Neff's (1996) research, a value of π =0.33 was used all the time.

Since the error term is assumed to be normal, the density function y_1 given M_i can be expressed as

$$p(y_1 | M_i, \beta_i) \propto \sigma^{-n_1} \exp[-(y_1 - X_i\beta_i)'(y_1 - X_i\beta_i)/2\sigma^2].$$

We assume that $\pi(\beta_{pri}) \propto 1$ and $\beta_{pot} \sim N(0, \tau^2\sigma^2 I)$, where 0 is a vector of $(p+q_i)$. The form (2.2) can be written as

$$p(y_1 | M_i) \propto \tau^{-q_i} |T_i + X_i'X_i|^{-1/2} (S(\hat{\beta}_i) + \hat{\beta}_i'T_i\hat{\beta}_i)^{-(n_1-1)/2},$$

where X_i is the first stage design in model M_i space, $\hat{\beta}_i = (T_i + X'X_i)^{-1}X_i'y_1$ for the model M_i, $S(\hat{\beta}_i)=(y_1-X_i\hat{\beta}_i)'(y_1-X_i\hat{\beta}_i)$ for the model M_i, and $T_i=\frac{1}{\tau^2}\begin{bmatrix}0_{pxp} & 0_{pxq_i}\\ 0_{q_ixp} & I_{q_ixq_i}\end{bmatrix}$.

The resulting posterior probability for model M_i becomes

$$p(M_i \mid y_1) = c\pi^{q_i}(1-\pi)^{q-q_i}\tau^{-q_i}\mid T_i + X`_iX_i\mid^{-1/2}(S(\hat{\beta}_i)+\hat{\beta}_iT_i\hat{\beta}_i)^{-(n_1-1)/2},$$

where c is a constant that forces all probabilities to sum to one.

Recall that in the first stage it is assumed that $\pi(\beta_{pri}) \propto 1$, $\beta_{pot} \sim N(0,\tau^2\sigma^2 I)$, and $y_1 \mid \beta,\sigma^2 \sim N(X_1\beta,\sigma^2 I)$, and the resulting posterior distribution of β for the first stage is normal,

$$\beta \mid y_1,\sigma^2 \sim N(b_1, V_1),$$

where

$$b_1=(X'_1X_1 + K/\tau^2)^{-1}X'_1y_1 \text{ and } V_1=\sigma^2(X'_1X_1 + K/\tau^2)^{-1}.$$

Let the second stage prior distribution of β be the first stage posterior. It can then be shown that the second stage posterior distribution of β is also normal,

$$\beta \mid y_1,y_2,\sigma^2 \sim N(b_2, V_2),$$

where

$$b_2=(X'_1X_1 + X'_2X_2 + K/\tau^2)^{-1}X'y \text{ and } V_2=\sigma^2(X'_1X_1 + X'_2X_2 + K/\tau^2)^{-1}.$$

Note that for all p+q parameters of the full model, a D-optimal design X_2 can be chosen to minimize $|n(X'_1X_1 + X'_2X_2 + K/\tau^2)^{-1}|$ over all possible designs. Consider any of the subset models M_i (i=0, 1, 2, …, m) as defined previously, with each model M_i identified by its parameters β_i. The posterior variance of β_i is

$$V_{2(i)} = \sigma^2 (X_{1(i)}' X_{1(i)} + X_{2(i)}' X_{2(i)} + K_{(i)} / \tau^2)^{-1},$$

where $X_{1(i)}$ and $X_{2(i)}$ are the first and second stage design matrices, respectively, expanded to model M_i space (containing only the $p+q_i$ regressors in M_i), and

$$K^{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & I_i \end{bmatrix}$$

is a $(p+q_i) \times (p+q_i)$ matrix. The second stage design in a Bayesian D-D optimal design for model M_i is the set of design points X_2 that minimizes $D_i = |V_{2(i)}|$ over all the possible designs.

Since the posterior probabilities above reflect the model importance, Neff (1996) utilized the Box and Meyer posterior probabilities for models $M_0, M_1, \ldots, M_m$ as weights to average the D criterion when the second stage design is selected. This is done by choosing the second stage design so as to minimize

$$\underset{X_2 \in X}{Min} \sum_{M_i} D_i \, p(M_i \mid y_1).$$

Chapter 3

Bayesian One-Stage D-Optimal Design for Mixture Models

DuMouchel and Jones (1994) and Andere-Rendon, Montgomery and Rollier (1997) use Bayesian modification to reduce the dependency of D-optimal designs on the assumed mixture experimental models. Compared to standard D-optimal designs, DuMouchel and Jones' approach is superior in terms of smaller prediction errors and more power to detect lack of fit. In terms of smaller bias errors and improved coverage of the factor space, Andere-Rendon et al.'s approach is superior. it also allows the fitting of a large number of higher order terms and maintains good variance properties. This chapter compares the designs of Andere-Rendon et al. (1997) with our Bayesian D-optimal designs. Also, several Bayesian D-optimal designs are analyzed and the choice of τ is discussed.

3.1 The Selection of Design Optimality

Cornell (1986) mentioned four properties that are particularly important in a mixture experiment. The design should:

1. have good coverage of the factor space throughout the region of interest,

2. ensure that the predicted $\hat{y}(x)$ is as close as possible to the true value $\eta(x)$,

3. provide good detectability of lack of fit, and

4. provide an internal estimate of error.

Bayesian D-optimal designs are proved to satisfy the above properties well by the studies of DuMouchel and Jones (1994) in regular linear models and Andere-Rendon et al. (1997) for mixture models. In a linear model situation, a Bayesian D-optimal design is chosen by maximizing the determinant of $(X'X + K/\tau^2)$. Here, K is a (p+r) $\times$ (p+r) diagonal matrix with the p diagonal elements associated with the primary terms equal to 0, the remaining r diagonal elements equal to 1 (p and r are the numbers of primary and potential terms, respectively), and X is the model matrix (see section 2.3). Notice that when τ is close to zero, the D-optimal design of the primary terms alone is obtained, and when τ is large and $n \geq$ p+r, the D-optimal design of the full model is generated

3.2 Existing Bayesian D-optimal Designs.

Andere-Rendon et al. (1997) developed the designs for mixture experiments by using Bayesian D-optimality. Their steps to maximize the determinant of $(X'X + K/\tau^2))$ were as follows:

1. The canonical model is transformed to a slack model.

2. When dealing with the Bayesian D-optimal design generation problem, the D-optimal design augmentation is used. The design to augment is represented by the matrix

$$W = \frac{1}{\tau}[0, I],$$

where 0 is an $(r \times p)$ matrix of zeros and I is the identity matrix of order r. The augmented design is

$$Z = \begin{bmatrix} X \\ W \end{bmatrix},$$

where X is the (n × (p+r)) matrix whose rows are the n augmented points. The Bayesian D-optimal design D^* is computed as the design that numerically maximizes the determinant of $Z'Z$ with respect to X.

3. The Bayesian D-optimal designs are generated by using the procedure of OPTEX from the SAS (1995) system.

3.3 Modified Bayesian D-optimal Designs.

Our steps, which use Fedorov's Algorithm to maximize the determinant of ($X'X + K/\tau^2$), are as follows:

1. The Fedorov algorithm starts with an n-points nonsingular design, $\zeta_1[n]$.

2. Fedorov showed that

$$\frac{|M(\zeta_{i+1}[n])|}{|M(\zeta_i[n])|} = 1 + \Delta_i(x_j, x).$$

During the i^{th} iteration a point, say x_j , is deleted and another point $x \in X$ is added if $\Delta_i(x_j, x)$ is maximal. Here, $|M(\zeta_i[N])|$ is the determinant of $X'X + K/\tau^2$ in i^{th} iteration.

3. Step (2) is repeated, and the procedure terminates whenever

$$\frac{(|M(\zeta_{i+1}[N])| - |M(\zeta_i[N])|)}{|M(\zeta_i[N])|} \le 10^{-5}.$$

We also can use the procedure OPTEX to generate Bayesian D-optimal designs (see page 695 in SAS/QC Software).

3.4 Example of Bayesian D-optimal Designs

In this example, it is assumed that the experimenter defines the following regressors to consider:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

The primary terms are $\{x_1, x_2, x_3\}$ with p=3 and $\{x_1x_2, x_1x_3, x_2x_3\}$ are potential terms with r=3. The sample size of eight is used. Bayesian D-optimal designs for different τ are developed and are presented in Figure 3.1 along with the designs of Andere-Rendon (1997) in Figure 3.2. The distribution of design points for different τ in Figure 3.1 is following. When τ is between 0 and 0.499, there are three vertices x_1 for estimating the coefficient β_1, three vertices x_2 for estimating the coefficient β_2, and two vertices x_3 for estimating the coefficient β_3. When τ is between 0.5 and 0.51, there are two vertices x_1 for estimating the coefficient β_1, three vertices x_2 for estimating the coefficient β_2, and two vertices x_3 for estimating the coefficient β_3, one binary blend for estimating the coefficient β_{13}. When τ is between 0.52 and 0.71, there are two vertices for each of three variables x_1 x_2, x_3 for estimating β_1, β_2, and β_3, and two binary blends for estimating β_{12} and β_{13}. When τ is between 0.72 and 1.05, there are two vertices x_1 for estimating β_1, one vertice x_2 for estimating β_2, two vertices x_3 for estimating β_3, and three binary blends for estimating β_{12}, β_{13}, and β_{23}. When τ is between 1.06 and 1.11, there are two vertices x_1 for estimating β_1, one vertice x_2 for estimating β_2, two vertices of x_3 for estimating β_3, three binary blends (where two of the three binary blends are equal proportion of the mixture) for estimating β_{12}, β_{13}, and β_{23}. When τ is

between 1.12 and ∞, there are two vertices x_1 for estimating β_1, one vertice x_2 for estimating β_2, two vertices x_3 for estimating β_3, and three binary blends (all equal proportion of the mixture) for estimating β_{12}, β_{13}, and β_{23}. Notice that as τ increases, the design reflects a stronger belief in higher order potential terms. Here, variables x_1, x_2 and x_3 are interchangeable.

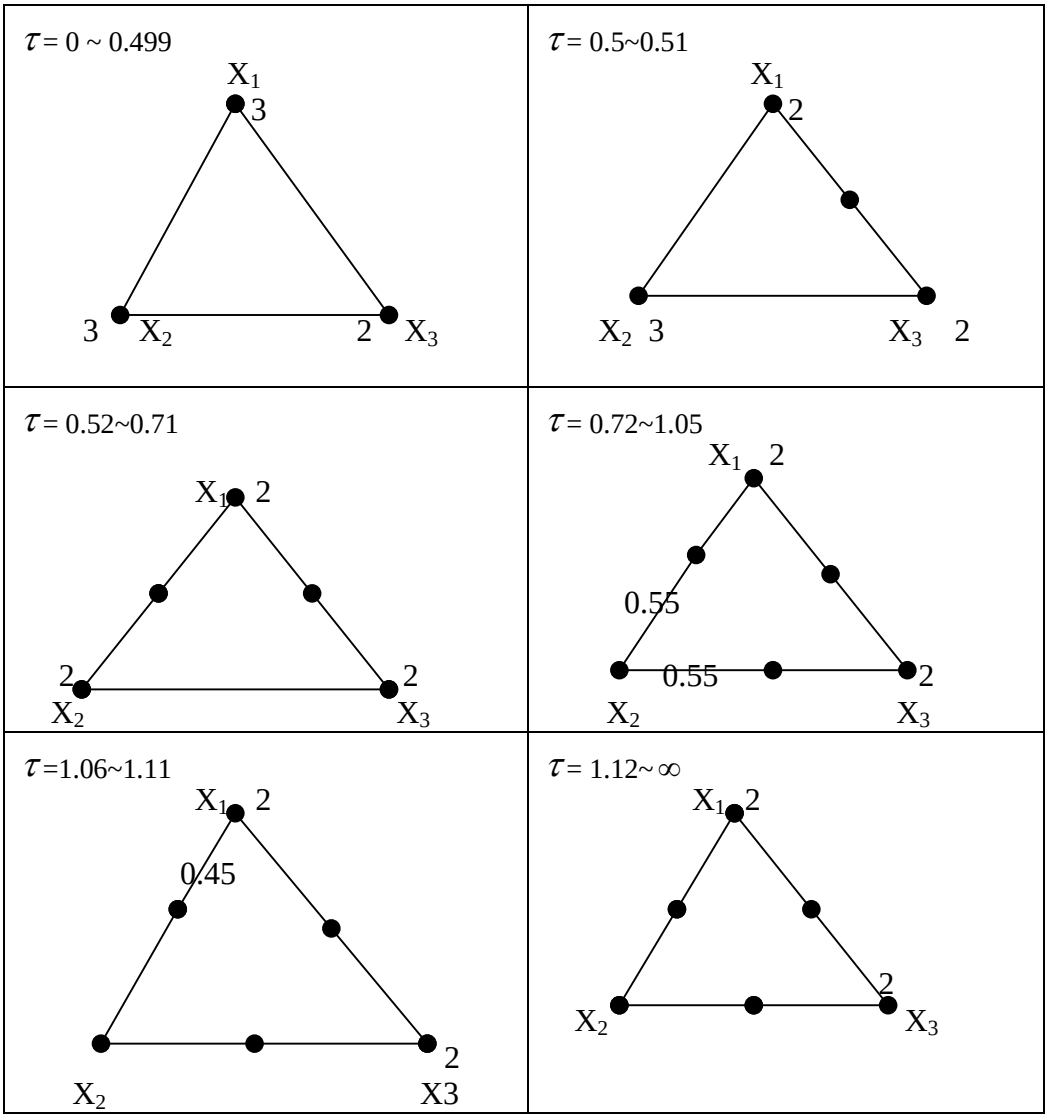


Figure 3.1. The sequence of designs with n=8 for a linear primary and a quadratic potential model in an unconstrained region (our designs).

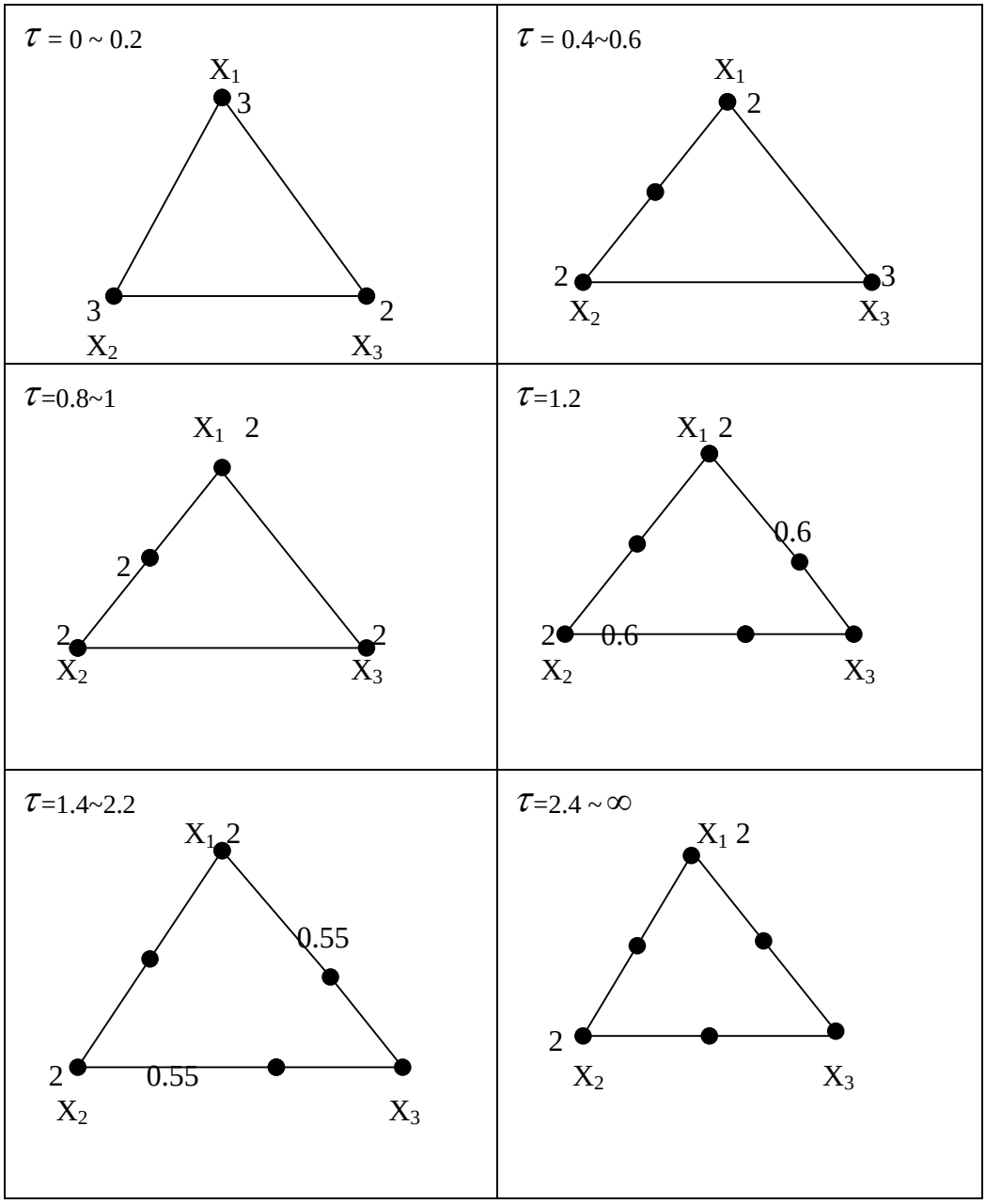


Figure 3.2. The sequence of designs with n=8 for a linear primary and a quadratic potential model in an unconstrained region (Andere-Rendon et al.'s designs).

3.5 Comparisons and Results

Since the Bayesian D-optimal design is generated by maximizing D = $|(X^{'}X + K/\tau^2)|$, a larger value of D means the design is more efficient to estimate parameters in the assumed fitting model. Therefore, the larger D= $|(X^{'}X + K/\tau^2)|$ is preferred. The determinant of $(X^{'}X + \frac{K}{\tau^2})$ for different values of τ is presented in Table 3.1, the best design in each column is marked with an asterisk.

Table 3.1. The determinant of D= $(X^{'}X + K/\tau^2)$

τ	0.1	0.4	0.6	1	1.2	2	5
Andere-Rendon et al.'s designs	18×10^6 *	3913.17	371.95	28	25.39	9.1	4.7*
Our designs	18×10^6 *	4394.53*	373*	50.11*	28.22*	9.45*	4.7*

It is clear to see that our Bayesian D-optimal designs are always at least as efficient as those of Andere-Rendon, et al. and are better in most cases. Especially when τ is one, our design is almost twice better than their design. By looking at the distribution of design points, we can also conclude that our designs are more desirable. For instance, when τ is one, there are two same binary blends in their design. These points are equal proportions of the mixture for estimating the second-order term β_{12}. Apparently there are no points to estimate the second-order terms β_{13}, β_{23}. This design should not be very efficient because it emphasizes the second-order term β_{12} and ignores the other second-order terms. However, in our design, there are three binary blends, one of them is an equal proportion point and the other two are not equal proportion points, to estimate the

second-order terms β_{12}, β_{13} and β_{23}. This makes perfect sense because there is no preliminary information for the potential terms.

In DuMouchel and Jones' paper, they recommended a scaling convention to eliminate the correlation between the primary and potential terms as much as possible. In Andere-Rendon, et al.'s procedure, transformation from canonical model to a slack model is necessary for the scaling convention. However, we get the same Bayesian D-optimal designs by using mixture canonical models and slack variable models. Therefore, we do not need to transform the canonical model to the slack model when using the scaling convention recommended by DuMouchel and Jones.

Figure 3.3 presents the sequence of designs for a linear primary and a second-order potential model with n=12 for four components. Figure 3.4 presents the sequence of the designs for a model that consists of a linear primary and both second order and special cubic potential terms with n=9 for the three components. In this case, we treat the linear terms as primary terms and both second-order and special cubic terms as potential terms. Figure 3.5 presents the sequence of designs for a second-order primary and a special cubic potential model with n=16 for the four components. Notice that all the notations $(X_1, X_2, X_3 \cdots)$ for vertices are interchangeable.

From Chapter 2, the coefficients of the potential terms follow a normal distribution with $\tau^2\sigma$ finite prior variance. The value of τ^2 will determine the choice of design, since the parameter τ^2 reflects the importance of the potential terms, relative to σ^2. Therefore, it is important for the experimenter to have some ideas about how to pick the value of τ. From Figures 3.2 to 3.5, we can say that τ from 0 to 0.3 is a small value since the Bayesian D-optimal design is the D-optimal design for the model of primary

terms. Also τ from 2 to ∞ is a large value since the Bayesian D-optimal design is the D-optimal design for the full model. Notice that there are vertices for estimating all the coefficients of primary terms and points for estimating part of the coefficients of potential terms when $\tau = 1$ in most Bayesian D-optimal designs.

Furthermore, the designs in Figures 3.2 to 3.5 are all quite desirable. For example, in Figure 3.3, when the value of τ is between 0~0.5 which means that the potential terms are almost not present, the design points are all in the vertices to estimate the primary terms (linear blending). When the value of τ is among 0.51~0.66, there are eight vertices to estimate the primary terms (linear blending) and four equal proportion binary blends to estimate the potential terms (second order terms). When τ is 0.67, there are seven vertices points to estimate the primary terms (linear blending) and five binary blends, two of which are equal proportion points and three of which are not equal proportion points, to estimate the potential terms (second order terms). When τ is between 0.68~1, six vertices points are used to estimate the primary terms (linear blending effects) and six binary blends, two of which are equal proportion points and four of which are not equal proportion points, to estimate the potential terms (second order terms). Finally, when τ is between $1.1 \sim \infty$, there are six vertices to estimate the primary terms (linear blending) and six equal proportion binary blends to estimate the potential terms (second-order terms). Since an increase in τ necessitates the use of more points to estimate the potential terms, the designs shown here look very desirable. Similar discussions can be applied to other designs in the other Figures.

Usually, the parameter τ, which reflects the degree of uncertainty of potential terms, is selected by the experimenter based on the prior information. If the experimenter

believes that the potential terms are unlikely to appear in the true model, then a small value of τ should be selected. If the experimenter believes that some of the potential terms are possibly more likely to be in the true model, then a large value of τ should be selected. If the experimenter has no idea about the importance of potential terms, then a moderate value of τ should be selected. The robust value of τ will be discussed in Chapter 5.

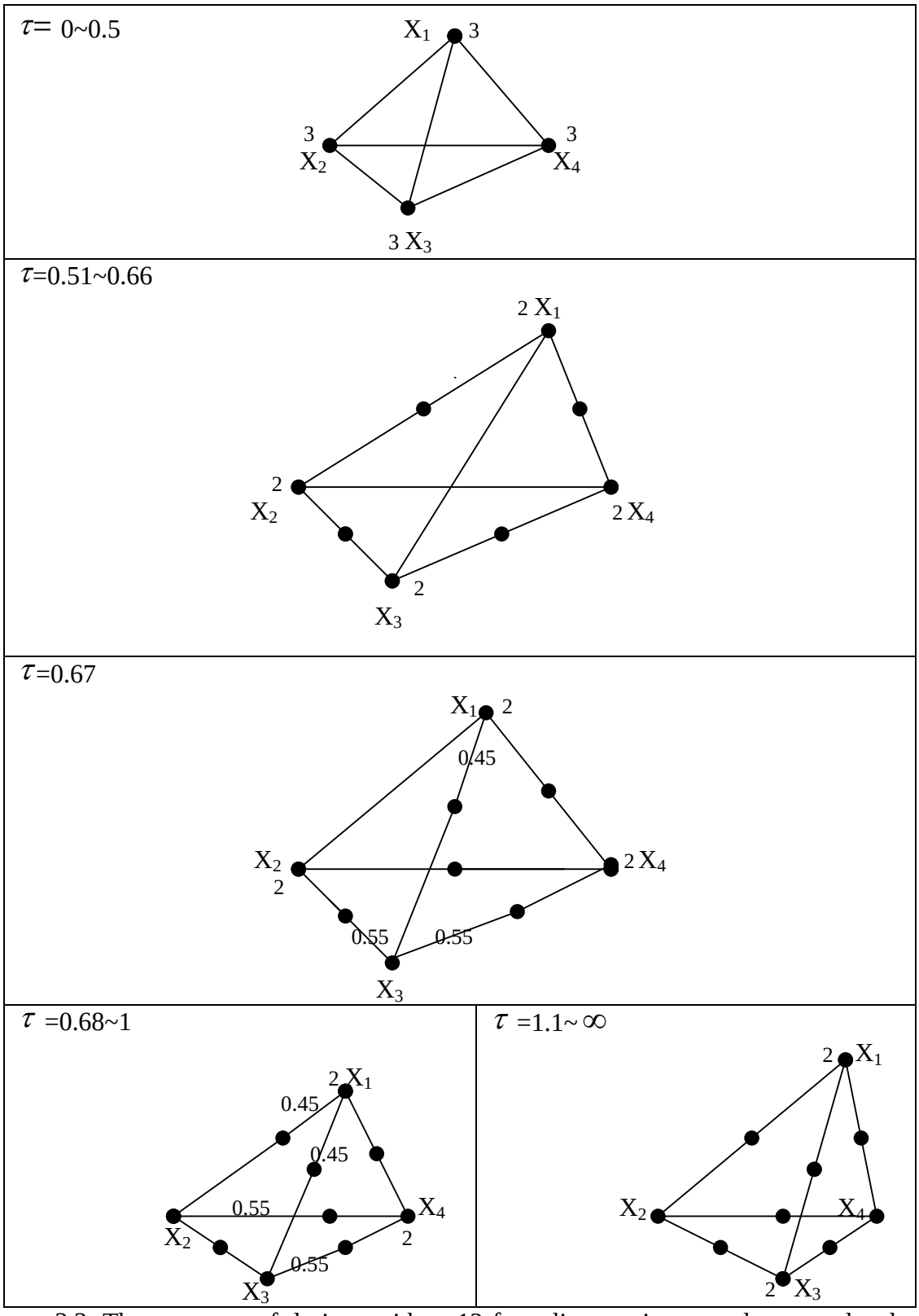


Figure 3.3. The sequence of designs with n=12 for a linear primary and a second-order potential model in an unconstrained region

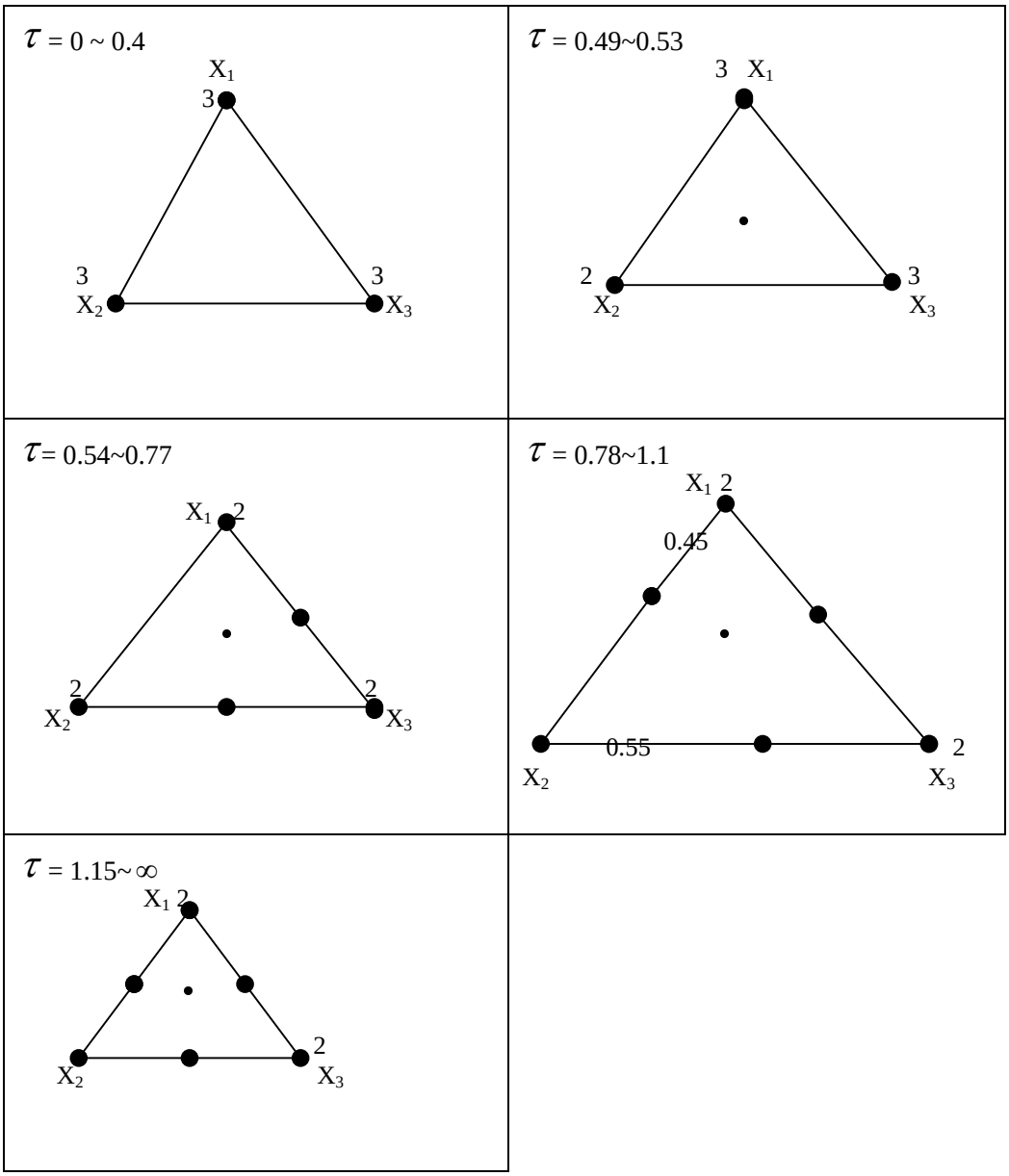


Figure 3.4. The sequence of designs with n=9 for a linear primary and both a second-order and a special cubic potential model in an unconstrained region

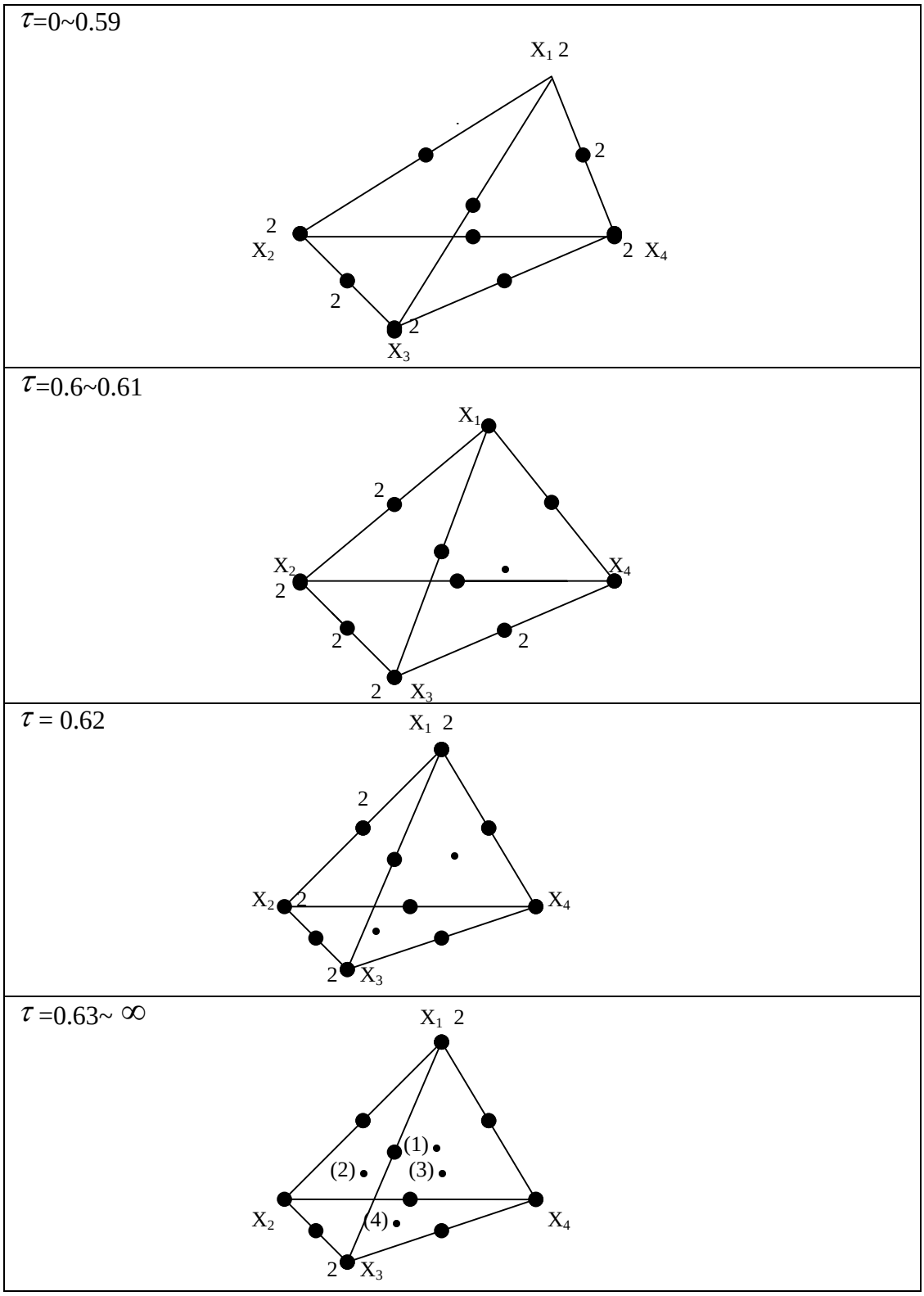


Figure 3.5. The sequence of designs with n=16 for a second-order primary and a special cubic potential model in an unconstrained region (1, 2, 3, 4 are the centers of $x_1x_2x_4$, $x_1x_2x_3$, $x_1x_3x_4$, $x_2x_3x_4$, respectively).

Chapter 4
Bayesian Two-Stage D-D
Optimal Design for Mixture Models

Bayesian two-stage D-D optimal design for mixture experiments under model uncertainty is developed in this chapter. A Bayesian D-optimality criterion is used in the first stage to minimize the determinant of the posterior variances of the parameters. The second stage design is then generated according to a D-optimality procedure that combines with the improved model from the first stage data. Examples of Bayesian two-stage designs are presented for different situations.

4.1 Selection and Analysis for the First Stage D-optimal Design

Consider the linear model $\mathbf{y} = X\beta + \varepsilon$, where $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ is the n dimensional vector of responses from a two-stage design, X is the $(n_1+n_2)\times(p+r)$ model matrix for the combined stage and $\beta = \begin{bmatrix} \beta_{pri} \\ \beta_{pot} \end{bmatrix}$ is the p+r vector of the parameters. It is assumed that

$y_i|\beta,\sigma^2 \sim N(X_i\beta,\sigma^2 I)$ for each stage i, with n_1 runs for the first stage and n_2 runs for the second stage.

The Bayesian D-optimality criterion that maximizes $|X_1'X_1+K/\tau^2|$ is used to select the first stage design, according to the method presented in section 2.3. Here, K is a (p+r) × (p+r) matrix, with p diagonal elements equal to 0 and the remaining r diagonal elements equal to 1 (p and r are the numbers of primary and potential terms, respectively). If the experimenter has no idea about the importance of the potential terms, then the value of τ is recommended to be one for the first stage, which is discussed in chapter 3 and 5. In this chapter, the parameter τ =1 is used for all of the examples.

Prior to observing the first stage data, a set of p+r regressors is defined by the experimenter for the full model, which consists of p primary terms and r potential terms. The model that best describes the relationship between the response and the regressors could be possibly one of the following three types: (1) primary terms only, (2) full model, (3) p primary terms and a subset of potential terms. The total number of possible models can be calculated by m= $\sum_{i=0}^{r}\binom{r}{i} = 2^r$. Each candidate model M_i contains all primary terms and r_i $(0 \le r_i \le r)$ potential terms.

The information from the analysis of the first stage data is used to reduce model uncertainty in the second stage. Model knowledge can be gained by scoring each of the candidate models according to how likely a particular model will predict the response behavior. The resulting scores can be used as weights to a second stage criterion, ensuring that the second stage data points could provide information on important terms.

The scores are obtained by calculating the posterior probabilities of these possible models.

The probability density of y_1 given M_i is expressed by

$$p(y_1|M_i, \sigma, \beta_i) \propto \sigma^{-n_1} \exp[-(y_1 - X_i\beta_i)'(y_1 - X_i\beta_i)/2\sigma^2].$$

Assume that $p(\beta_{pri}) \propto 1$, $\beta_{pot(i)} \sim N(0, \sigma^2\tau^2 I_{(p+r_i)})$ where 0 is a vector of $(p+r_i)$, and $\sigma^2 \sim IG(\alpha, \delta)$ with a small value of α and a large value of δ. Notice that for an inverse Gamma distribution, $\mu(\sigma^2) = \frac{1}{(\alpha-1)\delta}$, and $\text{var}(\sigma^2) = \frac{1}{\delta^2(\alpha-1)^2(\alpha-2)}$. Therefore, the density of σ^2 has a very heavy tail given the assumed values of α and δ. Under these conditions, the density of y_1 given candidate model M_i, $p(y_1|M_i)$ can be written (in the sense of a limiting procedure; see Appendix A)

$$p(y_1|M_i) \propto \tau^{-r_i} |T_i + X_i'X_i|^{-1/2} (S(\hat{\beta}_i) + \hat{\beta}_i'T_i\hat{\beta}_i)^{-n_1/2},$$

where X_i is the first stage design in model M_i space, $\hat{\beta}_i = (T_i + X_i'X_i)^{-1}X_i'y_1$ for the model M_i, $S(\hat{\beta}_i) = (y_1 - X_i\hat{\beta}_i)'(y_1 - X_i\hat{\beta}_i)$ for the model M_i, and

$$T_i = \frac{1}{\tau^2}\begin{bmatrix} 0_{pxp} & 0_{pxr_i} \\ 0_{r_ixp} & I_{r_ixr_i} \end{bmatrix}.$$

Suppose that the prior probability of each potential term to be active in the true model is π. The prior probability of M_i, containing r_i active potential terms, can be expressed by

$$p(M_i) = \pi^{r_i}(1-\pi)^{r-r_i}, \text{ i=0,1,…,m.}$$

The value π =0.33 is used in our study following Neff's (1996) suggestion. The resulting posterior probability for model M_i becomes

$$p(M_i|y_1)=c\,\pi^{r_i}(1-\pi)^{r-r_i}\tau^{-r_i}\,|T_i+X`_iX_i|^{-1/2}\,(S(\hat{\beta}_i)+\hat{\beta}_iT_i\hat{\beta}_i)^{-n_1/2},$$

where c is the normalization constant. Note that a large value of $p(M_i|y_1)$ means that M_i is a plausible model and a small value of $p(M_i|y_1)$ indicates that the model M_i is not likely in the presence of the first stage data. Analytically, the derivation of the parameter τ is not easy so empirical method will be used. The procedure for selecting the value of τ after first stage is discussed in section 4.2.

4.2 Uncertainty in τ^2

In the first stage, the parameter τ^2 is unknown to the experimenter. Once the first stage data is observed, a convenient and reasonable approximation of a value of τ is obtained by maximizing the posterior density $p(\tau^2|\mathbf{y_1})$. It needs to be mentioned that the estimation of τ^2 is empirical, which means we use a point estimate instead of a formal Bayesian posterior estimate. The reason to do this is to reduce the complexity of optimal procedure.

Setting the prior density $p(\tau^2)$ to be locally uniform (again, in the sense of using a limiting procedure), the posterior density $p(\tau^2|\mathbf{y_1})$ is given approximately by

$$p(\tau^2|y_1)\propto p(\tau^2)p(y_1|\tau^2)\approx p(y_1|\tau^2)=\sum_0^m p(M_i)\,p(y_1|M_i,\tau^2)$$

$$=\sum_0^m\left(\frac{\pi}{(1-\pi)\tau}\right)^{r_i}\frac{|X_0'X_0|^{1/2}}{|T_i+X_i'X_i|^{1/2}}\left(\frac{(y_1-X_i\hat{\beta}_i)'(y_1-X_i\hat{\beta}_i)+\hat{\beta}'_iT_i\hat{\beta}_i}{(y_1-X_0'\hat{\beta}_0)'(y_1-X_0'\hat{\beta}_0)}\right)^{-n_1/2},$$

where X_0 is the primary terms matrix and $\hat{\beta}_0 = (T_0 + X'_0 X_0)^{-1} X'_0 \mathbf{y_1}$.

Since the posterior mode $\hat{\tau}$ is the most likely to occur in the posterior density of τ given $\mathbf{y_1}$, $\hat{\tau}$ is a more reasonable value than the initial τ used in the first stage. Therefore, the value $\hat{\tau}$ is used for all the calculations after the first stage. Now $p(M_i|\mathbf{y_1})$ becomes

$$p(M_i|\mathbf{y_1}) = c\,\pi^{r_i}(1-\pi)^{r-r_i}\hat{\tau}^{-r_i}\,|T_i + X_i' X_i|^{-1/2}\,(S(\hat{\beta}_i) + \hat{\beta}_i' T_i \hat{\beta}_i)^{-n_1/2},$$

where c is a constant that forces all probabilities to sum to one, and

$$T_i = \frac{1}{\hat{\tau}^2}\begin{bmatrix} 0_{p\times p} & 0_{p\times r_i} \\ 0_{r_i\times p} & I_{r_i\times r_i} \end{bmatrix}..$$

4.3 Selection of the Second Stage Bayesian D-optimal Designs

Recall that the first stage posterior distribution of β is normal (see section 2.3),

$$\beta\,|y_1, \sigma^2 \sim N(b_1, V_1),$$

where

$$b_1 = (X'_1 X_1 + K/\hat{\tau}^2)^{-1} X'_1 y_1 \quad \text{and} \quad V_1 = \sigma^2 (X'_1 X_1 + K/\hat{\tau}^2)^{-1}.$$

Let the second stage prior distribution of β be the first stage posterior. It can then be shown that the second stage posterior distribution of β is also normal,

$$\beta\,|\mathbf{y_1}, \mathbf{y_2}, \sigma^2 \sim N(b_2, V_2),$$

where

$$b_2=(X'_1X_1 + X'_2X_2 + K/\hat{\tau}^2)^{-1} X' y \text{ and } V_2=\sigma^2 (X'_1X_1 + X'_2X_2 +K/\hat{\tau}^2)^{-1}.$$

Note that for all p+r parameters of the full model, a D-optimal design, X_2 can be chosen by minimizing $|n(X'_1X_1 + X'_2X_2 +K/\hat{\tau}^2)^{-1}|$ over all the possible designs.

Consider any of the subset models M_i (i = 0, 1, 2 , …, m) as defined previously, with each model M_i identified by its parameters β_i. The posterior variance of β_i is

$$V_{2(i)}=\sigma^2 (X_{1(i)}' X_{1(i)} + X_{2(i)}' X_{2(i)} + K_{(i)} / \hat{\tau}^2)^{-1}$$

where $X_{1(i)}$ and $X_{2(i)}$ are the first and second stage design matrices for model M_i space, respectively, and

$$K_{(i)} = \begin{bmatrix} 0 & 0 \\ 0 & I_i \end{bmatrix}$$

is a $(p+r_i) \times (p+r_i)$ matrix. A Bayesian D-D optimal design for model M_i is the set of design points that minimizes $D_i = |V_{2(i)}|$ over all the possible designs.

Since the posterior probabilities by Box and Meyer (1993) in section 2.4.2 reflect the model importance, we utilize the posterior probabilities for models M_0, M_1,…, M_m as weights to average the D criterion when the second stage design is selected. The objective is to choose the second stage design points so as to minimize D_i for each model M_i enjoying a high probability of being the best model. This is done by choosing the second stage design so as to minimize the average of D criteria

$$\underset{X_2 \in X}{Min} \sum\nolimits_{M_i} D_i \, p(M_i \mid y_1).$$

4.4 Examples of Bayesian D-D Optimal Designs

Examples of Bayesian D-D optimal designs for three and four component models are presented in this section. Section 4.4.1 represents the Bayesian D-D optimal design for three components with the linear primary and second-order potential model. Section 4.4.2 represents the Bayesian D-D optimal design for four components with second-order primary and special cubic potential model. These examples show that if the true model contains only primary terms, then the value of $\hat{\tau}$ is small, while if the true model contains primary terms and any of the potential terms, then the value of $\hat{\tau}$ is large. They also show that the true model enjoys the highest posterior probability $p(M_i|\mathbf{y_1})$, and the parameters in the true model are well estimated in the second stage design.

4.4.1 A Linear Primary and Second Order Potential Model for Three components

In this example, it is assumed that the experimenter defines the following regressors to consider:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

The primary terms are $\{x_1, x_2, x_3\}$ with p=3 and $\{x_1x_2, x_1x_3, x_2x_3\}$ are potential terms with r=3. The true model in this example is defined to contain the linear terms,

$$x^t = \{x_1, x_2, x_3\}.$$

The first stage data set is simulated from the true model:

$$y=7.24x_1 +9.57x_2 +5.66x_3 + \varepsilon,$$

where $\varepsilon \sim N(0,1)$. Note that the values of the coefficients of parameters in the true model are randomly selected and a large value of the coefficient means that term is very important, then second stage will provide more runs to estimate these important terms. In this example, the values of coefficients of x_1, x_2, x_3 are large since primary terms are important. On the other hand, no second-order terms appears in the true model, which means that second-order is not significantly important in the true model. Eight runs are allocated to each stage for the experiment, and the parameter τ =1 is used in the first stage.

The Bayesian D-optimality criterion presented in section 2.3 is used to select the first stage design with eight runs that are based on the full model by

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

The resulting design is shown in Figure 4.1.

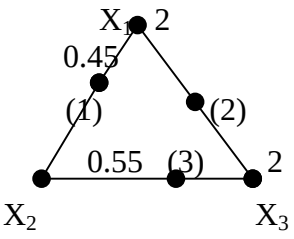


Figure 4.1. Bayesian first stage D-D optimal design for three components with a quadratic full model

This is a desirable design since there are five vertices for estimating the coefficients of linear blending $\beta_1, \beta_2, \beta_3$ and three binary blends 1, 2, and 3, where point 2 is equal proportion binary blend, and 1, 3 are not equal proportion binary blends, for estimating

the coefficients of second-order terms $\beta_{12}, \beta_{13}, \text{and } \beta_{23}$. According to section 4.2, the response data from the first stage experiment is used to get the value of $\hat{\tau}$, and to calculate the posterior probabilities associated with each of the candidate models. The following results are only for one simulation data set. Figure 4.2 represents the posterior density of τ^2, which shows that the value of $\hat{\tau}$ is close to zero. This makes good sense considering that the true data is from the linear blending model and that there are no potential terms in the true model. The first stage data and value of $\hat{\tau}$ are used to calculate posterior probabilities to weigh the importance of each candidate model. The posterior probabilities are included in Table 4.1. From these posterior probabilities, we can see that the true model enjoys the highest probability.

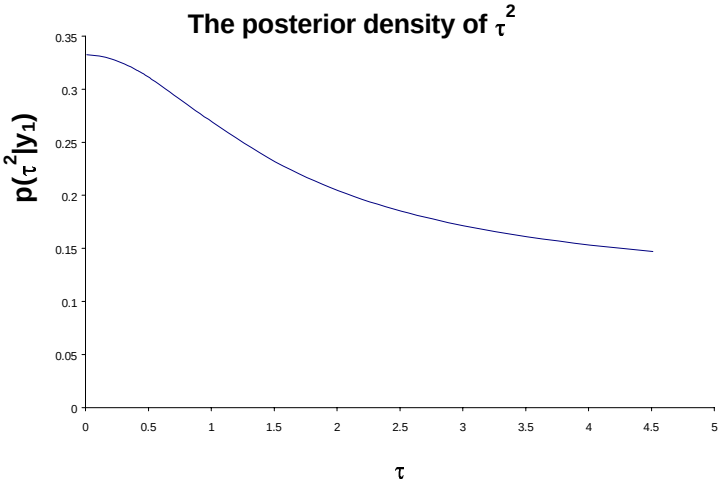


Figure 4.2. The posterior density of τ^2 given the true model contains only primary terms

Table 4.1. Posterior probabilities of candidate models for three components with linear primary and second-order potential terms model

Posterior Probability	Terms included in the candidate model
0.301*	x_1 x_2 x_3*
0.148	x_1 x_2 x_3 x_1x_2
0.148	x_1 x_2 x_3 x_1x_3
0.148	x_1 x_2 x_3 x_2x_3
0.073	x_1 x_2 x_3 x_1x_2 x_1x_3
0.073	x_1 x_2 x_3 x_1x_2 x_2x_3
0.073	x_1 x_2 x_3 x_1x_3 x_2x_3
0.036	x_1 x_2 $_3$ x_1x_2 x_1x_3 x_2x_3

For this simulation data set, eight runs of a second stage design are chosen. Figure 4.3 represents a second stage design points and shows that the design points are all in the vertices after the first stage design. This means that the second stage design is very efficient in selecting design points based on information from the first stage design, considering that the true model is a linear blending model. Combined first and second stage design is presented in Figure 4.4.

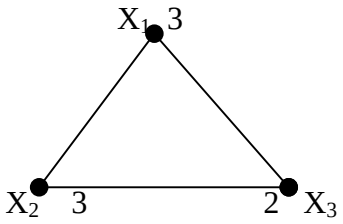


Figure 4.3. Second Stage D-D optimal design for three components given the true model contains only primary terms

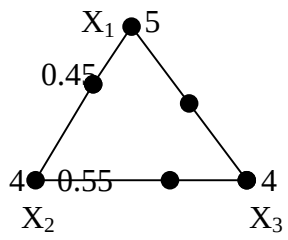


Figure 4.4. The combined two-stage Bayesian D-D design for three components given the true model contains only primary terms

4.4.2: A Second Order Primary and a Special Cubic Potential Model for Four Components

The experimenter may consider the following regressors:

$$x^{(f)} = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4\}.$$

The primary terms are $\{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}$ with p=10 and $\{x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4\}$ are potential terms with r=4. The true model in this example is defined to be a model containing the following regressors:

$$x^t = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1x_2x_3\}.$$

The first stage data set is simulated from the true model:

$$y = 17.24x_1+19.57x_2+16.66x_3+16.8x_4+17x_1x_2+18.51x_1x_3+17.64$$

$$+x_1x_4+18.21x_2x_3+15x_2x_4+17.6x_3x_4+15x_1x_2x_3+\varepsilon,$$

with $\varepsilon \sim N(0,1)$. Here, the coefficients of primary terms and the special cubic $x_1x_2x_3$ are large (again, the coefficients of the parameters are randomly picked), this means that special cubic term $x_1x_2x_3$ is significantly important as well as primary terms. Sixteen

runs are allocated to each stage of the experiment and the parameter τ =1 is used in the first stage design in this section. The Bayesian D-optimality criterion presented in section 2.3 is used to select the first stage design with sixteen runs that based on the full model defined by

$$x^{(f)} = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4\}.$$

The resulting design is shown in figure 4.5.

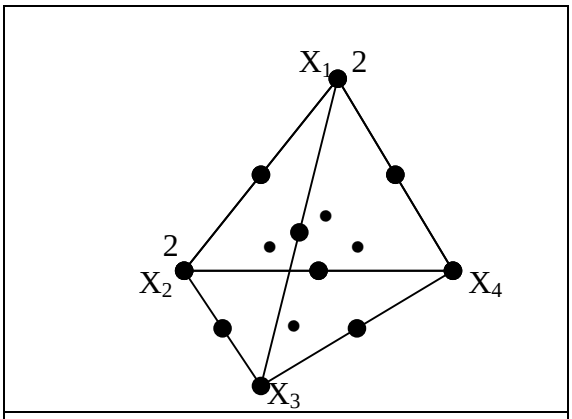


Figure 4.5. Bayesian first stage D-D optimal design for four components with a special cubic full model

In the first stage, six vertices estimate the coefficients of linear blending $\beta_1, \beta_2, \beta_3, \beta_4$; six equal proportion binary blends estimate coefficients of the second-order terms β_{12}, β_{13}, β_{14}, β_{23}, β_{24}, β_{34}; four centroid points estimate the coefficients of all three special cubic terms β_{123}, β_{124}, β_{134}, β_{234}. Therefore, in the first stage all the parameters in the full model can be estimated. After observing data from the first stage, we adjust the parameter τ. The following result is only for one simulation set. The posterior distribution $p(\tau^2/y_1)$ is shown in Figure 4.6 and shows the value of $\hat{\tau}$ is

close to 17. This makes good sense since there is one of the potential terms in the true model. The posterior probability of M_i given y_1 is presented in Table 4.2, where the true model has the highest posterior probability. The design points of the second stage are shown in Figure 4.7. Notice that there are two centroid points for estimating the coefficient of special cubic term $x_1x_2x_3$, which is contained in the true model. The combined design points are represented in Figure 4.8.

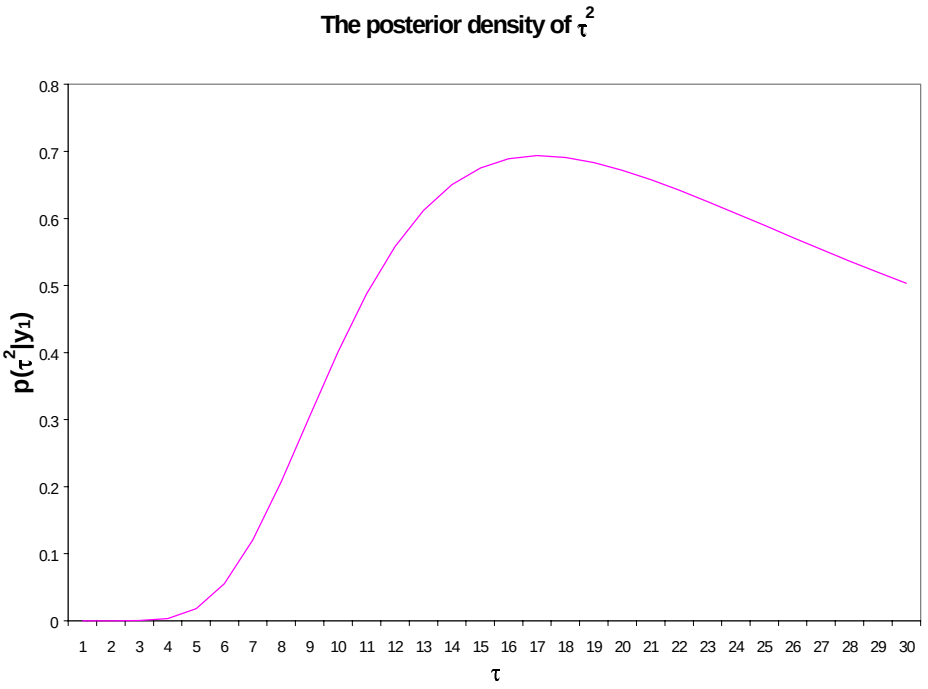


Figure 4.6. The posterior density of τ^2 given the true model contains one of potential terms

Table 4.2. Posterior probabilities of candidate models for four components with second-order primary and special cubic potential terms model

Primary terms is in the every model	$P(M_i\|y_1)$
Primary terms	0
$x_1x_2x_3$*	0.816*
$x_1x_2x_4$	0
$x_1x_3x_4$	0
$x_2x_3x_4$	0
$x_1x_2x_3$ $x_1x_2x_4$	0.053
$x_1x_2x_3$ $x_1x_3x_4$	0.052
$x_1x_2x_3$ $x_2x_3x_4$	0.066
$X_1x_2x_4$ $x_1x_3x_4$	0
$x_1x_2x_4$ $x_2x_3x_4$	0
$x_1x_3x_4$ $x_2x_3x_4$	0
$x_1x_2x_3$ $x_1x_2x_4$ $x_1x_3x_4$	0.003
$x_1x_2x_3$ $x_1x_2x_4$ $x_2x_3x_4$	0.004
$x_1x_2x_3$ $x_1x_3x_4$ $x_2x_3x_4$	0.004
$x_1x_2x_4$ $x_1x_3x_4$ $x_2x_3x_4$	0
$x_1x_2x_3$ $x_1x_2x_4$ $x_1x_3x_4$ $x_2x_3x_4$	0.002

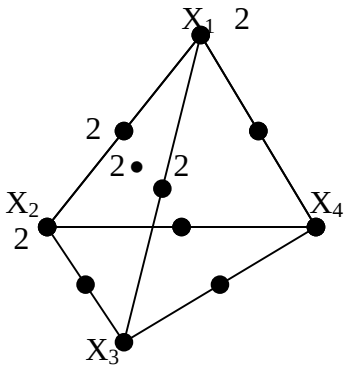


Figure 4.7. Bayesian second stage D-D optimal design for four components given the true model contains a cubic $x_1x_2x_3$ potential term

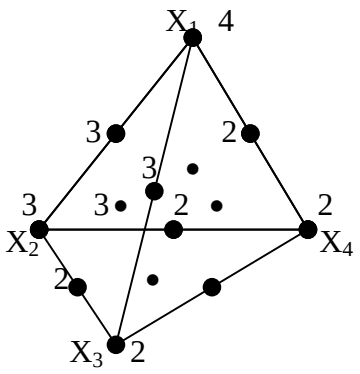


Figure 4.8. Combined stage of Bayesian D-D optimal design for four components given the true model contains a cubic $x_1x_2x_3$ potential term

4.5 Two Classes of Potential Terms

Sometimes we may also be interested in knowing if we have more than one class of potential terms. Therefore, it is natural to divide potential terms into two categories or even more. For example, the experimenter may know from past experience that linear blending terms must be included in the assumed model but is uncertain whether both of second-order and special cubic terms are needed for the correct model. Then we call linear terms the primary terms, second-order terms the first class of the potential terms, and special cubic terms the second class of the potential terms. Now the full model can be written as

$$y = X_{pri}\beta_{pri} + X_{pot(1)}\beta_{pot(1)} + X_{pot(2)}\beta_{pot(2)} + \varepsilon .$$

Again the prior variances of the coefficients of the primary terms are large, and variances of the coefficients of the two classes of potential terms are unknown to the experimenter. Let τ_1^2 be the parameter that reflects the importance of the first class of

potential terms, and let τ_2^2 be the parameter that reflects the importance of the second class of potential terms.

4.5.1 The Posterior Density of M_i Given y_1

Calculating posterior probabilities for candidate models using the Box and Meyer (1993) method, is as follows for this situation. The models can be labeled in order as M_0, $M_1,\ldots,Mm$, with M_0 denoting the model containing primary terms only. Each model M_i contains the parameters $\beta_i = \begin{pmatrix} \beta_{pri} \\ \beta_{pot1(i)} \\ \beta_{pot2(i)} \end{pmatrix}$, where β_{pri} is p primary parameters and $\beta_{pot1(i)}$ is a subset of $r_{1(i)}$ of the first class r_1 of potential parameters and $\beta_{pot2(i)}$ is a subset of $r_{2(i)}$ of the second class of r_2 potential parameters. Each candidate model M_i contains all primary terms and r_i $(0 \le r_i \le r_{1(i)}+r_{2(i)})$ two classes of potential terms.

Suppose that the prior probability of each potential term to appear in the true model is π. The prior probability of M_i, containing r_i active potential terms and $(r-r_i)$ inactive, can be expressed by

$$p(M_i)= \pi^{r_i}(1-\pi)^{r-r_i}, \quad i=0,1,\ldots,m.$$

Again the value π =0.33 is used. We assume the probability density of y_1 given M_i to be the usual normal linear model,

$$p(y_1 \mid M_i, \sigma, \beta_i) \propto \sigma^{-n_1} \exp[-(y_1 - X_i\beta_i)'(y_1 - X_i\beta_i)/2\sigma^2].$$

We assume $\pi(\beta_{pri}) \propto 1$, $\beta_{pot1} \sim N(0, \tau_1^2 \sigma^2 I)$, $\beta_{pot2} \sim N(0, \tau_2^2 \sigma^2 I)$ and $\sigma^2 \sim IG(\alpha, \delta)$ with a small value of α and a large value of δ. Under these assumptions, the posterior probability of model M_i given first stage data can be written as (see Appendix B),

$$p(M_i|y_1)=c\left(\frac{\pi}{1-\pi}\right)^{r_i}\left(\frac{1}{\tau_1}\right)^{r_{1(i)}}\left(\frac{1}{\tau_2}\right)^{r_{2(i)}} |K_i + X_i' X_i|^{-1/2} \left(\frac{S(\hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}}{S(\hat{\beta}_0)}\right)^{-n_1/2},$$

where

$$K_i=\begin{bmatrix} 0_{p \times p} & 0_{p \times r_{1(i)}} & 0_{p \times r_{2(i)}} \\ 0_{r_{1(i)} \times p} & \frac{1}{\tau_1^2} I_{(r_{1(i)})} & 0_{r_{1(i)} \times r_{2(i)}} \\ 0_{r_{2(i)} \times p} & 0_{r_{2(i)} \times r_{1(i)}} & \frac{1}{\tau_2^2} I_{(r_{2(i)})} \end{bmatrix}.$$

Here, r_i is the total potential terms in model M_i space such that $r_i = r_{1(i)}+r_{2(i)}$ where $r_{1(i)}$ is the total number of first class of potential terms in model M_i space such that $0 \le r_{1(i)} \le r_1$, and $r_{2(i)}$ is the total number of second class of potential terms in model M_i space such that $0 \le r_{2(i)} \le r_2$, $X_{1(i)}$= first stage design in model M_i space, $\hat{\beta}_i = (T_i + X_i' X_i)^{-1} X_i' y_1$ and $S(\hat{\beta}_i) = (y_1 - X_i \hat{\beta}_i)'(y_1 - X_i \hat{\beta}_i)$ for Model M_i, and c is a constant that forces all probabilities to sum to one. In the equation for $p(M_i|y_1)$, we do not know the values of τ_1 and τ_2. The procedure for selecting the values of τ_1 and τ_2 is discussed in section 4.5.2.

4.5.2 Uncertainty in τ_1^2 and τ_2^2

Recall in section 4.2 we discussed how to choose the value of $\hat{\tau}$. In this section, the idea of the estimations of τ^2_1 and τ_2^2 is the same as in estimation of τ^2, except this time we are discussing about the posterior distribution of τ^2_1 and τ_2^2, given the first stage data

y_1. A convenient and reasonable approximation values of $\hat{\tau}_1, \hat{\tau}_2$ are obtained by maximizing the posterior density $p(\tau^2_1, \tau_2{}^2 | y_1)$. Setting the prior density $p(\tau_1{}^2, \tau_2{}^2)$ to be locally uniform, the posterior density $p(\tau^2_1, \tau_2{}^2 | y_1)$ is given approximately by

$$P(\tau^2_1, \tau_2{}^2 / y_1) \propto P(\tau_1{}^2, \tau_2{}^2) P(y_1 / \tau_1{}^2, \tau^2_2) = P(y_1 / \tau_1{}^2, \tau_2{}^2)$$

$$= \sum_0^m p(M_i)\, p(y_1 \mid M_i, \tau_1{}^2, \tau_2{}^2)$$

$$= \sum_0^m \left(\frac{\pi}{1-\pi}\right)^{r_i} \left(\frac{1}{\tau_1}\right)^{r_{1(i)}} \left(\frac{1}{\tau_2}\right)^{r_{2(i)}} |K_i + X_i{}' X_i|^{-1/2} \left(\frac{S(\hat{\beta}_i) + \hat{\beta}_i{}' K_i \hat{\beta}}{S(\hat{\beta}_0)}\right)^{-n_1/2}.$$

After observing data from the first stage, we use the above method to get the most likely values of $\hat{\tau}_1$ and $\hat{\tau}_2$. The values of $\hat{\tau}_1$ and $\hat{\tau}_2$ are more reasonable values because they are updated from the first stage data. Therefore, the values of $\hat{\tau}_1$ and $\hat{\tau}_2$ are used in all the calculations after the first stage. Now $p(M_i|y_1)$ can be written as

$$p(M_i|y_1) = c \left(\frac{\pi}{1-\pi}\right)^{r_i} \left(\frac{1}{\hat{\tau}_1}\right)^{r_{1(i)}} \left(\frac{1}{\hat{\tau}_2}\right)^{r_{2(i)}} |K_i + X_i{}' X_i|^{-1/2} \left(\frac{S(\hat{\beta}_i) + \hat{\beta}_i{}' K_i \hat{\beta}}{S(\hat{\beta}_0)}\right)^{-n_1/2},$$

where

$$K_i = \begin{bmatrix} 0_{\,p \times p} & 0_{\,p \times r_{1(i)}} & 0_{\,p \times r_{2(i)}} \\ 0_{\,r_{1(i)} \times p} & \frac{1}{\hat{\tau}^2_1} I_{(r_{1(i)})} & 0_{\,r_{1(i)} \times r_{2(i)}} \\ 0_{\,r_{2(i)} \times p} & 0_{\,r_{2(i)} \times r_{1(i)}} & \frac{1}{\hat{\tau}_2{}^2} I_{(r_{2(i)})} \end{bmatrix}.$$

Here, c is a constant that forces all probabilities to sum to one.

4.5.3: An Example of A Linear Primary and Both Second-Order and Special Cubic Potential Model for Three Components

In this example, it is assumed that the experimenter defines the following regressors for consideration:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}.$$

The terms $\{x_1, x_2, x_3\}$ are primary terms with p=3, the terms $\{x_1x_2, x_1x_3, x_2x_3\}$ are the first class of potential terms with r_1=3, and $\{x_1x_2x_3\}$ is the second class of potential term with r_2=1. The true model in this example is defined to contain the following regressors:

$$x^{(t)} = \{x_1, x_2, x_3, x_1x_3, x_1x_2x_3\}.$$

The first stage data set is simulated from the true model:

$$y_1 = 17.24x_1 + 19.57x_2 + 15.66x_3 + 10x_1x_3 + 11x_1x_2x_3 + \varepsilon$$

with ε ~N(0, 1). Nine runs are allocated to each stage of the experiment. The parameters $\tau_1 = 1, \tau_2 = 1$ are used in the first stage. The Bayesian D-optimality criterion presented in chapter three is used to select the first stage design with nine runs that are based on the full model, described by

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}.$$

The resulting design is shown in Figure 4.9.

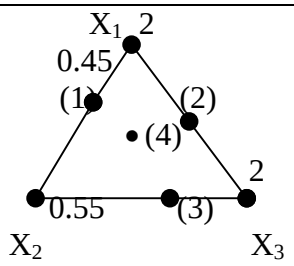


Figure 4.9. Bayesian first stage D-D optimal design for three components with two classes of potential terms

This is a desirable design since there are five vertices for estimating the coefficients of linear blending $\beta_1, \beta_2, \beta_3$, the three points 1, 2, and 3 for estimating the second-order terms $\beta_{12}, \beta_{13}, \beta_{23}$ and one point d for estimating the third-order term β_{123}. The response data from the first stage experiment is used to obtain the most likely values of $\hat{\tau}_1$ and $\hat{\tau}_2$, and to calculate the probabilities $p(M_i|y_1)$ associated with each of the candidate models. The following results are only for one simulation set. Figure 4.10 represents the posterior density of ${\tau_1}^2$ and ${\tau_2}^2$ that show that the value of $\hat{\tau}_1$ is close to 12.01 and the value of $\hat{\tau}_2$ is close to 15. These values of $\hat{\tau}_1$ and $\hat{\tau}_2$ are quite large. This makes good sense considering that the true model includes one term from each class of potential terms. The posterior probabilities are included in Table 4.3. From these posterior probabilities, we can see that the true model enjoys the highest probability.

For this example, nine runs of a second stage design are chosen. Figure 4.11 represents the second stage design points and shows that there are four vertices to estimate the coefficients of linear blending $\beta_1, \beta_2, \beta_3$, two equal proportion binary blends to estimate the second-order terms β_{13}, and three centroid points to estimate the third-order term β_{123}. This means that second stage design is very efficient in the

selection of design points based on information from the first stage design where one consider that the true model contains linear blending effects, the second-order x_1x_3 and the special cubic $x_1x_2x_3$. First stage and second stage design combined is presented in Figure 4.12.

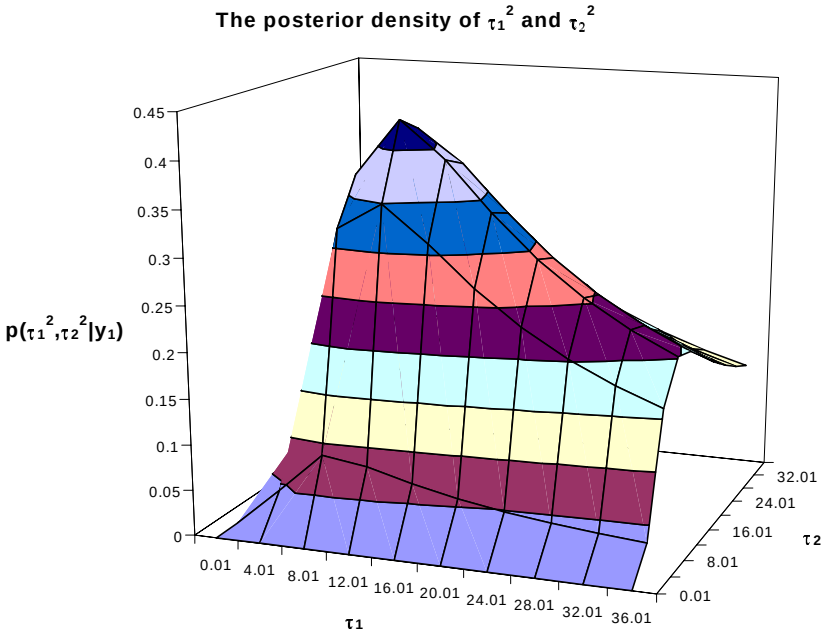


Figure 4.10. The posterior density of τ_1^2 and τ_2^2 and given the true model contains linear primary terms plus one from each of classes of potential terms

Table 4.3. Posterior probabilities of candidate models for three components with two classes of potential terms

Posterior Probability	Terms included in the candidate model
0	x_1 x_2 x_3
0	x_1 x_2 x_3 x_1x_2
0	x_1 x_2 x_3 x_1x_3
0	x_1 x_2 x_3 x_2x_3
0	x_1 x_2 x_3 $x_1x_2x_3$
0	x_1 x_2 x_3 x_1x_2 x_1x_3
0	x_1 x_2 x_3 x_1x_2 x_2x_3
0	x_1 x_2 x_3 x_1x_2 $x_1x_2x_3$
0	x_1 x_2 x_3 x_1x_2 x_2x_3
0.867*	x_1 x_2 x_3 x_1x_3 $x_1x_2x_3$*
0	x_1 x_2 x_3 x_2x_3 $x_1x_2x_3$
0	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3
0.072	x_1 x_2 x_3 x_1x_2 x_1x_3 $x_1x_2x_3$
0	x_1 x_2 x_3 x_1x_2 x_2x_3 $x_1x_2x_3$
0.052	x_1 x_2 x_3 x_1x_3 x_2x_3 $x_1x_2x_3$
0.009	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 $x_1x_2x_3$

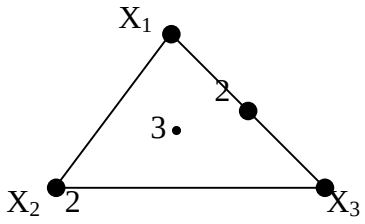


Figure 4.11. Bayesian second stage D-D optimal design for three components given the true model contains second order x_1x_3 and cubic $x_1x_2x_3$ potential terms

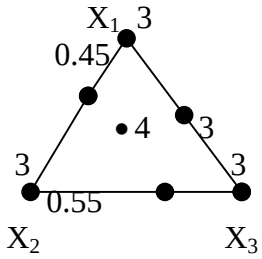


Figure 4.12. The combined Bayesian two-stage D-D optimal design for three components given the true model contains second order x_1x_3 and cubic $x_1x_2x_3$ potential terms

Chapter 5

Evaluation of the Bayesian Two-Stage D-D Design

The method of evaluating a Bayesian two-stage D-D optimal design will be presented in this chapter, as well as the results of evaluations of the Bayesian two-stage D-D optimal design versus both the Bayesian one-stage D-optimal designs and the non-Bayesian one-stage D-optimal designs. We will show that the Bayesian two-stage D-D optimal design is the most efficient when the true model contains fewer regressors than the full model.

5.1 Method of Evaluating Bayesian Two-Stage D-D optimal Designs

In the Bayesian two-stage D-D procedure, the second stage design points $x_{2(s|f)}$, and $\tau^2_{(s|f)}$ are random variables depending on the first stage parameter estimates, where $x_{2(s|f)}$ is the model matrix and $\tau^2_{(s|f)}$ is the parameter which reflects the importance of potential terms for the particular second stage design. Thus, the performance of the Bayesian two-stage D-D optimal design must take this fact into account.

By the laws of probability:

$$\mathrm{Var}(\hat{\beta}) = E_{d_2}[\mathrm{Var}(\hat{\beta} \mid d_2)] + \mathrm{Var}_{d_2}[E(\hat{\beta} \mid d_2)],$$

where $d_2 = [x_{2(s|f)}, \tau^2_{(s|f)}]$ is a vector of two parameters $x_{2(s|f)}$ and $\tau^2_{(s|f)}$. Here, $\hat{\beta}$ is the maximum likelihood estimate of β, and the expectation of $\hat{\beta}$ given d_2 is β. Asymptotically, $\mathrm{Var}_{d_2}[E(\hat{\beta} \mid d_2)] = 0$, and

$$\mathrm{Var}(\hat{\beta}) \approx E_{d_2}[\mathrm{Var}(\hat{\beta} \mid d_2)].$$

Therefore,

$$|\mathrm{Var}(\hat{\beta})| \approx \text{Average of } |[Var(\hat{\beta} \mid d_2)]|.$$

The covariance matrix $\mathrm{Var}(\hat{\beta} \mid d_2)$ for the Bayesian two-stage D-D optimal design in M_i is

$$\mathrm{Var}(\hat{\beta} \mid d_2) \approx n(X_{1(i)}'X_{1(i)} + X_{2(i)}'X_{2(i)})^{-1}.$$

Therefore, the performance of the Bayesian two-stage D-D optimal design is measured by

$$|\mathrm{Var}(\hat{\beta})| \approx \text{average of } |n(X_{1(i)}'X_{1(i)} + X_{2(i)}'X_{2(i)})^{-1}| \approx \frac{1}{m}\sum_{0}^{m} |n(X_{1(i)}'X_{1(i)} + X_{2(i)}'X_{2(i)})^{-1}|.$$

In the sections that follow, we use

$$D^* \approx |(\mathrm{Var}(\hat{\beta})| \approx \frac{1}{m}\sum_{0}^{m} |(X_{1(i)}'X_{1(i)} + X_{2(i)}'X_{2(i)})^{-1}|,$$

since the total sample size n is the same for the comparisons among different methods. Here, m is the number of simulation sets and the terms $X_{1(i)}$ and $X_{2(i)}$ contain the n_1 and n_2 points model matrix with traditional scaling (not the model matrix after scaling convention), expanded to contain only those regressors of the true model.

The reasonable alternatives to the Bayesian D-D design are the one-stage Bayesian D-optimal design and the non-Bayesian D-optimal design. The one-stage non-Bayesian D-optimal design and the Bayesian one-stage D-optimal design are not data dependent and can be evaluated by the following D* value for each of the true models.

$$D^* = |(X_i` X_i)^{-1}|.$$

To further support of Bayesian two-stage D-D optimal designs, even though the determinant criteria is used to compute design, A-criteria that deal with the individual variances of the regression coefficients are included. A-criteria for the Bayesian two-stage design and one-stage are defined by

$$A^* = \frac{1}{m}\sum_{0}^{m} \operatorname{tr}(X_{1(i)}'X_{1(i)} + X_{2(i)}'X_{2(i)})^{-1},$$

and

$$A^* = \operatorname{tr}((X_i'X_i)^{-1}),$$

respectively.

5.2 An Example of Results for Three Components

In this section, different initial values of τ used in the first stage are examined. In this example, the full model defined by the experimenter contains the following regressors:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

The primary terms are $\{x_1, x_2, x_3\}$ with p=3, while $\{x_1x_2, x_1x_3, x_2x_3\}$ are the potential terms where r=3. Four cases are considered for simulation purposes; the response data is simulated from each of the true models. It is assumed that $\varepsilon \sim N(0,1)$.

Case 1: Primary terms only

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + \varepsilon$$

Case 2: Primary terms plus one of potential terms

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 8x_1x_3 + \varepsilon$$

Case 3: Primary terms plus two of potential terms

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + \varepsilon$$

Case 4: Full model

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + 7x_2x_3 + \varepsilon$$

The following three methods are compared.

1. Bayesian two-stage D-D optimal design with $n_1 = 8$ and $n_2 = 8$ is generated by the method presented in Chapter 4.

2. Bayesian one-stage D-optimal design with n=16 is generated by the method presented in Chapter 3.

3. Non-Bayesian one stage design with n=16 is generated by using "Proc Optex" from the candidate points.

Two hundred data sets were simulated from each of the true models for the first stage data analysis, which leads to the second stage design. Notice that the design only slightly difference based on different simulation data set, therefore two hundred sets or fifty simulation data sets are enough. Tables 5.1, 5.2, 5.3 and 5.4 present the results of evaluation for different true models, respectively. The best design for D and A criterion in each row is marked with an asterisk, respectively. These tables show that the Bayesian

two-stage D-D optimal design is more D-efficiency than the traditional designs for the true model that contains fewer regressors than the full model when the different initial τ values are used. If the true model is the full model, then the one-stage non-Bayesian design is more D-efficiency for all the different τ values used in the first stage. Notice that when we increase the assumed value of τ used in the first stage, the D* for the Bayesian two-stage D-D optimal design comes closer to the one-stage non-Bayesian D-optimal design, where the full model is the true model. Also notice that the Bayesian two-stage D-D optimal designs are the most A-efficiency for all true models when τ is 1 and 5.

The most D-efficient design is marked with # for different values of τ used in the first stage. In Tables 5.1 and 5.3, the design for τ =1 is more efficient than the design for τ =5 and little bit less efficient than the design for τ =0.6. In Table 5.4, the design for τ =1 is more efficiency than the design for τ =0.6, and little less efficient than the design τ =5. In Table 5.2, the design for τ =1 is the most efficient. Therefore, the value of τ =1 is a robust value, and can be recommended for initial τ used in the first stage when the experimenter has no idea about the importance of the potential terms.

Table 5.1. Values of D^* and A* for three components with only primary terms in the true model

True Model: $\mu_y = 7.24x_1+9.57x_2+5.66x_3$						
τ	Bayesian D-D Procedure		1-Stage Bayesian D		1-Stage non-Bayesian D	
	D*	A*	D*	A*	D*	A*
0.6	0.0103*#	0.63*	0.013	0.718	0.0147	0.747
1	0.0106*	0.678*	0.0147	0.747	0.0147	0.747
5	0.0114*	0.672*	0.0147	0.747	0.0147	0.747

Table 5.2. Values of D^* and A* for three components with primary terms and one potential term in the true model

True Model: $\mu_y = 7.24x_1+9.57x_2+5.66x_3+8x_1x_3$						
τ	Bayesian D-D Procedure		1-Stage Bayesian D		1-Stage non-Bayesian D	
	D*	A*	D*	A*	D*	A*
0.6	0.1*	7.478*	0.1312	10.873	0.1513	11.142
1	0.09*#	7.721*	0.1513	11.142	0.1513	11.142
5	0.0926*	7.415*	0.1513	11.142	0.1513	11.142

Table 5.3. Values of D^* and A* for three components with primary terms and two potential terms in the true model

True Model: $\mu_y = 7.24x_1+9.57x_2+5.66x_3+6.8x_1x_2+8x_1x_3$						
τ	Bayesian D-D Procedure		1-Stage Bayesian D		1-Stage non-Bayesian D	
	D*	A*	D*	A*	D*	A*
0.6	0.9064*#	16.8*	1.3333	21.167	1.5802	21.78
1	0.9269*	16.85*	1.5802	21.78	1.5802	21.78
5	0.9337*	16.93*	1.5802	21.78	1.5802	21.78

Table 5.4. Values of D^* and A* for three components with full model

True Model: $\mu_y = 7.24x_1+9.57x_2+5.66x_3+6.8x_1x_2+8x_1x_3+7x_2x_3$						
τ	Bayesian D-D Procedure		1-Stage Bayesian D		1-Stage non-Bayesian D	
	D*	A*	D*	A*	D*	A*
0.6	16.81	35.53	14.2222	32.25	12.642*	30.33*
1	13.9127	29.78*	12.642*	30.33	12.642*	30.33
5	12.71#	29.75*	12.642*	30.33	12.642*	30.33

5.3 The Results of Evaluation for Four Components

5.3.1 A Linear Primary and Quadratic Potential Model

In this section, the full model defined by the experimenter is

$$x^{(f)} = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}.$$

The primary terms are $\{x_1, x_2, x_3, x_4\}$ where p=4, and $\{x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}$ are potential terms where r=6. Five cases are considered for simulation purposes, and the response data is simulated from each of the true models. The initial τ =1 is used in the first stage and $\varepsilon \sim N(0,1)$ is assumed.

Case 1: Primary terms only

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + \varepsilon$$

Case 2: Primary terms plus one of potential terms

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + \varepsilon$$

Case 3: Primary terms plus two of potential terms

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + 4.5x_1x_3 + \varepsilon$$

Case 4: Primary terms plus three of potential terms

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + 4.5x_1x_3 + 5x_2x_3 + 6x_3x_4 + \varepsilon$$

Case 5: Full model

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + 4.5x_1x_3 + 5x_2x_3 + 5.2x_1x_4 + 7x_2x_4 + 6x_3x_4 + \varepsilon$$

In this section, the following three methods are compared:

1. Bayesian two-stage D-D optimal design with n_1=12 and n_2 =12 is generated by the method represented in Chapter 4.

2. Bayesian one stage D-optimal design with n=24 is generated by the method which is presented in Chapter 3.

3. Non-Bayesian one stage design with n=24 is generated by using "Proc Optex" from the candidate points.

In this section, fifty data sets are simulated from each of the true models for the first stage data analysis, which leads to the second stage design. Table 5.5 presents the results of the evaluation for five different true models, along the value of A* (in parentheses). The table shows that Bayesian D-D optimal design is the best for all subsets of the full model, but not the full model itself, which is the true model. If the true model is the full model, then the one-stage non-Bayesian and Bayesian designs are somewhat better than the two-stage D-D optimal design.

Table 5.5. Values of D^* and A* (in parentheses) for four components with a linear primary and a second-order potential terms model

D* (A*)	Bayesian D-D Procedure	Bayesian One-stage D	Non-Bayesian One-stage D
Case 1	0.0025* (0.842*)	0.0026 (0.92)	0.0026 (0.92)
Case 2	0.0165* (7.53*)	0.0263 (11.08)	0.0263 (11.08)
Case 3	0.134* (16.307*)	0.2689 (21.47)	0.2689 (21.47)
Case 4	16.53* (36.43*)	28.9 (42.91)	28.9 (42.91)
Case 5	3815 (66.9)	3236* (65.33*)	3236* (65.33*)

Case 1: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + \varepsilon$

Case 2: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + \varepsilon$

Case 3: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + 4.5x_1x_3 + \varepsilon$

Case 4: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 5.5x_1x_2 + 4.5x_1 x_3 + 5x_2x_3 + 6x_3 x_4 + \varepsilon$

Case 5: $y = 17.24x1 + 19.57x2 + 15.66x3 + 14.8x4 + 5.5x1x2 + 4.5x1x3$

$4.5x_1x_3 + 5x_2x_3 + 6x_3 x_4 + 5.2x_1x_4 + 7x_2x_4 + \varepsilon$

5.3.2 A Second-Order Primary and Special Cubic Potential Model

In this section, the full model defined by the experimenter is

$$x^{(f)} = \{x_1,x_2,x_3,x_4,x_1x_2,x_1x_3,x_1x_4,x_2x_3,x_2x_4,x_3x_4,x_1x_2x_3,x_1x_2x_4,x_1x_3x_4,x_2x_3x_4\}.$$

The primary terms are $\{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}$ where p=10, and $\{x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4\}$ are potential terms where r=4. Four cases are considered for simulation purposes and the response data set is simulated from each of the true models. The initial τ =1 is used in the first stage and $\varepsilon \sim N(0,1)$ is assumed.

Case 1: Primary terms

$$y=17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 + 18.21x_2x_3$$

$$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 + \varepsilon$$

Case 2: Primary terms plus one of the potential terms

$$y=17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 + 18.21x_2x_3 +$$

$$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3 + \varepsilon$$

Case 3: Primary terms plus two of the potential terms

$$y = 17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 +18.21x_2x_3$$

$$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3 + 14.7x_2x_3x_4 + \varepsilon$$

Case 4: Full model

$$y = 17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3$$

$$+18.21x_2x_3+17.64x_1x_4 + 15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3$$

$$+14x1x2x4+15.7x1x3x4+ 14.7x_2x_3x_4+ \varepsilon$$

In this section, the following three methods are compared.

1. Bayesian two-stage D-D optimal design with n_1=16 and n_2 =16 is generated by the method presented in Chapter 4.

2. Bayesian one stage D-optimal design with n=32 is generated by the method presented in Chapter 3.

3. Non-Bayesian one stage design with n=32 is generated by using "Proc Optex" from the candidate list.

In this section, fifty data sets are simulated from each of the true models for the first stage data analysis, which leads to the second stage design. Table 5.6 presents the results of the evaluation for the four different true models, along the value of A* (in parentheses). The table shows that the Bayesian two-stage D-D optimal design is the most D-efficiency for all the subsets of the full model, except in case 4, where the two stage Bayesian D-D optimal design is as efficient as one-stage Bayesian and non-Bayesian designs. Notice that the Bayesian two-stage design is always more A-efficiency than the natural competitors.

Table 5.6. Values of D^* and A* (in parentheses) for four components with a second-order primary and a special cubic potential terms model

D^* (A*)	Bayesian D-D Procedure	Bayesian One-stage D	Non-Bayesian One-stage D
Case 1	371.1* (50.2*)	518.41 (50.83)	1330.68 (53.651)
Case 2	202449.1* (484.65*)	292922.05 (619.19)	587491.68 (499.72)
Case 3	15139146* (560.5*)	167345430 (1200.217)	263672873 (960.653)
Case 4	5.71e13* (1888.24*)	5.71e13* (2423.44)	5.71e13* (1964.07)

Case 1: $y=17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 + 18.21x_2x_3$

$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 + \varepsilon$

Case 2: $y=17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 + 18.21x_2x_3 +$

$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3 + \varepsilon$

Case 3: $y = 17.24x_1+ 19.57x_2 + 16.66x_3 + 14.8x_4 + 17x_1x_2 + 18.51x_1x_3 +18.21x_2x_3$

$17.64\ x_1x_4 + 15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3 + 14.7x_2x_3x_4 + \varepsilon$

Case 4: $y=17.24x_1+19.57x_2+16.66x_3+14.8x_4+17x_1x_2+18.51x_1x_3+18.21x_2x_3+17.64x_1x_4+$

$15x_2x_4 + 17.6x_3x_4 +15x_1x_2x_3 +14x1x2x4+15.7x1x3x4+ 14.7x_2x_3x_4+ \varepsilon$

5.4. An Example of the Results for Three Components with Two Classes of Potential Terms

In this section, two classes of potential terms are considered for three components. Also we show the results of the Bayesian two-stage D-D design vs. the Bayesian one-stage D-optimal design and the non-Bayesian one-stage D-optimal design.

In this example, the full model defined by the experimenter contains the following regressors:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}.$$

The primary terms are $\{x_1, x_2, x_3\}$ where p=3, the first class of potential terms is $\{x_1x_2, x_1x_3, x_2x_3\}$ where $r_1=3$ and $\{x_1x_2x_3\}$ is the second class of potential terms where $r_2=1$. Five cases are considered for simulation purposes and the response data set is simulated

from each of the true models. The error $\varepsilon \sim N(0,1)$ is assumed and initial $\tau = 1$ is used in the first stage.

Case 1: Primary terms only

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + \varepsilon$$

Case 2: Primary terms plus one of first class of potential terms

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + \varepsilon$$

Case 3: Primary terms plus one from each of class of potential terms

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 10x_1x_3 + 11x_1x_2x_3 + \varepsilon$$

Case 4: Primary terms and two of first class of potential terms and one of second class of potential term

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 9.5x_1x_2 + 10x_2x_3 + 9.8x_1x_2x_3 + \varepsilon$$

Case 5: Full model

$$y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 9.5x_1x_2 + 8.6x_1x_3 + 10x_2x_3 + 9.8x_1x_2x_3 + \varepsilon$$

In this section, the following three methods are compared.

1. Bayesian two-stage D-D optimal design with $n_1=9$ and $n_2=9$ is generated by the method presented in Chapter 4.

2. Bayesian one stage D-optimal design with n=18 is generated by the method presented in Chapter 3.

3. Non-Bayesian one-stage design with n=18 is generated by using "Proc Optex" from the candidate points.

In this section, two hundred data sets are simulated from each of the true models for the first stage data analysis, which leads to the second stage design. Table 5.7 presents the results of the evaluation for three components models with two classes of potential terms, along the value of A* (in parentheses). This table shows that the Bayesian two-stage D-D optimal design is the best for all the subsets of the full model, but not the full model itself, which is the true model. If the true model is the full model, then the one-stage non-Bayesian design and the one-stage Bayesian design are little better than the Bayesian two-stage D-D optimal design.

Table 5.7. Values of D* and A* (in parentheses) for three components with two classes of potential terms

D^* (A*)	D-D Procedure	Bayesian One-stage D	Non-Bayesian One-stage D
Case 1	0.009* (0.632*)	0.014 (0.738)	0.014 (0.738)
Case 2	0.069* (6.868*)	0.121 (9.79)	0.121 (9.79)
Case 3	24.953* (279.68*)	41.142 (348.53)	41.142 (348.53)
Case 4	336.12* (368.11*)	432 (416.44)	432 (416.44)
Case 5	4662.94 (558.16)	4608* (501*)	4608* (501*)

Case 1: $y = 17.24x1 + 19.57x2 + 15.66x3 + \varepsilon$

Case 2: $y = 17.24x1 + 19.57x2 + 15.66x3 + 8x1x2 + \varepsilon$

Case 3: $y = 17.24x1 + 19.57x2 + 15.66x3 + 10x1x3 + 11x1x2x3 + \varepsilon$

Case 4: $y = 17.24x1 + 19.57x2 + 15.66x3 + 9.5x1x2 + 10x2x3 + 9.8x1x2x3 + \varepsilon$

Case 5: $y = 17.24x1 + 19.57x2 + 15.66x3 + 9.5x1x2 + 8.6x1x3 + 10x2x3 + 9.8x1x2x3 + \varepsilon$

5.5 Conclusion

It is expected that the non-Bayesian one-stage D-optimal design is somewhat more D-efficiency than the Bayesian two-stage D-D optimal design when the true model is the full model, since the non-Bayesian one-stage design is generated based on the full model. The Bayesian two-stage D-D optimal design, however, shows more efficient using the D criterion than one-stage D-optimal designs when the true model contains fewer regressors than the full model. Also we see that Bayesian D-D designs are more A-efficient in most cases. Therefore, the Bayesian two-stage D-D optimal design is a good choice when one is uncertain about the right model.

Chapter 6

Distribution of the Sample Size between Two Stages

A reasonable ratio of the sample sizes between the two stages is discussed in this chapter. The sufficient sample size in the first stage is also investigated. We show that a ratio of 1:1 between the two stages is quite satisfactory and the sufficient amount of data for the first stage should be at least p+r+2, where p and r are the numbers of primary and potential terms, respectively.

Due to the difficulty to reach the goals analytically, simulations are carried out in this chapter. The full model is given by the following regressors:

$$x^{(f)} = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

The primary terms are $\{x_1, x_2, x_3\}$ with p=3, and the potential terms are $\{x_1x_2, x_1x_3, x_2x_3\}$ with r=3. In the simulation study, four cases are considered, and the response data is simulated from each of the following true models.

Case 1: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + \varepsilon$

Case 2: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + \varepsilon$

Case 3: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + 9x_1x_3 + \varepsilon$

Case 4: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + 10x_1x_3 + 9x_2x_3 + \varepsilon$

The initial parameter $\tau = 1$ is used in the first stage and $\varepsilon \sim N(0,1)$ is assumed. In this chapter two hundred data sets are generated from each of the true models for the first stage data analysis. The second stage design will be determined by using the first stage data. Again the prior probability of activity or importance of any potential terms is $\pi = 0.33$ in this chapter.

6.1 Ratio of the Sample Sizes between the Two Stages

The total sample size of 24 is used, and the following five methods are compared:

1. Bayesian two-stage D-D optimal design with $n_1=12$ and $n_2 =12$.
2. Bayesian two-stage D-D optimal design with $n_1=8$ and $n_2 =16$
3. Bayesian two-stage D-D optimal design with $n_1=16$ and $n_2 =8$
4. Bayesian one stage D-optimal design with n=24
5. Non-Bayesian one stage design with n=24

For simplicity, an abbreviation of n_1-n_2 is used to represent a partition of total size n, where n_1 and n_2 are the number of runs in the first stage and second stage respectively. Table 6.1 shows that the 12-12 design is the most efficient design among the three partitions of the 12-12, 8-16 and 16-8 designs. To obtain some indication of the effect of total sample size, the value of n is increased to 30. Again the following five methods are compared:

1. Bayesian two-stage D-D optimal design with $n_1=15$ and $n_2 =15$
2. Bayesian two-stage D-D optimal design with $n_1=10$ and $n_2 =20$
3. Bayesian two-stage D-D optimal design with $n_1=20$ and $n_2 =10$

4. Bayesian one stage D-optimal design with n=30

5. Non-Bayesian one stage design with n=30

Table 6.2 reports the results for the partitions for sample size n=30. It shows that the 10-20 design is the most efficient among different partitions for cases 1 and 2, while the 15-15 design is the most efficient among different partitions for cases 3 and 4. From above two studies, we may say that 1:1 is a reasonable ratio between the two stages. To have clear picture about a good ratio between the two stages, large sample sizes and small sample sizes are used for the different ratios between the two stages, all the results show that a ratio one is robust value between the two stages.

Table 6.1. Simulation study for data allotment with sample size 24

D^*	D-D procedure		D-D procedure		D-D procedure		D procedure
	n_1=12	n_2=12	n_1=8	n_2=16	n_1=16	n_2=8	n=24
Case 1	0.0026*		0.0027		0.003		0.05
Case 2	0.016*		0.017		0.02		0.029
Case 3	0.116*		0.13		0.131		0.167
Case 4	1*		1.29		1*		1*

Table 6.2. Simulation study for data allotment with sample size 30

D^*	D-D procedure		D-D procedure		D-D procedure		D procedure
	n_1=15	n_2=15	n_1=10	n_2=20	n_1=20	n_2=10	n=30
Case 1	0.0015		0.0014*		0.0017		0.0026
Case 2	0.0072		0.0071*		0.0079		0.0117
Case 3	0.04*		0.045		0.044		0.0546
Case 4	0.2621*		0.32		0.2621*		0.2621*

Case 1: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + \varepsilon$

Case 2: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + \varepsilon$

Case3: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + 9x_1x_3 + \varepsilon$

Case 4: $y = 17.244x_1 + 19.57x_2 + 15.66x_3 + 8x_1x_2 + 10x_1x_3 + 9x_2x_3 + \varepsilon$

6.2 Sufficient Sample Size for the First Stage Design

Obviously, in order to obtain accurate information for estimating the parameters, certain amount of data is necessary for the first stage design. For this reason, additional simulations are carried out in this section. In first simulation, the same full model as in section 6.1 is used. Three different true models are considered, and the 8-8 and 9-7 are partitioned in the sample size n=16.

Case 1: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + \varepsilon$

Case 2: $y = 7.244x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + \varepsilon$

Case 3: $y = 7.244x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + 7x_2x_3 + \varepsilon$

In second simulation, the 6-6. 8-4, and 9-3 are partitioned in the sample size n=12. Table 6.3 presents the D^* for the sample size n=16. Table 6.4 presents the D^* for the sample size n=12. Table 6.3 shows that 8 runs in the first stage is more efficient than 9 runs in the first stage when the true model is any subset of the full model, except in case 3 where the true model is the full model. Table 6.4 shows that 9 runs in the first stage is more efficient than 6 and 8 runs in the first stage when the true model is any subset of the full model, except in case 4 where true model is the full model.

From this study, we see that 9 runs in the first stage is better than 6 and 8 runs in the first stage for sample size 12, and 8 runs in the first stage is better than 9 runs in the

first stage for the sample size 16. In this example the number of primary terms and potential terms are p=3 and r=3, respectively. Therefore, we might say that a reasonable sample size for the first stage design should be at least p+r+2 (where p and r is the number of primary and potential terms, respectively).

Table 6.3. Simulation study for sufficient data in the first stage (n=16)

D^*	D-D procedure		D-D procedure		Bayesian D	No-Bayesian D
	n_1=9	n_2=7	n_1=8	n_2=8	n=16	n=16
Case 1	0.012		0.01*		0.0145	0.0145
Case 2	1.2		0.93*		1.58	1.58
Case 3	12.66		13.12		12.642*	12.642*

Case 1: $y= 7.24x_1 + 9.57x_2 + 5.66x_3$

Case 2: $y= 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + \varepsilon$

Case 3: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8x_1x_2 + 8x_1x_3 + 7x_2x_3 + \varepsilon$

Table 6.4. Simulation study for sufficient data for the first stage (n=12)

D^*	D-D procedure		D-D procedure		D-D procedure		One-stage D
	n_1=9	n_2=3	n_1=6	n_2=6	n_1=8	n_2=4	n=12
Case 1	0.027*		0.039		0.029		0.04
Case 2	0.337*		0.369		0.347		0.457
Case 3	4.601*		4.82		4.858		5.333
Case 4	67.4		64.21		72.84		64*

Case 1: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + \varepsilon$

Case 2: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 5x_1x_2 + \varepsilon$

Case 3: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 5x_1x_2 + 4.5x_1x_3 + \varepsilon$

Case 4: $y = 17.24x_1 + 19.57x_2 + 15.66x_3 + 5x_1x_2 + 4.5x_1x_3 + 5.5x_2x_3 + \varepsilon$

Chapter 7

Application with Restrictions on Component Proportions

We have focused on mixture experiments for simplex regions. However, in many mixture experiments there are restrictions on the component proportions, where the component proportion has lower or upper bound or both $(L_i \le x_i \le U_i)$. In this chapter we discuss the different types of restrictions; we also present two examples for cases where the feasible region for the design is not a simplex and the design is chosen using the method of Bayesian two-stage D-D optimality.

7.1 Case where Feasible Region is a Simplex

In this section, we consider a case where the feasible region is a simplex. The lower bound constrained mixture models satisfy $\sum x_i = x_1 + x_2 + \cdots x_q = 1$, where $L_i \le x_i \le 1$, i = 1, 2, …, q. The following is an example of a lower bound case with three components, where $0.3 \le x_1$, $0.4 \le x_2$, and $0.1 \le x_3$. The graph is shown in Figure 7.1.

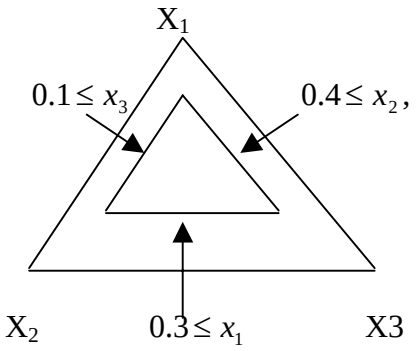


Figure 7.1. Example where the feasible region is a simplex.

In this case the feasible mixture space for three components with lower bounds is a simplex, which is always the case under this circumstance (Montgomery and Myers, 1995). As a result, it seems reasonable to define a new set of components that will take on the values 0 to 1 over the feasible region. This will then allow the use of designs in the metric of those discussed for a standard mixture situation. These redefined components are called L-pseudocomponents (Cornell, 1990). The pseudocomponents X_i are defined using the following transformation: $X_i = \frac{x_i - L_i}{1 - L}$, where $L = \sum_{i=1}^{q} L_i < 1$. In this situation, we can treat Bayesian two-stage D-D optimal design points as points of pseudocomponents. Using $x_i = L_i + (1 - L)X_i$, we can get real design points in the feasible region by applying Bayesian two-Stage D-D optimal design points.

There are terms when upper-bound constraints of the form $x_i \le U_i$ may be placed on the component proportions. One possible result is that the feasible region is an inverted simplex. An example would be the three-component problem $x_1 \le 0.4$, $x_2 \le 0.5$, and $x_3 \le 0.3$, shown in Figure 7.2.

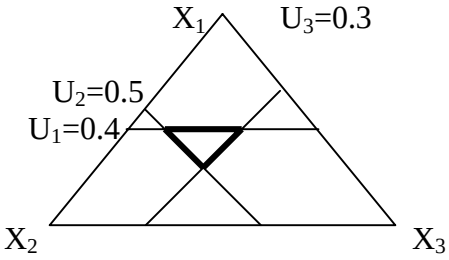


Figure 7.2. Example where the feasible region is an inverted simplex region.

For this case, we also can treat Bayesian two-stage D-D optimal design points as points of pseudocomponents. Using $x_i = U_i - \left(\sum_{i=1}^{q} U_i - 1\right)u_i$, where $\sum_{i=1}^{q} U_i > 1$ and u_i is the i^{th} design point from Bayesian two-stage D-D optimal design, we can get real design points in the feasible region by applying Bayesian two-stage D-D optimal design points.

7.2 Case in Which the Feasible Region is not a Simplex

Another possible situation is that the feasible region is not a simplex when there are upper bound constraints. The candidate points consist of extreme vertices, the middle points of the edges, the third points of the edges, constraint plane centroids, the axial check blends, the overall centroid, as well as a lattice grid where each proportion is a multiple of 0.05. The following are two examples of Bayesian two-stage D-D optimal designs for such a case. In the first example, the constraints on component proportions are $x_2 \leq 0.7$, $x_2 \leq 0.5$, $x_3 \leq 0.8$. It is assumed that the experimenter defines the following regressors:

$$x^{(f)} = \{ \underbrace{x_1, x_2, x_3}_{\text{Primary}}, \underbrace{x_1x_2, x_1x_3, x_2x_3}_{\text{Potential}} \}.$$

The true model is set to contain all primary terms and x_1x_2. The response is simulated from the true model:

$$y = 7.24x_1+9.57x_2+5.66x_3+6.8x_1x_2+\varepsilon ,$$

and eight runs are allocated in each stage. The following result is one possible design. The results may vary based on different simulated first stage data sets; however, most of the results observed are the same as this one. The first stage extreme vertices design is shown in Figure 7.3, and a second stage design is presented in Figure 7.4. In the first stage design, there are points a and c to estimate β_1 and $\beta_{2,}$ points b and d to estimate β_1 and β_3, and points e and f to estimate β_2 and β_3. Meanwhile, points a and b help to estimate the coefficients of the second order terms β_{12} and β_{13} respectively, points d and f help to estimate the coefficients of second order terms β_{13} and β_{23} respectively, and points c and e help to estimate the coefficients of second order β_{12} and β_{23} respectively. The first stage seems reasonable, since there is no preliminary information on potential terms. In the second stage design, point g helps to estimate β_{12} considering that the true model contains the term x_1x_2. Notice that the distance between a and c is too short to insert a binary blend.

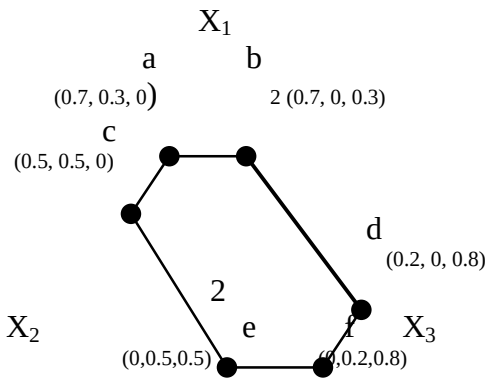


Figure 7.3. The first stage design of Bayesian two-stage design for three constraints components

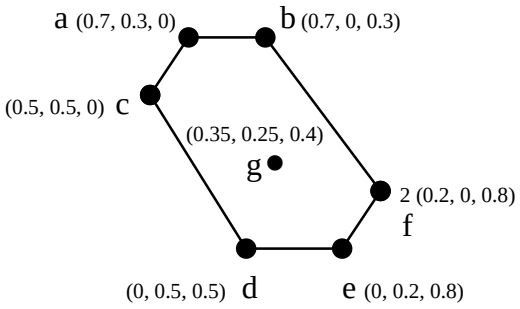


Figure 7.4. A second stage of Bayesian two-stage design for three constraints components

In the second example, constraints on the four components are chosen to be

$$0.4 \le x_1 \le 0.6,\ 0.1 \le x_2 \le 0.47,\ 0.1 \le x_3 \le 0.47,\ 0.3 \le x_3 \le 0.08.$$

It is assumed that the experimenter defines the following regressors:

$$x^{(f)} = \{ \underbrace{x_1, x_2, x_3, x_4}_{\text{Primary}},\ \underbrace{x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4}_{\textit{Potential}} \}.$$

The true model is set to contain all primary terms and potential terms x_1x_2 and x_1x_3. The response is simulated from the true model:

$$y = 1717.24x_1 + 19.57x_2 + 15.66x_3 + 14.8x_4 + 15.5x_1x_2 + 14.5x_1x_3 + \varepsilon,$$

and 12 runs are allocated in each stage. The first stage design and one second stage design are given in Figures 7.5 and 7.6, respectively. The first stage design allows

estimation of the coefficients of linear blending terms and second order potential terms. This makes good sense considering that there is no information on potential terms. In the second stage, the new points a* and b* help to estimate β_{12}, and new points c* and d* help to estimate β_{13}.

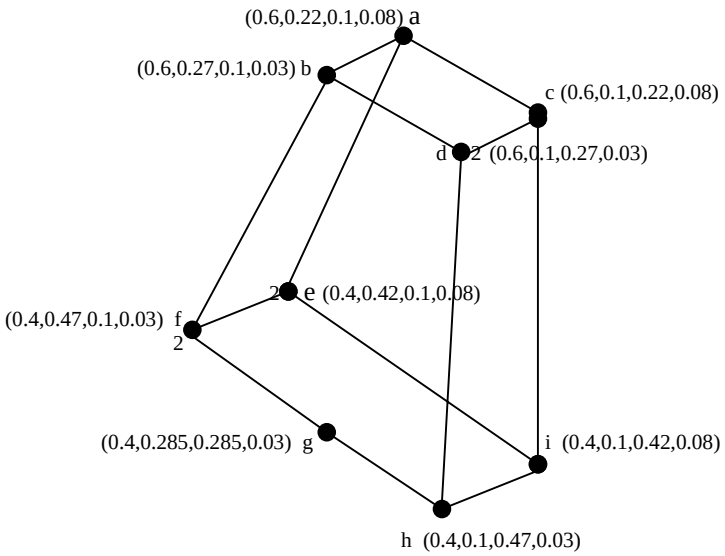


Figure 7.5. The first stage of Bayesian two-stage design for four constraints components

.

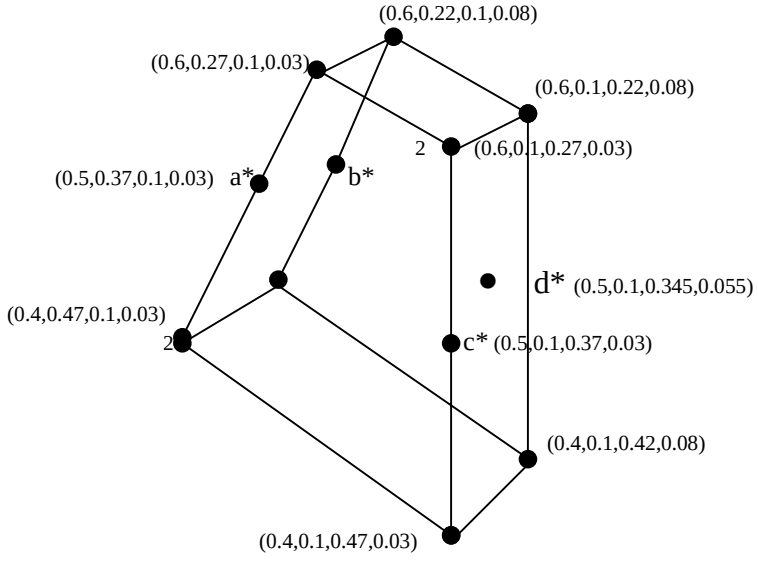


Figure 7.6. A second stage of Bayesian two-stage design for four constraints components

Chapter 8

Bayesian Two-Stage Designs for Mixture Experiments with Process Variables

8.1 Introduction

Cornell (1990) defined process variables as "factors in an experiment that do not form any portion of the mixture but whose levels, when changed, could affect the blending properties of the ingredients". Two experiments involving process variables from Cornell's book have been selected to explain what are called process variables. In the first example, the goal is to see whether a mixture of fuels A and B result in an increase in miles per gallon compared to the simple average of the mileage obtained with the separate fuels. When we compare miles per gallon between mixed A and B fuel and separate fuel A or B, usually we are talking about the same automobile driven at a constant speed. Now, suppose that, in addition to studying the joint blending of the two fuels, we are interested in studying the effect of different driving speed and different automobile sizes, particularly as these two factors (driving speed and automobile size) influence the blending behavior of the fuels. These latter two factors are called the

process variables of the experiment, since they do not form any portion of the mixtures and are therefore not components. In the second example, the goal of the experiment is to find the proportions of three types of fish (mullet, sheepshead, and croaker) that will give the best texture. Obviously the texture of the fish patties depends not only on the proportion of three fish species that are blended but also on cooking temperature, cooking time, and deep fat frying time. Here, cooking temperature, cooking time, and deep fat frying time are viewed as the process variables of the experiment.

The number of process variables can be any positive integer s in most practical situations. However, it would be reasonable if we can keep s small ($s \leq 6$) since the blending properties of the mixture components are also studied. Now both the mixture components and the process variables are simultaneously investigated. In this chapter, the process variables are to be studied at low, middle, and high levels. If quantitative, the settings of the i^{th} process variable are denoted by $z_l = -1$, $z_m = 0$ and $z_h = +1$, respectively, and the region of interest for the process variables is coded to be a s-dimensional hypercube. For the mixture components, the mixture region of interest is the (q-1)-dimensional simplex (q is the number of mixture components). Therefore, the combined region of interest for the mixture components and the process variables is of dimensionality q-1+s.

Historically, in the literature about mixtures with process variables, the mixture components have been of the most interest to the researchers, while the process variables have been treated as "noise" factors. The primary focus on interactions of the mixture by process variable has been on the effects of the process variables on the blending properties of the mixture components. However, in many industrial situations, the

interest in the process variables is at least equal to the interest in the mixture components. This study considers process variables to be as important as mixture components, therefore, we use Kowalski, Cornell and Vining's (1999) proposed mixture models with process variables. For example, the combined second-order model is

$$\eta(x,z) = \sum_{i=1}^{q} \beta^{*}_{i} x_i + \sum \sum_{i<j}^{q} \beta^{*}_{ij} x_i x_j + \sum_{k=1}^{n} \alpha_{kk} z_k^{2} + \sum \sum_{k<l}^{n} \alpha_{kl} z_k z_l + \sum_{i=1}^{q} \sum_{k=1}^{n} \gamma_{ik} x_i z_k ,$$

which includes the mixture model, plus pure quadratic as well as two-factor interaction effects among the process variables, and two-factor interactions between the linear blending terms of the mixture components and the main effect terms of the process variables. To efficiently estimate the coefficients in the mixture models with the process variables under model uncertainty, Bayesian two-stage D-D optimal designs are developed. Examples of Bayesian two-stage D-D optimal designs and comparison of its design versus both the one-stage D-optimal design and the one-stage Bayesian D-optimal design are made.

8.2 Examples of Bayesian Two-Stage Designs

Examples of Bayesian two-stage D-D optimal designs for three component models with one and two process variables are presented in this section. Section 8.2.1 represents a Bayesian two-stage D-D optimal design for three components with one process variable, where x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3 are treated as primary terms and interactions x_1z, x_2z and x_3z are treated as potential terms. Section 8.2.2 represents a Bayesian D-D optimal design for three components with two process variables, where x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3 and x_2x_3 are treated as primary terms, and interactions x_1z_1, x_2z_1 x_3z_1, x_1z_2, x_2z_2 and x_3z_2 are treated as potential terms. These examples show that the second stage design

provides more points to estimate parameters in the true model. These results show that Bayesian two-stage D-D optimal design is very efficient based on the information from the first stage data.

8.2.1 Three Components with One Process Variable

In this example, it is assumed that the following regressors are used:

$$x^{(f)} = \{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3, x_1z, x_2z, x_3z\}.$$

The primary terms are $\{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3\}$ with p=7 (p is the number of primary terms) and $\{x_1z, x_2z, x_3z\}$ are the potential terms with r=3 (r is the number of potential terms). The true model in this example is set to contain only the primary terms,

$$x^t = \{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3\}.$$

The first stage data set is simulated from the true model:

$$Y = 7.24x_1+9.57x_2+5.66x_3+6.8z^2+8x_1x_2+6.5x_1x_3+7x_2x_3+\varepsilon,$$

where $\varepsilon \sim N(0,1)$. Twelve runs are allocated to each stage of the experiment, and the parameter $\tau = 1$ is used in the first stage. Recall that the $\tau = 1$ is a robust value for the first stage.

The Bayesian D-optimality criterion is used to select the first stage design with twelve observations that are based on the full model by

$$x^{(f)} = \{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3, x_1z, x_2z, x_3z\}.$$

The resulting design is shown in Figure 8.1. In the first stage design, six points 1, 2, 7, 9, 11, and 12 estimate the coefficients of the linear blending $\beta_1, \beta_2, \beta_3$; six binary blends 5, 6, 3, 4, 8, and 10 estimate the coefficients of second-order terms $\beta_{12}, \beta_{13}, \text{and } \beta_{23}$; two points 9 and 11 estimate the coefficient of interaction x_2z; and two points 9 and 12 estimate the coefficient of interaction x_3z. Notice that all the parameters in the full model are estimated in the first stage.

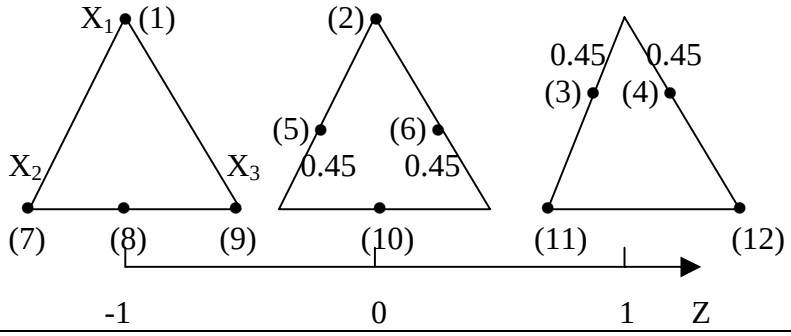


Figure 8.1. Bayesian first stage design for three components with one process variable (n_1=12)

Recall that the response data from the first stage experiment is used to estimate the value of τ, and to calculate the probabilities/weights associated with each of the candidate models. The following results are only for one set of simulation data. Figure 8.2 represents the posterior density of τ^2, which shows that $\hat{\tau}$ is close to zero. This makes good sense considering that the true data is from the model of the primary terms and that there are no potential terms in the true model. The first stage data and $\hat{\tau}$ are used to calculate posterior probabilities to weigh the importance of each candidate model. The posterior probabilities are included in Table 8.1. From these posterior probabilities, we can see that the true model enjoys the highest probability.

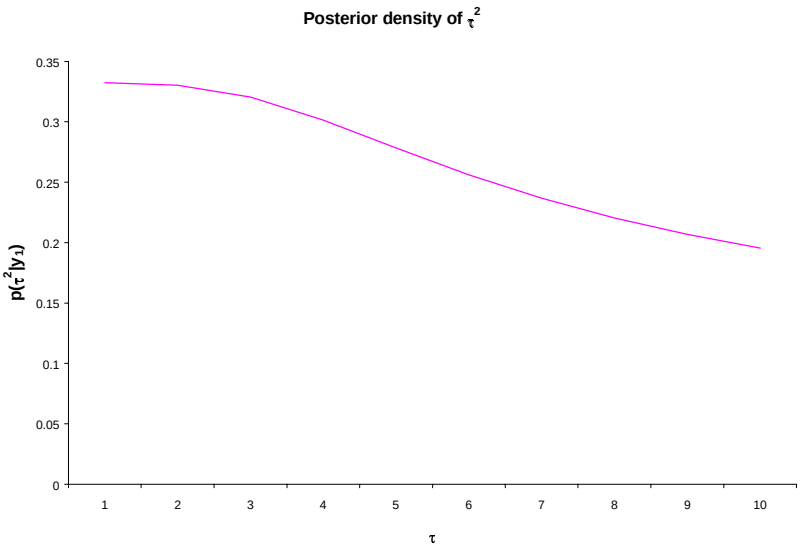


Figure 8.2. The posterior density of τ^2 given the true model contains the primary terms.

Table 8.1. Posterior probabilities of candidate models for three components with one process variable

Posterior Probability	Terms included in the candidate model
0.3008*	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3*
0.1481	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_1z
0.1481	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_2z
0.148	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_3z
0.073	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_1z x_2z
0.073	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_1z x_3z
0.073	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_2z x_3z
0.036	x_1 x_2 x_3 x_1x_2 x_1x_3 x_2x_3 x_1z x_2z x_3z

For this set of simulation data, twelve observations of a second stage design are selected. Figure 8.3 represents the second stage design points. Six points 3, 4, 7, and 9 estimate the coefficients of linear blending $\beta_1, \beta_2, \beta_3$; six points 1, 2, 5, 6, 8, and 10

estimate the coefficients of second-order terms β_{12}, β_{13}, and β_{23}; and two points 3 and 4 support estimation of the coefficient of interaction x_1z. Notice that there are more points to estimate primary terms in the second stage than in the first stage, and no points to estimate the coefficients of interaction x_2z and x_3z. This means that the second stage design is very efficient in selecting design points based on information from the first stage design, considering that the true model is a model of primary terms.

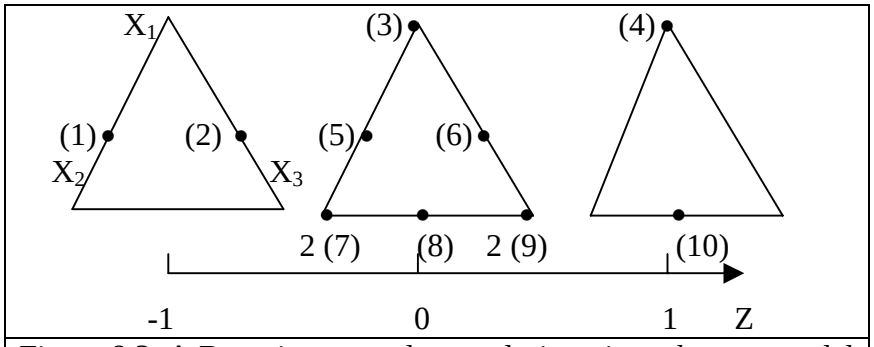


Figure 8.3. A Bayesian second stage design given the true model contains the primary terms (n_2=12)

8.2.2 Three Components with Two Process Variables

In this example, it is assumed that the experimenter defines the following regressors to consider:

$$x^{(f)} = \{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3, x_1z_1, x_2z_1, x_3z_1, x_1z_2, x_2z_2\ x_3z_2\}.$$

The primary terms are $\{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3\}$ with p=9, and $\{x_1z_1, x_2z_1, x_3z_1, x_1z_2, x_2z_2\ x_3z_2\}$ are potential terms with r=6. The most natural potential terms here are the interactions between mixture components and process variables. Often an uncertainty exists, and terms are important, because they completely dictate the

structure of the design. The true model in this example is set to contain the primary terms, and potential terms x_3z_1 and x_1z_2,

$$x^t = \{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3, x_3z_1, x_1z_2\}.$$

The first stage data set is simulated from the true model:

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2$$

$$+28x_1x_3+22x_2x_3+20x_3z_1+22x_1z_2+\varepsilon,$$

where $\varepsilon \sim N(0,1)$. Seventeen runs are allocated to each stage of the experiment, and the parameter $\tau = 1$ is used in the first stage.

The Bayesian D-optimality criterion is used to select the first stage design with seventeen observations that are based on the full model by

$$x^{(f)} = \{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3, x_1z_1, x_2z_1, x_3z_1, x_1z_2, x_2z_2, x_3z_2\}.$$

The resulting design is shown in Figure 8.4.

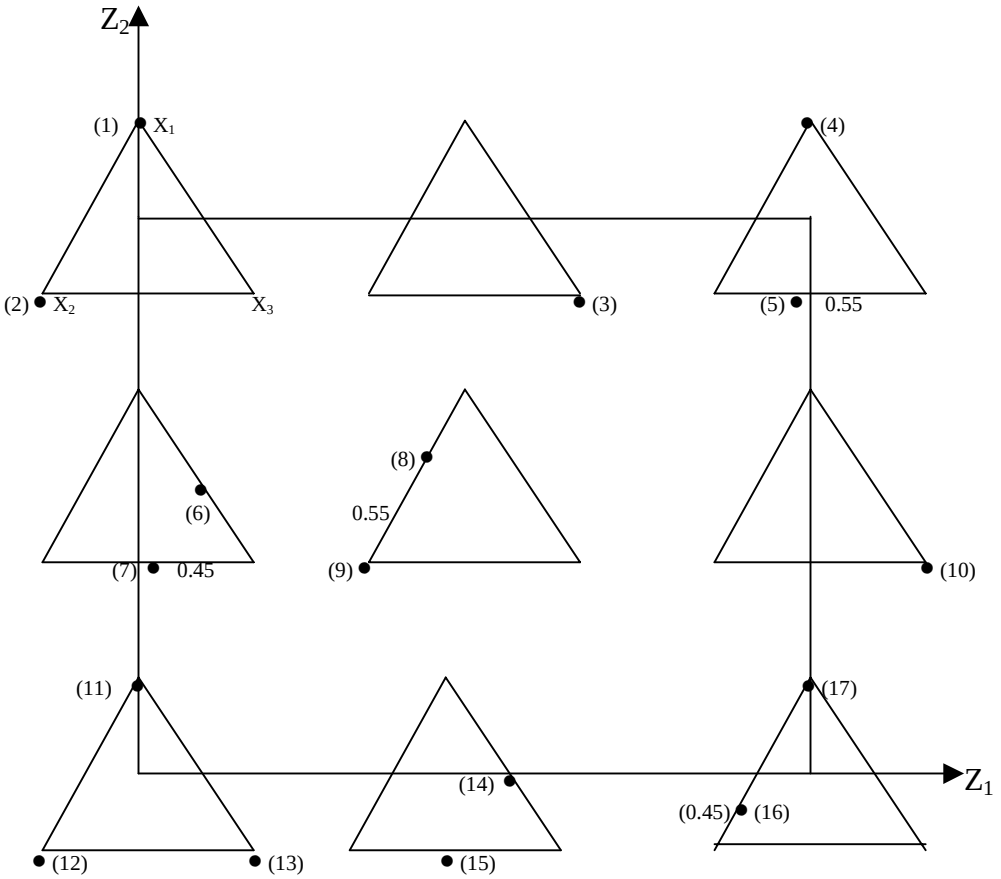


Figure 8.4. Bayesian first stage design for three components with two process variables.

In the first stage design, points 1, 2, 3, 4, 9, 10, 11, 12, 13, and 17 support to estimation of the coefficients of linear blending $\beta_1, \beta_2, \beta_3$; points 5, 6, 7, 8, 14, 15, and 16 support estimation of the coefficients of the second order terms $\beta_{12}, \beta_{13}, \text{and } \beta_{23}$; points 11 and 17 support estimation of the coefficient of z_1x_1; points 12 and 9 support estimation of the coefficient of x_2z_1; and points 13 and 10 support estimation of the coefficient of x_3z_1; points 11 and 1 support estimation of the coefficient of x_1z_2; points 12 and 2 support estimation of the coefficient of x_2z_2; points 13, 10, and 3 support estimation of the

coefficient of x_3z_2. Notice that all the parameters in the full model can be estimated. This is a reasonable design based on no preliminary information in the first stage. Seventeen runs of a second stage design are chosen. In this example, the density of posterior τ^2 and posterior probabilities for each possible model are not shown here, where $\hat{\tau}$ is large and the true model has the highest posterior probability. Figure 8.5 represents the most possible design. Points 1, 3, 15, 8, 7, 16, 2, 12, and 17 support estimation of the coefficients of linear blending $\beta_1, \beta_2, \beta_3$; points 4, 6, 14, 13, 9, 10, 5, and 11 support estimation of the coefficients of second order terms of $\beta_{12}, \beta_{13}, \text{and } \beta_{23}$; points 12, 15 and 16 support estimation of the coefficient of x_1z_1; points 7 and 8 support estimation of the coefficient of x_2z_1; points 1, 3, 2, and 17 support estimation of the coefficient of x_3z_1; points 15 and 16 support estimation of the coefficient of x_1z_2; points 8 and 7 support estimation of the coefficient of x_2z_2; and points 1, 3, 2, and 17 support estimation of the coefficient of x_3z_2. Notice that in the second stage, the coefficients of x_3z_1 and x_1z_2 are estimated well, since there are lowest and highest points of x_3 at lowest and highest z_1, and lowest and highest points of x_1 at lowest and highest z_2. This result suggests that the second stage design is very efficient in selecting design points based on information from the first stage design, considering that the true model contains the primary terms and potential terms x_3z_1 and x_1z_2.

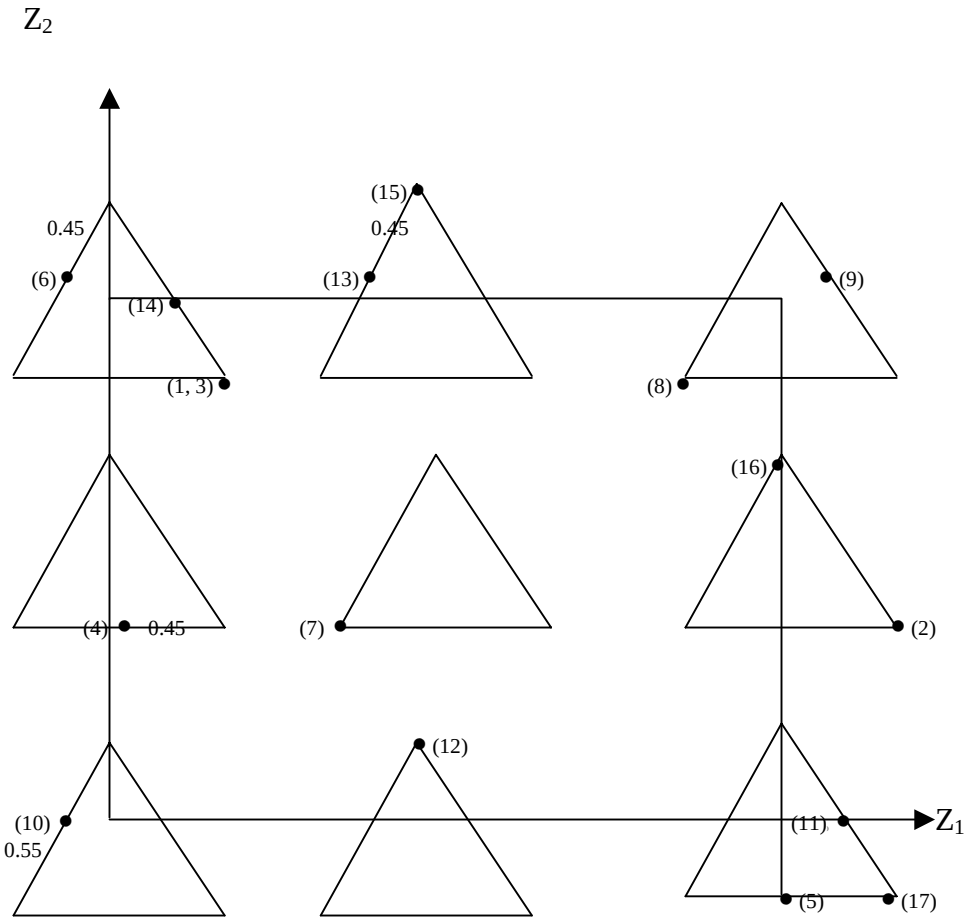


Figure 8.5. Bayesian second stage design for three components with two process variables.

8.3 Evaluation of the Designs

We want to evaluate, in general, our designs compared with two natural competitors, which are Bayesian one-stage D optimal design and one-stage non-Bayesian D-optimal design. In this evaluation, the performance of each design was measured by its determinant of information matrix D*, relative to the true models. The D* is defined as

$$D^* = \det(X_i'X_i).$$

Here, the X_i contains the $X_{1(i)}$ and $X_{2(i)}$ for the two-stage design in model M_i with the traditional scaling, expanded to include only those regressors of the true model. Notice that our measurement of performance has been changed here to avoid number that are exceeding small. Also note that "best design" is now the value with the largest criteria value. The performance of each Bayesian two-stage design is measured by its average of D* for each of the true models. Fifty data sets are simulated from each of the true models for the simulation purpose in this chapter. Therefore, the reported D* is actually the average of the D* from the simulation. Since single stage designs are not data dependent, they can be evaluated by a single D* for each of the true models. The results of evaluation of the Bayesian two-stage designs, one-stage Bayesian D-optimal design, and one-stage non-Bayesian D-optimal design are given in following sections.

8.4. Results of Bayesian Two-Stage D-D Designs

The results of evaluations of the Bayesian two-stage D-D design versus the Bayesian one-stage D-optimal designs and the non-Bayesian one-stage D-optimal designs for three examples of mixture experiments with process variables will be presented in the following section. We show that the Bayesian two-stage D-D optimal design is the most efficient design when the true model contains fewer regressors than the full model.

8.4.1. Three Components with One Process Variable

In this section, the full model defined by the experimenter is

$$x^{(f)} = \{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3, x_1z, x_2z, x_3z\}.$$

The primary terms are $\{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3\}$ with p=7, and $\{x_1z, x_2z, x_3z\}$ are the potential terms with r=3. Four cases are considered for simulation purposes, and the response data is simulated from each of the true models. The initial τ =1 is used in the first stage and $\varepsilon \sim N(0,1)$ is assumed.

Case 1: Primary terms only

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + \varepsilon$$

Case 2: Primary terms and one potential term

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 9x_2z + \varepsilon$$

Case 3: Primary terms and two of the potential terms

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 9x_2z + 10x_3z + \varepsilon$$

Case 4: Full model

$$y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 11x_1z + 9x_2z + 10x_3z + \varepsilon$$

In this section, the following three methods are compared:

1. Bayesian two-stage D-D optimal design with n_1=12 and n_2 =12,
2. Bayesian one stage D-optimal design with n=24, and
3. Non-Bayesian one stage design with n=24.

Table 8.2 presents the results of the evaluation for four different true models. The best design for each row is marked with an asterisk. The table shows that Bayesian D-D optimal design is the best for all of the subsets of the full model, but not the full model itself, which is the true model. Even if Bayesian D-D optimal design is not the best when true model is the full model, the Bayesian D-D optimal design is very efficient, since the D^* is very close to the best one.

Table 8.2. The values of D* for three components with one process variable

D*	Case 1	Case 2	Case 3	Case 4
D-D procedure	2.52*	7.01*	23.16*	69.88
Bayesian D procedure	2.08	6.11	20.04	70.92*
D procedure	2.08	6.11	20.04	70.92*

Case 1: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + \varepsilon$

Case 2: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 9x_2z + \varepsilon$

Case 3: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 9x_2z + 10x_3z + \varepsilon$

Case 4: $y = 7.24x_1 + 9.57x_2 + 5.66x_3 + 6.8z^2 + 11x_1z + 9x_2z + 10x_3z + \varepsilon$

8.4.2 Three Components with One Process Variable for Two Classes of Potential Terms

In this section, two classes of potential terms are considered for three components with one process variable. The results of the Bayesian two-stage D-D optimal design vs. the Bayesian one-stage D-optimal design and the non-Bayesian one-stage D-optimal design are also shown.

In this section, the full model defined by the experimenter contains the following regressors:

$$x^{(f)} = \{x_1, x_2, x_3, z^2, x_1x_2, x_1x_3, x_2x_3, x_1z, x_2z, x_3z\}.$$

The primary terms are $\{x_1, x_2, x_3, z^2\}$ where p=4, the first class of potential terms is $\{x_1x_2, x_1x_3, x_2x_3\}$ where $r_1=3$, and $\{x_1z, x_2z, x_3z\}$ is the second class of potential terms where $r_2=3$. Five cases are considered for simulation purposes and the response data set

is simulated from each of the true models. The error $\varepsilon \sim N(0,1)$ is assumed and initial $\tau = 1$ is used in the first stage.

Case 1: Primary terms only

$$y=17.24x_1+19.57x_2+15.66x_3+15z^2+\varepsilon$$

Case 2: Primary terms and one of the second class of potential terms

$$y=17.24x_1+19.57x_2+15.66x_3+15z^2+10x_1z+\varepsilon$$

Case 3: Primary terms and one from each class of potential terms

$$y=17.24x_1+19.57x_2+15.66x_3+15z^2+11x_1x_3+10x_1z+\varepsilon$$

Case 4: Primary terms and two terms from each class of potential terms

$$y=17.24x_1+19.57x_2+15.66x_3+15z^2+14x_1x_2+11x_1x_3+10x_1z+13x_3z+\varepsilon$$

Case 5: Full model

$$y=17.24x_1+19.57x_2+15.66x_3+15z^2+14x_1x_2+11x_1x_3+12x_2x_3+10x_1z+16x_2z+13x_3z+\varepsilon$$

In this section, the following three methods are compared.

1. Bayesian two-stage D-D optimal design with $n_1=12$ and $n_2=12$
2. Bayesian one-stage D-optimal design with n=24
3. NonBayesian one-stage design with n=24

Table 8.3 presents the D* and D-efficiency for Bayesian D-D optimal design and two natural competitors. This table shows that the Bayesian two-stage D-D optimal design is the best for all of the subsets of the full model, but not the full model itself, which is the true model. We can see that even if the true model is the full model, the Bayesian two-stage D-D design is still very efficient compared to two natural competitors, since the D-efficiency is close to one.

Table 8.3. The values of D* for three components with one process variable when full model contains two classes of potential terms

D*	Case 1	Case 2	Case 3	Case 4	Case5
D-D procedure	1656.52*	11574.37*	1813.2*	651.257*	430.15
Bayesian D	1190.13	5951.34	874.12	627.71	440.3*
D procedure	1190.13	5951.34	874.12	627.71	440.3*

Case 1: $y=17.24x_1+19.57x_2+15.66x_3+15z^2+\varepsilon$

Case 2: $y=17.24x_1+19.57x_2+15.66x_3+15z^2+10x_1z+\varepsilon$

Case 3: $y=17.24x_1+19.57x_2+15.66x_3+15z^2+11x_1x_3+10x_1z+\varepsilon$

Case 4: $y=17.24x_1+19.57x_2+15.66x_3+15z^2+14x_1x_2+11x_1x_3+10x_1z+13x_3z+\varepsilon$

Case 5: $y=17.24x_1+19.57x_2+15.66x_3+15z^2+14x_1x_2+11x_1x_3+12x_2x_3+10x_1z+16x_2z+13x_3z+\varepsilon$

8.4.3. Three Components with Two Process Variables

In this section, the full model defined by the experimenter is

$$x^{(f)} = \{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3, x_1z_1, x_2z_1, x_3z_1, x_1z_2, x_2z_2, x_3z_2\}.$$

The primary terms are $\{x_1, x_2, x_3, z_1^2, z_2^2, z_1z_2, x_1x_2, x_1x_3, x_2x_3\}$ with p=9, and $\{x_1z_1, x_2z_1, x_3z_1, x_1z_2, x_2z_2, x_3z_2\}$ are potential terms with r=6. Five cases are considered for simulation purposes, and the response data is simulated from each of the true models. The initial τ =1 is used in the first stage and $\varepsilon \sim N(0,1)$ is assumed.

Case 1: Primary terms only:

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+\varepsilon$$

Case 2: Primary terms and one potential term:

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+18x_1z_1+\varepsilon$$

Case 3: Primary terms and two of the potential terms:

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3 +20x_3z_1+22x_1z_2+\varepsilon$$

Case 4: Primary terms and two of the potential terms:

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+ 20x_3z_1+22x_1z_2+24.5x_3z_2+\varepsilon$$

Case 5: Full model

$$y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+ 18x_1z_1+19.8x_2z_1+20x_3z_1+22x_1z_2+21.5x_2z_2+24.5x_3z_2+\varepsilon$$

In this section, the following three methods are compared:

1. Bayesian two-stage D-D optimal design with $n_1=17$ and $n_2=17$,
2. Bayesian one stage D-optimal design with n=34, and
3. Non-Bayesian one stage design with n=34.

Table 8.4 presents the results of the evaluation for five different true models. The best design for each row is marked with an asterisk. The table shows that Bayesian two-stage D-D optimal design is the best for all of the subsets of the full model, but not the full model itself, which is the true model. Even if Bayesian two-stage D-D optimal design is not the best when true model is the full model, Bayesian D-D optimal design is still very efficient, since D^* is very close to the best one.

Table 8.4. The values of D* for three components with two process variables

D*	Case 1	Case 2	Case 3	Case 4	Case5
D-D procedure	264.91*	1186.52*	6865.36*	39567.24*	2993950
Bayesian D	153.24	797.9	4094.26	20754.65	2903516
D procedure	153.24	829.87	4249.1	21524.52	3078808*

.

Case 1: $y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+\varepsilon$

Case 2: $y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+$

$22x_2x_3+18x_1z_1+\varepsilon$

Case 3: $y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3$

$+20x_3z_1+22x_1z_2+\varepsilon$

Case 4: $y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+$

$20x_3z_1+22x_1z_2+24.5x_3z_2+\varepsilon$

Case 5: $y=17.24x_1+19.57x_2+15.66x_3+15z_1^2+14.2z_2^2+20z_1z_2+23x_1x_2+28x_1x_3+22x_2x_3+$

$18x_1z_1+19.8x_2z_1+20x_3z_1+22x_1z_2+21.5x_2z_2+24.5x_3z_2+\varepsilon$

8.5. Conclusion

For mixture models with model uncertainty, Bayesian two-stage D-D optimal design is the best when the true model contains fewer regressors than the full model, and Bayesian two-stage D-D optimal design is a little less efficient than the non-Bayesian one-stage design when true model is the full model. The performance of the Bayesian two-stage D-D optimal design for a mixture model with process variables is the similar as the

performance of the Bayesian two-stage D-D optimal design for a mixture model. If the experimenter knows the right model, one-stage D-optimal design should be used; otherwise, it is better to use Bayesian two-stage D-D optimal design, which has been proven to be a very robust procedure. Therefore, we believe that Bayesian two-stage D-D optimal design is a good choice under model uncertainty for mixture experiments with process variables.

Chapter 9

D_s-Optimal Designs for Mixture Models

While the effects of individual components are important for mixture models, the experimenter is often more interested in how the combination of the components works to produce a particular response. As a result, second or higher order effects can be more important than main effects. This chapter presented a discussion of D-optimal second or higher order designs, which result in parameter estimates as precise as possible in that the variance of second or higher order terms estimate is minimized. Such a D-optimal second or higher order design is denoted by a D_s-optimal design. Notice s stands for "subset" of parameters in the model, and in our case s stands for second-order or higher order terms. When the comparisons of the designs are made, the results show that the D_s-optimal design has almost the same efficiency as the D-optimal design. Meanwhile, the D_s-optimal design has a higher D_s-efficiency than the D-optimal design in most cases. Furthermore, the D_s–optimal design is superior to the D-optimal design in terms of the prediction variance along certain directions inside the region of the experiment.

9.1 Criteria for D_s-optimal Designs

Consider a mixture model in standard form, $y = X\beta + \varepsilon$, where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & . & . & . & x_{1k} \\ x_{21} & . & . & . & x_{2k} \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ x_{n1} & . & . & . & x_{nk} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_k \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}.$$

Let $X=[X_1|X_2]$, where X_1 contains linear blending terms, and X_2 contains second or higher order terms, then $M(\beta) = X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M'_{12} & M_{22} \end{bmatrix}$. In order to minimize the generalized variance of the coefficients of second or higher order terms, $M^{-1}(\beta)$ must be found in terms of its partitioned form. This is given by

$$M^{-1}(\beta) = \begin{bmatrix} C_{11}^{-1} & M_{11}^{-1}M_{12}C_{22} \\ C_{22}M`_{12}M^{-1}_{11} & C^{-1}_{22} \end{bmatrix},$$

where $C_{11}^{-1} = (M_{11}-M_{12}M^{-1}_{22}M`_{12})^{-1}$ and $C_{22}^{-1} = (M_{22}-M`_{12}M^{-1}_{11}M_{12})^{-1}$. Therefore, the D_s-optimal design is a subset of candidate points by maximizing the determinant of $(M_{22}-M`_{12}M^{-1}_{11}M_{12})$.

9.2 D_s-Optimal Designs for Three Components with Second-Order Model.

In this example, it is assumed that the experimenter defines regressors:

$$x = \{ x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3\}.$$

In this example, the number of parameters for the model k is equal to six, and the number of parameters of second order s is equal to three. Different sample sizes 6, 7, 9, 12, 14, and 16 are used for D_s–optimal designs and D-optimal designs. The D_s- and D-optimal designs for different sample sizes are presented in Figure 9.1. Notice that there are more

middle points of edges in D_s–optimal designs than in D-optimal designs for some sample sizes. This is expected since these types of points are the ones that enhance estimation of second order terms.

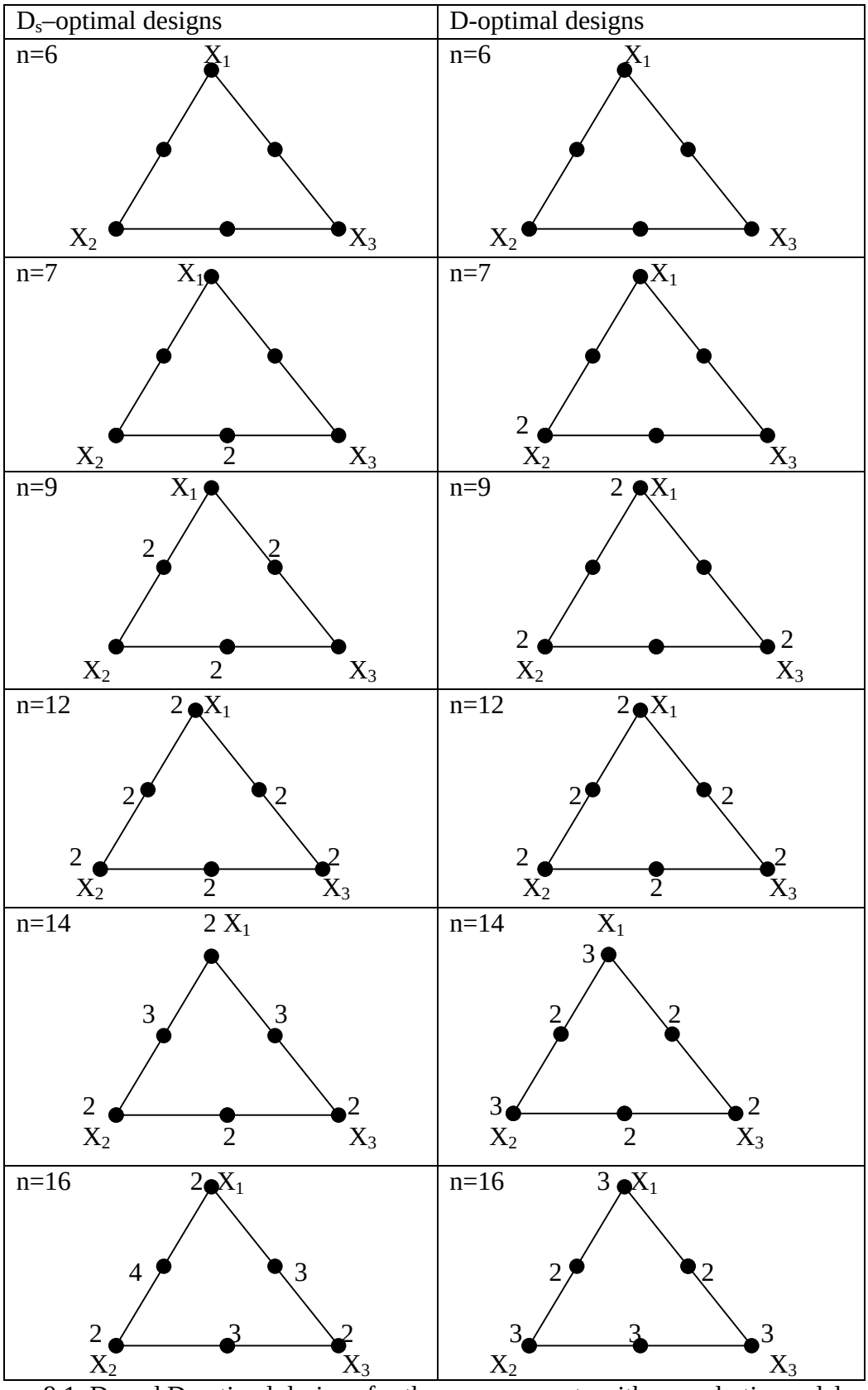


Figure 9.1. D_s and D optimal designs for three components with a quadratic model.

Since the D-optimal and D_s–optimal designs are generated by maximizing $D=|(X^{'}X)|$ and $|(M_{22}-M^{`}_{12}M^{-1}_{11}M_{12})|$ respectively, a larger value of D means the design is more efficient for estimating parameters. Therefore, the larger values of |(X'X)| and $|(M_{22}-M_{12}'M_{11}^{-1}M_{12})|$ are preferred. Comparisons of |(X`X)| and $|(M_{22}-M_{12}^{`}M_{11}^{-1}M_{12})|$ are presented in Tables 9.1 and 9.2. In Table 9.1, the determinants of the model information matrix are presented. Note that the best design in the column is marked with an asterisk. We can see that there are no differences between D_s–optimal designs and D-optimal designs for sample sizes 6, 7, 9, 12, and 14, and the D-optimal design is a little bit more efficient than D_s–optimal design for sample size 16. Table 9.2 shows the determinants of the second order information matrix. Notice that there are no differences between the D_s–optimal designs and the D-optimal designs for sample size 6 and 12, and the D_s-optimal designs are more efficient in second order estimates than the D-optimal designs for the sample sizes 7, 9, 14, and 16. For a stronger indication of which designs should be used between D_s–optimal design and D-optimal design for mixture experiments, more examples are investigated. Notice that D-optimal design is not unique.

Table 9.1. The determinant of D=|(X`X)| for three components with second-order model between D_s- and D-optimal designs

n	6	7	9	12	14	16
D_s	0.000244*	0.0005*	0.00195*	0.0156*	0.03515*	0.0703
D	0.000244*	0.0005*	0.00195*	0.0156*	0.03515*	0.079*

Table 9.2. The determinant of $|(M_{22}-M_{12}^{`}M_{11}^{-1}M_{12})|$ for three components with second-order model between D_s- and D-optimal designs

n	6	7	9	12	14	16
D_s	0.000078*	0.00012*	0.00029*	0.000625*	0.0011*	0.00165*
D	0.000078*	0.00009	0.000128	0.000625*	0.00077	0.00116

9.3 D_s-Optimal Design for Four Components with Second-Order Model

In this example, it is assumed that the experimenter defines regressors as follows:

$$x = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}.$$

In this case, the number of parameters in the model k is ten and the number of second order terms is six. The D_s and D-optimal designs are presented in Figure 9.2 for different sample sizes 12, 14, 20 and 40. Notice that there are more middle points of edges in the D_s–optimal designs than in D-optimal designs for some sample sizes.

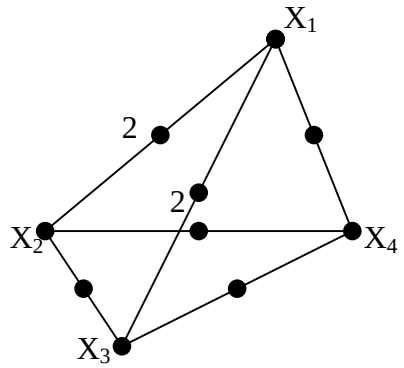
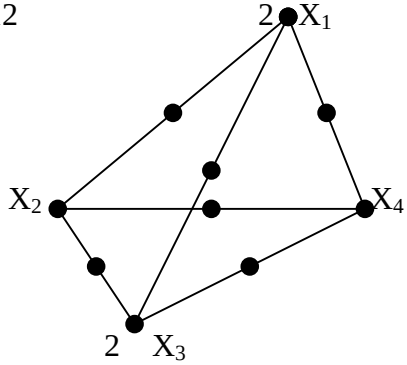
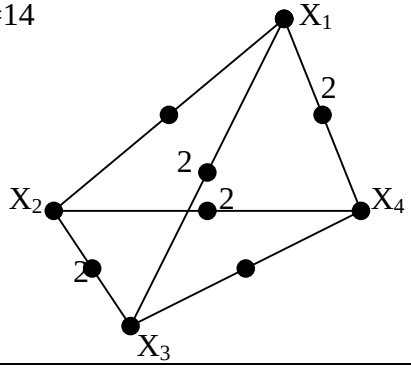
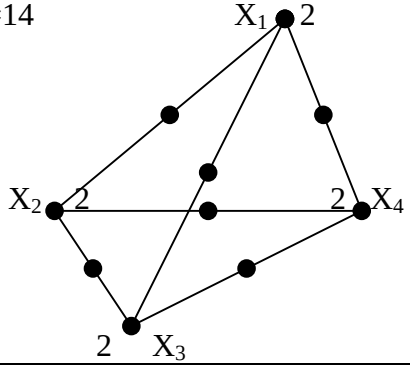
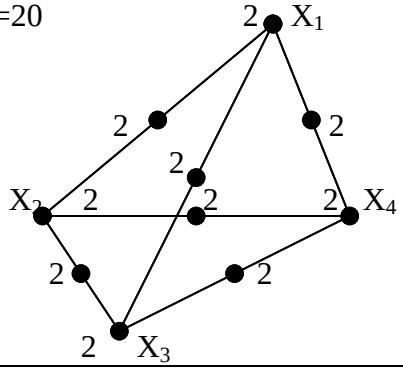
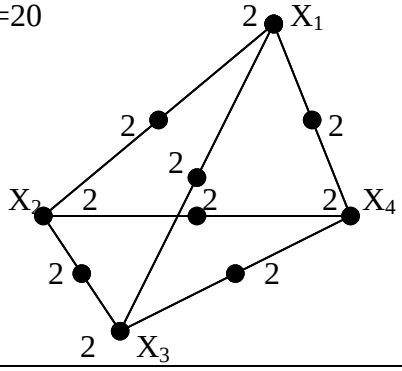
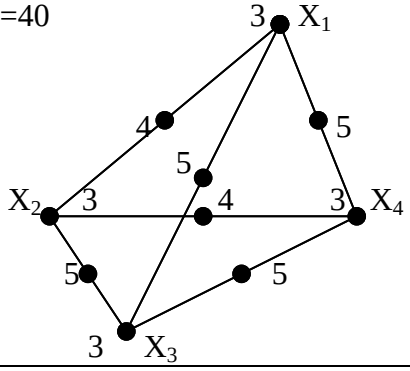
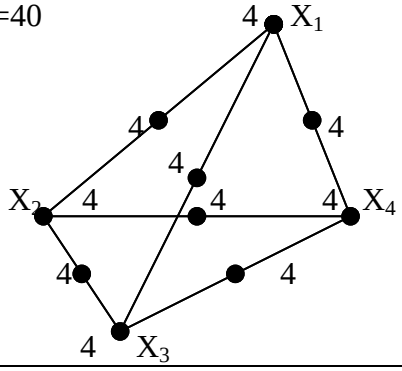
D_s-optimal designs	D-optimal designs
n=12 	n=12 
n=14 	n=14 
n=20 	n=20 
n=40 	n=40 

Figure 9.2. D_s- and D-optimal designs for four components with a quadratic model.

The comparisons of D_s–optimal designs and D-optimal designs are presented in Tables 9.3 and 9.4. In Table 9.3, the determinants of the model information matrix are calculated. Notice that the best design for each column is marked with an asterisk. There are no differences between D_s–optimal designs and D-optimal designs for sample sizes 12, 14, 20, and the D is a bit more efficient for sample size 40. In Table 9.4, the determinants of the second order information matrix are presented. There is no difference between the D_s–optimal design and the D-optimal design for sample size 20, and the D_s–optimal designs are more efficient for second order estimates than the D-optimal designs for the sample sizes 12, 14, and 40.

Table 9.3. The determinant of D=|(X`X)| for four components with second-order model between D_s- and D-optimal designs

n	12	14	20	40
D_s	2.384e-7*	9.537e-7*	0.000061*	0.051
D	2.384e-7*	9.537e-7*	0.000061*	0.0625*

Table 9.4. The determinant of ($|(M_{22}-M_{12}'M_{11}^{-1}M_{12})|$) for four components with second-order model between D_s- and D-optimal designs

n	12	14	20	40
D_s	1.78e-8*	4.541e-8*	4.521e-7*	0.000319*
D	1.105e-8	1.74e-8	4.521e-7*	0.0000289

9.4 D_s-Optimal Designs for Four Components with Special Cubic Model

In this example, the experimenter defines model regressors as follows:

$$x^{(f)} = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4\}.$$

In this example, the number of parameters in the model is equal to 14, and the number of second and cubic terms is equal to 10. Figure 9.3 below shows the D_s–optimal designs

and D-optimal designs for different sample sizes 16, 20, 30 and 40. Notice that there are more points of middle points of edges and centroids of planes in D_s-optimal designs than in D-optimal designs for some sample sizes.

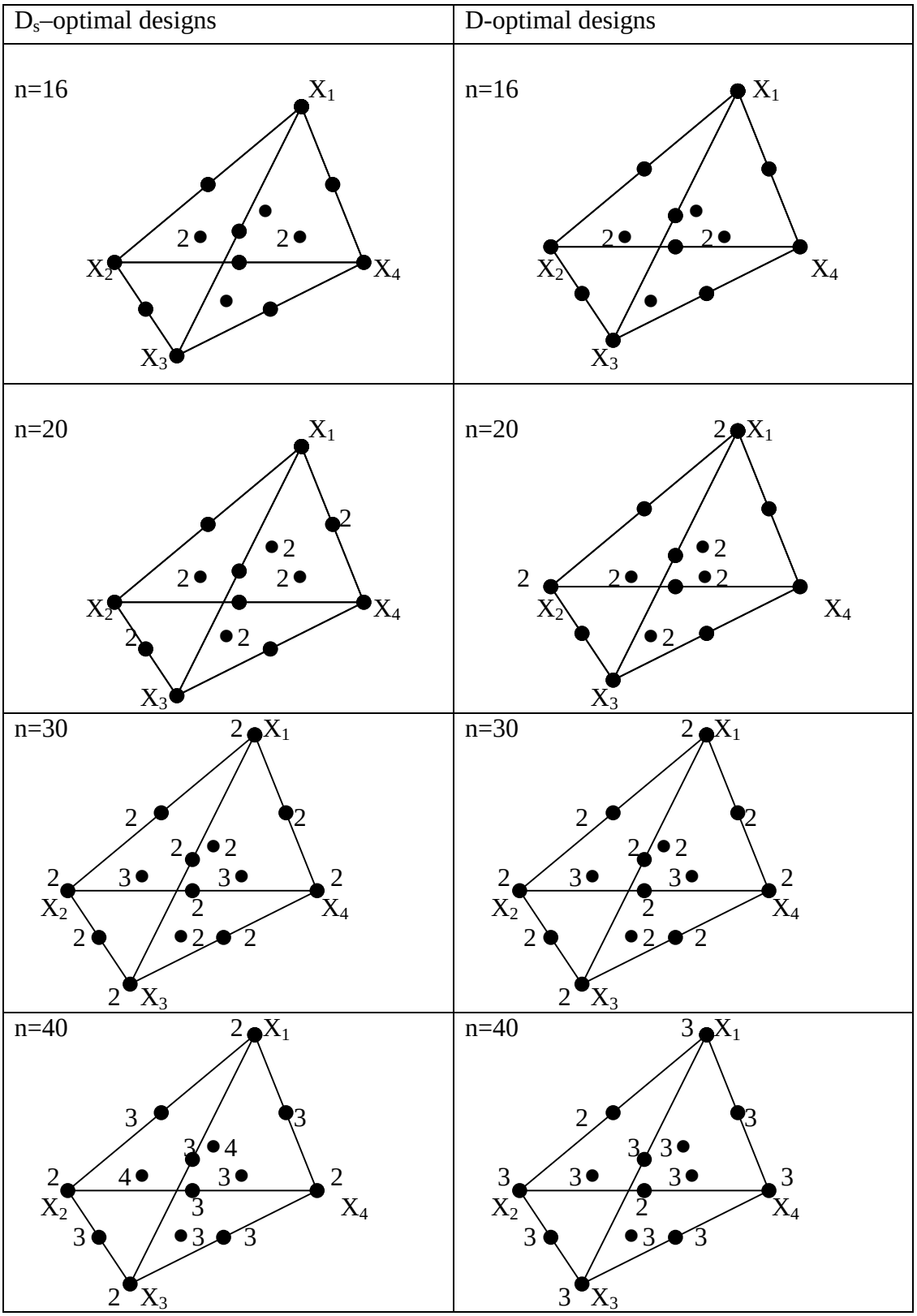


Figure 9.3. D_s and D-optimal designs for four components with a cubic model.

The comparisons of D_s-optimal designs and D-optimal designs are presented in Tables 9.5 and 9.6. Table 9.5 presents the determinants of the model information matrix. There are no differences between the D_s and D-optimal designs for sample sizes 16 and 30, and D-optimal design is a bit more efficient than D_s-optimal design for sample size 20 and 40. Table 9.6 presents the determinants of the second and cubic term information matrix. There are no differences between D_s-optimal designs and D-optimal designs for sample sizes 16 and 30, and D_s–optimal designs are more efficient in higher order estimates than D-optimal designs for sample sizes 20 and 40.

Table 9.5. The determinant D=|(X`X)| for four components with special cubic model between D_s- and D-optimal designs

n	16	20	30	40
D_s	8.241e-19*	1.319e-17	7.595e-15*	3.46e-13
D	8.241e-19*	1.32e-17*	7.595e-15*	4.38e-13*

Table 9.6. The determinant D=|(M_{22}-M_{12}`$M_{11}^{-1}M_{12}$)| for four components with special cubic model between D_s- and D-optimal designs

n	16	20	30	40
D_s	4.48e-20*	4.01e-19*	2.883e-17*	5.9e-16*
D	4.48e-20*	2.59e-19	2.883e-17*	4.33e-16

Based on the above results, we can see that D_s-optimal design is better than D-optimal design, since the D_s–optimal design is almost as efficient as the D-optimal design for model coefficient estimates, and D_s-optimal design is more efficient than D-optimal design in second or higher order estimates.

9.5 Comparing Designs between D_s and D through Plots along Prediction Rays

The experimenter is often interested in certain directions inside the experimental region. In particular, Hare (1985) and Cornell (1990) illustrated this using the response trace, which is a plot of predicted response values taken along certain directions that strengthening a 'reference' mixture by increasing the proportion of a single component at a time in the mixture, i.e. along the so-called Cox (1971) directions. To illustrate the Cox directions, Vining, Cornell, and Myers (1993) define a reference mixture $s = (s_1, s_2, ..., s_q)'$, where $\sum_{i=1}^{q} s_i = 1$. Typically, s will be the centroid of the region or the blend whose co-ordinates are the averages of the co-ordinates of the extreme vertices of the region of interest. Suppose that another blend is defined where the proportion of the i^{th} component is changed by an amount Δ_i (where $\Delta_i > 0$ or $\Delta_i < 0$), for example,

$$x_i = s_i + \Delta_i .$$

Adjusting the proportions of the remaining components in the new blend by using

$$x_j = s_j - \frac{\Delta_i s_j}{1 - s_i}, \; j = 1, 2, ..., q, \; j \neq i$$

ensures that the relative proportions of any two components j and k ($j \neq i$, $k \neq i$) in the new blend remain the same as their ratio at the reference mixture, i.e. $x_j/x_k = s_j/s_k$ $k \neq i$, which is known as Cox's direction when viewed from s to the $x_i=1$ vertex. Hereafter, Cox directions are referred to as 'prediction rays' because points (or blends) inside the experimental region along these rays are defined and the fitted model is used to predict the response at the points. Since there is a prediction ray for each component, there is a separate response trace for each component. Therefore, we will study the performance of D_s–optimal design and D-optimal design through plots along prediction rays. The

purpose of Mclean and Anderson's (1966) flare experiment was to find the combination of the proportions of magnesium (x_1), sodium nitrate (x_2), strontium nitrate (x_3) and binder (x_4) for creating flares with maximum illumination. Engineering experience indicates that the following constraints should be placed on each component: $0.4 \le x_1 \le 0.6$, $0.1 \le x_2 \le 0.5$, $0.1 \le x_3 \le 0.5$, $0.03 \le x_4 \le 0.08$, and defines the following regressors with k=10 and s=6:

$$x = \{x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4\}.$$

Comparisons of the D-optimal design and D_s–optimal design are presented in Tables 9.7 and 9.8, respectively. The prediction variance ($\sigma^2 x_0'(X'X)^{-1}x_0$) along rays at x_0 (the points along the prediction rays) for 15-point design employed by D-optimal design and D_s-optimal design is presented in Figure 9.4. In these traces, D_s-optimal design is clearly superior to the D-optimal design in most values of x's, because the D_s–optimal design performs significantly better than the D-optimal design over the majority of the ranges of these traces. Only at the boundaries of the x_4 region does the D-optimal design seem better, yet the improvement is rather marginal. Therefore, the D_s –optimal designs mainly perform better than the D-optimal design in terms of the prediction variance along rays.

Table 9.7. D-optimal design for four components with a second-order model (n=15)

x_1	x_2	x_3	x_4
0.4	0.1	0.47	0.03
0.6	0.1	0.27	0.03
0.4	0.47	0.1	0.03
0.6	0.27	0.1	0.03
0.4	0.1	0.42	0.08
0.4	0.42	0.1	0.08
0.6	0.1	0.22	0.08
0.6	0.22	0.1	0.08
0.5	0.345	0.1	0.055
0.4	0.285	0.285	0.03
0.4	0.26	0.26	0.08
0.6	0.1	0.245	0.055
0.32637	0.14523	0.4925	0.0359
0.4	0.2725	0.2725	0.055
0.32637	0.4925	0.14523	0.0359

Table 9.8. D_s-optimal design for four components with a second-order model (n=15)

X_1	X_2	X_3	X_4
0.4	0.1	0.47	0.03
0.6	0.1	0.27	0.03
0.4	0.47	0.1	0.03
0.6	0.27	0.1	0.03
0.4	0.1	0.42	0.08
0.4	0.42	0.1	0.08
0.6	0.1	0.22	0.08
0.6	0.22	0.1	0.08
0.5	0.345	0.1	0.055
0.4	0.285	0.285	0.03
0.4	0.26	0.26	0.08
0.5	0.1	0.345	0.055
0.32637	0.14523	0.4925	0.0359
0.4	0.2725	0.2725	0.055
0.32637	0.4925	0.14523	0.0359

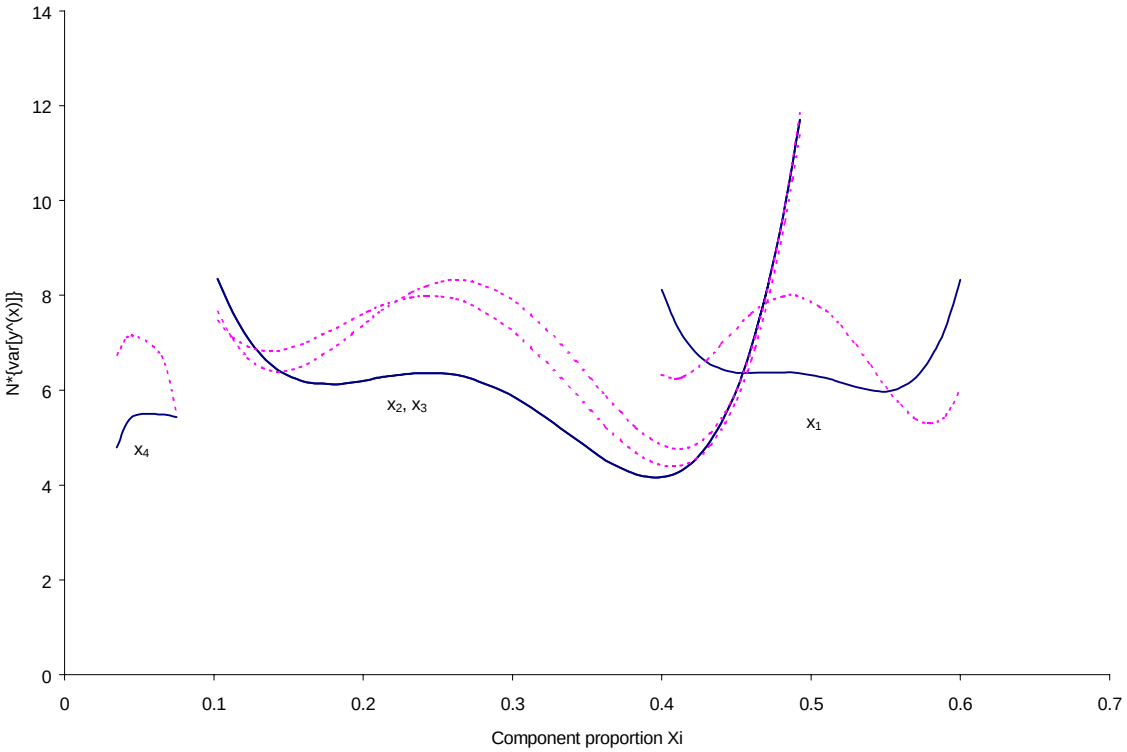


Figure 9.4. Plot of the prediction variance traces for the D (------) and D_s (———)designs: x_1, magnesium; x_2, sodium nitrate; x_3, strontium nitrate; x_4, binder.

To further investigate D_s–optimal design and D-optimal designs along prediction rays, another example is considered. The purpose of this example is to find the optimal combination of x_1, x_2, x_3, x_4, x_5 and x_6. The constraints for each component proportion are following: $x_1 \leq 0.2$, $x_2 \leq 0.6$, $x_3 \leq 0.15$, $x_4 \leq 0.65$, $x_5 \leq 0.15$, and $x_6 \leq 0.75$. The following regressors are defined:

$$x = \{x_1, x_2, x_3, x_4, x_5, x_6, x_1x_2, x_1x_3, x_1x_4, x_1x_5, x_1x_6, x_2x_3 , x_2x_4,$$

$$x_2x_5 , x_2x_6, x_3x_4 x_3x_5, x_3x_6, x_4x_5, x_4x_6, x_5x_6\}.$$

The designs of D and D_s are presented in Tables 9.7 and Table 9.8, respectively. The prediction variances along rays for 38-point design employed by D-optimal design

and D_s-optimal design are presented in Figures 9.5.1, 9.5.2, 9.5.3, 9.5.4, 9.5.5, and 9.5.6, where (------) represent D-optimal designs and (———) represent D_s–optimal designs. Looking at these traces, the D_s–optimal design is clearly superior in terms of the prediction variance trace. Only for large values of x_2 and x_5, the D-optimal design seems better, yet the improvement is rather marginal.

In sections 9.2, 9.3, and 9.4, we see that the D_s–optimal designs are more efficient in second or higher order coefficient estimates than the D-optimal designs, without sacrificing D-efficiency in all coefficient estimates. Furthermore, in this section we see that the prediction variances along rays in the D_s–optimal design is smaller than the prediction variances along rays in the D-optimal designs over the majority of ranges of these traces. Also, Wardrop and Myers (1990) found that the prediction variance along the rays is influenced more by higher order terms than by linear terms. Therefore, we may suggest that D_s–optimal designs can be used instead of D-optimal designs for mixture experiments in terms of variance of coefficients of second or higher order terms and variance of prediction variance along rays.

Table 9.9. D-optimal design for six components with second-order model (n=38)

x_1	x_2	x_3	x_4	x_5	x_6
0	0	0	0.25	0	0.75
0	0	0	0.65	0	0.35
0	0	0	0.65	0.15	0.2
0	0	0.1	0	0.15	0.75
0	0	0.15	0.28	0.08	0.49
0	0	0.15	0.65	0	0.2
0	0.05	0.15	0.65	0.15	0
0	0.1	0	0	0.15	0.75
0	0.1	0.15	0	0	0.75
0	0.25	0	0	0	0.75
0	0.275	0	0.65	0.075	0
0	0.35	0	0.65	0	0
0	0.6	0	0	0	0.4
0	0.6	0	0.25	0.15	0
0	0.6	0.075	0	0.075	0.25
0	0.6	0.075	0.325	0	0
0	0.6	0.15	0	0.15	0.1
0	0.6	0.15	0.25	0	0
0.08182	0.24545	0	0.21364	0.15	0.30909
0.0852	0.22752	0.06835	0.265	0.06835	0.28557
0.1	0	0	0	0.15	0.75
0.1	0	0.15	0	0	0.75
0.1	0.6	0	0	0	0.3
0.125	0	0.075	0.65	0.15	0
0.15	0.6	0.15	0	0.1	0
0.2	0	0	0	0.1	0.7
0.2	0	0	0.05	0	0.75
0.2	0	0	0.65	0.15	0
0.2	0	0.15	0	0.15	0.5
0.2	0	0.15	0.575	0.075	0
0.2	0	0.15	0.65	0	0
0.2	0.15	0	0.65	0	0
0.2	0.25	0.07	0	0	0.48
0.2	0.25	0.15	0.25	0.15	0
0.2	0.6	0	0	0.15	0.05
0.2	0.6	0	0.2	0	0
0.2	0.6	0.05	0	0.15	0
0.2	0.6	0.15	0	0	0.05

Table 9.10. D_s-optimal design for six components with a second-order model (n=38)

x_1	x_2	x_3	x_4	x_5	x_6
0	0	0	0.25	0	0.75
0	0	0	0.65	0	0.35
0	0	0	0.65	0.15	0.2
0	0	0.1	0	0.15	0.75
0	0	0.15	0.28	0.08	0.49
0	0	0.15	0.65	0	0.2
0	0.05	0.15	0.65	0.15	0
0.2	0	0.07	0.27	0	0.46
0	0.35	0.15	0	0	0.5
0	0.25	0	0	0	0.75
0	0.275	0	0.65	0.075	0
0	0.275	0.075	0.65	0	0
0	0.6	0	0	0	0.4
0	0.6	0	0.25	0.15	0
0	0.6	0.075	0	0.075	0.25
0	0.6	0	0.4	0	0
0	0.6	0.15	0	0.15	0.1
0	0.6	0.15	0.25	0	0
0.09583	0.24583	0	0.24375	0.07083	0.34375
0.09583	0.24583	0.07083	0.24375	0	0.34375
0.1	0	0	0	0.15	0.75
0.1	0	0.15	0	0	0.75
0.1	0.6	0	0	0	0.3
0.125	0	0.075	0.65	0.15	0
0.1	0.6	0.15	0	0.15	0
0.2	0	0	0	0.1	0.7
0.2	0.05	0	0	0	0.75
0.2	0	0	0.65	0.15	0
0.2	0	0.15	0	0.15	0.5
0.125	0	0.15	0.65	0.075	0
0.2	0	0.15	0.65	0	0
0.2	0.15	0	0.65	0	0
0	0.35	0	0	0.15	0.5
0.2	0.25	0.15	0.25	0.15	0
0.2	0.6	0	0	0.15	0.05
0.2	0.6	0	0.2	0	0
0.2	0.6	0.1	0	0.1	0
0.2	0.6	0.15	0	0	0.05

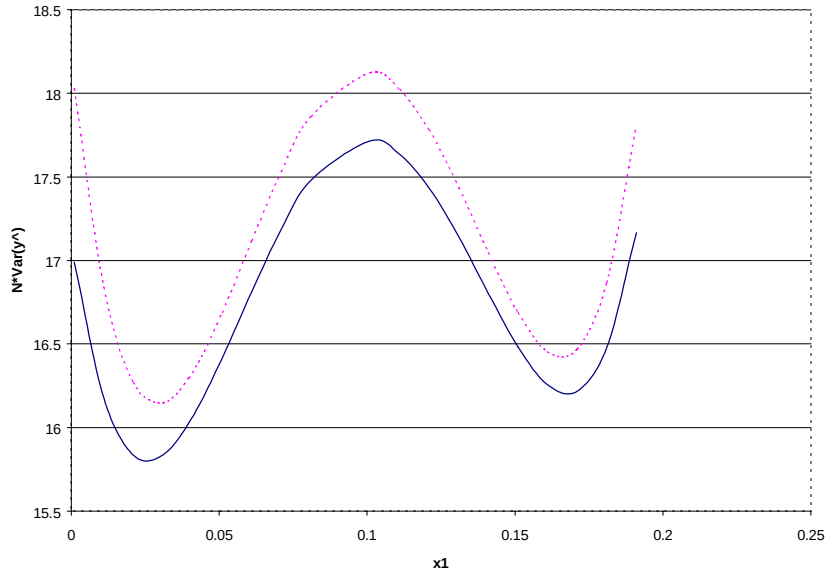


Figure 9.5.1. The prediction variance along X_1, where (⋯) represents D-optimal design and (____) represents D_s-optimal design.

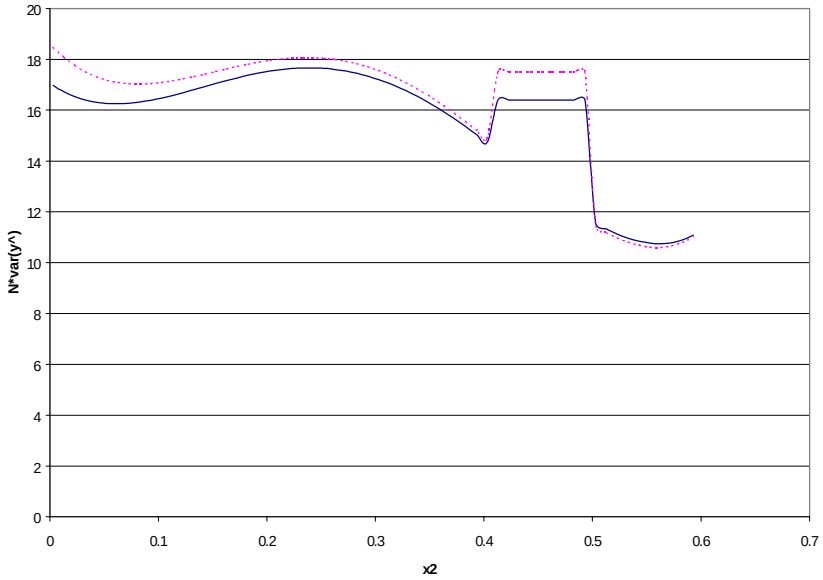


Figure 9.5.2. The prediction variance along X_2, where (⋯) represents D-optimal design and (──) represents D_s-optimal design.

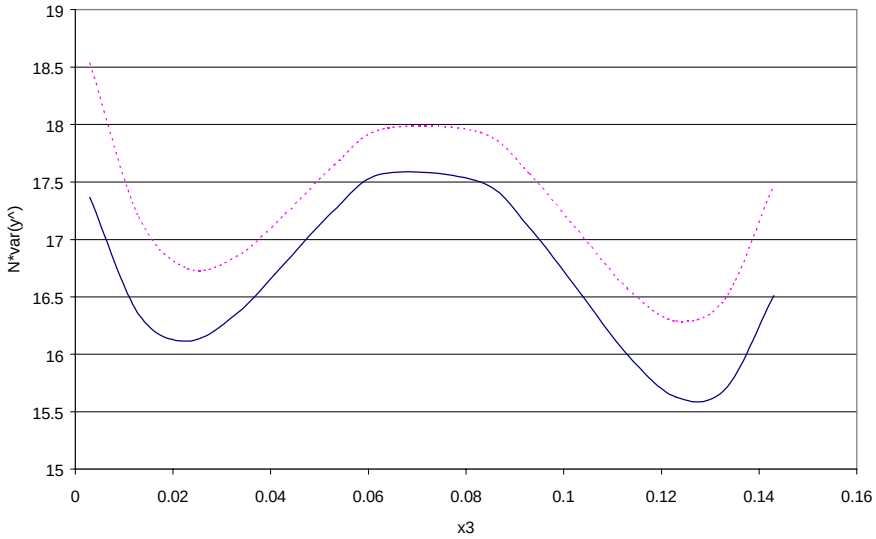


Figure 9.5.3. The prediction variance along X_3, where (⋯) represents D-optimal design and (—) represents D_s-optimal design.

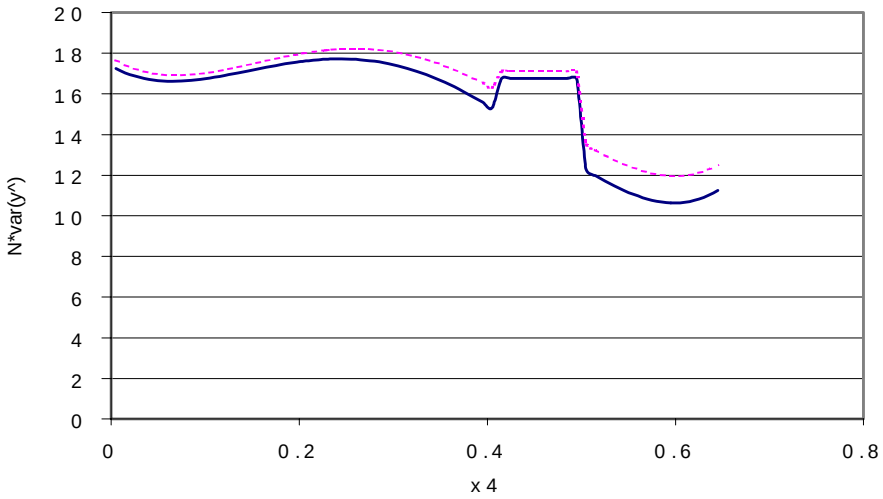


Figure 9.5.4. The prediction variance along X_4, where (⋯) represents D-optimal design and (—) represents D_s-optimal design.

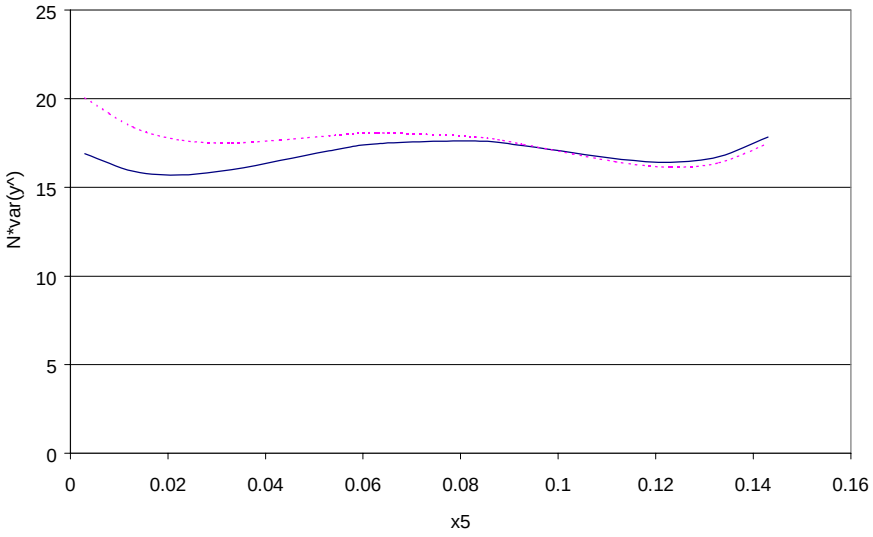


Figure 9.5.5. The prediction variance along X_5, where (⋯) represents D-optimal design and (—) represents D_s-optimal design.

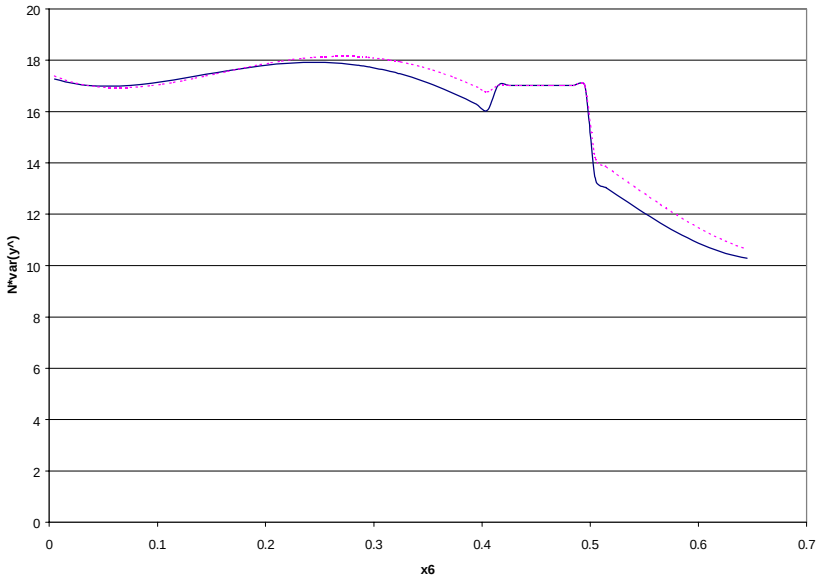


Figure 9.5.6. Prediction variance long X_6, where (⋯) represents D-optimal design and (—) represents D_s-optimal design.

Chapter 10

Discussions and Future Research

The primary goal of this research was to develop and study Bayesian two-stage D-D optimal designs for mixture models with or without process variables under model uncertainty. The secondary goal is to study the performance of D_s-optimal designs for mixture models.

In this dissertation, the Bayesian two-stage D-D procedure was developed. The motivation for the Bayesian approach comes from a need to deal with model uncertainty. The goal of two-stage design is to gain model information from the first stage data; the second stage design then takes into account the improved model information gained from the first stage data. The benefit of the using the Bayesian two-stage procedure for mixture models lies in the fact that there is likely more model uncertainty in mixture models, therefore, the use of prior information through Bayesian approach and an initial stage can be very beneficial. The complex model uncertainty results from the fact that higher order terms in mixture experiments (including cubic terms) are more complicated in the mixture situation, due to the use of canonical polynomial models. Thus, a priori

anticipation of model structure is difficult for users. We presented the results of an evaluation of the performance of Bayesian two-stage D-D optimal designs for mixture models with or without process variables when compared to two natural competitors that do not involve a two-stage approach. The results show that the Bayesian two-stage D-D optimal designs are quite appropriate for mixture models. We also discussed the idea of a reasonable ratio between the sample sizes for two stages and what is a sufficient amount of data for the first stage. We found that the ratio 1:1 is reasonable for the two stages and that a sufficient amount of data for the first stage should involve at least p+r+2 design points, where p and r are the number of primary terms and potential terms, respectively.

In developing Bayesian two-stage designs, we focused on mixture experiments for simplex regions. However, in many mixture experiments there are restrictions on the component proportions. Therefore, we showed how Bayesian two-stage D-D optimal designs can be applied in cases with constrained mixture regions.

While the effects of individual components are important for mixture models, the experimenter is often more interested in how the combination of the components works to produce a particular response, as a result, second or higher order effects can be more important than main effects. Hence, we discussed D-optimal second or higher order designs. We showed that D_s optimal designs are a reasonable alternative to D-optimal designs in terms of variance of parameter estimates and variance of prediction along the prediction rays.

Ideas for areas of extension and interest that have arisen from this research are listed below.

1) Extend the application of Bayesian two-stage designs to a nonlinear mixture model

with Binary or Poisson responses. One must first determine the nature of the single stage optimal design.

2) Multiple stage designs containing three or more stages are natural extensions of the two-stage design. The important issue here is whether there is sufficient improvement to warrant a third stage that may otherwise be impractical.

Bibliography

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

Andere-Rendon, J., Montgomery, D. C. and Rollier, D. A. (1997). "Design of Mixture Experiments Using Bayesian D-Optimality". Journal of Quality Technology, 29, 451-463.

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

Box, G. E. P. and Meyer, R. D. (1993). "Finding the Active factors in Fractionated Screening Experiments". Journal of Quality Technology, 25, 94-105.

Box G. E. P. and Draper. N. R. (1987). Empirical Model Building and Response Surfaces. John Wiley & Sons, Inc., New York, NY.

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

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

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

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

Chaloner, K. and Verdilli, I. (1995). "Bayesian Experimental Design: A Review". University of Minnesota technical report.

Cook, R. D. and Nachtsheim. C. J. (1980). "A Comparison of Algorithms for Constructing Exact D-optimal Designs". Technometrics 22, pp. 315-324.

Cornell. J. A. (1990). Experiments with Mixtures: Designs, Models and Analysis of Mixture Data, 2nd. ed., John Wiley & Sons, Inc., New York, NY.

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

DuMouchel, W and Jones, B. (1994). "A simple Bayesian Modification of D-optimal Designs to reduce Dependence on an Assumed Model". Technometrics, 36, 37-47.

Dykstra, O. Jr. (1971). "The Augmentation of Experimental Data to Maximize $|X'X|$". Technometrics 13, pp. 682-688.

Fedorov, V. V. (1972). Theory of Optimal Experiments. Academic Press, New York, NY.

Hill, W. J. and W. G. Hunter (1974). "Designs of Experiments for Subsets of Parameter". Technometrics 16, 425-434.

Huffman, J. (1998). "Optimal Experimental Design for Poisson Impaired Reproduction Studies". Ph. D. Thesis, Virginia Tech.

Kiefer, J. (1959). "Optimum Experimental Designs". Journal of the Royal Statistical Society, Series B, 21, 272-304.

Kiefer, J. (1961). "Optimum Designs in Regression Problems". Annals of Mathematical Statistics, 32, 298-325.

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

Kiefer, J. (1961). "Optimal Designs in Regression Problem, II". Ann. Math. Statist. 37, 783-815.

Kowalski, S., Cornell, J. A. and Vining, G. G. (1999). "A New Model and Class of Designs for Mixture Experiments with Process Variables".

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

Lindley, D. V. (1972). "Bayesian Statistics - A Review". SIAM, Philadelphia.

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

Mitchell, T. J. (1974a). "An algorithm for the Construction of D-optimal Experimental Designs". Technometrics 20, pp. 203-210.

Meyer, R. D. and Wilkinson, R. G. (1998). "Bayesian Variable Assessment". Communication Statistics – Theory Meth., 27(11), 2675-2705.

Myers, R. H. and Montgomery, D. C. (1995). Response Surface Methodology: Process and Product Optimization Using Designed Experiments. New York: John Wiley & Sons, Inc.

Myers, W. R., Myers, R. H., Jr. and White, K. L, Jr. (1996). "Two Stage Designs for the Logistic Regression Model in Singer Agent Bioassays". Journal of Biopharmaceutical Statistics, April issue.

Myers R. H. and Wardrop D. M. (1990). "Some Response Surface Designs for Finding Optimal Conditions". Journal of Statistical Planning and Inference 25 (1990) 29-34.

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

Neff A. R. (1996). "Bayesian Two Stage Design Under Model Uncertainty". Ph. D. Thesis, Virginia Tech.

Pilz, J. (1991). Bayesian Estimation and Experimental Design in Linear Regression Models. John wiley & Sons, Inc., New York, NY.

SAS Institute, Inc. (1996). SAS/QC Software: Usage and Reference, Version 6, First Edition, Volumes 1 and 2. SAS Institute Inc., Cary, NC.

Scheffe, H. (1958). "Experiments with Mixtures". Journal of the Royal Statistical Society, Series B, 20**,** 344-366.

Scheffe, H. (1959). "Reply to Mr. Quenouille's Comments About my Paper on Mixtures". Journal of the Royal Statistical Society, Series B, 23, 171-172.

Scheffe, H. (1965). "The Simplex-Centroid Design for Experiments with Mixtures". Journal of the Royal Statistical Society, B, 25, 235-263.

Vining, G. G., Cornell J. A. and Myers, R. H. (1993). "A Graphical Approach for Evaluation Mixture Designs". Applied Statistics, 42, No. 1, pp. 127-138.

Appendix

A. Calculations of $p(y|M_i)$ and $p(M_i|y)$ for section 4.1

The predictive density $p(y|M_i)$ can be obtained by using the following integration,

$$p(y \mid M_i) = \int_0^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y \mid M_i, \beta_i, \sigma) \times p(\beta_i \mid M_i, \sigma) p(\sigma \mid M_i) d\beta_i d\sigma ,$$

where

$$p(y \mid M_i, \beta_i, \sigma) = \frac{1}{(2\pi)^{n/2}} \sigma^{-n} \times \exp\left\{ \frac{-(y - X_i \beta_i)'(y - X_i \beta_i)}{2\sigma^2} \right\},$$

$$\beta_i \mid M_i, \sigma \sim N\left(0_{(p+r_i)\times 1}, \quad \sigma^2 \begin{pmatrix} \gamma^2 I_{(p\times p)} & 0_{p\times r_i} \\ 0_{r_i \times p} & \tau^2 I_{r_i \times r_i} \end{pmatrix} \right),$$

and

$$K_i^{-1} = \begin{pmatrix} \gamma^2 I_p & 0_{p\times r_i} \\ 0_{r_i \times p} & \tau^2 I_{r_i} \end{pmatrix} \text{ with a large value of } \gamma ,$$

where p and r_i are the numbers of primary and potential terms in model M_i, respectively,

Also

$$\mathrm{p}(\beta_i \mid M_i, \sigma) = \frac{e^{-\frac{1}{2\sigma^2}\beta_i' K_i \beta_i} \mid K_i \mid^{1/2}}{(2\pi\sigma^2)^{\frac{P+r_i}{2}}},$$

and the prior distribution of $\sigma^2 \sim IG(\alpha, \delta)$ with a small value of α and a large value δ,

$$\mathrm{p}(\sigma^2 \mid M_i) = \frac{e^{-\frac{1}{\sigma^2 \delta}}}{\Gamma(\alpha)\delta^{\alpha}(\sigma^2)^{\alpha+1}} .$$

Performing the integration with respect to β_i, σ ,

$$p(y \mid M_i) = \int_{\sigma^2} \int_{\beta_i} \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y - x_i\beta_i)'(y - x_i\beta_i)} \frac{e^{-\frac{1}{2\sigma^2}\beta_i' K_i \beta_i} \mid K_i \mid^{1/2}}{(2\pi\sigma^2)^{\frac{P+r_i}{2}}}$$

$$\times \frac{e^{-\frac{1}{\sigma^2\delta}}}{\Gamma(\alpha)\delta^{\alpha}(\sigma^2)^{\alpha+1}} d\beta_i d\sigma^2$$

$$= \frac{\mid K_i \mid^{1/2}}{\delta^{\alpha}(2\pi)^{\frac{n+p+r_i}{2}}\Gamma(\alpha)} \int_{\sigma^2} \frac{e^{-\frac{1}{\delta\sigma^2}}}{(\sigma^2)^{\frac{n+p+r_i}{2}+\alpha+1}}$$

$$\times \int_{\beta} \exp\{-\frac{1}{2\sigma^2}[y'y - 2y'x_i\beta_i + \beta_i'(x_i'x_i + K_i)\beta_i\} d\beta_i d\sigma^2,$$

where

$$\int_{\beta_i} \exp\{-\frac{1}{2\sigma^2}[y'y - 2y'x_i\beta_i + \beta_i'(x_i'x_i + K_i)\beta_i]\} d\beta_i$$

$$= \int_{\beta_i} \exp\{-\frac{1}{2\sigma^2}[(\beta_i - (x_i'x_i + K_i)^{-1}x_i'y)'(x_i'x_i + K_i)(\beta_i - (x_i'x_i + K_i)^{-1}x_i'y)]\}$$

$$\times e^{-\frac{1}{2\sigma^2}(y'y - y'x_i(x_i'x_i + K_i)^{-1}x_i'y} d\beta_i$$

$$= \frac{(2\pi\sigma^2)^{\frac{p+r_i}{2}}}{\mid x_i'x_i + K_i \mid^{1/2}} \exp[-\frac{1}{2\sigma^2} y'(I - x_i(x_i'x_i + K_i)^{-1}x_i')y].$$

Therefore,

$$\mathrm{p(y}\mid M_i) = \frac{\mid K_i \mid^{1/2}}{\delta^{\alpha} \mid x_i'x_i + K_i \mid^{1/2} (2\pi)^{n/2}\Gamma_{(\alpha)}} \int_{\sigma^2} \frac{e^{-\frac{1}{\sigma^2}\{[\frac{1}{\delta} + \frac{1}{2}y'(I - x_i(x_i'x_i + K_i)^{-1}x_i')y]\}}}{(\sigma^2)^{\frac{n}{2}+\alpha+1}} d\sigma^2,$$

$$= \frac{\Gamma_{(n/2+\alpha)} \, |K_i|^{1/2}}{\Gamma_{(\alpha)} \, |x_i' x_i + K_i|^{1/2} \, (2\pi)^{n/2} \delta^{\alpha} [\frac{1}{\delta} + \frac{1}{2} y`(I - x_i (x_i' x_i + K_i)^{-1} x_i' y]^{\frac{n}{2}+\alpha}} .$$

When $\alpha \to 0$ and δ is ∞,

$$p(y|M_i) \propto \frac{|K_i|^{1/2}}{|X_i`X_i + K_i|^{1/2} \, [y`(I - x_i (x_i`x_i + K_i)^{-1} x_i`y]^{n/2}} .$$

The expression of $p(y|M_i)$ also can be rewritten to the following form

$$p(y|M_i) \propto \frac{|K_i|^{1/2}}{|K_i + X_i' X_i|^{1/2} \, [(y - X_i \hat{\beta}_i)' (y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

Finally,

$$p(y|M_i) \propto \frac{\tau^{-r_i}}{|K_i + X_i' X_i|^{1/2} \, [(y - X_i \hat{\beta}_i)' (y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

and

$$p(M_i|y) \propto p(M_i) p(y|M_i)$$

$$= c (\frac{\pi}{1-\pi})^{r_i} \frac{\tau^{-r_i}}{|K_i + X_i' X_i|^{1/2} \, [(y - X_i \hat{\beta}_i)' (y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

where $\hat{\beta}_i = (K_i + X_i`X_i)^{-1} X_i`y$ and c is a constant that forces all the probabilities to sum to one. Notice that γ in the numerator is cancelled out due to the normalization of the probability. Letting $\gamma \to \infty$ in the denominator concludes the formula in section 4.1.

B. Calculations of p(y| M_i) and p(M_i |y) for section 4.5.1

The predictive density p(y|M_i) is derived by using the following integration,

$$p(y|M_i) = \int_0^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} p(y|M_i,\beta_i,\sigma)\times p(\beta_i|M_i,\sigma)p(\sigma|M_i)d\beta_i d\sigma ,$$

where

$$p(y|M_i,\beta_i,\sigma) = \frac{1}{(2\pi)^{n/2}}\,\sigma^{-n}\times\exp\left\{\frac{-(y-X_i\beta_i)'(y-X_i\beta_i)}{2\sigma^2}\right\},$$

$$\beta_i|M_i,\sigma \sim N\left(0_{(p+r_i)\times 1},\quad \sigma^2\begin{pmatrix}\gamma^2 I_p & 0_{p\times r_{1(i)}} & 0_{p\times r_{2(i)}}\\ 0_{r_{1(i)}\times p} & \tau_1^2 I_{r_{1(i)}} & 0_{r_{1(i)}\times r_{2(i)}}\\ 0_{r_{2(i)}\times p} & 0_{r_{2(i)}\times r_{1(i)}} & \tau_2^2 I_{r_{2(i)}}\end{pmatrix}\right),$$

let

$$K_i^{-1} = \begin{bmatrix}\gamma^2 I_{p\times p} & 0_{p\times r_{1(i)}} & 0_{p\times r_{2(i)}}\\ 0_{r_{1(i)}\times p} & \tau_1^2 I_{(r_{1(i)})} & 0_{r_{1(i)}\times r_{2(i)}}\\ 0_{r_{2(i)}\times p} & 0_{r_{2(i)}\times r_{1(i)}} & \tau_2^2 I_{(r_{2(i)})}\end{bmatrix}$$

with a large value of γ, where p, $r_{1(i)}$ and $r_{2(i)}$ are the numbers of primary, first class and second class of potential terms in model M_i, respectively, Also

$$p(\beta_i|M_i,\sigma) = \frac{e^{-\frac{1}{2\sigma^2}\beta_i' K_i\beta_i}\,|K_i|^{1/2}}{(2\pi\sigma^2)^{\frac{P+r_i}{2}}},$$

and the prior distribution of $\sigma^2 \sim IG(\alpha,\delta)$ with a small value of α and a large value δ,

$$p(\sigma^2|M_i) = \frac{e^{-\frac{1}{\sigma^2\delta}}}{\Gamma(\alpha)\delta^{\alpha}(\sigma^2)^{\alpha+1}}.$$

Performing the integration with respect to β_i, σ ,

$$p(y|M_i) = \int_{\sigma^2}\int_{\beta_i} \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y-x_i\beta_i)`(y-x_i\beta_i)} \frac{e^{-\frac{1}{2\sigma^2}\beta_i' K_i \beta_i}}{(2\pi\sigma^2)^{\frac{P+r_i}{2}} |K_i^{-1}|^{1/2}}$$

$$\times \frac{e^{-\frac{1}{\sigma^2\delta}}}{\Gamma(\alpha)\delta^\alpha(\sigma^2)^{\alpha+1}} d\beta_i d\sigma^2$$

$$= \frac{|K_i|^{1/2}}{\delta^\alpha(2\pi)^{\frac{n+p+r_i}{2}}\Gamma(\alpha)} \int_{\sigma^2} \frac{e^{-\frac{1}{\delta\sigma^2}}}{(\sigma^2)^{\frac{n+p+r_i}{2}+\alpha+1}}$$

$$\times \int_{\beta} \exp\{-\frac{1}{2\sigma^2}[y'y - 2y'x_i\beta_i + \beta_i'(x_i'x_i + K_i)\} d\beta_i d\sigma^2,$$

where

$$\int_{\beta_i} \exp\{-\frac{1}{2\sigma^2}[y'y - 2y`x_i\beta_i + \beta_i'(x_i'x_i + K_i)\beta_i]\} d\beta_i$$

$$= \int_{\beta_i} \exp\{-\frac{1}{2\sigma^2}[(\beta_i - (x_i`x_i + K_i)^{-1}x_i`y)`(x_i`x_i + K_i)(\beta_i - (x_i`x_i + K_i)^{-1}x_i`y)]\} \times$$

$$e^{-\frac{1}{2\sigma^2}(y'y - y'x_i(x_i'x_i + K_i)^{-1}x_i'y)} d\beta_i$$

$$= \frac{(2\pi\sigma^2)^{\frac{p+r_i}{2}}}{|x_i'x_i + K_i|^{1/2}} \exp[-\frac{1}{2\sigma^2} y'(I - x_i(x_i'x_i + K_i)^{-1}x_i')y].$$

Therefore,

$$p(y|M_i) = \frac{|K_i|^{1/2}}{\delta^\alpha |x_i'x_i + K_i|^{1/2} (2\pi)^{n/2}\Gamma_{(\alpha)}} \int_{\sigma^2} \frac{e^{-\frac{1}{\sigma^2}\{[\frac{1}{\delta} + \frac{1}{2}y'(I - x_i(x_i'x_i + K_i)^{-1}x_i')y]\}}}{(\sigma^2)^{\frac{n}{2}+\alpha+1}} d\sigma^2$$

$$= \frac{\Gamma_{(n/2+\alpha)} \mid K_i \mid^{1/2}}{\Gamma_{(\alpha)} \mid x_i' x_i + K_i \mid^{1/2} (2\pi)^{n/2} \delta^{\alpha} [\frac{1}{\delta} + \frac{1}{2} y'(I - x_i(x_i` x_i + K_i)^{-1} x_i' y]^{\frac{n}{2}+\alpha}}.$$

When $\alpha \to 0$ and δ is ∞,

$$p(y| M_i) \propto \frac{\mid K_i \mid^{1/2}}{\mid X_i' X_i + K_i \mid^{1/2} [y'(I - x_i(x_i' x_i + K_i)^{-1} x_i' y]^{n/2}}.$$

The expression of $p(y|M_i)$ can be rewritten to the following form

$$P(y| M_i) \propto \frac{\mid K_i \mid^{1/2}}{\mid K_i + X'_i X_i \mid^{1/2} [(y - X_i \hat{\beta}_i)'(y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

Finally,

$$p(y| M_i) \propto \frac{\tau_1^{-r_{1(i)}} \tau_2^{-r_{2(i)}}}{\mid K_i + X_i' X_i \mid^{1/2} [(y - X_i \hat{\beta}_i)'(y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

and

$$p(M_i|y) \propto p(M_i) p(y| M_i)$$

$$p(M_i|y) = c (\frac{\pi}{1-\pi})^{r_i} \frac{\tau_1^{-r_{1(i)}} \tau_2^{-r_{2(i)}}}{\mid K_i + X_i X_i \mid^{1/2} [(y - X_i \hat{\beta}_i)'(y - X_i \hat{\beta}_i) + \hat{\beta}_i' K_i \hat{\beta}_i]^{n/2}},$$

where $\hat{\beta}_i = (K_i + X_i' X_i)^{-1} X_i' y$, and c is the constant that forces all probabilities to sum to one. To deal with γ, we can use a similar argument as that in Appendix A.

Vita

Hefang Lin was born in 1966 in Zhejiang, China. In 1985, she entered Taizhou Teacher's College in Zhejiang, China to study physics and graduated in 1988. She began her graduate study in August 1995 at Virginia Polytechnic Institute and State University. In 1997 she was inducted in the national statistics honor society, Mu Sigma Rho, and received a Master of Science degree in Statistics. She received her Ph.D. in statistics in December 1999.