

Optimal Design of Single Factor cDNA Microarray Experiments and Mixed Models for Gene Expression Data

Xiao Yang

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

Doctor of Philosophy
in
Statistics

Ina Hoeschele, Co-chair
Keying Ye, Co-chair
Samantha Bates
Eric P. Smith
George Terrell

February 25, 2003
Blacksburg, Virginia

Keywords: Microarray Experiment, Mixed Models, Optimal Design
Copyright 2003, Xiao Yang

Optimal Design of Single Factor cDNA Microarray Experiments and Mixed Models for Gene Expression Data

Xiao Yang

(ABSTRACT)

Microarray experiments are used to perform gene expression profiling on a large scale. E- and A-optimality of mixed design was established for experiments with up to 26 different varieties and with the restriction that the number of arrays available is equal to the number of varieties. Because the IBD setting only allows for a single blocking factor (arrays), the search for optimal designs was extended to the Row-Column Design (RCD) setting with blocking factors dye (row) and array (column). Relative efficiencies of these designs were further compared under analysis of variance (ANOVA) models. We also compared the performance of classification analysis for the interwoven loop and the replicated reference designs under four scenarios. The replicated reference design was favored when gene-specific sample variation was large, but the interwoven loop design was preferred for large variation among biological replicates.

We applied mixed model methodology to detection and estimation of gene differential expression. For identification of differential gene expression, we favor contrasts which include both variety main effects and variety by gene interactions. In terms of t-statistics for these contrasts, we examined the equivalence between the one- and two-step analyses under both fixed and mixed effects models. We analytically established conditions for equivalence under fixed and mixed models. We investigated the difference of approximation with the two-step analysis in situations where equivalence does not hold. The significant difference between the one- and two-step mixed effects model was further illustrated through Monte Carlo simulation and three case studies. We implemented the one-step analysis for mixed models with the ASREML software.

Acknowledgments

I would like to thank Dr. Ina Hoeschele and Dr. Keying Ye, who have been there for me every step of the way, teaching, encouraging and helping. I greatly appreciate the time and effort that they have given to me and to my work, I have enjoyed working with them. What they have taught me in the last few years will benefit me all my life, and I am grateful to them forever.

I would also like to recognize the contribution of my committee members, Dr. Samantha Bates, Dr. Eric Smith and Dr. George Terrell. I am so thankful to them for agreeing to serve on my committee and for those helpful and constructive suggestions. I would also like to thank Dr. John P. Morgan, my former committee member, for those stimulating discussions about design of experiment with block size two and for providing references that have greatly improved this work.

Contents

1	Introduction	1
1.1	What is a DNA Microarray Experiment?	1
1.2	Statistical Design of cDNA microarray Experiment	3
1.3	Statistical Methods for Identifying Differentially Expressed Genes	6
1.4	Objectives	9
2	Optimal Design of Single Factor Microarray Experiments	11
2.1	Introduction	11
2.2	Optimality Results from Incomplete Block Design	12
2.3	Optimality Results from Row-Column Design	14
2.4	Applications of Optimal Designs to Microarray Experiment	16
2.4.1	Fixed Effects Models and Definition of Relative Efficiency	16
2.4.2	Relative Efficiencies of Mixed Designs	17
2.4.3	Relative Efficiencies of Row-Orthogonal Designs	17
2.4.4	Relative Efficiencies under Mixed Effects Models	18
2.5	Comparison of Classification Performance among Designs	22
2.6	Conclusion	32
3	Supplement to Optimal Design of Single Factor Microarray Experiments	34
3.1	Introduction	34
3.2	Optimality Results from Incomplete block Design	35
3.2.1	Theory on Incomplete Block Design	35

3.2.2	Construction of E-optimal Designs	36
3.2.3	E-efficiency of Mixed Designs	37
3.2.4	Proof of E-optimality for Certain Mixed Designs	38
3.2.5	A-efficiency of Mixed Designs	40
3.2.6	Comparison of Our Search with Kerr's Search	40
3.3	Optimality Results from Row-Column Design	42
3.3.1	Theory of Row-Column Design	42
3.3.2	Optimality Results for $b = 2\nu$ and $k = 2$	43
3.3.3	Optimality Results for $b = 3\nu$ and $k = 2$	44
3.3.4	Optimality Results for $b = 4\nu$ and $k = 2$	46
3.4	Comparison of Designs for Classifying Samples	50
3.4.1	Designed Experiments using Interwoven Loop Design	50
3.4.2	Comparison of Classification for Designs	52
4	Linear Mixed Effects Models for Microarray Gene Expression Data	55
4.1	Introduction	55
4.2	Hypothesis Testing under ANOVA Models	56
4.3	Linear Fixed Effects Models	58
4.3.1	Two-Step Analysis of Gene Expression Data	58
4.3.2	Global Factors Interact with Gene Factor (G)	58
4.3.3	General Case	59
4.4	Linear Mixed Effects Models	60
4.4.1	Two-Step Mixed Model Analysis	61
4.4.2	Modified Two-Step Mixed Model Analysis	63
4.5	Simulation Studies	63
4.5.1	Data Simulation and Analyses	64
4.5.2	Loop Design	64
4.5.3	Reference Design	65
4.6	Case Studies	66

4.6.1	The Mouse Model	67
4.6.2	Bovine Embryo Classification Study	68
4.6.3	Porcine Embryo Classification Study	69
4.6.4	Discussion	71
4.7	Conclusion	74
5	Future Research	79
5.1	Discussion	79
5.2	Future Research	79

List of Figures

2.1	Relative Efficiencies of Certain Designs including Loop and Reference Design under Fixed Effects Model M1	18
2.2	Relative Efficiencies of Replicated Reference Design under Three Scenarios	19
2.3	Relative Efficiencies of Mixed, Loop and Reference Designs under Mixed Effects Model with $\lambda = \sigma_e^2/\sigma_a^2$ set at 0.1, 0.33, 1 and 10.	20
2.4	Relative efficiencies of the replicated reference design, relative to optimal row-column design for different numbers of arrays (b) at any given number of varieties (ν) and for different ratios (λ) of array to residual variance.	21
2.5	Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 2\nu, \nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.	24
2.6	Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 3\nu, \nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.	25
2.7	Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 4\nu, \nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.	26
2.8	Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 5\nu, \nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.	27
2.9	Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=0.5$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.	28

2.10	Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=1$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.	29
2.11	Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=2$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.	30
2.12	Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=4$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.	31
3.1	E-Efficiency of certain mixed designs	38

List of Tables

2.1	Minimally Connected Design ($b = \nu - 1$)	12
2.2	Design Mix(3)	13
2.3	Design Mix(2)	13
2.4	A Special Case of Mixed Design: Mix(1)	13
2.5	Resolvable Inter-woven Loop Design ($\nu = 6$)	15
2.6	Average Number of Misclassifications for Different λ Values under Each Scenario: standard errors were reported in brackets.	23
2.7	Average Number of Misclassifications for Different δ values under Each Scenario: standard errors were reported in brackets.	24
3.1	Relative Efficiency of Certain Mixed Designs for $b = \nu$ and $k = 2$: when $\nu=12$, Mix(3) and Mix(4) are both A-optimal, which is denoted by (*). No Mix(4) exists for $\nu=3$, because only 3 varieties are available, and the maximum level of mixing is 3.	41
3.2	Labelling of 20 Samples in Classification Study	51
3.3	Interwoven Loop Design ($b = 40, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples	51
3.4	Interwoven Loop Design ($b = 60, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples	51
3.5	Interwoven Loop Design ($b = 80, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples	52
3.6	Summary of Variances of Pairwise Contrasts among Samples under Four Scenarios: $\nu=20$	53
3.7	Variances of Contrasts between Types (T) under Each Scenario	53

4.1	Gene Expression Values for Experiment A and B: expression values of both experiments are the same except for expression values of gene 1 in sample 2 for Experiment B, which are listed in the bracket.	57
4.2	Combination of Varieties with Dyes for the Loop Design	65
4.3	Summary of Simulation Results for the Loop Design ($\sigma_{AG}^2 = 0.5$): Median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.	65
4.4	Summary of Simulation Results for the Loop Design ($\sigma_{AG}^2 = 2$): Median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.	66
4.5	Combination of Varieties with Dyes for the Reference Design	66
4.6	Summary of Simulation Results for the Reference Design ($\sigma_{AG}^2 = 0.5$): median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.	67
4.7	Summary of Simulation Results for the Reference Design ($\sigma_{AG}^2 = 2$): median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.	67
4.8	Adjusted p -values for Differentially Expressed Apo AI. Genes: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.05, and gene selection from the step-down Bonferroni was based on p -value=0.05.	68
4.9	Adjusted p -values for Differentially Expressed Genes in Bovine Study: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.05, and gene selection from the step-down Bonferroni was based on p -value=0.05.	70
4.10	Adjusted p -values for Differentially Expressed Genes in Porcine Study: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.01, and gene selection from the step-down Bonferroni was based on p -value=0.01.	72

Chapter 1

Introduction

Genomic science is increasingly focused on understanding the genetic architecture of biochemical and developmental pathways underlying multi-factorial, quantitative, or complex traits, encouraged by the impressive technological advances in the field in the past decade or more. As a high throughput technology, DNA microarray technology not only provides biologists deep insights into the molecular functionality of genes through gene expression profiling, but also demands new statistical methods to handle the tremendous amount of information generated from these experiments. Development of statistical and probabilistic methods for the analysis and interpretation of microarray data is very important, as these data typically contain substantial amounts of measurement error and variability, and are highly dimensional with many observed and latent variables, which must be modelled for marginal inference about the quantities of interest in a probabilistic system.

1.1 What is a DNA Microarray Experiment?

The genome of an organism stores the ultimate genetic information, which determines what an organism is and how it functions. A genome can be comprised of tens of thousands of genes. The proteins drive and regulate the various metabolic pathways in the living cells. A gene is a fragment of deoxyribonucleic acid (DNA), a double-stranded helix. A DNA molecule consists of four nucleotides, and each nucleotide is made up of a deoxyribose, a phosphate group and one of the four nitrogen bases called adenine (A), guanine (G), cytosine (C) and thymine (T). The two strands of a DNA molecule are complementary to each other on the base-pairing rule: A is complementary to T and G to C. They attach to each other by hydrogen bonds of the complementary base pairing.

Gene expression is carried out in two steps, *transcription* and *translation*. During transcription, DNA is transcribed into messenger RNA (mRNA). mRNA is a complementary copy of one strand of the DNA, in which the deoxyribose is replaced by ribose and the base T is

replaced by base uracil (U). In the stage of translation, the transcribed mRNA is translated into protein, the final gene expression product to function in biological reaction. The proteins drive and regulate the various metabolic pathways in the living cells. Traditional molecular biology studies gene expression on a "one gene in one experiment" basis. However, the various metabolic pathways connect to each other and form a huge complicated metabolic network and functions as a whole dependable system. Therefore, genes and their products coordinate with each other. Study of all genes at one time allows biologists to investigate the gene interactions of living organisms under different conditions, which should provide an efficient way to study the complicated metabolic pathways and the relationship of genes to gain insight on gene functions (Lander, 1999).

DNA microarray experiments are part of a new class of biotechnologies which allow the monitoring of expression levels for thousands of genes simultaneously (Derisi *et al.*, 1997; Eisen *et al.*, 1998). Applications of microarrays range from the study of gene expression in plants under different environmental stress conditions to the comparison of gene expression profiles for tumors from cancer patients. Gene transcription experiments based on microarrays identify genes involved in complex disease or multifactorial traits by comparing transcription levels of genes in different tissue samples.

In a typical microarray experiment (*e.g.*, cDNA microarrays for gene transcription profiling), the technique takes the following steps. In the first step, samples of DNA clones with known sequences are spotted and immobilized onto a glass slide or other substrate. Secondly, tissue samples containing mRNA of currently transcribed genes are collected from different cell lines (*e.g.*, samples from trees under normal condition and trees under drought stress), each sample undergoes a process by which the mRNA is reverse-transcribed into cDNA. Subsequently, the two samples are labelled with two different fluorescent dyes (usually red fluorescent dye Cy5 and green fluorescent dye Cy3). The samples are then mixed and hybridized to the arrayed DNAs on the glass slide, and any unhybridized cDNA is washed off. Next, the slides are imaged using a laser scanner, and separate fluorescence intensity measurements are made for each dye on each spot on the array. Typically, the two measurements (one for red and the other for green) for each spot indicate the relative abundance of the corresponding mRNA in the two tissue samples. Therefore, these measured fluorescent intensities should approximately represent the gene transcriptional levels across different tissue samples.

Performing and interpreting cDNA microarray expression experiments is not efficient without statistical support at each step of the process. This includes experimental design (*e.g.*, Kerr and Churchill, 2001), image processing (*e.g.*, Yang *et al.*, 2000), identification of systematic sources of variation, identification of differentially expressed genes (*e.g.*, Dudoit *et al.*, 2000), clustering of genes and samples or class discovery (*e.g.*, Getz *et al.*, 2000), sample classification (*e.g.*, Dudoit *et al.*, 2001), and pathway and gene network inference (*e.g.*, Pe'er *et al.*, 2001; Henderson *et al.*, 2003). Among these steps, design of valid and efficient microarray experiments and identification of differentially expressed genes are crucial to subsequent statistical inferences, which will be addressed thoroughly in this dissertation.

1.2 Statistical Design of cDNA microarray Experiment

Design of experiment plays an important role in characterizing the process of science and technology (Hinkelmann and Kempthorne, 1994). The choice of a design depends on the availability of experimental units, the structure of these units, and the precision of estimation desired by the investigator. For this purpose, valid and efficient statistical designs are needed in order to identify and eliminate sources of systematic variations in microarray experiments. Kerr and Churchill (2001) and Kerr *et al.* (2000) are among the first to describe in detail the need for alternative designs and to give some recommendations on finding "good" designs for microarray experiments. These authors have also proposed to analyze gene expression data using the method of analysis of variance (ANOVA).

Kerr and Churchill (2001) discussed sources of variations from a microarray experiment, and also provided statistical insights into various forms of designs widely used in practice, especially the reference and loop design. In a reference design, the "reference sample" is labelled with a single dye, samples of interest are labelled with the secondary dye and individually hybridized to one or more arrays with the reference sample. The loop design is used to design microarray experiments in which the layout of dyes and varieties are balanced, in a sense that each variety is dyed with each dye (Cy3 or Cy5) equally often. The arrangement of varieties within each array takes place in a cyclic fashion, *i.e.*, for a given sample, on one array, the sample is dyed with one dye, and on the next array, it would be labelled with the secondary dye. Some variants of the loop design have been used in practice, especially for the case that there are only two varieties (or tissue samples) of interest. When there are two varieties and two or more microarrays, such a design is termed as Latin Square Design or a replicated Latin Square Design (Kerr *et al.*, 2000). These designs often have a nice orthogonal structure, which makes the analysis straightforward.

Kerr and Churchill (2001) extensively studied the properties of the reference and loop design. For the reference design, due to the peculiar combination of dyes and tissue samples (or varieties), dye effects and variety effects are completely confounded, and consequently the variety by gene interaction (VG) is also completely confounded with dye by gene interaction (DG). In ANOVA models, confounding means certain estimated factor effects are not distinguishable from other effects. As a result, the statistical power for testing these factor effects may be substantially reduced due to confounding. For the purpose of data normalization for microarray experiment, as dyes used in the labelling reactions introduce bias, dye effects have to be accounted for, thus reference design is not favored due to the confounding between dyes and samples of interest. There are other drawbacks to such a design (Jin *et al.*, 2001; Kerr and Churchill, 2001). First, it is inefficient since the reference sample does not supply any information of biological interest. Second, the use of different reference pools places a strong constraint on the potential for comparison of data generated by different investigators.

Compared to the reference design, the loop design has certain advantages, different sources

of variations can be accounted for by achieving the balance between two dyes and varieties of interest. Kerr and Churchill (2001) also found that the loop design is A-optimal when the number of samples is from two to eight. However, a potential problem with loop design is that it tends to be less efficient when the number of varieties exceeds eight. This prompts the need to search for a general class of optimal designs that can be used for design of larger microarray experiments with more than eight varieties. These authors also established the A-optimality of interwoven loop designs in the class of equireplicate designs where the number of arrays available is two times of that of varieties.

From the nature of the microarray experiment, since each microarray is probed with two differentially labelled cDNA samples, each array can be treated as the experimental blocks with block size two. When there are more than two cDNA samples of interest, not every sample can appear on each array. Therefore, the nature of the experimental design is essentially an incomplete block design with block size of two (Kerr and Churchill, 2001). However, the computer search strategy implemented by Kerr and Churchill (2001) becomes extremely inefficient when the number of varieties exceeds eight, which makes the search for optimal designs almost impossible. Therefore, it is desirable to have a more general and efficient strategy to search for optimal designs with more than eight varieties.

In a block design setting, another closely related question is the choice of optimality criterion. For a given design, its performance is assessed based on the formulated information matrix for estimating the treatments effects. Following this idea, Kiefer (1975) introduced various optimality criteria to measure the performance of a design. Each of these (A-, D- and E-) criteria amounts to the minimization of a particular convex function of the eigenvalues of the information matrix. For design of microarray experiment, two of these are quite relevant, the E- and A-optimality criteria. The E-optimality criterion was first introduced by Ehrenfeld (1955), and it is appropriate to use in experimental settings where it is desirable to estimate all treatment contrasts with as much precision as possible. It is known that a design is E-optimal if and only if it minimizes the maximum variance of the least squares estimators of normalized treatment contrasts, implying the E-optimality is a minimax criterion (Shah and Sinha, 1989). Another criterion, A-optimality has been used to minimize the average of variances of all pairwise treatment contrasts, which gives rise to designs with the maximum average precision when estimating an elementary treatment contrast. The choice of either criterion usually depends on the goal of the experiment.

Very little is known about the E- and A-optimality results regarding the incomplete block design with block size of two in current literature. Mitchell and John (1977) considered classes of such designs containing regular graph designs (RGD) where the number of varieties is less than thirteen. They used computer search techniques to find those RGDs which are A-optimal among the sets of RGD. Jacroux (1985) also considered the optimality of some of designs by John and Mitchell (1977) in the class of connected designs. Special cases have been studied in the literature. Cheng (1979) extensively studied the optimality aspects of incomplete block designs involving four varieties, and gave a general solution to optimal designs for any given number of blocks with block size of 2. In an attempt to contradict the

John-Mitchell Conjecture, Constantine (1986) proposed a class of E-optimal designs with block size of two. Bagchi and Cheng (1993) studied the properties of a class of E-optimal designs based on graph theory. However, these designs contain too many blocks, which makes them not practically useful in the context of DNA microarray experiment.

Optimality results by Kerr and Churchill (2001) were obtained within the framework of incomplete block design. In the literature of statistical design, block designs are used for experiments where it is important to eliminate heterogeneity in one direction. Therefore, the blocking factor considered is just array, and these optimality results are obtained without considering variation due to dyes. In a more realistic experimental situation, the combination of dyes with varieties within each array should be taken into account, because observed responses can be affected by different combinations of dyes and varieties (Kerr and Churchill, 2001; Yang *et al.*, 2000). In this experimental situation, the row-column designs should be used to eliminate heterogeneity in two directions (Shah and Sinha, 1989). In the context of microarray experiment, these two factors would be arrays and dyes. It should be noted that optimality results from block designs do not necessarily carry over to the row-column designs. Therefore, a natural extension would be to find out the optimality results under such settings. Optimal row-column designs with relatively small number of columns (arrays) and row size of two are yet to be developed. Jacroux (1985) considered a class of E-optimal designs with $k = 2$ based on group divisible design (GDD). Das (1991) established the E-optimality of some row-column designs through the augmentation process. Using graph theory, Bagchi and Cheng (1993) obtained another class of E-optimal rectangular designs with block size of two. In general, these optimal designs often contain many arrays, thus not practically useful for design of microarray experiments.

Another aspect of designing microarray experiments is to determine the number of replicates required for each gene in order to detect any significant difference of gene expression level across cell lines of interest. Lee *et al.* (2000) pointed out the importance of replication in gene expression microarray studies. Currently, three approaches are used in practice. Wolfinger *et al.* (2001) proposed to analyze the gene expression data using mixed effects model. Based on the classical t test in the context of analysis of variance, we can not only calculate the number of replicates needed for each gene in an experiment but also compare the performance of a particular design from a microarray experiment. Wolfinger's method can be applied to microarray experiments with any number of arrays. However, the properties of Wolfinger's method have not been stated clearly and more studies are needed to further explore these. Another method by Medvedovic (2001) is based on the classical t test and the number of replicates for each gene can also be calculated by the pre-specified power. However, this method requires the correct estimation of intra-gene variability, and also does not consider the possible correlation among these genes. Black and Doerge (2002) presented a method for determining the number of per gene replicate spots required in microarray experiments. Based on estimates of sampling variability from control microarrays, the number of replicates are calculated from specific contrasts using fixed effect models of Kerr and Churchill (2001).

1.3 Statistical Methods for Identifying Differentially Expressed Genes

Various statistical approaches have been suggested for identifying differential gene expression in the literature. Currently there are three types of methods available. A first group of methods include relatively simple classical approaches, and most of them require prior normalization of the raw data. These methods include the simple ratio-based approach (DeRisi *et al.*, 1996; Chen *et al.*, 1997), simple *t*-test (Claverie, 1999) and the two-stage *t*-test (Dudoit *et al.*, 2002, Yang *et al.*, 2001; see Pan (2002) for reviews of *t*-test based methods.), and some nonparametric tests. DeRisi *et al.* (1996) identified differentially expressed genes using a ± 3 cut-off for the log ratios of the fluorescence intensities, standardized with respect to the mean and standard deviation of the log ratios for a panel of 90 "housekeeping" genes (*i.e.*, genes believed not to be differentially expressed genes between the two cell lines of interest). Chen *et al.* (1997) propose a data dependent rule for choosing the cut-offs for the red and green intensity ratio R/G. Based on the assumption of normality and constant coefficient of variation for measurements from two dyes, a confidence interval has been constructed for R/G for each gene. The disadvantage of the ratio-based approach is that it only applies to experiment with single array and two tissue samples.

For the two-step *t*-test approach, at the first stage, the log-transformed ratios are normalized by robust locally weighted regression, which can be applied per array, print-tips, etc.). Then the standard *t*-statistic is calculated for two-group data (assuming there are two samples of interest), or an F-statistic can be calculated for multi-group data (assuming more than two samples). To assess the statistical significance, a bootstrapping is performed to make the test robust and to correct for the problem of multiple tests. The simple *t*-test by Claverie (1998) was similar to the second step of the approach by Dudoit *et al.* (2001), which also corrects for the problem of multiple testing. A clear advantage of the *t*-test approach is that it makes no parametric assumption regarding the distribution of the expression levels. It is also the first attempt to rigorously address the multiple comparison problem. In general, these methods do not perform very well, because they do not fully account for the systematic variations in the raw intensity measurements. Another limitation is that, because these methods all use ratios either directly or indirectly, they may not be adequate for testing differential expression when some measured intensity are relatively low (Rocke and Durbin, 2001).

Lin *et al.* (2001) focused on genes with low levels of transcriptional abundance, where there is a problem with background subtraction leading to negative fluorescent intensity values, which are set to zero or discarded. Instead of performing a usual log transformation, they compute normal scores from the ranks of the intensity measurements. Normal scores assume that some transformation of the intensity data to normality exists. Ranks are obtained separately for sets of observations obtained under the same conditions (*i.e.*, same array, same dye), unless these conditions are considered as random samples from a population of conditions. For each spot, a standardized normal score is computed using robust estimates of

center and scale of its average normal score. Standardized scores are assumed to be standard normally distributed (conditional on average) for all genes whose expression is not affected by condition or sample type. The resulting tests require adjustment for multiple testing as described earlier.

A more recent work is by Gibbons *et al.* (2002), which identified differentially expressed genes using the method of sequential sampling. Assume we have a replicated microarray experiment with two tissue samples: control versus treatment. The nonparametric prediction interval is computed as an order statistic of control measurements and is applied sequentially to a series of replicate subsets of experimental measurements. and is applied sequentially to a series of replicate sets of experimental measurements. After each subset is obtained, they determine which gene has a mean intensity outside of the prediction interval in that subset and all preceding subsets. This method continues until all genes are within the prediction interval in at least one subset or all the subsets have been visited once. Those genes which have mean intensities consistently above or below the prediction interval in all subsets are considered as differentially expressed. Compared to the traditional two sample comparison approach, this method provides a way to screen large number of genes when only small numbers of experimental samples are available, and is also a robust procedure which takes into account of the problem of multiple testing. The disadvantages of this approach is also obvious: it can only be applied to replicated microarray experiment with two samples, treatment versus control.

A second group of methods consist of Bayesian approaches, providing posterior belief probabilities of change (*e.g.*, Sapir and Churchill, 2000; Newton *et al.*, 2000; Baldi and Long, 2001; Theilhaber *et al.*, 2001; Efron *et al.*, 2001). Like some of the classical approaches, these approaches also rely on strong distributional and parametric assumptions (Newton *et al.*, 2000; Baldi and Long 2001), while the method of Sapir and Churchill (2000) is more empirical. Sapir and Churchill (2000) attempted to estimate the posterior belief probability of differential gene expression based on orthogonal linear regression of transformed intensity values from one channel on transformed intensity values from the other channel. The residuals of this regression are modelled with a two-component normal mixture distribution, one corresponding to differentially transcribed and the other to not differentially transcribed genes. Baldi and Long (2001) considered two samples (treatment and control) and repeated (and normalized) transcription values for each gene, which are assumed to be normally distributed within each sample. The authors discussed priors and point estimates for mean and variance of the normal distributions and Bayesian inferences to decide whether the two normal distributions for a given gene are different. A limitation of these two methods is that they both ignore the information contained in the product RG representing the overall transcriptional abundance. In view of this problem, Newton *et al.* (2001) considered a hierarchical model (Gamma-Gamma-Bernoulli model) for the measurements (R, G) and suggested identifying differentially expressed genes based on the posterior odds of change. The odds are functions of $(R + G)$ and RG and thus produce a rule which takes into account overall transcriptional abundance. Similar to the classical approaches, most Bayesian approaches

assume very strong distributional assumptions, without proper prior data normalization, their performance can be questionable under certain circumstances.

Limitations with the first two groups of methods include the need for prior data normalization of the raw intensity measurements, restricted applicability to experiments with only two samples (control and treatment) or to data without replication, and analysis of each array separately (Kerr *et al.*, 2002). In the context of gene expression data analysis, data normalization is used to remove systematic noises in the measurements, *i.e.*, effects of slides, dyes, print-tips, spatial variation. Because of the peculiar nature of DNA microarray experiments, every experimental step may contribute many noises to the gene expression data. These effects should be selectively excluded from the raw intensity measurements. The first of this kind is to normalize the measured intensities of genes of interest using the expression levels of "housekeeping" genes (DeRisi *et al.*, 1996). Housekeeping genes are genes whose expressions are not supposed to change across different tissue samples. This idea is conceptually interesting but hard to implement because it is hard to find certain known genes which are not sensitive to tissue samples. Schuchardt *et al.* (2000) suggested to remove these variations by simply averaging the raw measurements across arrays, while Yang *et al.* (2000) propose a robust procedure to account for intensity and spatial dependence in two color channels. In general, since these normalization strategies were developed under different contexts, their applicability and generality to other DNA microarrays are not clear, and these procedures deserves further investigation. Therefore, applications of classical tests or Bayesian methods are limited unless a more general and more appropriate normalization method can be found.

A third group of methods include fixed or mixed effects models which perform both normalization and identification of differentially transcribed genes, which are preferred to the above two groups. Kerr and Churchill (2001) gave a full account for sources of variation in the context of statistical design of experiment. By accounting for these variations, different effects due to slides, dyes, genes and tissue samples can be estimated and systematically removed by means of ANOVA approach. The data normalization and inference for differentially expressed genes have been integrated into a single step of linear fixed effects model. Although the ANOVA approach seems promising in theory, it is computationally intensive, and becomes almost infeasible when the number of genes in the model gets relatively large. Alternatively, a two-step procedure is given by Wu *et al.* (2003). Another limitation with this approach is that some of the factors in the model can be better modelled as random effects, such as the effects due to slides. In general, mixed model methodology provides both a formal framework and a flexible tool for identifying systematic sources of variation and differential gene expression.

An extension of the ANOVA approach in terms of mixed effect model is due to Wolfinger *et al.* (2001), which has been implemented to real applications (Jin *et al.*, 2001; Churchill and Oliver, 2001). This method uses two interconnected sequential linear mixed models. For the first stage of analysis, a linear mixed effect model for all genes is used to perform data normalization, which is essentially an additive model with only "global effects", such as effects of slides, dyes, tissue samples, print tips, etc. In the second step, another linear

mixed effects model is applied to estimate the intra-gene variability and to make inference for each gene, and responses are residuals obtained from the first step. Differential gene expression is detected based on t-statistics and confidence intervals for linear contrasts of the sample effects for each gene model. Yang *et al.* (2001) have made some modifications to the method by Wolfinger *et al.* (2001) by considering Bayesian analysis with a student t distribution. This method can be viewed as a combination of the ANOVA approach and the Bayesian approach, but it only applies to microarray experiments with two varieties. The full fixed or mixed model analysis of gene expression are computationally intensive, and the two-step analyses (Wolfinger *et al.*, 2001; Wu *et al.*, 2003) have been proposed as alternatives to the full model analysis. While the need for data transformations to adjust for intensity dependent dye effects and spatial variation and to stabilize variances prior to analyses is being recognized (see Cui *et al.* for a quite comprehensive discussion of suitable transformation), the consequences of replacing the single or full model analysis with a two-step approach have not yet been studied nor clearly stated. None of above authors have justified the two-step analysis, or stated clearly when single or full model and two-step model analyses provide the exact same tests for differential gene expression.

1.4 Objectives

The objective of this dissertation is to provide new and improved statistical and computational tools for designing more efficient microarray experiments and identifying differentially expressed genes from these experiments. This work includes:

- (i) A comparison and search for more efficient statistical designs which can be used to improve the efficiency of microarray experiments. We focus on the search for optimal designs suitable for single-factor cDNA microarray experiments. Optimality results will be obtained within the Incomplete Block Design (IBD) setting. Next, we extend our search for optimal designs to the Row-Column Design (RCD) setting under which variations due to arrays and dyes are fully accounted for.
- (ii) A comparison and further improvements of current statistical methods to identify differentially expressed genes. We first propose a new set of contrasts for testing hypotheses, then examine the equivalence between one- and two-step analyses under both fixed and mixed effects models. Three case studies will be used to illustrate our methods.

This dissertation is organized as follows. Statistical design of single-factor microarray experiments are discussed in Chapter 2. The optimality of mixed designs is established within incomplete block design, and the interwoven designs are found optimal under different scenarios under row-column design. The relative efficiency of these designs and the performance of sample classification for some designs are also studied. Chapter 3 supplements Chapter 2, where technical details of the design theory and optimality results are presented. In Chapter

4, linear mixed effects models for identification of differentially expressed genes are presented. A new set of contrasts are proposed, and the equivalence of the one- and two-step analyses under fixed and mixed effects models are examined. Finally, in Chapter 5, results from this contribution are summarized and implications for future research are discussed.

Chapter 2

Optimal Design of Single Factor Microarray Experiments

2.1 Introduction

Microarray gene expression experiments are used to perform gene expression profiling on a large scale (e.g., Derisi *et al.*, 1997; Eisen *et al.*, 1998). Performing and interpreting cDNA microarray expression experiments is not efficient without statistical support at each step of the process. This includes experimental design (e.g., Kerr and Churchill, 2001), image processing (e.g., Yang *et al.*, 2000), identification of systematic sources of variation, identification of differentially expressed genes (e.g., Dudoit *et al.*, 2000), clustering of genes and samples or class discovery (e.g., Getz *et al.*, 2000), sample classification (e.g., Dudoit *et al.*, 2001), and pathway and gene network inference (e.g., Pe'er *et al.*, 2002; Henderson *et al.*, 2003). Design of valid and efficient experiments plays a critical role in the interpretation of the data from microarray experiments, since efficiency and reliability of obtained gene expression data can potentially be greatly improved (Yang and Speed, 2002).

Statistical design recommendations for microarray experiments are still rather scarce. As noted by Smyth *et al.* (2002), it is not possible to give universal recommendations for all situations but the general principles of statistical experimental design apply to all microarray experiments. These authors point out that choice of design depends not only on statistical efficiency, but also on the goals of the experiment and on the available resources (number of arrays, amount of mRNA samples). While it is important to optimally design microarray experiments with factorial structure (Glonek and Solomon, 2002), here we focus on single factor microarray experiments with larger number of varieties. The reason for this focus is that in some microarray experiments we have been involved in, a single treatment or variety factor was of interest, but this factor had in the order of 10 to 25 levels (e.g., biological replicates of one type or of several unknown types). To investigate optimal design for such

experiments, we first identify optimal designs in the context of incomplete block design and row-column design using standard statistical A- and E-optimality criteria. Subsequently we compare the optimal design to the designs currently used (reference, loop and replicated reference) for detection of differential gene expression by fitting Analysis of Variance (ANOVA) models. Lastly, we compare the performance of optimal and replicated reference designs for sample classification.

2.2 Optimality Results from Incomplete Block Design

Kerr and Churchill (2001) studied optimal designs for cDNA microarray experiments in the context of incomplete block design, based on the observation that there is more variation between arrays than within arrays for gene expression data (Yang and Speed, 2002). Because only two differentially labelled mRNA samples can be hybridized on the same array or slide, arrays are experimental blocks of size two.

We now introduce some notation for statistical block design. Assume that ν varieties are to be compared via b blocks (arrays) of size $k = 2$. Any arrangement of the ν varieties into the bk experimental units is a design. It is known that for all the possible variety contrasts to be estimable, a design must be connected (Shah and Sinha, 1989). We denote by $D(\nu, b, k)$ the class of all connected block designs having ν varieties arranged in b blocks of size k .

In this dissertation E- and A-optimality criteria are used to find optimal designs. E-optimal designs are appropriate when all variety contrasts are to be estimated with as much precision as possible (Ehrendfeld, 1955). A-optimality minimizes the average variance of all pairwise variety contrasts. Incomplete block designs with $k = 2$ have been studied by Cheng (1979), Jacroux (1985), Constantine (1986) and Bagchi and Cheng (1993). When there is no limitation on the number of blocks (or arrays), a balanced incomplete block design (BIBD) is universally optimal. An augmentation or deletion process can be used to construct E-optimal designs (Jacroux, 1981a, 1982; Constantine, 1981, 1982; Sathe and Bapat, 1985; Das, 1991; Srivastav and Morgan, 1998). Methods of constructing A-optimal designs for $k = 2$ remain largely unknown, while computer algorithms can be used to search for optimal designs (Eccleston, 1980; Tjur, 1993; Venables and Eccleston, 1993; Nguyen, 1994).

We search for optimal design for $b = \nu$ and $k = 2$, which are design parameter values for the reference and loop designs (without replication).

Table 2.1: Minimally Connected Design ($b = \nu - 1$)

1	1	1	...	1	1
2	3	4	...	$\nu - 1$	ν

Based on the minimally-connected design in Table 2.1, optimal designs with $b = \nu$ can be constructed by applying the following augmentation process: (1) Add a block with two distinct varieties (e.g., block with varieties 2 and 3 only, see Table 2.2); (2) Add an identical block from design in Table 2.1 to itself (e.g., block with varieties 1 and 2 only, see Table 2.3); (3) Add a non-binary block with variety 1 only, see Table 2.4).

Table 2.2: Design Mix(3)

1	2	3	1	1	...	1
2	3	1	4	5	...	ν

Table 2.3: Design Mix(2)

1	1	1	...	1	2
2	3	4	...	ν	1

These designs can be viewed as a mixture of a loop and a reference component, and we therefore refer to these designs as mixed designs. We distinguish between different mixed designs by the level of mixing, which is defined as the number of varieties arranged in the loop component. For example, in the design of Table 2.2, the first three varieties are arranged in a loop, hence Mix(3). Note that the design in Table 2.4 is a special mixed design, Mix(1), and another special case is the loop design, Mix(ν). Kerr and Churchill (2001) referred to Mix(1) as the augmented reference design. For all designs in $D(\nu, b, k)$ where $b = \nu$ and $k = 2$, the loop design is E-optimal for $3 \leq \nu \leq 6$, while mixed designs Mix(1), Mix(2) and Mix(3) are all E-optimal for $\nu \geq 7$. A proof for these results is in Section 2.4 of Chapter 3.

Kerr and Churchill (2001) performed a computer search for $3 \leq \nu \leq 8$ and found that the loop design is A-optimal. We applied the algorithm by Tjur (1993) to search for A-optimal designs when $9 \leq \nu \leq 26$. The optimal designs are all from the class of mixed designs. In most cases, Mix(3) is A-optimal when $\nu \geq 9$, except that Mix(4) is A-optimal for $9 \leq \nu \leq 11$. Designs Mix(3) and Mix(4) both are A-optimal when $\nu = 12$. We also note that the A- and E-criteria do not produce the exact same optimal designs.

Table 2.4: A Special Case of Mixed Design: Mix(1)

1	1	1	1	...	1	1
1	2	3	4	...	$\nu - 1$	ν

2.3 Optimality Results from Row-Column Design

In row-column designs, the experimental units are grouped in two directions, i.e., two blocking factors are used with one factor representing the rows of the design and the other factor the columns (Shah and Sinha, 1989). The designs considered here have ν varieties arranged in k rows and b columns such that variety i is replicated r_i times. For the design of cDNA microarray experiments, the rows represent the two fluorescent dyes and columns represent arrays, and $k = 2$ always in a two-color experiment.

We now let $D(\nu, b, k)$ denote the class of connected designs having ν varieties arranged in b columns and k rows, and we search for A-optimal designs within this class. Very little is known about the A-optimality of row-column designs, although some sufficient conditions have been studied by Bagchi and Shah (1989), Jacroux and Ray (1991), Shah and Sinha (1993), but these results are not applicable to the case $k = 2$.

We focus our search for optimal row-column design on the following scenarios: (1) $b = \nu$; (2) $b = 2\nu$; (3) $b = 3\nu$; (4) $b = 4\nu$, where b is the number of columns or arrays and k is the number of rows (dyes) or the column block size. We chose these scenarios to allow for two-fold to four-fold replication, as the importance of replication is being increasingly recognized (Callow *et al.*, 2000; Pritchard *et al.*, 2001). We restricted our search to designs in which each variety is labelled with each dye equally often. Because varieties and rows are orthogonal, such a design is called row-orthogonal design (John and Williams, 1995). When there is balance between dyes and varieties, systematic bias in the red and green intensities can be reduced (Yang and Speed, 2002). A special case of row-orthogonal design is a dye-swap experiment where gene expression is compared across two varieties, and the hybridization is repeated with the dye reversed. Jacroux and Ray (1991) provide a result, which bridges the optimality results between block design and row-column design and can be used as follows. Within incomplete block design, if the A-optimal block design is equireplicate and the number of replicates is even (i.e., each variety is replicated a constant and even number of times), then the corresponding row-orthogonal design (which balances varieties with respect to rows) is the A-optimal row-column design. If the A-optimal block design is not equireplicate or even, then no corresponding row-orthogonal design exists and the optimal row-column design is unknown.

When blocks can be grouped such that each variety is replicated exactly once in each group, a resolvable row-column design is obtained (John and Williams, 1995). Methods for construction of resolvable designs can be found in Furino *et al.* (1996). Resolvable designs are important in practice, because it is often useful to perform an experiment one replicate at a time. Resolvability is important for microarray experiments where many arrays are used and all hybridizations cannot be completed at one time. A resolvable design accounts for temporal changes such as different batches of arrays or print-tips, and changes in the concentration of spotting material (Pritchard *et al.*, 2001). As an example, Table 2.5 presents an optimal design for $\nu = 6$ and $b = 12$, which is resolvable in each set of 3 arrays.

Table 2.5: Resolvable Inter-woven Loop Design ($\nu = 6$)

1	2	6	1	3	2	3	4	5	4	6	5
5	3	4	6	4	5	1	2	6	1	2	3

Based on row-orthogonality and resolvability, our strategy of searching for best designs is as follows: (1) We search for the A-optimal design in the context of block design for given b and ν , which is based on the optimization technique by Tjur (1993). If this design is not equireplicate, a row-orthogonal design cannot be constructed; otherwise we determine whether this design is resolvable. (2) Based on the results of Jacroux and Ray (1991), we treat columns as slides and re-arrange symbols of varieties in each slide so that each sample is replicated an equal number of times in each row (dye).

When $b = \nu$, the loop design is row-orthogonal and A-optimal under row-column design. For $\nu \geq 9$, the A-optimal (under block design) mixed designs are not equireplicate, hence optimal row-column designs are unknown. When $b = \nu$, there is no resolvable row-orthogonal design due to the limited number of blocks for the experiment.

When $b = 2\nu$, Kerr and Churchill (2001) found the interwoven loop design to be A-optimal for $\nu \leq 11$ under block design. Our search showed that for $\nu \geq 10$, the interwoven loop design is no longer A-optimal under block design (an example is provided in Section 2 of Chapter 3). Therefore, in the context of row-column design the interwoven design is very likely not A-optimal as $\nu > 10$. With the interwoven loop design being A-optimal for $3 \leq \nu \leq 9$, resolvable designs only exists when ν is even, and are listed in Section 3.1 of Chapter 3.

When $b = 3\nu$, certain designs are universally optimal, because they can be constructed as a generalized Youden design (GYD) (Shah and Sinha, 1989), which applies to cases $\nu = 3$, $\nu = 4$ and $\nu = 7$. Our search showed that the A-optimal block designs are row-orthogonal for $3 \leq \nu \leq 14$, which are listed in Section 3.2 of Chapter 3. Resolvable design again exists for even ν .

When $b = 4\nu$ and $3 \leq \nu \leq 20$, A-optimal block designs are row-orthogonal. Generalized Youden designs (GYDs) have been found for cases $\nu = 3$, $\nu = 5$ and $\nu = 9$, implying that these designs are universally optimal (Kiefer, 1975). Again resolvable designs exist for even ν . A complete catalogue of optimal designs is in Section 3.3 of Chapter 3.

2.4 Applications of Optimal Designs to Microarray Experiment

In Sections 2 and 3, the search for optimal designs within the framework of incomplete block design and row-column design was presented. We now apply these designs to microarray experiments using fixed and mixed linear models, and we evaluate relative efficiencies of these designs.

2.4.1 Fixed Effects Models and Definition of Relative Efficiency

We consider two fixed models along the lines of Kerr and Churchill (2001). The first model (M1) includes the main effects of arrays (A), dyes (D), varieties (V), genes (G), the variety-by-gene interactions (VG), and the array-by-gene interactions (AG). The second model (M2) omits the D factor and is suitable for analyzing data from a reference design.

$$y_{ijkgl} = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkgl} \quad (\text{M1})$$

$$y_{ikgl} = \mu + A_i + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \epsilon_{ikgl}. \quad (\text{M2})$$

Response y_{ijkgl} is the log-transformed fluorescence intensity for array i , dye j , variety k and gene g . There are N genes, b arrays and ν varieties, and ϵ_{ijkgl} are identically and independently normally distributed with mean zero and constant variance σ_e^2 . Furthermore, each gene is present only once on any array. We want to make inferences about differential expression of any gene g between any pair of varieties i and j . The test statistic for differential expression is based on the contrast $(\hat{V}_i - \hat{V}_j) + ((V\hat{G})_{ig} - (V\hat{G})_{jg})$, where $i \neq j$ and $i, j = 1, 2, \dots, \nu$. Note that this contrast differs from that of Kerr and Churchill (2001) for reasons discussed in Yang *et al.* (2003).

We also note that in loop and interwoven designs, the dye and variety effects are orthogonal, however, in the mixed designs they are not orthogonal. Hence, variety effects must be estimated by accounting for dye effects simultaneously.

Incomplete block designs are usually compared based on the *efficiency factor* (Shah and Sinha, 1989, pp. 45) which is the ratio of the average variance for a Randomized Complete Block Design (RCBD) to the average variance of the estimated pairwise or elementary variety contrasts. In the context of two-dye cDNA microarray experiments, calculation of average variance for a RCBD is not meaningful. Another measure is the notion of relative efficiency (RE), which has been used to measure the effectiveness of blocking for block designs (Hinkelmann and Kempthorne, 1994). Relative efficiency of design 1 (d_1) to design 2 (d_2) is defined as the ratio of the variance of a given contrast for d_1 to the variance of the same contrast for d_2 . But we assume that in a microarray experiment, we are (equally) interested in all pairwise contrasts. Therefore, we define a measure of performance for two microarray

designs d_1 and d_2 by combining the definition of efficiency factor and relative efficiency as follows:

$$RE(d_1 \text{ to } d_2) = \frac{Eff(d_1)}{Eff(d_2)} \quad (2.1)$$

where the efficiency of design d , $Eff(d)$, is defined under a particular model as:

$$Eff(d) = \left\{ \frac{2}{\nu(\nu-1)} \sum_{i < j} Var[(\hat{V}_i - \hat{V}_j) + ((V\hat{G})_{ig} - (V\hat{G})_{jg})] \right\}^{-1} \quad (2.2)$$

Note that with all genes represented equally on all arrays, $Eff(d)$ needs to be considered only for a single gene g . As a function of the number of genes (N) and the error variance (σ^2), $Eff(d)$ can be evaluated numerically using the information matrix under (M1) or (M2). We also note that the ratio RE in (2.1) does not depend on N and σ^2 . Finally, we note that had we used the comparison criterion of Kerr and Churchill (2001) instead of (2.2), the same relative efficiency values would have been obtained.

2.4.2 Relative Efficiencies of Mixed Designs

Figure 2.1 lists the relative efficiencies of the reference, loop and mixed designs, relative to the best design under model (M1) and for numbers of varieties ranging from 7 to 25. For the mixed designs, Mix(3) and Mix(4) perform much better than loop and reference designs when the number of varieties exceeds 8, but they perform worse than the loop design when $\nu \leq 8$. Mix(4) is superior to Mix(3) when $\nu = 9, 10, 11, 12$, while Mix(3) outperforms Mix(4) when the number of varieties exceeds 12. Finally, A-optimality in the block design setting of Mix(3) and Mix(4) is more consistent with these results than their E-optimality.

When $\nu \geq 9$, efficiency increases with decreasing level of mixing. The relative efficiencies of mixed designs and the reference design converge as ν becomes large. This finding implies that the reference design is a robust design for microarray experiments with many varieties, when the number of arrays is limited to a value close to the number of varieties, as conjectured by Churchill and Oliver (2001).

2.4.3 Relative Efficiencies of Row-Orthogonal Designs

For each of the three scenarios ($b = 2\nu$, $b = 3\nu$, $b = 4\nu$) considered earlier, we compute the relative efficiency of the replicated reference design for each ν , relative to the optimal row-column designs. These results are summarized in Figure 2.2.

In most cases, the efficiency of the replicated reference design is only between one third and one half of the efficiency of the optimal row-orthogonal design, but the relative efficiency of the reference design increases with the number of varieties. For much larger ν , the reference

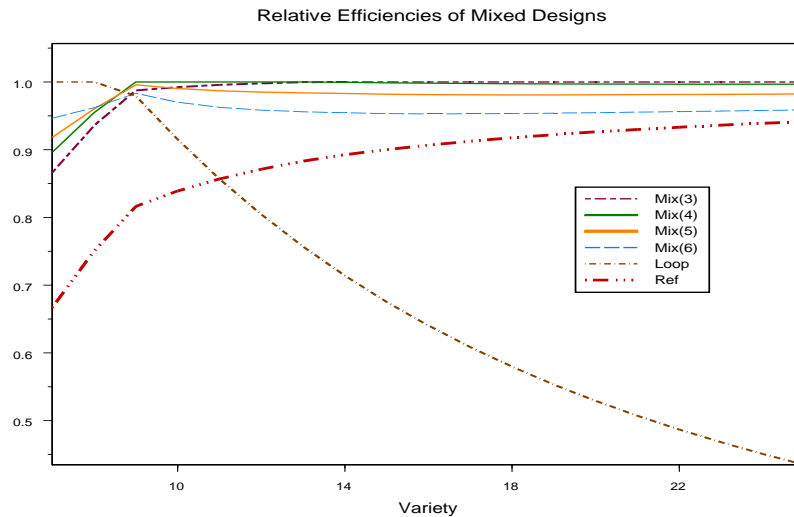


Figure 2.1: Relative Efficiencies of Certain Designs including Loop and Reference Design under Fixed Effects Model M1

design may again become efficient for a given scenario. For those values ν , where the optimal row-column design was not identified in Section 3, we still constructed the row-orthogonal interwoven design and computed the relative efficiency of the replicated reference design. Although not optimal, the interwoven designs still outperformed the replicated reference designs. For example, as $b = 2\nu$ and $\nu = 20$, the efficiency of the replicated reference is only about 58 percent. Furthermore, for fixed ν , as the number of arrays (b) increases, the relative efficiency of the replicated reference design decreases slightly.

2.4.4 Relative Efficiencies under Mixed Effects Models

Up to this point, the relative efficiencies of various designs have been compared under fixed effects ANOVA models, i.e., variety comparisons are based on within block (array) information only. However, the contrasts of interest can also be estimated using additional inter-block information by treating blocks (array) as random. Therefore, relative efficiencies of different designs are now compared based on mixed effects models, which are obtained by treating A and AG in models (M1) and (M2) as random factors, with A_i being iid $(0, \sigma_a^2)$ for all i , and $(AG)_{ig}$ iid $(0, \sigma_{ag}^2)$ for all i and g . We define $\lambda = \sigma_e^2 / \sigma_a^2$ to be the relative variability in arrays and hold $\sigma_e^2 / \sigma_{ag}^2$ constant at 1.

Four scenarios were considered ($\lambda=0.1, 0.33, 1$ and 10) for $7 \leq \nu \leq 25$. Results for the comparison among reference, loop, and mixed designs ($b = \nu$) are summarized in Figure 2.3. Except for a very few cases, the reference design is again outperformed by the other designs. As expected, the relative efficiencies of these designs depend on λ . When the variability of

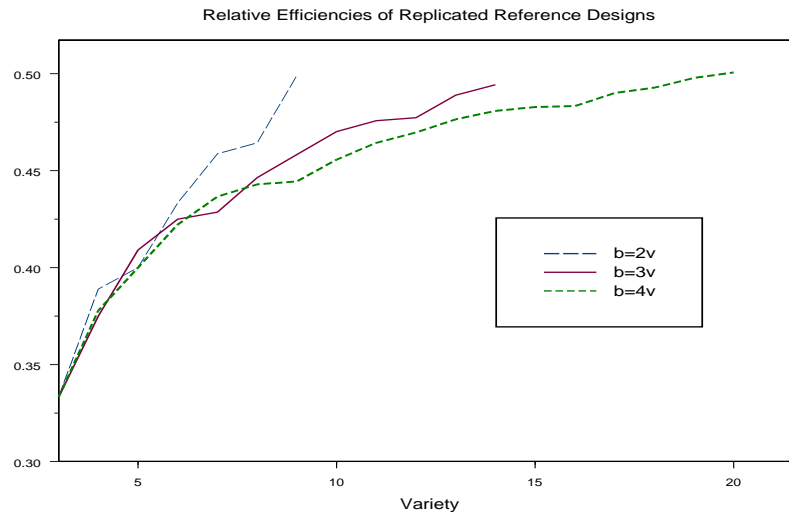


Figure 2.2: Relative Efficiencies of Replicated Reference Design under Three Scenarios

arrays is relatively small ($\lambda=1, 10$), the loop design is highly efficient even for large ν . The mixed designs with a high level of mixing (e.g., Mix(5), Mix(6)) have relative efficiency that decreases as ν increases. The relative efficiency of the reference design and mixed designs with a lower level of mixing (e.g., Mix(3)) increases as ν increases for $\lambda = 1$.

When $\lambda=0.33$, the loop design is still most efficient, but the relative efficiency of the reference and mixed designs approaches that of the loop design as ν increases. When $\lambda=0.1$, the loop design is highly efficient for $\nu \leq 18$, but its efficiency decreases drastically as $\nu > 18$. The efficiency of mixed designs converges to 1 very quickly. In this case, the block effects are large, and the amount of information recovered from inter-block comparisons is small. As λ decreases, the loop design becomes less efficient. We note that the relative efficiency of the reference design slightly improves as σ_{ag}^2 increases.

We also evaluated the relative efficiency of replicated reference designs, relative to optimal row-column designs as in Section 4.3, but under the mixed effects model. The results are summarized in Figure 2.4. In a given scenario (b, ν), the relative efficiency of the reference design decreases as λ increases, implying that the gain in efficiency from the random treatment of the blocking (array) factors is less for the reference design is less than that of the row-orthogonal designs. The reference design is more efficient under mixed effects model with relatively small σ_a^2 (or larger λ) than under fixed effects model for $b = 4\nu$ and a given ν . However, it is less efficient for $b = 2\nu$ and a given ν under the mixed effects model for given ν . Again, the relative efficiency of the replicated reference design slightly improves as σ_{ag}^2 increases, implying it is more robust to the increased spot effects.

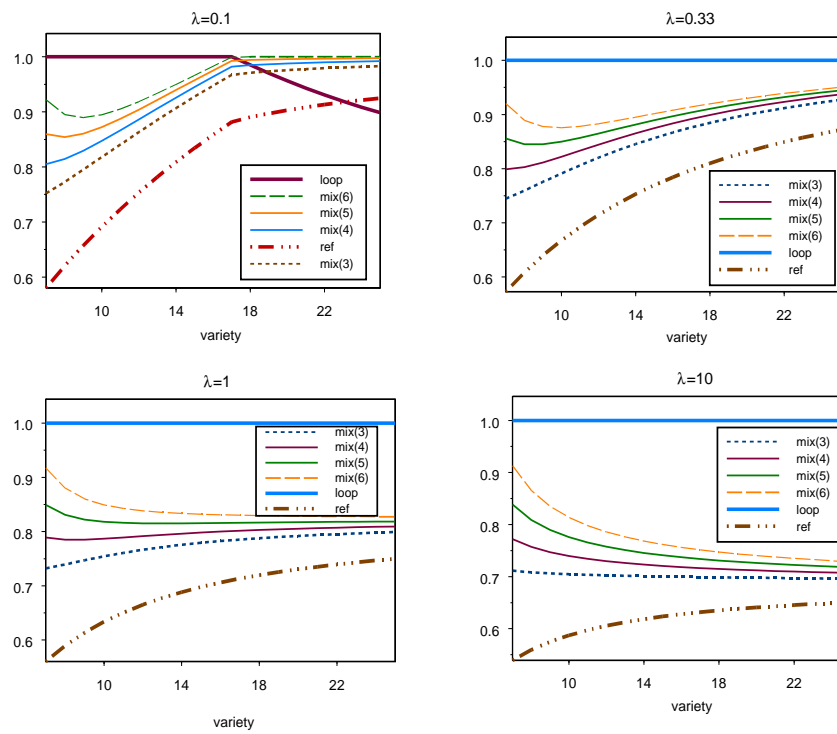


Figure 2.3: Relative Efficiencies of Mixed, Loop and Reference Designs under Mixed Effects Model with $\lambda = \sigma_e^2/\sigma_a^2$ set at 0.1, 0.33, 1 and 10.

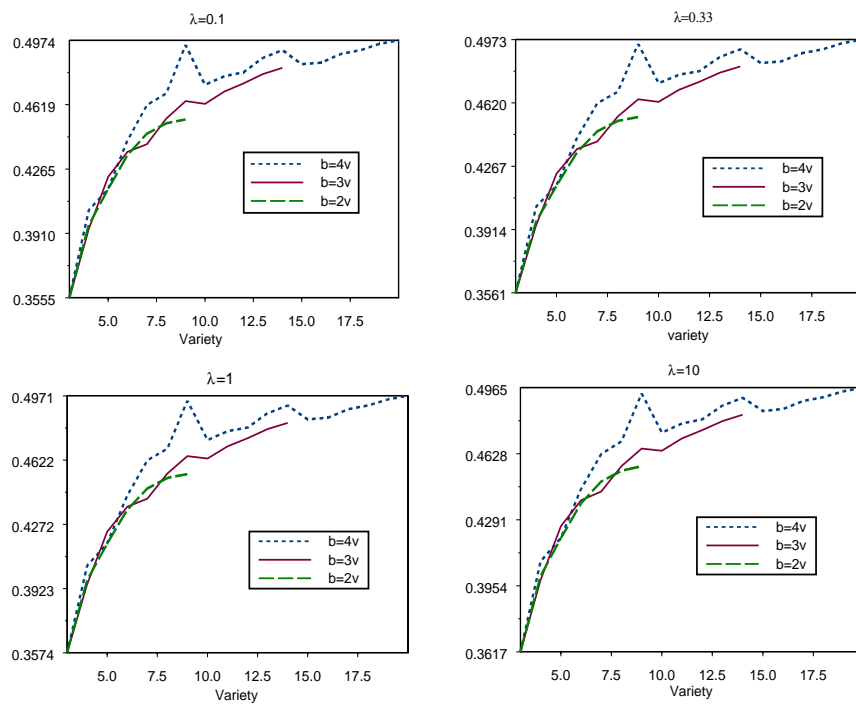


Figure 2.4: Relative efficiencies of the replicated reference design, relative to optimal row-column design for different numbers of arrays (b) at any given number of varieties (ν) and for different ratios (λ) of array to residual variance.

2.5 Comparison of Classification Performance among Designs

Until now, our comparison among designs has focused on identification of differential expression. We now turn to another goal of microarray experiments: sample classification. The (replicated) reference designs have been widely used in classification studies (e.g., Veer *et al.*, 2002). Although replicated reference designs were shown to be inferior to optimal row-column and interwoven designs for identification of differential expression, this result does not necessarily extend to the performance of these designs for sample classification. In fact, the loop design was found to outperform the reference design for $\nu \leq 8$, but for clustering of samples, the performance of the loop design has been shown to be inferior to that of the reference design (Dobbin and Simon, 2002). We now investigate the classification performances of optimal row-column or interwoven designs and of the replicated reference design under different scenarios.

Unlike the single factor experiment, microarray experiments used for classification profile samples taken from biological replicates of at least two different types of cell populations (e.g., cancerous tissues versus normal tissues). Therefore, in the linear fixed or mixed models the variety effect is structured as a fixed type (T) factor (with two or more levels) and random factor (S) representing the effects of biological replicates within each type. Then, the model for the optimal row-column or interwoven design (referred to as interwoven designs from here on) is

$$\begin{aligned} y_{ijklgm} &= \mu + A_i + D_j + T_k + S_l(T_k) + G_g \\ &+ (AG)_{ig} + (TG)_{kg} + G_g \times S_l(T_k) + \epsilon_{ijklgm} \end{aligned} \quad (2.3)$$

where y_{ijklgm} is the transformed intensity from array i , dye j , type k , biological replicate l and gene g . Under this model, dispersion assumptions are $A_i \sim \text{iid } N(0, \sigma_a^2)$, $S_l(T_k) \sim \text{iid } N(0, \sigma_{ts}^2)$, $AG_{ig} \sim \text{iid } N(0, \sigma_{ag}^2)$, $G_g S_l(T_k) \sim \text{iid } N(0, \sigma_{gts}^2)$, and $\epsilon_{ijklgm} \sim \text{iid } N(0, \sigma_e^2)$.

For the replicated reference design, a common reference sample for all hybridizations is often favored by biologists (Dudoit *et al.*, 2002). If the reference is treated as another type of cell population (T_0), then the above model can be used for the replicated reference design by dropping the term D_j . However, because variation among reference samples is to be minimized, variance of S within T_0 would be much smaller and have to be estimated separately from the variance of S within other levels of T. Here, we model the difference between the two log intensities (or log ratios) from the same spot, or $w_{klg} = y_{ijk(1)} - y_{ijk(0)}$, where subscript 0 denotes the reference sample and 1 an individual sample of interest. This would result in the following ANOVA model:

$$w_{klgm} = \mu + T'_k + S_l(T'_k) + G_g + (T'G)_{kg} + G_g \times S_l(T'_k) + \epsilon_{klgm} \quad (2.4)$$

where $T'_k = T_k - T_0$, and T_0 denotes the main effect for the reference samples. Under this model, dispersion assumptions are $S_l(T'_k) \sim \text{iid } N(0, \sigma_{ts}^2)$, $GS(T') \sim \text{iid } N(0, \sigma_{gts}^2)$, and $\epsilon_{klgm} \sim \text{iid } N(0, \sigma_e^2)$.

Table 2.6: Average Number of Misclassifications for Different λ Values under Each Scenario: standard errors were reported in brackets.

Scenarios	Designs	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$	$\lambda = 4$
$b = 2\nu$	Interwoven Loop	1.42 (1.24)	2.00 (1.28)	2.02 (1.31)	2.50 (1.41)
	Rep. Reference	0.48 (0.88)	0.77 (1.00)	0.95 (1.05)	2.02 (1.47)
$b = 3\nu$	Interwoven Loop	0.07 (0.22)	0.84 (1.08)	1.25 (1.13)	1.90 (1.41)
	Rep. Reference	0.45 (0.88)	0.73 (0.94)	1.05 (1.12)	1.58 (1.52)
$b = 4\nu$	Interwoven Loop	0.15 (0.46)	1.09 (1.21)	1.48 (1.19)	2.07 (1.33)
	Rep. Reference	0.51 (0.88)	0.92 (1.09)	1.20 (1.18)	1.95 (1.45)
$b = 5\nu$	Interwoven Loop	0.19 (0.49)	1.12 (1.22)	1.42 (1.08)	2.17 (1.45)
	Rep. Reference	0.54 (0.87)	0.91 (1.03)	1.13 (1.12)	2.10 (1.45)

We consider the comparison under four scenarios, $b = 2\nu$, $b = 3\nu$, $b = 4\nu$ and $b = 5\nu$, where $\nu = 20$. We assumed samples are from two classes or types of cell populations, each having 10 biological replicates. Expression data were simulated for 20 genes, 10 of which were differentially expressed between the two classes. Parameters used to generate the data were estimated from a subset of a bovine embryo classification study described in experimental designs in Yang *et al.* (2003). Other details of the simulation are provided in Section 4.2 of Chapter 4. We used Fisher’s linear discriminant method to obtain the classifier and the method of cross validation (*leave-one-out*) was used to measure performance by counting the number of misclassifications.

We first examined the impact of gene-specific sample variation (σ_{gts}^2) on the performance of both designs. Let $\lambda = \sigma_{gts}^2/\sigma_e^2$. We considered four cases of λ (0.5, 1, 2, 4) while other variance components were fixed. The results appear in Figures 2.5, 2.6, 2.7, 2.8. The classification performance of both designs depends on λ . For each scenario, their performances decline as λ increases. When $b = 2\nu$, the replicated reference design always performs better for the given range of λ (Figure 2.5). When $b = 3\nu$, 4ν and 5ν , the interwoven loop design outperforms the replicated reference design for $\lambda=0.5$, but the replicated reference design is superior to the interwoven loop design for other λ values (1, 2, 4) (Figures 2.6, 2.7, 2.8). Given λ , the performance of both designs varies as b changes. The number of misclassifications for each design declines from $b = 2\nu$ to $b = 3\nu$, but increases as b increases from 3ν to 5ν . This suggests that given the 20 samples to classify, the best scenario is $b = 3\nu$, because both designs outperform other scenarios for each λ .

Variation among biological replicates (σ_{ts}^2), also affects the classification performance of both designs. Define $\delta = \sigma_{ts}^2/\sigma_e^2$. We also considered four cases of δ (0.5, 1, 2, 4) while other variance components were fixed ($\lambda=1$). The results are summarized in Table 2.7 and Figures 2.9, 2.10, 2.11, 2.12. As expected, the number of misclassifications increases under each scenario, as δ increases. When $\delta=0.5$, the interwoven loop design is outperformed by

Table 2.7: Average Number of Misclassifications for Different δ values under Each Scenario: standard errors were reported in brackets.

Scenarios	Designs	$\delta=0.5$	$\delta=1$	$\delta=2$	$\delta=4$
$b = 2\nu$	Interwoven Loop	2.09 (1.33)	2.42 (1.39)	3.19 (1.49)	3.69 (1.33)
	Rep. Reference	1.22 (1.21)	2.10 (1.43)	2.74 (1.51)	3.63 (1.51)
$b = 3\nu$	Interwoven Loop	1.08 (1.13)	1.60 (1.35)	2.21 (1.51)	2.99 (1.58)
	Rep. Reference	1.11 (1.18)	1.90 (1.43)	2.77 (1.50)	3.41 (1.57)
$b = 4\nu$	Interwoven Loop	1.30 (1.24)	1.92 (1.50)	2.51 (1.58)	3.30 (1.54)
	Rep. Reference	1.28 (1.18)	2.02 (1.49)	2.93 (1.57)	3.69 (1.53)
$b = 5\nu$	Interwoven Loop	1.56 (1.30)	2.09 (1.48)	2.61 (1.49)	3.43 (1.57)
	Rep. Reference	1.42 (1.30)	2.21 (1.57)	3.05 (1.57)	3.66 (1.42)

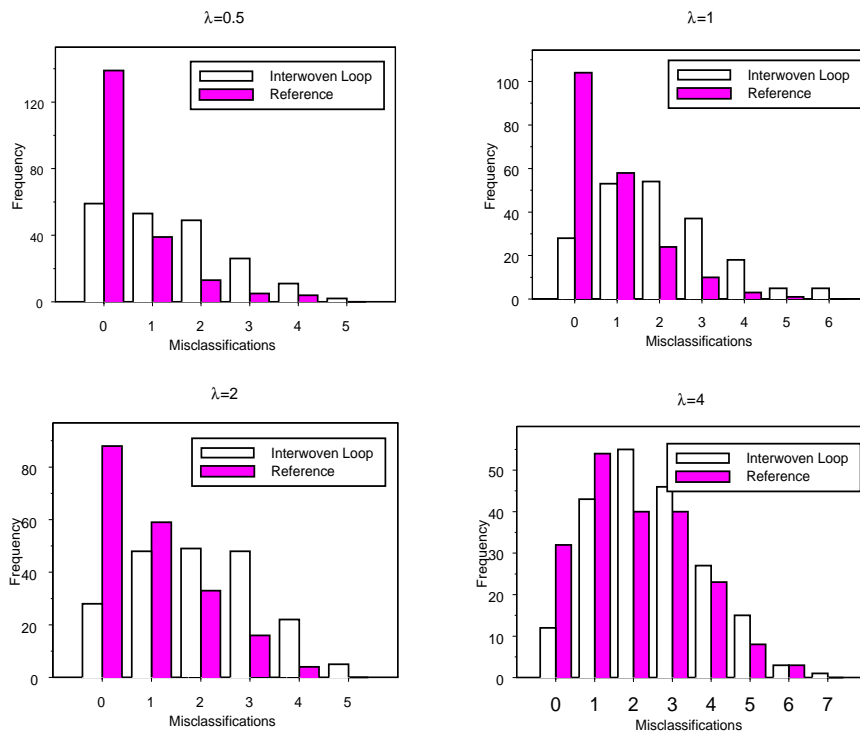


Figure 2.5: Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 2\nu$, $\nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.

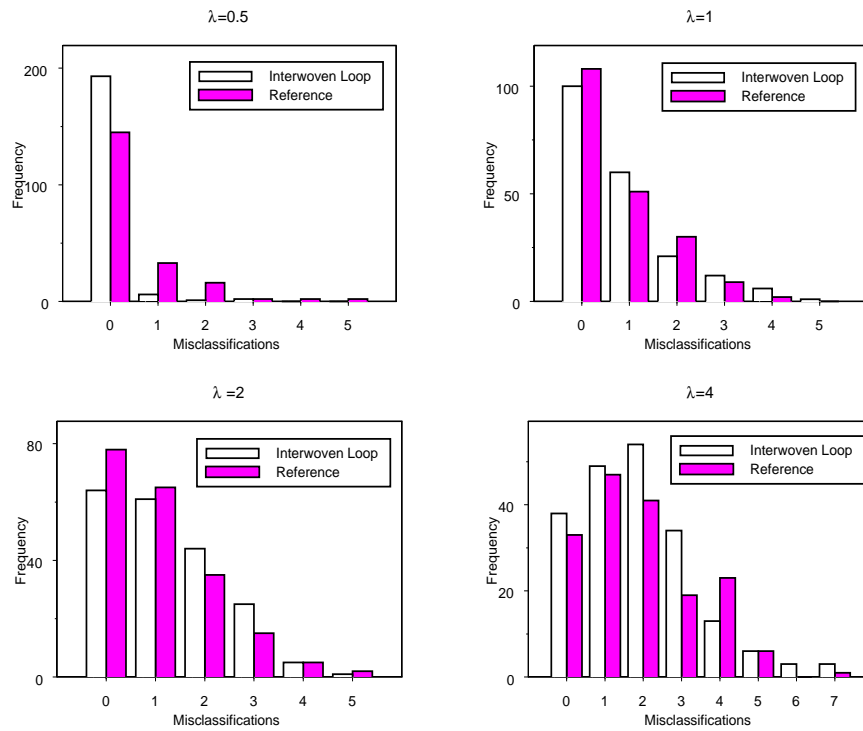


Figure 2.6: Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 3\nu$, $\nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.

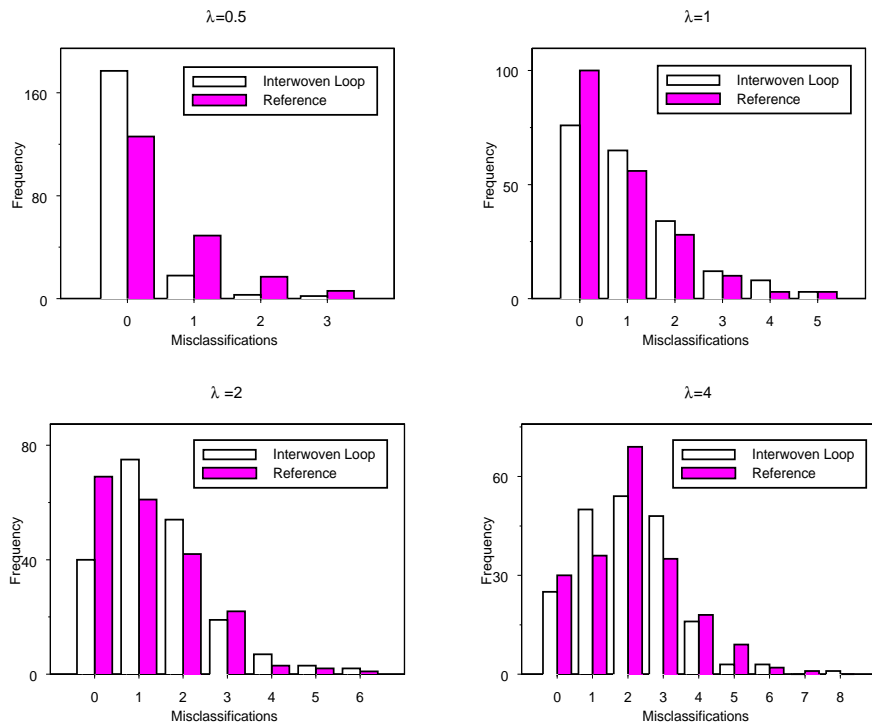


Figure 2.7: Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 4\nu$, $\nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.

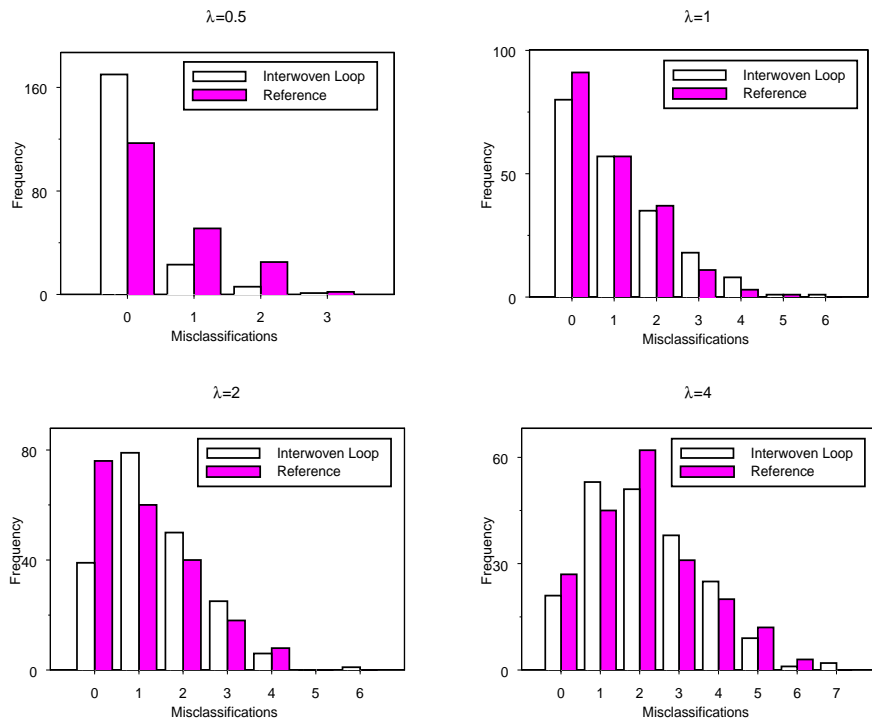


Figure 2.8: Frequency of misclassifications for the interwoven loop design and the replicated reference design ($b = 5\nu$, $\nu = 20$) for different λ values: 0.5, 1, 2, 4. Average number of misclassifications was reported in Table 2.6, and results were based on 200 replicated data sets.

the replicated reference design for $b = 2\nu$ and $b = 5\nu$, but both designs perform almost equally well for $b = 3\nu$ and $b = 4\nu$. For other δ values, the interwoven loop design is superior to the replicated reference design for $b = 3\nu$, $b = 4\nu$ and $b = 5\nu$, but is outperformed by the replicated reference design for $b = 2\nu$. Among all scenarios, $b = 3\nu$ still performs better than other scenarios, implying it is more robust to the increased σ_{ts}^2 . The results also show that the interwoven loop design is more robust to the increased sample variations as the range of average misclassifications is lower than in the reference design. This is likely because the interwoven loop design uses more observations than the replicated reference design to estimate the variation among biological replicates, resulting in a higher precision of estimating σ_{ts}^2 .

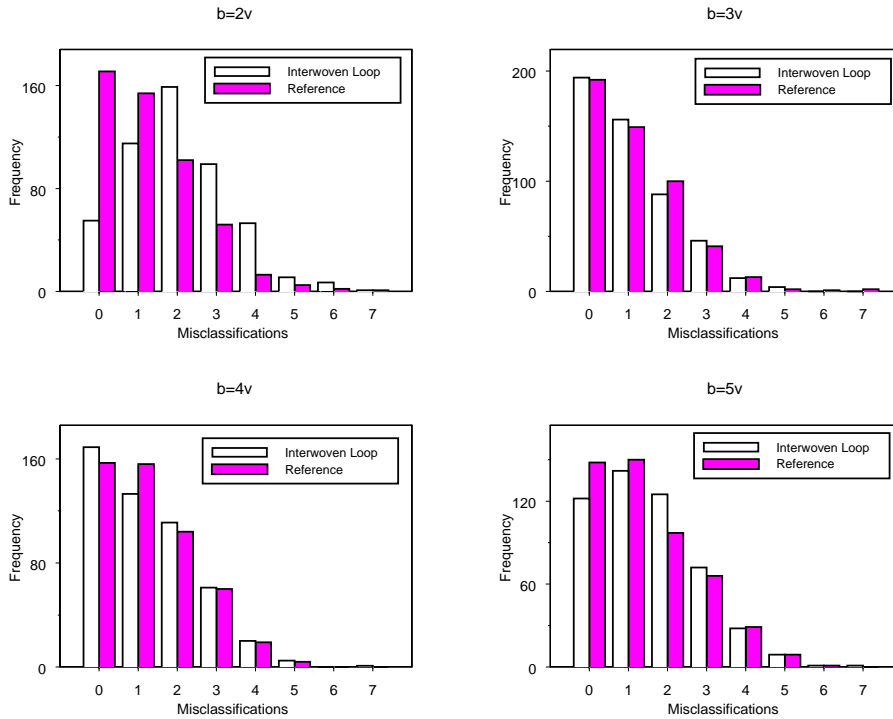


Figure 2.9: Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=0.5$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.

There are several reasons behind these phenomena. In classification analysis, the classifier is estimated using the measured distances among samples. The classification performance of designs also depends on the number of differentially expressed genes and how many of them are used for classification process. Including too many genes into the classification process introduces noise, and this worsens the classification performance.

In the true state, 10 genes were assumed to be differentially expressed across the cell popu-

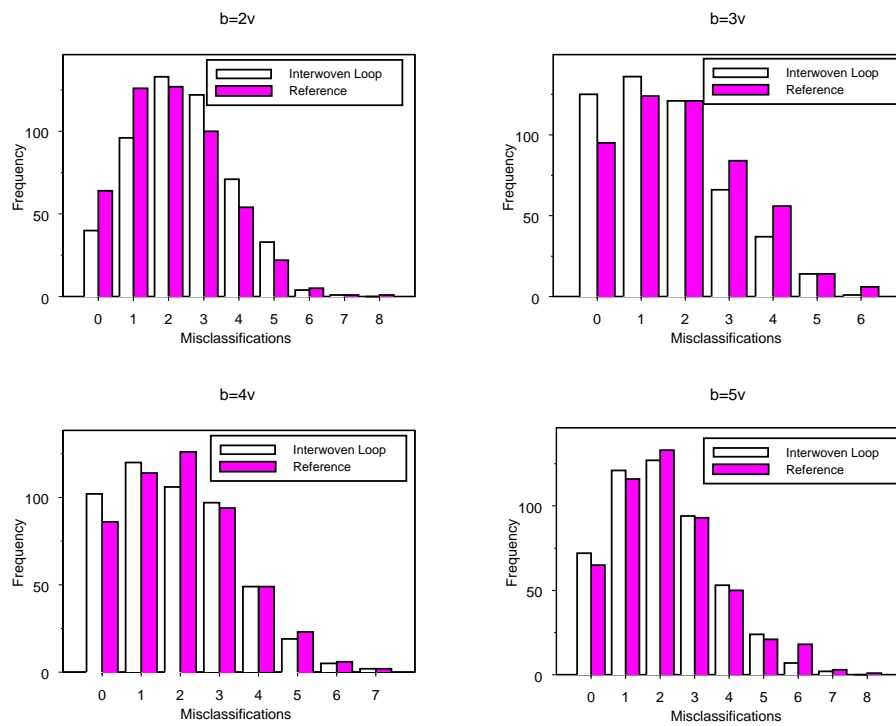


Figure 2.10: Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=1$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.

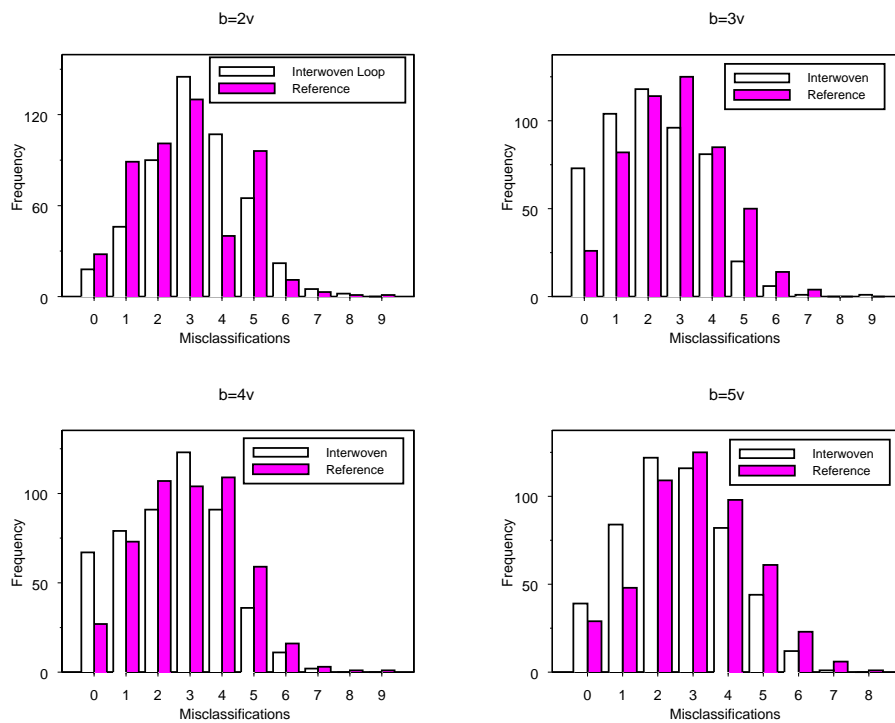


Figure 2.11: Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=2$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.

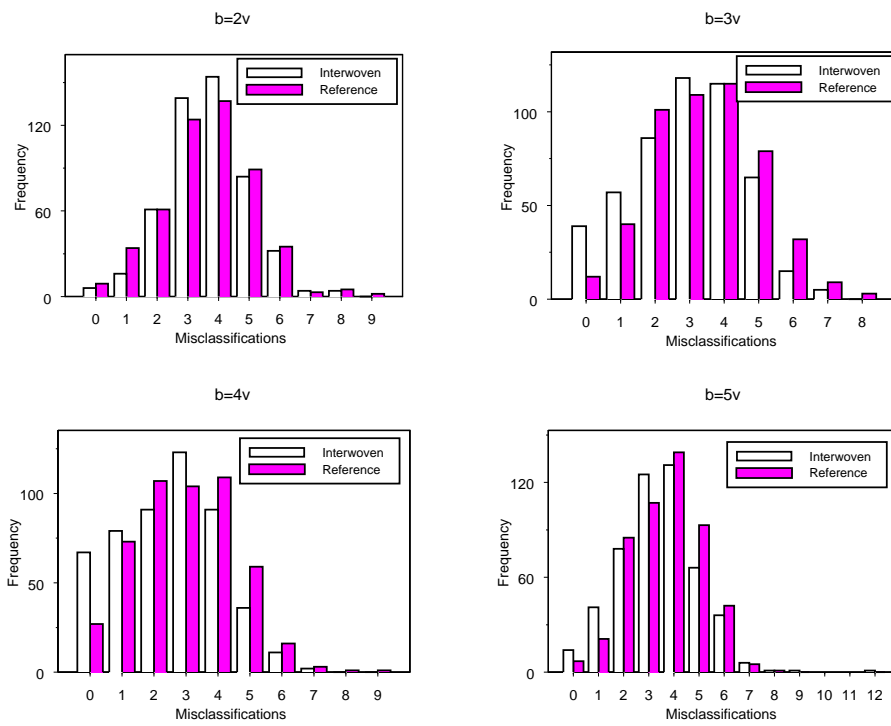


Figure 2.12: Frequency of misclassifications for the interwoven loop design and the replicated reference design under four scenarios when $\delta=4$. Average number of misclassifications was reported in Table 2.7, and results were based on 500 replicated data sets.

lations, but only a few genes (Gene 12, 13, 14, 16, 20) contribute the most to the separation of the cell populations. The interwoven loop design is certainly preferred for detecting gene differential expression, because the variances of pairwise contrasts among samples are always smaller than those for the replicated reference design (see the summary in Section 4.2 of Chapter 3). On the other hand, the variances of pairwise contrasts among samples are constant for the replicated reference design, thus the distances among samples are less variable than those under the interwoven loop design. Although the replicated reference design detects fewer differentially expressed genes, these genes are more likely to be those that contribute the most in the separation of the cell populations, thus the reference design performs better than the interwoven loop design. This suggests that the problem of gene selection in the context of sample classification is not identical to identification of differential expression. For sample classification, a subset of those differentially expressed genes may suffice. Our findings also imply that although statistical optimality criteria are useful for designing microarray experiments for detecting gene differential expression, design of microarray experiments for sample classification should be based on other criteria.

2.6 Conclusion

Kerr and Churchill (2001) established A-optimality for the loop design in the $b = \nu$ scenario, and for the interwoven design in the $b = 2\nu$ scenario for small number of varieties. Here, we searched for A-optimal designs for large numbers of varieties, higher levels of replication ($b = \nu, 2\nu, 3\nu, 4\nu$), and in the context of row-column design, which account for two blocking factors typically present in cDNA microarray experiments, arrays and dyes. We identified A-optimal designs for the case of $b = \nu$ in the context of incomplete block design, leading to the class of mixed designs. Subsequently, we identified A-optimal designs in the context of Row-Column design for the replication levels $b = 2\nu$, $b = 3\nu$ and $b = 4\nu$. We also determined which of the optimal design are resolvable and hence capable of accounting for temporal effects in large microarray experiments.

We subsequently compared the performance of the A-optimal designs to standard reference, loop and replicated reference designs for identification of differentially expressed genes, using fixed ANOVA models as well as mixed models treating array and array-by-gene interaction as random. In the $b = \nu$ scenario, the mixed designs outperformed reference and loop designs for $9 \leq \nu \leq 25$ under the fixed model analysis. However, in the mixed model analysis, with random A and AG effects, the loop design was superior to the reference and mixed designs when the array variance was small. For the replication scenarios, the A-optimal row-column or interwoven loop designs continued to clearly outperform the replicated reference design. Relative efficiency of the replicated reference design improved somewhat when the array variance was small and the level of replications was higher.

We have also shown in a simulation study that for classification analysis a replicated reference design outperforms interwoven designs when the gene-specific sample variation (σ_{gts}^2)

is large. The interwoven designs are superior to the replicated reference design when variation between biological replicates (σ_{ts}^2) is relatively large. The choice of either design for classification depends on the sizes of both variance components. This implies that although A-optimality is useful for designing microarray experiments for detecting gene differential expression, design of experiments for classification should be based on other criteria. Our findings also indicate that design optimality depends primarily on the goal of the experiment (identification of differential expressed genes or factors influencing gene expressions, versus sample classification), and secondarily on the treatment of array replicated factors as fixed or random in the mixed model analysis.

Chapter 3

Supplement to Optimal Design of Single Factor Microarray Experiments

3.1 Introduction

Microarray gene expression experiments are used to perform gene expression profiling on a large scale (e.g., Derisi *et al.*, 1997; Eisen *et al.*, 1998). Performing and interpreting cDNA microarray expression experiments is not efficient or possible without statistical support at each step of the process. This includes experimental design (e.g., Kerr and Churchill, 2001), image processing (e.g., Yang *et al.*, 2000), identification of systematic sources of variation, identification of differentially expressed genes (e.g., Dudoit *et al.*, 2000), clustering of genes and samples or class discovery (e.g., Getz *et al.*, 2000), sample classification (e.g., Dudoit *et al.*, 2001), and pathway and gene network inference (e.g., Pe'er *et al.*, 2002; Henderson *et al.*, 2002). Design of valid and efficient experiments plays a critical role in the interpretation of the data from microarray experiments, since efficiency and reliability of obtained gene expression data can potentially be greatly improved (Yang and Speed, 2002).

Statistical design recommendations for microarray experiments are still rather scarce. As noted by Smyth *et al.* (2002), it is not possible to give universal recommendations for all situations but the general principles of statistical experimental design apply to all microarray experiments. These authors point out that choice of design depends not only on statistical efficiency, but also on the goals of the experiment, which determines the importance of different comparisons, as well as on the available resources (number of arrays, amount of mRNA samples). While it is important to optimally design microarray experiments with factorial structure (Glonek and Solomon, 2002), here we focus on single factor microarray experiments with larger number of varieties. The reason for this focus is that in some microarray experiments we have been involved in, a single treatment or variety factor was of interest, but this factor had in the order of 10 to 25 levels, and in the initial exploratory phase

of the experiments only a very limited number of arrays was available. For the case of limited number of arrays, we assume that the number of arrays available (b) equals the number of varieties (ν), or $b = \nu$. Kerr and Churchill (2001) showed that for this case the loop design is A-optimal if $\nu \leq 8$. Here we show that for larger ν ($8 < \nu < 26$), designs in the class of mixed designs are optimal, although variety levels are not equally replicated in these designs. We then search for optimal designs in the context of row-column design with the constraints of row-orthogonality and equal replication of all variety levels, so that all contrasts are estimated with the same precision. We consider different scenarios for experimental resources in terms of the number of available arrays, $b = 2\nu$, $b = 3\nu$, and $b = 4\nu$. Resolvable designs are also determined, because use of these designs can significantly reduce the temporal effects of hybridizations performed at different times for large microarray experiments. The interwoven loop designs are useful for detecting gene differential expression.

3.2 Optimality Results from Incomplete block Design

3.2.1 Theory on Incomplete Block Design

In the context of statistical block designs, suppose ν varieties are to be compared via b blocks of size k with $k \leq \nu$ (varieties are often referred to as treatments in the design literature). Any arrangement of the ν varieties into the bk experimental units is called a design. The usual additive model specifies the expectation of an observation on variety i in block j to be $\alpha_i + \beta_j$, where α_i and β_j are unknown fixed effects of variety i and block j with $1 \leq i \leq \nu$, $1 \leq j \leq b$. Furthermore, the observations are assumed to be uncorrelated with common variance σ^2 .

For any design d , it is well known that the coefficient matrix of the reduced normal equation for estimating the variety effects is

$$C_d = \text{Diag}(r_{d1}, r_{d2}, \dots, r_{d\nu}) - (1/k)N_d N_d' \quad (3.1)$$

where r_{di} is the number of replications of variety i , and N_d is the variety-block incidence matrix, i.e., $N_d = (n_{ij})_{\nu \times b}$, where n_{ij} is the number of times that variety i appears in block j . The matrix C_d is symmetric, nonnegative definite and has zero row sums. It is also known that for all the possible treatment contrasts to be estimable, a design d must be connected, i.e., its information matrix C_d must have rank $\nu - 1$ (Shah and Sinha, 1989). Because connectedness is a desirable property for most block designs, we will restrict our attention to such designs. We denote by $D(\nu, b, k)$ the class of all connected block designs having ν treatments arranged in b blocks of size k .

Various optimality criteria can be defined in terms of C_d (in (3.1)), such as A-, D- and E-optimality. In the present paper only E- and A-optimality criteria will be considered. For each design $d \in D(\nu, b, k)$, let $0 < z_{d1} \leq z_{d2} \leq \dots \leq z_{d\nu-1}$ be the nonzero eigenvalues of C_d .

Definition 3.1. A design d is called E-optimal if it maximizes z_{d1} over $d \in D(\nu, b, k)$.

It is known that d^* is E-optimal if and only if it minimizes the maximum variance of the least squares estimators of normalized treatment contrasts, i.e., d^* minimizes $Var(\sum_{i=1}^{\nu} c_i \hat{\alpha}_i)$, subject to $\sum_{i=1}^{\nu} c_i^2 = 1$ and $\sum_{i=1}^{\nu} c_i = 0$, where $\hat{\alpha}_i$ are least squares estimates of α_i 's. The E-optimality criterion was introduced by Ehrenfeld (1955), and it is appropriate to use in experimental settings where it is desired to estimate all treatment contrasts with as much precision as possible. Thus, the E-optimality is somewhat a minimax criterion.

Another criterion, A-optimality, has been used to search for optimal designs in class $D(\nu, b, k)$. The A-criterion belongs to the more general family of *type-1* criteria, which was introduced by Cheng (1978).

Definition 3.2. $\phi_f(C_d) = \sum_{i=1}^{\nu-1} f(z_{di})$ is a type-1 criterion if f is a convex, real-valued function for which

- (i) f is continuously differentiable on $\left(0, \max_{d \in D(\nu, b, k)} tr(C_d)\right)$ with $f' < 0$, $f'' > 0$, and $f''' < 0$ on this range, and
- (ii) f is continuous at 0 or $\lim_{x \rightarrow 0} f(x) = f(0) = \infty$, where f' , f'' , f''' are the first, second and third order derivatives, respectively.

The A-criterion is a special case of the type-1 criterion in which $f(x) = 1/x$ in the above definition. Intuitively, A-optimality minimizes the average variance of all pairwise variety contrasts. In other words, an A-optimal design provides maximum average precision for the estimation of all variety contrasts.

3.2.2 Construction of E-optimal Designs

Apart from the use of balanced incomplete block design (BIBD), construction of E-optimal designs is either through the standard augmentation or the deletion process. In other words, deletion or addition of blocks of certain types to known E-optimal designs gives additional E-optimal designs. Jacroux (1981a) and Constantine (1981, 1982) gave some general rules for the augmentation process: Adding certain blocks to a BIBD or to a group divisible design (GDD) can result in another class of E-optimal designs. Along the same lines, Jacroux (1982) proposed the deletion process to construct E-optimal designs. In what follows, we first discuss the construction of BIBDs when $k = 2$, and subsequently we present some known construction methods based on BIBD.

In an ideal situation, where there is no constraint on the number of blocks available, a balanced incomplete block design (BIBD) is universally optimal, hence also E-optimal. For $k = 2$ and any number of varieties ν , the BIBD consists of all $\nu(\nu - 1)/2$ pairs of varieties.

Obviously the BIBD is not feasible in practice, because the number of blocks required by this design becomes very large as ν increases beyond 5.

When the number of blocks available is less than the number required by the BIBD, then the standard deletion process can be used to construct E-optimal designs. Several results have been obtained by Constantine (1981), Jacroux (1982), and Sathe and Bapat (1985). Several results are presented below, while the reader is referred to the original papers for proofs.

Method 1. Let $\nu/k^2 \leq s \leq \nu/k$ and s be an integer. If s disjoint binary blocks are deleted from the BIBD(b, ν, k), the resulting design is E-optimal in the class of $D(b-s, \nu, k)$. This result is due to Constantine (1981) and Jacroux (1982).

Method 2. Let $s \leq (\nu - \sqrt{\nu})/(\nu - k)$ be an integer. If any s disjoint binary blocks are deleted from the BIBD(b, ν, k), the resulting design is E-optimal in $D(b-s, \nu, k)$. This result is due to Sathe and Bapat (1985).

In addition to the above results related to BIBD, E-optimal designs can also be constructed from group divisible design (GDD), provided that certain conditions are satisfied. Some sufficient conditions are given by Das (1991) and by Srivastav and Morgan (1998), among others. However, all designs constructed from BIBD or GDD as described above require a large number of blocks.

3.2.3 E-efficiency of Mixed Designs

Because we are particularly interested in the E-optimality of certain mixed designs, we will explore further the properties of the mixed designs in terms of their E-efficiency. For any design $d \in D(\nu, b, 2)$, denote the E-optimal design as d^* . Furthermore, define z_{d1} and z_{d^*1} as the smallest positive eigenvalue of the information matrix for design d and d^* , respectively. Then the E-efficiency of any design $d \in D(\nu, b, 2)$ is defined as

$$E(d) = \frac{z_{d1}}{z_{d^*1}} \quad (3.2)$$

Based on (3.2), we are ready to study E-efficiency in the class of mixed designs for the case $b = \nu$. In Figure 3.1, the E-efficiencies for certain mixed designs are plotted for $7 \leq \nu \leq 26$. Obviously, with Mix(2) and Mix(3) being E-optimal, their E-efficiencies are both 100%. For other mixed designs, their E-efficiency decreases as the number of varieties increases. The E-efficiency of the loop design changes dramatically. Figure 3.1 shows that the E-efficiency of the loop design decreases rapidly with the number of varieties increasing above $\nu = 8$. In contrast, the E-efficiencies of non-optimal mixed designs with relatively low levels of mixing decline slowly as ν increases.

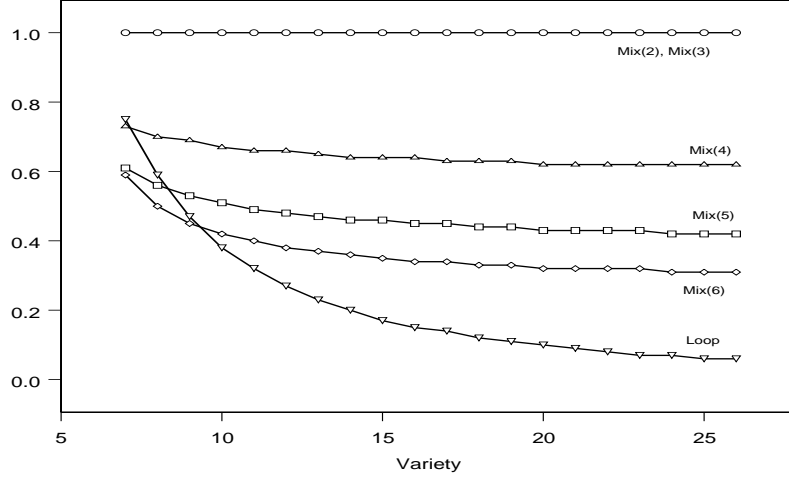


Figure 3.1: E-Efficiency of certain mixed designs

3.2.4 Proof of E-optimality for Certain Mixed Designs

Proof. For notations in this section, we refer to Shah and Sinha (1989) for details. To prove the E-optimality of certain design, several results about bounds for z_{d1} , where $d \in D(\nu, b, k)$, are needed. Let $r = bk/\nu$ and concurrence parameter $\lambda = [r(k-1)/(\nu-1)]$. The results are stated now without proofs.

$$(A1.1) \quad z_{d1} \leq \frac{(k-1)}{2k}(r_{d1} + r_{d2}) + \frac{\lambda_{d12}}{k}, \quad (\text{Jacroux, 1982})$$

$$(A1.2) \quad z_{d1} \leq \min C_{ii}, \quad (\text{Constantine, 1981})$$

$$(A1.3) \quad z_{d1} \leq \left(\frac{\nu}{m(\nu-m)}\right) \sum \sum_{i,i' \in M} C_{dii'}, \quad \text{where } M \text{ denotes some set of } m \text{ subscripts out of } 1, 2, \dots, \nu \text{ and } 1 \leq m \leq \nu. \quad (\text{Jacroux, 1991b})$$

$$(A1.4) \quad z_{d1} \leq \max\left[\frac{r(k-1)-\lambda-2}{(\nu-2)k}\nu, \frac{r(k-1)+\lambda-1}{k}\right], \quad \text{where } d \in D^*(\nu, b, k) \text{ but is not an RGD.} \quad (\text{Cheng, 1980})$$

To prove the E-optimality of Mix(1), Mix(2) and Mix(3) within $D(\nu, b, 2)$ for $\nu \geq 7$, we need to compute z_{d^*1} of the information matrix C_{d^*} , where d^* represents the E-optimal design. We only prove the case for Mix(3). The eigenvalues of Mix(3) are found to be $z_{d^*1} = z_{d^*2} = \dots = z_{d^*\nu-3} = 1/2$, $z_{d^*\nu-2} = 3/2$ and $z_{d^*\nu-1} = \nu/2$. Now we consider any other $d \in D(\nu, b, 2)$, and we want to prove that $z_{d1} \leq z_{d^*1}$. In what follows, parts (I) and (II) deal with optimality of Mix(3) within the class of binary designs, while part (III) considers optimality of Mix(3) in the class of non-binary design. For the class of binary designs, there are two general sub-classes of designs, equireplicate designs considered in part (I) and

nonequireplicate designs treated part (II). Part (IV) deals with the optimality of loop design as $3 \leq \nu \leq 6$.

(I). Equireplicate designs. The replication is $r_{d1} = r_{d2} = \dots = r_{d\nu} = 2$. Designs can be further identified as regular graph designs (RGD) and equireplicate but non-RGD designs.

For $d \in RGD$, we sketch the proof as follows. Due to the small number of blocks available RGD design takes the form of a loop design. The unique pattern of the information matrix C_d and Inequality (A1.3) implies $z_{d1} \leq z_{d*1} = 1/2$ as $\nu \geq 8$. For $\nu = 7$, it can be shown that $z_{d1} \leq 1/2$ by direct computation.

If d is equireplicate but not an RGD, $z_{d1} \leq \text{Max}[0, 1/2] = 1/2 = z_{d*1}$ by Inequality (A1.4).

(II). Non-equireplicate designs ($r_{d1} \leq r_{d2} \leq \dots \leq r_{d\nu}$ and $r_{d1} < r_{d\nu}$).

Two subcases exist: (1) designs with at least two varieties replicated once, and (2) designs with only one variety replicated once.

Case 1: Without loss of generality, we assume that $r_{d1} = r_{d2} = 1$. If varieties 1 and 2 are in the same block, then $\lambda_{12} = 1$; however, the resulting design is disconnected, so this case is ruled out. If they are not in the same block, then $\lambda_{12} = 0$ and upon using Inequality (A1.1), we have $z_{d1} \leq \frac{(r_{d1} + r_{d2})}{2 \times 2} = 1/2 = z_{d*1}$.

Case 2: Without loss of generality, we assume that $r_{d1} = 1$, $r_{d\nu} = 3$ and $r_{d2} = \dots = r_{d\nu-2}$. Then there are two cases: (i) Variety 1 is paired with treatment ν in a block; (ii) variety 1 is not paired with treatment ν .

For Case (i), in addition to one block with varieties 1 and ν , at least four varieties will be arranged in the same fashion as a RGD (see (I)). For this case, the proof will be similar to the case where d is a RGD.

For case (ii), without loss of generality, we assume that varieties 1 and 2 occur only once in a block. Therefore, certain elements of the information matrix will be as follows: $c_{11} = 1/2$, $c_{22} = 1$, and $c_{12} = -\lambda_{12}/2 = -1/2$. In this case, by Inequality (A1.3), it can be seen that $z_{d1} \leq z_{d*1} = 1/2$ for $\nu \geq 4$.

(III). For nonbinary design, it turns out that for $b = \nu$ and $k = 2$, one variety must be replicated at least three times, hence at least one of the other varieties is replicated only once. We distinguish between two cases:

Case 1: Non-binary designs with at least two varieties replicated once. Following Inequality (A1.1), it can be shown that $z_{d1} \leq z_{d*1} = 1/2$.

Case 2: Non-binary design with one variety replicated once. By some combinatorics, it can be seen that there is only one such design with the following layout (variety ν is replicated once):

1	1	2	3	...	$\nu - 2$	$\nu - 1$
1	2	3	4	...	$\nu - 1$	ν

From the information matrix and by using Inequality (A1.3), we can show that $z_{d1} \leq z_{d*1} = 1/2$ for $\nu \geq 7$.

(IV). When $\nu = 3$, the loop is in fact an BIBD, which is universally optimal (Kiefer, 1975). When $\nu = 4$, Cheng (1979) proved the E-optimality of loop design. When $\nu = 5$, the smallest eigenvalue for the loop design is $z_{d*1} = 0.691$, from Inequality (A1.2) we can see that $z_{d1} = 0.625 < z_{d*1}$. When $\nu = 6$, the proof is also similar to the case of (I), (II), and (III). \square

3.2.5 A-efficiency of Mixed Designs

In Table 3.1, we evaluate the *relative efficiency* of various designs from the class of mixed designs and for the nonbinary design Mix(1), relative to the optimal design for each value of ν . Columns contain the relative efficiencies of designs Mix(2), Mix(3), Mix(4), loop, and Mix(1), relative to the optimal design. We note that the A- and E-criteria do not agree with each other. The E-optimal designs Mix(2) and Mix(1) are not A-optimal, while Mix(4), although not E-optimal, is A-optimal when $\nu = 9, 10, 11$. This findings lead to the question about which optimality criterion to use when we search for the best design. In general, there is no consensus on this problem, it is a practical issue, which depends on the purpose of the experiment. We also note that the A-optimal designs listed were found using a computer search algorithm, and hence they may not be the actual optimal design, although they are expected to be highly efficient or near optimal.

3.2.6 Comparison of Our Search with Kerr's Search

Kerr and Churchill (2001) searched for the A-optimal design in the class of equireplicated designs. This is different from our search because we search over the much broader class of connected designs. The loop design is A-optimal within the class of connected designs for ν between 2 and 8 inclusive. Our results showed that the mixed designs are A-optimal when $\nu > 8$ in the class of connected designs. This is the case for $b = \nu$.

Similarly, the optimality results of the interwoven loop design when $b = 2\nu$ by Kerr and Churchill (2001) were obtained from the class of equireplicate designs. Our search showed when $\nu \leq 9$, the interwoven loop design is A-optimal for the class of connected designs, and other unequireplicate designs beat the interwoven loop designs (equireplicate designs) when $\nu \geq 10$. For example, when $b = 20$ and $\nu = 10$, the following design is more efficient than the interwoven loop design given by Kerr and Churchill (2001):

$\frac{1}{4}$	$\frac{1}{10}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{7}$	$\frac{1}{7}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{6}$	$\frac{1}{8}$	$\frac{1}{5}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{8}$
---------------	----------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	---------------	----------------	---------------	---------------	---------------	----------------	----------------	---------------	---------------

Table 3.1: Relative Efficiency of Certain Mixed Designs for $b = \nu$ and $k = 2$: when $\nu=12$, Mix(3) and Mix(4) are both A-optimal, which is denoted by (*). No Mix(4) exists for $\nu=3$, because only 3 varieties are available, and the maximum level of mixing is 3.

ν	Mix(2)	Mix(3)	Mix(4)	Loop	Mix(1)
3	0.7727	1	NA	1	0.5
4	0.7085	0.7895	1	1	0.5555
5	0.7378	0.7894	0.8696	1	0.625
6	0.7933	0.8332	0.8749	1	0.7
7	0.86	0.8936	0.9181	1	0.7779
8	0.9322	0.9619	0.9769	1	0.8572
9	0.9655	0.9914	1	0.9585	0.8987
10	0.9727	0.9954	1	0.8967	0.9135
11	0.9781	0.9983	1	0.8409	0.925
12*	0.9822	1	1	0.7904	0.9339
13	0.9837	1	0.9986	0.7436	0.9396
14	0.9849	1	0.9979	0.7018	0.9446
15	0.9862	1	0.9972	0.6642	0.949
16	0.9875	1	0.9971	0.6306	0.9528
17	0.9881	1	0.9968	0.5999	0.9557
18	0.9888	1	0.9964	0.5717	0.9584
19	0.9898	1	0.9964	0.5462	0.961
20	0.9901	1	0.996	0.5229	0.963
21	0.9908	1	0.9963	0.5015	0.9651
22	0.9911	1	0.9963	0.4814	0.9668
23	0.9918	1	0.9963	0.4631	0.9685
24	0.9922	1	0.9963	0.4461	0.9699
25	0.9922	1	0.9959	0.4303	0.9709

This shows that the optimality results of these designs depend on the scope of the search. If we consider the optimality of these designs only within the class of equireplicate designs, then the loop design is always A-optimal for $b = \nu$ and the interwoven loop design is always A-optimal for $b = 2\nu$.

3.3 Optimality Results from Row-Column Design

3.3.1 Theory of Row-Column Design

In row-column designs the experimental units are grouped in two directions, i.e., two blocking factors are used with one factor representing the rows of the design and the other factor representing columns. The designs considered here have ν treatments set out in an array of k rows and b columns such that the i^{th} treatment is replicated r_i times.

Let y_{ijm} be the response obtained when the i^{th} treatment is applied in the j^{th} row and m^{th} column. The usual fixed effects linear model is specified as

$$y_{ijm} = \mu + \tau_i + \rho_j + \gamma_m + \epsilon_{ijm} \quad (3.3)$$

where the parameters τ_i , ρ_j and γ_m measure the effect of the i^{th} treatment, j^{th} row and m^{th} column respectively for $i=1,2,\dots,\nu$; $j=1,2,\dots,k$; $m=1,2,\dots,b$. The error terms ϵ_{ijm} are assumed to be uncorrelated random variables each with mean zero and variance σ^2 (John and Williams, 1995). In such a situation there are two types of incidence matrix: *treatments vs. rows* and *treatments vs. columns*, which are denoted by $M(\nu \times k)$ and $N(\nu \times b)$ respectively. Under this set-up, the reduced normal equation for estimating the treatment effects is

$$C_d = \text{Diag}(r_{d1}, r_{d2}, \dots, r_{d\nu}) - k^{-1}N_dN'_d - b^{-1}M_dM'_d + b^{-1}k^{-1}r_d r'_d \quad (3.4)$$

where $r_d = M_d \mathbf{1} = N_d \mathbf{1}$, and $r'_d \mathbf{1} = bk$.

The C-matrix has all the usual properties of being symmetric nonnegative definite with zero row and column sums. Also, for all treatment contrasts to be estimable, it is necessary and sufficient that the underlying design be connected, i.e., the rank of C-matrix being $\nu - 1$. Henceforth, we shall let $D(\nu, b, k)$ denote the class of all treatment connected designs having ν treatments arranged in b columns and k rows. With each design, $d \in D(\nu, b, k)$, we associate the block designs d_N and d_M with incidence matrices N_d and M_d , respectively, i.e., d_N is that block design which can be obtained from d by treating the columns of d as blocks and ignoring the row effects, whereas d_M is that block design which can be obtained from d by treating the rows of d as blocks and ignoring the column effects. The reduced normal equations for estimating the treatment effects in d_N and d_M are, under the appropriate two-way ANOVA model,

$$C_d^N = \text{Diag}(r_{d1}, r_{d2}, \dots, r_{d\nu}) - (1/k)N_dN'_d \quad (3.5)$$

$$C_d^M = \text{Diag}(r_{d1}, r_{d2}, \dots, r_{d\nu}) - (1/b)M_dM_d' \quad (3.6)$$

Similar to the case of block design, the search for optimal row-column design is based on minimizing some convex function of the eigenvalues of C_d . The optimality criterion will be the same as before.

The augmentation process has been used to construct optimal row-column designs in current literature. Jacroux (1985) provided some sufficient conditions for design of E-optimal row-column designs based on group divisible designs including some row-column designs with $k = 2$. Similarly, Das (1991) came up with another class of E-optimal designs based on Balanced Incomplete Block Design (BIBD) or group divisible design (GDD). The most relevant work is due to Bagchi and Cheng (1993) who studied a class of E-optimal designs with $k=2$ based on graph theory. However, these designs often assume a relatively large number of blocks are available, which makes it almost impossible to apply them to design of microarray experiments, since researchers often face limited resources. Very little is known of the A-optimality of row-column designs, though some sufficient conditions have been studied by Bagchi and Shah (1989), Jacroux and Ray (1991) and Shah and Sinha (1993). A-optimality conditions applicable to the case of $k = 2$ are yet to be developed.

Optimal or near optimal row-column designs can also be constructed through a computer search when the number of varieties is not very large. A number of computer algorithms have been developed capable of producing good designs for a wide range of parameter values and for a wide range of blocking structures. The early algorithms by Jones (1979), Eccleston and Jones (1980), and Russell *et al*(1981) all have the problem that they could be trapped at local optimal designs. In view of this, Nguyen and Williams (1993) describe an iterative improvements algorithm which involves two steps and essentially avoids the problem of local optima. A second approach by Venables and Eccleston (1993) is simulated annealing. A random interchange process has been built into such an algorithm which also overcomes the problem of being trapped at local optima.

3.3.2 Optimality Results for $b = 2\nu$ and $k = 2$

The following optimality results were established for the class of connected designs that are equireplicate. If we only consider the optimality of designs within the class of equireplicate designs, then interwoven (loop) designs are always A-optimal. But these designs may not be A-optimal when the number of varieties exceeds a certain number under each scenario.

This section lists the optimal row-column designs for the case of $b = 2\nu$, $k = 2$, and $3 \leq \nu \leq 9$. Note the columns of each table correspond to arrays and the two rows correspond to the two dyes (Cy3 and Cy5). For resolvable designs as indicated, every $\nu/2$ blocks forms a replicate in which each treatment appears only once.

$\nu = 3$

1	2	3	2	3	1
2	3	1	1	2	3

$\nu = 4$ (Resolvable)

1	3	1	2	2	4	4	3
2	4	3	4	1	3	1	2

$\nu = 5$

1	1	4	5	2	2	5	3	3	4
2	3	1	1	3	4	2	4	5	5

$\nu = 6$ (Resolvable)

1	2	6	1	3	2	3	4	5	4	6	5
5	3	4	6	4	5	1	2	6	1	2	3

$\nu = 7$

1	1	2	2	3	3	4	4	5	5	6	6	7	7
4	7	4	1	5	2	6	7	6	2	3	1	3	5

$\nu = 8$ (Resolvable)

6	3	1	5	8	3	1	7	5	8	4	2	4	2	7	6
7	4	8	2	5	6	4	2	6	3	7	1	5	3	8	1

$\nu = 9$

1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	8	9
2	7	5	3	4	8	9	1	4	6	9	2	3	6	5	1	7	8

3.3.3 Optimality Results for $b = 3\nu$ and $k = 2$

This section lists the optimal row-column designs for the case of $b = 3\nu$, $k = 2$, and $3 \leq \nu \leq 10$. Note the columns of each table correspond to arrays and the two rows correspond to the two dyes (Cy3 and Cy5). For resolvable designs as indicated, every $\nu/2$ blocks forms a replicate in which each treatment appears only once.

$\nu = 3$

1	2	3	1	2	3	1	2	3
2	3	1	2	3	1	2	3	1

$\nu = 4$ (Resolvable)

3	1	4	2	3	1	2	4	2	1	3	4
4	2	1	3	4	2	3	1	4	3	1	2

$\nu = 5$

1	1	4	5	2	2	5	3	3	4	1	5	2	4	3
2	3	1	1	3	4	2	4	5	5	4	1	3	2	5

$\nu = 6$ (Resolvable)

3	2	1	4	1	5	6	2	1	3	4	2	5	4	6	3	6	5
4	6	5	2	3	6	4	5	3	5	6	1	2	1	3	2	1	4

$\nu = 7$

1	1	1	5	6	7	2	2	2	6	7	3	3	3	7	4	4	4	5	5	6
2	3	4	1	1	1	3	4	5	2	2	4	5	6	3	5	6	7	6	7	7

$\nu = 8$ (Resolvable)

1	2	5	7	8	2	4	3	6	2	7	1	4	5	6	3
4	3	8	6	1	5	6	7	3	8	4	5	2	7	1	8

1	8	4	5	3	7	6	8
2	7	3	6	5	1	2	4

$\nu = 9$

1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6
5	4	8	1	4	3	5	4	9	5	7	6	7	8	9	3	1	7

7	7	7	8	8	8	9	9	9
2	8	9	6	3	2	2	6	1

$\nu = 10$ (Resolvable)

9	3	5	2	7	5	8	10	1	2	4	7	3	6	1	9
6	4	8	10	1	9	6	3	4	7	8	5	9	10	2	2

7	6	10	1	4	6	8	2	10	4	5	8	9	3
3	5	4	8	9	7	3	5	1	7	10	2	1	6

$\nu = 11$

1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6
5	4	3	5	4	3	6	7	8	8	5	7	7	10	8	2	1	10

7	7	7	8	8	8	9	9	9	10	10	10	11	11	11
6	9	11	6	11	9	2	1	10	3	4	11	9	2	1

 $\nu = 12$ (Resolvable)

9	8	1	3	10	7	12	5	11	2	9	7	3	1	11	6	12	4
2	6	5	4	11	12	3	6	8	4	1	10	5	8	9	10	2	7

7	1	10	6	5	8	10	4	2	5	11	9	3	6	4	12	8	2
9	4	2	12	11	3	1	6	8	7	12	3	10	9	11	1	7	5

 $\nu = 13$

13	13	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7
1	2	11	3	5	8	9	4	7	5	11	13	9	6	7	7	10	8	12	13

7	8	8	8	9	9	9	10	10	10	11	11	11	12	12	12	13	1	1
8	4	1	9	12	13	11	5	3	12	10	6	2	4	2	1	10	3	6

 $\nu = 14$ (Resolvable)

4	9	12	13	11	5	3	2	5	12	6	3	10	14	1	6
6	10	2	1	7	14	8	4	11	13	9	7	1	8	2	14

7	8	9	4	10	6	8	7	9	13	10	3	4	13	12	11
12	11	13	5	3	11	12	1	2	14	4	5	8	3	6	2

7	14	1	14	13	2	1	11	8	5
9	10	5	7	4	3	6	10	9	12

3.3.4 Optimality Results for $b = 4\nu$ and $k = 2$

This section lists the optimal row-column designs for the case of $b = 4\nu$, $k = 2$, and $9 \leq \nu \leq 15$. Note the columns of each table correspond to arrays and the two rows correspond to the two dyes (Cy3 and Cy5). For resolvable designs as indicated, every $\nu/2$ blocks forms a replicate in which each treatment appears only once.

 $\nu = 3$

1	2	3	1	2	3	1	2	3	1	2	3
2	3	1	2	3	1	2	3	1	2	3	1

$\nu = 4$ (Resolvable)

3	1	4	3	3	2	1	2	2	4	2	3	1	4	1	4
4	2	2	1	4	1	4	3	3	1	4	1	2	3	3	2

$\nu = 5$

1	1	1	1	2	2	2	3	3	4	2	3	4	5	3	4	5	4	5	5
2	3	4	5	3	4	5	4	5	5	1	1	1	1	2	2	2	3	3	4

$\nu = 6$ (Resolvable)

2	4	3	2	6	5	2	6	4	5	3	1	5	1	4	3
6	1	5	3	4	1	5	3	1	6	4	2	6	3	2	6

1	2	1	5	6	4	6	3
5	4	3	2	4	5	1	2

$\nu = 7$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4
2	3	2	5	3	5	4	3	5	6	4	4	1	5	6	7

5	5	5	5	6	6	6	6	7	7	7	7
1	6	6	7	2	1	7	7	4	3	2	1

$\nu = 8$ (Resolvable)

7	8	3	1	4	2	5	6	2	8	3	7	1	7	6	3
2	6	5	4	8	3	7	1	4	1	6	5	5	8	2	4

4	1	6	2	8	4	5	3	2	5	7	8	5	1	6	4
5	3	7	8	1	6	2	7	4	6	1	3	8	2	3	7

$\nu = 9$

1	1	1	1	6	7	8	9	2	2	2	2	7	8	9	3	3	3
2	3	4	5	1	1	1	1	3	4	5	6	2	2	2	4	5	6

3	8	9	4	4	4	4	9	5	5	5	5	6	6	6	7	7	8
7	3	3	5	6	7	8	4	6	7	8	9	7	8	9	8	9	9

$\nu = 10$ (Resolvable)

8	6	4	9	2	6	2	8	5	1	1	4	7	2	5	3	6	9	7	8
1	7	10	3	5	9	3	10	7	4	3	6	9	8	10	4	1	10	2	5

1	4	10	9	3	10	4	2	5	3	9	6	10	3	7	7	1	5	10	8
5	8	7	2	6	1	7	6	9	8	4	8	2	5	1	3	2	4	6	9

 $\nu = 11$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5	6	6
3	2	4	11	7	6	3	5	6	4	5	7	8	6	5	10	8	10	7	9	10	7

6	6	7	7	7	7	8	8	8	8	9	9	9	9	10	10	10	10	11	11	11	11
8	9	1	8	9	11	2	1	9	10	4	3	1	11	3	2	1	11	6	5	4	2

 $\nu = 12$ (Resolvable)

8	6	1	7	4	3	8	9	3	7	10	1	8	10	2	5
10	12	2	9	5	11	2	6	4	11	12	5	9	11	7	6

1	12	1	6	2	11	3	4	12	6	9	4	7	11	9	2
3	4	9	8	10	12	5	7	1	2	3	10	8	5	11	4

12	6	5	3	5	4	10	7	12	11	11	8	2	9	5	10
7	1	10	8	8	6	9	1	3	2	6	12	3	4	7	1

 $\nu = 13$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
5	3	6	10	6	5	3	7	9	5	4	8	10	6	5	7	11	7	9	10

6	6	6	6	7	7	7	7	8	8	8	8	9	9	9	9	10	10	10	10
7	9	8	12	1	13	9	11	1	2	4	11	8	10	13	12	2	12	13	11

11	11	11	11	12	12	12	12	13	13	13	13
6	3	12	13	1	2	3	4	8	1	2	4

 $\nu = 14$ (Resolvable)

6	8	7	13	14	3	9	8	11	7	10	14	12	2	4	10	1	5	9	2
10	2	12	1	11	5	4	1	3	9	13	4	5	6	8	12	3	6	11	7

13	6	7	13	8	4	3	12	13	11	7	12	14	1	3	11	5	8	1	14
14	11	14	2	9	5	10	1	6	5	10	2	8	4	9	2	7	13	6	3

9	4	2	6	1	4	9	10	5	2	3	6	12	10	5	11
12	10	3	14	7	12	13	11	8	4	7	9	14	8	13	1

$\nu = 15$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5	5
2	9	8	12	10	9	8	14	9	10	5	8	7	12	8	13	10	12	9	11
6	6	6	6	7	7	7	7	8	8	8	8	9	9	9	9	10	10	10	10
14	11	12	13	10	13	11	14	6	5	7	15	6	7	4	15	1	4	6	15
11	11	11	11	12	12	12	12	13	13	13	13	14	14	14	14	15	15	15	15
4	1	2	3	2	7	3	15	5	3	1	2	5	4	1	3	14	11	13	6

 $\nu = 16$ (Resolvable)

3	14	8	11	1	16	12	9	14	13	1	4	3	6	5	12
6	4	7	10	5	13	2	15	10	15	2	11	16	9	8	7
2	4	8	1	10	9	5	11	4	9	11	15	12	6	10	7
3	7	14	6	15	16	12	13	5	14	1	3	16	8	2	13
2	16	10	14	13	7	8	15	4	9	15	16	13	3	7	12
9	4	6	1	5	3	11	12	6	11	1	8	2	5	10	14
10	5	15	14	11	1	2	6	2	16	3	8	7	6	13	5
16	9	4	3	12	7	8	13	4	1	11	15	9	12	14	10

 $\nu = 17$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5
2	3	6	16	9	13	14	7	16	14	5	17	12	5	17	13	11	6
5	5	6	6	6	6	7	7	7	7	8	8	8	8	9	9	9	9
8	15	9	14	7	13	4	8	10	16	1	12	14	9	3	4	15	10
10	10	10	10	11	11	11	11	12	12	12	12	13	13	13	13	14	14
5	1	12	17	9	7	1	12	6	3	2	15	11	10	8	3	10	11
14	14	15	15	15	15	16	16	16	16	17	17	17	17				
4	16	13	7	1	17	5	15	4	2	11	8	6	2				

 $\nu = 18$ (Resolvable)

5	7	9	10	2	8	11	18	14	2	13	4	7	11	9	6	15	14
6	16	13	1	4	17	15	12	3	3	1	10	8	17	12	16	18	5
8	4	6	5	3	12	10	16	2	18	12	8	16	15	1	3	14	5
14	17	11	13	7	1	18	15	9	6	11	13	2	4	7	10	9	17
6	17	15	13	8	1	12	16	4	17	12	13	10	7	11	18	5	3
9	18	3	11	2	14	5	10	7	1	4	16	8	9	2	14	15	6
11	4	2	9	1	3	18	15	17	6	1	9	7	17	14	16	10	13
14	13	5	10	6	12	7	8	16	8	2	15	5	3	4	12	11	18

$\nu = 19$

1	1	1	1	2	2	2	2	3	3	3	3	4	4	4	4	5	5	5
4	9	14	6	14	4	6	16	15	6	9	4	12	7	11	5	19	13	8
5	6	6	6	6	7	7	7	7	8	8	8	8	9	9	9	9	10	10
9	10	13	8	7	9	14	16	15	19	17	15	18	2	18	11	12	5	11
10	10	11	11	11	11	12	12	12	12	13	13	13	13	14	14	14	14	15
17	19	8	14	13	16	8	10	17	13	19	15	18	17	12	3	5	18	11
15	15	15	16	16	16	16	17	17	17	17	18	18	18	18	19	19	19	19
2	10	1	12	3	5	1	7	3	1	2	16	10	6	4	7	3	1	2

$\nu = 20$ (Resolvable)

10	1	15	11	8	12	16	6	3	4	14	17	5	8	19	1	11	7	18	13
19	9	20	17	18	14	2	7	5	13	20	2	9	12	4	6	15	10	3	16
2	18	5	11	12	20	7	19	1	13	4	2	3	7	19	16	15	17	5	9
9	6	15	14	17	4	8	16	3	10	11	8	13	14	6	12	18	1	10	20
18	9	10	11	2	6	15	17	13	19	6	9	2	20	16	12	4	18	8	14
4	12	1	16	3	20	7	5	8	14	11	13	15	3	7	10	5	17	19	1
1	20	6	7	17	14	16	10	3	12	9	5	13	14	4	20	10	3	8	15
4	8	13	9	19	18	5	2	11	15	19	6	17	2	12	16	18	7	11	1

3.4 Comparison of Designs for Classifying Samples

3.4.1 Designed Experiments using Interwoven Loop Design

Because we are comparing the interwoven loop design to the replicated reference design under different scenarios, we first design these experiments given the number of samples to classify. When $b = 2\nu$ and $b = 3\nu$ with $\nu = 20$, the interwoven loop design is constructed according to Churchill (2002) and are in Table 3.3 and 3.4, respectively. When $b = 4\nu$ and $\nu = 20$, the designed experiment is from the row-orthogonal design we have found by relabelling symbols and rearranging blocks. This design is in fact an interwoven loop design and is given in Table 3.5. Details of the interwoven loop design for $b = 5\nu$ and $\nu = 20$ can be found in the section on experimental designs in Yang *et al.* (2003). For the replicated reference sample, given 10 samples for each cell line (NT) and (IVF), 20 arrays will be hybridized by pairing the samples of interest and the reference sample. Such a pattern is replicated 2, 3, 4 or 5 times depending on the experimental resources. It is assumed that no dye reverse is used for the reference design.

Table 3.2: Labelling of 20 Samples in Classification Study

Sample	Symbol	Sample	Symbol
NT1	1	IVF1	11
NT2	2	IVF2	12
NT3	3	IVF3	13
NT4	4	IVF4	14
NT5	5	IVF5	15
NT6	6	IVF6	16
NT7	7	IVF7	17
NT8	8	IVF8	18
NT9	9	IVF9	19
NT10	10	IVF10	20

Table 3.3: Interwoven Loop Design ($b = 40, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples

Array	1	21	3	25	5	29	24	7	28	9	31	11	33	13	35	15	37	17	39	19
Cy3	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	10	10
Cy5	11	18	12	19	13	20	11	14	12	15	13	16	14	17	15	18	16	19	17	20
Array	2	23	4	27	22	6	26	8	30	10	32	12	34	14	36	16	38	18	20	40
Cy3	11	11	12	12	13	13	14	14	15	15	16	16	17	17	18	18	19	19	20	20
Cy5	2	9	3	10	1	4	2	5	3	6	4	7	5	8	6	9	7	10	1	8

Table 3.4: Interwoven Loop Design ($b = 60, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples

Array	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cy3	1	1	1	2	2	2	3	3	3	4	4	4	5	5	5	6	6	6	7	7
Cy5	15	13	14	16	14	15	17	15	16	18	16	17	19	17	18	20	18	19	11	19
Array	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Cy3	7	8	8	8	9	9	9	10	10	10	11	11	11	12	12	12	13	13	13	14
Cy5	20	12	20	11	13	11	12	14	12	13	1	4	2	2	5	3	3	6	4	4
Array	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Cy3	14	14	15	15	15	16	16	16	17	17	17	18	18	18	19	19	19	20	20	20
Cy5	7	5	5	8	6	6	9	7	7	10	8	1	8	9	2	9	10	3	10	1

Table 3.5: Interwoven Loop Design ($b = 80, \nu = 20$): symbols in rows Cy3 and Cy5 are the 20 samples

Array	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Cy3	1	6	14	10	13	2	12	4	20	16	14	12	6	17	3	11	19	18	13	5
Cy5	15	11	9	19	7	17	5	18	8	3	2	10	20	7	15	4	9	1	8	16
Array	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Cy3	12	4	9	2	17	3	7	15	8	1	19	7	16	18	4	17	9	10	20	14
Cy5	6	13	20	19	10	14	18	5	11	16	3	12	8	2	15	5	13	11	1	6
Array	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Cy3	13	6	11	5	8	14	18	20	7	2	19	16	9	8	5	1	3	10	15	11
Cy5	3	17	1	19	12	4	9	10	16	15	4	6	12	14	18	17	20	13	7	2
Array	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Cy3	3	7	16	6	15	2	20	1	8	9	15	4	10	12	17	5	13	18	19	11
Cy5	11	14	4	18	10	13	5	12	19	17	6	20	16	2	3	14	1	8	7	9

3.4.2 Comparison of Classification for Designs

For the interwoven loop design, the corresponding ANOVA model is

$$y_{ijklgm} = \mu + A_i + D_j + T_k + S_l(T_k) + G_g + (AG)_{ig} + (TG)_{kg} + G \times S_l(T_k) + \epsilon_{ijklgm} \quad (3.7)$$

where response y_{ijklgm} is the log-transformed fluorescence intensity for array i , dye j , cell line k , sample l , gene g and replicate m . For the replicated reference design, the reference sample is often prepared by pooling all samples of interest and is applied to each array. If the reference sample is treated as an independent sample from another cell line, then the above model can be used for the reference design by dropping the term D_j . However, because variation among reference samples is very small, use of such a model will underestimate the variability among samples of interest, hence it is not recommended for analysis of data from replicated reference designs. Here, we model the difference between the two log intensities (or log ratios) from the same spot, or $w_{klgm} = y_{ijk(1)} - y_{ijk(0)}$, where subscript 0 denotes the reference sample and 1 an individual sample of interest. This would result in the following ANOVA model:

$$w_{klgm} = \mu + T'_k + S_l(T'_k) + G_g + (T'G)_{kg} + G \times S_l(T'_k) + \epsilon_{ijklgm} \quad (3.8)$$

where $T'_k = T_k - T_0$, and T_0 denotes the reference sample. Therefore, detection of gene expression and sample classification can be done in a similar way as the interwoven loop design.

Under both models, terms $S_l(T_k)$ and $G \times S_l(T_k)$ are assumed to be random, and terms A and AG are also assumed random for the interwoven loop design. Dispersion assumptions are: $A_i \sim \text{iid } N(0, \sigma_A^2)$, $S_l(T_k) \sim \text{iid } N(0, \sigma_{TS}^2)$, $AG_{ig} \sim \text{iid } N(0, \sigma_{AG}^2)$, $GS(T) \sim \text{iid } N(0, \sigma_{gts}^2)$, and $\epsilon_{ijklgm} \sim \text{iid } N(0, \sigma_e^2)$. The values of these variance components were estimated from a subset of Infigen's bovine data, where $\sigma_A^2=0.2458$, $\sigma_{AG}^2=0.3558$, $\sigma_{TS}^2=0.03838$, $\sigma_{gts}^2=0.1248$, $\sigma_e^2=0.1295$.

In the true state, 20 genes were used (indexed as $1, \dots, 20$), 10 of which were differentially expressed (Gene 11, ..., Gene 20).

Table 3.6: Summary of Variances of Pairwise Contrasts among Samples under Four Scenarios: $\nu=20$.

Scenarios	Designs	Min	Max	Mean
$b = 2\nu$	Interwoven Loop	0.1517	0.2039	0.1852
	Rep. Reference	0.2544	0.2544	0.2544
$b = 3\nu$	Interwoven Loop	0.1036	0.1205	0.1131
	Rep. Reference	0.1696	0.1696	0.1696
$b = 4\nu$	Interwoven Loop	0.0784	0.8732	0.0821
	Rep. Reference	0.1272	0.1272	0.1272
$b = 5\nu$	Interwoven Loop	0.0630	0.0660	0.0644
	Rep. Reference	0.1018	0.1018	0.1018

Table 3.7: Variances of Contrasts between Types (T) under Each Scenario

Scenarios	Interwoven Loop	Rep. Reference
$b = 2\nu$	0.0359	0.0456
$b = 3\nu$	0.0348	0.0413
$b = 4\nu$	0.0343	0.0391
$b = 5\nu$	0.0339	0.0378

Tables 3.6 and 3.7 list the variance of contrasts of interest under both models. Under each scenario, the variance of contrast for the interwoven loop design is smaller than that for the replicated reference design. This implies that when the nested models are used for detecting gene differential expression between types of cell populations, the interwoven loop design has more power to detect those differentially expressed genes.

Our simulations followed these 3 steps:

Step 1: Expression values for the 20 genes were first simulated for the interwoven loop design and the replicated reference design with parameters estimated from Infigen's bovine classification study. The gene expression values for the interwoven loop design were simulated using model (3.7). The expression values for the reference design were simulated based on model (3.8) with the error variance two times of that for the interwoven loop design. Under the assumption of a single reference sample used, gene expression values for samples of interest can be simulated under model (3.7). Further, expression values for genes in the reference sample can be simulated using a model with only A, D, AG, error, and a constant (the expression values of genes in the reference sample), or

$$y_{ij0g} = \mu + A_i + D_j + T_0 + (AG)_{ig} + G_g + (GT)_{g0} + \epsilon_{ij0g} \quad (3.9)$$

If we ignore effects due to labelling reactions, T_0 will be constant, and $(GT)_{g_0}$ is also constant for each gene across arrays. If we analyze the data using log ratios, then both A and AG drop out (same spot), which produces model (3.8). So, the simulation for the reference design is simpler.

Step 2: Differential expression were detected using models (3.7) and (3.8) with the contrasts defined in Yang *et al.* (2002), and the estimated $G \times S(T)$ were adjusted by the estimated T , GT and $S(T)$, which were the inputs for the classification analysis;

Step 3: we used Fisher's linear discriminant method to estimate the discriminant function, and the training data set was cross-validated to estimate the misclassification error rate at each iteration.

Chapter 4

Linear Mixed Effects Models for Microarray Gene Expression Data

4.1 Introduction

Microarray gene expression experiments allow researchers to monitor the expression levels of thousands of genes simultaneously (Derisi *et al.*, 1997; Eisen *et al.*, 1998). Performing and interpreting (cDNA) microarray expression experiments is not efficient without statistical support at each step of the process. This includes experimental design (e.g., Kerr and Churchill, 2001), image processing (e.g., Yang *et al.*, 2000), identification of differentially expressed genes (e.g., Dudoit *et al.*, 2002), clustering of genes and samples or class discovery (e.g., Getz *et al.*, 2000), discriminant analysis or classification of samples into known classes (e.g., Dudoit *et al.*, 2001), and pathway and gene network inference (e.g., Pe'er *et al.*, 2001; Henderson *et al.*, 2003). In this paper, we focus on statistical methods for identifying differentially expressed genes.

Many statistical methods have been developed for identifying differentially expressed genes from microarray experiments. A first group of methods include the simple ratio-based approach (DeRisi *et al.*, 1996; Chen *et al.*, 1997), simple *t*-test (Claverie, 1999), the two-stage *t*-test (Dudoit *et al.*, 2002; Yang *et al.*, 2001) and nonparametric tests (Lin *et al.*, 2001; Gibbons *et al.*, 2002). See Pan (2002) for a review of *t*-test based methods. A second group of methods consist of Bayesian methods, providing posterior belief probabilities of fold change (e.g., Theilhaber *et al.*, 2001; Efron *et al.*, 2001; Newton *et al.*, 2000; Baldi and Long, 2001). Potential limitations with the first two groups of methods include the need for prior data normalization of the raw intensity measurements, restricted applicability to experiments with only two samples (control and treatment) or to data without replication, and analysis of each array separately (Kerr *et al.*, 2002). In the context of gene expression data analysis, data normalization refers to adjustment of the raw data for systematic technology biases,

e.g., effects of slides, dyes, print-tips and spatial variation (Schuchardt *et al.*, 2000; Yang *et al.*, 2001; Kerr and Churchill, 2001; Quackenbush, 2001).

A third group of methods include fixed or mixed effects models which perform both normalization and identification of differentially expressed genes. Kerr *et al.* (2000) and Kerr and Churchill (2001) were the first to propose to study gene expression data using analysis of variance (ANOVA) models. However, because fitting fixed or mixed models to data on thousands of genes is computationally intensive, a two-step approach has been suggested, which consists of a normalization step based on a global model, and a second step analyzing differential expression for each gene (Wolfinger *et al.*, 2001; Wu *et al.*, 2003). The two-step approach has been applied to real experiments (e.g., Jin *et al.*, 2001; Churchill and Oliver, 2001). A modification of the two-step approach by Yang *et al.* (2001) is essentially a combination of Bayesian methods and the ANOVA approach. While the need for data transformations to adjust for intensity-dependent dye effects and spatial variation and to stabilize variances prior to analyses is now recognized (see Cui *et al.* 2002 for a quite comprehensive discussion of suitable transformations), the consequences of replacing the single or full model analysis with a two-step approach have not yet been studied nor clearly stated. None of the above authors have justified the two-step analysis, or stated clearly when single or full model and two-step model analyses provide the exact same tests for differential gene expression. In this article, we investigate the rationale of the two-step model analysis analytically under both fixed and mixed models. For more complicated mixed models, simulation is used to study the advantages and disadvantages of the two-step analysis. Lastly, we apply the methods to three real microarray experiments.

4.2 Hypothesis Testing under ANOVA Models

Kerr and Churchill (2001) proposed to analyze gene expression using linear fixed effects models. A basic model of this type is expressed as

$$y_{ijkl} = \mu + A_i + D_j + V_k + G_g + (AG)_{ig} + (VG)_{kg} + \epsilon_{ijkl}, \quad (4.1)$$

where y_{ijkl} is the transformed fluorescent intensity taken from array i , dye j , variety k and gene g , with A , D , V and G defined as the main effects for array, dye, variety and genes, respectively; (VG) and (AG) are interactions between V , A and G , respectively. For a detailed discussion of these main and interaction effects, readers are referred to Kerr and Churchill (2001) and to Wu *et al.* (2003). Note that dye effects (and possibly DG effects) need to be included in the model for non-reference designs. Model (4.1) is appropriate for single variety factor experiments and on replication of genes on the same array (or averaging of replicate intensity within an array).

Researchers (Kerr and Churchill, 2001; Wolfinger *et al.*, 2001; Lee *et al.*, 2000) have stated that differential expression of gene g in samples k and k' should be assessed based on the

Table 4.1: Gene Expression Values for Experiment A and B: expression values of both experiments are the same except for expression values of gene 1 in sample 2 for Experiment B, which are listed in the bracket.

Expression Values	Gene 1	Gene 2	Gene 3
Sample 1	2, 3	2, 3	2,3
Sample 2	9,10 (2,3)	9,10	9,10

linear contrast $(VG)_{kg} - (VG)_{k'g}$. According to Winer (1962), the (VG) interaction is the effect attributable to the combination of V and G above and beyond that which can be predicted from the estimated main effects for V and G , thus $(V_k - V_{k'})$ is not taken into account. It should be pointed out that the contrast $(VG)_{kg} - (VG)_{k'g}$ is generally not estimable except in certain restricted models for balanced data.

An alternative contrast is

$$\begin{aligned} \mu_d &= (V_k + G_g + (VG)_{kg}) - (V_{k'} + G_g + (VG)_{k'g}) \\ &= ((VG)_{kg} - (VG)_{k'g}) + (V_k - V_{k'}). \end{aligned} \quad (4.2)$$

The difference between the two types of contrasts is now illustrated with an artificial example. Consider two microarray experiments with 3 genes and 2 samples (Table 4.1). In experiment *A*, all genes are differentially expressed, which is represented by contrasts in the form of (4.2) but not by the contrast in the (VG) effects only, because there is no interaction in experiment *A*. In experiment *B*, gene 1 is not differentially expressed, which is again consistent with $\mu_{11} - \mu_{21}$ being zero where μ_{ij} is the cell mean for gene j and sample i in Table 4.1. However, $(\hat{VG})_{11} - (\hat{VG})_{21}$ is not zero here and might lead to the wrong conclusion of gene 1 being differentially expressed.

If differences among the V main effects are negligible due to most genes not being differentially expressed or half of the genes being up- and the other half down-regulated, one might not expect much difference between the two types of contrasts. In our experience with real data, we have however observed a non-negligible level of difference. We favor the contrast in (4.2), despite concerns about the V main effects not clearly representing average (or across genes) mRNA expression levels in different cell populations, but being confounded with technology effects unique to individual samples and labelling reactions. Replication across arrays will reduce or eliminate the labeling effects. Often, however, only one mRNA sample is drawn from a given cell population (one extraction and amplification), so that mRNA samples effects are confounded with the variety effects. However, in some experiments there are biological replicates of the varieties, which should reduce or eliminate sample effects (see Section 4.2 and 4.3).

4.3 Linear Fixed Effects Models

In this section, we examine the consistency between one- and two-step analyses under fixed effects models.

4.3.1 Two-Step Analysis of Gene Expression Data

Although ANOVA methods are flexible tools to analyze gene expression data, they are computationally challenging, because thousands of genes and several tens or hundreds of samples are often involved in a single experiment. Consequently the dimension of the normal equations (NE) for a model such as (4.1) can be very large. Testing for differential gene expression requires the computation of test statistics $t = c/Var(c)$ with contrast $c = \mu_d$ given in (4.2). In general, while estimates of contrasts of interest can be estimated efficiently even for very large number of equations by iteration, computation of $Var(c)$ requires elements of the inverse of the coefficient matrix of NE which may not be feasible to compute. For specific designs (e.g., loop, reference) estimates of the VG contrasts have been derived in the form $c = l'y$ (e.g., Kerr and Churchill, 2001). To achieve computational efficiency for the general case, two-step fixed effects models (Wu *et al.*, 2002) and two-step mixed effects models (Wolfinger *et al.*, 2001; Jin *et al.*, 2001) have been proposed, but their equivalence or degree of approximation to the single-step full model analysis has not been studied. Therefore, we first present the equivalence between single and two-step analyses for fixed models. We partition the fixed model as follows

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon, \quad (4.3)$$

where \mathbf{X}_1 and \mathbf{X}_2 denote the design matrices for global factors (e.g., A , D , V) and gene-specific factors (e.g., G , AG , DG , VG), respectively. The two-step estimation procedure consists of a normalization step fitting a model with global factors only, or

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \epsilon_1. \quad (4.4)$$

The estimated residuals from (4.4) are used as the responses in the second step fitting a model with gene-specific factors, or

$$\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}_1\hat{\beta}_1 = \mathbf{X}_2\beta_2 + \epsilon_2. \quad (4.5)$$

The NE for this model are block-diagonal with blocks corresponding to genes. Hence, this model can be fit separately for each gene. We now describe under what conditions the full model and the two-step analyses yield the same test statistics.

4.3.2 Global Factors Interact with Gene Factor (G)

We first assume that all the global factors interact with the gene (G) factor, and that all other two-way or higher order interactions are not of interest. Without loss of generality, we

assume global factors are A and V only, and the gene-specific effects are G , AG and VG , or

$$y_{ijkl} = \mu + A_i + V_j + G_k + (VG)_{jk} + (AG)_{ik} + \epsilon_{ijkl}.$$

A full rank reparameterization of this over-parameterized model is

$$y_{ijkl} = \mu_{(VG)jk} + (\mu_{(AG)ik} - \mu_{(AG)1k}) + \epsilon_{ijkl},$$

where $\mu_{(VG)jk} = \mu'_k + V_j + G_k + (VG)_{jk}$, $\mu'_k = \mu + A_1 + (AG)_{1k}$, and $\mu_{(AG)ik} = A_i + G_k + (AG)_{ik}$. The NE for the reparameterized model are block-diagonal with each block corresponding to a level of G . Consequently, in this case the analysis can be performed with gene-specific models without a prior normalization step. Estimates of the contrasts of interest, $\hat{\mu}_d$, can be obtained by taking differences among the estimates of the $\mu_{(VG)jk}$ parameter estimates. This result holds for balanced and unbalanced data.

4.3.3 General Case

In the more general case, where the fixed effects model does not contain interactions between G and some global factors, it is not possible to find a full-rank reparameterized model with block-diagonal structure of the NE resulting in gene-specific analyses. For example, the model may include print-tip effects (P) but no interaction PG , or an interaction among global factors such as (AD) but no (ADG) interaction (Wu *et al.*, 2002). In this case, Σ -restrictions can be imposed on the solutions to the NE (e.g., $\sum_l \hat{G}_l = 0$, $\sum_i \hat{A}_i = 0$, $\sum_l (\hat{V}G)_{kl} = 0$, $\sum_k (\hat{V}G)_{kl} = 0$, and so on). A detailed discussion of Σ -restrictions can be found in Searle (1971). One way of obtaining a Best Linear Unbiased Estimator (*BLUE*) of β is to solve the following augmented NE,

$$\begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 & \mathbf{C}_1 & \mathbf{0} \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 & \mathbf{0} & \mathbf{C}_2 \\ \mathbf{C}'_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}'_2 & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{X}'_2 \mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \quad (4.6)$$

which is of dimension $2p_1 + 2p_2 - \text{rank}(\mathbf{X}_1) - \text{rank}(\mathbf{X}_2|\mathbf{X}_1)$, where p_i is the dimension of β_i for $i = 1, 2$, $\text{rank}(\mathbf{X}_2|\mathbf{X}_1)$ is rank of \mathbf{X}_2 after fitting β_1 ; λ_1 and λ_2 are vectors of Lagrange multipliers, and $\mathbf{C}'_i \hat{\beta}_i = \mathbf{0}$ ($i = 1, 2$) represents the Σ -restrictions. There are $p_1 - \text{rank}(\mathbf{X}_1)$ rows or restrictions in \mathbf{C}'_1 and $p_2 - \text{rank}(\mathbf{X}_2|\mathbf{X}_1)$ rows in \mathbf{C}'_2 . Due to the structure of $\mathbf{X}'_1 \mathbf{X}_2$ and the Σ -restrictions on the solutions, the equality $\mathbf{X}'_1 \mathbf{X}_2 \hat{\beta}_2 = \mathbf{0}$ holds. The equations in (4.6) can be decomposed into two systems of equations, which can be solved successively as follows:

$$\begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{C}_1 \\ \mathbf{C}'_1 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{y} \\ \mathbf{0} \end{pmatrix}, \quad (4.7)$$

and

$$\begin{pmatrix} \mathbf{X}'_2 \mathbf{X}_2 & \mathbf{C}_2 \\ \mathbf{C}'_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_2 \mathbf{y} - \mathbf{X}'_2 \mathbf{X}_1 \hat{\beta}_1 \\ \mathbf{0} \end{pmatrix}. \quad (4.8)$$

The equality $\mathbf{X}'_1 \mathbf{X}_2 \hat{\beta}_2 = \mathbf{0}$ holds when each gene is represented equally often on each array. In such cases all rows in $\mathbf{X}'_1 \mathbf{X}_2$ corresponding to one factor in β_1 (μ , A , D or V) are equal, and each row in $\mathbf{X}'_1 \mathbf{X}_2$ has several partitions corresponding to the factors in β_2 (G , VG , AG) with all elements equal within a partition. Imposing the simple (unweighted) Σ -restrictions on the solution results in the desired equality.

To obtain identical test statistics under the one-step and two-step analyses, not only must the estimates of $c = \hat{\mu}_d$ be the same, but also the variances of the contrast must be identical. The variance of the estimated contrast in (4.2) is

$$\begin{aligned} \text{Var}(\hat{\mu}_d) &= \text{Var}(\hat{V}_k - \hat{V}_{k'}) + \text{Var}((\hat{V}G)_{kl} - (\hat{V}G)_{k'l}) \\ &+ 2 \times \text{Cov}((\hat{V}_k - \hat{V}_{k'}), ((\hat{V}G)_{kl} - (\hat{V}G)_{k'l})). \end{aligned} \quad (4.9)$$

By the properties of the inverse of partitioned matrices and the imposed Σ -restrictions, the equivalence of test statistics can be established. A proof of these results appears in Section A of the Appendix.

In summary, under fixed effects models and with constant gene layout across arrays (any gene is replicated a constant number of times on all arrays, but different genes may have different numbers of replicates), a two-step approach with a global normalization model and individual gene model can be used to compute test statistics for differential gene expression very efficiently, and these test statistics are identical to those obtained from the full model analysis. We note that computationally the NE for the normalization model and the gene-specific models do not need to be solved with the use of Σ -restrictions as in (4.7) and (4.8), which are awkward to use in practice. Any solution for β_1 and β_2 , which leads to the same contrasts can be computed. The results for $\text{Var}(\hat{\mu}_d)$ hold irrespective of the particular solutions, because any solution vector for β_1 (β_2) is in the form $L'\hat{\beta}_1$ ($L'\hat{\beta}_2$) with $\hat{\beta}_1$ ($\hat{\beta}_2$) defined in (4.7) and (4.8).

4.4 Linear Mixed Effects Models

Compared to the fixed effects models discussed in Section 2, mixed models are more flexible and reasonable in analyzing gene expression data; see Wolfinger *et al.* (2001) for a discussion of the merits of these models. In general, a mixed effects model can be written as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon, \quad (4.10)$$

where \mathbf{X} and \mathbf{Z} denote the design matrices for fixed and random factors, respectively, and \mathbf{u} is vector of random effects. This model can be partitioned analogously to (4.3) as follows

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{X}_2\beta_2 + \mathbf{Z}_2\mathbf{u}_2 + \epsilon, \quad (4.11)$$

where \mathbf{X}_1 and \mathbf{Z}_1 are the design matrices for fixed and random global factors, respectively; \mathbf{X}_2 and \mathbf{Z}_2 are the design matrices for the fixed and random gene-specific factors, respectively.

For the variance-covariance structure of the random factor, assume that

$$\text{Var} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{pmatrix},$$

and for $k = 1, 2$,

$$\text{Var}(\mathbf{u}_k) = \mathbf{Var} \begin{pmatrix} \mathbf{u}_{k(1)} \\ \vdots \\ \mathbf{u}_{k(i)} \\ \vdots \\ \mathbf{u}_{k(q_k)} \end{pmatrix} = \text{Blockdiag}(\mathbf{I}_{n_{ki}} \sigma_{k_i}^2),$$

where n_{ki} is the number of levels of the random factor k_i ($k = 1, 2$, $i = 1, \dots, q_k$) or dimension of $\mathbf{u}_{k(i)}$, and q_k is the number of random factors in \mathbf{u}_k . Application of two-step mixed model analysis (MMA), as suggested by Wolfinger *et al.* (2001), yields the following global normalization model

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{Z}_1 \mathbf{u}_1 + \epsilon_1, \quad (4.12)$$

and the gene-specific models

$$\hat{\mathbf{r}}_j = X_{2j} \beta_{2j} + \mathbf{Z}_{2j} \mathbf{u}_{2j} + \epsilon_{2j}, \quad (4.13)$$

where $\hat{\mathbf{r}}_j$ is the sub-vector corresponding to gene j , of the vector of estimated residuals,

$$\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}_1 \hat{\beta}_1 - \mathbf{X}_2 \hat{\mathbf{u}}_1,$$

where $\mathbf{X}_1 \hat{\beta}_1$ is the BLUE of $\mathbf{X}_1 \beta_1$, and $\hat{\mathbf{u}}_1$ is the Best Linear Unbiased Prediction (BLUP) of \mathbf{u}_1 . For example, factor array (A) may be treated as random, and in this case we may have $\mathbf{u}_1 = \mathbf{A}$ and $\mathbf{u}_2 = \mathbf{AG}$. For mixed models, in particular with unknown variance components, it is far less obvious than for fixed models when equivalence of test statistics for differential expression from full model (4.11) and two-step (4.12, 4.13) analyses can be achieved or approximated. Here, we provide analytical and simulation-based results on the equivalence between full and two-step MMA.

4.4.1 Two-Step Mixed Model Analysis

Here we investigate whether it is possible to perform mixed model analysis (MMA) with a two-step approach yielding the same test statistics for differential expression as the corresponding full model analysis.

We first present the mixed model equations (MME) (Searle *et al.*, 1993) for the full model (4.11) (see Section B of Appendix for details). Without loss of generality, we assume that $\mathbf{u}_1 = \mathbf{A}$, $\mathbf{u}_2 = \mathbf{AG}$, $\text{Var}(\mathbf{A}) = \mathbf{U}_1 = \sigma_a^2 \mathbf{I}$ and $\text{Var}(\mathbf{AG}) = \mathbf{U}_2 = \sigma_{ag}^2 \mathbf{I}$. Using an approach

similar to that in Section 2.2.2, we first investigate whether models (4.12) and (4.13) yield the same estimates for β_1 and β_2 , and the same predictions for \mathbf{u}_1 and \mathbf{u}_2 as model (4.11). From Section 2.2 and the augmented MME in (4.20), this will hold under the following conditions:

$$\begin{aligned} \mathbf{X}'_1 \mathbf{X}_2 \hat{\beta}_2 &= \mathbf{0}, & \mathbf{Z}'_1 \mathbf{X}_2 \hat{\beta}_2 &= \mathbf{0}, \\ \mathbf{X}'_1 \mathbf{Z}_2 \hat{\mathbf{u}}_2 &= \mathbf{0}, & \mathbf{Z}'_1 \mathbf{Z}_2 \hat{\mathbf{u}}_2 &= \mathbf{0}. \end{aligned} \quad (4.14)$$

The Σ -restrictions ensure that the first two equations hold. The following properties of the random effects solutions to MME are useful: (i) For a random factor (\mathbf{u}) with n levels, $\sum_{i=1}^n \hat{u}_i = 0$; (ii) For a random interaction factor \mathbf{u} representing the interaction between a fixed (i with q_1 levels) and a random (j with q_2 levels) factor, we have $\sum_{j=1}^{q_2} \hat{u}_{ij} = 0$ for each i , and $\sum_{i=1}^{q_1} \hat{u}_{ij} = \frac{\lambda_1}{\lambda_2} \hat{u}_j$ for each j , where λ_1 refers to the random main factor and λ_2 to the random interaction. These properties hold for a general mixed model with a random interaction between a fixed and a random main factor for any design and unbalanced data, as long as no restrictions are imposed on the interaction effects, and only if the variance-covariance matrix of the random matrix of the random main and interaction factor is in the form $\mathbf{I}\sigma_1^2$.

By properties (i) and (ii) and the structure of the design matrices (any gene replicated an equal number of times on each array), the last two conditions in (4.14) do not hold. A detailed proof is provided in Section B of the Appendix. Therefore, for mixed models with random global effects and corresponding random interactions involving the gene factor, the one- and two-step analyses are not equivalent.

Several remarks:

- (1) In general, the conditions in (4.14) for the separation are sufficient but not necessary. There are certain balanced cases where the full model and two-step analyses yield the same estimates and test statistics even if some of the conditions in (4.14) are violated. Suppose we have a replicated dye-swap experiment and we fit model (4.1) with random A and AG. When using the properties of the random effects solutions and imposing Σ -restrictions on the fixed factors in the model, it can be shown that under the full model we have $\hat{G}_l + \hat{\mu} = \bar{y}_{...l}$ for $l = 1, 2, \dots, g$, $\hat{V}_k + \hat{\mu} = \bar{y}_{..k..}$ for $k = 1, 2$, $\hat{\mu} = \bar{y}_{.....}$, and $\hat{\mu} + \hat{V}_k + \hat{G}_l + (\hat{VG})_{kl} = \bar{y}_{..kl}$ for any combination of l and k . Under the two-step analysis, the normalization model yields estimates of V being $\hat{V}_k + \hat{\mu} = \bar{y}_{..k..}$ for $k = 1, 2$ and $\hat{\mu} = \bar{y}_{.....}$. Then for the l^{th} gene, the gene model yields estimates $(\hat{VG})_{kl} = \bar{y}_{..kl} - \bar{y}_{..k..} - \bar{y}_{...l} + \hat{\mu}$. Hence the one- and two-step analyses give the same parameter estimates, although some conditions are not satisfied. In the replicated dye-swap design, the estimation of these parameters is independent of the random effects due to the orthogonal structure. In addition, the variance of contrast (4.2) is $\frac{\sigma_\epsilon^2}{ar}$, which is also independent of other random factors. This result shows that

hypothesis tests under the replicated dye-swap experiment are identical between one- and two-step analyses.

- (2) The above discussions apply to unrestricted mixed models. When Σ -restrictions are imposed on the model including restrictions on the random interaction factor $\mathbf{u}_2 = \mathbf{AG}$ of the form $\sum_j (AG)_{ij} = 0$ (summation over the levels of the fixed factor) for all i 's, this introduces a covariance between two (AG) interaction effects pertaining to the same array but different genes (Searle *et al.*, 1993). Then the MME in the second stage would no longer be block diagonal, and hence would not separate into individual gene models.
- (3) If the mixed model contains random global factors which do not interact with the gene factor, then the two-step analysis produces the same test statistics as the full model analysis by the arguments for fixed models in Section 2.2.2.

4.4.2 Modified Two-Step Mixed Model Analysis

The preceding discussions have shown that a general mixed effects model cannot be split into two steps if the model contains random global factors which interact with the gene factor (G). Here we show how two-step analysis for a mixed model with at least one random factor interacting with the gene factor can be modified to provide the same test statistics as the full model analysis. The modification consists of including the random interactions factor into the residuals of the normalization model in the first step.

We first assume that only one random global factor (A) interacts with the gene factor (G). Terms $(AG)_{ig}$ and ϵ_{ijkgl} in (4.1) are joined to form a new error term $e_{ijkgl} = (AG)_{ig} + \epsilon_{ijkgl}$ with variance-covariance matrix $\mathbf{R}\sigma_\epsilon^2$ for $\mathbf{R} = \mathbf{I}_N + \mathbf{Z}_2\mathbf{Z}_2' \frac{1}{\lambda_{AG}}$. In order to split the augmented MME with Σ -restrictions into a first system based on a global normalization model in $\hat{\beta}_1$ and $\hat{\mathbf{u}}_1$ and into gene-specific systems in $\hat{\beta}_{2j}$, two conditions are needed (analogous to Section 3.1):

$$\mathbf{X}'_1\mathbf{R}^{-1}\mathbf{X}_2\hat{\beta}_2 = \mathbf{0}, \quad \mathbf{Z}'_1\mathbf{R}^{-1}\mathbf{X}_2\hat{\beta}_2 = \mathbf{0}. \quad (4.15)$$

Again, the equivalence of the test statistics can be established. A detailed proof is provided in Section C of the Appendix.

4.5 Simulation Studies

The performance of one- and two-step mixed model analysis can be further evaluated by Monte Carlo simulation. In this section, these two approaches are compared using two designs: the loop design and the reference design, which can be viewed as partially balanced and unbalanced designs, respectively.

4.5.1 Data Simulation and Analyses

The data were simulated using model (4.1). Factors A and AG were assumed to be random and i.i.d normal with mean 0 and variances $\sigma_A^2 = 1$ and $\sigma_{AG}^2 = 0.5$, respectively. The residuals were also i.i.d normal with mean 0 and variance $\sigma_e^2 = 1$. Five varieties of interest were used for each design (5 varieties of interest and a reference variety for the reference design). To generate the differential expression, the mean expression level of each gene in each variety was first generated, with differences as in (4.2) between 2 and 5 (randomly generated from a uniform distribution). 80 genes were assumed to be present on each array, and gene main effects (G) were derived from the VG cell means. In addition, we assume that in the true state, each gene is differentially expressed only between sample 1 and other tissue samples. Under these assumptions, among 800 pairwise comparisons, 320 would correspond to differential expression.

The simulated data were first analyzed with the full mixed effects model, and hypothesis tests for differential gene expression were based on the contrasts in (4.2). Next, the simulated data were analyzed with the original two-step approach by Wolfinger *et al.* (2001) using contrasts in the (VG) effects only. Thirdly, the data were analyzed with a modified two-step approach, which used contrast (2.2) and computed the standard error of the contrasts by assuming zero covariance between V and VG estimates and by using an average residual variance for the individual gene models. In either analysis, unknown variance components (e.g., σ_A^2 , σ_{AG}^2 , σ_e^2) were estimated by REML both in the full model and the two-step analyses. Variance component σ_{AG}^2 was estimated from the data on a single gene in the two-stage approach. The problem of multiple testing was accounted for by the Benferroni adjustment to the raw p -values ($p \leq 0.05$). A commercial statistical package, SAS (Littell *et al.*, 1996), was used to perform the simulation and the analyses.

The three methods were compared based on the number of false positives (genes not differentially expressed in true state but identified to be differentially expressed from simulated data), the number of false negatives (gene differentially expressed in true state but identified to be not differentially expressed from simulated data), and the total number of misclassifications (the sum of the numbers of false positives and false negatives). Results for each design are reported based on 100 replicated data sets.

4.5.2 Loop Design

The combination of the varieties with two dyes is listed in Table 4.2, where column, row and cell correspond to array, dye and variety, respectively.

The data from the simulated design were analyzed based on model (4.1). Results appear in Tables 4.3 and 4.4. When $\sigma_{AG}=0.5$, the modified two-step analysis performed about as well as the full model analysis, while the two-step approach of Wolfinger *et al.* (2001) performed much worse than the other methods. As σ_{AG}^2 increases to 2, the number of false positives

Table 4.2: Combination of Varieties with Dyes for the Loop Design

V_1	V_2	V_3	V_4	V_5
V_2	V_3	V_4	V_5	V_1

increases, and also shows larger variation. The discrepancy between the one-step and the modified two-step analyses also increases. We note that the performance of the modified two-step analysis declined, also producing a number of false positives as spot effects increased.

In the true state, the main effects for V were significant while the interaction VG was not. Therefore, the two-step analysis from Wolfinger *et al.* (2001) using contrasts of VG, produced a large number of false negatives. For either approach, very few false positives were reported, because the Bonferroni method was used to adjust the raw p -values. The size of σ_{AG}^2 also affected the performance of these analyses. As σ_{AG}^2 increased, the variance of a given contrast also increased, thus either approach produced more false negatives. However, the modified two-step approach also produced a few false positives, because correlations between estimated main effects V and VG were ignored when the test statistics were computed for the modified two-step analysis. We note that because the increase of σ_A^2 has similar impact on the variance of contrasts as that of σ_{AG}^2 (Yang *et al.*, 2003a), these methods perform similarly as the case of σ_{AG}^2 .

Table 4.3: Summary of Simulation Results for the Loop Design ($\sigma_{AG}^2 = 0.5$): Median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.

Methods	False Negatives	False Positives	Total Misclassification
One-step	17 (17.63, 5.61)	0 (0.07, 0.26)	17 (17.7, 5.61)
Two-step	319 (318.9, 1.11)	0 (0.18, 0.39)	319 (319.1, 1.56)
Modified Two-step	16.5(16.16, 4.04)	1 (0.76, 0.95)	17 (16.92, 4.19)

4.5.3 Reference Design

The simulated reference design included five varieties of interest (V_1, \dots, V_5), and the reference sample (V_6). The design was depicted in Table 4.5, where columns, rows, and cells represent arrays, dyes and varieties, respectively. It was assumed that each gene was spotted twice on each array. The data were analyzed with the mixed model as in the loop design, except that factor dye (D) was omitted. Simulation results appear in Table 4.6 and 4.7. When $\sigma_{AG}^2=0.5$, the modified two-step approach performed only slightly worse than the full

Table 4.4: Summary of Simulation Results for the Loop Design ($\sigma_{AG}^2 = 2$): Median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.

Methods	False Negatives	False Positives	Total Misclassification
One-step	38.5 (39.82, 17.47)	0 (0.26, 0.58)	40 (40.08, 17.37)
Two-step	320 (319.82, 0.69)	0 (0.32, 1.22)	320 (320.14, 1.14)
Modified Two-step	56 (55.26, 12.49)	5 (4.19, 2.42)	61 (59.45, 12.22)

Table 4.5: Combination of Varieties with Dyes for the Reference Design

V_1	V_2	V_3	V_4	V_5
V_6	V_6	V_6	V_6	V_6

model analysis, and the original two-step approach again performed considerably worse. As σ_{AG}^2 increased to 2, the number of false positives increased, and also showed larger variation. The discrepancy between the one-step and the modified two-step analyses also increased.

Again, the main effects for factor V were significant while VG was not significant in the true state. Therefore, use of VG contrasts resulted in a large number of false negatives for Wolfinger’s two-step analysis. The big difference between Wolfinger’s two-step analysis and the modified two-step analysis was due to use of different contrasts and the adjustment of standard errors of estimated contrasts. A number of false positives showed up for the modified two-step analysis, which was not surprising. The numerator of the t test statistic in second step was inflated for some genes, but the correlation between main effects V and the VG effects in the two-step analysis was not properly accounted for, thus producing these false positives. The estimates of unknown variance components played a role here, which affected not only estimates of adjustments for V, but also the estimates of VG in the second step. We also note that the reference design performed worse than the loop design under either method because of its lower statistical efficiency for estimating contrasts of interest (Yang *et al.*, 2003a).

4.6 Case Studies

In Section 3, we found that a general mixed effects model cannot be split into a two-step analysis. We further illustrate their differences by applying mixed models to three gene expression data sets. We now present results from the analyses.

Table 4.6: Summary of Simulation Results for the Reference Design ($\sigma_{AG}^2 = 0.5$): median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.

Methods	False Negatives	False Positives	Total Misclassification
One-step	37 (37.51, 7.82)	0 (0.33, 0.65)	37 (37.84, 7.93)
Two-step	320 (319.34, 2.81)	0 (0.36, 5.98)	320 (319.7, 1.58)
Modified Two-step	57 (56.18, 9.51)	11 (11.18, 2.84)	68 (67.36, 9.68)

Table 4.7: Summary of Simulation Results for the Reference Design ($\sigma_{AG}^2 = 2$): median number of misclassification is listed while mean and standard deviation are reported in brackets, respectively. Results are based on 100 replicated data sets.

Methods	False Negatives	False Positives	Total Misclassification
One-step	46 (45.44, 9.80)	0 (0.24, 0.53)	46 (45.68, 9.85)
Two-step	320 (319.24, 1.04)	0 (0.30, 0.58)	320 (319.54, 1.10)
Modified Two-step	74 (73.96, 9.77)	10 (9.88, 2.32)	84 (83.84, 10.19)

4.6.1 The Mouse Model

The first data set was from Dudoit *et al.* (2002). The goal of this study was to identify genes with altered expression in the livers of two lines of mice with very low HDL cholesterol levels compared to inbred control mice. The treatment group consisted of eight mice with the apo AI gene knocked-out and the control group consisted of eight normal C57Bl/6 mice. For each of these 16 mice, target cDNA was obtained from mRNA by reverse transcription and labeled using a red-fluorescent dye (Cy5). The reference sample (green-fluorescent dye Cy3) was prepared by pooling cDNA from the eight control mice. The target cDNA was hybridized to 16 arrays each consisting of 6384 spots. The following nested mixed model was used for reasons discussed in Yang *et al.* (2003b):

$$y_{ijkl} = \mu + T_j + M_k(T_j) + G_l + (TG)_{jl} + \epsilon_{ijkl},$$

where response y_{ijkl} was the log-ratio of intensities at each spot from array i , cell population j , mouse k and gene l . $M(T)$ was assumed random with dispersion assumption $M_k(T_j) \sim \text{iid } N(0, \sigma_{MT}^2)$, and $\epsilon_{ijkl} \sim \text{iid } N(0, \sigma_e^2)$.

As suggested by Dudoit *et al.* (2002), log-ratios from raw data were adjusted for intensity-dependent biases using a local regression function (LOESS). A statistical package ASREML (Gilmour *et al.*, 2002) was used to estimate the variance components and the test statistics. Combined with the sparse matrix technique, ASREML uses the algorithm of average information residual maximum likelihood (REML) for variance parameter estimation (Gilmour

Table 4.8: Adjusted p -values for Differentially Expressed Apo AI. Genes: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.05, and gene selection from the step-down Bonferroni was based on p -value=0.05.

Gene ID	Bonferroni adjustment ($p \leq 1.57E - 6$)	Step-down Bonferroni (p -value ≤ 0.01)
540	0.00E+00	0.00E+00
541	4.78E-07	3.05E-03
1496	2.56E-09	1.63E-05
1739	2.82E-08	1.80E-04
2149	0.00E+00	0.00E+00
2537	3.64E-11	2.32E-07
3729	1.11E-06	7.10E-03
4139	5.52E-12	3.52E-08
4941	6.55E-09	4.17E-05
5356	0.00E+00	0.00E+00
5986	1.25E-07	7.95E-04

et al., 1995). The null hypothesis was tested using contrasts in (4.2), and the problem of multiple testing was accounted for by adjusting raw p -values using the Bonferroni procedure and the step-down Bonferroni method of Holm (1979). The Bonferroni procedure makes single-step adjustments to the raw p -values regardless of the ordering of the unadjusted p -values, which is very conservative. The step-down Bonferroni procedure orders the p -values and makes successively smaller adjustments to p values, which is less conservative. A detailed discussion of their properties can be found in Westfall and Young (1993). Results appeared in Table 4.8. Eleven genes were found differentially expressed by both methods of adjustments, but 3 genes (541, 3729, 5896) were not reported in Dudoit *et al.* (2002). This result showed that the mixed effects model performed at least as well as the two-sample t tests in terms of detecting gene differential expression.

4.6.2 Bovine Embryo Classification Study

The second data set was from Infigen, a biotechnology company. Its bovine embryo classification study was to classify NT embryos developmentally incompetent against IVF embryos developmentally competent. Ten embryos were collected from each cell population (NT or IVF). An interwoven loop design was used, where 2640 genes were spotted twice on each of 100 arrays. The following nested mixed effects model was used for the analysis:

$$\begin{aligned}
 y_{ijklmn} &= \mu + A_i + D_j + (AD)_{ij} + T_k + E_l(T_k) \\
 &+ G_m + (TG)_{km} + (AG)_{im} + G_m E_l(T_k) + \epsilon_{ijklmn},
 \end{aligned}
 \tag{4.16}$$

where response y_{ijklmn} was the transformed intensity from array i , dye j , cell population k , embryo l and gene m . Effects of labelling reactions were estimated with factor AD . Factors A , (AD) , (AG) , $E(T)$ and $GE(T)$ were assumed random with dispersion assumptions $A_i \sim \text{iid } N(0, \sigma_A^2)$, $(AD)_{ij} \sim \text{iid } N(0, \sigma_{AD}^2)$, $E_l(T_k) \sim \text{iid } N(0, \sigma_{ET}^2)$, $(AG)_{im} \sim \text{iid } N(0, \sigma_{AG}^2)$, $G_m E_l(T_k) \sim \text{iid } N(0, \sigma_{EG}^2)$, and $e_{ijklmn} \sim \text{iid } N(0, \sigma_e^2)$.

The scatter plot of log ratios versus the average of log intensity (RI plots) revealed intensity-dependent biases in the two color channels. The raw data were log-transformed and normalized using LOESS and spatial LOESS functions (Cui *et al.*, 2002). Again, ASREML was used to estimate the variances components and test statistics, and the null hypothesis was tested using contrasts in (4.2). The problem of multiple testing was accounted for by adjusting the raw p -values using Bonferroni and the step-down Bonferroni methods. Nineteen genes were identified to be differentially expressed (Table 4.9), and the two methods of adjustments produced the same set of genes. Results were also obtained from the two-step analysis by Wolfinger *et al.* (2001) using contrasts of VG . Only one gene (Gene 329) was found differentially expressed, which was also verified from the full model analysis. The modified two-step analysis in Section 3.3 was also applied using contrasts in (4.2) and only gene 329 was found differentially expressed.

The difference between the full model analysis and the two-step analysis of Wolfinger *et al.* (2001) was mainly caused by different contrasts used, because there were significant main effects for factor Type (T) (p -value=0.02). In the modified two-step analysis, the standard errors of the contrasts were inflated due to the larger estimates of variance components in the second step, resulting in a smaller test statistic with fewer degrees of freedom (18). Therefore, fewer genes were identified to be differentially expressed.

4.6.3 Porcine Embryo Classification Study

The third data set was from Infigen's porcine embryo classification study whose goal was to classify NT embryos against IVF embryos developmentally competent. Ten embryos were collected from each cell line (NT or IVF). An interwoven loop design was used, where 2640 genes were spotted twice on each of 100 arrays. The nested mixed model in (4.16) was used for the analyses with the same dispersion assumptions.

The raw data were log-transformed and normalized using LOESS and spatial LOESS functions due to strong systematics biases in dye effects. Test statistics were estimated by ASREML, and the null hypothesis was tested using contrasts in (4.2). Raw p -values were adjusted similarly as in the bovine experiment. Differentially expressed genes were partially summarized in Table 4.10. 146 genes were found differentially expressed by the Bonferroni adjustment ($FWE \leq 0.01$), and two more genes (Gene 568 and 2582) were identified by the step-down Bonferroni method. The two-step analysis by Wolfinger *et al.* (2001) was also applied using contrasts of VG and only two genes (Gene 1311 and 1425) were identified to be differentially expressed, but none of them was verified by the full model analysis. The

Table 4.9: Adjusted p -values for Differentially Expressed Genes in Bovine Study: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.05, and gene selection from the step-down Bonferroni was based on p -value=0.05.

Gene ID	Bonferroni adjustment ($p \leq 1.89E - 05$)	Step-down Bonferroni ($p\text{-value} \leq 0.05$)
112	1.74E-05	4.55E-02
198	8.27E-06	2.17E-02
329	0.00E+00	0.00E+00
461	2.04E-08	5.37E-05
620	1.43E-05	3.76E-02
767	2.15E-07	5.66E-04
769	6.68E-11	1.76E-07
982	3.91E-06	1.03E-02
1067	4.43E-07	1.17E-03
1271	1.25E-07	3.31E-04
1288	1.85E-06	4.85E-03
1423	8.91E-08	2.35E-04
1539	4.89E-06	1.28E-02
1737	1.47E-06	3.88E-03
1778	5.51E-06	1.45E-02
1820	6.39E-07	1.68E-03
2295	2.50E-08	6.60E-05
2410	1.47E-05	3.85E-02
2547	1.41E-06	3.70E-03

modified two-step analysis in Section 3.3 was also applied using contrasts in (4.2), but none of the genes was found differentially expressed.

For the two-step analyses, because the main effects for factor Type (T) were not significant (p -value=0.36), results from the modified and the original two-step approaches were almost the same for either type of contrasts. However, the two-step analyses produced a large number of false negatives, because the estimates of variance components from the second step showed very large variations and standard errors of contrasts for many genes were inflated while a much smaller degree of freedom (18) was used for hypothesis testing. Take gene 6 for an example, in the full model analysis, the contrast was estimated to be 1.475 with a standard error of 0.30 and very large number of degrees of freedom, and this gene was found differentially expressed. But under Wolfinger's two-step analysis, the estimated contrast was only 1.256 with standard error of 0.314 and 18 degrees of freedom, resulting in a smaller test statistic and a larger cut-off value, thus leading to the conclusion of no differential expression.

4.6.4 Discussion

In the above analyses only a common error was assumed for all genes, implying the estimates of all contrasts have the same standard error, which seems unrealistic (Dudoit *et al.*, 2002). Ideally, we could assume an error variance for each gene, but this is computationally prohibitive and unnecessary. In the bovine and porcine experiments, groups of genes showed similar variation in expression values, but large variations were observed for highly expressed genes while lowly expressed genes showed much smaller variations. Therefore, models with a common error variance often overestimate the standard errors of contrasts for lowly expressed genes but underestimate standard errors of contrasts for highly expressed genes. Our recommendation is to group genes based on their variations and expression values. For the bovine experiment, we assigned genes into 9 different groups each with a separate error variance. Initial screening of normalized data revealed that 29 genes were lowly expressed with the average variance of 0.08, 58 genes are highly expressed with average variance of 1.5, and two groups were assigned to these genes. The variation of other genes ranged from 0.11 to 0.98, and were further divided into seven groups with average variance of 0.14, 0.18, 0.22, 0.28, 0.35, 0.45 and 0.62. As a result, 17 of 19 previously reported genes were identified except genes 198 and 2410. The raw data revealed that the two genes were highly expressed with large variances. Without grouping, their variability was underestimated when a common error variance was assumed. For the porcine experiment, we assigned genes into 7 groups based on their similarity in expression values and variations. From the normalized data, the average variance for each group was 0.56, 0.64, 0.79, 0.97, 1.04, 1.32, 1.86. One more gene (Gene 2582) was identified. For both experiments, the by-group residual plots were used to assess how grouping worked. Although the choice of the number of groups is somewhat arbitrary and certainly depends on features of a given data set, we have found the by-group residual plots particularly useful for the visual justification of the grouping.

Table 4.10: Adjusted p -values for Differentially Expressed Genes in Porcine Study: p -values from Bonferroni adjustment were obtained by controlling the familywise error rate at 0.01, and gene selection from the step-down Bonferroni was based on p -value=0.01.

Gene ID	Bonferroni adjustment ($p \leq 3.79E - 06$)	Step-down Bonferroni p value ≤ 0.01
6	8.19E-07	2.07E-03
24	9.84E-09	2.54E-05
41	3.61E-07	9.14E-04
48	2.00E-06	5.02E-03
51	1.14E-08	2.95E-05
55	1.34E-11	3.54E-08
66	1.73E-06	4.33E-03
75	4.38E-09	1.14E-05
95	2.54E-11	6.69E-08
100	9.94E-10	2.60E-06
104	4.08E-07	1.03E-03
137	1.82E-06	4.56E-03
144	2.71E-11	7.13E-08
178	1.64E-06	4.11E-03
181	1.39E-09	3.62E-06
195	2.46E-11	6.48E-08
217	1.70E-08	4.39E-05
242	3.01E-10	7.88E-07
266	2.86E-12	7.55E-09
309	1.18E-07	3.02E-04

The design of the bovine and porcine embryo experiments has a number of merits. First, dyes are always balanced with embryos, and systematic biases due to dyes can be taken into account in the modelling stage. Second, for both experiments, an interwoven loop design was used such that contrasts among samples within a cell population were balanced, and contrasts among samples across cell populations were also balanced, thus resulting in a higher efficiency for estimating contrasts of interest. Thirdly, because ten embryos were collected from each cell population, and each embryo was replicated with ten different labelling reactions, it is possible to take into account the variation between cell populations, variation due to biological replicates, and the technology variation (labelling effects). Therefore, the contrasts in (4.2) provide a clean measure of changes in expression values across cell populations. A natural question arising from this context is statistical design of microarray experiment. Very often, microarray experiments are designed such that variety main effects are confounded with sample or labelling effects, hence the contrasts in (4.2) are confounded with sample effects or labelling effects, and the changes in expression values across samples are not evaluated properly. We would recommend designs from which these variations can be fully accounted for.

Mixed models are promising for analyzing gene expression data, but they are computationally intensive for the full mixed model analysis. The modified two-step approach was proposed in Section 3.3 as an alternative to the full mixed model analysis. Our simulations showed that it seemed to work well for certain single-factor microarray experiments, but not for the nested mixed model analysis in Section 4. Several factors seem to affect its performance. First of all, statistical design affects its performance, which can be seen from our simulation studies in Section 3.3: the reference design performed worse than the loop design. Secondly, the estimates of unknown variance components have an impact on the performance. In the analyses of bovine and porcine experiments, the estimates of variance components from the second step often showed large variations, thus over- or underestimating standard errors of the estimated contrasts. Similar effects of unknown variance components on the performance of the modified two-step approach were also observed in our simulation studies. Thirdly, possible correlations between factor variety and the variety by gene interactions are ignored when adjustments are made in the modified two-step analysis, which also affects the performance of this method.

In the analyses of these data sets, prior normalization was done to remove systematic biases in the two fluorescent dyes and spatial heterogeneity, which seemed to work well with our ANOVA methods. Linear additivity is assumed for factors in mixed ANOVA models, which is useful for accounting for additive effects, but not adequate for correcting nonlinear systematic biases. For the bovine and porcine experiments, nonlinear intensity-dependent biases were detected from the two color channels. Without adjustment of these biases, 23 genes would have been found differentially expressed for the bovine experiment and 562 genes for the porcine experiments. With the combination of both types of adjustments, all sources of systematic variations can be identified and properly accounted for, which improves the efficiency of subsequent statistical analyses.

4.7 Conclusion

We first identified the correct set of contrasts to be used for testing gene differential expression under ANOVA methods. We analytically established the equivalence of the one- and two-step analysis under fixed effects ANOVA models, and showed that a general mixed model cannot be split into a two-step analysis as in Wolfinger *et al.* (2001). We also found that some mixed effects model with specific variance-covariance structures can be split into a two-step analysis, which also yields identical hypothesis testing. In general, a two-step analysis under mixed effects model may not be suitable for detecting gene differential expression. Our case studies demonstrated that statistical packages such as ASREML can be used to carry out a full mixed model analysis for some microarray data sets.

In our study, the equivalence of certain mixed effects models under the one- and two-step analyses was established under the assumption of known variance components. The presence of unknown variance components can influence the estimation of other effects including the contrasts of interest. Many methods (Searle *et al.*, 1992) can be used to estimate these variance components. Our further research is to examine the equivalence between one- and two-step analyses with unknown variance components, and to develop computationally more efficient algorithms for analysis of gene expression data using mixed ANOVA models.

Appendix

A: Equivalence Under Fixed Effects Models

The equivalence of test statistics under fixed effects models can be established as follows. From Equation (4.6) and the results on inverses of partitioned matrices, we find that $Var(\hat{\beta}_1)$ in the full model is the $p_1 \times p_1$ upper sub-matrix of the $(2p_1 - rank(X_1)) \times (2p_1 - rank(X_1))$ matrix

$$\begin{aligned} \begin{pmatrix} \mathbf{X}'_1\mathbf{X}_1 - \mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11}\mathbf{X}'_2\mathbf{X}_1 & \mathbf{C}_1 \\ \mathbf{C}'_1 & \mathbf{0} \end{pmatrix}^{-1} &= \begin{pmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{C}_1 \\ \mathbf{C}'_1 & \mathbf{0} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix}, \end{aligned} \quad (4.17)$$

if $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$ and where

$$\begin{pmatrix} \mathbf{X}'_2\mathbf{X}_2 & \mathbf{C}_2 \\ \mathbf{C}'_2 & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix}.$$

The variance-covariance matrix of $\hat{\beta}_2$ for the second step in (4.8) is

$$Var(\hat{\beta}_2) = Var(\mathbf{D}_{11}\mathbf{X}'_2(\mathbf{y} - \mathbf{X}_1\hat{\beta}_1))$$

$$\begin{aligned}
&= \mathbf{D}_{11}\mathbf{X}'_2\text{Var}(\mathbf{y} - \mathbf{X}_1\mathbf{B}_{11}\mathbf{X}'_1\mathbf{y})\mathbf{X}_2\mathbf{D}_{11} \\
&= \mathbf{D}_{11}\mathbf{X}'_2[\mathbf{I} - \mathbf{X}_1\mathbf{B}_{11}\mathbf{X}'_1]\mathbf{X}_2\mathbf{D}_{11}\sigma_\epsilon^2 \\
&= \mathbf{D}_{11}\sigma_\epsilon^2,
\end{aligned} \tag{4.18}$$

if $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$. Hence $\text{Var}(\hat{\beta}_1)$ and $\text{Var}(\hat{\beta}_2)$ under the full model can be computed from the inverse of the coefficient matrices of (4.7) and (4.8), respectively.

For the covariance matrix of $\hat{\beta}_1$ and $\hat{\beta}_2$, we have

$$\begin{aligned}
\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) &= \text{Cov}(\mathbf{B}_{11}\mathbf{X}'_1\mathbf{y}, \mathbf{y}'\mathbf{X}_2\mathbf{D}_{11}) - \text{Var}(\hat{\beta}_1)\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} \\
&= \mathbf{B}_{11}\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11}\sigma_\epsilon^2 - \text{Var}(\hat{\beta}_1)\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} \\
&= \mathbf{0},
\end{aligned} \tag{4.19}$$

if $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$. Hence $\text{Var}(\hat{\mu}_i)$ is as in (4.9) but with the covariance term equal to zero and with $\text{Var}(\hat{V}_k - \hat{V}_{k'})$ and $\text{Var}((\hat{V}G)_{kl} - (\hat{V}G)_{k'l})$ computed from the inverse coefficient matrix in (4.7) and (4.8). These results hold iff $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$. With the simple Σ -restrictions, for example, on the solutions for the levels of factor G , we have $\sum_i \hat{G}_i = 0$, implying $\text{Var}(\sum_i \hat{G}_i) = 0$. And if $\text{Var}(\hat{G}_i)$ is constant for all i then we have

$$\text{Var}(\hat{G}_i) + \sum_{j \neq i} \text{Cov}(\hat{G}_i, \hat{G}_j) = 0.$$

This result and similar results for other gene-specific factors, combined with the structure of $\mathbf{X}'_1\mathbf{X}_2$ prove that $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$, implying equations (4.17), (4.18) and (4.19) hold.

When each gene is replicated an equal number of times across all arrays but different genes have different numbers of replicates, then the desired equalities can be established by imposing weighted Σ -restrictions on the solutions. In this case, we have $\sum_i w_i \hat{G}_i = 0$ where w_i is the number of times gene i appears on an array, which implies that $\text{Var}(\sum_i w_i \hat{G}_i) = 0$ for all i , or

$$w_i \text{Var}(\hat{G}_i) + \sum_{j \neq i} w_j \text{Cov}(\hat{G}_i, \hat{G}_j) = 0.$$

The weighted Σ -restrictions and (4.18) with the structure of $\mathbf{X}'_1\mathbf{X}_2$, show that $\mathbf{X}'_1\mathbf{X}_2\hat{\beta}_2 = \mathbf{0}$ and $\mathbf{X}'_1\mathbf{X}_2\mathbf{D}_{11} = \mathbf{0}$, implying the two-step analysis yields tests identical to those from the full model.

B: Unequivalence Under General Mixed Effects Models

Under assumptions given in Section 3.1, the mixed model equations (MME) are:

$$\begin{pmatrix} \mathbf{X}'_1\mathbf{X}_1 & \mathbf{X}'_1\mathbf{Z}_1 & \mathbf{X}'_1\mathbf{X}_2 & \mathbf{X}'_1\mathbf{Z}_2 \\ \mathbf{Z}'_1\mathbf{X}_1 & \mathbf{Z}'_1\mathbf{Z}_1 + \lambda_1\mathbf{I}_a & \mathbf{Z}'_1\mathbf{X}_2 & \mathbf{Z}'_1\mathbf{Z}_2 \\ \mathbf{X}'_2\mathbf{X}_1 & \mathbf{X}'_2\mathbf{Z}_1 & \mathbf{X}'_2\mathbf{X}_2 & \mathbf{X}'_2\mathbf{Z}_2 \\ \mathbf{Z}'_2\mathbf{X}_1 & \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{X}_2 & \mathbf{Z}'_2\mathbf{Z}_2 + \lambda_2\mathbf{I}_{ag} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\mathbf{u}}_1 \\ \hat{\beta}_2 \\ \hat{\mathbf{u}}_2 \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{X}_1' \mathbf{y} \\ \mathbf{Z}_1' \mathbf{y} \\ \mathbf{X}_2' \mathbf{y} \\ \mathbf{Z}_2' \mathbf{y} \end{pmatrix}, \quad (4.20)$$

where $\lambda_1 = \sigma_\epsilon^2 / \sigma_a^2$ and $\lambda_2 = \sigma_\epsilon^2 / \sigma_{ag}^2$.

For the two-step analysis, to get the same estimates for β_1 and β_2 , and the same predictions for \mathbf{u}_1 and \mathbf{u}_2 as model (4.11), the four conditions in (4.14) must be satisfied. The Σ -restrictions ensure that the first two equations hold. By properties of solutions to the random effects from Section 3.1, we show that the last two conditions in (4.14) do not hold.

For $\mathbf{X}_1' \mathbf{Z}_2 \hat{\mathbf{u}}_2$, assuming $\mathbf{X}_1 = [\mathbf{1} \ \mathbf{X}_D \ \mathbf{X}_V]$, we have

$$\mathbf{X}_1' \mathbf{Z}_2 \hat{\mathbf{u}}_2 = \begin{pmatrix} \mathbf{1}' \mathbf{Z}_2 \hat{\mathbf{u}}_2 \\ \mathbf{X}_D' \mathbf{Z}_2 \hat{\mathbf{u}}_2 \\ \mathbf{X}_V' \mathbf{Z}_2 \hat{\mathbf{u}}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{k} \end{pmatrix}. \quad (4.21)$$

By properties (i) and (ii) and the structure of the design matrices (any gene replicated an equal number of times on each array), the first two sub-vectors of (4.21) are zero, while vector \mathbf{k} is not. One particular element in \mathbf{k} , corresponding to variety i , is equal to

$$\sum_{j(i)} \sum_k \hat{A}G_{j(i)k} = \frac{\lambda_a}{\lambda_{ag}} \sum_{j(i)} \hat{A}_{j(i)},$$

where $j(i)$ denotes any array j which is associated with variety i . Note that $\mathbf{k} = \mathbf{0}$ only if each variety appears on each array (by property (ii)). The latter only holds when $\nu = 2$ (two varieties) or for multi-dye experiments with number of dyes equal to number of varieties.

For $\mathbf{Z}_1' \mathbf{Z}_2 \hat{\mathbf{u}}_2$, using property (ii), we find that

$$\mathbf{Z}_1' \mathbf{Z}_2 \hat{\mathbf{u}}_2 = \mathbf{d} \begin{pmatrix} \sum_{l=1}^g (\hat{A}G)_{1l}^j \\ \sum_{l=1}^g (\hat{A}G)_{2l}^j \\ \vdots \\ \sum_{l=1}^g (\hat{A}G)_{al}^j \end{pmatrix} = \mathbf{d} \frac{\lambda_1}{\lambda_2} \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \\ \vdots \\ \hat{A}_a \end{pmatrix},$$

which is also not zero (as long as $\sigma_A^2 > 0$) and d is some constant. Therefore, the last two conditions do not hold.

C: Equivalence Under Modified Mixed Effects Models

Following the assumptions in Section 3.2, the MME for the full model, after adding Σ -restrictions for the fixed factor solutions as in Section 2.2.2 are

$$\begin{aligned}
 & \begin{pmatrix} \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{Z}_1 & \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{X}_2 & \mathbf{C}_1 & \mathbf{0} \\ \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{Z}_1 + \mathbf{I} \lambda_A & \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{X}_2 & \mathbf{0} & \mathbf{C}_2 \\ \mathbf{X}'_2 \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{R}^{-1} \mathbf{Z}_1 & \mathbf{X}'_2 \mathbf{R}^{-1} \mathbf{X}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{C}'_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}'_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\mathbf{u}}_1 \\ \hat{\beta}_2 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{X}'_2 \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}.
 \end{aligned}$$

But $\mathbf{R}^{-1} = \mathbf{I} - \delta \mathbf{Z}_2 \mathbf{Z}'_2$, where δ is some constant which depends on λ_{AG} and the number of times each gene is present on each array. Then $\mathbf{X}_1 \mathbf{R}^{-1} \mathbf{X}_2 = \mathbf{X}'_1 \mathbf{X}_2 - \delta \mathbf{X}'_1 \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{X}_2$, and $\mathbf{Z}_1 \mathbf{R}^{-1} \mathbf{X}_2 = \mathbf{Z}'_1 \mathbf{X}_2 - \delta \mathbf{Z}'_1 \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{X}_2$. Similar to the arguments in Section 2.2.2, because all rows in $\mathbf{X}_1 \mathbf{R}^{-1} \mathbf{X}_2$ and $\mathbf{Z}_1 \mathbf{R}^{-1} \mathbf{X}_2$ corresponding to one factor in β_1 are equal, and each row in $\mathbf{X}_1 \mathbf{R}^{-1} \mathbf{X}_2$ and $\mathbf{Z}_1 \mathbf{R}^{-1} \mathbf{X}_2$ has several partition corresponding to β_2 with all elements equal within a partition, then imposing the simple Σ -restrictions on the solutions results in the desired equalities. Then estimates and predictors computed from the augmented MME, can be alternatively computed with the two-step approach consisting of the global normalization model

$$\begin{pmatrix} \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{Z}_1 & \mathbf{C}_1 \\ \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{X}_1 & \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{Z}_1 + \mathbf{I} \lambda_A & \mathbf{0} \\ \mathbf{C}'_1 & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\mathbf{u}}_1 \\ \lambda_1 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}'_1 \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{0} \end{pmatrix},$$

and the gene model:

$$\begin{pmatrix} \mathbf{X}'_2 \mathbf{R}^{-1} \mathbf{X}_2 & \mathbf{C}_2 \\ \mathbf{C}'_2 & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_2 \mathbf{R}^{-1} \hat{\mathbf{r}} \\ \mathbf{0} \end{pmatrix},$$

where $\hat{\mathbf{r}} = \mathbf{y} - \mathbf{X}_1 \hat{\beta}_1 - \mathbf{Z}_1 \hat{\mathbf{u}}_1$. However, in the second step $\hat{\beta}_2$ can be computed equivalently and efficiently by fitting random AG effects and splitting the system into gene-specific systems in $\hat{\beta}_2$ and $\hat{\mathbf{u}}_2$:

$$\begin{pmatrix} \mathbf{X}'_2 \mathbf{X}_2 & \mathbf{X}'_2 \mathbf{Z}_2 & \mathbf{C}_2 \\ \mathbf{Z}'_2 \mathbf{X}_2 & \mathbf{Z}'_2 \mathbf{Z}_2 + \mathbf{I} \lambda_{AG} & \mathbf{0} \\ \mathbf{C}'_2 & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\mathbf{u}}_2 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_2 \hat{\mathbf{r}} \\ \mathbf{Z}'_2 \hat{\mathbf{r}} \\ \mathbf{0} \end{pmatrix}.$$

To obtain the identical test statistics under both approaches, recall that the estimates of β can also be obtained from the method of Generalized Least Squares (GLS), which is

equivalent to the MME by absorbing random effects (Searle *et al.*, 1993, pp. 276). With Σ -constraints imposed, the NE for estimating β_1 and β_2 are

$$\begin{pmatrix} \mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_2 & \mathbf{C}_1 & \mathbf{0} \\ \mathbf{X}'_2 \mathbf{V}^{-1} \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{V}^{-1} \mathbf{X}_2 & \mathbf{0} & \mathbf{C}_2 \\ \mathbf{C}'_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}'_2 & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{y} \\ \mathbf{X}'_2 \mathbf{V}^{-1} \mathbf{y} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{V} = \mathbf{I}\sigma_e^2 + \mathbf{Z}_1 \mathbf{Z}'_1 \sigma_a^2 + \mathbf{Z}_2 \mathbf{Z}'_2 \sigma_{ag}^2$, which is the variance-covariance matrix for \mathbf{y} . In addition, it can be shown that $\mathbf{V}^{-1} = \delta_1 \mathbf{I} - \delta_2 \mathbf{Z}_1 \mathbf{Z}'_1 - \delta_3 \mathbf{Z}_2 \mathbf{Z}'_2$, where δ_1 , δ_2 , and δ_3 are constants which depend on the sizes of variance components, the number of genes, and the number of times each gene is present on each array. Therefore, we have

$$\mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_2 = \delta_1 \mathbf{X}'_1 \mathbf{X}_2 - \delta_2 \mathbf{X}'_1 \mathbf{Z}_1 \mathbf{Z}'_1 \mathbf{X}_2 - \delta_3 \mathbf{X}'_1 \mathbf{Z}_2 \mathbf{Z}'_2 \mathbf{X}_2.$$

Similar to arguments in Section 2.2.2, simple Σ -restrictions ensure that $\mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_2 \hat{\beta}_2 = \mathbf{0}$ holds. The equivalence of test statistics under both approaches can be established if equality $\mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_2 \mathbf{D}_{11} = \mathbf{0}$ holds where

$$\begin{pmatrix} \mathbf{X}'_2 \mathbf{V}^{-1} \mathbf{X}_2 & \mathbf{C}_2 \\ \mathbf{C}'_2 & \mathbf{0} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix}.$$

Again, by the structure of $\mathbf{X}'_1 \mathbf{V}^{-1} \mathbf{X}_2$ and the imposed Σ -constraints, this condition holds, which implies zero covariance between $\hat{\beta}_1$ and $\hat{\beta}_2$ from the two-step analysis. Thus the two approaches produce identical test statistics.

Chapter 5

Future Research

5.1 Discussion

The goals of this research are optimal design of single factor microarray experiments and application of mixed effects models to gene expression data. For design of single factor microarray experiments, we have successfully established the E-optimality of certain mixed designs, and have found certain A-optimal or near A-optimal designs from the class of mixed designs when $b = \nu$ and $k = 2$ using computer algorithms. The search for optimal design has been extended to the framework of row-column design, where two blocking factors (dyes and arrays) are taken into account simultaneously. For sample classification, we have found that the performance of the replicated reference design and interwoven designs depends on the sample variations and the gene-specific variations. Statistical optimality criteria are useful for the design of experiments to detect gene differential expression, while design of experiments for sample classification should be based on other criteria.

For application of mixed model methodology to gene expression data, we have analytically established the equivalence between the one-step and two-step analyses under fixed models, and found that some mixed effects models with specific variance-covariance structures can be split into two steps. In general, a two-step mixed model analysis may not be suitable for analyzing gene expression data. Some statistical software (e.g., ASREML) can be used to carry out a full mixed model analysis.

5.2 Future Research

For design of microarray experiments, we have demonstrated that the choice of designs not only depends on statistical optimality criteria, but also on the goals of the experiments. Future work in this area includes:

- (i) Design of single factor microarray experiments with more varieties. In our work, the A-optimality of the mixed designs has been established with up to 26 varieties, and the A-optimality of interwoven designs under row-column design has been established with up to 20 varieties. However, with the current advances in microarray technology, larger experiments with tens or hundreds of varieties (or samples) are being applied in practice, optimality of these designs is needed in order to provide a solid foundation for statistical inferences. More computationally efficient algorithms are yet to be developed to search for optimal designs within this context.
- (ii) Design of multifactorial microarray experiments. For these experiments, two or more factors (e.g., cell lines, sexes, different time points) are often involved, and the contrasts of interest are likely to depend on the goals of specific experiments. Design of these experiments is quite challenging, because statistical theory for design of these complicated experiments has not yet been well developed.
- (iii) Design of control-test microarray experiments. Two set of varieties are used for these experiments, control and test varieties, while only the contrasts between control and test varieties are of interest but not contrasts among varieties within control or within test varieties. Some work has been done within the framework of row-column design (e.g., Majumdar and Tamhane, 1996), but applications of these designs to microarray experiments have not been fully studied.
- (iv) Optimality of designs under mixed effects models. In our study of the optimality of certain designs, the optimality of these designs was established under fixed effects models, where blocking factor arrays were treated as fixed. Results from Chapter 2 have shown that efficiencies of certain designs under mixed effects models are different from those under fixed effects models. Mixed effects models are preferred over fixed effects models, because inter-block information can be recovered, hence improving the efficiency of estimating contrasts of interest. Very little work has been done regarding the optimality of block designs with block size two, although the E-optimality of certain designs under mixed effects models has been developed (e.g., Bagchi, 1987; Jacrox, 1989). More research is needed in this area to establish the optimality of designs under mixed effects models, especially for designs with block of size two.
- (v) Comparison of designs in terms of other goals of microarray experiment: clustering, sample classification, etc. In Chapter 2, our comparison of classification performance among different designs has demonstrated that design of microarray experiments also depends on the goals of these experiments. Statistical efficiency criteria are useful for designing microarray experiments for identification of differential gene expression, but criteria for design of microarray experiments with goals other than that need to be further determined.

For applications of mixed models to identifying gene differential expression, future research includes:

- (i) The equivalence between the one- and two-step analyses under unknown variance components. In Chapter 4, the equivalence of certain mixed effects models under the one- and two-step analyses was established under the assumption of known variance components. Although many methods can be used to estimate unknown variance components, the equivalence of these analyses under unknown variance components is yet to be studied. Consequently, computationally more efficient algorithms can be developed for the mixed model analysis of gene expression data.
- (ii) The equivalence between one- and two-step mixed model analyses under the assumption of heterogeneous error variance structure. As pointed out by Dudoit *et al.* (2002), the assumption of a common error variance under ANOVA methods is unrealistic, because different genes might show different variations across different samples and arrays. In our case studies, we have grouped genes based on the similarity in variation of expression values, which seems to work well. However, more work is needed to study the equivalence of mixed effects model under one- and two-step analyses where heterogeneous error variance structure is assumed.
- (iii) Approximation to the mixed model approach in terms of detecting differential gene expression. When a two-step analysis is not suitable, a full mixed effects model has to be used for the analysis, which is computationally prohibitive. Alternatively, approximation methods can be used as an alternative to the full model analysis, such as the modified two-step analysis and the method by Oleksiak *et al.* (2002). The modified two-step analysis seemed to work well for the single factor microarray experiments under certain conditions, but performed worse in our cases studies. The properties of these methods and conditions under which these methods can approximate the full model analysis are yet to be studied.

Bibliography

- Alizadeh, A.A. and Eisen, M.B. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503-511.
- Baldi, P. and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17**, 509-519.
- Bagchi, S. (1987). On the E-optimality of certain asymmetrical designs under mixed effects model. *Metrika* **34**, 95-105.
- Bagchi, S., and Cheng, C.S. (1993). Some optimal designs of block size two. *Journal of Statistical Planning and Inference* **37**, 245-253.
- Brem, R.B., Yvert, G., Clinton, R. and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752-755.
- Cheng, C.S. (1979). Optimal incomplete block design with four varieties. *Sankhya, Series B* **41**, 1-14.
- Cheng, C.S. (1980). On the E-optimality of some block designs. *Journal of the Royal Statistical Society B* **41**, 199-204.
- Chen, Y., Dougherty, E. and Bittner M. (1997). Ratio-based decisions and the quantitative analysis of cDNA microarray images. *Journal of Biomedical Optics* **2**, 364-374.
- Churchill, G.A. and Oliver, B. (2001). Sex, flies and microarrays. *Nature Genetics* **29**, 355-356.
- Claverie, J. (1999). Computational methods for the identification of differential and coordinated gene expression. *Human Molecular Genetics* **8**, 1821-1832.
- Constantine, G.M. (1981). Some E-optimal block designs. *Annals of Statistics* **9**, 886-892.
- Constantine, G.M. (1982). On the optimality of block designs. *Annals of Institute of Statistical Mathematics* **38**, 161-174.
- Cui, X., Kerr, M.K. and Churchill, G. (2002). Data transformations for cDNA microarray data. (<http://www.jax.org/research/churchill/pubs/index.html>)
- Das, A. (1993). E-optimal block and row-column designs with unequal numbers of replicates. *Sankhya B* **55**, 77-90.

- Derisi, J.L., Iyer, V.R., and Brown, P.O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686.
- Dobbin, K. and Simon, R. (2002). Comparison of microarray designs for class comparison and class discovery. *Bioinformatics* **18**, 1438-1445.
- Dudoit, S., Fridlyand, J. and Speed, T.P. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association* **97**, 77-87.
- Dudoit, S., Yang, Y.H., Speed, T.P. and Callow, M.J. (2002) Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* **12**, 111-139.
- Eccleston, J.A. (1980). Recursive techniques for the construction of experimental designs. *Journal of Statistical Planning and Inference* **4**, 291-297.
- Efron, B., Tibshirani R., Storey J.D., and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. Technical Report, Division of Biostatistics, Stanford University. To appear in *Journal of American Statistical Association*.
- Ehrenfeld, S. (1955). On the efficiency of experimental designs. *Annals of Mathematical Statistics* **26**, 247-255.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences* **25**, 14863-14868.
- Furino, S., Miao, Y. and J. Yin (1996). Frames and resolvable designs: uses, constructions and existence. CRC Press, Inc.
- Getz, G., Levine, E. and Domany, E. (2000). Coupled two-way clustering analysis of gene microarray data. *Proceedings of the National Academy of Sciences* **97**, 12079-12084.
- Gibbons, R.D., Cox, D.R., Grayson, G.R., Bhaumik, D.K., Davis, J.M., and Sharma, R.P. (2002). Sequential prediction bounds for identifying differentially expressed genes in replicated microarray experiments. Technical report, Department of Biostatistics, University of Illinois at Chicago.
- Gilmour, A.R., Thompson, R. and Cullis, B.R. (1995). Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* **51**, 1440-1450.
- Gilmour, A.R., Cullis, B.R., Welham, S.J. and Thompson, R. (2002). ASREML reference manual. Printed by NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800, Australia.
- Glonek, G.F.V and Solomon, P.J. (2002). Factorial designs for microarray experiments. <http://mag.maths.adelaide.edu.au/documents.html>.

- Henderson, D.A., Mendes, P., de la Fuente, A. and Hoeschele, I. (2003) Factor analysis for the identification of metabolic pathways from microarray expression data. *Journal of Computational Biology* (submitted).
- Hinkelmann, K., and Kempthorne, O. (1994). *Design and analysis of experiments (Vol.1)*. John Wiley and Sons, New York.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65-70.
- Jacroux, M. (1980a). On the determination and construction of E-optimal block designs with unequal number of replicates. *Biometrika* **67**, 661-667.
- Jacroux, M. (1980b). On the E-optimality of regular graph designs. *Journal of the Royal Statistical Society B* **42**, 205-209.
- Jacroux, M. (1982). Some E-optimal designs for the One-Way and Two-Way elimination of heterogeneity. *Journal of the Royal Statistical Society B* **44**, 253-261.
- Jacroux, M. (1985). Some sufficient conditions for the Type 1 optimality of block designs. *Journal of statistical Planning and Inference* **11**, 385-398.
- Jacroux, M. (1989). On the E-optimality of block designs under the assumption of random block effects. *Sankhya B* **51**, 1-12.
- Jacroux, M. (1989). Some sufficient conditions for Type 1 optimality with applications to regular graph designs. *Journal of Statistical Planning and Inference* **23**, 193-215.
- Jacroux, M. (1991a). Some new methods for establishing the optimality of block designs having unequally replicated treatments. *Statistics* **22**, 33-48.
- Jacroux, M. (1991b). A note on the E-Optimality of block design. *Metrika* **38**, 203-214.
- Jacroux, M. and Ray, R.S. (1991). On the determination and construction of optimal row-column designs having unequal row and column sizes. *Annals of the Institute of Statistical Mathematics* **43**, 377-390.
- Jansen, R.C. and Nap, J.P. (2001). Genetical genomics: The added value from segregation. *Trends in Genetics* **17**, 388-391.
- Jin, W., Riley, R.M., Wolfinger, R.D., White, K.P., Passador-Gurgel, G. and Gibson, G. (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics* **29**, 389-395.
- John, J.A. and Mitchell, T.J. (1977). Optimal incomplete block designs. *Journal of Royal Statistical Society B* **39**, 39-43.
- Kerr, M. and Churchill, G. (2001). Experimental design for gene expression microarrays. *Biostatistics* **2**, 183-201.
- Kerr, M., Afshari, C.A., Bennett, L., Bushel, P., Martinez, J., Walker, N.J. and Churchill, G.A. (2002). Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica* **12**, 203-217.

- Kerr, M., Martin, M. and Churchill, G. (2000). Analysis of variance for gene expression microarray data. *Journal of Computational Biology* **7**, 819-837.
- Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models*. (J.N. Srivastava, ed.). Amsterdam: North-Holland Publishing Co., 333-353.
- Lander, E.S. (1999). Array of hope. *Nature Genetics Supplement* **21**, 3-4.
- Lee, M. T., Kuo, F.C., Whitmore, G.A. and Sklar, J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proceedings of National Academy of Science* **97**, 9834-9839.
- Lin, Y., Nadler, S.T., Attie, A.D. and Yandell, B.S. (2001) Mining for low-abundance transcripts in microarray data. Technical Report No. 1031, Department of Statistics, UW-Madison.
- Littell, R.C., Milliken, G.A., Stroup, W.W. and Wolfinger, R.D. (1996) *SAS System for Mixed Models*, Cary, NC: SAS Institute Inc..
- Majumdar, D. and Tamhane, A.C. (1996). Row-column designs for comparing treatments with a control. *Journal of statistical planning and Inference* **49**, 387-400.
- Medvedovic, M. (2001). Determining the number of replicates needed to detect differentially expressed genes in DNA array experiments. Technical report. Department of Environmental Health, University of Cincinnati Medical Center, Cincinnati, OH 25467.
- Morgan, J.P. and Srivastava, S.K. (2000). On the Type-1 optimality of nearly balanced incomplete block designs with small concurrence range. *Statistica Sinica* **10**, 1091-1116.
- Nguyen, N-Y (1994). Construction of optimal block designs by computer. *Technometrics* **36**, 300-307.
- Nguyen, N-Y (1996). Construction of optimal row-column designs with computer. *Computing Science and Statistics. Graph-Image-Vision. Proceedings of the 28th Symposium on the Interface*, 471-475.
- Nguyen, N-Y and Williams, E.R. (1993). An algorithm for constructing optimal resolvable row-column designs. *Australian Journal of Statistics* **35**, 363-370.
- Nguyen, N-K (1994). Construction of optimal block designs by computer. *Technometrics* **36**, 300-307.
- Oleksiak, M.F., Churchill, G.A. and Crawford, D.L. (2002). Variation in gene expression within and among natural populations. *Nature Genetics* **32**, 261-266.
- Pan, W. (2002). A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments. *Bioinformatics* **18**, 536-545.
- Patterson, L.J. (1983). Circuits and efficiency factors in incomplete block designs. *Biometrika* **70**, 215-226.

- Pe'er, D., Regev, A., Elidan, G. and Friedman, N. (2001). Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **S1**(1), 1-9.
- Pritchard, C.C., Hsu, L., Delrow, J. and Nelson, P.S. (2001). Project normal: defining normal variance in mouse gene expression. *Proceedings of National Academy of Science* **98**, 13266-13271.
- Quackenbush, J. (2001). Computational analysis of microarray data. *Nature Reviews* **2**, 418-428.
- Sapir, M. and Churchill, G. (2000). Estimating the posterior probability of differential gene expression from microarray data. Technical Report, The Jackson Laboratory.
- Sathe, Y.S. and Bapat, R.B. (1985). On the E-optimality of truncated BIBD. *Calcutta Statistical Association Bulletin* **34**, 113-117.
- Schuchhardt, J., Beule, D., Malik, A., Wolski, E., Eickhoff, H., Lehrach, H. and Herzog, H. (2000). Normalization strategies for cDNA microarrays. *Nucleic Acids Research* **28** (10): e47.
- Searle, S.R. (1971). *Linear models*. John Wiley & Sons, Inc., New York.
- Searle, S.R., Casella, G. and McCulloch, C.E. (1993). *Variance components*. John Wiley & Sons, Inc., New York.
- Shah, K.R. and Sinha, B.K. (1989). Theory of optimal designs. *Lecture notes in statistics*. Springer-Verlag.
- Smyth, G.K., Yang, Y.-H., Speed, T.P. (2002). Statistical issues in microarray data analysis. In: *Functional Genomics: Methods and Protocols*, M.J. Brownstein and A.B. Khodursky (eds.), Methods in Molecular Biology series, Humana Press, Totowa, NJ, 111-136.
- Srivastav, S.K. and Morgan, J.P. (1998). Optimality of designs with generalized group divisible structure. *Journal of Statistical Planning and Inference* **71**, 313-330.
- Theilhaber, J., Bushnell, S., Jackson, A., Fuchs, R. (2001). Bayesian estimation of fold-changes in the analysis of gene expression: the PFOLD algorithm. *Journal of Computational Biology* **8**(6), 585-614.
- Tjur, T. (1993). An algorithm for optimization of block designs. *Journal of Statistical Planning and Inference* **36**, 277-282.
- Veer, L.J. (2002). Gene expression profiling predict clinical outcome of breast cancer. *Nature* **415**, 530-536.
- Venables, W.N. and Eccleston, J.A. (1993). Randomized search strategies for finding near-optimal block and row-column designs. *Australian Journal of Statistics* **35**, 371-382.
- Westfall, P.H. and Young, S.S. (1993). *Resampling-based multiple testing*. John Wiley & Sons, Inc.

- Wildsmith, S.E., Archer, G.E., Winkley, A.J., Lane, P.W. and Bugelski, P.J. (2000). Maximization of signal derived from cDNA microarrays. *BioTechniques* **30**, 202-208.
- Winer, B.J. (1962). *Statistical principles in experimental design*. McGraw-Hill, New York.
- Wolfinger, R., Gibson, G., Wolfinger, E., Bennett, L., Hamadeh, H., Bushel, P., Afshari, C. and Paules, S. (2001). Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology* **8**, 625-637.
- Wu, H., Kerr, M.K. and Churchill, G.A. (2003). The analysis of gene expression data: methods and software. chapter MAANOVA: A Software Package for the Analysis of Spotted cDNA Microarray Experiments. Springer
- Yang, X., Ye, K. and Hoeschele, I. (2001). Identifying differentially expressed genes in cDNA microarray experiments: an application of Bayesian methods using non-informative priors. Paper presented at Sixth Workshop of Case Studies in Bayesian Statistics at Carnegie Mellon University.
- Yang, X., I. Hoeschele and K. Ye (2003). Mixed effects models for microarray gene expression data. (Manuscript), Department of Statistics, Virginia Tech.
- Yang, X., Ye, K. and Hoeschele, I. (2003a). Mixed effects models for microarray gene expression data. (Manuscript), Department of Statistics, Virginia Tech.
- Yang, X., Ye, K. and I. Hoeschele (2003b). Supplement to design optimality of single factor microarray experiments. Technical report, Department of Statistics, Virginia Tech.
- Yang, Y.H., M. Buckley, S. and T. Speed (2000). Comparison of methods for image analysis on cDNA microarray data.
(<http://www.stat.Berkeley.EDU/users/terry/zarray/Html/image.html>.)
- Yang, Y.H., T. Speed (2002). Design issues for cDNA microarray experiments. *Nature Reviews* **3**, 579-588.

Vita

Xiao Yang is the second son of Xi'ai Yang and Tangling Yuan, and was born on August 8, 1972 in Shandong Province, P.R. China. He earned a Bachelor of Science degree and a Master of Science degree in Economics from People's University of China in Beijing in July, 1994 and July, 1997, respectively. He continued his studies in the Statistics Department at Virginia Polytechnic Institute and State University where he received a Master of Science degree in Statistics in December of 2000. Xiao Yang is expected to complete his Ph.D. degree in Statistics in March, 2003.