2014

Contaminated Chi-square Modeling and Its Application in Microarray Data Analysis

Feng Zhou
University of Kentucky, fzh223@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Feng Zhou, Student

Dr. Richard Charnigo, Major Professor

Dr. Constance Wood, Director of Graduate Studies

2014

Contaminated Chi-square Modeling and Its Application in Microarray Data Analysis

Feng Zhou
University of Kentucky, fzh223@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies. I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Feng Zhou, Student

Dr. Richard Charnigo, Major Professor

Dr. Constance Wood, Director of Graduate Studies

Contaminated Chi-square Modeling and Its Application in Microarray Data Analysis

DISSERTATION

A dissertation submitted in partial
fulfillment of the requirements for
the degree of Doctor of Philosophy
in the College of Arts and Sciences
at the University of Kentucky

By
Feng Zhou
Lexington, Kentucky

Director: Dr. Richard Charnigo, Professor of Mathematics
Lexington, Kentucky 2014

ABSTRACT OF DISSERTATION

Contaminated Chi-square Modeling and Its Application in Microarray Data Analysis

Mixture modeling has numerous applications. One particular interest is microarray data analysis. My dissertation research is focused on the Contaminated Chi-Square (CCS) Modeling and its application in microarray. A moment-based method and two likelihood-based methods including Modified Likelihood Ratio Test (MLRT) and Expectation-Maximization (EM) Test are developed for testing the omnibus null hypothesis of no contamination of a central chi-square distribution by a non-central Chi-Square distribution. When the omnibus null hypothesis is rejected, we further developed the moment-based test and the EM test for testing an extra component to the Contaminated Chi-Square (CCS+EC) Model. The moment-based approach is easy and there is no need for re-sampling or random field theory to obtain critical values. When the statistical models are complicated such as large mixtures of dimensional distributions, MLRT and EM test may have better power than moment based approaches, and the MLRT and EM tests developed herein enjoy an elegant asymptotic theory.

KEYWORDS: Contaminated Chi-Square Model, EM Algorithm, MLRT, Microarray, Mixture Model, Moment-Based Method

Author's signature: Feng Zhou

Date: November 6, 2014

Contaminated Chi-square Modeling and Its Application in Microarray Data Analysis

By
Feng Zhou

Director of Dissertation: Richard Charnigo

Director of Graduate Studies: Constance Wood

Date: November 6, 2014

RULES FOR THE USE OF DISSERTATIONS

Unpublished dissertations submitted for the Doctor's degree and deposited in the University of Kentucky Library are as a rule open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but quotations or summaries of parts may be published only with the permission of the author, and with the usual scholarly acknowledgments.

Extensive copying or publication of the dissertation in whole or in part also requires the consent of the Dean of the Graduate School of the University of Kentucky.

A library that borrows this dissertation for use by its patrons is expected to secure the signature of each user.

Name	Date

DISSERTATION

Feng Zhou

The Graduate School
University of Kentucky
2014

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 A Review of Mixture Models and Their Applications

1.1 Microarray Data Analysis

Mixture modeling has numerous applications. One of particular interests to us is microarray data analysis. Indeed, since our review of mixture modeling in this chapter will prominently feature papers that have applied mixture modeling to microarray data analysis, and since we will interpret mixture model parameters in that context, we begin this chapter with a brief overview of microarray data analysis.

Biologists are becoming more and more interested in studying the genome-wide patterns of gene expression because these patterns can reveal the functional importance of correlations between gene expression and the development of a phenotype (Gibson 2002 [24]). DNA (deoxyribonucleic acid) microarrays have emerged as powerful tools to allow biologists to study the genome-wide patterns of gene expression across many conditions. Basically, there are two types of microarrays, cDNA (Complementary DNA) probes and oligonucleotide. For the cDNA arrays, cDNAs are spotted onto glass slides. The mRNA (Messenger RNA) samples labeled with one of the red or green fluorescent dyes are hybridized to the microarray. Then the expression levels can be derived for each spot on the microarray by analyzing the fluorescence signal intensity. The other class, oligonucleotide arrays, is thought to be more specific in the measurement by correcting for estimates of noise. The major difference of oligonucleotide arrays from cDNA arrays lies in how the genes are probed. Instead of the full length cDNAs, oligos are spotted onto the chips. Since

oligos are usually shorter, the high density of the chips makes them more sensitive to weakly expressed transcripts.

The DNA microarray technology makes it possible to view the expression of thousands of genes from an experimental sample simultaneously. Accompanying its unprecedented capability of measurement in the areas of medicine, genetics, molecular biology, and physiology, many statistical questions still remain unresolved about the appropriate analysis of microarray data. The main purposes of DNA microarray analysis may include but are not limited to answering the questions like whether the observed differences in expression are statistically significant or not, what proportion of genes are differentially expressed, what biological or physiological relationships there may be among differentially expressed genes? These generate a large-scale hypothesis testing problem corresponding to thousands of genes. Hence, there comes the first statistical question: how to decrease the genome-wise false negative rate (Type II error rate), while the genome-wise false positive rate (Type I error rate) is under control? Also, due to the fact that microarray data is always of high dimension, low replication and large variation, how to accommodate these properties into the statistical inference forms another important research question. There are many other statistical problems involved that are worth working on. For example, how to model differential expression, how to interpret parameters in such a model, and how to obtain point and interval estimates of such parameters?

Of course, the application of mixture modeling is not just limited to microarray data analysis. In the field of health sciences, Wilcox and Russell (1983) [43] proposed a contaminated normal model for birthweight, where a predominant normal distribution accounts for most birthweights while a contaminating residual distribution

yields most very low birthweights and extremely low birthweights. In the field of finance, hierarchical Markov normal mixture models can be used to discribe stock returns, dollar-pound returns and bond returns (Geweke and Amisano (2007) [23]). More introduction of mixture models will be presented in the next section.

1.2 Mixture Models

As mentioned in the first section, mixture models are widely used in many applications especially in biology, psychology and genetics (see, e.g., Titterington et al., (1985) [40], Lindsay, (1995) [32], McLachlan and Peel (2000) [36], Geweke and Amisano (2007) [23], Charnigo and Chesnut, LoBianco, and Kirby (2010) [10]). They are often used to determine whether data come from homogeneous or heterogeneous population. If there is heterogeneity, mixture modeling may be useful for describing the nature of that heterogeneity. For the application of mixture models in DNA microarray data, for instance, some genes may be differentially expressed while others are not. Formally a mixture model corresponds to a mixture distribution that represents the probability distribution of observations in the overall population. For finite mixture models, let $\{f(x;\phi) : \phi \in \Theta\}$ be a parametric family of probability density functions (PDFs). ϕ can be a scalar or vector. The corresponding finite mixture distributions have the form:

$$g(x;\phi_1,\ldots,\phi_p,\pi_1,\ldots,\pi_p) = \sum_{i=1}^{p} \pi_i f(x;\phi_i)$$

where x is a generic element of the support set of $X_1,\ldots,X_n$ and $\pi_i \in [0,1]$ and $\sum_{i=1}^{p} \pi_i = 1$. $\phi_1,\ldots,\phi_p,\pi_1,\ldots,\pi_p$ are often treated as unknown, although in some models (called "contamination models") ϕ_1 is treated as known while the others are

treated as unknown. Two examples are explained to see the applications of finite mixture models in DNA microarray data.

Beta contamination model for P-values

Blalock et al. (2003) [5] analyzed the hippocampal tissue of male Fischer rats to identify genes related to aging and cognition. Three groups of rats(10 old, 1o middle-aged, and 10 young) received 7-day memory training, after which hippocampal tissue was collected and analyzed on an individual microarray (one chip per rat). In total, 8,799 probe sets were scanned on each microarray chip. Let Y_{ijk} be the ith gene of kth mouse in jth group, where $i = 1 \ldots 8,799$, $j = 1 \ldots 3$, $k = 1 \ldots 10$. Assume $Y_{ijk} \sim N(\mu_{ij}, \sigma_i^2)$ independently. For each $i = 1 \ldots 8,799$, we wish to test the hypothesis:

$$H_0 : \quad \mu_{i1} = \mu_{i2} = \mu_{i3}$$

ANOVA is applied to all of the 8,799 hypothesis testing problems. A P-value can be computed from each hypothesis test.

A mixture of $Beta$ distributions can be used to model the P-values in such large-scale hypothesis testing. The P-values for genes without expression alterations are viewed as independently and identically distributed according to $Beta(1, 1)$, while the P-values for differentially expressed genes are treated as a sample from another $Beta$ distribution. In the hippocampal aging example, let $P_1, \ldots, P_{8799}$ be the P-values from 8,799 hypothesis tests. For $i = 1, \ldots, 8,799$, let T_i be the latent variable indicating the group membership of the ith gene. That is,

$$T_i = \begin{cases} 1 \text{ if the } i\text{th gene is differentially expressed} \\ 0 \text{ otherwise} \end{cases}$$

Conditioning on T_i, then

$$P_i|(T_i = 0) \sim Beta(1,1) \text{ and } P_i|(T_i = 1) \sim Beta(\alpha, \beta)$$

where $\alpha > 0$ and $\beta > 0$. Let $\pi \in [0,1]$ be the proportion of genes in the batch that are differentially expressed, and $1 - \pi$ be the proportion of genes without expression alterations. Hence the marginal distribution of P_i is a mixture $Beta$ distribution with the form:

$$(1-\pi)Beta(1,1) + \pi Beta(\alpha, \beta)$$

where $\pi \in [0,1], \alpha > 0$ and $\beta > 0$. Technically, the $Beta$ contamination model allows $Beta(\alpha, \beta) = Beta(1,1)$, but in practice we anticipate $\alpha \in (0,1)$ and $\beta \in (1, \infty)$. Since the $Beta$ contamination model technically allows $Beta(\alpha, \beta) = Beta(1,1)$, the omnibus null hypothesis is not as simple as $\pi = 0$ but rather has the form $\pi(\alpha - 1) = \pi(\beta - 1) = 0$.

Normal contamination model for Z statistics

The second example is about real microarray data for 10 SARS patients and 4 healthy controls. The data are available at the Gene Expression Omnibus of the National Center for Biotechnology Information. (see Dai and Charnigo (2010) [18]). For each of the 8,793 genes, one can compute the Z statistic Z_i or T statistic T_i for $i = 1 \ldots 8,793$ such that Z_i has a standard normal distribution or T_i has a T distribution on ν degrees of freedom for some $\nu > 0$ under the null hypothesis of no differential expression. Let $\Phi(\cdot)$ denote the cumulative distribution function for a standard normal random variable, and $\Psi(\cdot)$ denote the cumulative distribution function for a T random variable on ν degrees of freedom. Thus $\Psi(T_i)$ has a uniform distribution on $[0,1]$, and then $\Phi^{-1}(\Psi(T_i))$ has a standard normal distribution under

the null hypothesis. In analogy to the *Beta* contamination model, let $\pi \in [0,1]$ be the proportion of genes in the batch that are differentially expressed, and $1-\pi$ be the proportion of genes that are not differentially expressed. Also the Z statistics for the genes that are without expression alteration are $N(0,\sigma^2)$ for some $\sigma^2 > 0$, while the Z statistics for genes that are differentially expressed are $N(\mu,\sigma^2)$, where $\mu \in \mathbb{R}$. The normal contamination model technically allows $\mu = 0$, but in practice we anticipate $\mu \neq 0$. Each Z_i has a contaminated normal distribution:

$$(1-\pi)N(0,\sigma^2) + \pi N(\mu,\sigma^2)$$

σ^2 is a nuisance parameter common to both components of the normal contamination model and σ^2 may be treated as either known or unknown. Moreover, whether σ^2 is treated as known or unknown has implications for testing the omnibus null hypothesis. Since the normal contamination model technically allows $\mu = 0$, the omnibus null hypothesis is not as simple as $\pi = 0$ but rather has the form $\pi\mu = 0$.

1.3 Likelihood Ratio Test and Modified Likelihood Ratio Test

The ordinary likelihood ratio test (LRT) is widely used in parametric hypothesis testing problems. Under standard regularity conditions, it has a simple and elegant asymptotic χ^2 distribution under the null hypothesis (Wilks, 1938) [44]. But when it comes to the mixture model problems, most of the asymptotic results of LRT can't be applied. The three reasons that might cause complications of the asymptotic null distribution of the ordinary LRT, as Lemdani and Pons (1999) [28], Lindsay (1989) [31], and Zhu and Zhang (2004) [45] mentioned, are:

- The parameters are not identifiable when H_0 is true.

- Possible values for the parameters when H_0 is true are on the boundary of the parameter space.

- The Fisher information matrix may be singular.

A lot of research has been done to study the asymptotic properties of LRT for mixture models. Except where otherwise indicated all theoretical results mentioned in Chapter 1 regard tests of homogeneity (i.e., one component versus two) under the null hypothesis. Hartigan (1985) [26], Bickel and Chernoff (1993) [3] mentioned that the LRT goes to infinity with probability 1 if Θ is unbounded even under the normal kernel $N(\theta, 1)$. Chernoff and Lander (1995) [13] suggested an approach based on Kullback Leibler information when the Fisher information matrix is singular. Lemdani and Pons (1999) [28], Chen and Chen (2001) [15] proved that the limiting distribution of the likelihood ratio statistic is the squared supremum of a truncated standard Gaussian process. The reason why the result of Lemdani and Pons (1999) [28], Chen and Chen (2001) [15] contradicts with that of Hartigan (1985) [26] is that Lemdani and Pons, Chen and Chen took the parameter space Θ to be compact while Hartigan did not.

Lindsay (1989) [31] proposed the method of moments to solve the problem of determining an unknown mixing distribution. The bootstrap method was first derived by McLachlan (1987) [35] to assess the null distribution of the LRT for a single normal density versus a mixture of two normal densities in the univariate case. The asymptotic distributions of maximum likelihood estimators and likelihood ratio statistics were derived by Self and Liang (1987) [38] when the true parameter value may be on the boundary of the parameter space. Although the large sample behavior of

the likelihood ratio statistic for a mixture model has been studied extensively, it is still not easy to implement mainly because the determination of the critical value involves the supremum of a Gaussian process.

Chen et al. (2001) [14] proposed a modified likelihood ratio test (MLRT) for homogeneity by adding a penalty term $C\log(4\pi(1-\pi))$, where C is a positive constant, to the ordinary log likelihood function. The maximizers for the modified log likelihood function are called maximum modified likelihood estimators (MMLEs). The MMLEs are consistent under the null hypothesis $f(x;\phi_0)$, which means $\widehat{\phi_1}$ and $\widehat{\phi_2}$ converge in probability to ϕ_0. For $0<\pi<1, \phi_1, \phi_2 \in \Theta$, the modified likelihood is defined to be

$$l_n(\pi,\phi_1,\phi_2) = \sum_{i=1}^{n} \log\{(1-\pi)f(X_i;\phi_1) + \pi f(X_i;\phi_2)\} + C\log\{4\pi(1-\pi)\}.$$

The modified likelihood function is one method to address the non-identifiability problems. By adding the penalty term $C\log(4\pi(1-\pi))$, the estimator of π is bounded away from 0 and 1, so that a null hypothesis of $\pi(\phi_2-\phi_1)=0$ effectively becomes a null hypothesis of $\phi_2-\phi_1=0$.

The other method used to circumvent the non-identifiability problems is reparametrization. See Lemdani and Pons (1999) [28]. We do this reparametrization by defining $\upsilon := \pi(\phi_2-\phi_1)$. And then the null hypothesis

$$H_0: \ \pi = 0 \text{ or } \phi_1 = \phi_2$$

is changed into

$$H_0: \ \nu = \mathbf{0}$$

Then the null hypothesis under reparametrization is a point instead of a set of parameters. The *Beta* contamination model, for example, can be reparametrized by

taking $\nu = \pi((\alpha, \beta)^t - (1, 1)^t)$. As shown by Dai and Charnigo (2008) [18], one may also combine the modified likelihood and reparametrization approaches.

The modified likelihood function can be viewed from a Bayesian perspective. Remember the general relationship among a posterior distribution, a prior distribution, and a likelihood function,

$$h(\phi|X) \propto h(\phi)h(X|\phi),$$

where $X = (X_1, \ldots, X_n)^t$. Taking the logarithm on both sides, we get

$$\log h(\phi|X) = \log h(\phi) + \sum_{i=1}^{n} \log h(X_i|\phi) + g(X),$$

where $g(X)$ is a quantity not depending on ϕ. Let $\log h(\phi) = C \log(4\pi(1-\pi))$, $\sum_{i=1}^{n} \log h(X_i|\phi) = \sum_{i}^{n} \log[(1-\pi)f(X;\phi_1) + \pi f(X;\phi_2)]$, then the form of the posterior distribution coincides with the modified likelihood function. Let $(\hat{\pi}, \widehat{\phi_1}, \widehat{\phi_2})$ be the maximizer of $l_n(\pi, \phi_1, \phi_2)$ over the full parameter space, and let $\widehat{\phi_0}$ maximize $l_n(0.5, \phi_0, \phi_0)$ over the parameter space for the null hypothesis. The MLRT statistic is then defined as:

$$\lambda_n := 2l_n(\hat{\pi}, \widehat{\phi_1}, \widehat{\phi_2}) - 2l_n(0.5, \widehat{\phi_0}, \widehat{\phi_0})$$

where $\hat{\pi}, \widehat{\phi_1}, \widehat{\phi_2}$ are the MMLEs. The null hypothesis is rejected for large values of λ.

The asymptotic behavior of the MLRT has been studied for several different mixture models. For instance, let λ_N and λ_B be the test statistics for the normal contamination model and *Beta* contamination model respectively. Dai and Charnigo (2010) [19] showed that $\lambda_N \xrightarrow{L} \chi_1^2$ under the null hypothesis of no contamination and under any fixed alternative, $n^{-1}\lambda_N$ converges in probability to some positive constant. Dai and Charnigo (2008) [18] proved that λ_B has a limiting distribution of χ_2^2 under the null hypothesis. More generally, for any contamination density model,

when satisfying appropriate regularity conditions, the MLRT statistic converges in law to χ^2_k under the null hypothesis of no contamination, where $\Theta \subset \mathbb{R}^k$. They also established the asymptotic consistency of the MLRT under fixed alternatives and derived a limiting distribution under contiguous local alternatives, revealing that the MLRT is asymptotically locally unbiased.

1.4 D-test

D-test was first proposed by Charnigo and Sun (2004) [9], to test homogeneity in mixture models. It uses the Lebesgue-measure L^2 distance between a fitted homogeneous model under the null hypothesis and a fitted heteogeneous model under the alternative hypothesis to do the hypothesis testing problem. Large value of D-test statistic rejects the null hypothesis of homogeneity. The D-test statistic for a contamination model is

$$d := \int_{\mathbb{R}} w(x)([(1-\hat{\pi})f(x;\phi_0) + \hat{\pi}f(x;\hat{\phi})] - f(x;\phi_0))^2 dx$$

where $\hat{\phi}$ and $\hat{\pi}$ are estimators under the alternative hypothesis, ϕ_0 is the known parameter defining the homogeneous distribution under the null hypothesis. MMLE's are recommended for these estimators because that gives tractable asymptotic null distributions. $w(x)$ is positive, which is a weighting function. For $w(x) = 1$, it is an ordinary D-test statistic. For other $w(x)$, it is a weighted D-test statistic. The choice of the weighting function may have a big effect on the performance in many situations. A weighting function carefully chosen based on the anticipated shape of the contaminating density can enhance the power of the D-test while keeping the actual Type I error probability close to the nominal Type I error probability.

Sometimes the original data including some private information may not be released to the public. One advantage of the D-test statistic is that it does not depend on the whole data set directly. In particular, the D-test is purely a function of the parameter estimators, and often this function can be expressed in closed form (i.e., the integral defining the D-test statistic can be evaluated analytically).

Taking the *Beta* contamination model and Normal contamination model as examples, the D-test statistic for the *Beta* contamination model is

$$d := \int_0^1 w(x)([(1-\hat{\pi})f(x;\alpha_0,\beta_0) + \hat{\pi}f(x;\hat{\alpha},\hat{\beta})] - f(x;\alpha_0,\beta_0))^2 dx$$

Assuming that $w(x)$ has the form $x^{c_1}(1-x)^{c_2}$, for some constants c_1 and c_2, this D-test statistic can be expressed in closed form

$$d := \sum_{i=0}^{2}\sum_{j=0}^{2} \hat{\pi}_i\hat{\pi}_j \frac{B(c_1+\hat{\alpha}_i+\hat{\alpha}_j-1, c_2+\hat{\beta}_i+\hat{\beta}_j-1)}{B(\hat{\alpha}_i,\hat{\beta}_i)B(\hat{\alpha}_j,\hat{\beta}_j)}$$

where $\hat{\pi}_0 := -1, \hat{\pi}_1 := 1-\hat{\pi}, \hat{\pi}_2 := \hat{\pi}, \hat{\alpha}_0 = \hat{\alpha}_1 := \alpha_0, \hat{\alpha}_2 := \hat{\alpha}, \hat{\beta}_0 = \hat{\beta}_1 := \beta_0, \hat{\beta}_2 := \hat{\beta}$, and

$$B(\gamma,\eta) := \int_0^1 x^{\gamma-1}(1-x)^{\eta-1}dx$$

The D-test statistic for the Normal contamination model is

$$d := \int_{\mathbb{R}} w(x)([(1-\hat{\pi})f(x;\mu_0,\hat{\sigma}^2) + \hat{\pi}f(x;\hat{\mu},\hat{\sigma}^2)] - f(x;\mu_0,\hat{\sigma}_0^2))^2 dx$$

Assuming that $w(x) = \exp[c(x-\mu_0)^2]$, for some constant c, this D-test statistic can also be expressed in closed form

$$d := \sum_{i=0}^{2}\sum_{j=0}^{2} \frac{\pi_1\pi_2}{\sqrt{2\pi}\sqrt{\hat{\sigma}_i^2+\hat{\sigma}_j^2-2c\hat{\sigma}_i^2\hat{\sigma}_j^2}} \exp[-\frac{1}{2}\frac{(\hat{\mu}_i-\hat{\mu}_j)^2 - 2c\hat{\sigma}_i^2(\hat{\mu}_j-\mu_0)^2 - 2c\hat{\sigma}_j^2(\hat{\mu}_i-\mu_0)^2}{\hat{\sigma}_i^2+\hat{\sigma}_j^2-2c\hat{\sigma}_i^2\hat{\sigma}_j^2}]$$

with $\hat{\pi}_0 = -1, \hat{\pi}_1 = 1 - \hat{\pi}, \hat{\pi}_2 = \hat{\pi}, \hat{\mu}_0 = \hat{\mu}_1 = \mu_0, \hat{\mu}_2 = \hat{\mu}$, and $\hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \hat{\sigma}^2$.

The asymptotic behavior of the D-test statistic was also analyzed. Charnigo and Sun (2010) [11] provided asymptotic results for the D-test statistic when ϕ_1 was unknown, including the case of a normal kernel with unknown nuisance parameter σ^2 common to all components. Dai and Charnigo (2008) [18] provided asymptotic results when ϕ_1 was known, while Dai and Charnigo (2010) [19] covered the case of a normal kernel with unknown nuisance parameter σ^2 common to all components. To be more specific, Dai and Charnigo (2008) [18] established when $w(x)$ is defined to be $1/f(x;\phi_0)$, the weighted D-test statistic multiplied by the sample size nd converges in law to χ_k^2 under the null hypothesis of no contamination under regularity conditions, where $\Theta \subset \mathbb{R}^k$. The weighted D-test is asymptotically locally unbiased under contiguous local alternatives and consistent under fixed alternatives. Dai and Charnigo (2010) [19] proved for the Normal contamination model, when σ^2 is unknown, assuming that σ_0^2 is the true value of σ^2 and there is an open neighborhood about σ_0^2 in the parameter space, then $4\pi^{1/2} n \sigma_0 d$ converges in law to χ_1^2 under the null hypothesis.

Chapter 2 Moment-Based Inference for Contaminated Chi-Square Model

2.1 Introduction

Consider the mixture model (Titterington et al, 1985 [40]; Lindsay, 1995 [32]; McLachlan and Peel, 2000 [36]) with probability density function (pdf)

$$(1-\gamma)\chi^2_\nu(0) + \gamma\chi^2_\nu(\mu), \tag{2.1}$$

where $0 \leq \gamma \leq 1$, $\chi^2_\nu(0)$ denotes the central Chi-Square pdf on $\nu > 0$ degrees of freedom (df), and $\chi^2_\nu(\mu)$ denotes the Chi-Square pdf on ν df with non-centrality parameter $\mu \geq 0$. Recall that the pdf of $\chi^2_\nu(\mu)$ is $f_X(x;\nu,\mu) = \frac{1}{2}e^{-(x+\mu)/2}(\frac{x}{\mu})^{(\nu/4-1/2)}I_{\nu/2-1}(\sqrt{\mu x})$, where $I_\nu(z)$ is a modified Bessel function of the first kind. We assume that ν is known, while γ and μ are unknown. We refer to (2.1) as the Contaminated Chi-Square (CCS) model, since we regard $\chi^2_\nu(0)$ as being contaminated by $\chi^2_\nu(\mu)$.

In this chapter, we develop a convenient procedure for testing

$$H_0 : \gamma\mu = 0 \text{ vs. } H_1 : \gamma\mu > 0, \tag{2.2}$$

we analyze its asymptotic and finite-sample properties, and we propose estimators of these parameters in the event that H_0 is rejected. For a reason that will become apparent later, we refer to H_0 as the omnibus null hypothesis. The CCS model simplifies to $\chi^2_\nu(0)$ if and only if the omnibus null hypothesis is true.

The CCS model and the omnibus null hypothesis relate to large-scale ANOVA testing and can be used in the microarray data analysis. Consider the example in Blalock et al (2003) [5], the hippocampal tissue was collected and analyzed on an

individual microarray of three groups of rats (10 old, 10 middle-aged and 20 young). In total, 8,799 probe sets were scanned on each microarray chip. For each of 8,799 genes, a one-way ANOVA was conducted to compare expression levels across the three groups. If the F statistics from ANOVA for each gene are transformed to the Chi-Square statistics, for example, by using a probability integral transform: $X_i := cdf_{\chi_{\nu_1}}^{-1}(cdf_{F_{\nu_1,\nu_2}})$, where ν_1 and ν_2 are the numerator and denominator df for the F statistic, then we may use CCS model to test the omnibus null hypothesis.

The CCS model may be similarly applied in other scenarios involving large numbers of ANOVA tests. For instance, the CCS model may be employed to analyze data on copy number variation, transcript splicing variation, or DNA methylation (Breheny et al, 2012 [6], Vandiedonck et al, 2011 [41], Herman, 1995 [27]).

The Chi-Square statistics (obtained by transforming F statistics) associated with CCS model are introduced for two reasons. First, in the situation when the microarray data analyses compare more than two populations, the Chi-Square statistics should be applied here since ANOVA does not yield a Z score. In this case, the methodology of Dai and Charnigo (2010) [19] is inapplicable. However, the methodology proposed herein is applicable. In fact, the methodology proposed herein is still applicable when only two populations are compared, since a Z score may be converted to a Chi-Square statistic via squaring.

Second, since the omnibus null hypothesis for the beta mixture model has a two-sided alternative, the differential expression is not the only reason that may lead to the rejection of the uniform distribution from a beta mixture model for P-values. Recall that the beta mixture model studied by Dai and Charnigo (2008) [18] has the form $(1-\gamma)Beta(1,1)+\gamma Beta(\alpha,\beta)$, where $0 \leq \gamma \leq 1$ and $\alpha > 0, \beta > 0$. Another

possible reason for rejection of the uniform distribution is that there may be too many large P-values instead of too many small P-values as would be the case with 0.5 Beta(1,1) + 0.5 Beta(2,0.5). The tests of Dai and Charnigo (2008) [18] will detect an excess in either direction. Thus, the power to detect a specific alternative that is indicative of systematic differential expression may be lower than desired. But the Chi-Square statistics transform the alternative to the omnibus null hypothesis to a one-sided test, which overcomes the aforementioned limitation by rejecting the omnibus null hypothesis only when there is an excess of large Chi-Square statistics (or, equivalently, small P-values). As such, the test proposed herein may have better power to detect systematical differential expression than the tests of Dai and Charnigo (2008) [18].

2.2 Testing

Suppose that $X_1, X_2, \ldots, X_n$ are a random sample from the CCS model (2.1). Our procedure for testing the omnibus null hypothesis in (2.2) is an intersection-union test based on the method of moments. More specifically, let

$$S := n^{-1} \sum_{1 \leq k \leq n} X_k - \nu \text{ and } W := \nu^2 + 2\nu(1 - n^{-1} \sum_{1 \leq k \leq n} X_k) + n^{-1} \sum_{1 \leq k \leq n} X_k^2 - 4n^{-1} \sum_{1 \leq k \leq n} X_k. \tag{2.3}$$

Lemma 1. *S converges in probability to $\gamma\mu$ and W converges in probability to $\gamma\mu^2$.*

Proof. Since $X_1, X_2, \ldots, X_n$ are a random sample from the CCS model (2.1), we may get $E[X_1] = \nu + \gamma\mu$ and $E[X_1^2] = 2\nu + \nu^2 + 4\gamma\mu + 2\gamma\nu\mu + \gamma\mu^2$. Then $E[S] = E[n^{-1} \sum_{1 \leq k \leq n} X_k] - \nu$, which converges in probability to $\gamma\mu$ by the weak law of

large numbers. And $E[W] = \nu^2 + 2\nu(1 - E[n^{-1}\sum_{1\leq k\leq n} X_k]) + E[n^{-1}\sum_{1\leq k\leq n} X_k^2] - 4E[n^{-1}\sum_{1\leq k\leq n} X_k]$, which converges in probability to $\gamma\mu^2$ by the weak law of large numbers and slutsky's theorem. □

This motivates us to reject the omnibus null hypothesis if $S > s_{crit}$ and $W > w_{crit}$, where s_{crit} and w_{crit} are chosen to achieve the desired Type I error probability. Theorem 1 below indicates how s_{crit} and w_{crit} may be chosen. Before stating Theorem 1, we establish some notations.

Let ϕ denote the standard normal cumulative distribution function and z_c the c quantile of the same, i.e. $\phi(z_c) = c$. Let r_j denote the jth moment of $\chi^2_\nu(0)$ for $1 \leq j \leq 4$, $\mathbf{R}$ the 2 x 2 matrix whose ijth entry is $r_{i+j} - r_i r_j$, and $\mathbf{B}$ the 2 x 2 matrix whose first column is $(1, 0)$ and whose second column is $(-2\nu - 4, 1)$.

Theorem 1. *Let $0 < \delta < 1$ and $0 < \epsilon < 1$ satisfy $\delta\epsilon = \alpha$. Under the omnibus null hypothesis,*

$$lim_{n\to\infty} P[S > z_{1-\delta} n^{-1/2} a_{11}^{1/2} \text{ and } W > z_{1-\epsilon} n^{-1/2} a_{22}^{1/2}] = \alpha, \tag{2.4}$$

where a_{11} and a_{22} are the diagonal entries of the 2×2 matrix $\mathbf{A} := \mathbf{B^T R B}$.

Moreover, under any fixed alternative $(\gamma, \mu) = (c_1, c_2)$ with $0 < c_1 \leq 1$ and $c_2 > 0$,

$$lim_{n\to\infty} P[S > z_{1-\delta} n^{-1/2} a_{11}^{1/2} \text{ and } W > z_{1-\epsilon} n^{-1/2} a_{22}^{1/2}] = 1. \tag{2.5}$$

Proof. Under the omnibus null hypothesis, $\gamma\mu = \gamma\mu^2 = 0$ Since we know that $E[S] = E[n^{-1}\sum_{1\leq k\leq n} X_k] - \nu = \gamma\mu = 0$, and $E[W] = \nu^2 + 2\nu(1 - E[n^{-1}\sum_{1\leq k\leq n} X_k]) + E[n^{-1}\sum_{1\leq k\leq n} X_k^2] - 4E[n^{-1}\sum_{1\leq k\leq n} X_k] = \gamma\mu^2 = 0$, then S and W will both converge

in probability to 0. More specifically, $n^{1/2}(n^{-1}\sum_{1\leq k\leq n} X_k - nv, n^{-1}\sum_{1\leq k\leq n} X_k^2 - 2\nu - \nu^2)^T$ converges in law to the multivariate normal distribution with mean vector $(0,0)^T$ and covariance matrix $\mathbf{R}$ by the Central Limit Theorem. Then $(S,W)^T$ converges in law to the multivariate normal distribution with mean vector $(0,0)^T$ and covariance matrix $\mathbf{A}$ by Cramer's Theorem, since $\mathbf{B}$ contains the partial derivatives relating $(S,W)^T$ to $n^{1/2}(n^{-1}\sum_{1\leq k\leq n} X_k - nv, n^{-1}\sum_{1\leq k\leq n} X_k^2 - 2\nu - \nu^2)^T$. The key observation to obtain (2.4) is that the off-diagonal entries of $\mathbf{A}$ are 0, whence $P[S > z_{1-\delta}n^{-1/2}a_{11}^{1/2}$ and $W > z_{1-\epsilon}n^{-1/2}a_{22}^{1/2}]$ converges to $(1-\phi[z_{1-\delta}])(1-\phi[z_{1-\epsilon}]) = \delta\epsilon = \alpha$.

Under the fixed alternative $(\gamma,\mu) = (c_1,c_2)$, S converges in probability to $c_1c_2 > 0$ and W converges in probability to $c_1c_2^2 > 0$, so that $P[S \leq z_{1-\delta}n^{-1/2}a_{11}^{1/2}]$ and $P[W \leq z_{1-\epsilon}n^{-1/2}a_{22}^{1/2}]$ converge to 0. Since $P[S > z_{1-\delta}n^{-1/2}a_{11}^{1/2}$ and $W > z_{1-\epsilon}n^{-1/2}a_{22}^{1/2}] \geq 1 - P[S \leq z_{1-\delta}n^{-1/2}a_{11}^{1/2}] - P[W \leq z_{1-\epsilon}n^{-1/2}a_{22}^{1/2}]$, the former must converge to 1.

□

A few comments are in order. First, one may choose $\epsilon := 1$ (i.e.,choose $w_{crit} := -\infty$) and effectively base the test on only S rather than on both S and W. In this case, one may replace $z_{1-\delta}n^{-1/2}a_{11}^{1/2}$ by $n^{-1}q_{\nu n,1-\alpha} - \nu$, where $q_{\nu n,1-\alpha}$ denotes the $1-\alpha$ quantile of $\chi^2_{\nu n}(0)$. Then the Type I error probability is exactly α for all finite n, not just converging to α in the limit. However, a potential problem with this choice is that one may reject the omnibus null hypothesis when $W < 0$. Since W is a moment-based estimator of $\gamma\mu^2$, moment-based estimation of γ and μ when $W < 0$ leads to the estimator of γ and/or that of μ not belonging to the appropriate parameter space. Taking the hippocampal aging data as an example, the moment-based test only depending on S will lead to a negative estimation of either γ or μ , which is outside of the parameter space.

Second, choosing $\epsilon \leq 1/2$ and $\delta \leq 1/2$(i.e., choosing $w_{crit} \geq 0$ and $s_{crit} \geq 0$) guarantees that γ and μ may be estimated using moments when the omnibus null hypothesis is rejected. This is described in Theorem 2 and its Corollary below. More specific choices of ϵ and δ can be recommended based on power considerations. However, while S and W are asymptotically may be independent under the omnibus null hypothesis, they may be correlated when the omnibus null hypothesis is false. Thus, analytically evaluating the power in relation to ϵ and δ is difficult. However, we can gain some insights from simulation studies, which we pursue in Section 2.4.

Two advantages of the moment-based approach can be listed here. One is, in contrast with a likelihood ratio test for the number of components in a mixture model, the testing procedure of Theorem 1 does not require a compact parameter space; note that no upper bound for μ was assumed. The other one is, the critical value is known and thus need not be estimated via resampling or random field theory.

We remark that the problem in (2.2) is not, strictly speaking, determining the number of components in a mixture model. This is because, although (2.1) reduces to one components under the omnibus null hypothesis, (2.1) also reduces to one component when $\gamma = 1$ and $\mu > 0$.

2.3 Estimation

We have known that $E[S] = \gamma\mu$, $E[W] = \gamma\mu^2$. So intuitively, $\frac{S^2}{W}$ and $\frac{W}{S}$ can be used to estimate γ and μ respectively when the omnibus null hypothesis is false. We can theoretically prove that $\frac{S^2}{W}$ and $\frac{W}{S}$ are $n^{1/2}-$consistent estimators of γ and μ under the alternative hypothesis. To state Theorem 2, we introduce

some more notations. Let $m_j := E[X_1^j]$ for $1 \leq j \leq 4, \mathbf{M}$ the 2×2 matrix whose ijth entry is $m_{i+j} - m_i m_j$, and $\mathbf{D}$ the 2×2 matrix whose first column is $((m_1-\nu)(-m_2-4m_1-2\nu m_1), (m_1-\nu)^2)^T/(m_2+2\nu+\nu^2-4m_1-2\nu m_1)^2$ and whose second column is $(-m_2+2\nu+\nu^2, m_1-\nu)^T/(m_1-\nu)^2$.

Theorem 2. *Under any fixed alternative* $(\gamma, \mu) = (c_1, c_2)$ *with* $0 < c_1 \leq 1$ *and* $c_2 > 0$, $n^{1/2}(S^2/W - c_1, W/S - c_2)^T$ *converges in law to the multivariate normal distribution with mean vector* $(0,0)^T$ *and covariance matrix* $\mathbf{D^T M D}$.

Proof. By the Central Limit Theorem, $n^{1/2}(n^{-1}\sum_{1\leq k\leq n} X_k - \nu - c_1c_2, n^{-1}\sum_{1\leq k\leq n} X_k^2 - (\nu + c_1c_2)^2 - c_2^2c_1(1-c_1) - 4c_1c_2 - 2\nu)^T$ converges in law to the multivariate normal distribution with mean vector $(0,0)^T$ and covariance matrix $\mathbf{M}$. The desired result then follows from Cramer's Theorem. □

Although the probability that $S < 0$ or $W < 0$ is nonzero (in which case the estimator or γ and/or that of μ will not belong to the appropriate parameter space), with $\epsilon \leq 1/2$ and $\delta \leq 1/2$, this event is a subset of accepting the omnibus null hypothesis. Hence, if one agrees to take $\epsilon \leq 1/2$ and $\delta \leq 1/2$ as well as to estimate γ and μ only if the omnibus null hypothesis is rejected, then this event will not be encountered in practice. The following corollary, an immediate consequence of (2.5) from Theorem 1, also demonstrates that such an agreement does not disturb the conclusion of Theorem 2.

Corollary 1. *Under any fixed alternative* $(\gamma, \mu) = (c_1, c_2)$ *with* $0 < c_1 \leq 1$ *and* $c_2 > 0$, *the conditional distribution of* $n^{1/2}(S^2/W - c_1, W/S - c_2)^T$ *given that* $W > w_{crit}$

and $S > s_{crit}$ *converges in law to the multivariate normal distribution with mean vector* $(0,0)^T$ *and covariance matrix* $\mathbf{D^T M D}$.

Proof. By Bayes' Theorem, the conditional pdf of $n^{1/2}(S^2/W - c_1, W/S - c_2)^T$ given that $W > w_{crit}$ and $S > s_{crit}$ is the unconditional pdf of $n^{1/2}(S^2/W - c_1, W/S - c_2)^T$ multiplied by an indicator for these inequalities, divided by the probability that these inequalities hold. Since the denominator converges to 1 by Theorem 1, the result follows immediately from Theorem 2. □

2.4 Simulation Studies

To investigate the performance of the moment-based test, we conducted a number of simulation studies to assess the Type I error probability and the power of our testing procedure in finite samples. In Figure 2.1 and in the following text, we use this shorthand:

CCS1 The procedure for testing the omnibus null hypothesis in (2.2) is applied directly to a random sample $X_1, X_2, \ldots, X_n$ from the CCS model (2.1), with $\delta = 1/2$ and $\epsilon = 1/10$. These choices of δ and ϵ emphasize W over S for rejection of the omnibus null hypothesis, requiring only that the latter be positive.

CCS2 Proceed as above but with $\delta = \epsilon = \sqrt{0.05}$. These choices emphasize W and S equally.

CCS3 Proceed as above but with $\delta = 1/10$ and $\epsilon = 1/2$. These choices of δ and ϵ emphasize S over W for rejection of the omnibus null hypothesis, requiring

only that the latter be positive.

CB A random sample $X_1, X_2, \ldots, X_n$ from the CCS model (2.1) is transformed by the survival function of the central Chi-Square distribution on ν df to yield "p-values" $P_1, P_2, \ldots, P_n$. These are treated as if they had arisen from the Contaminated Beta (CB) model with pdf

$$(1-\gamma)1_{0<p<1} + \gamma 1_{0<p<1} p^{\alpha-1}(1-p)^{\beta-1}/B(\alpha,\beta), \tag{2.6}$$

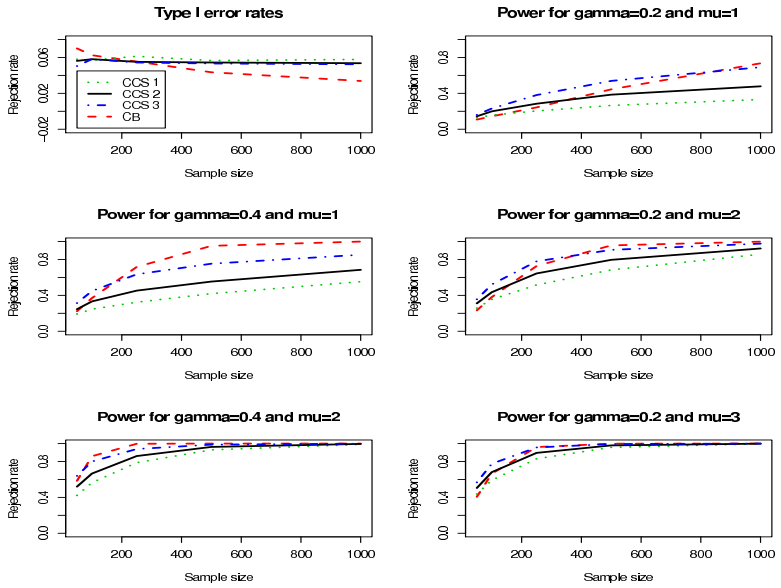
where $B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ The MLR test is applied to $P_1, P_2, \ldots, P_n$ to see whether the CB model can be reduced to a uniform distribution (Dai and Charnigo, 2008 [18]).

For each sample size n $\in \{50, 100, 250, 500, 1000\}$, we generated 10,000 random samples $X_1, X_2, \ldots, X_n$ from CCS model (2.1) with $\gamma\mu = 0$. Then we determine how many times out of 10,000 we reject H_0. These are Type I error rates, which are displayed in the top left panel of Figure 2.1. For methods CCS1, CCS2 and CCS3, the Type I error rates are between 0.0504 and 0.0613 at all n. Thus, the critical values for our testing procedure, which were based on the asymptotic result of Theorem 1, appear satisfactory for finite samples. For method CB, the calculated Type I error rates decrease from 0.0701 at n=50 to 0.0338 at n=1000, indicating that the MLR test applied to p-values is slightly anticonservative for small n.

The remaining panels of Figure 2.1 present the power curves for these five models $(\gamma,\mu) = (0.2,1), (\gamma,\mu) = (0.4,1), (\gamma,\mu) = (0.2,2), (\gamma,\mu) = (0.4,2), (\gamma,\mu) = (0.2,3)$ respectively. Power is calculated as the number of omnibus null hypothesis rejections divided by 10,000. In all scenarios, power increases with n for each method as what we anticipated.

In the panel of power for $\gamma = 0.2$ and $\mu = 1$, Method CCS3 exhibits better power than method CCS2, which in turn is more powerful than method CCS1. Method CB appears relatively strong for large n but comparatively weak for small n. The power of method CB for $\gamma = 0.4$ and $\mu = 1$ starts with weak power for small n but performs much better than other methods for large n. The powers for the three Contaminated Chi-Square methods are still in the order of CCS3, CCS2 and CCS1. Actually, when looking into the last three panels, all of these scenarios maintain the relative ordering of methods CCS3, CCS2, and CCS1. Roughly speaking, method CB fares well with larger γ, μ, and n but does not perform as well with smaller γ, μ, and n.

Figure 2.1: Type I Error Rates and Powers for CCS1, CCS2, CCS3 and CB Models



Based on the results of these simulation studies, CCS3 has better power than CCS1 and CCS2. So we recommend taking $\delta := 1/10$ and $\epsilon := 1/2$ when applying

our testing procedure. If n is large, or if γ and μ are anticipated to be large, then one may also wish to consider transforming Chi-Square statistics to P-values and then analyzing P-values using the CB model (2.6). However, referring to the problem we mentioned in section 2.1, a naive analysis of P-values may lead to an inappropriate declaration of systematic differential expression; therefore, care must be exercised in any decision to transform Chi-Square statistics to P-values.

We also note that our moment-based procedure has its own advantages of no resampling required and no compactness restriction on the parameter space. But it may be less powerful than other approaches yet to be developed. In particular, we will investigate in chapter 3 how the EM test (Chen and Li, 2009 [16]; Li, Chen, and Marriott, 2009 [29]) can be adapted to this setting.

2.5 Case Study

Dai and Charnigo (2008) [18] applied the CB model (6) to analyze the P-values generated from a microarray experiment conducted by Blalock and colleagues (2003) [5]. The hippocampal tissue was collected and analyzed on an individual microarray for three groups of rats (10 old, 10 middle-aged, and 10 young). In total, 8799 probe sets were scanned on each microarray chip. For each of 8,799 genes, a one-way ANOVA was conducted to compare expression levels across the three groups. This produced 8,799 F statistics, which in turn yielded the P-values. Blalock et al. (2003) [5] introduced a filtration procedure to reduce the number of gene probe sets and P-values by three steps:

1A Exclude all probe sets rated absent.

2A Exclude all present transcript sets representing unannotated expressed sequence tags.

3A Exclude genes for which the young and old groups did not differ by at least 75% of the maximal difference among groups.

One problem emerged when Dai and Charnigo (2008) [18] analyzed the P-values and, in particular, employed the MLR test (Chen, Chen, and Kalbfleisch, 2001 [14]) and D test (Charnigo and Sun, 2004 [10]) to see whether the CB model could be reduced to a uniform distribution. For the genes eliminated at step 3, the MLR test and D test decisively rejected the omnibus null hypothesis of a uniform distribution because there are fewer large P-values in the right tail. In this case, the departure from a uniform distribution may not indicate differential expression but rather, as suggested by Allison et al (2002) [2], correlations among the P-values corresponding to different genes. One can also see that the estimated values of parameters, $\hat{\gamma} = 0.696, \hat{\alpha} = 1.01$, and $\hat{\beta} = 1.28$, do not indicate an excess of small P-values. Thus, the alternative to the omnibus null hypothesis of a uniform distribution may be too broad if our main interest is in ascertaining differential expression.

To avoid this problem, Chi-Square statistics can be applied. Instead of analyzing P-values, we transform the F statistics to the Chi-Square statistics based on the probability integral transformation (Casella and Berger, 2002 [7]). More specifically, we first converted the F statistics to their corresponding P-values by successively applying the cumulative distribution function (cdf) of the central F distribution on 2 and 27 df. Since the P-values follow the uniform(0,1) distribution, under the om-

nibus null hypothesis the inverse cdf of the central Chi-Square distribution on 2 df leads to the Chi-Square statistics.

Figure 2.2 shows histograms of Chi-Square statistics for all 8,799 genes, for the genes eliminated in the first two steps, and for the genes remaining after each step. The fitted CCS model and the null model $\chi^2_2(0)$ are superimposed against each histogram. Also the parameter estimates are displayed in Table 2.1. From six panels, each fitted model is in much better concordance with its respective histogram than the null model, though the fitted model yields a smaller density between 0 and 2 but a larger density between 5 and 10 compared to the null model. This is most apparent in the last panel. Correspondingly, our procedure for testing the omnibus null hypothesis in (2.2) yields a P-value less than 0.0001, regardless of whether one defines the P-value by taking $\delta = 1/2, \epsilon = 2\alpha$ or $\delta = \epsilon = \alpha^{1/2}$ or $\delta = 2\alpha, \epsilon = 1/2$. Here, P-value is defined as the smallest α at which one rejects H_0. For example, consider the case in which $\delta = 1/2$ and $\epsilon = 2\alpha$, we have a series of choices of ϵ and corresponding α. For instance, we have $\epsilon_1 = 0.20, \alpha_1 = 0.10, \epsilon_2 = 0.10, \alpha_2 = 0.05, \epsilon_3 = 0.05, \alpha_3 = 0.025, \epsilon_4 = 0.02, \alpha_4 = 0.01, \epsilon_5 = 0.01, \alpha_5 = 0.005, \epsilon_6 = 0.005, \alpha_6 = 0.0025$. Suppose we reject the omnibus null hypothesis at $\epsilon_1 = 0.20, \alpha_1 = 0.10, \epsilon_2 = 0.10, \alpha_2 = 0.05, \epsilon_3 = 0.05, \alpha_3 = 0.025$, then the P-value is no larger than 0.025.

The top panel of Figure 2.3 shows a histogram of Chi-Square statistics for the 1,483 genes eliminated in step 3, along with the null model $\chi^2_2(0)$. A fitted CCS model is not superimposed because W is negative in the notation of Section 2.3. This precludes valid moment-based estimation of γ and μ. Even if a likelihood-based approach can be applied to estimate γ and μ, this is not called for because the omnibus null hypothesis is not rejected at any $\alpha \leq 0.25$, regardless of whether one takes

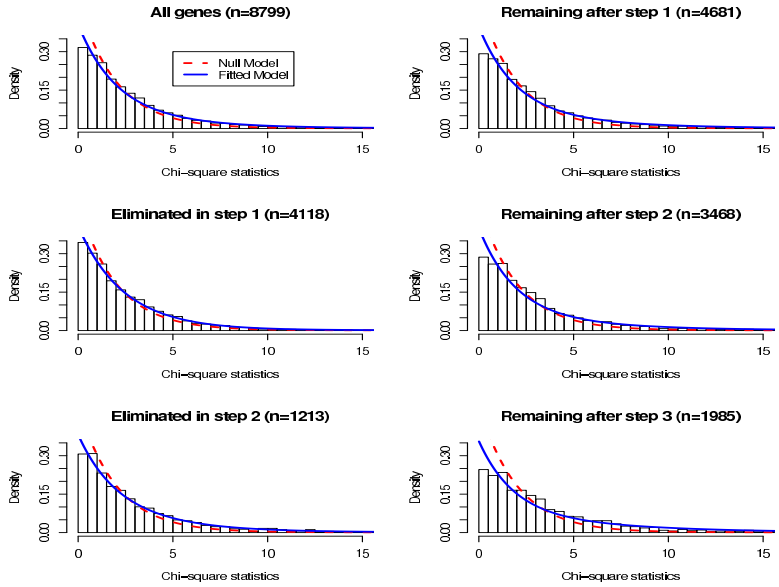
Table 2.1: Parameter Estimations for All Genes, Remaining Genes and Eliminated Genes Groups by CCS Model

Genes	Estimated γ	Estimated μ
All genes	0.231	3.25
Remaining after 1A	0.236	4.13
Eliminated in 1A	0.389	1.28
Remaining after 2A	0.223	4.54
Eliminated in 2A	0.314	2.77
Remaining after 3A	0.308	5.19

$\delta = 1/2, \epsilon = 2\alpha$ or $\delta = \epsilon = \alpha^{1/2}$ or $\delta = 2\alpha, \epsilon = 1/2$. In fact, the null model is not a bad fit to the histogram, except for overstating the number of very small Chi-Square statistics.

The bottom panel of Figure 2.3 shows a histogram of the P-values for the same

Figure 2.2: Null Models and CCS Models for All Genes, Remaining Genes and Eliminated Genes

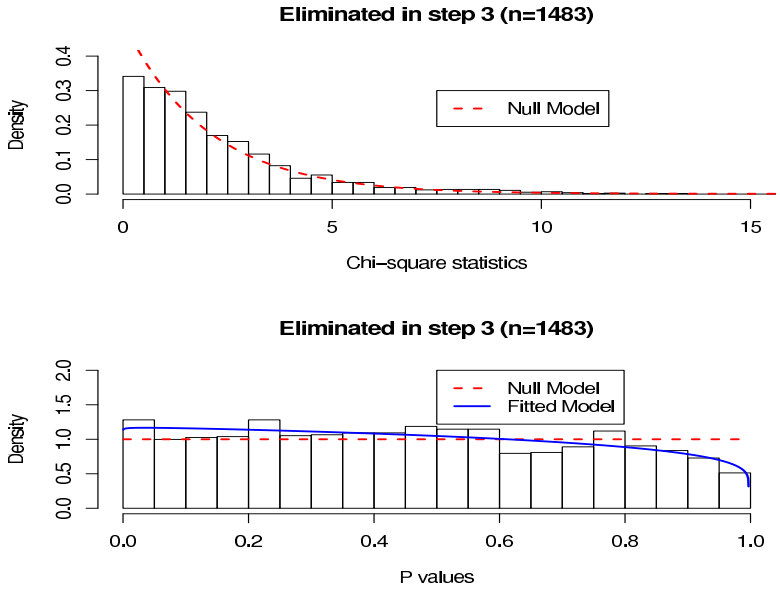


1,483 genes, along with the fitted CB model (2.6) and the null model of a uniform

distribution. From this panel, one may find there are noticeably fewer extremely large P-values than would be compatible with a uniform distribution. And both the MLR test and D test decisively reject the omnibus null hypothesis of a uniform distribution for this reason rather than a surplus of small P-values. This rejection is inappropriate insofar as one uses it to infer differential expression.

In summary, employing the CCS model to analyze Chi-Square statistics instead

Figure 2.3: Eliminated Genes in Step 3 fitted by CB Model and CCS Model



of the CB model to assess P-values resolves the aforementioned concern, because the omnibus null hypothesis from (2.2) is not rejected for the genes eliminated in step 3. Thus, using the CCS model avoided an inappropriate declaration of differential expression.

2.6 Discussion

In this chapter, we have developed a moment-based testing procedure for testing the omnibus null hypothesis of no contamination of a central Chi-Square distribution by a non-central Chi-Square distribution. This procedure is based on the first two sample moments, which permits critical values to be derived from quantiles of the standard normal distribution. One of the advantages of this moment-based testing procedure is that even for small sample sizes, there is excellent agreement between the nominal and actual significance levels. Another advantage is that since the asymptotic null distribution is not as complicated as the one from likelihood ratio tests for mixture models (Dacunha-Castelle and Gassiat, 1999 [17]; Chen and Chen, 2001 [14]; Liu and Shao, 2003 [33]; Chambaz, 2006 [8]), there is no need for re-sampling (McLachlan, 1987 [35]) or random field theory (Sun, 1993 [39]) to obtain critical values. Besides, the Chi-Square statistics can overcome an inappropriate declaration of differential expression which may not be avoided by analyzing P-values. Moreover, our simulation studies show that, under certain conditions, analysis of Chi-Square statistics may actually yield better power to detect differential expression than analysis of P-values.

As for the estimating procedure, moment-based estimators of the contamination fraction and non-centrality parameter of the contaminating distribution have been proposed when the omnibus null hypothesis is rejected. Provided that the fraction and parameter are both nonzero, our estimators are $n^{1/2}$-consistent. Moreover, our estimators have probability 1 of being positive, conditional on rejection of the omnibus null hypothesis, with a good choice of δ and ϵ. This differs from previous work

in that moment-based estimators in mixture models ordinarily do not belong to their respective parameter spaces with probability 1, as noted by Charnigo and colleagues (2013) [12] for another type of contamination model.

Our testing and estimation procedures are primarily motivated by the modeling of numerous Chi-Square statistics arising from microarray data analysis specifically or large-scale testing generally. A filtration process similar to that employed by Blalock et al (2003) [5], which can reduce Type II errors by justifying lighter controls for Type I errors can also be applied here.

While we only investigate the application of CCS model to Chi-Square statistics related to F statistics from one-way ANOVA, the potential applications of the CCS model are considerably broader. For example, if the normality and equal variance assumptions underlying one-way ANOVA are violated, then one may employ the nonparametric Kruskal-Wallis test for equal medians. Since the Kruskal-Wallis test statistic is distributed approximately $\chi^2_{k-1}(0)$ when the medians are equal, the CCS model can be applied in conjunction with Chi-Square statistics from Kruskal-Wallis tests as easily as with F statistics from one-way ANOVA.

Moreover, sophisticated experimental designs or sampling schemes may preclude using either one-way ANOVA or Kruskal-Wallis tests. For instance, Mao and colleagues (2005) [34] obtained multiple tissue samples from some of their subjects, so that linear mixed models were required to test genewise null hypotheses. However, as long as genewise null hypotheses are tested using Chi-Square or F statistics (or even Z or T statistics, since these can be squared), the CCS model remains applicable.

A number of promising avenues exist for future research. One of them is to investigate whether the EM test (Chen and Li, 2009 [16]; Li, Chen, and Marriott, 2009

[29]) can be profitably employed in the setting of the CCS model and, in particular, whether power to reject a false omnibus null hypothesis is improved; we will pursue this investigation in Chapter 3.

Another topic for future research is to generalize the CCS model to provide greater flexibility for describing real data. For instance, suppose that each X_i has its own non-centrality parameter μ_i under the genewise alternative hypothesis. Then we may consider a new model,

$$(1-\gamma)\chi^2_\nu(0) + \gamma \int \chi^2_\nu(\mu) dG(\mu), \tag{2.7}$$

where G is some cumulative distribution function defined on the nonnegative real numbers. Note that the first sample moment of data from (2.7) is ν if and only if (2.3) reduces to $\chi^2\nu(0)$, as both are equivalent to $\gamma\{1-G(0)\} = 0$. Thus, one obtains a consistent level α test for whether (2.7) reduces to $\chi^2_\nu(0)$ by asking whether the first sample moment exceeds $n^{-1}q_{\nu n,1-\alpha}$, where $q_{\nu n,1-\alpha}$ denotes the $1-\alpha$ quantile of $\chi^2_{\nu n}(0)$. However, the subsequent estimation of γ and G are anticipated to be considerably more delicate.

But one potential problem is that our testing and estimation procedures are based on iid data. In practice, $F_1, F_2, \ldots, F_n$ or $X_1, X_2, \ldots, X_n$ will not be independent. How robust are these testing procedures to the iid assumption? i.e. if $X_1, X_2, \ldots, X_n$ are correlated, will our tests still work? Will the type I or type II error go up?

Chapter 3 Modified Likelihood Ratio Test and EM Test for Contaminated Chi-Square Model

3.1 Introduction

In Chapter 2, we developed a moment-based method for inference in the Contaminated Chi-Square model. Generally speaking, a moment-based approach can yield consistent point estimators and hypothesis testing procedures under mild conditions. However, a moment-based approach is problematic for complicated statistical models such as large mixtures of high-dimensional distributions. The reason for that is the equations determining the estimators of parameters are typically based on moments of order equal to the number of model parameters, and high-order moments are exceedingly difficult to estimate accurately due to their large variances (Anandkumar et al, 2012) [1].

An alternative to the moment-based method for testing homogeneity in finite mixture models is the modified likelihood ratio method. Likelihood-based methods play a central role in parametric testing problems, and among these the likelihood ratio test (LRT) is often preferred (Chen, Chen and Kalbfleisch, 2001 [14]). Under standard regularity conditions, the LRT statistic has a simple and elegant asymptotic χ^2 distribution under the null hypothesis (Wilks, 1938 [44]). But most of the asymptotic results of LRT can't be applied to mixture models. In this chapter, we propose a modified likelihood ratio test (MLRT), implemented with the aid of the Expectation-Maximization (EM) algorithm (a local search heuristic for likelihood-

based estimation) (Dempster, Laird and Rubin, 1977 [20]), for the Contaminated Chi-Square model by introducing a penalty term $C \log(1 - |1 - 2\gamma|)$ to the log likelihood. We show in later sections that the MLRT has better power than the moment-based test and enjoys an elegant asymptotic theory. The MLRT is also appealing as a general strategy in that a large number of parameters in the model can be accommodated.

However, the EM algorithm has its limitations in practice, including slow convergence and suboptimal local optima (Redner and Walker, 1984 [37]). Recognizing these difficulties, we also develop an EM-test which shares the same simple limiting distribution with the MLRT. The EM-test requires no more than two to three iterations of the EM algorithm, which is time efficient. Simulation studies show that the EM-test has accurate type I error rates and appealing power.

Chapter 3 is organized as follows. Section 3.2 states the regularity conditions based on which the theorems and lemmas in Chapter 3 are built. Section 3.3 describes the major theorems and lemmas of inference via the MLRT and EM-test for CCS model. The asymptotic distribution of the MLRT statistic under the omnibus null hypothesis of homogeneity is investigated. In particular, the asymptotic null distribution of the MLRT is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. The estimator $\hat{\mu}$ is convergent in probability to 0 under the omnibus null hypothesis. Then the EM-test statistic is proved to have the same limiting distribution as the MLRT statistic. Section 3.4 presents the proofs regarding asymptotic distributions. To study the type I error rates and the power of these two test statistics, simulation studies are conducted. Section 3.5 presents simulation results and section 3.6 analyzes data from the microarray experiment by Blalock and colleagues (2003) [5]. These data have been studied by a moment-based

method in section 2.5, but now will be studied using the MLRT and the EM-test. Section 3.7 is a brief summary of the two new tests.

3.2 Regularity Conditions on the Kernel Function

Before introducing the MLRT statistic and the EM test statistic, five regularity conditions are stated. These conditions are adapted from Chen, Chen and Kalbfeisch (2001) [14]. All lemmas and theorems in Chapter 3 are built based on the following regularity conditions. The proofs that these five regularity conditions are statisfied for the CCS model appear in Appendix I. Below let $f_\nu(X;\mu) = \chi^2_\nu(\mu)$, denoting the pdf of the non-central Chi-Square distribution on $\nu > 0$ df with non-central parameter $\mu \in \Theta$, where Θ is defined to be $[0, M]$, M is a large positive constant.

Condition 1. Wald's integrability conditions. The kernel function $\chi^2_\nu(\mu)$ satisfies Wald's integrability conditions for consistency of the maximum likelihood estimator, i.e. for each $\mu \in \Theta$, (i) $E\|\log f_\nu(X;\mu)\| < \infty$, and (ii) for sufficiently small $\rho > 0$ the expected values $E \log f(X;\mu,\rho) < \infty$, where

$$f(X;\mu,\rho) = 1 + sup_{\|\mu'-\mu\|\leq\rho}\{f(X;\mu')\}.$$

Condition 2. Smoothness. The kernel function $\chi^2_\nu(\mu)$ has common support for all $\mu \in \Theta$ and is twice continuously differentiable with respect to μ.

Condition 3. Strong identifiability. For any two mixing distribution functions Ψ_1 and Ψ_2

such that

$$\int f_\nu(X;\mu)d\Psi_1(\mu) = \int f_\nu(X;\mu)d\Psi_2(\mu), \text{for all } x,$$

we must have $\Psi_1 = \Psi_2$. The mixing distribution Ψ is defined as

$$\Psi_(\mu) = (1-\gamma)I(\mu_1 \leq \mu) + \gamma I(\mu_2 \leq \mu).$$

Condition 4. Uniform strong law condition of large numbers. There exists integrable g with some $\delta > 0$ such that $\|Y_i(\mu)\|^3 \leq g(X_i)$ for all $\mu \in \Theta$, where $Y_i(\mu) = \frac{f_\nu(X_i;\mu)-f_\nu(X_i;0)}{\mu f_\nu(X_i;0)}$; $Y(0) = \frac{f'_\nu(X_i;0)}{f_\nu(X_i;0)}$.

Condition 5. Tightness. The processes $n^{-1/2}\sum Y_i(\mu)$ is tight.

3.3 Modified Likelihood Ratio Test and EM-test

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a two-component CCS model (2.1). We define the penalized log-likelihood function for CCS model as

$$pl_n(\gamma,\mu) = \sum_{i=1}^{n} \log\{(1-\gamma)f_\nu(X_i;0) + \gamma f_\nu(X_i;\mu)\} + p(\gamma) \tag{3.1}$$

where $p(\gamma)$ is a penalty function on γ. We choose $p(\gamma)$ to be $C\log(1-|1-2\gamma|)$ for some positive C. The penalty term is minimized at $\gamma = 0.5$ and $\hat{\gamma}$ is bounded away from zero or one in probability. Li, Chen and Marriott's paper in 2008 [29] found this penalty function could best balance the type I error rate and the power. However, we emphasize that the theoretical results of Li, Chen and Marriott are not applicable to the CCS model because they did not assume one of the mixture component parameters be to known nor did their mixture component parameters lie on the boundary of their parameter space under the null hypothesis.

The modified log-likelihood ratio statistic we proposed for testing homogeneity is

$$L_n(\hat{\gamma}, \hat{\mu}) = 2\{pl_n(\hat{\gamma}, \hat{\mu}) - pl_n(0.5, 0)\} \tag{3.2}$$

where $\hat{\gamma}$ amd $\hat{\mu}$ maximize the modified likelihood function (3.1) over the parameter space $\gamma \in (0, 1), \mu \in [0, M]$. The EM algorithm is used to iteratively approximate these maximizers. We assume that the largest number of iterations for the EM algorithm is K. Let

$$\gamma^{(1)} = \gamma_0,$$

where γ_0 is the initial value, $\gamma^{(1)}$ is the estimated value of γ in the 1st iteration, and k goes from 0 to K. Then we calculate

$$\mu^{(1)} = argmax_{\mu} ml(\gamma^{(1)}, \mu)$$

and

$$L_n^{(1)} = 2\{pl_n(\gamma^{(1)}, \mu^{(1)}) - pl_n(0.5, 0)\}.$$

The following procedure is the main part of the EM algorithm for k goes from 2 to K.

- Get the conditional expectation $w_i^{(k)}$ in E-step for each $i = 1, 2, \ldots, n$.

$$w_i^{(k)} = \frac{\gamma^{(k)} f_\nu(X_i; \mu^{(k)})}{\gamma^{(k)} f_\nu(X_i; \mu^{(k)}) + (1 - \gamma^{(k)}) f_\nu(X_i; 0)}$$

- Maximize the approximation to the complete data penalized log likelihood in M-step. Let

$$\gamma^{(k+1)} = argmax_{\gamma}\{(n - \sum_{i=1}^{n} w_i^{(k)}) \log(1-\gamma) + \sum_{i=1}^{n} w_i^{(k)} \log(\gamma) + C \log(1 - |1 - 2\gamma|)\},$$

and

$$\mu^{(k+1)} = argmax_{\mu}\{\sum_{i=1}^{n} w_i^{(k)} \log f(X_i;\mu)\}.$$

- Compute

$$L_n^{(k+1)}(\gamma^{(k+1)}, \mu^{(k+1)}) = 2\{pl_n(\gamma^{(k+1)}, \mu^{(k+1)}) - pl_n(0.5, 0)\}.$$

Let $k = k+1$ and repeat the above procedure until $k = K$ or $\left|L_n^{(k+1)}(\gamma^{(k+1)}, \mu^{(k+1)}) - L_n^{(k)}(\gamma^{(k)}, \mu^{(k)}\right.$ is less than a pre-specified tolerance.

Let $\hat{\gamma}$ and $\hat{\mu}$ maximize the penalized likelihood. These maximizers are numerically approximated by the EM algorithm with $\gamma^{(K)}$ and $\mu^{(K)}$ or $\gamma^{(k+1)}$ and $\mu^{(k+1)}$. Likewise, the MLRT statistic $2\{pl_n(\hat{\gamma}, \hat{\mu}) - pl_n(\frac{1}{2}, 0)\}$ is numerically approximated with $L_n^{(K)}(\gamma^{(K)}, \mu^{(K)})$ or $L_n^{(k+1)}(\gamma^{(k+1)}, \mu^{(k+1)})$. The following lemma and theorem show that the modified maximum likelihood estimator (MMLE) $\hat{\mu}$ converges in probability to 0 under the null hypothesis.

Lemma 2. *Under the null hypothesis,* $C \log(1 - |1 - 2\hat{\gamma}|) = O_p(1)$.

Proof. Let L_n be the MLRT statistic defined in (3.2). Let $R_n = 2\{l_n(\gamma^{(k)}, \mu^{(k)}) - l_n(0.5, 0)\}$ be the ordinary LRT statistic. First we need to prove that L_n is stachastically bounded. Since $C \log(1 - |1 - 2\gamma|)$ is non-positive for $\gamma \in (0, 1)$, it is obvious

$$0 \leq L_n \leq R_n. \tag{3.3}$$

By Theorem 1 in Di and Liang's paper (2011) [21], under the null hypothesis, $R_n = O_p(1)$, so $L_n = O_p(1)$. Since R_n is the maximum likelihood ratio,

$$0 \leq L_n - C \log(1 - |1 - 2\hat{\gamma}|) \leq R_n.$$

Thus $L_n - C\log(1-|1-2\hat{\gamma}|) = O_p(1)$, so $\log(1-|1-2\hat{\gamma}|) = O_p(1)$. □

Theorem 3. *Under the null hypothesis, $\hat{\mu} \to 0$ in probability as $n \to \infty$.*

Proof. Let

$$Q(\gamma,\mu) = E[\log\{(1-\gamma)f_\nu(X;0) + \gamma f_\nu(X;\mu)\} - \log f_\nu(X;0)].$$

According to the uniform strong law of large numbers in C4, under the null hypothesis,

$$\frac{1}{n}\{l_n(\gamma,\mu) - l_n(0.5,0)\} \to Q(\gamma,\mu), \tag{3.4}$$

almost surely and uniformly over $\delta \le \gamma \le 1-\delta$ and $\mu \in \Theta$. Let ω be a point in the sample space such that (3.3) is true. Then we get a set of all these points with probability 1. Then we prove the therem by contradiction. Follow Chen, Chen and Kalbfeisch, 2001's paper, we suppose there exists a ω such that the claim of the theorem is not true, i.e., $\hat{\mu}$ doesn't converge to 0. Then we can find a corresponding subsequence n' such that $\hat{\mu} \to \mu'$ and $\mu' \neq 0$. Consider $\Omega' = \{\mu : \delta \le \gamma \le 1-\delta, |\mu - 0| \ge \epsilon\}$, where $\epsilon = (\mu' - 0)/2$. Then for all large n', $(\hat{\gamma},\hat{\mu})$ at the sample point ω belongs to Ω'. In addition, $Q(\gamma,\mu) < 0$ for all $(\gamma,\mu) \in \Omega'$. So we have $l_{n'}(\hat{\gamma},\hat{\mu}) - l_{n'}(0.5,0) < 0$ for all large n'. Thus $(\hat{\gamma},\hat{\mu})$ can't be the MMLE and so $\hat{\mu} \to 0$ on ω. Therefore $\hat{\mu} \to 0$ a.s.

□

The other nice property of MLRT statistic is its simple limiting distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ under the null hypothesis, which makes the critical value easily accessed using a χ^2 table. This distribution can be established with the aid of the following result on the maximum modified likelihood point estimators.

Theorem 4. *Under the null hypothesis of homogeneity, we have*

$$\hat{\gamma} - \frac{1}{2} = o_p(1), \hat{\mu} = O_p(n^{-1/2}).$$

$\hat{\gamma}$ and $\hat{\mu}$ are the MMLE under the omnibus null hypothesis.

Theorem 5. *Under the null hypothesis of homogeneity, MLRT statistic converges in law to $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$.*

While the MLRT statistic enjoys a simple limiting null distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, the MLRT statistic is not time efficient since it is computed using the EM algorithm which is characterized by slow (numerical) convergence. To solve this practical problem, we may employ the EM-test statistic, which uses only a few iterations of the EM algorithm, and therefore is not impeded by slow convergence. The idea of the EM-test is that we choose J number of initial values instead of only one initial value for γ. And for each initial value, we repeat the pseudo code for the MLRT statistic in section 3.3 except that K is equal to a very small number like 1 or 2. Hence, we get J MLRT-like statistics which we may denote $L_n^{(K)}(\gamma_j, \mu_j), j = 1, 2, \ldots, J$. Define the EM-test statistic to be

$$EM_n^{(k)} = \max\{L_n^{(K)}(\gamma_j, \mu_j), j = 1, 2, \ldots, J\} \tag{3.5}$$

Theorem 6. *Under the null hypothesis of homogeneity, we have*

$$\gamma_j^{(k)} - \gamma_j = o_p(1), \mu_j^{(k)} = O_p(n^{-1/2}),$$

where γ_j is the jth initial value for γ, $\gamma_j^{(k)}$ and $\mu_j^{(k)}$ are the estimators from the kth iteration of EM algorithm.

Theorem 7. *For fixed finite K,*

$$EM_n^{(K)} \to \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2 \tag{3.6}$$

in law under the null hypothesis of homogeneity.

Theorem 7 shows that the EM-test statistic shares the same simple limiting null distribution with the MLRT statistic. In the simulation study in section 3.5, we will see that the EM-test has desirable Type I and Type II error properties with K as small as 1. Thus, the EM-test is more computationally efficient than the MLRT. However, the MLRT may be more appealing when there is interest not only in hypothesis testing but also in estimation, since with multiple initial values, there is ambiguity in defining point estimators.

3.4 Proofs of Asymptotic Null Distribution of L_n and $EM_n^{(K)}$

Proof. Proof of Theorem 4. The basic idea of the proof follows Chen, H., Chen, J., Kalbfleisch, J.D. (2001) [14]. It can be proved that

$$\begin{aligned}
0 &\leq -2C\log(1-|1-2\hat{\gamma}|) \\
&\leq -2(l_n(\hat{\gamma},\hat{\mu}) - l_n(0.5,\hat{\mu}(0.5))) \\
&= -2\{\sum_{i=1}^{n}\log f(X_i;0) - \sum_{i=1}^{n}\log\{(1-\hat{\gamma})f(X_i;0)+\hat{\gamma}f(X_i;\hat{\mu})\}\} \\
&+ 2\{\sum_{i=1}^{n}\log f(X_i;0) - \sum_{i=1}^{n}\log\{(1-0.5)f(X_i;0)+0.5f(X_i;\hat{\mu}(0.5))\}\} \\
&= \{\frac{(\sum Y_i^+)^2}{\sum Y_i^2} + o_p(1)\} - \{\frac{(\sum Y_i^+)^2}{\sum Y_i^2} + o_p(1)\} \\
&= o_p(1)
\end{aligned}$$

We have proved that

$$\begin{aligned} 2c &\le 2\{pl_n(\hat{\gamma},\hat{\mu}) - pl_n(0.5,0)\} \\ &\le 2\{l_n(\hat{\gamma},\hat{\mu}) - l_n(0.5,0)\} \\ &\le 2\sum_{i=1}^{n}\{\hat{m}Y_i\} - \sum_{i=1}^{n}\{\hat{m}Y_i\}^2\{1+o_p(1)\} + o_p(1) \\ &\le \frac{((\sum_{i=1}^{n} Y_i)^+)^2}{\sum_{i=1}^{n} Y_i^2} + o_p(1) \end{aligned} \tag{3.7}$$

where $\hat{m} = \hat{\gamma}\hat{\mu}$. From the above two inequality, it can be concluded that

$$2\hat{m}\sum_{i=1}^{n} Y_i - \hat{m}^2\{\sum_{i=1}^{n} Y_i^2\}\{1+o_p(1)\} = O_p(1)$$

Because $\sum_{i=1}^{n} Y_i = O_p(n^{\frac{1}{2}})$ and $\sum_{i=1}^{n} Y_i^2 = O_p(n)$, we get $\hat{m} = O_p(n^{-\frac{1}{2}})$. Because $\hat{\gamma}$ is bounded away from 0 and 1 in probability, $\hat{\mu} - 0 = O_p(n^{-\frac{1}{2}})$.

□

Next we show that the asymptotic null distribution of MLRT based on the estimators $\hat{\gamma}, \hat{\mu}$, which maximize the modified likelihood function pl_n is a mixture of χ^2 distributions. Let

$$R_n(\gamma,\mu) = 2[l_n(\gamma,\mu) - l_n(0.5,0)] + C\log(1-|1-2\gamma|).$$

Write $R_n(\gamma,\mu)$ in the form of

$$R_n(\gamma,\mu) = 2\sum_{i=1}^{n}\log(1+\delta_i) + C\log(1-|1-2\gamma|),$$

where $\delta_i = \gamma\{\frac{f_\nu(X_i;\mu)}{f_\nu(X_i;0)} - 1\}$. Put $Y_i = Y_i(0)$, then

$$\delta_i = mY_i + e_{in},$$

with $m = \gamma\mu$, and $e_{in} = \gamma(\frac{f_\nu(X_i;\mu) - \mu f'_\nu(X_i;0)}{f_\nu(X_i;0)} - 1)$. $Y_i(\mu)$ is defined to be $\frac{f_\nu(X_i,\mu) - f_\nu(X_i,0)}{\mu f_\nu(X_i,0)}$

Using the Taylor Expansion, we have

$$R_n(\gamma,\mu) \leq 2\sum_{i=1}^{n}\{mY_i + e_{in}\} - \sum_{i=1}^{n}\{mY_i + e_{in}\}^2 + \frac{2}{3}\sum_{i=1}^{n}\{mY_i + e_{in}\}^3.$$

Now several lemmas are stated to support the proof. The proof of lemmas follow the idea of Chen, Chen and Kalbfleisch (2001) [14].

Lemma 3. *Under the null hypothesis, uniformly in m,*

$$\frac{\sum_{i=1}^{n}\{mY_i\}^3}{\sum_{i=1}^{n}\{mY_i\}^2} = O_p(1)|m|.$$

Proof. From the uniform strong law condition of large numbers, which is satisfied for the family of Chi-Square distributions indexed by their non-centrality parameters,

$$n^{-1}\sum_{i=1}^{n}\{mY_i\}^k \to E[mY_1]^k,$$

almost surely and uniformly for $k = 2, 3$. From the strong identifiability condition, the limit of $n^{-1}\sum_{i=1}^{n}\{mY_i\}^k$ when $k = 2$ is a positive-definite quadratic form in m. So its smallest eigenvalue is positive. Then, $\sum_{i=1}^{n}\{mY_i\}^2 \geq m^2 O_p(1)$ uniformly in m. It follows that

$$\begin{aligned}\left|\frac{\sum_{i=1}^{n}\{mY_i\}^3}{\sum_{i=1}^{n}\{mY_i\}^2}\right| &= \left|\frac{E\{mY_i\}^3}{E\{mY_i\}^2}O_p(1)\right| \\ &\leq \frac{|m|^3}{m^2}O_p(1) \\ &= |m|O_p(1)\end{aligned}$$

□

Lemma 4. *Under the null hypothesis, uniformly in γ and μ,*

$$\left|\sum_{i=1}^{n} e_{in}\right| = n^{1/2}\gamma\mu O_p(1).$$

Proof.

$$\left|\sum_{i=1}^{n} e_{in}\right| = |\sum_{i=1}^{n} \delta_i - mY_i|$$
$$= \gamma\mu|\sum_{i=1}^{n} Y_i(\mu) - Y_i(0)|$$

We know from the tightness condition that $n^{-1/2}\sum_{i=1}^{n}\{Y_i(\mu) - Y_i(0)\} = O_p(1)$, it follows that $|\sum_{i=1}^{n} e_{in}| = n^{1/2}\gamma\mu O_p(1)$. □

Then $R_n(\gamma, \mu)$ can be expressed as

$$R_n(\gamma, \mu) \leq 2\sum_{i=1}^{n}\{mY_i\} - \sum_{i=1}^{n}\{mY_i\}^2 + \frac{2}{3}\sum_{i=1}^{n}\{mY_i\}^3 + O_p(\sum_{i=1}^{n} e_{in}). \tag{3.8}$$

Applying the previous two lemmas to (3.6) yields

$$R_n(\gamma, \mu) \leq 2\sum_{i=1}^{n}\{mY_i\} - \sum_{i=1}^{n}\{mY_i\}^2[1 + |m|O_p(1)]. \tag{3.9}$$

Let $\hat{m}$ be the MMLE of m. By Theorem 3, $\hat{m} = o_p(1)$ under the null hypothesis. Thus,

$$R_n(\gamma, \mu) \leq 2\sum_{i=1}^{n}\{\hat{m}Y_i\} - \sum_{i=1}^{n}\{\hat{m}Y_i\}^2\{1 + o_p(1)\} + o_p(1). \tag{3.10}$$

We can further prove that the right side of the above inequality is asymptotically no larger than the maximum of the following quadratic function

$$q(m) = 2\sum_{i=1}^{n}\{mY_i\} - \sum_{i=1}^{n}\{mY_i\}^2$$
$$= n - \sum_{i=1}^{n}(1 - mY_i)^2.$$

Since $m \geq 0$, the maximum is obtained at $\tilde{m} = \frac{(\sum Y_i)^+}{\sum Y_i^2}$. Thus an upper bound for $R_n(\hat{\gamma}, \hat{\mu})$ is established as follows:

$$R_n(\hat{\gamma}, \hat{\mu}) \leq q(\tilde{m}) + o_p(1) = \frac{((\sum Y_i)^+)^2}{\sum Y_i^2} + o_p(1). \tag{3.11}$$

Next we need to prove the upper bound is achievable. Let $\tilde{\delta}_i$ be the value of δ_i when $\gamma = \frac{1}{2}$, $\mu = \tilde{\mu}$, $\tilde{\mu}$ is determined by $\tilde{m}$. By Taylor Expansion,

$$\begin{aligned} R_n(\frac{1}{2}, \tilde{\mu}) &= 2\sum_{i=1}^{n} \log(1+\tilde{\delta}_i) \\ &= 2\sum_{i=1}^{n} \tilde{\delta}_i - \sum_{i=1}^{n} \tilde{\delta}_i^2 (1-\xi_i)^{-2}, \end{aligned}$$

where $|\xi_i| < |\tilde{\delta}_i|$. Because $\tilde{\delta}_i = \tilde{m}Y_i, |\tilde{\delta}_i| \leq |\tilde{m}Y_i|$. Then we have the following inequality,

$$\begin{aligned} \max|\tilde{\xi}| &\leq \max|\tilde{\delta}_i| \\ &= |\tilde{m}| \max|Y_i| \\ &= O_p(n^{\frac{1}{2}}) o_p(\sqrt{\log n}) \\ &= o_p(1). \end{aligned}$$

Thus

$$R_n(\frac{1}{2}, \tilde{\mu}) = 2\sum_{i=1}^{n} \tilde{\delta}_i - \sum_{i=1}^{n} \tilde{\delta}_i^2 (1+o_p(1)),$$

which means the upper bound (3.10) is achievable, i.e.

$$R_n(\hat{\gamma}, \hat{\mu}) = q(\tilde{m}) + o_p(1) = \frac{((\sum Y_i)^+)^2}{\sum Y_i^2} + o_p(1). \tag{3.12}$$

Therefore we have established the asymptotic null distribution of MLRT statistic. Based on the proof above, now we continue to prove the asymptotic null distribution of the EM-test statistic. Before proving Theorem 7, three lemmas are stated and proved as follows. The proof of the lemmas follow the idea of Li, Chen and Marriott (2009) [29]. The difference is that in their model, they have two unknown parameters except the proportion parameter. And none of their parameters are on the boundary

of the parameter space when the null hypothesis is true. While our model only has one unknown parameter, which can be on the boundary of the parameter space.

Lemma 5. *Let* $(\hat{\gamma}, \hat{\mu})$ *be some estimators of* (γ, μ) *such that* $\delta_1 \leq \hat{\gamma} \leq \delta_2$ *for some* $\delta_1 < \delta_2 \in (0, 1)$*, and*

$$l_n(\hat{\gamma}, \hat{\mu}) - l_n(0.5, 0) \geq c > -\infty$$

for some fixed c. Then under $H_0, \hat{\mu} - 0 = o_p(1)$.

Proof. The parameter space under the full model is $[0, 1] \times [0, M]$. The parameter space with the indicated restriction on $\hat{\gamma}$ is $\Lambda := \{(\gamma, \mu) : \delta_1 \leq \gamma \leq \delta_2, \mu \in [0, M]\}$. The parameter space of a null model with this restriction is $\{(\gamma, 0) : \delta_1 \leq \gamma \leq \delta_2\}$. For some positive constants ϵ and r, define

$$A(\gamma, \epsilon, r) = \{(\gamma', \mu) \in \Lambda; |\gamma' - \gamma| \leq \epsilon, |\mu| > r\},$$

and

$$\Psi(X; \gamma, \epsilon, r) = \sup\{\gamma' f_\nu(X; \mu') + (1 - \gamma') f_\nu(X; 0) : (\gamma', \mu') \in A(\gamma; \epsilon, r)\}.$$

By Wald's integrability condition and Smoothness condition, for small enough ϵ and large enough r, under the null hypothesis,

$$E[\log \Psi(X; \gamma, \epsilon, r)] < E[\log f_\nu(X; 0)].$$

Therefore, by the law of large numbers,

$$Pr[\sup\{l_n(\gamma', \mu') : (\gamma', \mu') \in A(\gamma; \epsilon, r)\} - l_n(\gamma, 0) > c] \to 0 \text{ for } \forall c > -\infty.$$

The above conclusion can be extended to

$$Pr[\sup\{l_n(\gamma', \mu') : (\gamma'; \mu') \in A\} - l_n(\gamma, 0) > c] \to 0,$$

where $A = \cup_{\delta_1 \leq \gamma \leq \delta_2} A(\gamma; \epsilon, r)$. Therefore the log-likelihood at any parameter point with large μ trails the log-likelihood at the true parameter point by an infinite amount accoding to Li, Chen and Marriott (2009) [29]. □

Lemma 6. *Let $(\hat{\gamma}, \hat{\mu})$ be some estimators of (γ, μ) such that under the null hypothesis, $\hat{\mu} - 0 = o_p(1)$ and $\delta_1 \leq \hat{\gamma} \leq \delta_2$ for some $\delta_1 < \delta_2 \in (0, 1)$. If*

$$pl_n(\hat{\gamma}, \hat{\mu}) - pl_n(0.5, 0) \geq c > -\infty,$$

then under the null hypothesis, $\hat{\mu} - 0 = O_p(n^{-\frac{1}{2}})$.

Proof. In theorem 4, we proved that for each initial value $\hat{\gamma}$,

$$2\hat{m} \sum_{i=1}^{n} Y_i - \hat{m}^2 \{\sum_{i=1}^{n} Y_i^2\}\{1 + o_p(1)\} = O_p(1)$$

Because $\sum_{i=1}^{n} Y_i = O_p(n^{\frac{1}{2}})$ and $\sum_{i=1}^{n} Y_i^2 = O_p(n)$, we get $\hat{m} = O_p(n^{-\frac{1}{2}})$. Due to the condition that $\delta_1 \leq \hat{\gamma} \leq \delta_2$ for some $\delta_1 < \delta_2 \in (0, 1)$, we further conclude that $\hat{\mu} - 0 = O_p(n^{-\frac{1}{2}})$. □

Now we show that under the null hypothesis, the EM-iteration changes the fitted value of γ by $o_p(1)$. Let $(\hat{\gamma}, \hat{\mu})$ be some estimators of (γ, μ) with the asymptotic properties as before, and let

$$w_i = \frac{\hat{\gamma} f_\nu(X_i; \hat{\mu})}{(1 - \hat{\gamma}) f_\nu(X_i; 0) + \hat{\gamma} f_\nu(X_i; \hat{\mu})}.$$

We further define $R_n(\gamma) = (n - \sum_{i=1}^{n} \hat{w}_i) \log(1 - \gamma) + \sum_{i=1}^{n} \hat{w}_i \log(\gamma)$. And $Q_n(\gamma) = R_n(\gamma) + p(\gamma)$. The EM algorithm updates γ by searching for $\hat{\gamma}^* = argmax Q_n(\gamma)$.

Lemma 7. *Suppose that $\hat{\gamma} - \gamma_0 = o_p(1)$ for some $\gamma_0 \in (0, 1)$. Under the null hypothesis, we have $|\hat{\gamma}^* - \gamma_0| = o_p(1)$.*

Proof. For $i = 1, 2, \ldots, n$, let

$$\hat{\delta}_i = \hat{\gamma}\hat{\mu}\frac{f_\nu(X_i;\hat{\mu}) - f_\nu(X_i;0)}{\hat{\mu} f_\nu(X_i;0)} = \hat{m} Y_i(\hat{\mu}).$$

Thus $\max_{1\le i\le n} |\hat{\delta}_i| = |\hat{m}| \max |Y_i(\hat{\mu})| \le |\hat{m}| \max_{1\le i\le n}\{\sup_{1\le i\le n, 0\le\mu\le M} Y_i(\mu)\}$.

By the uniform boundedness condition and a result on order statistic, we have $\max_{1\le i\le n} |\sup_{\mu\in N(0), 0\le\mu\le M} Y_i(\mu)| = o_p(n^{\frac{1}{2}})$. It follows that $max_{1\le i\le n}|\delta_i| = o_p(1)$. By Taylor's expansion of $f_\nu(X_i;\hat{\mu})$ at $\hat{\mu} = 0$, we get

$$\begin{aligned}\hat{w}_i - \hat{\gamma} &= \hat{\gamma}(1-\hat{\gamma})\frac{f_\nu(X_i;\hat{\mu}) - f_\nu(X_i;0)}{(1-\hat{\gamma}) f_\nu(X_i;0) + \hat{\gamma} f_\nu(X_i;\hat{\mu})} \\ &= \frac{\hat{\gamma}(1-\hat{\gamma})}{1+\hat{\delta}_i}\{\hat{\mu} Y_i(\hat{\mu})\}.\end{aligned}$$

Put $\tilde{\gamma} = \frac{1}{n}\sum_{i=1}^n \hat{w}_i$, we have

$$|\tilde{\gamma} - \hat{\gamma}| = \left|(\hat{\mu} - 0)\sum_{i=1}^n Y_i(\hat{\mu})\right| O_p(n^{-1}) = o_p(1).$$

Based on the above result and the assumption that $\hat{\gamma} - \gamma_0 = o_p(1)$, we have $\tilde{\gamma} - \gamma_0 = o_p(1)$. Thus we can get the conclusion that $\hat{\gamma}^* - \tilde{\gamma} = o_p(1)$.

For $\forall\epsilon > 0$, and $\gamma \ge \tilde{\gamma} + 2\epsilon$, it follows that

$$R_n(\gamma) - R_n(\tilde{\gamma}) \le R_n(\tilde{\gamma} + 2\epsilon) - R_n(\tilde{\gamma} + \epsilon) = \epsilon R_n'(\xi)$$

for some $\xi \in [\tilde{\gamma} + \epsilon, \tilde{\gamma} + 2\epsilon]$ by mean value theorem. It is easy to prove that $R_n'(\xi) \to -\infty$ as $n \to \infty$ uniformly for ξ in this range. On the other hand, we have

$$p(\gamma) - p(\tilde{\gamma}) = p(\gamma) - p(\gamma_0) + o_p(1) = O_p(1).$$

Thus,

$$Q_n(\gamma) - Q_n(\tilde{\gamma}) = R_n(\gamma) - R_n(\tilde{\gamma}) + \{p(\gamma) - p(\tilde{\gamma})\} \to -\infty$$

uniformly for any $\gamma > \tilde{\gamma} + 2\epsilon$ with the probability goes to 1. Hence, we must have $\hat{\gamma}^* < \tilde{\gamma} + 2\epsilon$ in probability. Similarly, $\hat{\gamma}^* > \tilde{\gamma} - 2\epsilon$ in probability. This completes the proof of Lemma 7. □

Proof. Proof of Theorem 6. By the property of EM algorithm, we have

$$pl_n(\gamma_j^{(k)}, \mu_j^{(k)}) \geq pl_n(\gamma_j, \mu_j^{(0)}) \geq pl_n(\gamma_j, 0).$$

It follows that $l_n(\gamma_j^{(k)}, \mu_j^{(k)}) - l_n(\gamma_j, 0) \geq p(\gamma_j) - p(\gamma_j^{(k)}) \geq p(\gamma_j) - p(0.5) > -\infty$. By Lemma 5 and $\gamma_j^{(0)} = \gamma_j$, it has been shown that $\mu_j^{(0)}$ is consistent for 0. Therefore the conclusion of Lemma 6 and Lemma 7 apply. Hence, we find $\gamma_j^{(1)} - \gamma_j = o_p(1), \mu_j^{(1)} - 0 = O_p(n^{-\frac{1}{2}})$.

This completes the proof for $k = 1$. Then by mathematical induction, the conclusion of the theorem is true for all finite k. □

Proof. Proof of Theorem 7. According to the results proved in Theorem 4, the inequality (3.7) is applicable. Hence for any k, we have

$$2\{pl_n(\gamma_j^{(k)}, \mu_j^{(k)}) - pl_n(0.5, 0)\} \leq \frac{((\sum_{i=1}^n Y_i)^+)^2}{\sum_{i=1}^n Y_i^2} + o_p(1).$$

At the same time, we can prove that the upper bound is achievable when $\gamma_j = 0.5$ by the same method in proof of lemma 4. Thus

$$EM_n^{(K)} = \frac{((\sum_{i=1}^n Y_i)^+)^2}{\sum_{i=1}^n Y_i^2} + o_p(1).$$

Therefore, the limiting null distribution is given by $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$. □

3.5 Simulation Studies

In this section, we investigate the performance of the MLRT and the EM-test by evaluating their Type I error rates and the power for sample sizes n $\in \{10, 50, 100, 200, 500, 1000\}$.

First 1,000 random samples $X_1, X_2, \ldots, X_n$ were generated from CCS model (2.1) under the null hypothesis of homogeneity. In the simulation studies, we chose $C = 1$ in the penalty term. The choice has been used in a variety of mixture models with satisfactory results (Chen and Li, 2009 [16]). The nominal Type I error is selected to be $\alpha = 0.05$. Based on 1,000 replications, the rejection rates of both MLRT and EM-tests for sample sizes greater than 50, are close to 0.05. For EM-tests, three initial values $0.1, 0.3, 0.5$ were used for γ. Table 3.1 presents the actual Type I error rates for MLRT, $EM_n^{(1)}$ and $EM_n^{(2)}$, where $EM_n^{(1)}$ is the EM-test with one iteration, and $EM_n^{(2)}$ is the EM-test with two iterations.

Panels in Figure 3.1 show both exact power curves and asymptotic power curves

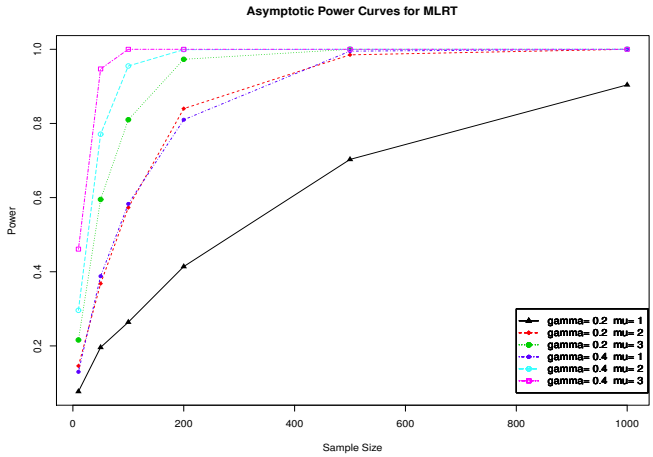
Table 3.1: Actual Type I Error Rates for MLRT, $EM_n^{(1)}$ and $EM_n^{(2)}$

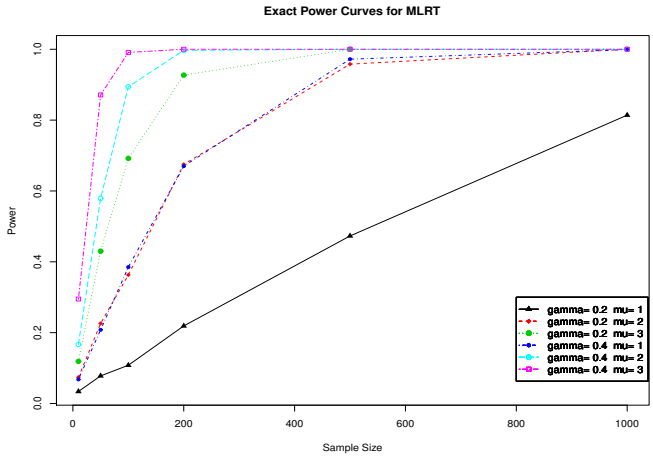
Model n	10	50	100	200	500	1000
MLRT	0.036	0.046	0.047	0.050	0.058	0.050
$EM_n^{(1)}$	0.051	0.044	0.045	0.050	0.058	0.050
$EM_n^{(2)}$	0.047	0.051	0.050	0.062	0.046	0.049

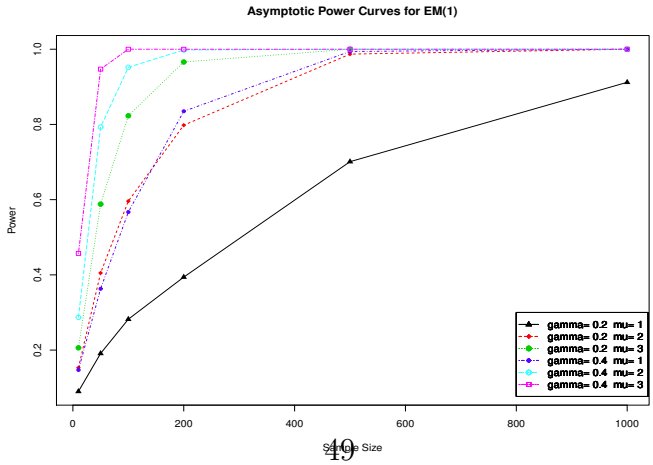
for MLRT, $EM_n^{(1)}$ and $EM_n^{(2)}$. The asymptotic power is based on the frequency with which the test statistic exceeds the .95 quantile of $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$, which is 2.70, whereas the exact power is based on the frequency with which the test statistic exceeds the .95 quantile of test statistics simulated under the null hypothesis of homogeneity.

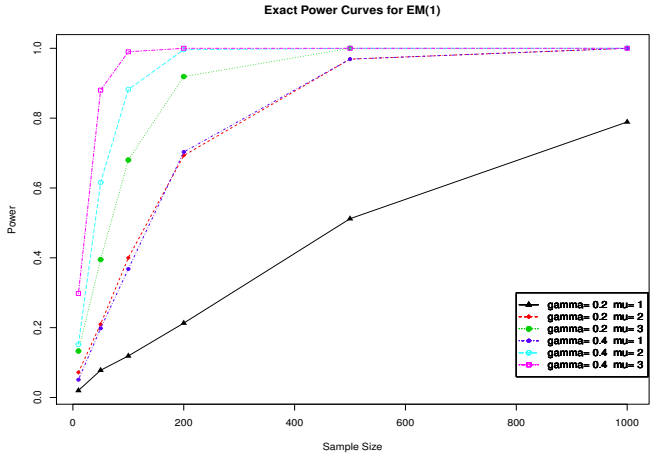
For power comparison, We chose several different models with $\mu = 1, 2, 3$, $\gamma = 0.2, 0.4$ and $\nu = 2$. The first two panels and the second two panels are the asymptotic and the exact powers for MLRT and $EM_n^{(1)}$ correspondingly. The power curves for $EM_n^{(2)}$ are displayed in the last two panels. In all six models with $\mu = 1, 2, 3$,

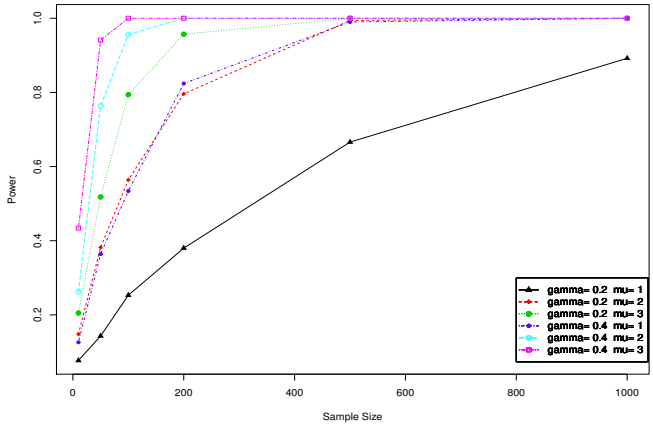
Figure 3.1: Exact Power Curves and Asymptotic Power Curves

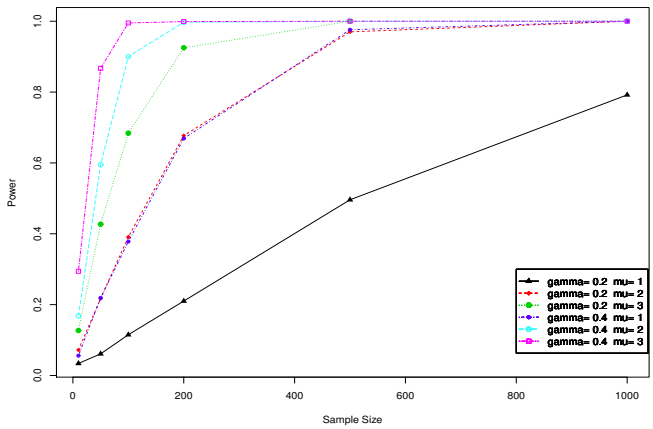







Exact Power Curves for EM(1)
Power
Sample Size
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
gamma= 0.2 mu= 1
gamma= 0.2 mu= 2
gamma= 0.2 mu= 3
gamma= 0.4 mu= 1
gamma= 0.4 mu= 2
gamma= 0.4 mu= 3


Asymptotic Power Curves for EM(2)
Power
Sample Size
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
gamma= 0.2 mu= 1
gamma= 0.2 mu= 2
gamma= 0.2 mu= 3
gamma= 0.4 mu= 1
gamma= 0.4 mu= 2
gamma= 0.4 mu= 3

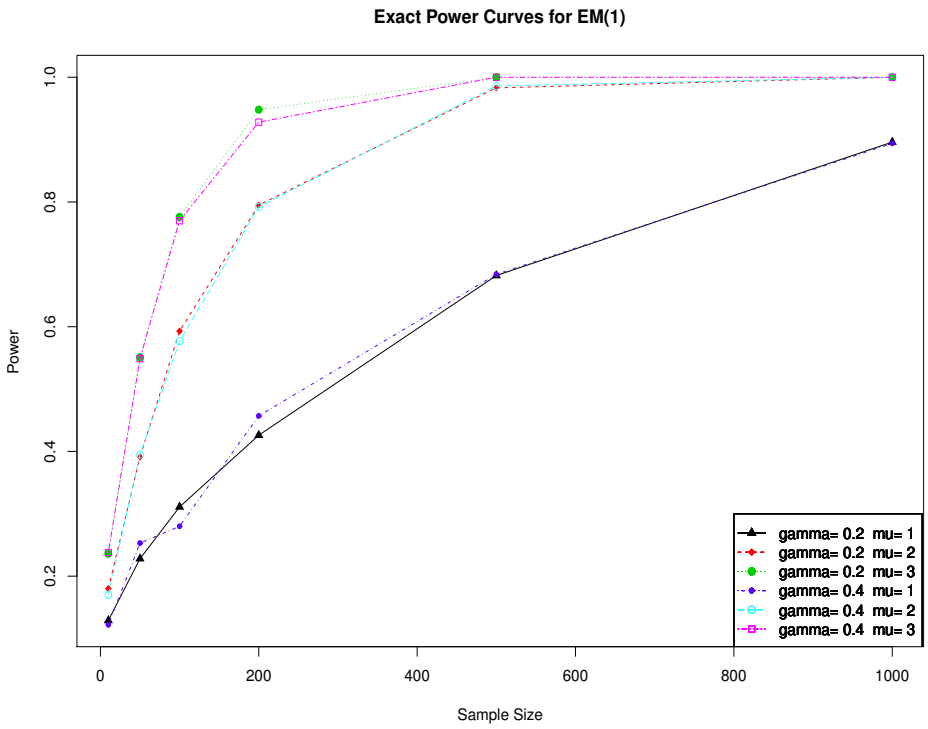

Exact Power Curves for EM(2)
Power
Sample Size
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600
800
1000
gamma= 0.2 mu= 1
gamma= 0.2 mu= 2
gamma= 0.2 mu= 3
gamma= 0.4 mu= 1
gamma= 0.4 mu= 2
gamma= 0.4 mu= 3

$\gamma = 0.2, 0.4$ and $\nu = 2$, the power of all three test statistics becomes stronger as n increases. The model with $\gamma = 0.4, \mu = 0.3$ always enjoys the best power, whereas the model with $\gamma = 0.2, \mu = 0.1$ always has the least power. Two things worth noticing are, first the larger the values of γ and μ the better the power. The other is that if the product of γ and μ is the same for two models, it is very possible the two models share similar power. We conclude this by comparing results from model $\gamma = 0.4, \mu = 1$ and model $\gamma = 0.2, \mu = 2$. The red line and the blue line corresponding to these models cross each other from $n = 10$ to $n = 1000$. Especially in the panel of exact power curves for $EM_n^{(2)}$, they nearly overlap.

Even though $EM_n^{(1)}$ entails only one iteration of the EM algorithm, it achieves power as good as the MLRT. But $EM_n^{(1)}$ runs much faster than the MLRT as the sample size n gets large. With one more iteration, there is no notable improvement in power from $EM_n^{(1)}$ to $EM_n^{(2)}$ in any case considered here.

In reality, if the null hypothesis of homogeneity is violated, then the distribution of F statistics from one-way ANOVA is contaminated F (CF) rather than Contaminated Chi-Square. To check whether our tests still work well for data from CF model, we did another simulation when data were from CF distribution. Since EM-test has about the same power as the MLRT, and it runs more efficiently, and we see there is little difference in power between $EM_n^{(1)}$ and $EM_n^{(2)}$, we only do simulation for exact power of $EM_n^{(1)}$. We transformed F statistics from ANOVA for each gene to the Chi-Square statistics by probability integral transform: $X_i := cdf_{\chi^2_{\nu_1}}^{-1}(cdf_{F_{\nu_1,\nu_2}}(F_i))$, where ν_1 and ν_2 are the numerator and denominator df for the F statistic F_i.

Figure 3.2: Exact Power Curves for $EM^{(1)}$ When Data is from Contaminated F Model



Compare with the exact power curve for $EM_n^{(1)}$ with data from CCS distribution, one may find the powers are similar for models with the same μ value when data are transformed from CF model to CCS model, which differs from the powers when data are originally from CCS model. But as the sample size gets larger, the powers will become close. These results are also noteworthy in that power at fixed n appeared to be mainly a function of μ when the EM-test for the CCS model is applied as an approximation to transformed data actually generated from the CF model, while power appeared to be mainly a function of the product $\gamma\mu$ when the EM-test for the CCS model was applied to data actually generated from the CCS model.

3.6 Case Study

In section 2.5, CCS model with the moment-based estimates was applied to analyze the microarray experiments on 8,799 genes from three groups of rats by Blalock and colleagues (2003) [5]. In this section, we study this data set again, still applying the filtration procedure, but with likelihood-based estimates. From the simulations in previous section, there is not much difference in power among $EM_n^{(1)}$, $EM_n^{(2)}$ and MLRT. And we know EM-test is not suitable when estimation is also of interest (i.e., not just testing). We chose MLRT here for real data analysis, which is accomplished by maximum modified likelihood estimators, since the hippocampal data are presented from CF distribution , we first converted the F statistics to their corresponding P-values, then converted P-values to Chi-Square statistics, and finally used MLRT and MMLE to do the estimation and hypothesis testing.

Table 3.2 shows the parameter estimates and the P-values of MLRT procedure for all 8,799 genes, the genes eliminated in each step, and for the genes remaining after each step. From Table 3.2, we see the P-values are all less than 0.0001, which means the omnibus null hypothesis needs to be rejected in each step. Correspondingly, Figure 3.3 displays histograms of Chi-Square statistics for all 8,799 genes, and the genes eliminated and remaining in each step. The fitted CCS model is superimposed on each histogram. From Figure 3.3, each fitted CCS model fits the corresponding histogram quite well.

Compared with the estimation from moment-based method, in general the estimates of γ from MMLE run larger and the estimates of μ run smaller. However,

the product of $\hat{\gamma}$ and $\hat{\mu}$ is similar to that from moment-based method. Another noticeable result is, the MLRT rejects the omnibus null hypothesis of no differential expression for the eliminated genes after the third step, unlike the moment-based test. One of the possible reasons is that since the MLRT uses more information than the moment-based method, the MLRT is more sensitive to departures from the null distribution, even departures that are not reflected in the first two sample moments or that may not be suggestive of differential expression.

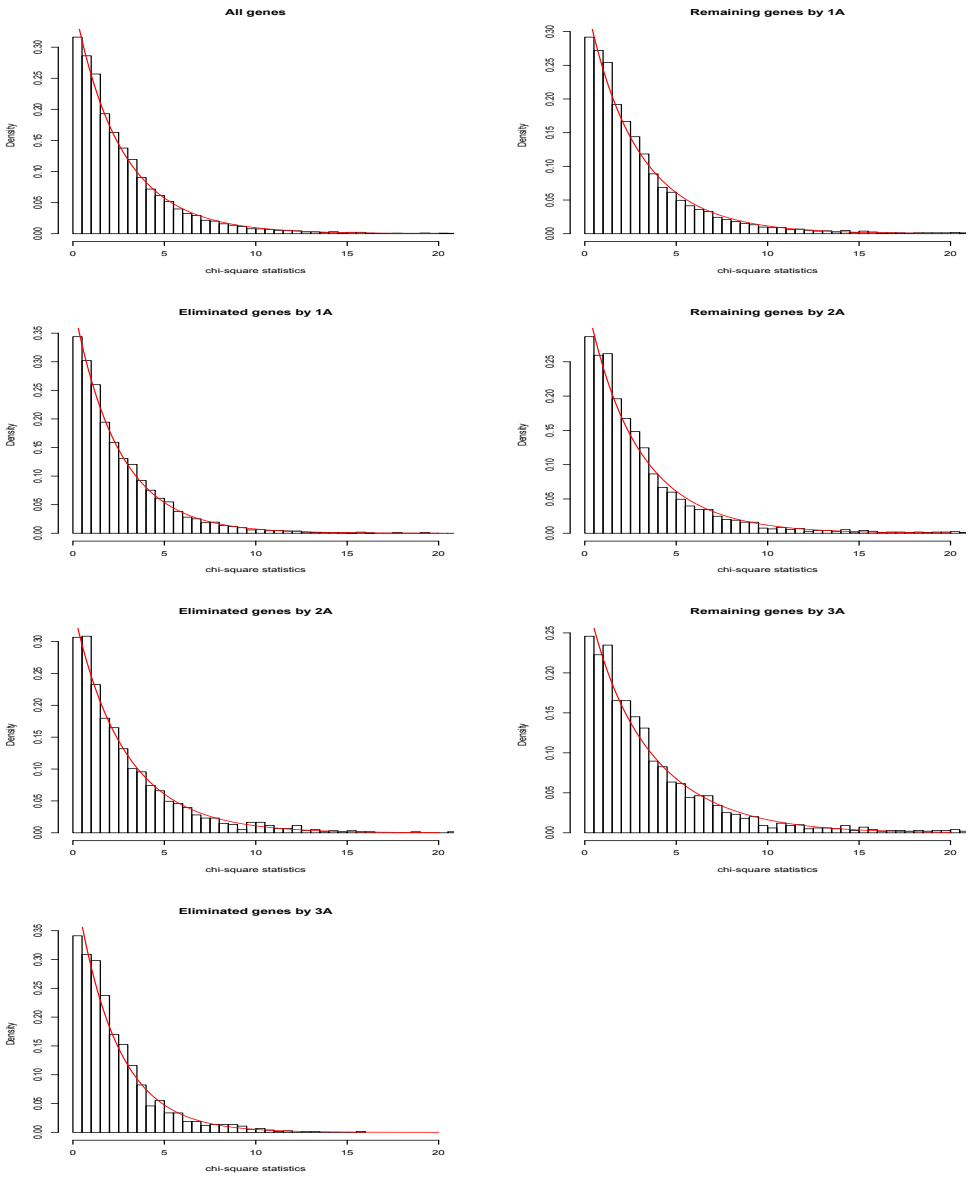
Table 3.2: P-values and Parameter Estimation by MLRT Method

Genes	Estimated γ	Estimated μ	Estimated $\gamma\mu$	P-value
All genes	0.41	1.75	0.72	$< .0001$
Remaining after 1A	0.46	1.97	0.91	$< .0001$
Eliminated in 1A	0.50	0.99	0.50	$< .0001$
Remaining after 2A	0.47	2.01	0.94	$< .0001$
Eliminated in 2A	0.50	1.73	0.87	$< .0001$
Remaining after 3A	0.54	2.60	1.66	$< .0001$
Eliminated after 3A	0.50	0.45	0.23	$< .0001$

3.7 Discussion

In Chapter 3, we developed two likelihood-based methods for hypothesis testing in Contaminated Chi-Square model, with accompanying estimation methodology available in one case. Likelihood methods in general compete with moment-based methods in that they can deal with large mixtures of high-dimensional distributions. For the CCS model which is neither a large nor a high-dimensional mixture, likelihood methods still may hold appeal due to greater power than moment-based methods, although this greater power can also manifest as sensitivity to departures from the null distribution that may not be indicative of differential expression. One method

Figure 3.3: Fitted Model by MLRT Method



we developed is modified likelihood ratio test. It enjoys a nice limiting distribution $\frac{1}{2}\chi_0^1+\frac{1}{2}\chi_0^2$ under the null and is asymptotically most powerful under local alternatives in other models where it has been employed previously (Chen, Chen and Kalbfleisch, 2001 [14]). However, the MLRT has its weak point. When the sample size is large, because computing the MLRT statistic requires the EM algorithm, which in turn requires many iterations to achieve convergence, the MLRT statistic may be computed quite slowly. To conquer this problem, an EM-test was developed. It is also based on EM algorithm and shares the same simple asymptotic null distribution with MLRT statistic. The difference is that we choose several initial values rather than one to start with. In our simulation studies, the EM-test has acceptable Type I and Type II error rates in only one iteration, which is much more time efficient than the MLRT. Even though the EM-test only iterates once or twice, it achieves power as good as MLRT. But due to the fact that it has more than one set of initial values, there is ambiguity in addressing the estimation problem when the EM test.

We developed the MLRT and EM-test procedure for data coming from Contaminated Chi-Square model. But a collection of ANOVA F statistics may follow the Contaminated F distribution when the omnibus null hypothesis is false. Therefore we also simulated the power of EM-test when the data is from contaminated F distribution instead of Contaminated Chi-Square distribution. In this case, we used a probability integral transformation to transform F statistics to Chi-Square statistics. Simulations showed that as the sample size gets large, the power of EM-test on data from Contaminated F distribution will get close to the power of EM-test on data from Contaminated Chi-Square distribution.

Chapter 4 Moment-Based Inference and EM Test for CCS+EC Model

4.1 CCS+EC Model

The CCS model presented in chapter 2 and 3 tests the homogeneity of expression, with an alternative hypothesis that all differentially expressed genes yield Chi-Square-scores (or transformed F-scores) from a single non-central Chi-Square distribution. In real data, it is very possible there are more than two latent groups of genes. For instance, some individuals may be overexpressed on a particular gene relative to normal individuals but underexpressed on another gene. Charnigo et al (2013) [11] proposed a Bilaterally Contaminated Normal (BCN) model for testing whether the three-component mixture may be reduced to a two-component mixture with one nonzero component mean. The BCN model can be used to describe z-scores derived from genewise hypothesis tests. It can also accommodate t-scores, since t-scores may be converted to z-scores via the map $T \mapsto \text{sign}(T)\Phi^{-1}[1 - p(T)/2] =: Z$, where $p(T)$ is the two-sided P-value associated with the t-score T and Φ is the standard normal cdf. However, genewise hypothesis tests are sometimes based on Chi-Square-scores or F-scores, especially when there are more than two groups of subjects or experimental conditions.

There are difficulties by simply transforming Chi-Square-scores or F-scores to z-scores for the purpose of employing the BCN model. One problem is that z-scores have both positive and negative values. For instance, a z-score of 5 obtained from transforming a Chi-Square-score or F-score is a strong signal of differential expres-

sion, corresponding to a P-value of 0.0000003. On the other hand, a z-score of -5 provides no signal of differential expression, corresponding to a P-value of 0.9999997. Indeed Chi-Square scores and F-scores are unsigned and thus cannot carry information about overexpression versus underexpression. Therefore, Chi-Square-scores and F-scores cannot be accommodated by the BCN model.

Thus we propose an alternative way to handle Chi-Square-scores and F-scores by adding extra components to the CCS model. One most parsimonious such model is $(1-\gamma_2-\gamma_3)\chi^2_\nu(0)+\gamma_2\chi^2_\nu(\mu_2)+\gamma_3\chi^2_\nu(\mu_3)$, in which Chi-Square-scores corresponding to differentially expressed genes arise from one of two non-central Chi-Squares distributions depending on the degree of differential expression. We label this extension as the "CCS+EC" ("EC" for extra component) model. For CCS model, we developed moment-based tests and two likelihood-based tests for homogeneity of differential expression. For CCS+EC model, likewise, we will also consider both moment-based tests and EM tests in the following sections. The CCS model of Chapters 2 and 3 allows us to test for differential expression; testing for homogeneity of differential expression is permitted by the CCS+EC model.

4.2 Moment-Based Inference for CCS+EC Model

Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (iid) according to the mixture probability density function (pdf)

$$(1-\gamma_2-\gamma_3)\chi^2_\nu(0)+\gamma_2\chi^2_\nu(\mu_2)+\gamma_3\chi^2_\nu(\mu_3), \tag{4.1}$$

where $\gamma_2, \gamma_3, 1-\gamma_2-\gamma_3, \mu_2, \mu_3$ are nonnegative unknown parameters with $\gamma_2, \gamma_3 \in [0,1]$ and $\mu_2, \mu_3 \in [0,+\infty)$, while $\nu \in (0,+\infty)$ is known. $\chi^2_\nu(0)$ denotes the central Chi-

Square model on ν degrees of freedom. $\chi^2_\nu(\mu_2)$ and $\chi^2_\nu(\mu_3)$ denote the noncentrality Chi-Square models with noncentral parameters μ_2 and μ_3 respectively on ν degrees of freedom. We refer (4.1) as the CCS+EC model.

An interesting problem here is to develop a moment-based test of the secondary null hypothesis that $\gamma_2\gamma_3\mu_2\mu_3(\mu_2-\mu_3)^2=0$. If the secondary null hypothesis is true, then the CCS+EC model reduces to the CCS model with pdf

$$(1-\gamma)\chi^2_\nu(0)+\gamma\chi^2_\nu(\mu),$$

with $\gamma\in[0,1]$, $\mu\in[0,+\infty)$ and known parameter $\nu\in(0,+\infty)$.

If the secondary null hypothesis is not rejected, we proceed to use one of the three methods introduced in chapters 2 and 3 to test the omnibus null hypothesis. Otherwise, we stop. An alternative way is to first test the omnibus null hypothesis. If the omnibus null hypothesis is rejected, then we continue testing the secondary null hypothesis.

Before stating the theorems, we establish some notations for convenience. Let $\mathbf{m}:=(m_1,m_2,m_3,m_4,m_5,m_6)^T$, $\hat{\mathbf{m}}:=(\hat{m}_1,\hat{m}_2,\hat{m}_3,\hat{m}_4,\hat{m}_5,\hat{m}_6)^T$, and $\mathbf{0}:=(0,0,0,0,0,0)^T$, where $m_t:=E[X_1^t]$ and $\hat{m}_t:=n^{-1}\sum_{i=1}^n X_i^t$. Define the 3×3 matrix $\mathbf{V}(\mathbf{m})$ to contain $m_{i+j}-m_im_j$ in row i and column j for $i,j\in\{1,2,3\}$. Put $g(\mathbf{m}):=2\nu^3+4\nu^2+\nu^2m_2-8\nu^2m_1+8\nu m_2-16\nu m_1-\nu m_3+m_1m_3-m_2^2-4m_1m_2+\nu m_1m_2-\nu^2m_1^2+2\nu m_1^2+8m_1^2$ and $\mathbf{h}(\mathbf{m}):=(-8\nu^2-16\nu+m_3-4m_2+\nu m_2-2\nu^2m_1+4\nu m_1+16m_1;\nu^2+8\nu-2m_2-4m_1+\nu m_1;-\nu+m_1)^T$, which is the derivative of $g(\mathbf{m})$ with respect to m_1,m_2,m_3.

Based on the assumption that the omnibus null hypothesis is false, $\mathbf{h}(\mathbf{m})\neq\mathbf{0}$. By Delta method, we know

$$\sqrt{n}(g(\hat{\mathbf{m}})-g(\mathbf{m}))\to N(0,\mathbf{h}(\mathbf{m})^T\mathbf{V}(\mathbf{m})\mathbf{h}(\mathbf{m})). \tag{4.2}$$

Then by Slutsky's Theorem,

$$\sqrt{n}\left(\frac{g(\hat{\mathbf{m}})}{\sqrt{\mathbf{h}(\hat{\mathbf{m}})^T\mathbf{V}(\hat{\mathbf{m}})\mathbf{h}(\hat{\mathbf{m}})}} - \frac{g(\mathbf{m})}{\sqrt{\mathbf{h}(\mathbf{m})^T\mathbf{V}(\mathbf{m})\mathbf{h}(\mathbf{m})}}\right) \to N(0,1). \tag{4.3}$$

Let $Z_n := \sqrt{\frac{n}{\mathbf{h}(\hat{\mathbf{m}})^T\mathbf{V}(\hat{\mathbf{m}})\mathbf{h}(\hat{\mathbf{m}})}}g(\hat{\mathbf{m}})$,$z_u$ denote the u quantile of the standard normal distribution, and $\mathbf{m}_a$ be the vector of moments implied by $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (\gamma_{2,a}, \gamma_{3,a}, \mu_{2,a}, 0)$ for fixed positive constants $\gamma_{1,a}, \gamma_{2,a}$ and $\mu_{2,a}$.

Theorem 8. *Suppose that $X_1, X_2, \ldots, X_n$ are iid according to model (4.1) and that the omnibus null hypothesis is false. Under the secondary null hypothesis, i.e. $\gamma_2\gamma_3\mu_2\mu_3(\mu_2 - \mu_3)^2 = 0$, and $\gamma_2\mu_2 + \gamma_3\mu_3 > 0$,*

$$\lim_{n\to\infty} P(Z_n > z_{1-\alpha}) = \alpha \tag{4.4}$$

Under the local alternative sequence $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (\gamma_{2,a}, \gamma_{3,a}, \mu_{2,a}, \tau n^{-0.5})$ for a fixed positive constant τ,

$$\lim_{n\to\infty} P(Z_n > z_{1-\alpha}) = \Phi\left(-z_{1-\alpha} + \frac{\gamma_{2,a}\gamma_{3,a}\mu_{2,a}^3\tau}{\sqrt{\mathbf{h}(\mathbf{m_a})^{\mathbf{T}}\mathbf{V}(\mathbf{m_a})\mathbf{h}(\mathbf{m_a})}}\right). \tag{4.5}$$

Under the fixed alternative $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (\gamma_{2,a}, \gamma_{3,a}, \mu_{2,a}, \mu_{3,a})$ for a fixed positive constant $\mu_{3,a}$,

$$\lim_{n\to\infty} P(Z_n > z_{1-\alpha}) = 1 \tag{4.6}$$

Proof. Under the secondary null hypothesis, since Z_n follows the standard normal distribution asymptotically, it is obvious that $\lim_{n\to+\infty} P(Z_n > z_{1-\alpha}) = 1 - \Phi(z_{1-\alpha}) = \alpha$. Under the local alternative sequence, combining (4.2) and (4.3),

$$\begin{aligned} P(Z_n > z_{1-\alpha}) &= P\left(Z_n - g(m_a^*)\sqrt{\frac{n}{\mathbf{h}(\mathbf{m_a^*})^T\mathbf{V}(\mathbf{m_a^*})\mathbf{h}(\mathbf{m_a^*})}} > z_{1-\alpha} - g(m_a^*)\sqrt{\frac{n}{\mathbf{h}(\mathbf{m_a^*})^T\mathbf{V}(\mathbf{m_a^*})\mathbf{h}(\mathbf{m_a^*})}}\right) \\ &\to 1 - \Phi\left[z_{1-\alpha} - \frac{\gamma_{2,a}\gamma_{3,a}\mu_{2,a}^3\tau}{\sqrt{\mathbf{h}(\mathbf{m_a})^T\mathbf{V}(\mathbf{m_a})\mathbf{h}(\mathbf{m_a})}}\right], \end{aligned}$$

where m_a^* is the vector implied by $(\gamma_{2,a}, \gamma_{3,a}, \mu_{2,a}, \tau n^{-0.5})$ and m_a is the vector implied by $(\gamma_{2,a}, \gamma_{3,a}, \mu_{2,a}, 0)$.

Under the fixed alternative, $1-P(Z_n \leq z_{1-\alpha}) = 1-P(Z_n - g(m)\sqrt{\frac{n}{\mathbf{h(m)}^T\mathbf{V(m)h(m)}}} \leq z_{1-\alpha} - g(m)\sqrt{\frac{n}{\mathbf{h(m)}^T\mathbf{V(m)h(m)}}}) \to 1$, because $g(\mathbf{m}) > 0$ under the H_a and $z_{1-\alpha} - g(m)\sqrt{\frac{n}{\mathbf{h(m)}^T\mathbf{V(m)h(m)}}} \to -\infty$.

□

4.3 Simulation Studies for Moment-Based Test

In this section, we evaluate the performance of moment-based test for CCS+EC model by assessing the Type I error probability and the power of our testing procedure in finite samples. The Type I error probability and the power were simulated under both nominal critical values and actual critical values for the following nine models:

m1: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 2, 0.2, 0.3)$

m2: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 4, 0.2, 0.3)$

m3: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (2, 4, 0.2, 0.3)$

m4: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 2, 0.3, 0.2)$

m5: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 4, 0.3, 0.2)$

m6: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (2, 4, 0.3, 0.2)$

m7: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 2, 0.25, 0.25)$

m8: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (1, 4, 0.25, 0.25)$

m9: $(\gamma_2, \gamma_3, \mu_2, \mu_3) = (2, 4, 0.25, 0.25)$

All these nine models are with degrees of freedom 2. The corresponding null model is defined as a model with the same γ_2, γ_3 and ν, but a weighted $\mu = \frac{\mu_2 * \gamma_2 + \mu_3 * \gamma_3}{\gamma_2 + \gamma_3}$. For instance, the null model of $(\gamma_2, \gamma_3, \mu_2, \mu_3, \nu) = (1, 2, 0.2, 0.3, 2)$ is $(1.6, 1.6, 0.2, 0.3, 2)$.

For each model, 100 random samples $X_1, X_2, \ldots, X_n$ were drawn from the corresponding null model to estimate the actual critical value with sample size $n \in \{500, 1000, 5000, 10000, 50000\}$. The critical value is defined as the 95th percentile of the test statistics based on the null model. In table 4.1, we can tell that all of the simulated critical values are smaller than 1.645. When the sample size gets larger, the actual critical value gets close to 1.645, but still does not achieve 1.645. This indicates us that the test based on simulated critical values will be a less conservative test compared with the test based on the nominal critical values.

The actual Type I error rates are defined to be the rejection rates under H_0

Table 4.1: Actual Critical Values for Different Sample Sizes

Alternative Model/Sample Size	500	1000	5000	10000	50000
(1,2,0.2,0.3,2)	1.0819	1.1013	1.2448	1.3076	1.4073
(1,4,0.2,0.3,2)	1.0705	1.1221	1.2453	1.3054	1.4259
(2,4,0.2,0.3,2)	1.0820	1.1310	1.2496	1.3092	1.4438
(1,2,0.3,0.2,2)	1.0497	1.1073	1.2573	1.2753	1.4087
(1,4,0.3,0.2,2)	1.0605	1.1343	1.2602	1.3151	1.3978
(2,4,0.3,0.2,2)	1.0698	1.1488	1.2550	1.3285	1.4285
(1,2,0.25,0.25,2)	1.0561	1.1065	1.2305	1.2792	1.4296
(1,4,0.25,0.25,2)	1.0789	1.1162	1.2601	1.3188	1.4349
(2,4,0.25,0.25,2)	1.0815	1.1497	1.2772	1.3194	1.4070

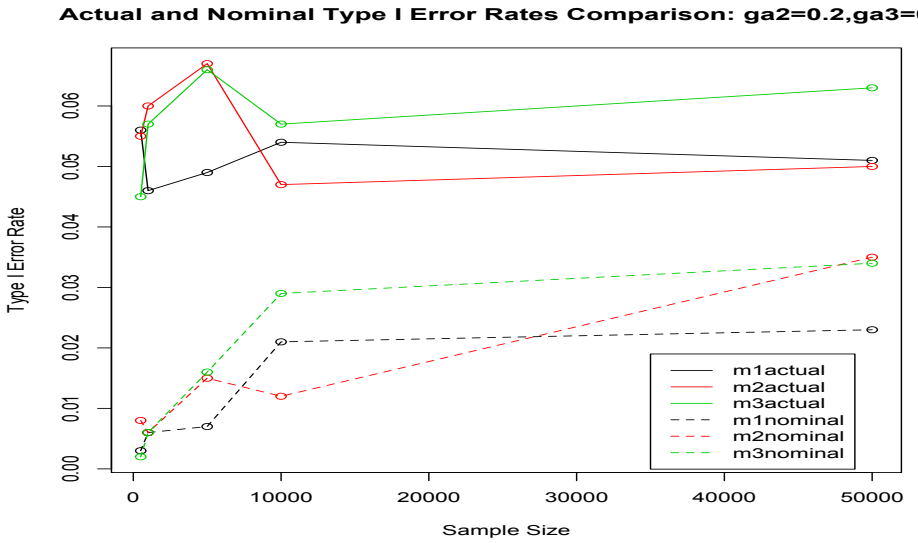
when the critical values are through simulation while the nominal Type I error rates are defined to be the rejection rates under H_0 when the critical value is 1.645 as derived from the standard normal distribution. For each sample size n, we generated

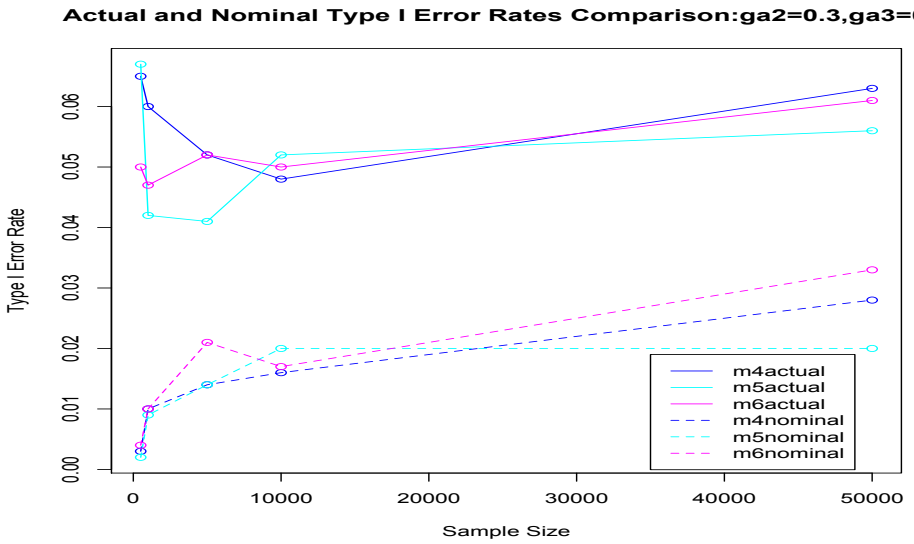
$1,000$ samples $X_1, X_2, \ldots, X_n$ from the nine corresponding null models. Then we determine how many times out of $1,000$ we reject H_0. Both nominal and actual Type I error rates are displayed in Figure 4.1. The models with $\gamma_2 = 0.2, \gamma_3 = 0.3$ are in the first panel. The models with $\gamma_2 = 0.3, \gamma_3 = 0.2$ are in the second panel and the models with $\gamma_2 = 0.25, \gamma_3 = 0.25$ are in the third panel. In all cases, the solid lines which represent the actual Type I error rates lay above the dashed lines which represent the nominal Type I error rates. This simulation result is consistent with the conclusion we get from the previous paragraph.

The power comparisons for nine models are presented in Figure 4.2. Power is calculated as the number of unilateral null hypothesis rejections divided by $1,000$ under the alternative hypothesis. Solid lines are for the actual power calculated from simulated critical values while dashed lines are for the nominal power calculated from the nominal critical value 1.645. All solid lines are above dashed lines showing that the test based on simulated critical values is more powerful. As the sample size gets to $50,000$, the gaps between actual power curves and nominal power curves become smaller. However, both types of powers hardly achieve 0.8, which reveals the weakness of our moment-based procedure.

It is mentioned in the second chapter that resampling is not required and there is no compactness restriction on the parameter space in the moment-based procedure for CCS model. Likewise, it is also true for CCS+EC model. But the power problem motivated us to investigate likelihood-based test for CCS+EC model.

Figure 4.1: Type I Error Rates Based on Actual and Nominal Critical Values





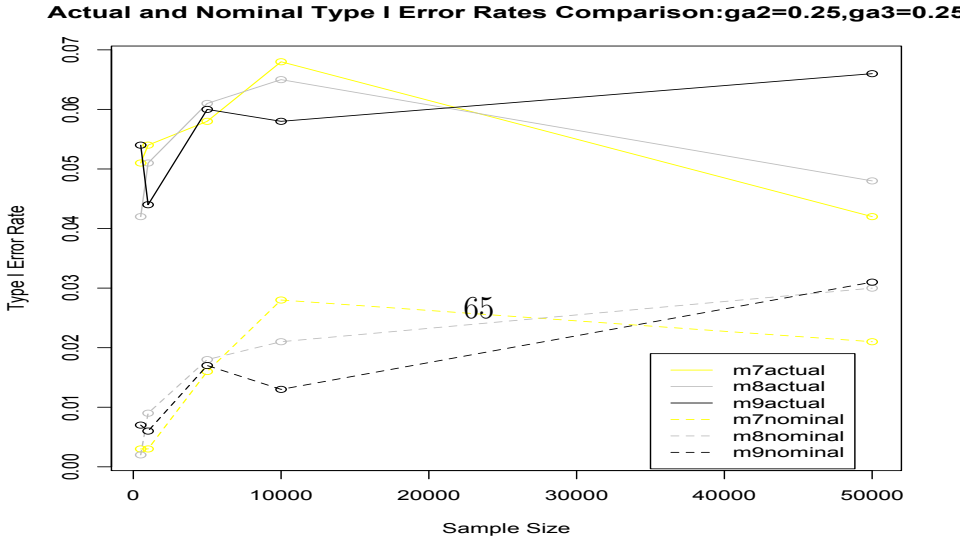
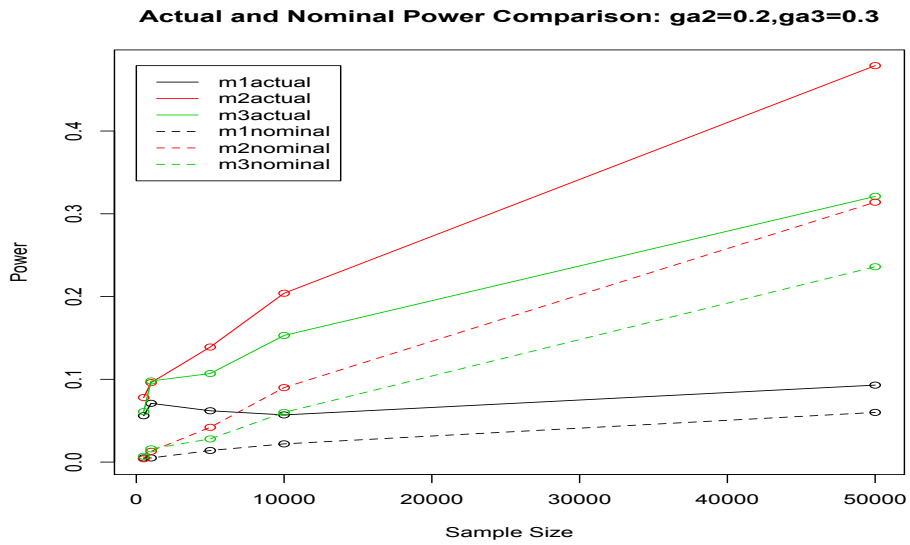
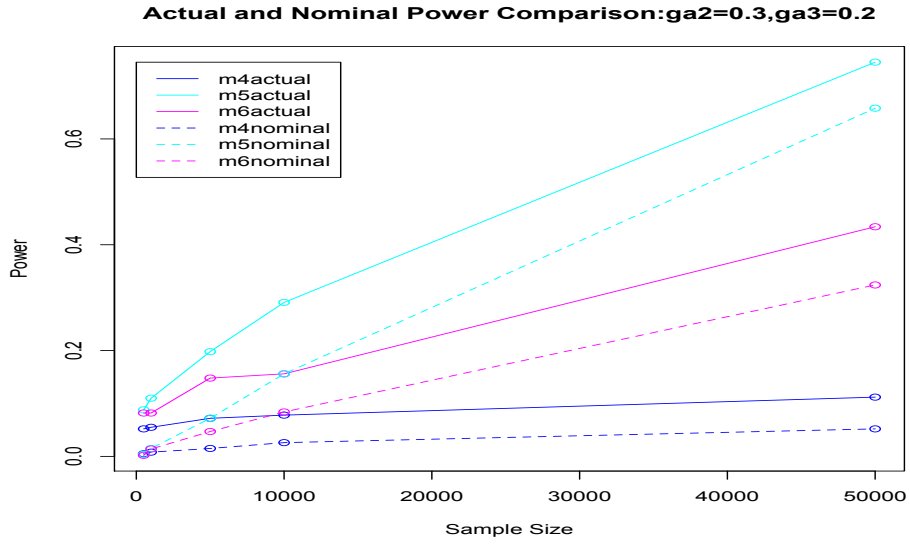
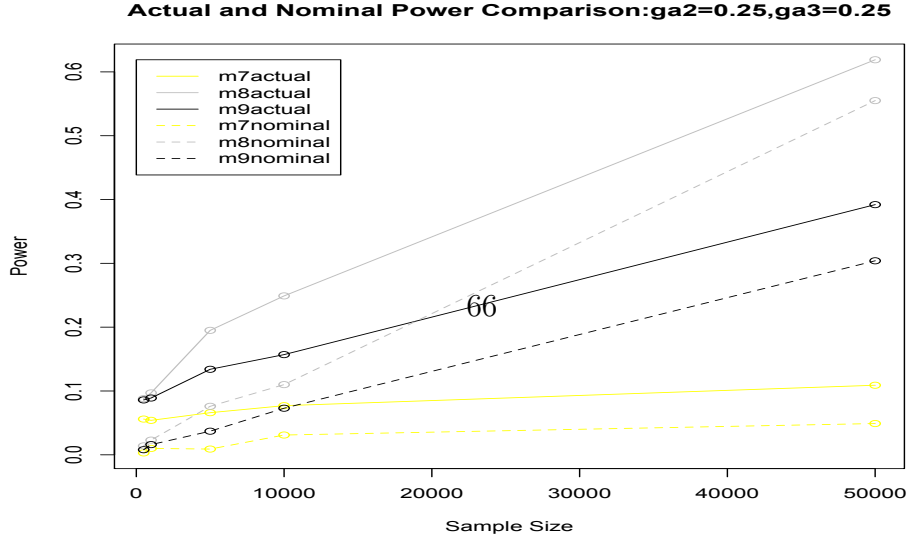


Figure 4.2: Powers Based on Actual and Nominal Critical Values







4.4 A Case Study for Moment-Based Test

In section 2.5, the moment-based test was applied to the microarray experiments on 8,799 genes from three groups of rats by Blalock and colleagues (2003) [5] to test if there existed a differential expression. The result showed that the genes were differentially expressed in the situations that all genes are considered, genes remaining after each step and the genes eliminated in step 1 and step 2. For the genes eliminated in step 3, the P-value is greater than 0.05, which means we fail to reject the omnibus null hypothesis. So we refrain from testing CCS model versus testing CCS+EC model for the genes eliminated in the third step. For other steps, we test whether the heterogeneity of differential expression is from one central Chi-Square distribution and one non-central Chi-Square distribution or from one central Chi-Square distribution and two different types of non-central Chi-Square distributions by using the moment-based test for CCS+EC model.

We conduct the case study under both actual critical value and nominal critical value. The two kinds of P-values are presented in Table 4.2. The P-values under nominal critical values in the second column are defined by $1 - \Phi(Z_n)$, when Z_n is the observed test statistic while the P-value under the actual critical value in the fourth column is calculated by the proportion of simulated test statistics under the unilateral null model greater than the observed test statistic. For all 8,799 genes, genes remaining after each step, both types of P-values reject the unilateral null hypothesis, indicating that there is an extra component of differential expression. For genes eliminated in the second step, both types of P-values fail to reject the unilateral null hypothesis which means the CCS model fits to the data adequately.

For genes eliminated in the first step, we fail to reject H_0 under the nominal critical value while we reject H_0 under the actual critical value but with a P-value very close to 0.05.

The moment-based method provides us a convenient procedure for testing

Table 4.2: Hypothesis Testing Results Based on Actual and Nominal Critical Values

Genes	Nominal P-values	Nominal Results	Actual P-values	Actual results
All genes	0.00015	Reject H_0	0	Reject H_0
Remaining after 1A	0.00013	Reject H_0	0	Reject H_0
Eliminated in 1A	0.1185	Fail to reject H_0	0.049	Reject H_0
Remaining after 2A	0.0002	Reject H_0	0	Reject H_0
Eliminated in 2A	0.1836	Fail to reject H_0	0.11	Fail to reject H_0
Remaining after 3A	0.0005	Reject H_0	0	Reject H_0

whether a central Chi-Square distribution is contamined by one or by two other non-central Chi-Square distributions. As exemplified in the hippocampal tissue example, the test can be applied to further assess whether the differetiatial expression levels are the same. Basically, the idea of the test is to see whether a quadratic function of the first three sample moments is sufficiently large. Therefore, resampling is not necessary if one is willing to use conservative nominal critical values. Moreover, through out our test, no compactness of parameter space is assumed, which is another advantage of the moment-based test. However, from the simulation study, the power is weak, which is a common limitation of moment-based tests. Besides, one may notice that we did not estimate the parameters when the unilateral null hypothesis was rejected in our case study. It has been stated that the method of moments may not work well when there are high dimensional parameter equations. For example, the first three moments are needed to get the test statistic for CCS+EC model. But we require the first six moments to get the variance covariance matrix.

For even higher dimensional statistical model, too many estimators for the moments would introduce more variance. A better way to estimate the parameters is to use an EM algorithm to approximate maximum likelihood for parameter estimation. We will next focus on the EM test for CCS+EC model.

4.5 EM Test for CCS+EC Model

In Chapter 3 we proposed an EM test for inference in the Contaminated Chi-Square model, motivated partly because the moment-based approach is problematic for complicated statistical models such as large mixtures of high-dimensional distributions. This problem becomes even apparent when an extra component is added to the Contaminated Chi-Square model. In section 4.4, we did not estimate the parameters when the unilateral null hypothesis was rejected because the parameter estimation of the moment-based method is apt to fail for complicated mixture models. In this section, we will look for an EM test for the CCS+EC model such that: (i) the asymptotic null distribution of the EM test statistic is analytically tractable; (ii) actual Type I error rates in finite samples are close to the corresponding nominal significance levels; and, (iii) Type II error rates in finite samples are low. Furthermore, the development of the EM test will yield as by-products parameter estimators for the CCS+EC model.

The rest of the chapter is organized as follows. Section 4.5 states the null hypothesis, the alternative hypothesis and the test statistic from the EM test for CCS+EC model. Section 4.6 investigates the asymptotic behavior of the proposed EM test under the unilateral null hypothesis. Section 4.7 presents simulation study results

and section 4.8 further analyzes whether the hipppocampal data by Blalock and colleagues (2003) [5] come from three different chi-square distributions. Section 4.8 also addresses the issue of parameter estimation.

Let

$$f(x;\Psi) = \gamma f(x;\mu_1) + (1-\gamma) f(x;\mu_2) = \int_{\Theta} f(x;\mu) d\Psi(\mu), \tag{4.7}$$

where $f(x;\mu)$ is a Chi-Square density function with noncentrality parameter μ on ν degrees of freedom, $\Psi(.)$ is defined as a mixing distribution in the form of

$$\Psi(\mu) = (1-\gamma) I(\mu_1 \leq \mu) + \gamma I(\mu_2 \leq \mu). \tag{4.8}$$

$I(.)$ is the indicator function, (μ_1, μ_2) are the nonnegative mixing parameters from the parameter space $\Theta := [0, \infty)$, and $\gamma \in [0, 1]$ is the mixing proportion.

An interesting problem remains from the third chapter is that, if the omnibus null hypothesis is rejected, (i.e., assume $\mu_1 = 0$ and reject $\gamma\mu_2 = 0$ in favor of $\gamma\mu_2 \neq 0$) are the differentially expressed genes truly arising from one noncentral Chi-Square distribution? It is also possible that the differential expression may come from two or more non-central Chi-Square distributions. The interpretation is that some genes may be highly over or under expressed while others may be only modestly over or under expressed. Here, we consider the situation whether there is one extra component to the CCS model.

Given a random sample $X_1, X_2, \ldots, X_n$ from $f(x;\Psi)$ with Ψ defined by $\Psi(\mu) = \sum_{j=1}^{m} \gamma_j I(\mu_j \leq \mu)$, the unilateral null hypothesis is that

$$H_0 : m = 2 \text{ vs. } H_A : m > 2, \tag{4.9}$$

where m is the number of components in the most parsimonious representation of the Contaminated Chi-Square model. $m = 2$ denotes CCS model we have already

studied while $m = 3$ denotes what we will hereafter call the CCS+EC model. When the unilateral null hypothesis is true,

$$f(x; \Psi_0) = (1 - \gamma) f(x; 0) + \gamma f(x; \mu),$$

with $\Psi_0(\mu^*) = (1-\gamma) I(0 \leq \mu^*) + \gamma I(\mu \leq \mu^*)$. μ^* may be regarded as a latent random variable.When the unilateral null hypothesis is false, since 0 is a known boundary point, we assume the parameter μ will split into μ_2 and μ_3, so we are in essence assuming $m = 3$ under the alternative. In Li and Chen (2010) [16], because they have two unknown parameters, they assume the alternative is $m = 2m_0$, where m_0 is an arbitrarily given positive integer specified for the null hypothesis. They state that any finite mixture model with $m_0 < m < 2m_0$ can be expressed as a mixture model of order $m = 2m_0$. The EM test would be less efficient if the data are from a model with $m > 2m_0$.

The log likelihood function is

$$l_n(\Psi) = \sum_{1}^{n} \log f(X_i; \Psi). \tag{4.10}$$

Let $\hat{\Psi}_0$ be the maximum likelihood estimator (MLE) of Ψ under the unilateral null hypothesis such that

$$\hat{\Psi}_0 = (1 - \hat{\gamma}) I(0 \leq \mu^*) + \hat{\gamma} I(\hat{\mu} \leq \mu^*). \tag{4.11}$$

The EM test statistic uses the EM algorithm to iteratively obtain the estimates and the test statistic. The following is the procedure to get the EM test statistic.

Step 1 Assuming that under the null hypothesis the MLE $\hat{\Psi}_0$ has been obtained using EM algorithm, then the parameter space Θ was divided into 2 intervals with $I_1 = (0, \eta]$ and $I_2 = [\eta, +\infty)$.

Step 2 For each $\beta \in (0,1)$, create a class of mixture distributions of order 3:

$$\Omega_3(\beta) = \{(1-\gamma)I(0 \le \mu^*) + \gamma\beta I(\mu_2 \le \mu^*) + \gamma(1-\beta)I(\mu_3 \le \mu^*) : \mu_2, \mu_3 \in I_2\},$$

where $\beta \in (0,1)$. Define the modified log-likelihood function to be

$$pl_n(\Psi) = l_n(\Psi) + p(\beta), \tag{4.12}$$

where $p(\beta)$ is a continuous penalty function maximized at 0.5 and goes to negative infinity as β goes to 0 or 1.

Step 3 J numbers of values for β are selected from $(0,1)$ to be $\boldsymbol{\beta}$, then $\boldsymbol{\beta}^2$ contains J numbers of β. Compute $\Psi^{(k)}(\beta^{(k)}) = argmax_{\gamma,\mu_2,\mu_3}\{pl_n(\Psi) : \Psi \in \Omega_3(\beta^{(k)})\}$. Define $\beta^{(1)} = \beta_0$.

Step 4 Get the conditional expectation $w_{i1h}^{(k)}$ and $w_{i2h}^{(k)}$ in E-step for each $i = 1, 2, \ldots, n$ and $h = 1, 2$. Let

$$\begin{aligned}
w_{i11}^{(k)} &= \frac{(1-\gamma^{(k)})f(X_i;0)}{f(X_i;\Psi^{(k)}(\beta_0))} \\
w_{i12}^{(k)} &= \frac{\gamma^{(k)}\beta^{(k)}f(X_i;\mu_2^{(k)})}{f(X_i;\Psi^{(k)}(\beta_0))} \\
w_{i21}^{(k)} &= 0 \\
w_{i22}^{(k)} &= \frac{\gamma^{(k)}(1-\beta^{(k)})f(X_i;\mu_3^{(k)})}{f(X_i;\Psi^{(k)}(\beta_0))}
\end{aligned}$$

Maximize the approximation to the complete data penalized log likelihood in M-step. Let

$$\gamma^{(k+1)} = n^{-1}\sum_{i=1}^{n}\{w_{i12}^{(k)} + w_{i22}^{(k)}\},$$

$$\mu_{j2}^{(k+1)} = argmax_{\mu}\{\sum_{i=1}^{n} w_{ij2}^{(k)} \log f(X_i;\mu)\}, j = 1, 2$$

and

$$\beta^{(k+1)} = argmax_{\beta}\{\sum_{i=1}^{n} w_{i12}^{(k)} \log(\beta) + \sum_{i=1}^{n} w_{i22}^{(k)} \log(1-\beta) + p(\beta)\}.$$

Compute

$$M_n^{(k)}(\beta_0) = 2\{pl_n(\Psi^{(k)}(\beta_0)) - l_n(\hat{\Psi}_0)\}.$$

Let $k = k+1$ and repeat Step 4 until $k = K$,which is a prespecified number of times.

Then the EM test statistic is defined as

$$EM_n^{(K)} = \max\{M_n^{(K)}(\beta_0) : \beta_0 \in \boldsymbol{\beta}^2\}. \tag{4.13}$$

The unilateral null hypothesis is rejected when $EM_n^{(K)}$ is greater than some critical value, which is determined by its limiting null distribution.

4.6 Asymptotic Behavior of EM Test Statistic

Before stating the lemmas and theorems, six regularity conditions are listed. These conditions are adapted from Li and Chen (2010) [30]. All lemmas and theorems in this Chapter are built based on the following regularity conditions. The proofs that these six regularity conditions are statisfied for the CCS model appear in Appendix II.

Condition 0. The penalty term $p(\beta)$ is a continuous function such that it is maximized at $\beta = 0.5$ and goes to negative infinity as β goes to 0 or 1.

Condition 1. The kernel function $f(X;\mu)$ is such that the mixture distribution $f(X;\mu)$ satisfies Wald's integrability conditions for consistency of the maximum likelihood estimator. For this, it suffices to require that

- $E|\log f(X;\Psi_0)| < \infty$.
- for sufficiently small ρ and for sufficiently large r, $E\log\{1+f(X;\mu,\rho)\} < \infty$ for $\mu \in \Theta$ and $E\log\{1+\phi(X;r)\} < \infty$, where $f(X;\mu,\rho) = \sup_{|\mu'-\mu|\leq\rho} f(X;\mu')$, $\phi(X;r) = \sup_{\mu\geq r} f(X;\mu)$ and $\Psi_0 := (1-\gamma_0)I(0 \leq \mu) + \gamma_0 I(\mu_0 \leq \mu)$.

Condition 2. The kernel function $f(X;\mu)$ has common support and is four times continuously differentiable with respect to μ.

Condition 3. For any two mixing distribution functios Ψ_1 and Ψ_2 such that $\int f(x;\mu)d\Psi_1(\mu) = \int f(x;\mu)d\Psi_2(\mu)$ for all x, we must have $\Psi_1 = \Psi_2$.

Condition 4. Let $N(\mu.\epsilon) = \{\mu' \in \Theta : |\mu'-\mu| \leq \epsilon\}$ for some positive ϵ. There exists an intergrable $g(.)$ and a small positive ϵ_0 such that $|\Delta_{ih}|^3 \leq g(X_i)$, $|Y_i(\mu)|^3 \leq g(X_i)$, $|Z_i^{(k)}(\mu)| \leq g(X_i)$, for $\mu \in N(\mu_{0h},\epsilon_0), h = 1,2$, and $k = 0,1,2$ with $Z_1^{(k)}(\mu)$ being the kth derivative, where $\Delta_{i1} = \frac{f(x_i;0)-f(x_i;\mu)}{f(x_1;\Psi_0)}, Y_i(\mu) = \frac{f'(x_i;\mu)}{f(x_i;\Psi_0)}, Z_i(\mu) = \frac{f''(x_i;\mu)}{2f(x_i;\Psi_0)}$.

Condition 5. The variance-covariance matrix $\mathbf{B}$ of $b_i = (\delta_{i1}, Y_i(0), Z_i(0), Z_i(\mu))^T$ is positive definite.

Theorem 9. *Let $f(x;\mu)$ be the Chi-Square distribution with noncentrality parameter μ on ν degrees of freedom. Let $p(\beta)$ be the penalty term satisfies the regularity conditions given in the Appendix II, and that $\Psi^{(k)}(\beta^{(k)})$ is as specified in Step 3. Under*

the null distribution $f(x;\Psi_0)$*, and for each given* $\beta_0 \in \boldsymbol{\beta}^2$*, we have*

$$\gamma^{(k)}-\gamma_0 = O_p(n^{-1/2}), \beta^{(k)}-\beta_0 = O_p(n^{-1/6}), \mu_2^{(k)}-\mu = O_p(n^{-1/4}), \mu_3^{(k)}-\mu = O_p(n^{-1/4}),$$

and $\mathbf{m_1}^{(k)} := \begin{pmatrix} m_{11}^{(k)} \\ m_{12}^{(k)} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta^{(1)}(\mu_2^{(1)}-\mu) + (1-\beta^{(1)})(\mu_3^{(1)}-\mu) \end{pmatrix} = O_p(n^{-1/2}).$

Notice that $\gamma^{(1)}, \beta^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}$ *are the results from the first iteration of EM algorithm.*

The proof of Theorem 9 consists of Lemma 8 to Lemma 12. Lemma 8 shows that the estimators $\gamma^{(k)}, \beta^{(k)}, \mu_2^{(k)}, \mu_3^{(k)}$ converge in probability to the true parameter under the null hypothesis when $k = 1$. Lemma 9 then strengthens Lemma 8 by providing the asymptotic orders. Lemma 10 shows that the convergence in probability of $\beta^{(k)}$ is also true for $\beta^{(k+1)}$. Lemma 11 proves that the large-sample properties of $\gamma^{(1)}, \beta^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}$ hold for $\gamma^{(k+1)}, \beta^{(k+1)}, \mu_2^{(k+1)}, \mu_3^{(k+1)}$. Lemma 12 completes the proof of Theorem 9 by providing the asymptotic orders. The proofs follow a similar structure to Li and Chen (2010) [30] but exhibit important differences that lead to a different limiting null distribution.

Lemma 8. *Let* $f(x;\mu)$ *be the Chi-Square distribution with noncentrality parameter* μ *on* ν *degrees of freedom. Let* $p(\beta)$ *be the penalty term that satisfies the regularity conditions C0-C3. Under the null distribution* $f(x;\Psi_0)$*, and for each* $\beta_0 \in \boldsymbol{\beta}^2$*, we have*

$$\gamma^{(k)} - \gamma_0 = o_p(1), \beta^{(k)} - \beta_0 = o_p(1), \mu_2^{(k)} - \mu = o_p(1), \mu_3^{(k)} - \mu = o_p(1).$$

Proof. For any given $\beta_0 \in \boldsymbol{\beta}^2$, we define

$$\hat{\Psi}_0 := (1-\gamma^{(k)})I(0 \le \mu^*) + \gamma^{(k)}\beta_0 I(\mu_2^{(k)} \le \mu^*) + \gamma^{(k)}(1-\beta_0)I(\mu_3^{(k)} \le \mu^*). \quad (4.14)$$

Hence $\hat{\Psi}_0 \in \Omega_3(\beta_0)$.

It is obvious that

$$pl_n(\Psi^{(1)}) = l_n(\Psi^{(1)}) + p(\beta_0). \geq pl_n(\hat{\Psi}_0) \geq l_n(\Psi_0) + p(\beta_0).$$

The first inequality is by the definition of $\Psi^{(1)}$, the second inequality is by the definition of MMLE. Therefore,

$$l_n(\Psi^{(1)}) \geq l_n(\Psi_0). \tag{4.15}$$

By Wald's result of the consistency of the maximum likelihood estimator (1949) [42], based on a set of iid observations from a distribution family parameterized by Ψ, if $\hat{\Psi}_n$ is an estimator of Ψ such that

$$l_n(\hat{\Psi}_n) \geq l_n(\Psi_0) - o(n)$$

as $n \to \infty$, then $\hat{\Psi}_n \to \Psi_0$. (Li and Chen 2010 [30]). Combine this result and (4.15), we can conclude that

$$\|\Psi^{(1)}(\beta_0) - \Psi_0\| = \int |\hat{\Psi}_n - \Psi_0| \exp^{-\mu} d\mu \to 0 \ a.s. \tag{4.16}$$

Therefore, the consistency result of (4.16) is possible only if the conclusions of the lemma are true. □

Lemma 9. *Let $f(x;\mu)$ be the Chi-Square distribution with noncentrality parameter μ on ν degrees of freedom. Let $p(\beta)$ be the penalty term that satisfies the regularity conditions C0-C5. Under the null distribution $f(x;\Psi_0)$, and for each $\beta_0 \in \boldsymbol{\beta}^2$, we have*

$$\gamma^{(1)} - \gamma_0 = O_p(n^{-1/2}), \beta^{(1)} - \beta_0 = O_p(n^{-1/6}), \mu_2^{(1)} - \mu = O_p(n^{-1/4}), \mu_3^{(1)} - \mu = O_p(n^{-1/4}).$$

$$\mathbf{m_1}^{(1)} := \begin{pmatrix} m_{11}^{(1)} \\ m_{12}^{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ \beta^{(1)}(\mu_2^{(1)} - \mu) + (1-\beta^{(1)})(\mu_3^{(1)} - \mu) \end{pmatrix} = O_p(n^{-1/2}).$$

Proof. Let $R_{1n}(\Psi^{(1)}(\beta_0)) = 2\{pl_n(\Psi^{(1)}) - l_n(\Psi_0)\} = 2\{l_n(\Psi^{(1)}) - l_n(\Psi_0)\} + 2p(\beta_0)$. It is obvious that a lower bound of R_{1n} is $2p(\beta_0)$. Now we need to find an upper bound of R_{1n}.

Since the penalty is negative, we have

$$\begin{aligned} R_{1n}(\Psi^{(1)}(\beta_0)) &\leq 2\{l_n(\Psi^{(1)}(\beta_0)) - l_n(\Psi_0)\} \\ &= 2\sum_{i=1}^{n} \log\left(1 + \frac{f(x_i;\Psi^{(1)}(\beta_0)) - f(x_i;\Psi_0)}{f(x_i;\Psi_0)}\right) \end{aligned}$$

Let

$$\begin{aligned} \delta_i &= \frac{f(x_i;\Psi^{(1)}(\beta_0)) - f(x_i;\Psi_0)}{f(x_i;\Psi_0)} \\ &= (\gamma^{(1)} - \gamma_0)\frac{f(x_i;\mu) - f(x_i;0)}{f(x_i;\Psi_0)} + \gamma^{(1)}\{\beta^{(1)}\frac{f(x_i;\mu_2^{(1)}) - f(x_i;\mu)}{f(x_i;\Psi_0)} + (1-\beta^{(1)})\frac{f(x_i;\mu_3^{(1)}) - f(x_i;\mu)}{f(x_i;\Psi_0)}\} \end{aligned}$$

By Taylor's expansion,

$$f(X_i;\mu_{j2}^{(1)}) - f(X_i;\mu) = (\mu_{j2}^{(1)} - \mu)f'(X_i;\mu) + \frac{1}{2}(\mu_{j2}^{(1)} - \mu)f''(X_i;\mu) + \epsilon_{i2},$$

where $j = 1, 2$, ϵ_{i2} is the remainder term, and differentiation is with respect to the parameters. With Taylor's expansion, we have

$$\delta_i = (\gamma^{(1)} - \gamma^{(0)})\Delta_{i1} + \gamma^{(1)}m_{12}^{(1)}Y_i(\mu) + \gamma^{(1)}m_{22}^{(1)}Z_i(\mu) + \epsilon_{i2},$$

where $\Delta_{i1} = \frac{f(X_i;0) - f(X_i;\mu)}{f(X_1;\Psi_0)}$, $Y_i(\mu) = \frac{f'(X_i;\mu)}{f(X_i;\Psi_0)}$, $Z_i(\mu) = \frac{f''(X_i;\mu)}{2f(X_i;\Psi_0)}$

$m_{12}^{(1)}$ has been defined in Theorem 9,

$m_{22}^{(1)} = \beta^{(1)}(\mu_2^{(1)} - \mu)^2 + (1-\beta^{(1)})(\mu_3^{(1)} - \mu)^2$,

and $\epsilon_{i2} = \gamma^{(1)}[\beta^{(1)}(\mu_2^{(1)} - \mu)^3 U_{i2}(\mu_2^{(1)}) + (1-\beta^{(1)})(\mu_3^{(1)} - \mu)^3 U_{i2}(\mu_3^{(1)})]$,
with $U_{i2}(\mu^*) = \frac{f(X_i;\mu^*) - f(X_i;\mu) - f'(X_i;\mu)(\mu^*-\mu) - \frac{f''(X_i;\mu)(\mu^*-\mu)^2}{2}}{f(X_i;\Psi_0)(\mu^*-\mu)^3}$.

Hence,

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} [(\gamma - \gamma^{(1)})\Delta_{i1} + \gamma^{(1)}(m_{12}^{(1)} Y_i(\mu)) + \gamma^{(1)}(m_{22}^{(1)} Z_i(\mu))] + \epsilon_n,$$

where

$$\begin{aligned}
|\epsilon_n| &= |\sum_{i=1}^{n} \epsilon_{i2}| \\
&\leq n^{1/2}\gamma^{(1)}[\beta^{(1)}(|\mu_2^{(1)} - \mu|)^3\{n^{-1/2}\sum_{i=1}^{n} U_{i2}(\mu_2^{(1)})\} + (1-\beta^{(1)})(|\mu_3^{(1)} - \mu|)^3\{n^{-1/2}\sum_{i=1}^{n} U_{i2}(\mu_3^{(1)})\}] \\
&= O_p(n^{1/2})[\beta^{(1)}(\mu_2^{(1)} - \mu)^3 + (1-\beta^{(1)})(\mu_3^{(1)} - \mu)^3] \\
&= o_p(n^{1/2}) m_{22}^{(1)} \\
&\leq o_p(1) + o_p(n)(m_{22}^{(1)})^2
\end{aligned}$$

Let

$$\bar{\mathbf{t}} = (\gamma - \gamma^{(1)}; 0, \gamma^{(1)} m_{12}^{(1)}; 0, \gamma^{(1)} m_{22}^{(1)})^T.$$

With the notations above, $\sum_{i=1}^{n} \delta_i$ can be simplified as

$$\sum_{i=1}^{n} \delta_i = \sum_{i=1}^{n} \bar{\mathbf{t}}^{\mathbf{T}} \mathbf{b_i} + \epsilon_n.$$

Using Taylor's expansion, $\log(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}$, then

$$\begin{aligned}
R_{1n}(\Psi^{(1)}(\beta_0)) &\leq 2\sum_{i=1}^{n}(1+\delta_i) \\
&\leq 2\sum_{i=1}^{n}\delta_i - \sum_{i=1}^{n}\delta_i^2 + \frac{2}{3}\sum_{i=1}^{n}\delta_i^3 \\
&= 2\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i}) - \sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i})^2 + \frac{2}{3}\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i})^3 + O_p(\epsilon_n).
\end{aligned}$$

Since $\mathbf{B}$ is positive definite, and by the law of large number, we have

$$\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i})^2 = n\bar{\mathbf{t}}^T\mathbf{B}\bar{\mathbf{t}}\{1+o_p(1)\},$$

$$\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i})^3 = o_p(n)\bar{\mathbf{t}}^T\bar{\mathbf{t}},$$

and

$$\epsilon_n = o_p(1) + o_p(n)\bar{\mathbf{t}}^T\bar{\mathbf{t}}.$$

Hence, the upper bound of $R_{1n}(\Psi^{(1)}(\beta_0))$ becomes

$$R_{1n}(\Psi^{(1)}(\beta_0)) \leq 2\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i}) - n\bar{\mathbf{t}}^T\mathbf{B}\bar{\mathbf{t}}\{1+o_p(1)\} + o_p(1). \tag{4.17}$$

We know $R_{1n}(\Psi^{(1)}(\beta_0))$ is bounded below by $2p(\beta_0)$ and bounded above by $2\sum_{i=1}^{n}(\bar{\mathbf{t}}^{\mathbf{T}}\mathbf{b_i}) - n\bar{\mathbf{t}}^T\mathbf{B}\bar{\mathbf{t}}\{1+o_p(1)\}+o_p(1)$. The inequality is possible only if $\bar{\mathbf{t}} = O_p(n^{-1/2})$. Since $\beta^{(1)}$ is defined to be equal to β_0, it is obvious that $\beta^{(1)} - \beta_0 = O_p(n^{-1/6})$. Consequently, $\Psi^{(1)}(\beta_0)$ must have the order claimed in this Lemma. □

Lemma 10. *Let $f(x;\mu)$ be the Chi-Square distribution with noncentrality parameter μ on ν degrees of freedom. Let $p(\beta)$ be the penalty term satisfies the regularity conditions C0-C5. If for some $k \geq 1$, under the null distribution $f(x;\Psi_0)$, and for each $\beta_0 \in \boldsymbol{\beta}^2$,*

$$\gamma^{(k)} - \gamma_0 = O_p(n^{-1/2}), \beta^{(k)} - \beta_0 = O_p(n^{-1/6}), \mu_2^{(k)} - \mu = O_p(n^{-1/4}), \mu_3^{(k)} - \mu = O_p(n^{-1/4}),$$

and $\mathbf{m}_1^{(k)} = O_p(n^{-1/2})$, then we have $\beta^{(k+1)} - \beta_0 = O_p(n^{-1/6})$.

Proof. Before proving this lemma, we first prove the equation below that will be applied to the lemma.

$$\sum_{i=1}^{n}\frac{f(X_i;\mu_{1h}^{(k)})}{f(X_i;\Psi^{(k)}(\beta_0))} = n\{1+O_p(n^{-1/6})\}, \tag{4.18}$$

for $h = 1$, $\mu_{11}^{(k)} = 0$, for $h = 2$, $\mu_{12}^{(k)} = \mu_2^{(k)}$. For $j = 1, 2$,

$$\frac{f(X_i; \mu_{j2}) - f(X_i; \mu)}{f(X_i; \Psi_0)} = Y_i(\mu)(\mu_{j2} - \mu) + Z_i(\tilde{\mu}_{j2})(\mu_{j2} - \mu)^2, \tag{4.19}$$

where $\mu_{12} = \mu_2, \mu_{22} = \mu_3$, $\tilde{\mu}_{j2}$ is a number in a small neighborbood of μ. Hence, we have,

$$\begin{aligned}\delta_i &= \frac{f(X; \Psi^{(k)}(\beta_0)) - f(X_i; \Psi_0)}{f(X_i; \Psi_0)} \\ &= (\gamma^{(k)} - \gamma_0)\Delta_{i1} + \gamma^{(k)} m_{12}^{(k)} Y_i(\mu) + \gamma^{(k)}\beta^{(k)}(\mu_2^{(k)} - \mu)^2 Z_i(\tilde{\mu_2}) + \gamma^{(k)}(1 - \beta^{(k)})(\mu_3^{(k)} - \mu)^2 Z_i(\tilde{\mu_3}).\end{aligned}$$

Because Δ_{i1} has a constant upper bound, we can conclude that

$$\max_{1 \le i \le n} |(\gamma^{(k)} - \gamma_0)\Delta_{i1}| = O_p(n^{-1/6}).$$

By the uniform condition in C4, we have

$$\max_{1 \le i \le n} |Y_i(\mu)| = O_p(n^{1/3}),$$

and for $j = 1, 2$,

$$\max_{1 \le i \le n} |Z_i(\tilde{\mu}_{jh}^{(k)})| = O_p(n^{1/3}).$$

In addition, because $m_{12}^{(k)} = O_p(n^{-1/2})$, the second term with $m_{12}^{(k)}$ must be $O_p(n^{-1/6})$ uniformly. According to the assumption of the lemma, we have the third term and the fourth term also $O_p(n^{-1/6})$. Thus, we can infer that $\max_{1 \le i \le n} |\delta_i| = O_p(n^{-1/6})$. Equivalently, uniformly in i,

$$\frac{f(X_i; \Psi_0)}{f(X_i; \Psi^{(k)}(\beta_0))} = 1 + O_p(n^{-1/6}).$$

Consequently, for $h = 2$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i;\mu_{1h}^{(k)})}{f(X_i;\Psi^{(k)}(\beta_0))} &= \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i;\mu_{1h}^{(k)})}{f(X_i;\Psi_0)}\times\frac{f(X_i;\Psi_0)}{f(X_i;\Psi^{(k)}(\beta_0))} \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{f(X_i;\mu_{1h}^{(k)})}{f(X_i;\Psi_0)}\{1+O_p(n^{-1/6})\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{f(X_i;\mu)}{f(X_i;\Psi_0)}+Y_i(\mu)(\mu_2^{(k)}-\mu)+Z_i(\tilde{\mu}_2)(\mu_2^{(k)}-\mu)^2\right\}\{1+O_p(n^{-1/6})\} \\
&= 1+O_p(n^{-1/6}).
\end{aligned}$$

This completes the proof of (4.18).

Next we need to show $\beta^{(k+1)}-\beta_0 = O_p(n^{-1/6})$. By definition, $\beta^{(k+1)}$ maximizes

$$Q_{nh}(\beta) = \sum_{i=1}^{n} w_{i1h}^{(k)}\log(\beta)+\sum_{i=1}^{n} w_{i2h}^{(k)}\log(1-\beta)+p(\beta)$$

with the weights $\sum_{i=1}^{n} w_{i1h}^{(k)}, \sum_{i=1}^{n} w_{i2h}^{(k)}$ for $h = 1, 2$ defined in Step 4.

Therefore, if letting

$$H_{nh}(\beta) = \sum_{i=1}^{n} w_{i1h}^{(k)}\log(\beta)+\sum_{i=1}^{n} w_{i2h}^{(k)}\log(1-\beta),$$

then $Q_{nh}(\beta)$ is maximized at

$$\hat{\beta} = \frac{\sum_{i=1}^{n} w_{i1h}^{(k)}}{n\gamma^{(k)}} = \beta^{(k)}\{1+O_p(n^{-1/6})\} = \beta_0\{1+O_p(n^{-1/6})\},$$

By using Taylor's expansion, if we constrain β^* within a very small neighborhood of $\hat{\beta}$, then

$$H_{nh}(\hat{\beta})-H_{nh}(\beta^*) \geq \epsilon\gamma^{(k)}n(\beta^*-\hat{\beta})^2$$

for some $\epsilon > 0$. In particular, if we set $|\beta^*-\hat{\beta}| \geq n^{-1/6}$, then

$$|\beta^*-\hat{\beta}| \geq \epsilon\gamma^{(k)}n^{2/3}$$

and therefore

$$Q_{n2}(\beta^*) - Q_{n2}(\hat{\beta}) \leq p(\beta^*) - p(\hat{\beta}) - \epsilon\gamma^{(k)}n^{2/3} < 0$$

when n is large enough. That is, the maximum point of Q_{n2} must be within an $O_p(n^{-1/6})$ neighborhood of $\hat{\beta}$. This completes the proof of the lemma. □

Lemma 11. *Assume the conditions of Lemma 10. Under the null distribution of* $f(x;\Psi_0)$*, we have*

$$\gamma^{(k+1)} - \gamma = o_p(1), \mu_2^{(k+1)} - \mu = o_p(1), \mu_3^{(k+1)} - \mu = o_p(1).$$

Proof. Define the mixing distribution obtained after a partial EM-iteration as

$$\Psi_2^{(k+1)}(\mu^*) = (1-\gamma^{(k)})I(0 \leq \mu^*) + \gamma^{(k)}\{\beta^{(k)}I(\mu_2^{(k+1)} \leq \mu^*) + (1-\beta^{(k)})I(\mu_3^{(k+1)} \leq \mu^*)\}.$$

Due to the property of EM-iteration that it always increases the likelihood, we get the following inequality

$$pl_n(\Psi_2^{(k+1)}) \geq pl_n(\Psi^{(k)}) \geq pl_n(\hat{\Psi}_0) = l_n(\hat{\Psi}_0) + p(\beta_0) \geq l_n(\Psi_0) + p(\beta_0).$$

Because the penalty term is less than or equal to zero,

$$l_n(\Psi_2^{(k+1)}) \geq l_n(\Psi_0) + p(\beta_0).$$

By classical result of Wald (1949) [42], this indicates the consistency of $\Psi_2^{(k+1)}$ for Ψ_0. Combined with the assumption that $\gamma^{(k)}, \beta^{(k)}, \mu_2^{(k)}, \mu_3^{(k)}$ are consistent, the consistency of $\Psi_2^{(k+1)}$ is possible only if we have $\mu_2^{(k+1)} = \mu + o_p(1)$ and $\mu_3^{(k+1)} = \mu + o_p(1)$. Plus the fact that $\beta^{(k+1)} = \beta_0 + o_p(1)$ in Lemma 10, the overall consistency of $\Psi^{(k+1)}$ implies that $\gamma^{(k+1)} = \gamma_0 + o_p(1)$. This completes the proof of Lemma 11. □

Lemma 12. *Assume the conditions of Lemma 10. Under the null distribution of* $f(x;\Psi_0)$*, we have*

$$\gamma^{(k+1)}-\gamma_0 = O_p(n^{-1/2}), \beta^{(k+1)}-\beta_0 = O_p(n^{-1/6}), \mu_2^{(k+1)}-\mu = O_p(n^{-1/4}), \mu_3^{(k+1)}-\mu = O_p(n^{-1/4}).$$

and $\mathbf{m}_1^{(k+1)} = O_p(n^{-1/2})$.

The proof of Lemma 12 can be achieved by following the proof of Lemma 9, replacing the first iteration with the $(k+1)$th iteration. And apply the result of Lemma 11.

Theorem 10. *Assume the same conditions as in Theorem 9. Under the null distribution* $f(x;\Psi_0)$*,and for any fixed finite* K*, as* $n \to \infty$*,*

$$EM_n^{(K)} = \sup_{\gamma m_{22}} \{2\gamma m_{22}(\sum_{i=1}^{n} \tilde{\mathbf{b}}_{\mathbf{2i}})_2 - n\gamma m_{22}(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}\gamma m_{22}\} + o_p(1), \tag{4.20}$$

with the new definitions

$$\mathbf{b_{1i}} = (\Delta_{i1}, Y_i(0), Y_i(\mu))^T,$$

and

$$\mathbf{b_{2i}} = (Z_i(0), Z_i(\mu))^T,$$

for $i = 1, 2, \ldots, n$.

For $j,k = 1,2$*, let the variance-covariance matrix* $\mathbf{B}_{jk} = E[\{\mathbf{b}_{ji} - E(\mathbf{b}_{ji})\}\{\mathbf{b}_{ki} - E(\mathbf{b}_{ki})\}^T]$*. Then we introduce* $\tilde{\mathbf{b}}_{\mathbf{2i}} = \mathbf{b_{2i}} - \mathbf{B_{21}}{\mathbf{B_{11}}}^{-1}\mathbf{b_{1i}}$ *by orhogonalizing* $\mathbf{b_{1i}}$ *and* $\mathbf{b_{2i}}$*. The variance-covariance matrix of* $\tilde{\mathbf{b}}_{\mathbf{2i}}$ *is defined as* $\tilde{\mathbf{B}}_{\mathbf{22}} = \mathbf{B_{22}} - \mathbf{B_{21}}{\mathbf{B_{11}}}^{-\mathbf{1}}\mathbf{B_{21}}$.

Proof. Let

$$\begin{aligned} M_n^{(k)}(\beta_0) &= 2\{pl_n(\Psi^{(k)}(\beta_0)) - l_n(\hat{\Psi}_0)\} \\ &= R_{1n}(\Psi^{(k)}(\beta_0)) - R_{0n}, \end{aligned}$$

where

$$R_{1n}(\Psi^{(k)}(\beta_0)) = 2\{pl_n(\Psi^{(k)}(\beta_0)) - l_n(\Psi_0)\},$$

and

$$R_{0n}(\Psi^{(k)}(\beta_0)) = 2\{l_n(\hat{\Psi}_0) - l_n(\Psi_0)\}.$$

Since R_{0n} is an ordinary LRT statistic under a regular model, an asymptotic approximation is

$$R_{0n} = (\sum_{i=1}^{n} \mathbf{b_{1i}})^T (n\mathbf{B_{11}})^{-1} (\sum_{i=1}^{n} \mathbf{b_{1i}}) + o_p(1).$$

Apply (4.17) here to get the upper bound of $R_{1n}(\Psi^{(k)}(\beta_0))$,

$$R_{1n}(\Psi^{(k)}(\beta_0)) \leq 2\sum_{i=1}^{n} \bar{\mathbf{t}}^T \mathbf{b_i} - n\bar{\mathbf{t}}^T \mathbf{B}\bar{\mathbf{t}}\{1 + o_p(1)\} + o_p(1),$$

with $\mathbf{b_i}$ defined in Condition 5.

Let

$$\mathbf{t_1} = (\gamma - \gamma_0, 0, \gamma m_{12})^T,$$

and

$$\mathbf{t_2} = (0, \gamma m_{22})^T.$$

Set $\mathbf{t} = (\mathbf{t_1}^T, \mathbf{t_2}^T)^T$. Then for any finite k, we find the upper bound of $M_n^{(k)}(\beta_0)$ is

$$R_{1n}(\Psi^{(k)}(\beta_0)) - R_{0n} \leq \sup_{\mathbf{t}} \left\{ 2\mathbf{t}^T \sum_{i=1}^{n} \mathbf{b_i} - n\mathbf{t}^T \mathbf{B}\mathbf{t} - (\sum_{i=1}^{n} \mathbf{b_{1i}})^T (n\mathbf{B_{11}})^{-1} (\sum_{i=1}^{n} \mathbf{b_{1i}}) + o_p(1) \right\}.$$

By the definition of $EM_n^{(k)}$, the upper bound above is also the one for $EM_n^{(k)}$. We then define $\tilde{\mathbf{t_1}}$ by orthogonalizing $\mathbf{t_1}$. Let $\tilde{\mathbf{t_1}} = \mathbf{t_1} + \mathbf{B_{11}}^{-11}\mathbf{B_{12}}\mathbf{t_2}$. We find the following equation

$$2\mathbf{t}^T \sum_{i=1}^{n} \mathbf{b_i} - n\mathbf{t}^T \mathbf{B}\mathbf{t} = 2\tilde{\mathbf{t}}_1^T (\sum_{i=1}^{n} \mathbf{b_{1i}}) - n\tilde{\mathbf{t}}_1^{\mathbf{T}} \mathbf{B_{11}} \tilde{\mathbf{t}}_1 + 2\mathbf{t_2}^T (\sum_{i=1}^{n} \mathbf{b_{2i}}) - n\mathbf{t_2^T} \tilde{\mathbf{B}}_{\mathbf{22}} \mathbf{t_2}.$$

Following Li and Chen (2010) [30],

$$EM_n^{(1)} \geq R_{1n}(\Psi^{(1)}(\beta_0)) - R_{0n} = \sup_{\mathbf{t_2}\geq 0}\{2\mathbf{t_2}^T(\sum_{i=1}^{n}\mathbf{b_{2i}}) - n\mathbf{t_2^T}\tilde{\mathbf{B}}_{\mathbf{22}}\mathbf{t_2}\} + o_p(1),$$

so that

$$\begin{aligned} EM_n^{(1)} &= \sup_{\mathbf{t_2}\geq 0}\{2\mathbf{t_2}^T(\sum_{i=1}^{n}\mathbf{b_{2i}}) - n\mathbf{t_2^T}\tilde{\mathbf{B}}_{\mathbf{22}}\mathbf{t_2}\} + o_p(1) \\ &= \sup_{\gamma m_{22}\geq 0}\{2\gamma m_{22}(\sum_{i=1}^{n}\mathbf{b_{2i}})_2 - n\gamma m_{22}(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}\gamma m_{22}\} + o_p(1) \\ &= \sup\{-n(\tilde{B}_{22})_{22}(\gamma m_{22} - \frac{(\sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}})_2}{n(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}})^2 + (\frac{(\sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}})_2^2}{n(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}})\}. \end{aligned}$$

By optimizing the quadratic equation, we can further conclude that

$$EM_n^{(1)} = \begin{cases} \frac{(\sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}})_2^2}{n(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}} + o_p(1), for \sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}} \geq 0 \\ 0 + o_p(1), for \sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}} < 0 \end{cases}$$

By the property of EM algorithm, $EM_n^{(K)} \geq EM_n^{(1)}$. Therefore,

$$EM_n^{(K)} = \begin{cases} \frac{(\sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}})_2^2}{n(\tilde{\mathbf{B}}_{\mathbf{22}})_{22}} + o_p(1), for \sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}} \geq 0 \\ 0 + o_p(1), for \sum_{i=1}^{n}\tilde{\mathbf{b}}_{\mathbf{2i}} < 0 \end{cases} \tag{4.21}$$

□

Theorem 11. *Assume the same conditions as in Theorem 9. Under the null distribution $f(x;\Psi_0)$ and for any fixed K, as $n \to \infty$,*

$$EM_n^{(K)} \xrightarrow{L} 0.5\chi_0^2 + 0.5\chi_1^2. \tag{4.22}$$

Proof. The asymptotic distribution of $EM_n^{(K)}$ can be concluded immediately from Theorem 10. □

4.7 Simulation Studies for EM Test

To assess the performance of EM test for CCS+EC model, a number of simulation studies were conducted. We evaluated the Type I error probability and the power of the EM testing procedure with sample sizes $n \in \{100, 500, 1000, 1500\}$. The EM test for CCS+EC model refers to the EM test with one iteration. Throughout our simulation studies in this section, the degree of freedom ν is set to be 3.

First, $1,000$ random samples $X_1, X_2, \ldots, X_n$ were generated from the null model

$$(1-\gamma)\chi^2_\nu(0) + \gamma\chi^2_\nu(\mu) \tag{4.23}$$

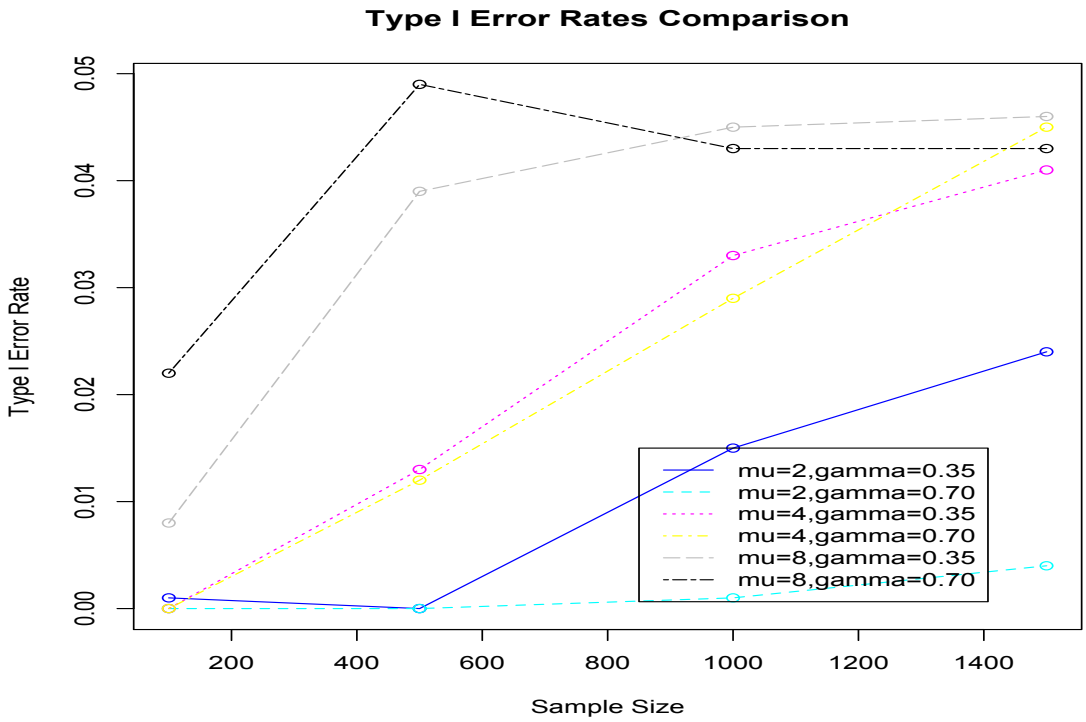
with parameters defined in section 4.5. Unber the null hypothesis, Type I error rates are defined as how many times out of 1,000 we reject H_0. We calculated the Type I error rates for 6 models with $\mu \in \{2, 4, 8\}$ and $\gamma \in \{0.35, 0.7\}$. These Type I error rates are displayed in the first panel of Figure 4.3. The Type I error rates started with numbers close to 0 for small sample sizes, while getting close to 0.05 when n is increased. However, the simulated Type I error rates for 6 models with $n \in \{100, 500, 1000, 1500\}$ are all below 0.05, indicating that the EM test for CCS+EC model is quite conservative.

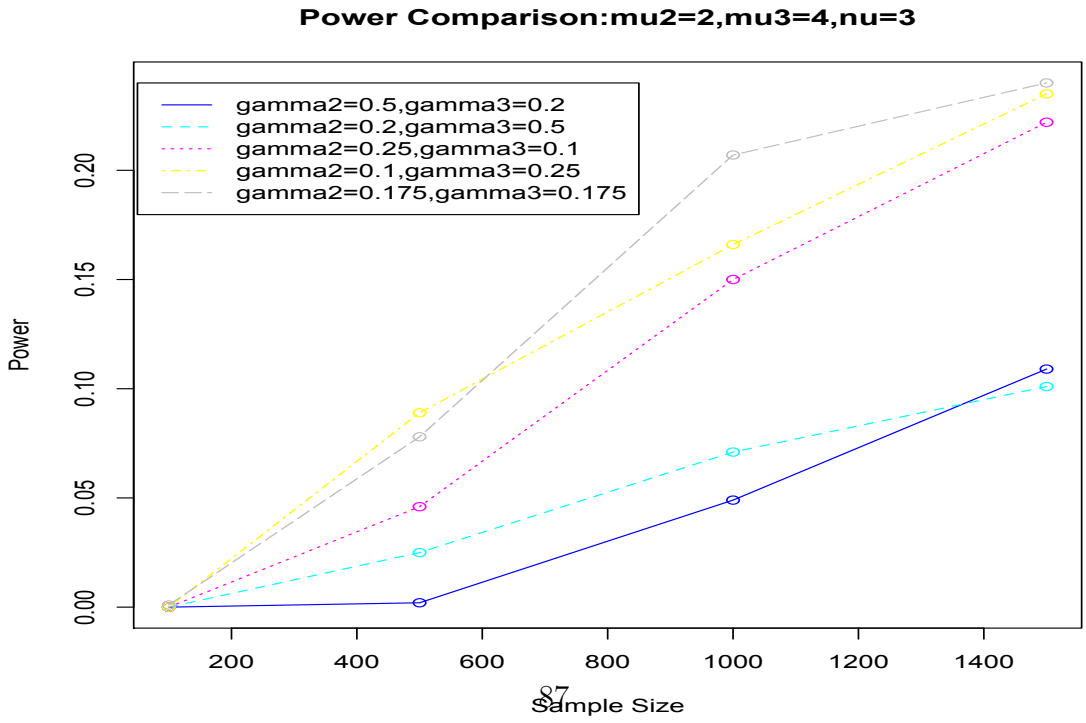
Next, we generated $1,000$ random samples $X_1, X_2, \ldots, X_n$ from the alternative model

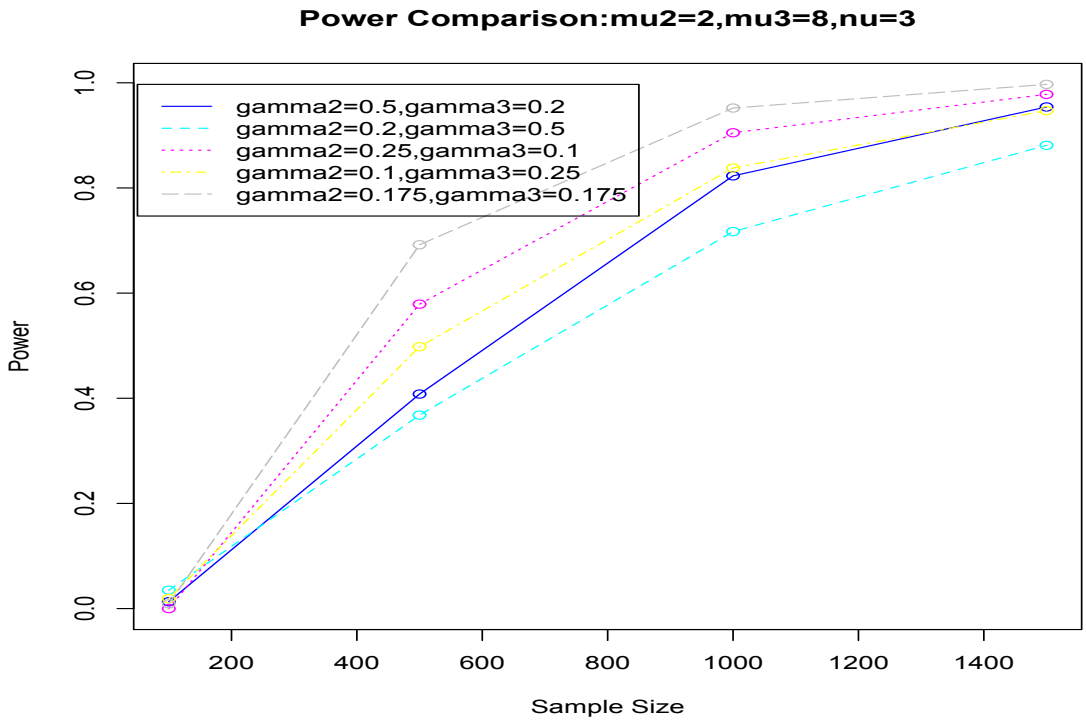
$$(1-\gamma)\chi^2_\nu(0) + \gamma\beta\chi^2_\nu(\mu_2) + \gamma(1-\beta)\chi^2_\nu(\mu_3) \tag{4.24}$$

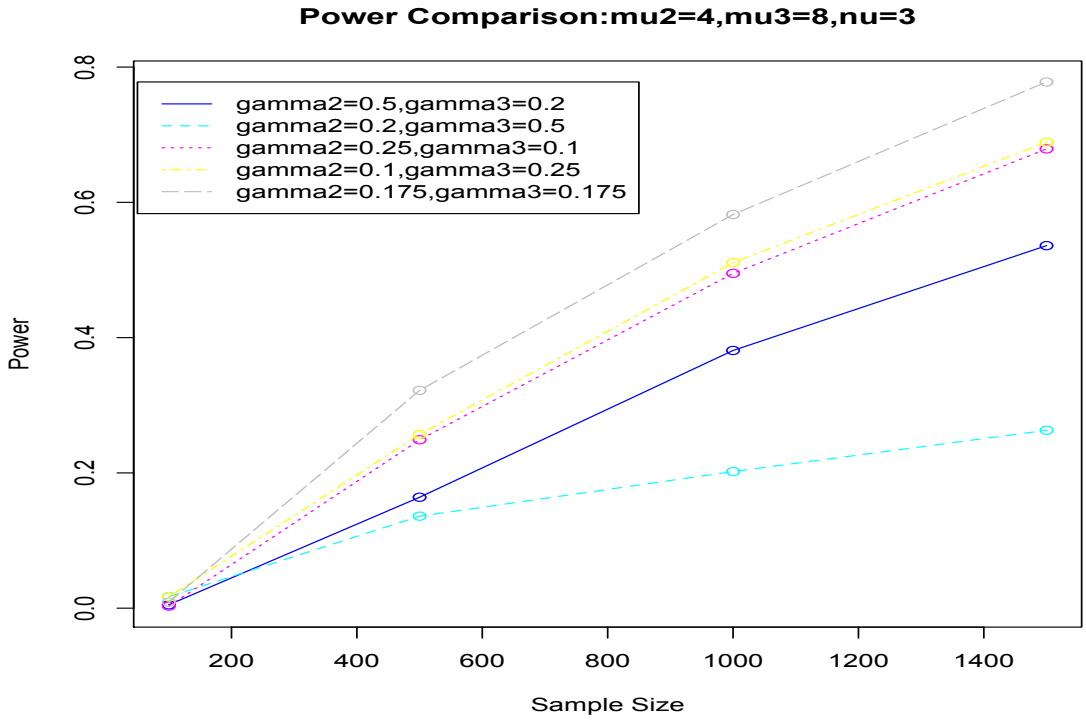
with parameters defined also in section 4.5. Powers are calcuated as the rejection rates of null hypothesis under H_A. Powers for various models are displayed in the remaining three panels of Figure 4.3.

Figure 4.3: Type I Error Rates and Powers for Different Models






Power Comparison:mu2=2,mu3=8,nu=3
gamma2=0.5,gamma3=0.2
gamma2=0.2,gamma3=0.5
gamma2=0.25,gamma3=0.1
gamma2=0.1,gamma3=0.25
gamma2=0.175,gamma3=0.175
Power
0.0
0.2
0.4
0.6
0.8
1.0
200
400
600
800
1000
1200
1400
Sample Size


Power Comparison:mu2=4,mu3=8,nu=3
gamma2=0.5,gamma3=0.2
gamma2=0.2,gamma3=0.5
gamma2=0.25,gamma3=0.1
gamma2=0.1,gamma3=0.25
gamma2=0.175,gamma3=0.175
Power
0.0
0.2
0.4
0.6
0.8
200
400
600
800
1000
1200
1400
Sample Size

In general, power increases with n for all models. For $\mu_2 = 2$ and $\mu_3 = 4$, the powers are between 0 and 0.24. This is anticipated, because we have conservative Type I error rates. But there is an apparent trend that power increases as n increases. The powers for models with $\mu_2 = 4$ and $\mu_3 = 8$ are better than the models with $\mu_2 = 2$ and $\mu_3 = 4$. And the powers for models with $\mu_2 = 2$ and $\mu_3 = 8$ are better than the models with $\mu_2 = 4$ and $\mu_3 = 8$. This is reasonable as the difference between the two contaminating components gets larger, it's easier to detect the extra contamination. In the meanwhile, if the proportion parameters γ_2 and γ_3 are large, it's also easier for us to observe the contamination.

4.8 A Case Study for EM Test

In section 4.4, CCS+EC model with moment-based method was applied to Blalock and colleagues' microarray experiments on 8,799 genes from three groups. The moment-based method has its own merits lying in that it is always easy to conduct a moment-based test. Besides, resampling and compactness are not necessary. However, a major weak point of the moment-based method is that when there are high dimensional parameter equations, the moment-based method cannot provide accurate estimation. In section 4.4, we did not estimate the parameters by the moment-based method. Now, we will apply the EM method to do the hypothesis testing and parameter estimation simultaneously.

As in section 3.6, we need to first convert the F statistics provided in the hippocampal data to corresponding P-values, then P-values to Chi-Square statistics. With Chi-Square statistics, we can apply EM method to do the hypothesis testing

and parameter estimation and we can also take filtration into account .

Although only 1 iteration is regulated for both null and alternative hypothesis testing, parameter estimation is based on 20 iterations of EM algorithm for both null hypothesis and alternative hypothesis, to provide more accurate estimation. The parameter estimates and the P-values for all genes, genes eliminated in each step and genes remaining after each step are shown in Table 4.3. From Table 4.3, we see the P-values for all genes, and genes remaining after each step are less than 0.05, which means CCS+EC model is more appropriate. For genes eliminated in each step, the P-values are greater than 0.05, indicating that CCS model fits the data better. This gives us some suggestion for how Blalock and his colleagues split the hippocampal data into remaining genes groups and eliminated genes groups, in that the latter groups had less heterogeneity. Figure 4.4 presents histograms for all genes and the genes remaining and eliminated in each step with fitted model under both null hypothesis and alternative hypothesis superimposed. It's quite hard to tell from the histogram which model fits better, since the figure cannot detect small differences between two models. But the P-values from EM testing procedure informed us which models fit better.

The fitted models for all genes and genes in each step suggested by EM testing and parameter estimation are displayed in Table 4.4. For genes eliminated in step 3A,in fact, it's not necessary to do further hypothesis testing with one extra component versus two extra components, because in chapters 2 and 3, the omnibus null hypothesis for genes eliminated in step 3A was not rejected. Therefore, since one extra component was not detected earlier there was no reason to anticipate detection

Figure 4.4: Fitted CCS Model vs. Fitted CCS+EC Model

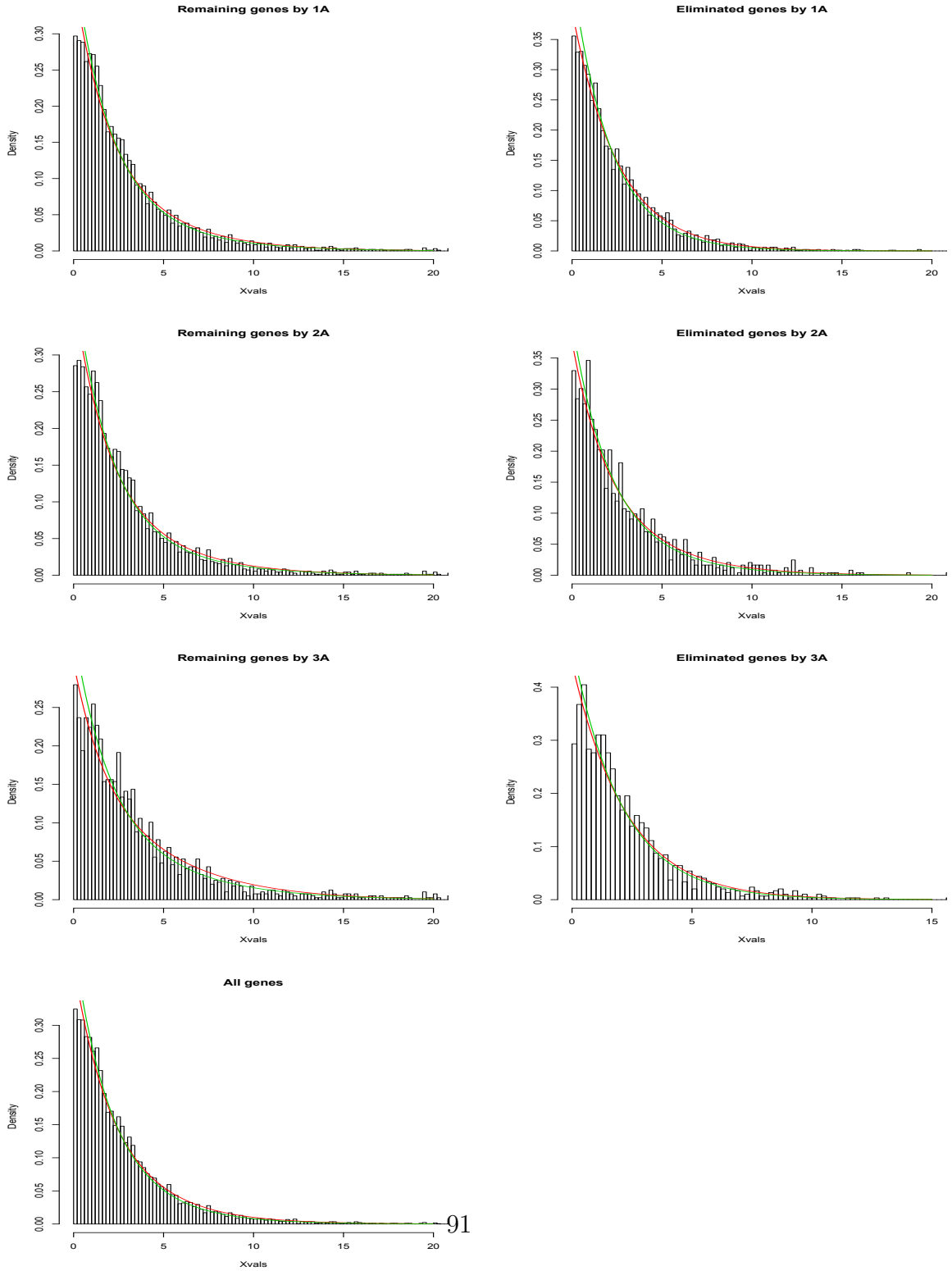


Table 4.3: Parameter Estimation by EM Method

Genes	$\hat{\mu}$	$\hat{\mu_2}$	$\hat{\mu_3}$	$\hat{\gamma}_0$	$\hat{\gamma}_A$	$\hat{\beta}$	$\hat{\gamma}_2$	$\hat{\gamma}_3$	P-value
All genes	2.4127	0.7037	2.9997	0.3054	0.2811	0.4980	0.1400	0.1411	0.0021
Remaining after 1A	3.1812	1.1294	4.2053	0.3048	0.2939	0.4967	0.1460	0.1479	0.0007
Eliminated in 1A	0.7063	0.1993	0.8991	0.6853	0.3686	0.4973	0.1833	0.1853	0.3372
Remaining after 2A	3.3651	1.1715	4.5678	0.2894	0.2805	0.4967	0.1393	0.1411	0.0006
Eliminated in 2A	2.700	0.7605	3.0914	0.3332	0.3348	0.4967	0.1663	0.1685	0.2308
Remaining after 3A	3.7485	1.1409	4.7513	0.4519	0.4490	0.4968	0.2230	0.2259	0.0004
Eliminated in 3A	0.2570	0.2376	0.2751	0.9012	0.3340	0.4973	0.1661	0.1679	0.5

of two extra components now.

Table 4.4: Fitted Model for All Genes, Remaining Genes and Eliminated Genes

Genes	Fitted Model	P-value
All genes	$0.7189\chi^2_2(0) + 0.1400\chi^2_2(0.7037) + 0.1411\chi^2_2(2.999)$	0.0021
Remaining after 1A	$0.7061\chi^2_2(0) + 0.1460\chi^2_2(1.1294) + 0.1479\chi^2_2(4.2053)$	0.0007
Eliminated in 1A	$0.6853\chi^2_2(0) + 0.3147\chi^2_2(0.7063)$	0.3372
Remaining after 2A	$0.2805\chi^2_2(0) + 0.1393\chi^2_2(1.1715) + 0.1411\chi^2_2(4.5678)$	0.0006
Eliminated in 2A	$0.3332\chi^2_2(0) + 0.6668\chi^2_2(2.7006)$	0.2308
Remaining after 3A	$0.4490\chi^2_2(0) + 0.2230\chi^2_2(1.1409) + 0.2259\chi^2_2(4.7513)$	0.0004
Eliminated in 3A	$0.9012\chi^2_2(0) + 0.0988\chi^2_2(0.2570)$	0.5

Chapter 5 Case Study: Gene Profile Analysis of sorted Sca1+/cKit- BM cells (BMCs)

5.1 Introduction

In this chapter, we consider the microarray data of Gifford AM et al. (2010) [25] for a practical application of the moment-based methods and the likelihood methods developed for CCS and CCS+EC models. The data are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25620. Gifford AM et al. analyzed the gene profile of sorted Sca1+/cKit- BMCs to identify their effects on the growth of responding tumors. Three groups of mice were studied: five mice bearing instigating tumors, five mice bearing non-instigating tumors, and four mice with matrigel. Systemic instigation is a process by which endocrine signals sent from certain tumors (instigators) simulate BMCs. (Meshe Elkabets et al. 2011 [22]). Matrigel is a control group with neither instigator nor non-instigator. In total, 22690 probe sets were scanned on each microarray chip.

In our case study, we performed both ANOVA F-tests on log transformed microarray data and Kruskal Wallis χ^2-tests on original microarray data to compare gene expression levels across three groups. Before data analysis, we reduced the total number of gene probe sets to be tested by excluding the genes for which the absolute difference between the instigating tumors group and the matrigel group did not differ by at least 91% of the maximal difference among groups for log transformed microarray data and at least 94% of the maximal difference among groups for original

microarray data. Then both the moment-based test and MLRT for CCS model were applied to test the omnibus null hypothesis of no contamination versus the alternative hypothesis of a central chi-square distribution contaminated by a non-central Chi-square distribution for the entire collection of genes and the genes remaining and excluded. For the groups of genes whose omnibus null hypotheses were rejected, we further tested whether there was an extra non-central Chi-Square component using moment-based test and EM test for CCS+EC model.

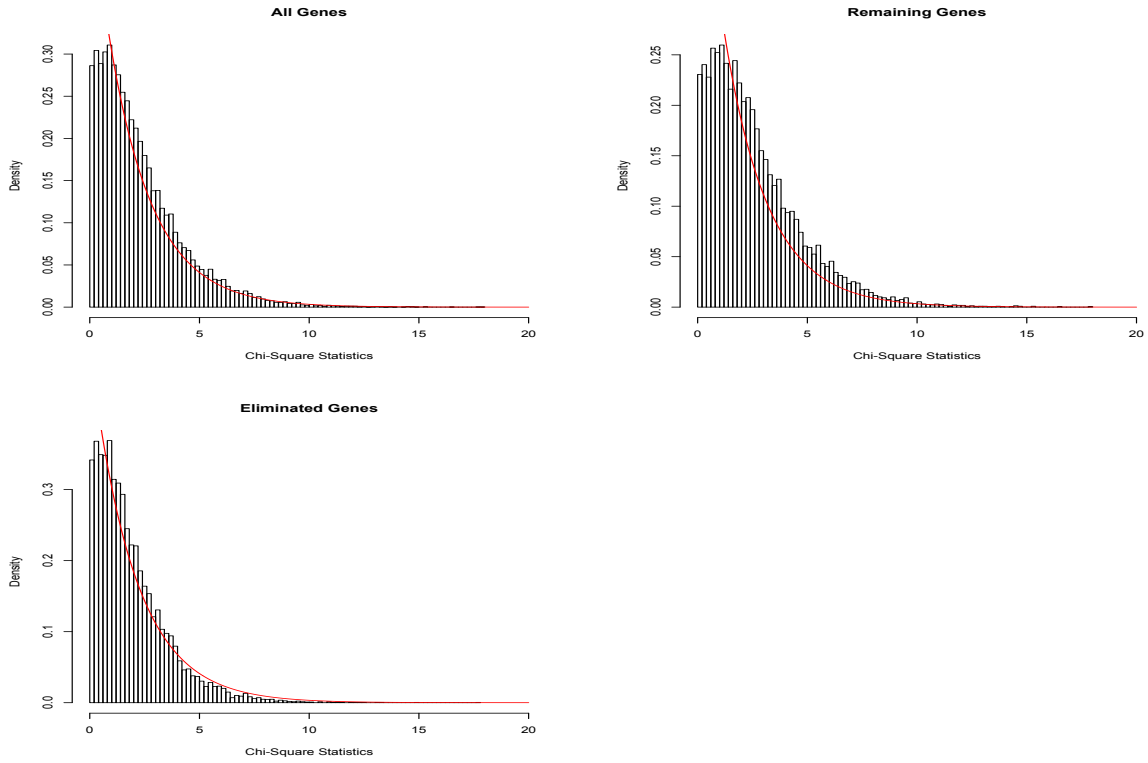
5.2 Moment-Based Tests and Likelihood-Based Tests for ANOVA F-Statistics

ANOVA test can be applied to test the means of the gene expression levels across the three groups of mice only if the assumptions for ANOVA test is satisfied. The Shapiro-Wilk test within each group of mice under each gene suggests that the genes are not normally distributed. Therefore, we made a log transformation on BMC microarray data. Then we applied the filtration criteria mentioned in the previous section to devide the entire collection of probe sets into two groups, a remaining group which included 11,366 genes whose log transformed absolute difference between the instigating tumors group and the matrigel group was greater than 91% of the log transformed maximal difference among groups, and an excluded group including 11,324 genes that were not in the remaining group. After filtration, one-way ANOVA F-tests were performed to compare expression levels across the three groups. Histograms of the χ^2-statistics with the fitted model from moment-based method for all genes, remaining genes and excluded genes are depicted in Figure 5.1,

with χ^2 statistics obtained from F statistics as first obtained in Chapter 2.

Now we apply the moment-based approach to test the homogeneity of all genes,

Figure 5.1: Histograms with the Fitted Model from Moment-Based Method for All Genes, Remaining Genes and Eliminated Genes



as well as the retained and the excluded genes. We set $\delta = \epsilon = \sqrt{0.05}$ (Chapter 2, CCS2 model). The results from moment-based test suggest that we should not reject the null hypothesis of homogeneity for all genes, retained genes and excluded genes, which means we may not conclude the genes were differentially expressed. By further investigating the moment-based approach in this case, we find that the reason for not rejecting the null hypotheses is that our W test statistic, which is used to ensure the positiveness of the estimators of μ and γ was negative. Thus any

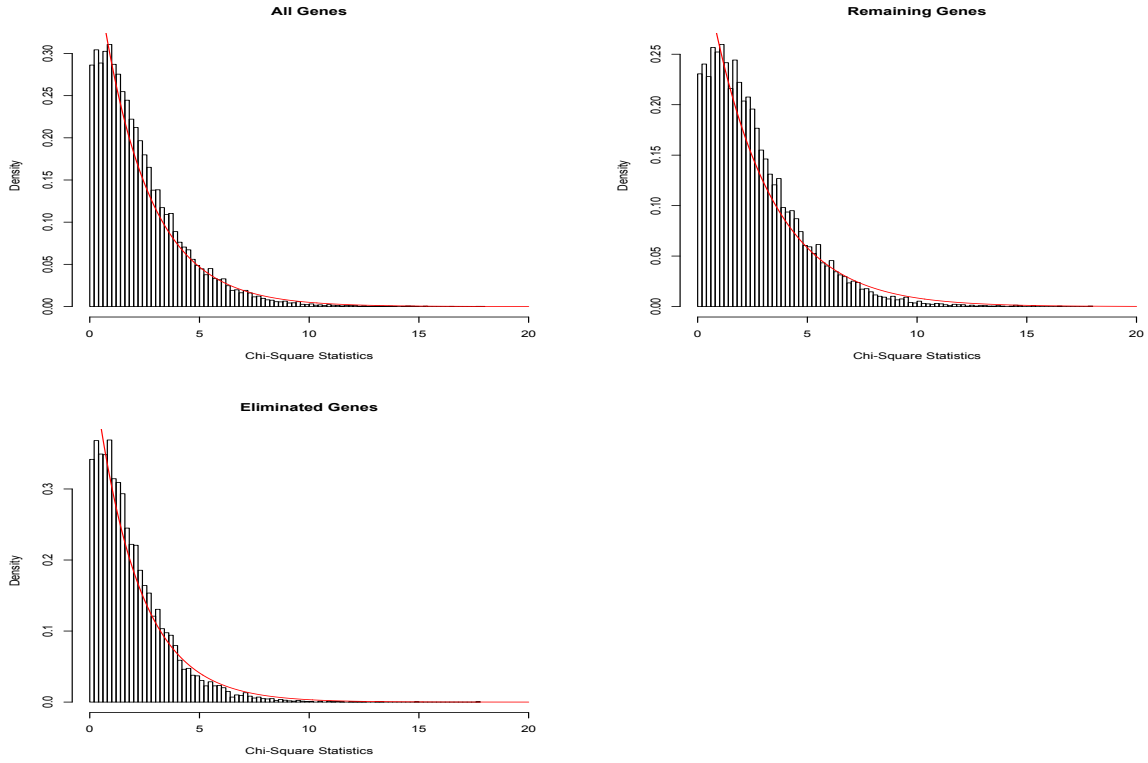
departure from a central χ^2 distribution suggested by the S statistic was driven by a factor other than differential expression (e.g., correlations between genes). Thus, we refrained from testing the second null hypothesis of one contamination of a central Chi-Square distribution versus the second alternative hypothesis of two contaminations of a central Chi-Square distribution.

Next MLRT with the penalty coefficient selected to be 10 is performed on the same microarray data. This time the P-values in Table 5.1 show that the omnibus null hypotheses of no contamination were rejected for all genes, eliminated genes but fail to be rejected for remaining genes. And we further estimate that 31.39% of 22690 genes were differentially expressed. For remaining genes, the percentage of differentially expressed genes is 71.19%. Therefore, the Contaminated Chi-Square models are more suitable for modeling the overall genes and the remaining genes, and the central Chi-Square model is used for the eliminated genes as judged by likelihood-based inference. The histograms of χ^2-statistics with the fitted Contaminated Chi-Square distributions superimposed for all genes and remaining genes and the histogram with the fitted central Chi-Square distribution superimposed for eliminated genes are presented in Figure 5.2. From the figure, the genes are well described by the corresponding distributions.

Table 5.1: Hypothesis Testing and Parameter Estimation from MLRT Method for central Chi-Square Model vs. CCS Model

	All Genes	Remaining Genes	Eliminated Genes
P-value	< 0.0001	< 0.0001	1
$\hat{\mu}$	0.6762	0.9056	$6.6 * 10^{-5}$
$\hat{\gamma}$	0.3139	0.7119	0.1752

Figure 5.2: Histograms with Fitted Model from MLRT Method for All Genes, Remaining Genes and Eliminated Genes on ANOVA F-statistics



Since the omnibus null hypotheses are easily rejected under significance level 0.05 for all genes and remaining genes, we are interested to know whether there will be an extra contamination involved. The EM tests with one iteration were conducted for testing problem while the EM tests with twenty iterations were performed for parameter estimation. The results in Table 5.2 indicate that we fail to reject the secondary null hypothesis for all genes and the eliminated genes with P-values equal to 0.2564 and 0.5 respectively. However, the secondary null hypothesis is rejected for the remaining genes. It can be noticed that the estimated value of μ_2, μ_3 are very close to each other, which seems to be a conflict with our hypothesis testing result.

This is mainly because of the difference of γ_{null} from γ_2 and γ_3. Hence, we can conclude there doesn't exist an extra contaminated component for all genes, remaining genes and eliminated genes.

Table 5.2: Hypothesis Testing and Parameter Estimation from EM Method for CCS Model vs. CCS+EC Model on ANOVA F-statistics

	All Genes	Remaining Genes	Eliminated Genes
P-value	0.2564	0.0050	0.5000
$\hat{\mu}$	0.3067	0.7423	0.0031
$\hat{\mu}_2$	0.3000	0.7196	0.0031
$\hat{\mu}_3$	0.3072	0.7210	0.0031
$\hat{\gamma}_{null}$	0.9092	0.9491	0.0969
$\hat{\gamma}_2$	0.2007	0.3573	0.0479
$\hat{\gamma}_3$	0.2024	0.3602	0.0490

By comparing the estimated values of parameters under the secondary null hypothesis with the estimated values under the omnibus alternative hypothesis, one may notice they are not consistent. This is because, for the omnibus hypothesis testing, we use MLRT which has only one set of initial values with many iterations. However, for the unilateral hypothesis testing problem, we apply the EM test which has several sets of initial values with only one iteration. Therefore, there is a difference between the estimaters.

5.3 Moment-Based Tests and Likelihood-Based Tests for Kruskal-Wallis χ^2-Statistics

The assumptions for ANOVA test are normality, independence and equal variance. But in real life, we rarely find a dataset that satisfies all the assumptions. Here the gene profile of Sca1+/cKit- BMCs does not fulfill the requirement of ANOVA test as well even after the log transformation. This might be the reason that the moment-based approach for testing the omnibus null hypothesis of homogeniety failed to reject the null. To avoid the strict assumptions for ANOVA, we turned to the Kruskal Wallis nonparametric method. Then the same filtration approach was also applied, but with a criterion that the absolute median difference between the instigating tumors group and the matrigel group did not difffer by at least 94% for the original microarray data of the maximal difference among groups. This split the overall 22690 genes into the remaining group with 11406 genes and the eliminated group with 11284 genes. Next, Kruskal Wallis χ^2-tests were conducted to compare the three group median expression levels.

This time we apply the moment-based method and the MLRT method to test the omnibus null hypothesis of no contamination versus the contaminated model on the Kruskal Wallis χ^2 statistics. The results of the moment-based method and MLRT method are in Table 5.3 and Table 5.4 respectively. From these two tables, the results based on Kruskal Wallis χ^2 statistics are very similar to those based on the ANOVA F statistics. For the moment-based method, the null hypotheses of homogeneity were all not rejected because of the negative values of W statistic. For the MLRT method, The null hypotheses were only rejected for all genes and remaining

genes, while the null hypothsis was not rejected for the eliminated genes. This is consistent with our expectation because the genes are less differentially expressed in the eliminated group.

Table 5.3: Hypothesis Testing and Parameter Estimation from Moment-Based Method for central Chi-Square Model vs. CCS Model on Kruskal Wallis χ^2-tests

	All Genes	Remaining Genes	Eliminated Genes
P-value	1	1	1
$\hat{\mu}$	-5.1147	-2.9800	-2.1601
$\hat{\gamma}$	-0.0459	-0.5644	0.5685

Table 5.4: Hypothesis Testing and Parameter Estimation from MLRT Method for central Chi-Square Model vs. CCS Model on Kruskal Wallis χ^2-tests

	All Genes	Remaining Genes	Eliminated Genes
P-value	< 0.0001	< 0.0001	0.9299
$\hat{\mu}$	0.6698	0.8722	$6.6 * 10^{-5}$
$\hat{\gamma}$	0.3007	0.5838	0.1750

Based on the testing results from the omnibus null hypotheses, we proceed to test the secondary null hypothesis of a CCS model versus a CCS+EC model only using the EM method. For all of the three batches, we failed to reject the secondary null hypotheses on account of large P-values. The only difference between the EM tests on Kruskal Wallis χ^2 statistics and ANOVA F statistics is that this time the null hypothesis was not rejected for the remaining genes with P-value equal to 0.5. Furthermore, we obtained close estimated values for μ and γ with the ones from the tests based on ANOVA F statistics. The detailed results are in Table 5.5.

Table 5.5: Hypothesis Testing and Parameter Estimation from EM Method for CCS Model vs. CCS+EC Model on Kruskal Wallis χ^2-tests

	All Genes	Remaining Genes	Eliminated Genes
P-value	0.3089	0.5	0.3784
$\hat{\mu}$	0.2881	2.1973	0.0031
$\hat{\mu}_2$	0.2879	1.8703	0.0031
$\hat{\mu}_3$	0.2852	2.2152	0.0031
$\hat{\gamma}_{null}$	0.9070	1	0.0933
$\hat{\gamma}_2$	0.1943	0.1983	0.0464
$\hat{\gamma}_3$	0.1959	0.8016	0.0469

5.4 Summary

In Chapter 5, we applied the moment-based methods and likelihood-based methods developed in previous chapters to the microarray data of Gifford AM et al. We analyzed the Chi-Square test statistics from both ANOVA P-values and Kruskal Wallis P-values. Both the moment-based method for the test of homogeneity on Chi-Square statistics from ANOVA P-values and Kruskal Wallis P-values produced negative estimates, which prevented us from further investigation. This means that we may still have departure from the central Chi-Square distribution, but it possibly comes from the dependence among genes within the same mice.

The results of MLRT tests for Chi-Square statistics from ANOVA P-values and Kruskal Wallis P-values both suggest that the genes are differentially expressed in all genes, remaining genes. But there is no sign of differentiation in the eliminated genes group. The estimates for all genes and remaining genes under the alternative hypotheses and the estimates for eliminated genes under the null hypotheses from both MLRT tests are very close to each other. This increases the credibility of our findings.

The secondary null hypotheses of one extra component failed to be rejected for all genes and eliminated genes with Chi-Square statistics from both ANOVA P-values and Kruskal Wallis P-values. However, it was rejected for the remaining genes when the statistics were transformed from ANOVA P-values, while failed to be rejected when the Chi-Square statistics were from nonparametric P-values. The result based on the nonparametric Kruskal-Wallis test is more reliable because the values of $\hat{\mu}_2$ and $\hat{\mu}_3$ from ANOVA are very close to $\hat{\mu}$, which suggests there is not really an extra component involved.

Based on all the analysis above, we can conclude that there exists a differentiation in the microarray. But we can only detect one extra differential component. Therefore, the CCS model is proposed to describe the distribution of the Sca1+/cKit- BMC genes, and we need likelihood based inference to do so

Gifford and colleagues (2010) [25] identified granulin (GRN) as the most upregulated gene in instigating Sca1+ cKit- BMCs. According to our results, the most differentially expressed genes are 1417033at and 1448309at based on Kruskal Wallis test and 1439075at based on ANOVA test. Our most differentially expressed gene is defined to be the one with the largest χ^2 statistics. We did post-hoc tests for multiple comparison of three groups of mice on 1439075at, 1417033at and 1448309at after ANOVA test and Kruskal Wallis test. The Bonferroni adjusted test after ANOVA shows that the tumor growth with BMCs of mice bearing instigating tumors is significantly different from the tumor growth with BMCs of mice bearing non-instigating tumors or matrigel. In contrast, Bonferroni-adjusted tests after Kruskal Wallis suggests that the tumor growth for mice with instigating tumors is significantly different only from mice in matrigel group. We may conclude that the instigating tumors have

greater impact than non-instigating tumors on expression of these genes.

Chapter 6 Future Work

My research is focused on the theory and methods for Chi-Square mixture modeling and its application to the microarray. Basically, three methods including a moment-based method and two likelihood based methods, the MLRT and the EM test are developed. The first two parts of my thesis consist of the application of these three methods for testing the omnibus null hypothesis of no contamination versus the alternative hypothesis of a central Chi-Square distribution contaminated by a non-central Chi-Square distribution. The second two parts of my work show that if the omnibus null hypothesis is rejected, how two of these methods can be developed and employed to test whether there are two non-central contaminated Chi-Square.

The contaminated Chi-Square model has been explored extensively in this dissertation. There are still some aspects remaining for future research.

First, in our work, we focused on testing the central Chi-Square model versus CCS model and CCS model versus CCS+EC model. For the real data which is more complicated, tests for even higher dimensional Chi-Square mixture models, i.e., the central Chi-Square distribution contaminted by more than two non-central Chi-Square distributions can be considered.

Second, the Chi-Square mixture models are based on the the Chi-Square statistics transformed from the ANOVA F statistics. We made the assumption of independence of each gene to perform the ANOVA test. In reality, the gene expression levels are correlated. While this assumption is violated, what action can be taken?

Third, given recent advances in univariate Chi-Square mixture models, we be-

lieve that this is an opportune time to develop new inferential tools for a "general" multivariate mixture model $X_1, \ldots, X_n \sim \sum_{j=1}^{k} \gamma_j f(x; \theta_j)$ in which $\theta_1, \ldots, \theta_k$ are not scalars. By general we mean that attention will not be confined to a single specific parametric family of multivariate pdfs. Rather, as various authors (including ourselves) have done for univariate mixture and/or contamination models, one may consider any parametric family satisfying appropriate assumptions. Moreover, the assumptions will be only those necessary to establish useful theorems on the behavior of test statistics and parameter estimators. As such, we anticipate that the new inferential tools will expedite the deployment of multivariate mixture modeling in a variety of practical applications, both inside the field of genetics (for example, consider both red and green intensities at n positions on a DNA microarray) and outside (for example, joint modeling of birthweight and gestational age data).

Chapter A Appendix

A.1 Appendix I: Proof of Regularity Conditions in Chapter 3

Condition 1. Wald's integrability conditions. The kernel function $\chi^2_\nu(\mu)$ satisfies Wald's integrability conditions for consistency of the maximum likelihood estimator, i.e. for each $\mu \in \Theta$, (i) $E\|\log f_\nu(X;\mu)\| < \infty$, and (ii) for sufficiently small $\rho > 0$ the expected values $E \log f(X;\mu,\rho) < \infty$, where

$$f(X;\mu,\rho) = 1 + sup_{\|\mu'-\mu\|\leq\rho}\{f(X;\mu')\}.$$

Proof.

$$\begin{aligned} E[\log f(X;\mu)] &= \int_0^\infty \log f(x;\mu) f(x;\mu) dx \\ &= \int_0^\infty \left(-\frac{x+\mu}{2} + (\frac{\nu}{4} - \frac{1}{2})\log\frac{x}{\mu} + \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x})\right) e^{-\frac{x+\mu}{2}} (\frac{x}{\mu})^{\frac{nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) dx. \end{aligned}$$

For (i), it suffices to show that,

1. $-\infty < E[X] < \infty$.
2. $-\infty < E[\log X] < \infty$.
3. $-\infty < E[\log I_{\frac{\nu}{2}-1}(\sqrt{\mu X})] < \infty$.

Since $E[X] = \mu + \nu$, it is easy to get 1.

$$\begin{aligned} E[\log X] &= \int_0^\infty \log x e^{-\frac{x+\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) dx \\ &< \int_0^\infty x e^{-\frac{x+\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) dx \\ &< \infty \end{aligned}$$

There exists $\epsilon > 0$ such that $I_{\frac{\nu}{2}-1}(\sqrt{\mu X}) \leq 2$ for all $x \in (0, \epsilon)$. Assume $\nu \geq 2$, then,

$$\begin{aligned} &\int_0^\epsilon \log x e^{-\frac{x+\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) dx \\ &\qquad \geq \int_0^\epsilon \log x e^{-\frac{x+\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} 2 dx \\ &\qquad \geq \int_0^\epsilon \log x e^{-\frac{\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} 2 dx \\ &\qquad = \int_0^\epsilon \log x dx e^{-\frac{\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} 2 \\ &\qquad = (\epsilon \log \epsilon - \epsilon)(e^{-\frac{\mu}{2}} (\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} 2) \\ &\qquad > -\infty. \end{aligned}$$

Then we need to prove that $-\infty < E[\log I_{\frac{1}{2}-1}(\sqrt{\mu X})] < \infty$. As $X \to 0^+$,

$$I_{\frac{\nu}{2}-1}(\sqrt{\mu X}) \approx \frac{1}{0!\Gamma(0 + \frac{\nu}{2} - 1 + 1)} (\frac{\mu X}{2})^{\frac{\nu}{2}-1}.$$

Since

$$\begin{aligned} E[\log I_{\frac{\nu}{2}-1}(\sqrt{\mu X})] &= \int_0^\infty \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) f(x;\mu) dx \\ &= \int_0^\epsilon \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) f(x;\mu) dx + \int_\epsilon^M \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) f(x;\mu) dx + \int_M^\infty \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) \end{aligned}$$

For $\nu \geq 2$, first consider

$$\int_0^\epsilon \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) f(x;\mu) dx \approx \int_0^\epsilon \log[\frac{1}{\Gamma(\frac{\nu}{2})} (\frac{\mu x}{2})^{\frac{\nu}{2}-1}] f(x;\mu) dx.$$

We have proved in 2. that $\log X$ is integrable, hence, the above equation is also integrable.

As $X \to \infty, I_{\frac{\nu}{2}-1}(X) \approx \frac{e^X}{\sqrt{2\pi X}}$, we have

$$\log I_{\frac{\nu}{2}-1}(\sqrt{\mu X}) \approx \sqrt{\mu X} - \frac{1}{2}\log(2\pi\sqrt{\mu X}).$$

Therefore,

$$\int_M^\infty \log I_{\frac{\nu}{2}-1}(\sqrt{\mu x}) f(x;\mu)dx \approx \int_M^\infty (x - \frac{1}{2}\log(2\pi\sqrt{\mu x})) f(x;\mu)dx,$$

which is integrable.

The proof of (ii) remains to be completed.

□

Condition 2. Smoothness. The kernel function $\chi^2_\nu(\mu)$ has common support for all $\mu \in \Theta$ and is twice continuously differentiable with respect to μ.

Proof. We know that for $\mu \in (0, \infty)$ and for fixed x,

$$f(x;\mu) \propto e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x}),$$

each part is twice continuously differentiable over $\mu \in (0, \infty)$. So the product is also twice continuously differentiable over $\mu \in (0, \infty)$.

When $\mu \to 0$,

$$\begin{aligned} f(X;\mu) - f(X;0) &\approx \frac{1}{\Gamma(\frac{k}{2}-1)2^{\frac{k}{2}}}[e^{-\frac{X+\mu}{2}}(\frac{X}{\mu})^{\frac{k}{4}-\frac{1}{2}}(\sqrt{\mu X})^{\frac{k}{2}-1} - e^{-\frac{X}{2}}X^{\frac{k}{2}-1}] \\ &\approx f(X;0)[1 - \frac{\mu}{2} - 1] \end{aligned}$$

From the equation above, we have

$$\frac{f(X;\mu) - f(X;0)}{\mu} \approx f(X;0) * -\frac{1}{2} := f'(X;0).$$

The existence of a continuous second derivative and the continuity of the derivative at 0 remains to be proved. □

Condition 3. Strong identifiability. For any two mixing distribution functions Ψ_1 and Ψ_2 such that

$$\int f_\nu(x;\mu)d\Psi_1(\mu) = \int f_\nu(x;\mu)d\Psi_2(\mu), \text{for all } x,$$

we must have $\Psi_1 = \Psi_2$. The mixing distribution Ψ is defined as

$$\Psi_(\mu) = (1-\gamma)I(\mu_1 \leq \mu) + \gamma I(\mu_2 \leq \mu).$$

Proof. Let

$$\int g(x)[(1-\alpha)\chi^2_\nu(0) + \alpha\chi^2_\nu(\mu_1)]dx = \int g(x)[(1-\beta)\chi^2_\nu(0) + \beta\chi^2_\nu(\mu_2)]dx$$

If either quantity is zero, then the proof is complete. Otherwise, choose $g(X) = X$, then

$$\int x[(1-\alpha)\chi^2_\nu(0) + \alpha\chi^2_\nu(\mu_1)]dx = \nu + \alpha\mu_1,$$

and

$$\int x[(1-\beta)\chi^2_\nu(0) + \beta\chi^2_\nu(\mu_2)]dx = \nu + \beta\mu_2.$$

$\Rightarrow \alpha\mu_1 = \beta\mu_2$. Choose $g(X) = X^2$, then

$$\int x^2[(1-\alpha)\chi^2_\nu(0) + \alpha\chi^2_\nu(\mu_1)]dx = 2\nu + \nu^2 + 4\alpha\mu_1 + 2\alpha\nu\mu_1 + \alpha\mu_1^2$$

and

$$\int x^2[(1-\beta)\chi^2_\nu(0) + \beta\chi^2_\nu(\mu_2)]dx = 2\nu + \nu^2 + 4\beta\mu_2 + 2\beta\nu\mu_2 + \beta\mu_2^2.$$

$\Rightarrow 2\nu + \nu^2 + 4\alpha\mu_1 + 2\alpha\nu\mu_1 + \alpha\mu_1^2 = 2\nu + \nu^2 + 4\beta\mu_2 + 2\beta\nu\mu_2 + \beta\mu_2^2$.

We can conclude that $\mu_1 = \mu_2$ and $\alpha = \beta$, where $\Phi_1 = \Psi_2$. □

Condition 4. Uniform strong law condition of large numbers. There exists integrable g with some $\delta > 0$ such that $\|Y_i(\mu)\|^3 \leq g(X_i)$ for all $\mu \in \Theta$, where $Y_i(\mu) = \frac{f_\nu(X_i;\mu)-f_\nu(X_i;0)}{\mu f_\nu(X_i;0)}$; $Y(0) = \frac{f'_\nu(X_i;0)}{f_\nu(X_i;0)}$.

Proof. By first order Taylor Expansion,

$$\frac{f_\nu(X_i;\mu) - f_\nu(X_i;0)}{\mu f_\nu(X_i;0)} = \frac{\frac{\partial}{\partial\mu}f(X;\tilde{\mu})}{f(X;0)} = \frac{f'(X;\tilde{\mu})}{f(X;0)},$$

thus

$$\begin{aligned}
f'(X_i;\mu) = \frac{\partial}{\partial\mu}f(X_i;\mu) &= \frac{\partial}{\partial\mu}\{\frac{1}{2}e^{-\frac{X_i+\mu}{2}}(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i})\} \\
&= \frac{\partial}{\partial\mu}\{\frac{1}{2}e^{-\frac{X_i+\mu}{2}}\}(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i}) \\
&+ \frac{1}{2}e^{-\frac{X_i+\mu}{2}}\frac{\partial}{\partial\mu}\{(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}\}I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i}) \\
&+ \frac{1}{2}e^{-\frac{X_i+\mu}{2}}(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}\frac{\partial}{\partial\mu}\{I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i})\}.
\end{aligned}$$

Let

$$\begin{aligned}
h_1(X_i;\mu) &:= \frac{\partial}{\partial\mu}\{\frac{1}{2}e^{-\frac{X_i+\mu}{2}}\}(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i}), \\
h_2(X_i;\mu) &:= \frac{1}{2}e^{-\frac{X_i+\mu}{2}}\frac{\partial}{\partial\mu}\{(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}\}I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i}), \\
h_3(X_i;\mu) &:= \frac{1}{2}e^{-\frac{X_i+\mu}{2}}(\frac{X_i}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}\frac{\partial}{\partial\mu}\{I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i})\}.
\end{aligned}$$

We want to find $g_1(X_i), g_2(X_i), g_3(X_i)$ such that

$$|h_1(X_i;\mu)| \leq g_1(X_i); |h_2(X_i;\mu)| \leq g_2(X_i); |h_3(X_i;\mu)| \leq g_3(X_i),$$

with

$$Eg_1(X_i) < \infty; Eg_2(X_i) < \infty; Eg_3(X_i) < \infty.$$

Recall for non-negative ν,

$$I_\nu(z) = (\frac{1}{2}z)^\nu \frac{(\frac{1}{4}z^2)^k}{k!\Gamma(\nu+k+1)}.$$

For small z, the first term in the summation is dominat. That is,

$$\begin{aligned} I_\nu(z) &= (\frac{1}{2}z)^\nu \sum_{k=0}^{\infty} \frac{(\frac{1}{4}z^2)^0}{0!\Gamma(\nu+1)}\{1+o(1)\} \\ &= \frac{(\frac{1}{2}z)^\nu}{\Gamma(\nu+1)}\{1+o(1)\}. \end{aligned}$$

So, given $\epsilon > 0$, there exists $\delta > 0$ such that $\forall z < \delta, |o(1)| < \epsilon$. Choose $\epsilon := 1$. Then for all $z < \delta$,

$$I_\nu(z) \leq \frac{(\frac{1}{2}z)^\nu}{\Gamma(\nu+1)}2.$$

Therefore, for $\sqrt{\mu X_i} < \delta$, we have

$$I_{\frac{\nu}{2}-1}(\sqrt{\mu X_i}) \leq \frac{(\frac{1}{2}\sqrt{\mu X_i})^{\frac{\nu}{2}-1}}{\Gamma(\nu+1)}2,$$

and thus

$$|h_1(X_i;\tilde{\mu})| \leq K(\nu),$$

for $\sqrt{\mu X_i} < \delta$, where $K(\nu)$ is some constant.
For $\sqrt{\mu X_i} > \delta$, we have $|h_1(X_i;\mu)| \leq L(\nu)\delta^{1-\frac{\nu}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max} X_i})$, where $L(\nu)$ is another constant.

Now put

$$g_1(X) := K(\nu)\delta^{1+\frac{nu}{2}} + L(\nu)\delta^{1-\frac{\nu}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max} X}).$$

Then $|h_1(X;\mu)| \leq g_1(X)$. Next we need to prove that $E_{H_0} g_1(X) < \infty$. This boils down to $E_{H_0} I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max} X_i}) < \infty$. It is equivalent to $\int_0^\infty I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max} X}) f(X;0) dX <$

∞.

As $X \to \infty$,

$$I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}X}) \approx \frac{e^{\sqrt{\mu_{max}X}}}{\sqrt{2\pi\sqrt{\mu_{max}X}}}$$

and

$$f_\nu(X;0) = X^{\frac{\nu}{2}-1}e^{-\frac{X}{2}}C,$$

where C is a constant. Then

$$I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}X})f(X;0) \approx CX^{\frac{\nu}{2}-\frac{5}{4}}e^{-\frac{X}{2}+(\mu_{\max}X)^{\frac{1}{2}}}.$$

Therefore,

$$E_{H_0}h_1(X_i;\mu) = \int -\frac{1}{2}\frac{f_\nu(x_i;\mu)}{f_\nu(x_i;0)}f_\nu(x_i;0)dx_i = -\frac{1}{2}.$$

Using similar method, we can find $g_2(X)$ and $g_3(X)$. So

$$\begin{aligned}
|Y_i(\mu)| &= |\frac{f(X_i;\mu) - f(X_i;0)}{\mu f(X_i;0)}| \\
&\leq |h_1(X_i;\mu)| + |h_2(X_i;\mu)| + |h_3(X_i;\mu)| \\
&\leq g_1(X_i) + g_2(X_i) + g_3(X_i) \\
&= g(X_i) \text{ for all } X_i \text{ and } \mu.
\end{aligned}$$

with $E[g(X_i)] = E[g_1(X_i) + g_2(X_i) + g_3(X_i)]$. We have shown $|Y_i(\mu)| \leq g(X_i)|$ with $E[g(X_i)] < \infty$. Now we need to show $|Y_i(\mu)|^3 \leq g(X_i)|$ with $E[g(X_i)] < \infty$. Since $\frac{e^{3\sqrt{\mu X}-\frac{X}{2}}}{e^{-\frac{X}{4}}} \to 0$, as $X \to \infty$. So for $X > x_0$,

$$e^{3\sqrt{\mu X}-\frac{X}{2}} < 2e^{-\frac{X}{4}}$$

$$\int_{x_0}^{\infty} \frac{e^{3\sqrt{\mu x}}}{(\mu x)^{\frac{3}{4}}}e^{-\frac{x}{2}}x^{\frac{\nu}{2}-1}dx < \int_{x_0}^{\infty} 2\frac{e^{-\frac{x}{4}}}{(\mu x)^{\frac{3}{4}}}e^{-\frac{x}{2}}x^{\frac{\nu}{2}-1}dx$$

Therefore,

$$\int_{x_0}^{\infty} \frac{e^{3\sqrt{\mu x}}}{(\mu x)^{\frac{3}{4}}} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1} dx < \infty$$
$$\int_{x_0}^{\infty} (I_{\frac{\nu}{2}-1}\sqrt{\mu_{max} x})^3 f(x;0) dx < \infty$$
$$\int_{0}^{\infty} (I_{\frac{\nu}{2}-1}\sqrt{\mu_{max} x})^3 f(x;0) dx < \infty$$
$$E[g_1(X)^3] < \infty.$$

So

$$\begin{aligned} |Y_i(\mu)|^3 &\leq |h_1(X_i;\mu)|^3 + |h_2(X_i;\mu)|^3 + |h_3(X_i;\mu)|^3 \\ &\leq g_1(X_i)^3 + g_2(X_i)^3 + g_3(X_i)^3 \text{ for all } X_i \text{ and } \mu. \end{aligned}$$

with $E[g_1(X)^3] < \infty$, $E[g_2(X)^3] < \infty$, $E[g_3(X)^3] < \infty$. □

Condition 5. Tightness. The process $n^{-1/2} \sum Y_i(\mu)$ is tight.

Proof. We know $E[Y_i(\theta)] = 0$ for any θ, so $E[n^{-\frac{1}{2}} \sum Y_i(\theta_2) - n^{-\frac{1}{2}} \sum Y_i(\theta_1)] = 0$ Thus

$$\begin{aligned} E[n^{-\frac{1}{2}} \sum Y_i(\theta_2) - n^{-\frac{1}{2}} \sum Y_i(\theta_1)]^2 &= V[n^{-\frac{1}{2}} \sum Y_i(\theta_2) - n^{-\frac{1}{2}} \sum Y_i(\theta_1)] \\ &= E[Y_i(\theta_2) - Y_i(\theta_1)]^2 \\ &= E[Y_i'(\tilde{\theta})(\theta_2 - \theta_1)^2]. \end{aligned}$$

The last equation is from the Mean Value Theorem. $\tilde{\theta}$ is between θ_2 and θ_1. The rest of proof remains to be completed.

□

A.2 Appendix II: Proof of Regularity Conditions in Chapter 4

Condition 0. The penalty term $p(\beta)$ is a continuous function such that it is maximized at $\beta = 0.5$ and goes to negative infinity as β goes to 0 or 1.

Proof. Condition 0 is obvious to see. □

Condition 1. The kernel function $f(X;\mu)$ is such that the mixture distribution satisfies Wald's integrability conditions for consistency of the maximum likelihood estimator. For this, it suffices to require that

a. $E|\log f(X;\Psi_0)| < \infty$.

b. for sufficiently small ρ and for sufficiently large r, $E\log\{1+f(X;\mu,\rho)\} < \infty$ for $\mu \in \Theta$ and $E\log\{1+\phi(X;r)\} < \infty$, where $f(X;\mu,\rho) = \sup_{|\mu'-\mu|\leq\rho} f(X;\mu')$ and $\phi(X;r) = \sup_{\mu\geq r} f(X;\mu)$.

c. $\lim_{|\mu|\to\infty} f(x;\mu) = 0$ for all x except on a set with probability 0.

Proof. We first prove a.part.

$$\begin{aligned} E[\log f(X;\Phi_0)] &= \int_0^\infty \log f_\nu(x;\Phi_0) f_\nu(x;\mu)dx \\ &= \int_0^\infty \log[(1-\gamma)\chi^2_\nu(0) + \gamma\chi^2_\nu(\mu)] f(x;\mu)dx \\ &= \int_0^\infty \log[(1-\gamma)\frac{x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{\nu}{2}}\Gamma(\frac{\nu}{2})} + \gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x})] \\ &\quad \times \frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx \end{aligned} \tag{A.1}$$

It is obvious that (A.1) is bounded below by

$$\int_0^\infty \log[0 + \gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}} I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx$$

and bounded above by

$$\int_0^\infty \log[1+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx.$$

It has been proved in Appendix I that

$$\int_0^\infty \log[0+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx$$

is bounded.

Now we need to prove that

$$\int_0^\infty \log[1+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx$$

is also bounded. We have

$$\begin{aligned}
&\int_0^\infty \log[1+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx\\
&\leq \int_0^\infty [1+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})]\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx\\
&\leq \int_0^\infty 1+\gamma\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu x})dx\\
&< \infty
\end{aligned}$$

where $\nu \geq 2$.

Therefore, $E|\log f(X;\Phi_0)| < \infty$.

Part b.

$E\log[1+f(X;\mu,\rho)] < \infty$ has been proved in Appendix I. Now we prove $E\log[1+\phi(X;r)] < \infty$. We have

$$\begin{aligned}
0 &\leq E\log[1+\Phi(X;r)]\\
&= E\log[1+\sup_{\mu\geq r} f(X;\mu)]\\
&\leq E\log[1+\frac{1}{2}e^{-\frac{X+r}{2}}(\frac{X}{r})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}X})]\\
&= \int_0^\infty \log[1+\frac{1}{2}e^{-\frac{x+r}{2}}(\frac{x}{r})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}x})]f(x;\mu)dx
\end{aligned}$$

Let

$$A := \{X : \frac{1}{2}e^{-\frac{X+\mu}{2}}(\frac{X}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}X}) \leq 1\}$$

Hence,

$$\begin{aligned}
&\int_0^\infty \log[1+\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}x})]f(x;\mu)dx \\
&= \int_A \log[1+\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}x})]f(x;\mu)dx \\
&+ \int_{A^c} \log[1+\frac{1}{2}e^{-\frac{x+\mu}{2}}(\frac{x}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu_{max}x})]f(x;\mu)dx
\end{aligned}$$

It can be proved that both A and A^c are integrable, this completes the proof of part b.

Part c.

There exists $\epsilon > 0$ such that

$$\begin{aligned}
P(f(X;\mu) \geq \epsilon) &= P(\frac{1}{2}e^{-\frac{X+\mu}{2}}(\frac{X}{\mu})^{\frac{\nu}{4}-\frac{1}{2}}I_{\frac{\nu}{2}-1}(\sqrt{\mu X}) \geq \epsilon) \\
&\leq P(I_{\frac{\nu}{2}-1}(\sqrt{\mu X} \geq 2\epsilon) \\
&\leq P(\frac{e^{\sqrt{\mu X}}}{\sqrt{2\pi\sqrt{\mu X}}} \geq 2\epsilon) \to 0 \text{ as } \mu \to \infty.
\end{aligned}$$

□

Condition 2. The kernel function $f(X;\mu)$ has common support and is four times continuously differentiable with respect to μ.

Proof. The proof of Condition 2 is possible using computations such as those in Appendix I. □

Condition 3. For any two mixing distribution functions Ψ_1 and Ψ_2 such that $\int f(x;\mu)d\Psi_1(\mu) = \int f(x;\mu)d\Psi_2(\mu)$ for all x, we must have $\Psi_1 = \Psi_2$.

Proof. The proof of Condition 3 can follow the proof of Condition 3 in Appendix I but with a three components model. □

Condition 4. Let $N(\mu, \epsilon) = \{\mu' \in \Theta : |\mu' - \mu| \leq \epsilon\}$ for some positive ϵ. There exists an intergrable $g(.)$ and a small positive ϵ_0 such that $|\delta_{ih}|^3 \leq g(X_i)$, $|Y_i(\mu)|^3 \leq g(X_i)$, $|Z_i^{(k)}(\mu)| \leq g(X_i)$, for $\mu \in N(\mu_{0h}, \epsilon_0), h = 1, 2$, and $k = 0, 1, 2$ with $Z_1^{(k)}(\mu)$ being the kth derivative.

Proof. The proof of Condition 4 here can be achieved by following the proof of Condition 4. in Appendix I. □

Condition 5. The variance-covariance matrix $\mathbf{B}$ of $b_i = (\delta_{i1}, Y_i(0), Z_i(0), Z_i(\mu))^T$ is positive definite.

Proof. The proof remains to be completed. □

Bibliography

[1] Anandkumar, A., Hsu, D., Kakade, S.M. (2012), "A Method of Moments for Mixture Models and Hidden Markov Models," *Journal of Machine Learning Research*, vol 23, 33.133.34.

[2] Allison, D.B., Gadbury, G.L., Heo, M., Fernandez, J.R., Lee, C., Prolla, T.A., Weindruch, R. (2002), "A Mixture Model Approach for the Analysis of Microarray Gene Expression Data," *Computational Statistics and Data Analysis*, 39 120.

[3] Bickel, P. and Chernoff, H. (1993), "Asymptotic Distribution of the Likelihood Ratio Statistic in a Prototypical Non-Regular Problem," *In Statistics and Probability: a Raghu Raj Bahadur Festschrift (eds J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao)*, pp. 8396. New Delhi: Wiley.

[4] Billingsley, P. (1968), "Convergence of Probability Measures", *Wiley, New York*.

[5] Blalock, E., Chen, K., Sharrow, K., Herman, J.P., Porter, N.M., Foster, T.C., Landfield, P.W. (2003), "Gene Microarrays in Hippocampal Aging: Statistical Profiling Identifies Novel Processes Correlated with Cognitive Impairment," *The Journal of Neuroscience*, 23(9): 38073819.

[6] Breheny, P., Li, Y., Charnigo, R. (2012), "Statistical Challenges and Opportunities in Copy Number Variant Association Studies," *Journal of Biometrics and Biostatistics*, 3: e118.

[7] Casella, G. and Berger, R. L. (2002), *Statistical Inference (2nd Edition).*, Duxbury.

[8] Chambaz, A. (2006), "Testing the Order of a Model," *Annals of Statistics*, 34, 1166-1203.

[9] Charnigo, R., Sun, J. (2004), "Testing Homogeneity in a Mixture Distribution via the L2 Distance between Competing Models," *Journal of the American Statistical Association*, Vol. 99, No. 466, pp. 488- 498.

[10] Charnigo, R., Chesnut, L.W., LoBianco, T., Kirby, R.S. (2012), "Thinking Outside the Curve, Part I: Modeling Birthweight Distribution", *BMC Pregnancy and Childbirth*, 10:37.

[11] Charnigo, R. and Sun, J. (2010), "Asymptotic Relationships Between the D-test and Likelihood Ratio-Type Tests for Homogeneity," *Statistica Sinica*, 20, 497-512.

[12] Charnigo, R., Fan, Q., Bittel, D., and Dai, H. (2013), "Testing Unilateral versus Bilateral Normal Contamination," *Statistics and Probability Letters*, 83, 163-167.

[13] Chernoff, H. and Lander, E. (1995), "Asymptotic Distribution of the Likelihood Ratio Test that a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 1940.

[14] Chen, H., Chen, J., Kalbfleisch, J.D. (2001),"A Modified Likelihood Ratio Test for Homogeneity in Fnite Mixture Models," *Royal Statistical Society*, 63, Part 1, pp. 19-29.

[15] Chen, H., Chen, J. (2001), "The Likelihood Ratio Test for Homogeneity in the Finite Mixture Models", *The Canadian Journal of Statistics*, Vol. 29, No. 2, 2001.

[16] Chen, J., Li, P. (2009), "Hypothesis test for normal mixture models: The EM approach," *Annals of Statistics*, Vol. 37, No. 5A, 2523-2542.

[17] Dacunha-Castelle, D., Gassiat, E. (1999), "Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary Arma Processes", *The Annals of Statistics*, Vol. 27, No. 4 (Aug., 1999), pp. 1178-1209.

[18] Dai, H., Charnigo, R. (2008), "Omnibus Testing and Gene Filtration in Microarray Data Analysis," *Journal of Applied Statistics*, 35:1, 31-47.

[19] Dai, H., Charnigo, R. (2010), "Contaminated Normal Modeling with Application to Microarray Data Analysis," *The Canadian Journal of Statistics*, Vol. 38, No. 3, 315332.

[20] Dempster, A.P., Laird, M.N., Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Vol. 39, No. 1, pp. 1-38.

[21] Di, C., Liang, K. (2011), "Likelihood Ratio Testing for Admixture Models with Application to Genetic Linkage Analysis," *Biometrics*, 67, 1249-1259.

[22] Elkabets, M., Gifford, A.M., Christina, S., Nilsson, B., Reinhardt, F., Bray, M., Carpenter, A.E., Jirstrom, K., Magnusson, K., Ebert, B.L. (2011), "Human Tumors Instigate Granulin-Expressing Hematopoietic Cells that Promote Malignancy by Activating Stromal Fibroblasts in Mice," *The Journal of Clinical Investigation*, Vol 121, Issue 2, 784-799.

[23] Geweke, J., Amisano, G. (2007), "Normal Mixture Models with Applications to Financial Asset Returns," *Working Paper Series*, No 831, November 2007.

[24] Gibson, G. (2002), " Microarrays in Ecology and Evolution: A Preview," *Mol. Ecol.*, 11, 1724.

[25] Gifford, A.M., Nilsson, B., McAllister, S.S., Ebert, B.L. (2010), "Gene Profile Analysis of Sorted Sca1+/cKit- BMCs Obtain from Mice Bearing Instigator, Non-Instigator and Matrigel", *National Center for Biotechnology Information.*

[26] Hartigan, J. (1985), "A Failure of Likelihood Asymptotics for Normal Mixtures.", *In Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer (L. M. Le Cam and R. A. Olshen, eds.)* , 2, 807-810, Wadsworth.

[27] Herman, J.G., Merio A., Mao L., Lapidus R.G., Issa, J.P., Davidson, N.E., Sidransky D., Baylin, S.B. (1995), "Inactivation of the CDKN2/p16/MTS1 Gene is Frequently Associated with Aberrant DNA Methlation in All Common Human Cancers," *Cancer Research* 55(20): 4525-30.

[28] Lemdani, M., Pons, O. (1999), "Likelihood Ratio Tests in Contamination Models," *International Statistical Institute (ISI) and Bernoulli Society for Mathematical Statistics and Probability*, Vol. 5, No. 4, pp. 705-719.

[29] Li, P., Chen, J., Marriott, P. (2009), "Non-Finite Fisher Information and Homogeneity: an EM Approach", *Biometrika*, Volume 96, Issue 2, pp. 411-426.

[30] Li, P., Chen, J. (2010), "Testing the Order of a Finite Mixture," *Journal of the American Statistical Association*, Vol 105 Issue 491.

[31] Lindsay, B. G. (1989), "Moment matrices: applications in mixtures", *Ann. Statist.*, 17, 722740.

[32] Lindsay, B. G. (1995), "Mixture Models: Theory, Geometry and Applications," *Institute of Mathematical Statistics*.

[33] Liu, X., Shao, Y. (2003), "Asymptotics for the Likelihood Ratio Test in a Two-Component Normal Mixture Model," *Journal of Statistical Planning and Inference*, 123 (2004) 6181.

[34] Mao, R., Wang, X., Spitznagel, E., Frelin, L., Ting, J., Ding, H., Kim, J., Ruczinski, I., Downey, T., and Pevsner, J. (2005), "Primary and Secondary Transcriptional Effects in the Developing Human Down Syndrome Brain and Heart", *Genome Biology*, 6, 107.

[35] McLachlan, G. (1987), "On Bootstrapping the Likelihood Ratio Test Statistic for the Number of Components in a Normal Mixture", *Appl. Statist*, 36, 318-324.

[36] McLachlan, G. and Peel, D. (2000), "Finite Mixture Models", *Wiley*.

[37] Redner, R.A., Walker, H.F. (1984), "Mixture Densities, Maximum Likelihood and the EM Algorithm", *Society for Industrial and Applied Mathematics*, Vol. 26, No. 2, pp. 195-239.

[38] Self, S.G., Liang, K. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions", *Journal of the American Statistical Association*, Vol. 82, No. 398 (Jun., 1987), pp. 605- 610.

[39] Sun, J. (1993), "Tail Probabilities of the Maxima of Gaussian Random Fields," *The Annals of Probability*, 21, 3471.

[40] Titterington, D. M., Smith A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Wiley.

[41] Vandiedonck, C., Taylor, M.S., Lockstone, H.E., Plant, K., Taylor, J.M., Durrant, C., Broxholme, J., Fairfax, B.P., Knight, J.C. (2011), "Pervasive Haplotypic Variation in the Spliceo-Transcriptome of the Human Major Histocompatibility Complex" *Genome Research*, 21(7): 1042-1054.

[42] Wald, A. (1949), "Note on the Consistency of the Maximum Likelihood Estimate", *The Annals of Mathematical Statistics*, Vol. 20, No. 4, pp. 595-601.

[43] Wilcox A.J., Russell I.T. (1983), "Birthweight and Perinatal Mortality: II. On Weight-Specific Mortality" *Int J Epidemiol*, 12:319-325.

[44] Wilks, S.S. (1938), "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," *The Annals of Mathematical Statistics*, Vol. 9, No. 1, pp. 60-62.

[45] Zhu, H. and Zhang, H. (2004), "Hypothesis Testing in Mixture Regression Models", *Royal Statistical Society*, 66, Part 1, pp. 316, 2004.

Vita

Education

- M.S. in Statistics, University of Kentucky, Aug 2009 - May 2011.
- B.S. in Statistics, East China Normal University, Sep 2005 - Jun 2009.

Employment

- Statistical Consultant, University of Kentucky Applied Statistics Lab, Aug 2012 - Jan 2013.
- Research Assistant, Department of Statistics, University of Kentucky, Jan 2011 - Aug 2012.
- Teaching Assistant, Department of Statistics, University of Kentucky, 2009-2010& 2013-2014.
- Intern, Pfizer Global R & D, China, Jul 2008 - Jun 2009.