Concordant Integrative Analysis of Multiple Gene Expression Data Sets

By Fanni Zhang

B.S. in Applied Mathematics, July 2008, Peking University, Beijing, China
M.S. in Statistics, May 2010, The George Washington University, Washington, DC

A Dissertation Submitted to

The Faculty of
Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 18, 2014

Dissertation directed by

Yinglei Lai
Associate Professor of Statistics

The Columbian College of Arts and Sciences of The George Washington University certifies that Fanni Zhang has passed the Final Examination for the degree of Doctor of Philosophy as of March 6, 2014. This is the final and approved form of the dissertation.

Concordant Integrative Analysis of Multiple Gene Expression Data Sets

Fanni Zhang

Dissertation Research Committee:

Yinglei Lai, Associate Professor of Statistics, Dissertation Director

Michael D. Larsen, Associate Professor of Statistics, Committee Member

Qing Pan, Assistant Professor of Statistics, Committee Member

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Yinglei Lai for guiding and supporting me over the three years. During these years, he has set an example of excellence as a researcher, mentor and instructor. He has showed me how to research a problem and achieve goals. He has always encouraged me to express my ideas. He has spent countless hours reviewing and proofreading my manuscripts, patiently correcting my writing, asking me questions to let me think harder, and supported me during the difficult times in my research. Our weekly progress emails and meetings have been our means of communication and way to keep in touch, which have given me the motivation to work harder and make the best effort possible. I would never have been able to finish my dissertation without his excellent guidance, caring, patience and encouragement.

I would also like to thank Professor Michael D. Larsen and Professor Qing Pan for being my readers. They have dedicated their time and effort to help me improve this document. Their interesting ideas, valuable suggestions and comments have undoubtedly augmented this work. I also wish to thank Dr. Paul Albert and Professor Hua Liang for being my examiners. Additionally, I am grateful to Professor Bose for being my dissertation defense committee chair. I truly appreciate all of their time and assistance.

Thanks also go to Professor Reza Modarres, Professor Hosam Mahmoud, Professor Zhaohai Li and Professor Tapan Nayak for their time and help in guiding me through every step of my master and doctoral programs.

I am also thankful to the Department of Statistics and Columbian College of Arts and Sciences for providing financial support for my master and doctoral program studies.

Sincere gratitude to Mr. David Harrington for all his friendly help during the six years. I cannot leave out my wonderful colleagues and friends since they not only gave me a lot of support but also made my long journey much more cheerful.

My special appreciations go to my family to whom I owe so much. I thank my parents for their love and endless support not only throughout the few years of Ph. D. study but throughout my entire life.

Finally, to my caring, loving, and supportive husband, Siyu Qing: my warmest thanks. I am truly grateful to meet my dearest Siyu in this department and share this entire wonderful journey with him. He is my most enthusiastic cheerleader; he is my best friend; and he is an amazing husband. Thanks for always standing by me, comforting me, believing in me and giving me the courage and motivation to keep going.

Abstract of Dissertation

Concordant Integrative Analysis of Multiple Gene Expression Data Sets

Microarray is an experimental method by which tens of thousands of genes can be printed on a small chip. This technology enables us to measure genome-wide expression profiles. The cost of a microarray experiment is still relatively high. Therefore, the sample size of a microarray experiment is still relatively small. For some important disease studies, microarray data have been collected by different laboratories. We expect to obtain more efficient analysis results if different data sets collected for the same or similar study can be integrated. However, due to many complicated experimental issues, it is necessary to evaluate the genome-wide concordance among these data sets before their integrative analysis. If the underlying behavior of a gene is consistent among different experiments, then the related expression profiles in different data sets will be concordant. Statistically, mixture models have been widely used to accommodate unobserved heterogeneities in a study population. A mixture model based method has been proposed for the integrative concordant analysis when there are two microarray data sets available for an integrative analysis. It is necessary to extend this approach for an integrative analysis of multiple data sets.

The general statistical framework for our integrative analysis is the partial concordance/discordance (PCD) model. Its related statistical estimation difficulty is that its parameter space increases exponentially with the number of data sets. Since the complete concordance model (CC) and the complete independence (CI) model are two basic statistical frameworks that can be derived from the PCD model, we propose a two-level mixture model to approximate the PCD model. It combines the basic CC and CI models and its parameter space increases linearly with the number

of data sets. We have implemented an expectation-maximization algorithm for the model parameter estimation. Simulation studies have been conducted to understand the performance of our method. We have also applied our method to a collection of microarray gene expression data sets for a lung cancer study.

Furthermore, we have also developed other approaches to decrease the parameter space of PCD model by simplifying the non-diagonal proportion parameters. The inspiration comes from the exchangeable structure and AR(1) structure in GEE, as well as the multiset coefficient in combinatorics. We still consider expectation-maximization algorithm to achieve the model fitting. The performance of the proposed methods is examined using simulation studies. We have also compared these methods with the two-level mixture model based method through applications to the same experimental data sets from the lung cancer study.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

DNA is the basic blueprint for making all living beings. It provides the starting template for every new life, and almost every cell of living organisms contains DNA. It is a long molecule, like a chain, composed of pieces called nucleotides (sometimes also called bases) [1]. There are four different types of nucleotides in DNA which are called Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). They are repeated over and over in pairs according to Watson-Crick complementary base paring rules: A pairs with T, G pairs with C [2]. The complementary structure of the DNA molecule allows every DNA molecule to make an exact copy of itself.

A gene is a specific contiguous sequence of DNA [3]. All the genetic information about living beings passed from parent to offspring is contained in genes. Because of this mechanism, family members tend to exhibit similar characteristics or traits. Some traits are part of an organism's physical appearance like a person's hair color or finger length, while some traits are not easily observed like blood types or predisposition toward certain diseases. Single gene or combination of different genes provides instructions for producing some property of its organism [3]. The entire set of genes in an organism is called its genome.

Although the genetic material in all the cells of the human body are identical, the genes which are active in some cells may not be active in some other cells [4]. Scientists have been studying gene activity in different cells to estimate how the cells function normally and how they are influenced by gene's performance. In the past, researchers have been able to conduct the genetic analyses on only a few genes at the same time. With the development of DNA microarray technology (see Section 1.1), scientists can now analyze thousands of genes in one experiment. This approach can examine the expression of many genes (see Section 1.2) in a single reaction quickly and in an efficient manner [5].

1.1 Microarray Technology

DNA microarrays technology provides a high-throughput method for measuring the expression levels of many genes simultaneously. It enables researchers to learn more about many different diseases, such as heart disease, mental illness and infectious diseases [6]. This technology has been intensely used in cancer study. In the past, researchers have classified different types of cancers based on the organs in which the tumors develop. While, they can further classify these types of cancers based on the patterns of gene activity in the tumor cells thanks to the microarray experiments [7]. Then treatment strategies can be devised to target each specific type of cancer. Moreover, by studying the differences in gene activity between untreated and treated tumor cells, scientists can analyze the different effects of therapies on tumors and be able to develop more effective treatment methods [8].

Each gene contains information for producing protein. The intermediate step between gene and protein is messenger RNA (mRNA). The fundamental idea behind microarrays is to exploit complementary base pairing to estimate the expression levels of genes by measuring the amount of the various types of mRNA molecules in a cell [9].

DNA Microarrays are small, solid supports onto which the sequences from thousands of different genes are attached or immobilized at fixed locations [10]. The supports themselves are usually glass microscope slides. The slide is theoretically divided into a rectangular grid, creating thousands of spots. Each of these spots will contain DNA which is representing the gene targeted. There are currently two types of DNA microarrays that are commercially available, two channel cDNA (complementary DNA) microarrays and high-density oligonucleotide microarrays [11].

To determine which genes are turned on and which are not in a given cell, a researcher must first collect the mRNA molecules present in that cell. In a cDNA microarray experiment, the researcher then labels each mRNA molecule by using a reverse transcriptase enzyme that generates a complementary DNA to the mRNA. During that process fluorescent nucleotides are attached to the cDNA. The test and the reference samples are labeled with different fluorescent dyes. Next, the researcher places the labeled cDNAs onto a DNA microarray slide. The labeled cDNAs that represent mRNAs in the cell will then hybridize or bind to their synthetic complementary DNAs attached on the microarray slide, leaving its fluorescent tag. A researcher must then use a special scanner to measure the fluorescent intensity for each spot on the microarray slide [12].

If a particular gene is very active, it produces many molecules of messenger RNA, thus, more labeled cDNAs, which hybridize to the DNA on the microarray slide and generate a very bright fluorescent area. Genes that are somewhat less active produce fewer mRNAs, thus, less labeled cDNAs, which results in dimmer fluorescent spots. If there is no fluorescence, none of the messenger molecules have hybridized to the

DNA, indicating that the gene is inactive. Researchers frequently use this technique to examine the activity of various genes at different times. When co-hybridizing test samples (Red Dye) and reference samples (Green dye) together, they will compete for the synthetic complementary DNAs on the microarray slide. As a result, if the spot is red, it means that the specific gene is more expressed in test samples than in reference ones (up-regulated, see Section 1.2). If a spot is Green, it means that the gene is more expressed in reference samples (down-regulated, see Section 1.2). If a spot is yellow that means the specific gene is equally expressed in reference and test samples.

In oligonucleotide microarrays, short DNA oligonucleotides are spotted onto the array [13]. It contains more than one probe per gene, which is the main feature of oligonucleotide microarray. The entirety of probes mapping to different regions of the gene is usually named probe set [13]. Typically, each gene will be represented by 16-20 pairs of oligonucleotides. Each probe pair consists of a perfect match oligonucleotide and a mismatch oligonucleotide. The perfect match probe can measure the expression of the specific gene, because the probe sequence is exactly complimentary to this gene. Commercial oligonucleotide chips are now widely used such as Affymetrix's GeneChip system. Their arrays contains hundreds of thousands of oligonucleotide probes. For technical reasons, only a single sample can be measured on one of these chips, thus oligonucleotides microarray experiments are often called single channel hybridization [14]. It implies that two separate hybridizations need to be performed to get data of differential expression between one test sample and the other reference sample. So, the expression level for a gene need to be summarized by combining the intensities of the probe set for the gene [14].

Due to the expense of such experiments, only a limited number of replicates are typically affordable for each condition. Small sample size may lead to a low statistical power. Nevertheless, the microarray technology is one of the excellent methods in the field of genomics, and becomes a major tool in our understanding and cataloging of the human genome.

1.2 Gene Expression and Differential Expression

Gene expression is the process by which information from a gene is used in the synthesis of a functional gene product, usually proteins. Genes encode proteins and proteins dictate cell functions [1]. Hence, what the cell can do and how the cell works are determined by the genes expressed in this cell.

Gene expression consists of two major steps: transcription and translation. During the process of transcription, the information carried in a gene's DNA is transferred to a similar molecule called RNA (ribonucleic acid) in the cell nucleus [15]. Similarly as DNA, RNA is also a long chain consisting of nucleotide bases, but they have slightly different chemical properties. There is one type of RNA called messenger RNA, because it carries the information for encoding a protein from DNA out of nucleus into cytoplasm. Translation, the second step is that mRNA functions to make a protein in the cytoplasm. When mRNA enters cytoplasm, it becomes associated with a specialized complex called a ribosome. Another type of RNA called transfer RNA carries a specific amino acid, pairs up with the mRNA bases inside the ribosome. Each sequence of three bases, called a codon, usually codes for one particular amino acid [15]. As the ribosome moves along the mRNA, the tRNA transfers its amino acid to produce the protein chain. Protein continues to assemble until the ribosome hits a "stop" codon [15]. This stop codon is a sequence of three bases that does not code for an amino acid. This whole process "DNA makes RNA makes protein" is called Central Dogma of Molecular Biology [16].

Only a fraction of the genes in a cell are expressed according to the needs at one time, the rest of the genes are repressed. The process of turning genes on and off is known as gene regulation [17]. Gene regulation also allows cells to react quickly to changes in their environments. Down-regulation is the process by which a cell decreases the quantity of a cellular component, such as RNA or protein, in response to an external variable. An increase of a cellular component is called up-regulation.

A gene is declared differentially expressed if an observed difference or change in expression between two experimental conditions is statistically significant. Frequently, to understand the effect of a drug we may concern about which genes are up-regulated or down-regulated between treatment and control group. Up-regulation occurs, for example, when a cell is deficient in some kind of receptor. In order to bring the cell back to the normal sensitivity level, more receptor protein is synthesized and transported to the membrane of the cell [1]. Down-regulation occurs, for example, when a cell is overstimulated by a neurotransmitter, hormone, or drug for a prolonged period of time, in order to protect the cell the expression of the receptor protein is decreased [1]. The expression levels for a gene in the treatment group and control group will be summarized as the mean of the expression levels in the group replicates. Thus, the analysis of differential expression problems is a comparison of means. An up-regulated gene has a positive population mean difference, while a down-regulated one has a negative population mean difference. When there are two sample groups, a t-test can be considered as one of the typical methods for comparing the means.

1.3 Motivation and Review of Integrative Analysis of Multiple Data Sets

1.3.1 Biological Review

The study of gene expression profiling has played a major role for discovery in biomedical research [18]. Microarray experiments give researchers the opportunity to simultaneously measure the expression level of thousands of genes within a particular mRNA sample. In microarray analysis, a goal is to identify the genes that are differentially expressed between the control and treated samples. Typically, statistical approach for two-sample comparison used to analyze microarray data is hypothesis testing. The use of DNA microarrays is highly beneficial to our understanding of genes and their influences on disease, drug discovery and development.

In the recent past, microarray technology has been one of the fastest-growing new technologies in the field of genetic research. But the cost of a microarray experiment is still relatively high, such as the initial cost of designing, fabricating custom, and the expense for replication of experiments to minimize statistical variability of results. As mentioned in Section 1.1, the sample size of a microarray experiment is still relatively small, which leads to a low detection power, particularly when differential expression signals are considerably weak [19].

As microarray technologies become routinely applied in genome laboratories, it is common that microarray experiments on identical or similar sets of genes are repeatedly conducted by various laboratories for different functional studies of these genes [20]. Because of the interest in increasing the detection power, several methods have been suggested to integrate different data sets in differential expression analysis.

Meta-analysis is a quantitative method of combining the results of individual research studies and synthesizing conclusions to evaluate the effectiveness of treatments or procedures [21]. This method is based on explicit and verifiable criteria, which may avoid selection bias [22]. Not only it can statistically test study heterogeneity and investigate explanatory variables, but also it can statistically summarize results to obtain an overall estimate of treatment effect. Compared with any single study, usually meta-analysis has increased power for statistical tests, and increased precision for confidence intervals. Since the relevant studies are carried out in different places, meta-analysis may also produce more generalizable results. Given the small sample size of many linkage and microarray experiments, meta-analysis might be considered a natural approach to the research integration. The general steps [21] involved in doing a meta-analysis:

1. Create a study design. Before collecting studies, it is important to decide which ones will be included and which will be excluded.

2. Search the literature. Once determining the scope of the meta-analysis, we need to locate all of the studies that fit within the scope.

3. Extract the information. Based on quality criteria, we need to select of specific studies on a well-specified study subject. We also need to decide whether unpublished studies should be included to avoid publication bias.

4. Analyze the meta-analytic database. In this step, statistical analyses should be performed to determine the overall strength and consistency of the effect of treatment.

5. Interpret the results. We need describe the implication of our analysis, report any limitation regarding the analysis and also suggest the future research.

In recent years, meta-analysis has been developed for microarray studies. There are two primary methods for data integration. One is combining gene expression measures across studies, the other is combining summary measures of expression such as p-values, probabilities or ranks [23]. Meta-analysis was applied to combine four datasets on prostate cancer to determine genes that are differentially expressed between clinically localized prostate and benign tissue [24]. Another meta-analysis was performed of four independent studies that applied high-density arrays for expression profiling of pancreatic cancer [25].

Additionally, many other methods have been illustrated for integrating multiple data sets. Combining both of advantages from meta-analytic and correlation-based techniques, a scalable method-Microarray Experiment Functional Integration Technology (MEFIT) has been presented for microarray datasets integration [26]. A Bayesian hierarchical model has been introduced to incorporate data from multiple independent cDNA microarray studies to identify highly expressed genes [27]. Two minimax concave penalty (MCP) based penalization approaches have been applied for marker selection under the heterogeneity model to the analysis of gene expression data on multiple cancers [28].

Genes may show same or different behavior in multiple experiments. What we concern about are those genes with same behavior in multiple experiments. We expect to obtain more efficient analysis results if different data sets collected for the same or similar study can be integrated. To avoid misleading results from many complicated experimental issues, it is necessary to evaluate the genome-wide concordance among these data sets before their integrative analysis. If the underlying behavior of a gene is consistent among different experiments, then the related expression profiles in different data sets will be concordant. By integrating them, the sample size can be increased and the noise impact can be reduced.

In the analysis of differential expression, a gene can show three distinct states: no change (null), positive change (up-regulated), and negative change (down-regulated). If we know its state, then the related statistical inference will be relatively simple. When the state information is unknown, a mixture model (see Section 1.3.2) can be generally considered. Statistically, mixture models have been widely used to accommodate unobserved heterogeneities in a study population over years. A mixture model based method has been proposed for the integrative concordant analysis when there are two microarray data sets available for an integrative analysis [32]. It is necessary to extend this approach for an integrative analysis of multiple data sets.

1.3.2 Statistical Review

Statistically, mixture models have been widely used to accommodate unobserved heterogeneities in a study population over years [29]. A mixture model based method has been proposed for the integrative concordant analysis when there are two microarray data sets available for an integrative analysis [19]. Expectation-Maximization algorithm has been implemented for the model parameter estimation.

Mixture Models

Mixture modeling provides a natural framework for unobserved heterogeneity in a population. There exist various features of finite mixture distributions, which made them useful in statistical modeling. On one hand, finite mixture distributions arise in a natural way as marginal distributions for statistical models involving discrete latent variables. On the other hand, it turns out that statistical models that are based on finite mixture distributions are able to capture many specific features of real data, such as multimodality, skewness, and kurtosis. Statistical modeling based on finite mixtures of normal distributions are frequently used in many areas of applied statistics such as biology, economics, marketing, medicine, or physics.

Assume all observations are a mixture of samples from different distributions with unknown component information. The density of a g-component mixture distribution can be written in the form of $f(x) = \sum_{i=1}^{g} \pi_i f_i(x)$, where π_i is defined as the proportion of the i-th component with constrains $0 \leq \pi_1, \pi_2, \ldots, \pi_g \leq 1$, $\sum_{i=1}^{g} \pi_i = 1$ [30]. A model must be identifiable, which means that parameters have unique solutions.

The shape of a mixture density is extremely flexible. A mixture density may be unimodal or multimodal, may exhibit considerable skewness. For this reason finite mixture distribution offer a flexible way to describe relatively unsmooth data structure [30]. Indeed, any continuous distribution can be approximated well by a finite mixture of normal densities with common variance [30].

Testing for the number of components g in a mixture is an important but very difficult problem which has not been completely resolved. Because it involves inference for an overfitted mixture model where the true number of components is less than the number of components in the fitted mixture model. Likelihood-based methods, including AIC, BIC as well as parametric bootstrap method have been put forward to deal with model specification uncertainty for a suitable density estimate [30] [31].

Existing Related Statistical Work

A three-component normal-mixture model approach has been presented to assess the concordance and discordance between two large-scale experiments with two-sample groups [32]. In this study, two separate experiments were conducted to identify whether each of m features can significantly distinguish a disease group from a normal group. After a normal-distribution quantile-based transformation of p-values of two hypothesis tests for each feature, they obtained paired z-scores. A three-component normal-mixture model for the joint distribution of paired z-scores was considered, which had the density as,

$$f_{PCD}(z^{(1)}, z^{(2)}) = \sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_i, \sigma_i^2}(z^{(1)}) \phi_{\nu_j, \tau_j^2}(z^{(2)}), \tag{1.1}$$

where ϕ_{μ,σ^2} is the normal probability distribution function with mean μ and variance σ^2. For $g = 3$, we use the second component ($i = 2$) to represent the null component (no population mean difference) with fixed parameters. Hence, we consider our models with constraints on means and variances of the null component: $\mu_2 = \nu_2 = 0$ and $\sigma_2^2 = \tau_2^2 = 1$. Indices 1 and 3 are used to represent the down-regulated and up-regulated components with constrains: $\mu_1, \nu_1 \leq 0$, $\mu_3, \nu_3 \geq 0$. π_{ij} is the proportion of genes consistent with component i in the first study and component j in the second study with constraint $\sum_{i,j=1}^{3} \pi_{ij} = 1$. The partial concordance/discordance (PCD) model could be reduced to the models for evaluating complete concordance (CC) or complete independence (CI) between the results from two experiments. The estimated parameters in the PCD, CC and CI models were used for calculation of likelihoods of three models. Based on these log-likelihoods, they conducted two likelihood ratio tests for PCD (H_1) against CC (H_0) models and PCD (H_1) against CI (H_0) models, respectively. With these tests, they evaluated the concordance and the discordance between these two experiments.

Later, a statistical method has been proposed for an integrative analysis of differential expression based on two microarray gene expression data sets [19]. First, for each gene, a hypothesis for differentially expresseion was tested through two-sample t-test. After a normal-distribution quantile-based transformation of p-values, they obtained paired z-scores. Then, the mixture-model based method proposed was simple and intuitive for testing genome-wide concordance and discordance before the data integration. As described in the paper, for the two lists of z-scores, first CI would be tested between them. If CI could not be rejected, then the data integration would be discouraged; otherwise, continue to test CC between them. If CC could not be rejected, then the concordant integrative scores would be calculated to prioritize genes based on the CC model; otherwise the concordant integrative scores would be calculated to prioritize genes based on the PCD model. This concordant integrative score was defined as the conditional probability of concordant differential expression under an appropriate mixture model:

Pr (concordant differential expression|observed pair of z-scores)
=[Pr(observed pair of z-scores both up-regulated) + Pr(observed pair of z-scores both down-regulated)]/Pr (observed pair of z-scores).

Currently, this framework for data integration focuses on two data sets with two sample groups. It is necessary to extend this approach for an integrative analysis of multiple data sets.

EM Algorithm for Normal Mixture Model

The EM algorithm is a general method for finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values [33]. It proceeds iteratively in two steps, E (for expectation) and M (for maximization). Given a statistical model consisting of a set of observed data $\mathbf{x}$, a set of unobserved latent data or missing values $\mathbf{z}$, and a vector of unknown parameters Θ. For this setting, the complete data are $(\mathbf{x}, \mathbf{z})$. Then we can derive the complete-data likelihood $L(\mathbf{x}, \mathbf{z}|\Theta)$. The EM algorithm first finds the expected value of the complete likelihood with respect to the unknown data $\mathbf{z}$ given the observed data $\mathbf{x}$ and the current parameter estimates. Define $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x}, \mathbf{z}|\Theta')|\Theta]$, where Θ are the current parameters estimates that we used to evaluate the expectation and Θ' are the new parameters that we optimize to increase Q. Then we need to find the parameters that maximize this Q function. Although the EM iteration does increase the observed data likelihood function, there is no guarantee that the sequence converges to a global maximum likelihood estimator. For multimodal distributions, this means that the EM algorithm may converge to a local maximum of the observed data likelihood function, depending on starting values. Also, in practice, the convergence of the EM algorithm may be desperately slow in some case because this algorithm converges linearly near the solution [34]. However, due to its main advantages - simplicity and ease of implementation, the EM algorithm is widely used for the fitting of finite mixture models, especially normal mixture models when we know the component information [35].

For the exsiting studies on concordant integrative analysis for two data sets [19][32], they assumed that a mixture of g normal components for each individual data set. For a better introduction of the extended method in Chapter 2, we need to reivew the base idea of their method for univariate and bivariate normal mixture model. For convenience, we use variable x instead of z to denote the z-scores obtained from the p-values of the tests. In addition, we use variable z to denote the missing values in the augmented data sets. Suppose each component is a normal distribution with mean μ_i and variance σ_i^2. The density of i^{th} component is $f_i(x) = \phi_{\mu_i,\sigma_i^2}(x)$. Let the weight of i^{th} component be π_i, $\sum_{i=1}^{g} \pi_i = 1$. The mixture density is

$$f(x) = \sum_{i=1}^{g} \pi_i f_i(x) = \sum_{i=1}^{g} \pi_i \phi_{\mu_i,\sigma_i^2}(x). \tag{1.2}$$

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ denote the n observations, then

$$L = f(\mathbf{x}) = \prod_{k=1}^{n} f(x_k) = \prod_{k=1}^{n} [\sum_{i=1}^{g} \pi_i \phi_{\mu_i,\sigma_i^2}(x_k)]. \tag{1.3}$$

The latent indicator variables $z'_{ik}s$ are defined as below,

$$z_{ik} = \begin{cases} 1 & \text{if } x_k \text{ is sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

It is known that $\pi_i = \Pr(z_{ik} = 1)$.

Then the complete-data likelihood and log-likelihood can be derived:

$$L^*(\mathbf{x}, \mathbf{z}|\Theta) = \prod_{k=1}^{n} \prod_{i=1}^{g} [\pi_i \phi_{\mu_i, \sigma_i^2}(x_k)]^{z_{ik}}, \tag{1.4}$$

$$l^* = \log L^* = \sum_{k=1}^{n} \sum_{i=1}^{g} z_{ik} \log[\pi_i \phi_{\mu_i, \sigma_i^2}(x_k)]. \tag{1.5}$$

The expected values of $z'_{ik}s$ can be obtained for the E-step:

$$\mathrm{E}(z_{ik}|\Theta) \quad = \quad \frac{\pi_i \phi_{\mu_i, \sigma_i^2}(x_k)}{\sum_{i=1}^{g} \pi_i \phi_{\mu_i, \sigma_i^2}(x_k)}, \tag{1.6}$$

where $\Theta = \{\pi_i, \mu_i, \sigma_i, \text{ for } i = 1, 2, \ldots, g\}$ are the current parameters estimates used to evaluate the expectation.

To maximize the log-likelihood with respect to π_i, μ_i and σ_i^2, the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x}, \mathbf{z}|\Theta')|\Theta]$ are taken to obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}{n},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n} [\mathrm{E}(z_{ik}|\Theta) x_k]}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},$$

$$\hat{\sigma}_i'^2 = \frac{\sum_{k=1}^{n} [\mathrm{E}(z_{ik}|\Theta)(x_k - \mu_i)^2]}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}. \tag{1.7}$$

The mathematical derivations for (1.6) and (1.7) can be found in Section 1.3.3.

They proposed their normal mixture model for a concordant analysis when two list of z-scores were obtained from two experiments. For convenience, we use variable x and y instead of z to denote the z-scores obtained from the p-values of the tests in two experiments. In addition, we use variable z and w to denote the missing values in the augmented data sets. Suppose there are n pairs of obseravations $(\mathbf{x}, \mathbf{y})$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. The joint normal probability distribution function for the k^{th} pair of observations of a g-component normal-mixture model is

$$f_{PCD}(x_k, y_k) = \sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k), \tag{1.8}$$

where π_{ij} is the proportion of that x_k belongs to the i^{th} component and y_k belongs to the j^{th} component. Here, x_k and y_k do not necessarily belong to the same component. This model is referred as partial concordance/discordance (PCD) model.

For PCD model, the latent variables z_{ik} and w_{jk} are defined as indicators as below,

$$z_{ik} = \begin{cases} 1 & \text{if } x_k \text{ is sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

$$w_{jk} = \begin{cases} 1 & \text{if } y_k \text{ is sampled from the } j^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

It is known that $\pi_{ij} = \Pr(z_{ik} = 1, w_{jk} = 1)$, $\Pr(z_{ik} = 1) = \sum_{j=1}^{g} \pi_{ij} = \pi_{i.}$, $\Pr(w_{jk} = 1) = \sum_{i=1}^{g} \pi_{ij} = \pi_{.j}$.

Then the complete-data likelihood and log likelihood of PCD model are derived:

$$L_{PCD}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta) = \prod_{k=1}^{n} \prod_{i=1}^{g} \prod_{j=1}^{g} [\pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)]^{z_{ik} w_{jk}}, \tag{1.9}$$

$$l_{PCD} = \sum_{k=1}^{n} \sum_{i=1}^{g} \sum_{j=1}^{g} z_{ik} w_{jk} \log[\pi_i \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)]. \tag{1.10}$$

The expected values of $z_{ik} w_{jk}$ are calculated for the E-step:

$$\mathrm{E}(z_{ik} w_{jk} | x_k, y_k, \Theta) = \frac{\pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}{\sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}, \tag{1.11}$$

where $\Theta = \{\pi_{ij}, \mu_i, \sigma_i^2, \nu_j, \tau_j^2, \text{ for } i, j = 1, 2, \ldots, g\}$ are the current parameters estimates used to evaluate the expectation.

To maximize the log-likelihood with respect to π_i, μ_i, σ_i^2, ν_j and τ_j^2, the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta')|\Theta]$ are taken to obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi'}_{ij} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk}|\Theta)}{n},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n} \sum_{j=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta) x_k}{\sum_{k=1}^{n} \sum_{j=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)},$$

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta) y_k}{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)},$$

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n} \sum_{j=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)(x_k - \mu_i')^2}{\sum_{k=1}^{n} \sum_{j=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)},$$

$$\hat{\tau_j'^2} = \frac{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)(y_k - \nu_j')^2}{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk}|\Theta)}. \tag{1.12}$$

If $(\mathbf{x}, \mathbf{y})$ are completely concordant, then each pair $(x_k,\ y_k)$ will belong to the same component. Therefore, we expect $\pi_{ij} = 0$ for $i \neq j$ and $\pi_{i.} = \pi_{.i} = \pi_{ii}$. The PCD model is reduced to the complete concordance (CC) model.

For CC model, the joint normal probability distribution function for the k^{th} pair of observations is

$$f_{CC}(x_k, y_k) = \sum_{i=1}^{g} \pi_i \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k). \tag{1.13}$$

The latent z_{ik} is defined as an indicator as below,

$$z_{ik} = \begin{cases} 1 & \text{if both } x_k \text{ and } y_k \text{ are sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

Then, let $\pi_{i.} = \pi_{.i} = \pi_{ii} = P(z_{ik} = 1)$. The complete-data likelihood and log likelihood of CC model are:

$$L_{CC}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta) = \prod_{k=1}^{n} \prod_{i=1}^{g} [\pi_i \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k)]^{z_{ik}}, \tag{1.14}$$

$$l_{CC} = \sum_{k=1}^{n} \sum_{i=1}^{g} z_{ik} \log[\pi_i \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k)]. \tag{1.15}$$

E-step:

$$\mathrm{E}(z_{ik}|x_k, y_k, \Theta) = \frac{\pi_{i.} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k)}{\sum_{i=1}^{g} \pi_{i.} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k)}, \tag{1.16}$$

where $\Theta = \{\pi_i, \mu_i, \sigma_i^2, \nu_i, \tau_i^2, \text{ for } i = 1, 2, \ldots, g\}$.

M-step:

$$\hat{\pi'}_{i.} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}{n},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta) x_k}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},$$

$$\hat{\sigma}_i'^2 = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)(x_k - \mu_i')^2}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},$$

$$\hat{\nu'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta) y_k}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},$$

$$\hat{\tau}_i'^2 = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)(y_k - \nu_i')^2}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}. \tag{1.17}$$

The mathematical derivations can be found in Section 1.3.3.

1.3.3 Mathematical Derivation

Formula (1.6) and (1.7) derivations for E and M steps in Section 1.3.2

$$\begin{aligned} Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x},\mathbf{z}|\Theta')|\Theta] \\ &= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i=1}^{g} z_{ik}\log(\pi_i'\phi_{\mu_i',\sigma_i'^2}(x_k))|\Theta] \\ &= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)\log(\pi_i') \\ &\quad +\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)\log(\phi_{\mu_i',\sigma_i'^2}(x_k)) \end{aligned}$$

E-step:

$$\begin{aligned} \mathrm{E}(z_{ik}|\Theta) &= 0\times \Pr(z_{ik}=0|x_k,\Theta)+1\times \Pr(z_{ik}=1|x_k,\Theta) \\ &= \Pr(z_{ik}=1|x_k,\Theta) \\ &= \frac{\Pr(z_{ik}=1,x_k,\mu_i,\sigma_i^2))}{\Pr(x_k,\Theta)} \\ &= \frac{\Pr(z_{ik}=1)\Pr(x_k,\Theta|z_{ik}=1)}{\sum_{i=1}^{g}\Pr(z_{ik}=1)\Pr(x_k,\Theta|z_{ik}=1)} \\ &= \frac{\pi_i\phi_{\mu_i,\sigma_i^2}(x_k)}{\sum_{i=1}^{g}\pi_i\phi_{\mu_i,\sigma_i^2}(x_k)}. \end{aligned}$$

Note that, $\{z_{ik}=1\}$ is a set of pairwise disjoint events for $i=1,2,\ldots,g$, and $\sum_{i=1}^{g}\Pr(z_{ik}=1)=1$, therefore, by the law of total probability,

$$\Pr(x_k,\Theta)=\sum_{i=1}^{g}\Pr(z_{ik}=1)\Pr(x_k,\Theta|z_{ik}=1).$$

M-step: The optimization of π_i, μ_i, σ_i^2 is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta,\Theta')}{\partial\pi_i'}=0,$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial\mu_i'}=0,$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \sigma_i'^2} = 0.$$

Subject to $\sum_{i=1}^{g} \pi_i = 1$, then $\pi_g = 1 - \pi_1 - \pi_2 - \ldots - \pi_{g-1}$, solve the system of equations above,

$$\hat{\pi'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}{n},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta) x_k}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},$$

$$\hat{\sigma}_i'^2 = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)(x_j - \mu_i)^2}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}.$$

Formula (1.11) and (1.12) derivations for E and M steps in Section 1.3.2

$$\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} z_{ik}w_{jk}\log(\pi'_{ij}\phi_{\mu'_i,\sigma_i'^2}(x_k)\phi_{\nu'_j,\tau_j'^2}(y_k))|\Theta] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\pi'_{ij}) \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma_i'^2}(x_k)) + \log(\phi_{\nu'_j,\tau_j'^2}(y_k))]
\end{aligned}$$

E-step:

$$\begin{aligned}
\mathrm{E}(z_{ik}w_{jk}|x_k,y_k,\Theta) &= 0 \times \Pr(z_{ik}w_{jk}=0|x_k,y_k,\Theta) + 1 \times \Pr(z_{ik}w_{jk}=1|x_k,y_k,\Theta) \\
&= \Pr(z_{ik}=1, w_{jk}=1|x_k,y_k,\Theta) \\
&= \frac{\Pr(z_{ik}=1, w_{jk}=1, x_k, y_k, \Theta)}{\Pr(x_k,y_k,\Theta)} \\
&= \frac{\Pr(z_{ik}=1,w_{jk}=1) \times \Pr(x_k,y_k,\Theta|z_{ik}=1,w_{jk}=1)}{\sum_{i=1}^{g}\sum_{j=1}^{g}\Pr(z_{ik}=1,w_{jk}=1)\Pr(x_k,y_k,\Theta|z_{ik}=1,w_{jk}=1)} \\
&= \frac{\pi_{ij}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{\sum_{i=1}^{g}\sum_{j=1}^{g}\pi_{ij}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.
\end{aligned}$$

Note that, for any fixed i, $\{z_{ik} = 1, w_{jk} = 1\}$ is a set of pairwise disjoint events for $j = 1, 2, \ldots, g$, then for any i and j, $\{z_{ik} = 1, w_{jk} = 1\}$ is a set of pairwise disjoint events. Moreover, $\sum_{i=1}^{g} \sum_{j=1}^{g} \Pr(\mathrm{z_{ik}} = 1, \mathrm{w_{jk}} = 1) = 1$. Therefore, by the law of total probability,

$$\Pr(x_k, y_k, \Theta) = \sum_{i=1}^{g} \sum_{j=1}^{g} \Pr(z_{ik} = 1, w_{jk} = 1) \Pr(x_k, y_k, \Theta | z_{ik} = 1, w_{jk} = 1). \quad (1.18)$$

The optimization of $\pi_{ij}, \mu_i, \sigma_i^2, \nu_j$, and τ_j^2 is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta, \Theta')}{\partial \pi'_{ij}} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk} | \Theta)}{\pi'_{ij}} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk} w_{gk} | \Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \pi'_{ij}} = 0,$$

for $i, j = 1, 2, \ldots, g-1$ the above equation holds.

Subject to $\sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} = 1$,

$$\begin{aligned}
\frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk} | \Theta)}{\pi'_{ij}} &= \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk} w_{gk} | \Theta)}{\pi'_{gg}} \\
&= \frac{\sum_{i=1}^{g} \sum_{j=1}^{g} \sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk} | \Theta)}{\sum_{i=1}^{g} \sum_{j=1}^{g} \pi'_{ij}} \\
&= \sum_{i=1}^{g} \sum_{j=1}^{g} \sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk} | \Theta) \\
&= \frac{\sum_{k=1}^{n} \sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}{\sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)} \\
&= n,
\end{aligned}$$

then,

$$\hat{\pi'}_{ij} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik} w_{jk} | \Theta)}{n}.$$

$$\frac{\partial Q(\Theta, \Theta')}{\partial \mu'_i} = \sum_{k=1}^{n} \sum_{j=1}^{g} \mathrm{E}(z_{ik} w_{jk} | \Theta) \frac{x_k - \mu'_i}{\sigma_i'^2} = 0,$$

then,

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)x_k}{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{1.19}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \sigma_i'^2} = \sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\frac{(x_k-\mu_i')^2}{2\sigma_i'^4} - \frac{1}{2\sigma_i'^2}] = 0,$$

then,

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)(x_k-\mu_i')^2}{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}.$$

By symmetry,

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)y_k}{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\tau}_j'^2 = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)(y_k-\nu_j')^2}{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}.$$

Formula (1.16) and (1.17) derivations for E and M steps in Section 1.3.2

$$\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z}}[\log L_{CC}(\mathbf{x},\mathbf{y},\mathbf{z}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\sum_{k=1}^{n}\sum_{i=1}^{g} z_{ik}\log(\pi_i'\phi_{\mu_i',\sigma_j'^2}(x_k)\phi_{\nu_i',\tau_i'^2}(y_k))|\Theta] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)\log(\pi_i') \\
&\quad +\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)[\log(\phi_{\mu_i',\sigma_i'^2}(x_k)) + \log(\phi_{\nu_i',\tau_i'^2}(y_k))]
\end{aligned}$$

E-step:

$$
\begin{aligned}
\mathrm{E}(z_{ik}|x_k, y_k, \Theta) &= 0 \times \Pr(z_{ik} = 0|x_k, y_k, \Theta) + 1 \times \Pr(z_{ik} = 1|x_k, y_k, \Theta) \\
&= \Pr(z_{ik} = 1|x_k, y_k, \Theta) \\
&= \frac{\Pr(z_{ik} = 1, x_k, y_k, \Theta)}{\Pr(x_k, y_k, \Theta)} \\
&= \frac{\Pr(z_{ik} = 1)\Pr(x_k, y_k, \Theta|z_{ik} = 1)}{\sum_{i=1}^{g} \Pr(z_{ik} = 1)\Pr(x_k, y_k, \Theta|z_{ik} = 1)} \\
&= \frac{\pi_{i.}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{\sum_{i=1}^{g} \pi_{i.}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}.
\end{aligned}
$$

Note that, $\{z_{ik} = 1\}$ is a set of pairwise disjoint events for $i = 1, 2, \ldots, g$, and $\sum_{i=1}^{g} \Pr(z_{ik} = 1) = 1$, therefore, by the law of total probability,

$$
\Pr(x_k, y_k, \Theta) = \sum_{i=1}^{g} \Pr(z_{ik} = 1)\Pr(x_k, y_k, \Theta|z_{ik} = 1). \tag{1.20}
$$

M-step: Subject to $\sum_{i=1}^{g} \pi_{i.} = 1$, the optimization of $\pi_{i.}, \mu_i, \sigma_i^2, \nu_i$, and τ_i^2 is a maximum likelihood estimation of the parameters as below,

$$
\hat{\pi'}_{i.} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}{n},
$$

$$
\hat{\mu'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)x_k}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},
$$

$$
\hat{\sigma}_i'^2 = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)(x_k - \mu_i')^2}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},
$$

$$
\hat{\nu'}_i = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)y_k}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)},
$$

$$
\hat{\tau}_i'^2 = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)(y_k - \nu_i')^2}{\sum_{k=1}^{n} \mathrm{E}(z_{ik}|\Theta)}.
$$

Chapter 2

Multilevel Mixture Model Based Methods

Recently, the mixture model approach has been proposed to test the concordance and discordance between two data sets with two-sample groups [32]. Motivated by this study, we expect to generalize the method for multiple data sets with multiple-sample groups. Actually, it is complicate to make this method applicable for multiple-sample groups since it is difficult to identify the concordance or discordance for multiple groups. Due to these multiple data sets, the z-scores obtained from the p-values of the tests will be considered as a multivariate normal vector. We establish a general g-component normal-mixture model for the joint distribution of z-scores. The PCD model can be reduced to the models for evaluating complete concordance (CC) or complete independence (CI). The estimated parameters in the PCD, CC and CI models through EM algorithm are used for calculation of likelihoods of three models. Notice that, when the dimension p of the data sets increases, the number of proportion parameters $\pi_{i_1 i_2 \ldots i_p}$ in PCD model will increase exponentially since $g^p - 1$ different proportion parameters in total need to be estimated. As a result, the estimation for parameters becomes much more challenging. To achieve a feasible method to estimate this model, we may consider a general PCD model as an approximation of combination of CC model and CI model. Based on this two-level mixture model, the original PCD model distribution density can be approximated. Also the number of parameters under the two-level mixture involves one of proportion parameter of CC model, $g-1$ of proportion parameters in CC model and $p \times (g-1)$ of proportion parameters in CI model, so totally $1+(g-1)+p(g-1) = pg+g-p$ different proportion parameters need to be estimated. It increases linearly as the dimension p of the data sets increases, which is much less than that under the original PCD model. In this Chapter, firstly, we extend the Model (1.8) in Section 1.3.2 to a p-dimensional normal mixture model (called the original PCD model). Then we start the multilevel mixture approach with the case of bivariate normal distribution, and parameters can be estimated by the EM algorithm. To illustrate the proposed method, simulation studies are performed and the results are presented for the evaluation of the method.

For convenience, we use variable x instead of z to denote the z-scores obtained from the p-values of the tests in multiple experiments. In addition, we use variable z to denote the missing values in the augmented data sets.

2.1 Multivariate Normal Mixture Model

Suppose there are n sets of obseravations on p-dimensional normal variables $(\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n})$, then the k^{th} set of observations is $(x_{k1}, x_{k2}, \ldots, x_{kp})$, for $k = 1, 2, \ldots, n$. We consider a g-component normal-mixture model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{PCD}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \cdots \sum_{i_p=1}^{g} \pi_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj}), \tag{2.1}$$

where $\pi_{i_1 i_2 \dots i_p}$ is the proportion of that x_{kj} belongs to the i_j-th component for all $j = 1, 2, \dots, p$. Here, $x_{k1}, x_{k2}, \dots, x_{kp}$ do not necessarily belong to the same component. We refer to this model as partial concordance/discordance (PCD) model.

For PCD model, we define z_{ji_jk} as indicators as below,

$$z_{ji_jk} = \begin{cases} 1 & \text{if } x_{kj} \text{ is sampled from the } i_j^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

So, $\pi_{i_1 i_2 \dots i_p} = \Pr(z_{1i_1k} = 1, z_{2i_2k} = 1, \dots, z_{pi_pk} = 1)$. We define $\pi_{\dots i_j \dots} = \Pr(z_{ji_jk} = 1) = \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \dots \sum_{i_{j-1}=1}^{g} \sum_{i_{j+1}=1}^{g} \dots \sum_{i_p=1}^{g} \pi_{i_1 i_2 \dots i_p}$.

Then we derive the complete-data likelihood and log likelihood of PCD model:

$$L_{PCD}(\mathbf{x}, \mathbf{z}|\Theta) = \prod_{k=1}^{n} \prod_{i_1=1}^{g} \prod_{i_2=1}^{g} \dots \prod_{i_p=1}^{g} [\pi_{i_1 i_2 \dots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj})]^{\prod_{l=1}^{p} z_{li_lk}}, \tag{2.2}$$

$$l_{PCD} = \log L_{PCD} = \sum_{k=1}^{n} \sum_{i_1=1}^{g} \dots \sum_{i_p=1}^{g} (\prod_{l=1}^{p} z_{li_lk}) \log[\pi_{i_1 \dots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj})]. \tag{2.3}$$

We need to calculate the following expected values for the E-step:

$$\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta) = \frac{\pi_{i_1 i_2 \dots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{k_j})}{\sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \dots \sum_{i_p=1}^{g} \pi_{i_1 i_2 \dots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{k_j})}, \tag{2.4}$$

where $\Theta = \{\pi_{i_1 i_2 \dots i_p}, \mu_{1i_1}, \mu_{2i_2}, \dots, \mu_{pi_p}, \sigma^2_{1i_1}, \sigma^2_{2i_2}, \dots, \sigma^2_{pi_p}$, for $i_j = 1, 2, \dots, g$ and $j = 1, 2, \dots, p\}$ are the current estimates of parameters which we use to evaluate the expectation.

To maximize the log-likelihood with respect to $\pi_{i_1 i_2 \dots i_p}$, μ_{ji_j}, $\sigma^2_{ji_j}$, we take the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z}}[\log L_{PCD}(\mathbf{x}, \mathbf{z}|\Theta')|\Theta]$. Let $\hat{u}_{i_1 i_2 \dots i_p k} = \mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)$, we can obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi'}_{i_1 i_2 \dots i_p} = \frac{\sum_{k=1}^{n} \hat{u}_{i_1 i_2 \dots i_p k}}{n},$$

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}},$$

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}(x_{kj}-\mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}}. \tag{2.5}$$

If different datasets are completely concordant, then for any set of observations, x_{k1}, $x_{k2}, \ldots,$ and x_{kp} will belong to the same component. Therefore, we expect $\pi_{i_1i_2\ldots i_p} = 0$ unless $i_1 = i_2 = \ldots = i_p$, let this number be i. We define $\pi_i = \pi_{i_1i_2\ldots i_p}|_{i_1=i_2=\ldots=i_p}$. The PCD model is reduced to the complete concordance (CC) model.

For CC model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{CC}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i=1}^{g}\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj}). \tag{2.6}$$

We define the indicator variables z_{ik} as below,

$$z_{ik} = \begin{cases} 1 & \text{if all } x_{kj}\text{'s, } j = 1, 2, \ldots, p, \text{ are sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

Let $\pi_i = P(z_{ik} = 1)$, and $\Theta = \{\pi_i, \mu_{1i}, \ldots, \mu_{pi}, \sigma_{1i}^2, \ldots, \sigma_{pi}^2, \text{ for } i = 1, 2, \ldots, g\}$.

Then the complete-data likelihood and log likelihood of CC model are:

$$L_{CC}(\mathbf{x}, \mathbf{z}|\Theta) = \prod_{k=1}^{n}\prod_{i=1}^{g}[\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})]^{z_{ik}}, \tag{2.7}$$

$$l_{CC} = \log L_{CC} = \sum_{k=1}^{n}\sum_{i=1}^{g} z_{ik}\log[\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})]. \tag{2.8}$$

We need to calculate the following expected values for the E-step:

$$\hat{z}_{ik} = \frac{\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}{\sum_{i=1}^{g}\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}. \tag{2.9}$$

To maximize the log-likelihood with respect to π_i, μ_{ji}, σ_{ji}^2, we take the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z}}[\log L_{CC}(\mathbf{x}, \mathbf{z}|\Theta')|\Theta]$ and obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi}_i = \frac{\sum_{k=1}^n \hat{z}_{ik}}{n},$$

$$\hat{\mu}_{ji} = \frac{\sum_{k=1}^n \hat{z}_{ik} x_{kj}}{\sum_{k=1}^n \hat{z}_{ik}},$$

$$\hat{\sigma}^2_{ji} = \frac{\sum_{k=1}^n \hat{z}_{ik}(x_k - \mu_{ji})^2}{\sum_{k=1}^n \hat{z}_{ik}}. \tag{2.10}$$

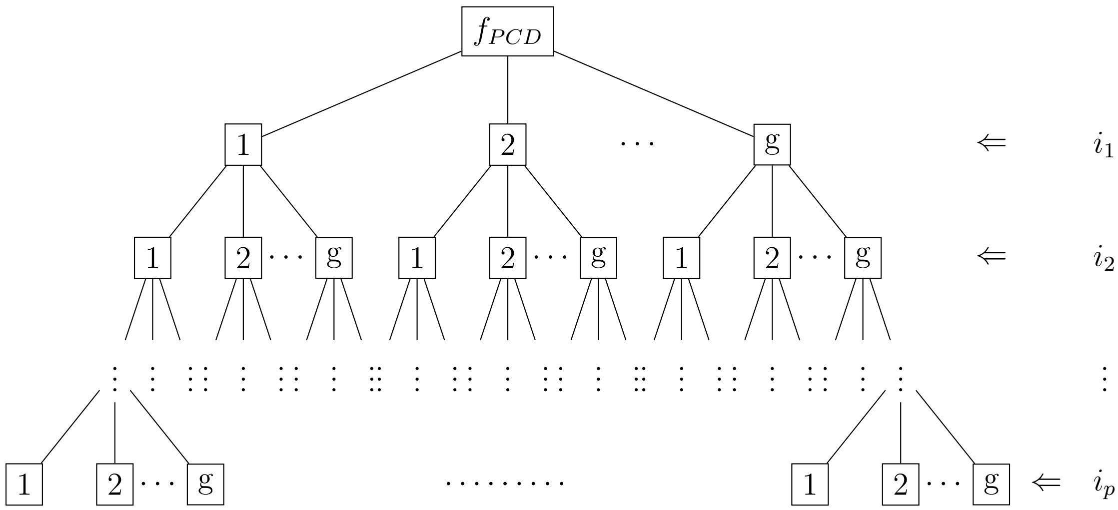
The mathematical derivations can be found in Section 2.4.

Implementation of computing programs

R is usually convenient for many aspects of statistical analysis but it is often criticized for running iterative programs slowly. To get the speed advantage of C programs, we implement the E-M algorithm in C language and then call the C functions in R by loading the .dll file compiled.

When writing the C code of E-M algorithm, we consider the dimension p as a general number. In Formula (2.1), we have p normal densities to multiply and p layers of summation. Since p is an unknown number, we can not simply do the iterations as usual. We adopt a recursive method to deal with this problem. Recursion is the process of repeating items in a self-similar way. In C, this takes the form of a function that calls itself [36]. For example, in Formula (2.1), the production $\prod_{j=1}^p \phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj}) = \phi_{\mu_{pi_p},\sigma^2_{pi_p}}(x_{kp}) \prod_{j=1}^{p-1} \phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj})$. Given information about μ, σ and $\mathbf{x}$, we define the function $Prod(p) = \prod_{j=1}^p \phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj})$, $t_p = \phi_{\mu_{pi_p},\sigma^2_{pi_p}}(x_{kp})$, then $Prod(p) = t_p Prod(p-1)$. So the recursion function can be written based on the recursive formula of $Prod(p)$.

It is much more complicated to calculate the p nested iterations of summations by recursive technique. The following tree gragh of data structure can help us to understand the recursion for these summations "$\sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_p=1}^{g}$".



There are p levels of nodes besides f_{PCD} in this tree, and each Level j can represent the sum process with respect to i_j. For each node of Level j, $j = 1, 2, \ldots, p-1$, there are g child nodes since the range of i_j is 1 to g. We start the calculation with the nodes in 1^{th} level. For Node i in this level, we add up all the g summations in Child Node 1, Child Node 2, ..., Child Node g respectively in Level 2 under Parent Node i in Level 1. The summation in Child Node i in Level 2 can be calculated by adding up the summations in Child Node i's g child nodes in Level 3. And so on, Child Node i in Level $p-1$ saves the sum of all values in its g Child Nodes in Level p. These nested summations can be converted into a recursive function, and the stop condition of loops is when j goes up to the last level p. There are totally g^p child nodes in Level p, and each end child node has a corresponding path to the upper levels, finally to the Level 1. Since we can record all these g^p paths from i_1, i_2 to i_p, then the corresponding $\pi_{i_1 i_2 \ldots i_p}$ can be found. By plugging-in all known parameters, the value of each endpoint node in Level p can be calculated from expression $\pi_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma_{ji_j}^2}(x_{kj})$. Then, we can obtain f_{PCD} based on Formula (2.1) by the recursive function. We have also applied this technique to the E-step and M-step, which has solved the difficulty of nested loops with unknown number of iterations.

2.2 Multilevel Mixture Model

2.2.1 Bivariate Normal Mixture Model

We have reviewed the g-component bivariate normal mixture model in Chapter 1, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{PCD}(x_{k1}, x_{k2}) = \sum_{i=1}^{g} \sum_{j=1}^{g} \pi_{ij} \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2}), \tag{2.11}$$

where π_{ij} is the proportion of that x_{k1} and x_{k2} belong to the i^{th} and j^{th} component, respectively. Here, x_{k1} and x_{k2} do not necessarily belong to the same component. This model is referred as partial concordance/discordance (PCD) model.

We have defined $z_{ijk}^{(PCD)}$ as an indicator as below,

$$z_{ijk}^{(PCD)} = \begin{cases} 1 & \text{if } x_{k1} \text{ is from the } i^{th} \text{ component and } x_{k2} \text{ is from the } j^{th}; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\Theta^{(f_{PCD})} = \{\pi_{ij}, \mu_{1i}, \mu_{2i}, \sigma_{1i}^2, \sigma_{2i}^2, \text{ for } i, j = 1, 2, \ldots, g\}$. Then the density function and complete-data likelihood of PCD model are:

$$f_{PCD}^{*}(x_{k1}, x_{k2}) = \prod_{i=1}^{g} \prod_{j=1}^{g} [\pi_{ij} \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{z_{ijk}^{(PCD)}}, \tag{2.12}$$

$$L_{f_{PCD}}^{*}(\mathbf{x}, \mathbf{z}|\Theta^{(f_{PCD})}) = \prod_{k=1}^{n} \prod_{i=1}^{g} \prod_{j=1}^{g} [\pi_{ij} \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{z_{ijk}^{(PCD)}}. \tag{2.13}$$

If different datasets are completely concordant, then each set of observations (x_{k1}, x_{k2}) will belong to the same component. Therefore, we expect $\pi_{ij} = 0$ unless $i = j$, let this number be i. Here, define $\pi_i = \pi_{ij}|_{i=j}$. The PCD model is reduced to the complete concordance (CC) model.

For the CC model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{CC}(x_{k1}, x_{k2}) = \sum_{i=1}^{g} \pi_i \phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1}) \phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2}). \tag{2.14}$$

Let $z_{ik}^{(CC)}$ be an indicator as below,

$$z_{ik}^{(CC)} = \begin{cases} 1 & \text{if } x_{k1} \text{ and } x_{k2} \text{ are sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

So we have $\pi_i = P(z_{ik}^{(CC)} = 1)$ and let $\Theta^{(CC)} = \{\pi_i, \mu_{1i}, \mu_{2i}, \sigma^2_{1i}, \sigma^2_{2i}, \text{ for } i = 1, 2, \ldots, g\}$. And the density function and complete-data likelihood of CC model are:

$$f^*_{CC}(x_{k1}, x_{k2}) = \prod_{i=1}^{g} [\pi_i \phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1}) \phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2})]^{z_{ik}^{(CC)}}, \tag{2.15}$$

$$L^*_{CC}(\mathbf{x}, \mathbf{z}^{(CC)} | \Theta^{(CC)}) = \prod_{k=1}^{n} \prod_{i=1}^{g} [\pi_i \phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1}) \phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2})]^{z_{ik}^{(CC)}}. \tag{2.16}$$

If different datasets are completely discordant, then each set of observations (x_{k1}, x_{k2}) can be randomly assigned to any component in both data sets. Therefore, any pair of (x_{k1}, x_{k2}) is independent. We expect $\pi_{ij} = \rho_{1i} \times \rho_{2j}$, where ρ_{1i} is the probability of the i^{th} component of x_{k1} and ρ_{2j} is the probability of the j^{th} component of x_{k2}. The PCD model is reduced to complete independence (CI) model.

For the CI model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{CI}(x_{k1}, x_{k2}) = \sum_{i=1}^{g} \sum_{j=1}^{g} \rho_{1i} \rho_{2j} \phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1}) \phi_{\mu_{2j},\sigma^2_{2j}}(x_{k2}). \tag{2.17}$$

We define $z_{i*k}^{(CI)}$ and $z_{*jk}^{(CI)}$ as two indicators as below,

$$z_{i*k}^{(CI)} = \begin{cases} 1 & \text{if } x_{k1} \text{ is sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{*jk}^{(CI)} = \begin{cases} 1 & \text{if } x_{k2} \text{ is sampled from the } j^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

So we have $\rho_{1i} = P(z_{i*k}^{(CI)} = 1)$, and $\rho_{2j} = P(z_{*jk}^{(CI)} = 1)$. Moreover, since $\mathbf{x_1}$ and $\mathbf{x_2}$ are independent, $\mathbf{z_1}^{(CI)}$ and $\mathbf{z_2}^{(CI)}$ are also independent. Let $\Theta^{(CI)} = \{\rho_{1i}, \rho_{2i}, \mu_{1i}, \mu_{2i}, \sigma_{1i}^2, \sigma_{2i}^2$, for $i = 1, 2, \ldots, g\}$, then we can derive the density function and complete-data likelihood of CI model:

$$f_{CI}^*(x_{k1}, x_{k2}) = \{\prod_{i=1}^{g}[\rho_{1i}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{i*k}^{(CI)}}\}\{\prod_{j=1}^{g}[\rho_{2j}\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{z_{*jk}^{(CI)}}\}, \tag{2.18}$$

$$L_{CI}^*(\mathbf{x}, \mathbf{z}^{(CI)}|\Theta^{(CI)}) = \prod_{k=1}^{n}\{\prod_{i=1}^{g}[\rho_{1i}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{i*k}^{(CI)}}\}\{\prod_{j=1}^{g}[\rho_{2j}\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{z_{*jk}^{(CI)}}\}. \tag{2.19}$$

We may consider the PCD model as a combination of CC model and CI model. Let the weight of CC model be λ. The mixture density is

$$g_{PCD}(x_{k1}, x_{k2}) = \lambda f_{CC}(x_{k1}, x_{k2}) + (1-\lambda) f_{CI}(x_{k1}, x_{k2}). \tag{2.20}$$

Let ω_k be an indicator as below,

$$\omega_k = \begin{cases} 1 & \text{if oberservation } \mathbf{x_k} \text{ is from CC model;} \\ 0 & \text{if oberservation } \mathbf{x_k} \text{ is from CI model.} \end{cases}$$

Given ω_k,

$$g_{PCD}^*(x_{k1}, x_{k2}) = [\lambda f_{CC}(x_{k1}, x_{k2})]^{\omega_k}[(1-\lambda) f_{CI}(x_{k1}, x_{k2})]^{1-\omega_k}. \tag{2.21}$$

Let $\Theta = \{\lambda, \pi_i, \rho_{1i}, \rho_{2i}, \mu_{1i}, \mu_{2i}, \sigma_{1i}^2, \sigma_{2i}^2$, for $i = 1, 2, \ldots, g\}$, also use $L_{g_{PCD}}^*$ to denote $L_{g_{PCD}}^*(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}^{(CC)}, \mathbf{z_1}^{(CI)}, \mathbf{z_2}^{(CI)}|\Theta)$. Then we derive the complete-data likelihood of PCD model:

$$\begin{aligned} L_{g_{PCD}}^* \quad = \quad & \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{\prod_{i=1}^{g}\pi_i^{z_{ik}^{(CC)}\omega_k}\rho_{1i}^{z_{i*k}^{(CI)}(1-\omega_k)}\rho_{2i}^{z_{*ik}^{(CI)}(1-\omega_k)} \\ & [\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{ik}^{(CC)}\omega_k + z_{i*k}^{(CI)}(1-\omega_k)}[\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})]^{z_{ik}^{(CC)}\omega_k + z_{*ik}^{(CI)}(1-\omega_k)}\}. \end{aligned}$$

We rewrite the $L^*_{f_{PCD}}$ as following,

$$L^*_{f_{PCD}} = \prod_{k=1}^{n}\{\prod_{i=1}^{g}\prod_{j=1}^{g}\pi_{ij}^{z_{ijk}^{(PCD)}}\}\{\prod_{i=1}^{g}[\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{i.k}^{(PCD)}}[\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})]^{z_{.ik}^{(PCD)}}\}$$

where $z_{i.k}^{(PCD)} = \sum_{j=1}^{g} z_{ijk}^{(PCD)}$, $z_{.ik}^{(PCD)} = \sum_{h=1}^{g} z_{hik}^{(PCD)}$.

Since we use g_{PCD} to approximate f_{PCD}, then each corresponding term in $L^*_{f_{PCD}}$ should approximate to that in $L^*_{g_{PCD}}$. Then we have the system of approximate equations as following,

$$z_{i.k}^{(PCD)} \approx \omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{i*k}^{(CI)}, \tag{2.22}$$

$$z_{.ik}^{(PCD)} \approx \omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{*ik}^{(CI)}, \tag{2.23}$$

where $i = 1, 2, \ldots, g$, for each $k = 1, 2, \ldots, n$.

Note that, for each k, the number of independent observations (z_{ijk}) is $g^2 - 1$ on the left side. For the right side, there is one ω_k, there are $g-1$ independent $z_{ik}^{(CC)}$'s, $(g-1)$ independent $z_{i*k}^{(CI)}$'s and $(g-1)$ independent $z_{*jk}^{(CI)}$'s, then we have $1+(g-1)+2(g-1) = 3g-2$ different parameters. When $g > 2$, $g^2 - 1$ is always greater than $3g - 2$.

We need to calculate the following expected values for the E-step:

$$\begin{aligned} E(\omega_k|\Theta) &= \lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\ &\quad \times[\lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\ &\quad +(1-\lambda)\sum_{i=1}^{g}\sum_{j=1}^{g}\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1} \end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) &= \lambda\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad\times[\lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad+(1-\lambda)\sum_{i=1}^{g}\sum_{j=1}^{g}\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1},
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big) &= (1-\lambda)\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2}) \\
&\quad\times[\lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad+(1-\lambda)\sum_{i,j=1}^{g}\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1},
\end{aligned}$$

$$\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}|\Theta\big) = \sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big),$$

$$\mathrm{E}\big((1-\omega_k)z_{*jk}^{(CI)}|\Theta\big) = \sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big). \tag{2.24}$$

To maximize the log-likelihood with respect to λ, π_i, ρ_{1h}, ρ_{2h}, μ_{1h}, μ_{2h}, σ_{1h}^2 and σ_{2h}^2, we take the partial derivatives of the following Q function,

$$Q(\Theta,\Theta') = \mathrm{E}_{\boldsymbol{\omega},\mathbf{z}^{(\mathrm{CC})},\mathbf{z_1}^{(\mathrm{CI})},\mathbf{z_2}^{(\mathrm{CI})}}[\log L(\mathbf{x},\boldsymbol{\omega},\mathbf{z}^{(CC)},\mathbf{z_1}^{(CI)},\mathbf{z_2}^{(CI)}|\Theta')|\Theta]$$

and obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\lambda}' = \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta),$$

$$\hat{\pi}_i' = \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)},$$

$$\hat{\rho'_{1h}} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)},$$

$$\hat{\rho'_{2h}} = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)},$$

$$\hat{\mu'_{1h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)]x_{k1}}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)},$$

$$\hat{\mu'_{2h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)]x_{k2}}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)},$$

$$\hat{\sigma'^2_{1h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)](x_{k1}-\mu'_{1h})^2}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)},$$

$$\hat{\sigma'^2_{2h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)](x_{k2}-\mu'_{2h})^2}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)}. \quad (2.25)$$

2.2.2 Multivariate Normal Mixture Model

We recall the g-component normal-mixture model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{PCD}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_p=1}^{g} \pi_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj}), \quad (2.26)$$

where $\pi_{i_1 i_2 \ldots i_p}$ is the proportion of that x_{kj} belongs to the i_j-th component for all $j = 1, 2, \ldots, p$. Here, $x_{k1}, x_{k2}, \ldots, x_{kp}$ do not necessarily belong to the same component, which means the component indices are not necessarily the same. We have defined $z_{ijk}^{(PCD)}$ in Section 3.1, similarly we also define $z_{i_1 i_2 \ldots i_p k}^{(PCD)}$ as an indicator as below,

$$z_{i_1 i_2 \ldots i_p k}^{(PCD)} = \begin{cases} 1 & \text{if } x_{kj} \text{ is from the } i_j^{th} \text{ component, for all } j = 1, 2, \ldots, p; \\ 0 & \text{otherwise.} \end{cases}$$

Let $\Theta = \{\pi_{i_1 i_2 \ldots i_p}, \mu_{1i_1}, \mu_{2i_2}, \ldots, \mu_{pi_p}, \sigma^2_{1i_1}, \sigma^2_{2i_2}, \ldots, \sigma^2_{pi_p},$ for $i_j = 1, 2, \ldots, g$ and $j = 1, 2, \ldots, p\}$. Then we can derive the density function and complete-data likelihood of PCD model:

$$f^*_{PCD}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \prod_{i_1=1}^{g} \prod_{i_2=1}^{g} \cdots \prod_{i_p=1}^{g} [\pi_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj})]^{z^{(PCD)}_{i_1 i_2 \ldots i_p k}}, \quad (2.27)$$

$$L^*_{f_{PCD}}(\mathbf{x}, \mathbf{z}|\Theta^{(f_{PCD})}) = \prod_{k=1}^{n} \prod_{i_1=1}^{g} \prod_{i_2=1}^{g} \cdots \prod_{i_p=1}^{g} [\pi_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj})]^{z^{(PCD)}_{i_1 i_2 \ldots i_p k}}. \quad (2.28)$$

For the CC model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{CC}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i=1}^{g} \pi_i \prod_{j=1}^{p} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj}). \quad (2.29)$$

We have defined $z^{(CC)}_{ik}$ as an indicator as below,

$$z^{(CC)}_{ik} = \begin{cases} 1 & \text{if all } x_{kj}\text{'s, } j = 1, 2, \ldots, p, \text{ are sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

So we have $\pi_i = P(z^{(CC)}_{ik} = 1)$ and let $\Theta = \{\pi_i, \mu_{1i}, \ldots, \mu_{pi}, \sigma^2_{1i}, \ldots, \sigma^2_{pi},$ for $i = 1, 2, \ldots, g\}$. Then we can derive the density function and complete-data likelihood of CC model:

$$f^*_{CC}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \prod_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]^{z^{(CC)}_{ik}}, \quad (2.30)$$

$$L^*_{CC}(\mathbf{x}, \mathbf{z}^{(CC)}|\Theta^{(CC)}) = \prod_{k=1}^{n} \prod_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]^{z^{(CC)}_{ik}}. \quad (2.31)$$

If different datasets are completely discordant, then for each set of observations, x_{k1}, x_{k2}, …, x_{kp} can be randomly assigned to any component in both data sets. Therefore, $x_{k1}, x_{k2}, \ldots, x_{kp}$ are independent. We expect $\pi_{i_1 i_2 \ldots i_p} = \prod_{j=1}^{p} \rho_{ji_j}$, where ρ_{ji_j} is the probability of the i_j^{th} component of x_{kj}. The PCD model is reduced to complete independence (CI) model.

For the CI model, the joint normal probability distribution function for the k^{th} set of observations is

$$f_{CI}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \cdots \sum_{i_p=1}^{g} [\prod_{j=1}^{p} \rho_{ji_j} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj})]. \qquad (2.32)$$

We can also consider this joint probability distribution function in another way. Since $x_{k1}, x_{k2}, \ldots, x_{kp}$ are independent and for each x_{kj}, it has a mixture density $\sum_{i=1}^{g} \rho_{ji} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})$. Then the joint probability distribution function for CI model can be also written as

$$f_{CI}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \prod_{j=1}^{p} [\sum_{i=1}^{g} \rho_{ji} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]. \qquad (2.33)$$

Similarly as $z^{(CI)}_{i*k}$ in Section 2.2.1, now we define indicators $z^{(CI)}_{i**\ldots**k}$, $z^{(CI)}_{*i**\ldots**k}$, $z^{(CI)}_{**i**\ldots**k}$, $\ldots$, $z^{(CI)}_{**\ldots**ik}$, then simplify $z^{(CI)}_{**\ldots**i**\ldots**k}$ as $z^{(CI)}_{ijk}$ as below,

$$z^{(CI)}_{ijk} = \begin{cases} 1 & \text{if } x_{kj} \text{ is sampled from the } i^{th} \text{ component;} \\ 0 & \text{otherwise.} \end{cases}$$

Here, $\rho_{ji} = P(z^{(CI)}_{ijk} = 1)$. Moreover, since $\mathbf{x_1}$, $\mathbf{x_2}$, $\ldots$, $\mathbf{x_p}$ are independent, $\mathbf{z_1}^{(CI)}$, $\mathbf{z_2}^{(CI)}, \ldots, \mathbf{z_p}^{(CI)}$ are also independent. Let $\Theta^{(CI)} = \{\rho_{ji}, \mu_{ji}, \sigma^2_{ji}$, for $i = 1, 2, \ldots, g, j = 1, 2, \ldots, p\}$, then we can derive the density function and complete-data likelihood of CI model:

$$f^*_{CI}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \prod_{j=1}^{p} \{\prod_{i=1}^{g} [\rho_{ji} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]^{z^{(CI)}_{ijk}}\}, \qquad (2.34)$$

$$L^*_{CI}(\mathbf{x}, \mathbf{z}^{(CI)} | \Theta^{(CI)}) = \prod_{k=1}^{n} \{\prod_{j=1}^{p} [\prod_{i=1}^{g} (\rho_{ji} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj}))^{z^{(CI)}_{ijk}}]\}. \qquad (2.35)$$

We may consider the PCD model as a combination of CC model and CI model. Let the weight of CC model be λ. The mixture density is

$$g_{PCD}(\mathbf{x_k}) = \lambda f_{CC}(\mathbf{x_k}) + (1 - \lambda) f_{CI}(\mathbf{x_k}). \qquad (2.36)$$

Let ω_k be an indicator as below,

$$\omega_k = \begin{cases} 1 & \text{if oberservation } \mathbf{x_k} \text{ is from CC model;} \\ 0 & \text{if oberservation } \mathbf{x_k} \text{ is from CI model.} \end{cases}$$

Given ω_k,

$$g^*_{PCD}(\mathbf{x_k}) = [\lambda f_{CC}(\mathbf{x_k})]^{\omega_k}[(1-\lambda)f_{CI}(\mathbf{x_k})]^{1-\omega_k}. \tag{2.37}$$

Let $\Theta = \{\lambda, \pi_i, \rho_{ji}, \mu_{ji}, \sigma^2_{ji}$, for $i = 1, 2, \ldots, g, j = 1, 2, \ldots, p\}$, also use $L^*_{g_{PCD}}$ to denote $L^*_{g_{PCD}}(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}^{(CC)}, \mathbf{z}^{(CI)}|\Theta)$. Then we derive the complete-data likelihood of PCD model:

$$\begin{aligned} L^*_{g_{PCD}} &= \lambda^{\sum_{k=1}^n \omega_k}(1-\lambda)^{\sum_{k=1}^n(1-\omega_k)} \prod_{k=1}^n \{\prod_{i=1}^g \pi_i^{z_{ik}^{(CC)}\omega_k}[\prod_{j=1}^p \rho_{ji}^{z_{ijk}^{(CI)}(1-\omega_k)} \\ &\quad (\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}\omega_k + z_{ijk}^{(CI)}(1-\omega_k)}]\}. \end{aligned}$$

We rewrite the $L^*_{f_{PCD}}$ as following,

$$\begin{aligned} L^*_{f_{PCD}} &= \prod_{k=1}^n \{\prod_{i_1=1}^g \prod_{i_2=1}^g \cdots \prod_{i_p=1}^g \pi_{i_1 i_2 \ldots i_p}^{z^{(PCD)}_{i_1 i_2 \ldots i_p k}}\} \\ &\quad \{\prod_{j=1}^p [\prod_{i=1}^g (\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{\sum^g_{i_1,i_2,\ldots,i_{j-1},i_{j+1},\ldots,i_p=1} z^{(PCD)}_{i_1 i_2 \ldots i_{j-1} i i_{j+1} \ldots i_p k}}]\}. \end{aligned}$$

Here, we define $z^{(PCD)}_{..i_{(j)}..k} = \sum^g_{i_1,i_2,\ldots,i_{j-1},i_{j+1},\ldots,i_p=1} z^{(PCD)}_{i_1 i_2 \ldots i_{j-1} i i_{j+1} \ldots i_p k}$. Since we use g_{PCD} to approximate f_{PCD}, then each corresponding term in $L^*_{f_{PCD}}$ should approximate to that in $L^*_{g_{PCD}}$. Then we have the system of approximate equations as following,

$$z^{(PCD)}_{..i_{(j)}..k} \approx \omega_k z^{(CC)}_{ik} + (1-\omega_k) z^{(CI)}_{ijk}, \tag{2.38}$$

where $i = 1, 2, \ldots, g$, for each $k = 1, 2, \ldots, n$.

Generally, for p-dimensional case, the number of independent observations $z_{i_1 i_2 \ldots i_p k}$ is $g^p - 1$ on the left side. For the right side, there is one ω_k, there are $g-1$ independent

$z_{ik}^{(CC)}$'s and $p(g-1)$ independent $z_{i_1 i_2 \ldots i_p k}^{(CI)}$'s, then we have $1+(g-1)+p(g-1) = pg+g-p$ different parameters. Here, we only care ab the cases when $p \geq 2, g \geq 3$. Therefore, $g^p - 1$ is always greater than $pg + g - p$.

We need to calculate the following expected values for the E-step:

$$\mathrm{E}(\omega_k|\Theta) = \frac{\lambda \sum_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})]}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})},$$

$$\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) = \frac{\lambda \pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})},$$

$$\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big) = \frac{(1-\lambda)\rho_{ji}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj}) \prod_{l=1,l\neq j}^{p} \sum_{h=1}^{g} \rho_{lh}\phi_{\mu_{lh},\sigma_{lh}^2}(x_{kl})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})}.$$

To maximize the log-likelihood with respect to λ, π_i, ρ_{ji}, μ_{ji} and σ_{ji}^2, we take the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\boldsymbol{\omega},\mathbf{z}^{(\mathrm{CC})},\mathbf{z}^{(\mathrm{CI})}}[\log L(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}^{(CC)}, \mathbf{z}^{(CI)})|\Theta]$ and obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\lambda'} = \frac{1}{n} \sum_{k=1}^{n} \mathrm{E}(\omega_k|\Theta),$$

$$\hat{\pi_i'} = \frac{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\sum_{k=1}^{n} \sum_{h=1}^{g} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)},$$

$$\hat{\rho_{ji}'} = \frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)}{\sum_{k=1}^{n} \sum_{h=1}^{g} \mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)},$$

$$\hat{\mu_{ji}'} = \frac{\sum_{k=1}^{n} [\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)] x_{kj}}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)},$$

$$\hat{\sigma_{ji}'^2} = \frac{\sum_{k=1}^{n} [\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)](x_{kj} - \mu_{ji}')^2}{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)}. \quad (2.39)$$

2.3 Simulation Studies

In this section, we describe the simulation studies based on both multilevel bivariate normal mixture model and multilevel multivariate normal mixture model. For each model, we consider two types of parameter configurations. One is moderate, which means the true values of parameters are under mild condition, neither too small nor too large for the proportion parameters and a relatively wide difference between means of components. The other is difficult, which means the true values are under extreme condition, relatively small or large for the proportion parameters and a relatively narrow gap between means of components.

For each setting of parameter initialization, we first generate normal data sets. To obtain the reasonable missing information z such as $\mathbf{z}^{(CC)}$ and $\mathbf{z}^{(CI)}$, we categorize them into three groups according to the data. If variable x_k is less than -1, then we suppose that it is sampled from the 1^{st} component; if it is greater than 1, we suppose that it is sampled from the 3^{rd} component; otherwise, we assume that it is sampled from the null Component 2. Then, we can get z values based on this classification. For example, when we have two data sets $\mathbf{x}$ and $\mathbf{y}$, if both of variable x_k and y_k are sampled from the same Component 3, then $z_{3k}^{(CC)} = 1$ and also $z_{1k}^{(CC)} = 0$, $z_{2k}^{(CC)} = 0$; if x_k is sampled from Component 1, and y_k is sampled from Component 2, then $z_{1*k}^{(CI)} = 1$, $z_{*2k}^{(CI)} = 1$ and also $z_{i*k}^{(CI)} = 0$, $z_{*jk}^{(CI)} = 0$, when $i \neq 1$, $j \neq 2$.

After calculation of all z values, we implement the E-M algorithm and save the estimation results. We set the maximum number of iterations as 10000. If the E-M algorithm does not achieve convergence in 10000 iterations, then we consider the iterative procedure does not converge. We present the summary table and box plots for each case to understand the performance of our models.

2.3.1 Bivariate Normal Mixture Model

To illustrate the proposed method based on the multilevel bivariate normal mixture model (2.20) in Section 2.2.1, experiments are performed with simulated datasets as following.

Case 1: we consider a moderate case (see Paragraph 1 in Section 2.3 for details) without repetitions (repetition number B=1) when there is no restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1$. We generate $N = 1000$ bivariate normal data according to our model (2.20). The estimated results are shown in Table 2.1, and we can see that some of them are in an acceptable range but not very close to the initial values.

Table 2.1: Parameters estimation in a moderate two-level model without restriction when $B = 1$, $N = 1000$, $p = 2$

Parameter	True Values	Estimates by EM
λ	0.6	0.5534
$\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$	(0.2, 0.55, 0.25)	(0.278, 0.511, 0.211)
$\boldsymbol{\rho_1} = (\rho_{11}, \rho_{12}, \rho_{13})$	(0.25, 0.5, 0.25)	(0.289, 0.430, 0.280)
$\boldsymbol{\rho_2} = (\rho_{21}, \rho_{22}, \rho_{23})$	(0.2, 0.6, 0.2)	(0.060, 0.799, 0.144)
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 3)	(-1.553, -0.098, 2.058)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-3, 0, 2)	(-2.648, -0.016, 2.167)
$\boldsymbol{\sigma^2_1} = (\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(2, 1, 2)	(3.083, 0.915, 1.796)
$\boldsymbol{\sigma^2_2} = (\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(2, 1, 2)	(1.946, 0.947, 1.527)

For comparison, we also perform experiments based on the original PCD model (2.11) in Section 2.2.1.

Case 2: similarly as case 1, we consider a moderate case without repetitions when there is no restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. We generate $N = 1000$ bivariate normal data according to our model (2.11). The results are obtained in Table 2.2, and similarly as those in Case 1, we can see that some of them are in an acceptable range but not very close to the initial values.

Table 2.2: Parameters estimation in a moderate PCD model without restriction when $B = 1$, $N = 1000$, $p = 2$

Parameter	True Values	Estimates by EM
$\boldsymbol{\pi}[\mathbf{1,}] = (\pi_{11}, \pi_{12}, \pi_{13})$	(0.05, 0.05, 0.15)	(0.049, 0.063, 0.149)
$\boldsymbol{\pi}[\mathbf{2,}] = (\pi_{21}, \pi_{22}, \pi_{23})$	(0.05, 0.40, 0.05)	(0.010, 0.420, 0.022)
$\boldsymbol{\pi}[\mathbf{3,}] = (\pi_{31}, \pi_{32}, \pi_{33})$	(0.15, 0.05, 0.05)	(0.152, 0.101, 0.032)
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.236, 0.059, 2.176)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-2.016, -0.057, 1.786)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.209, 1.119, 1.537)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.383, 0.868, 1.493)

For a comprehensive performance evaluation, we conduct simulation studies with repetitions to understand the estimation performance of the two-level model (2.20). For a reasonable parameter configuration, we repeat the estimation for 1000 times, obtain all these estimation results, and then generate boxplots of estimated parameters.

Case 3: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. In this case, the number of non-convergence (see Paragraph 3 in Section 2.3 for details) is 2 out of 1000. Based on the 1000 repetitions, Table 2.3 presents the summary, and Figure 2.1 and 2.2 display the box plots for each parameter. In Figure 2.1, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis). In Figure 2.2, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.

Comparing with the estimation in Case 1, the results are much more closer to the initial values with relatively small variances.

Table 2.3: Parameters estimation in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
λ	0.7	0.686	0.114
(π_1, π_2, π_3)	(0.25, 0.6, 0.15)	(0.250, 0.593, 0.157)	(0.099, 0.106, 0.080)
$(\rho_{11}, \rho_{12}, \rho_{13})$	(0.3, 0.5, 0.2)	(0.342, 0.418, 0.239)	(0.201, 0.241, 0.183)
$(\rho_{21}, \rho_{22}, \rho_{23})$	(0.2, 0.5, 0.3)	(0.242, 0.429, 0.329)	(0.191, 0.249, 0.199)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-1.999, 0, 1.955)	(0.203, 0, 0.344)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-2.001, 0, 1.989)	(0.245, 0, 0.288)
$(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.501, 1, 1.532)	(0.274, 0, 0.428)
$(\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.492, 1, 1.525)	(0.325, 0, 0.363)

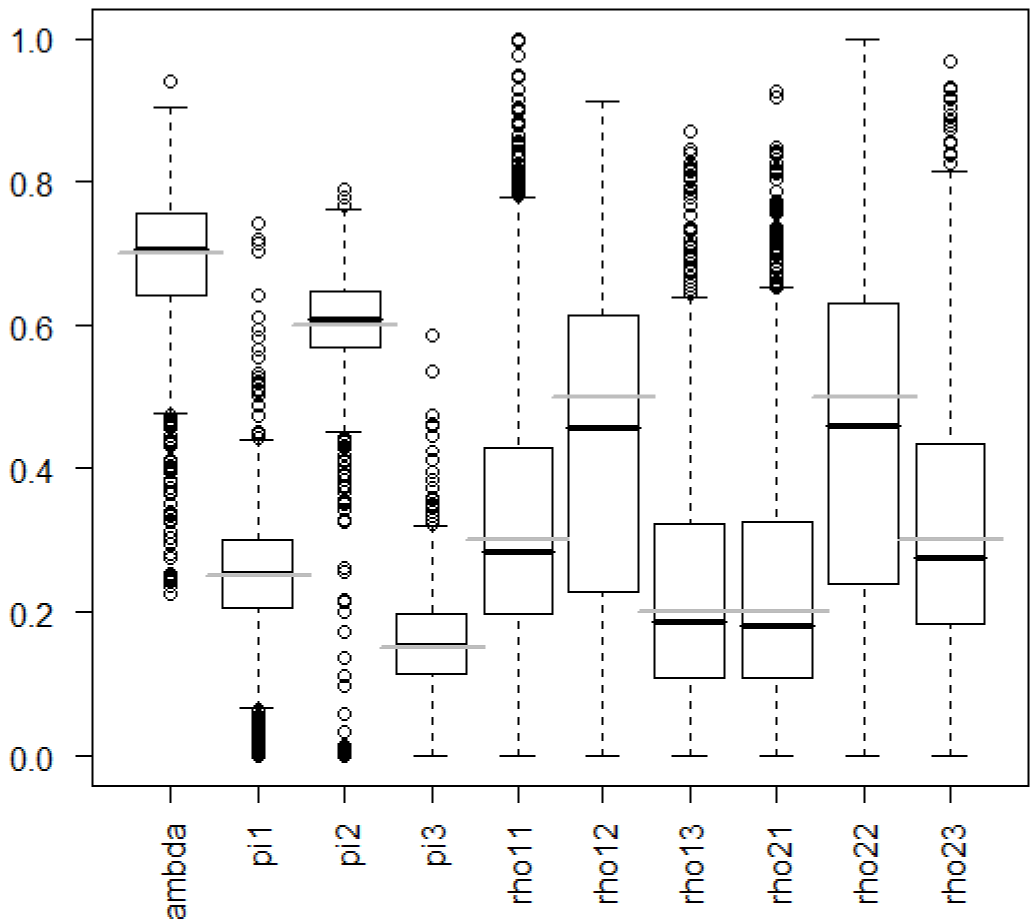


Figure 2.1: Boxplots of estimated proportion parameters in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis).

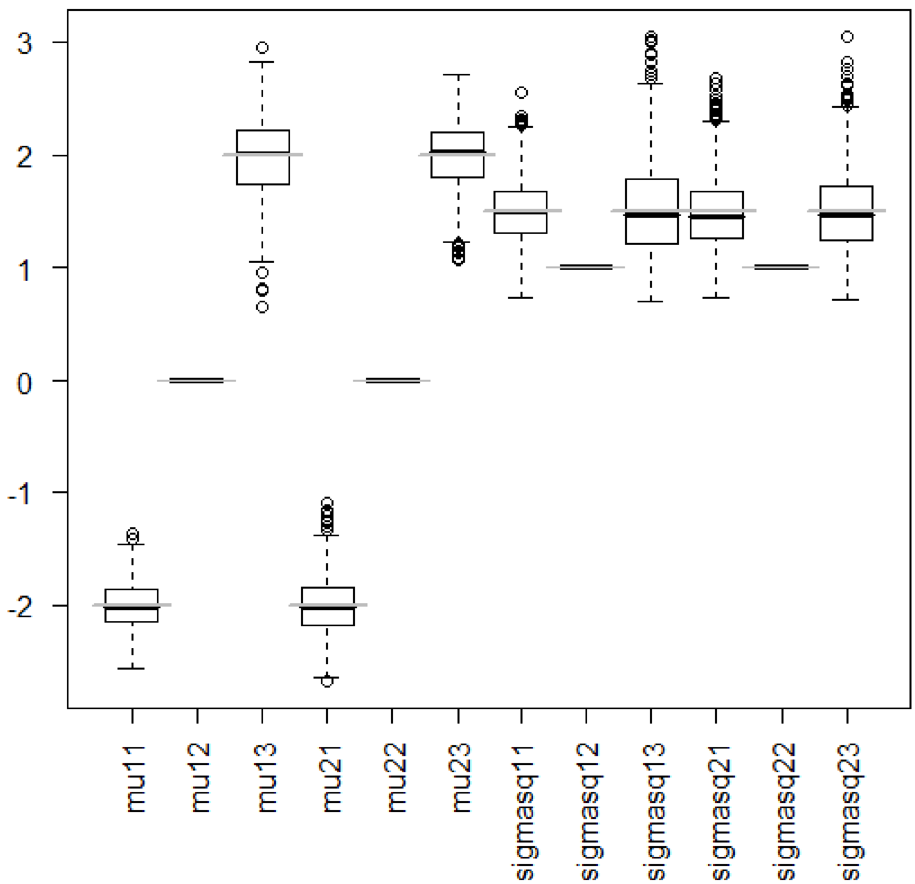


Figure 2.2: Boxplots of estimated means and variances in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.

Case 4: then we consider a difficult case (see Paragraph 1 in Section 2.3 for details) with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1$. In this case, the number of non-convergence is 1 out of 1000. Table 2.4 shows the summary, and Figure 2.3 and 2.4 display the box plots for each parameter. In Figure 2.3, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis). In Figure 2.4, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma^2_1}$, $\boldsymbol{\sigma^2_2}$ in horizontal sequence are with vertical scale from -4 to 6. Comparing with the estimation in Case 3, some of the estimated results are in an acceptable range but not very close to the initial values with a little greater variances than those in moderate case.

Table 2.4: Parameters estimation in a difficult two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
λ	0.15	0.229	0.233
$\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$	(0.1, 0.8, 0.1)	(0.243, 0.530, 0.237)	(0.310, 0.394, 0.314)
$\boldsymbol{\rho_1} = (\rho_{11}, \rho_{12}, \rho_{13})$	(0.1, 0.8, 0.1)	(0.129, 0.742, 0.128)	(0.145, 0.198, 0.143)
$\boldsymbol{\rho_2} = (\rho_{21}, \rho_{22}, \rho_{23})$	(0.1, 0.8, 0.1)	(0.131, 0.739, 0.130)	(0.148, 0.204, 0.144)
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.662, 0, 1.699)	(0.854, 0, 0.874)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.671, 0, 1.676)	(0.865, 0, 0.841)
$\boldsymbol{\sigma^2_1} = (\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.309, 1, 1.299)	(0.712, 0, 0.755)
$\boldsymbol{\sigma^2_2} = (\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.315, 1, 1.361)	(0.724, 0, 0.770)

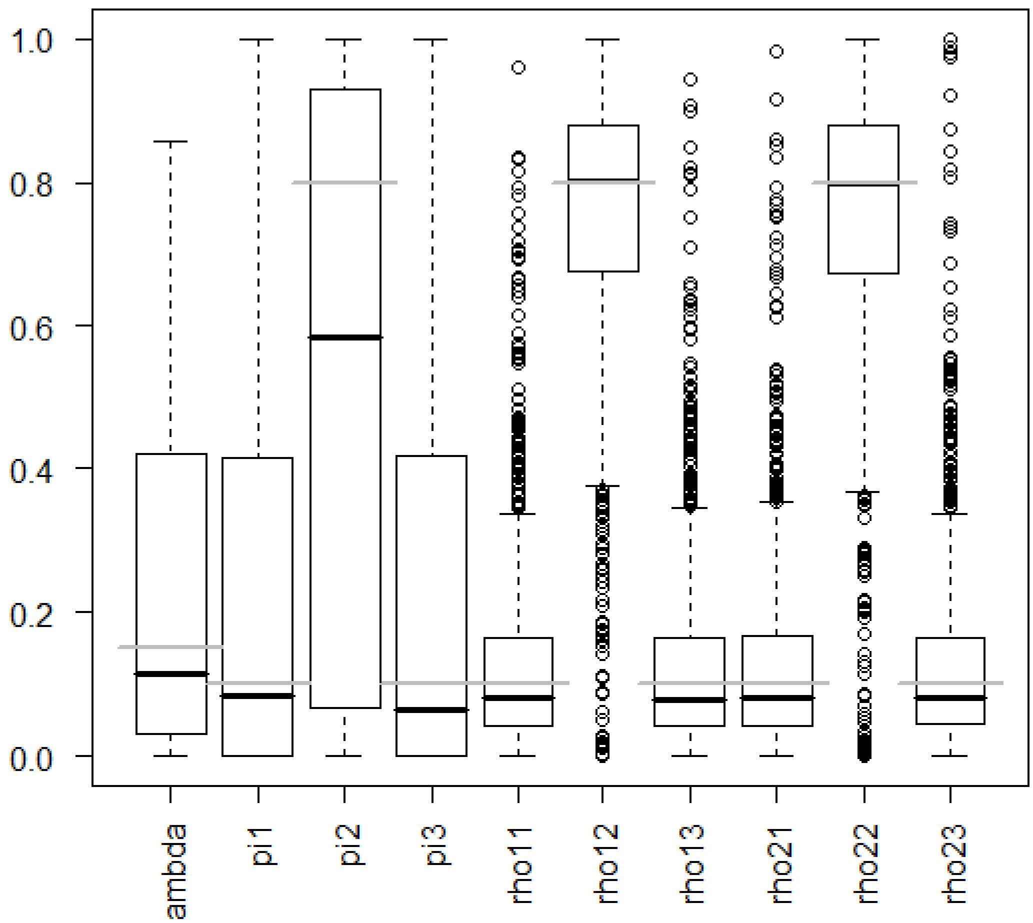


Figure 2.3: Boxplots of estimated proportion parameters in a difficult two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$. On the left, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis).

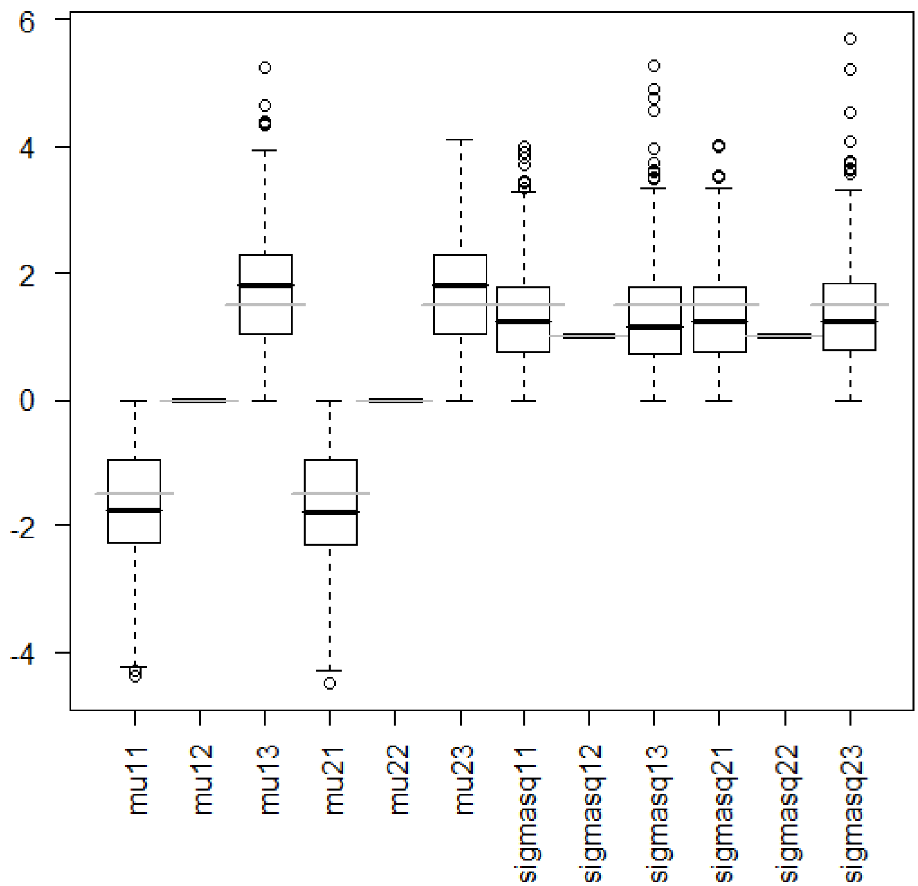


Figure 2.4: Boxplots of estimated means and variances in a difficult two-level model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -4 to 6.

For comparison, we also conduct simulation studies to understand the estimation performance of the original PCD model (2.11). For a reasonable parameter configuration, we repeat the estimation for 1000 times, save all these estimation results, and then generate boxplots of estimated parameters.

Case 5: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. In this case, the number of non-convergence is 0. Table 2.5 shows the summary, and Figure 2.5 and 2.6 display the box plots for each parameter. In Figure 2.5, the proportion parameters $\boldsymbol{\pi[1,]}$, $\boldsymbol{\pi[2,]}$, $\boldsymbol{\pi[3,]}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.5 (y-axis). In Figure 2.6, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3. Comparing with the estimation in Case 2, the results are much more closer to the initial values with relatively small variances.

Table 2.5: Parameters estimation in a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
$(\pi_{11}, \pi_{12}, \pi_{13})$	(0.05, 0.05, 0.05)	(0.051, 0.048, 0.157)	(0.039, 0.034, 0.018)
$(\pi_{21}, \pi_{22}, \pi_{23})$	(0.05, 0.40, 0.05)	(0.048, 0.396, 0.049)	(0.034, 0.040, 0.036)
$(\pi_{31}, \pi_{32}, \pi_{33})$	(0.15, 0.05, 0.05)	(0.155, 0.047, 0.051)	(0.018, 0.032, 0.038)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-1.988, 0, 1.996)	(0.287, 0, 0.280)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.994, 0, 1.979)	(0.282, 0, 0.289)
$(\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.507, 1, 1.500)	(0.372, 0, 0.355)
$(\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.504, 1, 1.518)	(0.343, 0, 0.370)

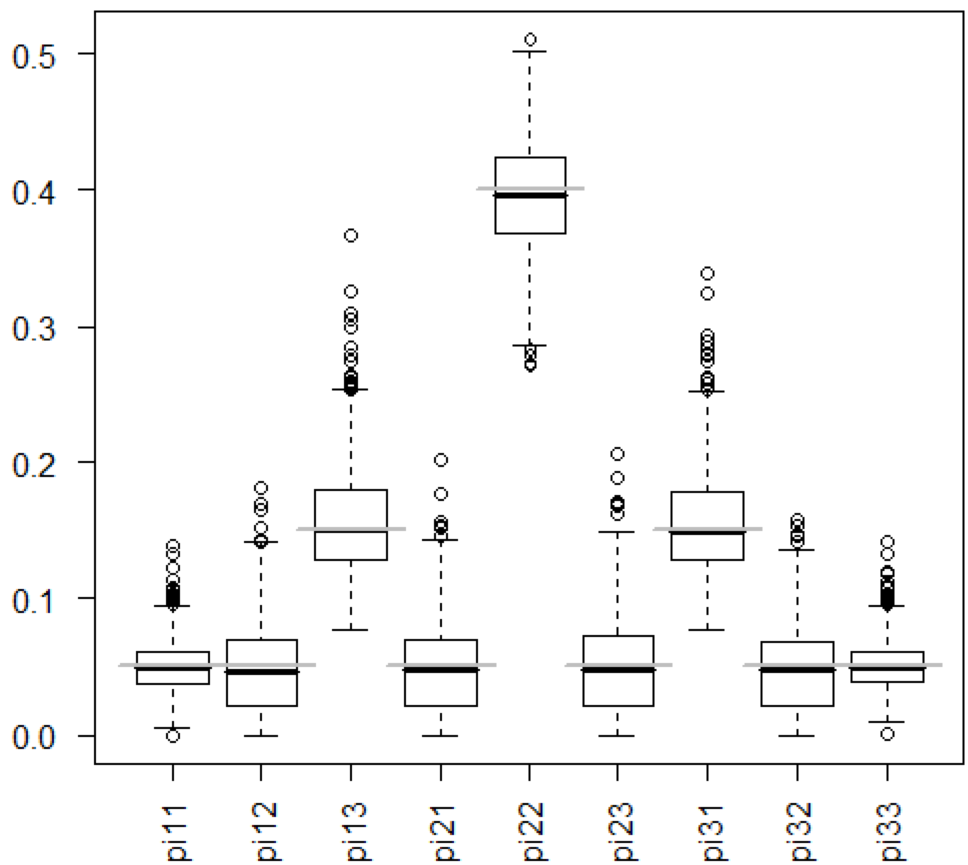


Figure 2.5: Boxplots of estimated proportion parameters in a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. On the left, the proportion parameters $\boldsymbol{\pi}[\mathbf{1},]$, $\boldsymbol{\pi}[\mathbf{2},]$, $\boldsymbol{\pi}[\mathbf{3},]$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.5 (y-axis).

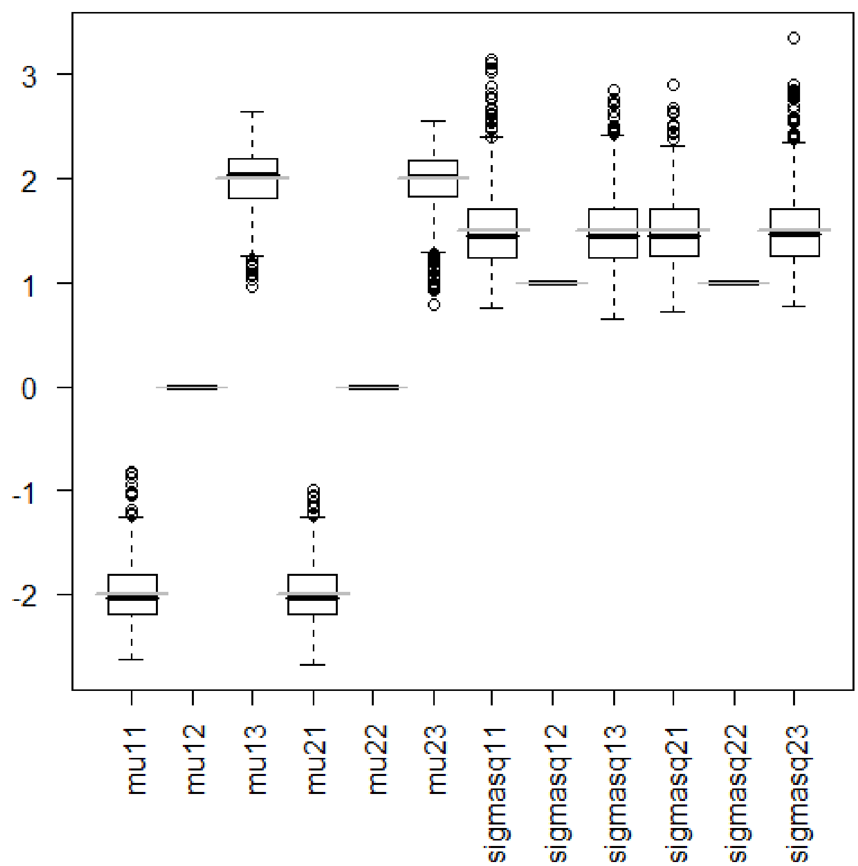


Figure 2.6: Boxplots of estimated means and variances in a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.

Case 6: then we consider a difficult case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. In this case, the number of non-convergence is 0. Table 2.6 shows the summary, and Figure 2.7 and 2.8 display the box plots for each parameter. In Figure 2.7, the proportion parameters $\boldsymbol{\pi}[\mathbf{1,}]$, $\boldsymbol{\pi}[\mathbf{2,}]$, $\boldsymbol{\pi}[\mathbf{3,}]$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.6 (y-axis). In Figure 2.8, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -3 to 3. Comparing with the estimation in Case 5, some of the estimated results are in an acceptable range but not very close to the initial values with slightly greater variances than those in moderate case.

Table 2.6: Parameters estimation in a difficult PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
$(\pi_{11}, \pi_{12}, \pi_{13})$	(0.05, 0.05, 0.15)	(0.045, 0.062, 0.183)	(0.103, 0.067, 0.039)
$(\pi_{21}, \pi_{22}, \pi_{23})$	(0.15, 0.2, 0.15)	(0.128, 0.184, 0.126)	(0.090, 0.092, 0.093)
$(\pi_{31}, \pi_{32}, \pi_{33})$	(0.15, 0.05, 0.05)	(0.174, 0.058, 0.041)	(0.035, 0.061, 0.098)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.382, 0, 1.430)	(0.580, 0, 0.558)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.546, 0, 1.526)	(0.385, 0, 0.376)
$(\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.598, 1, 1.564)	(0.540, 0, 0.525)
$(\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.463, 1, 1.483)	(0.382, 0, 0.382)

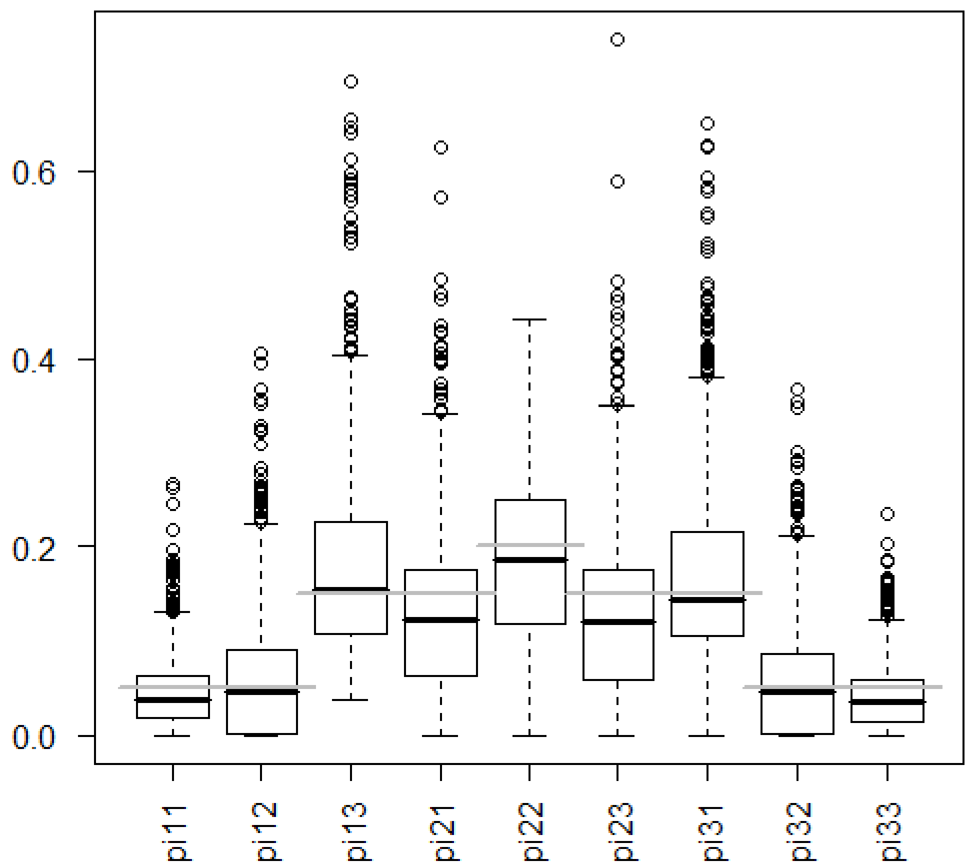


Figure 2.7: Boxplots of estimated proportion parameters in a difficult PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters $\boldsymbol{\pi}[\mathbf{1,}]$, $\boldsymbol{\pi}[\mathbf{2,}]$, $\boldsymbol{\pi}[\mathbf{3,}]$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.6 (y-axis).

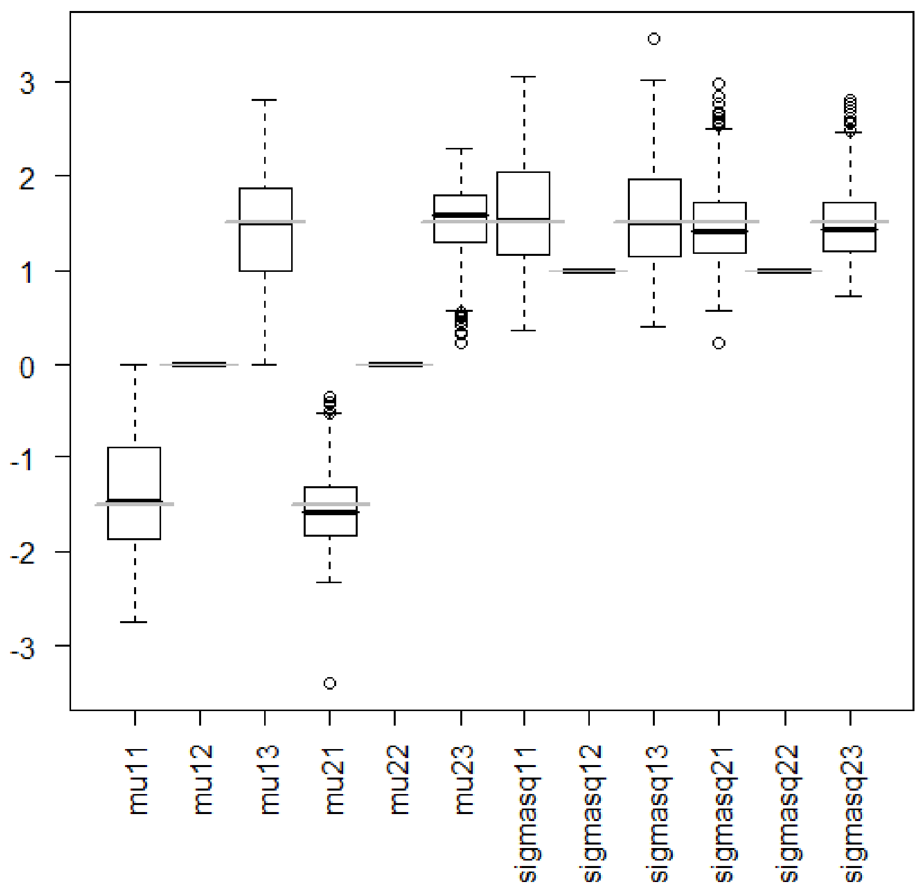


Figure 2.8: Boxplots of estimated means and variances in a difficult PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -3 to 3.

For further comparison, we also conduct simulation studies based on PCD model to understand the estimation performance of the two-level model. For a reasonable parameter configuration, we generate data based on the original PCD model with relatively mild initial values (2.11), but estimate the parameters based on the two-level model (2.20). We repeat the estimation for 1000 times, save all these estimation results, and then generate boxplots of estimated parameters.

Case 7: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. In this case, the number of non-convergence is 2. Table 2.7 shows the summary, and Figure 2.9 and 2.10 display the box plots for each parameter estimated by two-level model. In Figure 2.9, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1.0 (y-axis). In Figure 2.10, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.The estimated results for μ's and σ's are quite close to the true values, while some of those for proportions are not. We also calculate approximated $\boldsymbol{\pi}$ matrix in PCD model by the two level model. The estimated proportions are not far from the true. We present the box plots for converted results to approximate the proportion parameters in the original PCD model in Figure 2.11.

Table 2.7: Parameters estimation by a two-level model based on data from a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
PCD model			
$(\pi_{11}, \pi_{12}, \pi_{13})$	(0.15, 0.05, 0.05)	(0.156, 0.047, 0.051)	(0.036, 0.034, 0.018)
$(\pi_{21}, \pi_{22}, \pi_{23})$	(0.05, 0.40, 0.05)	(0.048, 0.397, 0.047)	(0.033, 0.041, 0.033)
$(\pi_{31}, \pi_{32}, \pi_{33})$	(0.05, 0.05, 0.15)	(0.051, 0.046, 0.157)	(0.019, 0.034, 0.039)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-1.995, 0, 1.984)	(0.267, 0, 0.289)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.992, 0, 1.982)	(0.275, 0, 0.275)
$(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.500, 1, 1.512)	(0.348, 0, 0.364)
$(\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.506, 1, 1.510)	(0.341, 0, 0.353)
Two-level model			
λ		0.550	0.065
(π_1, π_2, π_3)		(0.176, 0.642, 0.182)	(0.095, 0.062, 0.095)
$(\rho_{11}, \rho_{12}, \rho_{13})$		(0.356, 0.295, 0.349)	(0.141, 0.177, 0.140)
$(\rho_{21}, \rho_{22}, \rho_{23})$		(0.359, 0.289, 0.352)	(0.144, 0.182, 0.139)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-1.995, 0, 1.984)	(0.267, 0, 0.289)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.992, 0, 1.982)	(0.275, 0, 0.275)
$(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.500, 1, 1.512)	(0.348, 0, 0.364)
$(\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.506, 1, 1.510)	(0.342, 0, 0.353)

We use results of λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ estimated by the two-level model to approximate π_{ij} in orginal PCD model. When $i = j$, the $\pi_{ii} \approx \lambda\pi_i + (1-\lambda)\rho_{1i}\rho_{2i}$; when $i \neq j$, $\pi_{ij} \approx (1-\lambda)\rho_{1i}\rho_{2j}$. Each π_{ij} can be approximated for PCD model, and then the mean and standard deviation of π_{ij} can be calculated.

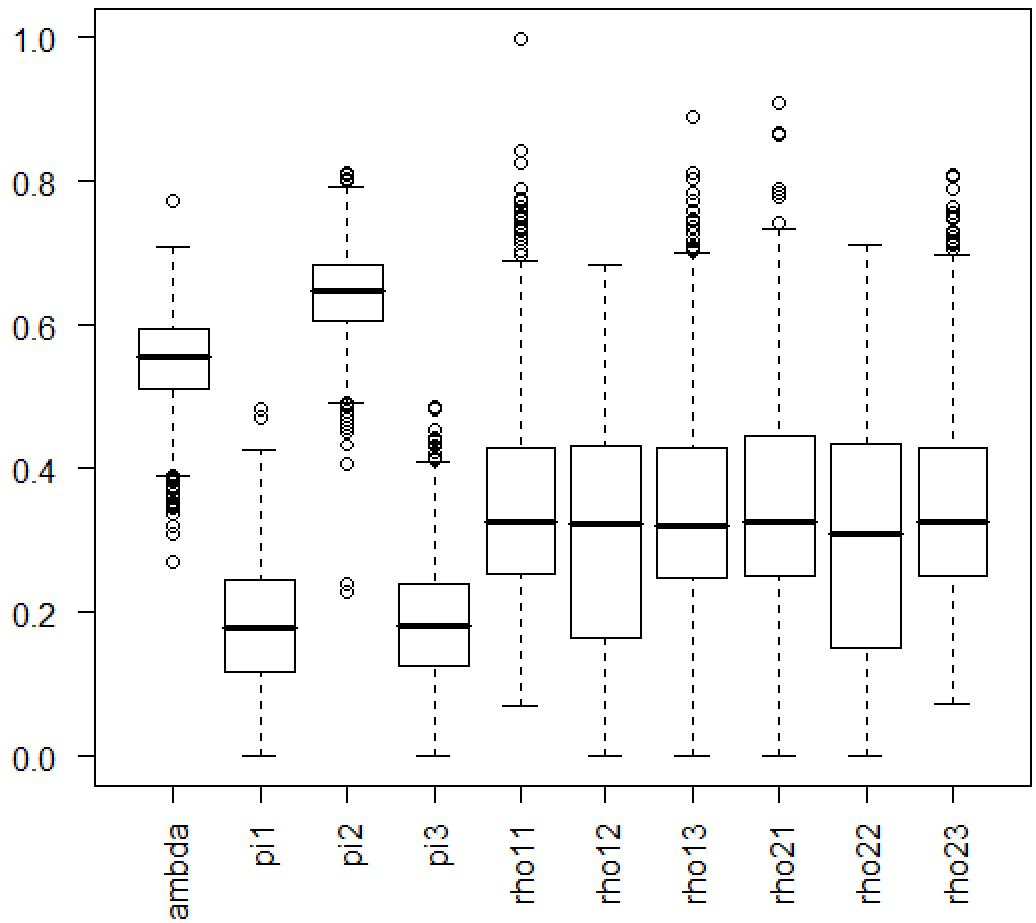


Figure 2.9: Boxplots of estimated proportion parameters by a two-level model based on data from a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1.0 (y-axis).

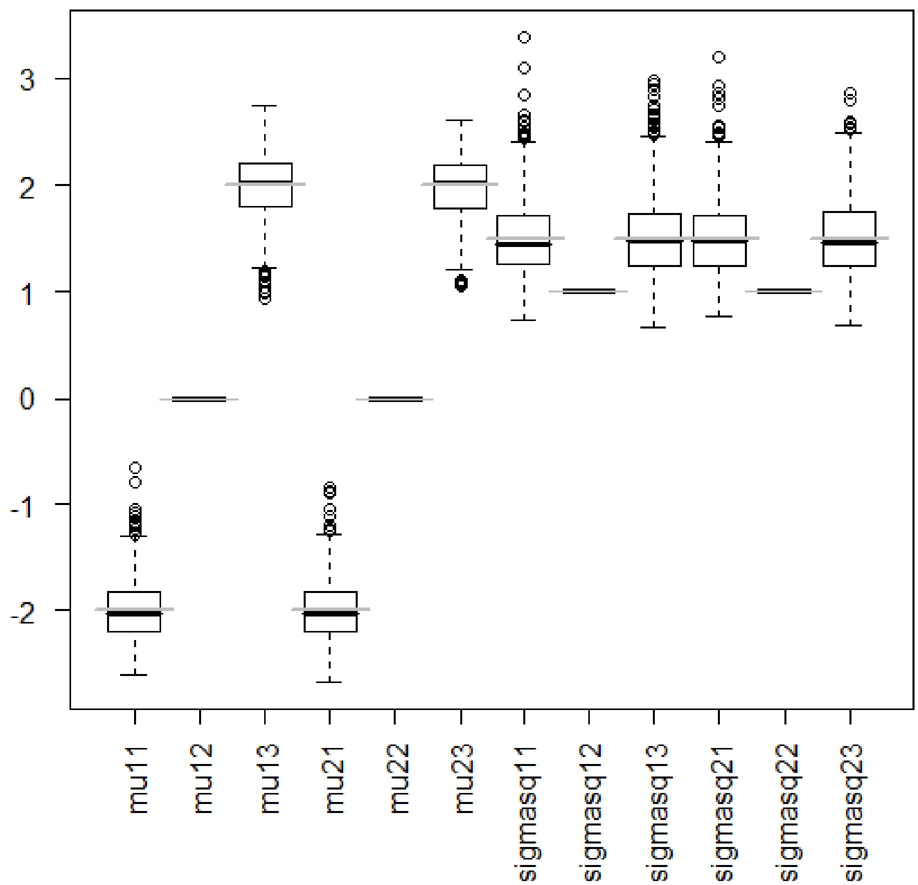


Figure 2.10: Boxplots of estimated means and variances by a two-level model based on data from a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.

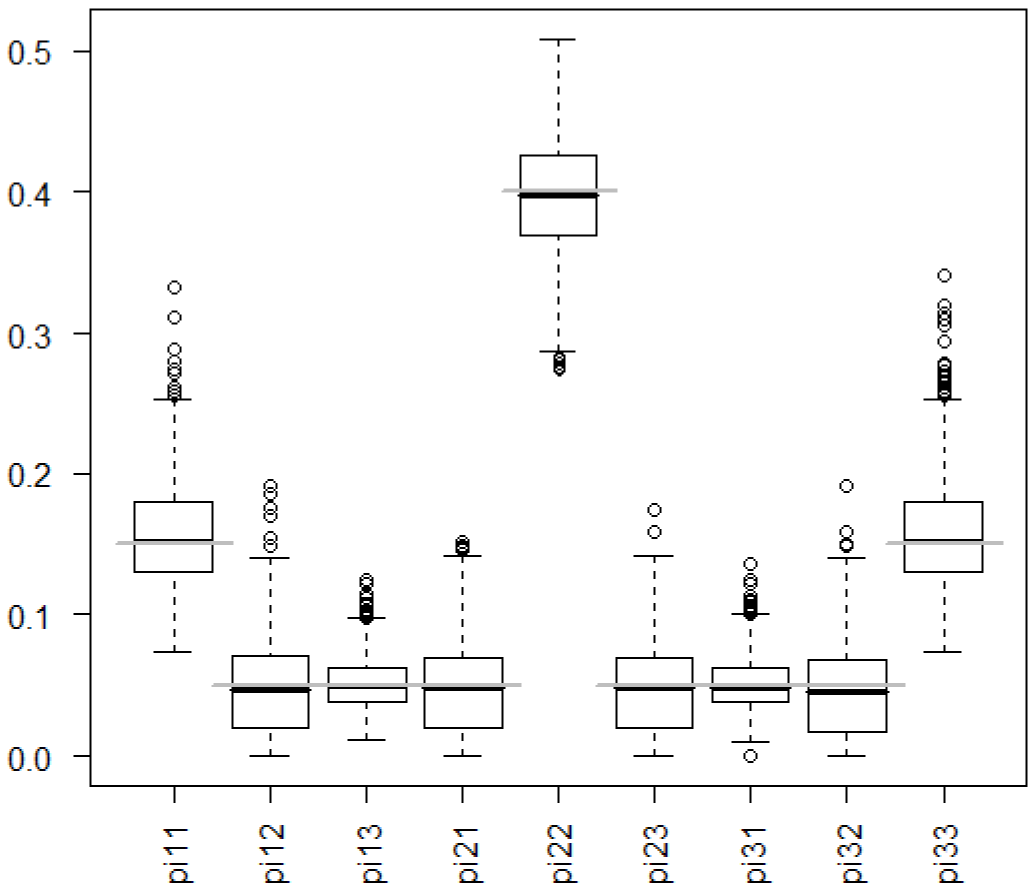


Figure 2.11: Boxplots of converted proportion parameters approximated by a two-level model based on data from a moderate PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters $\boldsymbol{\pi}[\mathbf{1,}]$, $\boldsymbol{\pi}[\mathbf{2,}]$, $\boldsymbol{\pi}[\mathbf{3,}]$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.5 (y-axis).

Case 8: we consider a difficult case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1$. In this case, the number of non-convergence is 16 out of 1000. Table 2.8 shows the summary, and Figure 2.12 and 2.13 display the box plots for each parameter. In Figure 2.12, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis). In Figure 2.13, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3. The estimated results for μ's and σ's are close to the true values, while some of those for proportions are not. We also calculate approximated $\boldsymbol{\pi}$ matrix in PCD model by the two level model. The estimated proportions are not far from the true. We present the box plots for converted results to approximate the proportion parameters in the original PCD model in Figure 2.14.

Table 2.8: Parameters estimation by a two-level model based on data from a difficult PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$

Parameter	True Values	Mean	Standard Deviation
PCD model			
$(\pi_{11}, \pi_{12}, \pi_{13})$	(0.15, 0.05, 0.05)	(0.171, 0.052, 0.048)	(0.089, 0.059, 0.037)
$(\pi_{21}, \pi_{22}, \pi_{23})$	(0.15, 0.20, 0.15)	(0.134, 0.192, 0.136)	(0.090, 0.094, 0.091)
$(\pi_{31}, \pi_{32}, \pi_{33})$	(0.05, 0.05, 0.15)	(0.046, 0.050, 0.172)	(0.036, 0.055, 0.095)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.474, 0, 1.481)	(0.5230, 0, 0.509)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.526, 0, 1.515)	(0.3740, 0, 0.385)
$(\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.5187, 1, 1.493)	(0.490, 0, 0.478)
$(\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.4736, 1, 1.485)	(0.373, 0, 0.391)
Two-level model			
λ		0.331	0.135
(π_1, π_2, π_3)		(0.388, 0.219, 0.392)	(0.246, 0.205, 0.253)
$(\rho_{11}, \rho_{12}, \rho_{13})$		(0.229, 0.549, 0.222)	(0.161, 0.212, 0.156)
$(\rho_{21}, \rho_{22}, \rho_{23})$		(0.348, 0.300, 0.352)	(0.183, 0.232, 0.184)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.474, 0, 1.481)	(0.523, 0, 0.509)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.526, 0, 1.514)	(0.374, 0, 0.385)
$(\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.519, 1, 1.493)	(0.490, 0, 0.478)
$(\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.474, 1, 1.485)	(0.373, 0, 0.391)

We use results of λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ estimated by the two-level model to approximate π_{ij} in orginal PCD model. When $i = j$, the $\pi_{ii} \approx \lambda\pi_i + (1-\lambda)\rho_{1i}\rho_{2i}$; when $i \neq j$, $\pi_{ij} \approx (1-\lambda)\rho_{1i}\rho_{2j}$. Each π_{ij} can be approximated for PCD model, and then the mean and standard deviation of π_{ij} can be calculated.

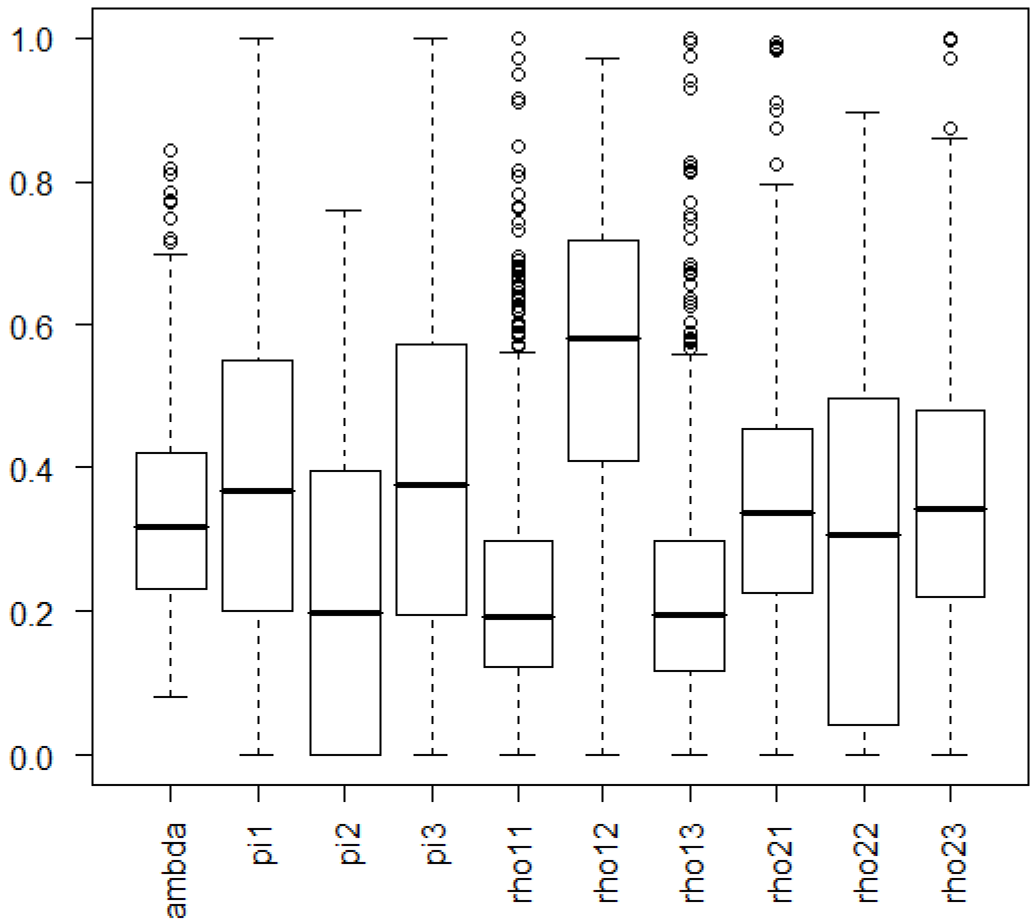


Figure 2.12: Boxplots of estimated proportion parameters by a two-level model based on data from a difficult PCD model with retriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 1 (y-axis).

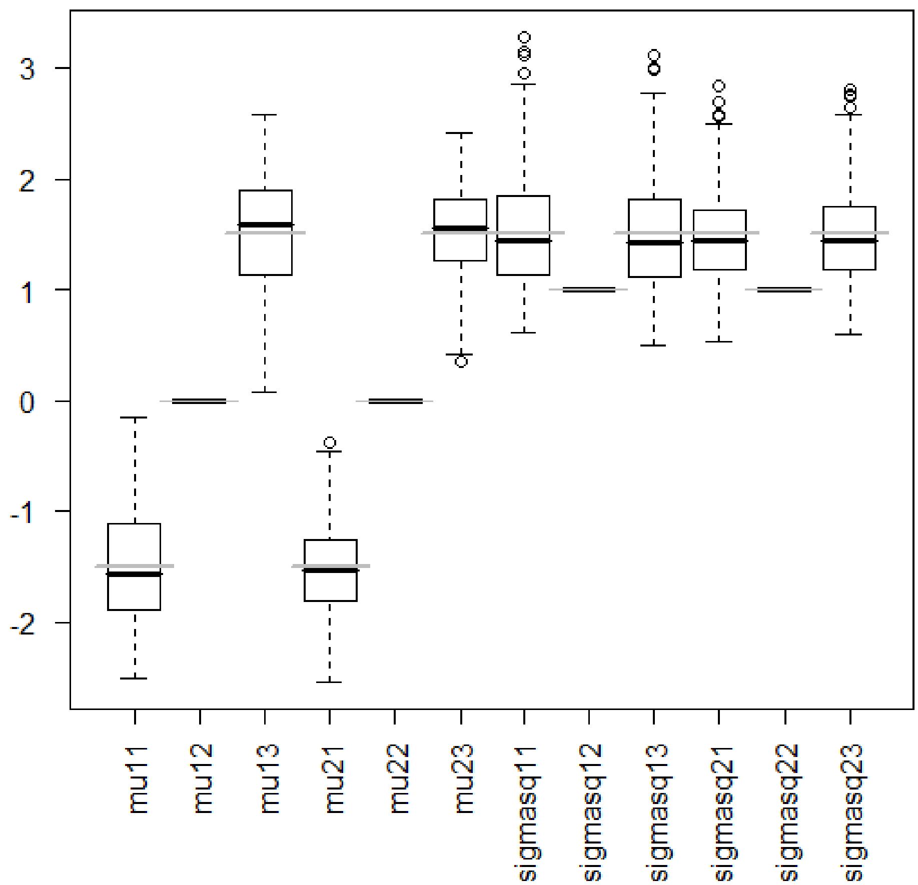


Figure 2.13: Boxplots of estimated means and variances by a two-level model based on data from a difficult PCD model with retriction when $B = 1000$, $N = 1000$, $p = 2$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$ in horizontal sequence are with vertical scale from -2 to 3.

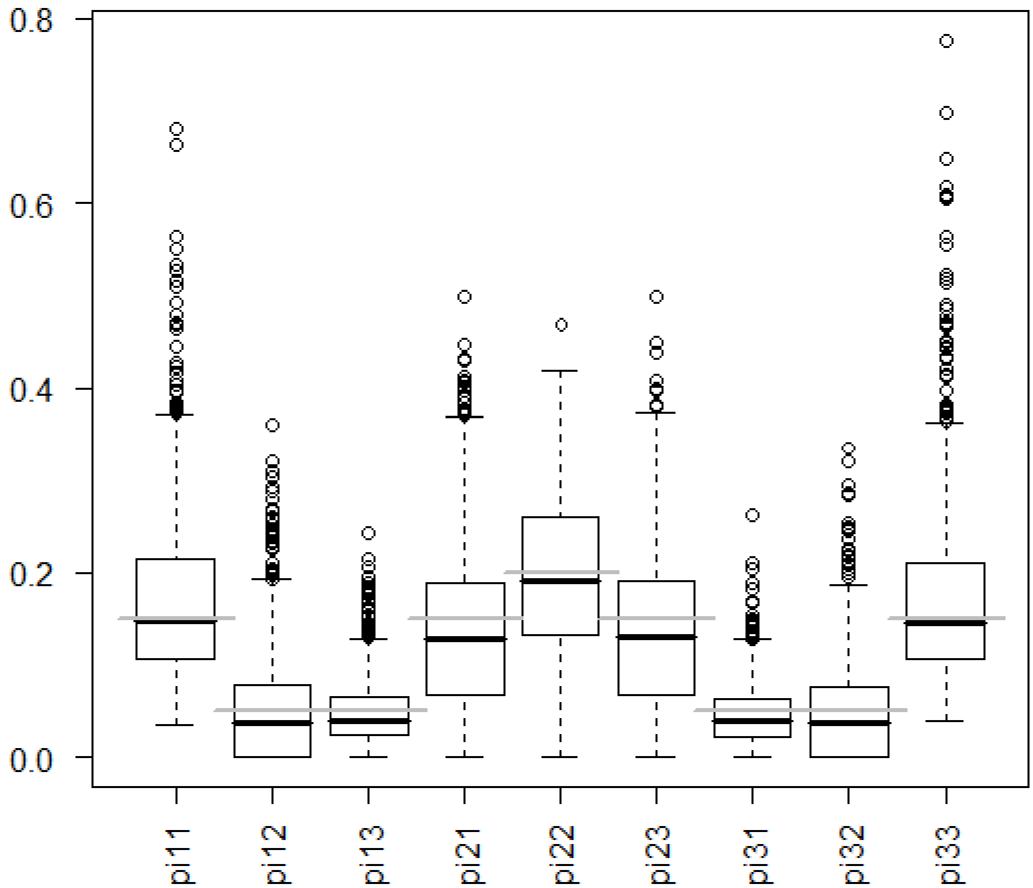


Figure 2.14: Boxplots of converted estimated proportion parameters approximated by a two-level model based on data from a difficult PCD model with restriction when $B = 1000$, $N = 1000$, $p = 2$. The proportion parameters $\boldsymbol{\pi}[\mathbf{1},]$, $\boldsymbol{\pi}[\mathbf{2},]$, $\boldsymbol{\pi}[\mathbf{3},]$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.8 (y-axis).

2.3.2 Multilevel Normal Mixture Model

To illustrate the proposed method based on the multilevel mixture model (2.36) for multiple dimensions in Section 2.2.2, experiments are performed with simulated datasets as following.

Case 1: we consider a moderate case without repetitions when there is a restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. We generate $N = 1000$ 3-dimensional normal data based on our model (2.36). The estimated results are shown in Table 2.9, and we can see that some of them are in an acceptable range but not very close to the initial values.

Table 2.9: Parameters estimation in a moderate two-level model with restriction when $B = 1$, $N = 1000$, $p = 3$

Parameter	True Values	Estimates by EM
λ	0.7	0.687
$\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$	(0.25, 0.6, 0.15)	(0.267, 0.561, 0.173)
$\boldsymbol{\rho_1} = (\rho_{11}, \rho_{12}, \rho_{13})$	(0.3, 0.5, 0.2)	(0.172, 0.637, 0.191)
$\boldsymbol{\rho_2} = (\rho_{21}, \rho_{22}, \rho_{23})$	(0.2, 0.5, 0.3)	(0.201, 0.440, 0.359)
$\boldsymbol{\rho_3} = (\rho_{31}, \rho_{32}, \rho_{33})$	(0.25, 0.5, 0.25)	(0.233, 0.601, 0.165)
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.129, 0, 2.006)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.913, 0, 1.633)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-1.912, 0, 2.127)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.300, 1, 1.592)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.227, 1, 1.578)
$\boldsymbol{\sigma_3^2} = (\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.219, 1, 1.445)

As what we have done for bivariate cases, for a comprehensive performance evaluations, we conduct simulation studies for multivariate mixture model to understand the estimation performance of the two-level model (2.36). For a reasonable parameter configuration, we repeat the estimation for 1000 times, save all these estimation results, and then generate boxplots of estimated parameters.

Case 2: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1, \sigma^2_{32} = 1$. In this case, the number of non-convergence is 0. Table 2.10 shows the summary, and Figure 2.15 and 2.16 displays the box plots for each parameter. In Figure 2.15, the proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ and $\boldsymbol{\rho_3}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.8 (y-axis). In Figure 2.16, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma^2_1}$, $\boldsymbol{\sigma^2_2}$, $\boldsymbol{\sigma^2_3}$ in horizontal sequence are with vertical scale from -2 to 2. Comparing with the estimation in Case 1, the results are much more closer to the initial values with relatively small variances.

Table 2.10: Parameters estimation in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 3$

Parameter	True Values	Mean	Standard Deviation
λ	0.7	0.7000	0.0510
(π_1, π_2, π_3)	(0.25, 0.6, 0.15)	(0.251, 0.595, 0.154)	(0.034, 0.033, 0.030)
$(\rho_{11}, \rho_{12}, \rho_{13})$	(0.3, 0.5, 0.2)	(0.308, 0.494, 0.198)	(0.089, 0.123, 0.075)
$(\rho_{21}, \rho_{22}, \rho_{23})$	(0.2, 0.5, 0.3)	(0.207, 0.488, 0.305)	(0.073, 0.122, 0.090)
$(\rho_{31}, \rho_{32}, \rho_{33})$	(0.25, 0.5, 0.25)	(0.256, 0.492, 0.252)	(0.078, 0.120, 0.081)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.004, 0, 1.999)	(0.144, 0, 0.216)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-2.000, 0, 1.986)	(0.152, 0, 0.191)
$(\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-2.007, 0, 1.995)	(0.142, 0, 0.197)
$(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.506, 1, 1.492)	(0.209, 0, 0.298)
$(\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.485, 1, 1.503)	(0.220, 0, 0.280)
$(\sigma^2_{31}, \sigma^2_{32}, \sigma^2_{33})$	(1.5, 1, 1.5)	(1.486, 1, 1.492)	(0.212, 0, 0.287)

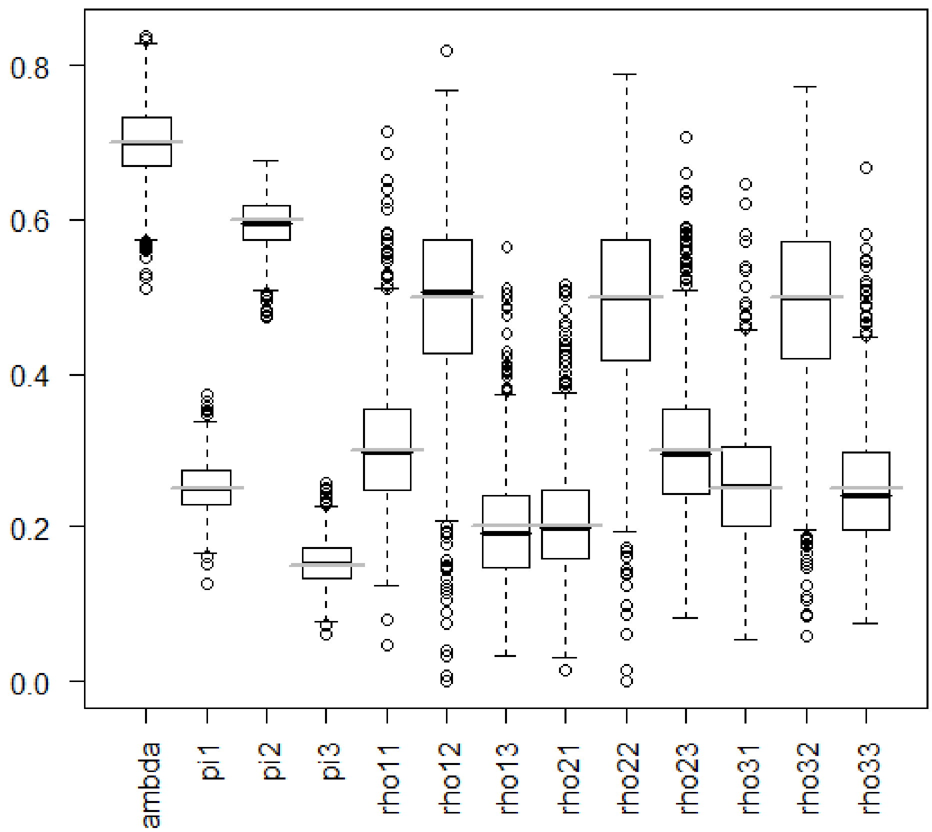


Figure 2.15: Boxplots of estimated proportion parameters in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 3$. The proportion parameters λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$ and $\boldsymbol{\rho_3}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.8 (y-axis).

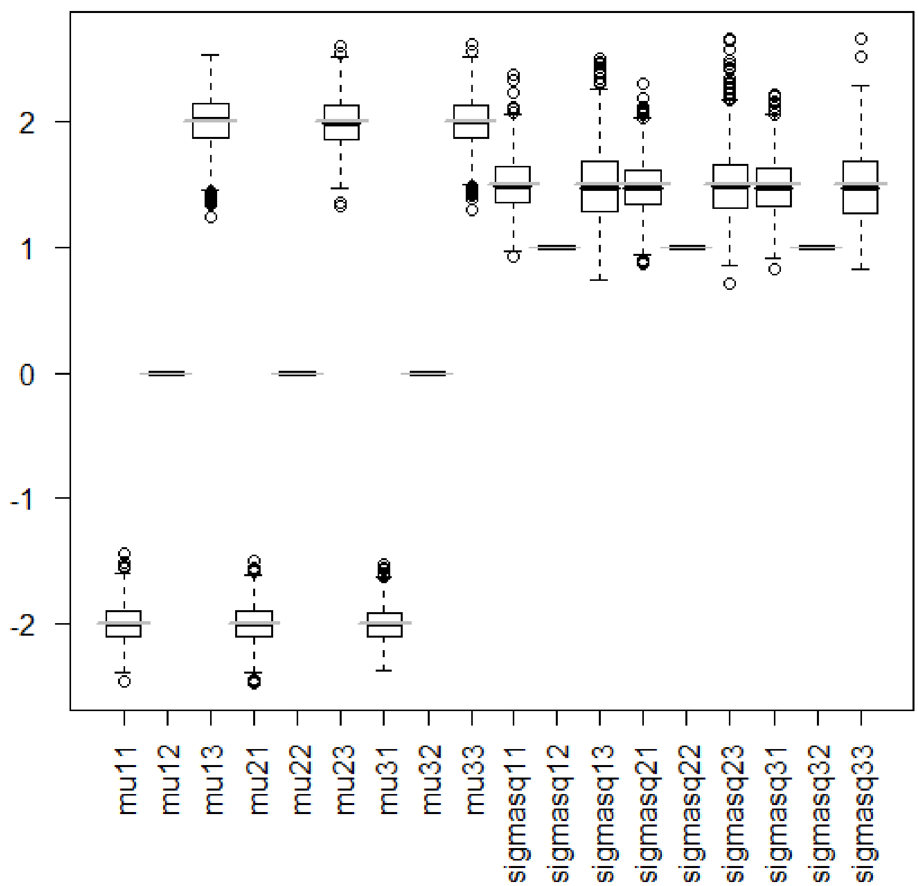


Figure 2.16: Boxplots of estimated means and variances in a moderate two-level model with restriction when $B = 1000$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2.

For comparison, we also conduct simulation studies to understand the estimation performance of the original PCD model (2.26) in Section 2.2.2. For a reasonable parameter configuration, we repeat the estimation for 100 times, save all these estimation results, and then generate boxplots of estimated parameters.

Case 3: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. In this case, the number of non-convergence is 0. Table 2.11 shows the summary, and the box plots for each parameter are shown in Figure 2.17 and 2.18. In Figure 2.17, the vector of proportion parameters $\boldsymbol{\pi}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.25 (y-axis); in Figure 2.18, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2. We can see that some of the estimated results are in an acceptable range but not very close to the initial values.

Table 2.11: Parameters estimation in a moderate PCD model with restriction when $B = 100$, $N = 1000$, $p = 3$

Parameter	True Values	Mean	Standard Deviation
$(\pi_{111}, \pi_{112}, \pi_{113})$	(0.1, 0.025, 0.025)	(0.098, 0.026, 0.023)	(0.023, 0.019, 0.013)
$(\pi_{121}, \pi_{122}, \pi_{123})$	(0.025, 0.025, 0.025)	(0.027, 0.026, 0.026)	(0.018, 0.019, 0.015)
$(\pi_{131}, \pi_{132}, \pi_{133})$	(0.025, 0.025, 0.025)	(0.024, 0.025, 0.024)	(0.012, 0.015, 0.012)
$(\pi_{211}, \pi_{212}, \pi_{213})$	(0.025, 0.025, 0.025)	(0.024, 0.027, 0.024)	(0.016, 0.020, 0.014)
$(\pi_{221}, \pi_{222}, \pi_{223})$	(0.025, 0.2, 0.025)	(0.028, 0.203, 0.028)	(0.018, 0.029, 0.019)
$(\pi_{231}, \pi_{232}, \pi_{233})$	(0.025, 0.025, 0.025)	(0.025, 0.025, 0.025)	(0.013, 0.017, 0.018)
$(\pi_{311}, \pi_{312}, \pi_{313})$	(0.025, 0.025, 0.025)	(0.024, 0.021, 0.023)	0.012, 0.015, 0.011)
$(\pi_{321}, \pi_{322}, \pi_{323})$	(0.025, 0.025, 0.025)	(0.027, 0.027, 0.028)	(0.017, 0.018, 0.020)
$(\pi_{331}, \pi_{332}, \pi_{333})$	(0.025, 0.025, 0.1)	(0.022, 0.025, 0.094)	(0.011, 0.018, 0.026)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.017, 0, 2.039)	(0.219, 0, 0.228)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-2.055, 0, 2.057)	(0.213, 0, 0.208)
$(\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-2.014, 0, 2.041)	(0.243, 0, 0.226)
$(\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.466, 1, 1.450)	(0.267, 0, 0.297)
$(\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.434, 1, 1.452)	(0.276, 0, 0.284)
$(\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.490, 1, 1.424)	(0.274, 0, 0.269)

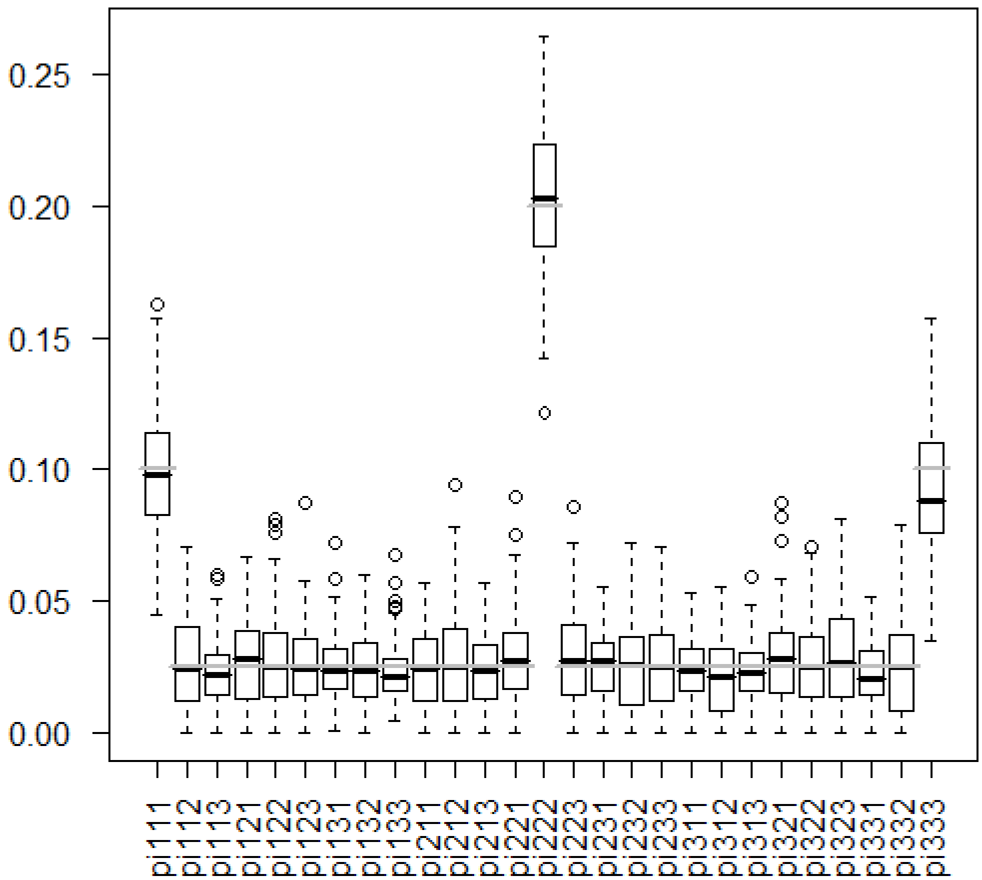


Figure 2.17: Boxplots of estimated proportion parameters in a moderate PCD model with restriction when $B = 100$, $N = 1000$, $p = 3$. The vector of proportion parameters $\boldsymbol{\pi}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.25 (y-axis).

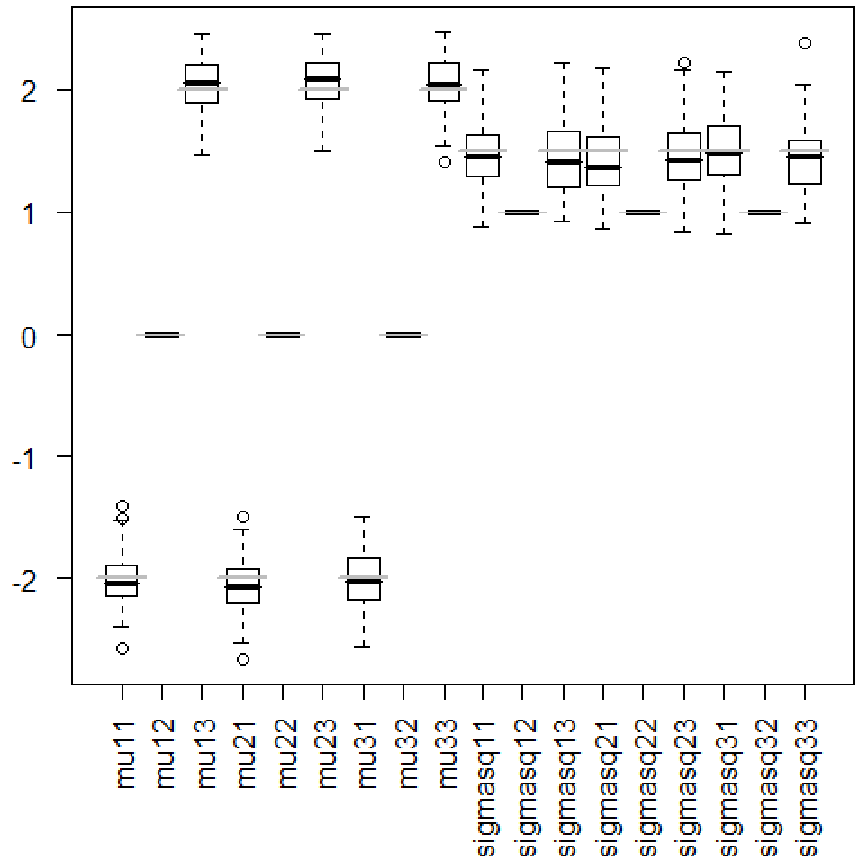


Figure 2.18: Boxplots of estimated means and variances in a moderate PCD model with restriction when $B = 100$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence (x-axis) are with vertical scale from -2 to 2 (y-axis).

Case 4: for large sample size N=10000, we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1, \sigma^2_{32} = 1$. In this case, the number of non-convergence is 0. Table 2.12 shows the summary, and the box plots for each parameter are shown in Figure 2.19 and Figure 2.20. In Figure 2.19, the vector of proportion parameters $\boldsymbol{\pi}$ in sequence horizontally (x-axis) are visualized with vertical scale from 0 to 0.20 (y-axis); in Figure 2.20, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -4 to 4. Comparing to the results in Case 3, the results are much more closer to the initial values with relatively small variances.

Table 2.12: Parameters estimation in a moderate PCD model with restriction when $B = 100$, $N = 10000$, $p = 3$

Parameter	True Values	Mean	Standard Deviation
$(\pi_{111}, \pi_{112}, \pi_{113})$	(0.1, 0.025, 0.025)	(0.010, 0.025, 0.026)	(0.004, 0.002, 0.002)
$(\pi_{121}, \pi_{122}, \pi_{123})$	(0.025, 0.025, 0.025)	(0.026, 0.025, 0.025)	(0.003, 0.003, 0.002)
$(\pi_{131}, \pi_{132}, \pi_{133})$	(0.025, 0.025, 0.025)	(0.024, 0.025, 0.026)	(0.002, 0.002, 0.002)
$(\pi_{211}, \pi_{212}, \pi_{213})$	(0.025, 0.025, 0.025)	(0.025, 0.025, 0.025)	(0.003, 0.002, 0.002)
$(\pi_{221}, \pi_{222}, \pi_{223})$	(0.025, 0.2, 0.025)	(0.025, 0.201, 0.025)	(0.003, 0.007, 0.003)
$(\pi_{231}, \pi_{232}, \pi_{233})$	(0.025, 0.025, 0.025)	(0.025, 0.025, 0.025)	(0.002, 0.002, 0.003)
$(\pi_{311}, \pi_{312}, \pi_{313})$	(0.025, 0.025, 0.025)	(0.025, 0.025, 0.025)	0.002, 0.002, 0.002)
$(\pi_{321}, \pi_{322}, \pi_{323})$	(0.025, 0.025, 0.025)	(0.026, 0.024, 0.025)	(0.003, 0.003, 0.003)
$(\pi_{331}, \pi_{332}, \pi_{333})$	(0.025, 0.025, 0.1)	(0.025, 0.025, 0.098)	(0.002, 0.003, 0.004)
$(\mu_{11}, \mu_{12}, \mu_{13})$	(-4, 0, 4)	(-3.991, 0, 3.998)	(0.054, 0, 0.067)
$(\mu_{21}, \mu_{22}, \mu_{23})$	(-4, 0, 4)	(-4.008, 0, 4.006)	(0.064, 0, 0.051)
$(\mu_{31}, \mu_{32}, \mu_{33})$	(-4, 0, 4)	(-3.997, 0, 4.016)	(0.077, 0, 0.066)
$(\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(3, 1, 3)	(3.016, 1, 2.996)	(0.134, 0, 0.138)
$(\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(3, 1, 3)	(3.010, 1, 2.986)	(0.148, 0, 0.123)
$(\sigma^2_{31}, \sigma^2_{32}, \sigma^2_{33})$	(3, 1, 3)	(3.018, 1, 2.944)	(0.143, 0, 0.141)

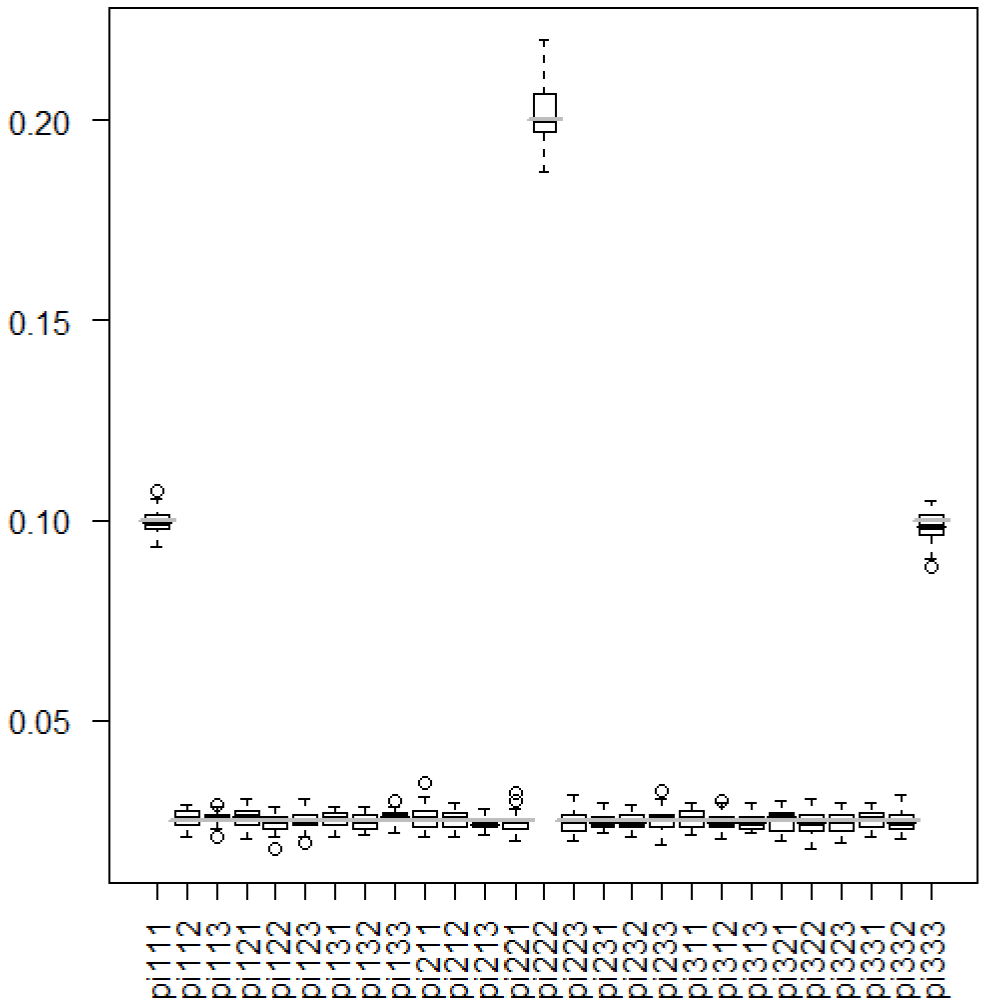


Figure 2.19: Boxplots of estimated proportion parameters in a moderate PCD model with restriction when $B = 100$, $N = 10000$, $p = 3$. The vector of proportion parameters $\boldsymbol{\pi}$ in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.20 (y-axis).

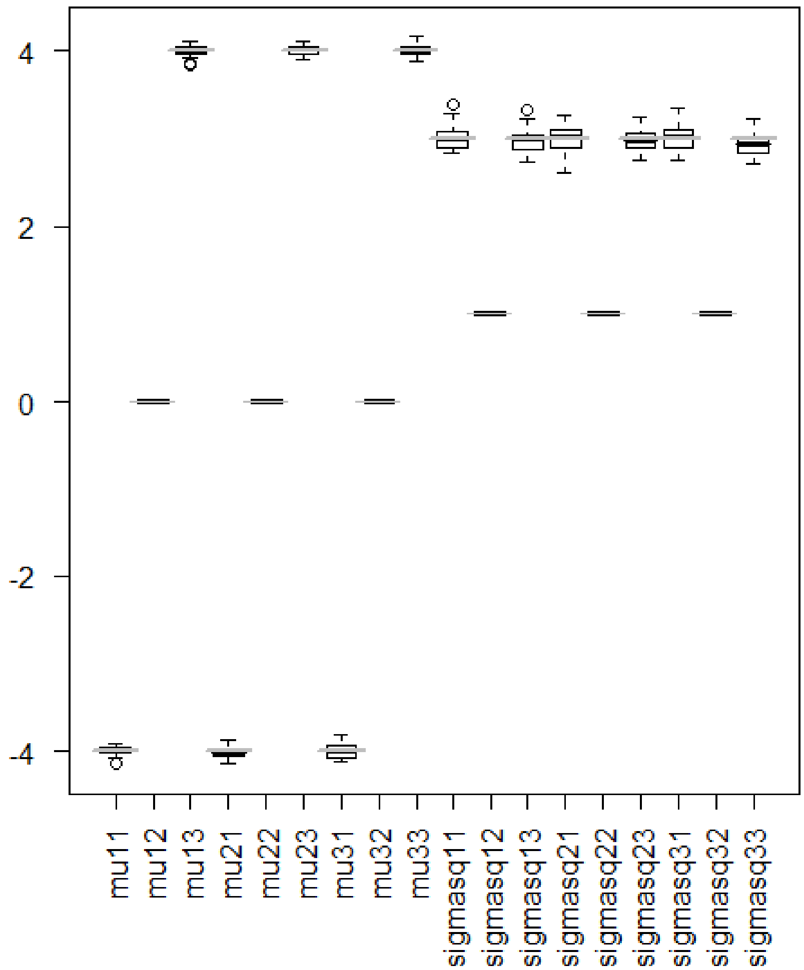


Figure 2.20: Boxplots of estimated means and variances in a moderate PCD model with restriction when $B = 100$, $N = 10000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence (x-axis) are with vertical scale from -4 to 4 (y-axis).

2.4 Mathematical Derivations

Formula (2.4) and (2.5) derivations for E and M steps in Section 2.1

$$\begin{aligned}
Q(\Theta, \Theta') &= \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x}, \mathbf{z}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj}))|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})|\Theta)] \\
&= \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj}))
\end{aligned}$$

E-step:

$$\begin{aligned}
&\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta) \\
&= 0\times\Pr(\prod_{l=1}^{p} z_{li_lk}=0|\mathbf{x_k},\Theta)+1\times\Pr(\prod_{l=1}^{p} z_{li_lk}=1|\mathbf{x_k},\Theta) \\
&= \Pr(\prod_{l=1}^{p} z_{li_lk}=1|\mathbf{x_k},\Theta) \\
&= \frac{\Pr(z_{1i_1k}=1,\ldots,z_{pi_pk}=1,\mathbf{x_k},\Theta)}{\Pr(\mathbf{x_k},\Theta)} \\
&= \frac{\Pr(z_{1i_1k}=1,\ldots,z_{pi_pk}=1)\times\Pr(\mathbf{x_k},\Theta|z_{1i_1k}=1,\ldots,z_{pi_pk}=1)}{\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\Pr(z_{1i_1k}=1,\ldots,z_{pi_pk}=1)\times\Pr(\mathbf{x_k},\Theta|z_{1i_1k}=1,\ldots,z_{pi_pk}=1)} \\
&= \frac{\pi_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{k_j})}{\sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_p=1}^{g}\pi_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{k_j})}.
\end{aligned}$$

Note that, since for any fixed $i_1, i_2, \ldots, i_{p-1}$, $\{z_{1i_1k}=1, z_{2i_2k}=1, \ldots, z_{pi_pk}=1\}$ is a set of pairwise disjoint events for $i_p = 1, 2, \ldots, g$. For example, the event $\{z_{1i_1k} = 1, z_{2i_2k}=1, \ldots, z_{pi_pk}=1\}$ is disjoint with the event $\{z_{1i_1k}=1, z_{2i_2k}=1, \ldots, z_{pi_{p'}k} = 1\}$, when $i_p \neq i_{p'}$. Then for any fixed $i_1, i_2, \ldots, i_{p-2}$, $\{z_{1i_1k}=1, z_{2i_2k}=1, \ldots, z_{pi_pk} =$

$1\}$ is a set of pairwise disjoint events for $i_{p-1} = 1, 2, \ldots, g$ and $i_p = 1, 2, \ldots, g$. And so force, for any $i_1, i_2, \ldots, i_p$, $\{z_{1i_1k} = 1, z_{2i_2k} = 1, \ldots, z_{pi_pk} = 1\}$ is a set of pairwise disjoint events. Moreover, $\sum_{i_1=1}^{g} \ldots \sum_{i_p=1}^{g} \Pr(z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1) = 1$. Therefore, by the law of total probability,

$$\Pr(\mathbf{x_k}, \Theta) = \sum_{i_1=1}^{g} \ldots \sum_{i_p=1}^{g} \Pr(z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1)\Pr(\mathbf{x_k}, \Theta | z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1).$$

M-step: Let $\hat{u}_{i_1 i_2 \ldots i_p k} = \mathrm{E}(\prod_{l=1}^{p} z_{li_lk} | \Theta)$, the optimization of $\pi_{i_1 i_2 \ldots i_p}$, μ_{ji_j}, $\sigma_{ji_j}^2$ is simply a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta, \Theta')}{\partial \pi'_{i_1 i_2 \ldots i_p}} = \frac{\sum_{k=1}^{n} \hat{u}_{i_1 i_2 \ldots i_p k}}{\pi'_{i_1 i_2 \ldots i_p}} - \frac{\sum_{k=1}^{n} \hat{u}_{gg \ldots gk}}{\pi'_{gg \ldots g}} = 0,$$

for $i_1, i_2, \ldots, i_p = 1, 2, \ldots, g-1$ the above equation holds.

Subject to $\sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \ldots \sum_{i_p=1}^{g} \pi_{i_1 i_2 \ldots i_p} = 1$,

$$\begin{aligned}
&\frac{\sum_{k=1}^{n} \hat{u}_{i_1 i_2 \ldots i_p k}}{\pi'_{i_1 i_2 \ldots i_p}} \\
= \;& \frac{\sum_{k=1}^{n} \hat{u}_{gg \ldots gk}}{\pi'_{gg \ldots g}} \\
= \;& \frac{\sum_{k=1}^{n} \hat{u}_{11 \ldots 1k}}{\pi'_{11 \ldots 1}} = \frac{\sum_{k=1}^{n} \hat{u}_{11 \ldots 2k}}{\pi'_{11 \ldots 2}} \\
= \;& \ldots = \frac{\sum_{k=1}^{n} \hat{u}_{g-1, g-1, \ldots, g-2, k}}{\pi'_{11 \ldots 1}} = \frac{\sum_{k=1}^{n} \hat{u}_{g-1, g-1 \ldots g-1, k}}{\pi'_{g-1, g-1 \ldots g-1}} \\
= \;& \frac{\sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \ldots \sum_{i_p=1}^{g} \sum_{k=1}^{n} \hat{u}_{i_1 i_2 \ldots i_p k}}{\sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \ldots \sum_{i_p=1}^{g} \pi'_{i_1 i_2 \ldots i_p}} \\
= \;& \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \ldots \sum_{i_p=1}^{g} \sum_{k=1}^{n} \hat{u}_{i_1 i_2 \ldots i_p k} \\
= \;& \frac{\sum_{k=1}^{n} \sum_{i_1=1}^{g} \ldots \sum_{i_p=1}^{g} \Pr(z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1) \cdot \Pr(\mathbf{x_k} | z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1)}{\sum_{i_1=1}^{g} \ldots \sum_{i_p=1}^{g} \Pr(z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1) \cdot \Pr(\mathbf{x_k} | z_{1i_1k} = 1, \ldots, z_{pi_pk} = 1)} \\
= \;& n,
\end{aligned}$$

then,

$$\hat{\pi'}_{i_1 i_2 \ldots i_p} = \frac{\sum_{k=1}^{n} \hat{u}_{i_1 i_2 \ldots i_p k}}{n}.$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \mu'_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}\frac{x_{kj}-\mu'_{ji_j}}{\sigma'^2_{ji_j}} = 0,$$

then,

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}}.$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \sigma'^2_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}[\frac{(x_{kj}-\mu'_{ji_j})^2}{2\sigma'^4_{ji_j}} - \frac{1}{2\sigma'^2_{ji_j}}] = 0,$$

then,

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}(x_{kj}-\mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}}.$$

Therefore,

$$\hat{\pi}_{i_1 i_2 \ldots i_p} = \frac{\sum_{k=1}^{n}\hat{u}_{i_1 i_2 \ldots i_p k}}{n}, \tag{2.40}$$

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}},$$

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}(x_{kj}-\mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2 \ldots i_p k}}.$$

Formula (2.9) and (2.10) derivations for E and M steps in Section 2.1

$$\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x},\mathbf{z}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i=1}^{g} z_{ik}\log\big(\pi_i'\prod_{j=1}^{p}\phi_{\mu_{ji}',\sigma_{ji}'^2}(x_{kj})\big)|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i=1}^{g}\big(z_{ik}\log(\pi_i') + z_{ik}\log(\prod_{j=1}^{p}\phi_{\mu_{ji}',\sigma_{ji}'^2}(x_{kj}))|\Theta\big)] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)\log(\pi_i') + \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu_{ji}',\sigma_{ji}'^2}(x_{kj}))
\end{aligned}$$

E-step:

$$\begin{aligned}
\hat{z}_{ik} &= \mathrm{E}(z_{ik}|\mathbf{x_k},\Theta) \\
&= 0\times\Pr(z_{ik}=0|\mathbf{x_k},\Theta) + 1\times\Pr(z_{ik}=1|\mathbf{x_k},\Theta) \\
&= \Pr(z_{ik}=1|\mathbf{x_k},\Theta) \\
&= \frac{\Pr(z_{ik}=1,\mathbf{x_k},\Theta)}{\Pr(\mathbf{x_k},\Theta)} \\
&= \frac{\Pr(z_{ik}=1)\Pr(\mathbf{x_k},\Theta|z_{ik}=1)}{\sum_{i=1}^{g}\Pr(z_{ik}=1)\Pr(\mathbf{x_k},\Theta|z_{ik}=1)} \\
&= \frac{\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}{\sum_{i=1}^{g}\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})},
\end{aligned}$$

Note that, $\{z_{ik}=1\}$ is a set of pairwise disjoint events for $i=1,2,\ldots,g$, and $\sum_{i=1}^{g}\Pr(z_{ik}=1)=1$, therefore, by the law of total probability,

$$\Pr(\mathbf{x_k},\Theta) = \sum_{i=1}^{g}\Pr(z_{ik}=1)\Pr(\mathbf{x_k},\Theta|z_{ik}=1).$$

M-step: Subject to $\sum_{i=1}^{g}\pi_i=1$, the optimization of π_i, μ_{ji}, σ_{ji}^2 is simply a maximum likelihood estimation of the parameters as below,

$$\hat{\pi}_i = \frac{\sum_{k=1}^{n}\hat{z}_{ik}}{n},$$

$$\hat{\mu}_{ji} = \frac{\sum_{k=1}^{n}\hat{z}_{ik}x_{kj}}{\sum_{k=1}^{n}\hat{z}_{ik}},$$

$$\hat{\sigma}_{ji}^2 = \frac{\sum_{k=1}^{n} \hat{z}_{ik}(x_k - \mu_{ji})^2}{\sum_{k=1}^{n} \hat{z}_{ik}}.$$

Formula (2.24) and (2.25) derivations for E and M steps in Section 2.2.1

$$\begin{aligned}
L_{g_{PCD}}^{*} &= \prod_{k=1}^{n} [\lambda f_{CC}^{*}(x_{k1}, x_{k2})]^{\omega_k} [(1-\lambda) f_{CI}^{*}(x_{k1}, x_{k2})]^{1-\omega_k} \\
&= \lambda^{\sum_{k=1}^{n} \omega_k} (1-\lambda)^{\sum_{k=1}^{n} (1-\omega_k)} \prod_{k=1}^{n} \{ [\prod_{i=1}^{g} \left(\pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})\right)^{z_{ik}^{(CC)}}]^{\omega_k} \\
&\quad [\prod_{h=1}^{g} (\rho_{1h} \phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1}))^{z_{h*k}^{(CI)}}]^{1-\omega_k} [\prod_{j=1}^{g} (\rho_{2j} \phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2}))^{z_{*jk}^{(CI)}}]^{1-\omega_k} \} \\
&= \lambda^{\sum_{k=1}^{n} \omega_k} (1-\lambda)^{\sum_{k=1}^{n} (1-\omega_k)} \\
&\quad \cdot \prod_{k=1}^{n} \{ [\prod_{i=1}^{g} \pi_i^{z_{ik}^{(CC)} \omega_k} [\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k1})]^{z_{ik}^{(CC)} \omega_k}] \\
&\quad \cdot [\prod_{h=1}^{g} \rho_{1h}^{z_{h*k}^{(CI)} (1-\omega_k)} (\phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1}))^{z_{h*k}^{(CI)} (1-\omega_k)}] \\
&\quad \cdot [\prod_{j=1}^{g} \rho_{2j}^{z_{*jk}^{(CI)} (1-\omega_k)} (\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2}))^{z_{*jk}^{(CI)} (1-\omega_k)}] \} \\
&= \lambda^{\sum_{k=1}^{n} \omega_k} (1-\lambda)^{\sum_{k=1}^{n} (1-\omega_k)} \\
&\quad \cdot \prod_{k=1}^{n} \{ \prod_{i=1}^{g} \pi_i^{z_{ik}^{(CC)} \omega_k} [\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k1})]^{z_{ik}^{(CC)} \omega_k} \\
&\quad \cdot \rho_{1i}^{z_{i*k}^{(CI)} (1-\omega_k)} [\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{i*k}^{(CI)} (1-\omega_k)} \\
&\quad \cdot \rho_{2i}^{z_{*ik}^{(CI)} (1-\omega_k)} [\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})]^{z_{*ik}^{(CI)} (1-\omega_k)} \} \\
&= \lambda^{\sum_{k=1}^{n} \omega_k} (1-\lambda)^{\sum_{k=1}^{n} (1-\omega_k)} \prod_{k=1}^{n} \{ \prod_{i=1}^{g} \pi_i^{z_{ik}^{(CC)} \omega_k} \rho_{1i}^{z_{i*k}^{(CI)} (1-\omega_k)} \rho_{2i}^{z_{*ik}^{(CI)} (1-\omega_k)} \\
&\quad \cdot [\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})]^{z_{ik}^{(CC)} \omega_k + z_{i*k}^{(CI)} (1-\omega_k)} [\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})]^{z_{ik}^{(CC)} \omega_k + z_{*ik}^{(CI)} (1-\omega_k)} \}
\end{aligned}$$

We rewrite the $L^*_{f_{PCD}}$ as following,

$$\begin{aligned}
L^*_{f_{PCD}} &= \prod_{k=1}^{n}\prod_{i=1}^{g}\prod_{j=1}^{g}[\pi_{ij}\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})\phi_{\mu_{2j},\sigma^2_{2j}}(x_{k2})]^{z^{(PCD)}_{ijk}} \\
&= \prod_{k=1}^{n}\prod_{i=1}^{g}\prod_{j=1}^{g}\pi_{ij}^{z^{(PCD)}_{ijk}}[\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})]^{z^{(PCD)}_{ijk}}[\phi_{\mu_{2j},\sigma^2_{2j}}(x_{k2})]^{z^{(PCD)}_{ijk}} \\
&= \prod_{k=1}^{n}\{\prod_{i=1}^{g}\prod_{j=1}^{g}\pi_{ij}^{z^{(PCD)}_{ijk}}\}\{\prod_{i=1}^{g}\prod_{j=1}^{g}[\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})]^{z^{(PCD)}_{ijk}}\} \\
&\quad \cdot\{\prod_{i=1}^{g}\prod_{j=1}^{g}[\phi_{\mu_{2j},\sigma^2_{2j}}(x_{k2})]^{z^{(PCD)}_{ijk}}\} \\
&= \prod_{k=1}^{n}\{\prod_{i=1}^{g}\prod_{j=1}^{g}\pi_{ij}^{z^{(PCD)}_{ijk}}\}\{\prod_{i=1}^{g}[\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})]^{z^{(PCD)}_{i.k}}\}\{\prod_{j=1}^{g}[\phi_{\mu_{2j},\sigma^2_{2j}}(x_{k2})]^{z^{(PCD)}_{.jk}}\} \\
&= \prod_{k=1}^{n}\{\prod_{i=1}^{g}\prod_{j=1}^{g}\pi_{ij}^{z^{(PCD)}_{ijk}}\}\{\prod_{i=1}^{g}[\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})]^{z^{(PCD)}_{i.k}}[\phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2})]^{z^{(PCD)}_{.ik}}\}
\end{aligned}$$

where, $z^{(PCD)}_{i.k} = \sum_{j=1}^{g} z^{(PCD)}_{ijk}$, $z^{(PCD)}_{.ik} = \sum_{h=1}^{g} z^{(PCD)}_{hik}$.

$$Q(\Theta,\Theta') = \mathrm{E}_{\boldsymbol{\omega},\mathbf{z}^{(\mathrm{CC})},\mathbf{z_1}^{(\mathrm{CI})},\mathbf{z_2}^{(\mathrm{CI})}}[\log L(\mathbf{x},\boldsymbol{\omega},\mathbf{z}^{(CC)},\mathbf{z_1}^{(CI)},\mathbf{z_2}^{(CI)}|\Theta')|\Theta],$$

where Θ are the current parameters estimates that we used to evaluate the expectation and Θ' are the new parameters that we optimize to increase Q. Recall $L^*_{g_{PCD}}$,

$$\begin{aligned}
L^*_{g_{PCD}} &= \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{\prod_{i=1}^{g}\pi_i^{z^{(CC)}_{ik}\omega_k}\rho_{1i}^{z^{(CI)}_{i*k}(1-\omega_k)}\rho_{2i}^{z^{(CI)}_{*ik}(1-\omega_k)} \\
&\quad [\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1})]^{z^{(CC)}_{ik}\omega_k+z^{(CI)}_{i*k}(1-\omega_k)}[\phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2})]^{z^{(CC)}_{ik}\omega_k+z^{(CI)}_{*ik}(1-\omega_k)}\},
\end{aligned}$$

$$\begin{aligned}
logL^*_{g_{PCD}} &= \sum_{k=1}^{n}\omega_k\log\lambda + \sum_{k=1}^{n}(1-\omega_k)\log(1-\lambda) \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}[\omega_k z^{(CC)}_{ik}\log\pi_i + (1-\omega_k)z^{(CI)}_{i*k}\log\rho_{1i} + (1-\omega_k)z^{(CI)}_{*ik}\log\rho_{2i} \\
&\quad + (\omega_k z^{(CC)}_{ik} + (1-\omega_k)z^{(CI)}_{i*k})\log\phi_{\mu_{1i},\sigma^2_{1i}}(x_{k1}) \\
&\quad + (\omega_k z^{(CC)}_{ik} + (1-\omega_k)z^{(CI)}_{*ik})\log\phi_{\mu_{2i},\sigma^2_{2i}}(x_{k2})].
\end{aligned}$$

$$
\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\boldsymbol{\omega},\mathbf{z}^{(\mathrm{CC})},\mathbf{z_1}^{(\mathrm{CI})},\mathbf{z_2}^{(\mathrm{CI})}}[\log L(\mathbf{x},\boldsymbol{\omega},\mathbf{z}^{(CC)},\mathbf{z_1}^{(CI)},\mathbf{z_2}^{(CI)}|\Theta')|\Theta] \\
&= \log\lambda'\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)+\log(1-\lambda')\sum_{k=1}^{n}\mathrm{E}(1-\omega_k|\Theta) \\
&\quad +\sum_{k=1}^{n}\sum_{i=1}^{g}\{\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\log\pi_i'+\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}|\Theta\big)\log\rho_{1i}' \\
&\quad +\mathrm{E}\big((1-\omega_k)z_{*ik}^{(CI)}|\Theta\big)\log\rho_{2i}' \\
&\quad +[\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)+\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}|\Theta\big)]\log\phi_{\mu_{1i}',\sigma_{1i}'^2}(x_{k1}) \\
&\quad +[\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)+\mathrm{E}\big((1-\omega_k)z_{*ik}^{(CI)}|\Theta\big)]\log\phi_{\mu_{2i}',\sigma_{2i}'^2}(x_{k2})\}
\end{aligned}
$$

E-step:

$$
\begin{aligned}
\mathrm{E}(\omega_k|\Theta) &= \Pr(\omega_k=1|\mathbf{x_k},\Theta) \\
&= \frac{\lambda f_{CC}(\mathbf{x_k})}{\lambda f_{CC}(\mathbf{x_k})+(1-\lambda)f_{CI}(\mathbf{x_k})} \\
&= \lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad \times[\lambda\sum_{i=1}^{g}\pi_i\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad +(1-\lambda)\sum_{i=1}^{g}\sum_{j=1}^{g}\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) &= \Pr(\omega_k z_{ik}^{(CC)}=1|\mathbf{x_k},\Theta) \\
&= \Pr(\omega_k=1,z_{ik}^{(CC)}=1|\mathbf{x_k},\Theta) \\
&= \Pr(\omega_k=1|\mathbf{x_k},\Theta)\times\Pr(z_{ik}^{(CC)}=1|\omega_k=1,\mathbf{x_k},\Theta).
\end{aligned}
$$

Here, $\omega_k = 1$ implies we are using CC model, then

$$\begin{aligned}
\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) &= \frac{\lambda f_{CC}(x_{k1}, x_{k2})}{\lambda f_{CC}(x_{k1}, x_{k2}) + (1-\lambda) f_{CI}(x_{k1}, x_{k2})} \\
&\quad \times \frac{\pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})}{\sum_{i=1}^{g} \pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})} \\
&= \frac{\lambda \pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2})}{\lambda f_{CC}(x_{k1}, x_{k2}) + (1-\lambda) f_{CI}(x_{k1}, x_{k2})} \\
&= \lambda \pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad [\lambda \sum_{i=1}^{g} \pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad + (1-\lambda) \sum_{i=1}^{g} \sum_{j=1}^{g} \rho_{1i} \rho_{2j} \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1}) \phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1}
\end{aligned}$$

Note that, $\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)$ can be zero. Since when we are using the CI model, *i.e.* $\omega_k = 0$, then $\omega_k z_{ik}^{(CC)}$ should be always 0.

$$\begin{aligned}
\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)}|\Theta\big) &= \Pr((1-\omega_k) z_{i*k}^{(CI)} = 1|\mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0, z_{i*k}^{(CI)} = 1|\mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0|\mathbf{x_k}, \Theta) \times \Pr(z_{i*k}^{(CI)} = 1|\omega_k = 0, \mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0|\mathbf{x_k}, \Theta) \times \Pr(z_{i*k}^{(CI)} = 1|\omega_k = 0, \mathbf{x_k}, \Theta) \\
&\quad \times \sum_{j=1}^{g} \Pr(z_{*jk}^{(CI)} = 1|\omega_k = 0, \mathbf{x_k}, \Theta) \\
&= \sum_{j=1}^{g} \Pr(\omega_k = 0|\mathbf{x_k}, \Theta) \times \Pr(z_{i*k}^{(CI)} = 1|\omega_k = 0, \mathbf{x_k}, \Theta) \\
&\quad \times \Pr(z_{*jk}^{(CI)} = 1|\omega_k = 0, \mathbf{x_k}, \Theta) \\
&= \sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big).
\end{aligned}$$

Since,

$$
\begin{aligned}
\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)} | \Theta\big) &= \Pr((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)} = 1 | \mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0, z_{i*k}^{(CI)} = 1, z_{*jk}^{(CI)} = 1 | \mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0 | \mathbf{x_k}, \Theta) \\
&\quad \times \Pr(z_{i*k}^{(CI)} = 1, z_{*jk}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 0 | \mathbf{x_k}, \Theta) \times \Pr(z_{i*k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
&\quad \times \Pr(z_{*jk}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta)
\end{aligned}
$$

We use $\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)} | \Theta\big)$ instead of $\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} | \Theta\big)$ because we have known the joint density $f_{CI}(x_{k1}, x_{k2})$ rather than the marginal one.

When $\omega_k = 0$, which implies we are using CI model, then

$$
\begin{aligned}
\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)} | \Theta\big) &= \frac{(1-\lambda) f_{CI}(x_{k1}, x_{k2})}{\lambda f_{CC}(x_{k1}, x_{k2}) + (1-\lambda) f_{CI}(x_{k1}, x_{k2})} \\
&\quad \times \frac{\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})}{\sum_{i,j=1}^{g} \rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})} \\
&= \frac{(1-\lambda)\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})}{\lambda f_{CC}(x_{k1}, x_{k2}) + (1-\lambda) f_{CI}(x_{k1}, x_{k2})} \\
&= (1-\lambda)\rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2}) \\
&\quad \times [\lambda \sum_{i=1}^{g} \pi_i \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2i},\sigma_{2i}^2}(x_{k2}) \\
&\quad + (1-\lambda) \sum_{i,j=1}^{g} \rho_{1i}\rho_{2j}\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})\phi_{\mu_{2j},\sigma_{2j}^2}(x_{k2})]^{-1}.
\end{aligned}
$$

By symmetry,

$$
\mathrm{E}\big((1-\omega_k) z_{*jk}^{(CI)} | \Theta\big) = \sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k) z_{i*k}^{(CD)} z_{*jk}^{(CI)} | \Theta\big).
$$

Note that, $\mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)} z_{*jk}^{(CI)} | \Theta\big)$ can be zero, since when we are using the CC model, $i.e.$ $1-\omega_k=0$, then $(1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}$ should be always 0.

M-step: The optimization of $\lambda, \pi_i, \rho_{1i}, \rho_{2i}, \mu_{1i}, \mu_{2i}, \sigma_{1i}^2$ and σ_{2i}^2 $(i=1,2,\ldots,g)$ is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta,\Theta')}{\partial\lambda'} = \frac{\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta)}{\lambda'} - \frac{\sum_{k=1}^n \mathrm{E}(1-\omega_k|\Theta)}{1-\lambda'} = 0,$$

then,

$$\begin{aligned}
\hat{\lambda}' &= \frac{\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta)}{\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta) + \sum_{k=1}^n \mathrm{E}(1-\omega_k|\Theta)} \\
&= \frac{\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta)}{\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta) + n - \sum_{k=1}^n \mathrm{E}(\omega_k|\Theta)} \\
&= \frac{1}{n}\sum_{k=1}^n \mathrm{E}(\omega_k|\Theta).
\end{aligned}$$

For $i=1,2,\ldots,g-1$ and subject to $\sum_{i=1}^g \pi_i = 1$, we have

$$\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial\pi_i'} &= \sum_{k=1}^n \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\pi_i'} + \sum_{k=1}^n \mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)\frac{1}{\pi_g'}\frac{\partial\pi_g'}{\partial\pi_i'} \\
&= \sum_{k=1}^n \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\pi_i'} - \sum_{k=1}^n \mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)\frac{1}{\pi_g'} = 0,
\end{aligned}$$

For each $i=1,2,\ldots,g-1$,the following equation holds,

$$\frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\pi_i'} = \frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)}{\pi_g'}$$

which means,

$$\begin{aligned}
\frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)}{\pi_g'} &= \frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{1k}^{(CC)}|\Theta)}{\pi_1'} = \frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{2k}^{(CC)}|\Theta)}{\pi_2'} \\
&= \ldots = \frac{\sum_{k=1}^n \mathrm{E}(\omega_k z_{g-1,k}^{(CC)}|\Theta)}{\pi_{g-1}'} \\
&= = \frac{\sum_{i=1}^g \sum_{k=1}^n \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\sum_{i=1}^g \pi_i'} = \sum_{i=1}^g \sum_{k=1}^n \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta),
\end{aligned}$$

then,

$$\hat{\pi}_i' = \frac{\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}.$$

Similarly, for $i = 1, 2, \ldots, g-1$ and subject to $\sum_{i=1}^{g} \rho_{1i} = 1$, we have

$$\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial \rho_{1h}'} &= \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)}|\Theta\big)\frac{1}{\rho_{1h}'} \\
&\quad + \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{g*k}^{(CI)}|\Theta\big)\frac{1}{\rho_{1g}'}\frac{\partial \rho_{1g}'}{\partial \rho_{1h}'} \\
&= \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{1}{\rho_{1h}'} \\
&\quad + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{g*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{1}{\rho_{1g}'}\frac{\partial \rho_{1g}'}{\partial \rho_{1h}'} \\
&= \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{1}{\rho_{1h}'} \\
&\quad - \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{g*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{1}{\rho_{1g}'} = 0,
\end{aligned}$$

For each $i = 1, 2, \ldots, g-1$,the following equation holds,

$$\frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)}{\rho_{1h}'} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{g*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)}{\rho_{1g}'}$$

which means,

$$
\begin{aligned}
& \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{g*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\rho'_{1g}} \\
= & \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{1*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\rho'_{11}} \\
= & \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{2*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\rho'_{12}} \\
= & \ldots = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{g-1,*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\rho'_{1,g-1}} \\
= & \frac{\sum_{i=1}^{g}\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\sum_{i=1}^{g}\rho'_{1i}} \\
= & \sum_{i=1}^{g}\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big),
\end{aligned}
$$

then,

$$
\hat{\rho'_{1h}} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}.
$$

By symmetry,

$$
\hat{\rho'_{2h}} = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}.
$$

$$
\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial \mu'_{1h}} &= \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}{\partial \mu_{1h}} \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}\big((1-\omega_k) z_{i*k}^{(CI)}|\Theta\big)\frac{1}{\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}{\partial \mu_{1h}} \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1i},\sigma_{1i}^2}(x_{k1})}{\partial \mu_{1h}} \\
&\quad + \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)}|\Theta\big)\frac{1}{\phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}{\partial \mu_{1h}} \\
&= \sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}{\partial \mu_{1h}} \\
&\quad + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{1}{\phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}\frac{\partial \phi_{\mu_{1h},\sigma_{1h}^2}(x_{k1})}{\partial \mu_{1h}} \\
&= \sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)\frac{(x_{k1}-\mu'_{1h})}{\sigma_{1h}'^2} \\
&\quad + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)\frac{(x_{k1}-\mu'_{1h})}{\sigma_{1h}'^2} \\
&= [\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)]\frac{(x_{k1}-\mu'_{1h})}{\sigma_{1h}'^2} \\
&= 0,
\end{aligned}
$$

then,

$$
\hat{\mu'_{1h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)]x_{k1}}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)}.
$$

$$
\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial \sigma_{1h}'^2} &= \sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)(\frac{(x_{k1}-\mu'_{1h})^2}{2\sigma_{1h}'^4} - \frac{1}{2\sigma_{1h}'^2}) \\
&\quad + \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)}|\Theta\big)(\frac{(x_{k1}-\mu'_{1h})^2}{2\sigma_{1h}'^4} - \frac{1}{2\sigma_{1h}'^2}) \\
&= [\sum_{k=1}^{n} \mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}\big((1-\omega_k) z_{h*k}^{(CI)} z_{*jk}^{(CI)}|\Theta\big)] \\
&\quad (\frac{(x_{k1}-\mu'_{1h})^2}{2\sigma_{1h}'^4} - \frac{1}{2\sigma_{1h}'^2}) = 0,
\end{aligned}
$$

then,

$$\hat{\sigma'^2_{1h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)](x_{k1}-\mu'_{1h})^2}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}\big((1-\omega_k)z_{h*k}^{(CI)}z_{*jk}^{(CI)}|\Theta\big)}.$$

By symmetry,

$$\hat{\mu'_{2h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)]x_{k2}}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)},$$

$$\hat{\sigma'^2_{2h}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)](x_{k2}-\mu'_{2h})^2}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta) + \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k)z_{i*k}^{(CI)}z_{*hk}^{(CI)}|\Theta\big)}.$$

Formula (2.39) derivations for E and M steps in Section 2.2.2

$$\begin{aligned}
L^*_{g_{PCD}} &= \prod_{k=1}^{n}[\lambda f^*_{CC}(\mathbf{x_k})]^{\omega_k}[(1-\lambda)f^*_{CI}(\mathbf{x_k})]^{1-\omega_k} \\
&= \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{[\prod_{i=1}^{g}(\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}}]^{\omega_k} \\
&\quad \{\prod_{l=1}^{p}[\prod_{h=1}^{g}(\rho_{lh}\phi_{\mu_{lh},\sigma^2_{lh}}(x_{kl}))^{z_{hlk}^{(CI)}}]\}^{1-\omega_k}\} \\
&= \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{\{\prod_{i=1}^{g}[\pi_i^{z_{ik}^{(CC)}\omega_k}\prod_{j=1}^{p}(\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}\omega_k}]\} \\
&\quad \{\prod_{h=1}^{g}[\prod_{l=1}^{p}(\rho_{lh}^{z_{hlk}^{(CI)}(1-\omega_k)}\phi_{\mu_{lh},\sigma^2_{lh}}(x_{kl})^{z_{hlk}^{(CI)}(1-\omega_k)})]\}\} \\
&= \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{\prod_{i=1}^{g}\pi_i^{z_{ik}^{(CC)}\omega_k}[\prod_{j=1}^{p}(\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}\omega_k}] \\
&\quad [\prod_{l=1}^{p}(\rho_{li}^{z_{ilk}^{(CI)}(1-\omega_k)}\phi_{\mu_{li},\sigma^2_{li}}(x_{kl})^{z_{ilk}^{(CI)}(1-\omega_k)})]\} \\
&= \lambda^{\sum_{k=1}^{n}\omega_k}(1-\lambda)^{\sum_{k=1}^{n}(1-\omega_k)}\prod_{k=1}^{n}\{\prod_{i=1}^{g}\pi_i^{z_{ik}^{(CC)}\omega_k}[\prod_{j=1}^{p}\rho_{ji}^{z_{ijk}^{(CI)}(1-\omega_k)} \\
&\quad (\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}\omega_k+z_{ijk}^{(CI)}(1-\omega_k)}]\}
\end{aligned}$$

We rewrite the $L^*_{f_{PCD}}$ as following,

$$L^*_{f_{PCD}}(\mathbf{x},\mathbf{z}|\Theta^{(f_{PCD})}) = \prod_{k=1}^{n}\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}[\pi_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj})]^{z^{(PCD)}_{i_1i_2\cdots i_pk}}. \quad (2.41)$$

$$\begin{aligned}
L^*_{f_{PCD}} &= \prod_{k=1}^{n}\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}[\pi_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj})]^{z^{(PCD)}_{i_1i_2\cdots i_pk}} \\
&= \prod_{k=1}^{n}\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}[\pi_{i_1i_2\ldots i_p}^{z^{(PCD)}_{i_1i_2\cdots i_pk}}\prod_{j=1}^{p}(\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj}))^{z^{(PCD)}_{i_1i_2\cdots i_pk}}] \\
&= \prod_{k=1}^{n}\{\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}\pi_{i_1i_2\ldots i_p}^{z^{(PCD)}_{i_1i_2\cdots i_pk}}\}\{\prod_{j=1}^{p}[\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}(\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj}))^{z^{(PCD)}_{i_1i_2\cdots i_pk}}]\} \\
&= \prod_{k=1}^{n}\{\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}\pi_{i_1i_2\ldots i_p}^{z^{(PCD)}_{i_1i_2\cdots i_pk}}\} \\
&\quad\cdot\{\prod_{j=1}^{p}[\prod_{i_j=1}^{g}\prod_{i_1,\ldots,i_{j-1},i_{j+1},i_p=1}^{g}(\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj}))^{z^{(PCD)}_{i_1i_2\cdots i_pk}}]\} \\
&= \prod_{k=1}^{n}\{\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}\pi_{i_1i_2\ldots i_p}^{z^{(PCD)}_{i_1i_2\cdots i_pk}}\} \\
&\quad\cdot\{\prod_{j=1}^{p}[\prod_{i=1}^{g}\prod_{i_1,\ldots,i_{j-1},i_{j+1},i_p=1}^{g}(\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{z^{(PCD)}_{i_1i_2\cdots i_{j-1}ii_{j+1}\cdots i_pk}}]\} \\
&= \prod_{k=1}^{n}\{\prod_{i_1=1}^{g}\prod_{i_2=1}^{g}\cdots\prod_{i_p=1}^{g}\pi_{i_1i_2\ldots i_p}^{z^{(PCD)}_{i_1i_2\cdots i_pk}}\} \\
&\quad\{\prod_{j=1}^{p}[\prod_{i=1}^{g}(\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj}))^{\sum_{i_1,i_2,\ldots,i_{j-1},i_{j+1},\ldots,i_p=1}^{g} z^{(PCD)}_{i_1i_2\cdots i_{j-1}ii_{j+1}\cdots i_pk}}]\}
\end{aligned}$$

Here, define $z^{(PCD)}_{..i_{(j)}..k} = \sum_{i_1,i_2,\ldots,i_{j-1},i_{j+1},\ldots,i_p=1}^{g} z^{(PCD)}_{i_1i_2\ldots i_{j-1}ii_{j+1}\ldots i_pk}$. Since we use g_{PCD} to approximate f_{PCD}, then each corresponding term in $L^*_{f_{PCD}}$ should approximate to that in $L^*_{g_{PCD}}$. Then we have the system of approximate equations as following,

$$z^{(PCD)}_{..i_{(j)}..k} \approx \omega_k z^{(CC)}_{ik} + (1-\omega_k)z^{(CI)}_{ijk}, \quad (2.42)$$

where $i = 1, 2, \ldots, g$, for each $k = 1, 2, \ldots, n$.

$$Q(\Theta, \Theta') = \mathrm{E}_{\boldsymbol{\omega}, \mathbf{z}^{(\mathrm{CC})}, \mathbf{z}^{(\mathrm{CD})}}[\log L(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}^{(CC)}, \mathbf{z}^{(CI)})|\Theta],$$

where Θ are the current parameters estimates that we used to evaluate the expectation and Θ' are the new parameters that we optimize to increase Q. Recall $L^*_{g_{PCD}}$,

$$\begin{aligned} L^*_{g_{PCD}} &= \lambda^{\sum_{k=1}^n \omega_k}(1-\lambda)^{\sum_{k=1}^n (1-\omega_k)} \prod_{k=1}^n \{\prod_{i=1}^g \pi_i^{z_{ik}^{(CC)}\omega_k} [\prod_{j=1}^p \rho_{ji}^{z_{ijk}^{(CD)}(1-\omega_k)} \\ & (\phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj}))^{z_{ik}^{(CC)}\omega_k + z_{ijk}^{(CI)}(1-\omega_k)}]\}, \end{aligned}$$

$$\begin{aligned} logL^*_{g_{PCD}} &= \sum_{k=1}^n \omega_k \log\lambda + \sum_{k=1}^n (1-\omega_k)\log(1-\lambda) \\ &+ \sum_{k=1}^n \sum_{i=1}^g [\omega_k z_{ik}^{(CC)} \log \pi_i + \sum_{j=1}^p (1-\omega_k) z_{ijk}^{(CI)} \log \rho_{ji} \\ &+ \sum_{j=1}^p (\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}) \log \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]. \end{aligned}$$

$$\begin{aligned} Q(\Theta, \Theta') &= \mathrm{E}_{\boldsymbol{\omega}, \mathbf{z}^{(\mathrm{CC})}, \mathbf{z}^{(\mathrm{CD})}}[\log L(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}^{(CC)}, \mathbf{z}^{(CI)}|\Theta')|\Theta] \\ &= \log\lambda' \sum_{k=1}^n \mathrm{E}(\omega_k|\Theta) + \log(1-\lambda') \sum_{k=1}^n \mathrm{E}(1-\omega_k|\Theta) \\ &+ \sum_{k=1}^n \sum_{i=1}^g [\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) \log \pi'_i + \sum_{j=1}^p \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big) \log \rho'_{ji} \\ &+ \sum_{j=1}^p \big(\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \mathrm{E}((1-\omega_k) z_{ijk}^{(CI)}|\Theta)\big) \log \phi_{\mu'_{ji}, \sigma'^2_{ji}}(x_{kj})]. \end{aligned}$$

E-step:

$$
\begin{aligned}
\mathrm{E}(\omega_k|\Theta) &= \Pr(\omega_k = 1|\mathbf{x_k}, \Theta) \\
&= \frac{\lambda f_{CC}(\mathbf{x_k})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})} \\
&= \lambda \sum_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})] \\
&\quad \times \{\lambda \sum_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})] + (1-\lambda) \prod_{j=1}^{p} [\sum_{i=1}^{g} \rho_{ji} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})]\}^{-1},
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) &= \Pr(\omega_k z_{ik}^{(CC)} = 1|\mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 1, z_{ik}^{(CC)} = 1|\mathbf{x_k}, \Theta) \\
&= \Pr(\omega_k = 1|\mathbf{x_k}, \Theta) \times \Pr(z_{ik}^{(CC)} = 1|\omega_k = 1, \mathbf{x_k}, \Theta).
\end{aligned}
$$

Here, $\omega_k = 1$ implies we are using CC model, then

$$
\begin{aligned}
\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) &= \frac{\lambda f_{CC}(\mathbf{x_k})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})} \\
&\quad \times \frac{\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}{\sum_{i=1}^{g} [\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})]} \\
&= \frac{\lambda \pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})} \\
&= \lambda \pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj}) \\
&\quad \times \{\lambda \sum_{h=1}^{g} [\pi_h \prod_{j=1}^{p} \phi_{\mu_{jh},\sigma_{jh}^2}(x_{kh})] + (1-\lambda) \prod_{j=1}^{p} [\sum_{h=1}^{g} \rho_{jh} \phi_{\mu_{jh},\sigma_{jh}^2}(x_{kj})]\}^{-1}
\end{aligned}
$$

Note that, $\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)$ can be zero. Since when we are using the CI model, *i.e.* $\omega_k = 0$, then $\omega_k z_{ik}^{(CC)}$ should be always 0.

$$
\begin{aligned}
&\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}|\Theta\big)\\
&= \Pr((1-\omega_k)z_{ijk}^{(CI)}=1|\mathbf{x_k},\Theta)\\
&= \Pr(\omega_k=0, z_{ijk}^{(CI)}=1|\mathbf{x_k},\Theta)\\
&= \Pr(\omega_k=0|\mathbf{x_k},\Theta)\times\Pr(z_{ijk}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&= \Pr(\omega_k=0|\mathbf{x_k},\Theta)\times\Pr(z_{ijk}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\times\sum_{i_1=1}^{g}\Pr(z_{i_11k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\ldots\\
&\quad\times\sum_{i_{j-1}=1}^{g}\Pr(z_{i_{j-1},j-1,k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\times\sum_{i_{j+1}=1}^{g}\Pr(z_{i_{j+1},j+1,k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\ldots\\
&\quad\times\sum_{i_p=1}^{g}\Pr(z_{i_ppk}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&= \sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\ldots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\ldots\sum_{i_p=1}^{g}\Pr(\omega_k=0|\mathbf{x_k},\Theta)\\
&\quad\times\Pr(z_{i_11k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\ldots\\
&\quad\times\Pr(z_{i_{j-1},j-1,k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\times\Pr(z_{ijk}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\times\Pr(z_{i_{j+1},j+1,k}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&\quad\ldots\\
&\quad\times\Pr(z_{i_ppk}^{(CI)}=1|\omega_k=0,\mathbf{x_k},\Theta)\\
&= \sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\ldots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\ldots\sum_{i_p=1}^{g}\mathrm{E}\Big((1-\omega_k)z_{ijk}^{(CI)}\prod_{l=1,l\neq j}^{p}z_{i_llk}^{(CI)}|\Theta\Big).
\end{aligned}
$$

Actually, this $\prod_{l=1}^{p} z_{i_l lk}^{(CI)}$ denotes $z_{i_1**\dots**k}^{(CI)} z_{*i_2**\dots**k}^{(CI)} z_{**i_3**\dots**k}^{(CD)} \dots z_{**\dots**i_p k}^{(CI)}$. Since,

$$
\begin{aligned}
& \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)} \prod_{l=1, l\neq j}^{p} z_{i_l lk}^{(CI)} | \Theta\big) \\
= \; & \Pr((1-\omega_k) z_{ijk}^{(CI)} \prod_{l=1, l\neq j}^{p} z_{i_l lk}^{(CI)} = 1 | \mathbf{x_k}, \Theta) \\
= \; & \Pr(\omega_k = 0, z_{i_1 k}^{(CI)} = 1, \dots, z_{i_p k}^{(CI)} = 1 | \mathbf{x_k}, \Theta) \\
= \; & \Pr(\omega_k = 0 | \mathbf{x_k}, \Theta) \\
& \times \Pr(z_{i_1 1k}^{(CI)} = 1, \dots, z_{i_{j-1}, j-1, k}^{(CI)} = 1, z_{ijk}^{(CI)} = 1, \\
& z_{i_{j+1}, j+1, k}^{(CI)} = 1, \dots, z_{i_p k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
= \; & \Pr(\omega_k = 0 | \mathbf{x_k}, \Theta) \times \Pr(z_{i_1 1k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \times \dots \\
& \times \Pr(z_{i_{j-1}, j-1, k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
& \times \Pr(z_{ijk}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
& \times \Pr(z_{i_{j+1}, j+1, k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \times \dots \\
& \times \Pr(z_{i_p k}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
= \; & \Pr(\omega_k = 0 | \mathbf{x_k}, \Theta) \times \Pr(z_{ijk}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta) \\
& \times \prod_{l=1, l\neq j}^{p} \Pr(z_{i_l lk}^{(CI)} = 1 | \omega_k = 0, \mathbf{x_k}, \Theta).
\end{aligned}
$$

Here, $\omega_k = 0$ implies we are using CI model, then

$$
\begin{aligned}
\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)} \prod_{l=1, l\neq j}^{p} z_{i_l lk}^{(CI)} | \Theta\big) &= \frac{(1-\lambda) f_{CI}(\mathbf{x_k})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})} \\
& \quad \times \frac{\rho_{ji} \phi_{\mu_{ji}, \sigma_{ji}^2}(x_{kj}) \prod_{l=1, l\neq j}^{p} \rho_{l i_l} \phi_{\mu_{l i_l}, \sigma_{l i_l}^2}(x_{kl})}{\sum_{i_1, \dots, i_p = 1}^{g} \prod_{l=1}^{p} \rho_{l i_l} \phi_{\mu_{l i_l}, \sigma_{l i_l}^2}(x_{kl})} \\
&= \frac{(1-\lambda) \rho_{ji} \phi_{\mu_{ji}, \sigma_{ji}^2}(x_{kj}) \prod_{l=1, l\neq j}^{p} \rho_{l i_l} \phi_{\mu_{l i_l}, \sigma_{l i_l}^2}(x_{kl})}{\lambda f_{CC}(\mathbf{x_k}) + (1-\lambda) f_{CI}(\mathbf{x_k})}.
\end{aligned}
$$

Note that, $\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)} \prod_{l=1, l\neq j}^{p} z_{i_l lk}^{(CI)} | \Theta\big)$ can be zero, since when we are using the CC model, *i.e.* $1 - \omega_k = 0$, then $(1-\omega_k) z_{ijk}^{(CI)} \prod_{l=1, l\neq j}^{p} z_{i_l lk}^{(CI)}$ should be always 0.

Therefore,

$$\begin{aligned}
&\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}|\Theta\big)\\
&= \sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}\prod_{l=1,l\neq j}^{p}z_{i_l lk}^{(CI)}|\Theta\big)\\
&= \frac{(1-\lambda)}{\lambda f_{CC}(\mathbf{x_k})+(1-\lambda)f_{CI}(\mathbf{x_k})}\\
&\quad\times\sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}[\rho_{ji}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})\prod_{l=1,l\neq j}^{p}\rho_{li_l}\phi_{\mu_{li_l},\sigma_{li_l}^2}(x_{kl})]\\
&= \frac{(1-\lambda)\rho_{ji}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{kj})\prod_{l=1,l\neq j}^{p}\sum_{h=1}^{g}\rho_{lh}\phi_{\mu_{lh},\sigma_{lh}^2}(x_{kl})}{\lambda f_{CC}(\mathbf{x_k})+(1-\lambda)f_{CI}(\mathbf{x_k})}.
\end{aligned}$$

M-step: The optimization of $\lambda, \pi_i, \rho_{ji}, \mu_{ji}$, and σ_{ji}^2 $(i=1,2,\ldots,g, j=1,2,\ldots,p)$ is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta,\Theta')}{\partial\lambda'}=\frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)}{\lambda'}-\frac{\sum_{k=1}^{n}\mathrm{E}(1-\omega_k|\Theta)}{1-\lambda'}=0,$$

then,

$$\begin{aligned}
\hat{\lambda}' &= \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)}{\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)+\sum_{k=1}^{n}\mathrm{E}(1-\omega_k|\Theta)}\\
&= \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)}{\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)+n-\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta)}\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{E}(\omega_k|\Theta).
\end{aligned}$$

For $i=1,2,\ldots,g-1$ and subject to $\sum_{i=1}^{g}\pi_i=1$, we have

$$\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial\pi_i'} &= \sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\pi_i'}+\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)\frac{1}{\pi_g'}\frac{\partial\pi_g'}{\partial\pi_i'}\\
&= \sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\pi_i'}-\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)\frac{1}{\pi_g'}=0.
\end{aligned}$$

For each $i=1,2,\ldots,g-1$,the following equation holds,

$$\frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\pi_i'} = \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)}{\pi_g'}$$

which means,

$$\begin{aligned}
\frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{gk}^{(CC)}|\Theta)}{\pi'_g} &= \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{1k}^{(CC)}|\Theta)}{\pi'_1} = \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{2k}^{(CC)}|\Theta)}{\pi'_2} \\
&= \ldots = \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{g-1,k}^{(CC)}|\Theta)}{\pi'_{g-1}} \\
&= \frac{\sum_{h=1}^{g}\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)}{\sum_{h=1}^{g}\pi'_h} \\
&= \sum_{h=1}^{g}\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta),
\end{aligned}$$

then,

$$\hat{\pi'_i} = \frac{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)}{\sum_{k=1}^{n}\sum_{h=1}^{g}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)}.$$

Similarly, for $i = 1, 2, \ldots, g-1, j = 1, 2, \ldots, p$ and subject to $\sum_{i=1}^{g}\rho_{ji} = 1$, we have

$$\begin{aligned}
\frac{\partial Q(\Theta, \Theta')}{\partial \rho'_{ji}} &= \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}|\Theta\big)\frac{1}{\rho'_{ji}} \\
&\quad + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{gjk}^{(CI)}|\Theta\big)\frac{1}{\rho'_{jg}}\frac{\partial \rho'_{jg}}{\partial \rho'_{ji}} \\
&= \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}|\Theta\big)\frac{1}{\rho'_{ji}} \\
&\quad - \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{gjk}^{(CI)}|\Theta\big)\frac{1}{\rho'_{jg}} = 0.
\end{aligned}$$

For each $i = 1, 2, \ldots, g-1$,the following equation holds,

$$\frac{\sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{ijk}^{(CI)}|\Theta\big)}{\rho'_{ji}} = \frac{\sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k)z_{gjk}^{(CI)}|\Theta\big)}{\rho'_{jg}},$$

which means,

$$\begin{aligned}
\frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{gjk}^{(CI)} | \Theta\big)}{\rho'_{jg}} &= \frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{1jk}^{(CI)} | \Theta\big)}{\rho'_{j1}} \\
&= \frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{2jk}^{(CI)} | \Theta\big)}{\rho'_{j2}} \\
&= \ldots = \frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{g-1,jk}^{(CI)} | \Theta\big)}{\rho'_{j,g-1}} \\
&= \frac{\sum_{h=1}^{g} \sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)} | \Theta\big)}{\sum_{h=1}^{g} \rho'_{jh}} \\
&= \sum_{k=1}^{n} \sum_{h=1}^{g} \mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)} | \Theta\big),
\end{aligned}$$

then,

$$\hat{\rho'_{ji}} = \frac{\sum_{k=1}^{n} \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)} | \Theta\big)}{\sum_{k=1}^{n} \sum_{h=1}^{g} \mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)} | \Theta\big)}.$$

$$
\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial \mu'_{jh}} &= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{ji},\sigma^2_{ji}}(x_{k2})}\frac{\partial \phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj})}{\partial \mu_{jh}} \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)\frac{1}{\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj})}\frac{\partial \phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj})}{\partial \mu_{jh}} \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj})}\frac{\partial \phi_{\mu_{ji},\sigma^2_{ji}}(x_{kj})}{\partial \mu_{jh}} \\
&\quad + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)\frac{1}{\phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}\frac{\partial \phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}{\partial \mu_{jh}} \\
&= \sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)\frac{1}{\phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}\frac{\partial \phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}{\partial \mu_{jh}} \\
&\quad + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)\times\frac{1}{\phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}\frac{\partial \phi_{\mu_{jh},\sigma^2_{jh}}(x_{kj})}{\partial \mu_{jh}} \\
&= \sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)\frac{(x_{kj}-\mu'_{jh})}{\sigma'^2_{jh}} \\
&\quad + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)\frac{(x_{kj}-\mu'_{jh})}{\sigma'^2_{jh}} \\
&= [\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)+\sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)]\times\frac{(x_{kj}-\mu'_{jh})}{\sigma'^2_{jh}}=0,
\end{aligned}
$$

then,

$$
\hat{\mu'_{ji}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)+\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CD)}|\Theta\big)]x_{kj}}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta)+\sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)}.
$$

$$
\begin{aligned}
\frac{\partial Q(\Theta,\Theta')}{\partial \sigma'^2_{jh}} &= \sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)(\frac{(x_{kj}-\mu'_{jh})^2}{2\sigma'^4_{jh}}-\frac{1}{2\sigma'^2_{jh}}) \\
&\quad + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)(\frac{(x_{kj}-\mu'_{jh})^2}{2\sigma'^4_{jh}}-\frac{1}{2\sigma'^2_{jh}}) \\
&= [\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{hk}^{(CC)}|\Theta)+\sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{hjk}^{(CI)}|\Theta\big)] \\
&\quad (\frac{(x_{kj}-\mu'_{jh})^2}{2\sigma'^4_{jh}}-\frac{1}{2\sigma'^2_{jh}})=0,
\end{aligned}
$$

then,

$$\hat{\sigma_{ji}^{'2}} = \frac{\sum_{k=1}^{n}[\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)](x_{kj} - \mu'_{ji})^2}{\sum_{k=1}^{n}\mathrm{E}(\omega_k z_{ik}^{(CC)}|\Theta) + \sum_{k=1}^{n}\mathrm{E}\big((1-\omega_k) z_{ijk}^{(CI)}|\Theta\big)}.$$

Chapter 3

Other Approaches for Model Reduction

As noted in Chapter 2, the number of proportion parameters $\pi_{i_1 i_2 \ldots i_p}$ in PCD model increases exponentially since $g^p - 1$ different proportion parameters in total need to be estimated. As a result, the estimation for parameters becomes much more challenging. In Chapter 2, we have described a general PCD model as an approximation of combination of CC model and CI model. Based on this two-level mixture model, the original PCD model distribution density can be approximated. Also the number of parameters under the two-level mixture involves one of proportion parameter of CC model, $g-1$ of proportion parameters in CC model and $p \times (g-1)$ of proportion parameters in CI model, so totally $1 + (g-1) + p(g-1) = pg + g - p$ different proportion parameters need to be estimated. Its parameter space increases linearly with the number of data sets, which is much less than that under the original PCD model as p increases. In the Chapter, we continue to consider other methods for reducing the number of parameters in the PCD model. For the model fitting, we still consider expectation-maximization algorithm. To illustrate the proposed methods, simulation studies are performed and the results are presented for the evaluation of the method.

3.1 Idea from Exchangeable Structure in GEE

Because we are more interested in the diagonal proportion parameters, we want to simplify the non-diagonal $\pi_{i_1 i_2 \ldots i_p} = \alpha$, i_j's are not all the same. Then the $d.f.$ of $\pi_{i_1 i_2 \ldots i_p}$ is g, and we have the following equation:

$$(g^p - g)\alpha + \sum_{i=1}^{g} \pi_i = 1, \qquad (3.1)$$

where π_i is the diagonal elements $\pi_{i_1 i_2 \ldots i_p | i_1 = i_2 = \ldots = i_p = i}$.

3.1.1 Bivariate Normal Mixture Model

We start this approach with $p = 2$. Let $\pi_{ij} = \alpha$, when $i \neq j$. Then the Equation (3.1) can be written as $(g^2 - g)\alpha + \sum_{i=1}^{g} \pi_{ii} = 1$. And the density function is

$$f_{PCD}(x_k, y_k) = \sum_{i=1}^{g} \pi_{ii} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \sigma_i^2}(y_k) + \alpha \sum_{i,j=1, i \neq j}^{g} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \sigma_j^2}(y_k). \qquad (3.2)$$

Based on Section 1.3.2, we need to calculate the following expected values in the E-step:

when $i = j$,

$$\begin{aligned} & \mathrm{E}(z_{ik} w_{ik} | x_k, y_k, \Theta) \\ = & \frac{\pi_{ii} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k)}{\sum_{i=1}^{g} \pi_{ii} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k) + \alpha \sum_{i,j=1, i \neq j}^{g} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}; \end{aligned}$$

when $i \neq j$,

$$\begin{aligned} & \mathrm{E}(z_{ik} w_{jk} | x_k, y_k, \Theta) \\ = & \frac{\alpha \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}{\sum_{i=1}^{g} \pi_{ii} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_i, \tau_i^2}(y_k) + \alpha \sum_{i,j=1, i \neq j}^{g} \phi_{\mu_i, \sigma_i^2}(x_k) \phi_{\nu_j, \tau_j^2}(y_k)}. \end{aligned} \qquad (3.3)$$

To maximize the log-likelihood with respect to π_{ii}, α, μ_i, ν_j, σ_i^2, τ_j^2, we take the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta')|\Theta]$ and obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi'}_{ii} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\alpha'} = \frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{(g^2-g)[\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)]},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)x_k}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)y_k}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(x_k-\mu_i')^2}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\tau_j'^2} = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(y_k-\nu_j')^2}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.4}$$

3.1.2 Multivariate Normal Mixture Model

We further apply this approach to multivariate normal mixture model. The density function is

$$f_{PCD}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i=1}^{g}[\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})] + \alpha \sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \prod_{j=1}^{p} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j}), \tag{3.5}$$

where set $A = \{i_j = 1, 2, \ldots, g, i_j$'s are not identical, for $j = 1, 2, \ldots, p\}$. Based on Section 2.2.1, we need to calculate the following expected values in the E-step:

when i_j's are identical,

$$\begin{aligned} & \mathrm{E}(\prod_{l=1}^{p} z_{lik}|\Theta) \\ = & \frac{\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})}{\sum_{i=1}^{g}[\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})] + \alpha \sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \prod_{j=1}^{p} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j})}; \end{aligned}$$

when i_j's are not identical,

$$\begin{aligned} & \mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta) \\ = & \frac{\alpha \prod_{j=1}^{p} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j})}{\sum_{i=1}^{g}[\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})] + \alpha \sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \prod_{j=1}^{p} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j})}. \end{aligned} \quad (3.6)$$

To maximize the log-likelihood with respect to π, α, μ, σ, we take the partial derivatives of $Q(\Theta, \Theta') = \mathrm{E}_{\mathbf{z}}[\log L_{PCD}(\mathbf{x}, \mathbf{z}|\Theta')|\Theta]$. Let $\hat{u}_{i_1 i_2 \ldots i_p k} = \mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)$, and $\hat{u}_{ik} = \hat{u}_{i_1 i_2 \ldots i_p k | i_1 = i_2 = \ldots = i_p = i}$.we can obtain the following maximum likelihood estimation of the parameters in the M-step:

$$\hat{\pi}_i' = \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}},$$

$$\hat{\alpha}' = \frac{\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{(g^p - g)[\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}]},$$

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g} \hat{u}_{i_1 i_2 \ldots i_p k} x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g} \hat{u}_{i_1 i_2 \ldots i_p k}},$$

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g} \hat{u}_{i_1 i_2 \ldots i_p k}(x_{kj} - \mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g} \hat{u}_{i_1 i_2 \ldots i_p k}}. \quad (3.7)$$

The mathematical derivations can be found in Section 3.5.

3.2 Idea from AR(1) Structure in GEE

3.2.1 Bivariate Normal Mixture Model

For the bivariate normal mixture model, we want to simplify the non-diagonal $\pi_{ij} = \alpha^{|j-i|}$, when $i \neq j$, and we have the following equation:

$$\sum_{i,j=1,i\neq j}^{g} \alpha^{|j-i|} + \sum_{i=1}^{g} \pi_{ii} = 1, \tag{3.8}$$

And the density function is

$$f_{PCD}(x_k, y_k) = \sum_{i=1}^{g} \pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\sigma_i^2}(y_k) + \sum_{i,j=1,i\neq j}^{g} \alpha^{|j-i|}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\sigma_j^2}(y_k). \tag{3.9}$$

E-step:

when $i = j$,

$$\begin{aligned}&\mathrm{E}(z_{ik}w_{ik}|x_k, y_k, \Theta)\\ =\ & \frac{\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k) + \alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)};\end{aligned}$$

when $i \neq j$,

$$\begin{aligned}&\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta)\\ =\ & \frac{\alpha\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k) + \alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.\end{aligned} \tag{3.10}$$

M-step: as discussed in Section 1.2, we usually consider a three-component mixure model ($g = 3$) since a gene has three kinds of expression in experiments: up-regulated, down-regulated or no differentially expressed.

When $g = 3$, let $f(\alpha) = \sum_{i,j=1,i\neq j}^{g} \alpha^{|j-i|} = 4\alpha + 2\alpha^2$.

Let $C_1 = \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{3} |j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta), C_2 = \sum_{k=1}^{n}\sum_{i=1}^{3}\mathrm{E}(z_{ik}w_{ik}|\Theta)$. The reasonable solution is

$$\hat{\alpha}' = \frac{-(2C_1 + 2C_2) + \sqrt{C_1^2 + C_1 + 5C_2^2}}{2C_1 + 4C_2}. \tag{3.11}$$

The mathematical derivations can be found in Section 3.5.

3.2.2 Multivariate Normal Mixture Model

We have simplified $\pi_{ij} = \alpha^{|j-i|}$ for model reduction when $p = 2$. As the number of data sets p increases, how to define a reasonable function $\pi_{i_1 i_2 \ldots i_p}$ with respect to α will be an issue. Although there are many ways to map α to $\pi_{i_1 i_2 \ldots i_p}$, we still want to simplify $\pi_{i_1 i_2 \ldots i_p}$ as a power function of α as we have done for bivariate normal mixture model. For example, if $p = 3$, when i_1, i_2 and i_3 are not identical, we try to simplify $\pi_{i_1 i_2 i_3}$=$\alpha^{|i_1 - i_2| + |i_2 - i_3| + |i_3 - i_1|}$. For a three-component mixure model, subject to $\sum_{i_1, i_2, i_3 = 1}^{3} \pi_{i_1 i_2 i_3} = 1$, let $A = \{i_j = 1,\ 2,\ 3,\ i_j\text{'s are not identical, for } j = 1,\ 2,\ 3\}$, we have the following equation:

$$\sum_{\{i_1, i_2, i_3\} \subset A} + \sum_{i=1}^{3} \pi_{iii} = 1. \tag{3.12}$$

Let $h(i_1, i_2, i_3) = |i_1 - i_2| + |i_2 - i_3| + |i_3 - i_1|$. Since $i_j = 1$, 2 or 3, $h(i_1, i_2, i_3)$ is some integer within interval [1, 6]. Thus, the left side of Equation (3.12) is a polynomial of degree 4 with respect to α. We consider the Equation (3.12) and other possible equations obtained from M-step as a high-order equation system, it is very challenging to derive the expressions of general solution for α and π_{iii}. In the future, we will try to solve this problem by programming. Also, if the dimension extends to a general dimension p, the functions which associate $\pi_{i_1 i_2 \ldots i_p}$ with α will be more complex to be defined. Moreover, because the order of the equation system will be relatively high as p goes up, to solve the equations will be much more complicated and difficult. We are going to explore such issues further for our future research.

3.3 Idea from Multiset Coefficient in Combinatorics

3.3.1 Bivariate Normal Mixture Model

Our goal is still to simplify the non-diagonal proportion parameter π_{ij}. We assume those π_{ij}'s with the same combination of index i and j have the same value. For example, when $g = 3$, let $\pi_{12} = \pi_{21} = \alpha$, $\pi_{23} = \pi_{32} = \beta$, $\pi_{13} = \pi_{31} = \gamma$, and we have the following equation:

$$2(\alpha + \beta + \gamma) + \sum_{i=1}^{g} \pi_{ii} = 1. \tag{3.13}$$

E-step:

when $i = j$,

$$\mathrm{E}(z_{ik}w_{ik}|x_k, y_k, \Theta) = \frac{\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{f_{PCD}(x_k, y_k)};$$

when $i, j \in \{1, 2\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\alpha\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)};$$

when $i, j \in \{2, 3\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\beta\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)};$$

when $i, j \in \{1, 3\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\gamma\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)}.$$

M-step:

$$\hat{\pi'}_{ii} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\alpha'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)]},$$

$$\hat{\beta'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)]},$$

$$\hat{\gamma'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)]},$$

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)x_k}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(x_k - \mu_i')^2}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)y_k}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)},$$

$$\hat{\tau_j'^2} = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(y_k - \nu_j')^2}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.14}$$

The mathematical derivations can be found in Section 3.5.

3.3.2 Multivariate Normal Mixture Model

Let set $A = \{i_j = 1, 2, \ldots, g, i_j$'s are not identical, for $j = 1, 2, \ldots, p\}$. We want to divide A into several groups, and in each group the numbers of occurrences of i_j are exactly the same. Let d_i be the number of occurrences of component i, the possible values of d_i are $0, 1, \ldots, p$. Then the number of groups is equal to the number of non-negative integer solutions to equation $d_1 + d_2 + \ldots + d_g = p$, then subtract g (All $i_j's$ are identical). This is a classic problem of combinations with repetition. The number of non-negative integer solutions is $\binom{p+g-1}{p}$, which is called the multiset coefficient in combinatorics.

For general p and g, let r_t be the number of elements in group A_t and d_{ti} be the number of repetition of component i. For group A_t, r_t is the number of ways that we choose d_{t1} of 1, d_{t2} of 2, ..., and $p - \sum_{i=1}^{g-1} d_{ti}$ of g. So the multinomial coefficient $r_t = \frac{p!}{\prod_{i=1}^{g} d_{ti}!}$, where d_{ti} is the number of repetition of component i in any element of this group. The restriction equation is

$$\sum_{t=1}^{\binom{p+g-1}{p}-g} r_t\alpha_t + \sum_{i=1}^{g} \pi_i = 1.$$

The density function is

$$f_{PCD}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \sum_{i=1}^{g}[\pi_i \prod_{j=1}^{p} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})] + \sum_{\{i_1,i_2,\ldots,i_p \subset A_t\}} \alpha_t \prod_{j=1}^{p} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j}). \tag{3.15}$$

E-step:

when i_j's are identical,

$$\mathrm{E}(\prod_{l=1}^{p} z_{lik}|\Theta) = \frac{\pi_i \prod_{j=1}^{g} \phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})}{f_{PCD}(x_1, x_2, \ldots, x_p)};$$

when $\{i_1, i_2, \ldots, i_p\} \subset A_t$,

$$\mathrm{E}(\prod_{l=1}^{g} z_{li_lk}|\Theta) = \frac{\alpha_t \prod_{j=1}^{g} \phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j})}{f_{PCD}(x_1, x_2, \ldots, x_p)}.$$

M-step:

$$\hat{\pi}_i' = \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}},$$

$$\hat{\alpha}_t' = \frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p \subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{r_t[\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p \subset A\}} \hat{u}_{i_1\ldots i_p k}]}. \tag{3.16}$$

The mathematical derivations can be found in Section 3.5.

3.3.3 Implementation of Computer Programming

The most problem of programming implemetation is the partition of $\pi_{i_1 \ldots i_p}$. As described, if all the groups A_t's and all the elements in A_t are known, then related variables can be calculated like r_t and corresponding variables can be obtained like $\hat{u}_{i_1 \ldots i_p k}$. For one $\pi_{i_1 \ldots i_p}$, we define g counter variables $d_i^{(i_1 \ldots i_p)}$, $i = 1, 2, \ldots, g$, to calculate the number of repetition of component i in the subscript of $\pi_{i_1 i_2 \ldots i_p}$. Let A_1 be a vector with length $\frac{p!}{\prod_{i=1}^g d_i^{(i_1 \ldots i_p)}!}$, we store $\pi_{i_1 i_2 \ldots i_p}$ in the set A_1.

Given a second $\pi_{i'_1 \ldots i'_p}$, we can also have g counter variables $d_i^{(i'_1 \ldots i'_p)}$, $i = 1, 2, \ldots, g$. Then we compare each $d_i^{(i'_1 \ldots i'_p)}$ with the corresponding $d_i^{(i_1 \ldots i_p)}$. If $d_i^{(i'_1 \ldots i'_p)}$=$d_i^{(i_1 \ldots i_p)}$, for all i's, then we can say that $\pi_{i'_1 \ldots i'_p}$ and $\pi_{i_1 \ldots i_p}$ are in the same set A_1. Otherwise, we generate a new vector A_2 with length $\frac{p!}{\prod_{i=1}^g d_i^{(i'_1 \ldots i'_p)}!}$, and store $\pi_{i'_1 \ldots i'_p}$ in the set A_2.

Given a third $\pi_{i''_1 \ldots i''_p}$, we compare each $d_i^{(i''_1 \ldots i''_p)}$ with both of $d_i^{(i_1 \ldots i_p)}$ and $d_i^{(i'_1 \ldots i'_p)}$. If $d_i^{(i''_1 \ldots i''_p)}$'s are the same as $d_i^{(i_1 \ldots i_p)}$'s, then we store $\pi_{i''_1 \ldots i''_p}$ in the set A_1; if $d_i^{(i''_1 \ldots i''_p)}$'s are the same as $d_i^{(i'_1 \ldots i'_p)}$'s, then we store $\pi_{i''_1 \ldots i''_p}$ in the set A_2; otherwise, we generate a new vector A_3 with length $\frac{p!}{\prod_{i=1}^g d_i^{(i''_1 \ldots i''_p)}!}$, and store $\pi_{i''_1 \ldots i''_p}$ in the set A_3.

And so forth, we can partition $A = \{\pi_{i_1 \ldots i_p}, i_j = 1, 2, \ldots, g, i_j$'s are not identical, for $j = 1, 2, \ldots, p\}$ into different $\binom{p+g-1}{p} - g$ sets A_t's, and in each set A_t the numbers of occurrences of subscript i_j are exactly the same. Since $r_t = \frac{p!}{\prod_{i=1}^g d_{ti}!}$, where d_{ti} is the number of repetition of component i in the subscript in set A_t, α_t's can be estimated by plugging-in all the corresponding parameters.

We may also apply recursion method to deal with the problem of nested loops with unknown number of iterations when programming for this multivariate normal mixture model.

3.4 Simulation Studies

In this section, we conduct simulation studies to illustrate approaches based on both exchangeable structure and multi-combination for model reduction. As mentioned in Section 2.3, there are two types for the model parameter configurations - moderate and difficult. We follow the similar simulation procedure as what we have done based on the multilevel mixture model: initializing z values, implementing EM algorithm, and saving the estimation results.

3.4.1 Exchangeable Structure Model

Case 1: we consider a moderate case without repetitions when there is a restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1, \sigma^2_{32} = 1$. We generate $N = 1000$ 3-dimensional normal data based on our model. The estimated results are shown in Table 3.1, and we can see that some of them are in an acceptable range but not very close to the initial values.

Table 3.1: Parameters estimation in a moderate PCD model based on exchangeable structure with restriction when $B = 1$, $N = 1000$, $p = 3$

Parameter	True Value	Estimate by EM
α	0.025	0.024
$\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)$	(0.1, 0.2, 0.1)	(0.120, 0.192, 0.107)
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.043, 0, 2.078)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.834, 0, 1.976)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-2.051, 0, 2.172)
$\boldsymbol{\sigma_1^2} = (\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.710, 1, 1.440)
$\boldsymbol{\sigma_2^2} = (\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.611, 1, 1.193)
$\boldsymbol{\sigma_3^2} = (\sigma^2_{31}, \sigma^2_{32}, \sigma^2_{33})$	(1.5, 1, 1.5)	(1.531, 1, 1.130)

For a comprehensive performance evaluations, we conduct simulation studies for multivariate mixture model to understand the estimation performance of the exchangeable structure model. For a reasonable parameter configuration, we repeat the estimation for 1000 times, save all these estimation results, and then generate boxplots of estimated parameters.

Case 2: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. In this case, the number of non-convergence is 0. Table 3.2 shows the summary, and Figure 3.1 and 3.2 display the box plots for each parameter. In Figure 3.1, on the left, the proportion parameters π_1, π_2 and π_3 in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.25 (y-axis); on the right,the proportion parameter α is visualized with vertical scale from 0.020 to 0.028. In Figure 3.2, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2. Comparing with the estimation in Case 1, the results are much more closer to the initial values with relatively small variances.

Table 3.2: Parameters estimation in a moderate PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$

Parameter	True Value	Mean	Standard Deviation
π_1	0.1	0.102	0.018
π_2	0.2	0.201	0.027
π_3	0.1	0.100	0.017
α	0.025	0.025	0.001
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-1.996, 0, 2.003)	(0.133, 0, 0.125)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.996, 0, 2.004)	(0.132, 0, 0.131)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-1.997, 0, 2.006)	(0.132, 0, 0.131)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.511, 1, 1.497)	(0.212, 0, 0.205)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.513, 1, 1.499)	(0.214, 0, 0.205)
$\boldsymbol{\sigma_3^2} = (\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.508, 1, 1.483)	(0.207, 0, 0.201)

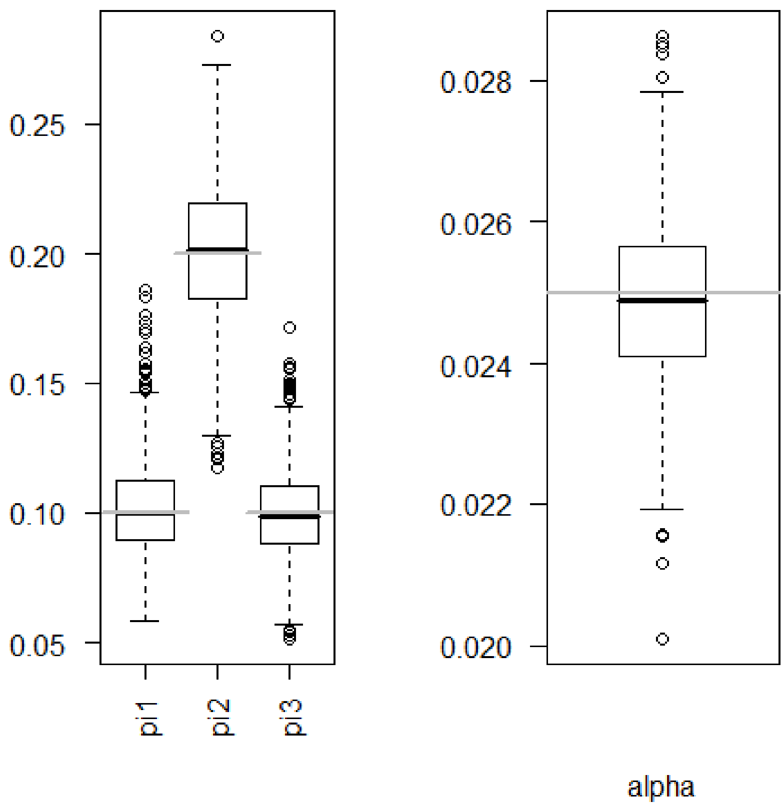


Figure 3.1: Boxplots of estimated proportion parameters in a moderate PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. On the left, the proportion parameters π_1, π_2 and π_3 in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.25 (y-axis); on the right,the proportion parameter α is visualized with vertical scale from 0.020 to 0.028.

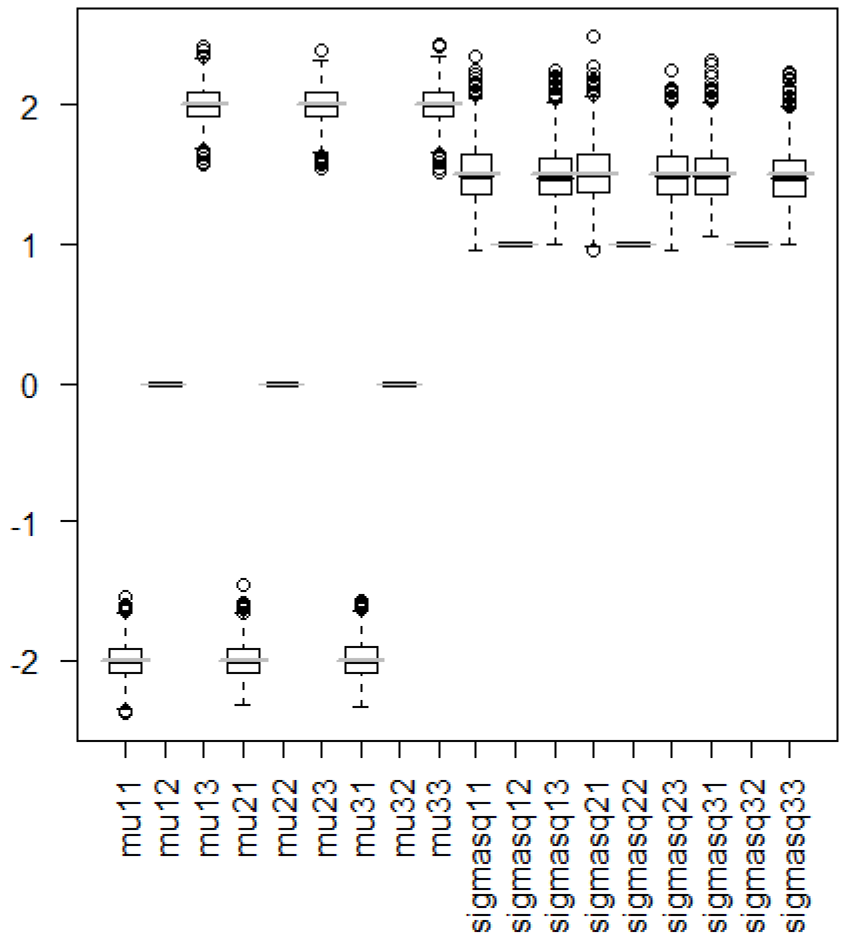


Figure 3.2: Boxplots of estimated means and variances in a moderate PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2.

Case 3: we consider a difficult case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma^2_{12} = 1, \sigma^2_{22} = 1, \sigma^2_{32} = 1$. In this case, the number of non-convergence is 4. Table 3.3 shows the summary, and Figure 3.3 and 3.4 display the box plots for each parameter. In Figure 3.3, on the left, the proportion parameters π_1, π_2 and π_3 in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.6 (y-axis); on the right, the proportion parameter α in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.02. In Figure 3.4, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma^2_1}$, $\boldsymbol{\sigma^2_2}$, $\boldsymbol{\sigma^2_3}$ in horizontal sequence are with vertical scale from -2 to 8. Comparing with the estimation in Case 2, the results are not far from to the initial values.

Table 3.3: Parameters estimation in a difficult PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$

Parameter	True Value	Mean	Standard Deviation
π_1	0.02	0.029	0.047
π_2	0.6	0.602	0.047
π_3	0.02	0.028	0.049
α	0.015	0.014	0.003
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.479, 0, 1.487)	(0.349, 0, 0.363)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.492, 0, 1.486)	(0.358, 0, 0.364)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-1.5, 0, 1.5)	(-1.466, 0, 1.483)	(0.352, 0, 0.359)
$\boldsymbol{\sigma^2_1} = (\sigma^2_{11}, \sigma^2_{12}, \sigma^2_{13})$	(1.5, 1, 1.5)	(1.534, 1, 1.532)	(0.435, 0, 0.452)
$\boldsymbol{\sigma^2_2} = (\sigma^2_{21}, \sigma^2_{22}, \sigma^2_{23})$	(1.5, 1, 1.5)	(1.515, 1, 1.522)	(0.487, 0, 0.457)
$\boldsymbol{\sigma^2_3} = (\sigma^2_{31}, \sigma^2_{32}, \sigma^2_{33})$	(1.5, 1, 1.5)	(1.527, 1, 1.511)	(0.434, 0, 0.438)

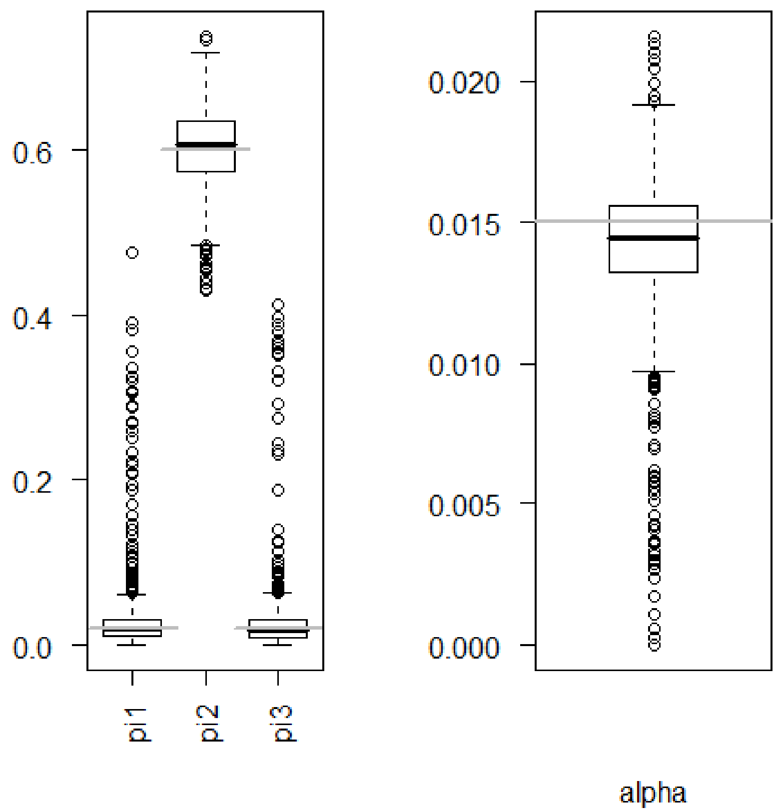


Figure 3.3: Boxplots of estimated proportion parameters in a difficult PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$.on the left, the proportion parameters π_1, π_2 and π_3 in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.6 (y-axis); on the right, the proportion parameter α in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.02.

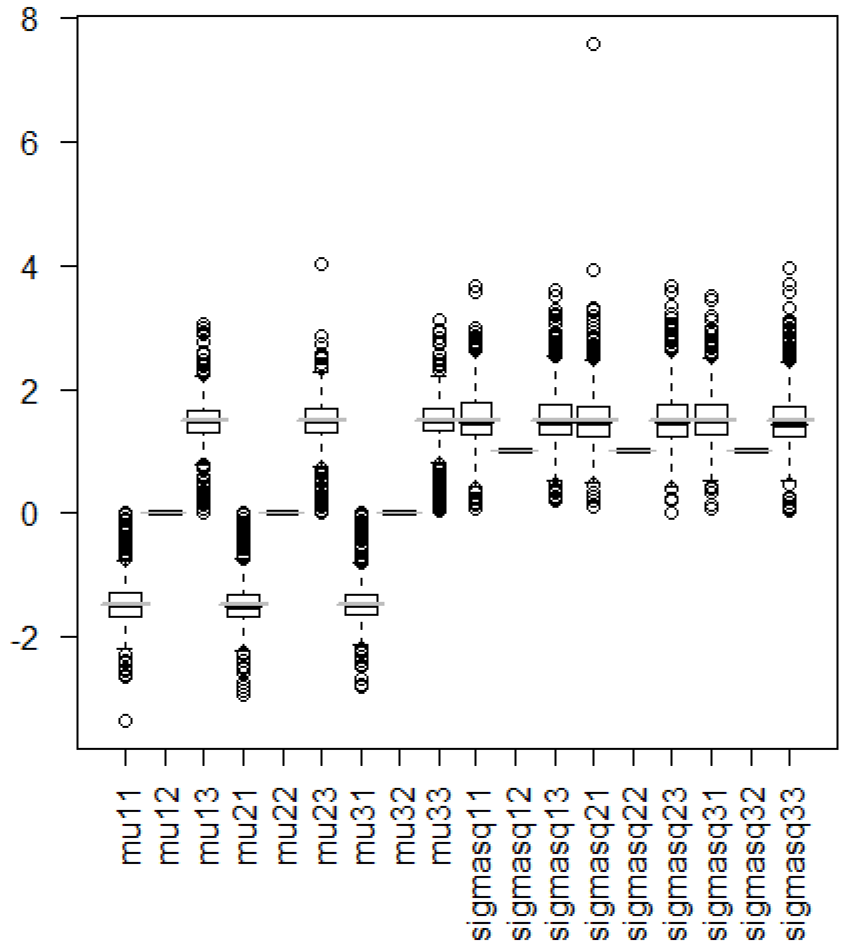


Figure 3.4: Boxplots of estimated means and variances in a difficult PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 8.

3.4.2 Multiset Coefficient Model

Case 1: we consider a moderate case without repetitions when there is a restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. We generate $N = 1000$ 3-dimensional normal data based on our model. The estimated results are shown in Table 3.4, and we can see that some of them are in an acceptable range but not very close to the initial values.

Table 3.4: Parameters estimation in a moderate PCD model based on combination structure with restriction when $B = 1$, $N = 1000$, $p = 3$

Parameter	True Value	Estimate by EM
π_1	0.1	0.102
π_2	0.2	0.220
π_3	0.1	0.113
α_1	0.025	0.025
α_2	0.02	0.015
α_3	0.035	0.033
α_4	0.03	0.029
α_5	0.025	0.031
α_6	0.015	0.007
α_7	0.02	0.020
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.157, 0, 2.040)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-1.915, 0, 2.005)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-2.177, 0, 2.113)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.562, 1, 1.486)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.485, 1, 1.467)
$\boldsymbol{\sigma_3^2} = (\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.485, 1, 1.189)

Case 2: we consider a moderate case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. In this case, the number of non-convergence is 0. Table 3.5 shows the summary, and Figure 3.5 and 3.6 display the box plots for each parameter. In Figure 3.5, on the left, the proportion parameters $\boldsymbol{\pi}$'s and in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.3 (y-axis); on the right, the proportion parameters $\boldsymbol{\alpha}$'s in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.08 (y-axis). In Figure 3.6, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$,

$\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2. Comparing with the estimation in Case 1, the results are much more closer to the initial values with relatively small variances.

Table 3.5: Parameters estimation in a moderate PCD model based on combination structure with restriction when $B = 1000$, $N = 1000$, $p = 3$

Parameter	True Value	Mean	Standard Deviation
π_1	0.1	0.102	0.028
π_2	0.2	0.203	0.030
π_3	0.1	0.100	0.026
α_1	0.025	0.024	0.011
α_2	0.02	0.020	0.008
α_3	0.035	0.036	0.014
α_4	0.03	0.029	0.006
α_5	0.025	0.025	0.008
α_6	0.015	0.016	0.011
α_7	0.02	0.019	0.011
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-2, 0, 2)	(-2.013, 0, 2.012)	(0.165, 0, 0.168)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-2, 0, 2)	(-2.010, 0, 2.018)	(0.163, 0, 0.161)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-2, 0, 2)	(-2.007, 0, 2.018)	(0.165, 0, 0.164)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.491, 1, 1.487)	(0.236, 0, 0.241)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.494, 1, 1.486)	(0.235, 0, 0.226)
$\boldsymbol{\sigma_3^2} = (\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.487, 1, 1.469)	(0.230, 0, 0.236)

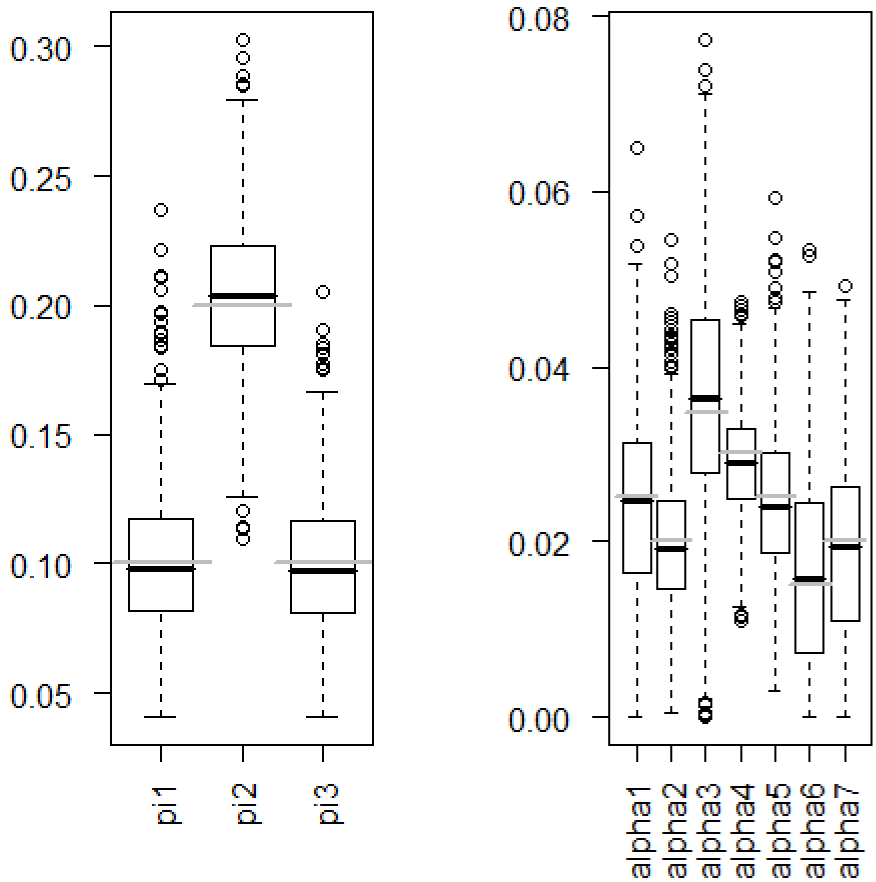


Figure 3.5: Boxplots of estimated proportion parameters in a moderate PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. On the left, the proportion parameters π's and in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.3 (y-axis); on the right, the proportion parameters α's in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.08 (y-axis).

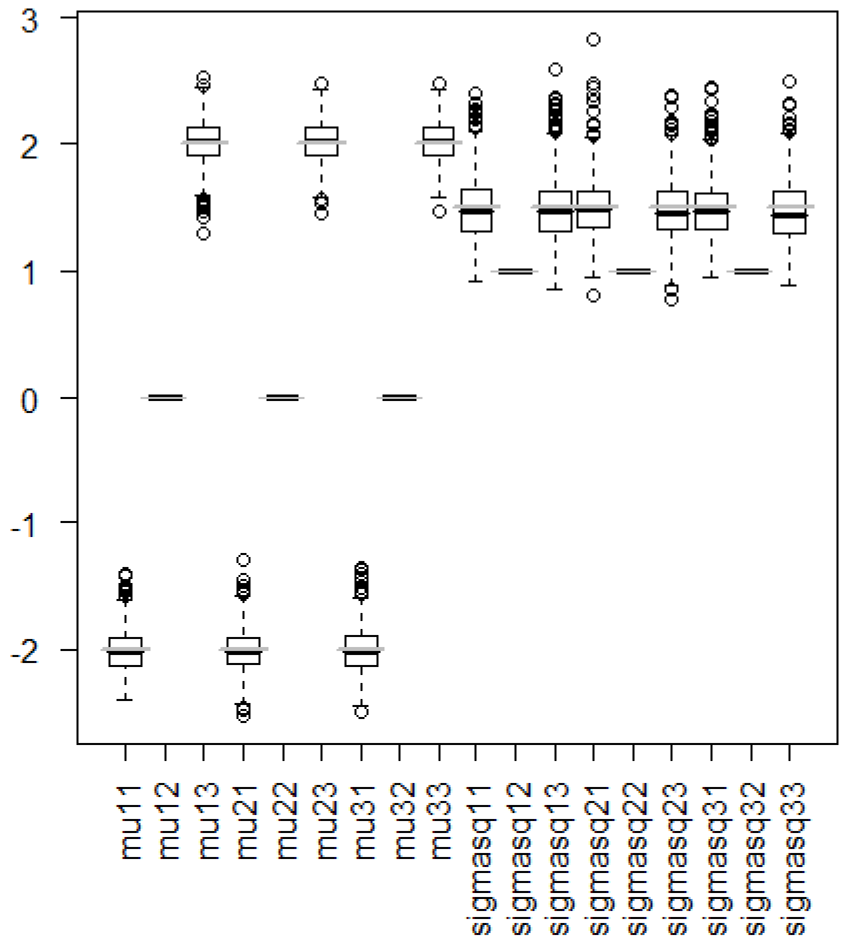


Figure 3.6: Boxplots of estimated means and variances in a moderate PCD model based on exchangeable structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 2.

Case 3: we consider a difficult case with restriction on $\mu_{12} = 0, \mu_{22} = 0, \mu_{32} = 0, \sigma_{12}^2 = 1, \sigma_{22}^2 = 1, \sigma_{32}^2 = 1$. Table 3.6 shows the summary, and Figure 3.7 and 3.8 display the box plots for each parameter. In Figure 3.4, on the left, the proportion parameters π's and in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.7 (y-axis); on the right, the proportion parameters α's in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.15. In Figure 3.8, the means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 4. Comparing with the estimation in Case 2, the results are not very close to the initial values with relatively large variances.

Table 3.6: Parameters estimation in a difficult PCD model based on combination structure with restriction when $B = 1000$, $N = 1000$, $p = 3$

Parameter	True Value	Mean	Standard Deviation
π_1	0.025	0.027	0.043
π_2	0.5	0.522	0.061
π_3	0.025	0.030	0.055
α_1	0.015	0.015	0.013
α_2	0.02	0.016	0.012
α_3	0.025	0.027	0.017
α_4	0.015	0.013	0.008
α_5	0.015	0.011	0.011
α_6	0.025	0.028	0.018
α_7	0.02	0.018	0.013
$\boldsymbol{\mu_1} = (\mu_{11}, \mu_{12}, \mu_{13})$	(-1.5, 0, 1.5)	(-1.629, 0, 1.625)	(0.420, 0, 0.423)
$\boldsymbol{\mu_2} = (\mu_{21}, \mu_{22}, \mu_{23})$	(-1.5, 0, 1.5)	(-1.642, 0, 1.616)	(0.432, 0, 0.432)
$\boldsymbol{\mu_3} = (\mu_{31}, \mu_{32}, \mu_{33})$	(-1.5, 0, 1.5)	(-1.625, 0, 1.627)	(0.428, 0, 0.424)
$\boldsymbol{\sigma_1^2} = (\sigma_{11}^2, \sigma_{12}^2, \sigma_{13}^2)$	(1.5, 1, 1.5)	(1.409, 1, 1.405)	(0.465, 0, 0.470)
$\boldsymbol{\sigma_2^2} = (\sigma_{21}^2, \sigma_{22}^2, \sigma_{23}^2)$	(1.5, 1, 1.5)	(1.397, 1, 1.407)	(0.473, 0, 0.469)
$\boldsymbol{\sigma_3^2} = (\sigma_{31}^2, \sigma_{32}^2, \sigma_{33}^2)$	(1.5, 1, 1.5)	(1.397, 1, 1.400)	(0.486, 0, 0.458)

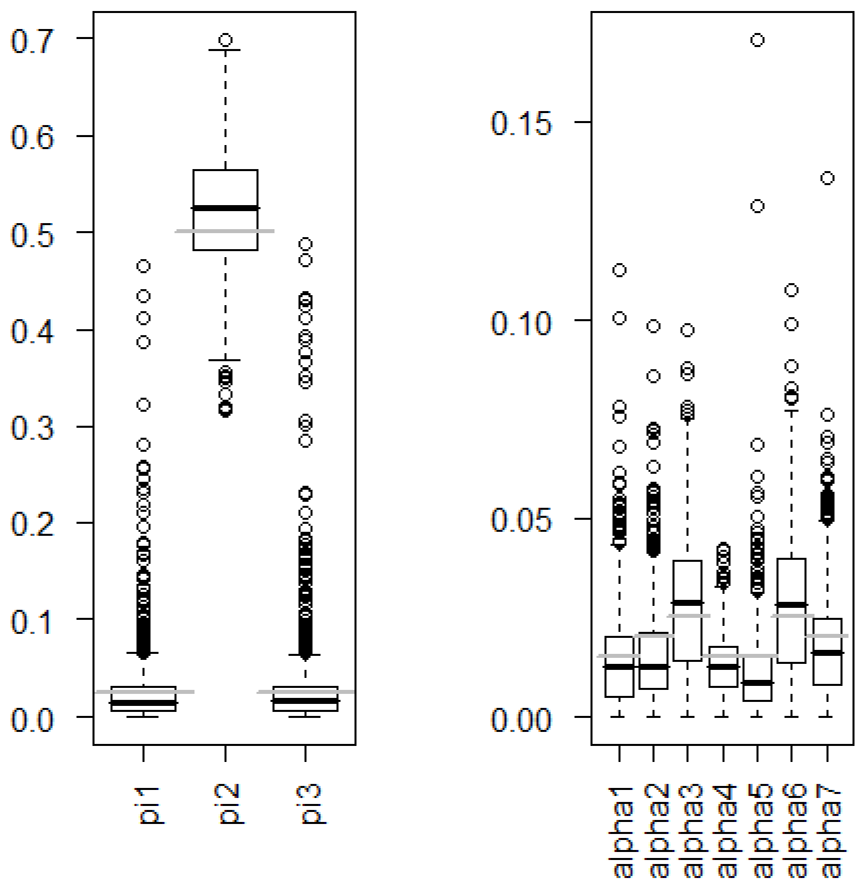


Figure 3.7: Boxplots of estimated proportion parameters in a difficult PCD model based on group structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. On the left, the proportion parameters π's and in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.7 (y-axis); on the right, the proportion parameters α's in horizontal sequence (x-axis) are visualized with vertical scale from 0 to 0.15.

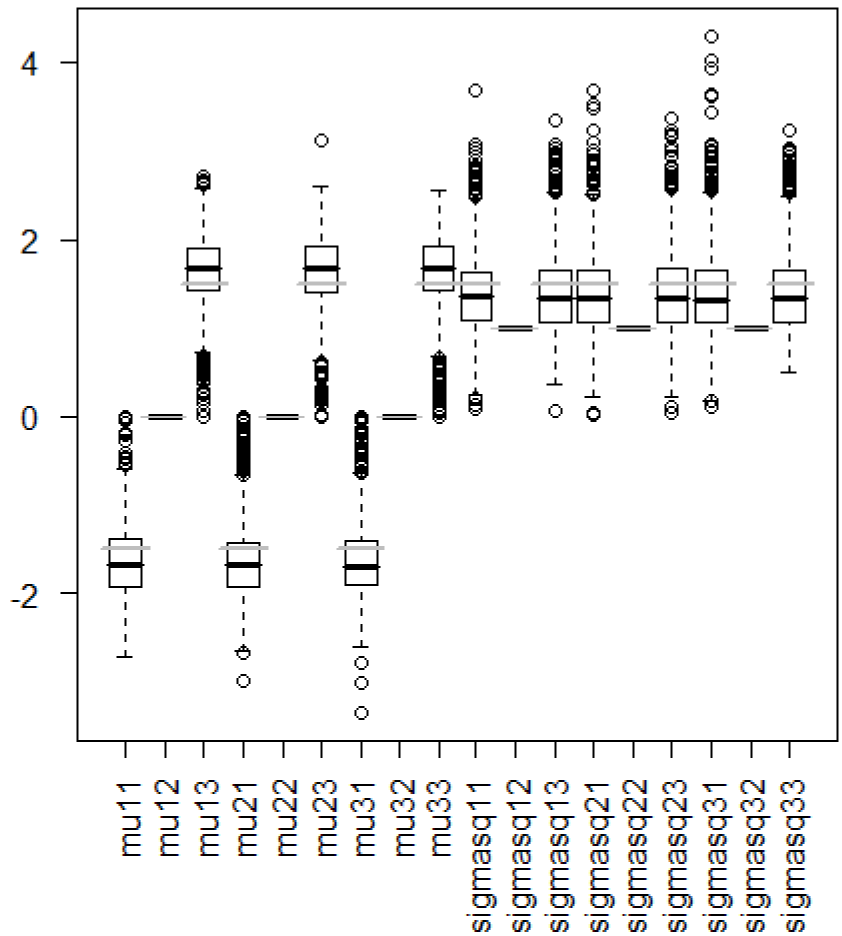


Figure 3.8: Boxplots of estimated means and variances in a difficult PCD model based on group structure with restriction when $B = 1000$, $N = 1000$, $p = 3$. The means $\boldsymbol{\mu_1}$, $\boldsymbol{\mu_2}$, $\boldsymbol{\mu_3}$ and variances $\boldsymbol{\sigma_1^2}$, $\boldsymbol{\sigma_2^2}$, $\boldsymbol{\sigma_3^2}$ in horizontal sequence are with vertical scale from -2 to 4.

3.5 Mathematical Derivations

Formula (3.3) and (3.4) derivations for E and M steps in Section 3.1.1

Based on Section 1.3.2, we rewrite the function $Q(\Theta, \Theta')$.

$$\begin{aligned}
Q(\Theta, \Theta') &= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} z_{ik}w_{jk}\log(\pi'_{ij}\phi_{\mu'_i,\sigma'^2_j}(x_k)\phi_{\nu'_j,\tau'^2_j}(y_k))|\Theta] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\pi'_{ij}) \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k)) + \log(\phi_{\nu'_j,\tau'^2_j}(y_k))] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)\log(\pi'_{ii}) \\
&\quad + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\alpha') \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k)) + \log(\phi_{\nu'_j,\tau'^2_j}(y_k))]
\end{aligned}$$

E-step:

When $i = j$,

$$\mathrm{E}(z_{ik}w_{ik}|x_k, y_k, \Theta) = \frac{\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k) + \alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.$$

When $i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\alpha\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k) + \alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.$$

M-step: The optimization of $\pi_{ii}, \alpha, \mu_i, \sigma_i^2, \nu_j$, and τ_j^2 is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta,\Theta')}{\partial \pi'_{ii}} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \pi'_{ii}} = 0, \tag{3.17}$$

for $i = 1, 2, \ldots, g-1$ the above equation holds.

$$\frac{\partial Q(\Theta,\Theta')}{\partial \alpha'} = \frac{\sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\alpha'} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \alpha'} = 0. \tag{3.18}$$

Subject to $(g^2-g)\alpha + \sum_{i=1}^{g} \pi_{ii} = 1$, $\frac{\partial \pi'_{gg}}{\partial \pi'_{ii}} = -1$, and $\frac{\partial \pi'_{gg}}{\partial \alpha'} = -(g^2-g)$. From Equations (3.17) and (3.18), we have the following equations:

$$\begin{aligned} \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} &= \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \qquad (3.19) \\ &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii}}, \end{aligned}$$

$$\frac{\sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{(g^2-g)\alpha'} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}. \tag{3.20}$$

Then

$$\begin{aligned} \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii}} \\ &= \frac{\sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{(g^2-g)\alpha'} \\ &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii} + (g^2-g)\alpha'} \\ &= \sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta). \end{aligned}$$

From (3.19) and (3.20), we have

$$\frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} = \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta),$$

$$\frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{(g^2-g)\alpha} = \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta).$$

So

$$\hat{\pi'}_{ii} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.21}$$

$$\hat{\alpha'} = \frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{(g^2-g)[\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)]}. \tag{3.22}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \mu'_i} = \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)\frac{x_k-\mu'_i}{\sigma_i'^2} = 0,$$

then,

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)x_k}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.23}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \sigma_i'^2} = \sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)[\frac{(x_k-\mu'_i)^2}{2\sigma_i'^4} - \frac{1}{2\sigma_i'^2}] = 0,$$

then,

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(x_k-\mu'_i)^2}{\sum_{k=1}^{n}\sum_{j=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.24}$$

By symmetry,

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)y_k}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}, \tag{3.25}$$

$$\hat{\tau}_j'^2 = \frac{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk} | \Theta)(y_k - \nu_j')^2}{\sum_{k=1}^{n} \sum_{i=1}^{g} \mathrm{E}(z_{ik} w_{jk} | \Theta)}. \tag{3.26}$$

Formula (3.6) and (3.7) derivations for E and M steps in Section 3.1.1

Based on Section 2.2.1, let set $A = \{i_j = 1, 2, \ldots, g, i_j$'s are not identical, for $j = 1, 2, \ldots, p\}$, we rewrite the function $Q(\Theta, \Theta')$.

$$
\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x},\mathbf{z}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj}))|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})|\Theta)] \\
&= \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})) \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{lik}|\Theta)\log(\pi'_i) \\
&\quad +\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\alpha') \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})).
\end{aligned}
$$

E-step:

When i_j's are identical,

$$
\mathrm{E}(\prod_{l=1}^{p} z_{lik}^{p}|\Theta) = \frac{\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma^2_{ji}}(x_{k_j})}{\sum_{i=1}^{g}[\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma^2_{ji}}(x_{k_j})]+\alpha\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{k_j})}.
$$

When i_j's are not identical,

$$
\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta) = \frac{\alpha\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{k_j})}{\sum_{i=1}^{g}[\pi_i\prod_{j=1}^{p}\phi_{\mu_{ji},\sigma^2_{ji}}(x_{k_j})]+\alpha\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\prod_{j=1}^{p}\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{k_j})}.
$$

M-step: Let $\hat{u}_{i_1 i_2 \ldots i_p k} = \mathrm{E}(\prod_{l=1}^{p} z_{i_l k}|\Theta)$, and $\hat{u}_{ik} = \hat{u}_{i_1 i_2 \ldots i_p k | i_1 = i_2 = \ldots = i_p = i}$, the optimization of π_i, α, $\mu_{j i_j}$, $\sigma^2_{j i_j}$ is simply a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta, \Theta')}{\partial \pi'_i} = \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\pi'_i} + \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} \frac{\partial \pi'_g}{\partial \pi'_i} = 0, \tag{3.27}$$

for $i = 1, 2, \ldots, g-1$ the above equation holds.

$$\frac{\partial Q(\Theta, \Theta')}{\partial \alpha} = \frac{\sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{\alpha'} + \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} \frac{\partial \pi'_g}{\partial \alpha'} = 0. \tag{3.28}$$

Subject to $(g^p - g)\alpha + \sum_{i=1}^{g} \pi_i = 1$, $\frac{\partial \pi'_g}{\partial \pi'_i} = -1$, and $\frac{\partial \pi'_g}{\partial \alpha'} = -(g^p - g)$. From Equations (3.27) and (3.28), we have the following equations:

$$\begin{aligned} \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\pi'_i} &= \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} \\ &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{i=1}^{g} \pi'_i}, \end{aligned} \tag{3.29}$$

$$\frac{\sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{(g^p - g)\alpha'} = \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g}. \tag{3.30}$$

Then

$$\begin{aligned} \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{i=1}^{g} \pi'_i} \\ &= \frac{\sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{(g^p - g)\alpha'} \\ &= \frac{\sum_{i=1}^{g} \sum_{k=1}^{n} \hat{u}_{ik} + \sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{\sum_{i=1}^{g} \pi'_i + (g^p - g)\alpha'} \\ &= \sum_{k=1}^{n} \sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}. \end{aligned}$$

From (3.29) and (3.30), we have

$$\frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\pi'_i} = \sum_{k=1}^{n} \sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A\}} \hat{u}_{i_1 i_2 \ldots i_p k}$$

$$\frac{\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1i_2\ldots i_pk}}{(g^p-g)\alpha'} = \sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1i_2\ldots i_pk}.$$

So

$$\hat{\pi_i'} = \frac{\sum_{k=1}^{n}\hat{u}_{ik}}{\sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1i_2\ldots i_pk}}, \tag{3.31}$$

$$\hat{\alpha'} = \frac{\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1i_2\ldots i_pk}}{(g^p-g)[\sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1i_2\ldots i_pk}]}. \tag{3.32}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial\mu'_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}\frac{x_{kj}-\mu'_{ji_j}}{\sigma'^2_{ji_j}} = 0,$$

then,

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}},$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial\sigma'^2_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}[\frac{(x_{kj}-\mu'_{ji_j})^2}{2\sigma'^4_{ji_j}}-\frac{1}{2\sigma'^2_{ji_j}}] = 0,$$

then,

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}(x_{kj}-\mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1i_2\ldots i_pk}}.$$

Formula (3.10) derivations for E and M steps in Section 3.2.1

$$
\begin{aligned}
\sum_{i,j=1,i\neq j}^{g} \alpha^{|j-i|} &= 2\sum_{i,j=1,i<j}^{g} \alpha^{j-i} \\
&= 2(\sum_{i<2}\alpha^{2-i} + \sum_{i<3}\alpha^{3-i} + \ldots + \sum_{i<g}\alpha^{g-i}) \\
&= 2[(g-1)\alpha + (g-2)\alpha^2 + \ldots + (g-(g-1))\alpha^{g-1}] \\
&= 2[g(\alpha + \alpha^2 + \ldots + \alpha^{g-1}) - (\alpha + 2\alpha^2 + \ldots + (g-1)\alpha^{g-1})] \\
&= 2[(g+1)\sum_{t=1}^{g-1}\alpha^t - \sum_{t=1}^{g-1}(t+1)\alpha^t] \\
&= 2[(g+1)\sum_{t=1}^{g-1}\alpha^t - \frac{d}{d\alpha}\sum_{t=1}^{g-1}\alpha^{t+1}] \\
&= 2[\frac{(g+1)\alpha(1-\alpha^{g-1})}{1-\alpha} - \frac{d}{d\alpha}(\frac{\alpha^2(1-\alpha^{g-1})}{1-\alpha})] \\
&= 2[\frac{(g+1)(\alpha-\alpha^g)}{1-\alpha} - \frac{d}{d\alpha}(\frac{\alpha^2-\alpha^{g+1}}{1-\alpha})] \\
&= 2[\frac{(g+1)(\alpha-\alpha^g)}{1-\alpha} - \frac{2\alpha-\alpha^2-(g+1)\alpha^g+g\alpha^{g+1}}{(1-\alpha)^2}] \\
&= \frac{2[(g-1)\alpha - g\alpha^2 + \alpha^{g+1}]}{(1-\alpha)^2}.
\end{aligned}
$$

Let $f(\alpha)=\sum_{i,j=1,i\neq j}^{g}\alpha^{|j-i|}=\frac{2[(g-1)\alpha-g\alpha^2+\alpha^{g+1}]}{(1-\alpha)^2}$. We have the following Q-function:

$$
\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x},\mathbf{y},\mathbf{z},\mathbf{w}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} z_{ik}w_{jk}\log(\pi'_{ij}\phi_{\mu'_i,\sigma'^2_j}(x_k)\phi_{\nu'_j,\tau'^2_j}(y_k))|\Theta] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\pi'_{ij}) \\
&\quad +\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k))+\log(\phi_{\nu'_j,\tau'^2_j}(y_k))] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)\log(\pi'_{ii}) \\
&\quad +\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}|j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\alpha') \\
&\quad +\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k))+\log(\phi_{\nu'_j,\tau'^2_j}(y_k))].
\end{aligned}
$$

E-step:

When $i=j$,

$$
\mathrm{E}(z_{ik}w_{ik}|x_k,y_k,\Theta) = \frac{\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)+\alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.
$$

When $i\neq j$,

$$
\mathrm{E}(z_{ik}w_{jk}|x_k,y_k,\Theta) = \frac{\alpha\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{\sum_{i=1}^{g}\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)+\alpha\sum_{i,j=1,i\neq j}^{g}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}.
$$

M-step: The optimization of $\pi_{ii},\alpha,\mu_i,\sigma_i^2,\nu_j$, and τ_j^2 is a maximum likelihood estimation of the parameters as below,

$$
\frac{\partial Q(\Theta,\Theta')}{\partial\pi'_{ii}}=\frac{\sum_{k=1}^{n}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}}+\frac{\sum_{k=1}^{n}\mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}\frac{\partial\pi'_{gg}}{\partial\pi'_{ii}}=0, \qquad (3.33)
$$

for $i = 1, 2, \ldots, g-1$ the above equation holds.

$$\begin{aligned}\frac{\partial Q(\Theta,\Theta')}{\partial \alpha'} &= \frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}|j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta)}{\alpha'} + \frac{\sum_{k=1}^{n}\mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}\frac{\partial \pi'_{gg}}{\partial \alpha'} \\ &= 0. \end{aligned} \tag{3.34}$$

Subject to $f(\alpha) + \sum_{i=1}^{g}\pi_{ii} = 1$, $\frac{\partial \pi'_{gg}}{\partial \pi'_{ii}} = -1$, and $\frac{\partial \pi'_{gg}}{\partial \alpha'} = -\frac{d}{d\alpha'}f(\alpha')$. From Equations (3.33) and (3.34), we have the following equations:

$$\begin{aligned}\frac{\sum_{k=1}^{n}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} &= \frac{\sum_{k=1}^{n}\mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \qquad (3.35)\\ &= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{g}\pi'_{ii}},\end{aligned}$$

$$\frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}|j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta)}{\alpha\frac{d}{d\alpha'}f(\alpha')} = \frac{\sum_{k=1}^{n}\mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}. \tag{3.36}$$

When $g = 3$, $f(\alpha) = 4\alpha + 2\alpha^2$, $f'(\alpha) = 4 + 4\alpha$. Then we have the following equations

$$\begin{aligned}\frac{\sum_{k=1}^{n}\mathrm{E}(z_{1k}w_{1k}|\Theta)}{\pi'_{11}} &= \frac{\sum_{k=1}^{n}\mathrm{E}(z_{2k}w_{2k}|\Theta)}{\pi'_{22}} \qquad (3.37)\\ &= \frac{\sum_{i=1}^{3}\sum_{k=1}^{n}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{3}\pi'_{ii}},\end{aligned}$$

So

$$\frac{\sum_{k=1}^{n}\mathrm{E}(z_{1k}w_{ik}|\Theta) + \sum_{k=1}^{n}\mathrm{E}(z_{1k}w_{ik}|\Theta)}{\pi'_{11}+\pi'_{22}} = \frac{\sum_{k=1}^{n}\sum_{i=1}^{3}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{1-f(\alpha')} \tag{3.38}$$

$$\begin{aligned}\frac{\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{3}|j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta)}{\alpha'(4+4\alpha')} &= \frac{\sum_{k=1}^{n}\mathrm{E}(z_{3k}w_{3k}|\Theta)}{\pi'_{33}} \qquad (3.39)\\ &= \frac{\sum_{k=1}^{n}\sum_{i=1}^{3}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{1-4\alpha'-2\alpha'^2}.\end{aligned}$$

Let $C_1 = \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{3}|j-i|\mathrm{E}(z_{ik}w_{jk}|\Theta), C_2 = \sum_{k=1}^{n}\sum_{i=1}^{3}\mathrm{E}(z_{ik}w_{ik}|\Theta)$. Equation

(3.39) can be written as

$$(2C_1 + 4C_2)\alpha'^2 + (4C_1 + 4C_2)\alpha' - C_1 = 0. \tag{3.40}$$

The reasonable solution to this equation is

$$\hat{\alpha}' = \frac{-(2C_1 + 2C_2) + \sqrt{C_1^2 + C_1 + 5C_2^2}}{2C_1 + 4C_2} \tag{3.41}$$

Formula (3.14) derivations for E and M steps in Section 3.3.1

Based on Section 1.3.2, we rewrite the function $Q(\Theta, \Theta')$.

$$\begin{aligned}
Q(\Theta, \Theta') &= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\log L_{PCD}(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z},\mathbf{w}}[\sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g} z_{ik}w_{jk}\log(\pi'_{ij}\phi_{\mu'_i,\sigma'^2_j}(x_k)\phi_{\nu'_j,\tau'^2_j}(y_k))|\Theta] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\pi'_{ij}) \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k)) + \log(\phi_{\nu'_j,\tau'^2_j}(y_k))] \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)\log(\pi'_{ii}) \\
&\quad + \sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\alpha') \\
&\quad + \sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\beta') \\
&\quad + \sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)\log(\gamma') \\
&\quad + \sum_{k=1}^{n}\sum_{i=1}^{g}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\log(\phi_{\mu'_i,\sigma'^2_i}(x_k)) + \log(\phi_{\nu'_j,\tau'^2_j}(y_k))].
\end{aligned}$$

E-step:

When $i = j$,

$$\mathrm{E}(z_{ik}w_{ik}|x_k, y_k, \Theta) = \frac{\pi_{ii}\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_i,\tau_i^2}(y_k)}{f_{PCD}(x_k, y_k)}.$$

When $i, j \in \{1, 2\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\alpha\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)}.$$

When $i, j \in \{2, 3\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\beta\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)}.$$

When $i, j \in \{1, 3\}, i \neq j$,

$$\mathrm{E}(z_{ik}w_{jk}|x_k, y_k, \Theta) = \frac{\gamma\phi_{\mu_i,\sigma_i^2}(x_k)\phi_{\nu_j,\tau_j^2}(y_k)}{f_{PCD}(x_k, y_k)}.$$

M-step: The optimization of $\pi_{ii}, \alpha, \mu_i, \sigma_i^2, \nu_j$, and τ_j^2 is a maximum likelihood estimation of the parameters as below,

$$\frac{\partial Q(\Theta, \Theta')}{\partial \pi'_{ii}} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \pi'_{ii}} = 0, \qquad (3.42)$$

for $i = 1, 2, \ldots, g-1$ the above equation holds.

$$\frac{\partial Q(\Theta, \Theta')}{\partial \alpha'} = \frac{\sum_{k=1}^{n} \sum_{i,j\in\{1,2\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\alpha'} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \alpha'} = 0. \qquad (3.43)$$

$$\frac{\partial Q(\Theta, \Theta')}{\partial \beta'} = \frac{\sum_{k=1}^{n} \sum_{i,j\in\{2,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\beta'} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \beta'} = 0. \qquad (3.44)$$

$$\frac{\partial Q(\Theta, \Theta')}{\partial \gamma'} = \frac{\sum_{k=1}^{n} \sum_{i,j\in\{1,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\gamma'} + \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \frac{\partial \pi'_{gg}}{\partial \gamma'} = 0. \qquad (3.45)$$

Subject to $2(\alpha + \beta + \gamma) + \sum_{i=1}^{g} \pi_{ii} = 1$, $\frac{\partial \pi'_{gg}}{\partial \pi'_{ii}} = -1$, and $\frac{\partial \pi'_{gg}}{\partial \alpha'} = -2$, $\frac{\partial \pi'_{gg}}{\partial \beta'} = -2$,

$\frac{\partial \pi'_{gg}}{\partial \gamma'} = -2$. From Equations (3.42)-(3.45), we have the following equations:

$$\begin{aligned} \frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} &= \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} \\ &= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii}}, \end{aligned} \tag{3.46}$$

$$\frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\alpha'} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}. \tag{3.47}$$

$$\frac{\sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\beta'} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}. \tag{3.48}$$

$$\frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\gamma'} = \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}}. \tag{3.49}$$

Then

$$\begin{aligned} \frac{\sum_{k=1}^{n} \mathrm{E}(z_{gk}w_{gk}|\Theta)}{\pi'_{gg}} &= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii}} \\ &= \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\alpha'} \\ &= \frac{\sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\beta'} \\ &= \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\gamma'} \\ &= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}{\sum_{i=1}^{g} \pi'_{ii} + 2\alpha' + 2\beta' + 2\gamma'} \\ &= \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta). \end{aligned}$$

From (3.46) and (3.49), we have

$$\frac{\sum_{k=1}^{n} \mathrm{E}(z_{ik}w_{ik}|\Theta)}{\pi'_{ii}} = \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{ik}|\Theta) + \sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta),$$

$$\begin{aligned}\frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\alpha'} &= \frac{\sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\beta'}\\ &= \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2\gamma'}\\ &= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)+\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta).\end{aligned}$$

So

$$\hat{\pi'}_{ii} = \frac{\sum_{k=1}^{n}\mathrm{E}(z_{ik}w_{ik}|\Theta)}{\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)+\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.50}$$

$$\hat{\alpha'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,2\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)+\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)]}. \tag{3.51}$$

$$\hat{\beta'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{2,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)+\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)]}. \tag{3.52}$$

$$\hat{\gamma'} = \frac{\sum_{k=1}^{n}\sum_{i,j\in\{1,3\},i\neq j}\mathrm{E}(z_{ik}w_{jk}|\Theta)}{2[\sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(z_{ik}w_{ik}|\Theta)+\sum_{k=1}^{n}\sum_{i,j=1,i\neq j}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)]}. \tag{3.53}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \mu_i'} = \sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)\frac{x_k-\mu_i'}{\sigma_i'^2} = 0,$$

then,

$$\hat{\mu'}_i = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)x_k}{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.54}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial \sigma_i'^2} = \sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)[\frac{(x_k-\mu_i')^2}{2\sigma_i'^4}-\frac{1}{2\sigma_i'^2}] = 0,$$

then,

$$\hat{\sigma_i'^2} = \frac{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)(x_k-\mu_i')^2}{\sum_{k=1}^{n}\sum_{j=1}^{g}\mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.55}$$

By symmetry,

$$\hat{\nu'}_j = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)y_k}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}, \tag{3.56}$$

$$\hat{\tau}_j'^2 = \frac{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)(y_k-\nu_j')^2}{\sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(z_{ik}w_{jk}|\Theta)}. \tag{3.57}$$

Formula (3.16) derivations for E and M steps in Section 3.3.2

We rewrite the function $Q(\Theta,\Theta')$.

$$\begin{aligned}
Q(\Theta,\Theta') &= \mathrm{E}_{\mathbf{z}}[\log L(\mathbf{x},\mathbf{z}|\Theta')|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj}))|\Theta] \\
&= \mathrm{E}_{\mathbf{z}}[\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}(\prod_{l=1}^{p} z_{li_lk})\log(\prod_{j=1}^{p}\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})|\Theta)] \\
&= \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\pi'_{i_1i_2\ldots i_p}) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})) \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{lik}|\Theta)\log(\pi'_i) \\
&\quad +\sum_{t=1}^{\binom{p+g-1}{p}-g}\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A_t\}}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\alpha'_t) \\
&\quad +\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g}\mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})).
\end{aligned}$$

When $p = 3$ and $g = 3$,

$$
\begin{aligned}
Q(\Theta, \Theta') &= \sum_{k=1}^{n} \sum_{i=1}^{3} \mathrm{E}(z_{ik}^3 | \Theta) \log(\pi_i') \\
&+ \sum_{t=1}^{7} \sum_{k=1}^{n} \sum_{\{i_1, i_2, i_3 \subset A_t\}} \mathrm{E}(\prod_{l=1}^{3} z_{li_lk} | \Theta) \log(\alpha_t') \\
&+ \sum_{k=1}^{n} \sum_{i_1=1}^{3} \ldots \sum_{i_3=1}^{3} \mathrm{E}(\prod_{l=1}^{3} z_{li_lk} | \Theta) \sum_{j=1}^{3} \log(\phi_{\mu'_{ji_j}, \sigma'^2_{ji_j}}(x_{kj})).
\end{aligned}
$$

E-step:

When i_j's are identical,

$$
\mathrm{E}(\prod_{l=1}^{3} z_{lik} | \Theta) = \frac{\pi_i \prod_{j=1}^{3} \phi_{\mu_{ji}, \sigma_{ji}^2}(x_{k_j})}{f_{PCD}(x_1, x_2, x_3)}.
$$

When $\{i_1, i_2, i_3\} \subset A_t$, $t = 1, 2, \ldots, 7$,

$$
\mathrm{E}(\prod_{l=1}^{3} z_{li_lk} | \Theta) = \frac{\alpha_t \prod_{j=1}^{3} \phi_{\mu_{ji_j}, \sigma_{ji_j}^2}(x_{k_j})}{f_{PCD}(x_1, x_2, x_3)}.
$$

M-step: Let $\hat{u}_{i_1 i_2 \ldots i_p k} = \mathrm{E}(\prod_{l=1}^{p} z_{li_lk} | \Theta)$, and $\hat{u}_{ik} = \hat{u}_{i_1 i_2 \ldots i_p k | i_1 = i_2 = \ldots = i_p = i}$, the optimization of π_i, α, μ_{ji_j}, $\sigma_{ji_j}^2$ is simply a maximum likelihood estimation of the parameters as below,

$$
\frac{\partial Q(\Theta, \Theta')}{\partial \pi_i'} = \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\pi_i'} + \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi_g'} \frac{\partial \pi_g'}{\partial \pi_i'} = 0, \tag{3.58}
$$

for $i = 1, 2, \ldots, g - 1$ the above equation holds.

$$
\frac{\partial Q(\Theta, \Theta')}{\partial \alpha_t'} = \frac{\sum_{k=1}^{n} \sum_{\{i_1, i_2, \ldots, i_p \subset A_t\}} \hat{u}_{i_1 i_2 \ldots i_p k}}{\alpha_t'} + \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi_g'} \frac{\partial \pi_g'}{\partial \alpha_t'} = 0. \tag{3.59}
$$

When $p = 3$ and $g = 3$, subject to $6\alpha_1 + 3\sum_{t=2}^{7} \alpha_t + \sum_{i=1}^{g} \pi_i = 1$, and $\frac{\partial \pi_g'}{\partial \alpha_1'} = -6$, $\frac{\partial \pi_g'}{\partial \alpha_t'} = -3$, $t = 2, \ldots, 7$. From Equations (3.58) and (3.59), we have the following equations:

$$\begin{aligned}\frac{\sum_{k=1}^n \hat{u}_{ik}}{\pi'_i} &= \frac{\sum_{k=1}^n \hat{u}_{gk}}{\pi'_g} \\ &= \frac{\sum_{i=1}^g \sum_{k=1}^n \hat{u}_{ik}}{\sum_{i=1}^g \pi'_i},\end{aligned} \tag{3.60}$$

$$\frac{\sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_1\}} \hat{u}_{i_1\ldots i_p k}}{6\alpha'_1} = \frac{\sum_{k=1}^n \hat{u}_{gk}}{\pi'_g}, \tag{3.61}$$

$$\frac{\sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{3\alpha'_t} = \frac{\sum_{k=1}^n \hat{u}_{gk}}{\pi'_g}, t = 2,\ldots,7. \tag{3.62}$$

Then

$$\begin{aligned}&\frac{\sum_{k=1}^n \hat{u}_{gk}}{\pi'_g} \\ &= \frac{\sum_{i=1}^g \sum_{k=1}^n \hat{u}_{ik}}{\sum_{i=1}^g \pi'_i} \\ &= \frac{\sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_1\}} \hat{u}_{i_1\ldots i_p k}}{6\alpha'_1} \\ &= \frac{\sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{3\alpha'_t}, t = 2,\ldots,7 \\ &= \frac{\sum_{t=2}^7 \sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{3\sum_{t=2}^7 \alpha'_t} \\ &= \frac{\sum_{i=1}^g \sum_{k=1}^n \hat{u}_{ik} + \sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_1\}} \hat{u}_{i_1\ldots i_p k} + \sum_{t=2}^7 \sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{\sum_{i=1}^g \pi'_i + 6\alpha'_1 + 3\sum_{t=2}^7 \alpha'_t} \\ &= \sum_{k=1}^n \sum_{i=1}^g \hat{u}_{ik} + \sum_{k=1}^n \sum_{\{i_1,i_2,\ldots,i_p\subset A\}} \hat{u}_{i_1 i_2\ldots i_p k}.\end{aligned}$$

From (3.60) - (3.62), we have

$$\frac{\sum_{k=1}^n \hat{u}_{ik}}{\pi'_i} = \sum_{k=1}^n \sum_{i=1}^g \hat{u}_{ik} + \sum_{k=1}^n \sum_{\{i_1,i_2,\ldots,i_p\subset A\}} \hat{u}_{i_1 i_2\ldots i_p k},$$

$$\frac{\sum_{k=1}^n \sum_{\{i_1,\ldots,i_p\subset A_1\}} \hat{u}_{i_1\ldots i_p k}}{6\alpha'_1} = \sum_{k=1}^n \sum_{i=1}^g \hat{u}_{ik} + \sum_{k=1}^n \sum_{\{i_1\ldots,i_p\subset A\}} \hat{u}_{i_1\ldots i_p k},$$

$$\frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}}\hat{u}_{i_1\ldots i_p k}}{3\alpha'_t} = \sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1\ldots,i_p\subset A_t\}}\hat{u}_{i_1\ldots i_p k}, t=2,\ldots,7.$$

So when $p = 3$ and $g = 3$,

$$\hat{\pi'_i} = \frac{\sum_{k=1}^{n}\hat{u}_{ik}}{\sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}}\hat{u}_{i_1 i_2\ldots i_p k}}, \tag{3.63}$$

$$\hat{\alpha'_1} = \frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_1\}}\hat{u}_{i_1\ldots i_p k}}{6[\sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A\}}\hat{u}_{i_1\ldots i_p k}]}, \tag{3.64}$$

$$\hat{\alpha'_t} = \frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}}\hat{u}_{i_1\ldots i_p k}}{3[\sum_{k=1}^{n}\sum_{i=1}^{g}\hat{u}_{ik}+\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A\}}\hat{u}_{i_1\ldots i_p k}]}, t=2,\ldots,7. \tag{3.65}$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial\mu'_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}\frac{x_{kj}-\mu'_{ji_j}}{\sigma'^2_{ji_j}} = 0,$$

then

$$\hat{\mu'}_{ji_j} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}x_{kj}}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}}.$$

$$\frac{\partial Q(\Theta,\Theta')}{\partial\sigma'^2_{ji_j}} = \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}[\frac{(x_{kj}-\mu'_{ji_j})^2}{2\sigma'^4_{ji_j}}-\frac{1}{2\sigma'^2_{ji_j}}] = 0,$$

then,

$$\hat{\sigma'^2_{ji_j}} = \frac{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}(x_{kj}-\mu'_{ji_j})^2}{\sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_{j-1}=1}^{g}\sum_{i_{j+1}=1}^{g}\cdots\sum_{i_p=1}^{g}\hat{u}_{i_1 i_2\ldots i_p k}}.$$

For general p and g, let r_t be the number of elements in group A_t and d_{ti} be the number of repetition of component i. For group A_t, firstly we consider all the possible permutations without repetition of these p components $(1, 2, \ldots, p)$, there are $p!$ ways totally. Then, if there are d_{ti} repetitions of component i, we need to divide $p!$ by $d_{ti}!$.

So $r_t = \frac{p!}{\prod_{i=1}^{g} d_{ti}!}$, where d_{ti} is the number of repetition of component i in any element of this group. The restriction equation is

$$\sum_{t=1}^{\binom{p+g-1}{p}-g} r_t\alpha_t + \sum_{i=1}^{g}\pi_i = 1.$$

The function $Q(\Theta, \Theta')$ is

$$\begin{aligned} Q(\Theta,\Theta') &= \sum_{k=1}^{n}\sum_{i=1}^{g} \mathrm{E}(\prod_{l=1}^{p} z_{lik}|\Theta)\log(\pi_i') \\ &+ \sum_{t=1}^{\binom{p+g-1}{p}-g}\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A_t\}} \mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\log(\alpha_t') \\ &+ \sum_{k=1}^{n}\sum_{i_1=1}^{g}\cdots\sum_{i_p=1}^{g} \mathrm{E}(\prod_{l=1}^{p} z_{li_lk}|\Theta)\sum_{j=1}^{p}\log(\phi_{\mu'_{ji_j},\sigma'^2_{ji_j}}(x_{kj})). \end{aligned}$$

E-step:

When i_j's are identical,

$$\mathrm{E}(z_{ik}^{g}|\Theta) = \frac{\pi_i\prod_{j=1}^{g}\phi_{\mu_{ji},\sigma_{ji}^2}(x_{k_j})}{f_{PCD}(x_1,x_2,\ldots,x_p)}.$$

When $\{i_1, i_2, \ldots, i_p\} \subset A_t$,

$$\mathrm{E}(\prod_{l=1}^{g} z_{i_lk}|\Theta) = \frac{\alpha_t\prod_{j=1}^{g}\phi_{\mu_{ji_j},\sigma_{ji_j}^2}(x_{k_j})}{f_{PCD}(x_1,x_2,\ldots,x_p)}.$$

M-step:

$$\frac{\partial Q(\Theta,\Theta')}{\partial\pi_i'} = \frac{\sum_{k=1}^{n}\hat{u}_{ik}}{\pi_i'} + \frac{\sum_{k=1}^{n}\hat{u}_{gk}}{\pi_g'}\frac{\partial\pi_g'}{\partial\pi_i'} = 0, \tag{3.66}$$

for $i = 1, 2, \ldots, g-1$ the above equation holds.

$$\frac{\partial Q(\Theta,\Theta')}{\partial\alpha_t'} = \frac{\sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A_t\}}\hat{u}_{i_1i_2\ldots i_pk}}{\alpha_t'} + \frac{\sum_{k=1}^{n}\hat{u}_{gk}}{\pi_g'}\frac{\partial\pi_g'}{\partial\alpha_t'} = 0. \tag{3.67}$$

Subject to $\sum_{t=1}^{\binom{p+g-1}{p}-g} r_t\alpha_t + \sum_{i=1}^{g} \pi_i = 1$, $\frac{\partial \pi'_g}{\partial \alpha'_t} = -r_t$. We have the following equations:

$$\frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\pi'_i} = \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} = \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{i=1}^{g} \pi'_i}, \tag{3.68}$$

$$\frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{r_t\alpha'_t} = \frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g}. \tag{3.69}$$

Then

$$\begin{aligned}
\frac{\sum_{k=1}^{n} \hat{u}_{gk}}{\pi'_g} &= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{i=1}^{g} \pi'_i} \\
&= \frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{r_t\alpha'_t} \\
&= \frac{\sum_{t=1}^{\binom{p+g-1}{p}-g}\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{\sum_{t=1}^{\binom{p+g-1}{p}-g} r_t\alpha'_t} \\
&= \frac{\sum_{i=1}^{g}\sum_{k=1}^{n} \hat{u}_{ik} + \sum_{t=1}^{\binom{p+g-1}{p}-g}\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{\sum_{i=1}^{g} \pi'_i + \sum_{t=1}^{\binom{p+g-1}{p}-g} r_t\alpha'_t} \\
&= \sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}} \hat{u}_{i_1 i_2\ldots i_p k}.
\end{aligned}$$

From (3.68) and (3.69),

$$\hat{\pi'_i} = \frac{\sum_{k=1}^{n} \hat{u}_{ik}}{\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,i_2,\ldots,i_p\subset A\}} \hat{u}_{i_1 i_2\ldots i_p k}},$$

$$\hat{\alpha'_t} = \frac{\sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A_t\}} \hat{u}_{i_1\ldots i_p k}}{r_t[\sum_{k=1}^{n}\sum_{i=1}^{g} \hat{u}_{ik} + \sum_{k=1}^{n}\sum_{\{i_1,\ldots,i_p\subset A\}} \hat{u}_{i_1\ldots i_p k}]}.$$

Chapter 4

Experiments

In Chapter 2 and Chapter 3, several different approaches have been proposed along with comprehensive simulation studies. For further evaluation, we also apply our approaches to a collection of microarray gene expression data sets for a lung cancer study in this chapter. We not only estimate the parameters in various models, but also obtain the 95% confidence intervals for parameters in the original PCD model and two-level model. We calculate the confidence intervals in two ways: Louis' method and Bonferroni adjusted method, respectively. Based on these results, we discuss and compare the performance of different methods.

4.1 Confidence Interval Calculation by Louis' Formula

As described in Chapter 2, the parameters in the PCD, CC and CI models can be estimated through EM algorithm. Also, the estimation results for simulation stuies have been presented in Section 2.3. Although we have obtained the means and variances based on the complete data, for a better understanding of the performance of our method, we are going to calculate the confidence intervals for the means of parameters based on the incomplete data (observed information) in our future research. A technique for extracting the observed Fisher information matrix when using EM algorithm has been proposed [38]. Let $\mathbf{x}$ be the complete data, $\mathbf{y}$ be the incomplete data and $\boldsymbol{\theta}$ be the parameter vector. According to this approach, the fisher information matrix can be derived by the first derivative and second derivative matrices of $\log\{f(x|\theta)\}$. Let $S(x,\theta)$ and $S^*(y,\theta)$ be the gradient vectors of $\log\{f(x|\theta)\}$ and $\log\{f_Y(y|\theta)\}$, and let $B(x,\theta)$ and $B^*(y,\theta)$ be the negative of the associated second derivative matrices. Then based on the following Louis' fomulas, the Fisher information matrix from observed data could be achieved.

$$S^*(y,\theta) = E_\theta[S(X,\theta)|X \in R], \tag{4.1}$$

$$S^*(y,\hat{\theta}) = 0, \tag{4.2}$$

$$I_Y(\theta) = E_\theta[B(X,\hat{\theta})|X \in R] - E_\theta[S(X,\hat{\theta})S^T(X,\hat{\theta})|X \in R] + S^*(y,\hat{\theta})S^{*T}(y,\hat{\theta}), \tag{4.3}$$

where $R = \{x : y(x) = y\}$ and $\hat{\theta}$ is the maximum likelihood estimates found by the EM method.

When X_1, X_2. $\ldots$, X_n are independent, as we now concern about, the formulas can be simplified as below,

$$S^*(y,\theta) = \sum_{i=1}^{n} S_i^*(y_i,\theta) = \sum_{i=1}^{n} E_\theta[S_i(X_i,\theta)|X_i \in R_i], \tag{4.4}$$

$$\begin{aligned} I_Y &= \sum_{i=1}^{n} E_\theta[B_i(X_i,\hat{\theta})|X_i \in R_i] - \sum_{i=1}^{n} E_\theta[S_i(X_i,\hat{\theta})S_i^T(X_i,\hat{\theta})|X_i \in R_i] \\ &\quad - \sum_{i,j=1,i\neq j}^{n} E_\theta[S_i(X_i,\hat{\theta})|X_i \in R_i]E_\theta[S_j(X_j,\hat{\theta})|X_j \in R_j]^T. \end{aligned} \tag{4.5}$$

Since the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$ is the inverse of Fisher information matrix

I_y, the standard deviation of each parameter can be calculated and the confidence interval can be reached as well. We apply this approach to our study.

4.2 Applications to Experimental Data

For our application, we consider the following three published experimental microarray datasets from lung cancer studies [37]. All of these studies analyzed gene espression data from tumor samples of patients with lung adenocarcinomas. There were n =62 samples collected from Boston, n =86 samples from Michigan, and n =24 samples from Stanford, and the clinical outcomes of these patients were classified as "good" or "poor" outcome. 2865 common genes were shared in three data sets. We apply the two-level model to analyze three lists of z-scores calculated from the p-values of differential expression tests. The results estimated based on two-level model are presented in Table 4.1.

Table 4.1: Parameters estimation in a moderate two-level model with restriction when $N = 2865$, $p = 3$

Parameter	Estimate by EM	95% Confidence Interval
λ	0.31	(0.20, 0.42)
π_1	0.53	(0.31, 0.75)
π_2	0	(0, 0.13)
π_3	0.47	(0.20, 0.74)
ρ_{11}	0.25	(0.09, 0.42)
ρ_{12}	0.01	(0, 0.18)
ρ_{13}	0.74	(0.53, 0.95)
ρ_{21}	0	the variance is relatively small
ρ_{22}	1	the variance is relatively small
ρ_{23}	0	the variance is relatively small
ρ_{31}	0.22	(0, 1.19)
ρ_{32}	0.60	(0, 1.68)
ρ_{33}	0.18	(0, 1.22)
μ_{11}	-0.65	(-0.79, -0.50)
μ_{12}	0	with restriction
μ_{13}	0.45	(0.38, 0.53)
μ_{21}	-1.20	(-1.40, -0.98)
μ_{22}	0	with restriction
μ_{23}	1.08	(0.62, 1.54)
μ_{31}	-0.59	(-0.77, -0.39)
μ_{32}	0	with restriction
μ_{33}	0.96	(0.69, 1.22)
σ^2_{11}	0.73	(0.63, 0.83)
σ^2_{12}	1	with restriction
σ^2_{13}	0.49	(0.44, 0.53)
σ^2_{21}	0.86	(0.70, 1.03)
σ^2_{22}	1	with restriction
σ^2_{23}	0.78	(0.49, 1.07)
σ^2_{31}	0.84	(0.61, 1.08)
σ^2_{32}	1	with restriction
σ^2_{33}	0.81	(0.65, 0.98)

We apply Louis' method to obtain the 95% confidence intervals for the means of parameters [38]. Since proportion parameters can not be negative, if the lower bound is below zero, we set 0 as the lower bound instead of the negative value for some proportion parameters such as ρ_{12} and ρ_{31}.

We notice that 31% of the genes are estimated from complete concordant (CC) model by our two-level model. Almost all of these genes are differentially expressed, 53% of them are down-regulated, and the rest are up-regulated. For those data estimated from complete independence (CI) model, among Boston data, only 1% genes have no significant change in the tests, and one fourth are down-regulated; among Michigan data, none of them are differential-expressed significantly; among Stanford data, 22% genes are down-regulated, and 18% genes are up-regulated. For all the estimated means of z-scores, those from Michigan data have the smallest negative mean and largest positive mean. Results from application to three real experimental data sets suggest that our model works reasonably well in practice.

4.3 A Comparison of Methods for Model Reduction

We also apply other approaches presented in Chapter 3 to the experimental data. We make a comparison of estimation performances in Table 4.2 and 4.3 based on different methods, which includes the original PCD method (PCD), the multilevel normal mixture method (MLM),the exchangeable structure method (ESM) and the multiset coefficient method (MCM). We apply both Louis' method (LCI) and Bonferroni adjusted way (BACI) to obtain the 95% confidence intervals for the means of parameters based on PCD model.

Table 4.2: Proportion parameters estimation in different models with restriction when $N = 2865$, $p = 3$

Parameter	PCD	95% LCI	95% BACI	MLM	ESM	MCM
π_{111}	0.1066	(0, 0.4422)	(0, 0.6015)	0.1628	0.2619	0.2083
π_{112}	0.0573	(0, 0.2356)	(0, 0.3203)	0	0.0044	0
π_{113}	0.0005	(0, 0.0557)	(0, 0.0819)	0	0.0044	0.0057
π_{121}	0	(0, 0.3426)	(0, 0.5053)	0.0383	0.0044	0
π_{122}	0.0950	(0, 0.4962)	(0, 0.6866)	0.1056	0.0044	0.0574
π_{123}	0.0253	(0, 0.1176)	(0, 0.1614)	0.0320	0.0044	0.0034
π_{131}	0.0015	(0, 0.1158)	(0, 0.1701)	0	0.0044	0.0057
π_{132}	0.0038	(0, 0.1815)	(0, 0.2659)	0	0.0044	0.0034
π_{133}	0	(0, 0.1156)	(0, 0.1705)	0	0.0044	0
π_{211}	0.0344	(0, 0.2187)	(0, 0.3062)	0	0.0044	0
π_{212}	0	(0, 0.3765)	(0, 0.5552)	0	0.0044	0.0574
π_{213}	0	(0, 0.0593)	(0, 0.0874)	0	0.0044	0.0034
π_{221}	0.1332	(0, 0.7044)	(0, 0.9756)	0.0007	0.0044	0.0574
π_{222}	0.2982	(0, 1)	(0,1)	0.0023	0.5231	0.4464
π_{223}	0.0642	(0, 0.1733)	(0, 0.2251)	0.0006	0.0044	0
π_{231}	0.0001	(0, 0.1730)	(0, 0.2550)	0	0.0044	0.0034
π_{232}	0.0398	(0, 0.3523)	(0, 0.5007)	0	0.0044	0
π_{233}	0.0126	(0, 0.1776)	(0, 0.2560)	0	0.0044	0
π_{311}	0.0044	(0, 0.1310)	(0, 0.1910)	0	0.0044	0.0057
π_{312}	0.0007	(0, 0.1544)	(0, 0.2275)	0	0.0044	0.0034
π_{313}	0	(0, 0.0525)	(0, 0.0774)	0	0.0044	0
π_{321}	0.0152	(0, 0.2394)	(0, 0.3458)	0.1118	0.0044	0.0034
π_{322}	0.0167	(0, 0.3536)	(0, 0.5135)	0.3079	0.0044	0
π_{323}	0	(0, 0.0291)	(0, 0.0429)	0.0934	0.0044	0
π_{331}	0	(0, 0.1255)	(0, 0.1851)	0	0.0044	0
π_{332}	0.0479	(0, 0.2635)	(0, 0.3659)	0	0.0044	0
π_{333}	0.0426	(0, 0.1404)	(0, 0.1869)	0.1446	0.1085	0.1356

We apply Louis' method to obtain the 95% confidence intervals for the means of parameters based on PCD model. Since proportion parameters can be neither negative nor greater than 1, if the lower bound is below zero, we set 0 as the lower bound instead of the negative value for some proportion parameters; if the upper bound is above 1, we set 1 as the upper bound. We use results of λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$and $\boldsymbol{\rho_3}$ estimated by the two-level model to approximate π_{ijk} in orginal PCD model. When $i = j = k$, the $\pi_{iii} \approx \lambda\pi_i + (1 - \lambda)\rho_{1i}\rho_{2i}\rho_{3i}$; when i, j and k are not identical, $\pi_{ijk} \approx (1 - \lambda)\rho_{1i}\rho_{2j}\rho_{3k}$. Each π_{ijk} can be approximated for PCD model.

Table 4.3: Mean and variance parameters estimation in different models with restriction when $N = 2865$, $p = 3$

Parameter	PCD	95% LCI	95% BACI	MLM	ESM	MCM
μ_{11}	-0.800	(-1.147, -0.454)	(-1.311, -0.289)	-0.646	-0.544	-0.651
μ_{12}	0	with restriction	with restriction	0	0	0
μ_{13}	0.998	(0.540, 1.456)	(0.323, 1.674)	0.454	0.795	0.732
μ_{21}	-1.183	(-2.414, 0)	(-2.999, 0.633)	-1.195	-1.017	-1.093
μ_{22}	0	with restriction	with restriction	0	0	0
μ_{23}	1.337	(1.250, 1.424)	(1.209, 1.465)	1.080	1.421	1.415
μ_{31}	-0.958	(-1.455, -0.460)	(-1.692, -0.224)	-0.587	-0.434	-0.529
μ_{32}	0	with restriction	with restriction	0	0	0
μ_{33}	1.603	(1.286, 1.920)	(1.136, 2.070)	0.959	0.932	0.799
σ^2_{11}	0.654	(0.391, 0.917)	(0.265, 1.042)	0.733	0.932	0.852
σ^2_{12}	1	with restriction	with restriction	1	1	1
σ^2_{13}	0.462	(0.140, 0.784)	(0, 0.937)	0.489	0.502	0.566
σ^2_{21}	0.855	(0.618, 1.091)	(0.506, 1.203)	0.862	0.859	0.796
σ^2_{22}	1	with restriction	with restriction	1	1	1
σ^2_{23}	0.627	(0.450, 0.804)	(0.366, 0.889)	0.783	0.571	0.570
σ^2_{31}	0.603	(0.013, 1.193)	(0, 1.474)	0.8419	0.988	0.910
σ^2_{32}	1	with restriction	with restriction	1	1	1
σ^2_{33}	0.426	(0.230, 0.621)	(0.137, 0.714)	0.813	0.922	0.986

In the PCD model, there are in total 26 independent proportion parameters to be estimated. We can see that 29.82% of genes have no significant change in the tests, 10.66% of genes show the same down-regulated change in three studies, and 4.26% are all up-regulated expressed. For all the estimated means of z-scores, those from Michigan data have the smallest negative mean and those from Stanford have the largest positive mean.

In the multilevel mixture model, we reduce the number of independent proportion parameters from 26 to 9. We convert the results of λ, $\boldsymbol{\pi}$, $\boldsymbol{\rho_1}$, $\boldsymbol{\rho_2}$and $\boldsymbol{\rho_3}$ to approximate π_{ijk} in orginal PCD model. When $i = j = k$, the $\pi_{iii} \approx \lambda\pi_i + (1-\lambda)\rho_{1i}\rho_{2i}\rho_{3i}$; when i, j and k are not identical, $\pi_{ijk} \approx (1-\lambda)\rho_{1i}\rho_{2j}\rho_{3k}$. Each π_{ijk} can be approximated for PCD model. We notice that only 0.23% of genes are not differentially expressed in all the three data sets, 16.28% have the same down-regulated change and 14.46% have the same up-regulated difference. Comparing the estimation by MLM with the estimation by PCD, the results of some proportion parameters by MLM are close to those by PCD like π_{122} and π_{123}, while some other proportions are not well approximated. And the means from Michigan data estimated by MLM are near the means estimated by PCD. All the variances are not far from those by PCD for Boston data and Michigan data.

In the exchangeable structure model, the number of independent proportion parameters is only 3 because all the non-diagonal parameters are fixed to the same value. Almost half of the genes have no significant difference among three studies, 26.19% of genes are all down-regulated and 10.85% are all up-regulated. The diagonal proportions differ a lot from those estimated by PCD. For the means, those from Michigan data have the smallest negative mean and the largest positive mean. Also, the variances from Michigan are quite close to the variances estimated by PCD.

In the multiset coefficient model, we have 9 independent proportion parameters to estimate. We notice that 44.64% of genes are estimated for showing consistent behavior among all the data sets, 20.83% have identical down-regulated difference and 13.56% have identical up-regulated difference.

In addiation, we observe that the proportion parameter estimates of π_{323} and π_{333}, the mean parameter estimates of μ_{13}, μ_{23} and μ_{33}, as well as the variance parameter estimate of σ_{33}^2 from MLM are not covered by 95% LCI. All the proportions estimates from ESM are inside the 95% LCI, but the mean estimates of μ_{31}, μ_{33}, and variance estimates of σ_{11}^2 and σ_{33}^2 lie outside the 95% LCI. For MCM, All of the estimated parameters except μ_{33} and σ_{33}^2 are between the confidence bounds of LCI. Furthermore, it can be seen that the 95% BACI is wider than 95% LCI and covers more estimated paramters. The proportion estimate of π_{323}, the mean parameter estimates of μ_{23} and

μ_{33}, as well as the variance parameter estimate of σ_{33}^2 from MLM do not lie within the estimated 95% BACI. For ESM and MCM, the only two estimated parameter which is not covered by BACI are the same, which are μ_{33} and σ_{33}^2. From this point of view, MCM seems the most similar to the general PCD model.

4.4 Mathematical Derivations

Confidence interval derivations for application in Section 4.2

We have derived the multivariate mixture density of two level model g_{PCD} in Section 2.2.2. According to the Louis' method, let

$$\begin{aligned} v(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}, \Theta) &= \log f^*(\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}|\Theta) \\ &= \omega_k \log \lambda + (1-\omega_k) \log (1-\lambda) \\ &\quad + \sum_{i=1}^{g} \omega_k z_{ik}^{(CC)} \log \pi_i + \sum_{i=1}^{g} \sum_{j=1}^{p} (1-\omega_k) z_{ijk}^{(CI)} \log \rho_{ji} \\ &\quad + \sum_{i=1}^{g} \sum_{j=1}^{p} [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \log \phi_{\mu_{ji}, \sigma_{ji}^2}(x_{kj}). \end{aligned}$$

In our application, the number of components $g = 3$. Since π_3 is dependent on π_1 and π_2, and ρ_{j3} is dependent on ρ_{j1} and ρ_{j2}, we should exclude these three parameters to derive the differentiations. Let $\Theta_1 = \Theta \setminus \{\pi_3, \rho_{13}, \rho_{23}, \rho_{33}\}$. Let $\mathbf{y}$ be the complete data $\{\mathbf{x}, \boldsymbol{\omega}, \mathbf{z}\}$, so the elements of the gradient vector $S_k(\mathbf{y_k}; \Theta_1)$ are

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \lambda} = \frac{\omega_k}{\lambda} - \frac{1-\omega_k}{1-\lambda},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_i} = \frac{\omega_k z_{ik}^{(CC)}}{\pi_i} - \frac{\omega_k z_{gk}^{(CC)}}{\pi_g},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \rho_{ji}} = \frac{(1-\omega_k) z_{ijk}^{(CI)}}{\rho_{ji}} - \frac{(1-\omega_k) z_{gjk}^{(CI)}}{\rho_{jg}},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji}} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \frac{x_{kj} - \mu_{ji}}{\sigma_{ji}^2},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}][\frac{(x_{kj} - \mu_{ji})^2}{2\sigma_{ji}^4} - \frac{1}{2\sigma_{ji}^2}],$$

The negative elements of the second derivative matrix $B_k(\mathbf{y_k}; \Theta_1)$ are

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \lambda} = \frac{\omega_k}{\lambda^2} + \frac{1-\omega_k}{(1-\lambda)^2},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \pi_i} = \frac{\omega_k z_{ik}^{(CC)}}{\pi_i^2} + \frac{\omega_k z_{gk}^{(CC)}}{\pi_g^2},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \rho_{ji}} = \frac{(1-\omega_k) z_{ijk}^{(CI)}}{\rho_{ji}^2} + \frac{(1-\omega_k) z_{gjk}^{(CI)}}{\rho_{jg}^2},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \mu_{ji}} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \frac{1}{\sigma_{ji}^2},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}][\frac{(x_{kj} - \mu_{ji})^2}{\sigma_{ji}^6} - \frac{1}{2\sigma_{ji}^4}],$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji} \partial \sigma_{ji}^2} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \frac{x_{kj} - \mu_{ji}}{\sigma_{ji}^4},$$

Since ω_k, $\omega_k z_{ik}^{(CC)}$ and $(1-\omega_k) z_{ijk}^{(CI)}$ are all indicators, we can get the elements of the matrix $S_k(\mathbf{y_k}; \Theta_1) S_k^T(\mathbf{y_k}; \Theta_1)$.

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \lambda} \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \lambda} = (\frac{\omega_k}{\lambda} - \frac{1-\omega_k}{1-\lambda})(\frac{\omega_k}{\lambda} - \frac{1-\omega_k}{1-\lambda}) \quad = \quad \frac{\omega_k}{\lambda^2} + \frac{1-\omega_k}{(1-\lambda)^2},$$

the second equation holds because $\omega_k(1-\omega_k) = 0$. Similarly,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \lambda} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_i} \quad = \quad \frac{\omega_k z_{ik}^{(CC)}}{\lambda \pi_i} - \frac{\omega_k z_{gk}^{(CC)}}{\lambda \pi_g},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \lambda}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \rho_{ji}} = -\frac{(1-\omega_k)z_{ijk}^{(CI)}}{(1-\lambda)\rho_{ji}}+\frac{(1-\omega_k)z_{gjk}^{(CI)}}{(1-\lambda)\rho_{jg}},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \lambda}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \mu_{ji}} = [\frac{\omega_k z_{ik}^{(CC)}}{\lambda}-\frac{(1-\omega_k)z_{ijk}^{(CI)}}{1-\lambda}]\frac{x_{kj}-\mu_{ji}}{\sigma_{ji}^2},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \lambda}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma_{ji}^2} = [\frac{\omega_k z_{ik}^{(CC)}}{\lambda}-\frac{(1-\omega_k)z_{ijk}^{(CI)}}{1-\lambda}][\frac{(x_{kj}-\mu_{ji})^2}{2\sigma_{ji}^4}-\frac{1}{2\sigma_{ji}^2}],$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i} = \frac{\omega_k z_{ik}^{(CC)}}{\pi_i^2}+\frac{\omega_k z_{gk}^{(CC)}}{\pi_g^2},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \rho_{ji}} = 0,$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \mu_{ji}} = \frac{\omega_k z_{ik}^{(CC)}}{\pi_i}\cdot\frac{x_{kj}-\mu_{ji}}{\sigma_{ji}^2},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \mu_{jg}} = -\frac{\omega_k z_{gk}^{(CC)}}{\pi_g}\cdot\frac{x_{kj}-\mu_{jg}}{\sigma_{jg}^2},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma_{ji}^2} = \frac{\omega_k z_{ik}^{(CC)}}{\pi_i}[\frac{(x_{kj}-\mu_{ji})^2}{2\sigma_{ji}^4}-\frac{1}{2\sigma_{ji}^2}],$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \pi_i}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma_{jg}^2} = -\frac{\omega_k z_{gk}^{(CC)}}{\pi_g}[\frac{(x_{kj}-\mu_{jg})^2}{2\sigma_{jg}^4}-\frac{1}{2\sigma_{jg}^2}],$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \rho_{ji}}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \rho_{ji}} = \frac{(1-\omega_k)z_{ijk}^{(CI)}}{\rho_{ji}^2}+\frac{(1-\omega_k)z_{gjk}^{(CI)}}{\rho_{jg}^2},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \rho_{ji}}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \mu_{ji}} = \frac{(1-\omega_k)z_{ijk}^{(CI)}}{\rho_{ji}}\cdot\frac{x_{kj}-\mu_{ji}}{\sigma_{ji}^2},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \rho_{ji}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{jg}} = -\frac{(1-\omega_k) z_{gjk}^{(CI)}}{\rho_{jg}} \cdot \frac{x_{kj} - \mu_{jg}}{\sigma_{jg}^2},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \rho_{ji}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = \frac{(1-\omega_k) z_{ijk}^{(CI)}}{\rho_{ji}} [\frac{(x_{kj} - \mu_{ji})^2}{2\sigma_{ji}^4} - \frac{1}{2\sigma_{ji}^2}],$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \rho_{ji}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = -\frac{(1-\omega_k) z_{gjk}^{(CI)}}{\rho_{jg}} [\frac{(x_{kj} - \mu_{jg})^2}{2\sigma_{jg}^4} - \frac{1}{2\sigma_{jg}^2}],$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji}} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \frac{(x_{kj} - \mu_{ji})^2}{\sigma_{ji}^4},$$

$$\begin{aligned} \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} &= [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \\ &\cdot \frac{x_{kj} - \mu_{ji}}{\sigma_{ji}^2} \cdot [\frac{(x_{kj} - \mu_{ji})^2}{2\sigma_{ji}^4} - \frac{1}{2\sigma_{ji}^2}], \end{aligned}$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = [\omega_k z_{ik}^{(CC)} + (1-\omega_k) z_{ijk}^{(CI)}] \cdot [\frac{(x_{kj} - \mu_{ji})^2}{2\sigma_{ji}^4} - \frac{1}{2\sigma_{ji}^2}]^2.$$

Since we have

$$S^*(\mathbf{x}, \Theta_1) = \sum_{k=1}^{n} S_k^*(\mathbf{x_k}, \Theta_1) = \sum_{k=1}^{n} E_{\Theta_1}[S_k(\mathbf{y_k}, \Theta_1)],$$

The Fisher information matrix based on incomplete data $\mathbf{x}$ can be calculated by plugging-in the vectors and matrices to the following formula,

$$\begin{aligned} I_x &= \sum_{k=1}^{n} E_{\Theta_1}[B_k(\mathbf{y_k}, \hat{\theta})] - \sum_{k=1}^{n} E_{\Theta_1}[S_k(\mathbf{y_k}, \hat{\theta}) S_k^T(\mathbf{y_k}, \hat{\theta})] \\ &- \sum_{k,l=1,k\neq l}^{n} E_{\Theta_1}[S_k(\mathbf{y_k}, \hat{\theta})] E_{\Theta_1}[S_l(\mathbf{y_l}, \hat{\theta})]^T, \end{aligned} \quad (4.6)$$

where $\hat{\theta}$ is the maximum likelihood estimates found by the EM method. Since the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$ is the inverse of Fisher information matrix I_x, the standard deviation of each parameter in Θ_1 can be calculated. The variances of

$\pi_3, \rho_{13}, \rho_{23}$ and ρ_{33} can also be derived from the variance-covariance matrix, for example,

$$\begin{aligned} Var(\pi_3) &= Var(1-\pi_1-\pi_2) \\ &= Cov(1-\pi_1-\pi_2, 1-\pi_1-\pi_2) \\ &= Cov(\pi_1+\pi_2, \pi_1+\pi_2) \\ &= Var(\pi_1)+Var(\pi_2)+2Cov(\pi_1,\pi_2). \end{aligned}$$

Then all the confidence intervals can be reached as well.

Confidence interval derivations for application in Section 4.3

We have derived the multivariate mixture density of original PCD model f_{PCD} in Section 2.1. According to the Louis' method, let

$$\begin{aligned} v(\mathbf{x},\mathbf{z},\Theta) &= \log f^*(\mathbf{x},\mathbf{z}|\Theta) \\ &= \sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_p=1}^{g}[z_{1i_1k}z_{2i_2k}\ldots z_{pi_pk}\log\pi_{i_1i_2\ldots i_p}] \\ &\quad +\sum_{i_1=1}^{g}\sum_{i_2=1}^{g}\cdots\sum_{i_p=1}^{g}[z_{1i_1k}z_{2i_2k}\ldots z_{pi_pk}\sum_{j=1}^{p}\log\phi_{\mu_{ji_j},\sigma^2_{ji_j}}(x_{kj})]. \end{aligned}$$

In our application, the number of components $g=3$. Since π_{333} is dependent on other $\pi_{i_1i_2i_3}$, when i_1, i_2 and i_3 are not all equal to 3. We should exclude π_{333} to derive the differentiations. Let $\Theta_1 = \Theta \setminus \{\pi_{333}\}$. Let $\mathbf{y}$ be the complete data $\{\mathbf{x},\mathbf{z}\}$, so the elements of the gradient vector $S_k(\mathbf{y_k};\Theta_1)$ are

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial\pi_{i_1i_2i_3}} = \frac{z_{1i_1k}z_{2i_2k}z_{3i_3k}}{\pi_{i_1i_2i_3}} - \frac{z_{1gk}z_{2gk}z_{3gk}}{\pi_{ggg}},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial\mu_{ji_j}} = [\sum_{i_1,i_2,i_3\setminus i_j} z_{1i_1k}z_{2i_2k}z_{3i_3k}]\frac{x_{kj}-\mu_{ji_j}}{\sigma^2_{ji_j}},$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma^2_{ji_j}} = [\sum_{i_1,i_2,i_3 \backslash i_j} z_{1i_1k} z_{2i_2k} z_{3i_3k}][\frac{(x_{kj} - \mu_{ji_j})^2}{2\sigma^4_{ji_j}} - \frac{1}{2\sigma^2_{ji_j}}],$$

The negative elements of the second derivative matrix $B_k(\mathbf{y_k}; \Theta_1)$ are

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \pi_{i_1i_2i_3}} = \frac{z_{1i_1k} z_{2i_2k} z_{3i_3k}}{\pi^2_{i_1i_2i_3}} + \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi^2_{ggg}},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial^2 \mu_{ji_j}} = [\sum_{i_1,i_2,i_3 \backslash i_j} z_{1i_1k} z_{2i_2k} z_{3i_3k}] \frac{1}{\sigma^2_{ji_j}},$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma^2_{ji_j}} = [\sum_{i_1,i_2,i_3 \backslash i_j} z_{1i_1k} z_{2i_2k} z_{3i_3k}][\frac{(x_{kj} - \mu_{ji_j})^2}{\sigma^6_{ji_j}} - \frac{1}{2\sigma^4_{ji_j}}],$$

$$-\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji_j} \partial \sigma^2_{ji_j}} = [\sum_{i_1,i_2,i_3 \backslash i_j} z_{1i_1k} z_{2i_2k} z_{3i_3k}] \frac{x_{kj} - \mu_{ji_j}}{\sigma^4_{ji_j}},$$

Since $z_{1i_1k} z_{2i_2k} z_{3i_3k}$ are all indicators, we can get the elements of the matrix $S_k(\mathbf{y_k}; \Theta_1) S_k^T(\mathbf{y_k}; \Theta_1)$.

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1i_2i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1i_2i_3}} = \frac{z_{1i_1k} z_{2i_2k} z_{3i_3k}}{\pi^2_{i_1i_2i_3}} + \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi^2_{ggg}}.$$

When i_1, i_2, i_3 and i are not equal to g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1i_2i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji}} = \frac{z_{1i_1k} z_{2i_2k} z_{3i_3k}}{\pi_{i_1i_2i_3}} \cdot \frac{x_{kj} - \mu_{ji}}{\sigma^2_{ji}}.$$

When i_1 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{gi_2i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{1g}} = [\frac{z_{1gk} z_{2i_2k} z_{3i_3k}}{\pi_{gi_2i_3}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot \frac{x_{k2} - \mu_{1g}}{\sigma^2_{1g}}.$$

When i_2 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1gi_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{2g}} = [\frac{z_{1i_1k} z_{2gk} z_{3i_3k}}{\pi_{i_1gi_3}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot \frac{x_{k2} - \mu_{2g}}{\sigma^2_{2g}}.$$

When i_3 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 i_2 g}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{3g}} = [\frac{z_{1i_1k} z_{2i_2k} z_{3gk}}{\pi_{i_1 i_2 g}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot \frac{x_{k3} - \mu_{3g}}{\sigma_{3g}^2}.$$

When i_1, i_2, i_3 are not g, but$i$ is equal to g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 i_2 i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{jg}} = -\frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}} \cdot \frac{x_{kj} - \mu_{jg}}{\sigma_{jg}^2}.$$

When i_1, i_2, i_3 and i are not equal to g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 i_2 i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{ji}^2} = \frac{z_{1i_1k} z_{2i_2k} z_{3i_3k}}{\pi_{i_1 i_2 i_3}} \cdot [\frac{(x_{kj} - \mu_{ji})^2}{2\sigma_{ji}^4} - \frac{1}{2\sigma_{ji}^2}].$$

When i_1 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{g i_2 i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{1g}^2} = [\frac{z_{1gk} z_{2i_2k} z_{3i_3k}}{\pi_{g i_2 i_3}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot [\frac{(x_{k1} - \mu_{1g})^2}{2\sigma_{1g}^4} - \frac{1}{2\sigma_{1g}^2}].$$

When i_2 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 g i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{2g}^2} = [\frac{z_{1i_1k} z_{2gk} z_{3i_3k}}{\pi_{i_1 g i_3}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot [\frac{(x_{k2} - \mu_{2g})^2}{2\sigma_{2g}^4} - \frac{1}{2\sigma_{2g}^2}].$$

When i_3 and i are g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 i_2 g}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{3g}^2} = [\frac{z_{1i_1k} z_{2i_2k} z_{3gk}}{\pi_{i_1 i_2 g}} - \frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}}] \cdot [\frac{(x_{k3} - \mu_{3g})^2}{2\sigma_{3g}^4} - \frac{1}{2\sigma_{3g}^2}].$$

When i_1, i_2, i_3 are not g, but$i$ is equal to g,

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \pi_{i_1 i_2 i_3}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \sigma_{jg}^2} = -\frac{z_{1gk} z_{2gk} z_{3gk}}{\pi_{ggg}} \cdot [\frac{(x_{kj} - \mu_{jg})^2}{2\sigma_{jg}^4} - \frac{1}{2\sigma_{jg}^2}].$$

$$\frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji_j}} \cdot \frac{\partial v(\mathbf{y_k}, \Theta_1)}{\partial \mu_{ji_j}} = [\sum_{i_1, i_2, i_3 \setminus i_j} z_{1i_1k} z_{2i_2k} z_{3i_3k}] \frac{(x_{kj} - \mu_{ji_j})^2}{\sigma_{ji_j}^4},$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \mu_{ji_j}}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma^2_{ji_j}} = [\sum_{i_1,i_2,i_3\backslash i_j} z_{1i_1k}z_{2i_2k}z_{3i_3k}] \cdot\frac{x_{kj}-\mu_{ji_j}}{\sigma^2_{ji_j}}\cdot[\frac{(x_{kj}-\mu_{ji_j})^2}{2\sigma^4_{ji_j}}-\frac{1}{2\sigma^2_{ji_j}}],$$

$$\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma^2_{ji_j}}\cdot\frac{\partial v(\mathbf{y_k},\Theta_1)}{\partial \sigma^2_{ji_j}} = [\sum_{i_1,i_2,i_3\backslash i_j} z_{1i_1k}z_{2i_2k}z_{3i_3k}]\cdot[\frac{(x_{kj}-\mu_{ji_j})^2}{2\sigma^4_{ji_j}}-\frac{1}{2\sigma^2_{ji_j}}]^2.$$

Since we have

$$S^*(\mathbf{x},\Theta_1)=\sum_{k=1}^{n} S^*_k(\mathbf{x_k},\Theta_1)=\sum_{k=1}^{n} E_{\Theta_1}[S_k(\mathbf{y_k},\Theta_1)],$$

The Fisher information matrix based on incomplete data $\mathbf{x}$ can be calculated by plugging-in the vectors and matrices to the following formula,

$$\begin{aligned} I_x &= \sum_{k=1}^{n} E_{\Theta_1}[B_k(\mathbf{y_k},\hat{\theta})]-\sum_{k=1}^{n} E_{\Theta_1}[S_k(\mathbf{y_k},\hat{\theta})S_k^T(\mathbf{y_k},\hat{\theta})] \\ &\quad -\sum_{k,l=1,k\neq l}^{n} E_{\Theta_1}[S_k(\mathbf{y_k},\hat{\theta})]E_{\Theta_1}[S_l(\mathbf{y_l},\hat{\theta})]^T, \end{aligned} \tag{4.7}$$

where $\hat{\theta}$ is the maximum likelihood estimates found by the EM method. Since the estimated covariance matrix of $\hat{\boldsymbol{\theta}}$ is the inverse of Fisher information matrix I_x, the standard deviation of each parameter in Θ_1 can be calculated. The variances of π_{333} can also be derived from the variance-covariance matrix. Then all the confidence intervals can be reached as well.

Chapter 5

Future Work

As mentioned in Chapter 1, a statistical framework has been presented for an integrative analysis of differential expression based on two microarray data sets [19]. Before the data integration, the mixture-model based method has been proposed to test genome-wide concordance and discordance [32]. It is necessary to generalize this method for multiple data sets since we are interested in identifying genes with concordant behavior among multiple experiments. In Chapter 2, we have extended the PCD model for multiple data sets and described a two-level mixture model to approximate the PCD model. In the future, we will propose a statistical approach to detect concordantly differentially expressed genes among multiple microarray data sets based on our two-level mixture model (see Section 5.1). Furthermore, because we are also interested in identifying gene sets with concordant behavior among different experiments, the detection of concordant gene set enrichment will be presented in our future work (see Section 5.2).

5.1 Detection of concordant differential expression

We have proposed several mixture models:

$$f_{PCD}(x_{k1}, x_{k2}, \dots, x_{kp}) = \sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \dots \sum_{i_p=1}^{g} \pi_{i_1 i_2 \dots i_p} \prod_{j=1}^{p} \phi_{\mu_{ji_j}, \sigma^2_{ji_j}}(x_{kj}), \tag{5.1}$$

$$f_{CC}(x_{k1}, x_{k2}, \dots, x_{kp}) = \sum_{i=1}^{g} \pi_i \prod_{j=1}^{p} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj}), \tag{5.2}$$

$$f_{CI}(x_{k1}, x_{k2}, \dots, x_{kp}) = \prod_{j=1}^{p} [\sum_{i=1}^{g} \rho_{ji} \phi_{\mu_{ji}, \sigma^2_{ji}}(x_{kj})]. \tag{5.3}$$

For $g = 3$, index $i = 2$ is used to represent the null component with fixed parameters. We consider our models with restriction on: $\mu_{j2} = 0$ and $\sigma^2_{j2} = 1$,for $j = 1, 2, \dots, p$. Indices 1 and 3 are used to represent the down-regulated and up-regulated components with constrains: $\mu_{j1} \leq 0, \mu_{j3} \geq 0$, for $j = 1, 2, \dots, p$. $\pi_{i_1 i_2 \dots i_p}$ is the proportion of genes belonging to the i_j-th component in the j-th data set. $\pi_i = \pi_{ii \dots ii}$ is the proportion of genes belonging to the same i-th component in all the data sets. ρ_{ji} is the marginal proportion of genes belonging to the i-th component in the j-th data set. x is the corresponding z-score calculated by the normal-distribution-quantile-based transformation $z = \phi^{-1}(1 - p)$, where p is one-sided and upper-tail from the two-sample t-test. The parameters in the above models can be estimated through E-M algorithm, which we have already discussed in Chapter 2.

With the assumption of independence among n sets of observations $(\mathbf{x_1}, \mathbf{x_2}, \dots, \mathbf{x_n})$, the likelihoods of PCD, CC and CI models are:

$$L_{PCD} = \prod_{k=1}^{n} f_{PCD}[x_{k1}, x_{k2}, \dots, x_{kp}];$$

$$L_{CC} = \prod_{k=1}^{n} f_{CC}[x_{k1}, x_{k2}, \ldots, x_{kp}];$$

$$L_{CI} = \prod_{k=1}^{n} f_{CI}[x_{k1}, x_{k2}, \ldots, x_{kp}].$$

Based on these likelihoods, we consider the following two log-likelihood ratio tests:

$$T_{CC} = log(L_{PCD}/L_{CC}) = log(L_{PCD}) - log(L_{CC});$$

$$T_{CI} = log(L_{PCD}/L_{CI}) = log(L_{PCD}) - log(L_{CI}).$$

One tests PCD (H_1) against CC (H_0), the other tests PCD (H_1) against CI (H_0).

If CI can not be rejected, then the data integration will be discouraged; otherwise, we can establish CC or PCD model. If CC can not be rejected, then we need to consider a more efficient data integration. The concordant integrative score has been brought up, which has been defined as the conditional probability of concordantly differential expression under an appropriate mixture model:

Pr [concordant differential expression|observed $\mathbf{x_k}$]
=[Pr(observed $\mathbf{x_k}$ all up-regulated) + Pr(observed $\mathbf{x_k}$ all down-regulated)]/Pr (observed $\mathbf{x_k}$).

Under the CC model, the concordant integrative score is calculated as:

$$S_{cc}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \frac{\sum_{i=1,3} \hat{\pi}_i \prod_{j=1}^{p} \phi_{\hat{\mu}_{ji}, \hat{\sigma}_{ji}^2}(x_{kj})}{\sum_{i=1}^{3} \hat{\pi}_i \prod_{j=1}^{p} \phi_{\hat{\mu}_{ji}, \hat{\sigma}_{ji}^2}(x_{kj})}. \tag{5.4}$$

Under the PCD model, it is calculated as:

$$S_{pcd}(x_{k1}, x_{k2}, \ldots, x_{kp}) = \frac{\sum_{i=1,3} \hat{\pi}_{i,i,\ldots,i} \prod_{j=1}^{p} \phi_{\hat{\mu}_{ji}, \hat{\sigma}_{ji}^2}(x_{kj})}{\sum_{i_1=1}^{g} \sum_{i_2=1}^{g} \cdots \sum_{i_p=1}^{g} \hat{\pi}_{i_1 i_2 \ldots i_p} \prod_{j=1}^{p} \phi_{\hat{\mu}_{ji_j}, \hat{\sigma}_{ji_j}^2}(x_{kj})}. \tag{5.5}$$

For the future research, we will continue to complete this integrative analysis based on our two-level mixture model. We also plan not only to perform computational simulation studies with simulated data, but also to apply this approach to real experimental data to illustrate this framework.

5.2 Detection of concordant gene set enrichment

Gene set enrichment analysis is a computational method to determine whether the expression of gene set shows statistically significant differences in biomedical research. It is necessary to consider a concordant enrichment of gene sets to avoid misleading results. For the concordant integrative analysis of gene set enrichment, we still consider multiple large-scale two-sample gene expression data sets. x is still the z-score inversed from one-sided upper-tailed p-value. Let p to be the number of the data sets and m be the number of common genes in the data sets. The concordant enrichment score (CES) has been developed to measure the concordant enrichment for given gene set S. It is the probability that gene set S is concordantly enriched given observed x matrix of gene set S, and the term "enriched" has been explained as "higher/better than expected". Based on the component information, the CES has been calculated as the probability that the number of events of interest is greater than expected given the x matrix. For a gene in a given gene set, an event of interest can be: (1) the gene is up-regulated; (2) the gene is down-regulated; (3) the gene is differentially expressed (either up-regualted or down-regulated).

For example, if we are interested in the event that the gene is up-regulated differentially expressed. We can derive the following $u_{S,l}$ for a gene $G_{S,l}$ in a give gene set $S = \{G_{S,l} : l = 1, 2, \ldots, m_S\}$, which is the probability that gene $G_{S,l}$ is concordantly up-regulated differentially expressed given observed x matrix:

$$u_{S,l} = \frac{\pi_3 \prod_{j=1}^{p} \phi_{\mu_{j3},\sigma_{j3}^2}(x_{S,l,kj})}{f_{PCD}(x_{S,l,k1}, x_{S,l,k2}, \ldots, x_{S,l,kp})}. \tag{5.6}$$

For $g = 3$, index $i = 2$ of π_i is used to represent the null component with fixed parameters. We consider our models with restriction on: $\mu_{j2} = 0$ and $\sigma_{j2}^2 = 1$,for $j = 1, 2, \ldots, p$. Indices 1 and 3 are used to represent the down-regulated and up-regulated components with constrains: $\mu_{j1} \leq 0, \mu_{j3} \geq 0$, for $j = 1, 2, \ldots, p$.

For each gene $G_{S,l}$, the event of our interest satisfies a Bernoulli distribtuion with probability $u_{S,l}$. This $u_{S,l}$ can be estimated as $\hat{u}_{S,l}$ based on the estimated parameters in the PCD model. Let $h_{S,l}$ be either 0 or 1. We enumerate all possible sets of values for $\mathbf{h_S} = (h_{S,1}, h_{S,2}, \ldots, h_{S,m_S})$. And for each set, we need to check the condition that $\sum_{l=1}^{m_S} h_{S,l}$ (the total number of events of interest) is greater than $m_S\hat{\pi}_2$ (expected number). Since we have assumed that x_{kj} from different genes are independent in each data set j, $j = 1, 2, \ldots, p$, then the CES for the gene set $S = \{G_{S,l} : l = 1, 2, \ldots, m_S\}$:

$$CES_S = \sum_{\sum_{l=1}^{m_S} h_{S,l} > m_S\hat{\pi}_2} [\prod_{l=1}^{m_S} \hat{u}_{S,l}^{h_{S,l}} (1 - \hat{u}_{S,l})^{1-h_{S,l}}]. \quad (5.7)$$

Since $u_{S,l}$'s are ordinarily different in practice and we have m_S trials with probability of success in each trial $p = u_{S,l}$, the CES is the upper tail probability of binomial distribution when the trial probabilities $u_{S,l}$ are unequal.

We will continue to investigate and present the method for the concordant integrative analysis of gene set enrichment in the future.

5.3 AR(1) Structure for Multivariate Normal Mixture

The model has been proposed based on AR(1) structure in Section 3.3 when there are two data sets. We have simplified $\pi_{ij} = \alpha^{|j-i|}$ for model reduction when $p = 2$. When $g = 3$, let $f(\alpha) = \sum_{i,j=1, i\neq j}^{g} \alpha^{|j-i|} = 4\alpha + 2\alpha^2$.

Let $C_1 = \sum_{k=1}^{n} \sum_{i,j=1,i\neq j}^{3} |j-i| \mathrm{E}(z_{ik}w_{jk}|\Theta), C_2 = \sum_{k=1}^{n} \sum_{i=1}^{3} \mathrm{E}(z_{ik}w_{ik}|\Theta)$. Then $\hat{\alpha}$ is the root of the following equation

$$(2C_1 + 4C_2)\alpha'^2 + (4C_1 + 4C_2)\alpha' - C_1 = 0. \tag{5.8}$$

The reasonable solution to this equation is

$$\hat{\alpha'} \;=\; \frac{-(2C_1 + 2C_2) + \sqrt{C_1^2 + C_1 + 5C_2^2}}{2C_1 + 4C_2}. \tag{5.9}$$

As the number of data sets p increases, how to define a reasonable function $\pi_{i_1 i_2 \ldots i_p}$ with respect to α will be an issue. Although there are many ways to map α to $\pi_{i_1 i_2 \ldots i_p}$, we still want to simplify $\pi_{i_1 i_2 \ldots i_p}$ as a power function of α as we have done for bivariate normal mixture model. For example, if $p = 3$, when i_1, i_2 and i_3 are not identical, we try to simplify $\pi_{i_1 i_2 i_3}$=$\alpha^{|i_1-i_2|+|i_2-i_3|+|i_3-i_1|}$. For a three-component mixure model, subject to $\sum_{i_1,i_2,i_3=1}^{3} \pi_{i_1 i_2 i_3} = 1$, let $A = \{i_j = 1, 2, 3,\ i_j\text{'s are not identical, for } j = 1, 2, 3\}$, we have the following equation:

$$\sum_{\{i_1,i_2,i_3\}\subset A} \alpha^{|i_1-i_2|+|i_2-i_3|+|i_3-i_1|} + \sum_{i=1}^{3} \pi_{iii} = 1. \tag{5.10}$$

Let $h(i_1, i_2, i_3) = |i_1 - i_2| + |i_2 - i_3| + |i_3 - i_1|$. Since $i_j = 1$, 2 or 3, $h(i_1, i_2, i_3)$ is some integer within interval [1, 6]. Thus, the left side of Equation (4.16) is a polynomial of degree 4 with respect to α. We consider the Equation (4.16) and other possible equations obtained from M-step as a high-order equation system, it is very challenging to derive the expressions of general solution for α and π_{iii}.

Further, if the dimension extends to a general dimension p, the functions which associate $\pi_{i_1 i_2 \ldots i_p}$ with α will be more complex to be defined. For example, if we simplify $\pi_{i_1 i_2 \ldots i_p}$ similarly as we have done for $p = 3$,

$$\pi_{i_1 i_2 \ldots i_p} = \alpha^{\sum_{j_1,j_2=1}^{p} |i_{j_1} - i_{j_2}|}. \tag{5.11}$$

Then the restriction is

$$\sum_{\{i_1,i_2,\ldots,i_p\}\subset A} \alpha^{\sum_{j_1,j_2=1}^{p}|i_{j_1}-i_{j_2}|} + \sum_{i=1}^{p} \pi_{ii\ldots i} = 1. \tag{5.12}$$

In the future, we will try to solve this problem by programming. Moreover, because the order of the equation system will be relatively high as p goes up, to solve the equations will be much more complicated and difficult. We are going to explore such issues further for our future research.

Bibliography

[1] Alberts, B., Johnson, A., Lewis, J., et al. (2002). *Molecular Biology of the Cell*, Garland Science, New York.

[2] Maxam, A. M., Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of National Academy of Sciences* **74** 560–564.

[3] Strachan, T., Read, A. P. (1999). *Human Molecular Genetics*, Wiley-Liss, New York.

[4] Adams, J. U. (2008). Transcriptome: Connecting the Genome to Gene Function. *Nature Education* **1**.

[5] Selvaraj, S., Natarajan, J. (2011). Microarray Data Analysis and Mining Tools. *Bioinformation* **6** 95–99.

[6] Debnath, M., Prasad, G. B. K. S., Bisen, P. S. (2010). *Molecular Diagnostics: Promises and Possibilities*, Springer.

[7] Ross, D. T., Scherf, U., Eisen, M. B., et al. (2000). Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genetics* **24** 227–235.

[8] Perez-Diez, A., Morgun, A. and Shulzhenko, N. (2007). Microarrays for Cancer Diagnosis and Classification. *Advances in Experimental Medicine and Biology* **593** 74–85.

[9] Naidu, C. K., Suneetha, Y. (2012). Current Knowledge on Microarray Technology - An Overview. *Tropical Journal of Pharmaceutical Research* **11** 153–164.

[10] Dey, A., Singh, S. (2009). Progress of Science form Microscopy to Microarrays: Diagnosis of Parasitic Diseases *Journal of Laboratory Physicians* **1** 2–6.

[11] Woo, Y., Affourtit, J., Daigle, S., et al. (2004). A Comparison of cDNA, Oligonucleotide, and Affymetrix GeneChip Gene Expression Microarray Platforms. *Journal of Biomolecular Techniques* **15** 276–284.

[12] Blalock, E. M. (2003). *A Beginner's Guide to Microarrays*, Springer.

[13] Sansonetti, P., Zychlinsky, A. (2002). Molecular Cellular Microbiology. *Methods in Microbiology* **31**.

[14] Amaratunga, D., Cabrera, J. (2003). *Exploration and Analysis of DNA Microarray and Protein Array Data*, Wiley.

[15] Freifelder, D., Malacinski, G. M. (2003). *Essentials of Molecular Biology*, Jones & Bartlett Learning, LLC.

[16] Crick, F. (1970). Central Dogma of Molecular Biology. *Nature* **227**.

[17] Yashon, R., Cummings, M. (2011). *Human genetics and society*, Cengage Learning.

[18] Tarca, A. L., Romero, R., Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. *American Journal of Obstetrics and Gynecology* **195** 373–388.

[19] Lai, Y., Eckenrode, S. E., She, J. (2009). A statistical framework for integrating two microarray data sets in differential expression analysis. *BMC Bioinformatics* **10**.

[20] Nq, S. K., Tan, S. H., Sundararajan, V. S. (2003). On combining multiple microarray studies for improved functional classification by whole-dataset feature selection. *Genome Informatics* **14** 44–53.

[21] Hartung, J., Knapp, G., Sinha, B. K. (2008). *Statistical Meta-Analysis with Applications*, Wiley.

[22] Sutton, A. J., Abrams, K. R., Jones, D. R., et al. (2000). *Methods for Meta-Analysis on Medical Research*, Wiley.

[23] Conlon, E. M., Song, J. J., Liu, A. (2007). Bayesian meta-analysis models for microarray data: a comparative study. *BMC Bioinformatics* **9**.

[24] Rhodes, D. R., Barrette, T. R., Rubin, M. A., et al. (2002). Meta-analysis of microarrays: inter-study validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Reserach* **62** 4427–4433.

[25] Grützmann, R., Boriss, H., Ammerpohl, O., et al. (2005). Meta-analysis of microarray data on pancreatic cancer denes a set of commonly dysregulated genes. *Oncogene* **24**.

[26] Huttenhower, C., Hibbs, M., Myers, C., et al. (2006). A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* **22**.

[27] Conlon, E. M., Song, J. J., Liu, A. (2006). Bayesian models for pooling microarray studies with multiple sources of replications. *BMC Bioinformatics* **7**.

[28] Liu, J., Huang, J., Ma, S. (2013). Integrative analysis of multiple cancer genomic datasets under the heterogeneity model. *Statistics in Medicine* **32** 3509–3521.

[29] Nylund, K. L., Asparouhov, T., Muthen, B. O. (2007). Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling: A Multidisciplinary Journal* **14** 535–569.

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

[31] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*, Springer.

[32] Lai, Y., Adam, B., Rodolsky, R. and She, J.(2007). A mixture model approach to the tests of concordance and discordance between two large-scale experiments with two-sample groups. *Bioinformatics* **23** 1243–1250.

[33] McLanchlan, G., Krishnan, T. (1996). *The EM Algorithm and Extensions*, Wiley.

[34] Wu, C. F. J. (1983). On the Convergence Properties of the EM Algorithm. *The Annals of Statistics* **11** 95–103.

[35] Dempster, A. P., Laird, N. M., Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society* **39** 1–38.

[36] Kernighan, B. W., Ritchie, D. M. (1988). *The C programming lanuage*, Prentice-Hull.

[37] Subramanian, A., Tamayo, P., Mootha, V. K. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102** 15545–15550.

[38] Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society* **44** 225-233.