Title

Statistical Analysis of RNA-Seq Alternative Splicing Data and Gas Chromatography-Mass Spectrometry Data

Author

Yi, Yi

Publication Date

2016

Peer reviewed|Thesis/dissertation

University of California

Los Angeles

Statistical Analysis of RNA-Seq Alternative Splicing Data and Gas Chromatography-Mass Spectrometry Data

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Statistics

by

Yi Yi

2016

Abstract of the Dissertation

Statistical Analysis of RNA-Seq Alternative Splicing Data and Gas Chromatography-Mass Spectrometry Data

by

Yi Yi

Doctor of Philosophy in Statistics

University of California, Los Angeles, 2016

Professor Yingnian Wu, Chair

With the blossom of bio-chemical technologies in recent years, large and diverse data from every branch of biology has been generated. These data contain insightful truth of science and always present challenges to modeling, computation and interpretation. In this work, I present statistical models for two types of bioinformatic data: RNA-Seq alternative splicing and GCMS metabolomics. R packages *grMATS* and *gcmsDecon* are available for download.

The next-generation sequencing produces rich RNA-Sequencing data, where we observe alternative splicing events. Replicate multivariate analysis of transcript splicing (rMATS) has shown advantages over other existing methods for detection of differential alternative splicing from replicate RNA-Seq data. However, the current framework of rMATS only deals with two-isoform splicing events, which limits its usage. In this paper, we present a generalized rMATS framework to deal with multiple isoform splicing events and the model could also be extended to compare differential splicing between multiple groups. We provide a generalized likelihood ratio test where the null hypothesis allows user-defined threshold of splicing change for isoforms. We show that our test statistic follow a mixture

of chi-square distributions where the coefficients depend on values of the true parameters and a least favorable test statistic is computed when true parameters are unknown. We show efficacy of our model in both 27+3 simulations and a real dataset. Due to the huge demand for methods on multiple isoform RNA-Seq data, our model will be useful in RNA-Seq research projects.

As a collection of metabolic end-products, metabolome reflects the overall activity of the metabolic network and has been playing an important role in modern bio-chemical researches. Monitoring metabolites and relating their changes to the influence of other factors is a major scientific interest. The technology of Gas Chromatograpy-Mass Spectrometry (GCMS) produces from biological samples a metabolomic data type where each metabolite is broken into different masses (their relative proportions form a mass spectrum s) and co-elute within a retention time range where the spectrum is unchanged. This unique signature data structure enables individual metabolite identification and allows library construction for the whole metabolome. However, GCMS is unable to clearly separate different metabolite elutions, which poses a challenging problem of deconvolution and library matching. In addition, studies of metabolome usually involve multiple biological samples in order to understand which metabolites are related to diseases. Building the multiple correspondence across all samples further complicates the task. We propose an automatic rank-based non-negative matrix factorization model to streamline the spectral deconvolution, multiple corrspondence, metabolite selection and library matching. We apply the program on 27 simulation datasets as well as 2 real contrived datasets. All results show superior strength of our model over existing software.

The dissertation of Yi Yi is approved.

Yi Xing

Frederic R. Paik Schoenberg

Hongquan Xu

Yingnian Wu, Committee Chair

University of California, Los Angeles

2016

To my parents and my brother.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my advisor Prof. Yingnian Wu, for his patient supervision and guidance throughout the years, not only on statistics, but also on soft research skills. The two big topics in my dissertation are all attributed to his efforts of collaboration across science departments. More importantly, it is his philosophy of intuitive thinking that assures my own natural instinct of trying to understand statistics and mathematics, which I doubted many times in earlier years.

I would like to thank Prof. Rick Schoenberg for his guidance ever since before I was admitted into the doctoral program. It was his kindness of meeting with me and appreciating my "naive" ideas back then that reinforced my passion of going further in statistics. I appreciated his generous agreement on being on my committee and suggestions he gave me along the way.

I would like to thank Prof. Hongquan Xu for agreeing on being on my committee, for his suggestions on my dissertation work and for his pinpoint comments during both the oral and final defenses. As an international student, I also need to thank him as the department vice chair, for his generous help on my funding during several academic quarters.

Due to the nature of my dissertation work, I would like to take the time and thank all my collaborators: Prof. Yi Xing, who is on my committee, for taking me into his lab and guiding me on the first part of my dissertation and for bearing with my mistakes and slow progresses. Dr. Shihao Shen, for his collaborative efforts on the first part of my dissertation. Prof. Kym Faull and Farbod Fazlollahi for coming to me and my advisor in the first place and providing us the opportunity of working on the amazing project as the second part of my dissertation.

Austin Quach for continuing the collaboration on the second part after Farbod left for medical school.

I would like to thank Dr. Beate Ritz, Dr. Simin Liu and Dr. Thomas Drake for their two and a half years' generous support of Burroughs Wellcome Fund Inter-school Training Program in Chronic Diseases (BWF-CHIP, previously BWF-IT-MD). It was also their instructions that developed my biological sense for my dissertation.

I would like to thank our department Student Affairs Officer Ms. Glenda Jones, for informing me beforehand of the good news of admission into the doctoral program to start with and taking care of countless favors I needed throughout the years. The graduate life would be much tougher if not for her constant help.

Last but not least, I would like to thank my family - my parents and my brother for the unconditional love at any given time, especially for understanding my being away from home for so many years during the graduate study.

Chapter One is from the following work that we will submit soon.
Y. Yi, S. Shen, Y. Wu, Y. Xing. *grMATS: Statistical Modeling and Testing for Detection of Differential Alternative Splicing in Multiple Isoforms Using RNA-Seq Data.*

Chapter Two is from the following work that we will submit soon.
Y. Yi, F. Fazlollahi, A. Quach, K. Faull, Y. Wu. *gcmsDecon: Localized and Simultaneous Non-Negative Matrix Factorization for Deconvolution of Multiple GCMS Signals.*

Vita

2010	B.S. in Applied Mathematics
	Shanghai University, China
2012	C.Phil in Statistics
	University of California, Los Angeles
2014	M.S. in Statistics
	University of California, Los Angeles
2012 - 2015	Teaching Fellow, Graduate Student Researcher and Special Reader
	Department of Statistics, University of California, Los Angeles

CHAPTER 0

Background

We live in an era when biological technologies are growing at a rocketing speed (Figure 0.1). Ever since the 70s, this industry has been through countless booms and busts, yet not showing any sign of stopping. The technology hardware and analytical software are constantly innovating over themselves.

Figure 0.1: Biotech global sales trends and future. Source: Internet.

A simplified version of biology could be divided into four levels and categories (Figure 0.2), my work of RNA-Seq alternative splicing and GCMS deconvolution happen to fall into two of these four categories. As broad as the category names themselves are, this dissertation only covers a tiny fraction of all possible topics regarding RNA (Figure 0.3) and metabolites (Figure 0.4).

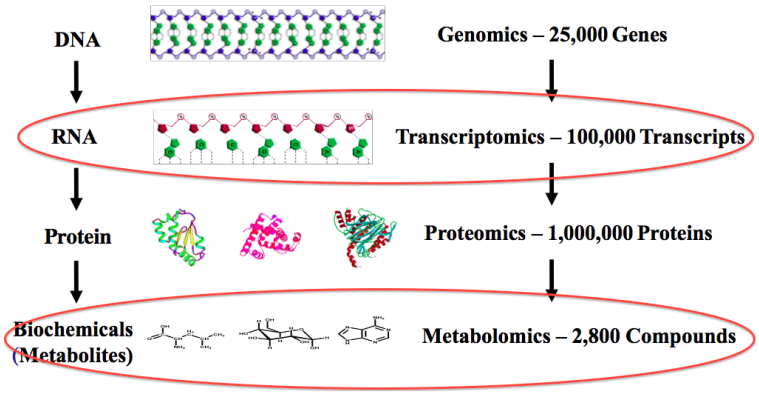


Figure 0.2: Four levels of biological world. Source: Internet.

⇓

Figure 0.3: RNA topics. Source: Internet.

Figure 0.4: Metabolomics topics. Source: Internet.

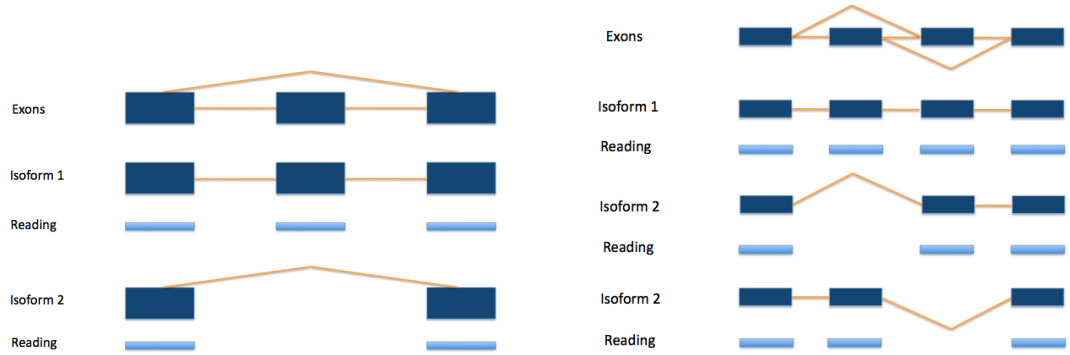
CHAPTER 1

grMATS: Statistical Modeling and Testing for Detection of Differential Alternative Splicing in Multiple Isoforms Using RNA-Seq Data

1.1 Introduction

The RNA sequencing (RNA-Seq) technology has been widely used for its powerful quantitative profiling of alternative splicing. As the cost of sequencing decreases, more and more replicate RNA-Seq data becomes available. Our interest here is to decide whether the isoform probability $\underline{\psi}_{\{I_f\}}$ differ between groups (usually case and control). A simple way of doing it is to pool all replicate data to fit one multinomial distribution. However, there are two issues unaddressed. First, since biological replicate comes from different patients, it is not wise to assume one single multinomial model with fixed parameters. Pooling loses individual information and those replicates with small total read counts would be under-represented in the estimation. Secondly, each replicate is likely to have its own baseline isoform proportions ψ's perturbed on its group level $\underline{\psi}_{\{I_f\}}$.

Based on a random-effect binomial model, rMATS [SPL+14] detects differential alternative splicing using RNA-Seq data of genes with two isoforms, and outperforms other existing methods with no such random effects. However, many genes have more than two isoforms (Figure 1.1), a more general model is in great demand.



(a) 2 isoforms (b) 3 isoforms

Figure 1.1: Alternative Splicing

The model we propose in this paper is a generalization of rMATS in terms of three aspects. First of all, we extend the 2-isoform 2-group model to any number of isoforms and groups. The random-effect model assumes in the first layer where multinomial logit transformation $\text{mlogit}(\psi_{I_f}) \sim N(\text{mlogit}(\underline{\psi}_{I_f}), \sigma^2)$ and in the second layer where read counts $R_{\{I_f\}}|\psi_{\{I_f\}} \sim MN(\underline{R}, \psi_{\{I_f\}})$. Secondly, the hypothesis testing framework is flexible to incorporate composite hypotheses of isoform probabilities between groups or within groups, i.e. $|\underline{\psi}_{g_1,I_f} - \underline{\psi}_{g_2,I_f}| \leq \Delta\underline{\psi}, \forall f, g_1, g_2$. Thirdly, besides the detection of significant genes, we identify which particular isoforms are significantly different between groups by comparing for each individual isoform probability $\underline{\psi}_{I_f}$, their unrestricted estimates between groups $\underline{\hat{\psi}}_{g,I_f}$ and testing with marginal hypotheses.

The accurate estimation of composite likelihood-ratio test distribution is essential for RNA-Seq data with multiple isoforms and groups. LRT statistic with equality constraints asymptotically follow χ^2 with degree of freedom as number of constraints. However, when the constraints become inequalities, the conventional LRT statistic distribution no longer holds. [Sha88] discusses LRT with various

types of cone constraints on the parameter space. [Che54] describes how the LRT with true parameter lying on constraint boundaries could be approximated by LRT with only cone constraints. [Sha87] defines and shows how the space around boundary point could be approximated by a cone. In our model, we combine these results and give the exact asymptotic composite LRT statistic distribution, which is a mixture of χ^2 with highest degree of freedom as the number of non-trivial inequality constraints.

We perform 27+3 simulations differing in terms of number of replicates, number of total read counts and logit variance levels. The model performances increase as the number of replicates increases and as logit variance decreases.

The model is also tested on the real dataset Hypoxia consisting of 10804 multiple-isoform genes from 3 replicates in control group under 20% oxygen and 3 in case group under 2% oxygen oxygen. With FDR cutoff 30% and isoform proportion difference $\Delta\underline{\psi} = 1\%$, grMATS identifies 671 genes with signifiant APA site shifts. Many of these genes are previously identified to be related to Hypoxia conditions.

1.2 Model

The RNA-Seq data consists of read counts from unpaired replicates of multiple groups, where each replicate has its own multiple-isoform read counts. Their total counts are usually different and assumed constant.

Notations:
n: gene index. $1 \leq n \leq N$. g: group index, $1 \leq g \leq G$, $G \geq 2$. k: replicate index, $1 \leq k \leq K_g$, the total number of replicates in group g. I_f: isoform index, $1 \leq f \leq F_n$, total number of isoforms for gene n. $\{\}$: The full set of values

spanning all possible indices inside it. e.g, $\{I_f\}$ means $\{I_1, ..., I_F\}$.

1.2.1 Hierarchical Model

In order to identify the multiple isoform proportion differences and address the random replicate effect at the same time, we propose a two-level hierarchical model with first layer as random replicate isoform proportions $\psi_{\{I_f\}}$ and second layer as the multinomial distribution based on the replicate isoform proportions.

For simplicity, we omit gene index n here. For k-th replicate of g-th group, f-th isoform, let R_{g,k,I_f} denote the read counts, $\underline{R}_{g,k} = \sum_{f=1}^{F} R_{g,k,I_f}$ is the total read counts of all isoforms in the gene n, considered as a non-random constant, $\underline{\mu}_{g,I_f} = \text{mlogit}(\underline{\psi}_{g,I_f})$ is group-level logit value, $\mu_{g,k,I_f} = \text{mlogit}(\psi_{g,k,I_f})$ as the corresponding random replicate logit value centered around the group-level logits. $p_{g,k,I_f} = \frac{l_{I_f}\psi_{g,k,I_f}}{\sum_{f=1}^{F} l_{I_f}\psi_{g,I_f}}$ is the isoform probability ψ_{g,k,I_f} adjusted by isoform length $l_{\{I_f\}}$. More details see appendix ***Logit Transformation***.

$$R_{g,k,\{I_f\}}|\psi_{g,k,\{I_f\}} \iff R_{g,k,\{I_f\}}|\mu_{g,k,\{I_f\}} \sim MN(\underline{R}_{g,k}, [p_{g,k,I_1}, ..., p_{g,k,I_F}]) \quad (1.1)$$

We assume the replicate multinomial logit follows normal distribution,

$$\mu_{g,k,I_f} \sim N(\underline{\mu}_{g,I_f}, \underline{\sigma}^2_{,I_f}), \quad 1 \leq f \leq F-1 \quad (1.2)$$

If $F = 2$, this model reduces to rMATS.

1.2.2 Likelihood Function

Omitting the gene index n and group index g. Combining the prior likelihood in (1.2) and conditional likelihood in (1.1) we have the joint likelihood, with which

we compute the marginal likelihood of replicate k

$$
\begin{aligned}
&P(R_{k,\{I_f\}}; \underline{\mu}_{k,\{I_f\}}, \underline{\sigma}^2_{k,\{I_f\}}) \\
&= \int P(R_{k,\{I_f\}}|\psi_{k,\{I_f\}})P(\psi_{k,\{I_f\}})d\psi_{k,\{I_f\}} = \int P(R_{k,\{I_f\}}|\mu_{k,\{I_f\}})P(\mu_{k,\{I_f\}})d\mu_{k,\{I_f\}} \\
&= \int \exp(h(\mu_{k,\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}))d\mu_{k,\{I_f\}} \qquad (1.3)
\end{aligned}
$$

where the logarithm of the joint density is

$$
\begin{aligned}
&h(\mu_{k,\{I_f\}}; \underline{\mu}_{k,\{I_f\}}, \underline{\sigma}^2_{k,\{I_f\}}) \\
&= \log(\frac{\underline{R}_k!}{\prod_{f=1}^{F} R_{k,I_f}!}) + \sum_{f=1}^{F-1} R_{k,I_f} \log(\underline{l}_{I_f} e^{\mu_{k,I_f}}) + R_{k,I_F} \log(\underline{l}_{I_F}) \\
&\quad - \underline{R}_k \log(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{k,I_f}} + \underline{l}_{I_F}) \quad + \left(\sum_{f=1}^{F-1} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\underline{\sigma}^2_{I_f}) - \frac{(\mu_{k,I_f} - \underline{\mu}_{I_f})^2}{2\underline{\sigma}^2_{I_f}} \right) \qquad (1.4)
\end{aligned}
$$

In order to compute (1.3) without the integration, we use Laplace approximation.

Theorem 1 [RYY00]:

Suppose $\mu \in \mathbb{R}^{F-1}$, $\hat{\mu}$ maximizes $h(\mu)$, then when $\underline{R}$ is big enough,

$$\int_{\mathbb{R}^{F-1}} \exp(h(\mu))d\mu \approx (\sqrt{2\pi})^{(F-1)} |-h^{(2)}(\hat{\mu})|^{-0.5} \exp(h(\hat{\mu})), \qquad (1.5)$$

where $h^{(2)}(\mu)$ is the second-order derivative of $h(\mu)$. Details see ***Laplace Approximation***.

Apply ***Theorem 1***,

$$P(R_{k,\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \approx (\sqrt{2\pi})^{F-1} |-h^{(2)}(\hat{\mu}_{k,\{I_f\}})|^{-0.5} \exp(h(\hat{\mu}_{k,\{I_f\}})) \qquad (1.6)$$

where $\hat{\mu}_{k,\{I_f\}} = \underset{\mu_{k,\{I_f\}}}{\operatorname{argmax}} h(\mu_{k,\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})$.

We assume independence of read counts between genes, groups and replicates.

With the gene index n and gene index g, the full likelihood function for n-th gene is,

$$
\begin{aligned}
&L(\underline{\mu}_{n,g,\{I_f\}}, \underline{\sigma}^2_{n,g,\{I_f\}}) = P(R_{n,\{g,k,I_f\}}; \underline{\mu}_{n,g,\{I_f\}}, \underline{\sigma}^2_{n,g,\{I_f\}}) \\
&= \prod_{g=1}^{G} \prod_{k=1}^{K_g} P(R_{n,g,k,\{I_f\}}; \underline{\mu}_{n,g,\{I_f\}}, \underline{\sigma}^2_{n,g,\{I_f\}}) \\
&\approx \prod_{g=1}^{G} \prod_{k=1}^{K_g} \Big\{ (\sqrt{2\pi})^{F-1} |-h^{(2)}(\hat{\mu}_{n,g,k,\{I_f\}}; \underline{\mu}_{n,g,\{I_f\}}, \underline{\sigma}^2_{n,g,\{I_f\}})|^{-0.5} \\
&\quad \exp(h(\hat{\mu}_{n,g,k,\{I_f\}}; \underline{\mu}_{n,g,\{I_f\}}, \underline{\sigma}^2_{n,g,\{I_f\}})) \Big\}
\end{aligned} \tag{1.7}
$$

Optimization on $\underline{\mu}_{n,\{g,I_f\}}, \underline{\sigma}^2_{n,\{g,I_f\}}$ should be applied to the smallest unit if possible for computational simplicity. In our case, we should optimize objective likelihood per gene&group $(n, g, \{k\}, \{I_f\})$ for unrestricted optimization, and per gene $(n, \{g\}, \{k\}, \{I_f\})$ for constrained optimization.

1.2.3 Composite Likelihood Ratio Test

Our interest is to test differences of $\underline{\psi}_{\{I_f\}}$ between all groups $\{g\}$ for one gene n. A typical problem of interest is to see if there is any isoform that satisfies inequality $|\underline{\psi}_{1,I_f} - \underline{\psi}_{2,I_f}| > \Delta\underline{\psi}$ instead of $\underline{\psi}_{1,I_f} \neq \underline{\psi}_{2,I_f}$. To solve this problem, we use composite likelihood-ratio test.

We note the feasible space for $\underline{\sigma}^2_{\{g,I_f\}}$ as Σ, where they take positive values, $\underline{\psi}_{\{g,I_f\}}$ as Ψ, and $\underline{\mu}_{\{g,I_f\}}$ as $\mathbb{M}$. Then

$$
\text{Problem: } \begin{cases} H_0 : \underline{\psi}_{\{g,I_f\}} \in \Psi_0 (\Leftrightarrow \underline{\mu}_{\{g,I_f\}} \in \mathbb{M}_0) \text{ And } \underline{\sigma}^2_{\{g,I_f\}} \in \Sigma \\ H_1 : \underline{\psi}_{\{g,I_f\}} \in \Psi_0^C \text{ And } \underline{\sigma}^2_{\{g,I_f\}} \in \Sigma \end{cases} \tag{1.8}
$$

$$
l(\underline{\mu}_{\{g,I_f\}}, \underline{\sigma}^2{}_{\{g,I_f\}}) = \log\ L(\underline{\mu}_{\{g,I_f\}}, \underline{\sigma}^2{}_{\{g,I_f\}}) = \log\ P(R_{\{g,k,I_f\}}; \underline{\mu}_{\{g,I_f\}}, \underline{\sigma}^2{}_{\{g,I_f\}}) \tag{1.9}
$$

LRT test statistic for gene n:

$$D = 2\Bigg(\sup\left\{ l(\underline{\mu}_{\{g,I_f\}}, \underline{\sigma}^2{}_{\{g,I_f\}}) : \underline{\mu}_{\{g,I_f\}} \in \mathbb{M}, \underline{\sigma}^2_{\{g,I_f\}} \in \Sigma \right\} \\ - \sup\left\{ l(\underline{\mu}_{\{g,I_f\}}, \underline{\sigma}^2{}_{\{g,I_f\}}) : \underline{\mu}_{\{g,I_f\}} \in \mathbb{M}_0, \underline{\sigma}^2_{\{g,I_f\}} \in \Sigma \right\} \Bigg) \tag{1.10}$$

Usually we let

$\Sigma = R_+^{G*(F-1)}$, $\mathbb{M} = R^{G*(F-1)}$, $\Psi = \{\underline{\psi}_{\{g,I_f\}} : \underline{\mu}_{\{g,I_f\}} \in \mathbb{M}\}$, $\Psi_0 = \{\underline{\psi}_{\{g,I_f\}} : \forall 1 \le g_1, g_2 \le G, \forall 1 \le f \le F, \ |\underline{\psi}_{g_1,I_f} - \underline{\psi}_{g_2,I_f}| \le \Delta\underline{\psi}, \quad \text{and } \forall 1 \le g \le G, \sum_{f=1}^{F} \psi_{g,I_f} = 1\}$, $\mathbb{M}_0 = \{\underline{\mu}_{\{g,I_f\}} : \underline{\psi}_{\{g,I_f\}} \in \Psi_0\}$

Let θ denote a parameter vector of dimension $m = G(2F-2)$,

$$\theta = [\underline{\mu}_{1,I_1}, \underline{\mu}_{1,I_2}, \cdots, \underline{\mu}_{1,I_{F-1}}, \cdots, \underline{\mu}_{G,I_1}, \underline{\mu}_{G,I_2}, \cdots, \underline{\mu}_{G,I_{F-1}}, \\ \underline{\sigma}^2_{1,I_1}, \underline{\sigma}^2_{1,I_2}, \cdots, \underline{\sigma}^2_{1,I_{F-1}}, \cdots, \underline{\sigma}^2_{G,I_1}, \underline{\sigma}^2_{G,I_2}, \cdots, \underline{\sigma}^2_{G,I_{F-1}}]^T \tag{1.11}$$

with the constrained space as $S = \mathbb{M}_0 \times \Sigma$ and its complement as $S^c = \mathbb{M}_0^c \times \Sigma$.

$$\text{Problem: } \begin{cases} H_0 : \theta \in S \\ H_1 : \theta \in S^c \end{cases} \qquad D = 2\left(\sup_{\theta \in S \cup S^c} l(\theta) - \sup_{\theta \in S} l(\theta) \right) \tag{1.12}$$

Constrained Parameter Space

The constraints on $\underline{\psi}$'s are usually written as a series of pairs of isoform probability difference inequalities for each isoform I_f in given group g_1 and g_2 smaller than some threshold $\Delta\underline{\psi}$.

$$|\underline{\psi}_{g_1,I_f} - \underline{\psi}_{g_2,I_f}| \le \Delta\underline{\psi} \iff \\ \begin{cases} s_{g_1,g_2,I_f}(\theta) = \frac{e^{\underline{\mu}_{g_1,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_1,I_f}} + 1} - \frac{e^{\underline{\mu}_{g_2,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_2,I_f}} + 1} - \Delta\underline{\psi} \le 0 \\ s_{g_2,g_1,I_f}(\theta) = \frac{e^{\underline{\mu}_{g_2,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_2,I_f}} + 1} - \frac{e^{\underline{\mu}_{g_1,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_1,I_f}} + 1} - \Delta\underline{\psi} \le 0 \end{cases} \tag{1.13}$$

This constrained parameter space for $\underline{\psi}$ is illustrated in Figure 1.2 and the corresponding logit $\underline{\mu}$ space is illustrated in Figure 1.3. Obviously, a boundary point

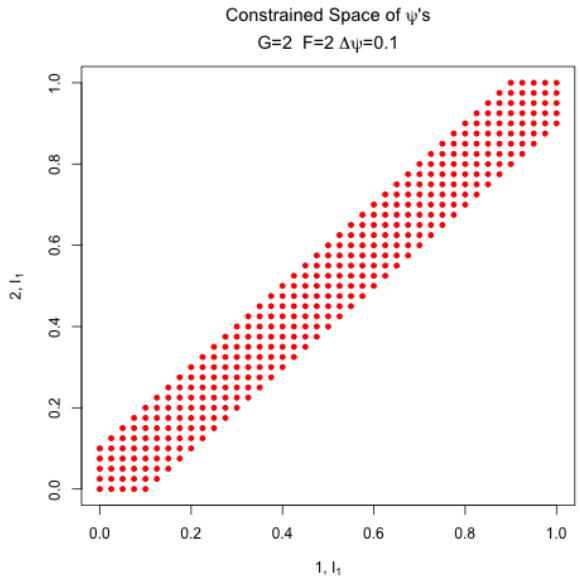


Figure 1.2: $G = 2,\ F = 2$. Constrained space for isoform probabilities. Same isoform different groups.

regarding isoform I_f inequalities can only locate in one of the constraints above. We do have non-negativity constraints on $\underline{\sigma}^2_{g,I_f}$. However, asymptotically the 0 is never to be touched and the MLE estimates of $\underline{\sigma}^2_{g,I_f}$ are not affected by non-negativity constraints. Thus we do not need to include them in the constraints above. It is also possible to have other constraints on $\underline{\sigma}^2_{g,I_f}$, but it is usually not of interest. The constrained space in (1.12) can be written as

$$S = \{\theta : s_{g_1,g_2,I_f}(\theta) \leq 0, 1 \leq f \leq F, 1 \leq g_1, g_2 \leq G\}, \tag{1.14}$$

Let Ξ denotes the constraints where true parameter θ_0 is located on the boundary

$$\Xi = \{([g_1, g_2], I_f) :\ s_{[g_1,g_2],I_f}(\theta_0) = 0, 1 \leq f \leq F-1,\ 1 \leq g_1 < g_2 \leq G\} \tag{1.15}$$

$[g_1, g_2]$ represents one the inequality boundary, either g_1, g_2 or g_2, g_1. $p = |\Xi|$, $m = |\theta_0| = G(2F-2)$ and because of the irrelevance of the order of elements in Ξ, we let $\{1, \cdots, p\}$ denote their indices. The boundary space of θ_0 is $\{\theta_0 : s_i(\theta_0) = 0, 1 \leq i \leq p\}$. Write

$$Q = [s_1'(\theta_0), ..., s_p'(\theta_0)]^T \tag{1.16}$$

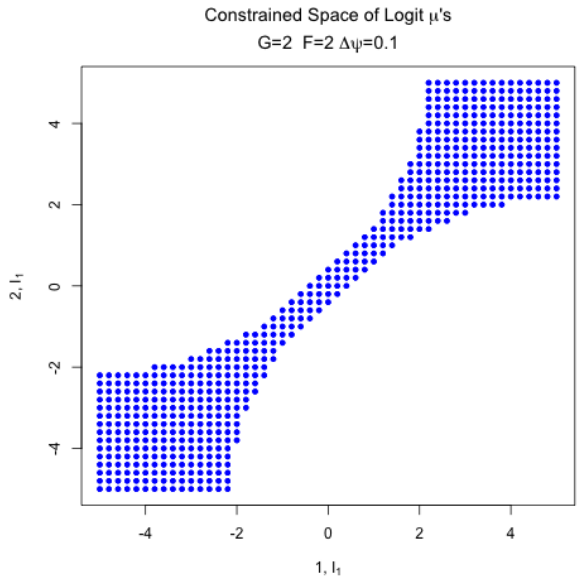


Figure 1.3: $G = 2$, $F = 2$. Constrained space for logit values. Same isoform different groups.

a matrix of $p \times m$, where $s_i'(\theta_0), 1 \leq i \leq p$ is the gradient of $s_i(\theta)$ at θ_0.

"Nominal" Replicates

Case 1. If $K_1 = \cdots = K_G$, then "nominal" replicates $1, \cdots, K$ are $i.i.d$ as $P(R_{\{g\},k,\{I_f\}}; \theta)$, the likelihood function is,

$$L = \prod_{k=1}^{K} P(R_{\{g\},k,\{I_f\}}; \theta) \tag{1.17}$$

Case 2. The orders of replicates do not matter since all of them are independent with each other.

If $K_1 = c_1 K, K_2 = c_2 K, \cdots, K_G = c_g K$, $c_1, \cdots, c_G$ are integers,

$$L = \prod_{k=1}^{K} \left(\prod_{g=1}^{G} \prod_{j=(k-1)c_g+1}^{kc_g} P(R_{g,j,\{I_f\}}; \theta) \right) \tag{1.18}$$

Case 3. Other forms of replicate number does not affect the parameter estimation of θ, but will require additional approximation in derivation of the asymptotic distribution of likelihood ratio test statistic.

We derive the exact asymptotically distribution of LRT statistic D in (1.12) based on asymptotic theories of maximum likelihood estimation, likelihood-ratio test and cone approximation. The following theorem is stated using Case 1 which is most common in our data. Similar results hold for Case 2.

Theorem 2 (Distribution of Composite Likelihood-ratio Test Statistic):

Fisher information for $P(R_{\{g\},k,\{I_f\}};\theta)$, $m \times m$ matrix,

$$\begin{aligned}\mathcal{I}_{\{g\}} &= -E\left[\frac{\partial^2}{\partial\theta^2}\log P(R_{\{g\},k,\{I_f\}};\theta)\right] \\ &= E\left[(\frac{\partial}{\partial\theta}\log P(R_{\{g\},k,\{I_f\}};\theta))(\frac{\partial}{\partial\theta}\log P(R_{\{g\},k,\{I_f\}};\theta))^T\right]\end{aligned} \tag{1.19}$$

with S, constrained space of θ_0 of size m as in (1.14), Ξ of size p, boundaries of constraints that θ_0 triggers as in (1.15), Q of dimension $p \times m$, gradients of constraints at θ_0 as in (1.16). We show asymptotically that the likelihood ratio test statistic D in (1.12)

$$D \xrightarrow{L} \sum_{i=0}^{m} w_i \chi_i^2 \tag{1.20}$$

$$\text{where} \quad w_i = \begin{cases} w_{p-i}(p, Q\mathcal{I}^{-1}Q^T) = w_i(p, (Q\mathcal{I}^{-1}Q^T)^{-1}) & 0 \le i \le p \\ 0 & p+1 \le i \le m \end{cases} \tag{1.21}$$

Let $Y \sim N(0, V)$, $V = Q\mathcal{I}^{-1}Q^T$, then the probability of Y falls in $\mathbb{R}_+^p$ is

$$w_j(p, V) = w_j(p, V, \mathbb{R}_+^p) = \sum_{|\alpha|=j} p(V_{\alpha'}^{-1}) p(V_{\alpha;\alpha'}), \tag{1.22}$$

Index α is a subset of $\{1, ..., p\}$, α' is its complement. They denote the indices of random variables in Y. $|\alpha|$ denotes the size of α. For example, if $\alpha = \{1, 2\}$, Y_α means $(Y_1, Y_2)^T$. $Y_\alpha \sim N(0, V_\alpha)$. V_α means the covariance matrix of Y_α. $V_{\alpha;\alpha'}$ means the conditional variance matrix of $Y_\alpha | Y_{\alpha'} = 0$. $P(V_\alpha) = P(Y_\alpha \ge 0), P(V_{\alpha;\alpha'}) = P(Y_\alpha \ge 0 | Y_{\alpha'} = 0)$.

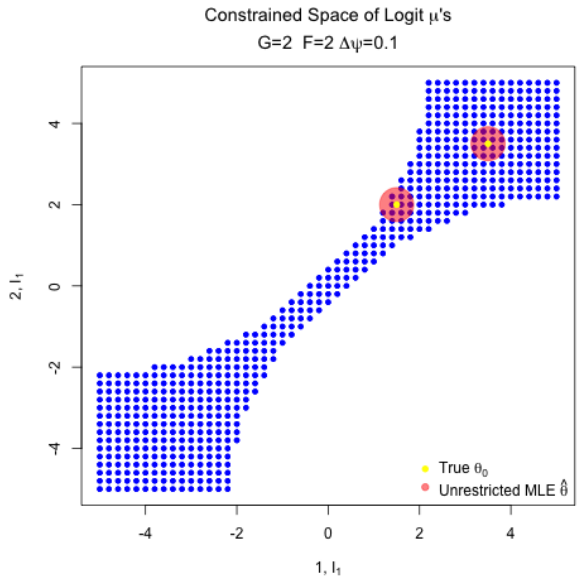


Figure 1.4: True value on the boundary and within the boundary.

Proof: see ***Proof of Theorem 2***.

Choice of Boundary Points - Least Favorable Test Statistic

However, in real data applications, we do not know if the true parameter is on the boundary (if it is inside the constrained space, and when sample size is big enough, the actual LRT test statistic is almost always 0, see Figure 1.4). A conservative choice is by assuming the true value on the boundary, instead of inside the constrained space. As shown in (1.20), the number of non-trivial boundary constraints determines the maximum number of degree of freedom in $\bar{\chi}^2$ distribution. We assume true θ_0 to be on boundary where most constraints are triggered, and each constraint is either one of the two in (1.13).

$$\mathbb{B}_{max} = \{\theta_0 : \ s_{[g_1,g_2],f}(\theta_0) = 0, ([g_1, g_2], f) \in \Xi\} \quad \text{where size of } \Xi \text{ is maximum} \tag{1.23}$$

On this boundary, the constrained MLE is most restricted, and the distribution of D is most left skewed and in general larger than the actual distribution (assuming

$Q\mathcal{I}^{-1}Q^T$ in (1.22) behave normally under different degree of freedoms). This way we keep the actual type I error small. We call the test statistic assuming true value on this boundary - the **least favorable test statistic**.

Usually $G = 2$, $|\Xi| = F - 1$, while in more general cases, the maximum number of possible non-redundant constraint intersections is $\binom{G}{2}(F-1)$. Besides, $G(F-1)$ is the total number of logit parameters, when G is big ($\binom{G}{2} > G$), it is not possible to have boundaries with $\binom{G}{2}(F-1)$ non-redundant constraints, we only need to choose $G(F-1)$ of them. For the least favorable test statistic, Q (1.16) is of dimension $[\min(\binom{G}{2}, G)(F-1)] \times G(2F-2)$ and Q has to be nonsingular.

With all the assumptions of the least favorable choice, we still need to estimate the true value on the assumed boundary, one way is to find a point on the boundary that is closest to our unrestricted MLE estimates.

$$\hat{\theta}_{Least} = \underset{\theta \in \mathbb{B}_{max}}{\operatorname{argmin}} \quad (\theta - \hat{\theta}_{MLE})^T (\theta - \hat{\theta}_{MLE}) \tag{1.24}$$

We let $\theta_0 = \hat{\theta}_{Least}$ and calculate the LRT test statistic D in (1.12) and its theoretical $\bar{\chi}^2$ distribution (1.20) using fisher information $\mathcal{I}$ (1.19), $\bar{\chi}^2$ weights (1.22).

1.2.4 Equal-weight $\bar{\chi}^2$ Test Statistic

We show that when the null hypothesis of likelihood ratio test is not linear equalities (e.g, $\underline{\psi}_{1,I_1} = \underline{\psi}_{2,I_1}$), but inequalities like $|\underline{\psi}_{1,I_1} - \underline{\psi}_{2,I_1}| \leq \Delta\underline{\psi}$, the likelihood ratio test statistic follows a mixture of χ^2 (*df* ranging from 0 to $p = |\Xi|$), instead of a conventional χ^2_p. The objective likelihood function involves parameters $\underline{\mu}_{g,I_1}, \cdots, \underline{\mu}_{g,I_{F-1}}$, $\underline{\sigma}^2_{g,I_1}, \cdots, \underline{\sigma}^2_{g,I_{F-1}}, 1 \leq g \leq G$, the weights for the $\bar{\chi}^2$ are determined by the location of true parameters. Because we do not know the true parameters, to use a most conservative, least favorable test statistic, we find a

point which triggers most constraints and that is closest to our unrestricted MLE estimators. The $\bar{\chi}^2$ weights are calculated for that point. For simplicity, the equal weight test statistic $\bar{\chi}^2_p = \sum_{i=0}^{p} w_i \chi^2_i$, $w_i = \frac{\binom{p}{i}}{2^p}$, usually is a good approximation to the actual test statistic when calculation of the least favorite test statistic is hard. We show in simulations both the least favorable test statistic and the equal-weight test statistic show good results of classification and inference. In practice, when isoform number $F \geq 8$), we use equal weight $\bar{\chi}^2$ and when $F \leq 7$, provide the option of using least favorable $\bar{\chi}^2$.

1.2.5 Detection of True Differential Isoforms

Our interest is not only in finding the significant genes, but also which specific isoforms are significantly differential. To address this, we further look at the marginal isoform significance within significant genes through hypothesis tests on individual isoforms $H_0^{I_f} : |\underline{\psi}_{g_1,I_f} - \underline{\psi}_{g_2,I_f}| \leq \Delta\underline{\psi}$. $H_1^{I_f}$: o.w.. We output the marginal p-values together with isoforms of large group differences based on estimated unrestricted $\hat{\underline{\psi}}_{\{g,I_f\}}$.

1.3 Simulation

1.3.1 Simulation of Asymptotic $\bar{\chi}^2$ Test Statistic

Here, to support all previous theoretical results, a 4-isoform gene for two groups is simulated, where $\underline{\psi}$'s between two groups are exactly with a difference $\Delta\underline{\psi} = 0.1$. We simulate this gene $N = 1000$ times, each time, every replicate has a total read counts $R = 100, 1000,$ or 10000 and replicate number fixed to be $K = 100, 500,$ or 2000 for both groups.

$$\text{variances } \underline{\sigma}^2_{\{g\},\{I_f\}} = 1, \quad \text{isoform lengths } \underline{l}_{\{I_f\}} = 1$$

	1	2	3	4	5	6	7	8	9
K	100	100	100	500	500	500	2000	2000	2000
R	100	1000	10000	100	1000	10000	100	1000	10000

Table 1.1: Combinations of K, R

$\underline{\psi}_{g,I_f}$	I_1	I_2	I_3	I_4
g = 1	0.08491507	0.3691922	0.2393482	0.3065446
g = 2	0.18491507	0.2691922	0.1393482	0.4065446

Table 1.2: True isoform proportions $\underline{\psi}$

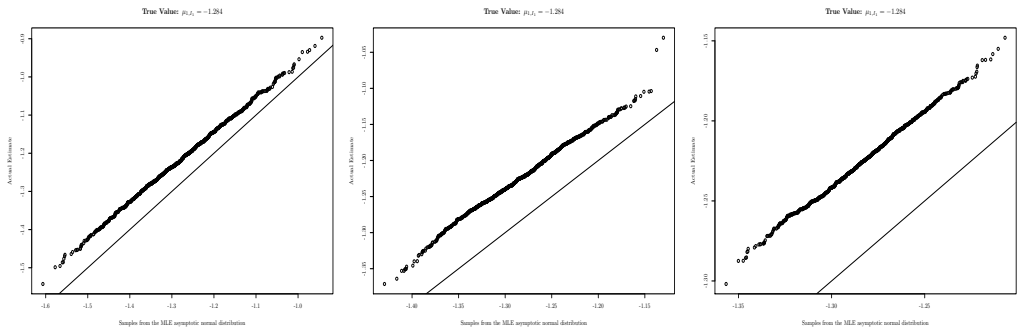
These parameters are on the boundary where most non-redundant equality constraints $|\underline{\psi}_{1,I_f} - \underline{\psi}_{2,I_f}| = 0.1$ are satisfied. Here we have 4 constraints, with one redundant. So the theoretical distribution of D is $\sum_{i=0}^{3} w_i \bar{\chi}_i^2$.

All theoretical quantiles are generated by taking 1000 samples from their asymptotic theoretical distribution.

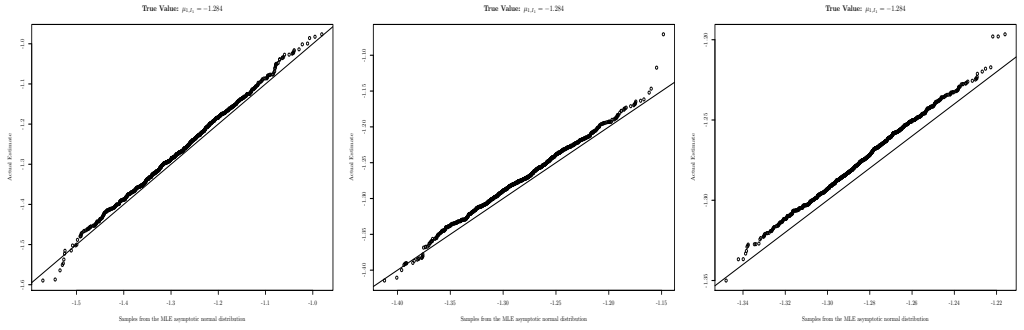
For cases where $R = 100$, unrestricted $\hat{\underline{\mu}}_{mle}$ converge to the true $\underline{\mu}$ ($\hat{\sigma^2}_{mle}$ to $\underline{\sigma}^2$) with a slight consistent difference. This difference doesn't improve much even with 500 replicates for both groups. This might be due to 1) the fact that Laplace approximation constant C not close enough to 1 when read counts is only 100, and C involves the parameters. 2) The existence of multiple local minima. This phenomenon disappears as total read counts R increases, where Laplace

$\underline{\mu}_{g,I_f}$	I_1	I_2	I_3
g = 1	-1.283712	0.1859540	-0.2474439
g = 2	-0.787797	-0.4122682	-1.0707179

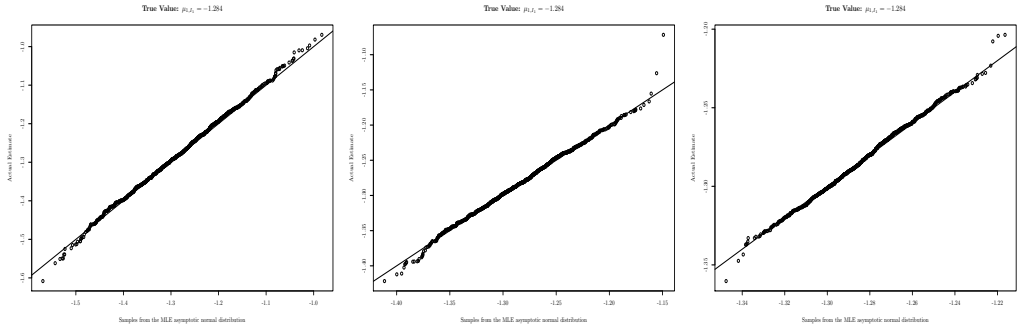
Table 1.3: True isoform logits $\underline{\mu}$



1. K = 100, R = 100 4. K = 500, R = 100 7. K = 2000, R = 100

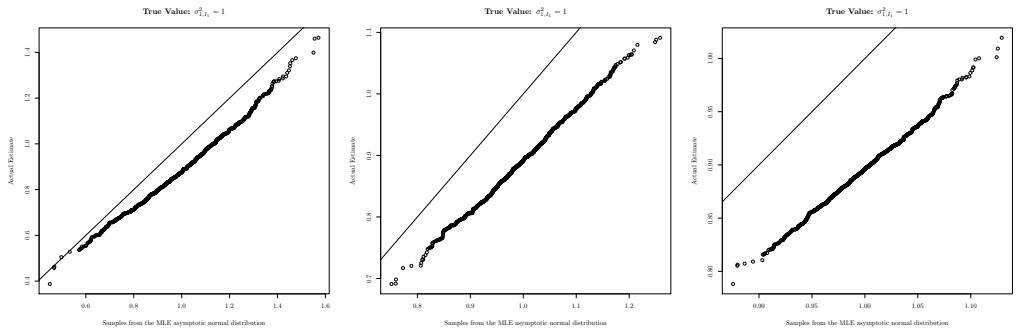


2. K = 100, R = 1000 5. K = 500, R = 1000 8. K = 2000, R = 1000

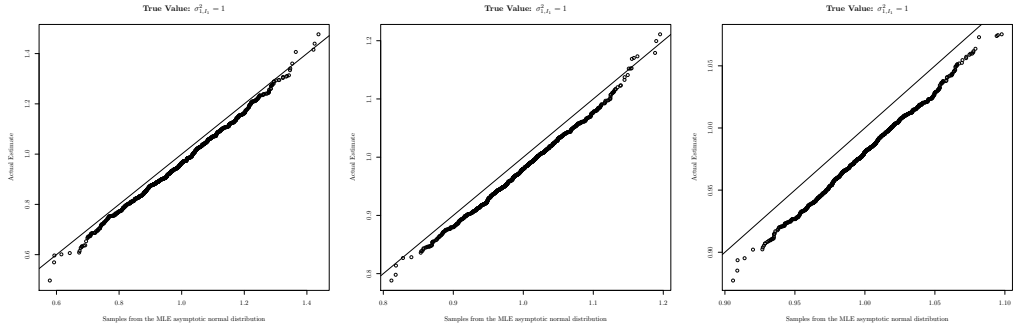


3. K = 100, R = 10000 6. K = 500, R = 10000 9. K = 2000, R = 10000

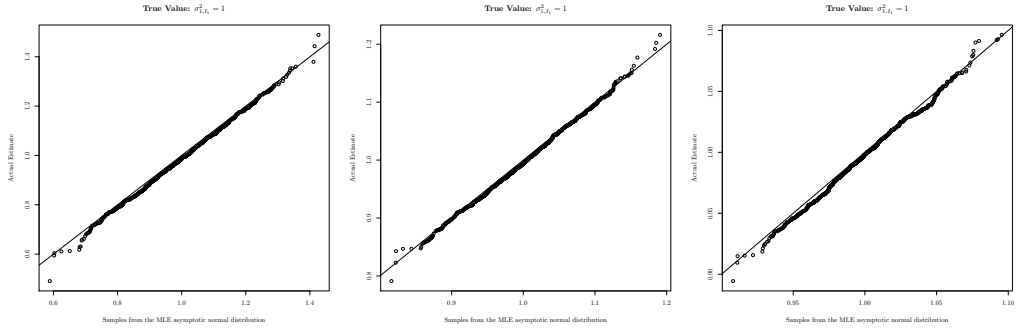
Figure 1.5: qq-plot of unrestricted $(\hat{\underline{\mu}}_{MLE})_{1,I_1}$ empirical against its theoretical asymptotic normal distribution.



1. K = 100, R = 100 4. K = 500, R = 100 7. K = 2000, R = 100



2. K = 100, R = 1000 5. K = 500, R = 1000 8. K = 2000, R = 1000



3. K = 100, R = 10000 6. K = 500, R = 10000 9. K = 2000, R = 10000

Figure 1.6: qq-plot of unrestricted $(\hat{\underline{\sigma}}^2_{MLE})_{1,I_1}$ empirical against its theoretical asymptotic normal distribution.

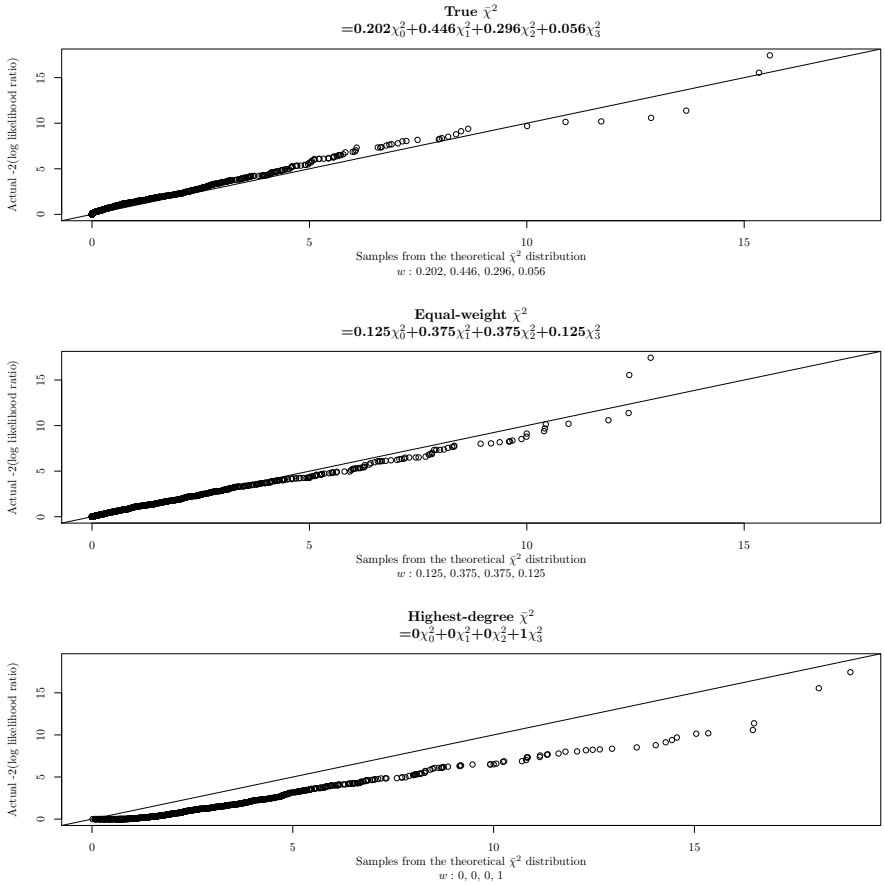
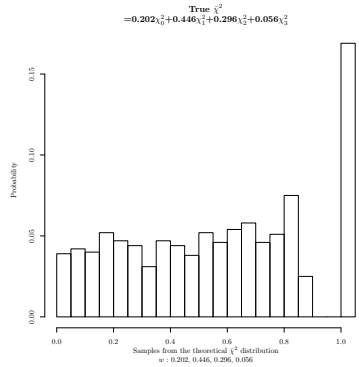
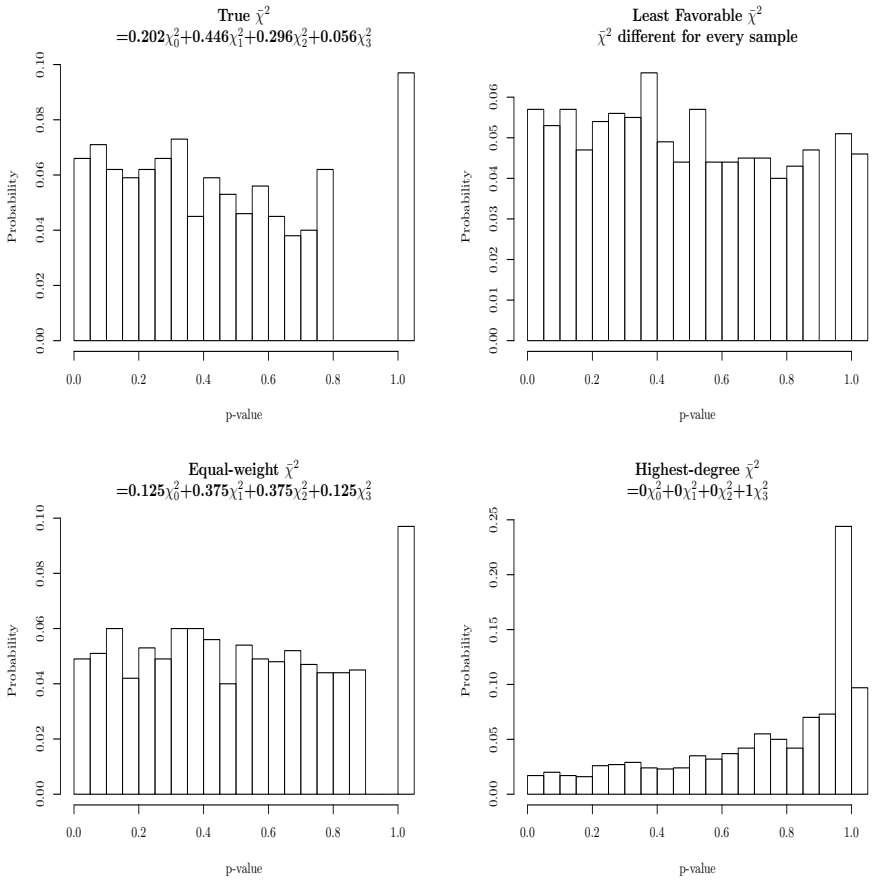


Figure 1.7: K=500, R=1000. qq-plot of pvalues of LRT statistic: under true $\bar{\chi}_3^2$, equal-weight $\bar{\chi}_3^2$ and χ_3^2

approximation is more accurate and the problem tends towards "convex". The estimates of parameters have no bias when cluster size $R = 1000$, because the Laplace approximation is quite accurate, the constant $C \approx 1$ has little to do with our parameters and prior distributions do not affect much in maximizing $h(\mu)$. See Figure 1.5 and 1.6.



(a) Histogram of pvalues of direct samples from true $\bar{\chi}^2_3$



(b) K=500, R=1000. histogram of pvalues of LRT statistic: under true $\bar{\chi}^2_3$, closest $\bar{\chi}^2_3$, equal-weight $\bar{\chi}^2_3$ and χ^2_3

Figure 1.8: Histogram of pvalues

In Figure 1.7, the top plot shows the qq-plot of the 1000 LRT statistic $-2\log\Lambda$ against 1000 random samples from the true theoretical $\bar{\chi}^2$ with the weight calculated through our method, where we used cones to approximate the surface around the true value. The middle plot shows the qq-plot of the 1000 LRT $-2\log\Lambda$ against 1000 random samples from the equal-weight $\bar{\chi}^2$, $w_i = \frac{\binom{F-1}{i}}{2^{F-1}}, 1 \leq i \leq F-1$ which is a good approximation the theoretical one if the theoretical weights don't deviate too much from the equal weights. The bottom plot shows the qq-plot of the 1000 LRT statistic D against 1000 random samples from χ^2_{F-1}, where degree of freedom is retrieved simply by counting the number of constraints. This χ^2_{F-1} distribution is too conservative and deviates a lot away from the true distribution.

Since in real applications, we are not able to know the true parameters. For each gene, we find on the boundary where most equality constraints are satisfied, the closest point to the unrestricted MLE estimation. We calculate $\bar{\chi}^2$ assuming true parameter is at this closest point (least favorable test statistic). If the distribution of p-values is approximately uniform except for a mass at 1, then this least favorable test statistic is a good approximation to the true test statistic. See Figure 1.8.

Scenarios where Accurate $\bar{\chi}^2$ Weight Estimation is Necessary

In previous simulations, the actual $\bar{\chi}^2$ does not show much superiority over the simple equal-weight $\bar{\chi}^2$, this is due to the off-diagonal elements in covariance matrix V in $w_j(p, V = R\mathcal{I}^{-1}R^T), j = 0, \cdots, p$ are not very big compared to the diagonal. Thus the weight calculation is very similar to equal-weight $\bar{\chi}^2$. If it is exactly a diagonal matrix, then the true distribution is equal-weight $\bar{\chi}^2$.

Under $R = 1000$,

$$V = \begin{bmatrix} 0.0333480519 & -0.0004903398 & 0.02465823 \\ -0.0004903398 & 0.0435565960 & 0.03892188 \\ 0.0246582305 & 0.0389218772 & 0.11243103 \end{bmatrix}$$

$\underline{\psi}_{g,I_f}$	I_1	I_2	I_3
g = 1	0.1740337	0.7170984	0.1088679
g = 2	0.0740337	0.8170984	0.1088679

Table 1.4: Situation where accurate weight calculation is necessary. $\underline{\psi}$.

$\underline{\mu}_{g,I_f}$	I_1	I_2
g = 1	0.4691137	1.885078
g = 2	-0.3856149	2.015624

Table 1.5: Situation where accurate weight calculation is necessary. $\underline{\mu}$.

The actual $\bar{\chi}_3^2 = 0.202\chi_0^2 + 0.446\chi_1^2 + 0.296\chi_2^2 + 0.056\chi_3^2$, which isn't much different from $\bar{\chi}_3^2 = 0.125\chi_0^2 + 0.375\chi_1^2 + 0.375\chi_2^2 + 0.125\chi_3^2$.

However, in situations where these two distribution differ a lot, the actual weight estimation becomes necessary. This following gene is one example where closest $\bar{\chi}^2$ is very different from equal-weight $\bar{\chi}^2$. It is easy to see that the off-diagonal elements of covariance matrix V are comparable to the diagonal. It is worth mentioning that the conventional χ_3^2 is spurious, and it is getting much worse compared to the true distribution as number of isoforms increase. It is not too off when the data only involves two isoforms.

$$V = \begin{bmatrix} 0.08472387 & 0.05756867 \\ 0.05756867 & 0.04539917 \end{bmatrix}$$

Let $K = 100, R = 1000$, $\underline{\sigma}^2_{\{g\},\{I_f\}} = 1$, $l_{\{I_f\}} = 1$, we generate data for this gene 1000 times. We can clearly see that equal-weight $\bar{\chi}_2^2$ is off the central line in Figure 1.9. We can also see p-value comparisons in Figure 1.10, where least favorable $\bar{\chi}^2$ resembles the theoretical p-value distribution the most.

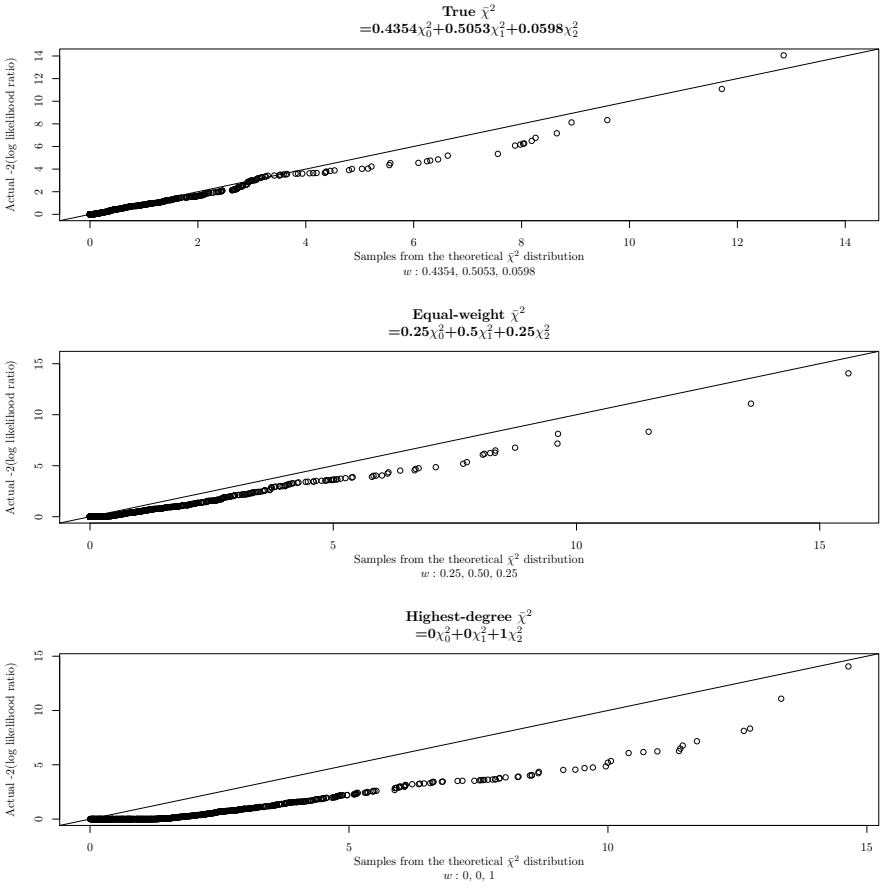
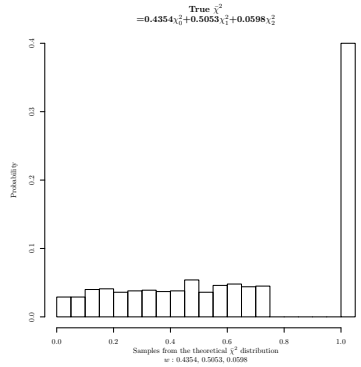
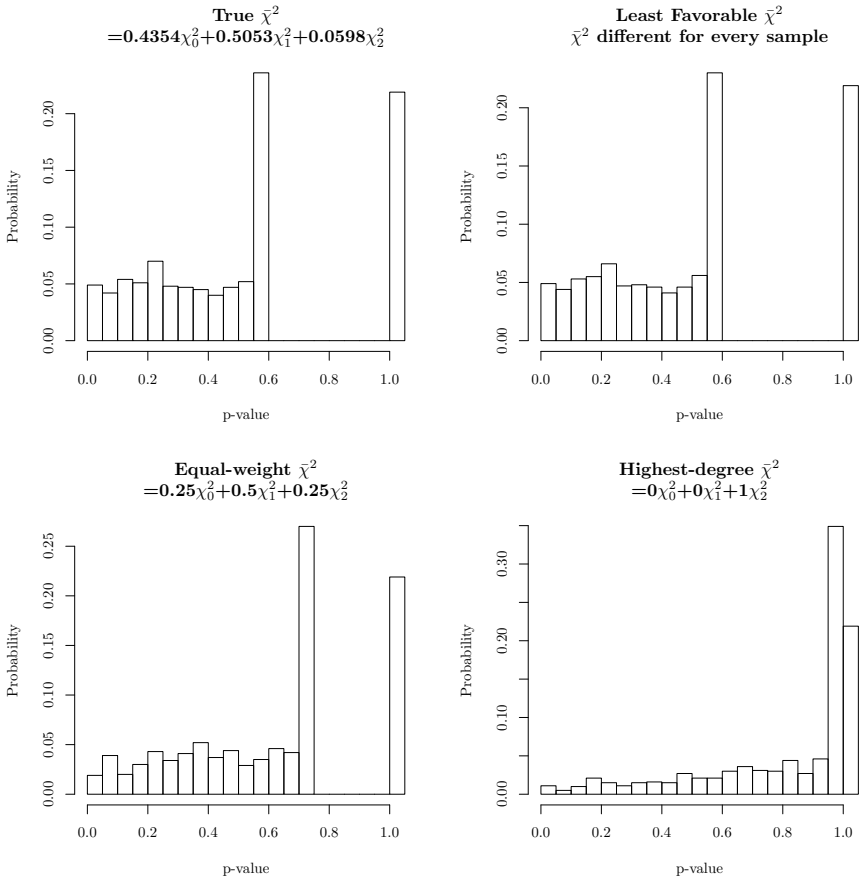


Figure 1.9: K=100, R=1000. qq-plot of LRT statistic: under true $\bar{\chi}_2^2$, equal-weight $\bar{\chi}_2^2$ and χ_2^2



(a) Histogram of pvalues of direct samples from true $\bar{\chi}^2_2$



(b) K=100, R=1000. histogram of pvalues of LRT statistic: under true $\bar{\chi}^2_2$, closest $\bar{\chi}^2_2$, equal-weight $\bar{\chi}^2_2$ and χ^2_2

Figure 1.10: Histogram of pvalues

1.3.2 Simulation Study of grMATS

We illustrate the performance of grMATS under 30 different scenarios of data of two groups.

Set-up:

The first 27 simulation datasets differ in terms of number of replicates ($K = 5, 10, \text{or } 20$), number of average total read counts ($\underline{R} = 20, 80, \text{or } 250$), and replicate variance ($\underline{\sigma}^2 = 0.07, 0.36, \text{or } 1.00$). The total read counts and replicate variances are respectively the 1st quartile, median, 3rd quartile of the average replicate total read counts of MAQC (14267 genes) and its corresponding estimated variances.

Besides the 27 simulations, we conduct additional 3 simulations of two groups with replicate $K = 5$, variance ($\underline{\sigma}^2 = 0.07, 0.36, \text{or } 1.00$), and the total read counts are simple random samples from the average replicate total read counts of MAQC.

For each simulation above, 5000 genes are simulated, of which 95% are generated under null hypothesis ($H_0 : |\underline{\psi}_{n,1,I_f} - \underline{\psi}_{n,2,I_f}| \leq \Delta\underline{\psi}, \forall 1 \leq f \leq F$) and 5% are under alternative hypothesis ($H_1 : \exists f, 1 \leq f \leq F, |\underline{\psi}_{n,1,I_f} - \underline{\psi}_{n,2,I_f}| > \Delta\underline{\psi}$). Isoform proportion difference is chosen as $\Delta\underline{\psi} = 10\%$.

Without any preference of true parameter distributions, we use flat Dirichlet distribution to uniformly simulate isoform proportions of two groups $\{\underline{\psi}_{g,\{I_f\}} : \underline{\psi}_{g,I_1} + \cdots + \underline{\psi}_{g,I_F} = 1, \underline{\psi}_{g,I_f} \geq 0, \forall 1 \leq f \leq F\}$. For each iteration, we independently sample two groups of isoform proportions, $\underline{\psi}_{1,\{I_f\}}$ and $\underline{\psi}_{2,\{I_f\}}$, if $|\underline{\psi}_{1,I_f} - \underline{\psi}_{2,I_f}| \leq \Delta\underline{\psi}, \forall f$, we assign them to one gene under H_0, otherwise we assign them as one gene under H_1. Repeat this process until enough number of genes under H_0 and H_1 are generated.

For every gene, the true isoform proportions $\underline{\psi}_{g,\{I_f\}}$ are converted to the logit scale $\underline{\mu}_{g,\{I_f\}} = \text{mlogit}(\underline{\psi}_{g,\{I_f\}})$. Independently K sets of replicate logit values are sampled $1 \leq k \leq K$, $\mu_{g,k,I_f} \sim N(\underline{\mu}_{g,I_f}, \underline{\sigma}^2), 1 \leq f \leq F-1$. $\psi_{g,k,\{I_f\}} = \text{mlogit}^{-1}(\mu_{g,k,\{I_f\}})$. A multinomial read counts vector $R_{g,k,\{I_f\}}$ is sampled from $MN(\underline{R}, \psi_{g,k,\{I_f\}})$ for each replicate k of each group g.

As a comparison, we pooled data from replicates and analyzed the pooled data using a reduced version of grMATS that used the same likelihood-ratio test with composite hypotheses H_0, H_1 as above.

Comparisons:

- ROC & PR. TPR, FPR.
- I_TPR, I_TNR. We pull all isoforms together to compute their isoform-level TPR and TNR. The significant isoforms of significant genes are detected as positive. The insignificant isoforms of significant genes are detected as negative. The isoforms of insignificant genes are detected as negative.

Results:

In all 27 simulations, grMATS outperformed the reduced (pooling all replicates) version.

Here we describe results for 3 simulations with 5 replicates, fixed total read counts 80, at 5% false positive rate, grMATS produced true positive rates 93.6% ($\underline{\sigma}^2 = 0.07$), 87.6% ($\underline{\sigma}^2 = 0.36$), 72.4% ($\underline{\sigma}^2 = 1$), while correspondingly the pooled model only had 86.4%, 68.4%, 50.4%. The true positive rate drop was more obvious as $\underline{\sigma}^2$ gets bigger. Thus we want to point out that the use of random effect in the model is crucial especially in studies with large between-replicate variation.

	σ^2	N.g	N.k.g	N.cts	I_1_TPR	I_2_TPR	I_3_TPR	I_4_TPR
④	0.07	2	(5, 5)	80	74.0%,134/181	80.1%,141/176	73.7%,123/167	83.0%,142/171
⑤	0.36	2	(5, 5)	80	62.9%,117/186	65.0%,106/163	58.1%,100/172	69.0%,120/174
⑥	1	2	(5, 5)	80	40.6%,71/175	35.3%,61/173	41.8%,76/182	55.4%,102/184

Table 1.6: Isoform Detection TPR

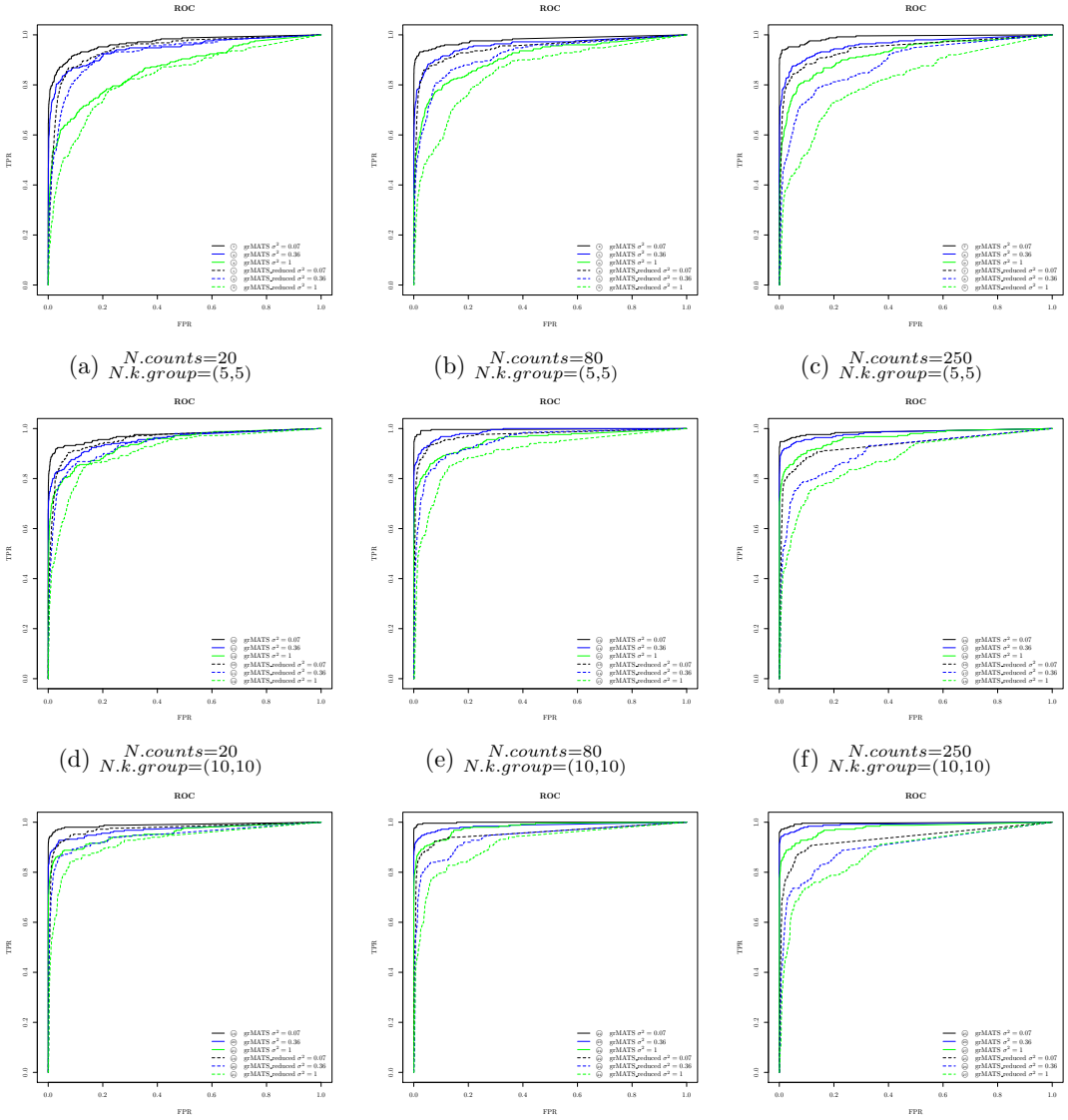
	σ^2	N.g	N.k.g	N.cts	I_1_TNR	I_2_TNR	I_3_TNR	I_4_TNR
④	0.07	2	(5, 5)	80	98.8%,4762/4819	99.3%,4789/4824	99.3%,4799/4833	99.3%,4796/4829
⑤	0.36	2	(5, 5)	80	98.3%,4732/4814	98.4%,4761/4837	98.3%,4746/4828	99.1%,4784/4826
⑥	1	2	(5, 5)	80	98.1%,4734/4825	98.3%,4743/4827	98.3%,4737/4818	98.9%,4764/4816

Table 1.7: Isoform Detection TNR

The I_TPR are $77.7\%, 63.7\%, 43.4\%$ and the I_TNR are $99.2\%, 98.5\%, 98.4\%$.

Besides the above scores, we show for these 3 simulations, the recovery of each individual isoform in Table 1.6 for TPR, Table 1.7 for TNR. It is a true negative if $|\underline{\psi}_{n,1,I_f} - \underline{\psi}_{n,2,I_f}| \leq \Delta\underline{\psi}$, and $|\hat{\underline{\psi}}_{n,1,I_f} - \hat{\underline{\psi}}_{n,2,I_f}| \leq \Delta\underline{\psi}$, and a true positive if $\underline{\psi}_{n,1,I_f} - \underline{\psi}_{n,2,I_f} > \Delta\underline{\psi}$ (or $\underline{\psi}_{n,1,I_f} - \underline{\psi}_{n,2,I_f} < -\Delta\underline{\psi}$), and $\hat{\underline{\psi}}_{n,1,I_f} - \hat{\underline{\psi}}_{n,2,I_f} > \Delta\underline{\psi}$ (or $\hat{\underline{\psi}}_{n,1,I_f} - \hat{\underline{\psi}}_{n,2,I_f} < -\Delta\underline{\psi}$).

In addition to the fixed total read counts simulations, we also performed 3 additional simulations with a similar set-up except that instead of fixed quartiles, the total read counts were empirically sampled from average total read counts of all multiple-isoform genes of MAQC. At 5% false positive rate, grMATS produced true positive rates 94.0% ($\underline{\sigma}^2 = 0.07$), 84.0% ($\underline{\sigma}^2 = 0.36$), 66.8% ($\underline{\sigma}^2 = 1$), in comparison with the pooled model 74.0%, 44.8%, 26.4%. The I_TPR are 76.4%, 62.7%, 44.7% and I_TNR are 98.8%, 98.4%, 98.2%. Figure 1.14a, 1.14b for ROC and PR, Table 1.14 for summary of results.

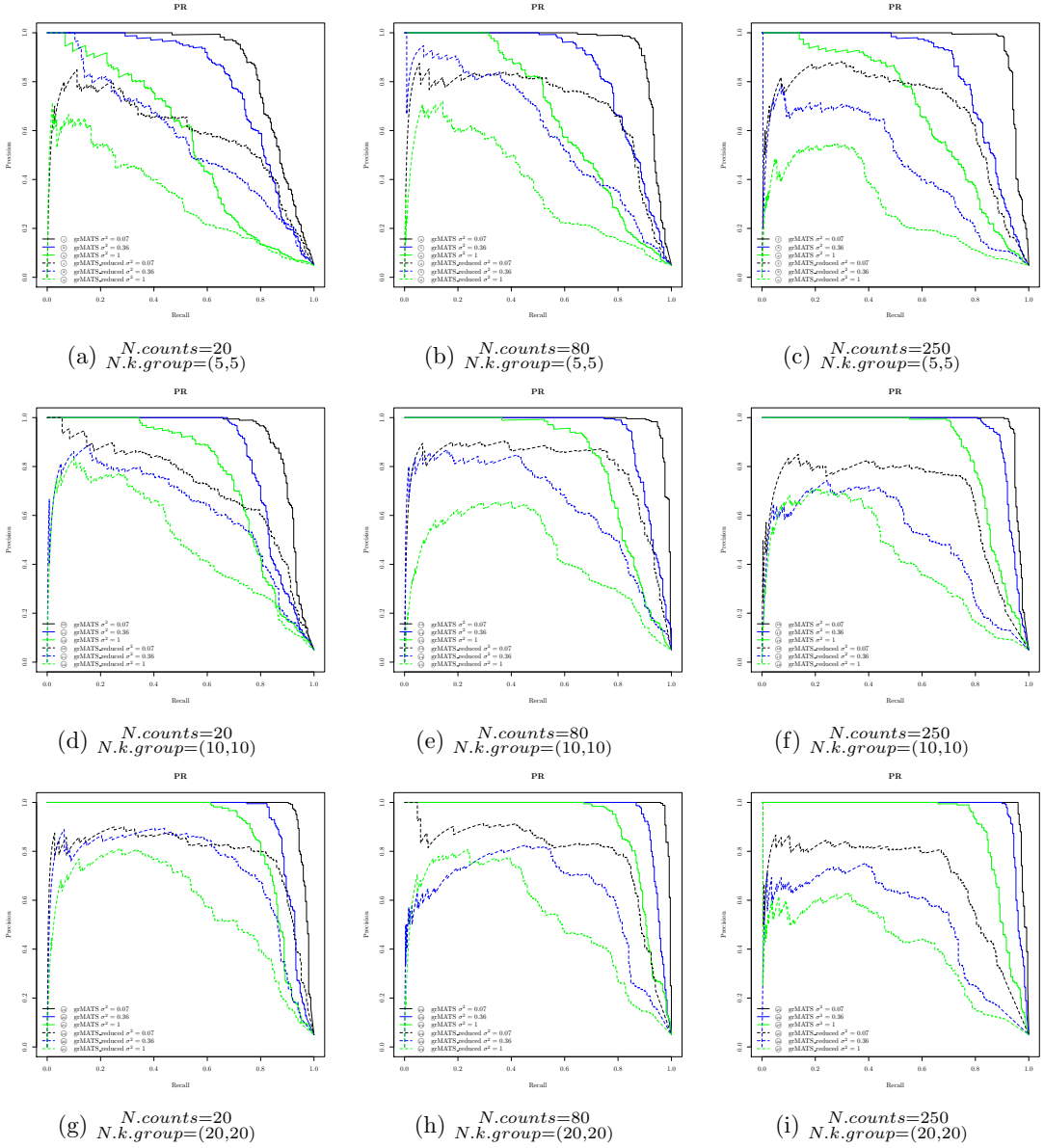


(a) $\substack{N.counts=20 \\ N.k.group=(5,5)}$ (b) $\substack{N.counts=80 \\ N.k.group=(5,5)}$ (c) $\substack{N.counts=250 \\ N.k.group=(5,5)}$

(d) $\substack{N.counts=20 \\ N.k.group=(10,10)}$ (e) $\substack{N.counts=80 \\ N.k.group=(10,10)}$ (f) $\substack{N.counts=250 \\ N.k.group=(10,10)}$

(g) $\substack{N.counts=20 \\ N.k.group=(20,20)}$ (h) $\substack{N.counts=80 \\ N.k.group=(20,20)}$ (i) $\substack{N.counts=250 \\ N.k.group=(20,20)}$

Figure 1.11: ROC of 27 simulations.



(a) $\substack{N.counts=20 \\ N.k.group=(5,5)}$ (b) $\substack{N.counts=80 \\ N.k.group=(5,5)}$ (c) $\substack{N.counts=250 \\ N.k.group=(5,5)}$

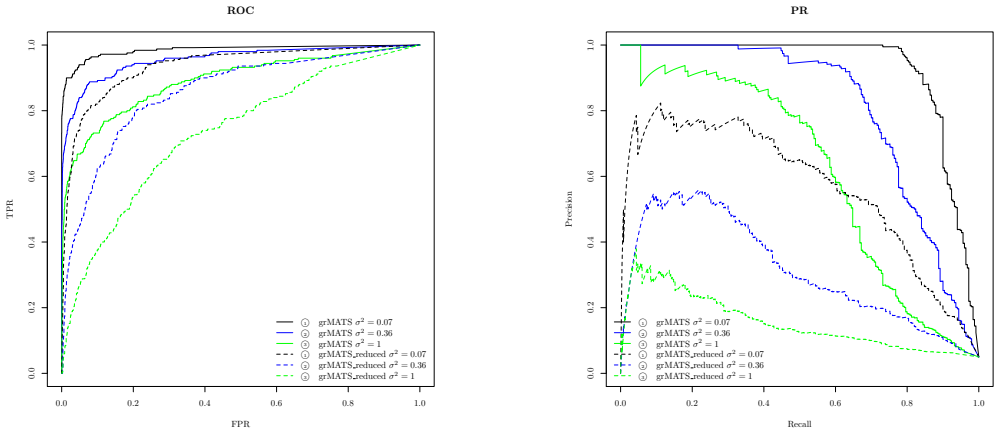
(d) $\substack{N.counts=20 \\ N.k.group=(10,10)}$ (e) $\substack{N.counts=80 \\ N.k.group=(10,10)}$ (f) $\substack{N.counts=250 \\ N.k.group=(10,10)}$

(g) $\substack{N.counts=20 \\ N.k.group=(20,20)}$ (h) $\substack{N.counts=80 \\ N.k.group=(20,20)}$ (i) $\substack{N.counts=250 \\ N.k.group=(20,20)}$

Figure 1.12: PR of 27 simulations.

	σ^2	N.g	N.k.g	N.cts	@FPR	TPR	AUC	I_TPR	I_TNR	@FPR_R	TPR_R	AUC_R
①	0.07	2	(5, 5)	20	5.01%	87.2%	0.968	63.6%,443/696	98.9%,19084/19304	5.01%	80.8%	0.945
②	0.36	2	(5, 5)	20	5.01%	82.4%	0.944	54.6%,394/721	98.6%,19010/19279	5.01%	70.4%	0.927
③	1	2	(5, 5)	20	5.01%	62.4%	0.865	39.4%,278/705	98.3%,18974/19295	5.01%	48.4%	0.832
④	0.07	2	(5, 5)	80	5.01%	93.6%	0.980	77.7%,540/695	99.2%,19146/19305	5.01%	86.4%	0.949
⑤	0.36	2	(5, 5)	80	5.01%	87.6%	0.960	63.7%,443/695	98.5%,19023/19305	5.01%	68.4%	0.920
⑥	1	2	(5, 5)	80	5.01%	72.4%	0.910	43.4%,310/714	98.4%,18978/19286	5.01%	50.4%	0.855
⑦	0.07	2	(5, 5)	250	5.01%	95.2%	0.991	82.7%,574/694	99.2%,19152/19306	5.01%	84.0%	0.945
⑧	0.36	2	(5, 5)	250	5.01%	87.6%	0.961	64.8%,448/691	98.5%,19026/19309	5.01%	61.2%	0.887
⑨	1	2	(5, 5)	250	5.01%	75.6%	0.924	47.7%,341/715	98.4%,18975/19285	5.01%	44.4%	0.815
⑩	0.07	2	(10, 10)	20	5.01%	92.4%	0.973	71.0%,487/686	99.2%,19156/19314	5.01%	86.4%	0.957
⑪	0.36	2	(10, 10)	20	5.01%	83.2%	0.954	64.8%,459/708	98.7%,19038/19292	5.01%	79.2%	0.939
⑫	1	2	(10, 10)	20	5.01%	78.4%	0.940	50.6%,353/698	98.6%,19026/19302	5.01%	61.2%	0.909
⑬	0.07	2	(10, 10)	80	5.01%	99.2%	0.997	85.4%,607/711	99.4%,19169/19289	5.01%	92.0%	0.973
⑭	0.36	2	(10, 10)	80	5.01%	93.2%	0.988	74.0%,510/689	98.9%,19091/19311	5.01%	81.2%	0.950
⑮	1	2	(10, 10)	80	5.01%	83.6%	0.954	55.6%,391/703	98.6%,19020/19297	5.01%	62.4%	0.906
⑯	0.07	2	(10, 10)	250	5.01%	96.4%	0.987	86.0%,608/707	99.5%,19198/19293	5.01%	82.8%	0.934
⑰	0.36	2	(10, 10)	250	5.01%	94.0%	0.982	73.4%,520/708	99.1%,19122/19292	5.01%	73.6%	0.907
⑱	1	2	(10, 10)	250	5.01%	85.6%	0.965	59.3%,409/690	98.6%,19046/19310	5.01%	58.4%	0.875
⑲	0.07	2	(20, 20)	20	5.01%	97.6%	0.990	83.8%,589/703	99.4%,19186/19297	5.01%	92.4%	0.973
⑳	0.36	2	(20, 20)	20	5.01%	92.8%	0.971	73.0%,498/682	99.1%,19150/19318	5.01%	86.8%	0.946
㉑	1	2	(20, 20)	20	5.01%	87.2%	0.955	63.1%,434/688	98.8%,19077/19312	5.01%	77.6%	0.924
㉒	0.07	2	(20, 20)	80	5.01%	99.6%	0.999	87.7%,614/700	99.6%,19230/19300	5.01%	88.8%	0.956
㉓	0.36	2	(20, 20)	80	5.01%	95.6%	0.986	79.5%,542/682	99.2%,19161/19318	5.01%	82.0%	0.940
㉔	1	2	(20, 20)	80	5.01%	90.8%	0.980	68.2%,471/691	98.8%,19084/19309	5.01%	70.4%	0.912
㉕	0.07	2	(20, 20)	250	5.01%	98.4%	0.997	87.4%,598/684	99.7%,19262/19316	5.01%	82.8%	0.936
㉖	0.36	2	(20, 20)	250	5.01%	96.0%	0.991	82.0%,584/712	99.3%,19150/19288	5.01%	73.6%	0.899
㉗	1	2	(20, 20)	250	5.01%	90.4%	0.976	67.8%,464/684	98.9%,19104/19316	5.01%	64.8%	0.878

Figure 1.13: Results of 27 simulations.



(a) ROC of 3 simulations with changing total read counts.

(b) PR of 3 simulations with changing total read counts.

	σ^2	N.g	N.k.g	N.cts	@FPR	TPR	AUC	I_TPR	I_TNR	@FPR_R	TPR_R	AUC_R
①	0.07	2	(5, 5)	changing	5.01%	94.0%	0.985	76.4%,519/679	98.8%,19087/19321	5.01%	74.0%	0.936
②	0.36	2	(5, 5)	changing	5.01%	84.0%	0.956	62.7%,432/689	98.4%,18995/19311	5.01%	44.8%	0.856
③	1	2	(5, 5)	changing	5.01%	66.8%	0.891	44.7%,298/666	98.2%,18990/19334	5.01%	26.4%	0.732

Figure 1.14: Results of 3 simulates with changing total read counts.

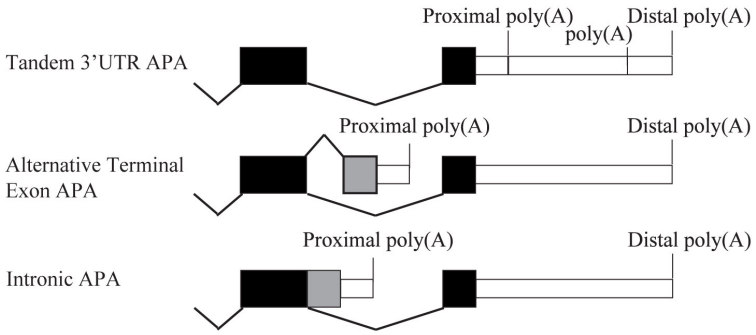


Figure 1.15: Hypoxia - APA sites

1.4 Real Data Application - Hypoxia

In the 3' end processing of most protein-coding genes, the 3' end of the mRNAs is cleaved and polyadenylated. The addition of the poly(A) tail is required for nuclear export, mRNA stability and efficient translation [Sac90]. A large proportion of genes contain multiple polyadenylation sites [Shi12], indicating that alternative polyadenylation (APA) is a widely used mechanism for gene regulation. APA sites can be classified into four categories (Figure 1.15): tandem 3' UTR (untranslated region) APA, the most frequent APA forms with multiple cleavage sites in 3' UTRs; alternative terminal exon APA, which involves usage of multiple terminal exons; and intronic APA. Through these types of events, APA contributes to the complexity of the gene expression by generating multiple mRNA forms that differ in cellular localization, stability and translation efficiency.

Widespread APA modulation is often associated with development, cellular differentiation and proliferation. A wide variety of APA events were observed in embryonic development and neuronal development [JLP+09]. The generation of iPSC (induced pluripotent stem cells) is often accompanied by global shift of poly(A) tails [JT09]. Moreover, in the T cell activation, widespread usage of proximal poly(A) sites was also observed [SNS+08]. Since pathways of cell differentiation and proliferation are often hijacked in cancers, the APA events are

often observed in cancer cells. It has been reported that compared to normal cells, cancer cells often expressed a variety of mRNAs with shorter 3' UTR from APA events [MB09]. The shorter mRNA forms exhibiting enhanced translational efficiency and stability, leading to more protein production. The high prevalence of APA in cancer cells suggests a role of APA in cancer development.

Hypoxia condition is often associated with tumor development [WH11]. Cancer cells usually use a shifted metabolic process from oxidative phosphorylation to altered glycolysis. The shifted metabolic process plays a central role in the development of solid tumors since it provides the necessary energy for tumor development. A large amount of the mRNA isoform changes have been observed in the hypoxia process, contributing to the hypoxia pathways or being results of the metabolic shifts Weigand, 2012 #1813. Here we use the PolyA-seq technique Derti, 2012 #1819, a stand-specific high-throughput sequencing analysis of 3' ends of polyadenylated transcripts, to conduct genome-wide analysis of APA events in hypoxia conditions compared to normal cell lines. A total of 3 replicates were generated under 2% oxygen chamber to represent the hypoxia condition, as a comparison to 3 controls under 20% oxygen condition.

Using the grMATS framework, we have identified a variety of APA site shifts between the hypoxia condition and normal condition. With FDR $\leq$ 30% and isoform ratio difference $>$ 1%, grMATS identified 671 genes with significant APA site shifts. We studied the gene functional enrichment among the 671 genes with significant APA shifts using DAVID (Supplemental Table Hypoxia). A variety of cancer related biological processes were enriched in the genes with APA shifts, such as 'DNA repair' (DAVID enrichment P = 2.5e-5), 'negative regulation of cell growth' (DAVID enrichment P = 2.8e-4), 'cellular response to stress' (DAVID enrichment P = 4.9e-4).

1.5 Summary

We propose a hierarchical model for RNA-seq alternative splicing data with replicates. In the first layer, the model uses gaussian logit values $\mu_{n,g,k,\{I_f\}}$ to describe biological replicate effect centered around the group-level isoform proportion logits $\mu_{n,g,\{I_f\}}$. In the second layer, a multinomial distribution $R_{n,g,k,\{I_f\}}$ is assumed to describe the replicate read counts given $\mu_{n,g,k,\{I_f\}}(\psi_{n,g,k,\{I_f\}})$. We compute the marginal likelihood of $P(R_{n,g,k,\{I_f\}})$ using laplace approximation. Besides the maximum likelihood estimation, we also provide the accurate asymptotic distribution of composite likelihood ratio test D, which follows a mixture of χ^2 with various degree of freedoms up to the number of constraints. We provide the least favorable test statistic in practice when true parameter is unknown. The model and program have been successfully run in our 27+3 simulation datasets as well as one real dataset Hypoxia.

Our work provides a R package *grMATS*.

1.6 Future Work

In this model, we assume that we observe the isoform read counts $R_{\{I_f\}}$ directly. However, in many other situations, what we observed are the read patterns from isoforms and different isoforms share some patterns (Figure 1.16). To address this ambiguity, we need another layer in the model. Let E_h denote the read pattern h, $\gamma_{E_h,I_f} = P(E = E_h|I = I_f)$ denote probability of read pattern h from isoform f, R_{E_h} as the read counts of these patterns. The marginally probability of observing read pattern h is

$$\phi_{E_h} = \sum_f P(E = E_h|I = I_f) * P(I = I_f) = \sum_f \gamma_{E_h,I_f}\psi_{I_f} \tag{1.25}$$

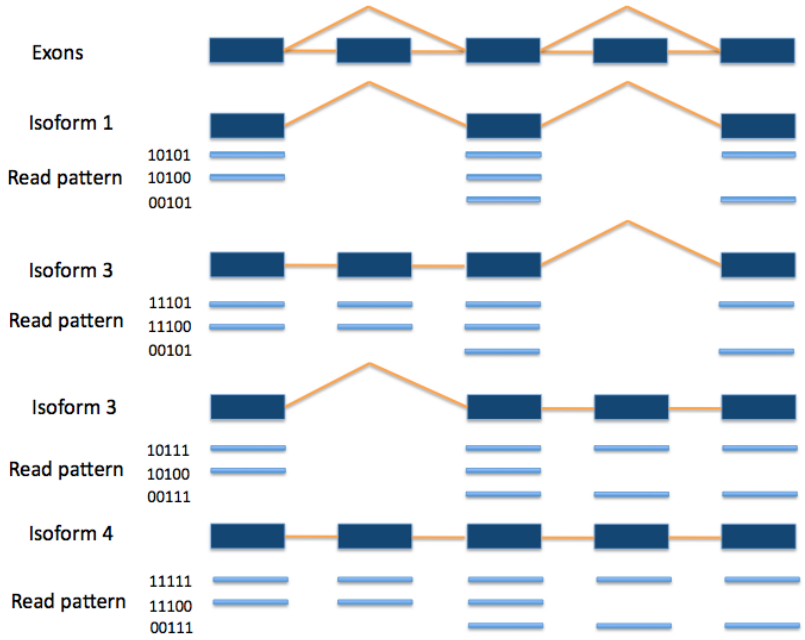


Figure 1.16: Example of alternative splicing with reading ambiguity. 4 isoforms, 8 possible read patterns E_h: 10101, 10100, 00101, 11101, 11100, 10111, 00111, 11111.

Conditional probability,

$$P(R_{\{E_h\}}|\psi_{\{I_f\}};\theta) = \prod_{h=1}^{H} \phi_{E_h}^{R_{E_h}} \tag{1.26}$$

Our likelihood function becomes

$$\begin{aligned} P(R;\theta) &= \int P(R|\psi_{\{I_f\}};\theta)P(\psi_{\{I_f\}};\theta)d\psi_{\{I_f\}} \\ &= \int \prod_{h=1}^{H} \phi_{E_h}^{R_{E_h}} P(\psi_{\{I_f\}};\theta)d\psi_{\{I_f\}} \\ &= \int \prod_{h=1}^{H} (\sum_{f} \gamma_{E_h,I_f}\psi_{I_f})^{R_{E_h}} P(\psi_{\{I_f\}};\theta)d\psi_{\{I_f\}} \end{aligned} \tag{1.27}$$

Difficulty might arise while computing this integral using Laplace approximation, since the convexity of this joint density needs further investigation.

1.7 Proof of Theorems

1.7.1 Logit Transformation

Logit Transformation between $\underline{\psi}_{\{I_f\}}$ *and* $\underline{p}_{k,\{I_f\}}$

The same transformation holds for r.v.s $\psi_{\{I_f\}}$ and $p_{k,\{I_f\}}$. $\sum_{f=1}^{F} \psi_{I_f} = 1$ are probabilities adding up to 1, with degree of freedom $F-1$.

Multinomial or categorical distributions belong to exponential families, it's natural to link the first $F-1$ ψ_{I_f}'s with $F-1$ multinomial logit values $\underline{\mu}_{I_f}$'s, while fixing $\underline{\mu}_{I_F} = 0$. $-\infty < \underline{\mu}_{I_f} < \infty, \quad 1 \le f \le F-1$.

$$\log(\frac{\underline{\psi}_{I_f}}{\underline{\psi}_{I_F}}) = \underline{\mu}_{I_f} \iff \underline{\psi}_{I_f} = \frac{e^{\underline{\mu}_{I_f}}}{1+\sum_{f=1}^{F-1} e^{\underline{\mu}_{I_f}}} \quad 1 \le f \le F \tag{1.28}$$

$$\{\underline{\psi}_{I_1}, \cdots, \underline{\psi}_{I_F}\} \longleftrightarrow \{\underline{\mu}_{I_1}, \cdots, \underline{\mu}_{I_{F-1}}, 0\} \tag{1.29}$$

Considering the isoform length $\underline{l}_{\{I_f\}}$, the actual observable multinomial probabilities,

$$\underline{p}_{I_f} = \frac{\underline{l}_{I_f}\underline{\psi}_{I_f}}{\sum_{f=1}^{F}\underline{l}_{I_f}\underline{\psi}_{I_f}} = \frac{\underline{l}_{I_f}e^{\underline{\mu}_{I_f}}}{\underline{l}_{I_F}+\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\underline{\mu}_{I_f}}}, \quad 1 \le f \le F \tag{1.30}$$

$$\{\underline{p}_{I_1}, \cdots, \underline{p}_{I_F}\} \overset{l_{\{I_f\}}}{\longleftrightarrow} \{\underline{\mu}_{I_1}, \cdots, \underline{\mu}_{I_{F-1}}, 0\} \tag{1.31}$$

Relationships between $\underline{\psi}_{n,g,\{I_f\}}$ and $\underline{\mu}_{n,g,\{I_f\}}$ ($\underline{p}_{n,g,k,\{I_f\}}$ and $\underline{\mu}_{n,g,k,\{I_f\}}$)
Assume $\underline{l}_{I_f} = 1, 1 \le f \le F$, then $\underline{p}_{I_f} = \underline{\psi}_{I_f}$, $R_{\{I_f\}} \sim MN(R, \underline{p}_{\{I_f\}} = \underline{\psi}_{\{I_f\}})$, as simple multinomial probability without replicates.

Multinomial distribution belongs to exponential family, we can write the probability as

$$\begin{aligned} P(R_{I_1}, \cdots, R_{I_F}; \underline{\psi}_{\{I_f\}}) &= \frac{R!}{\prod_{f=1}^{F} R_{I_f}!}\prod_{f=1}^{F}\underline{\psi}_f^{R_{I_f}} \\ &= \frac{R!}{\prod_{f=1}^{F} R_{I_f}!}(\prod_{f=1}^{F-1}\underline{\psi}_f^{R_{I_f}})(\underline{\psi}_{I_F})^{R-\sum_{f=1}^{F-1}R_{I_f}} \\ &= \exp\left([\sum_{f=1}^{F-1} R_{I_f}\log(\frac{\underline{\psi}_{I_f}}{\underline{\psi}_{I_F}})] + R\log(\underline{\psi}_{I_F}) + \log(\frac{R!}{\prod_{f=1}^{F} R_{I_f}!})\right) \end{aligned}$$

A very natural link function choice is (plug in n, g notation here)

$$\underline{\mu}_{n,g,I_f} = \begin{cases} \log(\frac{\underline{\psi}_{n,g,I_f}}{\underline{\psi}_{n,g,I_{F_n}}}) & 1 \le f \le F_n - 1 \\ 0 & f = F_n \end{cases} \tag{1.32}$$

$$\Updownarrow$$

$$\underline{\psi}_{n,g,I_f} = \begin{cases} \frac{e^{\underline{\mu}_{n,g,I_f}}}{1+\sum_{f=1}^{F_n-1} e^{\underline{\mu}_{n,g,I_f}}} & 1 \le f \le F_n - 1 \\ \frac{1}{1+\sum_{f=1}^{F_n-1} e^{\underline{\mu}_{n,g,I_f}}} & f = F_n \end{cases} \tag{1.33}$$

When isoform lengths $\underline{l}_{\{I_f\}}$ are different, $\underline{p}_{\{I_f\}}$ become weighted $\underline{\psi}_{\{I_f\}}$ as in (1.30).

1.7.2 More on Theorem 1 - Laplace Approximation

$$
\begin{aligned}
&\int_{\mathbb{R}^{F-1}} \exp(h(\mu))d\mu \\
&= (2\pi)^{(F-1)/2}|V|^{1/2}\exp(h(\hat{\mu}))E\left(\exp(\sum_{i=3}^{\infty}\frac{1}{i!}((\overset{i-1}{\otimes}(\mu-\hat{\mu})^T)h^{(i)}(\hat{\mu})(\mu-\hat{\mu})))\right) \\
&= (2\pi)^{(F-1)/2}|V|^{1/2}\exp(h(\hat{\mu}))E\left(\exp(\sum_{i=3}^{\infty}T_i)\right) \\
&= (2\pi)^{(F-1)/2}|V|^{1/2}\exp(h(\hat{\mu}))E\left(\exp(S)\right) \qquad (1.34)
\end{aligned}
$$

where $\hat{\mu} = \arg\max_{\mu}\exp(h(\mu)) = \arg\max_{\mu} h(\mu)$,

$$V = (-h^{(2)}(\hat{\mu}))^{-1}: \quad \dim\ (F-1)\times(F-1) \qquad (1.35)$$

$$h^{(k)}(\hat{\mu}) = \frac{\partial vec(h^{(k-1)}(\mu))}{\partial \mu^T}|_{\mu=\hat{\mu}} \qquad (1.36)$$

$vec()$ means reorganize the matrix into a vector by its columns.
$\otimes$ means Kronecker product, $\overset{i-1}{\otimes}(\mu-\hat{\mu})^T = (\mu-\hat{\mu})^T\otimes(\mu-\hat{\mu})^T\otimes\cdots(\mu-\hat{\mu})^T$, $i-1$ times.

$$A\otimes B = \begin{bmatrix} \mu_{11}B & \cdots & \mu_{1n}B \\ \vdots & \ddots & \vdots \\ \mu_{m1}B & \cdots & \mu_{mn}B \end{bmatrix} \qquad (1.37)$$

$E(\exp(S))$ is expectation on a function $\exp(S)$ where $S=\sum_{i=3}^{\infty}T_i$ involving only third or higher derivative of $h(\mu)$ at $\hat{\mu}$. (i.e. $\frac{\partial^3 h(\mu)}{\partial\mu_i\partial\mu_j\partial\mu_k}$, ...). Although third or higher order derivatives of $h(\mu)$ do not involve any parameters of prior distribution of μ, the maximizer $\hat{\mu}$ involves these parameters, so $E(\exp(S))$ involves all parameters. However, in the generalized linear model with random effects, as the cluster size gets bigger (here is the total read counts $\underline{R}$), $E(\exp(S))\approx 1$.

$$
\begin{aligned}
E(\exp(S)) &\approx 1 + E(T_4) + E(T_6) + \frac{1}{2}E(T_3^2) \\
&= 1 + O(R^{-1}) + O(R^{-2}) + O(R^{-1}) = 1 + O(R^{-1})
\end{aligned}
$$

We see as cluster size (total read counts) gets bigger, the $E(\exp(S)) \approx 1$. The log of our integral objective can be approximated by sum of linear additive terms without integrals, which is much easier to handle.

$$\int_{\mathbb{R}^{F-1}} \exp(h(\mu))d\mu \approx (2\pi)^{(F-1)/2}|V|^{1/2}\exp(h(\hat{\mu})) \tag{1.38}$$

As in (1.34), C is related to replicate counts $R_{\{I_f\}}$ and isoform lengths $\underline{l}_{\{I_f\}}$, and also third or higher derivatives of $h(\hat{\mu}_{\{I_f\}})$, which contains $\hat{\mu}_{\{I_f\}}$ and it relates to parameters $\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}$. Thus, C is related to unknown parameters of interest $\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}$, but when cluster size R is big enough, C is approximately 1. We consider R is big enough and optimize the likelihood over parameter $\underline{\mu}_{\{I_f\}}$, $\underline{\sigma}^2_{\{I_f\}}$, with $C \approx 1$.

1.7.3 Proof of Theorem 2

To prove the theorem, we mainly need to work on the constrained and unrestricted maximum likelihood estimation and cone approximation on the constrained space.

Log-likelihood $l(x;\theta)$

$X = [X_1, \cdots, X_n]^T$ are random observations, $X_i \sim f(x,\theta)$ and θ are parameters, θ_0 are true parameters.

$$\begin{aligned} l(X,\theta) &= \sum_{k=1}^{n} l(X_k,\theta) \\ &= l(X,\theta_0) + l'(X,\theta_0)^T(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^T l''(X,\theta_0)(\theta-\theta_0) + o(||\theta-\theta_0||_2^2) \end{aligned} \tag{1.39}$$

Let $A = \frac{1}{n}l'(X,\theta_0)$ be a vector where

$$A_i = \frac{1}{n}\frac{\partial l(X,\theta)}{\partial \theta_i}\Big|_{\theta=\theta_0} = \frac{1}{n}\sum_{k=1}^{n}\frac{\partial \log f(X_k,\theta)}{\partial \theta_i}\Big|_{\theta=\theta_0} \tag{1.40}$$

$B = \frac{1}{n} l''(X, \theta_0)$ be a matrix where

$$B_{i,j} = \frac{1}{n} \frac{\partial^2 l(X,\theta)}{\partial\theta_i \partial\theta_j}\Big|_{\theta=\theta_0} = \frac{1}{n}\sum_{k=1}^{n} \frac{\partial^2 \log f(X_k,\theta)}{\partial\theta_i \partial\theta_j}\Big|_{\theta=\theta_0} \tag{1.41}$$

$$E(B) = -\mathcal{I} = \Big[E(\frac{\partial^2 \log f(X,\theta)}{\partial\theta_i \partial\theta_j})\Big] \tag{1.42}$$

$$E(A) = 0, \quad Var(\sqrt{n}A) = \mathcal{I} \tag{1.43}$$

for details on calculation of $\mathcal{I}$ see appendix ***Fisher Information***. Rewrite

$$l(X,\theta) = l(X,\theta_0) + nA^T(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^T nB(\theta-\theta_0) + O_p(n)[(\theta-\theta_0)^T(\theta-\theta_0)]^{\frac{3}{2}} \tag{1.44}$$

Let $\hat{\theta}$ denote the MLE estimator for θ, vector θ^2 denote element-wise square of θ. Suppose $||\theta - \theta_0||_2$ is bounded by a finite number.

Under regularity conditions (P121 [Fer96]), the MLE estimator:

$$\hat{\theta} = \arg\max_{\theta} l(X,\theta) \tag{1.45}$$

$$\Rightarrow \hat{\theta} - \theta_0 = -B^{-1}A + o_p(\frac{1}{\sqrt{n}}) \tag{1.46}$$

$$-B \xrightarrow{P} \mathcal{I}, \quad -B^{-1} = \mathcal{I}^{-1} + o_p(1) \quad A = O_p(\frac{1}{\sqrt{n}}) \tag{1.47}$$

$$\hat{\theta} - \theta_0 = \mathcal{I}^{-1}A + o_p(\frac{1}{\sqrt{n}}) \tag{1.48}$$

By [Che54], if θ_0 is a limit point of parameter space, then for any estimator in this space, $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{\theta} - \theta_0 = O_p(\frac{1}{\sqrt{n}})$. For any MLE estimate of θ, we let

$$\hat{\theta} - \theta_0 = \mathcal{I}^{-1}A + \eta, \quad \eta = O_p(\frac{1}{\sqrt{n}}) \tag{1.49}$$

$$\begin{aligned}
&l(X,\theta) \\
&= n\left\{\frac{1}{n}l(X,\theta_0) + A^T(\theta-\theta_0) + \frac{1}{2}(\theta-\theta_0)^T B(\theta-\theta_0) + O_p(1)[(\theta-\theta_0)^T(\theta-\theta_0)]^{\frac{3}{2}}\right\} \\
&= n\left\{\frac{1}{n}l(X,\theta_0) + A^T\mathcal{I}^{-1}A + A^T\eta - \frac{1}{2}(\mathcal{I}^{-1}A+\eta)^T\mathcal{I}(\mathcal{I}^{-1}A+\eta) + O_p(n^{-\frac{3}{2}})\right\} \\
&= n\left\{\frac{1}{n}l(X,\theta_0) + \frac{1}{2}A^T\mathcal{I}^{-1}A - \frac{1}{2}\eta^T\mathcal{I}\eta + O_p(n^{-\frac{3}{2}})\right\}
\end{aligned} \tag{1.50}$$

Let $z = \sqrt{n}\mathcal{I}^{-1}A$

$$z \xrightarrow{L} N(0, \mathcal{I}^{-1}), \quad o_p(||z||^2) = o_p(1) \tag{1.51}$$

Likelihood ratio test statistic

For two constrained space in Null space ω and Alternative space τ. Plug (1.50) in.

$$\begin{aligned}
D &= -2\{\max_{\theta\in\omega} l(X;\theta) - \max_{\theta\in\omega\cup\tau} l(X;\theta)\} \\
&= -2n\left\{\max_{\theta\in\omega}(-\frac{1}{2}\eta^T\mathcal{I}\eta) - \max_{\theta\in\omega\cup\tau}(-\frac{1}{2}\eta^T\mathcal{I}\eta) + O_p(n^{-\frac{3}{2}})\right\} \\
&= n\left\{\min_{\theta\in\omega}\eta^T\mathcal{I}\eta - \min_{\theta\in\omega\cup\tau}\eta^T\mathcal{I}\eta\right\} + O_p(n^{-\frac{1}{2}}) \\
&= n\Big\{\min_{\theta\in\omega}[\mathcal{I}^{-1}A - (\theta-\theta_0)]^T\mathcal{I}[\mathcal{I}^{-1}A - (\theta-\theta_0)] \\
&\quad - \min_{\theta\in\omega\cup\tau}[\mathcal{I}^{-1}A - (\theta-\theta_0)]^T\mathcal{I}[\mathcal{I}^{-1}A - (\theta-\theta_0)]\Big\} + O_p(n^{-\frac{1}{2}}) \\
&= \Big\{\min_{\theta\in\omega}[z - \sqrt{n}(\theta-\theta_0)]^T\mathcal{I}[z - \sqrt{n}(\theta-\theta_0)] \\
&\quad - \min_{\theta\in\omega\cup\tau}[z - \sqrt{n}(\theta-\theta_0)]^T\mathcal{I}[z - \sqrt{n}(\theta-\theta_0)]\Big\} + O_p(n^{-\frac{1}{2}})
\end{aligned} \tag{1.52}$$

If space ω is not linear or a cone, we can use a cone approximate the space around θ_0.

Definition of a cone: $\mathcal{C} \in R^m$ is a cone if for any $x \in \mathcal{C}$ implies $ax \in \mathcal{C}$, $\forall a > 0$.

Let a closed and convex cone $\mathcal{C}_\omega$ approximates ω at θ_0. This cone is independent of choices of norms, because norms in R^p are equivalent. Here we use the norm $||x|| = \sqrt{x^T\mathcal{I}x}$,

$$\inf_{\theta\in\omega}||(\theta-\theta_0) - \theta_c|| = o(||\theta_c||), \theta_c \in \mathcal{C}_\omega \quad \inf_{\theta_c\in\mathcal{C}_\omega}||(\theta-\theta_0) - \theta_c|| = o(||\theta-\theta_0||), \theta \in \omega$$

According to [Sha87] Th2 and [Che54], the projections of y onto ω and $\mathcal{C}_\omega$, $\hat{\theta}$ and $\hat{\theta}_c$ have the relationship

$$||\hat{\theta} - \theta_0 - \hat{\theta}_c|| = o(||y||) \Rightarrow ||\hat{\theta} - \theta_0||^2 - ||\hat{\theta}_c||^2 = o(||y||^2)$$

$$\Rightarrow \inf_{\theta\in\omega}(y-(\theta-\theta_0))^T\mathcal{I}(y-(\theta-\theta_0)) = \inf_{\theta_c\in\mathcal{C}_\omega}(y-\theta_c)^T\mathcal{I}(y-\theta_c)+o(||y||^2) \quad (1.53)$$

Both ω or $\mathcal{C}_\omega$ are closed, so the notations "inf" can be replaced by "min",

$$\begin{aligned}
&\min_{\theta\in\omega}[z-\sqrt{n}(\theta-\theta_0)]^T\mathcal{I}[z-\sqrt{n}(\theta-\theta_0)] \\
&= n\left\{\min_{\theta\in\omega}[z/\sqrt{n}-(\theta-\theta_0)]^T\mathcal{I}[z/\sqrt{n}-(\theta-\theta_0)]\right\} \\
&= n\left\{\min_{\theta_c\in\mathcal{C}_\omega}[z/\sqrt{n}-\theta_c]^T\mathcal{I}[z/\sqrt{n}-\theta_c]+o_p(\frac{||z||^2}{n})\right\} \\
&= \min_{\theta_c\in\mathcal{C}_\omega}[z-\sqrt{n}\theta_c]^T\mathcal{I}[z-\sqrt{n}\theta_c]+o_p(||z||^2) \\
&\quad \{\theta_c:\theta_c\in\mathcal{C}_\omega\} \iff \{\sqrt{n}\theta_c:\theta_c\in\mathcal{C}_\omega\} \\
&\quad \text{because a cone is a positively homogeneous set } \theta_c\in\mathcal{C}_\omega \Rightarrow a\theta_c\in\mathcal{C}_\omega,\ \forall a>0 \\
&= \min_{\theta_c\in\mathcal{C}_\omega}[z-\theta_c]^T\mathcal{I}[z-\theta_c]+o_p(1)
\end{aligned}$$

Thus, [Che54]

$$D = \left\{\min_{\theta_c\in\mathcal{C}_\omega}[z-\theta_c]^T\mathcal{I}[z-\theta_c] - \min_{\theta_c\in\mathcal{C}_{\omega\cup\tau}}[z-\theta_c]^T\mathcal{I}[z-\theta_c]\right\}+o_p(1) \quad (1.54)$$

In our case, if $\omega\cup\tau$ constitutes the whole parameter space, $\mathcal{C}_{\omega\cup\tau}$ at θ_0 is also the whole parameter space,

$$\min_{\theta_c\in\mathcal{C}_{\omega\cup\tau}}[z-\theta_c]^T\mathcal{I}[z-\theta_c]=0$$

$$D = \min_{\theta_c\in\mathcal{C}_\omega}[z-\theta_c]^T\mathcal{I}[z-\theta_c]+o_p(1) \quad (1.55)$$

We can easily see the asymptotical distribution of D, if we rewrite $z\sim N(0,\mathcal{I}^{-1})$ precisely instead of the asymptotical expression in (1.51),

$$D \xrightarrow{L} \min_{\theta_c\in\mathcal{C}_\omega}[z-\theta_c]^T\mathcal{I}[z-\theta_c] \quad (1.56)$$

Cone Approximation at Boundary Points of Constrained Space

Our constrained logit value space $\mathbb{M}_0$ is not linear (constraints on $\underline{\psi}_{\{I_f\}}$ are linear), however, if the sample size is big enough, the MLE estimate gets very close to the

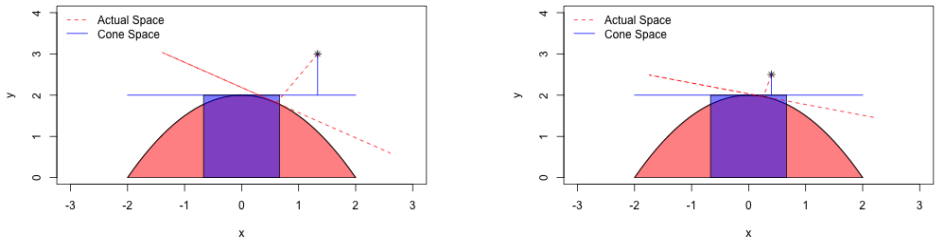


Figure 1.17: Illustration of cone approximation

small region around the true value, the constrained space can then be approximated by a cone at the true value.

Let the constrained space be $S = \{\theta : s(\theta) \leq 0\}$, S can be approximated at its boundary point θ_0 ($s(\theta_0) = 0$) by a cone $\mathcal{C}$, where $\mathcal{C} = \{\theta_c : s'(\theta_0, \theta_c) \leq 0\}$. One can imagine θ_c to be a vector where its origin is at θ_0.

$$s(\theta_0 + \theta_c) = s(\theta_0) + s'(\theta_0, \theta_c) + o(||\theta_c||_2) \tag{1.57}$$

$$\text{directional derivative:} \quad s'(\theta_0, \theta_c) = \lim_{t \to 0^+} \frac{s(\theta_0 + t\theta_c) - s(\theta_0)}{t} = \theta_c^T s'(\theta_0) \tag{1.58}$$

In our problem, for each gene, full parameter space size is $G * (2F - 2)$,

$$\begin{aligned} \theta = (&\underline{\mu}_{1,I_1}, \underline{\mu}_{1,I_2}, ..., \underline{\mu}_{1,I_{F-1}}, ..., \underline{\mu}_{G,I_1}, \underline{\mu}_{G,I_2}, ..., \underline{\mu}_{G,I_{F-1}}, \\ &\underline{\sigma}^2_{1,I_1}, \underline{\sigma}^2_{1,I_2}, ..., \underline{\sigma}^2_{1,I_{F-1}}, ..., \underline{\sigma}^2_{G,I_1}, \underline{\sigma}^2_{G,I_2}, ..., \underline{\sigma}^2_{G,I_{F-1}})^T \end{aligned} \tag{1.59}$$

Our constraints are usually written as a series of pairs of isoform probability difference inequalities for each isoform I_f in given group g_1 and g_2 smaller than some threshold ξ,

$$|\underline{\psi}_{g_1,I_f} - \underline{\psi}_{g_2,I_f}| \leq \xi \iff \begin{cases} s_{g_1,g_2,I_f}(\theta) = \frac{e^{\underline{\mu}_{g_1,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_1,I_f}} + 1} - \frac{e^{\underline{\mu}_{g_2,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_2,I_f}} + 1} - \xi \leq 0 \\ s_{g_2,g_1,I_f}(\theta) = \frac{e^{\underline{\mu}_{g_2,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_2,I_f}} + 1} - \frac{e^{\underline{\mu}_{g_1,I_f}}}{\sum_{f=1}^{F-1} e^{\underline{\mu}_{g_1,I_f}} + 1} - \xi \leq 0 \end{cases} \tag{1.60}$$

Obviously, a boundary point regarding to isoform I_f inequalities can only locate in one of the constraints above, thus in the following we wrote the inequality in

short as $s_{[g_1,g_2],I_f}(\theta)$, representing either one of the above. We do have constraints regarding to $\underline{\sigma}^2_{g,I_f}$, non-negativity. However, the 0 is never to be touched, so we do not need to put them in the constraints above.

If θ_0 is at the boundary of one constraint,

$$s'_{[g_1,g_2],I_f}(\theta_0,\theta_c) = \theta_c^T s'_{[g_1,g_2],I_f}(\theta_0) \leq 0 \tag{1.61}$$

Thus by [Sha87], constrained space S at point θ_0 can be approximated by cone $\mathcal{C} = \{\theta_c : s'_{[g_1,g_2],I_f}(\theta_0,\theta_c) \leq 0\}$ with the differentiability of norm $||x|| = \sqrt{x^T\mathcal{I}x}$ at θ_0, and this cone is actually a half-space of $\mathbb{R}^{G(2F-2)}$.

If x_0 is the intersection of multiple constraints (if $F \geq 3$ or $G \geq 3$), the constrained space at point x_0 can be approximated by intersection of $\binom{G}{2}(F-1)$ half-spaces. If $G = 2$, $[g_1, g_2]$ notations can be omitted.

$$S = \{\theta : s_{[g_1,g_2],I_f}(\theta) \leq 0, \forall 1 \leq f \leq F-1, \forall 1 \leq g_1 < g_2 \leq G\} \tag{1.62}$$

Let

$$s(\theta) = \max\{s_{[g_1,g_2],I_f}(\theta) : 1 \leq f \leq F-1, 1 \leq g_1 < g_2 \leq G\} \tag{1.63}$$

$$\Xi = \{([g_1,g_2],I_f) : \ s_{[g_1,g_2],I_f}(\theta_0) = s(\theta_0) = 0, 1 \leq f \leq F-1, 1 \leq g_1 < g_2 \leq G\} \tag{1.64}$$

Under regularity conditions [Sha87] (all satisfied in our case),

$$s'(\theta_0,\theta_c) = \max\{\theta_c^T s'_{[g_1,g_2],f}(\theta_0), ([g_1,g_2],f) \in \Xi\} \tag{1.65}$$

$$\begin{aligned}\mathcal{C} &= \{\theta_c : s'(\theta_0,\theta_c) \leq 0\} \\ &= \{\theta_c : s'_{([g_1,g_2],f)}(\theta_0,\theta_c) \leq 0, \forall\ ([g_1,g_2],f) \in \Xi\} \\ &= \underset{([g_1,g_2],f)\in\Xi}{\cap} \{\theta_c : s'_{[g_1,g_2],f}(\theta_0,\theta_c) \leq 0\}\end{aligned} \tag{1.66}$$

For notational simplicity, let $p = |\Xi|$, $m = G(2F-2)$ and let $Q = [s_1'(\theta_0), ..., s_p'(\theta_0)]^T$, $p \times m$. We can rewrite cone in $\mathbb{R}^m$ as

$$C = \{\theta_c : -Q\theta_c \geq 0\} \tag{1.67}$$

If $p > G(F-1)$, we need to keep only $G(F-1)$ of the inequalities to have a nonempty cone, because there're only $G(F-1)$ unknown $\underline{\mu}$'s. The choice is not unique though, depending on where the true parameter θ_0 locates. Besides, if Q is singular, we need to delete the redundant constraints, e.g. $F = 4$, $G = 2$, first 3 constraints $|\underline{\psi}_{1,I_f} - \underline{\psi}_{2,I_f}| \leq \xi$ would imply the forth, thus Q is singular if all constraints are kept.

The Distribution of Our Test Statistic given True Parameters Locating on the Boundary

Let $\mathcal{C}_\omega$ being a cone, $\mathcal{C}_\omega^0 = \{y : x^T\mathcal{I}y \leq 0, \forall x \in \mathcal{C}_\omega\}$ is its polar cone under inner product $(x, y) = x^T\mathcal{I}y$ and norm $||x|| = \sqrt{x^T\mathcal{I}x}$. Based on [Sha88], the likelihood ratio test statistic D in (1.55) asymptotically follows a mixture of chi-square distributions, let $z \sim N(0, \mathcal{I}^{-1})$

$$D \xrightarrow{L} \min_{\theta_c \in \mathcal{C}_\omega} [z-\theta_c]^T\mathcal{I}[z-\theta_c] = z^T\mathcal{I}z - \min_{\theta_c \in \mathcal{C}_\omega^0} [z-\theta_c]^T\mathcal{I}[z-\theta_c] \quad \text{(Pythagoras' theorem)} \tag{1.68}$$

$$D \xrightarrow{L} \bar{\chi}_m^2(\mathcal{I}^{-1}, \mathcal{C}_\omega^0) = \sum_{i=0}^{m} w_i\chi_i^2, \quad w_i = w_i(m, \mathcal{I}^{-1}, \mathcal{C}_\omega^0) \tag{1.69}$$

- The basic idea is for a random z, $\hat{\theta}_c$ could touch boundaries of the constrained space $\mathcal{C}_\omega^0$. The more constraints the boundary touches, the shorter the projection, the longer the distance. E.g., if the constrained space is $\mathcal{C}_\omega^0 = \{\theta_c : \theta_c = 0\}$, which triggers equality constraints of all dimensions $((\theta_c)_1 = (\theta_c)_2 = \cdots = (\theta_c)_m = 0)$, then the projection of z onto this space $\mathcal{C}_\omega^0$ would always be the shortest 0, and the distance from z to $\mathcal{C}_\omega^0$ would always be the longest $\min_{\theta_c \in \mathcal{C}_\omega^0} [z-\theta_c]^T\mathcal{I}[z-\theta_c] = z^T\mathcal{I}z$.

- The weights w_i are determined as long as the covariance matrix $\mathcal{I}^{-1}$ of $z_{m\times 1} \sim N(0, \mathcal{I}^{-1})$ and the cone $\mathcal{C}_\omega$ in space $\mathbb{R}^m$ are determined.
- If $\mathcal{C}_\omega^0$ is $\mathbb{R}_+^m = \{\theta_c : \theta_c \geq 0\}$, and $\mathcal{I}^{-1} = I$, then $w_i = \binom{m}{i} 2^{-m}$, $i = 0, \cdots, m$.

In our problem, $\mathcal{C}_\omega = \{\theta_c : -Q\theta_c \geq 0\}$, where Q is $p \times m$. Based on [Sha88] eq (5.5),

$$w_i(m, \mathcal{I}^{-1}, \mathcal{C}_\omega) = \begin{cases} w_{i-(m-p)}(p, Q\mathcal{I}^{-1}Q^T) & m-p \leq i \leq m \\ 0 & 0 \leq i \leq m-p-1 \end{cases} \tag{1.70}$$

Together with

$$w_i(m, \mathcal{I}^{-1}, \mathcal{C}_\omega^0) = w_{m-i}(m, \mathcal{I}^{-1}, \mathcal{C}_\omega) \tag{1.71}$$

We get

$$\Rightarrow w_i(m, \mathcal{I}^{-1}, \mathcal{C}_\omega^0) = \begin{cases} w_{p-i}(p, Q\mathcal{I}^{-1}Q^T) = w_i(p, (Q\mathcal{I}^{-1}Q^T)^{-1}) & 0 \leq i \leq p \\ 0 & p+1 \leq i \leq m \end{cases} \tag{1.72}$$

As to the actual calculation of the weight, let $Y \sim N(0, V)$, $V = Q\mathcal{I}^{-1}Q^T$,

$$w_j(p, V) = w_j(p, V, \mathbb{R}_+^p) = \sum_{|\alpha|=j} p(V_{\alpha'}^{-1}) p(V_{\alpha;\alpha'}) \tag{1.73}$$

Index α is a subset of $\{1, ..., p\}$, α' is its complement. They denote the indices of random variables in Y. $|\alpha|$ denotes the size of α. For example, if $\alpha = \{1, 2\}$, Y_α means $(Y_1, Y_2)^T$. $Y_\alpha \sim N(0, V_\alpha)$. V_α means the covariance matrix of Y_α. $V_{\alpha;\alpha'}$ means the conditional variance matrix of $Y_\alpha | Y_{\alpha'} = 0$. $P(V_\alpha) = P(Y_\alpha \geq 0), P(V_{\alpha;\alpha'}) = P(Y_\alpha \geq 0 | Y_{\alpha'} = 0)$.

Given V, it is possible to calculate the analytic result for these weights. The exact formulas for these gaussian probabilities are available in [Kud63] *Theorem* (3.1), yet complicated to carry out. For simplicity, we use MCMC sampling of normal distributions to approximate these weights at a very small computational cost.

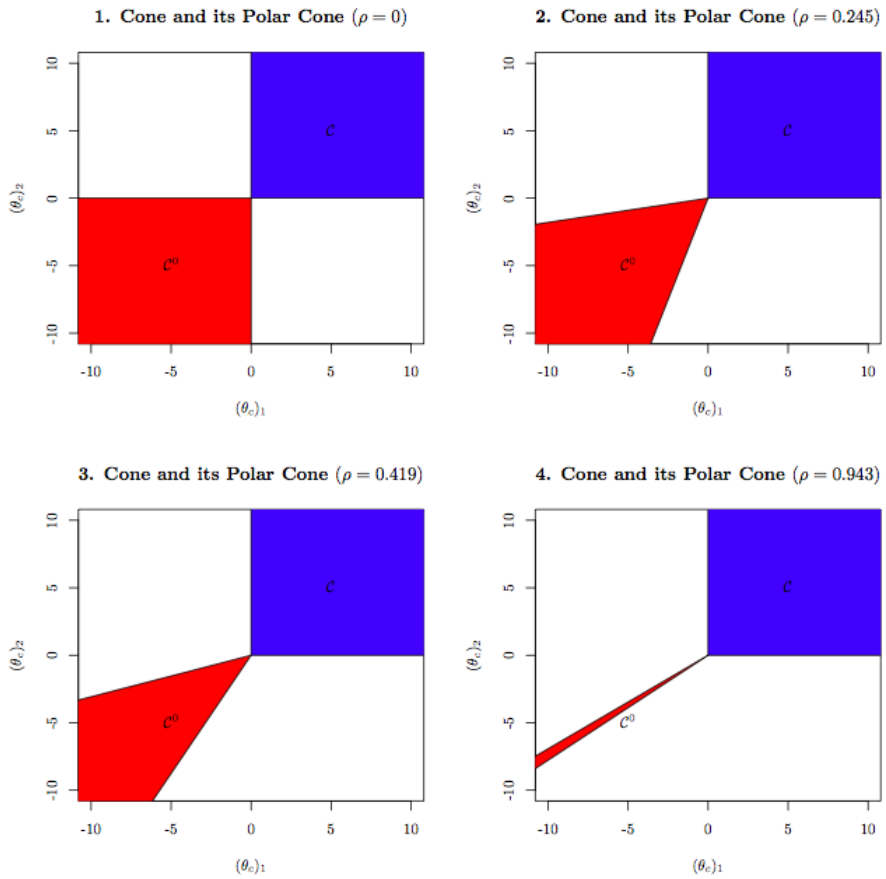


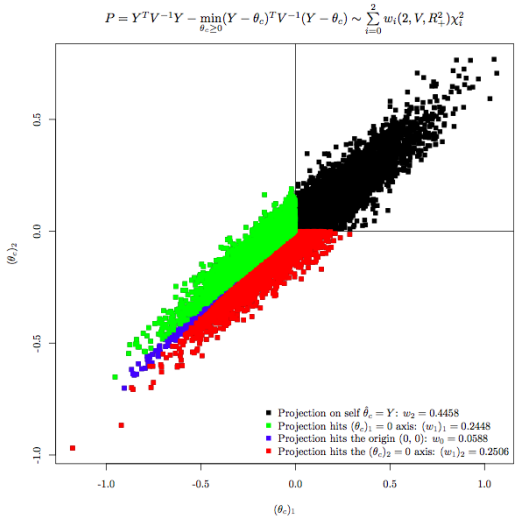
Figure 1.18: Cone and its polar cone

Example 1 (Fig 1.18):

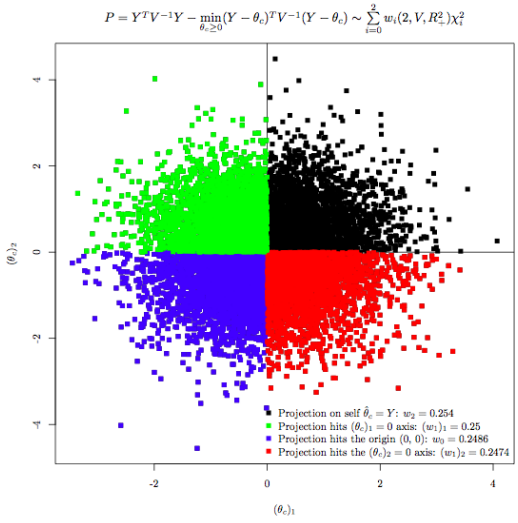
$$\mathcal{C}_\omega = \mathbb{R}^2_+ = \{\theta_c : \theta_c \geq 0\}, \quad \mathcal{C}^0_\omega = \{y : x^T V^{-1} y \leq 0, \forall x \in \mathcal{C}_\omega\}$$

$$V_1 = \begin{bmatrix} 0.078 & 0 \\ 0 & 0.042 \end{bmatrix}. \quad V_2 = \begin{bmatrix} 0.078 & 0.014 \\ 0.014 & 0.042 \end{bmatrix}. \quad V_3 = \begin{bmatrix} 0.078 & 0.024 \\ 0.024 & 0.042 \end{bmatrix}. \quad V_4 = \begin{bmatrix} 0.078 & 0.054 \\ 0.054 & 0.042 \end{bmatrix}$$

Example 2 (Fig 1.19):



(a) χ_2^2, projection on self (black) 44.84%; χ_1^2, projection on $x_1 = 0$ (green) 25.08%; χ_0^2, projection on the origin (blue) 5.6%; χ_1^2, Projection on $x_2 = 0$ (red) 24.42%



(b) χ_2^2, projection on self (black) 24.76%; χ_1^2, projection on $x_1 = 0$ (green) 25%; χ_0^2, projection on the origin (blue) 25%; χ_1^2, projection on $x_2 = 0$ (red) 25.24%

Figure 1.19: 10000 points sampled from $N(0, V)$ under Scenario 1 and Scenario 2. Different colors denote 4 different boundaries in constrained space that solutions $\hat{\theta}_c$ could touch: (1) not on boundaries. (2) $x_1 = 0$. (3) origin. (4) $x_2 = 0$.

$$Y = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N(0, V)$$

$$P = Y^T V^{-1} Y - \min_{\theta_c \geq 0} (Y-\theta_c)^T V^{-1} (Y-\theta_c) \sim \sum_{i=0}^{2} w_i(2, V, R_+^2)\chi_i^2 \tag{1.74}$$

Scenario 1: let $V = \begin{bmatrix} 0.078 & 0.054 \\ 0.054 & 0.042 \end{bmatrix}$, $Y \sim N(0, V)$, $w_0 \approx 0.056, w_1 \approx 0.504, w_2 \approx 0.44$.

Scenario 2: let $V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, obviously $w_0 = 0.25, w_1 = 0.5, w_2 = 0.25$.

1.7.4 Fisher Information $\mathcal{I}$

We denote all parameters $\theta = (\underline{\mu}_{1,\{I_f\}}, \underline{\mu}_{2,\{I_f\}}, \cdots, \underline{\mu}_{G,\{I_f\}}, \underline{\sigma}^2_{1,\{I_f\}}, \underline{\sigma}^2_{2,\{I_f\}}, \cdots, \underline{\sigma}^2_{G,\{I_f\}})$ as in (1.11). The fisher information matrix can be decomposed into 4 parts $\mathcal{I} = \begin{bmatrix} \mathcal{I}_{\underline{\mu}\underline{\mu}} & \mathcal{I}_{\underline{\mu}\underline{\sigma}^2} \\ \mathcal{I}_{\underline{\sigma}^2\underline{\mu}^2} & \mathcal{I}_{\underline{\sigma}^2\underline{\sigma}^2} \end{bmatrix}$. $\mathcal{I}$ is of dimension $G(2F-2)$. Use likelihood formula for a single replicate of one group of one gene in (1.6), and assume the replicate number is equal among all groups $K = K_1 = \cdots = K_G$, and total read counts $R_{g,1} = R_{g,2} = \cdots = R_{g,K_g}, \forall g$ among all replicates for each group. We randomly pair one replicate from all groups together as one replicate. By the independence assumptions across all replicates, without any loss, we simply pair them by their original indices. Our target probability function for each nominal replicate is

$$P(R_{\{g\},k,\{I_f\}};\theta) = \prod_{g=1}^{G} P(R_{g,k,\{I_f\}};\theta)$$
$$= \prod_{g=1}^{G} (\sqrt{2\pi})^{F-1} |-h^{(2)}(\hat{\mu}_{g,k,\{I_f\}}; \underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}})|^{-0.5} \exp(h(\hat{\mu}_{g,k,\{I_f\}}; \underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}})) \tag{1.75}$$

$$h(\mu_{g,k,\{I_f\}}; \underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}}) \tag{1.76}$$
$$= \log P(R_{g,k,\{I_f\}} | \mu_{g,k,\{I_f\}}; \underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}}) + \log f(\mu_{g,k,\{I_f\}}; \underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}})$$
$$= \log(\frac{R_{g,k}!}{\prod_{f=1}^{F} R_{g,k,I_f}!}) + \sum_{f=1}^{F-1} R_{g,k,I_f} \log\left(\frac{\underline{l}_{I_f} e^{\mu_{g,k,I_f}}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{g,k,I_f}} + \underline{l}_{I_F}}\right) +$$
$$R_{g,k,I_F} \log\left(\frac{\underline{l}_{I_F}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{g,k,I_f}} + \underline{l}_{I_F}}\right) +$$
$$\left(\sum_{f=1}^{F-1} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\underline{\sigma}^2_{g,I_f}) - \frac{(\mu_{g,k,I_f} - \underline{\mu}_{g,I_f})^2}{2\underline{\sigma}^2_{g,I_f}}\right) \tag{1.77}$$

All "nominal" replicates $1, \cdots, K$ are $i.i.d$ as $P(R_{\{g\},k,\{I_f\}};\theta)$, the likelihood function is,

$$L = \prod_{k=1}^{K} P(R_{\{g\},k,\{I_f\}};\theta) \tag{1.78}$$

We want fisher information $\mathcal{I}_{\{g\}}$,

$$\begin{aligned}\mathcal{I}_{\{g\}} &= -E\left[\frac{\partial^2}{\partial\theta^2}\log P(R_{\{g\},k,\{I_f\}};\theta)\right] \\ &= E\left[(\frac{\partial}{\partial\theta}\log P(R_{\{g\},k,\{I_f\}};\theta))(\frac{\partial}{\partial\theta}\log P(R_{\{g\},k,\{I_f\}};\theta))^T\right]\end{aligned}$$

The only probability difference between groups is their parameters $\underline{\mu}_{g,\{I_f\}}, \underline{\sigma}^2_{g,\{I_f\}}$. We pull all replicates from all groups together to form the nominal replicate, $P(R_{\{g\},k,\{I_f\}};\theta)$ is actually a direct product of these independent replicate likelihood in each group. There is no interaction between parameters of different groups in the fisher information matrix $\mathcal{I}_{\{g\}}$. We only need to calculate fisher information for each group $\mathcal{I}_g$, then fill them in their respective positions in $\mathcal{I}_{\{g\}}$. Ignore group and replicate indices,

$$\begin{aligned}\mathcal{I} = E\Bigg[&\left(\frac{\partial}{\partial(\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})}\log P(R_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})\right) \\ &\left(\frac{\partial}{\partial(\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})}\log P(R_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})\right)^T\Bigg]\end{aligned} \tag{1.79}$$

The random variable $R_{\{I_f\}}$ here is a multinomial distribution, we use MCMC sampling to approximate $\mathcal{I}$ in (1.79). When R is small, the Laplace approximated probabilities could be directly sum over all possible outcomes of $R_{\{I_f\}}$, otherwise, we sample random logit value $\mu_{\{I_f\}} \sim N(\underline{\mu}_{\{I_f\}}, \underline{\sigma})$ first, then sample the multinomial $R_{\{I_f\}}$ based on the sampled logit values.

$$\begin{aligned}&P(R_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\ &= C(\sqrt{2\pi})^{F-1}|-h^{(2)}(\hat{\mu}_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})|^{-0.5}\exp(h(\hat{\mu}_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}))\end{aligned} \tag{1.80}$$

$$\begin{aligned}&\log P(R_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\ &= \log C + (\frac{F-1}{2})\log(2\pi) - \frac{1}{2}\log(|-h^{(2)}(\hat{\mu}_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})|) + \end{aligned} \tag{1.81}$$

$$h(\hat{\mu}_{\{I_f\}};\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \tag{1.82}$$

$\log C = 1 + O(R^{-1})$ is approximately zero when total counts R is large enough.

$$\begin{aligned}
&\frac{\partial}{\partial \underline{\mu}_{\{I_f\}}} \log P(R_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\
&= -\frac{1}{2}\left((\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}})^T \frac{\partial \log|-h^{(2)}|}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} + \frac{\partial \log|-h^{(2)}(\hat{\mu}_{\{I_f\}})|}{\partial \underline{\mu}_{\{I_f\}}} \right) + \\
&\quad (\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}})^T h^{(1)}(\hat{\mu}_{\{I_f\}}) + \frac{\partial h(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{\{I_f\}}}
\end{aligned} \tag{1.83}$$

$$\begin{aligned}
&\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}} \log P(R_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\
&= -\frac{1}{2}\left((\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}})^T \frac{\partial \log|-h^{(2)}|}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} + \frac{\partial \log|-h^{(2)}(\hat{\mu}_{\{I_f\}})|}{\partial \underline{\sigma}^2_{\{I_f\}}} \right) + \\
&\quad (\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}})^T h^{(1)}(\hat{\mu}_{\{I_f\}}) + \frac{\partial h(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{\{I_f\}}}
\end{aligned} \tag{1.84}$$

Note that if there is notation conflict, the partial derivative on the left-side uses the chain rule, on the right hand-side it refers to partial derivative w.r.t. the position of the variable.

Given $\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}$, we want $\frac{\partial \log P(R_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})}{\partial (\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})}$, where $h^{(2)}$, h and $\hat{\mu}$ are all related to $\underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}$.

We need

1. $h^{(1)}(\mu_{\{I_f\}}) = \frac{\partial h}{\partial \mu_{\{I_f\}}}$, $h^{(2)}(\mu_{\{I_f\}}) = \frac{\partial^2 h}{\partial^2 \mu_{\{I_f\}}}$
2. $\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}}$
3. $\frac{\partial \log|-h^{(2)}|}{\partial \mu_{\{I_f\}}}$
4. $\frac{\partial \log|-h^{(2)}|}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial \log|-h^{(2)}|}{\partial \underline{\sigma}^2_{\{I_f\}}}$, $\frac{\partial h}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial h}{\partial \underline{\sigma}^2_{\{I_f\}}}$

1. $h^{(1)}(\mu_{\{I_f\}}) = \frac{\partial h}{\partial \mu_{\{I_f\}}}$, $h^{(2)}(\mu_{\{I_f\}}) = \frac{\partial^2 h}{\partial^2 \mu_{\{I_f\}}}$

$$
\begin{aligned}
&h(\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) = \log P(R_{\{I_f\}}|\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) + \log f(\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\
&= \frac{F-1}{2}\log(2\pi) - \frac{1}{2}\log|\underline{\sigma}| + \log P(R_{\{I_f\}}|\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) - \\
&\frac{1}{2}(\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})^T \underline{\sigma}^{-1}(\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})
\end{aligned} \tag{1.85}
$$

$$
\text{where } \mu_{\{I_f\}} \sim N(\underline{\mu}_{\{I_f\}}, \underline{\sigma}), \quad \underline{\sigma} = \begin{bmatrix} \underline{\sigma}^2_{I_1} & 0 & \cdots & 0 \\ 0 & \underline{\sigma}^2_{I_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \underline{\sigma}^2_{I_{F-1}} \end{bmatrix} \tag{1.86}
$$

Rewrite

$$
\begin{aligned}
&\log P(R_{\{I_f\}}|\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\
&= \sum_{f=1}^{F-1} R_{I_f}(\log \frac{\underline{l}_{I_f}}{\underline{l}_{I_F}} + \mu_{I_f}) + R\log(\frac{\underline{l}_{I_F}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F}}) + \log(\frac{R!}{\prod_{f=1}^{F-1} R_f!}) \\
&= y^T_{\{I_f\}} \eta(\mu_{\{I_f\}}) - \delta(\mu_{\{I_f\}}) + \gamma(y_{\{I_f\}})
\end{aligned} \tag{1.87}
$$

Let

$$
y_{\{I_f\}} = \begin{bmatrix} R_{I_1} \\ R_{I_2} \\ \vdots \\ R_{I_{F-1}} \end{bmatrix}, \quad \eta(\mu_{\{I_f\}}) = \begin{bmatrix} \log(\frac{\underline{l}_{I_1}}{\underline{l}_{I_F}}) + \mu_{I_1} \\ \log(\frac{\underline{l}_{I_2}}{\underline{l}_{I_F}}) + \mu_{I_2} \\ \vdots \\ \log(\frac{\underline{l}_{I_{F-1}}}{\underline{l}_{I_F}}) + \mu_{I_{F-1}} \end{bmatrix} = c_{\{I_f\}} + \mu_{\{I_f\}} \tag{1.88}
$$

$$
\delta(\mu_{\{I_f\}}) = -R\log(\frac{\underline{l}_{I_F}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F}}), \quad \gamma(y_{\{I_f\}}) = \log(\frac{R!}{\prod_{f=1}^{F-1} R_f!}) \tag{1.89}
$$

$$\frac{\partial \delta}{\partial \mu_{\{I_f\}}} = \frac{\partial \delta}{\partial \eta} = E(y_{\{I_f\}}|\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}})$$
$$= R \begin{bmatrix} \frac{\underline{l}_{I_1} e^{\mu_{I_1}}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F}} \\ \frac{\underline{l}_{I_2} e^{\mu_{I_2}}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F}} \\ \vdots \\ \frac{\underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}}}{\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F}} \end{bmatrix} = Rb(\mu_{\{I_f\}}) \tag{1.90}$$

$$\frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}} = \tag{1.91}$$

$$\begin{bmatrix} \frac{\underline{l}_{I_1} e^{\mu_{I_1}} (\sum_{f \neq 1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & -\frac{\underline{l}_{I_1} e^{\mu_{I_1}} \underline{l}_{I_2} e^{\mu_{I_2}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & \cdots & -\frac{\underline{l}_{I_1} e^{\mu_{I_1}} \underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} \\ -\frac{\underline{l}_{I_1} e^{\mu_{I_1}} \underline{l}_{I_2} e^{\mu_{I_2}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & \frac{\underline{l}_{I_2} e^{\mu_{I_2}} (\sum_{f \neq 2} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & \cdots & -\frac{\underline{l}_{I_2} e^{\mu_{I_2}} \underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\underline{l}_{I_1} e^{\mu_{I_1}} \underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & -\frac{\underline{l}_{I_2} e^{\mu_{I_2}} \underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}}}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} & \cdots & \frac{\underline{l}_{I_{F-1}} e^{\mu_{I_{F-1}}} (\sum_{f \neq F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})}{(\sum_{f=1}^{F-1} \underline{l}_{I_f} e^{\mu_{I_f}} + \underline{l}_{I_F})^2} \end{bmatrix} \tag{1.92}$$

$$\begin{aligned} & h(\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\ & = -\frac{F-1}{2} \log(2\pi) - \frac{1}{2} \log |\underline{\sigma}| + y^T_{\{I_f\}} \eta(\mu_{\{I_f\}}) - \delta(\mu_{\{I_f\}}) + \gamma(y_{\{I_f\}}) \\ & - \frac{1}{2} (\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})^T \underline{\sigma}^{-1} (\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \end{aligned} \tag{1.93}$$

$$\begin{aligned} & h^{(1)}(\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) \\ & = \frac{\partial}{\partial \mu_{\{I_f\}}} h(\mu_{\{I_f\}}; \theta) \\ & = \frac{\partial}{\partial \mu_{\{I_f\}}} \left(y^T_{\{I_f\}} \eta(\mu_{\{I_f\}}) - \delta(\mu_{\{I_f\}})) - \frac{1}{2} (\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})^T \underline{\sigma}^{-1} (\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \right) \\ & = y^T_{\{I_f\}} - Rb(\mu_{\{I_f\}}) - \underline{\sigma}^{-1} (\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \end{aligned} \tag{1.94}$$

$$h^{(2)}(\mu_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) = -R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}} - \underline{\sigma}^{-1} \tag{1.95}$$

2. $\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}}$

The maximizer $\hat{\mu}_{\{I_f\}}$ is the solution of $h^{(1)}(\hat{\mu}_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) = 0$. Use implicit differentiation by taking derivatives w.r.t. $\underline{\mu}_{\{I_f\}}$ and $\underline{\sigma}^2_{\{I_f\}}$ on both sides of this equation.

$$\frac{\partial}{\partial \underline{\mu}_{\{I_f\}}} h^{(1)}(\hat{\mu}_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) = 0$$

$$- R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}} - \underline{\sigma}^{-1} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}} + \underline{\sigma}^{-1} = 0 \tag{1.96}$$

$$\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\mu}_{\{I_f\}}} = (R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} + \underline{\sigma}^{-1})^{-1} \underline{\sigma}^{-1} \tag{1.97}$$

$$\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}} h^{(1)}(\hat{\mu}_{\{I_f\}}; \underline{\mu}_{\{I_f\}}, \underline{\sigma}^2_{\{I_f\}}) = 0$$

$$- R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}} - \frac{\partial (\underline{\sigma}^{-1}(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}))}{\partial \underline{\sigma}^2_{\{I_f\}}} = 0 \tag{1.98}$$

$$\underline{\sigma}^{-1}(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) = \begin{bmatrix} \frac{1}{\underline{\sigma}^2_{I_1}}(\hat{\mu}_{I_1} - \underline{\mu}_{I_1}) \\ \frac{1}{\underline{\sigma}^2_{I_2}}(\hat{\mu}_{I_2} - \underline{\mu}_{I_2}) \\ \vdots \\ \frac{1}{\underline{\sigma}^2_{I_{F-1}}}(\hat{\mu}_{I_{F-1}} - \underline{\mu}_{I_{F-1}}) \end{bmatrix} \tag{1.99}$$

The independence of prior probabilities of $\mu_{\{I_f\}}$ makes the calculation easy here, if $\underline{\sigma}$ is not diagonal, the derivative w.r.t. the variance-covariance terms is not limited to $\underline{\sigma}^2_{\{I_f\}}$, but all covariance elements in $\underline{\sigma}$, which is more complicated to

write down.

$$\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}} \underline{\sigma}^{-1}(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) =$$

$$\begin{bmatrix} -\frac{1}{\underline{\sigma}^4_{I_1}}(\hat{\mu}_{I_1} - \underline{\mu}_{I_1}) + \frac{1}{\underline{\sigma}^2_{I_1}} \frac{\partial \hat{\mu}_{I_1}}{\partial \underline{\sigma}^2_{I_1}} & \frac{1}{\underline{\sigma}^2_{I_1}} \frac{\partial \hat{\mu}_{I_1}}{\partial \underline{\sigma}^2_{I_2}} & \cdots & \frac{1}{\underline{\sigma}^2_{I_1}} \frac{\partial \hat{\mu}_{I_1}}{\partial \underline{\sigma}^2_{I_{F-1}}} \\ \frac{1}{\underline{\sigma}^2_{I_2}} \frac{\partial \hat{\mu}_{I_2}}{\partial \underline{\sigma}^2_{I_1}} & -\frac{1}{\underline{\sigma}^4_{I_2}}(\hat{\mu}_{I_2} - \underline{\mu}_{I_2}) + \frac{1}{\underline{\sigma}^2_{I_2}} \frac{\partial \hat{\mu}_{I_2}}{\partial \underline{\sigma}^2_{I_2}} & \cdots & \frac{1}{\underline{\sigma}^2_{I_2}} \frac{\partial \hat{\mu}_{I_2}}{\partial \underline{\sigma}^2_{I_{F-1}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\underline{\sigma}^2_{I_{F-1}}} \frac{\partial \hat{\mu}_{I_{F-1}}}{\partial \underline{\sigma}^2_{I_1}} & \frac{1}{\underline{\sigma}^2_{I_{F-1}}} \frac{\partial \hat{\mu}_{I_{F-1}}}{\partial \underline{\sigma}^2_{I_2}} & \cdots & -\frac{1}{\underline{\sigma}^4_{I_{F-1}}}(\hat{\mu}_{I_{F-1}} - \underline{\mu}_{I_{F-1}}) + \frac{1}{\underline{\sigma}^2_{I_{F-1}}} \frac{\partial \hat{\mu}_{I_{F-1}}}{\partial \underline{\sigma}^2_{I_{F-1}}} \end{bmatrix}$$

(Let $diag(x)$ denote a matrix where the diagonal elements is x)

$$= -\underline{\sigma}^{-2} diag(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) + \underline{\sigma}^{-1} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}} \tag{1.100}$$

Back to (1.98),

$$-R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}} + \underline{\sigma}^{-2} diag(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) - \underline{\sigma}^{-1} \frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}} = 0$$

$$\frac{\partial \hat{\mu}_{\{I_f\}}}{\partial \underline{\sigma}^2_{\{I_f\}}} = (R \frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} + \underline{\sigma}^{-1})^{-1} \underline{\sigma}^{-2} diag(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \tag{1.101}$$

3. $\frac{\partial \log|-h^{(2)}|}{\partial \mu_{\{I_f\}}}$

$$\frac{\partial \log|-h^{(2)}|}{\partial \mu_{\{I_f\}}}\Big|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} = \begin{bmatrix} tr((h^{(2)})^{-1} \frac{\partial h^{(2)}}{\partial \mu_{I_1}}) \\ tr((h^{(2)})^{-1} \frac{\partial h^{(2)}}{\partial \mu_{I_2}}) \\ \vdots \\ tr((h^{(2)})^{-1} \frac{\partial h^{(2)}}{\partial \mu_{I_{F-1}}}) \end{bmatrix}_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}} \tag{1.102}$$

$$\frac{\partial h^{(2)}}{\partial \mu_{I_1}} = (-R)\frac{-2\underline{l}_{I_1}e^{\mu_{I_1}}}{(\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F})^3} * \tag{1.103}$$

$$\begin{bmatrix} \underline{l}_{I_1}e^{\mu_{I_1}}(\sum_{f\neq 1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}) & -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_2}e^{\mu_{I_2}} & \cdots & -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} \\ -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_2}e^{\mu_{I_2}} & \underline{l}_{I_2}e^{\mu_{I_2}}(\sum_{f\neq 2}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}) & \cdots & -\underline{l}_{I_2}e^{\mu_{I_2}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} \\ \vdots & \vdots & \ddots & \vdots \\ -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} & -\underline{l}_{I_2}e^{\mu_{I_2}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} & \cdots & \underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}}(\sum_{f\neq F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}) \end{bmatrix} +$$

$$(-R)\frac{1}{(\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F})^2} *$$

$$\begin{bmatrix} \underline{l}_{I_1}e^{\mu_{I_1}}(\sum_{f\neq 1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}) & -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_2}e^{\mu_{I_2}} & \cdots & -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} \\ -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_2}e^{\mu_{I_2}} & \underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_2}e^{\mu_{I_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} & 0 & \cdots & \underline{l}_{I_1}e^{\mu_{I_1}}\underline{l}_{I_{F-1}}e^{\mu_{I_{F-1}}} \end{bmatrix} \tag{1.104}$$

$$= (-R)\frac{-2\underline{l}_{I_1}e^{\mu_{I_1}}}{\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}}\frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}} + (-R)\frac{1}{(\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F})^2}M_{I_1} \tag{1.105}$$

$$\frac{\partial h^{(2)}}{\partial \mu_{I_f}} = (-R)\frac{-2\underline{l}_{I_f}e^{\mu_{I_f}}}{\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F}}\frac{\partial b(\mu_{\{I_f\}})}{\partial \mu_{\{I_f\}}} + (-R)\frac{1}{(\sum_{f=1}^{F-1}\underline{l}_{I_f}e^{\mu_{I_f}}+\underline{l}_{I_F})^2}M_{I_f} \tag{1.106}$$

$$\begin{aligned} &(M_{I_f})_{f,j} = (M_{I_f})_{j,f} = -\underline{l}_{I_f}e^{\mu_{I_f}}\underline{l}_{I_j}e^{\mu_{I_j}}, \forall j \neq f \\ &(M_{I_f})_{f,f} = \underline{l}_{I_f}e^{\mu_{I_f}}(\sum_{ff\neq f}\underline{l}_{I_{ff}}e^{\mu_{I_{ff}}}+\underline{l}_{I_F}) \\ &(M_{I_f})_{j,j} = \underline{l}_{I_f}e^{\mu_{I_f}}\underline{l}_{I_j}e^{\mu_{I_j}}, \forall j \neq f \\ &(M_{I_f})_{i,j} = 0, \forall i,j, \text{other than positions above} \end{aligned} \tag{1.107}$$

We have

$$\left.\frac{\partial h^{(2)}}{\partial \mu_{I_f}}\right|_{\mu_{\{I_f\}}=\hat{\mu}_{\{I_f\}}}, 1 \leq f \leq F-1 \tag{1.108}$$

4. $\frac{\partial \log|-h^{(2)}|}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial \log|-h^{(2)}|}{\partial \underline{\sigma}^2_{\{I_f\}}}$, $\frac{\partial h}{\partial \underline{\mu}_{\{I_f\}}}$, $\frac{\partial h}{\partial \underline{\sigma}^2_{\{I_f\}}}$

$$\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{I_f}} = 0, \quad 1 \leq f \leq F-1 \tag{1.109}$$

$$\frac{\partial \log|-h^{(2)}(\hat{\mu}_{\{I_f\}})|}{\partial \underline{\mu}_{\{I_f\}}} = \begin{bmatrix} tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{I_1}}) \\ tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{I_2}}) \\ \vdots \\ tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{I_{F-1}}}) \end{bmatrix} = 0 \tag{1.110}$$

$$\frac{\partial h(\hat{\mu}_{\{I_f\}})}{\partial \underline{\mu}_{\{I_f\}}} = \underline{\sigma}^{-1}(\hat{\mu}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \tag{1.111}$$

$$[\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{I_f}}]_{f,f} = \frac{1}{\underline{\sigma}^4_{I_f}}, f = 1, \cdots, F-1, \quad 0 \text{ for any other entries in the matrix} \tag{1.112}$$

$$\begin{aligned} \frac{\partial \log|-h^{(2)}(\hat{\mu}_{\{I_f\}})|}{\partial \underline{\sigma}^2_{\{I_f\}}} &= \begin{bmatrix} tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{I_1}}) \\ tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{I_2}}) \\ \vdots \\ tr((h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}\frac{\partial h^{(2)}(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{I_{F-1}}}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\underline{\sigma}^4_{I_1}}[(h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}]_{1,1} \\ \frac{1}{\underline{\sigma}^4_{I_2}}[(h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}]_{2,2} \\ \vdots \\ \frac{1}{\underline{\sigma}^4_{I_{F-1}}}[(h^{(2)}(\hat{\mu}_{\{I_f\}}))^{-1}]_{F-1,F-1} \end{bmatrix} \end{aligned} \tag{1.113}$$

$$\frac{\partial h(\hat{\mu}_{\{I_f\}})}{\partial \underline{\sigma}^2_{\{I_f\}}} = -\frac{1}{2}\frac{1}{|\underline{\sigma}|}\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}}|\underline{\sigma}| - \frac{1}{2}\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}}\left((\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})^T\underline{\sigma}^{-1}(\mu_{\{I_f\}} - \underline{\mu}_{\{I_f\}})\right) \tag{1.114}$$

$$\frac{1}{|\underline{\sigma}|}\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}}|\underline{\sigma}| = \begin{bmatrix} \frac{1}{\underline{\sigma}^2_{I_1}} \\ \frac{1}{\underline{\sigma}^2_{I_1}} \\ \cdots \\ \frac{1}{\underline{\sigma}^2_{I_{F-1}}} \end{bmatrix} \tag{1.115}$$

$$\frac{\partial}{\partial \underline{\sigma}^2_{\{I_f\}}} \left((\hat{\underline{\mu}}_{\{I_f\}} - \underline{\mu}_{\{I_f\}})^T \underline{\sigma}^{-1} (\hat{\underline{\mu}}_{\{I_f\}} - \underline{\mu}_{\{I_f\}}) \right) = \begin{bmatrix} -\frac{(\hat{\mu}_{I_1} - \underline{\mu}_{I_1})^2}{\underline{\sigma}^4_{I_1}} \\ -\frac{(\hat{\mu}_{I_2} - \underline{\mu}_{I_2})^2}{\underline{\sigma}^4_{I_2}} \\ \vdots \\ -\frac{(\hat{\mu}_{I_{F-1}} - \underline{\mu}_{I_{F-1}})^2}{\underline{\sigma}^4_{I_{F-1}}} \end{bmatrix} \tag{1.116}$$

CHAPTER 2

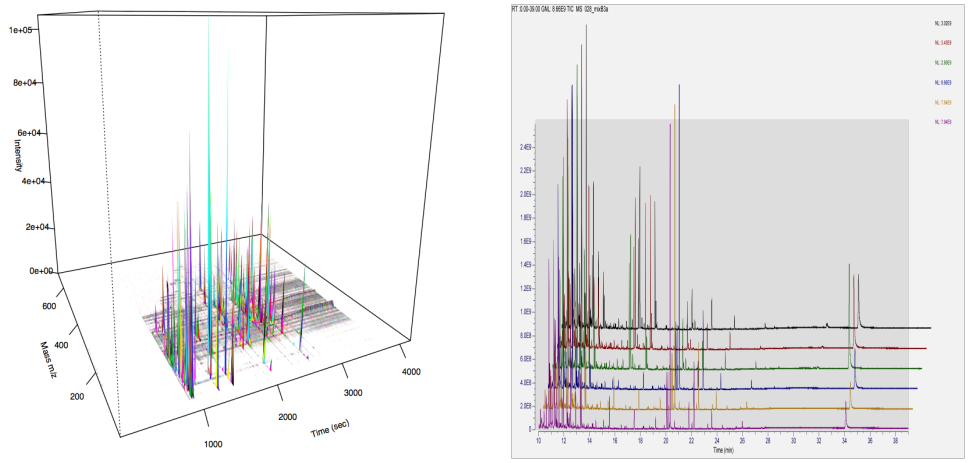
Localized and Simultaneous Non-Negative Matrix Factorization for Deconvolution of Multiple GCMS Signals

2.1 Introduction

Gas Chromatography - Mass Spectrometry (GCMS) is a technique to investigate the metabolome in bio-chemical research. The metabolome is the collection of metabolites and metabolic end-products in a biological system and it reflects the overall activity of the metabolic network that led to their formation by the combined net activity of the genome and proteome. Factors that affect transcription, translation and enzymatic activity will ultimately be reflected in the metabolome. It consists of a wide range of different classes of chemicals, generally less than 2 kDa in molecular weight, including charged and uncharged species, volatile and involatile molecules, lipids, carbohydrates, amino acids and their derivatives, acids, bases, etc. Figure 2.1 shows the GCMS machine.

The GCMS data has a three-dimensional (Figure 2.2) image-like structure: axes of time (chromatographic separation), mass (or more correctly the mass/charge ratio as mass spectrometers measure the mass/charge ratio of charged molecules) and signal intensity (amount of the metabolite in the sample). The data could be represented using a two-dimensional mass-time matrix where entries are intensities. Each metabolite has their own signature in the mass-intensity planes

Figure 2.1: GC and GC/MS with thermal desorption systems



(a) Illustration of one single GCMS data sample.

(b) Illustration of multiple GCMS data sample by merging all their masses.

Figure 2.2: Illustration of GCMS data

(Figure 2.3) in terms of the relative intensities in masses. The retention time of each metabolite is relatively stable but varies across machines and environments. The metabolite enters the map with increasing intensities and decreasing after the peak until the elution ends, which forms a chromatogram peak in the time-intensity planes (Figure 2.4). Metabolites are not 100% separated in GCMS data, thus their overlapping nature poses a deconvolution task.

Methods and software for this deconvolution and metabolite profiling have been around along with development of GCMS technology itself. Early in the 70's, [BB74] tries to identify spectrum by finding peaks across mass slices. This is a rough idea of simultaneously utilizing all masses because peaks from the same

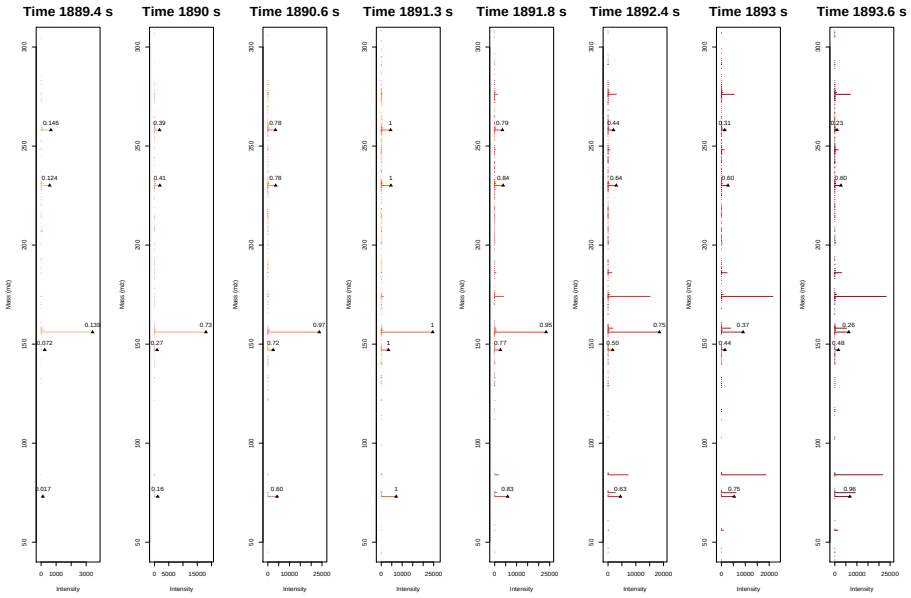


Figure 2.3: Spectrum scans at several retention times

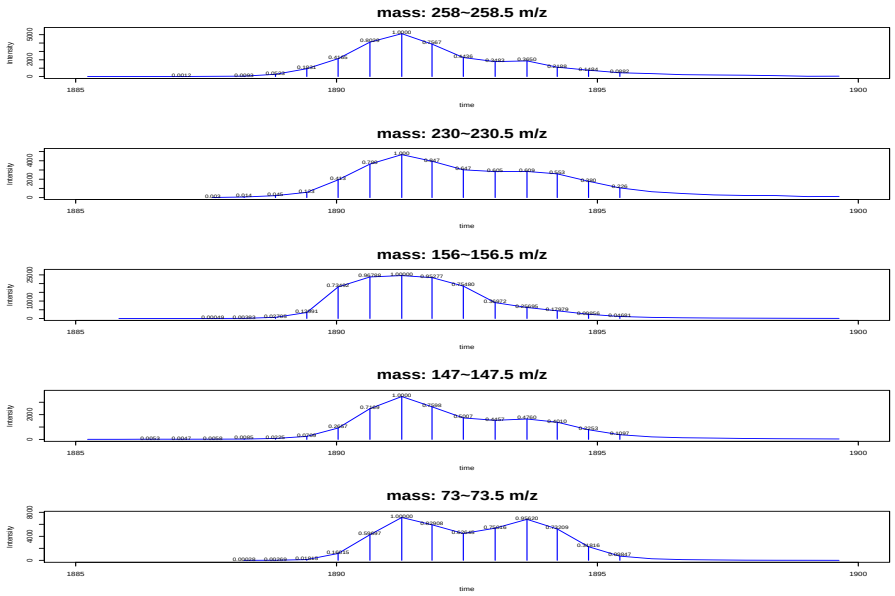


Figure 2.4: Chromatograms at several mass slices

spectra are assumed to have exactly the same shape. [DSRD76] assumes two metabolite spectra don't totally overlap and uses peak counts at every time scan to determine a peak area where there is only one clean spectrum (singlet). The full peak shapes are decided directly from observed data of this singlet spectrum. Similarly to [DSRD76], [Col92] finds groupings of a peak "centroid"s by incorporating all masses that contribute to the centroid using a computationally simpler algorithm to allow practical implementation. [PDV97] locates peaks by a back-folding algorithm which keeps subtracting the two sides of a peak to shorten the peak width until it is sharp enough. Extending [DSRD76], [Ste99] uses signal-to-noise ratio throughout the analysis process, to extract weak signals which would be neglected otherwise. It deals with uncertain peaks and those that are not consistent with model and also calculates a library matching factor for every imputed spectrum. This is an overall extension on every paper above and implemented in the free software **AMDIS**.

All the works above take advantage of clean spectrum scans and lack the ability of deconvolving closely overlapping spectra and of building multiple correspondence across samples. We need a solid model-based approach to address this task.

Matrix factorization methods such as singular value decomposition (SVD) do not serve our purpose as all spectra and chromatograms are non-negative. A reasonable model is non-negative matrix factorization (NMF). However, each sample alone usually has a matrix of size 1k*4k, with unknown true component number r over hundreds. It is impossible to achieve meaningful inference on one sample using direct NMF, not to mention the computational challenge of this task on all samples (1k*400k) simultaneously.

In addition, Three factors further complicate the modeling task. First, many

metabolites elute with very close peak times, which makes the deconvolution task hard. Secondly, for the same metabolite, there exist retention time shifts among various biological samples and replicates. Thirdly, the random noise is present at any given data point in this matrix.

One work that is worth noticing and related to ours is [JGN+04]. Implemented in **ChromaTof** of LECO Corporation, it uses non-negative matrix factorization on local windows to deconvolve the GCMS data matrix across multiple samples. However, the window choices are manually determined, rank choice of each window matrix has not been theoretically justified and there is no merging between windows. With all advantages considered, it still lacks the automation of window selection and merging, the ability of automatic detection of all hidden interesting metabolites and an overall theoretical justification for the model.

This motivates us to build a model based on non-negative matrix factorization and random matrix theory.

Ever since the work of [Wig55] motivated by applications in nuclear physics, there have been rapid developments in random matrix theory (RMT). Particular interests are focused on the eigenvalue distribution (spectral distribution) of various random matrix ensembles. The famous Wigner's semi-circle law serves as a foundation to these developments. In real cases, semi-circle law applies on Wigner matrices that are symmetric and whose elements from diagonal or above are i.i.d random variables with mean 0 and variance 1. It states that the empirical distribution of eigenvalues of a Wigner matrix follows the semicircle distribution. Well-understood results also have the joint eigenvalue distribution of gaussian ensembles. However, these random matrices occur in theoretical physics. One significant mathematical work was [MP67] on spectral distribution of large ran-

dom covariance matrices. A special case from this work was the Marcenko-Pastur Law which describes the empirical distribution of eigenvalues of $\frac{1}{n}X^TX$ where covariance matrix $\Sigma = I$. A special case of it is the quarter-circle law of empirical singular value distribution of X. It has been well studied on the distribution of eigenvalues of $\frac{1}{n}X^TX$ when covariance matrix $\Sigma \neq I$. [BBAP05, BS06], describe the phase transition phenomena of eigenvalues that correspond to the large diagonal elements in Σ (spiked population models where finite number of variables have larger variance). [BY08] provides the central limit theorem for large eigenvalues in spiked population models. [Joh01] discussed the distribution of largest eigenvalue of $\frac{1}{n}X^TX$, where X has i.i.d standard gaussian entries, which approximates the Tracy-Widom distribution [TW94].

[BGGM11] discusses the asymptotic positions and central limit theorem of r largest eigenvalues, and [BGN12] gives the asymptotic and central limit theorem on singular values and asymptotic lints on singular vectors of low-rank deformed rectangular random matrix $\sum_{i=1}^{r} \theta_i u_i v_i^T + X_n$, where u_i, v_i are assumed random. Under a slight change of assumptions, the asymptotic limiting results still hold when u_i, v_i are assumed deterministic, which is exactly the solution in our model. Furthermore, although we do not use it in our model, [CCHM12] addresses the central limit theorem of singular values under the same assumptions of this article.

Numerous computational methods for non-negative matrix factorization have been developed, within which the most popular one is alternate regression, [LS00], [CZPA09]. However, the uniqueness and correctness of the solution in general situations is still not theoretically justifiable. [DS03] suggests the uniqueness is guaranteed if the data is spread across the positive orthant. [LCP+08] provides a few sufficient and necessary conditions for unique NMF, however these conditions are hard to check in practice and are NP-hard [HSS14], [Vav09]. Despite

its lack of theoretical justification for uniqueness and noise further complicates this task, NMF has also shown great strength in practice due to the reasonable non-negativity constraints.

In this article, we propose a rank-based NMF model to streamline the GCMS metabolomic study. We theoretically justify each step of our model based on RMT and NMF. Our model able to automatically split and merge local time windows, estimate rank for each local window, deconvolve overlapping spectra and build multiple correspondence across samples. We demonstrates the strength and automation of our model over any existing GCMS deconvolution methods by showing successful results from our 27 simulations (vary in extent of spectra overlapping, between-sample chromatogram shifts and random noise) and 2 real experimental datasets.

The GCMS data we observe is assume to be a matrix plus random noise. The rank of this matrix is altered by the random noise, which results in full rank for the matrix. We repeatedly use rank estimation and evaluate the similarity between singular vectors of random matrices using RMT. The model procedures can be summarized as 9 major steps (Figure 2.5). 1. A parallel computation based on pseudo-rank estimation is performed on each sample to determine all the time points where two or more spectra start overlapping. We call $s * c_j$ a scan at time t_j. In this process, we call rank$=1$ spectrum scans "sea" and rank≥ 2 spectrum scans "island". We perform NMF on "island" scans that present consecutively in one cluster (no sea scan in between) to get spectra from islands. 2. Cluster "sea" and "island" spectra across all files. 3. We adaptively determine the window splitting process by dynamically computing the estimated rank of selected window and make sure the ranks are small (usually ≤ 6). Each window will incorporate any spectra clusters learned from step 2 and extend itself for each file. Once win-

dows are determined, we perform a combinatorial NMF within each window using individual sample rank and biological group information at this window. 4. A window merging process merges NMF results from all windows into one by measuring the inner product between spectra within or between windows. 5. A large sequential NMF is carried out, in each inner loop we only update one spectrum and chromatogram. 6. Similar nearby spectra are combined into one. 7. We fix our spectra matrix S, and extend the range of C for each sample until it reaches zero to get a full shape. 8. We split multiple peaks in C into different features. 9. After all features and learned, we deploy the multinomial logistic regression with a group lasso penalty used to select significant features that differentiate between groups.

The remainder of this article is organized as follows: 2. Modeling details. 3. Simulation study. 4. Analysis on real datasets. 5. Summary. 6. Future work. 7. Theorems and discussions related to random matrices and nonnegative matrix factorization during some of the modeling steps. 8. Lemmas on random matrices and other existing theorems on uniqueness of nonnegative matrix factorization.

2.2 Modelling

2.2.1 Notation

$\tilde{X} : n \times p$, observed GCMS data matrix.

$X : n \times p$, random noise matrix.

$S : n \times r$, spectrum matrix (dictionary).

$C : r \times p$, chromatogram matrix.

$P : n \times p$, $P = SC = U\Theta_{r\times r}V$, where $U\Theta V$ is a singular value decomposition of P.

$T : p$-dimensional vector denoting the retention time of each column of $\tilde{X}$, X, C.

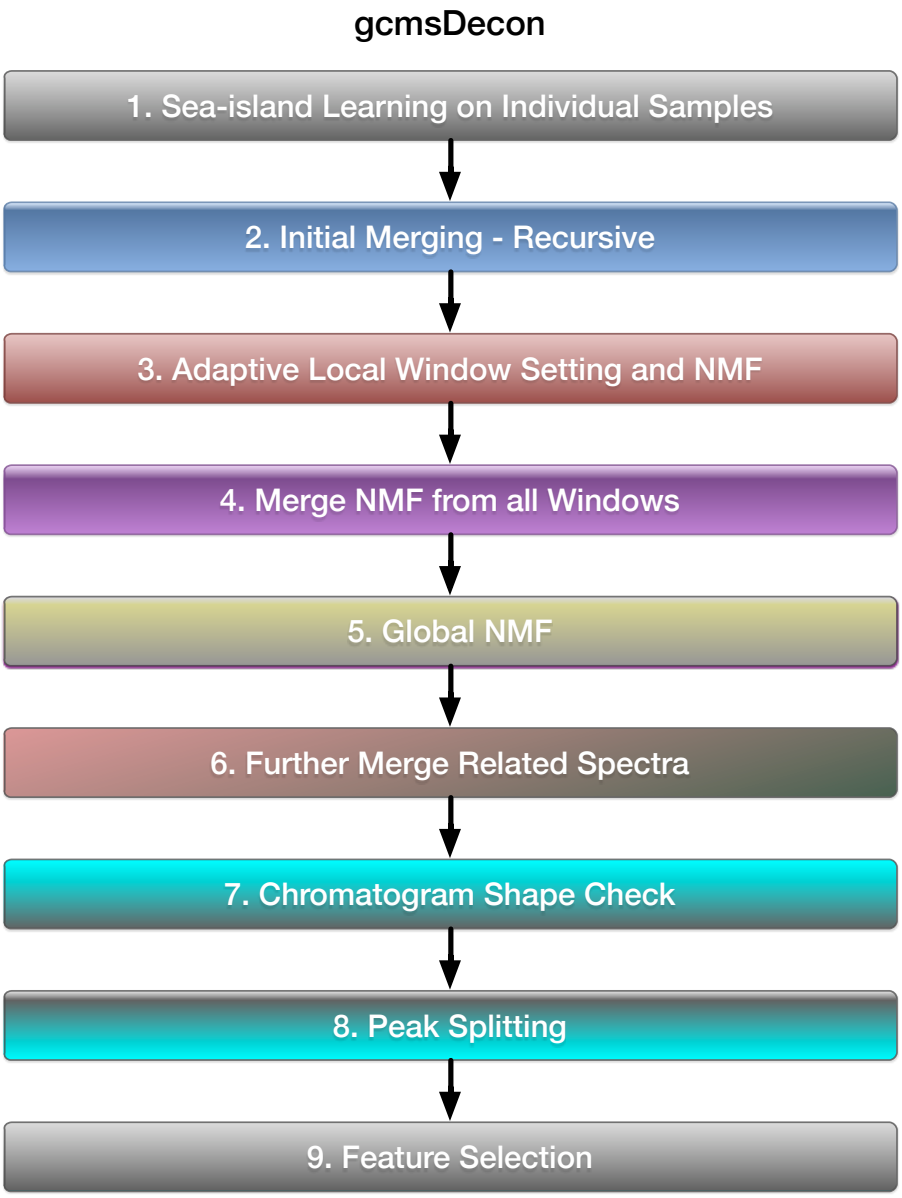


Figure 2.5: 9 Steps of gcmsDecon

Superscript g, k: group g and replicate k, e.g. $\tilde{X}^{g,k}$ is of dimensions $n \times p^{g,k}$.

2.2.2 Model

We assume a shared spectrum matrix S and our model is

$$\tilde{X}^{g,k} = P^{g,k} + X^{g,k} = SC^{g,k} + X^{g,k}$$

$$s.t. \begin{cases} S \geq 0, C^{g,k} \geq 0 & \forall 1 \leq g \leq G, 1 \leq k \leq K_g \\ \forall (g_1, k_1) \neq (g_2, k_2), \left| \max_{t \in \{t: C^{g_1,k_1}_{s,t} > 0\}} T^{g_1,k_1}_t - \min_{t \in \{t: C^{g_2,k_2}_{s,t} > 0\}} T^{g_2,k_2}_t \right| \leq \mathbf{T} & 1 \leq s \leq r \end{cases} \tag{2.1}$$

Since noise $X^{g,k}$ is present at any non-zero observation point, and $\tilde{X}^{g,k} \geq 0$, we assume

$$X^{g,k}_{m,t} \overset{iid}{\sim} \begin{cases} 0 & (SC^{g,k})_{m,t} = 0 \\ \frac{1}{1-\Phi(-(SC^{g,k})_{m,t})} \phi(\frac{x}{\sigma^2}) \mathbf{1}(x \geq -(SC^{g,k})_{m,t}) & (SC^{g,k})_{m,t} > 0 \end{cases}$$

Note: Empirically, most of truncation would be negligible since the variance σ^2 is small compared to $SC^{g,k}$. Thus the distribution $\tilde{X}^{g,k}$ resembles joint i.i.d gaussian $N(0, \sigma^2)$.

The log-likelihood:

$$l = \sum_{g,k} \log P(\tilde{X}^{g,k}; S, C^{g,k})$$

We want to maximize the log-likelihood l and accurately infer $S, C^{g,k}$ from observations $\tilde{X}^{g,k}$.

2.2.3 Estimation

The algorithm aims to maximize the likelihood and infer true $S, C^{g,k}$ by

$$\text{Obj 1} : \min \text{ncol}(S) \quad \text{Obj 2} : \min_{S,C^{\{g,k\}}} \sum_{g=1}^{G} \sum_{k=1}^{K_g} ||\tilde{X}^{g,k} - SC^{g,k}||_F^2 \tag{2.2}$$

$$\approx \min_{S,C^{\{g,k\}}} \sum_{g=1}^{G} \sum_{k=1}^{K_g} ||\tilde{X}^{g,k} - SC^{g,k}||_F^2 + \lambda_S \text{ncol}(S) \tag{2.3}$$

We want to maximize the likelihood as well as make accurate statistical inference. This problem is impossible to solve directly, but every step of our algorithm aims to take care of both objectives at the same time.

Preprocessing - Matrix forming

We bin every data file into a matrix $\tilde{X}^{g,k}$ of size $n * p^{\{g,k\}}$, each sample has slightly different retention time $T^{\{g,k\}}$, therefore this matrix size would slightly differ.

Rank Estimation

Throughout steps of the program, we always need a rank estimator for a given random matrix $\tilde{X} = P + X$, we want to estimate the true rank of P based on singular values of $\tilde{X}$. $\hat{r}^{rbt} = \sum_{i=1}^{n} \mathbf{1}(\tilde{\lambda}_i > \frac{\tilde{\lambda}_1}{\kappa_{rank}})$. Based on ***Theorem 6***, $\hat{r}^{rbt}$ might underestimate the true rank r_0 depending on the choice of κ_{rank} when ≥ 2 spectra with different magnitudes of intensities overlap, but it also is robust against violation of assumptions on noise X. Example see Figure 2.6.

Rank Estimation. (Figure 2.6)

function Rank($\tilde{X}, \kappa_{rank}$)

Calculate singular values of $\tilde{X} \to \tilde{\lambda}_1, \cdots, \tilde{\lambda}_{\min(n,p)}$

return $\sum_{i=1}^{\min(n,p)} \mathbf{1}(\tilde{\lambda}_i > \frac{\tilde{\lambda}_1}{\kappa_{rank}})$

end function

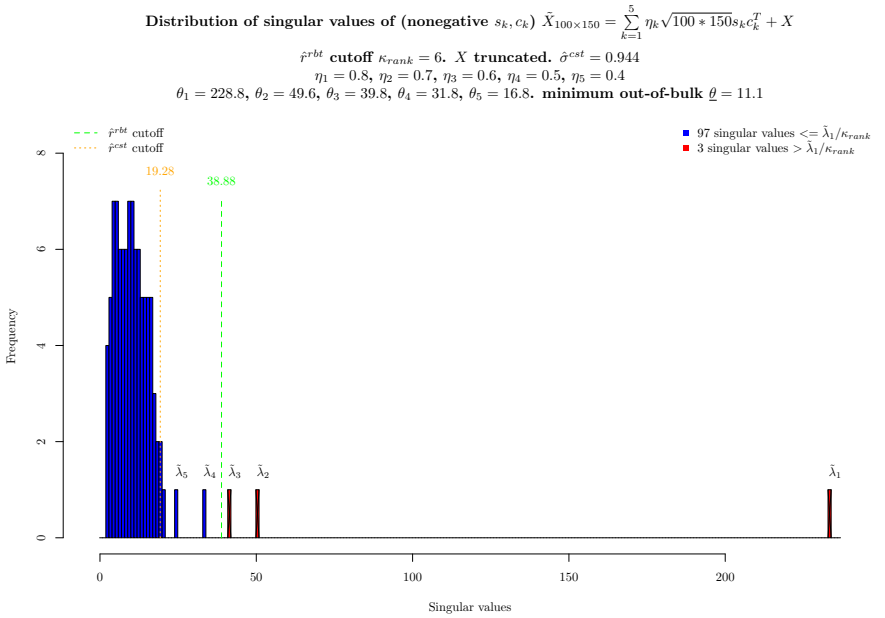


Figure 2.6: Rank estimation. In this example, true $r = 5$, estimated $\hat{r}^{rbt} = 3$.

Step 1. Sea-island Learning on Individual Samples

This step tries to cluster nearby scans if they are similar and represent them using as few spectra as possible. For each g, k,

$$\min \text{ncol}(S^{g,k}) \quad \text{And} \quad \min_{S^{g,k}, C^{g,k}} ||\tilde{X}^{g,k} - S^{g,k} C^{g,k}||_F^2$$

For each file $\tilde{X}^{(g,k)}$, we detect the scan clusters that share the same spectrum. If consecutive scans belong to one spectrum, they form their own cluster. If two spectra overlap, every scan of the overlapping parts should theoretically be its own cluster. We detect whether one scan belongs to previous cluster by computing the inner product between it and the normalized average of previous normalized scans in this cluster $\langle s_{pre}, \tilde{X}^{(g,k)}_{.,j} \rangle$, $s_{pre} = \frac{1}{t_2 - t_1} \sum_{t=t_1}^{t_2-1} \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||_2}$, t_1 is the start of current cluster, and t_2 is the current scan. If this inner product is below a threshold cutoff κ_{si}, we assign the current scan t_2 to a new cluster. Repeat this process until all scans have been processed. In the end, if the size of a cluster is 1, we call it a "island", otherwise a "sea". The process can detect majority of rank-1 and overlapping scans and misidentify only for scans with small intensities (usually in the

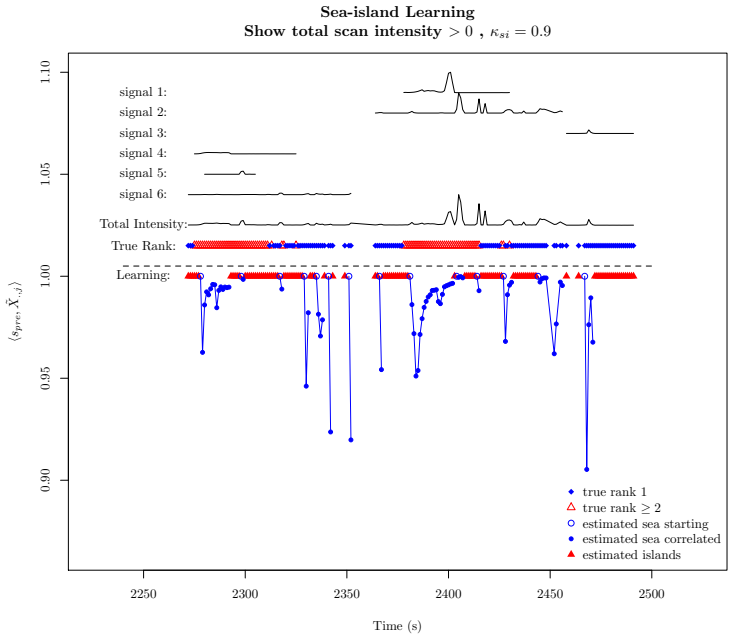


Figure 2.7: Sea-island learning. Connected scans are "sea" denoted by blue and "islands" are denoted by red. Under the dashed line is sea-island learning process. The blue dot clusters produce S_{sea}, and red dots clusters produce S_{island}. The correlation cutoff here is chosen to be 0.9.

beginning and end of a peak). Refer to ***Discussion 8***. Results of this step is to be used later in window selection process, where new joint spectra would be relearned for all files simultaneously. Example see Figure 2.7.

Step 2. Initial Merging

We learn the correspondence between files by find clusters of their sea-island spectra. $S_{sea}^{\{g,k\}}, T_{sea}^{\{g,k\}}, S_{island}^{\{g,k\}}, T_{island}^{\{g,k\}}$. We aim to merge similar close-by spectra as many as possible, because they are very likely to be the same compound. The time range of merged spectra are used to help window splitting. Let S_{sea_merge}, S_{island_merge} to be the merged spectra. Indicator matrix $\xi_{sea}^{g,k}$ of dimension $\text{ncol}(S_{sea\_merge}) * L_{sea}^{g,k}$, whose columns and rows have all 0 except for one 1, indicating which merged

Sea-island Learning. (Figure 2.7)

Require: Inner product cutoff κ_{si} (e.g., 0.95), κ_{rank} (e.g., 6), $\tilde{X}^{g,k}$ from Matrix Forming.

Ensure: "Sea" and "island" cluster spectrum matrices S_{sea}, $n \times L_{sea}$, S_{island}, $n \times L_{island}$ and their corresponding time range T_{sea}, $3 \times L_{sea}$, T_{island}, $3 \times L_{island}$.

for g in $1:G$, k in $1:K_g$ **do**

$p \leftarrow p^{g,k}$, $T \leftarrow T^{g,k}$, $\tilde{X} \leftarrow \tilde{X}^{g,k}$, initialize scan cluster labels $l = [l_1, \cdots, l_p] \leftarrow [1, \cdots, 1]$. Indices $t_1 \leftarrow 1$, $t_2 \leftarrow 2$, total number of clusters $L \leftarrow 1$. $S_{pool} = []$, $T_{pool} = []$.

while $t_2 \leq p$ **do**

$s_{pre} \leftarrow \frac{1}{t_2 - t_1} \sum_{j=t_1}^{t_2-1} \frac{\tilde{X}_{\cdot,j}}{||\tilde{X}_{\cdot,j}||_2}$, $s_{pre} \leftarrow \frac{s_{pre}}{||s_{pre}||_2}$

If $\langle s_{pre}, \frac{\tilde{X}_{\cdot,t_2}}{||\tilde{X}_{\cdot,t_2}||_2} \rangle \geq \kappa_{si}$, **then** $l_{t_2} \leftarrow L$, $s_{pre} \leftarrow \frac{1}{t_2 - t_1 + 1} \sum_{j=t_1}^{t_2} \frac{\tilde{X}_{\cdot,j}}{||\tilde{X}_{\cdot,j}||_2}$, $s_{pre} \leftarrow \frac{s_{pre}}{||s_{pre}||_2}$. **else** $S_{pool} \leftarrow [S_{pool}, s_{pre}]$, $l_{t_2} \leftarrow L + 1$, $L \leftarrow L + 1$, $t_1 \leftarrow t_2$.

$t_2 \leftarrow t_2 + 1$

end while

e.g. of labels, see Figure 2.8.

$S^{g,k}_{sea} \leftarrow$ columns of $[S_{pool}]_{n \times L}$ whose Freq > 1. $T^{g,k}_{sea} \leftarrow$ range of T by labels.

$S^{g,k}_{island} \leftarrow$ NMF on these cluster of scans with Freq 1 with the help of nearby sea scans, $\min_{[S_{nmf}]_{n \times r_0}, [C]_{r_0 \times (t_2 - t_1 - 1)}} ||\tilde{X}_{nmf} - [s_{left\_sea}, S_{nmf}, s_{right\_sea}]C||_F^2$.

end for

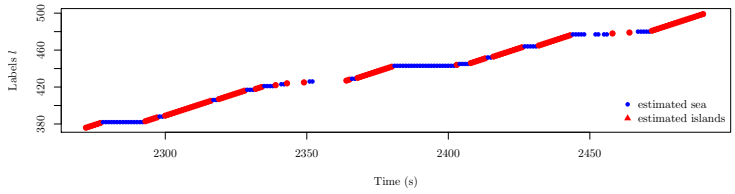


Figure 2.8: Labels of Figure 2.7 example.

spectrum the individual spectrum of $S_{sea}^{\{g,k\}}$ belongs to.

$$\min \text{ncol}(S_{sea\_merge}) \text{ AND } \min_{S_{sea\_merge}, \delta_{sea}^{g,k}} \sum_{g,k} ||S_{sea}^{g,k} - S_{sea\_merge}\xi_{sea}^{g,k}||_F^2$$

$$s.t. \quad \forall \text{ spectrum in } S_{sea\_merge}, \text{ the time range of its belonging } S_{sea}^{\{g,k\}} \text{is within } \mathbf{T}$$

Similar objective goes for island spectra $S_{island}^{\{g,k\}}$. This step is run in a recursive way. Each step we find a cluster and the spectra across files belonging to this cluster are killed off, then the next cluster is found within the rest of spectra. Each sample contributes once in one merged spectrum. Each time we find the biggest cluster from a joint matrix of one spectrum from every file, $Y = [S_{sea_{.,i^{1,1}}}^{1,1}, \cdots, S_{sea_{.,i^{G,K_G}}}^{G,K_G}]$, $r = \text{Rank}(Y)$, split Y into r clusters (this number does not have to be r) using k-means algorithm and pick out the largest cluster. This process runs until all sea-island spectra are in their own merged clusters. This result is prepared for later window splitting. Illustration please see Figure 2.9.

Step 3. Adaptive Local Window Setting and NMF

Dynamically split time range into windows (Figure 2.10), each data file in j-th window is $\tilde{X}_{local_j}^{g,k}$, and $\text{Rank}(\tilde{X}_{local_j}^{g,k}, \kappa_{rank}) \leq 6, \forall g, k$.

$$\forall \text{ window } j, \quad \min \text{ncol}(S_{local_j}) \quad \text{AND} \quad \min_{S, C^{\{g,k\}}} \sum_{g=1}^{G} \sum_{k=1}^{K_g} ||\tilde{X}_{local_j}^{g,k} - S_{local_j} C_{local_j}^{g,k}||_F^2 \tag{2.4}$$

We want to avoid computation on NMF with a big number of components, which is challenging both theoretically and empirically. We dynamically split retention time range (e.g. 0 4800 seconds) into multiple windows (e.g. [10, 30), [30, 50), etc.). We compute the rank of data $X_{local_j}^{g,k}$ within this window (refer to ***Theorem 6***). If any of them is larger than a threshold (usually 6), we split the windows until all window matrices have ranks $\leq$ threshold. This guarantees a low-rank fixed window scheme for all files. In order to learn full chromatogram shapes, we

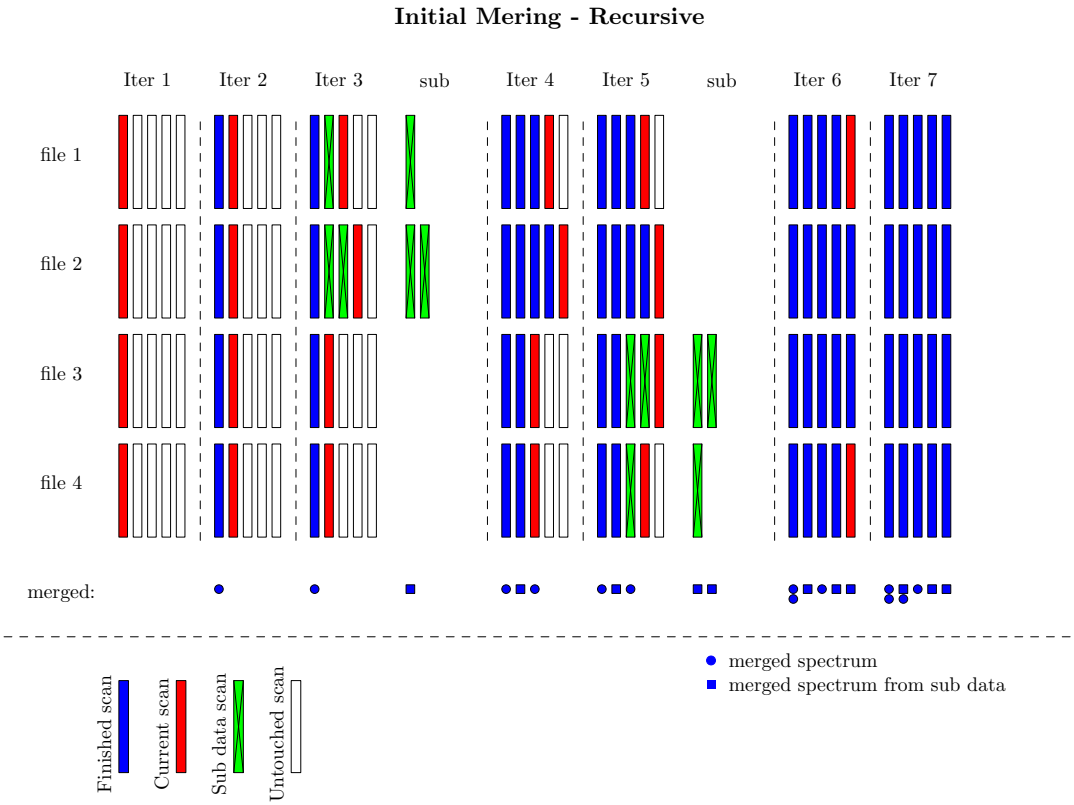


Figure 2.9: Initial recursive merging. Red scans Y are currently in computation, skipping of red scans (2nd to 3rd step) means searching with perturbation. Blue denotes computed scans and green denotes the scans forming a sub-data $S'^{\{k\}}$ in a recursive function call.

Initial Merging - Recursive. (Figure 2.9)

Require: $[S_{sea}^{\{g,k\}}]_{n\times L_{sea}^{\{g,k\}}}, T_{sea}^{\{g,k\}}, [S_{island}^{\{g,k\}}]_{n\times L_{island}^{\{g,k\}}}, T_{island}^{\{g,k\}}, \mathbf{T}$

Ensure: $S_{sea\_merge}, T_{sea\_merge}, S_{island\_merge}, T_{island\_merge}, \xi_{sea\_merge}^{\{g,k\}}, \xi_{island\_merge}^{\{g,k\}}$.

Take "sea" as an example. For simplicity, use k in $1:K$ instead of (g,k).

function InitialMerge($S_{sea}^{\{k\}}$, $T_{sea}^{\{k\}}$, $\mathbf{T}$)

$S_{sea\_merge} \leftarrow []$, $T_{sea\_merge} \leftarrow []$, $t_{sea\_merge}^{\{g,k\}} \leftarrow []$. Initialize $i^k = 1, 1 \le k \le K$.

while $\exists k$, $i^k \le L_{sea}^k$ **do**

$Y = [S_{sea.,i^1}^1, \cdots, S_{sea.,i^K}^K]$, $r = \text{Rank}(Y)$, group K spectra into r clusters using k-means algorithm $\rightarrow$ cluster index vector $\delta^h = [\delta_1^h, \cdots, \delta_{p_h}^h]$, $\sum_{h=1}^{r} p_h \le K$.

$\forall 1 \le h \le r$, $s_{curh} \leftarrow \frac{1}{p_h}\sum_{j=1}^{p_h} S_{sea_{.,i^{\delta_j^h}}}^{\delta_j^h}$, $s_{curh} \leftarrow \frac{s_{curh}}{||s_{curh}||_2}$, $T_{curh} \leftarrow \frac{1}{p_h}\sum_{j=1}^{p_h} T_{sea_{i^{\delta_j^h}}}^{\delta_j^h}$.

Base: if $L_{sea}^k \le 1$, $\forall k$, **return** $S_{merge\_sea} \leftarrow \{s_{cur}\}_{1:r}$, $T_{merge\_sea} \leftarrow \{T_{cur}\}_{1:r}$, $\xi_{sea\_merge}^k \leftarrow \{i^{\delta_h}\}_{1:r}$.

$h \leftarrow \text{argmax}\, p_h$. Ignore this subscript h for s_{curh}, T_{curh}, δ_h below.

$\eta \leftarrow \delta^c = [\eta_1, \cdots, \eta_{|\eta|}]$ denotes samples not in cluster h_{max} with their scans i^η. $\forall 1 \le j \le |\eta|$, search for $i_2^{\eta_j}$ in its neighborhood $\pm\mathbf{T}$ of T_{cur} such that $\langle s_{cur}, S_{sea_{.,i_2^{\eta_j}}}^{\eta_j}\rangle \ge \kappa_{im}$. Delete η_j from η if no scan located. Update $s_{cur} \leftarrow \frac{1}{p+|\eta|}(\sum_{j=1}^{p} s_{sea_{.,i^{\delta_j}}}^{\delta_j} + \sum_{j=1}^{|\eta|} S_{sea_{.,i_2^{\eta_j}}}^{\eta_j})$, $s_{cur} \leftarrow \frac{s_{cur}}{||s_{cur}||}$, $T_{cur} \leftarrow \frac{p}{p+|\eta|}T_{cur} + \frac{1}{p+|\eta|}\sum_{j=1}^{|\eta|} T_{sea_{2,i_2^{\eta_j}}}^{\eta_j}$, $\xi_{cur} \leftarrow \{i^k, k \in \delta; i_2^k, k \in \eta\}$.

Recursion: {Due to the scan skipping in this search for samples η, we define a submatrix $S'^k_{sea} = S^k_{sea.,i^k:(i_2^k-1)}$, $T'^k_{sea} = T^k_{sea.,i^k:(i_2^k-1)}$, $\forall k = 1, \cdots, K$. $S'^k_{sea} = []$ if $k \notin \eta$.

$S'_{sea\_merge}, T'_{sea\_merge} \leftarrow \text{InitialMerge}(S'^{\{k\}}_{sea}, T'^{\{k\}}_{sea})$. }

$S_{merge\_sea} \leftarrow [S_{merge\_sea}, S'_{merge\_sea}, S_{cur}]$, $T_{merge\_sea} \leftarrow [T_{merge\_sea}, T'_{merge\_sea}, T_{cur}]$, $t_{sea\_merge}^k \leftarrow [\xi_{sea\_merge}^k, \xi'^k_{sea\_merge}, \xi_{cur}]$.

end while

return S_{merge_sea}, T_{merge_sea}

end function

expand the window by incorporating the range of any sea-island spectrum that falls in this window. We check the ranks of expanded $X^{g,k}_{local_j}$, if any individual rank is above the threshold, we dynamically truncate its window size until the rank falls equal or below threshold.

After windows are determined this way, we compute the NMF with the individual file ranks. r is set to be the largest file rank. When computing the C step in NMF, we choose the median of its group file ranks r_g as the number of spectra in $S_{n\times r}$. Doing this would require $\binom{r}{r_g}$ combinations, which is the reason we want r to be small (≤ 6). The rationale behind this group learning is we assume samples from one group would have similar spectra. Sometimes a diseases group of samples would share a metabolite that does not exist in the control group.

Step 4. Merge NMF from all Windows

$$\min \text{ncol}(S_{joint}) \text{ AND } \min_{S_{joint}, C^{\{g,k\}}_{joint}} \sum_{g,k} ||\tilde{X}_{g,k} - S_{joint} C^{g,k}_{joint}||^2_F$$

$$\text{AND } \min_{S_{joint}, \xi_{joint}} \sum_{j} ||S_{local_j} - S_{joint} \xi^{j}_{joint}||^2_F$$

ξ^{j}_{joint} is the indicator matrix of dimension ncol(S_{joint})*ncol(S_{local_j}) whose columns have all zero but one 1, indicating which spectrum in S_{joint} this spectrum in S_{local_j} belongs to.

The way we split data into different time windows in previous step might truncate the chromatograms, so that we want to combine NMF results from different windows $S^{\{j\}}_{local}$, $C^{\{j\}}_{local}$ to recover the spectrum profile. We compute the inner products for spectra within its own window and between spectra of consecutive windows. Based on ***Discussion 7***, if there are spectra from consecutive windows $\geq \kappa_{mw}$ (e.g. 0.95), we identify them as one spectrum. The same spectrum in different windows differ slightly due to random noise and inner products between different

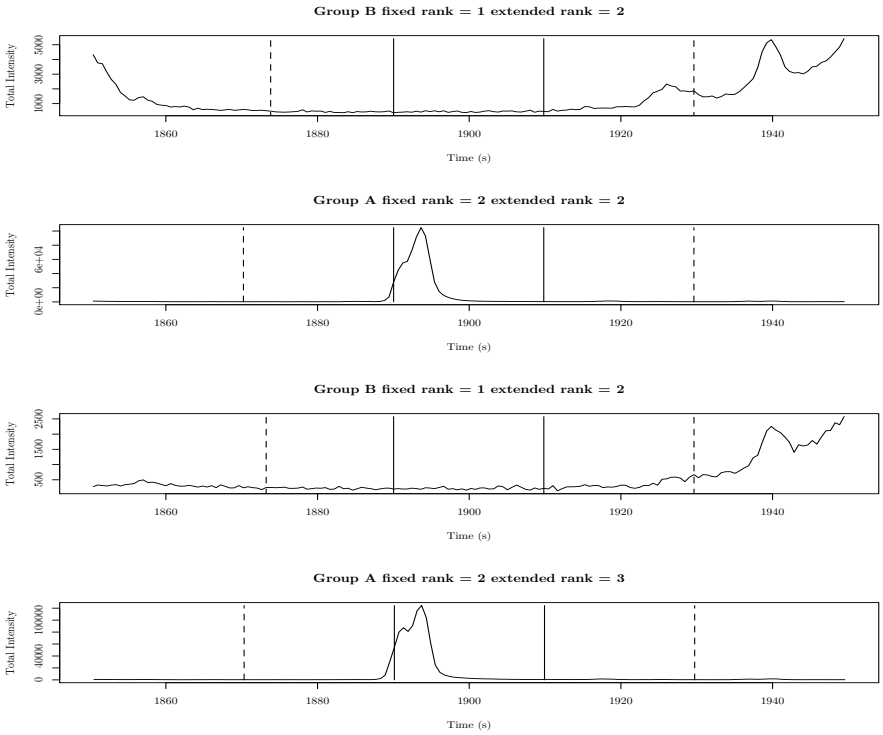


Figure 2.10: Adaptive local NMF. Solid line denotes fixed windows and dashed one denotes extended windows. NMF is calculated once in one window simultaneous for all samples. The window is extended due to the presence of S_{sea_merge}, S_{island_merge} from previous step within fixed windows (solid lines). The windows are extended under reasonable limits.

Adaptive Local NMF. (Figure 2.10)

Require: $\{S, T, \xi^{\{g,k\}}\}_{sea\_merge}, \{S, T, \xi^{\{g,k\}}\}_{island\_merge}$, $\tilde{X}^{\{g,k\}}$, $T^{\{g,k\}}$, $\mathbf{r} \leq 6$, $\mathbf{T}(e.g.10)$.

Ensure: $S^{\{j\}}_{local}$, $C^{\{j\}}_{local}$, T_{win}

Window width set as $2\mathbf{T}$. For simplicity, assume $\frac{T_0}{2\mathbf{T}}$ is an integer,

$$T_{win} = \begin{bmatrix} -\mathbf{T} & \mathbf{T} & \cdots & T_0 - \mathbf{T} \\ \mathbf{T} & 3\mathbf{T} & \cdots & T_0 + \mathbf{T} \end{bmatrix}_{2\times(\frac{T_0}{2\mathbf{T}}+1)}$$

for $j = 1 : \text{ncol}(T_{win})$ **do**

$$t^{g,k}_{local_{\cdot,j}} = \begin{bmatrix} \underset{t:\, T_{win1,j} \leq T^{g,k}_t \leq T_{win2,j}}{\text{argmin}} T^{g,k}_t & \underset{t:\, T_{win1,j} \leq T^{g,k}_t \leq T_{win2,j}}{\text{argmax}} T^{g,k}_t \end{bmatrix}^T$$

$$r^{g,k}_j = \text{Rank}(\tilde{X}^{g,k}_{\cdot, t^{g,k}_{local_{1,j}} : t^{g,k}_{local_{2,j}}}, \kappa_{rank})$$

5: **end for**

$T_{win} \leftarrow \{$**repeat**, if $\forall j$ that $r^{g,k}_j > \mathbf{r}$, break j-th window in half and expand the total number of windows by 1. Recalculate $r^{g,k}_{j_1}$, $r^{g,k}_{j_2}$. **until** $\forall j$, $r^{g,k}_j \leq \mathbf{r}\}$

for $j = 1 : \text{ncol}(T_{win})$ **do**

$\forall g, k$, $X^{g,k}_{local_j} \leftarrow$, scans in $\tilde{X}^{g,k}$ whose $T^{g,k} \in [T_{win1,j}, T_{win2,j})$. $r^{g,k}_j \leftarrow \text{Rank}(X^{g,k}_{local_j}, \kappa_{rank})$. Find i's whose range of $T_{merge\_sea_{,i}} \cap [T_{win1,j} - \mathbf{T}, T_{win2,j} + \mathbf{T}) \neq \emptyset$. The corresponding $\xi^{g,k}_{sea\_merge_i}$ points to extra times in $T^{g,k}$ that should be included in $X^{g,k}_{local_j}$. Dynamically shrink this expansion if new $r^{g,k}_j > \mathbf{r}$.

$r_{local_j} \leftarrow \max_{g,k} r^{g,k}_j \leq \mathbf{r}$, $r^g_{local_j} \leftarrow \underset{k}{\text{median}}\ r^{g,k}_j$, $g = 1 : G$. Perform localized NMF,

$$\min_{[S_{local_j}]_{n\times r_{local_j}}, C^{\{g,k\}}_{local_j}} \sum_{g,k} ||\tilde{X}^{g,k}_{local_j} - S_{local_j} C^{g,k}_{local_j}||^2_F$$

10: **Alternate Regression**: Randomize initial S_{local_j},

repeat

C-step: Update $C_{local\ j}^{g,\{k\}}$, for each group g simultaneously by comparing all $\binom{r_{local\,j}}{r_{local\,j}^{g}}$ possible combinations of columns of $S_{local\,j}$, $n \times r_{local\,j}$.

for all g **do**

Let δ^g be a combinatorial choice $r_{local\,j}^{g}$ out of 1 to $r_{local\,j}$, denoting nonzero rows in $C_{local\,j}^{g,k}$, compute $\hat{\delta}^g, \hat{C}_{local_j}^{g,\{k\}} \leftarrow \min_{\delta^g, C_{local\ j}^{g,\{k\}}} \sum_k ||\tilde{X}_{local_j}^{g,k} - [S_{local\,j}]_{.,\delta^g}[C_{local\,j}^{g,k}]_{\delta^g,.}||_F^2$

if $\exists k$ in group g whose $r_j^{g,k} > r_{local\,j}^{g}$ **then**

Recalculate $\hat{C}_{local_j}^{g,k}$. Pick additional $r_j^{g,k} - r_{local\,j}^{g}$ spectra in the rest columns of $S_{local\,j}$, $\hat{\eta}^g = \{i : i \notin \hat{\delta}^g, 1 \leq i \leq r_{local\,j}\}$. There are $\binom{r_{local\,j} - r_{local\,j}^{g}}{r_j^{g,k} - r_{local\,j}^{g}}$ all possible choices. Let $\zeta^{g,k} = [\zeta_1^{g,k}, \cdots, \zeta_{r_j^{g,k} - r_{local\,j}^{g}}^{g,k}]$ denote the chosen indices out of 1 to $r_{local\,j} - r_{local\,j}^{g}$. Compute $\hat{\zeta}^{g,}, \hat{C}_{local_j}^{g,\{k\}} \leftarrow \min_{\zeta^{g,k}, C_{local\,j}^{g,k}} ||\tilde{X}_{local_j}^{g,k} - [S_{local\,j}]_{.,(\hat{\delta}^g, \hat{\eta}_{\zeta^{g,k}}^{g})}[C_{local\,j}^{g,k}]_{(\hat{\delta}^g, \hat{\eta}_{\zeta^{g,k}}^{g}),.}||_F^2$

end if

end for. End of C-step

S-step: Compute $\hat{S}_{local_j}^{T} \leftarrow \min_{S_{local\,j}^{T}} \sum_{g,k} |||[\tilde{X}_{local_j}^{g,k}]^T - [C_{local\,j}^{g,k}]^T S_{local\,j}^{T}||_F^2$.

until Error small enough

end for

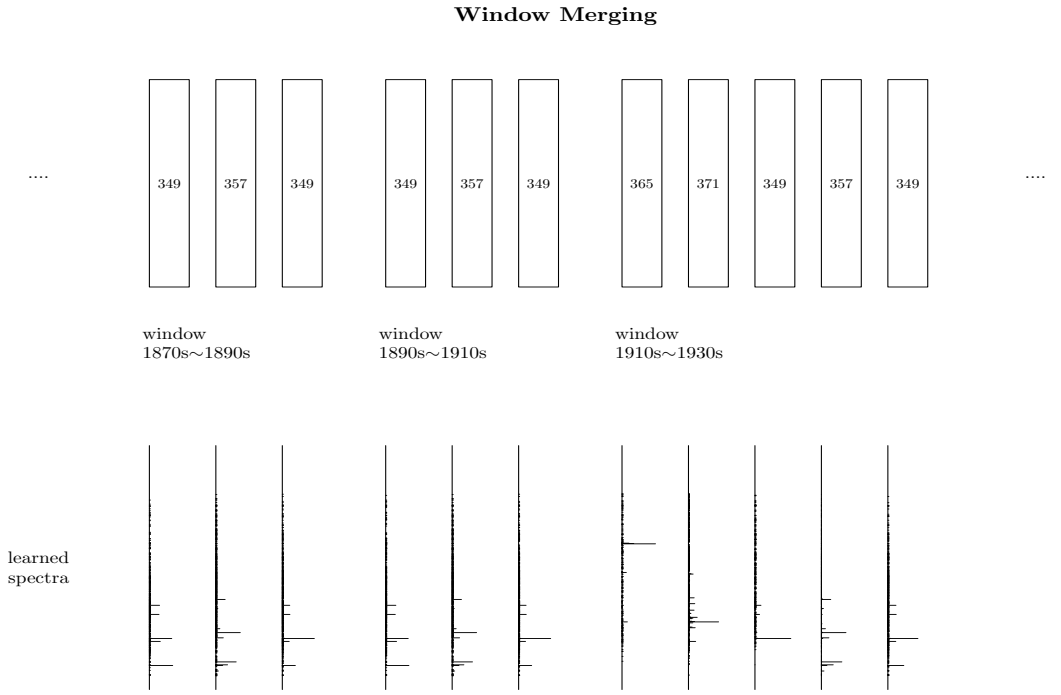


Figure 2.11: Merge local windows. The label numbers are results of spectrum inner products within and between windows. E.g., spectra with label 349 across the three windows are all learned differently but similar, they actually represent one spectrum with chromatogram across all three windows.

spectra are generally much lower. In our experience, we have not encountered a lot of false identifications. Doing this not only improves accuracy of spectrum learning, but also identifies the true spectrum chromatograms. See Figure 2.11.

Step 5. Global NMF

$$\min_{S_{global}, C_{global}^{\{g,k\}}} \sum_{g,k} ||\tilde{X}_{g,k} - S_{global} C_{global}^{g,k}||_F^2 \quad \text{with initial values } S_{global} = S_{joint}, C_{global}^{\{g,k\}} = C_{joint}^{\{g,k\}}$$

Once we combine all windows in to two large S, C matrices, with the original large data matrix $\tilde{X}$ of all files. We only need to minimize $||\tilde{X} - SC||_F^2$ in a sequential way starting with S, C we got from window merging. Sequential way means in

Merge NMF windows. (Figure 2.11)

Require: S_{local_j}, $C_{local_j}^{\{g,k\}}$, T_{win}, $\tilde{X}^{\{g,k\}}$, κ_{cor} (e.g. 0.95)

Ensure: S_{joint}, $C_{joint}^{\{g,k\}}$

for $j = 1$ to ncol(T_{win}) **do**

Indices of similar spectra within window j. $B_{self_j} \leftarrow \{(i,k) : \langle [S_{local_j}]_{.,i}, [S_{local_j}]_{.,k} \rangle \geq \kappa_{cor})$. Indices of similar spectra between window j and its next $j+1$. $B_{next_j} \leftarrow \{(i,k) : \langle [S_{local_j}]_{.,i}, [S_{local_{j+1}}]_{.,k} \rangle \geq \kappa_{cor}\}$

end for

Labeling: Any pair of spectra appears either in $B_{self_{\{j\}}}$ or $B_{next_{\{j\}}}$ are assigned the same label. If multiple pairs of spectra overlap, assign one label number for all of them. Every unique label represents a cluster (chain) of similar spectra learnt from all windows. $\rightarrow label_j = [l_{j1}, \cdots, l_{j_{ncol(S_{local_j})}}]$ for S_{local_j}, j = 1:ncol(T_{win}). $L_c = \max_{i,j} label_j[i]$.

5: Let $chain_i \leftarrow \{(w_p^i, m_p^i)\}$, $1 \leq i \leq L_c$ stores the window number and spectrum number for i-th chain. $profall_{i,p}^{g,k} \leftarrow [C_{local}{}_{w_p^i}^{g,k}]_{m_p^i,.}$, the chromatogram score of file g,k using p-th spectrum in $chain_i$. Reorganize $profall_{i,\{p\}}^{g,k}$ so that they can be written in one matrix.

$$profall_i^{g,k} = \begin{bmatrix} t_1 & t_2 & \cdots & \cdots \\ c_{1,1} & c_{1,2} & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots \\ c_{p,1} & c_{p,2} & \cdots & \cdots \end{bmatrix}.$$ To compute one spectrum for this $chain_i$, we need to subtract other chains from the data.

Initialize $S_{joint} \leftarrow 0$, $n \times L_c$, $C_{joint}^{\{g,k\}} \leftarrow 0$, $L_c \times p^{\{g,k\}}$.

Window Choice:. $win^{g,k} = [win_1^{g,k}, win_2^{g,k}, win_3^{g,k}, \cdots, win_{p^{g,k}}^{g,k}]$. $\forall g, k$

for t in 1:$p^{g,k}$ **do**

$win_t^{g,k} \leftarrow \underset{j}{\text{argmin}} \, ||\tilde{X}_{.,t}^{g,k} - S_{local_j}[C_{local_j}^{g,k}]_{.,t}||_F^2$

end for

To ensure continuity of window choosing, if $\exists t$, $win_t^{g,k} < win_{t-1}^{g,k}$, let $win_t^{g,k} \leftarrow win_{t-1}^{g,k}$.

for $i = 1, \cdots, L_c$ **do**

for all g, k **do**

For every scan $t \in [{t_{start}}_i^{g,k}, {t_{end}}_i^{g,k}]$ in $profall_i^{g,k}$, its best window choice $j \leftarrow win_t^{g,k}$ and all the window numbers in $chain_i$, $w_{\{p\}}^i$.

if $j \in w_{\{p\}}^i$, then its spectrum numbers $s \leftarrow \{m_p^i : p \text{ that } w_p^i = j\}$. Lets s^c denote the rest spectra in this window. Compute the residual $E_i^{g,k}{}_{.,t} \leftarrow \tilde{X}_{.,t}^{g,k} - S_{local_j}{}_{.,s^c} C_{local_j}{}_{s^c,.}^{g,k}$. Only need to keep these in $[t_{start}i, t_{end}i]$, $E_i^{g,k} \leftarrow [E_i^{g,k}]_{,{t_{start}}_i^{g,k}:{t_{end}}_i^{g,k}}$.

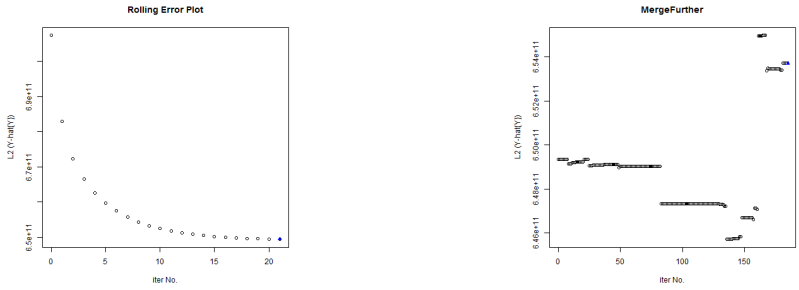
else, $E_i^{g,k} \leftarrow []$.

end for

Compute $\hat{s}_{target_i}, \hat{C}_{target_i}^{g,k} \leftarrow \min_{[s_{target_i}]_{m_0 \times 1}, C_{target_i}^{g,k}} \sum_{g,k} ||E_i^{g,k} - s_{target_i} C_{target_i}^{g,k}||$

$S_{joint}{}_{.,i} \leftarrow \hat{s}_{target_i}$, **for all** g, k, $C_{joint}{}_{i,t_{start_i}^{g,k}:t_{end_i}^{g,k}}^{g,k} \leftarrow \hat{C}_{target_i}^{g,k}$.

end for



(a) Global NMF error change (b) Further merging error change

Figure 2.12: Global NMF and Further Merging.

each inner loop i we $\min_{S_{.,i}C_{i,.}} ||\tilde{X} - S_{.,-i}C_{-i,.} - S_{.,i}C_{i,.}||_F^2$. This step usually take $\leq$100 outer loops to converge. Example of error changes see Figure 2.12a.

Step 6. Further Merge Related Spectra

$$\min_{S_{fur}, C_{fur}^{\{g,k\}}} \sum_{g,k} ||\tilde{X}_{g,k} - S_{fur}C_{fur}^{g,k}||_F^2 \quad \text{with initial values } S_{fur} = S_{global}, C_{fur}^{\{g,k\}} = C_{global}^{\{g,k\}}$$

$$\text{AND } \min \text{ncol}(S_{fur})$$

Within S, after the global NMF, there are chances they still resemble each other, we further merge this kind of spectra into one. The error might increase here since we are reducing the model complexity. Example of error changes see Figure 2.12b.

Step 7. Chromatogram Shape Check

$$\min_{C_{shape}^{\{g,k\}}} \sum_{g,k} ||\tilde{X}_{g,k} - S_{shape}C_{shape}^{g,k}||_F^2 \quad \text{with initial values } S_{shape} = S_{fur}, C_{shape}^{\{g,k\}} = C_{fur}^{\{g,k\}}$$

In the initial sea-island learning, this is a chance that overlapping spectra do not show as "island" if the intensities from one spectrum is too low. Here, after all spectra have been learned, we extend C of corresponding spectra for every file, to make sure the spectra explain as much data as possible. This step is also important to retrieve a full chromatogram peak shape. See Figure 2.13.

Global NMF. (Figure 2.12a)

Require: S_{joint}, $C_{joint}^{\{g,k\}}$

Ensure: S_{global}, $C_{global}^{\{g,k\}}$

Let $\delta = [\delta_1, \cdots, \delta_{L_c}]$, be the order of spectra in S_{joint} from big to small. The relative magnitude is measured by maximum in the corresponding row of $C_{joint}^{\{g,k\}}$.

Let $S_{global} \leftarrow S_{joint}$,

repeat{

for $i = 1, \cdots, L_c$ **do**

$$\min_{[s_{global}]_{m_0 \times 1}, c_{global}^{g,k}} \sum_{g,k} ||I^{g,k} - S_{joint}{}_{.,\delta_i^c} C_{joint}{}_{\delta^c,.}^{g,k} - s_{global} c_{global}^{g,k}||_F^2.$$

$S_{global}{}_{.,\delta_i} \leftarrow s_{global}, C_{joint}{}_{\delta,.}^{g,k} \leftarrow c_{global}^{g,k}$

end for

until Error change is small

Delete columns in S_{global} whose corresponding rows in $C_{global}^{g,k}$ are all zero.

Further Merge Related Spectra. (Figure 2.12b)

Require: S_{global}, $C_{global}^{\{g,k\}}$

Ensure: S_{fur}, $C_{fur}^{\{g,k\}}$

$S_{fur} \leftarrow S_{global}$, $C_{fur}^{g,k} \leftarrow C_{global}^{g,k}$

repeat

Calculate $\langle S_{fur}, S_{fur} \rangle$, pick the pair with highest correlation that is greater than a cutoff c_α, i_1, i_2,

if there is no (g,k) that the nonzero times of $C_{fur_{i_1,\cdot}}^{g,k}$ and $C_{fur_{i_2,\cdot}}^{g,k}$ are within distance $\mathbf{t}$, find the next highest correlation $> c_\alpha$

if no pair of spectra satisfy this condition **then**

the further merging is over, break;

else

$$[s_{fur}] \leftarrow \min_{[s_{fur}]_{n\times 1}, c_{fur}^{g,k}} \sum_{g,k \text{ that qualify}} ||\tilde{X}_{fur}^{g,k} - S_{fur_{\cdot,-(i_1,i_2)}} C_{fur_{-(i_1,i_2),\cdot}}^{g,k} - s_{fur} c_{fur}^{g,k}||_F^2$$

end if

$[S_{fur}]_{\cdot,i_1} \leftarrow s_{fur}$, $[C_{fur}^{g,k}]_{i_1,\cdot} \leftarrow c_{fur}^{g,k}$, $[C_{fur}^{g,k}]_{i_2,\cdot} \leftarrow 0$ for these g,k that qualify. Delete $[S_{fur}]_{,i_2}$ if $[C_{fur}^{g,k}]_{i_2,\cdot}$ are all 0.

until

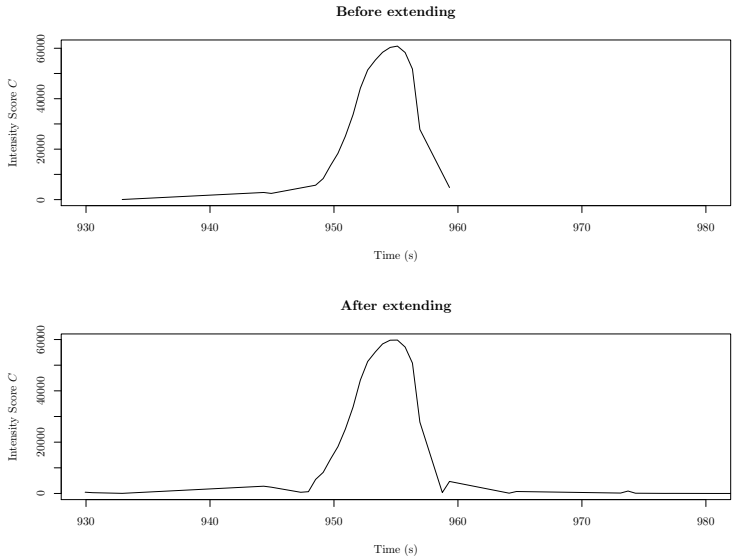


Figure 2.13: Shape checking

Step 8. Peak Splitting

Chromatogram with multiple peaks could be due to the chemical reaction that two similar spectra co-elute closely, which would be hard to detect in any computational algorithm, or it could be due to same spectra with multiple peaks. Either way, we want to split these peaks into different features. We used the template-based aligner (TBA) we developed for LCMS type of data to select template peaks using all samples. We make sure only peak clusters with sufficient large signals be selected as template features. See Figure 2.14. Every template feature represents a variable from all $\{g, k\}$ samples.

Step 9. Feature Selection

We use multinomial logistic regression with L_1 group penalties $\sqrt{\beta_{1,j}^2 + \cdots + \beta_{G,j}^2}$ on the coefficients belonging to one feature (all in or all out).

$$\min \sum_{g=1}^{G} \sum_{k=1}^{K_g} -\log\left(\frac{\exp(\sum_{j=1}^{J} \beta_{g,j} F_j^{g,k})}{\sum_{g=1}^{G} \exp(\sum_{j=1}^{J} \beta_{g,j} F_j^{g,k})}\right) + \lambda \sum_{j=1}^{J} \sqrt{\beta_{1,j}^2 + \cdots + \beta_{G,j}^2}$$

Shape Check. (Figure 2.13)

Require: $S_{fur}, C_{fur}^{\{g,k\}}$.

Ensure: S_{shape}, $C_{shape}^{\{g,k\}}$

$S_{shape} \leftarrow S_{fur}$, $C_{shape}^{g,k} \leftarrow C_{fur}^{g,k}$

Sort spectra by the biggest chromatogram response same as before, δ is the order

for $i = 1, \cdots, L_c$ **do**

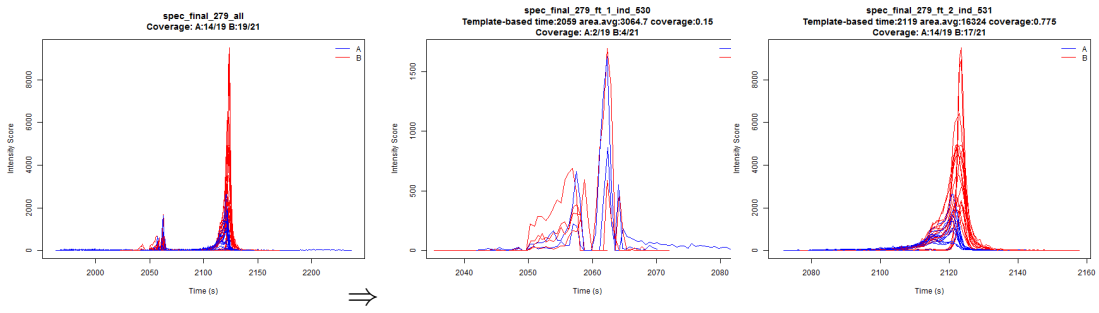
if the two endpoints of non-zero time region of $C_{shape_{\delta_i,.}}^{g,k}$ are bigger than c_{cut}, we extend the spectrum range, by $\mathbf{t}$, additional time points η,

5:

$$\forall g, k, \min_{[c_{shape}^{g,k}]_{1\times(|\eta|)}} ||I_{.\eta}^{g,k} - S_{shape_{.,-\delta_i}} C_{shape_{-\delta_i,\eta}}^{g,k} - S_{shape_{.,\delta_i}} c_{shape}^{g,k}||_F^2$$

$C_{shape_{\delta_i,\eta}}^{g,k} \leftarrow c_{shape}^{g,k}$

end for



(a) Pulling intensity scores C from all files. (b) Split feature one (c) Split feature two

Figure 2.14: Peak Splitting

Peak Splitting. (Figure 2.14)

Require: S_{shape}, $C_{shape}^{\{g,k\}}$

Ensure: Final feature matrix F, $J * \sum_{g=1}^{G} K_g$, $J \geq \text{ncol}(S_{shape})$

$F \leftarrow []$,

for i in $1 : ncol(S_{shape})$ **do**

Let peak function $Peak^{g,k}(\delta_t) \leftarrow [C_{shape}^{g,k}]_{i,.}$. δ_t represents discrete peak times. $Peak^{g,k}(y) = 0$ when $y \notin \delta_t$

repeat

$T_f(t) \leftarrow \sum_{g,k} \max_{|y-t| \leq T} Peak^{g,k}(y)$

feature_time $\leftarrow$ $\text{argmax}\, T_f(t)$, feature $\leftarrow$ **for all** g, k, $\max_{|y-feature\_time| \leq T} Peak^{g,k}(y)$, update peak function $Peak^{g,k}(y) \leftarrow$ subtract related peaks in $Peak^{g,k}(y)$.

$F \leftarrow \begin{bmatrix} F \\ \text{feature} \end{bmatrix}$

until Enough features selected or no more template peaks available

end for

Feature Selection

Require: Feature F, $J \times \sum_{g=1}^{G} K_g$

Ensure: F_{choice}

$F_{sparse} \leftarrow$ Keep the results of $\hat{\beta}_{.,j}$ which has at least one non-zero in the G-dimensional vector.

$F_{cor} \leftarrow$, keep these features whose correlation with at least one of F_{sparse} are $\geq c_\alpha$ from $\text{cor}(F_{sparse}, F)$.

$F_{choice} \leftarrow \begin{bmatrix} F_{sparse} \\ F_{cor} \end{bmatrix}$

2.3 Simulation

2.3.1 Set-up

We conduct 27 simulation studies which differ in terms of 3 levels of overlapping between metabolite chromatograms, 3 levels of between replicate chromatogram shifts, and 3 levels of random noise ϵ. Each simulation consists of two groups, 20 replicate samples each group. 5 true metabolites of interest are simulated, where their intensities significantly differentiate between the two groups. For each of the 5 true metabolites, we simulate another non-differential metabolite that elutes close to the true metabolite with a similar magnitude of intensity. 20 additional non-differential metabolites with smaller intensities are generated with their chromatogram peaks away from the 5 true metabolites but could affect their corresponding nearby metabolites.

Choices of metabolite spectra. All metabolites are selected and exported from NIST main EI MS library. 5 true metabolites are selected as Cholesterol-TMS, Citric-Acid-Tetra-TMS, Dibutylphthalate, Fumaricacidtetradecyltrans-hex-3-enylester, Glycine-tri-TMS. The 5 metabolites that closely co-elute are picked as those whose spectra overlap with the true metabolites for at least one major mass slice. Spectra with bigger inner product with the true spectra are selected at priority. They are chosen as 912-Octadecadienoicacid(ZZ)-trimethylsilylester, D-Psicopyranosepentakis(trimethylsilyl)ether(isomer2), Diethyl44'-azoxydibenzoate, cis-7cis-11-Hexadecadien-1-ylacetate, and Olean-12-ene-31516212228-hexol. Doing this would allow the overlapping occur within the same mass slices, thus increases the difficulty of deconvolution. The noise spectra are randomly picked from the library with no preferences.

Choices of chromatogram shapes. Chromatograms of spectra are selected from

previous modeling results on contrived datasets and a real dataset. For the 5 true metabolites and the 5 metabolites that closely co-elute, chromatograms with clean shapes are selected. The 5 true metabolite chromatogram peaks are scattered evenly across the time span, 480, 1440, 2400, 3360, and 4320 secs. The chromatograms of noise spectra are randomly selected.

3 levels of overlapping between metabolite chromatograms. Each true metabolite co-elutes with a nearby metabolite, their chromatogram peak distances determines the difficulty of the deconvolution task. We choose 1, 5, or 10 secs for this peak distance. The noise metabolite spectra co-elutes are intended to elute away from true metabolites, meaning their peaks are beyond the proximity (within 1, 5, or 10 secs) of true metabolite peaks. They might co-elute with the nearby metabolites though.

3 levels of replicate chromatogram perturbation. There exists time perturbation among replicate samples even for the same known metabolite. To take this factor into consideration, we add a random noise to the true peak locations, with the variance of noise to be 1, 5, and 10. This time shifts apply to all true, near and noise metabolites. The original chromatogram pictures might be broken by this replicate time shifts, and the resulting deconvolution task would be much harder when the noise is high.

3 levels of random matrix noise ϵ. Every data point in the time-mass intensity matrix we observe has its noise which could be a machine noise or compound interfering noise. We assume the noises follow independent normal distribution. Any negative data point is set to zero. Essentially, each noise follows a different truncated normal distribution, and majority of them are approximately non-truncated.

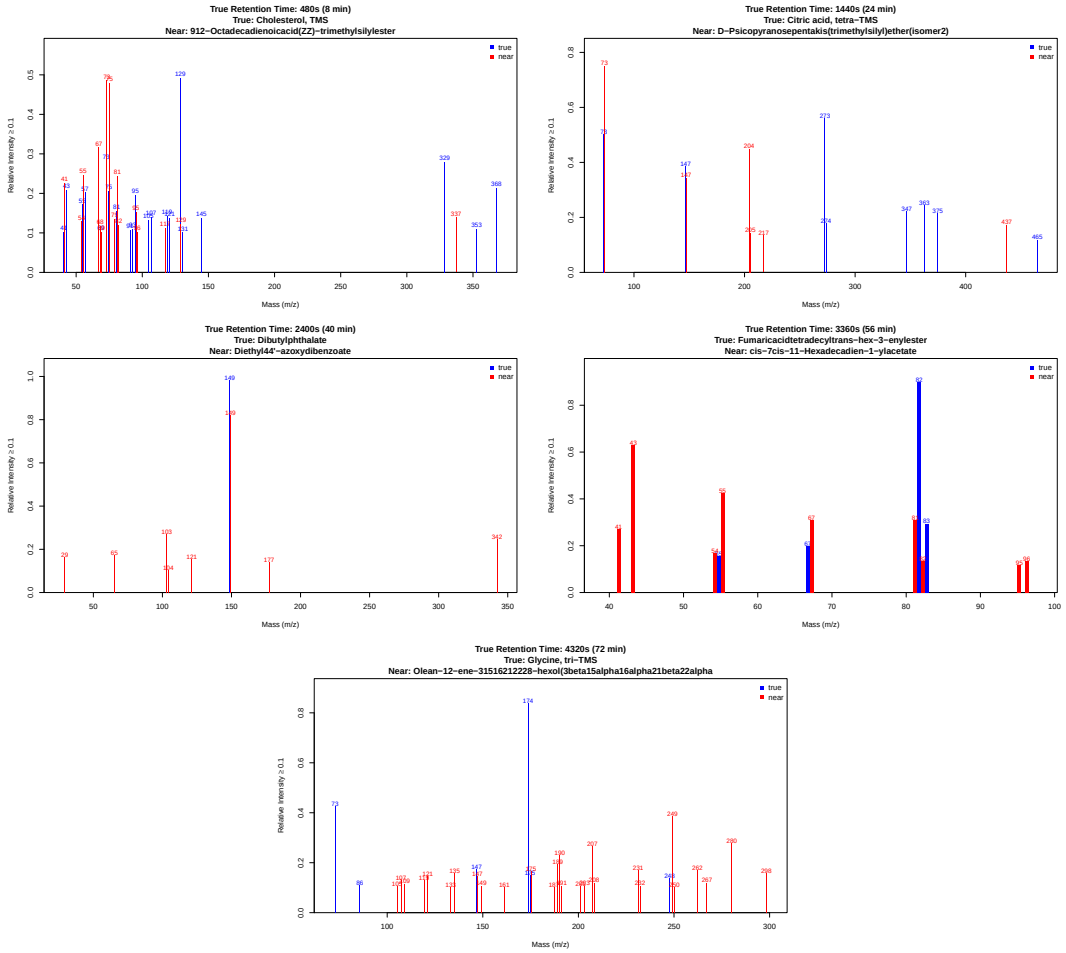


Figure 2.15: Simulated true and nearby spectra

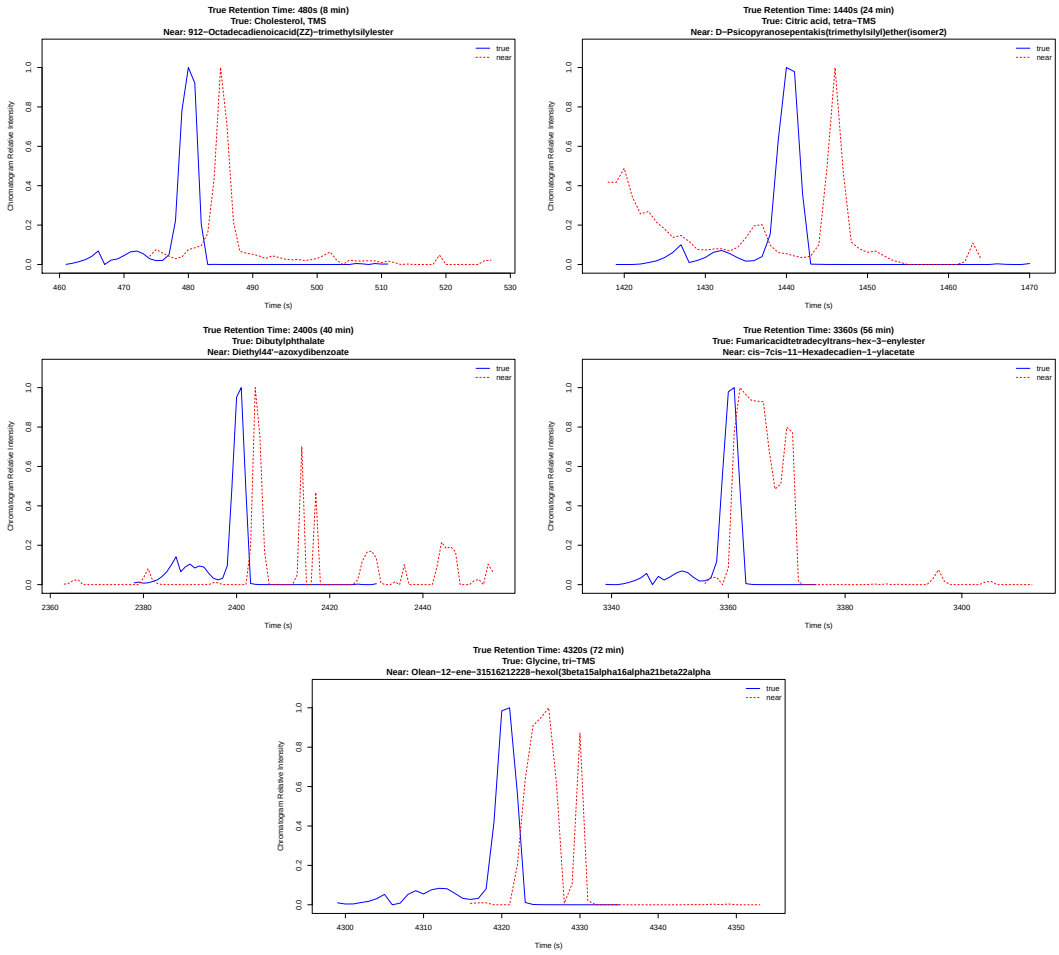
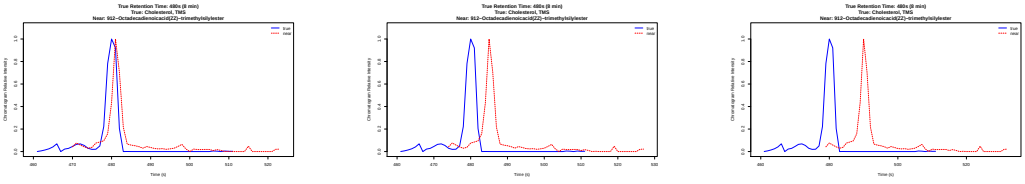
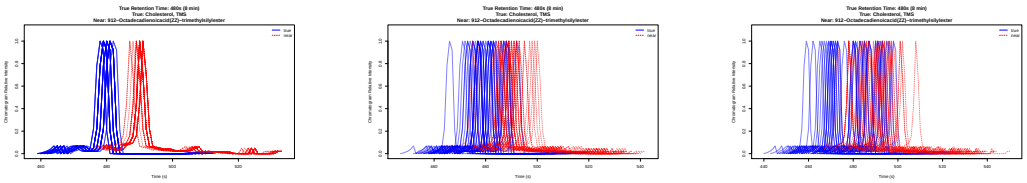


Figure 2.16: 5 Simulated true and nearby overlapping chromatograms with 5 secs peak distance.



(a) Time distance 1 sec (b) Time distance 5 secs (c) Time distance 10 secs

Figure 2.17: 3 levels of overlapping between metabolite chromatograms. Spectrum 1.



(a) sd_rep 1 second (b) sd_rep 5 seconds (c) sd_rep 10 seconds

Figure 2.18: 3 levels of standard deviation in times of replicate chromatograms. Spectrum 1, dist = 10s.

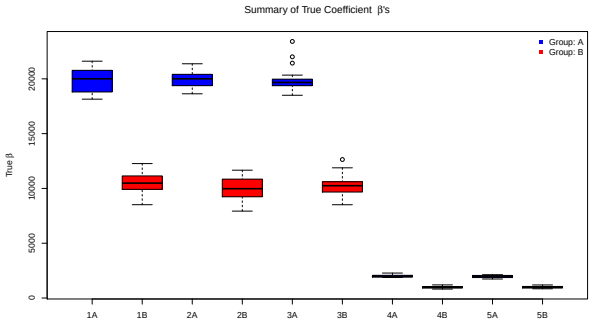
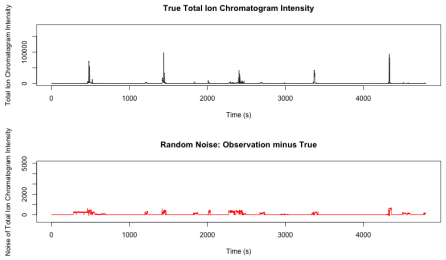
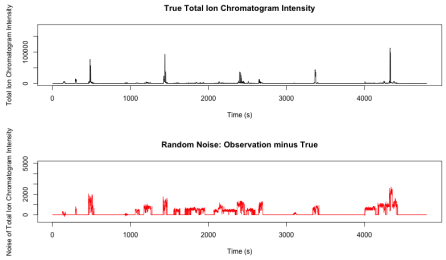


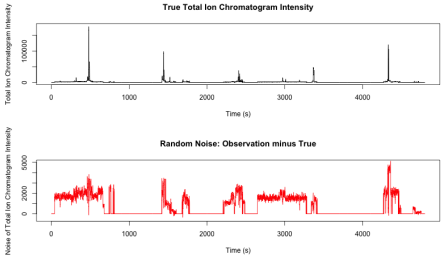
Figure 2.19: Boxplot of true β's. Dataset ⑭



(a) sd_noise 5. Dataset ① file 1.



(b) sd_noise 15. Dataset ② file 1.



(c) sd_noise 30. Dataset ③ file 1.

Figure 2.20: 3 levels of random noises. dist = 10s, sd_rep = 1s. Data is from the actual simulations.

2.3.2 Results

dist: the time difference between true spectra and corresponding nearby spectra.

sd_shift: variance of replicate peak time.

sd_noise: sd of random noise.

scr: inner product between true spectrum and learned spectrum.

cvrg: the proportion of samples that contain this learned spectrum.

	dist	sd_rep	sd_noise	scr_1	scr_2	scr_3	scr_4	scr_5	cvrg_1	cvrg_2	cvrg_3	cvrg_4	cvrg_5
①	10s	1s	5	1.000|1.000	1.000|0.999	1.000|1.000	1.000|1.000	1.000|1.000	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	82.5%|100.0%	100.0%|100.0%
②	10s	1s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.998	0.999|0.998	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%
③	10s	1s	30	1.000|1.000	1.000|1.000	1.000|1.000	0.997|0.995	0.997|0.999	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	75.0%|100.0%	92.5%|100.0%
④	10s	5s	5	1.000|1.000	1.000|1.000	1.000|1.000	1.000|0.999	1.000|1.000	97.5%|100.0%	100.0%|100.0%	100.0%|100.0%	75.0%|100.0%	85.0%|100.0%
⑤	10s	5s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.999	0.999|1.000	100.0%|100.0%	97.5%|100.0%	100.0%|100.0%	55.0%|100.0%	75.0%|100.0%
⑥	10s	5s	30	1.000|1.000	1.000|1.000	1.000|1.000	0.997|0.997	0.997|0.999	97.5%|100.0%	92.5%|100.0%	95.0%|100.0%	72.5%|100.0%	87.5%|100.0%
⑦	10s	10s	5	1.000|1.000	1.000|1.000	1.000|1.000	1.000|1.000	1.000|1.000	97.5%|100.0%	95.0%|100.0%	97.5%|100.0%	72.5%|100.0%	85.0%|100.0%
⑧	10s	10s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.999	0.999|0.998	95.0%|100.0%	95.0%|100.0%	100.0%|100.0%	65.0%|100.0%	82.5%|100.0%
⑨	10s	10s	30	1.000|1.000	1.000|1.000	1.000|1.000	0.997|0.998	0.997|0.999	87.5%|100.0%	95.0%|100.0%	97.5%|100.0%	62.5%|100.0%	75.0%|100.0%
⑩	5s	1s	5	1.000|1.000	1.000|1.000	1.000|1.000	0.996|1.000	1.000|1.000	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	35.0%|100.0%	100.0%|100.0%
⑪	5s	1s	15	1.000|1.000	1.000|1.000	1.000|0.998	0.997|0.998	0.999|0.991	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	15.0%|100.0%	82.5%|100.0%
⑫	5s	1s	30	1.000|0.999	1.000|1.000	1.000|1.000	0.995|0.996	0.995|0.992	100.0%|100.0%	100.0%|100.0%	97.5%|100.0%	15.0%|100.0%	90.0%|100.0%
⑬	5s	5s	5	1.000|1.000	1.000|1.000	1.000|0.996	1.000|0.996	1.000|1.000	100.0%|100.0%	100.0%|100.0%	100.0%|100.0%	47.5%|100.0%	90.0%|100.0%
⑭	5s	5s	15	1.000|0.996	1.000|1.000	1.000|1.000	0.999|0.997	0.999|0.999	95.0%|100.0%	100.0%|100.0%	100.0%|100.0%	60.0%|100.0%	80.0%|100.0%
⑮	5s	5s	30	1.000|0.997	1.000|1.000	1.000|1.000	0.997|0.994	0.997|0.996	87.5%|100.0%	97.5%|100.0%	95.0%|100.0%	37.5%|100.0%	60.0%|100.0%
⑯	5s	10s	5	1.000|1.000	1.000|1.000	1.000|1.000	1.000|0.999	1.000|1.000	97.5%|100.0%	97.5%|100.0%	100.0%|100.0%	65.0%|100.0%	70.0%|100.0%
⑰	5s	10s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.998	0.999|1.000	97.5%|100.0%	92.5%|100.0%	82.5%|100.0%	75.0%|100.0%	65.0%|100.0%
⑱	5s	10s	30	1.000|0.984	1.000|1.000	1.000|1.000	0.997|0.999	0.997|0.999	90.0%|100.0%	97.5%|100.0%	95.0%|100.0%	55.0%|100.0%	72.5%|100.0%
⑲	1s	1s	5	0.999|0.946	1.000|1.000	1.000|0.993	0.995|0.999	0.999|1.000	100.0%|100.0%	95.0%|100.0%	100.0%|100.0%	50.0%|100.0%	35.0%|100.0%
⑳	1s	1s	15	0.999|0.993	0.999|1.000	1.000|0.995	0.994|0.996	0.993|0.998	95.0%|100.0%	92.5%|100.0%	100.0%|100.0%	25.0%|100.0%	7.5%|100.0%
㉑	1s	1s	30	0.999|0.989	0.999|1.000	1.000|0.994	0.988|0.999	0.994|0.998	70.0%|100.0%	67.5%|100.0%	90.0%|100.0%	22.5%|100.0%	15.0%|100.0%
㉒	1s	5s	5	1.000|1.000	1.000|1.000	1.000|1.000	1.000|1.000	1.000|1.000	95.0%|100.0%	100.0%|100.0%	100.0%|100.0%	35.0%|100.0%	60.0%|100.0%
㉓	1s	5s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.994	0.999|0.999	95.0%|100.0%	97.5%|100.0%	97.5%|100.0%	25.0%|100.0%	57.5%|100.0%
㉔	1s	5s	30	1.000|0.964	1.000|1.000	1.000|1.000	0.995|0.994	0.997|0.995	97.5%|100.0%	95.0%|100.0%	100.0%|100.0%	30.0%|100.0%	52.5%|100.0%
㉕	1s	10s	5	1.000|0.999	1.000|1.000	1.000|1.000	1.000|1.000	1.000|1.000	92.5%|100.0%	97.5%|100.0%	95.0%|100.0%	72.5%|100.0%	75.0%|100.0%
㉖	1s	10s	15	1.000|1.000	1.000|1.000	1.000|1.000	0.999|0.998	0.999|0.999	92.5%|100.0%	92.5%|100.0%	95.0%|100.0%	60.0%|100.0%	72.5%|100.0%
㉗	1s	10s	30	1.000|1.000	1.000|1.000	1.000|0.997	0.996|0.999	0.997|0.999	87.5%|100.0%	97.5%|100.0%	92.5%|100.0%	60.0%|100.0%	62.5%|100.0%

Table 2.1: Recovery of 5 true spectra. Amdis | gcmsDecon.

For all 27 simulations, gcmsDecon is able to identify all 5 true spectra for every one of the 40 samples. We measure the error percentage as $\frac{\hat{\beta}_i - \beta_i}{\beta_i}, 1 \leq i \leq 5$, and error percentage histograms of the 4th spectrum from all simulation datasets are presented. gcmsDecon outperforms Amdis in every single case.

Take simulation ⑭ as an example, the chromatogram peak distance between true spectra and corresponding nearby overlapping spectra is 5 seconds, the replicate chromatogram has a standard deviation of 5 seconds, and random noise is set to be 15 for any positive intensity.

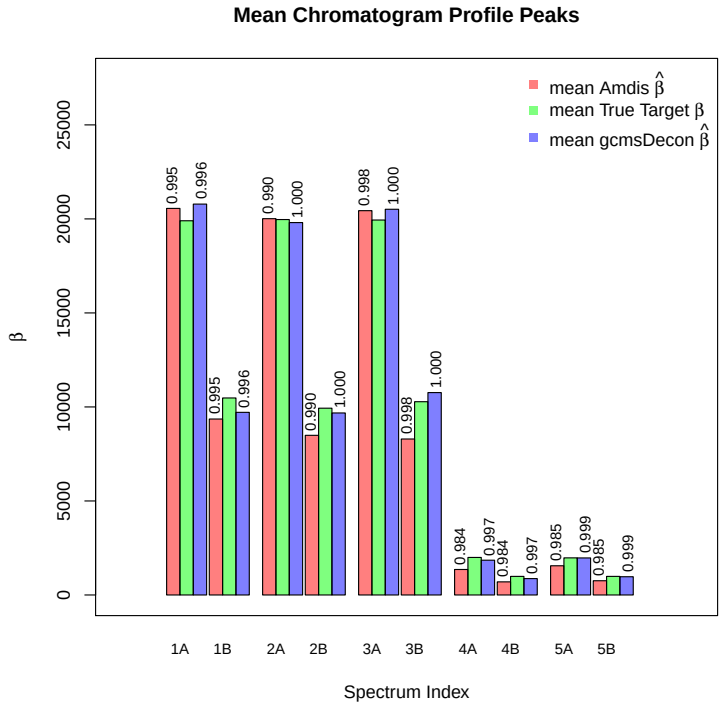


Figure 2.21: Dataset ⑭. Mean $\hat{\beta}_{Amdis}$, β and $\hat{\beta}_{gcmsDecon}$ of all 5 target spectra. Numbers on top of bars are the average inner products between learned spectra and true spectra. Dataset No.

2.4 Results on Real Data

We contrive two experimental data sets. In the first dataset, we intentionally add one compound into one group while the other group is missing this compound. In the second dataset, we add one compound to four groups with different amounts. We test the strength of our program by detecting this significantly differentiating compound and recover all other compounds in these samples.

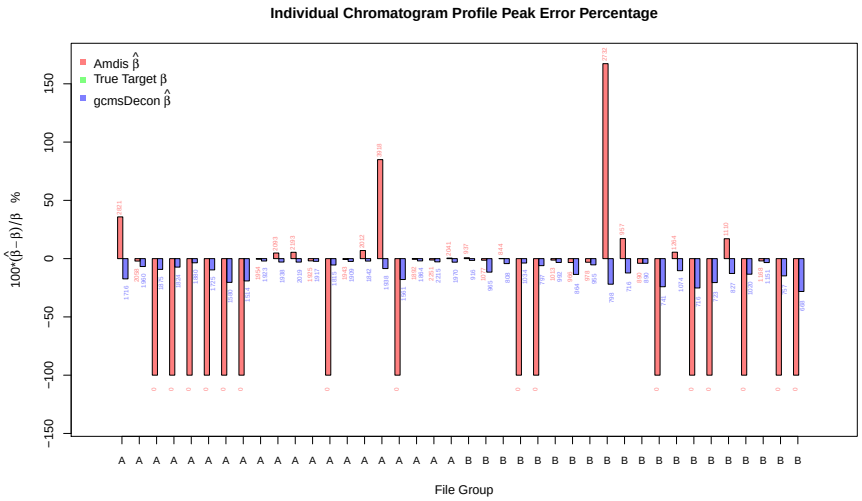


Figure 2.22: Dataset (14). $\frac{\hat{\beta}_{Amdis}-\beta}{\beta}$, $\frac{\hat{\beta}_{gcmsDecon}-\beta}{\beta}$ of true target spectrum 4. Numbers are $\hat{\beta}_{Amdis}$ and $\hat{\beta}_{gcmsDecon}$.

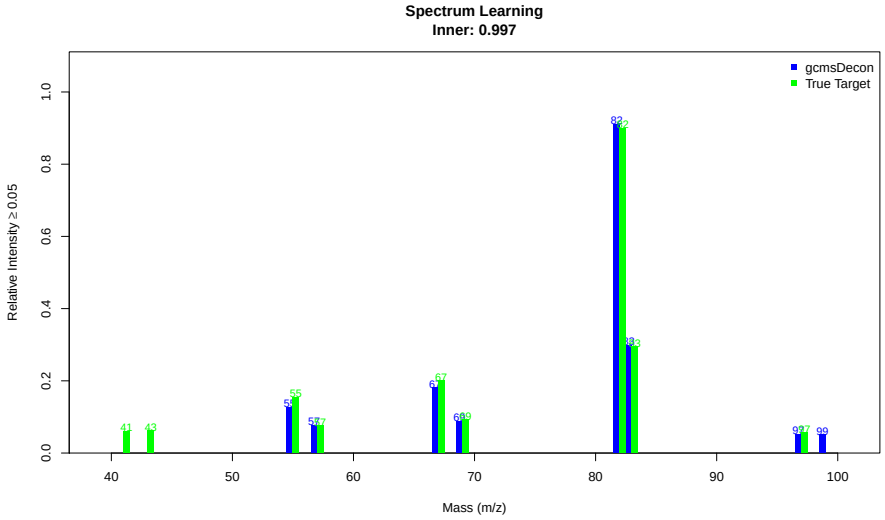


Figure 2.23: Dataset (14). gcmsDecon learned and true target spectrum 4.

2.4.1 Contrived I

In this study, 40 samples were prepared with 19 in group A and 21 in group B. All samples contained 7 compounds in equal amounts (leucine, syringic acid, tar-

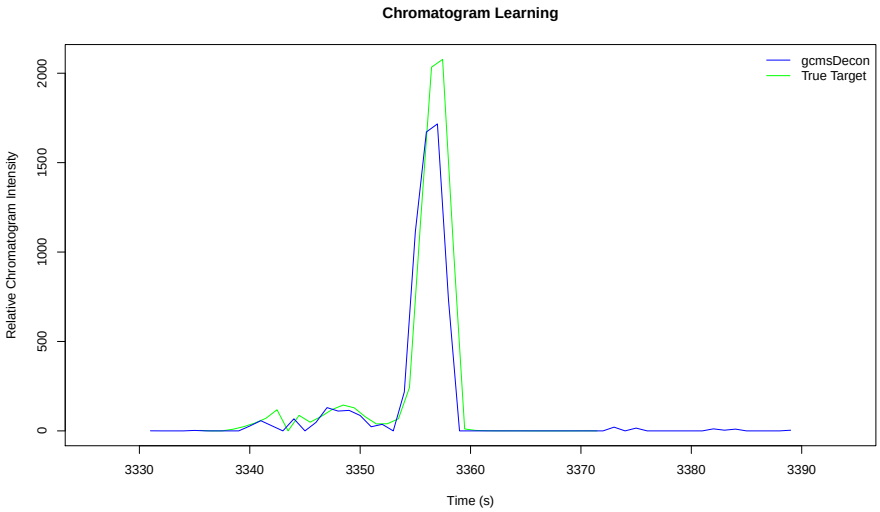


Figure 2.24: Dataset ⑭ file 1. gcmsDecon learned and true chromatogram 4.

taric acid, 2H5-3- hydroxyglutaric acid, methylnonadecanoic acid, nonadecanoic acid and myoinositol). In addition, glutamic acid was only added to one group. The samples were converted to their trimethylsilyl-methyl oxime derivatives and analyzed by GC/MS. The analysis is expected to pick out glutamic acid as the

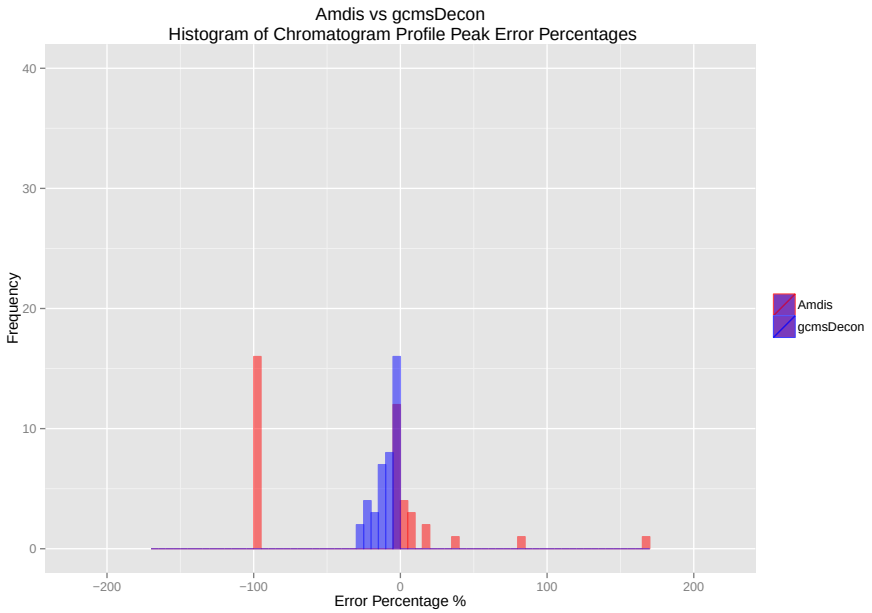
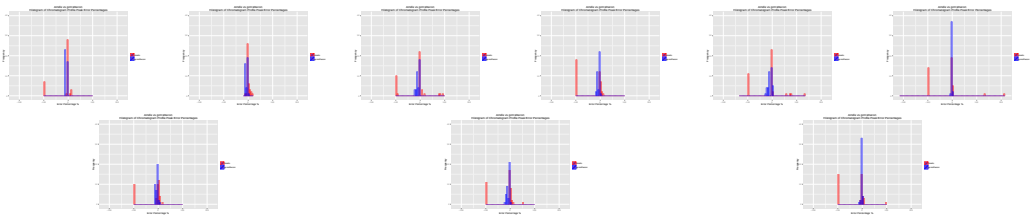
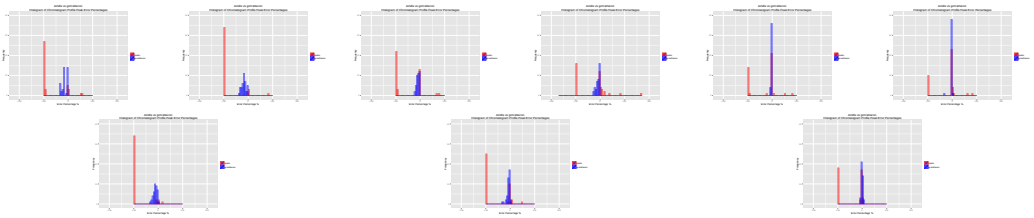


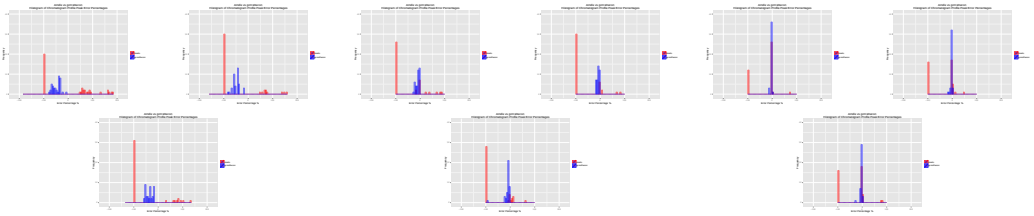
Figure 2.25: Dataset ⑭. Histogram of $\frac{\hat{\beta}_{Amdis}-\beta}{\beta}$, $\frac{\hat{\beta}_{gcmsDecon}-\beta}{\beta}$ of true target spectrum 4.



(a) dist 10s, sd_rep 1s (b) dist 10s, sd_rep 5s (c) dist 10s, sd_rep 10s



(d) dist 5s, sd_rep 1s (e) dist 5s, sd_rep 5s (f) dist 5s, sd_rep 10s



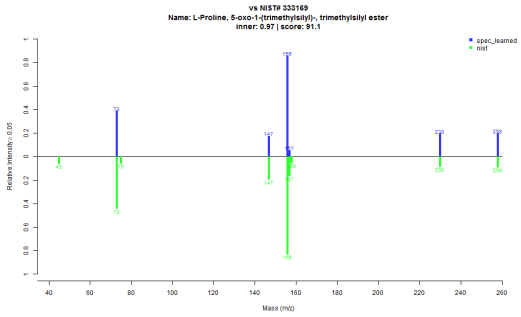
(g) dist 1s, sd_rep 1s (h) dist 1s, sd_rep 5s (i) dist 1s, sd_rep 10s

Figure 2.26: Error percentage Histograms for True Target Spectrum 4. sd_noise: 5 (left), 15 (right), 30 (bottom)

only significant difference between the two groups. In the system used, glutamic acid elutes at 31.56 minutes (1893.6 seconds).

We learn 394 spectra with 716 features in total, excluding 432 features with group coverage $< 50\%$ and feature AUC < 50. We use multinomial logistic regression with L_1 group penalties $\sqrt{\beta_{1,j}^2 + \cdots + \beta_{G,j}^2}$ on the coefficients (all in or all out). We are able to select 3 features that differentiate between the two groups.

All three are confirmed to clearly differentiate between groups. One of them is the intended compound, another one is an unexpected result from the chemical process, which also manifests the strength of our program. The third one is of



	1
Score	inner:0.970|score:91.1
Name	L-Proline, 5-oxo-1-(trimethylsilyl)-, trimethylsilyl ester
Synon_1	Proline, 5-oxo-1-(trimethylsilyl)-, trimethylsilyl ester
Synon_2	Pyroglutamic acid, bis(trimethylsilyl)-
Synon_3	Pyroglutamic acid, di(trimethylsilyl)-
Synon_4	Pyroglutamic acid, N,O-bis(trimethylsilyl)deriv.
Synon_5	Trimethylsilyl 5-oxo-1-(trimethylsilyl)-2-pyrrolidinecarboxylate, (2S)-
Synon_6	(2S)-2-Pyrrolidone-5-carboxylic acid, N-trimethylsilyl, trimethylsilyl ester
Synon_7	Pyroglutamic acid, (N,O-TMS)
Formula	C11H23NO3Si2
MW	273
ExactMass	273.121647
CAS#	30274-77-2
NIST#	333169
DB#	129206
Comments	NIST Mass Spectrometry Data Center

Figure 2.27: Contrived data I - 1st compound

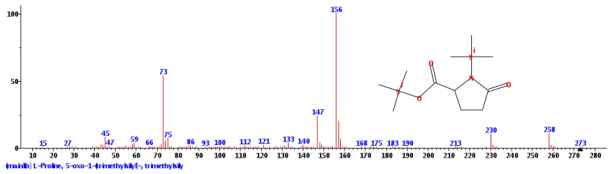


Figure 2.28: Contrived data I - 1st compound

small intensities, likely due to experiments related to the group assignment.

Derivatization of glutamate and or glutamine leads to the formation of pyroglutamic acid which is one of the peaks eluting at 31.56 minutes. There is another peak closely co-eluting, and the best library match is for an amidated 5-methoxy indole acetic acid derivative (N-(3- Hydroxypropyl)-2-(5-methoxy-1-methyl-1H-indol-2-yl)acetamide). The strength of the match is 0.883 for the inner product. Regardless of the true identity, the source of this compound is unclear. It could be an impurity in one of the glutamate standard used to make the mixture.

2.4.2 Contrived II

In this study, 40 samples were prepared with 10 each group. Syringic acid was added to all four groups, with amounts $A < B < C < D$ at retention time $39.7m$.

We learn 249 spectra with 496 features in total, excluding 224 features with group coverage $< 50\%$ and feature AUC < 50. We are able to select 21 features that

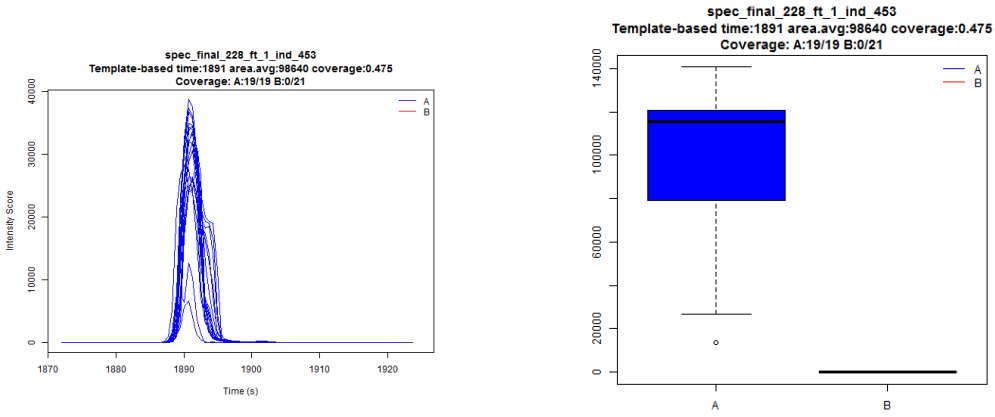
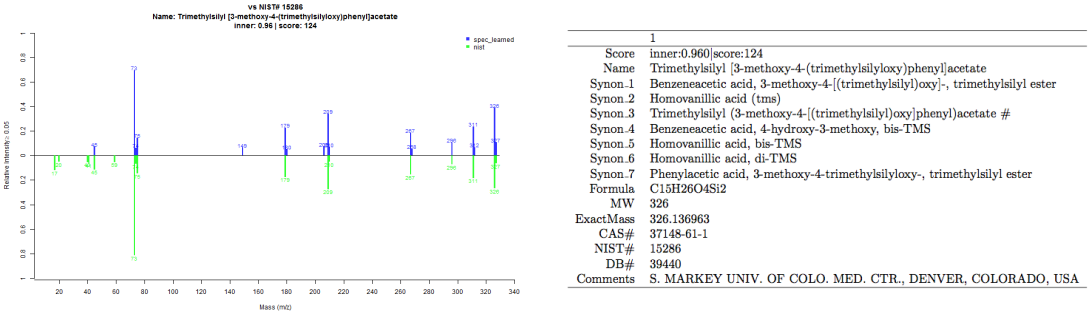


Figure 2.29: Contrived data I - 1st compound



	1
Score	inner:0.960|score:124
Name	Trimethylsilyl [3-methoxy-4-(trimethylsilyloxy)phenyl]acetate
Synon_1	Benzeneacetic acid, 3-methoxy-4-[(trimethylsilyl)oxy]-, trimethylsilyl ester
Synon_2	Homovanillic acid (tms)
Synon_3	Trimethylsilyl (3-methoxy-4-[(trimethylsilyl)oxy]phenyl)acetate #
Synon_4	Benzeneacetic acid, 4-hydroxy-3-methoxy, bis-TMS
Synon_5	Homovanillic acid, bis-TMS
Synon_6	Homovanillic acid, di-TMS
Synon_7	Phenylacetic acid, 3-methoxy-4-trimethylsilyloxy-, trimethylsilyl ester
Formula	C15H26O4Si2
MW	326
ExactMass	326.136963
CAS#	37148-61-1
NIST#	15286
DB#	39440
Comments	S. MARKEY UNIV. OF COLO. MED. CTR., DENVER, COLORADO, USA

Figure 2.30: Contrived data II - 1st compound

differentiate between the four groups.

Among these 21, there is the compound syringic acid at $39.7m$ that was intended to be different, as well as homovanillic acid at retention time $37.1m$, which is believed to be related to the group assignment and syringic acid.

2.5 Summary

GCMS deconvolution is a large-scale computational problem with significant biological interests. It also provides the potential for statistical modeling. Our model provides accurate results with complete automation.

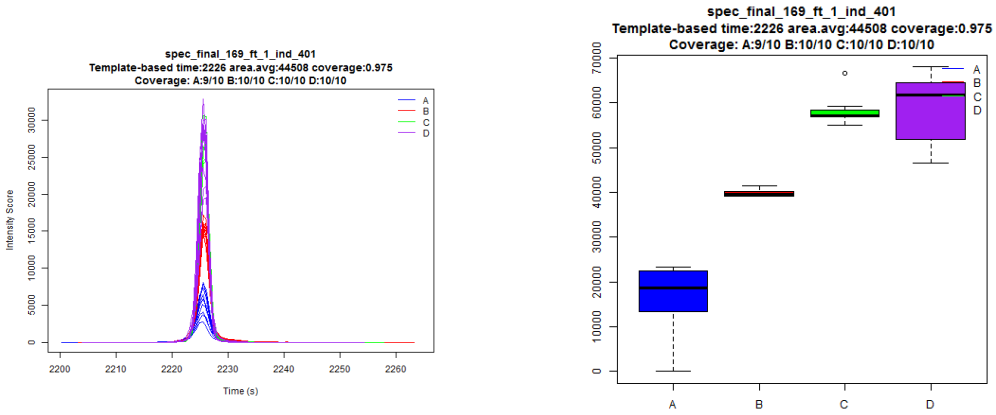


Figure 2.31: Contrived data II - 1st compound

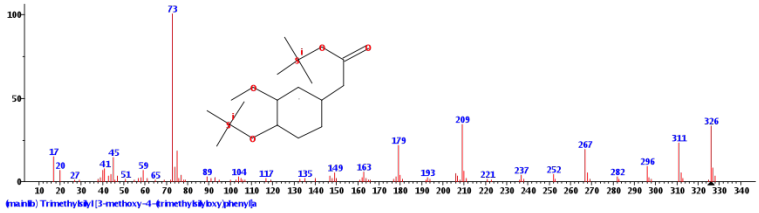
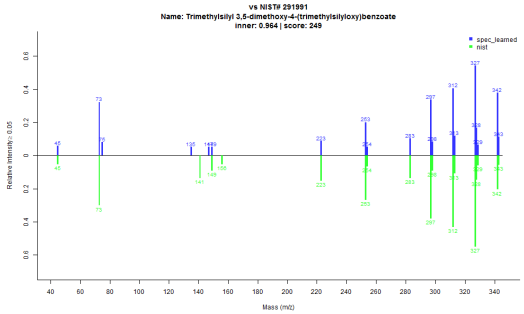


Figure 2.32: Contrived data II - 1st compound



	1
Score	inner:0.964|score:249
Name	Trimethylsilyl 3,5-dimethoxy-4-(trimethylsilyloxy)benzoate
Synon_1	Benzoic acid, 3,5-dimethoxy-4-[(trimethylsilyl)oxy]-, trimethylsilyl ester
Synon_2	Syringic acid d–tms
Synon_3	Syringic acid-di-TMS
Synon_4	Syringic acid, TMS
Formula	C15H26O5Si2
MW	342
ExactMass	342.131878
CAS#	10517-29-0
NIST#	291991
DB#	202500
Comments	NIST Mass Spectrometry Data Center, 1998.

Figure 2.33: Contrived data II - 2nd compound

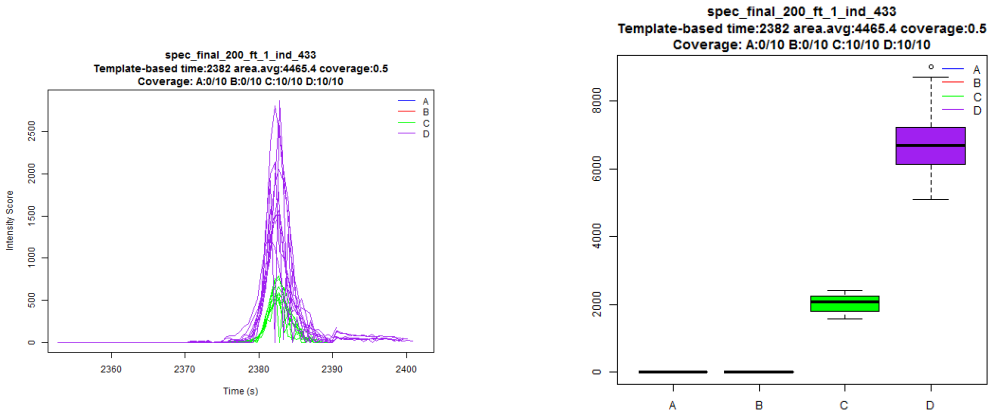


Figure 2.34: Contrived data II - 2nd compound

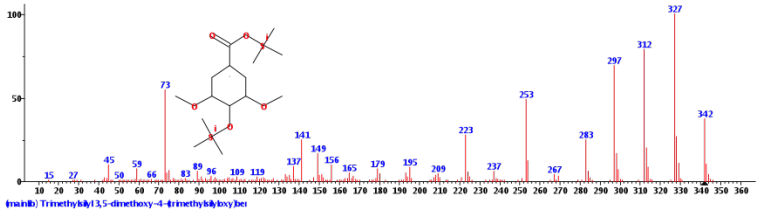


Figure 2.35: Contrived data II - 2nd compound

We repeatedly use rank estimation based on random matrix theory to compute the non-negative matrix factorization from small scales and eventually assemble them to build up the entire matrix deconvolution scheme. Besides the pre-processing, the total algorithm consists of 9 large steps: Sea-island learning, initial merging, adaptive local NMF, merging from local windows, global NMF, further merging, shape checking, peak splitting, and feature selection. We successfully demonstrate our model on 27 simulations with different settings in terms of 3 levels of chromatogram overlapping, 3 levels of replicate chromatogram perturbation, 3 levels of random matrix noise. The results are compared with the free GCMS deconvolution software AMDIS and our model outperforms it in every way. Besides, our model provides a complete streamlined version of solution including simultaneous learning of all samples which is crucial to identify compounds differentiating between disease groups.

This work provides a R package *gcmsDecon*.

2.6 Future Work

1. Scalability for higher resolution data.

   Higher resolution data becomes available and the computational capacity needs to improve. Our current data is of column size 1000, the new machines are able to generate data with much larger matrix columns. Some algorithm steps will fail due to the memory ceiling. We have to adjust the program for higher-resolution data.

2. The lack of library for high-resolution GCMS data.

   Current NIST library for GC-MS data is outdated and revolutionizing the whole library takes time and joint efforts from the whole community.

3. Random matrix behavior under non-negative matrix setting.

   Needless to say, theoretical justification of some non-negative matrix properties itself is already challenging and quite often leaves no analytical conclusions. Random matrix theory is developing at a fast pace, however, non-i.i.d random matrix is hard to analyze. More theories regarding non-negative random matrix would definitely benefit our model as well as other potential applications.

4. Scientific applications.

   We want to apply *gcmsDecon* for more biological research and applications. A great tool is only useful when the right people are using it to perform exciting tasks.

2.7 Main Theorems and Discussions

2.7.1 Random Matrix

Many random matrix theories regarding singular vectors (eigenvectors) can not be directly applied to NMF due to non-negative constraints. However, they help understand the similar process in non-negative random matrices. The rank detection based on RMT is still legitimate to our problem. Here I summarize a few related random matrix theories developed to justify and describe algorithm steps of *gcmsDecon*.

Theorem 1:

For any $n \times p$ random matrix $\widetilde{X}_n = X_n + P_n = X_n + \sum_{k=1}^{r} \theta_k u_k v_k^T = X_n + U\Theta V^T$, where $[X_n]_{i,j}$ are i.i.d r.v.s, $E([X_n]_{i,j}) = 0$, $Var([X_n]_{i,j}) = \frac{1}{\max(p,n)}$. $U^TU = I, V^TV = I$. $\lim_{n\to\infty} \frac{n}{p} = c$. $\theta_1 > \theta_2 > \cdots > \theta_r$, let r_0 be the largest subscript such that for $1 \le k \le r_0 \le r$, $\theta_k \ge c^{1/4}$. Let $\widetilde{\lambda}_1 \ge \widetilde{\lambda}_2 \ge \cdots \ge \widetilde{\lambda}_n$ denote the singular values of $\widetilde{X}_n$. $\mu_{X_n} = \frac{1}{n-r} \sum_{i=r+1}^{n} \delta_{\tilde{\lambda}_i}$. We have the asymptotic limit

$$\widetilde{\lambda}_k \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \rho_k = \begin{cases} \sqrt{\frac{(1+\theta_k^2)(\min(c,c^{-1})+\theta_k^2)}{\theta_k^2}} & 1 \le k \le r_0 \\ 1 + \sqrt{\min(c, c^{-1})} & r_0 < k \le r \end{cases} \tag{2.5}$$

$$\frac{1}{\min(n,p) - r} \sum_{i=r+1}^{\min(n,p)} \mathbf{1}(\tilde{\lambda}_i \le t) = \int_0^t \mu_{X_n}(t)dt \xrightarrow[n,p\to\infty]{P} \int_0^t \mu(t; \min(c, c^{-1}))dt, \tag{2.6}$$

where

$$\mu(t; d) = \frac{\sqrt{(t^2 - (1-\sqrt{d})^2)((1+\sqrt{d})^2 - t^2)}}{\pi d t}, \quad 1 - \sqrt{d} \le t \le 1 + \sqrt{d} \tag{2.7}$$

$$1 \le j, k \le r_0, \quad |\langle \tilde{u}_j, u_k \rangle|^2 \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \begin{cases} 1 - \frac{\min(c,1)(\min(1,c^{-1})+\theta_k^2)}{\theta_k^2(\theta_k^2+\min(c,1))} & j = k \\ 0 & j \ne k \end{cases} \tag{2.8}$$

$$1 \le j, k \le r_0, \quad |\langle \tilde{v}_j, v_k \rangle|^2 \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \begin{cases} 1 - \frac{\min(1,c^{-1})(\min(c,1)+\theta_k^2)}{\theta_k^2(\theta_k^2+\min(1,c^{-1}))} & j = k \\ 0 & j \ne k \end{cases} \tag{2.9}$$

Proof: This theorem combines multiple theorems (*Theorem 2.8, 2.9, 2.10*) in [BGN12] ($n \leq p$). Results for $n > p$ random matrix $\tilde{X}_n$ can be derived from $\tilde{X}_n^T$.

Theorem 2:

For any $n \times p$ random matrix $\widetilde{X}_n = X_n + P_n = X_n + \sum_{i=1}^{r} \theta_k(n,p) u_k v_k^T = X_n + U\Theta(n,p)V^T$, where $[X_n]_{i,j}$ are i.i.d r.v.s, $E([X_n]_{i,j}) = 0$, $Var([X_n]_{i,j}) = \sigma^2$, $\lim_{n\to\infty} \frac{\theta_i(n,p)}{\sqrt{np}\sigma} = \zeta_i$. $U^TU = I, V^TV = I$. $\lim_{n\to\infty} \frac{n}{p} = c$. $\theta_1 > \theta_2 > \cdots > \theta_r$, let r_0 be the largest subscript such that for $1 \leq k \leq r_0 \leq r$, $\sqrt{\min(n,p)}\zeta_k \geq c^{1/4}$. Let $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_n$ denote the singular values of $\widetilde{X}_n$. $\mu_{X_n} = \frac{1}{\min(n,p)-r} \sum_{i=r+1}^{\min(n,p)} \delta_{\tilde{\lambda}_i/(\sqrt{\max(p,n)}\sigma)}$, then

$$\frac{\widetilde{\lambda}_k}{\sqrt{\max(p,n)}\sigma} \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \rho_k = \begin{cases} \infty \approx \sqrt{\frac{(1+\min(n,p)\zeta_k^2)(\min(c,c^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2}} & 1 \leq k \leq r_0 \\ 1+\sqrt{\min(c,c^{-1})} & r_0 < k \leq r \end{cases} \tag{2.10}$$

$$\begin{aligned} &\frac{1}{\min(n,p)-r} \sum_{i=r+1}^{\min(n,p)} \mathbf{1}\left(\frac{\tilde{\lambda}_i}{\sqrt{\max(p,n)}\sigma} \leq t\right) \\ &= \int_0^t \mu_{X_n}(t)dt \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{P} \int_0^t \mu(t; \min(c,c^{-1}))dt, \end{aligned} \tag{2.11}$$

where

$$\mu(t;d) = \frac{\sqrt{(t^2-(1-\sqrt{d})^2)((1+\sqrt{d})^2-t^2)}}{\pi d t}, \quad 1-\sqrt{d} \leq t \leq 1+\sqrt{d} \tag{2.12}$$

$1 \leq j,k \leq r_0$,

$$|\langle \tilde{u}_j, u_k \rangle|^2 \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \begin{cases} 1 \approx 1 - \frac{\min(c,1)(\min(1,c^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(c,1))} & j = k \\ 0 & j \neq k \end{cases} \tag{2.13}$$

$$|\langle \tilde{v}_j, v_k \rangle|^2 \xrightarrow[n,p\to\infty, \frac{n}{p}\to c]{a.s.} \begin{cases} 1 \approx 1 - \frac{\min(1,c^{-1})(\min(c,1)+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(1,c^{-1}))} & j = k \\ 0 & j \neq k \end{cases} \tag{2.14}$$

Proof for case $n \leq p$:

$$\widetilde{X}_n = X_n + P_n = X_n + \sum_{i=1}^{r} \theta_k(p) u_k v_k^T = \sqrt{p}\sigma\left(\frac{1}{\sqrt{p}\sigma} X_n + \sum_{i=1}^{r} \frac{\theta_i(n,p)}{\sqrt{p}\sigma} u_k v_k^T\right)$$

then, when n, p are big enough, $\frac{\theta_i(n,p)}{\sqrt{p}\sigma} \approx \sqrt{n}\zeta_i$, let $Y_n = \frac{1}{\sqrt{p}\sigma} X_n$, then $E([Y_n]_{i,j}) = 0, Var([Y_n]_{i,j}) = \frac{1}{p}$,

$$\tilde{X}_n \approx \sqrt{p}\sigma(Y_n + \sum_{i=1}^{r} \sqrt{n}\zeta_k u_k v_k^T)$$

we can apply ***Theorem 1*** on random matrix $Y_n + \sum_{i=1}^{r} \sqrt{n}\zeta_k u_k v_k^T$. Theorem proved.

Remark: This is the set-up of real data we encounter. Throughout the paper, we assume this is the default model set up.

Theorem 3:

For any $n \times p$ random matrix $\widetilde{X}_n = X_n + P_n = X_n + \sum_{i=1}^{r} \theta_k u_k v_k^T = X_n + U\Theta V^T$, where p is finite, $\lim_{n\to\infty} \frac{n}{p} \to \infty$, $[X_n]_{i,j}$ are i.i.d r.v.s, $E([X_n]_{i,j}) = 0$, $Var([X_n]_{i,j}) = \frac{1}{n}$. $U^T U = I, V^T V = I$. $\theta_1 > \theta_2 > \cdots > \theta_r$. Let $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_n$ denote the singular values of $\widetilde{X}_n$, then

$$\tilde{\lambda}_k \xrightarrow[n\to\infty]{P} \rho_k = \begin{cases} \sqrt{1+\theta_k^2} & 1 \leq k \leq r \\ 1 & r+1 \leq k \leq p \text{ (if } p > r) \end{cases} \tag{2.15}$$

$$|\langle \tilde{u}_k, u_k \rangle|^2 \xrightarrow[n\to\infty]{a.s.} \frac{\theta_k^2}{\theta_k^2 + 1}, \quad |\langle \tilde{v}_k, v_k \rangle|^2 \xrightarrow[n\to\infty]{a.s.} 1, \quad 1 \leq k \leq r \tag{2.16}$$

Proof:

Results above could be intuitively derived from ***Theorem 1*** by letting $c \to 0$, however we do not have the condition that $p \to \infty$. We derive our results through determinant computation similar to [BGN12]. Based on [HJ85], non-zero singular values of $\tilde{X}$ are positive eigenvalues of matrix $\begin{bmatrix} 0 & \tilde{X} \\ \tilde{X}^T & 0 \end{bmatrix}$, with the determinant formula

$$\det(\begin{bmatrix} A & B \\ C & D \end{bmatrix}) = \det(D)\det(A - BD^{-1}C) = \det(A)\det(D - CA^{-1}B)$$

we have

$$\det(zI_{n+p} - \begin{bmatrix} 0 & \widetilde{X}_n \\ \widetilde{X}_n^T & 0 \end{bmatrix}) = det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix} - \begin{bmatrix} 0 & U_n\Theta V_p^T \\ V_p\Theta U_n^T & 0 \end{bmatrix})$$

$$= \det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix} - \begin{bmatrix} U_n & 0 \\ 0 & V_p \end{bmatrix} \begin{bmatrix} 0 & \Theta \\ \Theta & 0 \end{bmatrix} \begin{bmatrix} U_n^T & 0 \\ 0 & V_p^T \end{bmatrix})$$

$$= \det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix}) \det(\begin{bmatrix} 0 & \Theta \\ \Theta & 0 \end{bmatrix})$$

$$det(\begin{bmatrix} 0 & \Theta \\ \Theta & 0 \end{bmatrix}^{-1} - \begin{bmatrix} U_n^T & 0 \\ 0 & V_p^T \end{bmatrix} (zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix})^{-1} \begin{bmatrix} U_n & 0 \\ 0 & V_p \end{bmatrix})$$

$$= \det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix}) \times \prod_{i=1}^{r} \theta_i^2 \times \det(M_n(z)) \tag{2.17}$$

where

$$M_n(z) = \begin{bmatrix} U_n^T & 0 \\ 0 & V_p^T \end{bmatrix} (zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix})^{-1} \begin{bmatrix} U_n & 0 \\ 0 & V_p \end{bmatrix} - \begin{bmatrix} 0 & \Theta \\ \Theta & 0 \end{bmatrix}^{-1} \tag{2.18}$$

Here we assume $\det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix}) \neq 0$, in fact, $\det(zI_{n+p} - \begin{bmatrix} 0 & X_n \\ X_n^T & 0 \end{bmatrix}) = 0$ would correspond to non-zero singular values of noise X_n, which is the eigenvalues of $X_n^T X_n \xrightarrow[n\to\infty]{a.s} I_{p\times p}$. There are $p - r$ eigenvalues 1 (if $p > r$). Intuitively, the within-in-bulk singular values based on quarter-circle law, in this finite case, $c \to 0$, it is almost surely identical to 1. The rest singular values come from those z that satisfy $\det(M_n(z)) = 0$.

with the help of formula

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix} \tag{2.19}$$

$$M_n(z)$$
$$= \begin{bmatrix} U_n^T z(z^2 I_n - X_n X_n^T)^{-1} U_n & U_n^T X_n (z^2 I_p - X_n^T X_n)^{-1} V_p \\ V_p^T (z^2 I_p - X_n^T X_n)^{-1} X_n^T U_n & V_p^T z (z^2 I_p - X_n^T X_n)^{-1} V_p \end{bmatrix} - \begin{bmatrix} 0 & \Theta \\ \Theta & 0 \end{bmatrix}^{-1} \tag{2.20}$$

with $X_n^T X_n \xrightarrow[n\to\infty]{a.s} I_p$, and when n is large,

$$X_n X_n^T \approx Q \begin{bmatrix} I_p & \\ & 0_{n-p} \end{bmatrix} Q^T, \quad z^2 I_n - X_n X_n^T \approx Q \begin{bmatrix} (z^2-1) I_p & \\ & z^2 I_{n-p} \end{bmatrix} Q^T \tag{2.21}$$

$$(z^2 I_n - X_n X_n^T)^{-1} \approx Q \begin{bmatrix} \frac{1}{z^2-1} I_p & \\ & \frac{1}{z^2} I_{n-p} \end{bmatrix} Q^T \approx Q \frac{1}{z^2} I_n Q^T = \frac{1}{z^2} I_n \tag{2.22}$$

use the *Lindeberg-Feller Theorem* in [Fer96] P27, we have

$$U_n^T X_n V_p \xrightarrow[n\to\infty]{a.s.} 0 \tag{2.23}$$

Combine these results,

$$\det(M_n(z)) \xrightarrow[n\to\infty]{a.s.} \det\left(\begin{bmatrix} \frac{1}{z} & 0 \\ 0 & \frac{z}{z^2-1} \end{bmatrix} - \begin{bmatrix} 0 & \Theta^{-1} \\ \Theta^{-1} & 0 \end{bmatrix} \right)$$
$$= \frac{1}{z^r} \det\left(\frac{z}{z^2-1} I_p - \Theta^{-1} z \Theta^{-1} \right) = \prod_{k=1}^{r} \left(\frac{1}{z^2-1} - \frac{1}{\theta_k^2} \right) \tag{2.24}$$

$$\lambda_k = \sqrt{1+\theta_k^2}, \quad 1 \le k \le r \tag{2.25}$$

The proof of $|\langle \tilde{u}_k, u_k \rangle|^2$ is similar to the proof of *Theorem 2.9* in [BGN12], thus skipped here.

Theorem 4:

For any $n \times p$ random matrix $\widetilde{X}_n = X_n + P_n = X_n + \sum_{i=1}^{r} \theta_k(p) u_k v_k^T = X_n +$

$U\Theta(n,p)V^T$, where p is finite, $\lim_{n\to\infty}\frac{n}{p}\to\infty$, $[X_n]_{i,j}$ are i.i.d r.v.s, $E([X_n]_{i,j})=0$, $Var([X_n]_{i,j})=\sigma^2$, $\lim_{n\to\infty}\frac{\theta_i(n,p)}{\sqrt{np}\sigma}=\zeta_i$. $U^TU=I, V^TV=I$. $\zeta_1>\zeta_2>\cdots>\zeta_r$. Let $\widetilde{\lambda}_1\geq\widetilde{\lambda}_2\geq\cdots\geq\widetilde{\lambda}_n$ denote the singular values of $\widetilde{X}_n$, then

$$\frac{\tilde{\lambda}_k}{\sqrt{n}\sigma}\xrightarrow[n\to\infty]{P}\rho_k=\begin{cases}\sqrt{1+p\zeta_k^2} & 1\leq k\leq r\\ 1 & r+1\leq k\leq p\ (\text{ if } p>r)\end{cases}\tag{2.26}$$

$$|\langle\tilde{u}_k,u_k\rangle|^2\xrightarrow[n\to\infty]{a.s.}\frac{p\zeta_k^2}{p\zeta_k^2+1},\quad |\langle\tilde{v}_k,v_k\rangle|^2\xrightarrow[n\to\infty]{a.s.}1,\quad 1\leq k\leq r\tag{2.27}$$

Proof: the proof for this theorem is similar to that of ***Theorem 2***, thus skipped here.

Lemma 5 (Random Noise SD Estimator $\hat{\sigma}$):

Given conditions in ***Theorem 2***, we have the consistent estimator from [DS03] for unknown σ

$$\hat{\sigma}=\frac{1}{\sqrt{\max(p,n)}}\frac{\tilde{\lambda}_{med}}{\mu_{med}}\xrightarrow[n,p\to\infty]{P}\sigma.\tag{2.28}$$

Where μ_{med} is the median of quarter-circle law distribution, i.e. the x such that $\int_{1-\sqrt{\min(c,c^{-1})}}^{x}\mu(t)dt=\frac{1}{2}$.

Proof: Since the rank of signal r is finite, as n,p grow big, median of all singular values is asymptotically the median of the within-bulk singular values ($r+1\leq i\leq\min(n,p)$), together with ***Theorem 2***, this median $\frac{\tilde{\lambda}_{med}}{\sqrt{\max(p,n)}\sigma}\xrightarrow{P}\mu_{med}$. Re-organize the equation we proves the lemma.

Remarks:

1. when p is finite, based on ***Theorem 4***, $\mu_{med}=1$.

$$\frac{\tilde{\lambda}_{med}}{\sqrt{n}\sigma}\xrightarrow{P}\begin{cases}1 & r<\lceil\frac{p}{2}\rceil\\ (1+\sqrt{1+p\zeta_k^2})/2 & r=\lceil\frac{p}{2}\rceil\text{ and }p\text{ is even}\\ \sqrt{1+p\zeta_k^2},1\leq k\leq r & o.w.\end{cases}\tag{2.29}$$

In the latter two cases, using $\hat{\sigma} = \frac{1}{\sqrt{n}} \frac{\tilde{\lambda}_{med}}{\mu_{med}}$ will overestimate the σ.

2. When the i.i.d assumptions of X_n are violated, $\hat{\sigma} = \frac{1}{\sqrt{n}} \frac{\tilde{\lambda}_{med}}{\mu_{med}}$ would also be a bad estimator for variance.

Theorem 6 (Asymptotic Limits of Rank Estimators $\hat{r}^{rbt}$ and $\hat{r}^{cst}$): Given conditions in ***Theorem 2***. $\rho_1, \cdots, \rho_{r_0}$ are the asymptotic limits of top singular values over $\frac{\tilde{\lambda}_k}{\sqrt{\max(n,p)}\sigma}$ as in (2.10).

a) The robust rank estimator $\hat{r}_n^{rbt} = \sum_{i=1}^{\min(n,p)} \mathbf{1}(\tilde{\lambda}_i > \frac{\tilde{\lambda}_1}{\kappa_{rank}})$, where κ_{rank} is a constant.

$$\hat{r}_n^{rbt} \xrightarrow[n,p\to\infty]{P} \begin{cases} k-1 & 1 < \frac{\rho_1}{\rho_{k-1}} < \kappa_{rank} \leq \frac{\rho_1}{\rho_k},\ 1 < k \leq r_0 \\ r_0 & \frac{\rho_1}{\rho_{r_0}} < \kappa_{rank} \leq \frac{\rho_1}{(1+\sqrt{\min(c,c^{-1})})} \\ r_0 + (n-r_0) \int_{\frac{\rho_1}{\kappa_{rank}}}^{(1+\sqrt{\min(c,c^{-1})})} \mu(x)dx & \frac{\rho_1}{(1+\sqrt{\min(c,c^{-1})})} < \kappa_{rank} \end{cases} \tag{2.30}$$

Proof: By ***Theorem 2*** and (2.10), we have asymptotic locations of $\rho_1, \cdots, \rho_{r_0}$ and the bulk boundary $1+\sqrt{\min(c, c^{-1})}$. Compare them with κ_{rank} it is easy to see first and second cases. The third case becomes impossible as for a constant κ_{rank} it will asymptotically be smaller than $\frac{\rho_1}{(1+\sqrt{c})}$ where ρ_1 is of magnitude $\sqrt{\min(n,p)}$.

b) The consistent rank estimator $\hat{r}_n^{cst} = \sum_{i=1}^{\min(n,p)} \mathbf{1}\left(\frac{\tilde{\lambda}_i}{\sqrt{\max(n,p)}\hat{\sigma}} > 1 + \sqrt{\min(c, c^{-1})}\right)$, where $\hat{\sigma} = \frac{1}{\sqrt{\max(n,p)}} \frac{\tilde{\lambda}_{med}}{\mu_{med}}$ as in ***Lemma 5***.

$$\hat{r}_n^{cst} \xrightarrow[n,p\to\infty]{P} r_0 \tag{2.31}$$

Proof of this part is obvious by directly applying asymptotic limits of singular values in ***Theorem 2***.

Remarks:

1. The accurate rank estimation helps infer the hidden components. The robust rank estimator $\hat{r}^{rbt}$ is likely to underestimate the true rank, but robust in many

occasions. $\hat{r}^{cst}$ on the other hand, could either underestimate or overestimate (Figure 2.36) the true rank when noise assumptions are violated. We use the robust rank estimator $\hat{r}^{rbt}$ throughout the program and empirically $\kappa_{rank} = 6$ is a good solution.

2. Although there is a risk of underestimating the rank r_0 using $\hat{r}_n^{rbt}$, this estimator is robust against violated assumptions. Especially for elements in X_n that do not follow i.i.d, e.g. truncated normal. Problem with $\hat{r}_n^{cst}$ is to estimate accurately the $\hat{\sigma}$, and this boundary $1 + \sqrt{\min(c, c^{-1})}$ totally depends on the quarter-circle law, which might not follow if assumptions on X_n are violated.

3. Based on ***Lemma 5*** remarks, when $p \ll n$, $\hat{\sigma}$ will dramatically overestimate the σ. Using $\hat{r}_n^{cst}$ under this circumstance could result in underestimation (Figure 2.37) of the true rank r_0.

4. Although we want r, we can only accurately estimate the true r_0, number of true singular values θ_i in P_n that are greater than $c^{1/4}$ or $(c^{-1})^{1/4}$, instead of true r, number of all nonzero true singular in P_n. However in our model, if n, p get big enough, r will always be equal to r_0.

Discussion 7 (Window Merging Improves Spectrum Estimation):
Similar to conditions as in ***Theorem 2***. Assume a random matrix of size $n \times p_2$, $\tilde{X}_n^{(2)} = X_n^{(2)} + P_n^{(2)}$ and an extended random matrix of size $n \times p$, $\tilde{X}_n = [\tilde{X}_n^{(1)}, \tilde{X}_n^{(2)}, \tilde{X}_n^{(3)}]$, $P_n^{(i)} = U\Theta^{(i)}V^{(i)}$. The singular value and vector estimation improves from $\tilde{X}_n^{(2)}$ to $\tilde{X}_n$. More specifically, in terms of the biases of singular value ρ_k in (2.10), singular vector $\tilde{u}_k$ and $\tilde{v}_k$ to the true parameters ζ_k, u_k and v_k. $c = \frac{1}{1/c_1+1/c_2+1/c_3} < c_2$.

As we extend the window from $\tilde{X}_n^{(2)}$ to $\tilde{X}_n$.

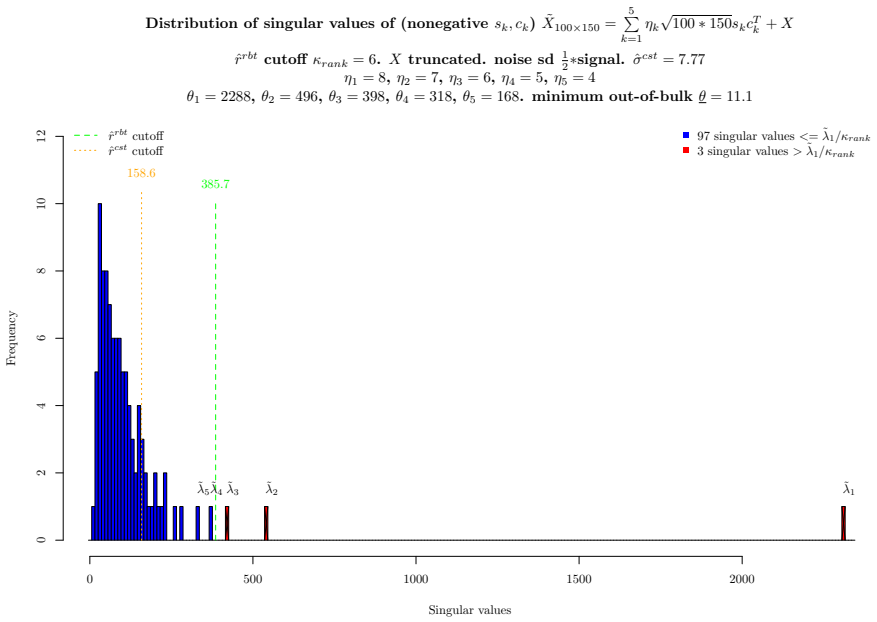


Figure 2.36: Rank estimation. In this example, true $r = 5$, noise sd is proportional to the true signal. Estimation $\hat{r}^{rbt} = 3$, $\hat{r}^{cst} > 10$.

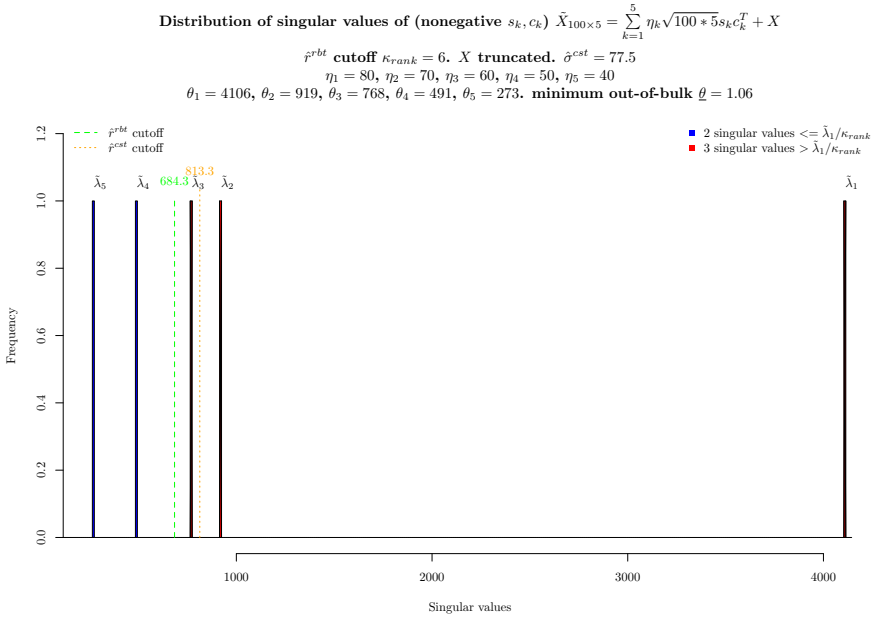


Figure 2.37: Rank estimation. In this example, $p = 5 \ll n = 100$, true $r = 5$, estimated $\hat{r}^{rbt} = 3$, $\hat{r}^{cst} = 2$.

The error ratio for the singular value limits ζ_k ($= \lim_{n,p\to\infty} \frac{\theta_k}{\sqrt{np}\sigma}$) satisfies,

$$
\begin{aligned}
&|\frac{(\rho_k - \sqrt{\min(n,p)}\zeta_k)/\sqrt{\min(n,p)}}{(\rho_k^{(2)} - \sqrt{\min(n,p_2)}\zeta_k)/\sqrt{\min(n,p_2)}}| \\
&= \left| \frac{\sqrt{\frac{(1+\min(n,p)\zeta_k^2)(\min(c,c^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2}} - \sqrt{\min(n,p)}\zeta_k}{\sqrt{\frac{(1+\min(n,p_2)\zeta_k^2)(\min(c_2,c_2^{-1})+\min(n,p_2)\zeta_k^2)}{\min(n,p_2)\zeta_k^2}} - \sqrt{\min(n,p_2)}\zeta_k} \right| \frac{\sqrt{\min(n,p_2)}}{\sqrt{\min(n,p)}} \quad < 1
\end{aligned}
\tag{2.32}
$$

The error ratio of left singular vectors satisfies,

$$
\begin{aligned}
&\frac{1-|\langle \tilde{u}_k, u_k\rangle|^2}{1-|\langle \tilde{u}_k^{(2)}, u_k\rangle|^2} \\
&= \frac{\min(c,1)(\min(1,c^{-1})+\min(n,p)\zeta_k^2)\min(n,p_2)\zeta_k^2(\min(n,p_2)\zeta_k^2+\min(c_2,1))}{\min(c_2,1)(\min(1,c_2^{-1})+\min(n,p_2)\zeta_k^2)\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(c,1))} \\
&< 1
\end{aligned}
\tag{2.33}
$$

The error ratio of right singular vectors satisfies,

$$
\begin{aligned}
&\frac{1-|\langle \tilde{v}_k[(p_1+1):(p_1+p_2)], v_k[(p_1+1):(p_1+p_2)]\rangle|^2}{1-|\langle \tilde{v}_k^{(2)}, v_k^{(2)}\rangle|^2} \\
&\approx \frac{\min(1,c^{-1})(\min(c,1)+\min(n,p)\zeta_k^2)\min(n,p_2)\zeta_k^2(\min(n,p_2)\zeta_k^2+\min(1,c_2^{-1}))}{\min(1,c_2^{-1})(\min(c_2,1)+\min(n,p_2)\zeta_k^2)\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(1,c^{-1}))} \\
&< 1
\end{aligned}
\tag{2.34}
$$

The difference between spectrum estimates in small window $\tilde{X}_n^{(2)}$ and extended window $\tilde{X}$ satisfies

$$
\begin{aligned}
&1 \geq |\langle \tilde{u}_k, \tilde{u}_k^{(2)}\rangle| \\
&\geq \sqrt{1 - \frac{\min(c,1)(\min(1,c^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(c,1))}} \\
&\sqrt{1 - \frac{\min(c_2,1)(\min(1,c_2^{-1})+\min(n,p_2)\zeta_k^2)}{\min(n,p_2)\zeta_k^2(\min(n,p_2)\zeta_k^2+\min(c_2,1))}} - \\
&\sqrt{\frac{\min(c,1)(\min(1,c^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(c,1))}} \sqrt{\frac{\min(c_2,1)(\min(1,c_2^{-1})+\min(n,p)\zeta_k^2)}{\min(n,p)\zeta_k^2(\min(n,p)\zeta_k^2+\min(c_2,1))}}
\end{aligned}
\tag{2.35}
$$

This lower bound could be much tighter, because of the dependence of $\tilde{X}_n^{(2)}$ and $\tilde{X}_n$. However, it is difficult to compute based on correlated random matrices.

Proof: The error formulas are derived using asymptotic limits of singular values and vectors in ***Theorem 2***. The proof of the inequality is through direct application of these formulas in three scenarios 1) $n \leq p_2 < p$. 2) $p_2 < n < p$. 3) $p_2 < p \leq n$. Here we only show the first scenario,

$$|\frac{(\rho_k - \sqrt{n}\zeta_k)/\sqrt{n}}{(\rho_k^{(2)} - \sqrt{n}\zeta_k)/\sqrt{n}}| = \frac{\sqrt{(\frac{1}{n} + \zeta_k^2)(\frac{1}{p} + \zeta_k^2)} - \zeta_k^2}{\sqrt{(\frac{1}{n} + \zeta_k^2)(\frac{1}{p_2} + \zeta_k^2)} - \zeta_k^2} < 1 \tag{2.36}$$

$$\frac{1 - |\langle \tilde{u}_k, u_k \rangle|^2}{1 - |\langle \tilde{u}_k^{(2)}, u_k \rangle|^2} = \frac{c(n\zeta_k^2 + c_2)}{c_2(n\zeta_k^2 + c)} < 1 \tag{2.37}$$

$$\frac{1 - |\langle \tilde{v}_k[(p_1+1):(p_1+p_2)], v_k[(p_1+1):(p_1+p_2)] \rangle|^2}{1 - |\langle \tilde{v}_k^{(2)}, v_k^{(2)} \rangle|^2} \approx \frac{c + n\zeta_k^2}{c_2 + n\zeta_k^2} < 1 \tag{2.38}$$

$$1 \geq |\langle \tilde{u}_k, \tilde{u}_k^{(2)} \rangle| \geq \tag{2.39}$$

$$\sqrt{1 - \frac{c(1+n\zeta_k^2)}{n\zeta_k^2(n\zeta_k^2 + c)}}\sqrt{1 - \frac{c_2(1+n\zeta_k^2)}{n\zeta_k^2(n\zeta_k^2 + c_2)}} - \sqrt{\frac{c(1+n\zeta_k^2)}{n\zeta_k^2(n\zeta_k^2 + c)}}\sqrt{\frac{c_2(1+n\zeta_k^2)}{n\zeta_k^2(n\zeta_k^2 + c_2)}} \tag{2.40}$$

Discussion 8 (Model Behavior during Sea-island Process):
During sea-island learning, our model sequentially compute each scan to the previous to decide whether it represent a new spectrum or not. We show that this process can recover the true model with false identification only for scans whose spectrum overlapping with others and its intensity is too small to stand out.

Procedure:
Average of normalized previous scans in the current cluster (same spectrum scans)

starting at j_0 (initialized as 1, to be updated during the process),

$$u_{pre} = \frac{1}{j - j_0} \sum_{i=j_0}^{j-1} \frac{\tilde{X}_{.,i}}{||\tilde{X}_{.,i}||} \tag{2.41}$$

A key step of the process computes for a threshold κ_{si} (usually $\in [0.9, 1)$) if

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \geq \kappa_{si} \tag{2.42}$$

then the scan j is assigned to the current cluster, otherwise it starts a new cluster.

Statement:

$$\tilde{X}^{(w)}_{.,j} = X^{(w)}_{.,j} + \sum_{k=1}^{r_w} \theta^{(w)}_k u^{(w)}_k v^{(w)}_k[j] \tag{2.43}$$

$$\tilde{X} = [\tilde{X}^{(1)}_{.,j}, \tilde{X}^{(2)}_{.,j}, \cdots, \tilde{X}^{(W)}_{.,j}] \tag{2.44}$$

Let $\xi(i)$ denote the index i of $\tilde{X}$ corresponding to that in window $\tilde{X}^{(w_i)}$, i.e. $\tilde{X}_{.,i} = \tilde{X}^{(w_i)}_{.,\xi^{w_i}(i)}$. Given conditions in ***Theorem 4*** where p is finite (a cluster is usually of size smaller than column size), we show that

$$\begin{aligned}
&||u_{pre}|| \xrightarrow{P} \frac{1}{(j-j_0)^2}(j - j_0 + 2 \sum_{j_0 \leq i < m \leq j-1} \frac{\tilde{X}_{.,i}}{||\tilde{X}_{.,i}||} \frac{\tilde{X}_{.,m}}{||\tilde{X}_{.,m}||}) \\
&= \frac{1}{(j-j_0)^2} \Bigg(j - j_0 + \qquad (2.45) \\
&\quad 2 \sum_{j_0 \leq i < m \leq j-1} \frac{\sum_{k_1,k_2} \zeta^{(w_i)}_{k_1} \zeta^{(w_m)}_{k_2} v^{(w_i)}_{k_1}[\xi(i)] v^{(w_m)}_{k_2}[\xi(m)] [u^{(w_i)}_{k_i}]^T u^{(w_m)}_{k_m}}{\sqrt{\sigma^2 + \sum_{k=1}^{r_{w_i}} (\zeta^{(w_i)}_k)^2 v^{(w_i)}_k[\xi(i)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{r_{w_m}} (\zeta^{(w_m)}_k)^2 v^{(w_m)}_k[\xi(m)]^2}} \Bigg) \\
&\leq 1 \qquad (2.46)
\end{aligned}$$

The equality approximately holds, when all $j_0 : (j-1)$ scans belong to one spectrum and in comparison $\sum_{k=1}^{r_{w_i}} (\zeta^{(w_i)}_k)^2 v^{(w_i)}_k[\xi(i)]^2$ and $\sum_{k=1}^{r_{wm}} (\zeta^{(w_m)}_k)^2 v^{(w_m)}_k[\xi(m)]^2$ are much larger than σ^2.

Illustration:

We assume the sample data is composed of two overlapping spectra. Assign all overlapping scans to $\tilde{X}_2$.

$$\tilde{X} = [\tilde{X}_1, \tilde{X}_2, \tilde{X}_3] \tag{2.47}$$

$$\begin{aligned} &\tilde{X}_1 = X_1 + \theta_1^{(1)} u_1 [v_1^{(1)}]^T, \quad \tilde{X}_2 = X_2 + \theta_1^{(2)} u_1 [v_1^{(2)}]^T + \theta_2^{(2)} u_2 [v_2^{(2)}]^T, \\ &\tilde{X}_3 = X_3 + \theta_2^{(3)} u_2 [v_2^{(3)}]^T \end{aligned} \tag{2.48}$$

Scan index: $1, 2, \cdots, p_1, p_1+1, p_1+2, \cdots, p_1+p_2, p_1+p_2+1, \cdots, p_1+p_2+p_3$

Truth: all scans in $\tilde{X}_1$ and $\tilde{X}_3$ are sea scans (rank 1), and all in $\tilde{X}_2$ are island scans.

Estimation: we are able to correctly identify most sea scans in $\tilde{X}_1$ and $\tilde{X}_3$ and two ends of $\tilde{X}_2$ are likely to be mistaken as "sea".

We show that (Figure 2.7)

1. If $1 < j \leq p_1$,

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \to \tag{2.49}$$

$$\frac{1}{j-1} \sum_{i=1}^{j-1} \frac{(\zeta_1^{(1)})^2 v_1^{(1)}[\xi(i)] v_1^{(1)}[\xi(j)]}{\sqrt{\sigma^2 + (\zeta_1^{(1)})^2 v_1^{(1)}[\xi(i)]^2} \sqrt{\sigma^2 + (\zeta_1^{(1)})^2 v_1^{(1)}[\xi(j)]^2}} \frac{1}{||u_{pre}||} \tag{2.50}$$

   This value would be very close to 1 which shows our algorithm in general will not miss identify clean spectrum scans.

2. If $j = p_1 + 1$,

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \to \tag{2.51}$$

$$\frac{1}{p_1} \sum_{i=1}^{p_1} \frac{\zeta_1^{(1)} \zeta_1^{(2)} v_1^{(2)}[\xi(i)] v_1^{(2)}[\xi(j)]}{\sqrt{\sigma^2 + (\zeta_1^{(1)})^2 v_1^{(1)}[\xi(i)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(j)]^2}} \frac{1}{||u_{pre}||} \tag{2.52}$$

If $v_2^{(2)}[\xi(j)]$ is still too small, this ratio would still be very close to 1, which means the beginning of intersection between two spectra is generally hard to be detected.

3. If $p_1 + 1 < j \leq p_1 + p_2$,

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \rightarrow \tag{2.53}$$

$$\Big(\frac{1}{j-1} \sum_{i=1}^{p_1} \frac{\zeta_1^{(1)} \zeta_1^{(2)} v_1^{(1)}[\xi(i)] v_1^{(2)}[\xi(j)]}{\sqrt{\sigma^2 + (\zeta_1^{(1)})^2 v_1^{(1)}[\xi(i)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(j)]^2}} +$$

$$\frac{1}{j-1} \sum_{i=p_1+1}^{j-1} \frac{\sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(i)] v_k^{(2)}[\xi(j)]}{\sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(i)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(j)]^2}} \Big) \frac{1}{||u_{pre}||} \tag{2.54}$$

as j increases, $v_1^{(2)}[\xi(j)]$ gets smaller, $v_2^{(2)}[\xi(j)]$ gets bigger, both terms of the summation above will decrease. With enough scans from $X^{(2)}$, this inner product will decrease below a certain threshold, which is what we use to determine an occurrence of island (a scan where spectra overlap). **Before this happens, a few island scans are misidentified as continuing sea scans from the previous window.**

If the inner product below the threshold, then we denote j as j_0, let $j = j_0 + 1$, $u_{pre} = \frac{\tilde{X}_{.,j_0}}{||\tilde{X}_{.,j_0}||}$,

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \tag{2.55}$$

$$\rightarrow \frac{1}{j-j_0} \sum_{i=j_0}^{j-1} \frac{\sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(i)] v_k^{(2)}[\xi(j)]}{\sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(i)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(j)]^2}} \frac{1}{||u_{pre}||} \tag{2.56}$$

This new inner product is still very likely to be below threshold and forming a new island because both spectra have relatively large intensities in this region. Thus it is very likely to observe a few consecutive islands in this

stage. In the end of this window, it is possible spectrum 2 totally disappears and spectrum 3 dominates. **This is where misidentification most likely happens**.

4. If $j \geq p_1 + p_2 + 1$, similar to case 2 and 3, we just briefly mention when $j > p_1 + p_2 + 1$ and $j_0 \geq p_1 + 1$.

$$\langle \frac{u_{pre}}{||u_{pre}||}, \frac{\tilde{X}_{.,j}}{||\tilde{X}_{.,j}||} \rangle \rightarrow$$
$$(\frac{1}{j-j_0} \sum_{i=j_0}^{p_1+p_2} \frac{\zeta_2^{(2)}\zeta_2^{(3)} v_2^{(2)}[\xi(i)] v_2^{(3)}[\xi(j)]}{\sqrt{\sigma^2 + \sum_{k=1}^{r_2} (\zeta_k^{(2)})^2 v_k^{(2)}[\xi(i)]^2} \sqrt{\sigma^2 + (\zeta_2^{(3)})^2 v_2^{(3)}[\xi(i)]^2}}$$
$$+ \frac{1}{j-j_0} \sum_{i=p_1+p_2+1}^{j-1} \frac{(\zeta_2^{(3)})^2 v_2^{(3)}[\xi(i)] v_2^{(3)}[\xi(j)]}{\sqrt{\sigma^2 + (\zeta_2^{(3)})^2 v_2^{(3)}[\xi(i)]^2} \sqrt{\sigma^2 + (\zeta_2^{(3)})^2 v_2^{(3)}[\xi(j)]^2}}) \frac{1}{||u_{pre}||} \tag{2.57}$$

This inner product is mostly likely going up as j increases so that we will eventually assign all remaining clean spectrum 3 scans into one cluster.

Proof of the statement:

$$\tilde{X}_{.,j}^{(w)} = X_{.,j}^{(w)} + \sum_{k=1}^{r_w} \theta_k^{(w)} u_k^{(w)} v_k^{(w)}[j], \quad \tilde{X} = [\tilde{X}_{.,j}^{(1)}, \tilde{X}_{.,j}^{(2)}, \cdots, \tilde{X}_{.,j}^{(W)}] \tag{2.58}$$

$\xi(i)$ denotes the index i of $\tilde{X}$ corresponding to that in window $\tilde{X}^{(w_i)}$,

$$\frac{\tilde{X}_{.,j}^{(w)}}{||\tilde{X}_{.,j}^{(w)}||} = \frac{\tilde{X}_{.,j}^{(w)}/\sqrt{n}}{\sqrt{\frac{1}{n}[X_{.,j}^{(w)}]^T X_{.,j}^{(w)} + \sum_{k=1}^{r} \frac{(\theta_k^{(w)})^2}{n} v_k^{(w)}[\xi(j)]^2 + \frac{1}{n} \sum_{k=1}^{r} \theta_k^{(w)} v_k^{(w)}[\xi(j)][X_{.,j}^{(w)}]^T u_k}}$$
$$\xrightarrow[n\to\infty]{P} \frac{\tilde{X}_{.,j}^{(w)}/\sqrt{n}}{\sqrt{\sigma^2 + \sum_{k=1}^{r} (\zeta_k^{(w)})^2 v_k^{(w)}[\xi(j)]^2}} \tag{2.59}$$

For $w_1 \neq w_2$ or $j \neq m$,

$$\begin{aligned}
&\langle \frac{\tilde{X}^{(w_1)}_{.,j}}{||\tilde{X}^{(w_1)}_{.,j}||}, \frac{\tilde{X}^{(w_2)}_{.,m}}{||\tilde{X}^{(w_2)}_{.,m}||} \rangle \\
&= \frac{1}{\sqrt{\sigma^2 + \sum_{k=1}^{r_{w_1}} (\zeta_k^{(w_1)})^2 v_k^{(w_1)}[\xi(j)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{r_{w_2}} (\zeta_k^{(w_2)})^2 v_k^{(w_2)}[\xi(m)]^2}} \\
&\left(\frac{1}{n} X^{(w_1)}_{.,j} X^{(w_2)}_{.,m} + \frac{1}{n} \sum_{i=1}^{r_{w_2}} \theta_k^{(w_2)} v_k^{(w_2)}[\xi(m)][X^{(w_1)}]^T_{.,j} u_k^{(w_2)} \right. \\
&\left. + \frac{1}{n} \sum_{i=1}^{r_{w_1}} \theta_k^{(w_1)} v_k^{(w_1)}[\xi(j)][X^{(w_2)}]^T_{.,m} u_k^{(w_1)} + \frac{1}{n} \sum_{k_1,k_2} \theta_{k_1}^{(w_1)} \theta_{k_2}^{(w_2)} v_{k_1}^{(w_1)}[\xi(j)] v_{k_2}^{(w_2)}[\xi(m)][u^{(w_1)}]^T_{k_1} u_{k_2}^{(w_2)} \right) \\
&\to \frac{\sum_{k_1,k_2} \zeta_{k_1}^{(w_1)} \zeta_{k_2}^{(w_2)} v_{k_1}^{(w_1)}[\xi(j)] v_{k_2}^{(w_2)}[\xi(m)][u^{(w_1)}]^T_{k_1} u_{k_2}^{(w_2)}}{\sqrt{\sigma^2 + \sum_{k=1}^{r_{w_1}} (\zeta_k^{(w_1)})^2 v_k^{(w_1)}[\xi(j)]^2} \sqrt{\sigma^2 + \sum_{k=1}^{r_{w_2}} (\zeta_k^{(w_2)})^2 v_k^{(w_2)}[\xi(m)]^2}}
\end{aligned} \tag{2.60}$$

The proof of this step uses P27 [Fer96] the *Lindeberg-Feller Theorem*.

Proof of the illustration:

Here in our set-up, $r_{w_1} = 1$, $r_{w_2} = 2$, $r_{w_3} = 1$. Apply (2.60) to each case above we prove the illustration.

2.7.2 Non-negative Matrix Factorization

Theorem 9:

If rank of s nonnegative matrix P is r, $P = S_{true} C_{true}$, where non-negative S_{true} is of size $n \times r$, non-negative C_{true} is of size $r \times p$, and P has $r - 1$ different non-overlapping (single component with coefficient instead of linear combination of components) columns (or rows) then

$$\min_{S_{n\times r}, C_{r\times p}} ||P - SC||_2^2 + \sum_j |C_{.,j}|_0 + \sum_i |S_{i,.}|_0$$

would uniquely recover S_{true}, C_{true}.

Proof:

Let $r = 2$, it is easy to see without noise, $S_{true}C_{true}$ achieves the minimal $||P - SC||_2^2$, so does $S_{true}Q^TQC_{true}$, Q is a $2 * 2$ orthonormal matrix (rotation matrix when $r = 2$). In many cases, optimizing $||P-SC||_2^2$ alone would not give a unique non-negative solution. However,

1. If P has non-overlapping columns. Let $C_{1,\cdot} = [c_1^T, 0_{1*p_0}]$, $C_{2,\cdot} = [0_{1*p_0}, c_2^T]$. Among all possible solutions only $[S_{true}]_{\cdot,1}$ can minimize $\sum_{j=1}^{p_0} |C_{\cdot,j}|_0 = p - p_0$ and $[S_{true}]_{\cdot,2}$, minimize $\sum_{j=p-p_0+1}^{p} |C_{\cdot,j}|_0 = p - p_0$. Thus S_{true} would be learned. After this, solution of C is unique, which is C_{true}.

2. If P has non-overlapping rows. Let $S_{,1} = [s_1^T, 0_{1*n_0}]^T$, $S_{,2} = [0_{1*n_0}, s_2^T]^T$. Similar to above, only C_{true} can minimize $\sum_{i=1}^{n_0} |S_{i,\cdot}|_0 = n_0$ and $\sum_{i=n-n_0+1}^{n} |S_{i,\cdot}|_0 = n_0$. Given C_{true}, we can solve $S = S_{true}$

Remark. We do not use this optimization framework in the program because it adds complexity to the computation. In the future work, it is possible to add computationally feasible penalties (L_1, etc.) to get more sparse NMF results.

2.8 Lemmas and Related Theorems

2.8.1 Random Matrix Lemmas

Lemma 10: For a $n \times p$ matrix X_n, where $[X_n]_{i,j}$ are i.i.d r.v.s, $E([X_n]_{i,j}) = \mu$, $Var([X_n]_{i,j}) = \sigma^2$, $u = [\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \cdots, \frac{1}{\sqrt{n}}]^T$, $v = [\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}, \cdots, \frac{1}{\sqrt{p}}]^T$, then, as n or $p \to \infty$

$$(u^T X_n v - \sqrt{np}\mu) \xrightarrow[\infty]{L} N(0, \sigma^2) \tag{2.61}$$

Proof:

Assume $p \to \infty$, $u^T X_n v = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sqrt{p}} \sum_{j=1}^{p} X_{i,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{p} \bar{X}_{i,\cdot}$, because for $o_p(1)$ in terms of p, $\sqrt{p}\bar{X}_{i,\cdot} = N(\sqrt{p}\mu, \sigma^2) + o_p(1)$, in terms either both n and p, we have $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sqrt{p}\bar{X}_{i,\cdot} = N(\sqrt{np}\mu, \sigma^2) + o_p(1)$.

Lemma 11 (A Collection of Some Useful Results):

For a $n \times p$ matrix X_n, where $[X_n]_{\{i,j\}} \overset{i.i.d}{\sim} N(0, \sigma^2)$, and a $p \times r$ deterministic matrix $V = [v_1, v_2, \cdots, v_r]$, $n \times r$ deterministic matrix $U = [u_1, u_2, \cdots, u_r]$, then

$$X_n^T X_n \sim \mathcal{W}(\sigma^2 I_{p\times p}, n), \quad X_n X_n^T \sim \mathcal{W}(\sigma^2 I_{n\times n}, p) \tag{2.62}$$

$\mathcal{W}()$ is a Wishart distribution.

$$V^T X_n^T X_n V \sim \mathcal{W}(\sigma^2 V^T V, n), \quad V^T V = \begin{bmatrix} v_1^T v_1 & v_1^T v_2 & \cdots & v_1^T v_r \\ v_2^T v_1 & v_2^T v_2 & \cdots & v_2^T v_r \\ \vdots & \vdots & \ddots & \vdots \\ v_r^T v_1 & v_r^T v_2 & \cdots & v_r^T v_r \end{bmatrix}, \tag{2.63}$$

$$u_i^T X_n v_j \sim N((\sum_{k=1}^{n} u_{ki} * \sum_{k=1}^{p} v_{kj})\mu, \sigma^2), \tag{2.64}$$

marginally,

$$v_i^T X_n^T X_n v_i \sim (v_i^T v_i)\sigma^2 \chi_n^2, \quad 1 \le i \le r \tag{2.65}$$

$$u_i^T X_n X_n^T u_i \sim (u_i^T u_i)\sigma^2 \chi_p^2 \quad 1 \le i \le r \tag{2.66}$$

This can also be derived through the n-dimension r.v. $X_n v_i \sim N(0, v_i^T v_i \sigma^2)$ and that $v_i^T X_n^T X_n v_i \sim (v_i^T v_i)\sigma^2 \chi_n^2$.

$$\begin{aligned} Y_{ij} &= v_i^T X_n^T X_n v_j \\ &\sim (\frac{1}{\sigma^4 ||v_i||^2 ||v_j||^2 (1-\rho_{ij}^2)})^{\frac{n}{2}} \frac{|y_{ij}|^{\frac{n-1}{2}}}{\Gamma(\frac{n}{2})\sqrt{2^{n-1}\pi(1-\rho_{ij}^2)(\sigma^2||v_i||||v_j||)^{n+1}}} \\ &K_{\frac{n-1}{2}}(\frac{|y_{ij}|}{\sigma^2||v_i||||v_j||(1-\rho_{ij}^2)}) \exp(\frac{\rho_{ij} y}{\sigma^2||v_i||||v_j||(1-\rho_{ij}^2)}). \end{aligned} \tag{2.67}$$

where $\rho_{ij} = \frac{v_i^T v_j}{||v_i||||v_j||}$, $K_{\frac{n-1}{2}}$ is the modified Bessel function of the second kind.

$$\begin{aligned} E(Y_{ij}) &= 0 + 2 * \frac{\rho_{ij}}{\sigma^2||v_i||||v_j||(1-\rho_{ij}^2)} * \frac{n}{2} / \frac{1}{\sigma^4||v_i||^2||v_j||^2(1-\rho_{ij}^2)} \\ &= n\rho_{ij}||v_i||||v_j||\sigma^2 = n\sigma^2 v_i^T v_j \end{aligned} \tag{2.68}$$

$$\begin{aligned} Var(Y_{ij}) &= \frac{2n(1 + 2(\frac{\rho_{ij}}{\sigma^2||v_i||||v_j||(1-\rho_{ij}^2)})^2 / \frac{1}{\sigma^4||v_i||^2||v_j||^2(1-\rho_{ij}^2)})}{\frac{1}{\sigma^4||v_i||^2||v_j||^2(1-\rho_{ij}^2)}} \\ &= 2n(1+\rho_{ij}^2)\sigma^4||v_i||^2||v_j||^2 = 2n(1 + (\frac{v_i^T v_j}{||v_i||||v_j||})^2)\sigma^4||v_i||^2||v_j||^2 \\ &= 2n\sigma^4(||v_i||^2||v_j||^2 + (v_i^T v_j)^2) \end{aligned} \tag{2.69}$$

If we let U and V be orthonormal matrices, the formula is simplified to

$$Y_{ij} = v_i^T X_n^T X_n v_j \sim \frac{|y_{ij}|^{\frac{n-1}{2}}}{\Gamma(\frac{n}{2})\sqrt{2^{n-1}\pi(\sigma^2)^{n+1}}} K_{\frac{n-1}{2}}(\frac{|y_{ij}|}{\sigma^2}) \tag{2.70}$$

Similarly,

$$Z_{ij} = u_i^T X_n X_n^T u_j \sim \frac{|z_{ij}|^{\frac{p-1}{2}}}{\Gamma(\frac{p}{2})\sqrt{2^{p-1}\pi(\sigma^2)^{p+1}}} K_{\frac{p-1}{2}}(\frac{|z_{ij}|}{\sigma^2}) \tag{2.71}$$

Lemma 12: For a $n \times p$ matrix X_n, where $[X_n]_{\{i,j\}} \overset{i.i.d}{\sim} N(0, \frac{\sigma^2}{n})$, and a $p \times r$ deterministic matrix $V = [v_1, v_2, \cdots, v_r]$, $n \times r$ deterministic matrix $U = [u_1, u_2, \cdots, u_r]$, then

$$Y = V^T X_n^T X_n V \xrightarrow[n\to\infty]{P} (V^T V)\sigma^2 \tag{2.72}$$

Proof:

Direct application of ***Lemma 11***, with variance $\frac{\sigma^2}{n}$ instead of σ^2. The diagonal elements of Y, $v_i^T X_n^T X_n v_i \sim v_i^T v_i \chi_n^2 \frac{\sigma^2}{n} \to v_i^T v_i \sigma^2$. The off-diagonal elements have $E(Y_{ij}) = n\frac{\sigma^2}{n} v_i^T v_j = v_i^T v_j \sigma^2$, $Var(Y_{ij}) = 2n\frac{\sigma^4}{n^2}(||v_i||^2||v_j||^2 + (v_i^T v_j)^2) \xrightarrow[\infty]{n} 0$, thus $Y_{ij} \xrightarrow{P} v_i^T v_j \sigma^2$

2.8.2 Other Existing Theorems in the Literature on Uniqueness of NMF

Theorem of [DS03]:

A non-negative matrix P has a unique NMF solution, if it satisfies the 3 conditions:

- Generative model. The actual data P,

$$P = S_{true}C_{true}, \quad S_{true} \geq 0, C_{true} \geq 0 \tag{2.73}$$

- Seperability. Every spectrum has their unique non-zero masses.

$$\forall 1 \leq m \leq n, \quad \sum_{j=1}^{p} \mathbf{1}([S_{true}]_{m,j} > 0) \leq 1. \tag{2.74}$$

  Suppose there are r spectra in S_{true}, forming G groups with equal size r/G.

$$S_{true} = [S_{true}^{(1)}, S_{true}^{(2)}, \cdots, S_{true}^{(G)}] \tag{2.75}$$

- Complete Factorial Sampling. For each group of spectra, there is only one spectrum included and each group has to be present, so that there are $G^{r/G}$ possible combinations.

  Let i_g be the spectrum index in S_{true} which belongs to group g. For any $i_1, i_2, \cdots, i_G$, there exists $1 \leq j \leq p$, that $C_{i_g,j} > 0, \forall 1 \leq g \leq G$.

Remark: The theorem is a sufficient condition which is easily violated in our data, where spectra overlap.

Definition: A simplicial cone generated by vectors $\{\phi_1, \phi_2, \cdots, \phi_r\}$ is $\Gamma = \{x : x = \sum_{j=1}^{r} c_j\phi_j, c_j \geq 0\}$.
Definition: An extreme ray of a convex cone Γ is the ray $R_x = \{cx : x \geq 0\}$, where x can not be a combination of two points which don't belong to the ray.
Definition: Let $\mathbb{A}^*$ denote the dual to $\mathbb{A}$, $\mathbb{A}^* = \{x : x^T a \geq 0, \forall a \in \mathbb{A}\}$.

Theorem 1 of [LCP+08]:

NMF on $P = SC$ is unique if and only if $\mathbb{A} = \mathbb{R}^+_r$ is the only simplicial cone with r extreme rays such that $span^+(S^T) \subset \mathbb{A} \subset span^+(C)^*$.

Remark: The condition is extremely hard to check in practice, is NP-hard as *Remark 1* in [HSS14] and [Vav09] pointed out on NMF rank determination.

References

[AW13] E. Alan and Y. Wang. Random matrix theory and its innovative applications. *Advances in Applied Mathematics, Modeling, and Computational Science. Springer US*, pages 91–116, 2013.

[Bai99] Z. Bai. Methodologies in spectral analysis of large-dimensional random matrices, a review. *Statist. Sinica*, 9(3):611–677, 1999.

[BB74] J. Biller and K. Biemann. Reconstructed mass spectra, a novel approach for the utilization of gas chromatograph-mass spectrometer data. *Analatical Letters*, 7(7):515–528, 1974.

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

[BGGM11] F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *arXiv.org*, 1009.0145v4, 2011.

[BGN12] F. Benaych-Georges and R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *arXiv.org*, 1103.2221v2, 2012.

[BS06] Jinho Baik and Jack Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[BY08] Z. Bai and J. Yao. Central limit theorems for eigenvalues in a spiked population model. *Annales de l'IHP Probabilités et statistiques*, 44(3):447–474, 2008.

[CCHM12] Francois Chapon, Romain Couillet, Walid Hachem, and Xavier Mestre. The outliers among the singular values of large rectangular random matrices with additive fixed rank deformation. *arXiv preprint arXiv:1207.0471*, 2012.

[Che54] H. Chernoff. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578, 1954.

[Col92] B. Colby. Spectral deconvolution for overlapping gc/ms components. *Journal of the American Society for Mass Spectrometry*, 3:558–562, 1992.

[CZPA09] A. Cichocki, R. Zdunnek, A. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations*. John Wiley & Sons, Ltd, West Sussex, United Kingdom, first edition, 2009.

[DS03] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts?. *Advances in neural information processing systems*, 2003.

[DSRD76] R. Dromey, M. Stefik, T. Rindfleisch, and A. Duffield. Extraction of mass spectra free of background and neighboring component contributions from gas chromatography/mass spectrometry data. *Analatical Chemistry*, 48(9):1369–1375, 1976.

[Fer96] T. Ferguson. *A Course in Large Sample Theory*. Chapman & Hall/CRC, Boca Raton, Florida, first edition, 1996.

[GD14] M. Gavish and D. Donoho. Optimal shrinkage of singular values. *arXiv.org*, 1405.7511v2, 2014.

[HJ85] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 1985.

[HSS14] K. Huang, N. Sidiropoulos, and A. Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 62(1):211–224, 2014.

[JGN+04] P. Jonsson, J. Gullberg, A. Nordstrom, M. Kusano, M. Kowalczyk, M. Sjostrom, and T. Moritz. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Analatical Chemistry*, 76(6):1738–1745, 2004.

[JLP+09] Zhe Ji, Ju Youn Lee, Zhenhua Pan, Bingjun Jiang, and Bin Tian. Progressive lengthening of 3? untranslated regions of mrnas by alternative polyadenylation during mouse embryonic development. *Proceedings of the National Academy of Sciences*, 106(17):7028–7033, 2009.

[Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[JT09] Zhe Ji and Bin Tian. Reprogramming of 3? untranslated regions of mrnas by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One*, 4(12):e8419, 2009.

[Kud63] A. Kudo. A multivariate analogue of the one-sided test. *Biometrika*, 50(3/4):403–418, 1963.

[LCP+08] H. Laurberg, M. Christensen, M. Plumbley, L. Hansen, and S. Jensen. Theorems on positive data: On the uniqueness of nmf. *Computational Intelligence and Neuroscience*, 2008(764206), 2008.

[LS00] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562. MIT Press, 2000.

[MB09] Christine Mayr and David P Bartel. Widespread shortening of 3? utrs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell*, 138(4):673–684, 2009.

[MP67] V. A. Marcenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Math. USSR Sbornik*, 1(4):457–483, 1967.

[PDV97] W. Pool, J. Deleeuw, and B. VanDeGraaf. Automated extraction of pure mass spectra from gas chromatographic/mass spectrometric data. *Journal of Mass Spectrometry*, 32:438–443, 1997.

[RYY00] S. Raudenbush, M. Yang, and M. Yosef. Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate laplace approximation. *Journal of Computational and Graphical Statistics*, 9(1):141–157, 2000.

[Sac90] A Sachs. The role of poly (a) in the translation and stability of mrna. *Current opinion in cell biology*, 2(6):1092–1098, 1990.

[Sha87] A. Shapiro. On differentiability of metric projections in R^n, 1:boundary case. *Proceedings of The American Mathematical Society*, 99(1), 1987.

[Sha88] A. Shapiro. Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, 56(1):4962, 1988.

[Shi12] Yongsheng Shi. Alternative polyadenylation: new insights from global analyses. *Rna*, 18(12):2105–2117, 2012.

[SNS+08] Rickard Sandberg, Joel R Neilson, Arup Sarma, Phillip A Sharp, and Christopher B Burge. Proliferating cells express mrnas with shortened 3'untranslated regions and fewer microrna target sites. *Science*, 320(5883):1643–1647, 2008.

[SPL+14] S. Shen, J. Park, Z. Lu, L. Lin, M. Henry, Y. Wu, Q. Zhou, and Y. Xing. rmats: Robust and flexible detection of differential alternative splicing from replicate rna-seq data. *Proceedings of the National Academy of Sciences of the United States of America*, 111(51):E5593E5601, 2014.

[Ste99] S. Stein. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, 10:770–781, 1999.

[TW94] Craig A Tracy and Harold Widom. Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics*, 159(1):151–174, 1994.

[Vav09] Stephen A Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.

[WH11] William R Wilson and Michael P Hay. Targeting hypoxia in cancer therapy. *Nature Reviews Cancer*, 11(6):393–410, 2011.

[Wig55] Eugene P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.