THE UNIVERSITY OF CHICAGO

MODELS AND INFERENCE FOR MICROBIOME DATA

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF STATISTICS

BY
YUNFAN TANG

CHICAGO, ILLINOIS
JUNE 2018

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Dan Nicolae, for his immensely helpful guidance and support that direct my research throughout the graduate education. His keen insights and rigorous approaches to statistical modelling have inspired many of the ideas in this work, and they will continue to have profound influences on my future endeavors.

I thank my other committee members, Matthew Stephens and Rina Barber, for their helpful feedbacks that improve the quality of my work. I also thank all faculty members at Department of Statistics for providing an intellectually stimulating environment throughout various seminars and classes. In particular, I would like to thank Rina Barber, Yali Amit and John Lafferty, with whom I had brief reading and research courses early in my PhD education, for helping me explore multiple research frontiers and ultimately find my own interest.

I also thank Katie Igartua, Carol Ober and members of the Copenhagen Prospective Study on Asthma in Childhood (COPSAC) group for providing microbiome data and offering helpful advices from geneticists' perspectives. These collaborations have deepened my understanding of statisticians' roles in applied sciences through clarifying the real scientific question of interest.

Finally, I thank my parents Lan and Xiaoyong. None of this would have been possible without their love, support and encouragement throughout my life.

ABSTRACT

Microbiome refers to the full collection of all microorganisms in a community. Recent advances in sequencing technologies have allowed scientists to quantify the microbiome compositions at an unprecedented resolution. However, high-throughput sequencing data have unique characteristics such as high dimensionality, sparseness, and compositional nature. Moreover, the phylogenetic tree that quantifies the evolutionary similarity among all taxa offers unique modelling challenges. This thesis presents novel statistical models to analyze microbiome data by leveraging these unique characteristics. The first problem we consider is testing equality of microbial compositions among groups of populations. We apply the Dirichlet-tree multinomial distribution (DTM), a generalization of the traditional Dirichlet-multinomial (DM) distribution, and design a scan statistic that takes advantage of the phylogenetic relations among taxa. We provide an upper bound of p-value using this scan statistic and show that this method has improved power in an empirical dataset and simulation. The second problem continues the investigation of DTM vs DM by introducing a penalization method that selects the best model along the DM-DTM spectrum. The last problem we address is estimating heritability when the input is a matrix of pairwise dissimilarities calculated from beta diversities. Beta diversity is an ecologically meaningful way to measure pairwise compositional differences by taking variations in all taxa into account. We extend the traditional ACE variance component model to the matrix case using Wishart distribution. We also present a new beta diversity measurement, named root-Unifrac, that matches the positive definiteness requirement of the Wishart distribution. This Wishart ACE model allows us to directly measure community heritability, which quantifies the contribution of additive genetics to overall variations in beta diversity.

CHAPTER 1

INTRODUCTION

Microbiome refers to the full collection of genes of all microbes in a community; for example, all bacteria in a fecal sample from a healthy individual. The total number of cells for human microbiome has been estimated to exceed the total number of human cells by a factor 10. The advent of next generation sequencing technologies, such as Illumina Solexa, has allowed researchers to investigate the microbiome communities at an unprecedented level of quantification while avoiding laborious cultivation of individual organisms. There have been burgeoning efforts devoted to the study of human microbiome in the past decade. For example, researchers have discovered that microbiome is associated with obesity (Devaraj et al., 2013; Tilg and Kaser, 2011), diabetes (Giongo et al., 2011; Kostic et al., 2015) and inflammatory bowel disease (Manichanh et al., 2006). Many studies are very large-scale initiatives with more than hundreds of samples, such as TwinsUK (Goodrich et al., 2016), PopGen (Krawczak et al., 2006), American Gut (McDonald et al., 2015), Human Microbiome Project (Methé et al., 2012), and FoCus (Müller et al., 2015). These studies jointly point to the fact that microbiome plays an integral part to our health, and much still remains to be explored in this area.

1.1 The role of microbiome in human health

Thanks to the increasing amount of effort devoted to microbiome research, one of the most drastic changes in our understanding of microbiome is that they are indispensable in maintaining our body's fundamental ecosystem. Typical services provided by microbiome include producing important resources and bioconversion of nutrients (Young, 2017). For example, the gut microbiome is capable of fermenting dietary polysaccharides (fiber) that cannot be digested by the host (Cockburn and Koropatkin, 2016). They also digest endogenous glycans such as those that decorate the host mucosal layer. Metabolites produced from digesting

these substances are mostly short-chain fatty acids, and they benefit both microbes themselves and the human host by promoting mucin secretion and barrier function. Another example of host-microbe co-metabolism is the biotransformations of bile salts and bile acids in the gut to generate secondary bile acids (Ridlon et al., 2006). These secondary bile acids, if accumulated to high levels in the intestine, may contribute to the pathogenesis of colon cancer, gallstones, and other gastrointestinal (GI) diseases. As a result, humans have evolved to recognize and respond to these bacterially generated secondary compounds, similar to the way we respond to the short chain fatty acids generated from complex carbohydrates. In particular, the metabolism pathway of bile acid is closely related to those of cholesterol and lipid, since bile acids are the end products of cholesterol metabolism (Wahlström et al., 2016). Therefore, alterations in the gut microbiome are regarded as potential treatments for various metabolic disorders such as obesity and insulin resistance to liver fibrosis.

The human microbiome also plays a vital role in shaping the development and activity of the host's immune system, protecting the host from colonization of pathogenic microbes (Shi et al., 2017). A large part of this function is provided through releasing beneficial nutrients such as vitamins and short chain fatty acids. In return, the immune system has evolved to regulate the homeostasis of these highly diverse and evolving microbes (Belkaid and Hand, 2014), and may even alter the structure and function of the microbiome. Studies in germ-free animals have revealed that the lack of gut microbiota caused a significant immune system deficiency. Overuse of antibiotics and changes in diet may also compromise the immune system through inducing microbial imbalance that lacks the resilience and diversity required to maintain balanced immune responses.

1.2 Microbiome analysis pipeline

So far there are two main sequencing methods to quantify the microbial composition: targeted amplicon sequencing and shotgun metagenomic sequencing. The focus of this thesis is on targeted amplicon sequencing, but the methods introduced here can be easily extended

to metagenomic data. For targeted amplicon sequencing, the 16S rRNA gene is frequently used since it is present within the entire bacteria kingdom and contains highly variable regions (V1 - V9). These variable regions serve as unique markers to identify different types of bacteria and recover their phylogenetic relationships. Since it is assumed that each bacterial cell has roughly equal number of copies of this gene, the relative abundance of a particular bacterial taxon can be inferred by the proportion of its sequences.

The first step in 16S rRNA analysis is to sequence one or a few of its variable regions on the Illumina MiSeq or HiSeq platform. The raw sequencing data usually contains tens of millions, or even billions, of DNA fragments in a single run. A common software to process such raw sequencing data is QIIME (Caporaso et al., 2010), which integrates a number of third-party tools to facilitate the analysis. QIIME pipeline starts from demultiplexing and quality filtering the raw sequences in order to link back each sequence to the individual sample it comes from. After this step, QIIME uses a reference database, such as Greengenes (DeSantis et al., 2006), to cluster the sequences into operational taxonomic units (OTU), assign taxonomy to each OTU, and construct the phylogenetic tree. As the name suggests, OTU is merely an operational definition used to classify groups of closely related sequences. By default, QIIME clusters 16S rRNA sequences into OTUs using 97% as sequence similarity threshold. Each OTU cluster has a single representative sequence to be used for all subsequent analysis such as taxonomy assignment and phylogeny construction. The main advantage of OTU clustering is computational (Nguyen et al., 2016), since it can compresses millions of raw sequences into only thousands of OTU. It should be noted that results from OTU analysis can be affected by OTU picking strategies (closed reference vs open reference), sequence dissimilarity for alignment or software implementation. Let Ω be the set of OTUs and $K = |\Omega|$. The final output from OTU picking is a count vector $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iK})$ for ith sample to denote the number of sequences inside each of the K OTUs. In addition, every $\omega \in \Omega$ has an assigned taxonomy. It should be noted that 16S rRNA data, in most cases, only have enough resolution for taxonomy assignment down to the genus level in most

cases. QIIME will also output a phylogenetic tree, which can be used for diversity analysis or directly incorporated into the model, as shown in Chapter 2 .

1.3 Relative abundance and compositional data

In microbiome analysis, the main question of interest is to quantify the compositional relationship among bacterial taxa. This composition, otherwise called relative abundance, of K OTUs for ith sample is written as $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, ..., \pi_{iK})$ with $\sum_{j=1}^{K} \pi_{ij} = 1$ and $\pi_{ij} \geq 0$. The unit sum and non-negativity constraint of means that $\boldsymbol{\pi}_i$ lies on a simplex $\mathbb{S}^K$:

$$\mathbb{S}^K = \{(\pi_1, \pi_2, ..., \pi_K) : \pi_k \geq 0 \text{ for } k = 1, 2, ..., K \text{ and } \sum_{k=1}^{K} \pi_k = 1\}$$

As a result, most of the traditional analysis based on Euclidean geometry does not perfectly fit in this context. For example, the correlation of relative abundances of two taxa, even though their raw abundance are independent, is slightly negative due to the unit sum constraint. For a general review of compositional geometry and relevant transformations, we refer readers to the work of Aitchison (1986). In addition to being compositional, microbiome data are high-dimensional and highly sparse (Li, 2015). Components among the K dimensions are also correlated due to taxa relatedness. Jointly considering all of the aforementioned factors is known to produce vast challenges in statistical modeling. The most common choice to model microbial compositions is the Dirichlet-class distributions. In particular, a Dirichlet distribution defined on $\mathbb{S}^K$ with parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_K)$ admits the following density function:

$$f(\boldsymbol{\pi}) = \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \pi_k^{\alpha_k}, \quad \boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_K) \in \mathbb{S}^K$$

where $\Gamma(\cdot)$ is the gamma function. Dirichlet distributions have been to applied to microbiome data for testing equality of compositions (La Rosa et al., 2012) and clustering (Holmes et al.,

2012). In addition, it can be easily used for regression and variable selection, in which the question of interest is to identify the association of particular microbial taxa with a set of covariates such as dietary and life habits (Chen and Li, 2013; Wadsworth et al., 2017). In this case, the parameters $\boldsymbol{\alpha}$ are modeled as a linear function on the covariates, and l_1 penalizations are applied on the regression coefficients to conduct variable selection.

The simplest way to estimate $\boldsymbol{\pi}_i$ from sequencing data is to divide the OTU counts by their sum within each sample:

$$\hat{\boldsymbol{\pi}}_i = \boldsymbol{x}_i / N_i \tag{1.1}$$

where $N_i = \sum_{j=1}^{K} x_{ij}$ is called sequencing depth or library size. However, results produced from Illumina sequencing platform usually has N_i varying by orders of magnitude. This causes drastically different sampling variabilities for $\hat{\boldsymbol{\pi}}_i$, potentially producing spurious results for downstream analysis that are sensitive to rare taxa since they have higher chances to be observed at higher sequencing depth. Consequently, a common strategy is to shrink the sample sequences down to a common library size by sampling a subset of sequences from each $\boldsymbol{x}_i$. This additional step to even the sequencing depth across all samples is called subsampling or rarefaction. Although it brings the sampling variability of all compositions to a uniform level, the subsampling step discards useful data and is statistically inadmissible (McMurdie and Holmes, 2014). Therefore, it is advantageous to use model raw counts $\boldsymbol{x}_i$ without normalization.

Since raw counts are known to be over-dispersed when modeled by a simple multinomial distribution (La Rosa et al., 2012), various attempts have been made to jointly model the count and the underlying variability in microbial compositions. Examples include Dirichlet-multinomial, negative-binomial and logistic-multinomial distributions. We have introduced the Dirichlet distribution in the previous paragraph. The Dirichlet-multinomial can also be extended to incorporate the phylogenetic tree structure (Wang and Zhao, 2017). Negative binomial distributions have been extensively applied to differential abundance analysis for RNA-Seq data (Anders and Huber, 2010; Robinson et al., 2010). However, one shortcom-

ing of applying negative-binomial distribution on microbiome data is that it targets each taxa individually, making it impossible to take taxa relatedness or their correlation into account. Logistic-multinomial has been used by Xia et al. (2013) for regression and variable selection, but its main drawbacks are computational difficulties (since normal and binomial distributions are not conjugate) and the need to choose one particular taxon for the log-ratio transformation.

1.4 Community biodiversity analysis: alpha and beta diversities

The concept of alpha and beta diversities are proposed in Whittaker (1960) to quantify components of the total species diversity in a landscape. Specifically, alpha diversity measures species diversity in a single ecosystem, whereas beta diversity characterizes compositional dissimilarity between two different ecosystems. These species diversity metrics are widely used in ecological analysis since they provide simple and interpretable summary of the entire communities. As before, let $\boldsymbol{\pi}_i, \boldsymbol{\pi}_j \in \mathbb{S}^K$ be the taxa composition of ith and jth sample, respectively. An example of alpha diversity is the Shannon diversity index $H_i = -\sum_{k=1}^K \pi_{ik} \log \pi_{ik}$, which is equivalent to the Shannon entropy. An example of beta diversity is the Bray-Curtis dissimilarity $\mathrm{BC}_{ij} = \sum_{k=1}^K |\pi_{ik} - \pi_{jk}|/(\pi_{ik} + \pi_{jk})$. It should be noted that beta diversity needs not to satisfy the condition to be a metric.

The alpha and beta diversity can be defined using the phylogeny, namely the evolutionary history and relationships, among bacterial taxa. To see why this important, consider a toy example with three taxa: human, monkey and fish. If two given communities have major difference in human to monkey ratio but the sum of human and monkey abundances stay the same, then their difference should be considered less prominent than if they have major difference in human to fish ratio. This is because the evolutionary distance between mammals (including human and monkey) and fish is much higher than between human and monkey. Specifically, a phylogeny over K taxa can be represented as an evolutionary tree spanning the K taxa as leaves. We defer formal introduction of phylogenetic tree to Chapter 2.

QIIME will output the phylogenetic tree based on multiple sequence alignment using the representative sequences for each OTU. This phylogenetic tree is further used to produce phylogenetic biodiversity metrics such as Unifrac (Lozupone and Knight, 2005; Lozupone et al., 2007; Chen et al., 2012).

Biodiversity indexes can be incorporated into statistical analysis in various ways. Since alpha diversity provides a univariate summary of each community, it can be directly modeled as a response. Example studies of microbial alpha diversity are focused on obesity (Sze and Schloss, 2016), diet (Menni et al., 2017) and the usage of antibiotics (Wipperman et al., 2017). A particularly interesting area is to understand the development of gut bacteria in newborns within the first few years during which significant colonization from environmental bacteria occurs (Gritz and Bhandari, 2015). This colonization period consequently leads to an increasing alpha diversity of gut microbiome. The neonatal microbiome has been known to be sensitive to mode of delivery, antibiotics (Yang et al., 2016) and malnutrition (Subramanian et al., 2014). Abnormalities in the newborns' microbiome development can have a significant impact on the child's health.

Analysis of beta diversity are slightly more complicated since it requires an $n \times n$ matrix as input, where n is the sample size. More specifically, let $\boldsymbol{B}$ be such matrix where B_{ij} the beta diversity between ith and jth sample. By definition of beta diversity, $\boldsymbol{B}$ quantifies the pairwise sample dissimilarities in an ecologically meaningful way. The most common method to incorporate $\boldsymbol{B}$ is to apply a certain dimension reduction technique, such as principal coordinate analysis (PCoA) or non-metric multidimensional scaling (NMDS), to find the best low-dimensional representation of $\boldsymbol{B}$. The low-dimensional representation themselves are further used as response for downstream analysis. Since dimension reduction methods only finds an approximation to the underlying beta diversity, it can be advantageous to bypass this step and directly incorporate $\boldsymbol{B}$ into the statistical model. For example, Zhao et al. (2015) uses kernel regression where the kernel matrix is induced by the dissimilarity matrix $\boldsymbol{B}$. A complete discussion about the relation between kernel based methods and

distance based methods can be found in Sejdinovic et al. (2013). It should be noted that there is no guarantee that the induced kernel is positive semidefinite. Therefore, corrections must be invoked to prevent existence of negative eigenvalues in the kernel matrix. In Chapter 4, we prove that the square root transformation of weighted Unifrac has an exact embedding into the Euclidean and its induced kernel matrix is a positive semidefinite. Using such beta diversity avoids the nuisance of eigenvalue corrections and facilitates statistical modelling.

CHAPTER 2

A PHYLOGENETIC SCAN TEST ON DIRICHLET-TREE MULTINOMIAL MODEL FOR MICROBIOME DATA

2.1 Introduction

The vast improvement in microbiome sequencing tools in the past decade contrasts with the slower development of statistical methods to analyze microbiome data. Typically, the majority of taxa can be observed in only a very small subset of samples, which causes the data table to be highly sparse. In addition, the within-group heterogeneity among samples leads to pronounced overdispersion in taxa proportions. Since standard multinomial distributions fail at capturing these features, Dirichlet-multinomial (DM) has been used as a natural extension. DM was originally proposed by Mosimann (1962) and introduced into the microbiome context by La Rosa et al. (2012) and Holmes et al. (2012). Applying DM to test cross-group variation suffers from a number of drawbacks such as inability to localize any signal to a subgroup of taxa and reduced test power when a large number of taxa is present. Recent efforts to tackle these issues focus on incorporating phylogenetic tree into the model (Tang et al., 2017; Silverman et al., 2017; Wang and Zhao, 2017). In particular, Wang and Zhao (2017) applied an extension of DM, namely Dirichlet-tree multinomial (DTM), first proposed by Dennis III (1991) under the name hyper-Dirichlet type 1 distribution. DTM is based on a decomposition of the sample space through a cascade of nested partitions similar to a Polya tree process (Lavine, 1992). Instead of placing a single global DM on all taxa, DTM consists of a collection of independent local DMs, each corresponding to a particular internal node on the phylogenetic tree. Since descendants of each internal node on the phylogenetic tree share a certain degree of evolutionary affinity, such decomposition strategy allows one to assign meaningful interpretation to each of the local DM distributions. An additional benefit is that the local DMs target only particular groups of taxa and consequently enjoy much lower degrees of freedom. This breaking down of the global distribution on all taxa

counts allows testing each branch of the phylogeny individually, hence locating the signals to a certain taxonomic rank. The global cross-group test is therefore represented by a number of independent and biologically relevant constituents. For more general application of the Polya tree decomposition to hypothesis testing, see Ma and Wong (2011), Chen and Hanson (2014), Holmes et al. (2015) and Soriano and Ma (2017).

Although standard multiple testing procedures could be applied to results from testing all nodes, it is usually not the best practice to treat each hypothesis as a segregated entity. Soriano and Ma (2017) pointed out that cross-group distributional variations tend to cluster, which causes hypotheses defined on nearby and/or nested windows more likely to be jointly true or false. This observation also holds in the microbiome data; cross-group differences in a certain ancestor node are more frequently accompanied with similar differences in its descendants. To take advantage of this structure and optimize test power, we adopt ideas from scan tests through constructing a collection of triplet statistics, each incorporating evidence from an internal node on the phylogenetic tree along with its parent and one of its children.

The maximum of all these triplet statistics is used to test the global null hypothesis. Since the exact distribution of maximum statistic is intractable, we derive an upper and lower bound on its tail probability based on existing results on union probability (Hunter, 1976; Efron, 1997; Taylor et al., 2007). Our improved strategy first finds a subset consisting of independent components from the union, followed by bounding the probability of remaining components conditioned on the complement of that subset. A decay rate of the relative error of our approximation is also provided.

Section 2.2 briefly reviews the DM model. Section 2.3 formulates the DTM model and establishes its relation to the DM. Section 2.4 develops p-value approximation on the scan statistic for the DTM and verifies the result through simulation. Section 2.5 applies the DTM model on the American Gut dataset to test the association of gut microbiome with a number of dietary habits. It also empirically demonstrates improvement of DTM over DM

through likelihood ratio tests and comparing simulated test power. Section 2.6 concludes with further discussions on potential DTM extensions. Section 2.7 presents theorem proofs.

2.2 Dirichlet-multinomial for microbiome data

In this section we briefly recap the cross-group testing procedures on microbiome data using the Dirichlet-multinomial model, as presented in La Rosa et al. (2012). Consider a microbiome dataset with n samples and let Ω be the collection of a total of $K = |\Omega|$ OTUs. Without loss of generality, we assume $\Omega = \{1, 2, ..., K\}$. Each sample is a K-dimensional count vector representing the number of sequences in each of the K OTUs. Let $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{iK})$ be the taxa count vector of the ith sample for $i = 1, 2, ..., n$. In addition, define $N_{i.} = \sum_{j=1}^{K} x_{ij}$ to be the total number of sequences in the ith sample, $N_{.j} = \sum_{i=1}^{n} x_{ij}$ to be the total number of sequences in the jth OTU, and $N_{..} = \sum_{i=1}^{n} N_{i.} = \sum_{j=1}^{K} N_{.j}$. The Dirichlet-multinomial (DM) model assumes

$$\boldsymbol{q}_i \overset{i.i.d.}{\sim} \text{Dir}(\nu\boldsymbol{\pi}), \quad \boldsymbol{x}_i|\boldsymbol{q}_i \sim \text{Multinomial}(N_{i.}, \boldsymbol{q}_i),$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_K)$ satisfies $\sum_{j=1}^{K} \pi_j = 1$, $\pi_j > 0$ denotes the mean taxa proportions and $\nu > 0$ is a dispersion parameter that controls the level of variation across samples. Alternatively one may use $\theta = \frac{1}{1+\nu}$ to parametrize the dispersion so that $0 \leq \theta < 1$. Integrating out the $\boldsymbol{q}_i$ gives

$$f(\boldsymbol{x}_i) = \binom{N_{i.}}{\boldsymbol{x}_i} \frac{\Gamma(\nu)}{\Gamma(N_{i.} + \nu)} \prod_{j=1}^{K} \frac{\Gamma(x_{ij} + \nu\pi_j)}{\Gamma(\nu\pi_j)}. \tag{2.1}$$

Throughout this paper we use $f(\cdot)$ exclusively to denote the DM probability mass function. When $\nu = \infty$ ($\theta = 0$), the DM degenerates to the standard multinomial distribution. Smaller values of ν indicates larger degrees of overdispersion. Assuming $\boldsymbol{x}_i$'s are independent, the likelihood function is simply the product of probabilities over all samples:

$$\mathcal{L}(\boldsymbol{\pi},\nu)=\prod_{i=1}^{n}\left[\binom{N_{i.}}{\boldsymbol{x}_i}\frac{\Gamma(\nu)}{\Gamma(N_{i.}+\nu)}\prod_{j=1}^{K}\frac{\Gamma(x_{ij}+\nu\pi_j)}{\Gamma(\nu\pi_j)}\right] \tag{2.2}$$

As is shown in Weir and Hill (2002), the method of moments (MoM) estimates of the mean proportion $\boldsymbol{\pi}$ and dispersion θ are respectively

$$\hat{\boldsymbol{\pi}}=(\hat{\pi}_1,\hat{\pi}_2,...,\hat{\pi}_K) \text{ with } \hat{\pi}_j=N_{.j}/N_{..}$$

$$\hat{\theta}=\frac{\sum_{j=1}^{K}(S_j-G_j)}{\sum_{j=1}^{K}\big(S_j+(N_c-1)G_j\big)},$$

where we have $N_c=\frac{N_{..}-(N_{..})^{-1}\sum_{i=1}^{n}N_{i.}^2}{n-1}$, $S_j=\frac{\sum_{i=1}^{n}N_{i.}(\hat{\pi}_{ij}-\hat{\pi}_j)^2}{n-1}$, and $G_j=\frac{\sum_{i=1}^{n}N_{i.}\hat{\pi}_{ij}(1-\hat{\pi}_{ij})}{\sum_{i=1}^{n}(N_{i.}-1)}$ with $\hat{\pi}_{ij}=x_{ij}/N_{i.}$.

For hypothesis testing, suppose we collect G groups and the gth group data is given by $\boldsymbol{x}_1^{(g)},\boldsymbol{x}_2^{(g)},...,\boldsymbol{x}_{n_g}^{(g)}$ with $N_{i.}^{(g)}=\sum_{j=1}^{K}x_{ij}^{(g)}$ and $N_{..}^{(g)}=\sum_{i=1}^{n}N_{i,}^{(g)}$. Similarly we define the gth group parameters as $\boldsymbol{\pi}^{(g)},\nu^{(g)}$ with $\theta^{(g)}=\frac{1}{1+\nu^{(g)}}$. We wish to test the equality of mean proportion across all groups:

$$H_0:\boldsymbol{\pi}^{(1)}=\boldsymbol{\pi}^{(2)}=...=\boldsymbol{\pi}^{(G)} \text{ vs } H_a: \text{otherwise}$$

Let $\hat{\boldsymbol{\pi}}^{(g)}$ and $\hat{\theta}^{(g)}$ be the MoM estimates of $\boldsymbol{\pi}^{(g)}$ and $\theta^{(g)}$, respectively. The cross-group pooled estimate of $\boldsymbol{\pi}$ is $\hat{\boldsymbol{\pi}}^{(Pool)}=\sum_{g=1}^{G}\bar{s}_g\hat{\boldsymbol{\pi}}^{(g)}$ with:

$$\bar{s}_g=\frac{(N_{..}^{(g)})^2C(\hat{\theta}^{(g)},N_{..}^{(g)})^{-1}}{\sum_{r=1}^{G}(N_{..}^{(r)})^2C(\hat{\theta}^{(r)},N_{..}^{(r)})^{-1}},$$

where

$$C(\hat{\theta}^{(g)},N_{..}^{(g)})=\hat{\theta}^{(g)}\Big(\sum_{i=1}^{n_g}(N_{i.}^{(g)})^2-N_{..}^{(g)}\Big)+N_{..}^{(g)}.$$

Finally, the test statistic is defined as

$$T = \sum_{g=1}^{G} (\hat{\boldsymbol{\pi}}^{(g)} - \hat{\boldsymbol{\pi}}^{(Pool)})^T (\bar{S}_g)^{-1} (\hat{\boldsymbol{\pi}}^{(g)} - \hat{\boldsymbol{\pi}}^{(Pool)}), \tag{2.3}$$

where $\bar{S}_g$ is a diagonal matrix given by

$$\bar{S}_g = \left((N_{\cdot\cdot}^{(g)})^2 C(\hat{\theta}_g, N_{\cdot\cdot}^{(g)})^{-1} \right)^{-1} D(\hat{\boldsymbol{\pi}}^{(Pool)}),$$

and $D(\hat{\boldsymbol{\pi}}^{(Pool)})$ is also diagonal with diagonal elements given by $\hat{\boldsymbol{\pi}}^{(Pool)}$. The asymptotic distribution of T under H_0 is $\chi^2_{(K-1)(G-1)}$ as $n_g \to \infty$ for all g.

2.3 Dirichlet-tree multinomial and hypothesis testing

To incorporate the phylogenetic tree into the model, Wang and Zhao (2017) considered an extension named Dirichlet-tree multinomial (DTM). DTM allows us to separately test cross-group differences in each internal node, locating the source of overall difference within particular subgroups of OTUs. Each of the local test, by design, has the benefit of reduced degrees of freedom.

2.3.1 Model formulation

Let $\mathcal{T} = (\Omega, \mathcal{I})$ be a rooted phylogenetic tree where the set of OTUs Ω are placed on the leaves and $\mathcal{I}$ is the set of all internal nodes. We represent the elements in $\mathcal{I}$ to be subsets of Ω since each internal node is uniquely identified by the subset of all OTUs underneath it, and vice versa. Each subset of OTU that corresponds to an internal node shares a hypothetical ancestor along the lineage. Additionally, each leaf node is uniquely identified by a singleton set consisting of that particular OTU.

Figure 2.1 shows an example of a simple phylogenetic tree over 5 OTUs and 4 internal nodes. This tree has $\Omega = \{1, 2, 3, 4, 5\}$ and $\mathcal{I} = \big\{\{1, 2, 3, 4, 5\}, \{1, 2, 3\}, \{4, 5\}, \{2, 3\}\big\}$.

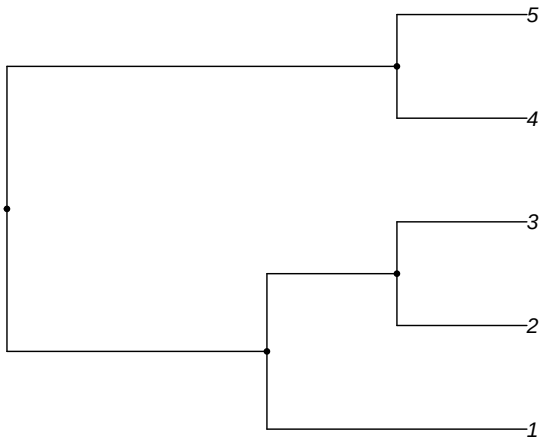

Figure 2.1: An example of a phylogenetic tree with five OTUs.

Now for $\forall A \in \mathcal{I}$, let $\mathcal{C}(A)$ be the collection of A's child nodes in $\mathcal{T}$. The elements of $\mathcal{C}(A)$ are also subsets of Ω. Also $\forall A \in \mathcal{I} \cup \big\{\{\omega\} | \omega \in \Omega\big\}, A \neq \Omega$, let $R(A)$ denote the parent node of A. In Figure 1, for example, $\mathcal{C}(\{1,2,3\}) = \big\{\{1\},\{2,3\}\big\}$ and $R(\{1,2,3\}) = \{1,2,3,4,5\} = \Omega$. Notice that certain $\mathcal{C}(A)$'s contain singletons of Ω since some children are leaves. Let $k(A) = |\mathcal{C}(A)|$ be the number of children under A and write $\mathcal{C}(A) = \{\mathcal{C}(A)_1, \mathcal{C}(A)_2, ..., \mathcal{C}(A)_{k(A)}\}$. For each $i = 1, 2, ..., n$ and $j = 1, 2, ..., k(A)$, let

$$x_{ij}(A) = \sum_{\omega \in \mathcal{C}(A)_j} x_{i\omega}$$

be the count of the jth child of A in the ith sample. The count vector associated with A is therefore

$$\boldsymbol{x}_i(A) = \big(x_{i1}(A), x_{i2}(A), ..., x_{ik(A)}(A)\big)$$

with the sum $N_i(A) = \sum_{j=1}^{k(A)} x_{ij}(A) = \sum_{\omega \in A} x_{i\omega}$. It is straightforward to see that $N_i\big(\mathcal{C}(A)_j\big) = x_{ij}(A)$ for $j = 1, 2, ..., k(A)$. In addition, we always have $N_i(\Omega) = N_{i.}$.

The DTM distribution separately models the count vector $\boldsymbol{x}_i(A)$ conditional on $N_i(A)$

for each A. Specifically for $\forall A \in \mathcal{I}$,

$$\boldsymbol{q}_{A,i} \overset{i.i.d.}{\sim} \text{Dir}(\nu_A \boldsymbol{\pi_A}), \quad \boldsymbol{x}_i(A)|N_i(A), \boldsymbol{q}_{A,i} \sim \text{Multinomial}(N_i(A), \boldsymbol{q}_{A,i}) \tag{2.4}$$

where $\nu_A > 0$ is the overdispersion parameter of the counts of A's children and $\boldsymbol{\pi}_A = (\pi_{A,1}, \pi_{A,2}, ..., \pi_{A,k(A)})$ satisfying $\sum_{i=1}^{k(A)} \pi_{A,i} = 1$ denotes their mean proportion. The Dirichlet prior distribution of all A's are mutually independent. Integrating out $q_{A,i}$ gives

$$f\big(\boldsymbol{x}_i(A)|N_i(A)\big) = \binom{N_i(A)}{\boldsymbol{x}_i(A)} \frac{\Gamma(\nu_A)}{\Gamma(N_i(A)+\nu_A)} \prod_{j=1}^{k(A)} \frac{\Gamma(x_{ij}(A)+\nu_A \pi_{A,j})}{\Gamma(\nu_A \pi_{A,j})}, \tag{2.5}$$

which ultimately yields

$$f_T(\boldsymbol{x}_i) = \prod_{A \in \mathcal{I}} f\big(\boldsymbol{x}_i(A)|N_i(A)\big) \tag{2.6}$$

and likelihood function

$$\mathcal{L}_T(\{(\nu_A, \boldsymbol{\pi}_A) : A \in \mathcal{I}\}) = \prod_{i=1}^{n} \prod_{A \in \mathcal{I}} f\big(\boldsymbol{x}_i(A)|N_i(A)\big) \tag{2.7}$$

with $f_T(\cdot)$ and $\mathcal{L}_T(\cdot)$ denoting the DTM probability mass function and likelihood function respectively. The representations in (2.5) and (2.6) naturally lead to a top-down generative scheme of the count data on the nodes, as each layer of DM models a subset of OTU counts at increased level of resolution conditioned on their sum.

Interestingly, Dennis III (1991) showed that the global DM distribution on OTU counts is nested in the DTM family. A simple explanation of this relation is that both the global Dirichlet prior and the multinomial probabilities can be factorized over $\mathcal{I}$, i.e.

$$\boldsymbol{q}_i \sim \text{Dir}(\nu \boldsymbol{\pi}) \Leftrightarrow \forall A \in \mathcal{I}, \quad \frac{\boldsymbol{q}_i(A)}{\sum_{\omega \in A} q_{i\omega}} \sim \text{Dir}\Big(\nu \sum_{\omega \in A} \pi_\omega \cdot \frac{\boldsymbol{\pi}(A)}{\sum_{\omega \in A} \pi_\omega}\Big) \text{ independently}$$

$$\boldsymbol{x}_i|\boldsymbol{q}_i \sim \text{Multinomial}(N_{i\cdot}, \boldsymbol{q}_i) \Leftrightarrow$$

$$\forall A \in \mathcal{I}, \quad \boldsymbol{x}_i(A)|\boldsymbol{q}_i, N_i(A) \sim \text{Multinomial}(N_i(A), \frac{\boldsymbol{q}_i(A)}{\sum\limits_{\omega \in A} q_{i\omega}})$$

where we similarly defined

$$q_{ij}(A) = \sum_{\omega \in \mathcal{C}(A)_j} q_{i\omega}, \boldsymbol{q}_i(A) = (q_{i1}(A), q_{i2}(A), ..., q_{ik(A)}(A))$$

$$\pi_j(A) = \sum_{\omega \in \mathcal{C}(A)_j} \pi_\omega, \boldsymbol{\pi}(A) = (\pi_1(A), \pi_2(A), ..., \pi_{k(A)}(A))$$

In the DTM representation of global DM, the overdispersion and mean proportion of the counts of A's children are respectively $\nu_A = \nu \sum_{\omega \in A} \pi_\omega$ and $\boldsymbol{\pi}_A = \boldsymbol{\pi}(A)/\sum_{\omega \in A} \pi_\omega$. It is not hard to notice that there is a bijective correspondence between $\boldsymbol{\pi}$ and $\{\boldsymbol{\pi}_A = \boldsymbol{\pi}(A)/\sum_{\omega \in A} \pi_\omega : A \in \mathcal{I}\}$. In addition, all of these local dispersions are governed by a single global ν, which is highly restrictive as it does not allow any node-specific characterization of within-group variation. Section 2.5.2 provides likelihood ratio test results supporting this claim.

2.3.2 Hypothesis testing

The DTM model in (2.6) and (2.7) motivates a node-by-node testing strategy for cross-group comparison. To compare the proportion across G groups of observations, we carry out an MoM test using (2.3) individually for each A, i.e.

$$H_{0,A} : \boldsymbol{\pi}_A^{(1)} = \boldsymbol{\pi}_A^{(2)} = ... = \boldsymbol{\pi}_A^{(G)} \text{ vs } H_{a,A} : \text{otherwise}$$

Each of the MoM test statistic are calculated conditional on $\{N_i^{(g)}(A)|1 \leq g \leq G, 1 \leq i \leq n_g\}$, where $N_i^{(g)}(A)$ is the sum of OTU counts under A in the ith sample of gth group. The

test statistic for $H_{0,A}$ has degrees of freedom $(G-1)(k(A)-1)$, much smaller than the degrees of freedom for DM test as $(G-1)(K-1)$. The local DM test is therefore more powerful than the global DM test, provided that the extent of cross-group difference on the internal nodes is not diluted too much as we group multiple OTUs together. Obviously, the extent of dilution is largely determined by the tree structure. The ideal scenario is that OTUs placed under the same internal node A demonstrate increasing or decreasing abundance simultaneously for all samples in a certain group, so $H_{0,R(A)}$ will be most effective. This also motivates using the phylogenetic tree to carry out the decomposition, as functionally similar OTUs tend to exhibit similar abundance changes within the same group.

The mean proportion of all OTUs across G groups are equal if and only if $H_{0,A}$ is true for all $A \in \mathcal{I}$. Therefore, we define the global null as $H_0 = \cap_{A\in\mathcal{I}} H_{0,A}$. Controlling the Type-I error on the global null is simply equivalent to controlling the family-wise error rate (FWER) across $H_{0,A}$'s.

The following theorem makes controlling FWER straightforward:

Theorem 1. *Let p_A be the MoM p-value for testing $H_{0,A}$. Under the global null $H_0 = \cap_{A\in\mathcal{I}} H_{0,A}$, p_A's are asymptotically mutually independent as the number of subjects in each group goes to infinity.*

Section 2.7 has a proof of this theorem.

The independence of p-value under the null grants one of the following procedures to control the exact FWER at level α: (i) Sidak's procedure, in which one assigns equal Type I error $\alpha(A) = 1 - (1-\alpha)^{1/\mathcal{I}}$ to all A's (ii) allocate α_A according to the tree structure while constraining $1 - \prod_{A\in\mathcal{I}}(1-\alpha_A) = \alpha$. After choosing the individual Type I error thresholds, one can report the collection of nodes $\{A : p_A < \alpha_A, A \in \mathcal{I}\}$ as being significant.

2.4 PhyloScan: scan statistic over the tree tuples

Cross-group difference in distributions of taxa counts often occurs in clusters or chains on the phylogenetic tree. If one internal node exhibits significant difference in relative proportion across several groups, then this is often associated with signals from at least one of its children or parent. Figure 2.4 shows four examples of signal clusters on American Gut data using the top 100 OTUs with the highest counts. In each graph, subjects are divided into two groups according to different ingestion frequencies in one of the following diets: milk and cheese, seafood, sugary sweets and vegetable. (details in Section 2.5.1). The size of the circle on internal node A is proportional to -log(p_A) from the cross-group comparison (the circle colors are irrelevant here). It is apparent that large circles tend to form in chaining patterns, which motivates scanning for signals in chains or clusters instead of on each node separately. Moreover the partitioning nature of the phylogenetic tree always leads to much smaller sample size on the bottom nodes (farthest from the root placed on top). Sharing information across nodes would alleviate the limitation to detect distributional differences on the bottom level.

Without prior knowledge of the length and shape of signal clusters, we focus on only triplets formulated by a certain internal node, its parent and one of its children. Each triplet has its own statistic defined as the sum of all the node statistics within, pooling signal strength from its members. The maximum of these statistics on all the triplets is then used to test the global null hypothesis. Our method belongs to the class of scan statistics (Glaz et al., 2001), in which one searches for signals over varying sizes of windows. In our case, each window denotes a particular branch of the phylogenetic tree. The shape of our designed triplet reflects our knowledge of correlated signals on the tree, while the size of the triplet achieves a compromise between signal pooling around neighboring nodes and the ability to detect alternatives in short chains. Since the exact distribution of the maximum statistic is unknown, we design a novel method to calculate the upper and lower bound of its tail probability using low dimensional integrals that can be efficiently evaluated through

standard numerical integration techniques. Since this entire hypothesis testing procedure is established on the phylogenetic tree decomposition, we call it PhyloScan.

2.4.1 Overview

For each $A \in \mathcal{I}$ such that $R(A) \in \mathcal{I}$ and $\mathcal{C}(A) \cap \mathcal{I} \neq \emptyset$, we define a triplet to be the set of three consecutive internal nodes $\{A, R(A), \mathcal{C}(A)_i\}$ where $i \in \{1, 2, ..., k(A)\}$ satisfies $\mathcal{C}(A)_i \in \mathcal{I}$. Let $\mathcal{B}$ be the set of all such triplets, and without loss of generality we write $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, ..., \mathcal{B}_b\}$ where each $\mathcal{B}_i$ is a triplet and $b = |\mathcal{B}|$ depends on both K and the structure of the tree. We assume the ordering of elements in $\mathcal{B}$ obeys the following rule: $\{A, R(A), \mathcal{C}(A)_i\}$ always has a smaller index than (or appear in front of) $\{\tilde{A}, R(\tilde{A}), \mathcal{C}(\tilde{A})_j\}$ if $\tilde{A} \subset A$. Now we proceed to define the test statistic for $\mathcal{B}_i$ as follows. First, each of the p-values on the internal nodes can be inverted to a chi-square random variable with 1 degree of freedom, namely

$$Z_A = F_1^{-1}(p_A) \text{ for all } A \in \mathcal{I},$$

where F_j denotes the cumulative distribution function (CDF) of χ_j^2 distribution. Theorem 4 states that under the global null H_0, Z_A's are asymptotically mutually independent. In order to test the following hypothesis on each triplet

$$H_{0,\mathcal{B}_i} = \bigcap_{A \in \mathcal{B}_i} H_{0,A} \text{ vs } H_{a,\mathcal{B}_i} : \text{otherwise},$$

we define the statistic to be the sum of Z_A's within:

$$W_i = \sum_{A \in \mathcal{B}_i} Z_A \text{ for } i = 1, 2, ..., b. \tag{2.8}$$

It is apparent that each $W_i \sim \chi_3^2$ under $H_{0,\mathcal{B}_i}$. For the global null hypothesis $H_0 =$

$\cap_{A\in\mathcal{I}} H_{0,A} = \cap_{i=1}^{b} H_{0,\mathcal{B}_i}$, we use the maximum of W_i's as the test statistic:

$$W = \max_{1\le i\le b} W_i. \tag{2.9}$$

Since $\mathcal{B}_i$'s overlaps with each other, W_i's are heavily correlated and the exact distribution of W is hard to derive. For testing purposes, it suffices to calculate the tail probability of W. Suppose our observed value of the maximum statistic is w, and let $B_i(w) = \{W_i > w\}$ be the event of ith triplet statistic exceeding w . Without incurring any confusion, we may drop w and simply write B_i. We are mainly interested in the global p-value $P(\bigcup_{i=1}^{b} B_i)$, which boils down to the problem of bounding the union probability.

The simplest upper bound of the union probability is the Bonferroni inequality:

$$P(\bigcup_{i=1}^{b} B_i) \le \sum_{i=1}^{b} P(B_i)$$

Several authors have provided sharper bounds over the Bonferroni inequality in the past few decades. The results in Hunter (1976), Worsley (1982) and Efron (1997) suggest the following improvement:

$$\begin{aligned} P(\bigcup_{i=1}^{b} B_i) &= P(B_1) + P(B_2 \cap B_1^c) + P(B_3 \cap B_1^c \cap B_2^c) + ... \\ &\le P(B_1) + \sum_{i=2}^{b} \min_{j<i} P(B_i \cap B_j^c) \end{aligned} \tag{2.10}$$

In particular, $\min_{j<i} P(B_i \cap B_j^c)$ is achieved at $j = i-1$ when the neighboring variables (W_{i-1}, W_i) have the highest pairwise correlation. Each of the term inside summation can be easily evaluated by numerical integration. It can be easily generalized to the union of more than two sets to improve approximation. More generally, the above inequality belongs

to the class of approximations with the following representation:

$$P(\bigcup_{i=1}^{b} B_i) \leq \sum_{J \in S} (-1)^{|J|-1} f(J) P(\bigcap_{j \in J} B_j) \tag{2.11}$$

where $f(J)$ is some non-negative function on subset of $S = \{1, 2, ..., b\}$. Naiman and Wynn (1992) and Naiman et al. (1997) gave results regarding when (2.11) achieves equality. Following their work, Dohmen (2000) and Dohmen (2002) gave further improvement on the Bonferroni inequalities. There is also research on Bonferroni inequalities for particular applications, such as Dohmen and Tittmann (2004) on partition lattice and Taylor et al. (2007) on maxima over Gaussian random fields.

2.4.2 Bounding the union probability

Our upper bound of the union probability involves a decomposition of $\bigcup_{i=1}^{b} B_i$ into (i) a union of independent events and (ii) their complement in $\bigcup_{i=1}^{b} B_i$. The probability of the union of independent events can be exactly evaluated, while a similar strategy to (2.10) is applied to estimate its complement.

Specifically, let $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_m\}$ be a class of disjoint nonempty subsets of $\mathcal{I}$ satisfying $\forall i \leq m, \exists j \leq b$ s.t $\mathcal{M}_i \subset \mathcal{B}_j$. For each i, define $M_i = \{\sum_{A \in \mathcal{M}_i} Z_A > w\}$ to be the exceeding event on $\mathcal{M}_i$. It follows that $\forall i \leq m, \exists j \leq b$ s.t $M_i \subset B_j$. Write $M = \bigcup_{i=1}^{m} M_i$. This leads to

$$P(\bigcup_{i=1}^{b} B_i) = P(M) + P(M^c) \cdot P(\bigcup_{i=1}^{b} B_i | M^c) \tag{2.12}$$

since $M \subset \bigcup_{i=1}^{b} B_i$. The independence of M_i's leads to a straightforward calculation of $P(M)$ as $P(M) = 1 - F_1(w)^{t_1} F_2(w)^{t_2} F_3(w)^{t_3}$ where $t_l = |\{i \leq m : |\mathcal{M}_i| = l\}|$ and $F_i(\cdot)$ is the CDF of χ^2_i distribution. Next, we approximate $P(\bigcup_{i=1}^{b} B_i | M^c)$ using a similar strategy to (2.10). It is apparent that enlarging M will always decrease $P(\bigcup_{i=1}^{b} B_i \cap M^c)$ and most likely the error of its upper bound, which makes our strategy superior to directly applying

(2.10) to the B_i's.

The next question is how to choose an M as large as possible. An obvious optimality condition is that $\bigcup_{i=1}^{m} \mathcal{M}_i = \mathcal{I}$, because otherwise we can always enlarge M by $\{Z_A > w\}$ for a certain $A \in \mathcal{I} \setminus \bigcup_{i=1}^{m} \mathcal{M}_i$. Moreover, the elements in $\mathcal{M}$ should not be able to combine together and still belong to a certain element in $\mathcal{B}$, i.e. $\forall i_1, i_2 \leq m, \nexists j \leq b$ s.t. $\mathcal{M}_{i_1} \cup \mathcal{M}_{i_2} \subset \mathcal{B}_j$. This is because merging $\mathcal{M}_{i_1}$ and $\mathcal{M}_{i_2}$ enlarges M, i.e. $M_{i_1} \cup M_{i_2} \subset \{\sum_{A \in \mathcal{M}_{i_1} \cup \mathcal{M}_{i_2}} Z_A > w\}$. Since an exhaustive search over all combinations is computationally infeasible for large trees, we propose the following greedy algorithm that satisfies these optimality conditions:

(a) Order the elements in $\mathcal{I}$ as $A_1, A_2, ..., A_{|\mathcal{I}|}$ such that each internal node always appears in front of its children.

(b) Set $\mathcal{M} = \emptyset$ and $i = 1$.

(c) For each $i = 1, 2, ..., |\mathcal{I}|$, sequentially go through the following steps:

  (i) If $\exists j \leq m$ s.t. $A_i \in \mathcal{M}_j$, set $i \leftarrow i + 1$ and go back to the beginning of step (c).

  (ii) If $\exists j_1, j_2$ s.t. $\mathcal{C}(A_i)_{j_1} \in \mathcal{I}$ and $\mathcal{C}\big(\mathcal{C}(A_i)_{j_1}\big)_{j_2} \in \mathcal{I}$, set $\mathcal{M} \leftarrow \mathcal{M} \cup \{A_i, \mathcal{C}(A_i)_{j_1}, \mathcal{C}\big(\mathcal{C}(A_i)_{j_1}\big)_{j_2}\}$ and $i \leftarrow i + 1$. Go back to the beginning of step (c).

  (iii) If $\exists j_1$ s.t. $\mathcal{C}(A_i)_{j_1} \in \mathcal{I}$, set $\mathcal{M} \leftarrow \mathcal{M} \cup \{A_i, \mathcal{C}(A_i)_{j_1}\}$ and $i \leftarrow i + 1$. Go back to the beginning of step (c).

  (iv) Otherwise, set $\mathcal{M} \leftarrow \mathcal{M} \cup \{A_i\}$ and $i \leftarrow i + 1$. Go back to the beginning of step (c).

The above greedy algorithm seeks to incorporate the longest chain (with maximum of 3 nodes) starting from A_i and use its descendants as subsequent nodes, if A_i has not been included in $\mathcal{M}$ so far. Since the parent node is always considered ahead of its children, the resulting $\mathcal{M}$ will always satisfy the two aforementioned optimality conditions. As the algorithm prioritizes longer chains at each step, it effectively produces a large M that yields

relatively accurate estimates of the union probability for our applications (numerical results to be shown later).

Figure 2.2 shows an example of $\mathcal{M}$ on a simple phylogenetic tree with $K = 13$ OTUs. Each internal node in $\mathcal{M}_i$ is assigned the same number i for $i = 1, 2, ..., 7$.

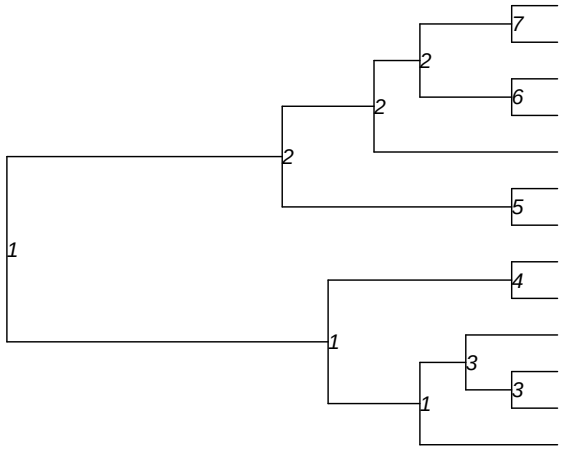

Figure 2.2: Example configuration of $\mathcal{M}$ using the greedy algorithm.

The remaining task is to put an upper bound on $P(\bigcup_{i=1}^{b} B_i|M^c)$. For each $i \leq b$, let $\mathcal{N}_i = \{j : |\mathcal{B}_j \cap \mathcal{B}_i| = 2 \text{ and } j < i\}$. Apparently $|\mathcal{N}_i| \leq 2$ for all i because of the ordering of $\mathcal{B}_i$'s. Write $B_{\mathcal{N}_i} = \bigcup_{j\in\mathcal{N}_i} B_j$ for short. Now we proceed as follows:

$$P(\bigcup_{i=1}^{b} B_i|M^c) \leq \sum_{i=1}^{b} P(B_i \cap B_{\mathcal{N}_i}^c|M^c) \tag{2.13}$$

The equation in (2.13) is very similar to (2.10) in that for each B_i, it includes only the highest correlated events $B_{\mathcal{N}_i}$, which will minimize the right side of the equation. It is worth noting that $P(B_i|M^c) = 0$ if $\mathcal{B}_i \in \mathcal{M}$, hence the strategy of prioritizing triplets while constructing $\mathcal{M}$. To efficiently evaluate each of the terms in the right side of (2.13), notice that the distributions of Z_A conditioned on M^c are the same as the product of a truncated chi-square and an independent Dirichlet random variable, so their density function can be expressed using chi-square CDFs. Let $f_i(\cdot)$ and $F_i(\cdot)$ denote the density and CDF of χ_i^2

distribution respectively, then the marginal density of Z_A conditional on M^c becomes

$$f_A(z|M^c) = \begin{cases} \frac{f_1(z)}{F_1(w)}, \text{ if } |\mathcal{M}(A)| = 1 \wedge w \leq z \\ \frac{F_1(w-z)f_1(z)}{F_2(w)}, \text{ if } |\mathcal{M}(A)| = 2 \wedge w \leq z \\ \frac{F_2(w-z)f_1(z)}{F_3(w)}, \text{ if } |\mathcal{M}(A)| = 3 \wedge w \leq z \\ 0, \text{ otherwise} \end{cases} \tag{2.14}$$

where we define $\mathcal{M}(A) = \mathcal{M}_i$ if $A \in \mathcal{M}_i$. The existence and uniqueness of $\mathcal{M}(A)$ is guaranteed by the fact that $\{\mathcal{M}_1, \mathcal{M}_2, ..., \mathcal{M}_m\}$ are disjoint with $\bigcup_{i=1}^m \mathcal{M}_i = \mathcal{I}$.

The joint density of Z_{A_1} and Z_{A_2} for $\forall A_1, A_2 \in \mathcal{I}$ is

$$f_{A_1,A_2}(z_1, z_2|M^c) = \begin{cases} f_{A_1}(z_1|M^c) f_{A_2}(z_2|M^c), \text{ if } \mathcal{M}(A_1) \cap \mathcal{M}(A_2) = \emptyset \\ \frac{\prod_{i=1}^2 f_1(z_i)}{F_2(w)}, \text{ if } \mathcal{M}(A_1) = \mathcal{M}(A_2) \wedge |\mathcal{M}(A_1)| = 2 \\ \qquad \wedge \sum_{i=1}^2 z_i \leq w \\ \frac{F_1(w - \sum_{i=1}^2 z_i) \prod_{i=1}^2 f_1(z_i)}{F_3(w)}, \text{ if } \mathcal{M}(A_1) = \mathcal{M}(A_2) \\ \qquad \wedge |\mathcal{M}(A_1)| = 3 \wedge \sum_{i=1}^2 z_i \leq w \\ 0, \text{ otherwise} \end{cases} \tag{2.15}$$

Given w, we pre-calculate the density functions in (2.14) and (2.15), and the CDFs of Z_A using (2.14) and of $Z_{A_1} + Z_{A_2}$ using (2.15) up to a certain precision and store them into the memory. This turns each term in the right side of (2.13) into at most two-dimensional integrals. We evaluate these integrals using the functions `cuhre` and `suave` in R package `R2Cuba` (Hahn, 2005).

Substituting (2.13) into (2.12) gives

$$P_0 = P(\bigcup_{i=1}^{b} B_i) \leq P_U = P(M) + P(M^c) \cdot \sum_{i=1}^{b} P(B_i \cap B^c_{\mathcal{N}_i} | M^c), \tag{2.16}$$

where P_0 is the actual p-value and P_U is its upper bound. Let $\epsilon_U = P_U - P_0$ be the error of our approximation. Using a similar strategy to Theorem A1 in Taylor et al. (2007), it follows that

$$\begin{aligned}
\epsilon_U &= P(M^c) \sum_{i=1}^{b} \big(P(B_i \cap B^c_{\mathcal{N}_i} | M^c) - P(B_i \cap B^c_{i-1} \cap B^c_{i-2} \cap ... B^c_1 | M^c) \big) \\
&= P(M^c) \sum_{i=1}^{b} P(B_i \cap B^c_{\mathcal{N}_i} \cap \big(\bigcup_{j<i, j \notin \mathcal{N}_i} B_j \big) | M^c) \\
&\leq P(M^c) \sum_{i=1}^{b} P(\bigcup_{j<i, j \notin \mathcal{N}_i} (B_i \cap B_j) | M^c) \\
&\leq P(M^c) \sum_{i=1}^{b} \sum_{j<i, j \notin \mathcal{N}_i} P(B_i \cap B_j | M^c)
\end{aligned} \tag{2.17}$$

Each term in (2.17) can be evaluated by at most three dimensional numerical integral using the pre-calculated densities and CDFs. This also establishes

$$P_0 \in \big(P_U - P(M^c) \sum_{i=1}^{b} \sum_{j<i, j \notin \mathcal{N}_i} P(B_i \cap B_j | M^c), P_U \big) \tag{2.18}$$

In the next section we give the numerical results of P_U and upper bound of ϵ_U using the phylogenetic tree from the American Gut dataset. In addition, we have the following theorem on the convergence rate of the relative error with regards to the observed statistic w.

Theorem 2. *Given the set of all triplets $\mathcal{B}$ and the partition $\mathcal{M}$ on the internal nodes $\mathcal{I}$, define the following quantities:*

- $\xi_1 = |\{(i,j) : \mathcal{B}_i \cap \mathcal{B}_j = \emptyset, \mathcal{B}_i \notin \mathcal{M}, \mathcal{B}_j \notin \mathcal{M} \text{ and } 1 \le j < i \le b|$
- $\xi_2 = |\{(i,j) : |\mathcal{B}_i \cap \mathcal{B}_j| = 1, \mathcal{B}_i \notin \mathcal{M}, \mathcal{B}_j \notin \mathcal{M} \text{ and } 1 \le j < i \le b|$
- $\xi_3 = |\{i : |\mathcal{M}_i| = 3 \text{ and } 1 \le i \le m\}|$

Then under the condition that $(\xi_3 - 1)\big(1 - F_3(w_T)\big) < 0.1$ *and* $w_T \ge 12$*, we have*

$$\begin{aligned}\frac{\epsilon_U}{P_U} &< \frac{\xi_1}{0.95\xi_3 + \xi_T}\big(1 - F_3(w)\big) + \frac{0.9\xi_2}{0.95\xi_3 + \xi_T}\sqrt{\frac{\pi}{2}} \cdot \frac{1}{w} \\ &= \mathcal{O}(e^{-\frac{w}{6}}) + \mathcal{O}(w^{-1})\end{aligned}$$

for all $w \ge w_T$*, where*

$$\xi_T = \frac{\sum_{i=1}^{b} P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c)}{1 - F_3(w)} \text{ is evaluated at } w = w_T$$

See Section 2.7 for the proof.

2.4.3 Comparison with Monte-Carlo simulation

We compared the lower and upper bound in (2.18) with Monte-Carlo simulated p-values. Each round of simulation produces 5×10^4 simulated maximum triplet statistics, and we use their proportion of exceeding w as the estimated p-value. The maximum triplet statistic is simulated through generating the χ^2_1 distributed Z_A's for all $A \in \mathcal{I}$ and then applying (2.8) and (2.9). We draw the comparison for a variety of scenarios with different numbers of OTU $K \in \{50, 100\}$ and different observed statistic $w \in \{15, 20, 25\}$. Given K, the tree structure is obtained from keeping the top K OTUs with the highest count in all feces samples from the American Gut dataset (introduced later). In each scenario, we provide the histogram of simulated p-values over 5000 rounds.

Figure 2.3 shows that our bounds consistently contain the center of the simulated p-values. Since the Monte Carlo p-values are merely binomial proportions, the ratio of their

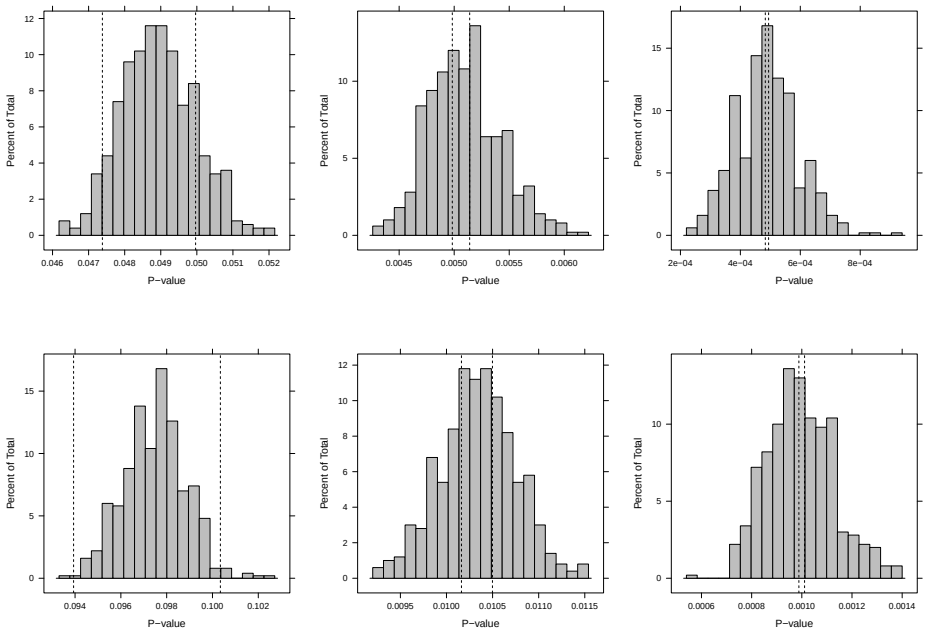

Figure 2.3: Comparison between the interval bound and simulated p-values. Each simulated p-value is the proportion exceeding w over 5×10^4 runs. Dashed lines indicate the upper and lower bound as in (2.18). Top row and bottom row indicate $K = 50$ and $K = 100$ respectively. Left column: $w = 15$, middle column: $w = 20$, and right column: $w = 25$.

spread (measured by standard deviation) to P_0 goes to infinity as $P_0 \to 0$. In contrast, our method gives a ratio that tends to zero by Theorem 2. This makes our approach particularly useful for scenarios where a large number of tests leads to very small p-value threshold after multiple testing correction. In order to keep a fixed relative error, the computation time of Monte Carlo method needs to scale up much faster with w than our method.

2.5 Application to American Gut Project

American Gut Project [McDonald et al. (2015)] is an open-access and crowd-sourced initiative that involves the public into the research of human microbiome and aims at providing a much more comprehensive reference set than the previous Human Microbiome Project (Methé et al., 2012). After contributing to the project fund, participants complete a questionnaire and ship their microbiome sample to the sequencing lab currently located at Uni-

versity of California, San Diego. The questionnaire covers a wide range of topics regarding demographic information, diet, lifestyle, etc. Sampling sites include skin, tongue and feces, although the vast majority of participants provided the feces sample. The samples are sequenced on 16s rRNA and further processed by QIIME (Caporaso et al., 2010) pipeline to produce the OTUs and the phylogenetic tree. The 2016 May 16 cohort of public dataset includes more than eight thousands of subjects, with median of sequences per individual as 14680 and standard deviation as 32455.

2.5.1 Cross-group comparison

Our focus is comparison of the feces microbiome across different diet habits. We pick the top 100 OTUs with the highest count summing over all feces samples. The phylogenetic tree on these OTUs is fully binary. We also select a total of seven categories of diet from the questionnaire. Each diet divides the samples into two groups; group 1 consists of individuals with ingestion rate less than three times per week, and group 2 corresponds to more than or equal to three times per week. Since the questions are not compulsory, a large number of subjects do not leave any response. The diet names and their sample sizes in both groups are as follows: fermented plant (880 vs 3024), fruit (2336 vs 1660), milk and cheese (1743 vs 2261), poultry (1421 vs 2611), seafood (556 vs 3452), sugary sweet (1542 vs 2493) and vegetable (3422 vs 577).

For each diet type, we test the equality of mean proportions between two groups using three methods: DTM with PhyloScan, DTM with Sidak correction and global DM. Table 2.1 presents their p-values using the 100 OTUs. The DTM(PhyloScan) column contains P_U and the upper bound of its error ϵ_U in the parenthesis, both derived in Section 2.4.2. DTM(Sidak) column is calculated as the Sidak multiple testing correction $1 - (1 - \min_{A \in \mathcal{I}} p_A)^{|\mathcal{I}|} \approx |\mathcal{I}| \min_{A \in \mathcal{I}} p_A$. We also provide DM p-values after grouping the 100 OTUs into family and class levels, respectively. The grouping operation based on taxonomy is a common practice in recent papers including La Rosa et al. (2012) and Chen and Li (2013). At each taxonomic

level, all OTUs with missing taxa information are placed into the same group. This leads to a total of 22 categories on family level and 9 categories on class level. The DM p-values are calculated using the R package `HMP`.

Table 2.1: DM and DTM p-values for testing microbiome compositions across different diet habits. DTM(PhyloScan) contains P_U and the upper bound on ϵ_U shown in parenthesis. DTM(Sidak) contains the Sidak-corrected p-values $1-(1-\min_{A\in\mathcal{I}} p_A)^{|\mathcal{I}|}$. For DM, we provide p-values directly on the 100 OTUs as well as after grouping the 100 OTUs into family and class levels, respectively.

	DTM		**DM**		
Diet	**PhyloScan**	**Sidak**	**OTU**	**Family**	**Class**
Fermented plant	0.308 (0.036)	0.239	0.377	0.147	0.038
Fruit	8.75×10^{-5} (1.52×10^{-6})	1.64×10^{-4}	2.81×10^{-3}	0.012	0.218
Milk and cheese	1.07×10^{-4} (1.86×10^{-6})	6.48×10^{-3}	0.029	0.262	0.285
Poultry	0.023 (8.71×10^{-4})	0.111	0.158	0.287	0.691
Seafood	6.85×10^{-5} (1.17×10^{-6})	6.40×10^{-3}	1.75×10^{-4}	0.194	0.772
Sugary sweet	5.13×10^{-3} (1.43×10^{-4})	0.015	0.719	0.558	0.815
Vegetable	7.39×10^{-5} (1.27×10^{-6})	4.79×10^{-5}	3.77×10^{-3}	1.88×10^{-3}	0.014

All diet habit comparisons exhibit significant DTM(PhyloScan) p-values at 0.05 level except fermented plant. This is consistent with the findings in Turnbaugh et al. (2009) and David et al. (2014), both of which established that the human gut microbiome is highly sensitive to the dietary nutrient composition. DTM(Sidak) also produces similar significance results, although in five out of seven comparisons its p-values are larger than PhyloScan. The largest relative difference occurs at seafood comparison (Sidak p-value about 100 times greater than PhyloScan). The remaining two comparisons (fermented plant and vegetable) are likely to have either a single dominating signal or weak clustering pattern, both of which hurt testing power after signal pooling. Still, PhyloScan has only mildly larger p-values under these circumstances. This data analysis concludes that PhyloScan is superior to Sidak correction in most cases. Note that p-values of DM on OTUs fail to reach significance for

fermented plant, poultry and sugary sweet. This happens in even more comparisons on family and class levels.

We further visualize the significant internal nodes in Figure 2.4 for four of the diet comparisons: milk and cheese, seafood, sugary sweets and vegetable. Using a simple binary search, we find that $w = 16.579$ yields $P_U = 0.05$ with $\epsilon_U \leq 2.43 \times 10^{-3}$. All triplets with test statistics greater than the above threshold, i.e. $W_i > 16.579$, are plotted in dark gray. In some cases the triplets overlap with each other, leading to a much longer chain than the original setup. We also provide the taxonomy for all internal nodes that belong to a certain significant triplet in Table 2.2. The internal node taxon is defined according to the following algorithm: starting from kingdom, repeatedly decrease the rank by one level until the descendant OTUs of that particular internal node no longer share the same taxa on the next lower rank (missing taxa on OTUs are excluded). The algorithm then picks the common taxon of the descendant OTUs on the rank at which the algorithm stops. In other words, the internal node taxon reflects the finest classification upon which all of its descendant OTUs agree.

2.5.2 DM vs DTM test

We can also test the model fit of DTM against DM directly on the OTUs. Since DM is nested in the DTM family, we can use the likelihood ratio test (LRT) for

$$H_0 : \exists \nu > 0 \text{ s.t. } \forall A \in \mathcal{I}, \nu_A = \nu \sum_{\omega \in A} \pi_\omega \text{ (DM) vs } H_a : \text{otherwise (DTM)}$$

with the test statistic defined as

$$\Lambda(\boldsymbol{x}) = -2 \log \frac{\mathcal{L}(\hat{\nu}, \hat{\boldsymbol{\pi}})}{\mathcal{L}_T(\{(\hat{\nu}_A, \hat{\boldsymbol{\pi}}_A) : A \in \mathcal{I}\})} \sim \chi^2_{|\mathcal{I}|-1} \text{ under } H_0, \tag{2.19}$$

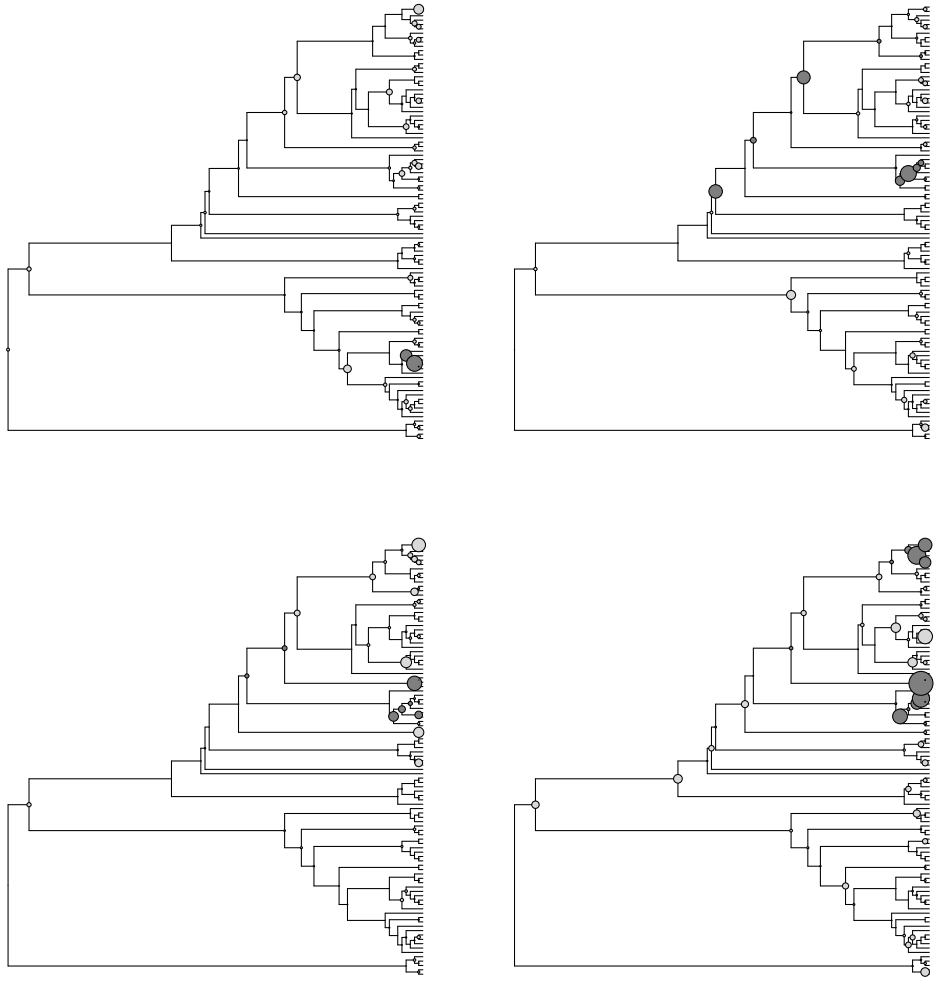

Figure 2.4: Significant triplets from DTM testing. Top left: Milk and cheese, top right: seafood, bottom left: sugary sweets and bottom right: vegetable. The size of the circle on internal node A is proportional to $-\log(p_A)$. Triplets with $W_i > 16.579$ are plotted in dark gray.

where $(\hat{\nu}, \hat{\boldsymbol{\pi}})$ in the numerator of (2.19) are MLEs of the DM model, and each $(\hat{\nu}_A, \hat{\boldsymbol{\pi}}_A)$ in the denominator are obtained through maximizing the DTM conditional likelihood (2.5). We use the low-storage BFGS optimization implemented in package `nloptr` to calculate the MLE estimates. The degrees of freedom in (2.19) is $|\mathcal{I}| - 1$ for a binary phylogenetic tree since (i) $\dim(\boldsymbol{\pi}) = \dim(\{\boldsymbol{\pi_A} : A \in \mathcal{I}\}) = K - 1$, and (ii) $\dim(\{\nu_A : A \in \mathcal{I}\}) = |\mathcal{I}|$.

Table 2.3 shows the LRT result. The test is separately applied to male and female Caucasians living in a variety of geographic regions. Each region consists of certain states in

Table 2.2: Taxa on significant triplets from PhyloDM hypothesis testing for each diet comparison. Each internal node that belongs to a certain significant triplet is assigned a taxon based on its descendant OTUs (details described in section 2.5.1). Only the lowest level taxon is reported for each internal node. We omit the class rank since there are no significant internal nodes on such level in any cross-group comparisons.

Diet	**Phylum**	**Order**	**Family**	**Genus**
Fermented plant	—	—	—	—
Fruit	Firmicutes	Clostridiales	Clostridiaceae	Clostridium
	—	—	Ruminococcaceae	Faecalibacterium
Milk and cheese	—	—	—	Bacteroides
Poultry	—	Clostridiales	—	Coprococcus
Seafood	Firmicutes	Clostridiales	Lachnospiraceae	Coprococcus
	—	—	Ruminococcaceae	Ruminococcus
Sugary sweet	Firmicutes	Clostridiales	Lachnospiraceae	Coprococcus
	—	—	Ruminococcaceae	—
Vegetable	Firmicutes	Clostridiales	Lachnospiraceae	Blautia
	—	—	Ruminococcaceae	Coprococcus
	—	—	—	Lachnospira

the U.S. defined according to Bureau of Economic Analysis. The degree of freedom for all tests is 98 since our phylogenetic tree is binary and $|\mathcal{I}| = |K| - 1 = 99$. All scenarios yield LRT p-values less than 10^{-10}, which indicates significantly improved fit on the data using DTM. We also note that $\Lambda(x)$ in general increases with the sample size, as evidence towards heterogeneity in OTU dispersion strengthens with more available data.

2.5.3 Simulation

We use two simulation strategies to evaluate the power of PhyloScan test under various conditions. From American Gut dataset, we extracted a total of 662 individuals who identified themselves as male Caucasian living in the far west (Alaska, California, Hawaii, Nevada, Oregon and Washington). In each round of simulation, these selected samples are randomly divided into two equal-sized groups to generate data under the global null. For data under the alternative, the first simulation strategy randomly selects an OTU and increases its

Table 2.3: Likelihood ratio test for DM vs DTM. The test is separately applied to male and female Caucasians in a variety of geographic regions. Each test is accompanied by the LRT statistic $\Lambda(x)$ and the sample size.

	Male		**Female**	
Region	$\boldsymbol{\Lambda(x)}$	**Sample size**	$\boldsymbol{\Lambda(x)}$	**Sample size**
Far West	14179.76	663	15768.10	775
Great Lakes	4025.36	180	5541.45	276
Mideast	7497.80	328	8112.80	396
New England	5630.21	239	5793.38	269
Rocky Mountain	6084.90	244	6676.25	300
Southeast	7030.43	324	7383.72	366
Southwest	3442.69	153	3675.52	189

count by a fixed percentage for all samples in the second group, whereas the second simulation strategy random selects an internal node and increases the count all of its descendant OTUs equally by a fixed percentage for all samples in the second group. We use the same 100 OTUs as before and produce 5000 rounds of simulation.

Figure 2.5 demonstrates the distribution of DM and DTM p-values under the global null. We fit three separate DM on 100 OTUs, family level and class level. The histogram of DTM p-values is produced by using all the p-values on the internal nodes. Surprisingly, the distribution of p-values for DM on the OTUs is far from being uniform on (0,1), which leads to conservative inference and loss of power. The discrepancy alleviates as we group more OTUs into family or class level, although its empirical distribution is still noticeably skewed. This phenomenon reflects the fact that DM is severely under-parametrized for microbiome data even in low dimensions, as it fits a single dispersion parameter that simultaneously controls all categories. In contrast, DTM solves this issue through fitting a family of dispersion parameters $\{\nu_A : A \in \mathcal{I}\}$ that leads to better calibrated p-values.

Figure 2.6 and 2.7 show the ROC and power curves when we use the first simulation strategy to increase the count of a random OTU. We provide the result for (i) DM on the OTUs (ii) DTM using the maximum of the single node statistic, or $\max_{A\in\mathcal{I}} Z_A$ and (iii) DTM using the maximum of the triplet statistic, or $\max_{i\leq b} W_i$. The last strategy is the

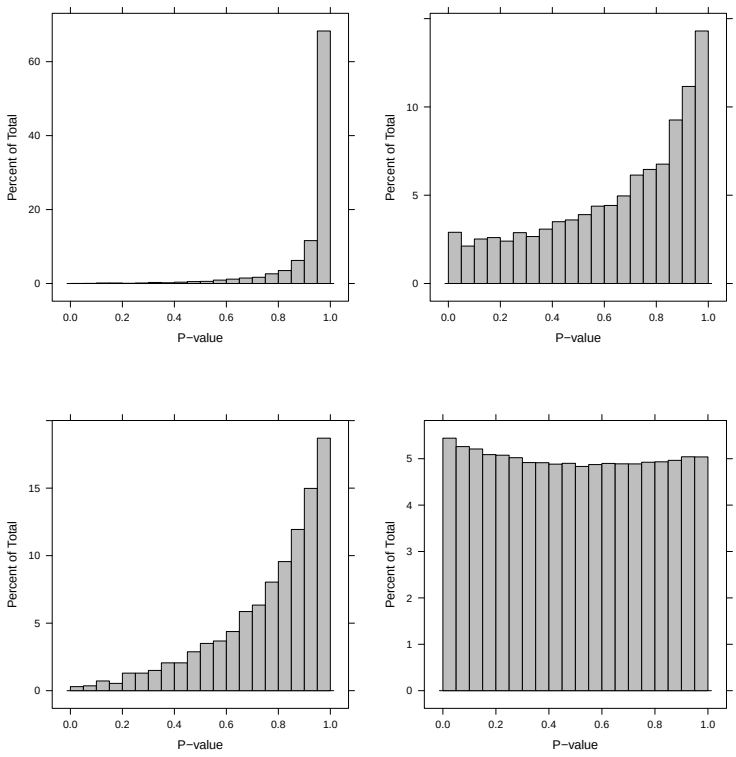

Figure 2.5: P-value histograms under the global null. We randomly place the 662 samples from Caucasian male living in far west into two equal-sized groups and produce their p-values for 5000 rounds. Top left is DM on the OTUs, top row right is DM on family level, bottom left is DM on class level, and bottom right is DTM .

one employed in PhyloScan procedure. Both DTM methods give improved performance compared to DM due to highly localized signal.

Figure 2.8 and 2.9 show the ROC and power curves when we use the second simulation strategy to increase the count of all OTUs under a random internal node. The minimum number of OTUs under the randomly selected internal node controls degree of localization in the signal. This simulation setup reflects the more biologically meaningful scenario in which a number of taxa exhibit differences in the between-group comparison. In all cases, DTM consistently provides higher power than DM. The DTM 3-node method also provides higher power than DTM 1-node at moderate increment levels. When the increment level is high, there will be a certain Z_A whose value dominates all other node statistics, so the extra gain from pooling signal strength within triplets diminishes.

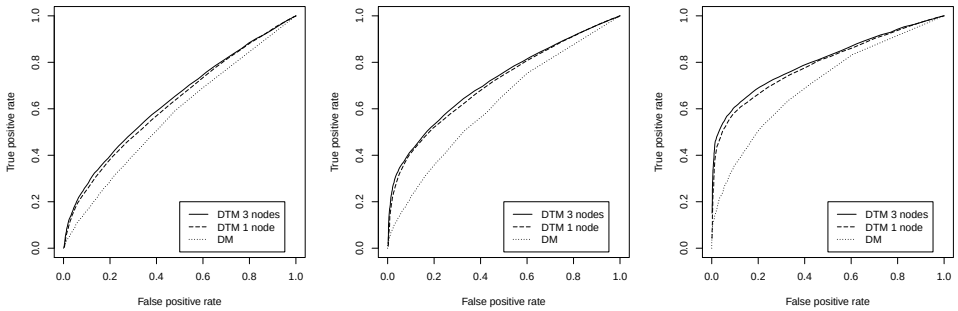

Figure 2.6: ROC curves from increasing the count of a random OTU. For left to right, the percentage increment is set as 100%, 150% and 250%

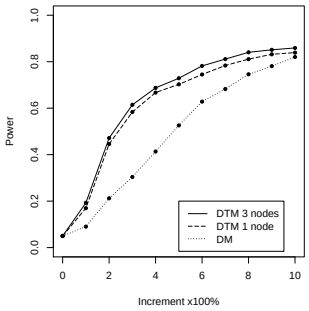

Figure 2.7: Power of DM and DTM with regard to different increment in a random OTU at false positive rate = 0.05.

2.6 Discussion

DTM models the microbiome data through a cascade of local DMs with varying degrees of resolutions on the phylogenetic tree. We take advantage of the correlated signals on the tree through a scan statistic approach and provide upper and lower bound on its tail probability for testing cross-group differences. Both empirical results on American Gut data and simulations demonstrated the efficiency and accuracy of our method.

DTM is a generalization of DM with $|\mathcal{I}| - 1$ more dispersion parameters. An interesting question is whether one could stepwise tune the model (hence the number of parameters) from DM to DTM. To start, the DTM representation in (3.1) shrinks to the degenerate DM if $\exists \nu > 0$ s.t. $\nu_A = \nu \sum_{w \in A} \pi_w$ for all $A \in \mathcal{I}$. This condition is equivalent to $\nu_A = \nu_{R(A)} \pi_{R(A),i}$ for $\forall A \in \mathcal{I} \backslash \{\Omega\}$ with $A = \mathcal{C}(R(A))_i$. Stepwise tuning can be achieved through requiring only $\nu_A = \nu_{R(A)} \pi_{R(A),i}$ to hold over $A \in \tilde{\mathcal{I}}$ where $\tilde{\mathcal{I}} \subset \mathcal{I} \backslash \{\Omega\}$ controls the effective

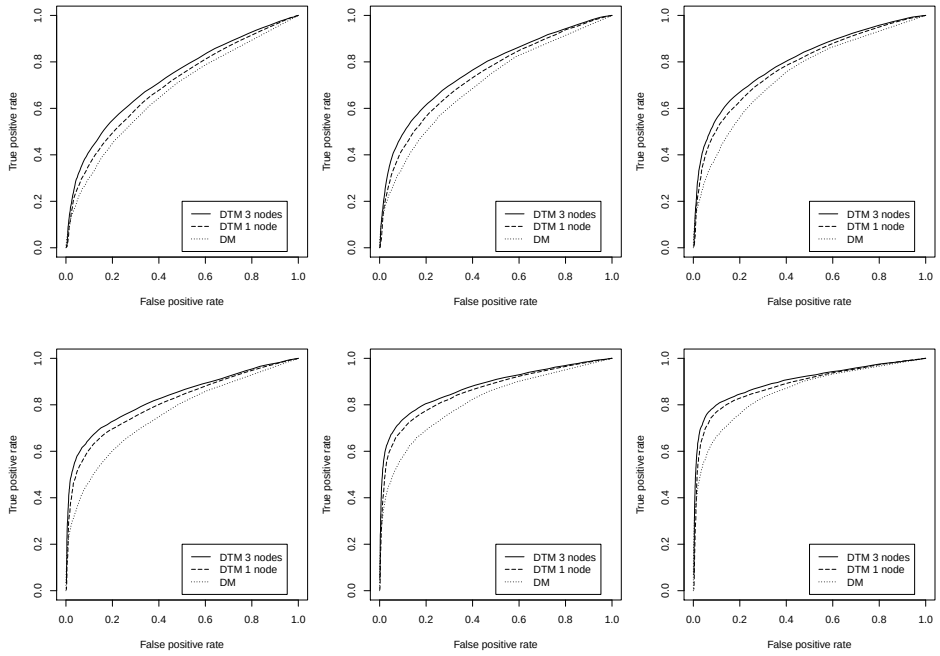

Figure 2.8: ROC curves from increasing the count of all OTUs under a random internal node. The top row and the bottom row have the percentage increment set as 50% and 75% respectively. From left to right column, the minimum number of OTUs under the chosen internal node is 2, 3 and 5.

degrees of freedom. Apparently $\tilde{\mathcal{I}} = \emptyset$ leads to DTM and $\tilde{\mathcal{I}} = \mathcal{I}\backslash\{\Omega\}$ leads to DM, so any choice of $\tilde{\mathcal{I}}$ in the middle yields a model between the two extremes. Standard model selection techniques such as information criterion or cross validation can then be applied. Although the existence of such spectrum grants substantial flexibility, we note that it can be computationally infeasible to examine the model fit of all $2^{|\mathcal{I}|-1}$ possible configurations. A potential workaround is to enlarge $\mathcal{T}$ stepwise by a greedy algorithm or use dynamic programming, but it is not clear under which conditions we are guaranteed to recover the global optimum.

Since our PhyloScan procedure requires only the p-value as input, it can be easily applied to any extensions or other distributions. For example, the DTM framework can be adapted to incorporate continuous variable of interest and adjust for the effects of confounders. When the tree is fully binary, we let $\lambda_A = \pi_{A,1}$ to fully represent $\boldsymbol{\pi}_A = (\pi_{A,1}, 1 - \pi_{A,1})$ in (3.1).

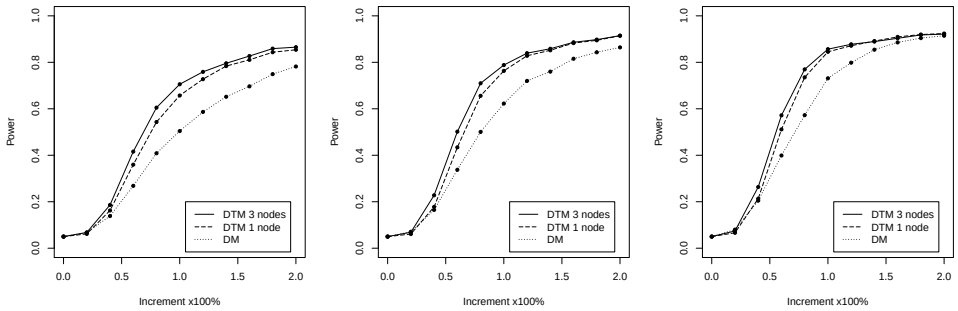

Figure 2.9: Power of DM and DTM with regard to different increment in all OTUs under a random internal node at false positive rate = 0.05. From left to right, the minimum number of OTUs under the chosen internal node is 2, 3 and 5.

Then we can build separate logistic regression models for each A:

$$\log \frac{\lambda_A}{1-\lambda_A} = \beta_{A,0} + \beta_{A,1} u + \sum_{i=1}^{s} \beta_{A,s+1} c_s \tag{2.20}$$

where $\beta_{A,i}$ is the ith regression coefficient, u denotes the continuous variable of interest and c_1, c_2, ..., c_s are the confounders. After obtaining maximum likelihood estimates of the coefficients as well as ν_A, we test the significance of u's coefficient to produce p-values and use them as input to PhyloScan in order to borrow strength from neighboring nodes. Another possible extension is related to the issue of zero-adjustment. In Figure 2.5, DM p-values exhibit apparent right skew under the global null hypothesis. A follow-up inspection shows that MoM estimation tends to produce higher expected zero count than observed. Both right-skewness and zero-deflation of the global DM are likely caused by underestimation of ν, which makes the Dirichlet prior more dispersed. In DTM, we still observe mild level of zero-deflation, although the extent is much less severe than global DM. It is also possible to have zero-inflation when one switches to a different sequencing technology or OTU construction algorithm. Incorporating zero-adjustment into existing model can lead to significantly better fit while easily handled by PhyloScan.

2.7 Theorem proofs

2.7.1 Proof of Theorem 1

We provide the proof when the phylogenetic tree is binary. The result can be easily generalized to an arbitrary tree. The elements in $\mathcal{I}$ can be ordered as $A_1, A_2, ..., A_{|\mathcal{I}|}$ such that each parent node always appears in front of its children. Let p_{A_l} be the p-value for testing H_{0,A_l}. Without loss of generality, let us assume $A_{|\mathcal{I}|} = \{K-1, K\}$. For any subject with OTU counts $\boldsymbol{x}$, the probability function (2.6) can be written as

$$\begin{aligned} f_T(\boldsymbol{x}) &= \prod_{l=1}^{\mathcal{I}} f\big(\boldsymbol{x}(A_l)|N(A_l)\big) \\ &= \prod_{l=1}^{\mathcal{I}-1} f\big(\boldsymbol{x}(A_l)|N(A_l)\big) \cdot f\big(\boldsymbol{x}(A_{|\mathcal{I}|})|N(A_{|\mathcal{I}|})\big) \\ &= f_T\big(x_1, x_2, ..., x_{K-2}, N(A_{|\mathcal{I}|})\big) f\big(\boldsymbol{x}(A_{|\mathcal{I}|})|N(A_{|\mathcal{I}|})\big), \end{aligned}$$

which yields

$$f_T\big(\boldsymbol{x}|N(A_{|\mathcal{I}|})\big) = f_T\big(x_1, x_2, ..., x_{K-2}, N(A_{|\mathcal{I}|})|N(A_{|\mathcal{I}|})\big) f\big(\boldsymbol{x}(A_{|\mathcal{I}|})|N(A_{|\mathcal{I}|})\big).$$

Since the above conditional independence relationship holds for all subjects, it follows that $p_{A_{|\mathcal{I}|}}$ is independent of all other p_{A_l}'s conditional on $N(A_{|\mathcal{I}|})$. Therefore under H_0,

$$\begin{aligned}
P(\bigcap_{l=1}^{|\mathcal{I}|}\{p_{A_l} \le \alpha_l\}) &= E\Big(P(\bigcap_{l=1}^{|\mathcal{I}|}\{p_{A_l} \le \alpha_l\}|N(A_{|\mathcal{I}|}))\Big) \\
&= E\Big(P(\bigcap_{l=1}^{|\mathcal{I}|-1}\{p_{A_l} \le \alpha_l\}|N(A_{|\mathcal{I}|}))P\big(p_{A_{|\mathcal{I}|}} \le \alpha_{|\mathcal{I}|}|N(A_{|\mathcal{I}|})\big)\Big) \\
&= \alpha_{|\mathcal{I}|}E\Big(P(\bigcap_{l=1}^{|\mathcal{I}|-1}\{p_{A_l} \le \alpha_l\}|N(A_{|\mathcal{I}|}))\Big), \text{ asymptotically} \\
&= \alpha_{|\mathcal{I}|}P(\bigcap_{l=1}^{|\mathcal{I}|-1}\{p_{A_l} \le \alpha_l\}).
\end{aligned}$$

where the second last equation requires the asymptotic distribution of (2.3) to hold so that $P\big(p_{A_{|\mathcal{I}|}} \le \alpha_{|\mathcal{I}|}|N(A_{|\mathcal{I}|})\big)$ becomes very close to $\alpha_{|\mathcal{I}|}$.

Repeating the above procedures iteratively for $A_{|\mathcal{I}|-1}$, $A_{|\mathcal{I}|-2}$, ... A_1 gives

$$P(\bigcap_{l=1}^{|\mathcal{I}|}\{p_{A_l} \le \alpha_l\}) = \prod_{l=1}^{|\mathcal{I}|}\alpha_l.$$

2.7.2 Proof of Theorem 2

By (2.17),

$$\begin{aligned}
\epsilon_U &\le \sum_{i=1}^{b}\sum_{j<i, j\notin\mathcal{N}_i} P(B_i \cap B_j \cap M^c) \\
&= \sum_{i\le b, \mathcal{B}_i\notin\mathcal{M}}\ \sum_{j<i, j\notin\mathcal{N}_i, \mathcal{B}_j\notin\mathcal{M}} P(B_i \cap B_j \cap M^c).
\end{aligned}$$

The elements inside the summation sign above fall in one of the following two categories

(i) $|\mathcal{B}_i \cap \mathcal{B}_j| = 0$ which means $P(B_i \cap B_j \cap M^c) < P(B_i \cap B_j) = \big(1 - F_3(w)\big)^2$

(ii) $|\mathcal{B}_i \cap \mathcal{B}_j| = 1$ so that $P(B_i \cap B_j \cap M^c) \leq P(\sum_{i=1}^{3} Y_i > w, \sum_{i=3}^{5} Y_i > w, Y_i < w \text{ for all } i)$, where Y_i's are i.i.d. chi-square distributed with 1 degree of freedom.

Let $f_1(y) = \frac{y^{-\frac{1}{2}} e^{-\frac{y}{2}}}{\sqrt{2\pi}}$ be the density function of χ_1^2. Conditioning on Y_3 gives the following upper bound on category (ii):

$$\begin{aligned} P(\sum_{i=1}^{3} Y_i > w, \sum_{i=3}^{5} Y_i > w, Y_i < w) &< \int_0^w f_1(y) P(\sum_{i=1}^{3} Y_i > w | Y_3 = y) \cdot \\ &\quad P(\sum_{i=3}^{5} Y_i > w | Y_3 = y) dy \\ &= \int_0^w \frac{y^{-\frac{1}{2}} e^{-\frac{y}{2}}}{\sqrt{2\pi}} \big(1 - F_2(w-y)\big)^2 dy \\ &= \frac{1}{\sqrt{2\pi}} \int_0^w y^{-\frac{1}{2}} e^{-\frac{y}{2}} e^{-(w-y)} dy \\ &= \frac{e^{-w}}{\sqrt{2\pi}} \int_0^w y^{-\frac{1}{2}} e^{\frac{y}{2}} dy \\ &< 0.9 w^{-\frac{1}{2}} e^{-\frac{w}{2}} \text{ for } w \geq 12, \end{aligned}$$

where the last line is deduced by noticing that the function $h(w) = \int_0^w y^{-\frac{1}{2}} e^{\frac{y}{2}} dy - 2.25 w^{-\frac{1}{2}} e^{\frac{w}{2}}$ satisfies 1) $h(12) < 0$ and 2) $h'(w) = -0.125 w^{-\frac{1}{2}} e^{\frac{w}{2}} + 1.125 w^{-\frac{3}{2}} e^{\frac{w}{2}} = e^{\frac{w}{2}} \cdot w^{-\frac{1}{2}} (1.125 w^{-1} - 0.125) < 0$ for $w \geq 12$. Together they establish $\int_0^w y^{-\frac{1}{2}} e^{\frac{y}{2}} dy < 2.25 w^{-\frac{1}{2}} e^{\frac{w}{2}}$ for $w \geq 12$.

Since there are ξ_1 terms in (i) and ξ_2 terms in (ii), we have

$$\epsilon_U < \xi_1 \big(1 - F_3(w)\big)^2 + 0.9 \xi_2 w^{-\frac{1}{2}} e^{-\frac{w}{2}}. \tag{2.21}$$

Our next step is to put a lower bound on P_U. According to (2.16),

$$P_U = P(M) + \sum_{i=1}^{b} P(B_i \cap B_{\mathcal{N}_i}^c \cap M^c). \tag{2.22}$$

The lower bound of first term in the right side of (2.22) is obtained by considering only

the triplets in $\mathcal{M}$:

$$\begin{aligned} P(M) &\geq 1 - F_3(w)^{\xi_3} \\ &= 1 - \Big(1 - \big(1 - F_3(w)\big)\Big)^{\xi_3} \\ &\geq \xi_3\big(1 - F_3(w)\big) - \frac{\xi_3(\xi_3 - 1)}{2}\big(1 - F_3(w)\big)^2, \text{ by Taylor expansion} \\ &\geq 0.95\xi_3\big(1 - F_3(w)\big), \text{ as long as } \xi_3\big(1 - F_3(w)\big) < 0.1. \end{aligned} \tag{2.23}$$

Next, we get the lower bound of the second term in the right side of (2.22). For any fixed i, let Y_1, Y_2 and Y_3 denote the i.i.d. χ^2_1 variables included in the event B_i so that $B_i = \{\sum_{j=1}^3 Y_j > w\}$. Then we have

$$\begin{aligned} P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c) &= \int_{\sum y_i > w} f_1(y_1)f_1(y_2)f_1(y_3)P(B^c_{\mathcal{N}_i} \cap M^c | y_1, y_2, y_3)d\boldsymbol{y} \\ &= \int_{\sum y_i > w} f_1(\boldsymbol{y})h(\boldsymbol{y}, w)d\boldsymbol{y} \end{aligned}$$

where we define $f_1(\boldsymbol{y}) = f_1(y_1)f_1(y_2)f_1(y_3)$ and $h(\boldsymbol{y}, w) = P(B^c_{\mathcal{N}_i} \cap M^c | y_1, y_2, y_3)$. In addition, let $V(w) = \{\boldsymbol{y} : \sum_{i=1}^3 y_i > w\}$ denote the region of integration. By Reynolds transport theorem (or multidimensional Leibniz's rule),

$$\begin{aligned} \frac{d}{dw}P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c) = &\int_{V(w)} f_1(\boldsymbol{y})\frac{\partial}{\partial w}h(\boldsymbol{y}, w)d\boldsymbol{y} \\ &+ \int_{\partial V(w)} (\boldsymbol{v}_b \cdot \boldsymbol{n})f_1(\boldsymbol{y})h(\boldsymbol{y}, w)dA \end{aligned} \tag{2.24}$$

where $\partial V(w)$ is the boundary of $V(w)$, $\boldsymbol{v}_b$ is the Eulerian velocity of the boundary, $\boldsymbol{n}$ is the outward-pointing unit-normal, and dA is the surface element.

On the other hand,

$$\begin{aligned}\frac{d}{dw}P(B_i) &= \frac{d}{dw}\int_{V(w)} f_1(\boldsymbol{y})d\boldsymbol{y} \\ &= \int_{\partial V(w)} (\boldsymbol{v}_b \cdot \boldsymbol{n}) f_1(\boldsymbol{y}) dA \end{aligned} \tag{2.25}$$

Since both $B^c_{\mathcal{N}_i}$ and M^c strictly enlarges as w increases, $h(\boldsymbol{y}, w)$ is an increasing function of w. This leads to $\frac{\partial}{\partial w}h(\boldsymbol{y}, w) > 0$. Moreover, $h(\boldsymbol{y}, w) < 1$ from definition. Lastly, the fact that $V(w_1) \subset V(w_2)$ for any $w_1 > w_2$ gives $\boldsymbol{v}_b \cdot \boldsymbol{n} \leq 0$. These altogether establish $\frac{d}{dw}P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c) > \frac{d}{dw}P(B_i)$ as we compare the expression in (2.24) and (2.25). With the apparent relation $P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c) < P(B_i)$, we conclude that

$$\frac{d}{dw} \cdot \frac{P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c)}{P(B_i)} > 0 \text{ for all } i \tag{2.26}$$

Therefore if we only focus on calculating ϵ_U/P_U for $w \geq w_T$ where w_T is a pre-fixed value, then we can first evaluate $\sum_{i=1}^b P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c)/\big(1 - F_3(w)\big) = \xi_T$ at $w = w_T$. By (2.26),

$$\sum_{i=1}^{b} P(B_i \cap B^c_{\mathcal{N}_i} \cap M^c) \geq \xi_T\big(1 - F_3(w)\big) \text{ for } w \geq w_T \tag{2.27}$$

Plugging in (2.23) and (2.27) into (2.22) gives

$$P_U \geq (0.95\xi_3 + \xi_T)\big(1 - F_3(w)\big) \tag{2.28}$$

The upper bound on ϵ_U in (2.21) and the lower bound on P_U in (2.28) yield

$$\begin{aligned}\frac{\epsilon_U}{P_U} &< \frac{\xi_1\big(1-F_3(w)\big)^2 + 0.9\xi_2 w^{-\frac{1}{2}} e^{-\frac{w}{2}}}{(0.95\xi_3+\xi_T)\big(1-F_3(w)\big)} \\ &= \frac{\xi_1\big(1-F_3(w)\big)}{0.95\xi_3+\xi_T} + \frac{0.9\xi_2 w^{-\frac{1}{2}} e^{-\frac{w}{2}}}{(0.95\xi_3+\xi_T)\big(1-F_3(w)\big)} \\ &< \frac{\xi_1\big(1-F_3(w)\big)}{0.95\xi_3+\xi_T} + \frac{0.9\xi_2}{0.95\xi_3+\xi_T}\sqrt{\frac{\pi}{2}}\cdot\frac{1}{w} \end{aligned} \tag{2.29}$$

for all $w > w_T \geq 12$. The last line comes from substituting the following equation:

$$1-F_3(w) = \frac{1}{\sqrt{2\pi}}\int_w^\infty \sqrt{y} e^{-\frac{y}{2}} dy > \frac{\sqrt{w}}{\sqrt{2\pi}}\int_w^\infty e^{-\frac{y}{2}} dy = \sqrt{\frac{2w}{\pi}} e^{-\frac{w}{2}}$$

Lemma 1 in Laurent and Massart (2000) gives $1-F_3(3+2\sqrt{3t}+2t) \leq e^{-t}$ for all $t > 0$. Since $3+2\sqrt{3t} < 4t$ when $t \geq 2$, we have that $1-F_3(6t) < e^{-t}$ or $1-F_3(w) < e^{-\frac{w}{6}}$ for $w \geq 12$. Therefore, the rate of decay for ϵ_U/P_U is $\mathcal{O}(e^{-\frac{w}{6}}) + \mathcal{O}(\frac{1}{w})$.

CHAPTER 3

DTM MODEL SELECTION BY SMOOTH APPROXIMATION TO TREE-BASED FUSION PENALTY

3.1 Tree-based fusion penalty

In the previous chapter we discussed that the Dirichlet-tree multinomial (DTM) family is a generalization of the global Dirichlet-multinomial (DM) model. In fact, there exists a spectrum of models from global DM to DTM. This chapter presents a penalization algorithm for optimal model selection along this spectrum.

We start from reviewing the notation introduced in the previous chapter:

1. $\mathcal{T} = (\Omega, \mathcal{I})$ is a rooted phylogenetic tree where the set of OTUs Ω are placed on the leaves and $\mathcal{I}$ is the set of all internal nodes. Also, let $\tilde{\mathcal{I}} = \mathcal{I} \backslash \{\Omega\}$ be the set of internal nodes excluding the root.

2. $\forall A \in \mathcal{I}$, $\mathcal{C}(A) = \{\mathcal{C}(A)_1, ..., \mathcal{C}(A)_{J_A}\}$. is collection of A's child nodes in $\mathcal{T}$, and let $R(A)$ denote the parent node of A. $J_A = |\mathcal{C}(A)|$ is the number of children under A, and $J = \sum_{A \in \mathcal{I}} J_A$. Without loss of generality, we assume $J_A = 2$ (i.e. binary tree) in this chapter.

3. For each $i = 1, 2, ..., n$ and $j = 1, 2, ..., J_A$, $x_{ij}(A) = \sum_{\omega \in \mathcal{C}(A)_j} x_{i\omega}$ is the count of the jth child of A in the ith sample. Also, define $\boldsymbol{x}_i(A) = \big(x_{i1}(A), x_{i2}(A), ..., x_{iJ_A}(A)\big)$.

4. $N_i(A) = \sum_{j=1}^{J_A} x_{ij}(A) = \sum_{\omega \in A} x_{i\omega}$ is the total number of sequence under A for the ith sample.

The DTM assumes that, $\forall A \in \mathcal{I}$,

$$\boldsymbol{q}_{A,i} \overset{i.i.d.}{\sim} \text{Dir}(\boldsymbol{\nu}_A), \quad \boldsymbol{x}_i(A)|N_i(A), \boldsymbol{q}_{A,i} \sim \text{Multinomial}(N_i(A), \boldsymbol{q}_{A,i}) \tag{3.1}$$

where $\boldsymbol{\nu}_A = (\nu_{A1}, \nu_{A2}, ..., \nu_{AJ_A})$ are the Dirichlet dispersion parameters. Notice that this is a different parametrization from the previous chapter. As we mentioned before, DTM is equivalent to global DM if

$$\forall A \in \tilde{\mathcal{I}},\ \sum_{j=1}^{J_A} \nu_{Aj} = \nu_{R(A)s_A} \tag{3.2}$$

where A is the s_Ath child of $R(A)$, or equivalently $A = \mathcal{C}(R(A))_{s_A}$. The flexibility comes from allowing (3.2) to be satisfied for only a subset of $\tilde{\mathcal{I}}$, yielding a full spectrum of models with number of parameters ranging from $\sum_{A\in\mathcal{I}} J_A$ (full DTM, (3.2) not satisfied for any $A \in \tilde{\mathcal{I}}$) to $|\Omega|$ (global DM, (3.2) satisfied for all $A \in \mathcal{I}$).

Let $\boldsymbol{\nu}$ be the concatenation of all $\boldsymbol{\nu}_A$'s into a single vector. Define $l(\boldsymbol{\nu}; \boldsymbol{x})$ as the sum of log likelihoods on all internal nodes. The log likelihood of DTM takes the following form (up to an irrelevant constant):

$$l(\boldsymbol{\nu}; \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{A\in\mathcal{I}} \Big(\sum_{j=1}^{J_A} \sum_{\xi=0}^{x_{ij}(A)-1} \log(\nu_{Aj} + \xi) - \sum_{\xi=0}^{N_i(A)-1} \log(\nu_{A\cdot} + \xi) \Big)$$

where $\nu_{A\cdot} = \sum_{j=1}^{J_A} \nu_{Aj}$. From now on we shall write $l(\boldsymbol{\nu})$ for short. The fact that (3.2) controls the model complexity suggests we use the tree-based fusion penalty for model selection:

$$\min_{\boldsymbol{\nu}} -l(\boldsymbol{\nu}) + \lambda \sum_{A\in\tilde{\mathcal{I}}} |\nu_{R(A)s_A} - \sum_{j=1}^{J_A} \nu_{Aj}| \tag{3.3}$$

where $\lambda \geq 0$ penalizes deviation from the global DM model. The tuning parameter λ can be chosen by minimizing the Bayesian information criterion (BIC) or cross-validation.

We show an example of the tree-guided fusion penalty on the dispersion in Figure 3.1. Here $\mathcal{I} = \{A_1, A_2, A_3\}$ with $A_1 = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, $A_2 = \{\omega_1, \omega_2\}$ and $A_3 = \{\omega_3, \omega_4\}$. The value of dispersion parameters are $\boldsymbol{\nu}_{A_1} = (2, 4)$, $\boldsymbol{\nu}_{A_2} = (1, 1)$, and $\boldsymbol{\nu}_{A_1} = (0.5, 1)$. Therefore,

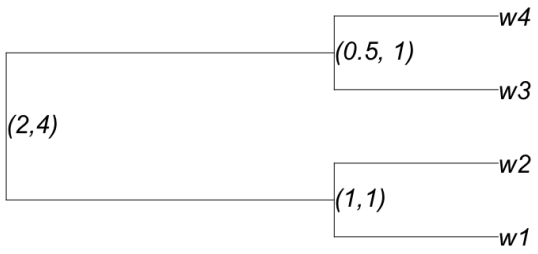

Figure 3.1: Example of DTM dispersion. For each internal node, the first and second number in the parenthesis denote the dispersion for its left child and right child, respectively.

the class of fusion penalties is calculated as

$$\lambda\big(|2-1-1|+|4-0.5-1|\big)$$

The penalty in (3.3) belongs to the class of fusion penalty first introduced by Tibshirani et al. (2005), although the parameters of interest are drastically different (dispersion versus fixed effect). The original problem considered by Tibshirani et al. (2005) is regressing y_i against covariates $x_{i1}, ..., x_{ip}$ when there is a natural ordering of the covariate dimensions:

$$\hat{\beta} = \text{argmin}_\beta \sum_i \big(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\big)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|$$

The fusion penalty induces sparseness in differences of consecutive β_j's and therefore leads to piece-wise constant solution. This fusion penalty can be easily generalized to the scenario where there exists an external graph on the covariates (Chen et al., 2012). In this case, the fusion penalty encourages fixed effects that are close on the graph to take similar values. This graph-based fusion penalty has been used in a variety of applications, including image classification (Yang et al., 2013), medical imaging (Wu et al., 2016) and regression on genomic data (Omranian et al., 2016).

3.2 Smooth approximation to fusion penalty

We follow Chen et al. (2012) to approximate the l_1 penalty term in (3.3). To start, notice that each $|\nu_{R(A),s_A} - \sum_{j=1}^{\mathcal{C}(A)} \nu_{A,j}|$ is a linear combination of elements from $\boldsymbol{\nu} \in \mathbb{R}^J$. Therefore, let $C \in \mathbb{R}^{|\tilde{\mathcal{I}}| \times J}$ be such that

$$||C\boldsymbol{\nu}||_1 = \lambda \sum_{A \in \tilde{\mathcal{I}}} |\nu_{R(A)s_A} - \sum_{j=1}^{J_A} \nu_{Aj}| \tag{3.4}$$

Using the property of dual norm, we have $||C\boldsymbol{\nu}||_1 = \max_{||\boldsymbol{\alpha}||_\infty \leq 1} \boldsymbol{\alpha}' C\boldsymbol{\nu}$. Now, define

$$f_\mu(\boldsymbol{\nu}) = \max_{||\boldsymbol{\alpha}||_\infty \leq 1} (\boldsymbol{\alpha}' C\boldsymbol{\nu} - \frac{\mu}{2}||\boldsymbol{\alpha}||_2^2) \tag{3.5}$$

as a smooth approximation to $||C\boldsymbol{\nu}||_1$. Obviously, $f_0(\boldsymbol{\nu}) = ||C\boldsymbol{\nu}||_1$ and $f_\mu(\boldsymbol{\nu}) \leq f_0(\boldsymbol{\nu})$. In addition, we have

$$f_0(\boldsymbol{\nu}) - f_\mu(\boldsymbol{\nu}) \leq \max_{||\boldsymbol{\alpha}||_\infty \leq 1} \frac{1}{2}\mu||\boldsymbol{\alpha}||_2^2 = \frac{\mu|\tilde{\mathcal{I}}|}{2}$$

This means that we simply need to take $\mu = 2\epsilon/|\tilde{\mathcal{I}}|$ to achieve $f_0(\boldsymbol{\nu}) - f_\mu(\boldsymbol{\nu}) \leq \epsilon$.

According to Theorem 1 and Proposition 2 in Chen et al. (2012), $f_\mu(\boldsymbol{\nu})$ is convex with the gradient as follows:

$$\nabla f_\mu(\boldsymbol{\nu}) = C'\boldsymbol{\alpha}^*, \quad \boldsymbol{\alpha}^* = S(\frac{C\boldsymbol{\nu}}{\mu}) \tag{3.6}$$

where $S(x) = \max(\min(x, 1), -1)$ applies to a vector element-wisely. Furthermore, $\nabla f_\mu(\boldsymbol{\nu})$ is Lipschitz continuous with the Lipschitz constant $L_\mu = ||C||^2/\mu$, where $||C||$ is the spectral norm of C.

We use a simple example to demonstrate the smooth approximation to fusion penalty. Let $\boldsymbol{\nu} = (\nu_1, \nu_2)$ and $C = (1, -1)$, which means that $||C\boldsymbol{\nu}||_1 = |\nu_1 - \nu_2|$. We plot $||C\boldsymbol{\nu}||_1$ and $f_\mu(\boldsymbol{\nu})$ in Figure 3.2 when $\mu = 0.01$. Using the definition of $S(\cdot)$, it is easy to deduce

that $\boldsymbol{\alpha}^* = \text{sign}(C\boldsymbol{\nu})$ when $|\nu_1 - \nu_2| > \mu$ and $\boldsymbol{\alpha}^* = C\boldsymbol{\nu}$ when $|\nu_1 - \nu_2| \leq \mu$. Therefore,

$$f_\mu(\boldsymbol{\nu}) = \begin{cases} |\nu_1 - \nu_2| - \frac{\mu}{2}, \text{ if } |\nu_1 - \nu_2| > \mu \\ \frac{(\nu_1 - \nu_2)^2}{2\mu}, \text{ if } |\nu_1 - \nu_2| \leq \mu \end{cases}$$

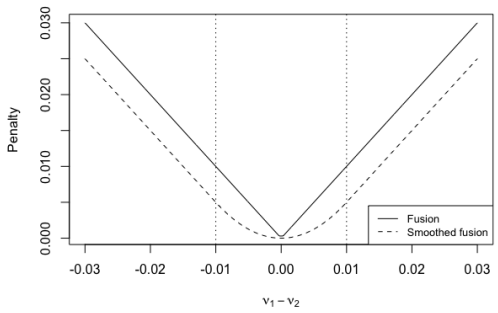

Figure 3.2: Plot of $||C\boldsymbol{\nu}||_1$ (solid line) and $f_\mu(\boldsymbol{\nu})$ (dashed line) versus $\nu_1 - \nu_2$ when $\mu = 0.01$. The vertical dotted line are plotted at $\pm\mu$.

Using $f_\mu(\boldsymbol{\nu})$ to approximate $||C\boldsymbol{\nu}||_1$ in (3.3), we minimize the following objective instead:

$$h_\mu(\boldsymbol{\nu}) = -l(\boldsymbol{\nu}) + f_\mu(\boldsymbol{\nu}) \tag{3.7}$$

Unfortunately, calculating $-l(\boldsymbol{\nu})$ and $-\nabla l(\boldsymbol{\nu})$ involves $\mathcal{O}(\sum_{i=1}^{n} \sum_{A \in \mathcal{I}} N_i(A))$ time complexity. For high-throughput sequencing data, $N_i(A)$ could easily reach tens of thousands especially for A close to the root, leading to considerable amount computational cost for each round of iteration. We present an accurate approximation algorithm with $\mathcal{O}(n|\mathcal{I}|)$ time complexity in Section 3.6.

3.3 Minimizing the objective function

3.3.1 Accelerated gradient descent

Chen et al. (2012) applied FISTA (Beck and Teboulle, 2009) to optimize the objective func-

tion with an additional non-smooth l_1 sparsity penalty in the linear regression case:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^J} \tilde{f}(\boldsymbol{\beta}) = \frac{1}{2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}||^2 + f_\mu(\boldsymbol{\beta}) + \lambda_1||\boldsymbol{\beta}||_1 \tag{3.8}$$

In comparison, our optimization problem is the following:

$$\min_{\boldsymbol{\nu}\in\mathbb{R}^J_+} h_\mu(\boldsymbol{\nu}) = -l(\boldsymbol{\nu}) + f_\mu(\boldsymbol{\nu}) \tag{3.9}$$

where $\mathbb{R}^J_+$ is the space of positive real numbers with J dimensions.

There are a few major differences between our case and Chen et al. (2012). First of all, because of the absence of the non-smooth sparseness penalty on $\boldsymbol{\nu}$ in our objective function, there is no need to use the proximal gradient. Therefore, FISTA becomes the accelerated gradient descent in Nesterov (1983). Second, our problem has a positivity constraint $\boldsymbol{\nu} \in \mathbb{R}^J_+$. As a result, we add a simple line search to make sure that at each iteration step, the current estimate of $\boldsymbol{\nu}$ satisfy the aforementioned constraint. Third, the non-convexity of $-l(\boldsymbol{\nu})$ means that the solution to (3.9) may not be unique. Last, the Lipschitz constant of $h_\mu(\boldsymbol{\nu})$, which is used to determine the step size in each round of iteration, is hard to calculate in close form due to presence of $l(\boldsymbol{\nu})$. However, we find that using $L^{(t)} = \lambda_{\max}\big(-\nabla^2 l(\boldsymbol{\nu}^{(t)})\big) + L_\mu$, where $\lambda_{\max}(\cdot)$ extracts the largest eigenvalue of the matrix in the parenthesis and ∇^2 is the Hessian operator, can successfully decrease the objective function value when the initial estimate is chosen by warm start. In the case of linear regression, the above definition of $L^{(t)}$ is simply $\lambda_{\max}(\boldsymbol{X}'\boldsymbol{X}) + L_\mu$, which coincides with the one used in Chen et al. (2012).

The accelerated gradient descent algorithm to optimize (3.9) is presented as follows:

Algorithm 1 Accelerated gradient descent for smoothed fusion penalty on DTM dispersion parameters

---

Input: Count data $\boldsymbol{x}$, fusion penalty λ, initial estimate $\boldsymbol{\nu}^{(0)}$, desired accuracy ϵ and line

search parameters ρ_ν, $\rho_w \in (0,1)$.

Initialization: Set $\mu = \epsilon/|\tilde{\mathcal{I}}|$, C according to (3.4), $L_\mu = ||C||^2/\mu$ and $\boldsymbol{w}^{(0)} = \boldsymbol{\nu}^{(0)}$.

Iterate: For $t = 0, 1, 2, ...$, do the following until $\boldsymbol{\nu}^{(t)}$ converges:

1. Compute $\nabla h_\mu(\boldsymbol{\nu}^{(t)})$ using (3.6) and the appendix. Also, calculate $L^{(t)} = \lambda_{\max}\big(-\nabla^2 l(\boldsymbol{\nu}^{(t)})\big) + L_\mu$.

2. Find the smallest positive integer z_ν such that $\boldsymbol{\nu}^{(t+1)} = \boldsymbol{w}^{(t)} - \rho_\nu^{z_\nu} \nabla h_\mu(\boldsymbol{\nu}^{(t)})/L^{(t)} \in \mathbb{R}^J_+$

3. Find the smallest positive integer z_w such that $\boldsymbol{w}^{(t+1)} = \boldsymbol{\nu}^{(t+1)} + \rho_w^{z_w} \frac{t}{t+3}(\boldsymbol{\nu}^{(t+1)} - \boldsymbol{\nu}^{(t)}) \in \mathbb{R}^J_+$

Output: $\hat{\boldsymbol{\nu}}_\lambda = \boldsymbol{\nu}^{(t+1)}$

Nesterov (1983) proved the following result for accelerated gradient descent. Let $f(\boldsymbol{\beta})$ be convex and smooth where $\nabla f(\boldsymbol{\beta})$ is Lipschitz continuous with Lipschitz constant L. Then the accelerated gradient descent satisfies

$$f(\boldsymbol{\beta}^{(t)}) - f(\boldsymbol{\beta}) \leq \frac{2L||\boldsymbol{\beta}^{(0)} - \boldsymbol{\beta}||^2}{t^2} \tag{3.10}$$

for any $\boldsymbol{\beta}$. In our case, however, the non-convexity of the objective function along with the usage of line search to enforce the positivity constraint makes (3.10) inapplicable. Furthermore, we have avoided explicitly calculating the Lipschitz constant of $h_\mu(\boldsymbol{\nu})$ and uses a heuristic instead. Despite these differences, we observe that the accelerated gradient descent is still much faster than the naive gradient descent.

Since $f_\mu(\boldsymbol{\nu})$ is a smooth approximation to $||C\boldsymbol{\nu}||_1$, the optimal solution $\hat{\boldsymbol{\nu}}_\lambda$ from Algorithm 1 will not yield a sparse $C\hat{\boldsymbol{\nu}}_\lambda$. Therefore, a threshold $D > 0$ must be chosen such that any elements in $C\hat{\boldsymbol{\nu}}$ that are within $[-D, D]$ are treated as if there were precisely zero. Here we provide a heuristic to determine the value of D. For simplicity of illustration, let ζ_μ be the set of indices in $C\hat{\boldsymbol{\nu}}_\lambda$ that are treated as if they are zero. Also, define $[\cdot]_i$ to be the operator

that extracts the ith element. When $\mu = \epsilon/|\tilde{\mathcal{I}}| = 0$, we have $\boldsymbol{\alpha}^* = \text{sign}(C\hat{\boldsymbol{\nu}}_\lambda)$ according to (3.6). In this case, ζ_0 is unambiguously defined as $\{i : [\boldsymbol{\alpha}^*]_i = 0\}$. Since elements in $\boldsymbol{\alpha}^*$ can only take values in $\{-1, 0, 1\}$, we can alternatively write $\zeta_0 = \{i : [\boldsymbol{\alpha}^*]_i \in (-\theta, \theta)\}$ for some fixed value of $\theta \in (0, 1]$. When $\mu > 0$, we have $\boldsymbol{\alpha}^* = S(C\hat{\boldsymbol{\nu}}_\lambda/\mu)$ and the above definition of ζ_μ becomes

$$\zeta_\mu = \{i : [S(\frac{C\hat{\boldsymbol{\nu}}_\lambda}{\mu})]_i \in (-\theta, \theta)\} = \{i : [C\hat{\boldsymbol{\nu}}_\lambda]_i \in (-\mu\theta, \mu\theta)\}$$

Therefore, it is clear that we should set $D = \mu\theta$. In practice, we simply choose $\theta = 1$, which leads to $D = \mu$.

Algorithm 1 can be further modified by replacing $L^{(t)}$ in step 2 with step size R that is dynamically determined by another backtracking line search algorithm (Chen et al., 2012). This can help accelerating the initial descent especially if L_μ is very large. Specifically, let

$$Q_R(\boldsymbol{\nu}; \boldsymbol{w}^{(t)}) = h_\mu(\boldsymbol{w}^{(t)}) + (\boldsymbol{\nu} - \boldsymbol{w}^{(t)})'\nabla h_\mu(\boldsymbol{w}^{(t)}) + \frac{R}{2}||\boldsymbol{\nu} - \boldsymbol{w}^{(t)}||^2$$

Apparently, the minimizer of Q_R is $\boldsymbol{\nu}^{(t+1)} = \boldsymbol{w}^{(t)} - \nabla h_\mu(\boldsymbol{w}^{(t)})/R$. To guarantee the convergence rate of Algorithm 1, we need the following condition to hold:

$$h_\mu(\boldsymbol{\nu}^{(t+1)}) \leq Q_R(\boldsymbol{\nu}^{(t+1)}; \boldsymbol{w}^{(t)}) = h_\mu(\boldsymbol{w}^{(t)}) - \frac{||\nabla h_\mu(\boldsymbol{w}^{(t)})||^2}{2R} \tag{3.11}$$

Therefore, we could start from setting R to be a small positive constant and $\rho_R > 1$ as the backtracking parameter. At iteration t, we repeatedly set $R \leftarrow \rho_R R$ until (3.11) is satisfied. This line search is used in conjunction with searching for z_v to ensure that $\boldsymbol{\nu}^{(t+1)} \in \mathbb{R}^J_+$.

Using line search to determine R can help speed up the initial descent since the local Lipschitz constant of $\nabla f_\mu(\boldsymbol{\nu})$ within a connected set B can be much smaller than the global Lipschitz constant, L_μ. The reason behind this claim is that $S(x) = \max(\min(x, 1), -1)$ is constant for $|x| \geq 1$. In particular, consider three different cases of B:

1. $B_1 = \{\boldsymbol{\nu} : |[C\boldsymbol{\nu}]_i| \leq \mu \text{ for all } i\} \Rightarrow |\zeta_\mu| = |\tilde{\mathcal{I}}|$

2. $B_2 = \{\boldsymbol{\nu} : [C\boldsymbol{\nu}]_i > \mu \text{ for all } i\} \Rightarrow |\zeta_\mu| = 0$

3. $B_3 = \{\boldsymbol{\nu} : [C\boldsymbol{\nu}]_i < -\mu \text{ for all } i\} \Rightarrow |\zeta_\mu| = 0$

In the first case, $\boldsymbol{\alpha}^* = S(C\boldsymbol{\nu}/\mu) = C\boldsymbol{\nu}/\mu$ according to (3.6) and the definition of the thresholding function $S(\cdot)$. Therefore, $\nabla f_\mu(\boldsymbol{\nu}) = C'C\boldsymbol{\nu}/\mu \Rightarrow \nabla f_\mu(\boldsymbol{\nu})$ has Lipschitz constant $||C||^2/\mu = L_\mu$ within B_1. In the other two cases, however, we have $\boldsymbol{\alpha}^* = \mathbf{1}$ (case 2) or $\boldsymbol{\alpha}^* = -\mathbf{1}$ (case 3). Therefore, $\nabla f(\boldsymbol{\nu}) = C'\mathbf{1}$ is constant within B_2 or B_3, hence its local Lipschitz constant being zero. Obviously, the more elements of $\boldsymbol{\alpha}^*$ being constant within B, the smaller local Lipschitz constant of $\nabla f_\mu(\boldsymbol{\nu})$ becomes. During initial rounds of iteration, it is very likely that most elements of $C\boldsymbol{\nu}^{(t)}$ are much larger than μ or much smaller than $-\mu$. Therefore, $1/L_\mu$ is too small as the step size for gradient descent, and the line search is useful to speed up the convergence.

3.3.2 Partial Newton-Raphson method

The objective function in (3.7) contains a negative log likelihood that is second-order differentiable and a smoothed fusion penalty that is not second-order differentiable. It is worth investigating whether we can use the Hessian of log likelihood to speed up the convergence.

For the log likelihood, we use a second-order expansion on $l(\boldsymbol{\nu})$ around $\boldsymbol{\nu}^{(t)}$ to find the estimate at next step:

$$\boldsymbol{\nu}_l^{(t+1)} = \text{argmin}_{\boldsymbol{\nu}} - \big(l(\boldsymbol{\nu}^{(t)}) + (\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})'\nabla l(\boldsymbol{\nu}^{(t)}) + \frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})'\nabla^2 l(\boldsymbol{\nu}^{(t)})(\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})\big) \quad (3.12)$$

where ∇^2 is the Hessian operator. Setting the gradient of (3.12) to zero yields $\boldsymbol{\nu}_l^{(t+1)} = \tilde{\boldsymbol{\nu}}^{(t)} - \big(\nabla^2 l(\boldsymbol{\nu}^{(t)})\big)^{-1}\nabla l(\boldsymbol{\nu}^{(t)})$, which is the same as the Newton-Raphson update.

For the smoothed fusion penalty, we optimize the following function for a given L:

$$\boldsymbol{\nu}_{f_\mu}^{(t+1)} = \text{argmin}_{\boldsymbol{\nu}}\big(f_\mu(\boldsymbol{\nu}^{(t)}) + (\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})'\nabla f_\mu(\boldsymbol{\nu}^{(t)}) + \frac{L}{2}||\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)}||_2^2\big) \quad (3.13)$$

Setting the gradient (3.13) to zero yields $\boldsymbol{\nu}_{f_\mu}^{(t+1)} = \boldsymbol{\nu}^{(t)} - \nabla f_\mu(\boldsymbol{\nu}^{(t)})/L$, which is equivalent to gradient descent with step size L.

Since the objective function $h_\mu(\boldsymbol{\nu})$ is just the sum of the negative log likelihood and smoothed fusion penalty, we combine (3.12) and (3.13) to get

$$\begin{aligned}\boldsymbol{\nu}^{(t+1)} = \text{argmin}_{\boldsymbol{\nu}}\Big(\big(-l(\boldsymbol{\nu}^{(t)}) + f_\mu(\boldsymbol{\nu}^{(t)})\big) + (\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})'(-\nabla l(\boldsymbol{\nu}^{(t)}) + \nabla f_\mu(\boldsymbol{\nu}^{(t)})) \\ + \frac{1}{2}(\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})'\big(-\nabla^2 l(\boldsymbol{\nu}^{(t)}) + LI_J\big)(\boldsymbol{\nu} - \boldsymbol{\nu}^{(t)})\Big) \end{aligned} \tag{3.14}$$

where $I_J \in \mathbb{R}^{J\times J}$ is an identity matrix. Setting the gradient of (3.14) to zero gives

$$\boldsymbol{\nu}^{(t+1)} = \boldsymbol{\nu}^{(t)} - \big(-\nabla^2 l(\boldsymbol{\nu}^{(t)}) + LI_J\big)^{-1}\big(-\nabla l(\boldsymbol{\nu}^{(t)}) + f_\mu(\boldsymbol{\nu}^{(t)})\big) \tag{3.15}$$

L can be fixed as $L = L_\mu$ or dynamically determined by a line search algorithm. Due to the same arguments provided in the last paragraph of the previous section, we use the line search algorithm to select L. Since this method only uses Hessian matrix for part of the objective function, we call it the partial Newton-Raphson method.

3.3.3 Combining partial Newton-Raphson with accelerated gradient descent

Obviously, the above algorithm is similar to Newton-Raphson when L is small, which makes the objective function descend much faster than the gradient-based algorithms. As L increases to $L \gg \lambda_{\max}(-\nabla^2 l(\boldsymbol{\nu}^{(t)}))$, however, the update in (3.15) becomes more similar to a naive gradient descent with step size L. In such case, the partial Newton-Raphson algorithm is inferior to the accelerated gradient descent (Algorithm 1). Therefore, it is important to distinguish when to use the partial Raphson or accelerated gradient descent. Suppose $\hat{\boldsymbol{\nu}}$ is the optimal solution and let $B_{\hat{\boldsymbol{\nu}}}$ be a small neighborhood around $\hat{\boldsymbol{\nu}}$,. There are two possibilities:

1. If $|[C\hat{\boldsymbol{\nu}}]_i| > \mu$ for all i (which most likely happens when λ is small), then $\hat{\boldsymbol{\alpha}}^* = S(C\hat{\boldsymbol{\nu}}/\mu)$

is constant within which means that $\nabla f_\mu(\boldsymbol{\nu})$ has local Lipschitz constant 0 within $B_{\hat{\boldsymbol{\nu}}}$. Therefore, the partial Newton-Raphson will behave very similar to the raw Newton-Raphson method.

2. If $\exists i$ such that $|[C\hat{\boldsymbol{\nu}}]_i| < \mu$, then the local Lipschitz $\nabla f_\mu(\boldsymbol{\nu})$ within $B_{\hat{\boldsymbol{\nu}}}$ is on the order of λ^2/μ. If μ is very small, then $\left(l(\boldsymbol{\nu}^{(t)}) + LI_J\right)^{-1} \approx (LI_J)^{-1}$. As a result, the partial Newton-Raphson will behave more similar to the gradient descent method.

We use a simple example to illustrate the different convergence behavior between partial Newton-Raphson method and the accelerated gradient descent. We simulate a DTM dataset using the following tree with 6 leaves $\Omega = \{1, 2, 3, 4, 5, 6\}$. The internal node configuration $\mathcal{I}$ and the true value of dispersion parameters on each node are set as in Figure 3.3. In this case, there are three nodes, namely $\{1, 2, 3\}$, $\{4, 5, 6\}$ and $\{5, 6\}$, that satisfy the constraint in (3.2).

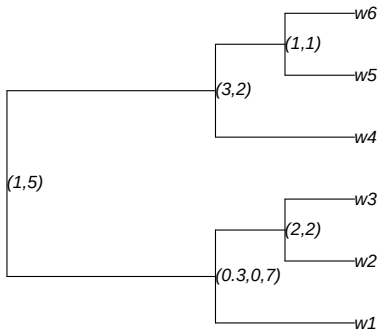

Figure 3.3: Phylogenetic tree used for simulation and its dispersion parameters. For each internal node, the first and second number in the parenthesis denote the dispersion for its left child and right child, respectively.

We use the above configuration to simulate 100 microbial samples, each of which has 1000 counts in total. Then, we optimize (3.3) at $\mu = 10^{-4}$ with three different choices of $\lambda \in \{0.5, 1, 2\}$. The reason we choose these values of λ is that $|\zeta_\mu| = 0, 1$ and 2 respectively

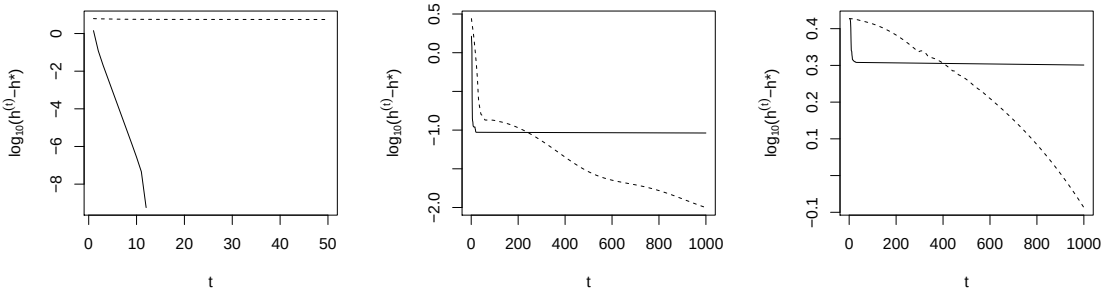

Figure 3.4: Comparison of convergence speed between accelerated gradient descent and partial Newton-Raphson method where t is the step size, $h^{(t)} = h(\boldsymbol{\nu}^{(t)})$ and $h^* = h(\hat{\boldsymbol{\nu}}_\lambda)$. Here $\lambda = 0.5$ (left), $\lambda = 1$ (middle) and $\lambda = 2$ (right). Solid line denotes partial Newton-Raphson method, and dashed line denotes accelerated gradient descent with line search for the step size.

in these three cases. In each scenario, two different strategies are used to optimize the objective function: 1) accelerated gradient descent (Algorithm 1) but with additional line search for step size, and 2) partial Newton-Raphson. The logarithm of difference between $h^{(t)} = h(\boldsymbol{\nu}^{(t)})$ and $h^* = h(\hat{\boldsymbol{\nu}}_\lambda)$ is plotted against time step t in Figure 3.4. The initial estimate $\boldsymbol{\nu}^{(0)}$ is chosen by warm start, i.e. $\boldsymbol{\nu}^{(0)} = \hat{\boldsymbol{\nu}}_{\tilde{\lambda}}$ where $\tilde{\lambda}$ is the largest available penalization parameter that is smaller than the current one.

We can clearly observe the difference of convergence rate between partial Newton-Raphson and accelerated gradient descent in each of three cases. When $\lambda = 0.5$, we have $|\zeta_\mu| = 0$, which means that the local Lipschitz constant of $\nabla f_\mu(\boldsymbol{\nu})$ is zero within a small neighborhood of $\hat{\boldsymbol{\nu}}_{0.5}$. Therefore, the partial Newton-Raphson method becomes the original Newton-Raphson, resulting in a much faster convergence rate than accelerated gradient descent. When $\lambda = 1$, however, the fact that $|\zeta_\mu| > 0$ makes the partial Newton Raphson method quickly turn into gradient descent at around $t = 15$, converging very slowly afterwards. The accelerated gradient descent, albeit initially slower than partial Newton-Raphson, quickly catches up at around $t = 300$. Similar pattern is observed at $\lambda = 2$. These demonstrates the necessity of choosing the right algorithm to minimize the objective function.

In practice, we don't know a priori the relative scale between L_μ and $\lambda_{\max}(-\nabla^2 l(\hat{\boldsymbol{\nu}}_\lambda))$ before obtaining $\hat{\boldsymbol{\nu}}_\lambda$. Therefore we use a short pilot-run of the partial Newton-Raphson method. If the line search parameter L is larger than a predefined multiple of $\lambda_{\max}(-\nabla^2 l(\boldsymbol{\nu}^{(t)}))$, we switch to accelerated gradient descent. This yields the Algorithm 2 below:

Algorithm 2 Combining partial Newton-Raphson method with accelerated gradient descent

---

Input: Count data $\boldsymbol{x}$, fusion penalty λ, initial estimate $\boldsymbol{\nu}^{(0)}$, desired accuracy ϵ, line search parameter $\rho_h \in (0,1)$ and $\rho_L > 1$, initial $L > 0$, threshold $U > 0$ for the relative multitude between L and $\lambda_{\max}(-\nabla^2 l(\boldsymbol{\nu}^{(t)}))$.

Iterate: For $t = 0, 1, 2, ...t_{\max}$, do the following until $\nu^{(t)}$ converges:

1. Compute $\nabla l(\boldsymbol{\nu}^{(t)}), \nabla^2 l(\boldsymbol{\nu}^{(t)})$ and $\nabla f_\mu(\boldsymbol{\nu}^{(t)})$ using (3.6) and the appendix.
2. Calculate the descent direction $\Delta\boldsymbol{\nu} = \big(-\nabla^2 l(\boldsymbol{\nu}^{(t)}) + LI_J\big)^{-1}\big(-\nabla l(\boldsymbol{\nu}^{(t)}) + f_\mu(\boldsymbol{\nu}^{(t)})\big)$
3. Find the smallest positive integer z_h such that $\boldsymbol{\nu}^{(t+1)} = \boldsymbol{\nu}^{(t)} - \rho_h^{z_h}\Delta\boldsymbol{\nu} \in \mathbb{R}_+^J$
4. If $h(\boldsymbol{\nu}^{(t+1)}) \geq h(\boldsymbol{\nu}^{(t)})$, set $L \leftarrow \rho_L L$ and go back to step 2.
5. If $L > U\lambda_{\max}(-\nabla^2 l(\boldsymbol{\nu}^{(t)}))$, terminate and switch to Algorithm 1.

Output: $\hat{\boldsymbol{\nu}}_\lambda = \boldsymbol{\nu}^{(t+1)}$.

3.4 Empirical results

In this section we provide use simulations as well as a real dataset to demonstrate the profiles of the fusion penalty as well as the cross-validated log likelihood as λ varies. For simulation, we use the same setup as Figure 3.3, with $n = 100$, $\mu = 10^{-4}$ and 1000 sequences within each sample. We run Algorithm 2 to optimize (3.9) using $\lambda \in \{0.5, 1, 2, 4, 8, 16, 32, 64, 128\}$.

For each element in $C\hat{\boldsymbol{\nu}}_\lambda$, we plot its value against $\log_{10}(\lambda)$ in the left plot Figure 3.5 to demonstrate the solution path with regard to the strength of penalization. Each color corresponds to a particular element in $C\hat{\boldsymbol{\nu}}_\lambda$. To relate these colors to penalization on the nodes, recall that each element in $C\boldsymbol{\nu}$ is, by definition, analytically equivalent to $\nu_{R(A)s_A} - \sum_{j=1}^{J_A} \nu_{A_j}$ for a certain A. The correspondence between the colors and internal nodes are as follows: black $A = \{1, 2, 3\}$, red $A = \{2, 3\}$, blue $A = \{4, 5, 6\}$ and green $A = \{5, 6\}$. The right plot in Figure 3.5 is the negative cross-validated likelihood versus $\log_{10}(\lambda)$ using 5-folds cross validation. The cross-validated likelihood, $\tilde{l}(\cdot)$, is defined as the sum of the log likelihood on each test set (5 in total with equal sizes) using the optimal solution of (3.9) on respective training sets in each fold. Since we are only interested in the relative difference among $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda)$'s in order to choose the optimal λ, the y-axis is chosen as $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda) + \tilde{l}(\hat{\boldsymbol{\nu}}_0)$ to better display the numerical scale. At $\lambda = 2$, which yields the minimum of cross-validated likelihood, there are two nodes that have their fusion penalty value shrunken within $[-\mu, \mu]$: black $\{1, 2, 3\}$ and $\{4, 5, 6\}$. Recall from our previous discussion that since we use a smooth approximation to the fusion penalty, elements of $C\hat{\boldsymbol{\nu}}_\lambda$ that fall within $[-\mu, \mu]$ are treated as if there were exactly zero. Therefore, the final selected model has 2 less parameters than the full DTM model. In comparison, the ground truth model has 3 less parameters than the full DTM model.

We next apply the same algorithm to the American Gut dataset introduced in the previous chapter. We use 104 fecal microbial samples from all males living in the Plains region with their 20 most common OTUs. This gives median of sequencing depth as 6301. Similar to Figure 3.5, we plot the profiles of fusion penalty as well as cross-validated likelihood, as λ varies, using American Gut data in Figure 3.6. The final selected model with the minimal cross-validated likelihood has $\lambda = 0.5$, corresponding to three of the elements in $C\hat{\boldsymbol{\nu}}_{0.5}$ within $[-\mu, \mu]$ and thus three less parameters than the full DTM model. Furthermore, we plot the values of $u_A = \nu_{R(A)s_A} - \sum_{j=1}^{J_A} \nu_{A_j}$ for each $A \in \tilde{\mathcal{I}}$, using the solution at $\lambda = 0.5$, on the phylogenetic tree in Figure 3.7. The three nodes with $u_A \in [-\mu, \mu]$ are identified as light

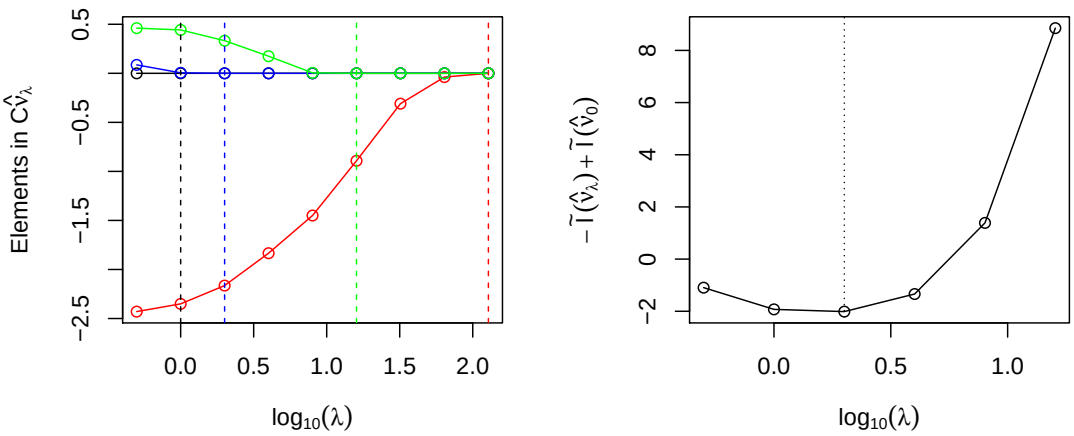

Figure 3.5: Profiles of fusion penalty as well as the cross-validated likelihood as λ varies on a simulation dataset. Left plot: solution path of elements in $C\hat{\boldsymbol{\nu}}_\lambda$ with regard to $\log_{10}(\lambda)$. Each color corresponds to a particular element in $C\hat{\boldsymbol{\nu}}_\lambda$. Vertical dashed lines are plotted according to the smallest λ such that the particular element in $C\hat{\boldsymbol{\nu}}_\lambda$ corresponding to that color shrinks to within $[-\mu, \mu]$. For example, the blue dashed line is at $\lambda = 2$, which means that the element in $C\hat{\boldsymbol{\nu}}_\lambda$ corresponding to blue color stays within $[\mu, \mu]$ for $\lambda \geq 2$. Right plot: $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda) + \tilde{l}(\hat{\boldsymbol{\nu}}_0)$ versus $\log_{10}(\lambda)$. Dotted line is plotted at the minimizer of $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda) + \tilde{l}(\hat{\boldsymbol{\nu}}_0)$, which equals to $\lambda = 2$ in this case.

blue squares on the tree.

3.5 Understanding DTM model selection through covariance structures

We end this chapter by presenting an alternative interpretation of DTM model selection through covariance representations. Since the tree-based decomposition of multinomial distribution (second part of (3.1)) is the same for both DM and DTM, we only focus on the Dirichlet part. The discussion here explores the dependence among K OTUs when the underlying compositions are generated by a Dirichlet distribution or a Dirichlet-tree distribution.

First of all, we introduce the concept of absolute abundance, which measures the actual

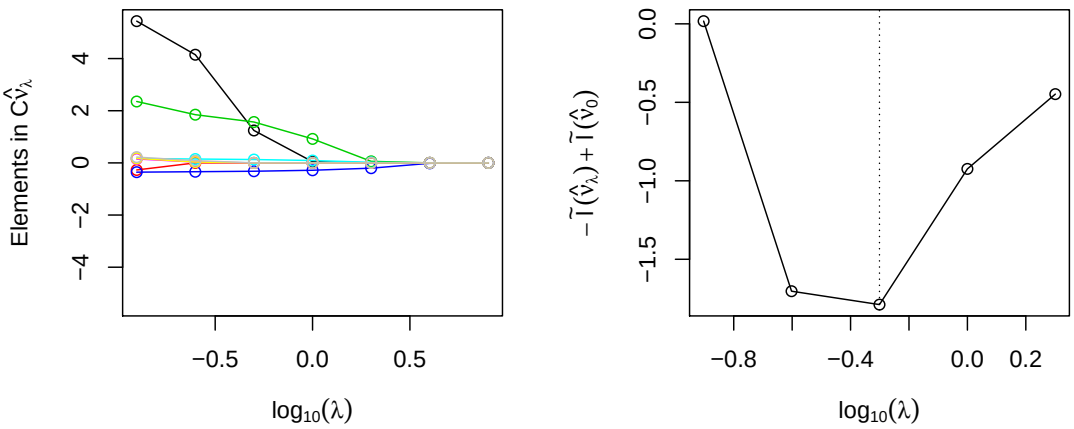

Figure 3.6: Profiles of fusion penalty as well as the cross-validated likelihood as λ varies using the American Gut dataset. Left plot: solution path of eight elements in $C\hat{\boldsymbol{\nu}}_\lambda$ with regard to $\log_{10}(\lambda)$. Only the eight elements that reaches within $[-\mu, \mu]$ prior to or at $\lambda = 2$ are plotted here. Each color corresponds to a particular element in $C\hat{\boldsymbol{\nu}}_\lambda$. Right plot: $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda) + \tilde{l}(\hat{\boldsymbol{\nu}}_0)$ versus $\log_{10}(\lambda)$. Dotted line is plotted at the minimizer of $-\tilde{l}(\hat{\boldsymbol{\nu}}_\lambda) + \tilde{l}(\hat{\boldsymbol{\nu}}_0)$, which equals to $\lambda = 0.5$ in this case.

density of bacterias on a certain sampling cite. The relation between absolute abundance and relative abundance is

$$\boldsymbol{q} = \frac{\boldsymbol{a}}{\sum_{k=1}^{K} a_k}$$

for all $i = 1, 2, ..., n$, where $\boldsymbol{q} = (q_1, q_2, ..., q_K)$ is the relative abundance and $\boldsymbol{a} = (a_1, a_2, ..., a_K)$ is the absolute abundance. Note that absolute abundance is not the same as sequence counts. The sum of absolute abundance depends on host-specific environments such as bio-availability of certain nutrients, whereas the sum of sequence counts depends on sequencing machine configurations. Typically it is very hard to directly observe absolute abundance, and for high throughput sequencing data, such value is regarded as completely unobserved.

Measuring the raw covariance among compositional data is subject to negative bias due to the unit sum constraint $\sum_{k=1}^{K} q_k = 1$. A detailed investigation into such spurious dependence can be found at Friedman and Alm (2012). As a result, these authors use the following

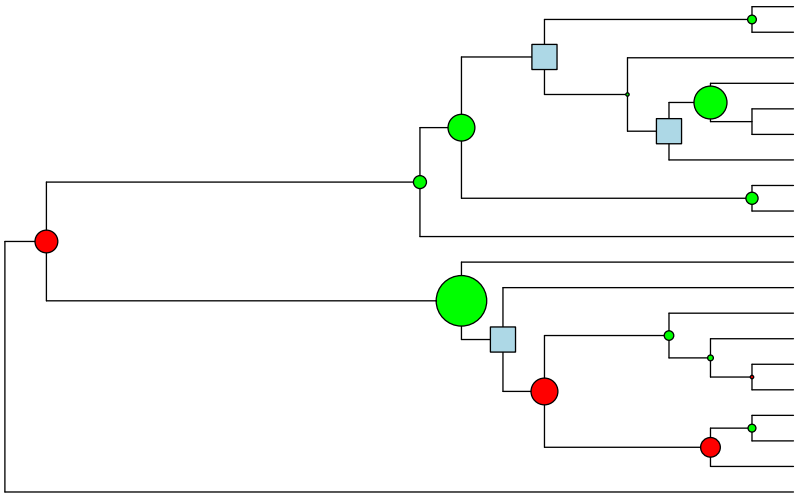

Figure 3.7: Visualization of fusion penalty for each internal node using the optimal solution from cross-validated likelihood, i.e. $\lambda = 0.5$. The size, shape and color of the object on internal node $A \in \mathcal{I}$ is determined by its fusion penalty $u_A = \nu_{R(A)s_A} - \sum_{j=1}^{J_A} \nu_{A_j}$. There are three different possibilities: 1) if $u_A > \mu$, then the object is a red circle with its size proportional to $\log(1 + u_A)$; 2) if $u_A < -\mu$, then the object is a green circle with its size proportional to $\log(1 - u_A)$; 3) if $|u_A| \leq \mu$, then the object is a light blue square with constant size. A total of three blue squares are visible from the plot.

quantity instead

$$\sigma_{sr} = \mathrm{Cov}(\log a_s, \log a_r) \tag{3.16}$$

for $1 \leq s, r \leq K$. The absence of unit-sum constraint on the absolute abundance avoids the negative spurious covariances. In addition, the value of (3.16) is unchanged when $\boldsymbol{a}$ is multiplied by an arbitrary constant. This avoids the identifiability issue since one can only observe $\boldsymbol{q}$, which is also invariant to multiplicative transform on $\boldsymbol{a}$. There is a relationship between the sample covariance of $\log \boldsymbol{a}$ and $\log \boldsymbol{q}$ using fact that $\log a_r - \log a_s = \log q_r - \log q_s$, which, together with sparsity assumptions on $\{\sigma_{rs}\}$, can lead to covariance estimates as in SparCC (Friedman and Alm, 2012) and CCLasso (Fang et al., 2015).

Now we explore the value of (3.16) when the compositional data is generated from a Dirichlet or Dirichlet-tree model. For ease of illustration, we use the same tree structure as in Figure 3.1 so that $K = 4$. For Dirichlet distribution, it is well known that

$$\boldsymbol{q} \sim \text{Dir}(\boldsymbol{\nu}) \Leftrightarrow a_k \sim \text{Gamma}(\nu_k, 1) \text{ independently for } 1 \leq k \leq 4 \tag{3.17}$$

where $\boldsymbol{\nu} = (\nu_1, ..., \nu_4)$. Therefore, $\sigma_{sr} = 0$ from the independence of a_k's. The scaling parameter of the gamma distributions can be any positive value other than 1 as long as they are the same for all a_k's, and they have no impact on the covariance after log transform.

For Dirichlet-tree distribution, $\boldsymbol{q}$ can be alternatively expressed as follows:

1. Node A_1: $a_1 + a_2 = g\theta_{A_1}$ and $a_3 + a_4 = g(1 - \theta_{A_1})$, where $g \sim \text{Gamma}(\nu_{A_1 1} + \nu_{A_1 2}, 1)$ and $\theta_{A_0} \sim \text{Dir}(\nu_{A_1 1}, \nu_{A_1 2})$.
2. Node A_2: $a_1 = (a_1 + a_2)\theta_{A_2}$, where $\theta_{A_2} \sim \text{Dir}(\nu_{A_2 1}, \nu_{A_2 2})$.
3. Node A_3: $a_3 = (a_3 + a_4)\theta_{A_3}$, where $\theta_{A_3} \sim \text{Dir}(\nu_{A_3 1}, \nu_{A_3 2})$.
4. $g, \theta_{A_1}, \theta_{A_2}$ and θ_{A_3} are all independent.

Altogether, they lead to the following representation on absolute abundances:

1. $a_1 = g\theta_{A_1}\theta_{A_2}$
2. $a_2 = g\theta_{A_1}(1 - \theta_{A_2})$
3. $a_3 = g(1 - \theta_{A_1})\theta_{A_3}$
4. $a_4 = g(1 - \theta_{A_1})(1 - \theta_{A_3})$

Before we calculate the values of σ_{sr}, we need the following facts. Let $\psi(\cdot)$ be the first-order polygamma function, which is defined as the first order derivative of the logarithm of gamma function. Then 1) $g \sim \text{Gamma}(\alpha, 1) \Rightarrow \text{Var}(\log g) = \psi(\alpha)$; 2) $\theta \sim \text{Dir}(\nu_1, \nu_2) \Rightarrow \text{Var}(\log \theta_i) = \psi(\nu_1) - \psi(\nu_1 + \nu_2)$ and $\text{Cov}\big(\log \theta, \log(1 - \theta)\big) = -\psi(\nu_1 + \nu_2)$.

Now we can easily deduce the covariances as follows:

$$\begin{aligned}\sigma_{13} = \sigma_{14} = \sigma_{23} = \sigma_{24} &= \mathrm{Var}(\log g) + \mathrm{Cov}\big(\log\theta_{A_1}, \log(1-\theta_{A_1})\big) \\ &= \psi(\nu_{A_1 1} + \nu_{A_1 2}) - \psi(\nu_{A_1 1} + \nu_{A_1 2}) = 0 \end{aligned} \tag{3.18}$$

$$\begin{aligned}\sigma_{12} &= \mathrm{Var}(\log g) + \mathrm{Var}(\log\theta_{A_1}) + \mathrm{Cov}\big(\log\theta_{A_2}, \log(1-\theta_{A_2})\big) \\ &= \psi(\nu_{A_1 1} + \nu_{A_1 2}) + \big(\psi(\nu_{A_1 1}) - \psi(\nu_{A_1 1} + \nu_{A_1 2})\big) - \psi(\nu_{A_2 1} + \nu_{A_2 2}) \\ &= \psi(\nu_{A_1 1}) - \psi(\nu_{A_2 1} + \nu_{A_2 2})\end{aligned}$$

$$\begin{aligned}\sigma_{34} &= \mathrm{Var}(\log g) + \mathrm{Var}\big(\log(1-\theta_{A_1})\big) + \mathrm{Cov}\big(\log\theta_{A_3}, \log(1-\theta_{A_3})\big) \\ &= \psi(\nu_{A_1 1} + \nu_{A_1 2}) + \big(\psi(\nu_{A_1 2}) - \psi(\nu_{A_1 1} + \nu_{A_1 2})\big) - \psi(\nu_{A_3 1} + \nu_{A_3 2}) \\ &= \psi(\nu_{A_1 2}) - \psi(\nu_{A_3 1} + \nu_{A_3 2})\end{aligned}$$

The implications of Dirichlet-tree distribution on covariances is now clear. Take σ_{12} as an example. When $\nu_{A_1 1} = \nu_{A_2 1} + \nu_{A_2 2}$, which means that the equality condition in (3.2) is satisfied for A_2, we have $\sigma_{12} = 0$, same as Dirichlet model. When $\nu_{A_1 1} < \nu_{A_2 1} + \nu_{A_2 2}$, however, $\sigma_{12} > 0$ since $\psi(\cdot)$ is monotonically decreasing on the positive axis. Similarly $\sigma_{12} < 0$ when $\nu_{A_1 1} > \nu_{A_2 1} + \nu_{A_2 2}$. Therefore, the covariance σ_{12} is inherently controlled by the difference between $\nu_{A_1 1}$ and $\nu_{A_2 1} + \nu_{A_2 2}$. This is where the extra degrees of freedom in Dirichlet-tree distribution contribute to. These pairwise covariances will have a much more complicated form for a large tree, but in general, we can regard the full Dirichlet-tree model as an $|\mathcal{I}| - 1$ degrees of freedom representation of all the $K(K-1)/2$ off-diagonal elements on the covariance matrix $\{\sigma_{rs}\}$. Testing Dirichlet vs Dirichlet-tree distribution is therefore equivalent to finding best covariance representation when the degrees of freedom varies from 0 (fully independent) to $|\mathcal{I}| - 1$.

The covariance interpretation of Dirichlet-tree model has other interesting implications.

For example, we can maximize the Dirichlet-tree likelihood over not only the dispersion parameters but also the tree structure. A special property of Dirichlet-tree induced covariance is that if there exists a common ancestor node A of four certain nodes $(A_{r_1}, A_{s_1}, A_{r_2}, A_{s_2})$ such that $A_{r_1}, A_{r_2} \in \mathcal{C}(A)_1$ and $A_{s_1}, A_{s_2} \in \mathcal{C}(A)_2$, then $\sigma_{r_1 s_1} = \sigma_{r_2 s_2}$. In other words, the cross-covariance between any of A's left nodes and any of A's right nodes is the same for fixed A. When $A = \Omega$ as the root of the tree, then such cross-covariance is simply zero as we showed in (3.18). The optimal tree structure that maximizes the likelihood function therefore induces the best hierarchy of binary partitions with constant cross-covariance at each level. The resulting partition can reveal whether the relations among groups of bacterial taxa, at various levels, are dominated by mutual benefits or competitive exclusion. However, finding the best tree structure is a combinatorial optimization problem that is computationally prohibitive. We can use similar approximation techniques in the field of phylogeny inference (Felsenstein and Felenstein, 2004; Price et al., 2009; Stamatakis, 2014). For example, a candidate tree is first constructed by a certain agglomerative clustering method, and the topology is further refined by local searches such as nearest neighbor interchange.

3.6 Appendix: DTM fast approximation

By definition of DTM likelihood,

$$l(\boldsymbol{\nu}; \boldsymbol{x}) = \sum_{i=1}^{n} \sum_{A \in \mathcal{I}} l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big)$$

where $l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big)$ is the log likelihood on node A for the ith sample:

$$l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big) = \sum_{j=1}^{J_A} \sum_{\xi=0}^{x_{ij}(A)-1} \log(\nu_{Aj} + \xi) - \sum_{\xi=0}^{N_i(A)-1} \log(\nu_{A\cdot} + \xi) \tag{3.19}$$

up to an irrelevant constant, with $\nu_{A\cdot} = \sum_{j=1}^{J_A} \nu_{Aj}$. Due to the summation operation, evaluating each $l_{ij}(\theta_A)$ involves $\mathcal{O}(N_i(A))$ computation cost. To reduce this, notice that

all summations in (3.19) take the form $g_0(\alpha, k) = \sum_{\xi=0}^{k-1} \log(\alpha + \xi)$. We use Taylor series approximation to fast compute $g_0(\alpha, k)$ for arbitrary values of α and k based on expansion on integer grids.

To start, let $[\alpha]$ denote the closest integer to α and define $\epsilon = \alpha - [\alpha]$. Also, let $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ be the floor and ceiling operators, respectively. When $[\alpha] + \xi > 0$, $\log(\alpha + \xi)$ can be expanded as

$$\log(\alpha + \xi) = \log([\alpha] + \xi) + \frac{\epsilon}{[\alpha] + \xi} - \frac{\epsilon^2}{2([\alpha] + \xi)^2} + ... \tag{3.20}$$

In order to simply the demonstration, suppose only quadratic expansion is used. We first calculate the values of the following three functions for all $k = 2, ..., N_i(A)$ and store them into memory:

$$S_0(k) = \sum_{\xi=1}^{k-1} \log(\xi), \;\; S_1(k) = \sum_{\xi=1}^{k-1} \frac{1}{\xi}, \;\; S_2(k) = \sum_{\xi=1}^{k-1} \frac{1}{\xi^2} \tag{3.21}$$

Next, we choose an integer T such that the approximation (3.20) is invoked only when $\alpha + \xi \geq T$. This turns the original function value $g_0(\alpha, k)$ into

$$\begin{aligned} g_0(\alpha, k) = & \sum_{\xi=0}^{\lfloor T-\alpha \rfloor} \log(\alpha + \xi) + \sum_{\xi=\lceil T-\alpha \rceil}^{k-1} \log(\alpha + \xi) \\ \approx & \sum_{\xi=0}^{\lfloor T-\alpha \rfloor} \log(\alpha + \xi) + \big(S_0([\alpha] + k) - S_0([\alpha] + \lceil T - \alpha \rceil)\big) + \epsilon\big(S_1([\alpha] + k) \\ & - S_1([\alpha] + \lceil T - \alpha \rceil)\big) - \frac{\epsilon^2}{2}\big(S_2([\alpha] + k) - S_2([\alpha] + \lceil T - \alpha \rceil)\big) \end{aligned}$$

assuming that α is not an integer. With $\mathcal{O}(N_i(A))$ memory complexity, calculating $g_0(\alpha, k)$ only has $\mathcal{O}(1)$ time complexity. As a result, (3.19) is approximated by

$$l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big) \approx \big(\sum_{j=1}^{J_A} g(\nu_{Aj}, x_{ij}(A))\big) - g\big(\nu_A, N_i(A)\big)$$

In practice, we choose $T = 10$ and fourth order Taylor expansion, which yields relative error

less than 10^{-8},

The gradient of (3.19) can be calculated using a similar approximation.

$$\frac{\partial l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big)}{\partial \nu_{Aj}} = \sum_{\xi=0}^{x_{ij}(A)-1} \frac{1}{\nu_{Aj}+\xi} - \sum_{\xi=0}^{N_i(A)-1} \frac{1}{\nu_A+\xi} \tag{3.22}$$

$$\frac{\partial^2 l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big)}{\partial \nu_{Aj}^2} = -\sum_{\xi=0}^{x_{ij}(A)-1} \frac{1}{(\nu_{Aj}+\xi)^2} + \sum_{\xi=0}^{N_i(A)-1} \frac{1}{(\nu_A+\xi)^2} \tag{3.23}$$

$$\frac{\partial^2 l_A\big(\boldsymbol{\nu}_A; \boldsymbol{x}_i(A)\big)}{\partial \nu_{Aj}\partial \nu_{Aj'}} = \sum_{\xi=0}^{N_i(A)-1} \frac{1}{(\nu_A+\xi)^2} \tag{3.24}$$

The techniques for calculating the log likelihood introduced above are still applicable to gradient calculation. Each term inside (3.22) is of the form $g_1(\alpha, k) = \sum_{\xi=0}^{k-1} 1/(\alpha+\xi)$. Each term inside (3.23) and (3.24) is of the form $g_2(\alpha, k) = \sum_{\xi=0}^{k-1} 1/(\alpha+\xi)^2$. Using the same expansion on integer grids yields

$$\frac{1}{\alpha+\xi} = \frac{1}{[\alpha]+\xi} - \frac{\epsilon}{([\alpha]+\xi)^2} + \frac{\epsilon^2}{([\alpha]+\xi)^3} + ...$$

$$\frac{1}{(\alpha+\xi)^2} = \frac{1}{([\alpha]+\xi)^2} - \frac{2\epsilon}{([\alpha]+\xi)^3} + \frac{3\epsilon^2}{([\alpha]+\xi)^4} + ...$$

as long as $[\alpha]+\xi > 0$. The rest details are omitted.

CHAPTER 4

MICROBIOME COMMUNITY HERITABILITY BY A VARIANCE COMPONENT MODEL USING WISHART DISTRIBUTION

4.1 Introduction

Genetic variation can lead to differences in food preferences, enzyme activity or immune response, hence having a nontrivial impact on microbial compositions. Therefore, quantifying the contribution of overall genetic effects to microbiome variation is of great scientific interest. In most studies of microbiome heritability, different taxon abundances are treated as separate phenotypes, each of which is analyzed individually through a standard statistical model such as the Additive Genetics, Common Environment, Unique Environment (ACE) variance component model (Eaves et al., 1978). Heritability for each taxon is then defined as the proportion of phenotypic variance that can be explained by the variance of genotypic values. The ACE variance component model has long been used for heritability estimation in family studies and recently applied to unrelated individuals (Yang et al., 2011) for a number of univariate traits such as height (Yang et al., 2010), inflammatory bowel diseases (Chen et al., 2014) and diabetes (Bonnefond and Froguel, 2015). However, the principal shortcoming of applying ACE variance model to each individual taxon is that it ignores taxa relatedness, as the model is unable to differentiate whether the change in a certain taxon's relative abundance is caused by its constituent taxa that are phylogenetically similar or different. Furthermore, van Opstal and Bordenstein (2015) argues that this approach leads to an unidirectional interpretation of genetics-microbiome interaction, in which host genetics regulates colonization. Considering that microbiome is a collection of organisms, each with its unique interplay with the host genetics, a more comprehensive view requires capturing the entire microbial community with the human host. Consequently, it is advisable to directly

measure the community heritability, which quantifies the contribution of genetic variation towards compositional differences, i.e. beta diversity, between microbiome communities.

Calculating community heritability based on beta diversity has a major difficulty in statistical modeling since the response variable is a matrix of pairwise dissimilarity. To comply with the traditional heritability models, ordination methods such as non-metric multidimensional scaling (NMDS) and principal coordinate analysis (PCoA) must be applied to transform the dissimilarity matrix into into a univariate response that best preserves the original pairwise distance. Obviously, different ordination standards can lead to different heritability results. In addition, the recovered univariate response usually represents only a fraction of total variation in the dissimilarity matrix and has unclear biological meanings. These difficulties altogether point to the necessity of a statistical model capable of decomposing total variation in a dissimilarity matrix into genetics and environmental components without any transformation.

A crucial property of a dissimilarity or distance matrix, as pointed out by Gower (1966), is that it can be transformed into an outer product matrix. This outer product matrix can be conveniently modeled by the Wishart distribution, which has a straightforward analogy to the univariate ACE model, hence definition of heritability, by imposing a similar additive form on the covariance matrix parameter. However, Wishart distribution is only applicable when the response is a positive definite matrix, a requirement not satisfied by most dissimilarity measurements that are ecologically meaningful. A major contribution of this chapter is that we prove this property for a particular beta-diversity measurement, the square root transformation of weighted Unifrac. Unifrac (Lozupone and Knight, 2005; Lozupone et al., 2007) incorporates phylogenetic information among bacterial species and has been extensively applied to a number of microbiome studies. To the best of our knowledge, no prior work exists to model the entire variation in Unifrac matrix for heritability analysis. Our result provides an easily justifiable statistical framework, essentially answering the following question: to what extent does genetically similar subjects carry phylogenetically similar

microbial communities? We also provide an extension to Wishart ACE model that directly incorporates the effect of sequencing noise on beta diversity, thus avoiding the need to rarefy microbiome samples.

The rest of this chapter is organized as follows. Section 4.2 introduces the Wishart variance component model with ACE formulation. Section 4.3 proves that the square root of weighted Unifrac is applicable for the Wishart distribution. Section 4.4 provides empirical results using TwinsUK fecal microbiome data and from simulation. Section 4.5 concludes this paper with further discussions.

4.2 Wishart distribution with variance components

We start from reviewing the ACE variance component model (A for additive genetics, C for common environment and E for unique environment) for heritability analysis on univariate traits (Eaves et al., 1978). This models assumes additive random effects from genetic factors, common environments and unique environments. Let n be the number of samples, $\boldsymbol{y}$ be an $n \times 1$ vector of their univariate traits, $\boldsymbol{X}$ be a $n \times m$ matrix of covariates such as age, sex and weight, and $\boldsymbol{\beta}$ be an $m \times 1$ vector of fixed effects. Furthermore, let $\boldsymbol{A}$ be an $n \times n$ genetic relationship matrix (GRM), $\boldsymbol{C}$ be an $n \times n$ matrix that quantifies shared environments, and $\boldsymbol{E} = \boldsymbol{I}_n$ be an $n \times n$ identity matrix for the unique environment effects. The GRM quantifies additive genetic covariance among individuals. In familial studies, $\boldsymbol{A}$ is twice the kinship matrix. For example, $\boldsymbol{A}_{i,j} = 1$ for monozygotic twins and $\boldsymbol{A}_{i,j} = 1/2$ for dizygotic twins. Furthermore, $\boldsymbol{C}_{i,j} = 1$ if and only if ith and jth individual share the same household. The diagonal entries of $\boldsymbol{A}$ and $\boldsymbol{C}$ are all set to one. In genome wide association studies on unrelated individuals, $\boldsymbol{A}$ can be estimated by SNP data (Yang et al., 2011) and the shared environment matrix is usually omitted.

The ACE variance component model takes the following form:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{c} + \boldsymbol{e}, \quad \boldsymbol{g} \sim N(0, \sigma_A^2 \boldsymbol{A}), \quad \boldsymbol{c} \sim N(0, \sigma_C^2 \boldsymbol{C}), \quad \boldsymbol{e} \sim N(0, \sigma_E^2 \boldsymbol{E}) \tag{4.1}$$

where $\boldsymbol{g}$, $\boldsymbol{c}$ and $\boldsymbol{e}$ are assumed to be mutually independent.

Heritability (h) is defined as the proportion of total variance that is due to genetic factors:

$$h = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_C^2 + \sigma_E^2} \tag{4.2}$$

A common approach to estimate $\sigma^2 = (\sigma_A^2, \sigma_C^2, \sigma_E^2)$ and hence h uses residual maximum likelihood (REML) (Yang et al., 2011; Zhou and Stephens, 2012). Let $\boldsymbol{L}$ be the $(n-m) \times n$ matrix with its rows spanning the kernel space of X'. Left multiplying (4.1) by $\boldsymbol{L}$ leads to $\boldsymbol{Ly} \sim N(\boldsymbol{0}, \boldsymbol{L\Sigma L'})$ where $\boldsymbol{\Sigma} = \sigma_A^2\boldsymbol{A} + \sigma_C^2\boldsymbol{C} + \sigma_E^2\boldsymbol{E}$, after which one maximizes its likelihood to obtain REML estimates $\hat{\sigma}^2$ and $\hat{h}$. The REML likelihood takes the following form:

$$l(\sigma^2; \boldsymbol{y}) = -\frac{n-m}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\boldsymbol{y'L'\Sigma^{-1}Ly} \tag{4.3}$$

We are interested in extending the ACE framework to the case where we can only observe an outer product matrix or covariance matrix, instead of the raw values of the univariate traits. To start, notice that the ACE model implies that

$$E(\boldsymbol{Lyy'L'}) = \boldsymbol{L\Sigma L'} \tag{4.4}$$

where $\boldsymbol{yy'}$ serves as a sample outer product matrix. Now suppose we can only observe an observed outer product matrix $\boldsymbol{M}$ but not $\boldsymbol{y}$. This happens when one analyze a dataset by measuring its pairwise dissimilarities and apply principal coordinate analysis (details provided in the next subsection). Since both $\boldsymbol{yy'}$ and $\boldsymbol{M}$ have the same interpretation, an analogy to (4.4) would be to assume

$$E(\boldsymbol{LML'}) = \boldsymbol{L\Sigma L'} \tag{4.5}$$

A similar analogy is used by McArdle and Anderson (2001) to derive a pseudo F-statistic to

test fixed effects when the observation is a pairwise dissimilarity matrix. The effect of $\boldsymbol{L}$ is, similar to the univariate case, to remove the fixed effects in the outer product matrix $\boldsymbol{M}$. One can interpret this by expanding $\boldsymbol{M}$ into the sum of rank 1 matrices: $\boldsymbol{M} = \sum_{i=1}^{\text{rank}(M)} \boldsymbol{M}_i \boldsymbol{M}_i'$ with $\boldsymbol{M}_i \in \mathbb{R}^n$, and imposing that $E(\boldsymbol{M}_i) = \boldsymbol{X}\boldsymbol{\beta}_i$.

In particular, (4.5) suggests that we can use a Wishart distribution to model $\boldsymbol{Z} = \boldsymbol{L}\boldsymbol{M}\boldsymbol{L}'$: $\boldsymbol{Z} \sim W(\boldsymbol{T}/q, q)$ where $\boldsymbol{T} = \boldsymbol{L}\boldsymbol{\Sigma}\boldsymbol{L}'$ in order to align with the form of expectation in (4.5). Without affecting heritability estimates, we can further remove q from the scale matrix and simply write $\boldsymbol{Z} \sim W(\boldsymbol{T}, q)$, leading to the following log likelihood:

$$\begin{aligned} l(q, \sigma^2; \boldsymbol{Z}) = &-\frac{q}{2}\log|\boldsymbol{T}| - \frac{1}{2}\text{tr}(\boldsymbol{T}^{-1}\boldsymbol{Z}) + \frac{q-(n-m)-1}{2}\log|\boldsymbol{Z}| - \frac{q(n-m)}{2}\log 2 \\ &- \log\Gamma_{n-m}(\frac{q}{2}) \end{aligned} \tag{4.6}$$

where $\Gamma_{n-m}(\cdot)$ is the multivariate gamma function and q can be any real number larger than $n - m - 1$. Maximizing (4.6) leads to the MLEs $(\hat{q}, \hat{\sigma}^2)$ and hence $\hat{h}$ from (4.2) by using $\hat{\sigma}^2$. The gradient of (4.6) with respect to σ^2 is very similar to the case of ACE model with normal distribution, and the partial derivate of q is straightforward to obtain. We use gradient based optimization, such as L-BFGS (Liu and Nocedal, 1989), to obtain the MLEs $(\hat{q}, \hat{\sigma}^2)$.

The log likelihood (4.6) is only applicable when $\boldsymbol{Z}$ is positive definite. The next subsection will prove this condition when $\boldsymbol{Z}$ is calculated from a particular microbiome beta-diversity metric.

4.3 Community heritability by root-Unifrac and Wishart distribution

4.3.1 Root-Unifrac and positive definiteness

Unifrac (Lozupone and Knight, 2005; Lozupone et al., 2007) is one of the most popular metrics to quantify pairwise dissimilarities among microbial communities. A common way to incorporate Unifrac into the ACE model (4.1) is to apply principal coordinate analysis (PCoA) on the $n \times n$ Unifrac dissimilarity matrix and then use each of the principal eigenvectors separately as a univariate response (Goodrich et al., 2014, 2016; Quigley et al., 2017). Specifically, let $u(i, j)$ be the Unifrac dissimilarity between ith and jth sample, and $\boldsymbol{D}$ satisfying $\boldsymbol{D}_{i,j} = -u(i, j)^2/2$. Also, define $\boldsymbol{J} = \boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n'/n$ where $\mathbf{1}_n$ is a unit vector of length n. PCoA first calculates Gower's centered matrix as $\boldsymbol{M} = \boldsymbol{JDJ}$ (Gower, 1966) , which turns a dissimilarity matrix into a centered outer product matrix. This is because if $u(i, j)$ is the Euclidean distance between the pair of vectors $\boldsymbol{c}_i$ and $\boldsymbol{c}_j$ for $1 \leq i \leq n$, then it is easy to deduce that

$$\boldsymbol{M} = \boldsymbol{JCC'J}, \quad \boldsymbol{C} = \begin{pmatrix} \boldsymbol{c}_1' \\ \boldsymbol{c}_2' \\ \cdots \\ \boldsymbol{c}_n' \end{pmatrix}$$

After this step, one applies principal component analysis (PCA) on $\boldsymbol{M}$ to obtain the principal eigenvectors to use as univariate traits.

Although the principal eigenvectors of $\boldsymbol{M}$ can quantify community information to some extent, they are hard to interpret and usually express only a fraction of the total variation in the Unifrac matrix. An alternative is to directly use $\boldsymbol{M}$ as the observation, which has been applied to nonparametrically testing fixed effects (McArdle and Anderson, 2001) and widely used in a number of microbial studies (Chen et al., 2012; Wang et al., 2016). Given the nature of $\boldsymbol{M}$ being an outer product matrix, it is reasonable to model $\boldsymbol{Z} = \boldsymbol{LML'}$ as generated

from a Wishart distribution, allowing us to maximize (4.6) to obtain the heritability estimate. However, the central difficulty is that $\boldsymbol{Z}$ may not be positive definite and therefore its log determinant in (4.6) can be undefined.

In this section, we present our result stating that using the square root transformation of the weighted Unifrac (Lozupone et al., 2007), which we call root-Unifrac, will guarantee that $\boldsymbol{LML'}$ is positive definite under a very relaxed condition. Suppose that we have a rooted phylogenetic tree with K branches. Let b_k be the length of kth branch and $p_{i,k}$ be the taxa proportions descending from the branch kth in the ith sample for $1 \leq i \leq n$ and $1 \leq k \leq K$. For 16S rRNA data clustered into operational taxonomic units (OTUs), $p_{i,k}$ is calculated by the sum of relative abundances of all OTUs under branch k. Let $\boldsymbol{x}_i = (x_{i1}, x_{i2}, ..., x_{in_o})$ denote the number of sequences belonging to each of the n_o OTUs in the ith microbial sample, then the OTU relative abundances $\boldsymbol{\theta}_i$, is $\boldsymbol{x}_i / \sum_{o=1}^{n_o} x_{io}$. The root-Unifrac is defined as

$$u(i,j) = \sqrt{\sum_{k=1}^{K} b_k |p_{i,k} - p_{j,k}|} \tag{4.7}$$

Similar to the original Unifrac, the root-Unifrac takes phylogenetic information and account for taxa relatedness while comparing different communities. It is simple to show that the root-Unifrac satisfies non-negativity, symmetry and triangle inequality. Therefore, we can define a finite metric space (Ω, u) where $\Omega = (\omega_1, \omega_2, ..., \omega_n)$ correspond to the n microbial samples and $u(\omega_i, \omega_j) = u(i,j)$ according to (4.7). We prove the positive definiteness of $\boldsymbol{LML'}$ by showing that (Ω, u) has an isometric embedding into an Euclidean space with dimension at least $n-1$, as long as there exists a k^* such that $\{p_{i,k^*}\}_i$ are all different. We will need the following results from Morgan (1974):

Definition 1. (Morgan, 1974) *Consider an ordered tuple* $(\kappa_0, \kappa_1, ..., \kappa_N)$ *whose elements are from a metric space* (Ω, d). *Define an* $N \times N$ *matrix* $\boldsymbol{V}$ *such that* $\boldsymbol{V}_{i,j} = \big(d^2(\kappa_i, \kappa_0) + d^2(\kappa_j, \kappa_0) - d^2(\kappa_i, \kappa_j)\big)/2$. (Ω, d) *is called flat if* $|\boldsymbol{V}| \geq 0$ *for any ordered tuple. Furthermore, the dimension of* (Ω, d), *provided that it is flat, is the largest number* N *such that there exists*

a tuple of size $N+1$ *with* $|\boldsymbol{V}|>0$.

Theorem 3. (Morgan, 1974) *A metric space can be embedded into an* n *dimensional Euclidean space if and only if the metric space is flat and of dimension less than or equal to* n.

Our main results are the following:

Theorem 4. *Let* $u(i,j)=\sqrt{\sum_{k=1}^{K} b_k|p_{i,k}-p_{j,k}|}$ *and define an* $n\times n$ *matrix* $\boldsymbol{D}$ *satisfying* $\boldsymbol{D}_{i,j}=-u(i,j)^2/2$. *Then the Gower's centered matrix,* $\boldsymbol{M}=\boldsymbol{JDJ}$, *is positive semidefinite with rank* $n-1$ *as long as* $\exists k^*\in\{1,2,...,K\}$ *such that* $\{p_{i,k^*}\}_i$ *are all different.*

Corollary 1. *Assume that the covariate matrix* $\boldsymbol{X}$ *includes the intercept. Under the condition of Theorem 4,* $\boldsymbol{LML'}$ *is positive definite.*

We present the proofs of Theorem 4 and Corollary 1 in the Appendix. Corollary 1 guarantees the applicability of Wishart likelihood to $\boldsymbol{Z}=\boldsymbol{LML'}$. In addition, Theorem 4 shows that there will be no negative eigenvalues in $\boldsymbol{M}$, hence no imaginary coordinates present in PCoA.

Our proof of positive definiteness only uses the fact that $u(i,j)$ is the square root of sum of absolute values. Therefore, it is still applicable if we only sum over a subset of branches for the distance. This is useful when we are only concerned with the portion of community difference that comes from a particular taxa such as Firmicutes or Bacteroidetes. Suppose we have R taxa in total ranging from kingdom to genus level. For the rth taxa, let $\mathcal{T}_r\subset\{1,2,...,K\}$ consists of the branches with all of its descendant OTUs belonging to the rth taxa. Define

$$u_{\mathcal{T}_r}(i,j)=\sqrt{\sum_{k\in\mathcal{T}_r} b_k|p_{i,k}-p_{j,k}|} \tag{4.8}$$

as the root-Unifrac contributed by only the rth taxa. Using Corollary 1 and substituting $\boldsymbol{Z}=\boldsymbol{LML'}$ into the Wishart variance component model (4.6), we can obtain the a heritability estimate $\hat{h}_{\mathcal{T}_r}$ for each taxa.

4.3.2 Assessing significance and confidence interval of heritability estimator

The p-value for testing the null hypothesis $H_0 : h = 0$ against $H_a : h > 0$ on the community heritability can be obtained by using likelihood ratio test or permutation. In the latter case, we randomly permute the rows and columns of the GRM matrix $\boldsymbol{A}$ for a total of n_{perm} rounds and record heritability estimates $\hat{h}^{(1)}, \hat{h}^{(2)}, ..., \hat{h}^{(n_{\text{perm}})}$ in each round. Then the p-value is simply defined as

$$\frac{|\{i : \hat{h}^{(i)} > \hat{h}\}|}{n_{\text{perm}}}$$

Since Schweiger et al. (2016) shows that asymptotic-based methods yield inaccurate coverage probability in the normal distribution model, we use bootstrapping to obtain nonparametric confidence interval (CI) for family based studies. When $\boldsymbol{A}$ is estimated by pedigree information and $\boldsymbol{C}$ does not contain shared environment between different families, observations across different families are assumed independent in our Wishart model. This implies that the observed Unifrac matrix is generated hierarchically in which each family is regarded as a group. Therefore, we can apply classical bootstrap techniques for hierarchical data by resampling families (groups). Since Theorem 4 requires existence of an edge that have different abundance in all samples, we need to further resample the sequences within each sample. Altogether, this is very similar to the two-stage bootstrap as described in Davison and Hinkley (1997).

Formally, suppose there are n_f families in total and, for simplicity, that the number of observations within each family are all equal to n_e. For illustration purposes, we take $n_e = 2$, which is the most common case of twin-based studies. Next, reorder the microbiome samples $\{\boldsymbol{x}_j\}_j$ so that the ith family corresponds to $\boldsymbol{x}_{2i-1}$ and $\boldsymbol{x}_{2i}$. For each bootstrap round $b = 1, 2, ..., n_{\text{boot}}$, execute the following steps:

1. Sample with replacement from the set $\{1, 2, ..., n_f\}$ for n_f times to obtain $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_{n_f})$ as the vector of resampled family labels.

2. Build the bootstrap GRM matrix $\boldsymbol{A}^*$ by setting setting all its diagonal elements equal to one and $\boldsymbol{A}^*_{2i-1,2i} = \boldsymbol{A}^*_{2i,2i-1} = \boldsymbol{A}_{2\lambda_i-1,2\lambda_i}$ for all $1 \leq i \leq n_f$. All other entries of $\boldsymbol{A}^*$ are set to zero. Building the bootstrap common environment matrix $\boldsymbol{C}^* = \boldsymbol{C}$.

3. Construct the bootstrapped microbial samples $\boldsymbol{x}^*_1, \boldsymbol{x}^*_2, ..., \boldsymbol{x}^*_n$ in the following way. For each $i = 1, 2, ..., n_f$, let $\boldsymbol{x}^*_{2i-1}$ be the result of sampling with replacement from $\boldsymbol{x}_{2\lambda_i-1}$ while keeping the same sequencing depth. Similarly, let $\boldsymbol{x}^*_{2i}$ be the result of sampling with replacement from $\boldsymbol{x}_{2\lambda_i}$ while keeping the same sequencing depth.

4. Calculate the root-Unifrac according to (4.7) or (4.8) using $\{\boldsymbol{x}^*_i\}_i$. Apply Wishart variance component model using $\boldsymbol{A}^*$ and $\boldsymbol{C}^*$ and estimate the bootstrap heritability.

Let $\hat{h}^*$ be the vector of all bootstrapped estimates of heritability. Then the $1-\alpha$ level confidence interval is constructed as

$$\left(\hat{h} - z_{1-\alpha/2}\text{se}(\hat{h}^*), \hat{h} + z_{1-\alpha/2}\text{se}(\hat{h}^*)\right) \tag{4.9}$$

where $\text{se}(\hat{h}^*)$ is the sample standard error of $\hat{h}^*$ and z_α is the α quantile of normal distribution.

The coverage probability of this confidence interval can be evaluated by simulation. Here the original microbiome samples $\{\boldsymbol{x}_i\}_i$ are regarded as population and their heritability estimate is regarded as ground truth. We randomly select a small portion, such as a half, of this population and build the bootstrap CI based on this portion of data. Repeat this process for a given number of times, and the simulated coverage probability (SCP) is simply the fraction of times that these bootstrap CIs contain the ground truth heritability. These SCPs are reported on a real microbiome dataset in Table 4.1.

4.4 Empirical results

4.4.1 Heritability estimates from TwinsUK

Goodrich et al. (2014) examined the influence of host genetics on fecal microbiome from a large twin-based study (TwinsUK). The TwinsUK population has more than 1000 16S rRNA microbial samples including 416 twin pairs. These sequences are processed by QIIME v1.9.1 (Caporaso et al., 2010) to produce the OTUs at 97% similarity level and the phylogenetic tree. Samples with sequencing depth less than 10000 are discarded. We do not apply any rarefaction or subsampling before calculating the taxon abundances. Since the overwhelming majority of observations are from females (1061 females vs 20 males), we remove all male observations to avoid too much variability in the sex effect. In the case of longitudinal observations for the same individual, only the first observation is used. This leaves 186 dizygotic (DZ) and 126 monozygotic (MZ) twin pairs. Similar to Goodrich et al. (2014), OTUs that appear in less than 50% of the microbial samples are excluded. The total number of remaining OTUs is 705. We also introduce a pseudo count in each OTU in all samples.

We apply the aforementioned ACE model with Wishart distribution on these microbial samples using the following covariates: age, body mass index, identity of technician (two), sequencing run (16 instrument runs) and shipment batch (8 shipments). These technical covariates are chosen according to Goodrich et al. (2014). The root-Unifrac matrix is calculated by (4.8) only for those taxa with at least 5 descendant OTUs, since our main concern is on the community level. To eliminate the burden of multiple hypothesis testing, if a higher level taxon (e.g. phylum Firmicutes) has more than 95% of its sequences belonging to one of its lower level taxon (e.g. order Clostridia), then the higher order taxon is excluded. This leaves a total of 26 taxa, each with its own root-Unifrac dissimilarity matrix and heritability estimate.

As described in Section 4.3.2, we permute the rows and columns of $\boldsymbol{A}$ for 10^4 times to test the null hypothesis of non-zero heritability for each of these 26 taxa. A total of 6 taxa

have p-values smaller than the Bonferroni threshold at 0.05 global Type-I error. Notice that this is a conservative correction due to the correlations among taxa abundances. We further calculate the 95% bootstrap confidence interval using (4.9) for these significant taxa along with their simulated coverage probabilities (SCP) from randomly drawing half of the samples for 200 times, as described previously. These results are reported in Table 4.1. The Bifidobacterium genus is also reported with significant heritability in Goodrich et al. (2014), although these authors use the traditional scalar ACE model, i.e. optimize (4.1) with $\boldsymbol{y}$ equal to the taxa relative abundance after Box-Cox transformation. Other significant taxa in Goodrich et al. (2014) that are related to our findings include Ruminococcaceae genus and Clostridiaceae family.

Table 4.1: Heritability estimates, p-values, 95% bootstrap confidence intervals (CI) and simulated coverage probabilities (SCP) for taxa that are globally significant at 0.05 level under Bonferroni correction. Taxon names are provided at their finest (lowest) possible rank: kingdom (k), phylum (p), class (c), order (o), family (f) and genus (g). P-value and CI are computed using ACE method.

Taxa	$\hat{h}$	**P-value**	**CI**	**SCP**
Actinobacteria (p)	0.223	$< 10^{-4}$	(0.091, 0.355)	92%
Clostridiales (o)	0.110	$< 10^{-4}$	(0.071, 0.149)	97%
Christensenellaceae (f)	0.185	10^{-4}	(0.075, 0.294)	92%
Rikenellaceae (f)	0.149	2×10^{-4}	(0.044, 0.254)	96%
Ruminococcaceae (f)	0.093	$< 10^{-4}$	(0.049, 0.137)	96%
Bifidobacterium (g)	0.231	$< 10^{-4}$	(0.096, 0.366)	93%

In order to compare our result with traditional heritability analysis on univariate traits, we calculate the first three principal coordinates on the root-Unifrac matrix for each taxon reported in Table 4.1. Each principal coordinate is used as a separate univariate response and substituted into (4.3) for heritability estimate. We report these heritability estimates as well as proportion of variance explained in Table 4.2. Interestingly, Actinobacteria, Christensenellaceae, Rikenellaceae and Bifidobacterium have their principal coordinates explaining more than 40% of total variation along with non-trivial heritability estimates. One way to interpret this phenomenon is that these principal coordinates could represent variation in certain

genetically encoded traits that strongly impact the microbiome. For example, it has been established that Bifidobacterium is significantly associated with the LCT gene in two large cohort studies (Wang et al., 2016; Goodrich et al., 2016). LCT gene encodes an enzyme that is capable of decomposing lactose, which is also metabolized by Bifidobacterium. Therefore, the first principal coordinate of the root-Unifrac matrix on Bifidobacterium might recover the availability of lactose in the intestine. However, results from second and third principal coordinates are more ambiguous and harder to explain. Another difficulty is that top principal components can exhibit distinctive sinusoidal patterns when the covariance matrix has spatial autocorrelation (Novembre and Stephens, 2008). This property is most likely to be shared by principal coordinates, making it even harder to interpret their heritability in a biologically meaningful way.

Table 4.2: Heritability estimates of the first three principal coordinates of each root-Unifrac disimilarity matrix. Taxon names are provided at their finest (lowest) possible rank similar to Table 4.1. Proportion of variance explained (Var prop) is obtained by dividing a particular principal eigenvalue over the sum of all eigenvalues.

Taxa	$\hat{h}_1^{\mathrm{PC}}$	**Var prop**	$\hat{h}_2^{\mathrm{PC}}$	**Var prop**	$\hat{h}_3^{\mathrm{PC}}$	**Var prop**
Actinobacteria (p)	0.392	50.0%	0.077	14.7%	0.136	6.8%
Clostridiales (o)	0.244	9.1%	0.229	6.5%	0.174	4.9%
Christensenellaceae (f)	0.347	51.9%	0.104	15.3%	0.000	7.1%
Rikenellaceae (f)	0.347	44.2%	0.161	13.5%	0.000	6.5%
Ruminococcaceae (f)	0.000	10.4%	0.000	8.0%	0.244	6.7%
Bifidobacterium (g)	0.401	52.3%	0.09	15.2%	0.143	7.0%

4.4.2 Effect of sequencing noise

Calculating the Unifrac or root-Unifrac requires relative abundances as input. For each sample, these relative abundances are obtained by normalizing the ith taxa sequences $\boldsymbol{x}_i$ over their sum $N_i = \sum_{o=1}^{n_o} x_{io}$, the latter conventionally called library size or sequencing depth. This normalization step introduces an extra layer of data uncertainty that is not modeled by any of the variance components in the Wishart ACE model. As a result, estimates of σ_A^2, σ_C^2 and σ_E^2, hence heritability, can be biased. Larger sequencing depth will mostly likely lead

to small sequencing variability and thus reduce the bias in the ACE variance component estimates. Although it is hard to deduce the closed form of this bias, the fact that such noises caused by normalization are independent across samples can more likely lead to an inflated $\hat{\sigma}_E^2$ and thus a downwards biased $\hat{h}$.

Here we inspect the bias of heritability estimates caused by sequencing noise through simulation using the same TwinsUK dataset. For each sample i, we first calculate the observed relative abundance $\boldsymbol{\theta}_i = \boldsymbol{x}_i/N_i$. These relative abundances are treated as if they were the true relative abundances for simulation purpose. After this step, we obtain $\tilde{\boldsymbol{x}}_i$ by subsampling $\boldsymbol{x}_i$ down to sequencing depth ξ, where ξ is 2500, 5000, 7500 or 10000. The ith simulated relative abundance is therefore $\tilde{\boldsymbol{\theta}}_i = \tilde{\boldsymbol{x}}_i/\xi$. In each simulation round, we calculate Wishart heritability estimates using $\{\tilde{\boldsymbol{\theta}}_i\}_i$ for the root-Unifrac metric on the six significant taxa reported in Table 4.1. The ground truth heritability, on the other hand, is obtained by using $\{\boldsymbol{\theta}_i\}_i$ to calculate the root-Unifrac metric. For each value of ξ, a total of 100 simulation rounds are conducted. We demonstrate the boxplot of these simulated heritability estimates and compare them against the ground truth heritability (dashed line) in Figure 4.1. The negative bias of simulated heritability estimates is present in all cases, and they decrease to zero at increasing levels of ξ. At $\xi = 10000$, the simulated estimates are all very close to the ground truth, with error less than 0.01. Since the mean and standard deviation of actual sequencing depth in TwinsUK dataset is 56911 and 18461, respectively, we conclude that the negative biases on the heritability estimates reported in Table 4.1 are negligible.

4.4.3 Simulation of type-I error and power

We simulate microbiome data to compare the type-I error and power of our method with those from analyzing principal coordinates. The simulation dataset contains 50 MZ twins and 50 DZ twins (200 microbiome samples in total), with the set of OTUs defined to be the five OTUs belonging to Bifidobacterium genus in the TwinsUK dataset. Phylogenetic tree on these five OTUs is constructed by pruning the TwinsUK phylogenetic tree down to only these

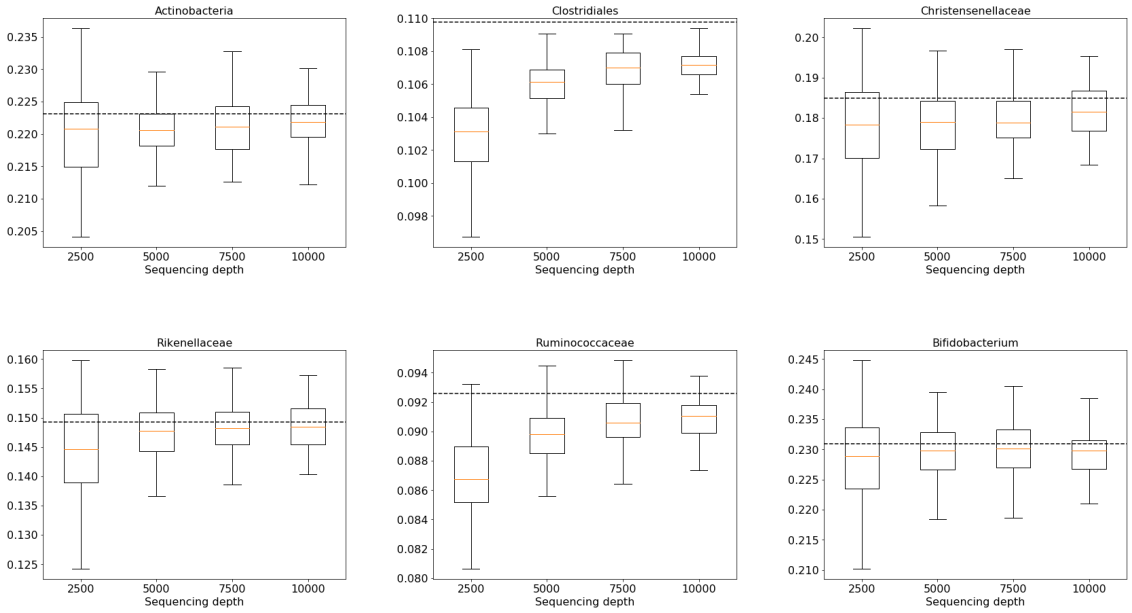

Figure 4.1: Boxplot of simulated heritability estimates from ACE models. In each round, the ith sequencing data are produced by subsampling $\boldsymbol{x}_i$ to a certain sequencing depth ξ. A total of 100 simulation rounds are conducted for each value of $\xi \in \{2500, 5000, 7500, 10000\}$. Dashed lines in each plot correspond to the ground truth heritability calculated using $\{\boldsymbol{\theta}_i\}_i$ as input data.

five OTUs. Since we conclude sequencing noise does not play a major impact in the previous section, we ignore it in this part and only simulate the relative abundance $\boldsymbol{\theta}_i$ for each i, which is assumed to follow independent and identical Dirichlet distribution with parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_5)$. This can be reparamatrized as $\boldsymbol{\alpha} = \alpha_+ \boldsymbol{\pi}$ with $\boldsymbol{\pi} = (\pi_1, ..., \pi_5)$ and $\sum_{o=1}^{5} \pi_o = 1$. Since we have $E(\boldsymbol{\theta}_i) = \boldsymbol{\pi}$, we set $\boldsymbol{\pi}$ to be proportional to the relative abundances of OTUs under Bifidobacterium genus calculated from all samples in TwinsUK, subject to the unit sum constraint on its components. The value of α_+ is fixed at 1.

Data generated under the null hypothesis of zero heritability is simply conducted by generating 200 microbiome samples from aforementioned Dirichlet distribution. For the alternative hypothesis, a signal strength parameter $\eta \in [0, 1]$ (smaller means stronger signal) is first chosen. After generating the Dirichlet distributed samples, we shrink each pair of relative abundances towards their family mean by a ratio of η for MZ twins and $\eta/2$ for

DZ twins. Evidently, $\eta = 0$ is equivalent to the null hypothesis case and $\eta = 1$ yields relative abundances within each MZ pair being the same. A total of three batches of data under the alternative hypothesis are generated with $\eta = 0.25, 0.33$ or 0.5. In each scenario, we produce 500 rounds of simulated data and apply our Wishart ACE model along with the traditional principal coordinate (PC) based methods (up to three PC), using the root-Unifarc for pairwise dissimilarities, to calculate heritability in each round. For all of these methods, p-values are calculated by permuting rows and columns of the genetic relationship matrix for 100 times according to the method in Section 4.3.2. Type-I error or power are obtained by counting the proportion of p-values that falls below 0.05 within these 500 rounds of simulation. The results are illustrated in Table 4.3. The Wishart method produces type-I error close to the nominal level 0.05, but all three PCs has lower type-I error than Wishart, suggesting that the test is more conservative than desired. PC-based methods also have much reduced power at all three levels of $\eta < 1$ compared to the Wishart method.

Table 4.3: Simulated type-I error and power from testing the null hypothesis of zero heritability, both calculated by the proportion of permutation p-values below 0.05 (nominal level). Larger level of η indicates greater signal strength. Under the Method column, Wishart indicates our method (4.6), and PC1-PC3 indicate using the traditional univariate method (4.3) on the first, second or third principal coordinates of the root-Unifrac matrix.

	Type-I error	**Power**		
Method	$\eta = 0$ (null)	$\eta = 0.25$	$\eta = 0.33$	$\eta = 0.5$
Wishart	0.042	0.186	0.308	0.744
PC1	0.02	0.112	0.164	0.492
PC2	0.038	0.084	0.114	0.304
PC3	0.022	0.1	0.156	0.354

4.5 Discussion

In this paper, we propose the Wishart variance component model to estimate microbiome community heritability when the microbiome data are summarized by their root-Unifrac dissimilarities. We rigorously prove that the root-Unifrac matrix always has an isometric

Euclidean embedding and therefore is adequate for REML estimation with the Wishart distribution. So far, almost all of the studies on community heritability rely on certain dimension reduction techniques to find the best univariate approximations to the pairwise dissimilarity matrix. Our work allows researchers to bypass this approximation step and directly analyze all the variations present in the dissimilarity matrix.

In Section 4.4.2 we inspected the negative biases of heritability estimates caused by sequencing noise. Although we concluded that such biases are negligible at the large sequencing depth of TwinsUK data, a better approach is to directly model the sequencing noise component as follows:

$$\boldsymbol{\Sigma} = \sigma_A^2 \boldsymbol{A} + \sigma_C^2 \boldsymbol{C} + \sigma_E^2 \boldsymbol{E} + \boldsymbol{S} \tag{4.10}$$

where $\boldsymbol{S} = \text{diag}(\sigma_{S_1}^2, \sigma_{S_2}^2, ..., \sigma_{S_n}^2)$ captures sequencing noise for each sample, and heritability is still defined as $h = \sigma_A^2/(\sigma_A^2 + \sigma_C^2 + \sigma_E^2)$. Therefore, this model makes it explicit that heritability is not dependent on sequencing noise.

Unfortunately, using (4.10) leads to identifiability issues among $\sigma_{S_i}^2$'s and σ_E^2. One possible way to avert this obstacle is to separately estimate $\sigma_{S_i}^2$ by exploring the variability of sequences within $\boldsymbol{x}_i$. Suppose the $\boldsymbol{\pi}_i$ is the true relative abundance for ith individual. If we can find a reasonable distribution to model $\boldsymbol{x}_i|\boldsymbol{\pi}_i$, then we can generate independent and identically distributed bootstrap samples, $\{\boldsymbol{x}_i^{[1]}, \boldsymbol{x}_i^{[2]}, ..., \boldsymbol{x}_i^{[B]}\}$, from such distribution by using $\hat{\boldsymbol{\pi}}_i = \boldsymbol{x}_i/N_i$. The sequencing depth of these bootstrap samples are kept at the same level at the original sample, i.e. $\boldsymbol{x}_i^{[b]} = (x_{i1}^{[b]}, ..., x_{in_o}^{[b]})$ and $\sum_{u=1}^{n_o} x_{io}^{[b]} = \sum_{u=1}^{n_o} x_{io}$ for all b. Since $\{\boldsymbol{x}_i^{[b]}\}_b$ share the same effect from covariates, genetics, common environment and unique environment, we can use a single intercept to model their total effect. This leaves the independent and identical sequencing noise the only remaining component that explains their variability:

$$\boldsymbol{L}_1 \boldsymbol{M}_i \boldsymbol{L}_1' \sim W(\sigma_{S_i}^2 \boldsymbol{L}_1 \boldsymbol{L}_1', q_i) \tag{4.11}$$

where $\boldsymbol{M}_i$ is the $B \times B$ Gower's centered matrix from calculating root-Unifrac on $\boldsymbol{x}_i^{[1]}, ..., \boldsymbol{x}_i^{[B]}$, and $\boldsymbol{L}_1$ is the $(B-1) \times B$ matrix that removes only the intercept effect. The estimated $\hat{\sigma}^2_{S_i}$ from maximizing the Wishart log likelihood of (4.11) can be used for (4.10), hence avoiding the identifiability issue.

4.6 Theorem proofs

We first prove the the following lemma:

Lemma 1. *For $a_1 > a_2 > ... > a_n > 0$,*

$$\det \begin{pmatrix} a_1 & a_2 & a_3 & ... & a_n \\ a_2 & a_2 & a_3 & ... & a_n \\ a_3 & a_3 & a_3 & ... & a_n \\ ... & ... & & & \\ a_n & a_n & a_n & ... & a_n \end{pmatrix} > 0$$

Proof. We prove by induction. Let $\boldsymbol{C}_i$ be the upper-left $i \times i$ corner of the matrix above. Evidently, $\det(\boldsymbol{C}_1) > 0$ and $\det(\boldsymbol{C}_2) > 0$.

Now assume that $\det(\boldsymbol{C}_{N-1}) > 0$ for some $N \geq 2$, we can write $\boldsymbol{C}_N$ as

$$\boldsymbol{C}_N = \begin{pmatrix} \boldsymbol{C}_{N-1} & a_N \mathbf{1}_{N-1} \\ a_N \mathbf{1}'_{N-1} & a_N \end{pmatrix}$$

Using the block formula for determinants, we have $\det(\boldsymbol{C}_N) = \det(\boldsymbol{C}_{N-1} - a_N \mathbf{1}_{N-1} \mathbf{1}'_{N-1}) a_N$. Notice that $\boldsymbol{C}_{N-1} - a_N \mathbf{1}_{N-1} \mathbf{1}'_{N-1}$ also assumes the form of $\boldsymbol{C}_{N-1}$ except that a_i is substituted by $a_i - a_N$ for $1 \leq i \leq N-1$. Since $a_1 - a_N > a_2 - a_N > ... > a_{N-1} - a_N$, we know from the induction assumption that $\det(C_{N-1} - a_N \mathbf{1}_{N-1} \mathbf{1}'_{N-1}) > 0$, and therefore $\det(C_N) > 0$. □

4.6.1 Proof of Theorem 4

Proof. We first prove that the metric space (Ω, u) has an isometric embedding into $n-1$ dimensional space by looking at each branch k separately. For an arbitrary value of $k \in \{1, 2, ..., K\}$, define $u_k(i,j) = \sqrt{b_k|p_{i,k} - p_{j,k}|}$. Obviously (Ω, u_k) is also a metric space. We shall prove that (Ω, u_k) has an isometric embedding into the Euclidean space. Using Theorem 3, we need to show the following two conditions are met:

1. *Flatness:* Take an arbitrary ordered tuple with size $N \leq n$ from (Ω, u_k). Without loss of generality, we assume that the tuple consists of the first N samples in Ω, i.e. $(\omega_1, \omega_2, ..., \omega_N)$. This means that the ith sample in the tuple has $p_{i,k}$ as its taxa proportion descending from branch k. According to Definition 1, $\boldsymbol{V}$ is defined as

$$\boldsymbol{V}_{i,j} = b_k\big(|p_{i+1,k} - p_{1,k}| + |p_{j+1,k} - p_{1,k}| - |p_{i+1,k} - p_{j+1,k}|\big)/2 \tag{4.12}$$

   For flatness we need to show $|\boldsymbol{V}| \geq 0$. There are three possibilities on $p_{i,k}$'s:

   (a) If there exists i such that $p_{i+1,k} = p_{1,k}$, then $\boldsymbol{V}_{i,j} = 0$ for all $j \Rightarrow |\boldsymbol{V}| = 0$.

   (b) If there exists i and j such that $p_{i+1,k} = p_{j+1,k}$, then the the ith and jth row of $\boldsymbol{V}$ are identical, leading to $|\boldsymbol{V}| = 0$.

   (c) If neither of the above is true, define a bijective sorting function $\tau : \{1, 2, ..., N-1\} \to \{1, 2, ..., N-1\}$ such that $p_{\tau(1)+1,k} < p_{\tau(2)+1,k} < \cdots < p_{\tau(N-1)+1,k}$. Furthermore, let $t = |\{p_{i+1,k} : p_{i+1,k} < p_{1,k} \text{ and } 1 \leq i \leq N-1\}|$.

   Let $\tilde{\boldsymbol{V}}$ be the matrix such that $\tilde{\boldsymbol{V}}_{i,j} = \boldsymbol{V}_{\tau(i),\tau(j)}$. Obviously $|\tilde{\boldsymbol{V}}| = |\boldsymbol{V}|$ and $\tilde{\boldsymbol{V}}$ is symmetric. Using (4.12) and the definition of τ, we see that the upper triangle of $\tilde{\boldsymbol{V}}$ satisfies the following properties:

   i. If $i = j$, then $\tilde{\boldsymbol{V}}_{i,j} = b_k|p_{\tau(i)+1,k} - p_{1,k}|$

   ii. If $i < j \leq t$, then $p_{\tau(i)+1,k} - p_{1,k} < p_{\tau(j)+1,k} - p_{1,k} < 0 \Rightarrow \tilde{\boldsymbol{V}}_{i,j} = b_k|p_{\tau(j)+1,k} - p_{1,k}|$

iii. If $i \leq t < j$, then $(p_{\tau(i)+1,k} - p_{1,k})(p_{\tau(j)+1,k} - p_{1,k}) < 0 \Rightarrow \tilde{\boldsymbol{V}}_{i,j} = 0$.

iv. If $t < i < j$, then $0 < p_{\tau(i)+1,k} - p_{1,k} < p_{\tau(j)+1,k} - p_{1,k} \Rightarrow \tilde{\boldsymbol{V}}_{i,j} = b_k|p_{\tau(i)+1,k} - p_{1,k}|$

Combining the above properties of $\tilde{\boldsymbol{V}}$, we can write it in block form:

$$\tilde{\boldsymbol{V}} = \begin{pmatrix} \tilde{\boldsymbol{V}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{V}}_2 \end{pmatrix}$$

where $\tilde{\boldsymbol{V}}_1 \in \mathbb{R}^{t \times t}$ and $\tilde{\boldsymbol{V}}_2 \in \mathbb{R}^{(N-1-t)\times(N-1-t)}$. According to Lemma 1, $|\tilde{\boldsymbol{V}}_1| > 0$ and $|\tilde{\boldsymbol{V}}_2| > 0$. Therefore, $|\tilde{\boldsymbol{V}}| = |\boldsymbol{V}| > 0$

2. *Minimum dimension* The minimum dimension of such embedding is simply the largest N such that $|\boldsymbol{V}| > 0$ for a certain tuple $(\kappa_0, \kappa_1, ..., \kappa_N)$ from (Ω, u_k). Notice that if all $p_{i,k}$'s are equal, then $N = 0$, leading to a trivial embedding into 0-dimensional space.

Now suppose k^* satisfies that p_{i,k^*} are all different for $1 \leq i \leq n$. According to the arguments above, (Ω, u_{k^*}) has an isometric embedding into an Euclidean space. Furthermore, the minimum dimension of such embedding is $n-1$ since $|\boldsymbol{V}| > 0$ for the tuple $(\omega_1, \omega_2, ..., \omega_n)$ due to the argument in 1(c).

So far we have proven the existence of Euclidean embedding for each (Ω, u_k). Let $\boldsymbol{\gamma}_{k1}, ..., \boldsymbol{\gamma}_{kn}$ be the Euclidean vectors that embeds (Ω, u_k) with minimum dimension. For each i, we define $\boldsymbol{\zeta}_i$ by concatenating all $\boldsymbol{\gamma}_{ki}$ for $1 \leq k \leq K$:

$$\boldsymbol{\zeta}_i = (\boldsymbol{\gamma}'_{1i}, \boldsymbol{\gamma}'_{2i}, ..., \boldsymbol{\gamma}'_{Ki})'$$

Since $u^2(i,j) = \sum_{k=1}^{K} u_k^2(i,j)$ for all i and j, it follows that $(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, ..., \boldsymbol{\zeta}_n)$ would be the embedded Euclidean vectors that preserve the metric u. Furthermore, $\text{rank}(\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, ..., \boldsymbol{\zeta}_n) \geq \text{rank}(\boldsymbol{\gamma}_{k^*1}, \boldsymbol{\gamma}_{k^*2}, ..., \boldsymbol{\gamma}_{k^*n}) = n-1$. It follows that the minimum dimension of (Ω, u)'s embedding is $n-1$. Now choose $\boldsymbol{\zeta}_1$ as the origin so that embedded vector of ith element becomes

$\tilde{\boldsymbol{\zeta}}_i = \boldsymbol{\zeta}_i - \boldsymbol{\zeta}_1$. Since $\text{rank}(\mathbf{0}, \tilde{\boldsymbol{\zeta}}_2, \tilde{\boldsymbol{\zeta}}_3, ..., \tilde{\boldsymbol{\zeta}}_n) = n-1$, we can orthogonally project them onto $\mathbb{R}^{n-1}$, hence the existence of an Euclidean embedding with $n-1$ dimensions.

Let $\boldsymbol{Q}$ be an $n \times (n-1)$ matrix with ith row denoting the $n-1$ dimensional embedding of ith element in (Ω, u). Furthermore, assume each column of $\boldsymbol{Q}$ has mean zero, which has no impact on the Euclidean distance induced by $\boldsymbol{Q}$. The arguments provided in the previous paragraph shows that $\text{rank}(\boldsymbol{Q}) = n-1$. By definition, $\boldsymbol{D}_{i,j} = -\sum_{z=1}^{n-1}(\boldsymbol{Q}_{iz} - \boldsymbol{Q}_{jz})^2/2$, so

$$\boldsymbol{M} = \boldsymbol{JDJ} = \boldsymbol{JQQ'J} = \boldsymbol{QQ'}$$

is positive semidefinite with rank $n-1$.

□

4.6.2 Proof of Corollary 1

Proof. Let $\boldsymbol{Q}$ be the same $n \times (n-1)$ matrix as defined above. For an arbitrary $\boldsymbol{v} \in \mathbb{R}^{n-m}$ and $\boldsymbol{v} \neq 0$, consider $\boldsymbol{v'LML'v} = \boldsymbol{(L'v)'M(L'v)}$. By definition of $\boldsymbol{L}$, we have $\mathbf{1}_n \in ker(\boldsymbol{L}) = im(\boldsymbol{L'})^{\perp} \Rightarrow im(\boldsymbol{L'}) \subset \mathbf{1}^{\perp}$.

Moreover, $\boldsymbol{M}\mathbf{1}_n = \boldsymbol{QQ'}\mathbf{1}_n = 0$ since $\boldsymbol{Q}$ is column-centered. Given that $\text{rank}(\boldsymbol{M}) = n-1$ from Theorem 4, it follows that $\mathbf{1}_n$ is the only eigenvector of $\boldsymbol{M}$ corresponding to zero eigenvalue.

Combining the above two observations, we see that $im(\boldsymbol{L'})$ is a subspace of the space spanned by all eigenvectors of $\boldsymbol{M}$ that correspond to positive eigenvalues. Therefore, we have $\boldsymbol{(L'v)'M(L'v)} > 0 \Rightarrow \boldsymbol{LML'}$ is positive definite.

□

REFERENCES

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall London.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology* **11,** R106.

Beck, A. and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences* **2,** 183–202.

Belkaid, Y. and Hand, T. W. (2014). Role of the microbiota in immunity and inflammation. *Cell* **157,** 121–141.

Bonnefond, A. and Froguel, P. (2015). Rare and common genetic events in type 2 diabetes: what should biologists know? *Cell Metabolism* **21,** 357–368.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pena, A. G., Goodrich, J. K., Gordon, J. I., et al. (2010). Qiime allows analysis of high-throughput community sequencing data. *Nature Methods* **7,** 335.

Chen, G.-B., Lee, S. H., Brion, M.-J. A., Montgomery, G. W., Wray, N. R., Radford-Smith, G. L., Visscher, P. M., and Consortium, I. I. G. (2014). Estimation and partitioning of (co) heritability of inflammatory bowel disease from gwas and immunochip data. *Human Molecular Genetics* **23,** 4710–4720.

Chen, J., Bittinger, K., Charlson, E. S., Hoffmann, C., Lewis, J., Wu, G. D., Collman, R. G., Bushman, F. D., and Li, H. (2012). Associating microbiome composition with environmental covariates using generalized unifrac distances. *Bioinformatics* **28,** 2106–2113.

Chen, J. and Li, H. (2013). Variable selection for sparse dirichlet-multinomial regression with an application to microbiome data analysis. *The Annals of Applied Statistics* **7,** 418–442.

Chen, X., Lin, Q., Kim, S., Carbonell, J. G., Xing, E. P., et al. (2012). Smoothing proximal gradient method for general structured sparse regression. *The Annals of Applied Statistics* **6,** 719–752.

Chen, Y. and Hanson, T. E. (2014). Bayesian nonparametric k-sample tests for censored and uncensored data. *Computational Statistics & Data Analysis* **71,** 335–346.

Cockburn, D. W. and Koropatkin, N. M. (2016). Polysaccharide degradation by the intestinal microbiota and its influence on human health and disease. *Journal of Molecular Biology* **428,** 3230–3252.

David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., Ling, A. V., Devlin, A. S., Varma, Y., Fischbach, M. A., et al. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505,** 559.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*, volume 1. Cambridge University Press.

Dennis III, S. Y. (1991). On the hyper-dirichlet type 1 and hyper-liouville distributions. *Communications in Statistics-Theory and Methods* **20,** 4069–4081.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., and Andersen, G. L. (2006). Greengenes, a chimera-checked 16s rrna gene database and workbench compatible with arb. *Applied and Environmental Microbiology* **72,** 5069–5072.

Devaraj, S., Hemarajata, P., and Versalovic, J. (2013). The human gut microbiome and body metabolism: implications for obesity and diabetes. *Clinical Chemistry* **59,** 617–628.

Dohmen, K. (2000). Improved bonferroni inequalities via union-closed set systems. *Journal of Combinatorial Theory, Series A* **92,** 61–67.

Dohmen, K. (2002). Improved inclusion-exclusion identities and bonferroni inequalities with reliability applications. *SIAM Journal on Discrete Mathematics* **16,** 156–171.

Dohmen, K. and Tittmann, P. (2004). Bonferroni-galambos inequalities for partition lattices. *Electronic Journal of Combinatorics* **11,** 85.

Eaves, L. J., Last, K. A., Young, P. A., and Martin, N. G. (1978). Model-fitting approaches to the analysis of human behaviour. *Heredity* **41,** 249.

Efron, B. (1997). The length heuristic for simultaneous hypothesis tests. *Biometrika* **84,** 143–157.

Fang, H., Huang, C., Zhao, H., and Deng, M. (2015). CCLasso: correlation inference for compositional data through lasso. *Bioinformatics* **31,** 3172–3180.

Felsenstein, J. and Felenstein, J. (2004). *Inferring Phylogenies*. Sinauer associates Sunderland, MA.

Friedman, J. and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Computational Biology* **8,** e1002687.

Giongo, A., Gano, K. A., Crabb, D. B., Mukherjee, N., Novelo, L. L., Casella, G., Drew, J. C., Ilonen, J., Knip, M., Hyöty, H., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. *The ISME Journal* **5,** 82.

Glaz, J., Naus, J. I., Wallenstein, S., Wallenstein, S., and Naus, J. I. (2001). *Scan Statistics*. Springer.

Goodrich, J. K., Davenport, E. R., Beaumont, M., Jackson, M. A., Knight, R., Ober, C., Spector, T. D., Bell, J. T., Clark, A. G., and Ley, R. E. (2016). Genetic determinants of the gut microbiome in uk twins. *Cell Host & Microbe* **19,** 731–743.

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., Beaumont, M., Van Treuren, W., Knight, R., Bell, J. T., et al. (2014). Human genetics shape the gut microbiome. *Cell* **159,** 789–799.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53,** 325–338.

Gritz, E. C. and Bhandari, V. (2015). The human neonatal gut microbiome: a brief review. *Frontiers in Pediatrics* **3,** 17.

Hahn, T. (2005). Cuba - a library for multidimensional numerical integration. *Computer Physics Communications* **168,** 78–95.

Holmes, C. C., Caron, F., Griffin, J. E., Stephens, D. A., et al. (2015). Two-sample bayesian nonparametric hypothesis testing. *Bayesian Analysis* **10,** 297–320.

Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PloS One* **7,** e30126.

Hunter, D. (1976). An upper bound for the probability of a union. *Journal of Applied Probability* **13,** 597–603.

Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host & Microbe* **17,** 260–273.

Krawczak, M., Nikolaus, S., von Eberstein, H., Croucher, P. J., El Mokhtari, N. E., and Schreiber, S. (2006). Popgen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Public Health Genomics* **9,** 55–61.

La Rosa, P. S., Brooks, J. P., Deych, E., Boone, E. L., Edwards, D. J., Wang, Q., Sodergren, E., Weinstock, G., and Shannon, W. D. (2012). Hypothesis testing and power calculations for taxonomic-based human microbiome data. *PloS One* **7,** e52078.

Laurent, B. and Massart, P. (2000). Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics* **28,** 1302–1338.

Lavine, M. (1992). Some aspects of polya tree distributions for statistical modelling. *The Annals of Statistics* **22,** 1222–1235.

Li, H. (2015). Microbiome, metagenomics, and high-dimensional compositional data analysis. *Annual Review of Statistics and Its Application* **2,** 73–94.

Liu, D. C. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming* **45,** 503–528.

Lozupone, C. and Knight, R. (2005). Unifrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology* **71,** 8228–8235.

Lozupone, C. A., Hamady, M., Kelley, S. T., and Knight, R. (2007). Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* **73,** 1576–1585.

Ma, L. and Wong, W. H. (2011). Coupling optional pólya trees and the two sample problem. *Journal of the American Statistical Association* **106,** 1553–1565.

Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., Nalin, R., Jarrin, C., Chardon, P., Marteau, P., et al. (2006). Reduced diversity of faecal microbiota in crohns disease revealed by a metagenomic approach. *Gut* **55,** 205–211.

McArdle, B. H. and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology* **82,** 290–297.

McDonald, D., Birmingham, A., and Knight, R. (2015). Context and the human microbiome. *Microbiome* **3,** 52.

McDonald, D., Hornig, M., Lozupone, C., Debelius, J., Gilbert, J. A., and Knight, R. (2015). Towards large-cohort comparative studies to define the factors influencing the gut microbial community structure of asd patients. *Microbial Ecology in Health and Disease* **26,** 26555.

McMurdie, P. J. and Holmes, S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology* **10,** e1003531.

Menni, C., Jackson, M. A., Pallister, T., Steves, C. J., Spector, T. D., and Valdes, A. M. (2017). Gut microbiome diversity and high-fibre intake are related to lower long-term weight gain. *International Journal of Obesity* **41,** 1099.

Methé, B. A., Nelson, K. E., Pop, M., Creasy, H. H., Giglio, M. G., Huttenhower, C., Gevers, D., Petrosino, J. F., Abubucker, S., Badger, J. H., et al. (2012). A framework for human microbiome research. *Nature* **486,** 215–221.

Morgan, C. (1974). Embedding metric spaces in euclidean space. *Journal of Geometry* **5,** 101–107.

Mosimann, J. E. (1962). On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika* **49,** 65–82.

Müller, N., Schulte, D. M., Türk, K., Freitag-Wolf, S., Hampe, J., Zeuner, R., Schröder, J. O., Gouni-Berthold, I., Berthold, H. K., Krone, W., et al. (2015). IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *Journal of Lipid Research* **56,** 1034–1042.

Naiman, D. Q. and Wynn, H. P. (1992). Inclusion-exclusion-bonferroni identities and inequalities for discrete tube-like problems via euler characteristics. *The Annals of Statistics* **20,** 43–76.

Naiman, D. Q., Wynn, H. P., et al. (1997). Abstract tubes, improved inclusion-exclusion identities and inequalities and importance sampling. *The Annals of Statistics* **25,** 1954–1983.

Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate O($1/k^2$). In *Soviet Mathematics Doklady*, volume 27, pages 372–376.

Nguyen, N.-P., Warnow, T., Pop, M., and White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms and Microbiomes* **2,** 16004.

Novembre, J. and Stephens, M. (2008). Interpreting principal component analyses of spatial population genetic variation. *Nature Genetics* **40,** 646.

Omranian, N., Eloundou-Mbebi, J. M., Mueller-Roeber, B., and Nikoloski, Z. (2016). Gene regulatory network inference using fused lasso on multiple data sets. *Scientific Reports* **6,** 20533.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26,** 1641–1650.

Quigley, K. M., Willis, B. L., and Bay, L. K. (2017). Heritability of the symbiodinium community in vertically-and horizontally-transmitting broadcast spawning corals. *Scientific Reports* **7,** 8219.

Ridlon, J. M., Kang, D.-J., and Hylemon, P. B. (2006). Bile salt biotransformations by human intestinal bacteria. *Journal of Lipid Research* **47,** 241–259.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140.

Schweiger, R., Kaufman, S., Laaksonen, R., Kleber, M. E., März, W., Eskin, E., Rosset, S., and Halperin, E. (2016). Fast and accurate construction of confidence intervals for heritability. *The American Journal of Human Genetics* **98,** 1181–1192.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics* **41,** 2263–2291.

Shi, N., Li, N., Duan, X., and Niu, H. (2017). Interaction between the gut microbiome and mucosal immune system. *Military Medical Research* **4,** 14.

Silverman, J. D., Washburne, A. D., Mukherjee, S., and David, L. A. (2017). A phylogenetic transform enhances analysis of compositional microbiota data. *Elife* **6,**.

Soriano, J. and Ma, L. (2017). Probabilistic multi-resolution scanning for two-sample differences. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79,** 547–572.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313.

Subramanian, S., Huq, S., Yatsunenko, T., Haque, R., Mahfuz, M., Alam, M. A., Benezra, A., DeStefano, J., Meier, M. F., Muegge, B. D., et al. (2014). Persistent gut microbiota immaturity in malnourished bangladeshi children. *Nature* **510,** 417–421.

Sze, M. A. and Schloss, P. D. (2016). Looking for a signal in the noise: revisiting obesity and the microbiome. *MBio* **7,** e01018–16.

Tang, Z.-Z., Chen, G., Alekseyenko, A. V., and Li, H. (2017). A general framework for association analysis of microbial communities on a taxonomic tree. *Bioinformatics* **33,** 1278–1285.

Taylor, J. E., Worsley, K. J., and Gosselin, F. (2007). Maxima of discretely sampled random fields, with an application to 'bubbles'. *Biometrika* **94,** 1–18.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67,** 91–108.

Tilg, H. and Kaser, A. (2011). Gut microbiome, obesity, and metabolic dysfunction. *The Journal of Clinical Investigation* **121,** 2126–2132.

Turnbaugh, P. J., Ridaura, V. K., Faith, J. J., Rey, F. E., Knight, R., and Gordon, J. I. (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine* **1,** 6ra14.

van Opstal, E. J. and Bordenstein, S. R. (2015). Rethinking heritability of the microbiome. *Science* **349,** 1172–1173.

Wadsworth, W. D., Argiento, R., Guindani, M., Galloway-Pena, J., Shelburne, S. A., and Vannucci, M. (2017). An integrative bayesian dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data. *BMC Bioinformatics* **18,** 94.

Wahlström, A., Sayin, S. I., Marschall, H.-U., and Bäckhed, F. (2016). Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metabolism* **24,** 41–50.

Wang, J., Thingholm, L. B., Skiecevičienė, J., Rausch, P., Kummen, M., Hov, J. R., Degenhardt, F., Heinsen, F.-A., Rühlemann, M. C., Szymczak, S., et al. (2016). Genome-wide association analysis identifies variation in vitamin d receptor and other host factors influencing the gut microbiota. *Nature Genetics* **48,** 1396–1406.

Wang, T. and Zhao, H. (2017). A dirichlet-tree multinomial regression model for associating dietary nutrients with gut microorganisms. *Biometrics* **73,** 792–801.

Weir, B. S. and Hill, W. G. (2002). Estimating F-statistics. *Annual Review of Genetics* **36,** 721–750.

Whittaker, R. H. (1960). Vegetation of the siskiyou mountains, oregon and california. *Ecological Monographs* **30,** 279–338.

Wipperman, M. F., Fitzgerald, D. W., Juste, M. A. J., Taur, Y., Namasivayam, S., Sher, A., Bean, J. M., Bucci, V., and Glickman, M. S. (2017). Antibiotic treatment for tuberculosis induces a profound dysbiosis of the microbiome that persists long after therapy is completed. *Scientific Reports* **7,** 10767.

Worsley, K. (1982). An improved bonferroni inequality and applications. *Biometrika* **69,** 297–302.

Wu, G., Shen, D., and Sabuncu, M. (2016). *Machine Learning and Medical Imaging*. Academic Press.

Xia, F., Chen, J., Fung, W. K., and Li, H. (2013). A logistic normal multinomial regression model for microbiome compositional data analysis. *Biometrics* **69,** 1053–1063.

Yang, I., Corwin, E. J., Brennan, P. A., Jordan, S., Murphy, J. R., and Dunlop, A. (2016). The infant microbiome: implications for infant health and neurocognitive development. *Nursing Research* **65,** 76–88.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42,** 565–569.

Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88,** 76–82.

Yang, X., Zhang, T., Xu, C., and Xu, M. (2013). Graph-guided fusion penalty based sparse coding for image classification. In *Pacific-Rim Conference on Multimedia*, pages 475–484. Springer.

Young, V. B. (2017). The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ* **356,** j831.

Zhao, N., Chen, J., Carroll, I. M., Ringel-Kulka, T., Epstein, M. P., Zhou, H., Zhou, J. J., Ringel, Y., Li, H., and Wu, M. C. (2015). Testing in microbiome-profiling studies with mirkat, the microbiome regression-based kernel association test. *The American Journal of Human Genetics* **96,** 797–807.

Zhou, X. and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44,** 821–824.