

Article

Dependence Structure Analysis and Its Application in Human Microbiome

Shilan Li ^{1,2}, Jianxin Shi ² , Paul Albert ² and Hong-Bin Fang ^{1,*} 

¹ Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University Medical Center, Washington, DC 20057, USA

² Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD 20892, USA

* Correspondence: hf183@georgetown.edu

Abstract: The human microbiome has been recently shown to be associated with disease risks and has important implications in risk stratification and precision medicine. Due to abundant taxa in the human body, microbiome data are high-dimensional and compositional. Dirichlet distributions and their generalization are used to characterize the dependence structures of microbial data. Another existing method for fitting microbiome data employed Gaussian graphical model using the centered log-transformation (CLR). However, Dirichlet distributions are not able to infer networks or to estimate some extremely rare probabilities. On the other hand, it is hard to interpret the network analysis results using CLR. Furthermore, the data analysis showed that there is a lack of efficient multivariate distributions for fitting microbiome data, which results in inadequate statistical inferences. In this paper, we propose new multivariate distributions for modeling the dependence structures of the high dimensional and compositional microbiome data using inverse gamma distributions and copula techniques. The data analysis in the American gut project shows our proposed methods perform well.

Keywords: copula; dependence analysis; goodness-of-fit test; Kendall's correlation coefficient; microbiome; multivariate distribution; parametric estimation

MSC: 62E10



Citation: Li, S.; Shi, J.; Albert, P.; Fang, H.-B. Dependence Structure Analysis and Its Application in Human Microbiome. *Mathematics* **2023**, *11*, 9. <https://doi.org/10.3390/math11010009>

Academic Editor: Kai-Tai Fang

Received: 29 November 2022

Revised: 13 December 2022

Accepted: 14 December 2022

Published: 20 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human microbiome is a gathering of microbes that inhabit different sites of the human body, including bacteria, archaea, and fung. The research on low-price high-throughput sequencing analyses has substantially facilitated the characterization of the compositions and diversities of the human microbiome. It provides a chance to carry out large-scale population-based microbiome studies to identify novel risk factors and to improve the risk prediction of diseases. Over the past two decades, increased research has focused on studying the structure, function, and dynamics of microbiome data, referring to all of the taxa and their genes in a well-defined environment [1]. Technical advancements and reduced costs in sequencing have led to large-scale studies such as the human microbiome project (HMP) and the American gut project (AGP), which characterize the microbiome of largely healthy individuals [2,3]. Most of the early research focused on microbial diversity and taxonomic classification. More recently, research has focused on differential abundance analysis and understanding how the host environment is associated with the microbiome [4–7]. The human microbiome is known to be associated with complex diseases, such as obesity, inflammatory bowel disease, and rheumatoid arthritis [8–10].

Despite these advances, little is known about inter-microbial interactions. Much of the lack of information on microbial interactions comes from the fact that most standard statistical methods for correlation or network analysis cannot be directly applied to such high-dimensional, compositional, and sparse data. To analyze microbiome data, there are

two methods in the literature for modeling the distributions of the relative abundance data of many taxa. One method is utilizing multivariate normal distributions to model correlation structures of taxa by means of the centred log ratio transformation (CLR) or alternative log ratio transformations (ALR), and the other is using Dirichlet multinomial distribution to model taxa counts [11,12]. Dirichlet distributions are not quite flexible enough to fit many real data sets in microbiome studies. Moreover, the lack of dependence structure analysis hinders further statistical inferences in microbiome studies [13].

For modeling microbial dependence structures with multivariate normal distributions, two of the most common methods, SparCC and SPIEC-EASI, are used. SparCC uses the additive log-ratio transformations and assumes that the underlying unobserved counts are correlated through log-ratio [14]. In contrast, SPIEC-EASI uses the CLR and assumes that the network of interactions is generated from a precision matrix [15]. While the log-ratio transformations are common in compositional data analysis, they are not particularly fitting for data with excessive zeros [16], as is the case with microbial sequencing data. The normality assumption of such data often does not hold (see, e.g., Figure 1a below), and they require the use of pseudo-counts, thus forcing the assumption that the true absolute abundance (AA) for every taxa is non-zero in each sample. Figure 1 shows that there are large departures between the estimated contour plots of samples and the contour plots based on the estimated normal distribution with CLR or the estimated Dirichlet distribution. Furthermore, the corresponding results from such transformations are difficult to interpret and are sensitive to the choice of reference groups; as Feng et al. [17] pointed out, the results of standard statistical tests performed on log-transformed data are often not relevant for the original, non-transformed data.

To model the high-dimensional and compositional data in microbiome studies, we propose two classes of multivariate distributions on a simplex. A new class of multivariate distributions on a simplex is composed of variable transformation of inverse gamma distributions since most datasets show that a single taxonomy can be characterized by an inverse Gamma distribution. Another new class of multivariate distributions on simplex is constructed by means of copula techniques. Copulas are functions that join univariate uniform distribution functions to form multivariate distributions and are powerful for characterizing the dependence structures of random variables [18].

Furthermore, human microbiota can be influenced by environment changes and lifestyle changes. Microbiome data are compositional and sparse. A few taxa are abundant, most taxa are rare (present in <10% subjects), and there are zero counts for some taxa. This could be caused by sampling zero (low sequence depth) or structural zeros (absent taxa). To deal with sampling zeros, we could enlarge the sequencing depth. However, in our settings, we ignore sampling zeros and only deal with structural zeros by zero-inflated distributions.

The remainder of this paper is organized as follows. Section 2 constructs the new class of multivariate distributions on a simplex by variable transformation of inverse gamma distributions. In Section 3, using the copula techniques, we propose a new class of multivariate distributions on a simplex with a Gaussian copula and a Farlie–Gumbel–Morgenstern copula. The real data from the American gut project (AGP) is analyzed to demonstrate the effectiveness of the proposed new multivariate distributions in Section 4, and some conclusions are given in Section 5.

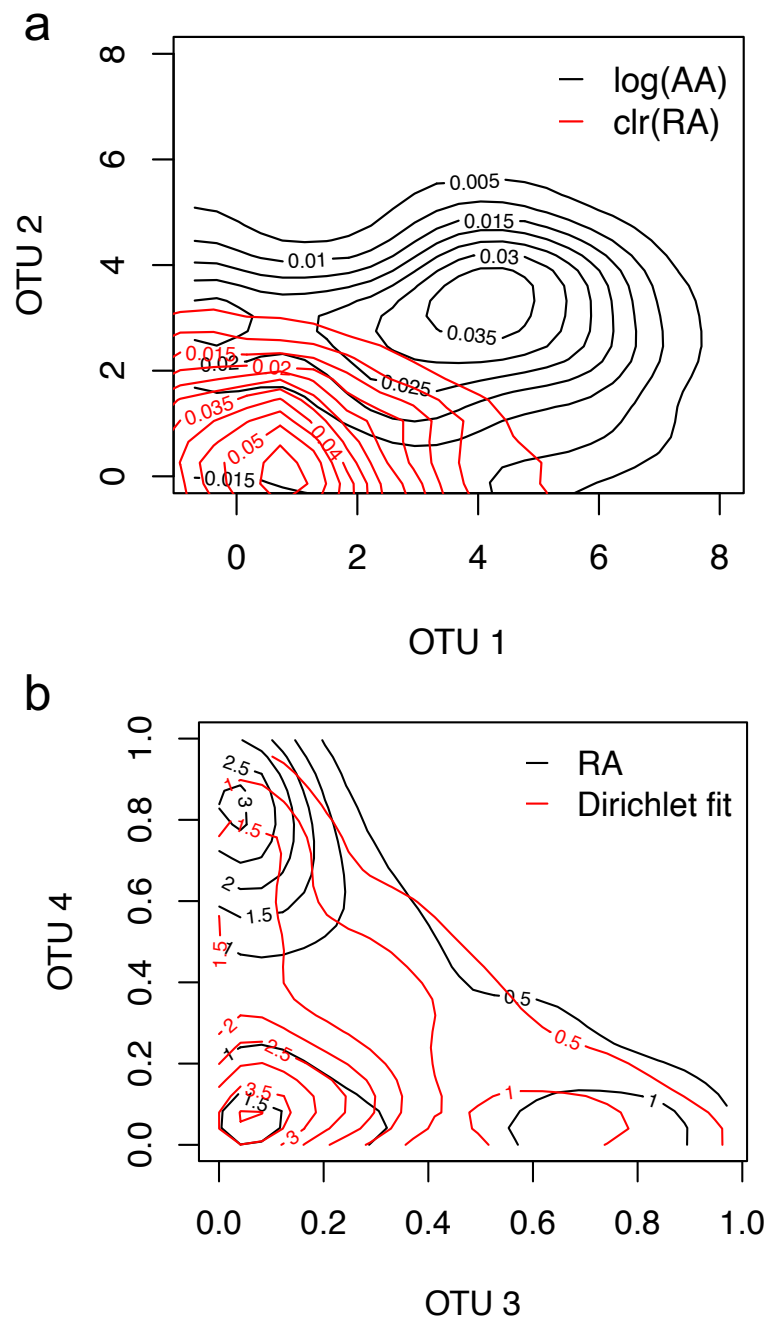


Figure 1. The contour plots of Faecalibacterium (OTU1) and Bacteroides (OTU2) from AGP in the upper panel and taxa Clostridium (OTU3) and Escherichia (OTU4) on the lower panel. In (a), AA is the absolute abundance that describes the OTU counts for taxa. The black lines are the estimated contour plots of samples, and the red lines are the contour plots of estimated distributions, which shows that there are large departures between the estimated contour plots of samples and the contour plots based on the estimated normal distribution with CLR (a) or the estimated Dirichlet distribution (b).

2. Multivariate Distributions on Simplex via Inverse Gamma Distributions

Suppose there are $k + 1$ microbial taxa, and denote $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T$ as the true relative abundance (RA) of microbial taxa. The RA has a compositional structure with $\sum_{j=1}^{k+1} Y_j = 1$, and the vector \mathbf{Y} lies on the k -dimensional simplex \mathcal{S}^k ,

$$\mathcal{S}^k := \left\{ (y_1, \dots, y_k)^T : y_i > 0, i = 1, \dots, k, \sum_{i=1}^k y_i \leq 1 \right\} \tag{1}$$

To construct the distributions of \mathbf{Y} , the inverse Gamma distributions on $(0, \infty)$ are employed since most datasets show that a single taxonomy can be characterized by an inverse Gamma distribution (IG). Let X_1, \dots, X_{k+1} be independent random variables on $(0, \infty)$ and have the inverse Gamma distributions $X_j \sim IG(\alpha, \beta_j)$ with its density function

$$f(x; \alpha, \beta_j) = \frac{\beta_j^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta_j}{x}\right), \text{ for } x > 0, \tag{2}$$

where $\alpha > 0$ is the shape parameter and $\beta_j > 0$ are the scale parameters for $j = 1, \dots, k + 1$. For a given subject, we assume that the k taxa have the same shape parameter α , but the scale parameters β_j vary. With the following transformation,

$$Y_j = \frac{X_j}{\sum_{j=1}^{k+1} X_j}, \quad Y_{k+1} = 1 - \sum_{j=1}^k Y_j, \text{ for } j = 1, \dots, k, \tag{3}$$

The RA of microbial taxa \mathbf{Y} has the multivariate distribution on the k -dimensional simplex \mathcal{S}^k with its density function as

$$f(\mathbf{y}; \alpha, \beta_1, \dots, \beta_{k+1}) = \frac{\Gamma((k+1)\alpha)}{\Gamma(\alpha)^{k+1}} \prod_{j=1}^{k+1} \beta_j^\alpha \prod_{j=1}^{k+1} y_j^{-\alpha-1} \left(\sum_{j=1}^{k+1} \frac{\beta_j}{y_j} \right)^{-(k+1)\alpha}, \tag{4}$$

for $\mathbf{y} = (y_1, \dots, y_{k+1})^T$ and $0 < y_j < 1, \sum_{j=1}^{k+1} y_j = 1$ for $j = 1, \dots, k + 1$. We denote $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T \sim MSIG(\alpha; \beta_1, \dots, \beta_{k+1})$. The special case of this multivariate distribution when $\alpha = \frac{1}{2}$ was first introduced by Carlton [19] and further discussed in Favaro et al. [20].

For the density function (4), the marginal density functions are

$$f(y_j; \alpha; \beta_1, \dots, \beta_{k+1}) = \frac{\beta_j^\alpha}{\Gamma(\alpha)} \frac{(1 - y_j)^{\alpha+3}}{y_j^{\alpha+1}} \int_0^\infty \dots \int_0^\infty \frac{\exp\left[\frac{-\beta_j(1-y_j)}{y_j(\sum_{i \neq j} x_i)} - \sum_{i \neq j} \frac{\beta_i}{x_i}\right]}{(\sum_{i \neq j} x_i)^\alpha \prod_{i \neq j} x_i^{\alpha+1}} \prod_{i \neq j} dx_i, \tag{5}$$

for $j = 1, \dots, k + 1$. Figure 2 shows the estimated marginal density functions by (5) are fitting well for the sample density functions of Faecalibacterium (OTU1) and Bacteroides (OTU2) from AGP.

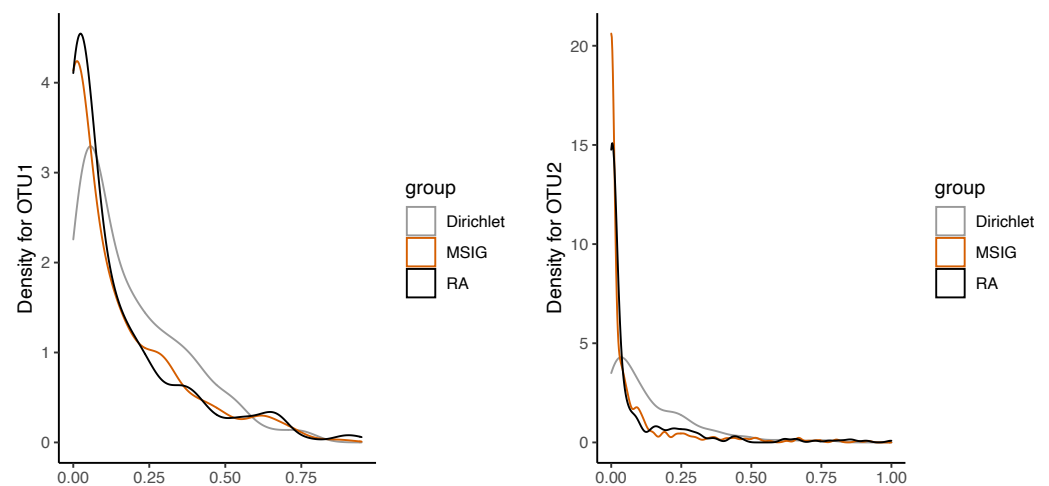


Figure 2. The estimated marginal density functions: the black line is estimated by the samples, the grey line by Dirichlet distributions, and the red line is by MSIG distribution.

In the studies of the human microbiome, it is more expedient to consider the average total positivity [21] (and reverse regularity) of order two microbial taxa. Kendall’s correlation coefficient is a measure of the difference between the probabilities of concordance and discordance for two independent and identically distributed pairs of random variables. It is the difference between the probabilities of concordance and discordance for two independent and identically distributed pairs of random variables [22,23]. Let $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T \sim MSIG(\alpha; \beta_1, \dots, \beta_{k+1})$ with the density given in (4). Because of the transformation (3), the corresponding Kendall correlation coefficient of Y_i and Y_j is given by

$$\tau(Y_i, Y_j) = E(\text{sign}(Y_i - Y_{i'})\text{sign}(Y_j - Y_{j'})) = E(\text{sign}(X_i - X_{i'})\text{sign}(X_j - X_{j'})) \quad (6)$$

where $(Y_{i'}, Y_{j'})$ is an independent copy of (Y_i, Y_j) and $(X_{i'}, X_{j'})$ is an independent copy of (X_i, X_j) for $i \neq j, i, j = 1, \dots, k + 1$. Since

$$P(X_i - X_{i'} < 0) = \int_0^\infty \int_x^\infty \frac{\beta_i^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta_i}{x}\right) \frac{\beta_i^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp\left(\frac{-\beta_i}{y}\right) dy dx, \quad (7)$$

then, the Kendall correlation coefficient of Y_i and Y_j is the same as that of two independent IG random variables X_i and X_j , and we have that

$$\begin{aligned} \tau(Y_i, Y_j) &= P(X_i > X_{i'})P(X_j > X_{j'}) + P(X_i < X_{i'})P(X_j < X_{j'}) \\ &\quad - P(X_i > X_{i'})P(X_j < X_{j'}) - P(X_i < X_{i'})P(X_j > X_{j'}) \\ &= 2 \left(1 - \frac{2^{-2\alpha}}{\Gamma(\alpha)} \sum_{l=0}^\infty \frac{\Gamma(2\alpha + l)}{2^l \Gamma(\alpha + l + 1)} \right) \left(1 - \frac{2^{-2\alpha+1}}{\Gamma(\alpha)} \sum_{l=0}^\infty \frac{\Gamma(2\alpha + l)}{2^l \Gamma(\alpha + l + 1)} \right), \end{aligned} \quad (8)$$

which is only dependent on the parameter α , for $i \neq j, i, j = 1, \dots, k + 1$.

2.1. Parameter Estimation

Let $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}$ be *i.i.d* samples from $\mathbf{Y} \sim MSIG(\alpha; \beta_1, \dots, \beta_{k+1})$. The log-likelihood function of parameters $\Theta = (\alpha; \beta_1, \dots, \beta_{k+1})$ is given by

$$\begin{aligned} \ell(\Theta | \mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}) &\propto n\alpha \sum_{j=1}^{k+1} \log \beta_j + n \log(\Gamma((k+1)\alpha)) - n(k+1) \log(\Gamma(\alpha)) \\ &\quad - \sum_{i=1}^n \left[(\alpha + 1) \sum_{j=1}^{k+1} \log y_j^{(i)} + (k+1)\alpha \log \left(\sum_{j=1}^{k+1} \frac{\beta_j}{y_j^{(i)}} \right) \right], \end{aligned}$$

and we obtain the score functions

$$k\beta_h = \sum_{i=1}^n \left(\frac{\sum_{j \neq h}^{k+1} \frac{\beta_j}{y_j^{(i)}}}{\beta_h + \sum_{j \neq h}^{k+1} \frac{\beta_j}{y_j^{(i)}}} \right) \left[\sum_{i=1}^n \frac{1}{\beta_h + \sum_{j \neq h}^{k+1} \frac{\beta_j}{y_j^{(i)}}} \right]^{-1}, h = 1, \dots, k + 1; \tag{9}$$

$$n(k + 1) \left(\frac{\Gamma'((k+1)\alpha)}{\Gamma((k+1)\alpha)} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) = \sum_{i=1}^n \left[\sum_{j=1}^{k+1} \log y_j^{(i)} + \log \left(\sum_{j=1}^{k+1} \frac{\beta_j}{y_j^{(i)}} \right) (k + 1) \right] - n \sum_{j=1}^{k+1} \log \beta_j,$$

where $\Gamma'(\cdot)$ is the derivative function of $\Gamma(\cdot)$. The maximum likelihood estimation (MLE) of $\Theta = (\alpha; \beta_1, \dots, \beta_{k+1})$ can be obtained by an iterative algorithm.

2.2. Multivariate Goodness-of-Fit Testing

For given n samples $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}$ in the k -dimensional simplex \mathcal{S}^k , we can obtain the estimated parameters $\hat{\Theta} = (\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$. A goodness-of-testing procedure is necessary to test the hypothesis that the samples $\{\mathbf{y}^{(i)}\}_{i=1}^n$ come from the distribution $\mathbf{Y} \sim MSIG(\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$. For one-dimensional or low-dimensional data, there are a wide variety of useful and powerful goodness-of-fit and two-sample testing procedures, such as Kolmogorov–Smirnov (K-S) test. However, for high-dimensional ($k > 3$) data, those tests rapidly lose statistical power because all finite samples are sparse in high-dimensional settings owing to the “curse of dimensionality” [24]. In this paper, a machine learning method for two-sample testing is employed, which was proposed by Friedman [24] and further developed by Gretton et al. [25]. For the given observed samples $\{\mathbf{y}^{(i)}\}_{i=1}^n$, we can obtain a set of score values $\{v_i = F(\mathbf{y}^{(i)})\}_{i=1}^n$, where $F(\mathbf{y})$ is the cdf of $MSIG(\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$. At same time, with the “Monte Carlo” samples $\{\mathbf{z}^{(i)}\}_{i=1}^n$ from the distribution $MSIG(\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$, we can obtain another set of score values $\{w_i = F(\mathbf{z}^{(i)})\}_{i=1}^n$. Then, a univariate two-sample test, such as the K-S test, chi-squared test, or Mann–Whitney test, can be used to test the goodness of fit based on the scores. The testing procedure is as follows:

- Step1. Generating n samples $\{x_{ij}\}_{i=1}^n$ from the distribution $IG(\hat{\alpha}, \hat{\beta}_j)$ for $j = 1, \dots, k + 1$.
 With the transform (3), we can obtain n samples $\{\mathbf{z}^{(i)}\}_{i=1}^n$ from the distribution $MSIG(\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$ such that $\mathbf{z}^{(i)} = (z_{i1}, \dots, z_{i,k+1})^T$, $z_{ij} = \frac{x_{ij}}{\sum_{j=1}^{k+1} x_{ij}}$, $z_{i,k+1} = 1 - \sum_{j=1}^k z_{ij}$, for $j = 1, \dots, k$, and $i = 1, \dots, n$.
- Step2. By denoting $F(\mathbf{y})$ the cumulative distribution function (cdf) of $MSIG(\hat{\alpha}; \hat{\beta}_1, \dots, \hat{\beta}_{k+1})$, we obtain two samples $\{w_i = F(\mathbf{z}^{(i)})\}_{i=1}^n$ and $\{v_i = F(\mathbf{y}^{(i)})\}_{i=1}^n$.
- Step3. Using the one-dimensional two-sample test, such as the K-S test, chi-squared test, or Mann–Whitney test, to test the hypothesis that the two samples $\{w_i\}_{i=1}^n$ and $\{v_i\}_{i=1}^n$ have the same distribution.
- Step4. Repeating N (for example, $N = 5000$) times of Steps 1–3, we cannot reject the null hypothesis if the percent of rejecting times of testing in Step 3 (i.e., the p -value < 0.05) is less than 5%.

Friedman showed that the power of this testing procedure can be highly sensitive to the learning machine methods that are employed [24].

2.3. Zero-Inflated Distribution with Zero Counts for Absent Taxa

The above multivariate distribution assumes that all taxa have positive proportions. However, for a given taxonomy j , they may be absent in some subjects. To consider the zero-inflated structures, we introduce the zero-inflated inverse Gamma distributions [26].

Let X_1^*, \dots, X_{k+1}^* be independent random variables on $[0, \infty)$ and $X_j^* = (1 - \Delta_j) + \Delta_j X_j$, where Δ_j follows a Bernoulli distribution $\Delta_j \sim \text{Bernoulli}(\pi_j)$ with $P(\Delta_j = 1) = \pi_j$ and $P(\Delta_j = 0) = 1 - \pi_j$, and X_j follows an inverse Gamma distribution $X_j \sim \text{IG}(\alpha, \beta_j)$ for $j = 1, \dots, k + 1$. Then, X_j^* has a zero-inflated inverse Gamma distribution $X_j^* \sim \text{ZIIG}(\pi_j, \alpha, \beta_j)$ with the probability density

$$f(x|\pi_j, \alpha, \beta_j) = \pi_j^\delta \left((1 - \pi) \frac{\beta_j^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta_j}{x}\right) \right)^{1-\delta}, \quad \delta = I\{x = 0\}, \tag{10}$$

for $x \in [0, \infty)$. With the transformation

$$Y_j = \frac{X_j^*}{\sum_{j=1}^{k+1} X_j^*}, \quad Y_{k+1} = 1 - \sum_{j=1}^k Y_j, \quad \text{for } j = 1, \dots, k, \tag{11}$$

The RA of microbial taxa \mathbf{Y} has the multivariate distribution on the k -dimensional simplex \mathcal{S}^k with its density function as

$$f(\mathbf{y}; \alpha, \beta_1, \dots, \beta_{k+1}) = \prod_{j=1}^{k+1} \pi_j^{\delta_j} (1 - \pi_j)^{1-\delta_j} \frac{\Gamma(\sum_{j=1}^{k+1} (1-\delta_j)\alpha)}{\Gamma(\alpha)^{\sum_{j=1}^{k+1} (1-\delta_j)}} \prod_{j=1}^{k+1} \beta_j^{(1-\delta_j)\alpha} \prod_{j=1}^{k+1} y_j^{(-\alpha-1)(1-\delta_j)} \left(\sum_{j=1}^{k+1} \frac{(1-\delta_j)\beta_j}{y_j} \right)^{-\sum_{j=1}^{k+1} (1-\delta_j)\alpha}, \tag{12}$$

where $\delta_j = I\{y_j = 0\}$ for $\mathbf{y} = (y_1, \dots, y_{k+1})^T$ and $0 \leq y_j < 1, \sum_{j=1}^{k+1} y_j = 1$ for $j = 1, \dots, k + 1$.

3. Multivariate Distributions on Simplex Using Copula Dependence Structures

Since the $k + 1$ microbial taxa $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T$ lie on the k -dimensional simplex \mathcal{S}^k defined in (1), \mathbf{Y} is a k -dimensional random vector. To characterize the dependence structure of the k -dimensional random vector, copula techniques are usually employed [18,23]. Copulas are multivariate distributions with uniformly distributed marginals on $[0, 1]$. Utilizing copulas, the density function of a multivariate distribution can be decomposed into the density weighting function and the product of marginal densities [27]. For a d -variate random vector $\mathbf{X} = (X_1, \dots, X_d)$ with joint cdf $F(\cdot)$ and marginal strictly increasing cdf $F_i(\cdot)$ ($i = 1, \dots, d$), Sklar’s Theorem [28] states that there a d -variate copula cdf $C(\cdot)$ on $[0, 1]^d$ exists such that

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)), \tag{13}$$

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)),$$

for all $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$ and $\mathbf{u} = (u_1, \dots, u_d)^T \in [0, 1]^d$, where $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Let $f(x_1, \dots, x_d)$ and $f_i(x_i)$ ($i = 1, \dots, d$) be the density function and the marginal density functions of $\mathbf{X} = (X_1, \dots, X_d)$, respectively. The corresponding copula density and the density function of \mathbf{X} have the following relationships:

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d} = \frac{f(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d))}{f_1(F_1^{-1}(u_1)) \dots f_d(F_d^{-1}(u_d))}, \tag{14}$$

$$f(x_1, \dots, x_d) = c(F_1(x_1), \dots, F_d(x_d)) \prod_{j=1}^d f_j(x_j).$$

The distributions of individual taxonomy are easily estimated, for example, Qiu et al. [13] proposed two-parametric simplex distribution with its density function,

$$f(y) = (2\pi s(y(1-y))^3)^{-1/2} \exp\left(-\frac{(y-m)^2}{2m(1-m)^2 y(1-y)s}\right)$$

where (s, m) are shaping parameters for $0 < y < 1$. In this paper, we propose two families of copulas, Gaussian and Farlie–Gumbel–Morgenstern copulas, to characterize the dependence structures of microbial taxa such that the corresponding multivariate distributions are good-of-fit for a given microbial data.

3.1. Gaussian Copula

Many new multivariate distributions are constructed based on the dependence structures of multivariate normal distributions, which are called Gaussian copulas. Let $\mathbf{Z} = (Z_1, \dots, Z_k)^T \sim N(\mathbf{0}, R)$ with mean vector $\mathbf{0}$ and the $k \times k$ covariance matrix R ,

$$R = \{\rho_{ij} : \rho_{ii} = 1, -1 < \rho_{ij} < 1 \text{ for } i \neq j, \rho_{ij} = \rho_{ji}; i, j = 1, \dots, k\}. \tag{15}$$

Denote the cdf of \mathbf{Z} by $F_{\mathbf{Z}}(z_1, \dots, z_k)$; the corresponding copula of \mathbf{Z} is given by

$$C_{\mathbf{Z}}(u_1, \dots, u_k) = F_{\mathbf{Z}}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k)), \tag{16}$$

for $(u_1, \dots, u_k)^T \in (0, 1)^k$, where $\Phi(\cdot)$ is the cdf of standard normal distribution $N(0, 1)$.

Suppose that the marginal distributions of the $k + 1$ microbial taxa $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T$ are $F_j(y_j)$ the cdf of Y_j for $j = 1, \dots, k$, and $Y_{k+1} = 1 - \sum_{j=1}^k Y_j$, the distribution function of \mathbf{Y} can be constructed with the copula (16) to be

$$\begin{aligned} F_{\mathbf{Y}}(y_1, \dots, y_k; R) &= P\{Y_1 \leq y_1, \dots, Y_k \leq y_k\} \\ &= F_{\mathbf{Z}}(\Phi^{-1}(F_1(y_1)), \dots, \Phi^{-1}(F_k(y_k))). \end{aligned} \tag{17}$$

Fang et al. [18] show that Kendall’s correlation coefficient $\tau(Y_i, Y_j)$ of Y_i and Y_j is given by

$$\tau(Y_i, Y_j) = \frac{2}{\pi} \arcsin(\rho_{ij}), \text{ for } i \neq j, i, j = 1, \dots, k, \tag{18}$$

which depends only on ρ_{ij} . Thus, the estimates of ρ_{ij} can be obtained by the estimators of Kendall’s correlation coefficient $\tau(Y_i, Y_j)$ of Y_i and Y_j .

3.2. Farlie–Gumbel–Morgenstern (FGM) Copula

As an alternative to a multivariate normal distribution, the FGM copula has been widely studied and applied to statistical modeling in various research fields [29–31]. Let $\mathbf{U} = (U_1, \dots, U_k)^T$ be a random vector that follows a k -dimensional FGM copula with its cdf $C_{FGM}(u_1, \dots, u_k; \Theta)$,

$$\begin{aligned} C_{FGM}(u_1, \dots, u_k; \Theta) &= P(U_1 \leq u_1, \dots, U_k \leq u_k) \\ &= \prod_{j=1}^k u_j \left(1 + \sum_{i=2}^k \sum_{1 \leq j_1 < \dots < j_i \leq k} \theta_{j_1 \dots j_i} (1 - u_{j_1}) \cdots (1 - u_{j_i}) \right), \end{aligned} \tag{19}$$

where $(u_1, \dots, u_k)^T \in [0, 1]^k$ and $\Theta = \{\theta_{j_1 \dots j_i} : 1 \leq j_1 < \dots < j_i \leq k, i = 2, \dots, k\}$ are dependence parameters since $\Theta \equiv 0$ means that the random variables U_1, \dots, U_k are obviously independent. According to Johnson and Kotz [30], the $2^k - k - 1$ parameters Θ have the following constraints:

$$1 + \sum_{i=2}^k \sum_{1 \leq j_1 < \dots < j_i \leq k} \theta_{j_1 \dots j_i} \delta_{j_1} \cdots \delta_{j_i} \geq 0, \delta_j = \pm 1, \tag{20}$$

for $j = 1, \dots, k$. With the cdf (19), we have the joint-density function of \mathbf{U} given by

$$\begin{aligned}
 c_{FGM}(u_1, \dots, u_k; \Theta) &= \frac{\partial^k}{\partial u_1 \dots \partial u_k} C_{FGM}(u_1, \dots, u_k; \Theta) \\
 &= 1 + \sum_{i=2}^k \sum_{1 \leq j_1 < \dots < j_i \leq k} \theta_{j_1 \dots j_i} (1 - 2u_{j_1}) \dots (1 - 2u_{j_i}).
 \end{aligned}
 \tag{21}$$

For the given marginal distributions $F_j(y_j)$ of the cdf of Y_j for $j = 1, \dots, k$, the distribution function of microbial taxa $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T$ can be constructed with the copula (19) to be

$$\begin{aligned}
 F_{\mathbf{Y}}(y_1, \dots, y_k; \Theta) &= P\{Y_1 \leq y_1, \dots, Y_k \leq y_k\} \\
 &= \prod_{j=1}^k F_j(y_j) \left(1 + \sum_{i=2}^k \sum_{1 \leq j_1 < \dots < j_i \leq k} \theta_{j_1 \dots j_i} (1 - F_{j_1}(y_{j_1})) \dots (1 - F_{j_i}(y_{j_i})) \right),
 \end{aligned}
 \tag{22}$$

where $0 < y_j < 1$ for $j = 1, \dots, k$, and $y_{k+1} = 1 - \sum_{j=1}^k y_j$. The Kendall’s correlation coefficient $\tau(Y_i, Y_j)$ of Y_i and Y_j is given by

$$\begin{aligned}
 \tau(Y_i, Y_j) &= 4 \iint_{[0,1]^2} C(u_i, u_j) dC(u_i, u_j) - 1 \\
 &= \frac{\theta_{ij}}{9} (\theta_{ij} - 8), \text{ for } i \neq j, i, j = 1, \dots, k,
 \end{aligned}
 \tag{23}$$

which depends only on θ_{ij} .

Thus, the estimates of θ_{ij} can be obtained by the estimators of Kendall’s correlation coefficient $\tau(Y_i, Y_j)$ of Y_i and Y_j , but all of the parameters $\Theta = \{\theta_{j_1 \dots j_i} : 1 \leq j_1 < \dots < j_i \leq k, i = 2, \dots, k\}$ should satisfy the constraints in (20). Shih et al. [32,33] provided the maximum likelihood estimates of Θ with the constrains in (20).

3.3. Multivariate Distributions with Zero-Inflated Marginal Distributions

Let the $k + 1$ microbial taxa $\mathbf{Y} = (Y_1, \dots, Y_{k+1})^T$ lie on the k -dimensional simplex S^k defined in (1). When there are proportions with zero counts for absent taxa, the zero and the non-zero responses are assumed to occur from two different processes; then, we could use two-part models to characterize the marginal distributions. Define that $\Delta_j = I\{Y_j = 0\}$ and $Y_j^* = Y_j | (Y_j > 0)$. Let $f_j^*(y, \omega_j)$ and $F_j^*(y, \omega_j)$ be the pdf and cdf of Y_j^* , respectively, where ω_j is a parameter (or a parametric vector). Assume that $\Delta_j \sim \text{Bernoulli}(\pi_j)$ with $P(\Delta_j = 1) = \pi_j$ and $P(\Delta_j = 0) = 1 - \pi_j$. Then, the marginal random variable Y_j has the probability density,

$$f_j(y) = \pi_j^{\delta_j} \left((1 - \pi_j) f_j^*(y, \omega_j) \right)^{1 - \delta_j}, \quad \delta_j = I\{y = 0\},
 \tag{24}$$

for $j = 1, \dots, k$. To construct the multivariate distribution of \mathbf{Y} , we assume that $\Delta_1, \dots, \Delta_k$ have the multinomial distribution with its probability mass function

$$\begin{aligned}
 g(\delta_1, \dots, \delta_k; \pi_1, \dots, \pi_k) &= P\{\Delta_1 = \delta_1, \dots, \Delta_k = \delta_k\} \\
 &= \prod_{j=1}^k \pi_j^{\delta_j} \times \prod_{j=1}^k (1 - \pi_j)^{1 - \delta_j},
 \end{aligned}
 \tag{25}$$

and the joint cdf of (Y_1^*, \dots, Y_k^*) is given by

$$f_{\mathbf{Y}^*}(y_1, \dots, y_k; \omega_1, \dots, \omega_k) = c(F_1^*(y_1, \omega_1), \dots, F_k^*(y_k, \omega_k)) \prod_{j=1}^k f_j^*(y_j, \omega_j),
 \tag{26}$$

where $c(u_1, \dots, u_k)$ is the density function of a copula $C(u_1, \dots, u_k)$. Denote $Y = (\pi_1, \dots, \pi_k; \omega_1, \dots, \omega_k; \Theta)$ or $Y = (\pi_1, \dots, \pi_k; \omega_1, \dots, \omega_k; R)$, where Θ and R are the parameters in the given copula function. Thus, the joint probability function of Y is given by

$$f_Y(y_1, \dots, y_k; Y) = \prod_{j=1}^k \pi_j^{\delta_j} + \prod_{j=1}^k (1 - \pi_j)^{1-\delta_j} \prod_{j=1}^k f_j^*(y_j, \omega_j)^{1-\delta_j} \tilde{c}([F_1^*(y_1, \omega_1)]^{1-\delta_1}, \dots, [F_k^*(y_k, \omega_k)]^{1-\delta_k}), \tag{27}$$

where $\delta_j = I\{y_j = 0\}$ for $j = 1, \dots, k$, $\tilde{c}([F_1^*(y_1, \omega_1)]^{1-\delta_1}, \dots, [F_k^*(y_k, \omega_k)]^{1-\delta_k})$ is the density function of $C(u_1^{1-\delta_1}, \dots, u_k^{1-\delta_k})$ at $u_j = F_j^*(y_j, \omega_j)$ for $\delta_j = 0$ and $j = 1, \dots, k$. Note that $a^0 = 1$ and the marginal cdf $C(u_1^{1-\delta_1}, \dots, u_k^{1-\delta_k})$ are also a copula.

3.4. Parameter Estimation and Goodness-of-Fit Testing

Let $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(n)}$ be *i.i.d* samples from the multivariate distribution with its pdf in (27). To obtain the maximum likelihood estimates of parameters Y , a two-stage approach is employed [34]. The marginal distributions are first estimated based on partial maximum likelihood methods. Given these marginal fits, the estimated cumulative distribution functions are then obtained using the estimates of the parameters in copula. Denote $\delta_{ij} = I\{y_j^{(i)} = 0\}$ for $j = 1, \dots, k, i = 1, \dots, n$. The estimates of π 's can be obtained by

$$\hat{\pi}_j = \sum_{i=1}^n \delta_{ij} / n, \quad j = 1, \dots, k, \tag{28}$$

and the estimates of ω 's maximize the log-likelihood functions.

$$\ell(\omega_j) = \sum_{i=1}^n (1 - \delta_{ij}) \log(f_j^*(y_j^{(i)}, \omega_j)), \quad j = 1, \dots, k. \tag{29}$$

Then, the estimates of the parameters in copula can be obtained by maximizing the likelihood function

$$Lik(Y) = \prod_{i=1}^n \tilde{c}([F_1^*(y_1^{(i)}, \hat{\omega}_1)]^{1-\delta_{i1}}, \dots, [F_k^*(y_k^{(i)}, \hat{\omega}_k)]^{1-\delta_{ik}}). \tag{30}$$

When $C(u_1, \dots, u_k)$ is an FGM copula, Shih et al. [33] provides the maximum likelihood estimates of Θ . If $C(u_1, \dots, u_k)$ is a Gaussian copula, the parameters $R = (\rho_{ij})$ can be estimated by the corresponding estimates of Kendall's correlation coefficient $\tau(Y_i^*, Y_j^*)$ based on (23).

When we obtain the estimates of parameters Y , the machine learning method given in Section 2.1 can be used for goodness-of-fit testing of the proposed multivariate distributions. Shih et al. [33] provides a procedure of generating samples from FGM copulas, and Fang et al. [18] gives a procedure of generating samples from Gaussian copulas.

4. Real Data Analysis

Now we apply our proposed methods to the datasets from AGP, a self-selected, open-platform cohort [2]. The cohort consists of individuals mostly from the United States, with some from the United Kingdom and Australia, who opted into the study by providing informed consent and paying a fee to offset the cost of processing and sequencing. The data, both 16S rRNA gene sequencing and self-reported meta-data, are publicly available in The European Bioinformatics Institute repository under the accession ERP012803.

We focus on the fecal microbiome samples and restrict our samples to individuals who reported not having inflammatory bowel disease or diabetes. We filtered data that showed that OTUs were removed if present in fewer than 37% of samples. Samples were removed if total sequencing depth fell below 10,800 sequence reads [15]. After filtering, there were 127 OTUs and 289 samples remaining in our dataset.

To illustrate our methods, the analysis results of two OTUs are shown only in this section. As shown in Figure 2, the sample density functions of *Faecalibacterium* (OTU1) and *Bacteroides* (OTU2) from AGP are fitted well by the marginal distributions of MSIG distribution, where the parameters $(\alpha, \beta_1, \beta_2)$ are estimated to be $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2) = (0.351, 0.089, 0.036)$. The parameters for the fitted Dirichlet distribution are estimated to be $(\hat{\alpha}_1, \hat{\alpha}_2) = (0.0791, 0.0562)$. Figure 3 shows the contour plots from the estimated Dirichlet distribution (DD) and MSIG, respectively. The p-values for the GOF test are 0.001 versus 0.567 between DD and MSIG, which show that the proposed MSIG is goodness-of-fitting the data. Furthermore, the Kendall's correlation coefficient between OTU1 and OTU2 is estimated to be $\tau = -0.061$ based on the samples, but the corresponding Kendall's τ are calculated to be -0.102 with DD and -0.057 with MSIG, respectively.

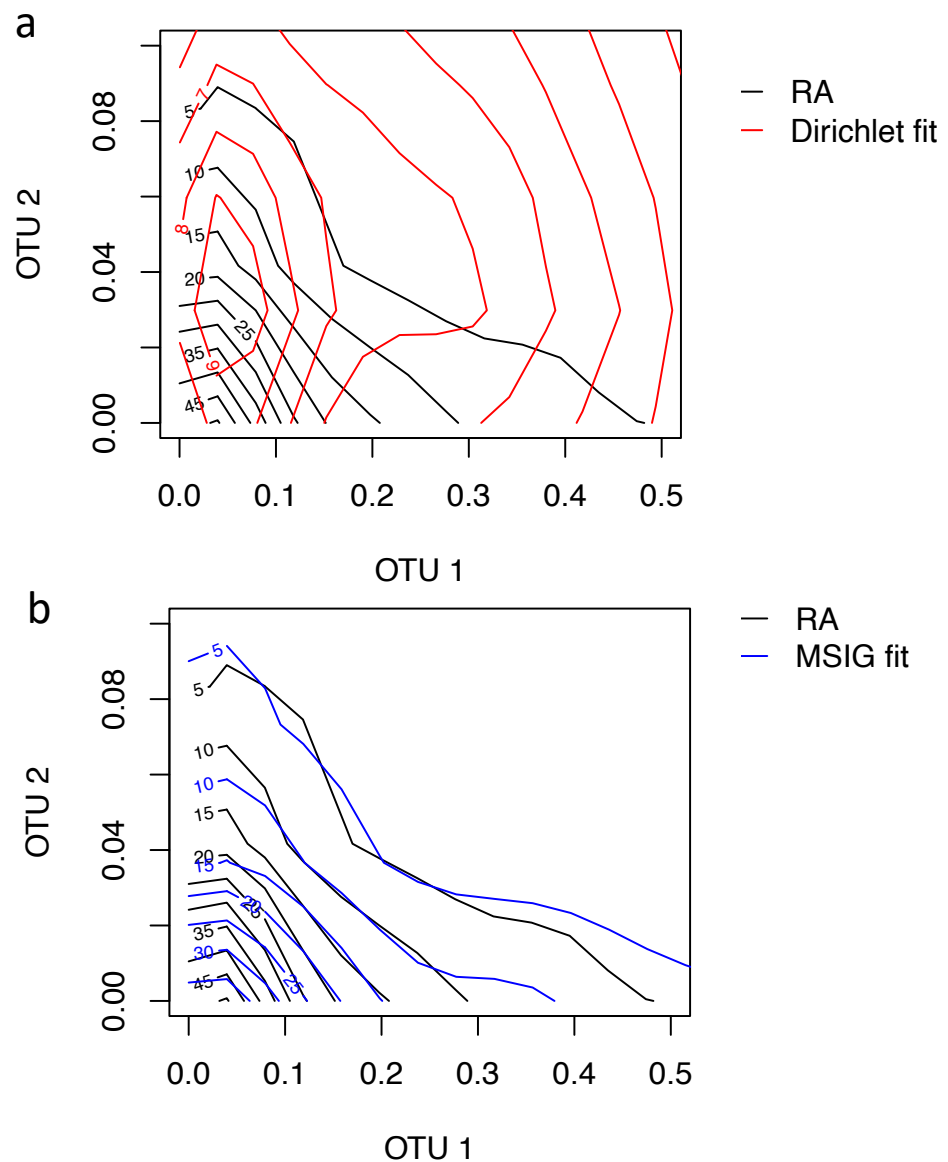


Figure 3. A comparison of contour plots between MSIG and Dirichlet distribution. The black lines are the estimated contour plots of samples. The red lines are the contour plots of the estimated Dirichlet distribution (a). The blue lines are the contour plots of the estimated MSIG (b).

Black lines in both panels are the estimated contour plots for samples. Red lines in the left panel are the contour plots of the estimated Dirichlet distribution. Blue lines in the

right panel are the contour plots of the estimated MSIG. These figures show that MSIG is a goodness-of-fit for the data.

Some individual taxa can not be fitted by the marginal distribution of MSIG. Figure 4a,b shows that taxa Clostridium (OTU3) and Escherichia (OTU4) are fitted by the two-parametric simplex distributions [13], where the corresponding parameters are estimated to be $(\hat{m}_1, \hat{s}_1) = (0.249, 5.801)$ and $(\hat{m}_2, \hat{s}_2) = (0.125, 8.610)$. The Kendall's correlation coefficient between OTU3 and OTU4 is estimated to be $\tau = 0.448$ based on the samples. We use the Gaussian copula (GC) to fit the joint distribution of OTU3 and OTU4 with $\hat{\rho} = 0.071$. The contour plots in Figure 4c show that the proposed multivariate distribution via Gaussian copula fitted the joint distribution of OTU3 and OTU4 well, and the p -value for GOF test is 0.647. However, Figure 1 (the right panel) shows that the contour plots of samples are a departure from that of the Dirichlet distribution, of which the corresponding parameters are estimated to be $(\hat{\alpha}_1, \hat{\alpha}_2) = (0.444, 0.556)$; the Kendall's $\tau = -0.407$, and the p -value for the GOF test is 0.001

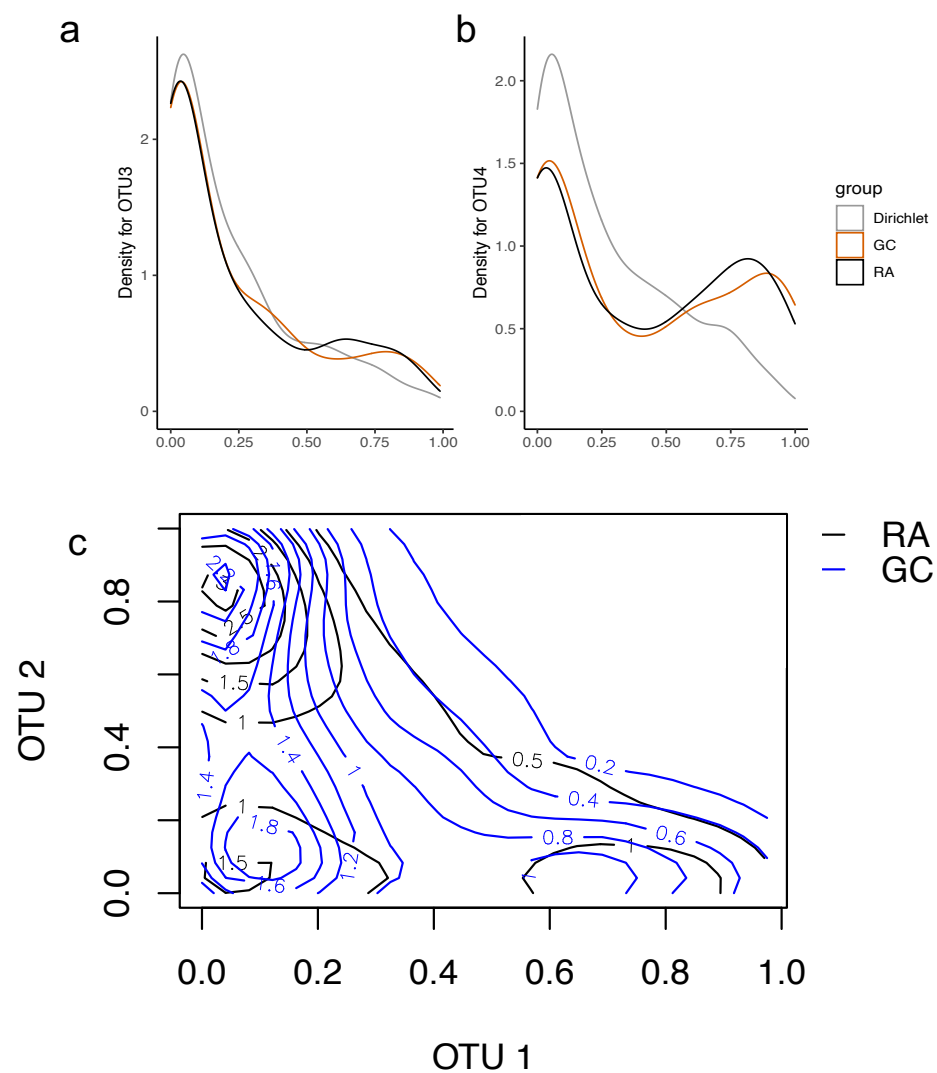


Figure 4. Multivariate distribution via Gaussian copula of taxa Clostridium and Escherichia. The upper panel (a,b) illustrates the estimated marginal density functions: the black line is estimated by the samples, the gray line by Dirichlet distributions, and the orange line is by Gaussian copulas. The lower panel (c) shows the contour plots of the estimated Gaussian copulas.

5. Discussions

The human microbiome has been recently shown to be associated with disease risks and has important implications in risk stratification and precision medicine. The existing methods for modeling microbiome dynamics focus on abundant taxa and ignore rare taxa. Modeling both will be crucial to provide information about microbial causal interactions and information for designing effective large-scale epidemiology. In this paper, we propose two classes of multivariate distributions to characterize the dependence structures among taxa in microbiome studies. Multivariate distributions on simplex via inverse Gamma distributions (MSIG) are useful in regression analysis due to the close-form of their density functions, but all of the Kendall's correlation coefficients between two taxa are assumed to be the same. The copula techniques are very useful to characterize the dependence structures of random variables. We proposed two most popular copulas, Gaussian and Farlie–Gumbel–Morgenstern copulas, to construct the multivariate distributions of taxa. The data analysis shows that our methods perform well.

In this paper, we consider only the dependence structures of taxa at the fixed time points. Some new methodologies for longitudinal observations to study the dynamics of the microbiome community have to be developed, which will be investigated in the future.

Author Contributions: Conceptualization, H.-B.F. and P.A.; methodology, S.L.; validation, J.S.; formal analysis, S.L. and J.S.; writing—original draft preparation, S.L.; and writing—review and editing, H.-B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Cancer Institute (NCI) grant P30CA051008 and the intramural Research Program of National Institute of Health.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lederberg, J.; McCray, A.T. “Ome Sweet” Omics—a genealogical treasury of words. *Scientist* **2001**, *15*, 8.
2. McDonald, D.; Vázquez-Baeza, Y.; Koslicki, D.; McClelland, J.; Reeve, N.; Xu, Z.; Gonzalez, A.; Knight, R. Striped UniFrac: Enabling microbiome analysis at unprecedented scale. *Nat. Methods* **2018**, *15*, 847–848. [[CrossRef](#)] [[PubMed](#)]
3. Turnbaugh, P.J.; Ley, R.E.; Hamady, M.; Fraser-Liggett, C.M.; Knight, R.; Gordon, J.I. The human microbiome project. *Nature* **2007**, *449*, 804–810. [[CrossRef](#)] [[PubMed](#)]
4. McMurdie, P.J.; Holmes, S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput. Biol.* **2014**, *10*, e1003531. [[CrossRef](#)] [[PubMed](#)]
5. Paulson, J.N.; Stine, O.C.; Bravo, H.C.; Pop, M. Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* **2013**, *10*, 1200–1202. [[CrossRef](#)]
6. Peng, S.; Yin, J.; Liu, X.; Jia, B.; Chang, Z.; Lu, H.; Jiang, N.; Chen, Q. First insights into the microbial diversity in the omasum and reticulum of bovine using Illumina sequencing. *J. Appl. Genet.* **2015**, *56*, 393–401. [[CrossRef](#)]
7. White, J.R.; Nagarajan, N.; Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput. Biol.* **2009**, *5*, e1000352. [[CrossRef](#)]
8. Greenblum, S.; Turnbaugh, P.J.; Borenstein, E. Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 594–599. [[CrossRef](#)]
9. Scher, J.U.; Abramson, S.B. The microbiome and rheumatoid arthritis. *Nat. Rev. Rheumatol.* **2011**, *7*, 569–578. [[CrossRef](#)]
10. Taneja, V. Arthritis susceptibility and the gut microbiome. *FEBS Lett.* **2014**, *588*, 4244–4249. [[CrossRef](#)]
11. Harrison, J.G.; Calder, W.J.; Shastry, V.; Buerkle, C.A. Dirichlet-multinomial modeling outperforms alternatives for analysis of microbiome and other ecological count data. *Mol. Ecol. Resour.* **2020**, *20*, 481–497. [[CrossRef](#)] [[PubMed](#)]
12. Metwally, A.A.; Aldirawi, H.; Yang, J. A review on probabilistic models used in microbiome studies. *Commun. Inf. Syst.* **2018**, *18*, 173–191. [[CrossRef](#)]
13. Qiu, Z.; Song, P.X.K.; Tan, M. Simplex mixed-effects models for longitudinal proportional data. *Scand. J. Stat.* **2008**, *35*, 577–596. [[CrossRef](#)]
14. Friedman, J.; Alm, E.J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **2012**, *8*, e1002687. [[CrossRef](#)]

15. Kurtz, Z.D.; Müller, C.L.; Miraldi, E.R.; Littman, D.R.; Blaser, M.J.; Bonneau, R.A. Sparse and compositionally robust inference of microbial ecological networks. *PLoS Comput. Biol.* **2015**, *11*, e1004226. [\[CrossRef\]](#)
16. Aitchison, J. The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B* **1982**, *44*, 139–160. [\[CrossRef\]](#)
17. Feng, C.; Wang, H.; Lu, N.; Tu, X.M. Log transformation: Application and interpretation in biomedical research. *Stat. Med.* **2013**, *32*, 230–239. [\[CrossRef\]](#)
18. Fang, H.B.; Fang, K.T.; Kotz, S. The meta-elliptical distributions with given marginals. *J. Multivar. Anal.* **2002**, *82*, 2017. [\[CrossRef\]](#)
19. Carlton, M.A. A family of densities derived from the three-parameter Dirichlet process. *J. Appl. Probab.* **2002**, *39*, 764–774. [\[CrossRef\]](#)
20. Favaro, S.; Hadjicharalambous, G.; Prünster, I. On a class of distributions on the simplex. *J. Stat. Plan. Inference* **2011**, *141*, 2987–3004. [\[CrossRef\]](#)
21. Erdélyi, A.; Karlin, S. Total Positivity, Vol. I Stanford University Press; London: Oxford University Press, xi 576 pp. 166s. 6d. *Proc. Edinb. Math. Soc.* **1968**, *17*, 110. [\[CrossRef\]](#)
22. Kruskal, W.H. Ordinal measures of association. *J. Am. Stat. Assoc.* **1958**, *53*, 814–861. [\[CrossRef\]](#)
23. Nelsen, R.B. *An Introduction to Copulas*; Springer: Berlin, Germany, 2006.
24. Friedman, J. On multivariate goodness-of-fit and two-sample testing. In *Technical Report SLAC PUB-10325*; University of Stanford Statistics Department: Stanford, CA, USA, 2003.
25. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A kernel two-sample test. *J. Mach. Learn. Res.* **2012**, *13*, 723–773.
26. Liu, L.; Shih, Y.-C.T.; Strawderman, R.L.; Zhang, D.; Johnson, B.A.; Chai, H. Statistical analysis of zero-inflated nonnegative continuous data: A review. *Stat. Sci.* **2019**, *34*, 253–279. [\[CrossRef\]](#)
27. Dall’Áglio, G.; Kotz, S.; Salinetti, G. *Advance in Probability Distributions with Given Marginals*; Kluwer Academic: Dordrecht, The Netherlands, 1991.
28. Sklar, A. Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **1959**, *8*, 229–231.
29. Jaworski, P.; Durante, F.; Härdle, W.; Rychlik, T. *Copula Theory and Its Applications*; Springer: Berlin/Heidelberg, Germany, 2010.
30. Johnson, N.L.; Kotz, S. On some generalized Farlie–Gumbel–Morgenstern distributions. *Commun. Stat.* **1975**, *4*, 415–427. [\[CrossRef\]](#)
31. Navarro, J.; Durante, F. Copula-based representations for the reliability of the residual lifetimes of coherent systems with dependent components. *J. Multivar. Anal.* **2017**, *158*, 87–102. [\[CrossRef\]](#)
32. Ota, S.; Kimura, M. Effective estimation algorithm for parameters of multivariate Farlie–Gumbel–Morgenstern copula. *Jpn. J. Stat. Data Sci.* **2021**, *4*, 1049–1078. [\[CrossRef\]](#)
33. Shih, J.H.; Chang, Y.T.; Konno, Y.; Emura, T. Estimation of a common mean vector in bivariate meta-analysis under the FGM copula. *Statistics* **2019**, *53*, 673–695. [\[CrossRef\]](#)
34. Joe, H.; Xu, J.J. The estimation method of inference functions for margins for multivariate models. In *Technical Report 166*; Department of Statistics, University of British Columbia: Vancouver, BC, Canada, 1996.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.