# Power-Enhanced Two-Sample Mean Tests for High-Dimensional Compositional Data with Application to Microbiome Data Analysis*

Danning Li[1], Lingzhou Xue[2], Haoyi Yang[2], and Xiufan Yu[3]

[1]Northeast Normal University, [2]Penn State University,

and [3]University of Notre Dame

First Version: August 2023; This Version: April 2024

**Abstract**

Testing differences in mean vectors is a fundamental task in the analysis of high-dimensional compositional data. Existing methods may suffer from low power if the underlying signal pattern is in a situation that does not favor the deployed test. In this work, we develop two-sample power-enhanced mean tests for high-dimensional compositional data based on the combination of $p$-values, which integrates strengths from two popular types of tests: the maximum-type test and the quadratic-type test. We provide rigorous theoretical guarantees on the proposed tests, showing accurate Type-I error rate control and enhanced testing power. Our method boosts the testing power towards a broader alternative space, which yields robust performance across a wide range of signal pattern settings. Our theory also contributes to the literature on power enhancement and Gaussian approximation for high-dimensional hypothesis testing. We demonstrate the performance of our method on both simulated data and

*The preliminary result of this paper was included in the National Institutes of Health (NIH) grant proposal (1R01GM152812). Lingzhou Xue and Xiufan Yu's research has been supported in part by the NSF grant DMS-1811552 and NIH grant 1R01GM152812.

real-world microbiome data, showing that our proposed approach improves the testing power substantially compared to existing methods.

**Key Words**: High-dimensional hypothesis testing, Cauchy combination test, Fisher's method, Microbiome compositional data, Power enhancement.

# 1 Introduction

Compositional data analysis has been receiving increasing attention in several research fields such as business analytics, ecology, and microbiome over the past few years. Especially, high-dimensional compositional data are becoming increasingly available in microbiome research, ecology, and business analytics. One common problem of interest in analyzing high-dimensional compositional data is to test for differences in composition between different samples or experimental groups. It naturally formulates a hypothesis test on the mean of compositional data.

Over the past decades, researchers have devoted significant efforts to the development of testing procedures for high-dimensional mean vectors (Bai and Saranadasa, 1996; Chen and Qin, 2010; Cai et al., 2014; Wang et al., 2015; Xu et al., 2016; Liu et al., 2022, 2024; Yu et al., 2023). These methods cannot be directly applied to compositional data due to the compositional nature of the data. Compositional data are characterized by the fact that the components of the data represent proportions that sum to one. The sum-to-unity constraint implies that the covariance matrix of the data is singular, which violates the eigenvalue assumptions required by most existing high-dimensional mean tests. Directly applying methods designed for unconstrained data to compositional data can lead to inaccurate or misleading conclusions (Aitchison, 1982; Li, 2015).

A variety of mean tests on compositional data have emerged since Aitchison (1982), such as Srivastava et al. (2007); Cuesta-Albertos et al. (2009); Tsagris et al. (2017); Cao et al. (2018) and others. In particular, Cao et al. (2018) extended the *maximum-type tests* in Cai et al. (2014) for high-dimensional compositional data. In recent years, it has been acknowledged that the testing power of many high-dimensional testing methods often depends on the sparsity level of the signal (Fan et al., 2015; Li and Xue, 2015; Xu et al., 2016; He et al., 2021; Chen et al., 2023a,b; Yu et al., 2023, 2024a,b), which is the difference be-

tween two mean vectors in the problem of mean tests. In specific, the tests based on the $L_\infty$-norm of the signal (called the *maximum-type tests*) tend to be powerful under sparse alternatives (Cai et al., 2014; Cao et al., 2018) when only a small proportion of the covariates drive differentiation between the two mean vectors, while the tests based on the (squared) $L_2$-norm of the signal (called the *quadratic-type tests*) appears to exhibit strong performance under dense alternatives (Chen and Qin, 2010; Chen et al., 2019) when the signal resides in a large number of covariates though weak within each covariate. In practical data analysis, the pattern of the underlying signal is unknown in advance and misspecification can limit discovery power. It is of great importance to develop robust testing procedures that remain powerful under a variety of signal patterns.

In this work, we propose a new power-enhanced two-sample mean test for high-dimensional compositional data. After extending the *quadratic-type tests* in Chen and Qin (2010) for high-dimensional compositional data, we prove that the *quadratic-type tests* and the *maximum-type tests* in Cao et al. (2018) are asymptotically independent when the dimension can be on the nearly exponential order of the sample size. Due to this important result, we propose the use of Fisher's method (Fisher, 1925) or the Cauchy combination test (Liu and Xie, 2020) to relax the assumptions on the signal density and improve testing power. We further show that the proposed power-enhanced tests asymptotically achieve the target size and have consistent asymptotic power under mild theoretical conditions. Moreover, we examine the finite-sample performance through numerical studies and a real-world application to a human microbiome study examining changes in the host-microbiome community in individuals with inflammatory bowel disease (IBD).

Our theory also contributes to the literature on power enhancement and Gaussian approximation for high-dimensional hypothesis testing. Specifically, we relax the Gaussian or sub-Gaussian assumption and do not need the pseudo-independence structure used in Chen and Qin (2010); Xu et al. (2014); Li and Xue (2015); Yu et al. (2023) and many others. We use the high-dimensional Gaussian approximation theory, instead of the martingale limit theory as in Chen and Qin (2010); Li and Xue (2015); Yu et al. (2023), to prove the asymptotic null distribution of the proposed power-enhanced tests. Also, our theory can deal with the challenges caused by the singular variance of high-dimensional compositional

data. Moreover, although the *quadratic-type* and *maximum-type tests* do not satisfy the bivariate normality assumption, we provide some insights on the non-asymptotic Cauchy approximation for the tail of the null distribution (see Remark 4 in Section 3), which may have its independent significance for future research.

The rest of this paper is organized as follows. Section 2 introduces our proposed power-enhanced mean tests for high-dimensional compositional data. Section 3 presents the asymptotic independence of maximum and quadratic-type tests and the asymptotic properties of the proposed tests. Section 4 carries out simulations to validate the finite-sample properties, and Section 5 applies the test to real-world microbiome datasets. Section 6 concludes the paper with a brief discussion. The proofs are presented in the supplement.

# 2 Methodology

## 2.1 Mean Tests for High-Dimensional Compositional Data

Let $\xi^{(k)} = (\xi_1^{(k)}, \ldots, \xi_{n_k}^{(k)})^\top \in \mathbb{R}^{n_k \times p}$ denote the individual compositional data matrices for each group $k \in \{1, 2\}$. By the compositional nature of the observations, each row of $\xi^{(k)}$ sums to one. Therefore, each row $\xi_i^{(k)}$ lie in the Aitchison simplex (Aitchison, 1982):

$$\mathcal{S}^{p-1} = \{(\xi_{i1}, .., \xi_{ip}) : \xi_{ij} > 0 \text{ for } j \in \{1, ..., p\}, \sum_{j=1}^{p} \xi_{ij} = 1\}.$$

Assume $\xi_i^{(k)} \overset{iid}{\sim} F_\xi^{(k)}$, where $F_\xi^{(k)}$ is a distribution with mean $\mu_\xi^{(k)} \in \mathbb{R}^p$ and covariance $\Sigma_\xi^{(k)} \in \mathbb{R}^{p \times p}$. A natural question of interest is to test for differences in microbial community compositions between different groups. The compositional constraint imposes dependencies between the multiple components of the data, which makes classical statistical tests inapplicable. To be specific, traditional testing strategies place restrictions on the covariance structure of the test samples. For example, most tests require the covariance matrix to be positive definite, which is violated in the compositional space due to the singularity in the covariance matrix of compositional data. The complex composition of the data presents a significant challenge.

To assuage this issue, we exploit the framework of compositional data analysis to shift

the assumption burden off of the observed compositional data and into a latent variable space. As a common practice in compositional data analysis, we assume the compositional variables are driven by a set of latent variables, which are known as the basis, denoted by $\eta^{(k)} = (\eta_{ij}^{(k)}) \in \mathbb{R}^{n_k \times p}$ with $\eta_{ij}^{(k)} > 0$ for $k = 1, 2$. Given the basis, we can reconstruct the associated compositional data via normalization as follows:

$$\xi_{ij}^{(k)} = \frac{\eta_{ij}^{(k)}}{\sum_{j=1}^{p} \eta_{ij}^{(k)}} \quad \text{for } \; i = 1, \ldots, n_k; \; j = 1, \ldots, p; \; k = 1, 2. \tag{2.1}$$

Let $\Delta^{(k)} = (\delta_{ij}^{(k)}) \in \mathbb{R}^{n_k \times p}$ in which $\delta_{ij}^{(k)} = \log \eta_{ij}^{(k)}$ denote the *log-basis* variables. We assume that $\delta_i^{(k)} \overset{iid}{\sim} F_\delta^{(k)}$, where $F_\delta^{(k)}$ is distribution with associated mean vector $\nu^{(k)} = \{\nu_1^{(k)}, .., \nu_p^{(k)}\}$ and covariance $\Omega^{(k)}$. A natural test in this log-basis space is to simply test the equality of $\nu^{(1)}$ and $\nu^{(2)}$; however, different basis vectors $\eta^{(k)}$ can lead to identical compositional $\xi^{(k)}$. By the compositional relationship (2.1), given $\xi$, one is only able to recover $\eta$ up to a multiplicative constant. In Aitchison (1982), this indicates all the bases arising from a given compositional vector can be enumerated as $\mathcal{B}(\xi) = \{t\xi; t > 0\}$. Alternatively, this yields an additive many-to-one relationship in the log-basis space $\mathcal{A}(\xi) = \{\log \xi + c1_p; c \in \mathbb{R}\}$. Thus, the natural testing scheme must be adapted slightly to account for this additive relationship. Two log-basis vectors $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are compositionally equivalent if their components only differ by a constant $c \in \mathbb{R}$. That is both log-bases lie within the same equivalence class $\mathcal{A}(\xi)$ (Cao et al., 2018). Let $1_p \in \mathbb{R}^p$ denote a vector of only 1s. This yields the compositional testing framework:

$$H_0 : \nu^{(1)} = \nu^{(2)} + c1_p \quad vs \quad H_1 : \nu^{(1)} \neq \nu^{(2)} + c1_p. \tag{2.2}$$

To construct a test on (2.2), we need to transform the compositional data into an easy-to-operate form. As discussed above, operating on raw compositional variables is challenging in practice. The downstream effects of the compositional constraint are often very restrictive as individual components are bounded between $(0, 1]$. As such, a common practice is to first transform the compositional data via a log-ratio transformation to relax the range constraint on individual components. One of the most commonly used log-ratio transformations is the

*centered log-ratio (clr)* transformation. For the observed compositional vectors $\xi_i^{(1)}$ and $\xi_i^{(2)}$, the associated clr-transformed variables are $X_i = clr(\xi_i^{(1)}) = \log\left(\frac{\xi_{i1}^{(1)}}{g(\xi_i^{(1)})}, ..., \frac{\xi_{ip}^{(1)}}{g(\xi_i^{(1)})}\right)$, for $i = 1, \ldots, n_1$, and $Y_i = clr(\xi_i^{(2)}) = \log\left(\frac{\xi_{i1}^{(2)}}{g(\xi_i^{(2)})}, ..., \frac{\xi_{ip}^{(2)}}{g(\xi_i^{(2)})}\right)$, for $i = 1, \ldots, n_2$, where $g(\xi_i^{(k)}) = (\prod_{j=1}^{p} \xi_{ij}^{(k)})^{\frac{1}{p}}$ denotes the geometric mean of $\xi_i^{(k)}$. The centered log-ratio transformation is appealing for several reasons. The individual components of $X$ and $Y$ are no longer constrained within $(0, 1]$ and are instead $X_{ij}$ and $Y_{ij} \in \mathbb{R}$ allowing us to employ traditional statistical machinery. Further, the unit constraint on $\xi$ is now a sum-to-zero constraint on the components of $X$ and $Y$. However, most useful for testing purposes is the distributional relationship between $X$ or $Y$ and the log-basis $\boldsymbol{\delta}$. The centered log-ratio transformation is scale-invariant, allowing us to substitute $(\xi_i^{(1)}, \xi_i^{(2)})$ for $(X_i, Y_i)$ and yielding the following relationship:

$$X_i = G\delta_i^{(1)},\ Y_i = G\delta_i^{(2)} \tag{2.3}$$

where $G = I_p - \frac{1}{p} 1_p 1_p^T$ with $I_p$ denoting the $p \times p$ identity matrix and $1_p \in \mathbb{R}^p$ denoting a vector of only 1s. Thus the distributions of $X_i$ and $Y_i$ can be completely characterized by the distributions of $\delta_i^{(1)}$ and $\delta_i^{(2)}$ via linear transformations. Suppose $X_i \overset{iid}{\sim} H^{(1)}$ and $Y_i \overset{iid}{\sim} H^{(2)}$, where $H^{(k)}$ characterizes a distribution with the centered log-ratio mean vector and centered log-ratio covariance matrix such that:

$$\mu^{(k)} = G\nu^{(k)},\ \Sigma^{(k)} = G\Omega^{(k)}G^\top. \tag{2.4}$$

By construction, $G$ is a rank $p-1$ matrix with associated null space $\mathcal{N}(G) = \{x \in \mathbb{R}^p : Gx = 0\} = \{c1_p : c \in \mathbb{R}\}$. Using this relationship, the test on (2.2) is equivalent to a two-sample test of means as follows (Cao et al., 2018):

$$H_0 : \mu^{(1)} = \mu^{(2)} \quad \text{versus} \quad H_1 : \mu^{(1)} \neq \mu^{(2)}. \tag{2.5}$$

Cao et al. (2018) proposed the maximum compositional equivalence test assuming that $\Sigma^{(1)} = \Sigma^{(2)}$ and proved its asymptotic null distribution as a Gumbel distribution as $n_1, n_2, p \to$

$\infty$.

## 2.2 The Proposed Power-Enhanced Mean Tests

The next step is to construct a more powerful test for the problem of interest (2.5) against a broader alternative space. It has been well-studied that the performance of various tests fundamentally relies on the underlying signal sparsity pattern (Fan et al., 2015; Li and Xue, 2015; Xu et al., 2016; He et al., 2021; Yu et al., 2023, 2024a,b). When the signal pattern in $\mu^{(1)} - \mu^{(2)}$ varies, different tests may yield distinct performance. Broadly speaking, two types of test statistics are prevalent in high-dimensional hypothesis tests: the *maximum-type tests* and the *quadratic-type tests*. The maximum-type mean tests construct the test statistics based on estimates of the $L_\infty$-norm of the difference in mean vectors, i.e., $\|\mu^{(1)}-\mu^{(2)}\|_\infty$. The quadratic-type mean tests design the test statistics by utilizing estimates of the squared $L_2$-norm of the mean difference, i.e., $\|\mu^{(1)} - \mu^{(2)}\|_2^2$. The maximum-type tests tend to be more powerful than quadratic-type tests under the sparse alternatives (Cai et al., 2014; Cao et al., 2018) when $\mu^{(1)} - \mu^{(2)}$ have only a few non-zero components that distinguish between groups, whereas the quadratic-type tests are more powerful under the dense alternatives (Chen and Qin, 2010; Chen et al., 2019, 2023a) which assume the differentiation between groups is caused by several components. When the sparsity of the alternative hypothesis is well-suited to the choice of the test statistic, there is a gain in discovery power. However, mismatches between sparsity assumptions and test statistic choice can negatively impact power substantially.

We propose combining maximum-type and quadratic-type test statistics to develop a testing framework that is more robust to improper sparsity assumptions on alternative hypotheses. Thus, our proposed strategy is more flexible in practical scenarios where there may not be strong evidence to assume a given sparsity level apriori. To begin with, we first detail the maximum-type and quadratic-type test statistics of interest.

To encompass scenarios of sparse alternatives, we follow the philosophy of Cao et al. (2018) to employ the following maximum-type statistic, where we do not need to assume

$\Sigma^{(1)} = \Sigma^{(2)}$:

$$M_{n_1,n_2} = n_1 \max_{1\le j\le p} \frac{(\overline{X}_j - \overline{Y}_j)^2}{\hat{\gamma}_j}. \tag{2.6}$$

Here, $\overline{X}_j = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{ij}$ and $\overline{Y}_j = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{ij}$ are the sample mean of the clr transformed samples. Define $\widehat{\gamma}_j = \frac{\sum_{i=1}^{n_1}(X_{ij}-\overline{X}_j)^2}{(n_1-1)} + \frac{\sum_{k=1}^{n_2}(Y_{kj}-\overline{Y}_j)^2}{n_2(n_2-1)/n_1}$ as the corresponding sample variance after the centered log-ratio transformation, which is an estimate of $\gamma_j = \sigma_{jj}^{(1)} + \frac{n_1}{n_2}\sigma_{jj}^{(2)}$. Similar to Cao et al. (2018), it can be proved that under the null hypothesis, $M_{n_1,n_2} - 2\log p + \log\log p$ converges to a Gumbel distribution as $n_1, n_2, p \to \infty$. The associated $\alpha$-level test is defined as

$$\Phi_\alpha^M = I(M_{n_1,n_2} \ge q_\alpha^M + 2\log p - \log\log p), \tag{2.7}$$

where $I(\cdot)$ denotes the indicator function and $q_\alpha^M$ is the upper $\alpha$-quantile of the Gumbel distribution. Therefore, the null hypothesis in (2.5) is rejected by the maximum-type test when $\Phi_\alpha^M = 1$. As noted previously, the statistic $M_{n_1,n_2}$ possesses high power in the sparse alternative setting but as explored in Section 4, this gain in power rapidly disappears as the signal becomes denser.

To account for the dense setting, we consider the quadratic-type statistic $Q_{n_1,n_2}$ developed by Chen and Qin (2010) to the clr transformed samples $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$. Let

$$\begin{aligned}
T_{n_1,n_2} &= \frac{\sum_{i\ne j} X_i^T X_j}{n_1(n_1-1)} + \frac{\sum_{i\ne j} Y_i^T Y_j}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_i^T Y_j}{n_1 n_2}, \\
\widehat{\sigma}^2_{n_1,n_2} &= \frac{2}{n_1(n_1-1)}\widehat{\mathrm{tr}((\Sigma^{(1)})^2)} + \frac{2}{n_2(n_2-1)}\widehat{\mathrm{tr}((\Sigma^{(2)})^2)} + \frac{4}{n_1 n_2}\widehat{\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})},
\end{aligned}$$

where the form of $\widehat{\sigma}^2_{n_1,n_2}$ is presented in the supplement. It can be verified that $T_{n_1,n_2}$ and $\widehat{\sigma}^2_{n_1,n_2}$ are unbiased estimates of $\|\mu^{(1)} - \mu^{(2)}\|_2^2$ and the variance of $T_{n_1,n_2}$ (i.e., $\sigma^2_{n_1,n_2} = \frac{2}{n_1(n_1-1)}\mathrm{tr}((\Sigma^{(1)})^2) + \frac{2}{n_2(n_2-1)}\mathrm{tr}((\Sigma^{(2)})^2) + \frac{4}{n_1 n_2}\mathrm{tr}(\Sigma^{(1)}\Sigma^{(2)})$). The test statistic is defined as

$$Q_{n_1,n_2} = \frac{T_{n_1,n_2}}{\widehat{\sigma}_{n_1,n_2}}. \tag{2.8}$$

The associated $\alpha$-level test is, therefore, defined as

$$\Phi_\alpha^Q = I(Q_{n_1,n_2} \geq q_\alpha^Q), \tag{2.9}$$

where $q_\alpha^Q$ denotes the upper $\alpha$-quantile of the standard Gaussian distribution. As such, the null hypothesis in (2.5) is rejected by the quadratic-type test when $\Phi_\alpha^Q = 1$. As shown in Section 4, the quadratic-type test statistic $Q_{n_2,n_2}$ achieves satisfactory power when the underlying signal density is dense and the associated performance suffers as the signal becomes more sparse.

In what follows, we leverage the power of both test statistics without placing stringent assumptions on signal sparsity. To this end, we construct our power-enhanced tests by employing $p$-value combination approaches, including Fisher's combination method (Fisher, 1925) and the Cauchy combination method (Liu and Xie, 2020). These combination methods aggregate information from the maximum-type test and the quadratic-type test to combine their respective strengths.

Let $p_M$ be the $p$-value of $M_{n_1,n_2}$ and $p_Q$ the $p$-value of $Q_{n_2,n_2}$. Fisher combination test statistic, denoted by $F_{n_1,n_2}$, combines both $p$-values as

$$F_{n_1,n_2} = -2(\log p_M + \log p_Q). \tag{2.10}$$

Theorem 1 of Section 3 proves the maximum-type statistic $M_{n_1,n_2}$ and the quadratic-type statistic $Q_{n_1,n_2}$ are asymptotically independent. Thus, under the null hypothesis, $F_{n_1,n_2}$ converges to the $\chi_4^2$ distribution as $n_1, n_2, p \to \infty$. The associated $\alpha$-level test is, therefore, defined as

$$\Phi_\alpha^F = I(F_{n_1,n_2} \geq q_\alpha^F), \tag{2.11}$$

where $q_\alpha^F$ is the upper $\alpha$-quantile of the $\chi_4^2$ distribution. Fisher combination test rejects the null hypothesis in (2.5) when $\Phi_\alpha^F = 1$.

We denote the Cauchy combination test statistic as $C_{n_1,n_2}$ and define it as follows

$$C_{n_1,n_2} = \omega_M \tan\{(0.5 - p_M)\pi\} + \omega_Q \tan\{(0.5 - p_Q)\pi\}, \tag{2.12}$$

where $\omega_M$ and $\omega_Q$ are non-negative weights for the maximum-type and quadratic-type test statistics respectively, and $\omega_M + \omega_Q = 1$. Under the null distribution, $p_M$ and $p_Q$ follow a Unif$(0, 1)$ distribution, thus $\tan\{(0.5 - p_M)\pi\}$ and $\tan\{(0.5 - p_Q)\pi\}$ follow a standard Cauchy distribution. Together with the asymptotic independence shown in Theorem 1, $C_{n_1,n_2}$ converges to a standard Cauchy distribution as $n_1, n_2, p \to \infty$. Therefore, the associated $\alpha$-level test is

$$\Phi_\alpha^C = I(C_{n_1,n_2} \geq q_\alpha^C), \tag{2.13}$$

where $q_\alpha^C$ is the upper $\alpha$-quantile of a standard Cauchy distribution. The Cauchy combination test rejects the null hypothesis (2.5) when $\Phi_\alpha^C = 1$.

As we will study theoretical properties in Section 3 and numerical properties in Section 4, both power-enhanced tests retain the appropriate $\alpha$-level type I error rate, and achieve improved power than the maximum and quadratic tests while agnostic to the underlying signal density.

# 3 Theoretical Properties

In this section, we first introduce four assumptions and then present the theoretical properties.

**Assumption 1.** For $k \in \{1, 2\}$, let $\delta^{(k)} = (\delta_1^{(k)}, \ldots, \delta_p^{(k)})$ be a $p$-dimensional random vector with mean $\nu^{(k)}$ and covariance $\Omega^{(k)}$ and satisfy the following conditions:

(i) there is a constant $K_4$ such that, for any $\alpha \in R^p$,

$$\{E|\alpha'(\delta^{(k)} - \nu^{(k)})|^4\}^{1/4} \leq K_4\{E|\alpha'(\delta^{(k)} - \nu^{(k)})|^2\}^{1/2}.$$

(ii) $\mathrm{E}(\max_j |\delta_j^{(k)}|^3) \leq M^3(\log p)^{\frac{3}{2}}$ with $(\log p)^{10} M^6 = o(n_1 + n_2)$.

**Remark 3.1.** Assumption 1(i) on the fourth moment was used in Giessing and Fan (2020), and Assumption 1(ii) was used in Chernozhukov et al. (2017). Assumption 1 relaxes the Gaussian or sub-Gaussian assumption, and it also relaxes the linear multivariate model assumption widely used in the theoretical analysis of quadratic-type statistics including Chen and Qin (2010); Xu et al. (2014); Yu et al. (2023) and so on. Note that the pseudo-independence structure in (3.2) of Chen and Qin (2010) or (4.1) of Xu et al. (2014) does not allow for an elliptical model. Assumption 1 includes the elliptical model with a finite fourth moment as a special example, where $M$ equals to a constant $C$ (or $C\sqrt{\log p}$) when $\delta_j^{(k)}$'s follow sub-Gaussian (or sub-exponential) distributions (Chernozhukov et al., 2017).

**Assumption 2.** For $k \in \{1, 2\}$, the covariance matrix of $\delta^{(k)}$ satisfy the following conditions:

(i) there is a constant $C$ such that $1/C \leq \omega_{jj}^{(k)} \leq C$ for $1 \leq i \leq p$.

(ii) $\lambda_1(\Omega^{(k)})/\sqrt{\mathrm{tr}\{(\Omega^{(k)})^2\}} = o((\log p)^{-1-\alpha_0})$ for a constant $\alpha_0 > 0$ and $\lambda_{p-q}(\Omega^{(k)}) > 0$ with $q = o(p)$, where $\lambda_p(\Omega^{(k)}) \leq \ldots \leq \lambda_1(\Omega^{(k)})$ are the eigenvalues of $\Omega^{(k)}$.

**Remark 3.2.** Assumption 2(i) was used in Cao et al. (2018) to bound the variances away from zero and infinity. The condition $\lambda_1(\Omega^{(k)})/\sqrt{\mathrm{tr}\{(\Omega^{(k)})^2\}} \to 0$ is sufficient for establishing the central limit theorem for quadratic-type test statistics, which was used in Chen and Qin (2010). By requiring Assumption 2(ii), we can derive the asymptotic independence between maximum-type and quadratic-type test statistics without Gaussian assumption.

**Assumption 3.** As $\min\{n_1, n_2\} \to \infty$, $n_1/(n_1 + n_2) \to c$, for some constant $c \in (0, 1)$.

This is a common assumption for the theoretical analyses of high-dimensional two-sample tests, see Chen and Qin (2010); Cao et al. (2018); Chen et al. (2019); Yu et al. (2024a, 2023).

Before proceeding to the next assumption, we define some useful notations. Denote the correlation matrices of $\delta^{(k)}$ by $(\tau_{ij}^{(k)})_{p\times p}$. For any set $\mathcal{A}$, $\mathrm{card}(\mathcal{A})$ denotes the cardinality of $\mathcal{A}$. For $0 < r < 1$, let $\mathcal{V}_i(\tau, r) = \left\{1 \leq j \leq p : |\tau_{ij}^{(1)}| \geq r \text{ or } |\tau_{ij}^{(2)}| \geq r\right\}$ be the set of indices $j$ such that $\delta_j^{(k)}$ is highly correlated (whose correlation $> r$) with $\delta_i^{(k)}$ for a given $i \in \{1, \ldots, p\}$. And for any $\alpha > 0$, let $s_i(\tau, \alpha) = \mathrm{card}(\mathcal{V}_i(\tau, (\log p)^{-1-\alpha}))$, $i = 1, \ldots, p$ be the number of indices $j$ in the set $\mathcal{V}_i(\tau, (\log p)^{-1-\alpha})$. Moreover, define $\mathcal{W}(\tau, r) = \{1 \leq i \leq p : \mathcal{V}_i(\tau, r) \neq \varnothing\}$ such that, $\forall i \in \mathcal{W}(\tau, r)$, $\delta_i^{(k)}$ is highly correlated with some other variables of $\delta^{(k)}$.

**Assumption 4.** The correlation matrix of $\delta^{(k)}$ satisfies the following conditions.

(i) For $\alpha_0$ given in above and all $\kappa > 0$, $\max_{1\le i\le p, i\notin\Upsilon} s_i(\tau, \alpha_0) = o(p^\kappa)$.

(ii) There exist a constant $0 < r_0 < 1$, $\mathrm{card}(\mathcal{W}(\tau, r_0)) = 0(p)$.

**Remark 3.3.** Cao et al. (2018) obtained the limiting distribution of the maximum test statistic under the assumptions $\max_i \sum_{j=1}^p |\tau_{ij}^{(k)}|^2 \le r_2$ and $\max_{i,j} |\tau_{ij}^{(k)}| \le r_1 < 1$ (see their Conditions 2 and 3). Note that Assumption 4 includes them as the special example when $s_i(\tau, 1) = C(\log p)^2$.

After introducing these assumptions, we can present the main results in Theorem 1.

**Theorem 1.** Given Assumptions 1- 4, under the null hypothesis $H_0$, we have

$$P\left(Q_{n_1,n_2} \le x,\ M_{n_1,n_2} - 2\log p + \log\log p \le y\right) \xrightarrow{d} \Phi(x) \cdot F(y) \tag{3.1}$$

for any $x, y \in \mathbb{R}$, as $n_1, n_2, p \to \infty$. $F(y) = \exp\left(-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{y}{2}\right)\right)$ is the cdf of a Gumbel distribution, $\Phi(x)$ is the cdf of standard normal distribution.

To the best of our knowledge, this is the first proof of the asymptotic independence result of the quadratic form statistic of Chen and Qin (2010) and the maximum statistic of Cao et al. (2018) for testing two-sample mean vectors of high-dimensional compositional data. It is worth pointing out that our results allow the dimension to be on the nearly exponential order of the sample size and do not require the Gaussian or pseudo-independence assumption. Also, our theory can deal with the challenges caused by the singular variance of high-dimensional compositional data.

Given the explicit joint distribution of $Q_{n_1,n_2}$ and $M_{n_1,n_2}$, we proceed to present the asymptotic properties of our proposed Fisher's test and the Cauchy combination test.

On top of the asymptotic independence established above and by simple probability transformation, it's easy to obtain the null distribution of $F_{n_1,n_2}$ and $C_{n_1,n_2}$, and therefore, the asymptotic size of the test. Specifically, Theorem 2 proved the correct asymptotic size for Fisher's method (Fisher, 1925) and the Cauchy combination test (Liu and Xie, 2020).

**Theorem 2** (Asymptotic Size)**.** Under the same assumptions as in Theorem 1, the Fisher's combined probability test and the Cauchy combination test for high-dimensional compositional data achieve the accurate asymptotic size, that is, under the null hypothesis,

$$P\left(F_{n_1,n_2} > q_\alpha^F\right) \to \alpha \quad \text{and} \quad P\left(C_{n_1,n_2} > q_\alpha^C\right) \to \alpha \quad \text{as } n_1, n_2, p \to \infty.$$

**Remark 3.4.** In the literature, Pillai and Meng (2016) proved a surprising result that the ratio of dependent Gaussian random variables follows a standard Cauchy distribution under an arbitrary covariance matrix, and Liu and Xie (2020) studied the non-asymptotic approximation for the tail of the null distribution of the Cauchy combination test under the bivariate normality assumption with arbitrary correlation structures. Although $Q_{n_1,n_2}$ and $M_{n_1,n_2}$ do not satisfy the bivariate normality assumption, we want to provide some insights on the non-asymptotic Cauchy approximation for the tail of the null distribution of $C_{n_1,n_2}$ using the results in Theorems 1 and 2. Given the asymptotic independence result in Theorem 1, when $n_1, n_2, p$ are large enough, we have

$$\frac{P(p_Q \le \alpha, p_M \le \alpha)}{P(p_Q \le \alpha) + P(p_M \le \alpha)} \approx \frac{P(p_Q \le \alpha)P(p_M \le \alpha)}{P(p_Q \le \alpha) + P(p_M \le \alpha)} \approx \frac{\alpha}{2}. \tag{3.2}$$

Thus, when $n_1, n_2, p$ are large enough and $\alpha$ tends to 0, we have

$$P(p_Q \le \alpha, p_M \le \alpha) \approx o(P(p_Q \le \alpha) + P(p_M \le \alpha)),$$

which provides a key result in the proof of Theorem 1 in Liu and Xie (2020) (see Step 2 in the supplemental file of Liu and Xie (2020)). This result indicates that the probability of $Q_{n_1,n_2}$ and $M_{n_1,n_2}$ simultaneously reaching extreme values is dominated by the probability of one of them doing so when $n_1, n_2, p$ are large enough. As a result, when $n_1, n_2, p$ are large enough and $t$ tends to $\infty$, we can follow the proof of Liu and Xie (2020) to show that

$$\frac{P\left(C_{n_1,n_2} > t\right)}{P(W_0 > t)} \approx 1,$$

where $W_0$ is a standard Cauchy random variable. This result implies that $C_{n_1,n_2}$ can still have

an approximately Cauchy tail under the null hypothesis in the non-asymptotic setting. Such insights are supported by the numerical properties that we will explore in Section 4, where the Cauchy combination test more closely achieves the desired size than Fisher's method.

Next, we will study the asymptotic power. Let $\Omega = \Omega^{(1)} + \frac{n_1\Omega^{(2)}}{n_2}$. For any fixed $\epsilon_0 > 0$, define the dense alternative $\mathcal{G}_d$ and the sparse alternative $\mathcal{G}_s$ respectively in the following:

$$\mathcal{G}_d(\epsilon_0) = \left\{(\nu^{(1)}, \nu^{(2)}) : \frac{n_1^2\|G(\nu^{(1)} - \nu^{(2)})\|^4}{n_1(\nu^{(1)} - \nu^{(2)})'G\Omega G(\nu^{(1)} - \nu^{(2)}) + \mathrm{tr}\{(G\Omega G)^2\}} \geq \epsilon_0 \log n\right\}; \quad (3.3)$$

$$\mathcal{G}_s(\epsilon_0) = \left\{(\nu^{(1)}, \nu^{(2)}) : \max_j \frac{|(G(\nu^{(1)} - \nu^{(2)}))_j|}{\{(G\Omega G)_{jj}\}^{1/2}} \geq \sqrt{\frac{(2+\epsilon_0)\log p}{n_1}}\right\} \quad (3.4)$$

with $(G(\nu^{(1)} - \nu^{(2)}))_j$ and $(G\Omega G)_{jj}$ denote the $j$-th element of the vector $G(\nu^{(1)} - \nu^{(2)})$ and the $j$-th diagonal element of the matrix $G\Omega G$.

In what follows, we provide a new Gaussian approximation result for the quadratic-type statistics $T_{n_1,n_2}$ and $Q_{n_1,n_2}$, which will be pivotal to prove the consistent asymptotic power without assuming the pseudo-independence structure of Chen and Qin (2010).

**Proposition 1.** Given Assumptions 1(i), 2 and 3, when $K_4$ is bounded, we have

$$P\left(\frac{T_{n_1,n_2} - \|\mu^{(1)} - \mu^{(2)}\|^2}{\sigma_{n,a}} \leq x\right) \to \Phi(x) \quad \text{as } n_1, n_2, p \to \infty \quad (3.5)$$

with $\sigma_{n,a}^2 = \sigma_{n_1,n_2}^2 + 4\frac{\mu'\Sigma\mu}{n_1}$, and $\Sigma = \Sigma^{(1)} + \frac{n_1\Sigma^{(2)}}{n_2}$. If $\frac{n^2\|\mu\|^4}{\max(n\mu'\Sigma\mu, \mathrm{tr}(\Sigma^2))} \to \infty$, then

$$P\left(Q_{n_1,n_2} > z_\alpha\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty, \quad (3.6)$$

with $z_\alpha$ be the upper $\alpha$-quantile of the standard normal distribution.

**Remark 3.5.** Xu et al. (2014) and Giessing and Fan (2020, 2023) studied the Gaussian approximation for the $\ell_2$-statistics in the one-sample mean test for high-dimensional data. However, as pointed by Chen and Qin (2010), the $\ell_2$-statistics is in fact a biased estimator for $\|\mu^{(1)} - \mu^{(2)}\|^2$. As a result, Xu et al. (2014) and Giessing and Fan (2020, 2023) all required either $p/n$ to be bounded or much stronger moment conditions to handle the bias term. For example, Xu et al. (2014) used the pseudo-independence structure in the linear process

to justify their moment conditions, Giessing and Fan (2020) assumed a low-rank structure within their models, and Giessing and Fan (2023) required that $p = o(n^{\frac{1}{3}})$ with an identity covariance matrix. Our new result in Proposition 1 provides an enhanced flexibility and applicability in the high-dimensional setting where the dimension is on the nearly exponential order of the sample size.

Now, using this new Gaussian approximation result, we show that Fisher's combined probability test $F_{n_1,n_2}$ can successfully boost the power against either dense or sparse alternatives. It is known that Fisher's method enjoys the Bahadur efficiency when combining independent tests. Thus, we only focus on the asymptotic power of Fisher's method in the following theorem.

**Theorem 3** (Asymptotic Power)**.** Under the same assumptions as in Proposition 1, Fisher's method achieves consistent asymptotic power, that is, for any $\epsilon_0 > 0$,

$$\inf_{(\nu^{(1)},\nu^{(2)})\in\mathcal{G}_d(\epsilon_0)\cup\mathcal{G}_s(\epsilon_0)} P\left(F_{n_1,n_2} \geq q_\alpha^F\right) \to 1 \quad \text{as } n_1, n_2, p \to \infty.$$

# 4 Numerical Properties

In this section, we evaluate the empirical power and size under several different testing settings. There are two types of covariance structures, under three various $(n, p)$ combinations, with two different data-generating methods. Aside from our proposed Fisher-based (F) test (2.11) and Cauchy-based (C) test (2.13), we compare these methods to the Maximum (M) test (2.7) and Quadratic (Q) test (2.9). Finally, the Bonferroni (BF) method is a commonly used method for combining multiple tests and we include it as an additional benchmark.

We consider two different data-generating frameworks: Gaussian distributions and Gamma distributions. Under the Gaussian framework, we generate the log-basis vectors from

$$\delta_i^{(1)} \sim N_p(\nu^{(1)}, \Omega) \text{ (for } i = 1, \ldots, n_1), \quad \delta_i^{(2)} \sim N_p(\nu^{(2)}, \Omega) \text{ (for } i = 1, \ldots, n_2).$$

Under the Gamma framework, we generate the log-basis vectors from

$$\delta_i^{(1)} = \nu^{(1)} + FU_i^{(1)}/\sqrt{10} \text{ (for } i = 1, \ldots, n_1), \quad \delta_i^{(2)} = \nu^{(2)} + FU_i^{(2)}/\sqrt{10} \text{ (for } i = 1, \ldots, n_2),$$

where $U_i^{(k)}$ are independent standard gamma with shape parameter ten, $F = QS^{\frac{1}{2}}$ in which $Q$ and $S$ are determined by $\Omega = QSQ^T$ via singular value decomposition. Given the log-basis vectors $\delta_i^{(k)}$, the centered log-ratio transformed data is obtained as described in Section 2. That is, $X_i = G\delta_i^{(1)}$, $Y_i = G\delta_i^{(2)}$ with $G = I_p - \frac{1}{p}1_p 1_p^T$.

The above $\Omega$ are set to mimic two different dependent structures. The first is an AR(1) matrix, $\Omega^{AR} = (\Omega_{ij}^{AR})_{1\leq i,j\leq p}$ with $\Omega_{ij}^{AR} = 0.5^{|i-j|}$. The AR(1) matrix yields a dense covariance structure, though ensures that the magnitude of the covariances rapidly decays (Bickel and Levina, 2008; Cai and Liu, 2011). The second is a random sparse block covariance matrix $\Omega^{BS} = (A_1, A_2)$, where $A_1 = B + \varepsilon I_q$ in which $B$ is a symmetric matrix with the lower-triangular entries drawn from $[-1, -0.5] \cup [0.5, 1]$ uniformly with probability 0.5, and $A_2 = I_{p-q}$ where $q = \lfloor 3\sqrt{p} \rfloor$. We let $\varepsilon$ be $\max\{-\lambda_{min}(B), 0\} + 0.05$ to ensure the positive definiteness of the covariance matrix. The random block matrix allows for sparsity to be embedded within the covariance structure and mimics the standard structure in Cao et al. (2019); Li et al. (2023b,a).

We aim to design the mean vectors $(\nu^{(1)}, \nu^{(2)})$ in such a way that the signals are comparable across all configurations of sample size and dimensionality. To accomplish this, we set $\nu^{(1)} = 0_p$ and choose the non-zero entries of $\nu^{(2)}$ to be equal and generated to satisfy: $||\nu^{(1)} - \nu^{(2)}||^2/\sqrt{\text{tr}(\Omega^2)} = 0.1$. We evaluate sparsity levels across $(0.01p, 0.05p, 0.2p, \text{ and } 0.5p)$. To select the non-zero entries of $\nu^{(2)}$, we randomly select entries of $\nu^{(2)}$ depending on the sparsity level of interest. As $\nu^{(2)}$ becomes denser, the individual signal strength of each element decreases. To evaluate the testing size, we include a case where $\nu^{(1)} = \nu^{(2)} = 0_p$ for the null setting.

We set $n_1 = n_2 = n$. We generate the samples under three $(n, p)$-configurations: $(n, p) \in (100, 200)$, $(100, 500)$, $(100, 1000)$. We evaluate each test at the significance levels of $\alpha = 0.05$ and $\alpha = 0.01$. The empirical percentages of rejections are reported in Tables 1–2, respectively.

Table 1: The empirical percentages of rejection at the significance level of $\alpha = 0.05$.

| $\Omega$ | $(n, p)$ | Sparsity | Gaussian | | | | | Gamma | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | M | Q | F | C | BF | M | Q | F | C | BF |
| $\Omega^{AR}$ | (100, 200) | 0.00 | 0.058 | 0.061 | 0.078 | 0.068 | 0.005 | 0.059 | 0.059 | 0.096 | 0.077 | 0.002 |
| | | 0.01 | 1.000 | 0.944 | 1.000 | 1.000 | 0.912 | 1.000 | 0.938 | 1.000 | 1.000 | 0.911 |
| | | 0.05 | 0.916 | 0.920 | 0.980 | 0.963 | 0.761 | 0.935 | 0.954 | 0.993 | 0.982 | 0.796 |
| | | 0.20 | 0.296 | 0.850 | 0.823 | 0.799 | 0.175 | 0.327 | 0.856 | 0.845 | 0.816 | 0.185 |
| | | 0.50 | 0.133 | 0.502 | 0.488 | 0.440 | 0.050 | 0.153 | 0.562 | 0.542 | 0.506 | 0.060 |
| | (100, 500) | 0.00 | 0.075 | 0.062 | 0.096 | 0.080 | 0.004 | 0.056 | 0.049 | 0.067 | 0.062 | 0.002 |
| | | 0.01 | 1.000 | 0.950 | 1.000 | 1.000 | 0.911 | 1.000 | 0.944 | 1.000 | 1.000 | 0.907 |
| | | 0.05 | 0.795 | 0.937 | 0.976 | 0.957 | 0.610 | 0.780 | 0.940 | 0.978 | 0.956 | 0.629 |
| | | 0.20 | 0.241 | 0.860 | 0.835 | 0.823 | 0.133 | 0.233 | 0.859 | 0.835 | 0.809 | 0.122 |
| | | 0.50 | 0.147 | 0.516 | 0.484 | 0.454 | 0.066 | 0.143 | 0.517 | 0.495 | 0.461 | 0.045 |
| | (100, 1000) | 0.00 | 0.067 | 0.049 | 0.073 | 0.064 | 0.001 | 0.063 | 0.054 | 0.072 | 0.057 | 0.001 |
| | | 0.01 | 1.000 | 0.948 | 1.000 | 1.000 | 0.912 | 1.000 | 0.959 | 1.000 | 1.000 | 0.930 |
| | | 0.05 | 0.635 | 0.943 | 0.973 | 0.947 | 0.448 | 0.619 | 0.937 | 0.954 | 0.939 | 0.439 |
| | | 0.20 | 0.167 | 0.858 | 0.812 | 0.801 | 0.086 | 0.183 | 0.865 | 0.840 | 0.820 | 0.092 |
| | | 0.50 | 0.124 | 0.545 | 0.515 | 0.488 | 0.040 | 0.113 | 0.537 | 0.505 | 0.464 | 0.036 |
| $\Omega^{BS}$ | (100, 200) | 0.00 | 0.060 | 0.055 | 0.066 | 0.066 | 0.001 | 0.078 | 0.060 | 0.086 | 0.076 | 0.003 |
| | | 0.01 | 0.999 | 0.960 | 0.999 | 0.999 | 0.919 | 0.999 | 0.950 | 0.998 | 0.999 | 0.907 |
| | | 0.05 | 1.000 | 0.944 | 1.000 | 1.000 | 0.894 | 1.000 | 0.946 | 1.000 | 1.000 | 0.904 |
| | | 0.20 | 0.808 | 0.860 | 0.963 | 0.931 | 0.539 | 0.826 | 0.853 | 0.969 | 0.925 | 0.560 |
| | | 0.50 | 0.345 | 0.495 | 0.636 | 0.544 | 0.097 | 0.290 | 0.507 | 0.614 | 0.522 | 0.081 |
| | (100, 500) | 0.00 | 0.070 | 0.061 | 0.079 | 0.074 | 0.004 | 0.054 | 0.061 | 0.073 | 0.066 | 0.002 |
| | | 0.01 | 1.000 | 0.958 | 1.000 | 1.000 | 0.920 | 1.000 | 0.963 | 1.000 | 1.000 | 0.937 |
| | | 0.05 | 1.000 | 0.947 | 1.000 | 1.000 | 0.895 | 1.000 | 0.951 | 1.000 | 1.000 | 0.900 |
| | | 0.20 | 0.670 | 0.870 | 0.956 | 0.919 | 0.415 | 0.713 | 0.856 | 0.956 | 0.916 | 0.443 |
| | | 0.50 | 0.294 | 0.529 | 0.628 | 0.560 | 0.079 | 0.271 | 0.514 | 0.614 | 0.535 | 0.066 |
| | (100, 1000) | 0.00 | 0.089 | 0.043 | 0.072 | 0.071 | 0.004 | 0.067 | 0.066 | 0.080 | 0.075 | 0.003 |
| | | 0.01 | 1.000 | 0.972 | 1.000 | 1.000 | 0.924 | 1.000 | 0.958 | 1.000 | 1.000 | 0.906 |
| | | 0.05 | 1.000 | 0.955 | 1.000 | 1.000 | 0.914 | 1.000 | 0.946 | 1.000 | 1.000 | 0.909 |
| | | 0.20 | 0.582 | 0.870 | 0.944 | 0.909 | 0.343 | 0.580 | 0.868 | 0.953 | 0.908 | 0.320 |
| | | 0.50 | 0.217 | 0.539 | 0.602 | 0.532 | 0.069 | 0.217 | 0.533 | 0.620 | 0.536 | 0.043 |

Across all covariance and data-generating setups, our proposed method is robust to the true underlying sparsity of the signal. As noted previously, the Maximum test has increased power when the signal is extremely sparse while the Quadratic test has improved power in the dense signal setting. This can be seen as the Maximum test tends to have improved power to the Quadratic test when the sparsity level is at the 1% and 5% level. However, at the 20% and 50% the Maximum test drastically loses power as expected. The inverse holds for the Quadratic test which has a substantial power increase at the denser signal settings but sacrifices a modicum of power in the sparse setting. However, the power loss of the quadratic test in the sparse setting is much smaller than the power loss of the maximum-type test in the dense settings.

Table 2: The empirical percentages of rejection at the significance level of $\alpha = 0.01$.

| $\Omega$ | $(n, p)$ | Sparsity | Gaussian | | | | | Gamma | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | M | Q | F | C | BF | M | Q | F | C | BF |
| $\Omega^{AR}$ | (100, 200) | 0.00 | 0.010 | 0.021 | 0.032 | 0.021 | 0.000 | 0.013 | 0.019 | 0.034 | 0.024 | 0.000 |
| | | 0.01 | 1.000 | 0.849 | 1.000 | 1.000 | 0.784 | 1.000 | 0.833 | 1.000 | 1.000 | 0.761 |
| | | 0.05 | 0.713 | 0.780 | 0.932 | 0.861 | 0.503 | 0.716 | 0.828 | 0.954 | 0.871 | 0.519 |
| | | 0.20 | 0.103 | 0.642 | 0.642 | 0.576 | 0.058 | 0.099 | 0.646 | 0.667 | 0.585 | 0.059 |
| | | 0.50 | 0.025 | 0.279 | 0.280 | 0.232 | 0.006 | 0.043 | 0.316 | 0.322 | 0.263 | 0.016 |
| | (100, 500) | 0.00 | 0.010 | 0.019 | 0.027 | 0.013 | 0.001 | 0.016 | 0.010 | 0.026 | 0.011 | 0.001 |
| | | 0.01 | 1.000 | 0.854 | 1.000 | 1.000 | 0.792 | 1.000 | 0.835 | 1.000 | 1.000 | 0.779 |
| | | 0.05 | 0.465 | 0.814 | 0.912 | 0.806 | 0.317 | 0.509 | 0.813 | 0.911 | 0.822 | 0.343 |
| | | 0.20 | 0.070 | 0.666 | 0.655 | 0.609 | 0.026 | 0.065 | 0.650 | 0.648 | 0.580 | 0.034 |
| | | 0.50 | 0.035 | 0.287 | 0.294 | 0.245 | 0.013 | 0.034 | 0.285 | 0.288 | 0.234 | 0.005 |
| | (100, 1000) | 0.00 | 0.016 | 0.013 | 0.021 | 0.009 | 0.000 | 0.014 | 0.007 | 0.016 | 0.012 | 0.000 |
| | | 0.01 | 1.000 | 0.850 | 1.000 | 1.000 | 0.797 | 1.000 | 0.872 | 1.000 | 1.000 | 0.815 |
| | | 0.05 | 0.318 | 0.819 | 0.895 | 0.815 | 0.186 | 0.309 | 0.823 | 0.881 | 0.805 | 0.196 |
| | | 0.20 | 0.045 | 0.665 | 0.618 | 0.582 | 0.023 | 0.057 | 0.682 | 0.630 | 0.600 | 0.029 |
| | | 0.50 | 0.033 | 0.292 | 0.278 | 0.242 | 0.009 | 0.030 | 0.282 | 0.277 | 0.231 | 0.006 |
| $\Omega^{BS}$ | (100, 200) | 0.00 | 0.008 | 0.017 | 0.012 | 0.014 | 0.001 | 0.015 | 0.016 | 0.020 | 0.017 | 0.000 |
| | | 0.01 | 0.995 | 0.848 | 0.998 | 0.996 | 0.776 | 0.995 | 0.841 | 0.997 | 0.996 | 0.776 |
| | | 0.05 | 1.000 | 0.824 | 1.000 | 1.000 | 0.739 | 1.000 | 0.816 | 1.000 | 1.000 | 0.727 |
| | | 0.20 | 0.468 | 0.641 | 0.869 | 0.722 | 0.227 | 0.490 | 0.659 | 0.879 | 0.730 | 0.235 |
| | | 0.50 | 0.106 | 0.269 | 0.376 | 0.272 | 0.020 | 0.087 | 0.289 | 0.365 | 0.263 | 0.016 |
| | (100, 500) | 0.00 | 0.019 | 0.019 | 0.031 | 0.025 | 0.001 | 0.009 | 0.014 | 0.018 | 0.014 | 0.000 |
| | | 0.01 | 1.000 | 0.849 | 1.000 | 1.000 | 0.779 | 1.000 | 0.866 | 1.000 | 1.000 | 0.805 |
| | | 0.05 | 1.000 | 0.812 | 0.999 | 0.999 | 0.731 | 1.000 | 0.807 | 1.000 | 1.000 | 0.734 |
| | | 0.20 | 0.336 | 0.674 | 0.840 | 0.705 | 0.149 | 0.324 | 0.642 | 0.827 | 0.675 | 0.133 |
| | | 0.50 | 0.087 | 0.280 | 0.367 | 0.256 | 0.013 | 0.072 | 0.268 | 0.340 | 0.245 | 0.014 |
| | (100, 1000) | 0.00 | 0.023 | 0.010 | 0.023 | 0.018 | 0.001 | 0.016 | 0.018 | 0.026 | 0.021 | 0.000 |
| | | 0.01 | 1.000 | 0.852 | 1.000 | 1.000 | 0.784 | 1.000 | 0.834 | 1.000 | 1.000 | 0.774 |
| | | 0.05 | 0.997 | 0.822 | 1.000 | 0.999 | 0.745 | 0.999 | 0.842 | 0.999 | 0.996 | 0.751 |
| | | 0.20 | 0.221 | 0.658 | 0.799 | 0.651 | 0.081 | 0.219 | 0.668 | 0.806 | 0.652 | 0.081 |
| | | 0.50 | 0.059 | 0.294 | 0.351 | 0.255 | 0.014 | 0.060 | 0.270 | 0.331 | 0.230 | 0.007 |

Combination tests afford means of protection against improperly assumed sparsity. In the most sparse setting, where the Maximum test is the most effective, we note that Fisher and Cauchy combination tests have improved power over the Quadratic test and are as powerful as the maximum-type test. While the Maximum test drastically reduces in power as the signal becomes denser, in these cases, the combination tests do not suffer due to the improved performance of the Quadratic tests. Thus, in the densest setting, we observe that the power of these methods is extremely comparable to the Maximum test and even has a higher power to the Quadratic test in sparse covariance settings. The Cauchy combination test appears to be slightly more conservative than the Fisher combination test; though, also boasts a somewhat tighter control over the type I error. The power differential between the

two methods is small and the Cauchy combination method is still greatly more powerful than the maximum-type test in a dense signal setting. As expected, the Bonferroni method has the lowest power as it is the most conservative with the most stringent control over the type I error.

# 5 Real Data Analysis

We demonstrate the effectiveness of the combination testing framework through application to a gut microbiome study carried out by Morgan et al. (2015). This study investigates changes in host gene response linked to the development of inflammatory bowel disease based on alterations in an individual's gut microbiome composition. In this study, 16S rRNA is sequenced and the resulting reads are clustered at a 97% similarity level to form the resulting operational taxonomic units (OTUs), which are proxies for the underlying microbial taxa. These OTU counts are inferred by sequencing 16S RNA and clustering the resulting reads to serve as proxies for the underlying taxa. This procedure identified over 7,000 unique bacterial taxa, though many were at extremely low abundance levels. As if common in microbiome analysis, these OTUs were then aggregated at the genera level to form 303 genera counts. We focus on taxa of interest that had at least 10 total counts. Due to the heavy zero inflation common in microbiome data, many potential taxa of interest have zero observed counts, which need to be first omitted. This preliminary cleaning step yields $p = 226$ genera of interest. From this, we are able to construct a series of binary tests of means from the following set of associated metadata.

Below, we summarize the three metadata values of interest:

1. Antibiotic (A): Did the individuals take antibiotics (YES/NO)

2. Location (L): Location of Pouch vs Pre-pouch ileum (Pouch/PPI)

3. Class (C): Classification of familial adenomatous polyposis(FAP) vs non-FAP (FAP/nFAP)

Given these metadata values, we construct twelve different two-way interaction tests. Note that we omit the tests for main effects as the signals for these main effects were relatively strong, and therefore, all tests indicate a relevant signal. Instead, it is more interesting

to compare situations where the Max test and Quadratic test disagree. The results are summarized in Table 3.

Table 3: Comparisons of testing two-way interactions of metadata using various methods.

| Two-way Interaction Tests | | | | | | | |
|---|---|---|---|---|---|---|---|
| Test Name | $n_1$ | $n_2$ | M | Q | BF | F | C |
| Fixing C at: nFAP, Testing: L | 122 | 28 | 0 | 0 | 0 | 0 | 0 |
| Fixing C at: nFAP, Testing: A | 96 | 54 | 1 | 1 | 1 | 1 | 1 |
| Fixing C at: FAP, Testing: L | 74 | 31 | 0 | 0 | 0 | 0 | 0 |
| Fixing C at: FAP, Testing: A | 93 | 12 | 1 | 1 | 1 | 1 | 1 |
| Fixing L at: PPI, Testing: C | 122 | 74 | 1 | 1 | 1 | 1 | 1 |
| Fixing L at: PPI, Testing: A | 144 | 52 | 1 | 1 | 1 | 1 | 1 |
| Fixing L at: Pouch, Testing: C | 28 | 31 | 1 | 1 | 1 | 1 | 1 |
| Fixing L at: Pouch, Testing: A | 45 | 14 | 1 | 1 | 1 | 1 | 1 |
| Fixing A at: NO, Testing: C | 96 | 93 | 1 | 1 | 1 | 1 | 1 |
| **Fixing A at: NO, Testing: L** | 144 | 45 | 1 | 0 | 0 | 1 | 1 |
| **Fixing A at: YES, Testing: C** | 54 | 12 | 0 | 1 | 0 | 1 | 1 |
| Fixing A at: YES, Testing: L | 52 | 14 | 0 | 0 | 0 | 0 | 0 |

Note: The tests are constructed by fixing one metadata variable at a specific level, testing a second, while averaging over the third metadata value. A value of 0 indicates that the test fails to reject the null, while a value of 1 indicates the test rejects the null.

Of note is that in all cases where either the Max or the Quadratic test detects a signal, both Cauchy and Fisher tests also detect a signal. There are two settings where they disagree.

The first is when the antibiotic status (A) is fixed to NO antibiotic regimen and the location (L) is tested. The Maximum test identifies a signal while the Quadratic test does not. This suggests that the underlying signal may be sparse as the Maximum test tends to have higher power than the Quadratic test in a sparse setting. As noted by Morgan et al. (2015), only a few microbial clades are differentially expressed when comparing the Pouch location to PPI. While the authors note that the transcriptome may be greatly shifted between locations, the difference between the microbiome in each location is small; thus, the Maximum test is well capable of capturing this sparse signal. In this setting, both Fisher and Cauchy's methods agree with the Maximum test in capturing this sparse underlying signal.

The second is when the antibiotic status (A) is fixed to YES and the class status (C) is tested. The Maximum test is significant while the Quadratic test is not. Morgan et al. (2015) noted that after accounting for antibiotic use, FAP and non-FAP classification is driven by

the Escherichia, Actinobacteria, and Sutterella abundance. However, the paper also notes that generally higher levels of the Bacteroidetes phylum are linked to differences between the classes. The Bacteroidetes phylum constitutes a large portion of the gut microbiome (Wexler, 2007), and Morgan et al. (2015) notes that broad changes in this phylum are linked to developing the FAP class instead of the non-FAP class. Further, while antibiotics are known to decrease microbial diversity (Dudek-Wicher et al., 2018), members of the Bacteroidetes phylum are known to commonly carry some measure of antibiotic resistance. Therefore, it is likely that many members of the Bacteroidetes phylum will remain active, continuing to drive the distinction between FAP and non-FAP individuals. This larger range of microbiota driving the FAP vs non-FAP classification suggests a dense underlying signal, thus showing the higher power of the Quadratic test. By using the Fisher or Cauchy combination test, we have the increased discovery power of the Quadratic test in capturing the dense underlying signal. Improperly assuming the signal is sparse and applying the Maximum test would miss this outcome.

# 6 Conclusion

We proposed a power-enhanced two-sample mean test for high-dimensional compositional data. Fisher's method (Fisher, 1925) and the Cauchy method (Liu and Xie, 2020) provide a useful tool to drop restrictive signal assumptions. In doing so, we can leverage the power of maximum-type tests under sparse signals and quadratic-type tests for dense signals to improve the power in the compositional data analysis. Through novel theoretical derivation, we have shown that the maximum-type equivalence test introduced by Cao et al. (2018) and the quadratic-type hypothesis test introduced by Chen and Qin (2010) are asymptotically independent, where the dimension can be on the nearly exponential order of the sample size. We further develop a new Gaussian assumption result to prove the correct asymptotic size and consistent asymptotic power of the proposed power-enhanced tests without requiring the pseudo-independence structure. We validate the numerical properties through simulation studies and demonstrate its real application to study the effects of microbiome dysbiosis on host gene expression.

# References


Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 44(2):139–160.

Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statistica Sinica*, 6:311–329.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics*, 36(6):2577–2604.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684.

Cai, T., Liu, W., and Xia, Y. (2014). Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(2):349–372.

Cao, Y., Lin, W., and Li, H. (2018). Two-sample tests of high-dimensional means for compositional data. *Biometrika*, 105(1):115–132.

Cao, Y., Lin, W., and Li, H. (2019). Large covariance estimation for compositional data via composition-adjusted thresholding. *Journal of the American Statistical Association*, 114(526):759–772.

Chen, S. X., Guo, B., and Qiu, Y. (2023a). Testing and signal identification for two-sample high-dimensional covariances via multi-level thresholding. *Journal of Econometrics*, 235(2):1337–1354.

Chen, S. X., Li, J., and Zhong, P.-S. (2019). Two-sample and anova tests for high dimensional means. *The Annals of Statistics*, 47(3):1443–1474.

Chen, S. X. and Qin, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, 38(2):808–835.

Chen, S. X., Qiu, Y., and Zhang, S. (2023b). Sharp optimality for high-dimensional covariance testing under sparse signals. *The Annals of Statistics*, 51(5):1921–1945.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability*, 45(4):2309–2352.

Cuesta-Albertos, J. A., Cuevas, A., and Fraiman, R. (2009). On projection-based tests for directional and compositional data. *Statistics and Computing*, 19:367–380.

Dudek-Wicher, R., Junka, A., and Bartoszewicz, M. (2018). The influence of antibiotics and dietary components on gut microbiota. *Gastroenterology Review*, 13(2):85–92.

Fan, J., Liao, Y., and Yao, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica*, 83(4):1497–1541.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*, volume 1. Edinburgh by Oliver and Boyd.

Giessing, A. and Fan, J. (2020). Bootstrapping $\ell_p$-statistics in high dimensions. *arXiv preprint arXiv:2006.13099*.

Giessing, A. and Fan, J. (2023). A bootstrap hypothesis test for high-dimensional mean vectors. *arXiv preprint arXiv:2309.01254*.

He, Y., Xu, G., Wu, C., and Pan, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *The Annals of Statistics*, 49(1):154–181.

Li, D., Srinivasan, A., Chen, Q., and Xue, L. (2023a). Robust covariance matrix estimation for high-dimensional compositional data with application to sales data analysis. *Journal of Business & Economic Statistics*, 41(4):1090–1100.

Li, D., Srinivasan, A., Xue, L., and Zhan, X. (2023b). Robust shape matrix estimation for high-dimensional compositional data with application to microbial inter-taxa analysis. *Statistica Sinica*, 33:1577–1602.

Li, D. and Xue, L. (2015). Joint limiting laws for high-dimensional independence tests. *arXiv preprint arXiv:1405.7244*.

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

Liu, W., Yu, X., and Li, R. (2022). Multiple-splitting projection test for high-dimensional mean vectors. *Journal of Machine Learning Research*, 23(71):1–27.

Liu, W., Yu, X., Zhong, W., and Li, R. (2024). Projection test for mean vector in high dimensions. *Journal of the American Statistical Association*, 119(545):744–756.

Liu, Y. and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529):393–402.

Morgan, X. C., Kabakchiev, B., Waldron, L., Tyler, A. D., Tickle, T. L., Milgrom, R., Stempak, J. M., Gevers, D., Xavier, R. J., Silverberg, M. S., and Huttenhower, C. (2015). Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biology*, 16(1):67.

Pillai, N. S. and Meng, X.-L. (2016). An unexpected encounter with cauchy and lévy. *The Annals of Statistics*, 44(5):2089–2097.

Srivastava, D. K., Boyett, J. M., Jackson, C. W., Tong, X., and Rai, S. N. (2007). A comparison of permutation hotelling's t 2 test and log-ratio test for analyzing compositional data. *Communications in Statistics—Theory and Methods*, 36(2):415–431.

Tsagris, M., Preston, S., and Wood, A. T. (2017). Nonparametric hypothesis testing for equality of means on the simplex. *Journal of Statistical Computation and Simulation*, 87(2):406–422.

Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *Journal of the American Statistical Association*, 110(512):1658–1669.

Wexler, H. M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. *Clinical Microbiology Reviews*, 20(4):593–621.

Xu, G., Lin, L., Wei, P., and Pan, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika*, 103(3):609–624.

Xu, M., Zhang, D., and Wu, W. B. (2014). $l^2$ asymptotics for high-dimensional data. *arXiv preprint arXiv:1405.7244*.

Yu, X., Li, D., and Xue, L. (2024a). Fisher's combined probability test for high-dimensional covariance matrices. *Journal of the American Statistical Association*, 119(545):511–524.

Yu, X., Li, D., Xue, L., and Li, R. (2023). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. *Journal of the American Statistical Association*, 118(544):2548–2561.

Yu, X., Yao, J., and Xue, L. (2024b). Power enhancement for testing multi-factor asset pricing models via fisher's method. *Journal of Econometrics*, 239(2):105458.