Statistica Sinica Preprint No: SS-2020-0236

Title	A new nonparametric extension of ANOVA via projection mean variance measure
Manuscript ID	SS-2020-0236
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202020.0236
Complete List of Authors	Jicai Liu, Yuefeng Si, Wenchao Xu and Riquan Zhang
Corresponding Author	Riquan Zhang
E-mail	zhangriquan@163.com

Notice: Accepted version subject to English editing.

Statistica Sinica

A NEW NONPARAMETRIC EXTENSION OF ANOVA VIA A PROJECTION MEAN VARIANCE MEASURE

Jicai Liu[1,2],Yuefeng Si[3], Wenchao Xu[1] and Riquan Zhang[4]

1. Shanghai Lixin University of Accounting and Finance

2. KLATASDS-MOE, East China Normal University

3. University of Hong Kong

4. East China Normal University

Abstract: In this paper, we introduce a novel projection mean variance (PMV) measure to construct a nonparametric test for the multisample hypothesis of equal distributions for univariate or multivariate responses. The proposed PMV measure generalizes the mean variance index via the projection technique. We obtain the theoretical properties of the PMV measure and its empirical counterpart. The PMV measure can yield an analogous variance component decomposition. Through this decomposition, an ANOVA F statistic is derived to test the multisample problem. The proposed test is statistically consistent against the general alternatives and robust to heavy-tailed data. The test is free of tuning parameters and does not require moment conditions on the response. The simulation results demonstrate that the PMV test has higher power than the classical Wilks-type methods and DISCO test, especially when the dimension of the response is relatively large or the moment conditions required by the DISCO test are violated. We further illustrate our method by empirical analyses of two real datasets.

Key words and phrases: Projection, multivariate multisample problem, nonparametric tests, independence test, nonparametric ANOVA extension.

1. Introduction

The multisample problem, i.e., testing whether the underlying distributions of two or more populations are the same, is a classical topic in statistics and arises in many modern scientific applications. For example, in genomics research, we wish to explore whether gene expression levels differ among distinct predefined patient groups to identify disease-associated gene expression. In data integration for bioinformatics, it is of interest to know whether datasets from different labs are distributed identically to synthesize information across labs (Borgwardt et al., 2006).

Let $F_k(\mathbf{z})$ be the distribution function of p-variate continuous random variable $\mathbf{Z}_k$, for $k = 1, \cdots, K$. The multisample problem is concerned with testing the null hypothesis

$$H_0 : F_1(\mathbf{z}) = \cdots = F_K(\mathbf{z}) \equiv F(\mathbf{z}), \qquad \text{for all } \mathbf{z} \in \mathcal{R}^p, \tag{1.1}$$

against the alternative hypothesis $H_1 : F_k(\mathbf{z}) \neq F_j(\mathbf{z})$ for some $k \neq j \in$

$\{1, \cdots, K\}$. When the distributions $F_k(\mathbf{z})$ are normal with constant variance, two widely used methods for testing the problem (1.1) are the analysis of variance (ANOVA) for univariate data and the multivariate analysis of variance (MANOVA) for multivariate data. These methods can effectively detect the location difference among K independent samples. However, the normality and common variance assumptions are usually violated in most applications. Thus, much effort has been devoted to exploring nonparametric test approaches without specific distribution assumptions. For example, Kruskal and Wallis (1952) proposed a rank-based test procedure, Kiefer (1959) introduced the K-sample Kolmogorov-Smirnov and Cramér-von Mises tests, and Scholz and Stephens (1987) extended the Anderson-Darling test to the K-sample setting.

In general, the above nonparametric test methods are limited to dealing with univariate data and are not easily extendable to multivariate settings. In this paper, we propose a novel nonparametric test for the multivariate multisample problem. The proposed method is based on the fact that the K-sample problem (1.1) is equivalent to an independence test between a continuous random vector and a categorical variable. Specifically, we introduce a latent categorical variable Y with K categories, denoted by $\{y_1, \cdots, y_K\}$. Then, a new random vector $(\mathbf{X}, Y)$ can be defined by $\mathbf{X} = \mathbf{Z}_k$

if $Y = y_k$. In this way, it is easy to see that the original variables $\mathbf{Z}_k$, $k = 1, \cdots, K$ are one-to-one transformed to the new variables $(\mathbf{X}, Y)$. Thus, the multisample problem has the following equivalent form:

$$H_0 : \text{pr}\{\mathbf{X} \leq \mathbf{x} | Y = y_k\} = F(\mathbf{x}), \quad \text{for all } \mathbf{x} \in \mathcal{R}^p \text{ and } k = 1, \cdots, K.$$

compared to the alternative hypothesis $H_1 : \text{pr}\{\mathbf{X} \leq \mathbf{x} | Y = y_k\} \neq \text{pr}\{\mathbf{X} \leq \mathbf{x} | Y = y_j\}$, for some $k \neq j \in \{1, \cdots, K\}$. This yields that (1.1) is equivalent to the following problem:

$$H_0 : \mathbf{X} \text{ and } Y \text{ are independent versus } H_1 : \mathbf{X} \text{ and } Y \text{ are dependent.} \quad (1.2)$$

In the following context, we shall mainly restrict our attention to inferring the independence test problem (1.2).

Recently, Cui et al. (2015) proposed a mean variance (MV) index for feature screening in high-dimensional discriminant analysis. The MV index can quantify the dependence between a continuous random variable and a categorical variable. This measure has also been applied to test the problem (1.1) for univariate data by Cui and Zhong (2019). In general, the above methods cannot effectively handle the multivariate multisample problem (1.1) or the independent test problem (1.2). The main reason is that the MV index is substantially rank-based and computationally expensive to implement when the dimension of $\mathbf{X}$ is moderate or high.

In this paper, we shall generalize the univariate MV index to an arbitrary dimension by a projection technique. The projection method is a useful tool for multivariate statistical inference, which can be found in Baringhaus and Franz (2004), Escanciano (2006), and Zhu et al. (2017), among others. The new measure has many nice properties. First, it is equal to zero if and only if $\mathbf{X}$ and Y are independent. Second, it has a closed-form expression and can be easily estimated from the data. Third, it does not require any moment condition and is easily applicable in arbitrary dimensions of $\mathbf{X}$. Finally, it is robust to heavy-tailed data and outliers.

The proposed measure can provide an analogous variance component decomposition. Thus, we can derive a nonparametric extension of the typical ANOVA and MANOVA. Based on this extension, an analog to the ANOVA F statistic is obtained to test hypothesis (1.1). A related research topic is the distance components (DISCO) test proposed by Rizzo and Székely (2010), who used all pairwise distances between-sample elements and obtained an analog to ANOVA decomposition of distances. An important difference between our method and the DISCO test is that the latter requires the moment condition $E[\|\mathbf{X}\|] < \infty$. Recently, Zhu et al. (2017) and Kim et al. (2020) have demonstrated that the distance-based statistics, distance covariance (DCOV, Székely et al. (2007, 2009)) and energy statistic

(Székely and Rizzo, 2013b) may suffer from low power when the moment condition is violated or when extreme observations exist. Thus, it is not difficult to imagine that the distance-based DISCO test may also inherit this shortcoming in certain settings. However, such data subject to heavy-tailed errors are often encountered in various areas of science, especially in the big data era. Examples include high-frequency financial data, fMRI data, and gene expression data. Thus, our aim is to develop new robust methods to tackle the multisample problem for heavy-tailed high-dimensional data.

The rest of the paper is organized as follows. In Section 2, we introduce the projection mean variance measure and its sample counterpart; the theoretical properties of the proposed estimators are established. In Section 3, we present some new interpretations of the MV index. In Section 4, we describe the PMV-based test. PMV decomposition for multifactor models follows in Section 5. The results from our numerical studies are reported in Sections 6 and 7. We provide some discussion in Section 8. All technical proofs are arranged in a supplementary file.

2. Projection mean variance measure

To facilitate the presentation, we first review the MV index. Let the latent group variable Y be a categorical variable with K classes $\{y_1, y_2, \cdots, y_K\}$.

When X is univariate, Cui et al. (2015) proposed the MV index for feature screening in high-dimensional discriminant analysis, given by

$$\mathrm{MV}(X|Y) := E_X[\mathrm{var}_Y(F(X|Y))], \tag{2.1}$$

where $F(x|Y) = \mathrm{pr}\{X \leq x|Y\}$. Cui et al. (2015) further showed that

$$\mathrm{MV}(X|Y) = \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} [F_k(x) - F(x)]^2 dF(x), \tag{2.2}$$

where $p_k = \mathrm{pr}\{Y = y_k\}$, $F_k(x) = \mathrm{pr}\{X \leq x|Y = y_k\}$ and $F(x) = \mathrm{pr}\{X \leq x\}$.

It follows from (2.2) that the MV index can be viewed as the weighted average of Cramér-von Mises distances between conditional and unconditional distribution functions. This indicates that $\mathrm{MV}(X|Y) = 0$ if and only if the distributions of the K populations are identical. Thus, $\mathrm{MV}(X|Y)$ is a natural measure to test the independent problem (1.2).

We next extend the univariate MV index to the setting where the dimensionality of X is arbitrary by the integration over all one-dimensional projections. Let $\mathbb{S}^{p-1} = \{\beta \in \mathcal{R}^p : \|\beta\| = 1\}$ be the unit hypersphere in $\mathcal{R}^p$ for any $p > 1$. Our approach relies on the following lemma:

Lemma 1. *Let $\boldsymbol{X}$ be a p-dimensional random vector and Y be a categorical variable. Then, we have that*

$$\mathbf{X} \perp\!\!\!\perp Y \Longleftrightarrow \beta^T \mathbf{X} \perp\!\!\!\perp Y, \ \textit{for any } \beta \in \mathbb{S}^{p-1}, \tag{2.3}$$

where " $\Longleftrightarrow$ *" stands for "equivalent to", and "* $\perp\!\!\!\perp$ *" indicates independence.*

This result in (2.3), together with (2.1), motivates us to propose the following projection mean variance:

Definition 1. Let $\mathbf{X}$ be a p-dimensional random vector and Y be a categorical random variable with K classes $\{y_1, y_2, \cdots, y_K\}$. The projection mean variance (PMV) index between Y and $\mathbf{X}$ is defined by

$$\mathrm{PMV}(\mathbf{X}|Y) := c_p^{-1} \int_{\mathbb{S}^{p-1}} E_{\beta^T \mathbf{X}}[\mathrm{var}_Y(F_{\beta^T \mathbf{X}}(\beta^T \mathbf{X}|Y))] d\beta, \tag{2.4}$$

where $F_{\beta^T \mathbf{X}}(u|Y) = \mathrm{pr}\{\beta^T \mathbf{X} \le u|Y\}$, $c_p = \pi^{p/2-1}/\Gamma(p/2)$, and $\Gamma(\cdot)$ is the gamma function.

By the definition in (2.4), we can see that $\mathrm{PMV}(\mathbf{X}|Y)$ is the integration of the MV index between the projected random variables $\beta^T \mathbf{X}$ and Y. Generally, it is difficult to compute such an integral over the p-dimensional unit sphere. Fortunately, $\mathrm{PMV}(\mathbf{X}|Y)$ has a closed-form expression owing to the following lemma:

Lemma 2. *(Escanciano, 2006) For any two nonzero vectors* $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{R}^p$*, we have that*

$$\int_{\mathbb{S}^{p-1}} I(\beta^T \mathbf{v}_1 \le 0) I(\beta^T \mathbf{v}_2 \le 0) d\beta = c_p\{\pi - \mathrm{ang}(\mathbf{v}_1, \mathbf{v}_2)\}, \tag{2.5}$$

where $\mathrm{ang}(\mathbf{v}_1, \mathbf{v}_2) := \arccos\left\{\frac{\mathbf{v}_1^T \mathbf{v}_2}{\|\mathbf{v}_1\| \|\mathbf{v}_2\|}\right\}$ *is the angle between* $\mathbf{v}_1$ *and* $\mathbf{v}_2$*.*

Let $F_{\beta^T\mathbf{X}}(u)=\text{pr}\{\beta^T\mathbf{X}\leq u\}$ and $F_{\beta^T\mathbf{X}}(u|Y=y_k)=\text{pr}\{\beta^T\mathbf{X}\leq u|Y=y_k\}$. By Lemma 2, we provide some useful properties for $\text{PMV}(\mathbf{X}|Y)$ as follows:

Theorem 1. *If* $p_k=\text{pr}\{Y=y_k\}>0$, *for* $k=1,\cdots,K$, *then we have that*

(i) $\text{PMV}(\mathbf{X}|Y)=c_p^{-1}\sum_{k=1}^K p_k\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}[F_{\beta^T\mathbf{X}}(u|Y=y_k)-F_{\beta^T\mathbf{X}}(u)]^2dF_{\beta^T\mathbf{X}}(u)d\beta;$

(ii) $\text{PMV}(\mathbf{X}|Y)=0$ *if and only if* $\mathbf{X}$ *and* Y *are statistically independent;*

(iii) $\text{PMV}(\mathbf{X}|Y)=E[\text{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)]-\text{PS}_\text{W}(\mathbf{X}|Y)$, *where* $(\mathbf{X}_1,Y_1)$, $(\mathbf{X}_2,Y_2)$ *and* $(\mathbf{X}_3,Y_3)$ *are i.i.d. copies of* $(\mathbf{X},Y)$ *and*

$$\text{PS}_\text{W}(\mathbf{X}|Y):=\sum_{k=1}^K p_k^{-1}E[I(Y_1=y_k,Y_2=y_k)\text{ang}(\mathbf{X}_1-\mathbf{X}_3,\mathbf{X}_2-\mathbf{X}_3)];$$

(iv) $\text{PMV}(\boldsymbol{a}+c\mathbf{A}\mathbf{X}|Y)=\text{PMV}(\mathbf{X}|Y)$, *where* $\mathbf{A}\in\mathcal{R}^{p\times p}$ *is any orthonormal matrix,* $\boldsymbol{a}\in\mathcal{R}^p$ *and* $c\in\mathcal{R}$.

We present some remarks on Theorem 1. Property (i) indicates that $\text{PMV}(\mathbf{X}|Y)$ can also be represented as a weighted average of the distances, such as the MV index in (2.2). Property (ii) implies that $\text{PMV}(\mathbf{X}|Y)$ is generally applicable as an index to measure the dependence between a continuous random vector and a categorical one. Property (iii) indicates that $\text{PMV}(\mathbf{X}|Y)$ has a closed form and is thus easily estimated from the data. Property (iv) suggests that PMV is invariant with respect to the group of orthogonal transformations.

Note that the integration over $\mathbb{S}^{p-1}$ in (2.4) implicitly requires $p > 1$. By the property (iii) of Theorem 1, we can extend the original definition of PMV in (2.4) to the one-dimensional setting. With slight abuse of notation, we still define the generalized PMV index by $\mathrm{PMV}(\mathbf{X}|Y)$, given by

$$\mathrm{PMV}(\mathbf{X}|Y) \;\; := \;\; E[\mathrm{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)] - \mathrm{PS}_W(\mathbf{X}|Y). \qquad (2.6)$$

When $p = 1$, the following result establishes the relationship between $\mathrm{MV}(X|Y)$ and $\mathrm{PMV}(X|Y)$:

Corollary 1. *Assume that X is univariate. If $p_k = \mathrm{pr}\{Y = y_k\} > 0$ for all $k = 1, \cdots, K$, then we have that* $\mathrm{PMV}(X|Y) = 2\pi\mathrm{MV}(X|Y)$.

Corollary 1 indicates that $\mathrm{PMV}(X|Y)$ is proportional to $\mathrm{MV}(X|Y)$ for one-dimensional random variable X. This property, together with Theorem 1, suggests that $\mathrm{PMV}(\mathbf{X}|Y)$ can measure independence for any $p \geq 1$.

We next develop the empirical estimate of $\mathrm{PMV}(\mathbf{X}|Y)$. Suppose that $\{(\mathbf{X}_i, Y_i),\ i = 1, \cdots, n\}$ is a random sample of $(\mathbf{X}, Y)$. To simplify the notations, we denote

$$\hat{p}_k := n^{-1}\sum_{i=1}^{n} I(Y_i = y_k), \quad g_U^n(u) := \widehat{\mathrm{pr}}\{\beta^T\mathbf{X} \leq u\} = n^{-1}\sum_{i=1}^{n} I(U_i \leq u),$$

$$g_{U,Y}^n(u; y_k) := \widehat{\mathrm{pr}}\{\beta^T\mathbf{X} \leq u|Y = y_k\} = \hat{p}_k^{-1}n^{-1}\sum_{i=1}^{n} I(U_i \leq u, Y_i = y_k),$$

where $U_i := \beta^T \mathbf{X}_i$. By property (i) in Theorem 1, we can give a straightforward plug-in estimator of $\mathrm{PMV}(\mathbf{X}|Y)$ as follows:

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) := \frac{1}{nc_p}\sum_{k=1}^{K}\hat{p}_k\sum_{i=1}^{n}\int_{\mathbb{S}^{p-1}}\left\{g^n_{U,Y}(\beta^T\mathbf{X}_i;y_k) - g^n_U(\beta^T\mathbf{X}_i)\right\}^2 d\beta.$$

Note that the above plug-in estimator is intractable. To put $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$ into practice, we shall present two equivalent forms in the following theorem: For $i,j,r = 1,2,\cdots,n$ and $k = 1,2,\cdots,K$, denote

$$\widetilde{A}_{jr;i} := a_{jri} - \frac{1}{n}\sum_{j=1}^{n}a_{jri} - \frac{1}{n}\sum_{r=1}^{n}a_{jri} + \frac{1}{n^2}\sum_{j,r=1}^{n}a_{jri},$$
$$\widetilde{B}_{ij;k} := b_{ij;k} - \frac{1}{n}\sum_{i=1}^{n}b_{ij;k} - \frac{1}{n}\sum_{j=1}^{n}b_{ij;k} + \frac{1}{n^2}\sum_{i,j=1}^{n}b_{ij;k},$$

where $a_{jri} := \mathrm{ang}(\mathbf{X}_j - \mathbf{X}_i, \mathbf{X}_r - \mathbf{X}_i)$, $b_{ik} := I(Y_i = y_k)$, $b_{ij;k} := b_{ik}b_{jk}$. Here, define $\arccos\{\frac{0}{0}\} = 0$. Then, we can obtain the following results:

Theorem 2. *For a given random sample $\{(\mathbf{X}_i, Y_i), i = 1,\cdots,n\}$, then we have that*

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) = -\frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}\widetilde{A}_{jr;i}\widetilde{B}_{jr;k} \quad (2.7)$$
$$= \frac{1}{n^3}\sum_{i,j,r=1}^{n}a_{ijr} - \frac{1}{n^3}\sum_{k=1}^{K}\hat{p}_k^{-1}\sum_{i,j,r=1}^{n}b_{ik}b_{jk}a_{ijr}. \quad (2.8)$$

Using (2.7), it is easy to compute $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$ in practice. A further discussion on its implementation is given in Section 4.2. The result in (2.8) is useful for studying the theoretical property of $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$. In fact, each term on the right side of (2.8) can easily be expressed in U-statistics. Then, we can establish their tail probability inequalities by the theory of U-statistics (Serfling, 1980) and obtain the following result, the proof of which can be found in the Supplementary Materials:

Theorem 3. *Assume that there exist two positive constants c_1 and c_2, such that $c_1/K \leq \min_{1\leq k\leq K} p_k \leq \max_{1\leq k\leq K} p_k \leq c_2/K$ and $K = O(n^\kappa)$ for some $0 \leq \kappa < 1/6$. Then, for any $\alpha \in (0,1)$ and sufficiently large n, there exists a positive constant c_0, such that*

$$\mathrm{pr}\left\{|\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)| \leq c_0\sqrt{\frac{K^6}{n}\log(K/\alpha)}\right\} \geq 1-\alpha.$$

The condition $c_1/K \leq \min_{1\leq k\leq K} p_k \leq \max_{1\leq k\leq K} p_k \leq c_2/K$ is also used in Cui et al. (2015). When K is fixed, the condition is automatically satisfied and $|\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)| = O(n^{-1/2})$. Theorem 3 suggests that $\lim_{n\to\infty} \widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) = \mathrm{PMV}(\mathbf{X}|Y)$ if $K = O(n^\kappa)$ with $0 \leq \kappa < 1/6$. However, when X is univariate, we can obtain from Lemma A.4 in Cui et al. (2015) that $\lim_{n\to\infty} \widehat{\mathrm{MV}}_n(X|Y) = \mathrm{MV}(X|Y)$ if $K = o(n)$. Thus, the order κ in Theorem 3 may further be relaxed to $0 \leq \kappa < 1$. This is beyond the scope of this work but is an interesting topic for future research.

3. Extension of ANOVA via MV index

In this section, we illustrate that the MV index can provide a decomposition similar to the variance components in ANOVA. Then, in the next section, we generalize this decomposition to the PMV index to construct an analogous ANOVA F statistic for the testing problem (1.1).

Note that the definition in (2.1) is formally similar to the quantities $E[\mathrm{var}(X|Y)]$ and $\mathrm{var}(\mathrm{E}[X|Y])$, both of which appear in the basic variance decomposition formula

$$\mathrm{var}(X) = E[\mathrm{var}(X|Y)] + \mathrm{var}(\mathrm{E}[X|Y]). \tag{3.1}$$

After some algebra, we can obtain that

$$\mathrm{var}(E[X|Y]) = \sum_{k=1}^{K} p_k(\mathrm{E}[X|Y = y_k] - E[X])^2, \tag{3.2}$$

$$E[\mathrm{var}(X|Y)] = \sum_{k=1}^{K} p_k E[(X - E[X|Y = y_k])^2|Y = y_k]. \tag{3.3}$$

From the above two equations, we can see that $E[\mathrm{var}(X|Y)]$ and $\mathrm{var}(\mathrm{E}[X|Y])$ are able to describe the population between and within the group variation.

From (2.2) and (3.2), we can see that $\mathrm{MV}(X|Y)$ and $\mathrm{var}(\mathrm{E}[X|Y])$ have similar forms. This motivates us to obtain a similar variance decomposition for $\mathrm{MV}(X|Y)$. This result is provided in the following theorem:

Theorem 4. *If $p_k = \mathrm{pr}\{Y = y_k\} > 0$, for $k = 1, \cdots, K$, then we have that*

$$E[I(X_1 > X_3)I(X_2 \leq X_3)] = \mathrm{MV}(X|Y) + \mathrm{S_W}(X|Y), \tag{3.4}$$

where $\mathrm{S_W}(X|Y) := \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} E[\{I(X \leq x) - F_k(x)\}^2 | Y = y_k] dF(x)$.

We next provide some intuition to explain the connection between (3.4) and the population ANOVA decomposition in (3.1). First, $\mathrm{MV}(X|Y)$ and (3.2) have a similar form, which can describe differences among groups, and $\mathrm{S_W}(X|Y)$ and (3.3) also enjoy a common property, which can measure differences within each of the groups. Next, consider the following decomposition: $I(X \leq x) - F(x) = [F_k(x) - F(x)] + [I(X \leq x) - F_k(x)]$, for any $x \in \mathcal{R}$. Then, it is easy to obtain that

$$\mathrm{var}(I(X \leq x)) = \sum_{k=1}^{K} p_k [F_k(x) - F(x)]^2 + \sum_{k=1}^{K} p_k E[\{I(X \leq x) - F_k(x)\}^2 | Y = y_k].$$

Integration over $x \in [-\infty, \infty]$ and simple calculations yield that

$$\begin{aligned} \int_{-\infty}^{\infty} \mathrm{var}(I(X \leq x)) dF(x) &= E[I(X_1 > X_3)I(X_2 \leq X_3)] \\ &= \mathrm{MV}(X|Y) + \mathrm{S_W}(X|Y). \end{aligned} \tag{3.5}$$

Thus, (3.5) can be viewed as a direct nonparametric extension of (3.1) by replacing X and its total variation $\mathrm{var}(X)$ by the binary variables $I(X \leq x)$ and the cumulative total variation $\int_{-\infty}^{\infty} \mathrm{var}(I(X \leq x)) dF(x)$.

In summary, from (3.5) and Theorem 4, we can obtain a nonparametric extension of the typical ANOVA as follows:

Total variation: $\int_{-\infty}^{\infty} \mathrm{var}(I(X \le x))dF(x) = E[I(X_1 > X_3)I(X_2 \le X_3)]$;

Between-group variation: $\mathrm{MV}(X|Y) = \sum_{k=1}^{K} p_k \int_{-\infty}^{\infty}[F_k(x)-F(x)]^2 dF(x)$;

Within-group variation: $\mathrm{S_W}(X|Y)=\sum_{k=1}^{K} p_k \int_{-\infty}^{\infty} E[\{I(X \le x)-F_k(x)\}^2|Y = y_k]dF(x)$.

As mentioned above, this decomposition is similar to that in ANOVA, except that it does not rely on assumptions on the distribution of the population. Thus, it would be a useful tool and have many statistical applications.

4. The PMV tests of equal distributions

4.1 Method

We first show that the PMV index also has an interpretation similar to that in (3.5). By Lemma 2, it can be shown that

$$\begin{aligned} c_p^{-1}\int_{\mathbb{S}^{p-1}}\int_{-\infty}^{\infty}\mathrm{var}(I(\beta^T\mathbf{X} \le u))dF_{\beta^T\mathbf{X}}(u)d\beta &= E[\mathrm{ang}(\mathbf{X}_1-\mathbf{X}_3, \mathbf{X}_2-\mathbf{X}_3)], \\ c_p^{-1}\int_{\mathbb{S}^{p-1}} \mathrm{S_W}(\beta^T\mathbf{X}|Y)d\beta &= \mathrm{PS_W}(\mathbf{X}|Y). \end{aligned}$$

These, together with the definition of $\mathrm{S_W}(\beta^T\mathbf{X}|Y)$, indicate that $E[\mathrm{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)]$ and $\mathrm{PS_W}(\mathbf{X}|Y)$ can be viewed as the population total variability and within-group variation. Thus, (2.6) suggests that $E[\mathrm{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)]$ can be decomposed into two sources: within-group variation $\mathrm{PS}_{\mathrm{W}}(\mathbf{X}|Y)$ and between-group variation $\mathrm{PMV}(\mathbf{X}|Y)$. That is, (2.6) can naturally provide a nonparametric analysis of variance decomposition.

Note that (2.6) is a population decomposition, and its empirical counterpart can be obtained by (2.8). By the notations of the classical ANOVA, we rewrite (2.8) as

$$\mathrm{SS}_T = \mathrm{SS}_B + \mathrm{SS}_W, \tag{4.1}$$

where $\mathrm{SS}_T = \frac{1}{n^3}\sum_{i,j,r=1}^n a_{ijr}$, $\mathrm{SS}_W = \frac{1}{n^3}\sum_{k=1}^K \hat{p}_k^{-1}\sum_{i,j,r=1}^n b_{ik}b_{jk}a_{ijr}$ and $\mathrm{SS}_B = \widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$. Then, an analog to the ANOVA F statistic can be derived as follows:

$$F_n = \frac{\mathrm{SS}_B/(K-1)}{\mathrm{SS}_W/(n-K)} = \frac{\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)/(K-1)}{(\mathrm{SS}_T - \mathrm{SS}_B)/(n-K)}.$$

The larger value of F_n presents stronger evidence to support the alternative hypothesis. We name the new test as the PMV test of equal distributions. Generally, F_n does not have a F distribution, and the following result presents its asymptotic null distribution when K is fixed:

Theorem 5. *Under the null hypothesis H_0, we have that*

$$F_n = \frac{\mathrm{SS}_B/(K-1)}{\mathrm{SS}_W/(n-K)} \xrightarrow{d} \sum_{j=1}^{\infty}\lambda_j\eta_j^2, \quad n\to\infty,$$

where η_j are independent standard normal random variables and λ_j are nonnegative constants and depend on the distribution of $(\mathbf{X}, Y)$.

When $\mathbf{X}$ is univariate, Theorem 3.1 in Cui and Zhong (2019) suggests that the λ_j in Theorem 5 has a simple closed form. However, in general, the λ_j do not necessarily have such a good form by the definition in (S1.17) and the Hilbert-Schmit theory of integral equation (Kuo, 1975). This leads to the asymptotic null distribution of F_n being computationally infeasible. To implement the PMV test in practice, we approximate the asymptotic null distribution through a random permutation approach. The permutation method is referred to in Section 4.2.

Next, we can further study the asymptotic performance of $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y)$ under the alternative hypothesis.

Theorem 6. *Under the alternative hypothesis, we have that*

$$\sqrt{n}(\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) - \mathrm{PMV}(\mathbf{X}|Y)) \xrightarrow{d} N\left(0, \sigma^2\right),$$

where $\sigma^2 = \mathrm{var}\left[\Phi(\mathbf{X}_i, Y_i)\right]$, *in which* $\Phi(X, Y)$ *is given in* (S1.20).

From Theorem 6 and Sultsky's theorem, we can easily obtain that F_n converges weakly to a normal distribution. This result shows that the PMV test can detect all types of differences between distributions as follows:

Corollary 2. *The PMV test of hypothesis* (1.1) *is consistent against all alternatives.*

From the above theoretical results, we can see that the main difference between the PMV test and the DISCO test is that the PMV test does not require any moment condition. This advantage shall be further demonstrated by numerical simulations.

4.2 Implementation

In this section, we discuss the implementation of the PMV test in practice. For any given $i \in \{1, 2, \cdots, n\}$ and $k \in \{1, 2, \cdots, K\}$, let $\mathbf{A}_i = (a_{jri})_{n\times n}$ and $\mathbf{B}_k = (b_{jr;k})_{n\times n}$ be $n\times n$ matrices with entries a_{jri} and $b_{jr;k}$, respectively. From the definitions of $\widetilde{A}_{jr;i}$ and $\widetilde{B}_{jr;k}$ and (2.7), we obtain that

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y) = -\frac{1}{n^2}\mathrm{Tr}\Big([\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i]\mathbf{H}[\sum_{k=1}^{K}\hat{p}_k^{-1}\mathbf{B}_k]\mathbf{H}\Big),\ \mathrm{SS}_T = \frac{1}{n^2}\mathbf{1}_n^T\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i\Big)\mathbf{1}_n,$$

where $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$, $\mathbf{I}_n$ is the identity matrix and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. Here, we use the property $\mathbf{H}^2 = \mathbf{H}$. Thus, the PMV test statistic is easily implemented by computing matrices $\mathbf{A}_i$ and $\mathbf{B}_k$.

To put the proposed test into practice, we apply the permutation method to approach the asymptotic null distribution in Theorem 5. The permutation approach can yield a valid level α test for finite sample size. It has been shown to be effective; see the DCOV test, the DISCO test, and the projection correlation-based test (Zhu et al., 2017).

The permutation test procedure is as follows:

Step 1. Compute F_n and $\widehat{\mathrm{SS}}_T$ for the observed data $\{(\mathbf{X}_i, Y_i), i = 1, \cdots, n\}$;

Step 2. For each replicate, indexed $b \in \{1, \cdots, B\}$, generate a random permutation $\boldsymbol{\pi}_b = (\pi_{b,1}, \ldots, \pi_{b,n})$ of $\{1, \ldots, n\}$, and compute the estimator of $\mathrm{PMV}(\mathbf{X}|Y)$ using the permuted sample $(\mathbf{X}, Y_{\pi_b}) := \{(\mathbf{X}_i, Y_{\pi_{b,i}}), i = 1, \cdots, n\}$, denoted by $\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y_{\pi_b})$. Calculate the test statistic

$$F_n^{(b)} = \frac{\widehat{\mathrm{PMV}}_n(\mathbf{X}|Y_{\pi_b})/(K-1)}{(\mathrm{SS}_T - \widehat{\mathrm{PMV}}_n(\mathbf{X}|Y_{\pi_b}))/(n-K)};$$

Step 3. Compute the empirical p-value by

$$\hat{p} = \frac{1}{B+1}\Big\{1 + \sum_{b=1}^{B} \mathrm{I}(F_n^{(b)} \geq F_n)\Big\}.$$

5. The PMV decomposition in the general case

Following our approach to the one-way PMV decompositions in (2.8) and (4.1), we can generalize it to the general factorial design case by analogy. Here, we focus on the full factorial two-level design. Suppose that there are K_A levels of factor A and K_B levels of factor B and that R independent observations can be observed at each of the $K_A K_B$ combinations of levels.

Using the classical ANOVA formula notation from linear models, we specify the corresponding two-way additive model as $\mathbf{X} \sim A + B$, and the two-way design with interaction as $\mathbf{X} \sim A + B + A * B$, where $A * B$ is the interaction term between factor A and factor B. Let $A : B$ be the crossed factors A and B with $K_A K_B$ levels. For the above two-factor models, we have the following two-way PMV decompositions in the population:

Theorem 7. (i) *For model* $\mathbf{X} \sim A + B$, *we have that*

$$E[\text{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)] = \text{PMV}(\mathbf{X}|A) + \text{PMV}(\mathbf{X}|B) + \sigma^2_{\text{E},1}; \quad (5.1)$$

(ii) *For model* $\mathbf{X} \sim A + B + A * B$, *we have that*

$$E[\text{ang}(\mathbf{X}_1 - \mathbf{X}_3, \mathbf{X}_2 - \mathbf{X}_3)] = \text{PMV}(\mathbf{X}|A) + \text{PMV}(\mathbf{X}|B) + \text{PMV}(\mathbf{X}|A * B) + \sigma^2_{\text{E},2}; \quad (5.2)$$

(iii) $\text{PMV}(\mathbf{X}|A * B) = \text{PMV}(\mathbf{X}|A : B) - \text{PMV}(\mathbf{X}|A) - \text{PMV}(\mathbf{X}|B)$,

where $\sigma^2_{\text{E},1}$, $\sigma^2_{\text{E},2}$ *and* $\text{PMV}(\mathbf{X}|A * B)$ *are defined in* (S1.26), (S1.28) *and* (S1.29), *respectively.*

In a manner analogous to (4.1), we can obtain the empirical counterparts of (5.1) and (5.2), given by

$$\text{SS}_\text{T} = \widehat{\text{PMV}}_n(\mathbf{X}|A) + \widehat{\text{PMV}}_n(\mathbf{X}|B) + \text{SS}_{\text{E},1}, \quad (5.3)$$

for model $\mathbf{X} \sim A + B$; and

$$\text{SS}_\text{T} = \widehat{\text{PMV}}_n(\mathbf{X}|A) + \widehat{\text{PMV}}_n(\mathbf{X}|B) + \widehat{\text{PMV}}_n(\mathbf{X}|A * B) + \text{SS}_{\text{E},2}, \quad (5.4)$$

for model $\mathbf{X} \sim A + B + A * B$, where

$$\widehat{\mathrm{PMV}}_n(\mathbf{X}|A * B) = \widehat{\mathrm{PMV}}_n(\mathbf{X}|A : B) - \widehat{\mathrm{PMV}}_n(\mathbf{X}|A) - \widehat{\mathrm{PMV}}_n(\mathbf{X}|B),$$

$\mathrm{SS}_{\mathrm{E},1}$ and $\mathrm{SS}_{\mathrm{E},2}$ are the plug-in estimators of $\sigma^2_{\mathrm{E},1}$ and $\sigma^2_{\mathrm{E},2}$.

From (5.3) and (5.4), we can see that SS_{T} has similar two-way ANOVA decompositions. In Table 1, we summarize the PMV analysis for the two-way design with interaction. For factorial designs on three or more factors, we can obtain similar results.

Table 1. PMV analysis for the two-factor model with interaction.

Factor	df	Dispersion	F-ratio
A	$K_A - 1$	$\widehat{\mathrm{PMV}}_n(\mathbf{X}\|A)$	$\frac{\widehat{\mathrm{PMV}}_n(\mathbf{X}\|A)}{K_A-1} / \frac{\mathrm{SS}_{\mathrm{E},2}}{K_A K_B (R-1)}$
B	$K_B - 1$	$\widehat{\mathrm{PMV}}_n(\mathbf{X}\|B)$	$\frac{\widehat{\mathrm{PMV}}_n(\mathbf{X}\|B)}{K_B-1} / \frac{\mathrm{SS}_{\mathrm{E},2}}{K_A K_B (R-1)}$
A*B	$(K_A - 1)(K_B - 1)$	$\widehat{\mathrm{PMV}}_n(\mathbf{X}\|A * B)$	$\frac{\widehat{\mathrm{PMV}}_n(\mathbf{X}\|A*B)}{(K_A-1)(K_B-1)} / \frac{\mathrm{SS}_{\mathrm{E},2}}{K_A K_B (R-1)}$
Error	$K_A K_B (R - 1)$	$\mathrm{SS}_{\mathrm{E},2}$	
Total	$K_A K_B R - 1$	SS_{T}	

6. Monte Carlo simulations

In this section, several simulations are conducted to assess the finite sample performance of the proposed PMV test. We compare our results with the DISCO test, the Wilks' lambda test (Wilks) in Wilks (1932) and the rank transformed Wilks' lambda method (RankWilks) in Nath and Pavur (1985). All the numerical studies described in this paper have been implemented using R software. The relevant codes are available on the second author's GitHub page: `https://github.com/Oliver9803/PMV_code`.

Throughout our experiments, the p-value of the PMV or DISCO test is obtained by $B = 199$ permutations. We repeat each setting 1000 times and report the empirical power or type-I error rate of different tests. In each example, we consider a balanced design with four groups, where the common sample size is denoted by n.

Example 1. *Data are generated from distributions with identical independent marginals. The following two settings are studied:*

Case (i): *The data are generated from Example 3 in Rizzo and Székely (2010). Group 1 is noncentral* $t(4)$ *with noncentrality parameter* δ*. Groups 2-4 each have central* $t(4)$ *distributions.*

Case (ii): *This is identical to Case (i), except that group 1 is from the noncentral* $t(2)$*, and groups 2-4 are from the central* $t(2)$ *distribution.*

Table 2 reports the empirical type-I error rate of each test at significance levels $\alpha = 0.01$, 0.05 and 0.1 with $p = 10$ and $n = 30, 50$. From Table 2, it can be seen that each test achieves approximately the three nominal significance levels under the null hypothesis in Cases (i) and (ii).

Table 2. Example 1: Empirical type-I error rate with $p = 10$.

		$n = 30$			$n = 50$		
Setting	Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
Case (i)	PMV	0.015	0.050	0.097	0.012	0.049	0.093
	DISCO	0.007	0.050	0.103	0.013	0.045	0.086
	Wilks	0.012	0.049	0.096	0.010	0.049	0.095
	RankWilks	0.009	0.050	0.103	0.012	0.055	0.102
Case (ii)	PMV	0.014	0.054	0.107	0.007	0.044	0.095
	DISCO	0.012	0.052	0.102	0.006	0.039	0.097
	Wilks	0.002	0.042	0.097	0.010	0.043	0.095
	RankWilks	0.008	0.051	0.102	0.011	0.052	0.101

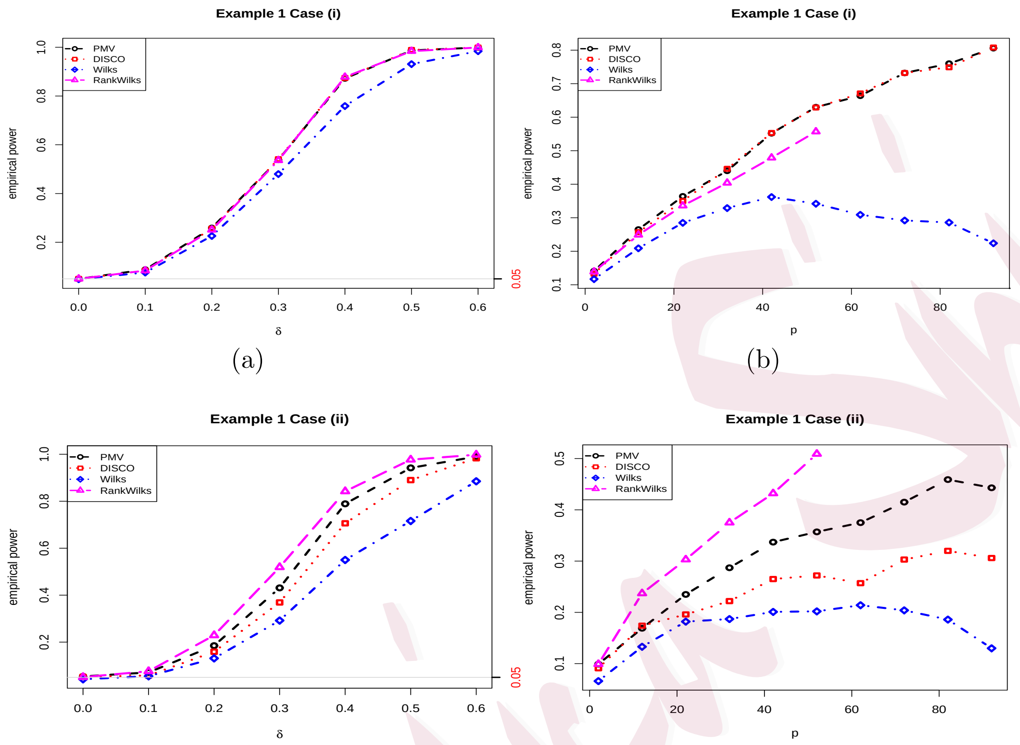
An empirical power comparison is displayed in Figure 1. Figures 1(a) and (c) show the plots of the power curve against the noncentrality parameter δ with dimensions fixed at $p = 10$. The results from Figure 1(a) suggest that the PMV, DISCO and RankWilks tests have similar performances and are slightly more powerful than the Wilks test in Case (i). Figure 1(c) indicates that the DISCO test is inferior to the PMV and RankWilks tests in Case (ii) where the data have heavy tails. This may be because the DISCO test is sensitive to the heavy-tailed data.

Figures 1(b) and (d) show the plots of the power curve against the dimension at the significance level $\alpha = 0.05$ and $\delta = 0.2$. Figure 1(b) illustrates that the PMV and DISCO tests perform comparably and are increasingly superior to Wilks and RankWilks as the dimension increases. For the dimension $p \geq 60$, the RankWilks test fails due to the dimension restriction, and thus, the power is missing in Figures 1(b) and (d). Thus, although the RankWilks test exhibits good power when p is small, it becomes practically infeasible for the large p. Again, Figure 1(d) suggests that the PMV test is still more powerful than the DISCO test in Case (ii). Thus, from Figure 1, we can see that the PMV test is robust to heavy-tailed data and can be applied in arbitrary dimensions, regardless of sample size.

Example 2. *Samples 2-4 have i.i.d. marginal Cauchy$(0,1)$ distributions. Sample 1 is the mixture distribution $0.5\,Cauchy(\delta,1)+0.5\,Cauchy(-\delta,1)$ with noncentrality parameter δ.*

The empirical type-I error rates for Example 2 are summarized in Table 3. The empirical type-I error rates of the PMV, DISCO and RankWilks tests are under reasonable control. It is also shown that the Wilks test fails to control the type-I error, mainly because the usual assumption of normality is not satisfied.

Figure 2(a) displays power curves with respect to δ. The results illustrate that the DISCO test has lower power than the PMV test. This may be because the condition $E[\|\mathbf{X}\|] < \infty$ required by DISCO is violated.



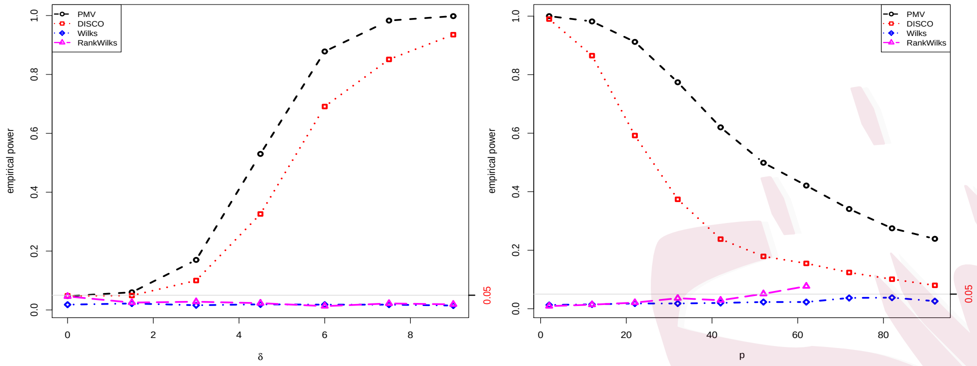
(a) (b)

(c) (d)

Figure 1. Example 1: Empirical power comparisons at the 0.05 significance level for $n = 30$: (a) δ varies with $p = 10$ for Case (i); (b) p varies and $\delta = 0.2$ for Case (i); (c) and (d): As in (a) and (b) but for Case (ii).

Table 3. Example 2: Empirical type-I error rate with $p = 10$.

	$n = 30$			$n = 50$		
Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
PMV	0.011	0.047	0.106	0.013	0.047	0.094
DISCO	0.014	0.048	0.095	0.010	0.045	0.093
Wilks	0.002	0.018	0.062	0.001	0.017	0.061
RankWilks	0.014	0.046	0.091	0.006	0.036	0.086



(a) (b)

Figure 2. Example 2: Empirical power comparisons at the 0.05 significance level for $n = 30$: (a) δ varies with $p = 10$; (b) p varies and $\delta = 8$.

The results suggest that our test is very robust in the setting. It might be surprising to see that the RankWilks test fails in the location model.

In Figure 2(b), the noncentrality parameter is fixed at $\delta = 8$, and the power varies with dimension. Figure 2(b) indicates that the PMV test has less power loss than the DISCO test as the dimension increases. In contrast to Figure 1(b), the power curve in Figure 2(b) decreases with respect to the dimension p. The phenomenon also occurred in Zhu et al. (2017) (see their simulations). This problem is another interesting topic in high-dimensional statistical analysis; see Székely and Rizzo (2013a) and Kim et al. (2020).

Example 3. *The marginal distributions are independent of Cauchy distributions. Sample 1 is Cauchy*$(0, \delta)$ *with the scale parameter* δ. *Samples 2-4* *each have standard Cauchy*$(0, 1)$.

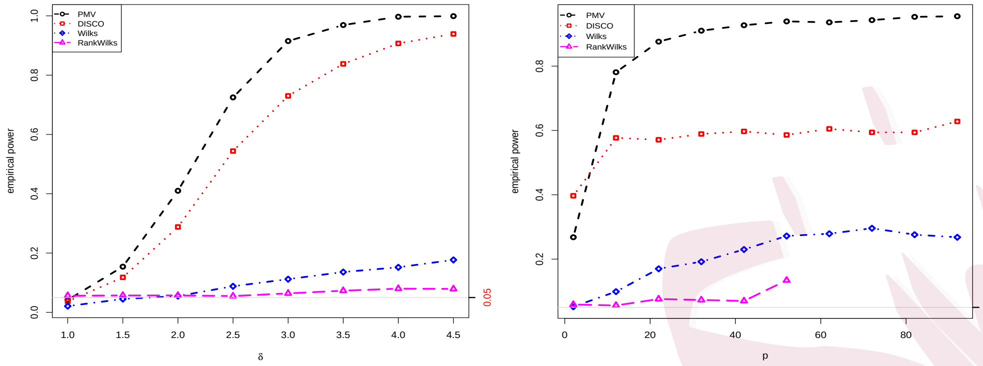
Example 3 is designed to evaluate the finite sample performance of our method for the K-sample hypothesis test of equal scale parameters. The results in Table 4 indicate that the empirical sizes of the PMV, DISCO and RankWilks tests are very close to the significance levels.

Table 4. Example 3: Empirical type-I error rate with $p = 10$.

	$n = 30$			$n = 50$		
Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
PMV	0.009	0.042	0.091	0.012	0.046	0.097
DISCO	0.007	0.037	0.079	0.016	0.054	0.109
Wilks	0.003	0.021	0.055	0.001	0.026	0.066
RankWilks	0.010	0.056	0.112	0.010	0.047	0.096

From Figure 3(a), it can be seen that the PMV test still has superior performance over the other three methods. As expected, the Wilks and RankWilks tests lose efficiency in such a scale model. Figure 3(b) suggests that the power of the PMV test is increasingly superior relative to the other methods as the dimension increases. In addition, we can see that the power of the DISCO test is increasing slowly in Figure 3(b), partly because $E[\|\mathbf{X}\|] < \infty$ is not satisfied.

Example 4. *In Sample 1, the marginal distributions are independent of*



(a) (b)

Figure 3. Example 3: Empirical power comparisons at the 0.05 significance level for $n = 30$: (a) δ varies with $p = 10$; (b) p varies and $\delta = 2.5$.

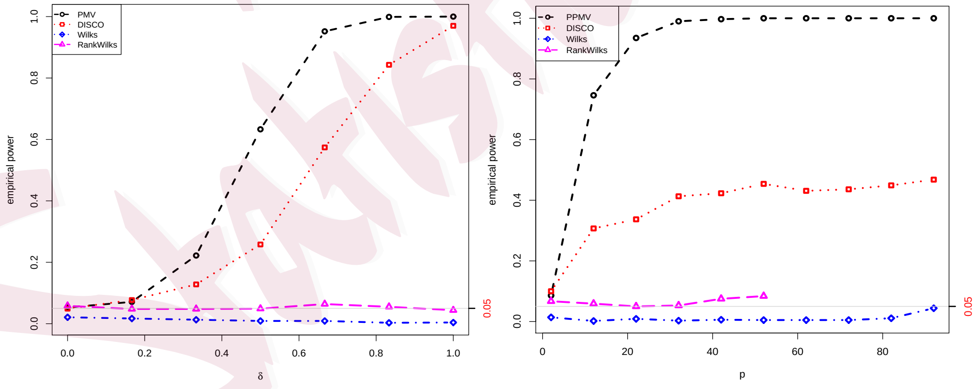
the mixture distributions $\delta N(0, 1) + (1 - \delta)\,Cauchy(0, 1)$, $\delta \in [0, 1]$. *Samples 2-4 each have* $Cauchy(0, 1)$ *distributions.*

From Example 4, the mixing weight $\delta = 0$ indicates that H_0 is true, and $\delta \neq 0$ suggests that H_0 is false. The simulation results are summarized in Table 5 and Figure 4. The results again indicate that the PMV test can roughly achieve the nominal significance levels at $\delta = 0$ and has almost the highest power at $\delta \neq 0$ when the dimension is fixed or increases.

Example 5. *Rizzo and Székely (2010) generalized the original DISCO decomposition to the* α*-DISCO decomposition by replacing the* $\|\cdot\|$*-norm with the* $\|\cdot\|_\alpha$*-norm for* $\alpha \in (0, 2]$*. For convenience, we call it the* α*-DISCO test. They proved that the* α*-DISCO tests work if* $E[\|\boldsymbol{X}\|^\alpha] < \infty$*. In this example,*

Table 5. Example 4: Empirical type-I error rate with $p = 10$.

	$n = 30$			$n = 50$		
Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
PMV	0.011	0.053	0.110	0.009	0.046	0.110
DISCO	0.009	0.049	0.096	0.006	0.046	0.098
Wilks	0.001	0.021	0.058	0.001	0.011	0.045
RankWilks	0.013	0.058	0.110	0.009	0.064	0.106



(a)

(b)

Figure 4. Example 4: Empirical power comparisons at the 0.05 significance level for $n = 30$: (a) δ varies with $p = 10$; (b) p varies and $\delta = 0.5$.

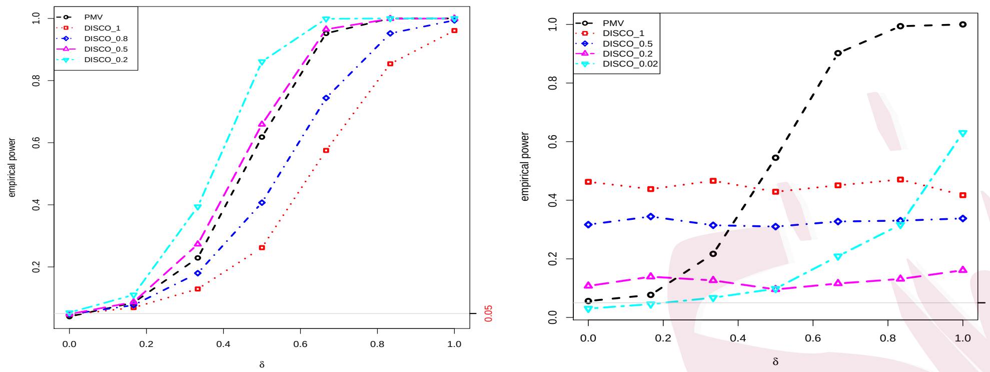
we would like to compare PMV with the α-DISCO test. The following two settings are studied:

Case (i): *The data are generated from Example 4;*

Case (ii): *In Sample 1, the marginal distributions are independent of the mixture distributions $\delta\,Cauchy(0,1) + (1-\delta)\exp\{Cauchy(0,1)\}$, $\delta \in [0,1]$. Samples 2-4 each have $\exp\{Cauchy(0,1)\}$ distributions.*

The simulation results for Example 5 are summarized in Figure 5 and Table 6, where DISCO_1, DISCO_0.8, DISCO_0.5, DISCO_0.2 and DISCO_0.02 represent the α-DISCO test with $\alpha = 1, 0.8, 0.5, 0.2$ and 0.02, respectively. For any $\alpha \in (0,1)$, it is easy to see that $E[\|\mathbf{X}\|^{\alpha}] < \infty$ but $E[\|\mathbf{X}\|] = \infty$ in Case (i) and $E[\|\mathbf{X}\|^{\alpha}] = \infty$ in Case (ii). Figure 5 Case (i) indicates that the α-DISCO tests work well for the empirical type-I error rate and empirical power in Case (i), which is consistent with Rizzo and Székely (2010). We can also see that DISCO_0.2 performs best, followed by PMV and DISCO_0.5 and then DISCO_0.8 and DISCO_1.

Table 6 illustrates that DISCO_1 and DISCO_0.5 cannot control the empirical type-I error rate in Case (ii). Figure 5 Case (ii) shows that the PMV test works the best, whereas DISCO_0.2 and DISCO_0.02 have inferior powers. In Figure 5 Case (ii), we only report the α-DISCO test where the



Case (i) Case (ii)

Figure 5. Example 5: Empirical power comparisons at the 0.05 significance level for $n = 30$ and $p = 10$.

minimum value of α is set to 0.02. A smaller α has also been considered, and it has been found that our method is still better than the α-DISCO test in this setting (here, we do not report the results).

Table 6. Example 5 Case (ii): Empirical type-I error rate with $p = 10$.

	$n = 30$			$n = 50$		
Method	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
PMV	0.006	0.056	0.096	0.009	0.046	0.095
DISCO_1	0.438	0.463	0.499	0.433	0.451	0.476
DISCO_0.5	0.287	0.317	0.333	0.256	0.293	0.323
DISCO_0.2	0.074	0.107	0.149	0.024	0.049	0.104
DISCO_0.02	0.014	0.030	0.083	0.000	0.030	0.104

From the above results, the finite sample performance of the PMV test is quite encouraging. In Example 1 Case (i), where data follow from $t(4)$ distributions, the PMV and DISCO tests behave comparably well. However, the PMV test outperforms the DISCO test in Example 1 Case (ii), where $E[\|\mathbf{X}\|] < \infty$ but $E[\|\mathbf{X}\|]$ is large. In Examples 2–4, where the data are generated from heavy-tailed distributions with infinite moments, our test exhibits superior performance over the other tests. In Example 5, the PMV and α-DISCO tests perform basically comparably in Case (i), and the PMV test works best than the α-DISCO tests with different α in Case (ii). Our limited experience demonstrates that the PMV test is very effective when the moments are large or data include outliers.

7. Real data analysis

The section illustrates our method by empirical analysis of two real datasets.

Example 6 (Michigan lung cancer data)**.** *This example considers Michigan lung cancer data, which has been analyzed by Subramanian et al. (2005). The dataset consists of observations of 86 samples on 5,217 gene expression levels from two classes: 62 in the "good outcomes" class and 24 in the "poor outcomes" class. The dataset is available at* `http://statweb.stanford.edu/~ckirby/brad/LSI/datasets-and-programs/datasets.html`.

We apply the proposed method to measure the differences between the "good outcomes" and "poor outcomes" classes. Since the dataset contains 86 samples, the statistical inference becomes a $p \gg n$ problem for which the Wilks-type methods fail. The PMV and DISCO tests with $B = 999$ permutations are listed in Table 7. The results suggest that both of them can detect significant differences between the good and poor outcome groups.

We also perform PMV and DISCO tests on subsets of the original data to provide power comparisons. Specifically, for some given subsample size, we pick a subsample from the full data uniformly at random. Then, we repeat each resampling 200 times and obtain the empirical power of each test method. In Figure 6, we conduct resamplings with subsample sizes from 30 to 86 and report empirical powers with $B = 199$ permutations at significance levels of 0.05 and 0.1. Figure 6 shows that the proposed method significantly outperforms DISCO in the dataset.

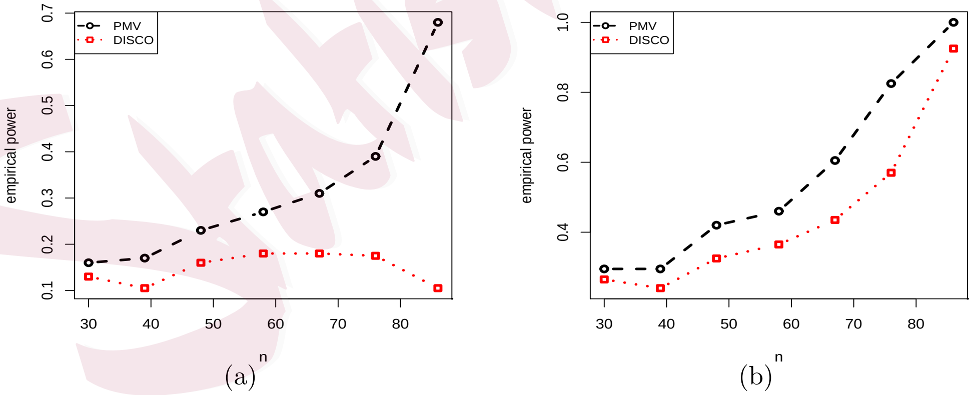
Example 7 (Prostate data)**.** *In this example, we consider the prostate dataset in the MultNonParam package for R. The dataset consists of 101 prostate cancer patients and 5 features for each patient. The 5 feature variables are hospital in which the patient is hospitalized (hosp), stage of the cancer (stage), used to help evaluate the prognosis of the cancer (gleason), prostate-specific antigen (psa) and age of the patient (age).*

hosp is a factor variable that consists of three levels: A, B and C hospi-

Table 7. Analysis of Michigan lung cancer data.

Methods	Source	Df	Sum	Mean	F-ratio	p-value
PMV	Between	1	106.647[1]	106.647	1.215	0.048
	Within	84	7370.344[2]	87.742		
	Total	85	7476.991[3]			
DISCO	Between	1	23.338	23.338	1.172	0.086
	Within	84	1672.408	19.910		
	Total	85	1695.746			

[1] $n^2\mathrm{SS}_B$ [2] $n^2\mathrm{SS}_W$ [3] $n^2\mathrm{SS}_T$.



(a) (b)

Figure 6. Michigan lung cancer data: Empirical power comparisons (a) at the 0.05 significance level; (b) at the 0.1 significance level.

tals. In the analysis, our interest is to check whether there is heterogeneity between the three hospitals. To this end, we test independence between $\mathbf{X} = (gleason, psa, age)^T$ and $Y = hosp$. Table 8 reports the p-values of the PMV, DISCO, Wilks and RankWilks tests, where $B = 999$ permutation replicates are carried out for the PMV and DISCO tests. From Table 8, it can be seen that the DISCO fails to detect the difference between three hospitals, whereas PMV, as well as Wilks and RankWilks, is able to reveal significant distinctions between hospitals. The reasonability of the result is supported by the boxplots of the data in Figure 7.

Table 8. Analysis of prostate data.

Methods	Source	Df	Sum	Mean	F-ratio	p-value
PMV	Between	2	342.534[1]	171.267	1.675	**0.056**
	Within	98	10018.56[2]	102.230		
	Total	100	10361.1[3]			
DISCO	Between	2	17.390	8.695	1.427	**0.15**
	Within	98	597.179	6.093		
	Total	100	614.570			
Wilks		Df	Wilks	approx F		p-value
	Between	2	0.810	3.547		**0.002**
RankWilks			Wilks	Chi2-Value		
			0.8134	20.034		**0.003**

[1] $n^2\mathrm{SS}_B$ [2] $n^2\mathrm{SS}_W$ [3] $n^2\mathrm{SS}_T$.

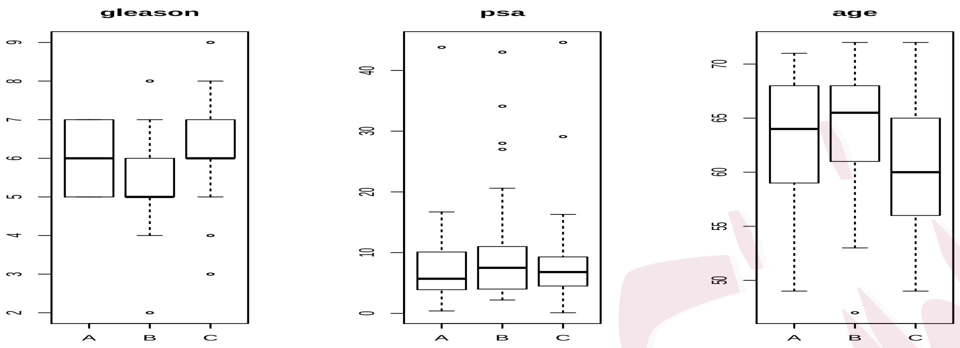


Figure 7. Boxplots for prostate data.

8. Discussion

In this paper, we propose a novel nonparametric multivariate multisample test based on the projection method and the mean variance index. The proposed method is equivalent to testing the independence between a continuous random vector and a categorical variable. The proposed test is consistent against all fixed alternatives and robust to heavy-tailed data, applicable in arbitrary dimensions, regardless of sample size.

Note that the time complexity for the DISCO statistic is $O(K^2pn^2)$, but that for the PMV statistic is $O(Kpn^3)$. Thus, DISCO may be faster than PMV for a small K. In fact, all projection-based methods, such as PC test (Zhu et al., 2017) and multivariate CvM test (Kim et al., 2020), also suffer from problems. However, we think that it may be significantly improved by the sketch approach (Pham and Pagh, 2012), which can easily be extended to our method. As suggested by Pham and Pagh (2012), it is a near-linear time approximation algorithm, which needs further research.

Although our theoretical results are obtained only for the setting p is fixed, we evaluate the finite sample performance in both small p and large p settings in our numerical studies. Thus, it is desirable to establish similar theoretical properties in the large p setting, such as the consistency of the PMV test and the limiting distributions of F_n under H_0 and H_1. We can refer to the works in Székely and Rizzo (2013a) and Kim et al. (2020). This is our future research.

Supplementary Materials

The online supplementary materials contain proofs of the theoretical results.

Acknowledgments

We sincerely thank the editor, the associate editor and the reviewers for their insightful suggestions, which have improved the manuscript significantly. Jicai Liu's research was supported by the National Science Foundation of China (11426156, 11501372, 11971018) and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (East China Normal University), Ministry of Education

(KLATASDS2001). Riquan Zhang's research was supported by National Natural Science Foundation of China (11571112, 11701360).

References

Baringhaus, L. and Franz, C. (2004). On a new multivariate two-sample test. *Journal of Multivariate Analysis*, 88(1):190–206.

Borgwardt, K. M., Gretton, A., Rasch, M. J., Kriegel, H. P., Schölkopf, B., and Smola, A. J. (2006). Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):e49–e57.

Cui, H., Li, R., and Wei, Z. (2015). Model-free feature screening for ultrahigh dimensional discriminant analysis. *Journal of the American Statistical Association*, 110(510):630–641.

Cui, H. and Zhong, W. (2019). A distribution-free test of independence based on mean variance index. *Computational Statistics & Data Analysis*, 139:117 – 133.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, 22(6):1030–1051.

Kim, I., Balakrishnan, S., and Wasserman, L. (2020). Robust multivariate nonparametric tests via projection-averaging. *The Annals of Statistics to appear*.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *Annals of Mathematical Statistics*, 30(2):420–447.

Kuo, H. H. (1975). *Gaussian measures in Banach spaces.* Springer-Verlag.

Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621.

Nath, R. and Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics & Data Analysis*, 2(4):297–315.

Pham, N. and Pagh, R. (2012). A near-linear time approximation algorithm for angle-based outlier detection in high-dimensional data. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 877–885. ACM.

Rizzo, M. L. and Székely, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *The Annals of Applied Statistics*, 4(2):1034–1055.

Scholz, F. W. and Stephens, M. A. (1987). K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82(399):918–924.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794.

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied*

Statistics, 3(4):1236–1265.

Székely, G. J. and Rizzo, M. L. (2013a). The distance correlation-test of independence in high dimension. *Journal of Multivariate Analysis*, 117(3):193–213.

Székely, G. J. and Rizzo, M. L. (2013b). Energy statistics: A class of statistics based on distances. *Journal of Statistical Planning and Inference*, 143(8):1249–1272.

Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, 24:471–494.

Zhu, L., Xu, K., Li, R., and Zhong, W. (2017). Projection correlation between two random vectors. *Biometrika*, 104(4):829–843.

School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance

KLATASDS-MOE, East China Normal University

E-mail: liujicai1234@126.com

Department of Statistics and Actuarial Science, University of Hong Kong

E-mail: u3006932@connect.hku.hk

School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance

E-mail: wcx_stat@126.com

Department of Statistics, East China Normal University

E-mail: zhangriquan@163.com