Statistica Sinica Preprint No: SS-2017-0236	
Title	A test for equality of two distributions via integrating characteristic functions
Manuscript ID	SS-2017-0236
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202017.0236
Complete List of Authors	Yiming Liu Zhi Liu and Wang Zhou
Corresponding Author	Zhi Liu
E-mail	liuzhi@umac.mo
Notice: Accepted version subject to English editing.	

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

A TEST FOR EQUALITY OF TWO DISTRIBUTIONS VIA INTEGRATING CHARACTERISTIC FUNCTIONS

Yiming LIU[1], Zhi LIU[2] and Wang ZHOU[3]

[1]*Nanyang Technological University*

[2]*University of Macau*

[3]*National University of Singapore*

Abstract: In this paper, we investigate the problem of testing the equality of two distributions by integrating the squared norm of the difference between two corresponding empirical characteristic functions. This results in a linear combination of three different U-statistics. So the original testing problem is reduced to testing whether this linear combination is zero or not. We apply the Jackknife Empirical Likelihood (JEL) method to the new hypothesis testing problem. It has been proved that, under multivariate case, the log JEL statistic after scaling tends to a chi-square distribution with degree of freedom 1. Simulation studies are presented to assess the finite-sample performance of our method.

Key words and phrases: Characteristic function; Jackknife Empirical Likelihood; U-statistic.

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

1. Introduction

In statistics, it is always interesting to know whether two multivariate populations are the same or have some common characteristics, such as population means and covariance matrices. The reason is that the comparison of multivariate populations has many real applications. One typical example lies in the case-control study in biomedicine. For example, assume that there are two sets of patients suffering from a particular sickness. The patients in one group are treated by a new cure and the other group is given the usual treatment. People would like to know if the new cure has any advantage by studying two sets of data which are often presented in the form of gene expressions of patients. Clearly one should apply some suitable statistical methods to make comparisons.

In the literature, there are some papers on this multivariate two-sample problem under the parametric setting. One can refer to Bai and Saranadasa (1996), Chen and Qin (2010), Li and Chen (2012), Wang et al. (2013) and references therein. However, under the nonparametric setting, this multivariate two-sample problem becomes more difficult. The classical methods, such as the nonparametric Kolmogorov-Smirnov test, the Wald-Wolfowitz runs test and Wilcoxon rank test, which are commonly used in the univariate population, no longer work. But there are still some works on the test of

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

two multivariate populations, such as Liu and Modarres (2011), Biswas and Ghosh (2014) and Liu et al. (2015). In the first two papers, the authors proposed two different nonparametric methods respectively to test the equality of two distributions based on one theoretical observation about inter-point distance by Maa et al. (1996). In Liu et al. (2015), a jackknife empirical likelihood (JEL) test, which works well in the case of small sample and asymmetric data, was proposed through incorporating characteristic functions. The inference based on empirical characteristic function was also investigated by Kellermeier (1980) and Fernandez et al. (2008).

In this paper, we will consider the integration of the difference of two empirical characteristic functions which could be expressed as a linear combination of three U-statistics. So the original testing problem is reduced to testing whether this linear combination equals zero or not. A similar idea appeared in Fernandez et al. (2008).

The U-statistics, including functions of U-statistics have been well studied in the literature. We refer to Bentkus et al. (2009) for more details for the classical results. Although one may attempt to apply the fundamental results of U-statistics, such as normal approximations to our testing problem, there is a better method, called JEL in the literature. The JEL method, proposed in Jing et al. (2009), works very well in the case of U-

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

statistics. There are many papers on applications of JEL. To name but a few, we mention Gong et al. (2010), Adimari and Chiogna (2012), Wang et al. (2013), and Zhao et al. (2015). The JEL approach not only inherits nice properties of Owen's EL (1988), such as range preserving and transformation respecting, but also is a very powerful tool to handle nonlinear statistics. The key idea of JEL is to turn the statistic of interest into a sample mean based on jackknife pseudo-values Quenouille (1956). If we can show that these pseudo-values are asymptotically independent, we can apply Owen's EL for the mean of the jackknife pseudo-values. Fortunately, the pseudo-values for U-statistics are asymptotically independent. In this paper, we will apply JEL to test if two multivariate distributions are equal or not. To the best of our knowledge, this is the first time to adapt JEL for functions of U-statistics.

The rest of this paper is organized as follows. In Section 2, we state our methodology and main results. Simulation studies are presented in Section 3. All the technical proofs are postponed to the Appendix.

2. Methodology

Suppose that $\mathbf{X} = (X_1, \ldots, X_p)^\top$ and $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$ are two p dimensional populations having distributions F and G, respectively. An impor-

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

tant issue in hypothesis testing is to consider the following problem:

$$\mathrm{H}_0 : F = G \quad \text{versus} \quad \mathrm{H}_1 : F \neq G. \tag{2.1}$$

In practice, we can only collect two groups of sample observations

$$\{\mathbf{X}_k = (X_{k1}, \cdots, X_{kp})^\top, 1 \leq k \leq n_1\} \text{ and } \{\mathbf{Y}_j = (Y_{j1}, \cdots, Y_{jp})^\top, 1 \leq j \leq n_2\}$$

from $\mathbf{X}$ and $\mathbf{Y}$, respectively, where n_1 and n_2 are the corresponding sample sizes.

The distribution of $\mathbf{X}$ is mutually determined by its characteristic functions (c.f.). So testing (2.1) is equivalent to testing whether $\phi(\mathbf{t}) = \psi(\mathbf{t})$ for all $\mathbf{t} := (t_1, t_2, \cdots, t_p)^\top \in R^p$, where $\phi(\mathbf{t}) = E(\mathrm{e}^{i\mathbf{t}^\top \mathbf{X}})$ and $\psi(\mathbf{t}) = E(\mathrm{e}^{i\mathbf{t}^\top \mathbf{Y}})$ are the c.f.'s of $\mathbf{X}$ and $\mathbf{Y}$, respectively. Hence, the problem (2.1) is reformulated by testing the hypothesis

$$\mathrm{H}_0 : \phi(\mathbf{t}) = \psi(\mathbf{t}) \text{ for all } \mathbf{t} \text{ v.s. } \mathrm{H}_1 : \phi(\mathbf{t}) \neq \psi(\mathbf{t}) \text{ for some } \mathbf{t} \neq 0. \tag{2.2}$$

2.1 Integration with respect to Lebesgue measure

By Riemann-Lebesgue's Lemma, $\lim_{|\mathbf{t}|\to\infty} \phi(\mathbf{t}) = 0$ for most of distributions. Hence when t is large, one can't tell $\phi(\mathbf{t})$ from $\psi(\mathbf{t})$. Therefore, instead of considering (2.2), one can consider a weaker hypothesis,

$$\mathrm{H}_0' : \phi(\mathbf{t}) = \psi(\mathbf{t}) \text{ for } |\mathbf{t}| \leq t_0 \text{ v.s. } \mathrm{H}_1' : \phi(\mathbf{t}) \neq \psi(\mathbf{t}) \text{ for some } |\mathbf{t}| \leq t_0. \tag{2.3}$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

where t_0 is some positive constant, and $j = 1, 2, \cdots, p$. One can refer to Section 3 for more discussions on t_0. Clearly H_0' is equivalent to

$$\int_{\Pi} \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 \, d\mathbf{t} = 0, \tag{2.4}$$

where $\Pi = [-t_0, t_0] \times \cdots \times [-t_0, t_0] \subset R^p$ and $\|\cdot\|$ denotes the norm of a complex number. Since

$$\begin{aligned} & \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 \\ = \; & E(\mathrm{e}^{i\mathbf{t}^\top(\mathbf{X}_1 - \mathbf{X}_2)}) + E(\mathrm{e}^{i\mathbf{t}^\top(\mathbf{Y}_1 - \mathbf{Y}_2)} - E(\mathrm{e}^{i\mathbf{t}^\top(\mathbf{X}_1 - \mathbf{Y}_1)}) - E(\mathrm{e}^{i\mathbf{t}^\top(\mathbf{Y}_1 - \mathbf{X}_1)}), \end{aligned}$$

one has

$$\begin{aligned} & \frac{1}{2}\int_{\Pi} \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 \, d\mathbf{t} \\ = \; & \frac{1}{2} E\Big(\int_{\Pi} \big(e^{i\mathbf{t}^\top(\mathbf{X}_1 - \mathbf{X}_2)} + e^{i\mathbf{t}^\top(\mathbf{Y}_1 - \mathbf{Y}_2)} - e^{i\mathbf{t}^\top(\mathbf{X}_1 - \mathbf{Y}_1)} - e^{i\mathbf{t}^\top(\mathbf{Y}_1 - \mathbf{X}_1)} \big) d\mathbf{t} \Big) \\ = \; & E[h(\mathbf{X}_1, \mathbf{X}_2)] + E[h(\mathbf{Y}_1, \mathbf{Y}_2)] - 2E[h(\mathbf{X}_1, \mathbf{Y}_1)], \end{aligned}$$

where

$$h(\mathbf{X}_1, \mathbf{X}_2) = \prod_{j=1}^{p} \frac{\sin[t_0(X_{1j} - X_{2j})]}{X_{1j} - X_{2j}}. \tag{2.5}$$

Define

$$\begin{aligned} U \; = \; & \frac{2}{n_1(n_1-1)} \sum_{1 \le k_1 < k_2 \le n_1} h(\mathbf{X}_{k_1}, \mathbf{X}_{k_2}) + \frac{2}{n_2(n_2-1)} \sum_{1 \le j_1 < j_2 \le n_2} h(\mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}) \\ & - \frac{2}{n_1 n_2} \sum_{k=1}^{n_1} \sum_{j=1}^{n_2} h(\mathbf{X}_k, \mathbf{Y}_j) \\ := \; & U_1 + U_2 - 2U_3. \end{aligned} \tag{2.6}$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Remark 2.1. The above statistic can also be expressed as

$$U = \binom{n_1}{2}^{-1} \binom{n_2}{2}^{-1} \sum_{1\le k_1<k_2\le n_1} \sum_{1\le j_1<j_2\le n_2} h(\mathbf{X}_{k_1}, \mathbf{X}_{k_2}; \mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}),$$

where

$$\begin{aligned} h(\mathbf{X}_{k_1}, \mathbf{X}_{k_2}; \mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}) &= h(\mathbf{X}_{k_1}, \mathbf{X}_{k_2}) + h(\mathbf{Y}_{j_1}, \mathbf{Y}_{j_2}) - \frac{1}{2}h(\mathbf{X}_{k_1}, \mathbf{Y}_{j_1}) \\ &\quad -\frac{1}{2}h(\mathbf{X}_{k_2}, \mathbf{Y}_{j_2}) - \frac{1}{2}h(\mathbf{X}_{k_1}, \mathbf{Y}_{j_2}) - \frac{1}{2}h(\mathbf{X}_{k_2}, \mathbf{Y}_{j_1}). \end{aligned}$$

This is a two-sample U-statistic of degree $(2, 2)$. However, it is degenerate under H_0. It is interesting to investigate this degenerate two-sample U-statistic. One can refer to Fernandez et al. (2008) for a similar version of (2.6) and its limiting properties.

Remark 2.2. Interestingly the test statistics proposed in Liu and Modarres (2011) and Biswas and Ghosh (2014) are two-sample U-statistics too.

So U is an unbiased estimator of $2^{-1}\int_{\Pi} \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 \, d\mathbf{t}$ under H_0. It is a linear combination of three different U-statistics, U_1, U_2 and U_3. We note that the hypothesis

$$\mathrm{H}_0'' : E(U_1) = E(U_2) = E(U_3) \tag{2.7}$$

implies H_0'. On the other hand, under H_0, we obviously have (2.7). So

$$\mathrm{H}_0 \Rightarrow \mathrm{H}_0'' \Rightarrow \mathrm{H}_0'.$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Based on the discussions between formulae (2.2) and (2.3), H_0 and H_0' are very close. Hence from now on, we will not distinguish these three hypotheses. Our test statistic will be U.

2.2 Integration with respect to probability measure

Let $f(t)$ be any probability density function supported on the real line. Suppose

$$f(t) > 0, \quad f(t) = f(-t)$$

for any $t \in R$. Define

$$f(\mathbf{t}) = \prod_{j=1}^{p} f(t_j),$$

which is a probability density functions supported on R^p. Clearly H_0 is equivalent to

$$\int_{R^p} \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 f(\mathbf{t}) d\mathbf{t} = 0, \tag{2.8}$$

A direct calculation shows that

$$\int_{R^p} \|\phi(\mathbf{t}) - \psi(\mathbf{t})\|^2 f(\mathbf{t}) d\mathbf{t} = E[h(\mathbf{X}_1, \mathbf{X}_2)] + E[h(\mathbf{Y}_1, \mathbf{Y}_2)] - 2E[h(\mathbf{X}_1, \mathbf{Y}_1)],$$

where

$$h(\mathbf{X}_1, \mathbf{X}_2) = \prod_{j=1}^{p} \int_R \cos\left[t_j(X_{1j} - X_{2j})\right] f(t_j) dt_j. \tag{2.9}$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Based on the above formula, we can propose another test statistic for H_0,

$$\begin{aligned} U &= \frac{2}{n_1(n_1-1)} \sum_{1\le k_1<k_2\le n_1} h(\mathbf{X}_{k_1},\mathbf{X}_{k_2}) + \frac{2}{n_2(n_2-1)} \sum_{1\le j_1<j_2\le n_2} h(\mathbf{Y}_{j_1},\mathbf{Y}_{j_2}) \\ &\quad -\frac{2}{n_1 n_2}\sum_{k=1}^{n_1}\sum_{j=1}^{n_2} h(\mathbf{X}_k,\mathbf{Y}_j) \\ &=: U_1+U_2-2U_3. \end{aligned} \tag{2.10}$$

In practise one can let $f(t) = \mathrm{e}^{-t^2/(2p)}/\sqrt{2\pi}$, the normal density function since in this case the kernel function has a simple expression,

$$h(\mathbf{X}_1,\mathbf{X}_2) = \prod_{j=1}^{p} \exp\big[-\frac{1}{2p}(X_{1j}-X_{2j})^2\big].$$

Remark 2.3. In the previous subsection,

$$h(\mathbf{X}_1,\mathbf{X}_2) = \prod_{j=1}^{p} \frac{\sin[t_0(X_{1j}-X_{2j})]}{X_{1j}-X_{2j}}.$$

Letting $f(t) = 2^{-1}I(-t_0 < t < t_0)$, one can see that it is a special case of the current one (2.9). This is the reason that we use one notation to denote different kernels. There will be no confusion because one can tell the meaning of $h(\mathbf{X}_1,\mathbf{X}_2)$ from the text. On the other hand, the motivations behind (2.5) and (2.9) are different. So we put these two statistics in two subsections although these two statistics have the same form.

2.3 JEL

Since U is a degenerate U-statistics whose asymptotic distribution contains unknown variance, we will have to further provide consistent estimates of

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

variance so that we can use the limiting distribution of U to do the hypothesis test. However, it is not easy to provide consistent estimates for degenerate U-statistics. On the other hand, it will decrease the inference accuracy if one has to estimate many parameters. Therefore, we employ the JEL approach, whose nice properties are already discussed in Section 1. Now we start to describe the detailed JEL procedure. Write

$$\begin{aligned} U_1 &= \binom{n_1}{2}^{-1} \sum_{1\le k<j\le n_1} h(\mathbf{X}_k, \mathbf{X}_j) \\ U_2 &= \binom{n_2}{2}^{-1} \sum_{1\le k<j\le n_2} h(\mathbf{Y}_k, \mathbf{Y}_j) \\ U_3 &= \binom{n}{2}^{-1} \sum_{1\le k<j\le n} \tilde{h}(\mathbf{Z}_k, \mathbf{Z}_j), \end{aligned}$$

where $n = n_1 + n_2$, $(\mathbf{Z}_1, \cdots, \mathbf{Z}_n) = (\mathbf{X}_1, \cdots, \mathbf{X}_{n_1}, \mathbf{Y}_1, \cdots, \mathbf{Y}_{n_2})$,

$$\tilde{h}(\mathbf{Z}_k, \mathbf{Z}_j) = \frac{1}{n_1 n_2} \binom{n}{2} h(\mathbf{X}_k, \mathbf{Y}_{j-n_1}),$$

for $1 \le k \le n_1 < j \le n_1 + n_2$, and 0 otherwise. Here $h(\mathbf{X}_1, \mathbf{X}_2)$ could be (2.5) or (2.9), or any kernel function satisfying the following **C2** and **C3**. The corresponding jackknife pseudo-values are defined by

$$V_i^{(s)} = n_s U_s - (n_s - 1) U_{s,n_s-1}^{(-i)},$$

for $s = 1, 2, 3$, $i = 1 \cdots n_s$, $n_3 = n$, where

$$U_{1,n_1-1}^{(-i)} = \binom{n_1 - 1}{2}^{-1} \sum_{(n_1-1,2)}^{(-i)} h(\mathbf{X}_k, \mathbf{X}_j),$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

$$U_{2,n_2-1}^{(-i)} = \binom{n_2-1}{2}^{-1} \sum_{(n_2-1,2)}^{(-i)} h(\mathbf{Y}_k, \mathbf{Y}_j),$$

$$U_{3,n_3-1}^{(-i)} = \binom{n_3-1}{2}^{-1} \sum_{(n_3-1,2)}^{(-i)} \tilde{h}(\mathbf{Z}_k, \mathbf{Z}_j).$$

Here $\sum_{(n_s-1,2)}^{(-i)}$ is the summation over $\{(j,k) : 1 \le j < k \le n_s-1, j \ne i, k \ne i\}$. It is well-known that $U_s = n_s^{-1} \sum_{i=1}^{n_s} V_i^{(s)}$, for $s = 1, 2, 3$. Therefore, under H_0'', $E(V_i^{(1)}) = \theta_0, E(V_j^{(2)}) = \theta_0$, and

$$E(V_k^{(3)}) = \frac{n_3\theta_0}{n_3-2}\Big[\frac{n_2-1}{n_1} I(0 \le k \le n_1) + \frac{n_1-1}{n_2} I(n_1+1 \le k \le n)\Big],$$

where $\theta_0 = E[h(\mathbf{X}_1, \mathbf{X}_2)]$.

We now apply JEL to these three sets of jackknife pseudo values. Let $\mathbf{p} = (p_1, \cdots, p_{n_1})$, $\mathbf{q} = (q_1, \cdots, q_{n_2})$ and $\mathbf{r} = (r_1, \cdots, r_n)$ be three probability vectors. Hence, the empirical likelihood is

$$L = \sup_{(\mathbf{p},\mathbf{q},\mathbf{r},\theta)} \Big(\prod_{i=1}^{n_1} p_i\Big)\Big(\prod_{j=1}^{n_2} q_i\Big)\Big(\prod_{k=1}^{n} r_i\Big), \tag{2.11}$$

subject to the following constraints

$$\sum_{i=1}^{n_1} p_i(V_i^{(1)} - \theta) = 0, \ \sum_{i=1}^{n_2} q_i(V_i^{(2)} - \theta) = 0, \ \sum_{i=1}^{n} r_i(V_i^{(3)} - E(V_i^{(3)})) = 0.$$

So the corresponding empirical log likelihood ratio statistic is

$$l = -2\Big(\sum_{i=1}^{n_1} \log(n_1 p_i) + \sum_{j=1}^{n_2} \log(n_2 q_j) + \sum_{k=1}^{n} \log(n r_k)\Big) \tag{2.12}$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

with

$$p_i = \frac{1}{n_1} \cdot \frac{1}{1+\lambda_1(V_i^{(1)}-\theta)}, \quad i = 1, \cdots n_1,$$
$$q_j = \frac{1}{n_2} \cdot \frac{1}{1+\lambda_2(V_i^{(2)}-\theta)}, \quad j = 1, \cdots n_2,$$
$$r_k = \frac{1}{n} \cdot \frac{1}{1+\lambda_3(V_i^{(3)}-EV_i^{(3)})}, \quad k = 1, \cdots n,$$

where $(\lambda_1, \lambda_2, \lambda_3, \theta)$ are solutions to the following four score equations

$$\frac{\partial}{\partial\theta} l = \lambda_1 \sum_{i=1}^{n_1} \frac{-1}{1+\lambda_1(V_i^{(1)}-\theta)} + \lambda_2 \sum_{i=1}^{n_2} \frac{-1}{1+\lambda_2(V_i^{(2)}-\theta)} + \tag{2.13}$$
$$\lambda_3 \sum_{i=1}^{n_1} \frac{-\frac{n}{n-2} \cdot \frac{n_2-1}{n_1}}{1+\lambda_3(V_i^{(3)}-E(V_i^{(3)}))} + \lambda_3 \sum_{i=1}^{n_2} \frac{-\frac{n}{n-2} \cdot \frac{n_1-1}{n_2}}{1+\lambda_3(V_{n_1+i}^{(3)}-E(V_{n_1+i}^{(3)}))} = 0,$$
$$\frac{\partial}{\partial\lambda_1} l = \sum_{i=1}^{n_1} \frac{V_i^{(1)}-\theta}{1+\lambda_1(V_i^{(1)}-\theta)} = 0, \tag{2.14}$$
$$\frac{\partial}{\partial\lambda_2} l = \sum_{i=1}^{n_2} \frac{V_i^{(2)}-\theta}{1+\lambda_2(V_i^{(2)}-\theta)} = 0, \tag{2.15}$$
$$\frac{\partial}{\partial\lambda_3} l = \sum_{i=1}^{n} \frac{V_i^{(3)}-E(V_k^{(3)})}{1+\lambda_3(V_i^{(3)}-E(V_i^{(3)}))} = 0. \tag{2.16}$$

To get Wilks' theorem for l, we need the following conditions:

- **C1.** As $\min(n_1, n_2) \to \infty$, $n_j/(n_1+n_2+n) \to \gamma_j > 0$, $j = 1, 2$, where $\gamma_1 + \gamma_2 = 1/2$. We also denote $\gamma_3 = 1/2$, so that $\gamma_1 + \gamma_2 + \gamma_3 = 1$.

- **C2.** $\sigma_{g_1} > 0$ and $\sigma_{g_2} > 0$.

- **C3.** $\sigma_{1,0} > 0$ and $\sigma_{0,1} > 0$, where the definitions of σ_{g_1}, σ_{g_2}, $\sigma_{1,0}$ and $\sigma_{0,1}$ are given at the beginning of the Appendix.

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Theorem 2.4. Under **C1-C3** and H_0, as $\min(n_1, n_2) \to \infty$, the empirical log likelihood ratio $\omega^{-1}l$ converges in distribution to χ_1^2, where χ_1^2 is a chi-square random variable with one degree of freedom, and

$$\omega = \frac{1}{16\gamma_1\gamma_2} - \frac{2}{1+16\gamma_1\gamma_2} + \frac{64\gamma_1\gamma_2}{(1+16\gamma_1\gamma_2)^2}.$$

Remark 2.5. The limiting distribution in Theorem 2.4 contains unknown parameters γ_1 and γ_2. In practice, one can estimate them by n_1/N and n_2/N respectively, where $N := n_1 + n_2 + n$.

3. Simulation

In this section, we investigate the finite sample performance of our proposed test statistics and compare with other existing methods in literature.

3.1 Simulation design

We generate the p-dimensional samples $\{\mathbf{X}_k, k = 1, \cdots, n_1\}$ and $\{\mathbf{Y}_j, j = 1, \cdots, n_2\}$ from the distributions in the following cases.

Case I: Normal distribution with non-diagonal covariance matrix. We generate the $\mathbf{X}_k, k = 1, 2, \cdots, n_1$ from the multi-variate normal distribution $N(\mathbf{0}, \Sigma)$, where Σ is a $p \times p$ matrix whose elements in the diagonal are 1 and in the off-diagonal are 0.5, and $\mathbf{Y}_j, j = 1, 2, \cdots, n_2$ are generated from $N(\mathbf{0}, (1+\delta)^2 \cdot \Sigma)$, where δ is a real number;

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Case II: Exponential distribution. We generate the $\mathbf{X}_k, k = 1, 2, \cdots, n_1$ from the standard multi-variate exponential distribution with independent components, namely, the components of $\mathbf{X}_k$ are independently generated from $Exp(1)$, and the components of $\mathbf{Y}_j, j = 1, 2, \cdots, n_2$ are independently generated from $Exp(1+\delta)$, where δ is a real number.

For all cases above, we consider $p = 2, 3$, $(n_1, n_2) = (50, 50), (40, 60)$ and (100,100), $\delta = 0, 0.2, 0.4, 0.6, 0.8, 1$, $\delta = 0$ is used to assess the size of the test statistics and the other cases are used to assess the power. For Case I, we also consider $p = 10$.

We compute the statistics in Sections 2.1 and 2.2, that is, $U^{(t_0)}$ with kernel function

$$h(\mathbf{X}_1, \mathbf{X}_2) = \prod_{j=1}^{p} \frac{\sin[t_0(X_{1j} - X_{2j})]}{X_{1j} - X_{2j}}$$

which is the statistic in (2.6) of Section 2.1 with $t_0 = 1, 1.5, 2$, and U with kernel function

$$h(\mathbf{X}_1, \mathbf{X}_2) = \prod_{j=1}^{p} \exp\{-\frac{1}{2p}(X_{1j} - X_{2j})^2\}$$

which is the statistic in (2.10) of Section 2.2). For the purpose of comparison, we compute the test statistics of Liu et al. (2015) (LXZ) with $t = 1$ which is the best performed case in their paper, and the test statistics pro-

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

posed by Biswas and Ghosh (2014) (BG). We refer to Theorem 2.1 in Liu et al. (2015) and Theorem 4.1 in Biswas and Ghosh (2014), respectively. For Case I, we also compare the proposed method with Box's test. Specifically, define

$$S = \frac{1}{n-2}\{(n_1-1)S_1 + (n_2-1)S_2\},$$

$$M = (n-2)\log\det(S) - (n_1-1)\log\det(S_1) - (n_2-1)\log\det(S_2),$$

$$c = \frac{2p^2+3p-1}{6(p-1)}(\frac{1}{n_1-1} + \frac{1}{n_2-1} - \frac{1}{n-2}),$$

where S_1 and S_2 are sample covariance matrices corresponding to $\mathbf{X}$ and $\mathbf{Y}$, respectively. Then Box's test tells one that when $\mathbf{X}$ and $\mathbf{Y}$ are multivariate normal,

$$(1-c)M \sim \chi^2_{p(p+1)/2}.$$

3.2 Simulation results

The simulation is repeated 10000 times. Tables 1, 3 and 5 display the empirical sizes and powers at the nominal significance levels 0.1 and 0.05 under H_0 for Case I. Tables 2 and 4 show the empirical sizes and powers for the Case II.

We have the following observations from simulation results:

(a) From Table 1 to Table 5, we can see that the proposed statistic $\omega^{-1}l$ in Section 2 with $t_0 = 1$ performs better than $t_0 = 1.5$ and 2 in terms of

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

power and size. This observation is consistent with our expectation. The reason is as follows. Large t_0 may cause cumulative errors. A small disturbance of data could lead to large deviations of $t_0\mathbf{X}$ and $t_0\mathbf{Y}$ even if $\mathbf{X}$ and $\mathbf{Y}$ are close to each other. On the other hand, small t would make $t_0\mathbf{X}$ and $t_0\mathbf{Y}$ close to zero even if $\mathbf{X}$ and $\mathbf{Y}$ are totally different. Thus, choosing $t_0 = 1$ is reasonable. We also conduct experiments for small t_0 and large t_0 (e.g., $t_0 = 0.2$, $t_0 = 8$), and the corresponding results further confirm our conclusion. So we suggest to let $t_0 = 1$ for the proposed statistic $\omega^{-1}l$ in Section 2 in practice.

(b) In Table 1, we see that both our methods ($U^{(1)}$ and U) and LXZ perform much better than BG in terms of size, and ours and LXZ are very competitive. The sizes of ours and LXZ are very close to the nominal one. In Tables 2 and 4, the sizes of our two methods are closer to the nominal level than LXZ and BG. From Tables 1, 3 and 5, we see that our two methods perform better than LXZ in terms of power. From Table 2 and 4, our proposed methods perform better than LXZ in terms of powers as well. On the other hand, when δ is small (for example, $\delta = 0.2$), the empirical powers of BG are slightly better than ours. But when δ becomes large, ours dominate BG again.

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

(c) When $\mathbf{X}$ and $\mathbf{Y}$ are normal, $U^{(1)}$ and U perform competitively with Box's test in terms of size. However when δ is nonzero, the powers of $U^{(1)}$ and U are better than Box's test. Of course as δ becomes large, this advantage disappears gradually.

From the above numerical observations, we can conclude that our two methods perform better than the existing methods in the literature.

4. Application

In this part, we apply the proposed method to a gene data set of 6033 gene expressions, which were obtained from 102 independent observations consisting of 52 prostate cancer patients and 50 healthy men. Our purpose is to detect all the genes which have different distributions in two groups. So it is natural to consider genes individually.

This data set was analyzed in Liu et al. (2015), and they asserted that 14 common genes are significant with Bonferroni correction, i.e., D_1=$\{$332, 377, 610, 905, 1082, 1113, 1458, 1557, 1589, 1620, 1647, 2450, 3439, 4405$\}$. Thus, we would like to discover whether there are any other relevant genes that can be identified by our proposed methods.

We apply our proposed two methods to each gene variable individually. For the method of integration with respect to Lebesgue measure, we set

Table 1: Powers at nominal level $\alpha = 0.1$ and 0.05 for Case I ($p = 2$).

δ	$\alpha = 0.1$							$\alpha = 0.05$						
	$(n_1, n_2) = (50, 50)$													
	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX
0	0.102	0.099	0.090	0.099	0.108	0.167	0.091	0.054	0.047	0.038	0.046	0.059	0.115	0.044
0.2	0.211	0.207	0.188	0.204	0.165	0.272	0.164	0.130	0.125	0.113	0.125	0.092	0.196	0.093
0.4	0.453	0.439	0.398	0.437	0.290	0.482	0.342	0.334	0.311	0.276	0.313	0.191	0.379	0.235
0.6	0.690	0.663	0.620	0.666	0.441	0.676	0.562	0.569	0.539	0.479	0.542	0.314	0.577	0.431
0.8	0.849	0.825	0.788	0.832	0.596	0.811	0.760	0.761	0.725	0.671	0.730	0.472	0.733	0.643
1	0.944	0.929	0.895	0.933	0.723	0.887	0.892	0.890	0.863	0.810	0.866	0.608	0.829	0.811
	$(n_1, n_2) = (40, 60)$													
0	0.098	0.103	0.040	0.093	0.119	0.159	0.097	0.046	0.055	0.117	0.044	0.062	0.102	0.048
0.2	0.204	0.203	0.087	0.192	0.172	0.306	0.159	0.121	0.119	0.123	0.109	0.105	0.219	0.091
0.4	0.440	0.428	0.237	0.413	0.286	0.597	0.331	0.313	0.299	0.257	0.286	0.196	0.483	0.217
0.6	0.675	0.653	0.436	0.643	0.443	0.814	0.552	0.540	0.518	0.453	0.506	0.319	0.723	0.414
0.8	0.839	0.812	0.633	0.816	0.590	0.928	0.740	0.742	0.713	0.646	0.708	0.469	0.878	0.616
1	0.931	0.915	0.779	0.919	0.715	0.971	0.877	0.878	0.840	0.781	0.844	0.605	0.949	0.783
	$(n_1, n_2) = (100, 100)$													
0	0.098	0.096	0.095	0.098	0.101	0.125	0.095	0.048	0.042	0.046	0.045	0.053	0.072	0.044
0.2	0.329	0.323	0.303	0.323	0.214	0.451	0.255	0.225	0.221	0.200	0.216	0.135	0.336	0.162
0.4	0.717	0.705	0.668	0.701	0.450	0.853	0.600	0.595	0.583	0.539	0.574	0.330	0.773	0.468
0.6	0.927	0.912	0.891	0.914	0.711	0.983	0.870	0.867	0.855	0.820	0.851	0.590	0.962	0.794
0.8	0.986	0.981	0.978	0.984	0.881	0.998	0.971	0.972	0.960	0.945	0.964	0.803	0.996	0.945
1	0.998	0.993	0.995	0.997	0.955	1.000	0.995	0.993	0.986	0.987	0.992	0.910	0.999	0.989

$t = 1$ and denote it as Method 1. For simplicity, we also denote the method of integration respect to probability measure as Method 2. The experiment is conducted under the significance level $\alpha = 0.1$ with Bonferroni correction.

We denote D_2 and D_3 as the index sets of genes that are selected by Methods 1 and 2, respectively, where $D_2 = \{2, 332, 377, 398, 905, 1082, 1113,$

Table 2: Powers at nominal level $\alpha = 0.1$ and 0.05 for Case II ($p = 2$).

δ	$\alpha = 0.1$						$\alpha = 0.05$					
	$(n_1, n_2) = (50, 50)$											
	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG
0	0.105	0.091	0.085	0.100	0.132	0.174	0.058	0.044	0.041	0.055	0.073	0.121
0.2	0.289	0.283	0.283	0.299	0.243	0.369	0.194	0.182	0.180	0.198	0.155	0.277
0.4	0.609	0.609	0.611	0.635	0.523	0.649	0.490	0.482	0.484	0.504	0.408	0.556
0.6	0.843	0.852	0.855	0.870	0.801	0.838	0.754	0.755	0.756	0.781	0.702	0.765
0.8	0.949	0.955	0.957	0.964	0.941	0.920	0.901	0.915	0.914	0.925	0.897	0.883
1	0.985	0.988	0.988	0.990	0.990	0.941	0.970	0.975	0.975	0.978	0.974	0.929
	$(n_1, n_2) = (40, 60)$											
0	0.110	0.093	0.092	0.108	0.140	0.180	0.055	0.047	0.043	0.052	0.078	0.126
0.2	0.288	0.285	0.293	0.297	0.223	0.345	0.190	0.186	0.192	0.196	0.143	0.263
0.4	0.599	0.603	0.614	0.626	0.495	0.636	0.483	0.480	0.488	0.504	0.364	0.541
0.6	0.844	0.843	0.843	0.863	0.773	0.839	0.748	0.752	0.755	0.781	0.665	0.768
0.8	0.947	0.946	0.952	0.960	0.922	0.914	0.903	0.901	0.911	0.921	0.871	0.881
1	0.985	0.986	0.988	0.991	0.981	0.935	0.968	0.970	0.971	0.979	0.959	0.921
	$(n_1, n_2) = (100, 100)$											
0	0.103	0.103	0.100	0.106	0.108	0.145	0.054	0.055	0.047	0.055	0.053	0.086
0.2	0.431	0.439	0.444	0.457	0.338	0.485	0.312	0.320	0.317	0.334	0.224	0.372
0.4	0.856	0.864	0.864	0.880	0.778	0.884	0.764	0.776	0.781	0.800	0.671	0.810
0.6	0.981	0.980	0.984	0.987	0.972	0.987	0.959	0.964	0.961	0.973	0.942	0.971
0.8	0.998	0.997	0.998	0.999	0.998	0.996	0.995	0.995	0.996	0.999	0.996	0.995
1	1.000	0.999	0.998	1.000	1.000	0.996	1.000	0.998	0.997	1.000	1.000	0.996

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Table 3: Powers at nominal level $\alpha = 0.1$ and 0.05 for Case I ($p = 3$).

δ	$\alpha = 0.1$							$\alpha = 0.05$						
	$(n_1, n_2) = (50, 50)$													
	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX
0	0.090	0.085	0.080	0.087	0.122	0.166	0.103	0.045	0.042	0.039	0.043	0.063	0.115	0.049
0.2	0.266	0.253	0.219	0.256	0.174	0.300	0.173	0.173	0.159	0.126	0.161	0.102	0.224	0.103
0.4	0.576	0.566	0.502	0.563	0.312	0.546	0.378	0.436	0.432	0.363	0.428	0.211	0.448	0.265
0.6	0.833	0.812	0.742	0.811	0.472	0.770	0.631	0.729	0.707	0.623	0.707	0.353	0.668	0.506
0.8	0.952	0.935	0.890	0.938	0.652	0.885	0.831	0.901	0.883	0.810	0.885	0.527	0.822	0.735
1	0.986	0.979	0.956	0.982	0.781	0.922	0.937	0.971	0.957	0.909	0.962	0.676	0.906	0.886
	$(n_1, n_2) = (40, 60)$													
0	0.099	0.094	0.087	0.095	0.120	0.175	0.106	0.051	0.043	0.040	0.044	0.060	0.122	0.055
0.2	0.247	0.232	0.188	0.235	0.182	0.278	0.171	0.157	0.139	0.102	0.142	0.115	0.211	0.098
0.4	0.548	0.534	0.458	0.532	0.325	0.524	0.355	0.420	0.397	0.314	0.402	0.223	0.419	0.241
0.6	0.798	0.778	0.701	0.780	0.482	0.734	0.601	0.696	0.665	0.567	0.670	0.367	0.645	0.468
0.8	0.934	0.918	0.858	0.919	0.626	0.864	0.803	0.873	0.846	0.759	0.852	0.517	0.796	0.696
1	0.984	0.975	0.939	0.979	0.751	0.910	0.924	0.962	0.943	0.875	0.949	0.649	0.885	0.859
	$(n_1, n_2) = (100, 100)$													
0	0.102	0.104	0.095	0.104	0.111	0.140	0.098	0.052	0.048	0.046	0.049	0.060	0.087	0.048
0.2	0.409	0.406	0.374	0.402	0.218	0.402	0.263	0.296	0.295	0.261	0.291	0.135	0.303	0.166
0.4	0.845	0.842	0.801	0.836	0.489	0.803	0.661	0.753	0.756	0.695	0.747	0.361	0.702	0.532
0.6	0.982	0.977	0.962	0.976	0.741	0.963	0.928	0.958	0.953	0.929	0.953	0.627	0.933	0.876
0.8	0.999	0.998	0.995	0.998	0.909	0.994	0.993	0.995	0.994	0.988	0.995	0.848	0.991	0.982
1	1.000	1.000	0.998	1.000	0.971	0.997	0.999	1.000	0.999	0.997	0.999	0.945	0.996	0.999

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Table 4: Powers at nominal level $\alpha = 0.1$ and 0.05 for Case II ($p = 3$).

δ	$\alpha = 0.1$						$\alpha = 0.05$					
	$(n_1, n_2) = (50, 50)$											
	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG
0	0.108	0.097	0.125	0.105	0.134	0.176	0.055	0.044	0.084	0.053	0.078	0.119
0.2	0.360	0.351	0.386	0.376	0.289	0.420	0.251	0.235	0.271	0.259	0.193	0.333
0.4	0.759	0.747	0.758	0.779	0.638	0.752	0.641	0.628	0.638	0.671	0.525	0.659
0.6	0.941	0.941	0.950	0.958	0.907	0.903	0.890	0.888	0.897	0.923	0.839	0.866
0.8	0.991	0.989	0.989	0.996	0.984	0.941	0.976	0.975	0.976	0.988	0.966	0.929
1	0.999	0.999	0.996	1.000	0.999	0.943	0.997	0.997	0.992	0.999	0.996	0.940
	$(n_1, n_2) = (40, 60)$											
0	0.111	0.091	0.130	0.104	0.158	0.183	0.059	0.043	0.087	0.047	0.095	0.125
0.2	0.360	0.363	0.415	0.380	0.260	0.406	0.256	0.254	0.295	0.269	0.163	0.317
0.4	0.740	0.735	0.778	0.769	0.600	0.739	0.638	0.623	0.667	0.662	0.467	0.663
0.6	0.938	0.941	0.960	0.957	0.875	0.901	0.887	0.887	0.912	0.916	0.797	0.853
0.8	0.990	0.992	0.991	0.997	0.974	0.937	0.976	0.978	0.981	0.990	0.948	0.924
1	0.997	0.998	0.992	1.000	0.996	0.938	0.994	0.996	0.988	0.999	0.991	0.937
	$(n_1, n_2) = (100, 100)$											
0	0.107	0.101	0.100	0.103	0.121	0.147	0.057	0.053	0.059	0.055	0.065	0.091
0.2	0.581	0.576	0.582	0.607	0.399	0.612	0.446	0.447	0.452	0.478	0.279	0.493
0.4	0.948	0.953	0.951	0.967	0.888	0.956	0.901	0.913	0.910	0.935	0.813	0.917
0.6	0.995	0.995	0.996	0.999	0.995	0.994	0.989	0.991	0.993	0.997	0.989	0.991
0.8	0.999	0.998	0.995	1.000	1.000	0.995	0.999	0.997	0.994	1.000	1.000	0.995
1	1.000	0.999	0.995	1.000	1.000	0.994	1.000	0.999	0.995	1.000	1.000	0.994

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Table 5: Powers at nominal level $\alpha = 0.1$ and 0.05 for Case I ($p = 10$).

δ	$\alpha = 0.1$							$\alpha = 0.05$						
	$(n_1, n_2) = (50, 50)$													
	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX	$U^{(1)}$	$U^{(1.5)}$	$U^{(2)}$	U	LXZ	BG	BOX
0	0.090	0.088	0.081	0.094	0.344	0.163	0.102	0.038	0.042	0.047	0.052	0.238	0.113	0.056
0.2	0.544	0.470	0.415	0.415	0.396	0.360	0.168	0.424	0.339	0.307	0.294	0.292	0.283	0.095
0.4	0.939	0.883	0.867	0.847	0.582	0.665	0.394	0.886	0.800	0.769	0.756	0.474	0.578	0.260
0.6	0.999	0.980	0.950	0.979	0.682	0.828	0.714	0.990	0.957	0.909	0.965	0.586	0.777	0.574
0.8	0.999	0.984	0.986	0.999	0.825	0.872	0.920	0.999	0.980	0.930	0.997	0.750	0.857	0.858
1	1.000	0.985	0.991	1.000	0.903	0.882	0.988	1.000	0.975	0.987	1.000	0.848	0.877	0.965
	$(n_1, n_2) = (40, 60)$													
0	0.080	0.086	0.090	0.091	0.389	0.165	0.097	0.051	0.037	0.042	0.046	0.293	0.117	0.047
0.2	0.511	0.410	0.368	0.400	0.457	0.358	0.150	0.369	0.357	0.333	0.271	0.356	0.277	0.092
0.4	0.917	0.840	0.797	0.821	0.622	0.657	0.364	0.857	0.858	0.828	0.736	0.520	0.564	0.220
0.6	0.996	0.976	0.954	0.974	0.742	0.828	0.662	0.987	0.989	0.971	0.950	0.660	0.771	0.524
0.8	0.999	0.995	0.975	0.998	0.840	0.878	0.894	0.999	0.999	0.972	0.996	0.768	0.860	0.816
1	0.997	0.994	0.998	1.000	0.900	0.892	0.977	0.999	0.997	0.987	1.000	0.860	0.884	0.950
	$(n_1, n_2) = (100, 100)$													
0	0.102	0.094	0.090	0.107	0.169	0.151	0.096	0.050	0.054	0.043	0.054	0.096	0.095	0.046
0.2	0.781	0.735	0.641	0.647	0.294	0.578	0.280	0.680	0.624	0.593	0.518	0.186	0.467	0.172
0.4	0.999	0.991	0.916	0.982	0.526	0.927	0.784	0.995	0.972	0.867	0.960	0.405	0.877	0.663
0.6	1.000	0.994	0.988	1.000	0.808	0.972	0.988	1.000	0.990	0.973	1.000	0.705	0.969	0.971
0.8	1.000	0.988	0.998	1.000	0.938	0.972	1.000	1.000	0.984	0.994	1.000	0.888	0.972	1.000
1	1.000	0.991	0.998	1.000	0.983	0.972	1.000	1.000	0.985	0.996	1.000	0.964	0.972	1.000

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

$1139, 1142, 1169, 1185, 1557, 1584, 1620, 1647, 3439\}$ and $D_3 = \{332, 377, 905, 1082, 1113, 1142, 1620\}$. We see that all the genes selected by Method 2 are contained in D_2. So, we consider the genes belonging to D_2 in the following. The number of genes in D_2 and the corresponding observed statistic of each gene are illustrated in Table 6.

In order to consolidate our method, boxplots of genes indexed by $D_2 \backslash D_1$ are also given in Figure 1. From Figure 1, one can clearly see that these genes of two groups have different distributions.

In addition, in order to make the hypothesis more powerful, we combine our method with false discovery rate (FDR) control (Benjamini and Hochberg (1995)) at nominal level $\alpha = 0.1$. The procedure is conducted as follows.

- Sort the corresponding P-values in ascending order, denoted as $P_{(1)}, \ldots, P_{(p)}$.
- Let k denote the largest index such that $P_{(i)} \leq \alpha \times i/p$ for all $i \leq k$.
- Declare all tests with P-values $P_{(1)}, \ldots, P_{(k)}$ significant.

By this procedure, we can find that all genes in $D_1 \cup D_2$ can be recruited.

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

Table 6: Observed statistics for genes in D_2, $t_0 = 1$

No.gene	value $\omega^{-1}l$	No.gene	value $\omega^{-1}l$
2	89.51	1142	27.95
332	68.84	1169	25.69
377	20.39	1185	19.02
398	19.55	1557	19.63
905	24.64	1584	20.85
1082	23.98	1620	28.75
1113	28.58	1647	23.84
1139	20.41	3439	26.26

5. Conclusion

In this paper, we investigate the problem of testing the equality of two distributions by using JEL approach. We prove that the scaled log JEL statistic tends to a chi square distribution with degree of freedom 1, which is easy to be implemented in practice. Extensive simulation studies confirm that our proposed method performs much better than its competitors. It is interesting to derive the limiting distribution of the scaled log JEL statistic when p tends to infinity.

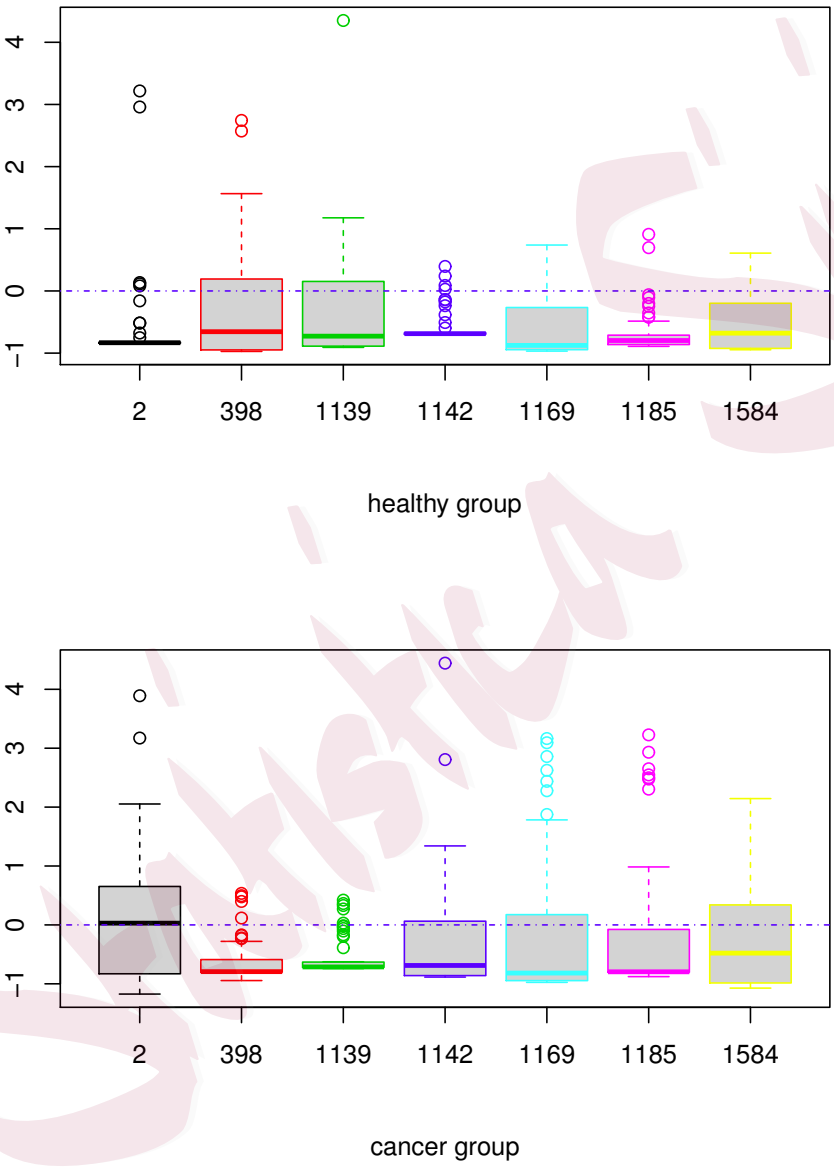

Figure 1: Expression levels for genes in $D_2 \backslash D_1$.

Appendix

In this appendix, we present the proof of Theorem 2.4. We begin with the following notations:

$$
\begin{aligned}
Q_{1n}(\theta,\lambda_1,\lambda_2,\lambda_3) &= \frac{1}{N}\sum_{i=1}^{n_1}\frac{V_i^{(1)}-\theta}{1+\lambda_1(V_i^{(1)}-\theta)},\\
Q_{2n}(\theta,\lambda_1,\lambda_2,\lambda_3) &= \frac{1}{N}\sum_{i=1}^{n_2}\frac{V_i^{(2)}-\theta}{1+\lambda_2(V_i^{(2)}-\theta)},\\
Q_{3n}(\theta,\lambda_1,\lambda_2,\lambda_3) &= \frac{1}{N}\sum_{i=1}^{n_3}\frac{V_i^{(3)}-EV_i^{(3)}}{1+\lambda_3(V_i^{(3)}-EV_i^{(3)})},\\
Q_{4n}(\theta,\lambda_1,\lambda_2,\lambda_3) &= \frac{1}{N}\Big[\lambda_1\sum_{i=1}^{n_1}\frac{-1}{1+\lambda_1(V_i^{(1)}-\theta)}+\lambda_2\sum_{i=1}^{n_2}\frac{-1}{1+\lambda_2(V_i^{(2)}-\theta)}\\
&\quad +\lambda_3\sum_{i=1}^{n_1}\frac{-\frac{n}{n-2}\cdot\frac{n_2-1}{n_1}}{1+\lambda_3(V_i^{(3)}-EV_i^{(3)})}\\
&\quad +\lambda_3\sum_{i=n_1+1}^{n_3}\frac{-\frac{n}{n-2}\cdot\frac{n_1-1}{n_2}}{1+\lambda_3(V_i^{(3)}-EV_i^{(3)})}\Big],
\end{aligned}
$$

where $N=2n=n_1+n_2+n$. Also, we define

$$g_1(x)=E[h(x,\mathbf{X}_2)]-\theta_0,\ \sigma_{g_1}^2=var(g_1(\mathbf{X}_1)),\ g_2(y)=E[h(y,\mathbf{Y}_2)]-\theta_0,$$

$$\sigma_{g_2}^2=\mathrm{Var}[g_1(\mathbf{Y}_1)],\ g_{1,0}(x)=E[h(x,\mathbf{Y}_1)]-\theta_0,\ g_{0,1}(y)=E[h(\mathbf{X}_1,y)]-\theta_0$$

$$\sigma_{1,0}^2=\mathrm{Var}[g_{1,0}(\mathbf{X}_1)],\ \sigma_{0,1}^2=\mathrm{Var}[g_{0,1}(\mathbf{Y}_1)].$$

Under H_0, $g_1(x)=g_2(x)=g_{1,0}(x)=g_{0,1}(x)$. $\to^D$ denotes the weak convergence.

Lemma 5.1. (Hoeffding (1948)) If $E[h^2(\mathbf{X}_1,\mathbf{X}_2)]<\infty$, $E[h^2(\mathbf{Y}_1,\mathbf{Y}_2)]<\infty$, $\sigma_{g_1}>0$ and $\sigma_{g_2}>0$, then

(a) $\sqrt{n_1}(U_1 - \theta_0)/(2\sigma_{g_1}) \to^D \mathbf{N}(0,1)$ as $n_1 \to \infty$;

(b) $\sqrt{n_2}(U_2 - \theta_0)/(2\sigma_{g_2}) \to^D \mathbf{N}(0,1)$ as $n_2 \to \infty$.

Lemma 5.2. (Arvesen (1969)) If $E[h^2(\mathbf{X}, \mathbf{Y})] < \infty$, $\sigma^2_{1,0} > 0$ and $\sigma^2_{0,1} > 0$, let $S^2_{n_1,n_2} = \sigma^2_{1,0}/n_1 + \sigma^2_{0,1}/n_2$, then, as $\min(n_1, n_2) \to \infty$, we have

$$\frac{U_3 - \theta_0}{S_{n_1,n_2}} \to^D \mathbf{N}(0,1).$$

Remark 5.3. To simplify our notation, we write

$$\sigma_1^2 = 4\sigma^2_{g_1}, \quad \sigma_2^2 = 4\sigma^2_{g_2}, \quad \sigma_3^2 = \sigma^2_{g_1}/(2\gamma_1) + \sigma^2_{g_2}/(2\gamma_2).$$

Under the null hypothesis, we denote $\sigma := \sigma_1 = \sigma_2$.

Lemma 5.4. (Jing et al. (2009)) Let $S_1 = n_1^{-1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0)^2$, $S_2 = n_2^{-1}\sum_{i=1}^{n_2}(V_i^{(2)} - \theta_0)^2$ and $S_3 = n^{-1}\sum_{i=1}^{n}(V_i^{(3)} - E(V^{(3)}))^2$. Under the conditions of Lemmas 5.1 and 5.2,

(a) $S_1 = \sigma_1^2 + o_P(1)$ as $n_1 \to \infty$,

(b) $S_2 = \sigma_2^2 + o_P(1)$ as $n_2 \to \infty$,

(c) $S_3 = \sigma_3^2 + o_P(1)$ as $\liminf_{n\to\infty} \min(n_1, n_2)/\max(n_1, n_2) > 0$.

Lemma 5.5. Under **C1-C3** and H_0, with probability tending to one as $\min(n_1, n_2) \to \infty$, there exists a root $\tilde{\theta}$ of (2.13)-(2.16), such that $|\tilde{\theta} - \theta_0| < \delta$, where $\delta = n^{-1/3}$.

Proof. For $\theta \in \{\theta : |\theta - \theta_0| < \delta\}$, $E(V_i^{(s)} - \theta) = \theta_0 - \theta$ for $s = 1, 2$, and $E(V_i^{(3)} - E(V_i^{(3)}|_\theta)) = \frac{n(\theta_0-\theta)}{n-2}\Big[\frac{n_2-1}{n_1} I(0 \le i \le n_1) + \frac{n_1-1}{n_2} I(n_1+1 \le i \le n)\Big]$, (here $E(V_i^{(3)}|_\theta)$ means the expectation of $V_i^{(3)}$ when $E[h(\mathbf{X}_1, \mathbf{X}_2)] = \theta$.)

$$\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta) = U_s - \theta_0 - (\theta - \theta_0) = O_P(\delta + n_s^{-1/2}),$$
$$\frac{1}{n}\sum_{i=1}^{n}\big(V_i^{(3)} - E(V_i^{(3)}|_\theta)\big) = U_3 - \theta_0 - (\theta - \theta_0) = O_P(\delta + n^{-1/2}).$$

Similarly, we have

$$\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta)^2 = S_s + O_P(\delta + n^{-1/2}) = \sigma^2 + O_P(\delta + n^{-1/2}) + o_P(1),$$
$$\frac{1}{n}\sum_{i=1}^{n}\big(V_i^{(3)} - E(V_i^{(3)}|_\theta)\big)^2 = S_3 + O_P(\delta + n^{-1/2}) + o_P(1).$$

In particular,

$$\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta_0 - \delta) = -\delta + O_P(n_s^{-1/2}),$$
$$\frac{1}{n}\sum_{i=1}^{n}\big(V_i^{(3)} - E(V_i^{(3)}|_{\theta_0+\delta})\big) = -\delta + O_P(n^{-1/2}),$$
$$\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta_0 - \delta)^2 = \sigma^2 + \delta^2 + O_p(n_s^{-1/2}) + o_P(1),$$
$$\frac{1}{n}\sum_{i=1}^{n}\big(V_i^{(3)} - E(V_i^{(3)}|_{\theta_0+\delta})\big)^2 = S_3 + \delta^2 + O_p(n^{-1/2}) + o_P(1).$$

Expanding (2.14)-(2.16), we have

$$\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta) - \lambda_s \frac{1}{n_s}\sum_{i=1}^{n_s}\frac{(V_i^{(s)} - \theta)^2}{1 + \lambda_s(V_i^{(s)} - \theta)} = 0, \tag{5.1}$$

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

$$\frac{1}{n}\sum_{i=1}^{n}(V_i^{(3)} - E(V_i^{(3)}|_\theta)) - \lambda_3\frac{1}{n}\sum_{i=1}^{n}\frac{(V_i^{(3)} - E(V_i^{(3)}|_\theta))^2}{1+\lambda_s(V_i^{(3)} - E(V_i^{(3)}|_\theta))} = 0,$$

and define $Z_n^{*(s)} = \max_{1\le i\le n_s}|V_i^{(s)} - \theta|$ for $s = 1, 2$, and

$$Z_n^{*(3)} = \max_{1\le i\le n}|V_i^{(3)} - E(V_i^{(3)}|_\theta)|.$$

Since the kernel function h is bounded, there exists a constant c such that $0 < Z_n^{*(s)} < c$ for $s = 1, 2, 3$. Thus, we have

$$\lambda_s|_\theta = O_P(\delta + n_s^{-1/2}) \tag{5.2}$$

uniformly in $\theta \in \{\theta : |\theta - \theta_0| < \delta\}$ and $\lambda_s|_{\theta_0+\delta} = O_P(\delta + n_s^{-1/2})$. Similarly, $\lambda_s|_{\theta_0-\delta} = O_P(\delta + n_s^{-1/2})$.

For simplicity, we write

$$\begin{aligned}
\Delta_{\delta 1} &= \delta + n_1^{-1/2},\\
H_1(\theta) &= \frac{1}{n_1}\sum_{i=1}^{n_1}\log[1+\lambda_1(V_i^{(1)} - \theta)],\\
H_2(\theta) &= \frac{1}{n_2}\sum_{i=1}^{n_2}\log[1+\lambda_2(V_i^{(2)} - \theta)],\\
H_3(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\log[1+\lambda_3(V_i^{(3)} - E(V_i^{(3)}|_\theta))],\\
H(\theta) &= n_1H_1(\theta) + n_2H_2(\theta) + nH_3(\theta).
\end{aligned}$$

We consider H_1 first. By the Taylor expansion and the order of λ_1,

$$\begin{aligned}
H_1(\theta_0+\delta) &= \lambda_1|_{\theta_0+\delta}\frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0 - \delta) - \frac{1}{2}\lambda_1|^2_{\theta_0+\delta}\frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0 - \delta)^2\\
&\quad + O_P(\Delta_{\delta 1}^3).
\end{aligned}$$

By $1/(x+1) = 1 - x + x^2/(1+x)$ and (5.1), we also have

$$\begin{aligned} 0 &= \frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0 - \delta) - \lambda_1|_{\theta_0+\delta}\frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0 - \delta)^2 \\ &\quad +\lambda_1^2|_{\theta_0+\delta}\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{(V_i^{(1)} - \theta_0 - \delta)^3}{1 + \lambda_1|_{\theta_0+\delta}(V_i^{(1)} - \theta_0 - \delta)}. \end{aligned}$$

Combining this with the order of $\lambda_1|_{\theta_0+\delta}$ yields

$$\lambda_1|_{\theta_0+\delta} = \frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \theta_0 - \delta)\sigma_\delta^{-2} + O_P(\Delta_{\delta 1}^2).$$

where $\sigma_\delta^2 = \sigma^2 + \delta^2 + O_P(n_s^{-1/2}) + o_P(1)$. Substituting $\lambda_1|_{\theta_0+\delta}$ into $H_1(\theta_0+\delta)$,

$$\begin{aligned} H_1(\theta_0 + \delta) &= \frac{1}{2}\sigma_\delta^{-2}\big[\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta_0 - \delta)\big]^2 + O_P(\Delta_{\delta 1}^3) \\ &= \frac{1}{2\sigma_\delta^2}\big[\delta^2 + O_p(n_s^{-1})\big] + O_P(\Delta_{\delta 1}^3). \end{aligned}$$

Letting $\delta = 0$, we also have

$$\begin{aligned} H_1(\theta_0) &= \frac{1}{2}\sigma_\delta^{-2}\big[\frac{1}{n_s}\sum_{i=1}^{n_s}(V_i^{(s)} - \theta_0)\big]^2 + O_P(\Delta_{\delta 1}^3) \\ &= \frac{1}{2\sigma_\delta^2}\big[O_p(n_s^{-1/2})\big]^2 + O_P(\Delta_{\delta 1}^3). \end{aligned}$$

Thus, we have that $H_1(\theta_0 + \delta) \geq H_1(\theta_0)$ with probability tending to one. Similarly, we are able to show that $H_1(\theta_0 - \delta) \geq H_1(\theta_0)$ with probability tending to one, and so do H_2 and H_3. So, $H(\theta_0 \pm \delta) \geq H(\theta_0)$ with probability tending to one. By the continuity of H in $\{\theta : |\theta - \theta_0| < \delta\}$, with probability tending to one $H(\theta)$ achieves its minimum in $\{\theta : |\theta - \theta_0| < \delta\}$, denoted as $\tilde{\theta}$, which is given by the root of (2.13)-(2.16). □

Now, we assume $\tilde{\beta} = (\tilde{\theta}, \tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{\lambda}_3)^\top$ is the solution to $Q_{in}(\theta, \lambda_1, \lambda_2, \lambda_3) = 0$, for $i = 1, 2, 3, 4$. Thus, we expand them at $\beta_0 = (\theta_0, 0, 0, 0)^\top$,

$$\begin{aligned} 0 &= Q_{in}(\beta_0) + \frac{\partial Q_{in}}{\partial \theta}(\beta_0)(\tilde{\theta} - \theta_0) + \frac{\partial Q_{in}}{\partial \lambda_1}(\beta_0)\tilde{\lambda}_1 + \frac{\partial Q_{in}}{\partial \lambda_2}(\beta_0)\tilde{\lambda}_2 \\ &\quad + \frac{\partial Q_{in}}{\partial \lambda_3}(\beta_0)\tilde{\lambda}_3 + R_{in}, \end{aligned} \tag{5.3}$$

where $R_{in} = \frac{1}{2}(\tilde{\beta} - \beta_0)^\top \frac{\partial^2 Q_{in}(\beta^*)}{\partial \beta \partial \beta^\top}(\tilde{\beta} - \beta_0)$, and one can check that $R_{in} = o_P(n^{-1/2})$ by Lemma 5.5 and (5.2).

Lemma 5.6. Under H_0, $Cov(U_1, U_2) = 0$, $Cov(U_1, U_3) = 2n_1^{-1}\sigma_{g_1}^2$ and $Cov(U_2, U_3) = 2n_2^{-1}\sigma_{g_1}^2$.

Proof. This follows from the Hoeffding decomposition of U-statistics. □

Proof of Theorem 2.4. By Lemma 5.4 and (5.3),

$$\begin{pmatrix} Q_{1n}(\beta_0) \\ Q_{2n}(\beta_0) \\ Q_{3n}(\beta_0) \\ 0 \end{pmatrix} = W \begin{pmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \\ \tilde{\lambda}_3 \\ \tilde{\theta} - \theta_0 \end{pmatrix} + o_P(n^{-1/2}),$$

where

$$W = \begin{pmatrix} \gamma_1\sigma_1^2 & 0 & 0 & \gamma_1 \\ 0 & \gamma_2\sigma_2^2 & 0 & \gamma_2 \\ 0 & 0 & \frac{1}{2}\sigma_3^2 & \frac{1}{2} \\ \gamma_1 & \gamma_2 & \frac{1}{2} & 0 \end{pmatrix}.$$

Thus, we can get

$$\begin{pmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \\ \tilde{\lambda}_3 \\ \tilde{\theta} - \theta_0 \end{pmatrix} = W^{-1} \begin{pmatrix} Q_{1n}(\beta_0) \\ Q_{2n}(\beta_0) \\ Q_{3n}(\beta_0) \\ 0 \end{pmatrix} + o_P(n^{-1/2}).$$

In particular

$$\tilde{\theta} - \theta_0 = d_1 Q_{1n}(\beta_0) + d_2 Q_{2n}(\beta_0) + d_3 Q_{3n}(\beta_0) + o_P(n^{-1/2}), \quad (5.4)$$

where

$$d_1 = d_2 = \frac{2}{1 + 16\gamma_1\gamma_2},\ d_3 = \frac{32\gamma_1\gamma_2}{1 + 16\gamma_1\gamma_2}.$$

From (2.14), we have

$$\frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \tilde{\theta}) - \tilde{\lambda}_1\frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \tilde{\theta}) + \tilde{\lambda}_1^2\frac{1}{n_1}\sum_{i=1}^{n_1}\frac{(V_i^{(1)} - \tilde{\theta})^3}{1 + \tilde{\lambda}_1(V_i^{(1)} - \tilde{\theta})} = 0.$$

This implies that

$$\tilde{\lambda}_1 = \frac{U_1 - \tilde{\theta}}{\tilde{S}_1} + o_P(n^{-1/2}),$$

where $\tilde{S}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}(V_i^{(1)} - \tilde{\theta})^2$. Similarly,

$$\tilde{\lambda}_2 = \frac{U_2 - \tilde{\theta}}{\tilde{S}_2} + o_P(n^{-1/2}),\ \tilde{\lambda}_3 = \frac{U_3 - \tilde{\theta}}{\tilde{S}_3} + o_P(n^{-1/2}),$$

where $\tilde{S}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}(V_i^{(2)} - \tilde{\theta})^2$ and $\tilde{S}_3 = \frac{1}{n}\sum_{i=1}^{n}(V_i^{(3)} - E(\tilde{V}_i^{(3)}))^2$ with

$$E(\tilde{V}_i^{(3)}) = \frac{n\tilde{\theta}}{n-2}\Big[\frac{n_2 - 1}{n_1}I(0 \le i \le n_1) + \frac{n_1 - 1}{n_2}I(n_1 + 1 \le i \le n)\Big].$$

It is easy to check $\tilde{S}_j = \sigma_j^2 + o_P(1)$ via Lemmas 5.4 and 5.5, $j = 1, 2, 3$. Now we plug the above formulas of $\tilde{\lambda}_j$ into the expansion of $l(\theta_0)$ to get

$$l(\theta_0) = \Big[n_1\frac{(U_1 - \tilde{\theta})^2}{\sigma_1^2} + n_2\frac{(U_2 - \tilde{\theta})^2}{\sigma_2^2} + n\frac{(U_3 - \tilde{\theta})^2}{\sigma_3^2}\Big]\big(1 + o_P(1)\big).$$

By (5.4), we have

$$\begin{aligned} & n_1\frac{(U_1 - \tilde{\theta})^2}{\sigma_1^2} + n_2\frac{(U_2 - \tilde{\theta})^2}{\sigma_2^2} + n\frac{(U_3 - \tilde{\theta})^2}{\sigma_3^2} \\ = \ & \Big(\sqrt{N}Q_{1n}(\beta_0), \sqrt{N}Q_{3n}(\beta_0), \sqrt{N}Q_{2n}(\beta_0)\Big) \\ & \times\mathcal{A}^\top\mathcal{D}\mathcal{A}\Big(\sqrt{N}Q_{1n}(\beta_0), \sqrt{N}Q_{2n}(\beta_0), \sqrt{N}Q_{3n}(\beta_0)\Big)^\top + o_p(1), \end{aligned}$$

where $\mathcal{D} = \begin{pmatrix} \gamma_1/\sigma_1^2 & 0 & 0 \\ 0 & \gamma_2/\sigma_2^2 & 0 \\ 0 & 0 & \gamma_3/\sigma_3^2 \end{pmatrix}$.

By Lemmas 5.1, 5.2 and 5.6, we have

$$\sqrt{N}\begin{pmatrix} Q_{1n}(\beta_0) \\ Q_{2n}(\beta_0) \\ Q_{3n}(\beta_0) \end{pmatrix} \to^D \mathbf{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \gamma_1\sigma_1^2 & 0 & \sigma_{g_1}^2 \\ 0 & \gamma_2\sigma_2^2 & \sigma_{g_2}^2 \\ \sigma_{g_1}^2 & \sigma_{g_2}^2 & \frac{1}{2}\sigma_3^2 \end{pmatrix}\right), \quad (5.5)$$

which is a degenerate three dimensional normal distribution. Note that under H_0, $\sigma_{g_1}^2 = \sigma_{g_2}^2 = \sigma^2/4$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, $\sigma_3^2 = \sigma^2/(16\gamma_1\gamma_2)$. Hence under H_0 the empirical log likelihood ratio l converges in distribution to $\sum_{i=1}^3 \omega_i \chi_i^2$, where $\chi_i^2, i = 1, 2, 3$ are three independent chi-square random variables with one degree of freedom, $\omega_i, i = 1, 2, 3$ are three eigenvalues of $\Omega^{1/2}\mathcal{A}^\top \mathcal{D}_0 \mathcal{A}\Omega^{1/2}$ with $\Omega = \begin{pmatrix} \gamma_1 & 0 & 1/4 \\ 0 & \gamma_2 & 1/4 \\ 1/4 & 1/4 & 1/(32\gamma_1\gamma_2) \end{pmatrix}$,

$$\mathcal{A} = \begin{pmatrix} \gamma_1^{-1} - d_1 & -d_2 & -d_3 \\ -d_1 & \gamma_2^{-1} - d_2 & -d_3 \\ -d_1 & -d_2 & \gamma_3^{-1} - d_3 \end{pmatrix}, \ \mathcal{D}_0 = \begin{pmatrix} \gamma_1 & 0 & 0 \\ 0 & \gamma_2 & 0 \\ 0 & 0 & \gamma_3 \end{pmatrix}.$$

By an elementary but tedious calculation one can check that both Ω and $\mathcal{A}$ have one zero eigenvalue with eigenvectors $(\gamma_1^{-1}, \gamma_2^{-1}, -4)^\top$ and $(2\gamma_1, 2\gamma_2, 1)^\top$ respectively. These two eigenvectors are orthogonal. So $\Omega^{1/2}\mathcal{A}^\top \mathcal{D}_0 \mathcal{A}\Omega^{1/2}$ has only one nonzero eigenvalue, which is

$$\text{trace}(\Omega^{1/2}\mathcal{A}^\top \mathcal{D}_0 \mathcal{A}\Omega^{1/2}) = \frac{1}{16\gamma_1\gamma_2} - \frac{2}{1 + 16\gamma_1\gamma_2} + \frac{64\gamma_1\gamma_2}{(1 + 16\gamma_1\gamma_2)^2}.$$

Now we can complete the proof. □

Acknowledgements

The authors would like to thank the Editor, the Associate Editor, and

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

two anonymous referees for their extensive and constructive suggestions that helped to improve the paper considerably. Zhi Liu's research is supported by FDCT of Macau No.127/2016/A3. Wang Zhou's research is partially supported by the MOE Tier 2 grant MOE2015-T2-2-039 (R-155-000-171-112) at the National University of Singapore.

References

Adimari G. and Chiogna M. (2012). Jackknife empirical likelihood based confidence intervals for partial areas under ROC curves. *Statistica Sinica* **22**, 1457–1477.

Arvesen J.N. (1969). Jackknife U-staitstics. *The Annals of Mathematical Statistics* **40**, 2076–2100.

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

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society: Series B* **57**, 289–300.

Bentkus V., Jing B.-Y. and Zhou W. (2009). On normal approximations to U-statistics. *Annals of Probability* **37,** 2174–2199.

Biswas M. and Ghosh A.K. (2014). A nonparametric two-sample test applicable to high dimensional data. *Journal of Multivariate Analysis* **123**, 160–171.

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

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**, 808–835.

Fernandez V. A., Gamero M.D.J and Garcia J.M. (2008). A test for the two-sample problem based on empirical characteristic functions. *Computational Statistics and Data Analysis* **52**, 3730–3748.

Gong Y., Peng L. and Qi Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *Journal of Multivariate Analysis* **101**, 1520–1531.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19**, 293–325.

Jing B.Y., Yuan J.Q. and Zhou W. (2009). Jackknife empirical likelihood. *Journal of the American Statistical Association* **104**, 1224–1232.

Kellermeier J. (1980). The Empirical Characteristic Function and Large Sample Hypothesis Testing. *Journal of Multivariate Analysis* **10**, 78–87

Li J. and Chen S.X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics* **40**, 908–940.

Liu Z., Xia X.C. and Zhou W. (2015). A test for equality of two distributions via jackknife empirical likelihood and characteristic functions. *Computational Statistics and Data Analysis* **92**, 97–114.

Liu Z.Y. and Modarres R. (2011). A triangle test for equality of distribution functions

Statistica Sinica: Newly accepted Paper
(accepted version subject to English editing)

in high dimensions. *Journal of Nonparametric Statistics* **23**, 605–615.

Lee A.J. (1990). *U-statistics: theory and practice*. Marcel Dekker, New York.

Maa J. F., Pearl D. K. and Bartoskzyński R. (1996). Reducing multidimensional two-sample data to one-dimensional interpoint comparisons. *The Annals of Statistics* **24**, 1069–1074.

Owen A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Quenouille M. H. (1956). Notes on Bias in Estimation. *Biometrika* **43**, 353–360.

Wang R.D., Peng L. and Qi Y. (2013). Jackknife empirical likelihood test for equality of two high dimensional means. *Statistica Sinica* **23**, 667–690.

Zhao Y., Meng X. and Yang H. (2015). Jackknife empirical likelihood inference for the mean absolute deviation. *Computational Statistics and Data Analysis* **91**, 92–101.

School of Physical and Mathematical Sciences, Nanyang Technological University.
E-mail: liuy0135@e.ntu.edu.sg

Department of Mathematics, University of Macau.
E-mail: liuzhi@umac.mo
Tel: +853-88224494
Fax: +853-88222426

Department of Statistics and Applied Probability, National University of Singapore.
E-mail: stazw@nus.edu.sg