应用概率统计 第 34 卷
第 5 期 2018 年 10 月

Chinese Journal of Applied Probability and Statistics
Oct., 2018, Vol. 34, No. 5, pp. 463-474
doi: 10.3969/j.issn.1001-4268.2018.05.003

Two Notes on Energy Distance *

CHEN Minqiong
(*Xinhua College of Sun Yat-Sen University, Guangzhou, 510520, China*)

Abstract: The definition of Brownian distance is presented and it's proved that Brownian distance coincides with the energy distance with respect to Brownian motion. Energy distance for dependent random vectors is also given and the asymptotic distribution is derived under null hypothesis. A simple numerical simulation result shows that the method for paired-sample test based on energy distance can distinguish the distributions of the paired variables more effectively than the classical t-test and Wilcoxon signed rank test.

Keywords: energy distance; Brownian distance; paired-sample test; asymptotic distribution; numerical simulation

2010 Mathematics Subject Classification: 62G99

Citation: CHEN M Q. Two notes on energy distance [J]. Chinese J Appl Probab Statist, 2018, 34(5): 463–474.

§1. Introduction: Energy Distance

Székely [1] introduced a new concept named energy distance to measure the difference between two independent probability distributions. If X and Y are independent random vectors in $\mathbb{R}^p$ with cumulative distribution functions (cdf) F and G respectively, then the energy distance between the distributions F and G is defined as

$$\varepsilon(F,G) = 2\mathsf{E}\|X-Y\| - \mathsf{E}\|X-X'\| - \mathsf{E}\|Y-Y'\|, \tag{1}$$

where X' is an i.i.d. copy of X, and Y' is an iid copy of Y. E is the expected value, and $\|\cdot\|$ denotes the Euclidean norm. One can also write $\varepsilon(F,G)$ as $\varepsilon(X,Y)$, and call it be the energy distance of X and Y. Székely [1] proved that for real-valued random variables this distance is exactly twice Harald Cramér's distance, that is

$$2\int_{-\infty}^{\infty}[F(t)-G(t)]^2\mathrm{d}t = 2\mathsf{E}\|X-Y\| - \mathsf{E}\|X-X'\| - \mathsf{E}\|Y-Y'\|.$$

*The project was supported by the Natural Science Foundation of Guangdong Province for Young Innovative Talents (Grant No. 2014KQNCX253).

E-mail: C_skirt@163.com.

Received December 13, 2016. Revised August 30, 2017.

In higher dimensions, however, the two distances are different because the energy distance is rotation invariant while Cramér's distance is not. The equality becomes

$$2\mathsf{E}\|X-Y\|-\mathsf{E}\|X-X'\|-\mathsf{E}\|Y-Y'\|=\frac{1}{c_p}\int_{\mathbb{R}^p}\frac{|\widehat{f}(t)-\widehat{g}(t)|^2}{\|t\|^{p+1}}\mathrm{d}t, \tag{2}$$

where $\widehat{f}$ is X's characteristic function and $\widehat{g}$ be Y's characteristic function, $c_p=\pi^{(p+1)/2}/\Gamma[(p+1)/2]$. Thus $\varepsilon(F,G)\geqslant 0$ with equality to zero if and only if $F=G$. This property makes it possible to use $\varepsilon(F,G)$ for testing goodness-of-fit, homogeneity, etc. in a consistent way. Rizzo [2] discussed the theoretical background of homogeneity test for two multivariate populations based on energy distance. He considered the test

$$H_0: F_1=F_2$$

versus alternative $F_1\neq F_2$. Suppose $\boldsymbol{X}=\{X_1,X_2,\ldots,X_m\}$ be an i.i.d. sample from the distribution of F_1 in $\mathbb{R}^p$, and $\boldsymbol{Y}=\{Y_1,Y_2,\ldots,Y_n\}$ be an i.i.d. sample from the distribution of F_2 in $\mathbb{R}^p$. The sample version of $\varepsilon(F_1,F_2)$ was defined as:

$$V_{m,n}:=\frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\|X_i-Y_j\|-\frac{1}{m^2}\sum_{k,l=1}^{m}\|X_k-X_l\|-\frac{1}{n^2}\sum_{u,v=1}^{n}\|Y_u-Y_v\|. \tag{3}$$

If one denotes

$$h(x_1,x_2;y_1,y_2):=\|x_1-y_1\|+\|x_2-y_2\|-\|x_1-x_2\|-\|y_1-y_2\|,$$

then, obviously $V_{m,n}$ is a two-sample V-statistic with kernel h

$$V_{m,n}=\frac{1}{m^2n^2}\sum_{i,k=1}^{m}\sum_{j,l=1}^{n}h(X_i,X_k;Y_j,Y_l).$$

Under the hypothesis of $F_1=F_2$, $V_{m,n}$ is first-order degenerated, which leads $[mn/(m+n)]V_{m,n}$ to converges in distribution to a quadratic form of centered Gaussian random variables [2]. So one can choose $\varepsilon_{m,n}:=[mn/(m+n)]V_{m,n}$ as the test statistic for H_0, and reject H_0 for large values of $\varepsilon_{m,n}$. Rizzo [2] also presented the procedure for practical implementation via bootstrap method.

For more applications of energy distance on the classical statistic problems such as multivariate normality test, hierarchical clustering, dependence test and extension of analysis of variance, etc., see [3–6]. Székely and Rizzo summarized all those applications in [7]. Contrary to the classical methods, these methods based on energy distance are simple to calculate, applicable for more widely distributed types of data and can deal with multi-variables.

Rizzo[2] tells us that one can distinguish any two probability distributions on Euclidean space in a very simple way, just by computing the difference of the "average between sample distances" and "average within sample distances" of the two samples $\boldsymbol{X} = \{X_1, X_2, \ldots, X_m\}$ and $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$. Here the distance only refers to the Euclidean distance. Lyons[8] generalized the notion of energy distance to probability distributions on metric spaces. Let (M, d) be a metric space with its Borel sigma algebra $\mathscr{B}(M)$, $\mathscr{P}(M)$ denotes the collection of all probability measures on the measurable space $(M, \mathscr{B}(M))$. If μ and ν are probability measures in $\mathscr{P}(M)$, then the energy distance of μ and ν can be defined as

$$\varepsilon(\mu, \nu) = 2\mathsf{E}[d(X, Y)] - \mathsf{E}[d(X, X')] - \mathsf{E}[d(Y, Y')],$$

provided that these expectations exist, where $X, X' \overset{\text{i.i.d.}}{\sim} \mu$ and $Y, Y' \overset{\text{i.i.d.}}{\sim} \nu$. However the result "$\varepsilon(\mu, \nu) \geqslant 0$ with equality to zero if and only if $\mu = \nu$" does not always hold. Lyons[8] indicated that it holds unless the metric space (M, d) is strong negative. We describe this point more clearly as following proposition.

Proposition 1 If (M, d) is a metric space of negative type, then, one has $\varepsilon(\mu, \nu) \geqslant 0$. Moreover, if (M, d) is strong negative, then $\varepsilon(\mu, \nu) = 0$ if and only if $\mu = \nu$.

We give a explicit proof of this proposition.

Proof If (M, d) is a metric space of negative type, then there exists a Hilbert space H and a map $\phi : M \to H$ such that $\forall\, x, x' \in M$, $d(x, x') = \|\phi(x) - \phi(x')\|^2$ holds[10], so

$$\begin{aligned}
\mathsf{E}d(X, Y) &= \mathsf{E}\|\phi(X) - \phi(Y)\|^2 \\
&= \mathsf{E}\|\phi(X) - \mathsf{E}[\phi(X)] + \mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)] + \mathsf{E}[\phi(Y)] - \phi(Y)\|^2 \\
&= \mathsf{E}\{\|\phi(X) - \mathsf{E}[\phi(X)]\|^2 + \|\mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)]\|^2 + \|\phi(Y) - \mathsf{E}[\phi(Y)]\|^2 \\
&\quad + 2\langle \phi(X) - \mathsf{E}[\phi(X)], \mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)]\rangle \\
&\quad + 2\langle \phi(X) - \mathsf{E}[\phi(X)], \mathsf{E}[\phi(Y)] - \phi(Y)\rangle \\
&\quad + 2\langle \mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)], \mathsf{E}[\phi(Y)] - \phi(Y)\rangle\} \\
&= \mathsf{E}\|\phi(X) - \mathsf{E}[\phi(X)]\|^2 + \mathsf{E}\|\mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)]\|^2 + \mathsf{E}\|\phi(Y) - \mathsf{E}[\phi(Y)]\|^2,
\end{aligned}$$

and obviously

$$\begin{aligned}
\mathsf{E}\|\phi(X) - \phi(X')\|^2 &= 2\mathsf{E}\|\phi(X) - \mathsf{E}[\phi(X)]\|^2, \\
\mathsf{E}\|\phi(Y) - \phi(Y')\|^2 &= 2\mathsf{E}\|\phi(Y) - \mathsf{E}[\phi(Y)]\|^2.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&2\mathsf{E}d(X,Y) - \mathsf{E}d(X,X') - \mathsf{E}d(Y,Y') \\
&= 2\mathsf{E}\|\phi(X) - \phi(Y)\|^2 - \mathsf{E}\|\phi(X) - \phi(X')\|^2 - \mathsf{E}\|\phi(Y) - \phi(Y')\|^2 \\
&= 2\mathsf{E}\|\mathsf{E}[\phi(X)] - \mathsf{E}[\phi(Y)]\|^2,
\end{aligned}$$

which means $\varepsilon(X,Y) \geqslant 0$.

Moreover, if (M,d) is strong negative, then, one can derive that the map $\beta_\phi(\mu) : \mu \mapsto \int \phi(x)\mathrm{d}\mu(x)$ is injective [8], which implies $\mathsf{E}[\phi(X)] = \mathsf{E}[\phi(Y)]$ if and only if $\mu = \nu$. □

Lyons [8] illustrated that all Euclidean spaces have strong negative type and proved that every separable Hilbert space is of strong negative type. Thus even if the observations are complex objects, like functions, texts, and graphs etc., one can use their real-valued nonnegative distances for inference.

In this paper, we give the next two notes on energy distance: Brownian distance and energy distance for dependent random vectors.

§2. Brownian Distance

The notion of covariance with respect to a stochastic process, named Brownian distance covariance, was introduced by Székely and Rizzo [9], and it was shown that the population distance covariance coincides with Brownian covariance. Similarly, we will present the notion of Brownian distance, and show the consistent of Brownian distance and energy distance.

Definition 2 Let X and Y be two $\mathbb{R}^p$-valued random variables, with distributions F and G respectively. Suppose W be a Brownian motion with expectation zero and covariance function

$$\mathsf{Cov}\,(W(t), W(s)) = \|s\| + \|t\| - \|s - t\| \tag{4}$$

on $\mathbb{R}^p$, independent of X and Y. The Brownian distance between X and Y is the non-negative number whose square is:

$$\mathrm{BD}_W^2(X,Y) = \mathsf{E}\{[W(X) - W(Y)][W(X') - W(Y')]\}, \tag{5}$$

where X' is an iid copy of X, and Y' is an i.i.d. copy of Y.

For two independent real-valued random variables X, Y, the square of their mean's difference is

$$(\mathsf{E}X - \mathsf{E}Y)^2 = (\mathsf{E}X - \mathsf{E}Y)(\mathsf{E}X' - \mathsf{E}Y') = \mathsf{E}(X - Y)\mathsf{E}(X' - Y') = \mathsf{E}(X - Y)(X' - Y').$$

So, Brownian distance can be understood as the generalized definition of square of mean difference by replacing identity process $I(t)$ with Brownian motion $W(t)$.

Theorem 3 If X and Y are two independent $\mathbb{R}^p$-valued random variables, with $\mathsf{E}\|X\| + \mathsf{E}\|Y\| < \infty$, then $\mathsf{E}\{[W(X) - W(Y)][W(X') - W(Y')]\}$ is nonnegative and finite, and

$$\mathrm{BD}_W^2(X, Y) = 2\mathsf{E}\|X - Y\| - \mathsf{E}\|X - X'\| - \mathsf{E}\|Y - Y'\|.$$

Proof Observe that

$$\begin{aligned}\mathrm{BD}_W^2(X, Y) &= \mathsf{E}\{[W(X) - W(Y)][W(X') - W(Y')]\} \\ &= \mathsf{E}\{\mathsf{E}\{[W(X) - W(Y)][W(X') - W(Y')] \,|\, W\}\} \\ &= \mathsf{E}\{\mathsf{E}\{[W(X) - W(Y)] \,|\, W\}\mathsf{E}\{[W(X') - W(Y')] \,|\, W\}\} \\ &= \mathsf{E}\{\mathsf{E}[W(X) \,|\, W] - \mathsf{E}[W(Y) \,|\, W]\}^2,\end{aligned}$$

where $\mathsf{E}[W(X) \,|\, W] = \int_{R^d} W(t)\mathrm{d}F(t)$ and $F(t)$ is the distribution function of X, and similar means for $\mathsf{E}[W(Y) \,|\, W]$. So, $\mathsf{E}\{[W(X)-W(Y)][W(X')-W(Y')]\}$ is always nonnegative. For finiteness, note that

$$\mathsf{E}[W^2(X)] = \mathsf{E}\{\mathsf{E}[W^2(X) \,|\, X]\} = \mathsf{E}(2\|X\|) = 2\mathsf{E}\|X\| < \infty.$$

Therefore, we have

$$\begin{aligned}\mathrm{BD}_W^2(X, Y) &= \mathsf{E}\{\mathsf{E}[W(X) \,|\, W] - \mathsf{E}[W(Y) \,|\, W]\}^2 \\ &\leqslant 2\mathsf{E}\{\{\mathsf{E}[W(X) \,|\, W]\}^2 + \mathsf{E}[W(Y) \,|\, W]^2\} \\ &\leqslant 2\mathsf{E}\{\mathsf{E}[W^2(X) \,|\, W] + \mathsf{E}[W^2(Y) \,|\, W]\} \\ &= 2\{\mathsf{E}[W^2(X)] + \mathsf{E}[W^2(Y)]\} = 4(\mathsf{E}\|X\| + \mathsf{E}\|Y\|) < \infty.\end{aligned}$$

On the other hand,

$$\begin{aligned}&\mathrm{BD}_W^2(X, Y) = \mathsf{E}\{[W(X) - W(Y)][W(X') - W(Y')]\} \\ &= \mathsf{E}[W(X)W(X') + W(Y)W(Y') - W(X)W(Y') - W(Y)W(X')] \\ &= \mathsf{E}\{\mathsf{E}\{[W(X)W(X') + W(Y)W(Y') - W(X)W(Y') - W(Y)W(X')] \,|\, X, X', Y, Y'\}\} \\ &= \mathsf{E}[\|X\| + \|X'\| - \|X - X'\| + \|Y\| + \|Y'\| - \|Y - Y'\| - 2(\|X\| + \|Y'\| - \|X - Y'\|)] \\ &= 2\mathsf{E}\|X - Y\| - \mathsf{E}\|X - X'\| - \mathsf{E}\|Y - Y'\|.\end{aligned}$$

This completes the proof. □

Theorem 3 indicates that the properties of Brownian distance for random vectors X and Y are therefore the same properties for energy distance. We have the result as follows.

Remark 4 Let X and Y be two $\mathbb{R}^p$-valued random variables, with distributions F and G respectively, W be a Brownian motion with expectation zero and covariance function (4) on $\mathbb{R}^p$, independent of X and Y, then, one has:

$$\mathrm{BD}_W(X,Y)=0 \text{ if and only if } X\overset{\mathrm{D}}{=}Y,$$

where $X\overset{\mathrm{D}}{=}Y$ refers to X and Y are identically distributed.

§3. Energy Distance for Paired Variables

From the introduction in section one, we can see that although the concept of energy distance is defined for independent variables, it's also applicable for dependent variables. In this section, we discuss the energy distance for paired variables which is a special case of dependent variables.

Definition 5 Let X and Y be two $\mathbb{R}^p$-valued random variables, suppose (X,Y) has joint distribution H, with marginal distribution F on X and G on Y respectively. Assume that $\mathsf{E}\|X\|+\mathsf{E}\|Y\|<\infty$, the energy distance between X and Y is defined as

$$\epsilon(X,Y):=\mathsf{E}\|X-Y'\|+\mathsf{E}\|Y-X'\|-\mathsf{E}\|X-X'\|-\mathsf{E}\|Y-Y'\|, \tag{6}$$

where (X',Y') is an i.i.d. copy of (X,Y).

Similar to equality (2) we have following result.

Theorem 6 Denote $\widehat{f}$ as X's characteristic function and $\widehat{g}$ as Y's characteristic function, then, one has the equality

$$\mathsf{E}\|X-Y'\|+\mathsf{E}\|Y-X'\|-\mathsf{E}\|X-X'\|-\mathsf{E}\|Y-Y'\|=\frac{1}{c_p}\int_{\mathbb{R}^p}\frac{|\widehat{f}(t)-\widehat{g}(t)|^2}{\|t\|^{p+1}}\mathrm{d}t, \tag{7}$$

where $c_p=\pi^{(p+1)/2}/\Gamma[(p+1)/2]$. Thus $\epsilon(X,Y)\geqslant 0$ with equality to zero if and only if X and Y are identically distributed.

Proof Note that $\widehat{f}(t)=\mathsf{E}\mathrm{e}^{it'X}$, $\widehat{g}=\mathsf{E}\mathrm{e}^{it'Y}$, so

$$\begin{aligned}
|\widehat{f}(t)-\widehat{g}(t)|^2&=[\widehat{f}(t)-\widehat{g}(t)][\overline{\widehat{f}(t)}-\overline{\widehat{g}(t)}]\\
&=\widehat{f}(t)\overline{\widehat{f}(t)}+\widehat{g}(t)\overline{\widehat{g}(t)}-\widehat{f}(t)\overline{\widehat{g}(t)}-\overline{\widehat{f}(t)}\widehat{g}(t)\\
&=\mathsf{E}\mathrm{e}^{it'X}\mathsf{E}\mathrm{e}^{-it'X'}+\mathsf{E}\mathrm{e}^{it'Y}\mathsf{E}\mathrm{e}^{-it'Y'}-\mathsf{E}\mathrm{e}^{it'X}\mathsf{E}\mathrm{e}^{-it'Y'}-\mathsf{E}\mathrm{e}^{-it'X'}\mathsf{E}\mathrm{e}^{it'Y}\\
&=\mathsf{E}\mathrm{e}^{it'(X-X')}+\mathsf{E}\mathrm{e}^{it'(Y-Y')}-\mathsf{E}\mathrm{e}^{it'(X-Y')}-\mathsf{E}\mathrm{e}^{it'(Y-X')}\\
&=1-\mathsf{E}\mathrm{e}^{it'(X-Y')}+1-\mathsf{E}\mathrm{e}^{it'(X'-Y)}-(1-\mathsf{E}\mathrm{e}^{it'(X-X')})-(1-\mathsf{E}\mathrm{e}^{it'(Y-Y')}).
\end{aligned}$$

Therefore, by integral on both sides, we get

$$\frac{1}{c_p}\int_{\mathbb{R}^p}\frac{|\widehat{f}(t)-\widehat{g}(t)|^2}{\|t\|^{p+1}}\mathrm{d}t = \mathsf{E}\|X-Y'\|+\mathsf{E}\|Y-X'\|-\mathsf{E}\|X-X'\|-\mathsf{E}\|Y-Y'\|,$$

via the integral equation [7]

$$\int_{\mathbb{R}^p}\frac{1-\cos\langle t,x\rangle}{\|t\|^{p+1}}\mathrm{d}t = c_p\|x\|.$$

Thus, we have $\epsilon(X,Y)\geqslant 0$ with equality to zero if and only if X and Y are identically distributed. □

Remark 7 Let X be a $\mathbb{R}^p$-valued random variable, with distribution F, suppose $\mathsf{E}\|X\|<\infty$, then $\mathsf{E}\|X+X'\|\geqslant\mathsf{E}\|X-X'\|$, where X' is an i.i.d. copy of X, $\mathsf{E}\|X+X'\|=\mathsf{E}\|X-X'\|$ if and only if X is diagonally symmetric.

Proof Wang et al. [12] gives the proof of this result for X be univariate. Generally, we can put $Y=-X$ to get

$$\begin{aligned}\epsilon(X,Y)&=\epsilon(X,-X)\\&=\mathsf{E}\|X-(-X')\|+\mathsf{E}\|X'-(-X)\|-\mathsf{E}\|X-X'\|-\mathsf{E}\|-X-(-X')\|\\&=2(\mathsf{E}\|X+X'\|-\mathsf{E}\|X-X'\|).\end{aligned}$$

Thus, according to Theorem 6 we have $\mathsf{E}\|X+X'\|-\mathsf{E}\|X-X'\|\geqslant 0$ and $\mathsf{E}\|X+X'\|-\mathsf{E}\|X-X'\|=0$ if and only if X and $-X$ are identically distributed, which means that X is diagonally symmetric. □

We now give the empirical energy distance for paired variables.

Definition 8 Let $W_i=(X_i,Y_i)$, $i=1,2,\ldots,n$ be a sample from the distribution of (X,Y), and denote $\boldsymbol{W}=(\boldsymbol{X},\boldsymbol{Y})=\{W_1,W_2,\ldots,W_n\}$, the sample energy distance of (X,Y) is given as

$$\epsilon_n(\boldsymbol{X},\boldsymbol{Y}):=\frac{1}{C_n^2}\sum_{i<j}(\|X_i-Y_j\|+\|Y_i-X_j\|-\|X_i-X_j\|-\|Y_i-Y_j\|).$$

If we denote

$$h(w_1,w_2):=h((x_1,y_1),(x_2,y_2))=\|x_1-y_2\|+\|y_1-x_2\|-\|x_1-x_2\|-\|y_1-y_2\|,$$

then $\epsilon_n(\boldsymbol{X},\boldsymbol{Y})$ has the form of an U-statistic, with kernel h

$$\epsilon_n(\boldsymbol{X},\boldsymbol{Y})=\frac{1}{C_n^2}\sum_{i<j}h(W_i,W_j).$$

Obviously, $\epsilon_n(\boldsymbol{X},\boldsymbol{Y})$ is an unbiased estimator of $\epsilon(X,Y)$.

Proposition 9 The empirical energy distance $\epsilon_n(\boldsymbol{X},\boldsymbol{Y})$ almost surely converges to the energy distance $\epsilon(X,Y)$. That is, $\epsilon_n(\boldsymbol{X},\boldsymbol{Y}) \xrightarrow[n\to\infty]{\text{a.s.}} \epsilon(X,Y)$.

Proof As to $h(W_1,W_2)$, we have

$$\mathsf{E}|h(W_1,W_2)| \leqslant 4 * (\mathsf{E}\|X_1\| + \mathsf{E}\|Y_1\|) < \infty,$$

due to the suppose $\mathsf{E}\|X\| + \mathsf{E}\|Y\| < \infty$ and $\mathsf{E}[h(W_1,W_2)] = \epsilon(X,Y)$. According to [11], we can obtain that

$$\epsilon_n(\boldsymbol{X},\boldsymbol{Y}) \xrightarrow[n\to\infty]{\text{a.s.}} \epsilon(X,Y).$$

Hence, the result follows. □

Theorem 10 Let X and Y be two $\mathbb{R}^p$-valued random variables, with $\mathsf{E}\|X\|+\mathsf{E}\|Y\| < \infty$, then

(i) If X and Y are identically distributed, then

$$n\epsilon_n(\boldsymbol{X},\boldsymbol{Y}) \xrightarrow[n\longrightarrow\infty]{\mathscr{D}} \sum_{v=1}^{\infty} \lambda_v(Z_v^2 - 1),$$

where Z_v are independent standard normal random variables,λ_v are nonnegative constants that depend on the distribution of (X,Y).

(ii) If X and Y are not identically distributed, then

$$n\epsilon_n(\boldsymbol{X},\boldsymbol{Y}) \xrightarrow[n\longrightarrow\infty]{\text{a.s.}} \infty.$$

Proof (i) When X and Y are identically distributed, we have

$$\mathsf{E}h(W_1,W_2) = \mathsf{E}\|X_1 - Y_2\| + \mathsf{E}\|X_2 - Y_1\| - \mathsf{E}\|X_1 - X_2\| - \mathsf{E}\|Y_1 - Y_2\| = 0,$$

and

$$\begin{aligned}
\overline{h}_1(w_1) &:= \mathsf{E}h(w_1,W_2) = \mathsf{E}h((x_1,y_1),(X_2,Y_2)) \\
&= \mathsf{E}\|x_1 - Y_2\| + \mathsf{E}\|X_2 - y_1\| - \mathsf{E}\|x_1 - X_2\| - \mathsf{E}\|y_1 - Y_2\| = 0, \\
\overline{h}_1(w_2) &:= \mathsf{E}h(W_1,w_2) = 0, \\
\overline{h}_2(w_1,w_2) &= \|x_1 - y_2\| + \|x_2 - y_1\| - \|x_1 - x_2\| - \|y_1 - y_2\|,
\end{aligned}$$

which means that $\epsilon_n(\boldsymbol{X},\boldsymbol{Y})$ is a degenerate U-statistic of order 1, therefore, it converges in distribution to a quadratic form:

$$n\epsilon_n(\boldsymbol{X},\boldsymbol{Y}) \xrightarrow[n\longrightarrow\infty]{\mathscr{D}} \sum_{v=1}^{\infty} \lambda_v(Z_v^2 - 1),$$

where Z_v are independent standard normal random variables, λ_v are the eigenvalues of the equation $\int \overline{h}_2(w_1, w_2) f(w_2) \mathrm{d}H(w_2) = \lambda f(w_1)$.

(ii) The result is obviously because $\epsilon_n(\boldsymbol{X}, \boldsymbol{Y}) \xrightarrow[n\to\infty]{\text{a.s.}} \epsilon(X, Y) > 0$ when X and Y are not identically distributed. □

Theorem 10 implies that we can reject the null hypothesis of X and Y are identically distributed when $n\epsilon_n(\boldsymbol{X}, \boldsymbol{Y})$ is too large.

§4. Paired-Sample Test Based on Energy Distance

In this section, we consider the the problem of testing the equality of distributions for paired variables based on the results we discussed in Section 3. For example, if we give two different treatments A, B to the same research object, denote $X = (X_1, X_2, \ldots, X_p)$ as the records of p features under A treatment, and $Y = (Y_1, Y_2, \ldots, Y_p)$ as the records under B treatment. If we want to compare the effect of this two different treatment, then, we have to test the equality of distributions of X and Y. In another case, if we give a treatment to a research object, also denote X the records before treatment, and Y the records after treatment. Again, to test whether the treatment is effective or not, we need to test the equality of distributions of X and Y.

The classical methods for testing equality of distributions for paired-sample include univariate t-test, Hotelling T^2 test, sign test, and Wilcoxcon signed rank test, etc. Univariate t-test and Hotelling T^2 test are applicable to (X, Y) with joint distribution of two-dimensional or multidimensional normal distributions, and in fact are testing the equality of means for X, Y. When (X, Y) does not meet the joint normal distribution assumption, one can use nonparametric methods such as sign test and Wilcoxon signed rank test. As we all know, the sign test is actually to test whether the distribution of $X - Y$ is with zero as median or not, while Wilcoxon signed rank test, is to test the symmetry of distribution of $X - Y$. So they are not consistent tests. Moreover, sign test and Wilcoxcon signed rank test are only applicable for X, Y be both univariate.

The statistic

$$\epsilon_n(\boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{C_n^2} \sum_{i<j} h(W_i, W_j)$$

presented in Section 3 gives a new idea to test the equality of distributions for paired-sample of arbitrary dimensions. Theorem 10 shows that $n\epsilon_n(\boldsymbol{X}, \boldsymbol{Y})$ converges in distribution to a quadratic form quadratic form $\sum\limits_{v=1}^{\infty} \lambda_v (Z_v^2 - 1)$ under null hypothesis of X and Y are identically distributed. But the λs depend on the distribution of (X, Y), so it's

difficult to compute the p-values by this result. Moreover, we need to be cautious of its serious limitations in practice, because the sample size in practice may be too small.

In practice, we consider the bootstrap method as an alternative approximation to the null distribution of $n\epsilon_n(\boldsymbol{X}, \boldsymbol{Y})$. Noticing that under H_0, (X, Y) has the same distribution as (Y, X), so we resample from the sample $D_n := \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n), (Y_1, X_1), (Y_2, X_2), \ldots, (Y_n, X_n)\}$. Denote $\boldsymbol{W}_n^* = \{(X_i^*, Y_i^*)\}_{i=1}^n$ the bootstrap sample obtained by sampling with replacement from D_n, and the bootstrap statistic is $n\epsilon_n^*$. Repeat this procedure for B times to obtain $n\epsilon_{nk}^*$, $k = 1, 2, \ldots, B$, then the p-value of the test is given by

$$p = \Big[1 + \sum_{k=1}^{B} I(|n\epsilon_{nk}^*| > |n\epsilon_n|)\Big] / (B + 1).$$

We conduct a simple Monte Carlo simulation to demonstrate the performance of the method we proposed here in comparison to t-test and Wilcoxcon signed rank test for paired sample. Consider the following 6 examples:

Example 1: $(X, Y) \sim \mathrm{N}_2(\mu, \Sigma)$, with $\mu = (0, 0)'$, $\Sigma = \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$;

Example 2: $X \sim \mathrm{U}(0, 6)$, $Y = 6 - X$;

Example 3: $X \sim \mathrm{B}(10, 0.5)$, $Y = 10 - X$;

Example 4: $(X, Y) \sim \mathrm{N}_2(\mu, \Sigma)$, with $\mu = (1, 1)'$, $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 4 \end{pmatrix}$;

Example 5: $X \sim \mathrm{Exp}(4)$, $Y = 0.5\mathrm{e}^{-4X}$;

Example 6: $X \sim \mathrm{U}(0, 1)$, $Z \sim \mathrm{N}(0, 1)$, $Y = X - Z$.

Examples 1 – 3 consider the cases that X and Y are identically distributed. Examples 4 – 6 are the cases that X and Y are not identically distributed, with Example 4 be the case that X and Y have the same expectation, and Example 6 be the case $X - Y$ are symmetric distributed. We conduct the simulation with sample sizes $n = 30, 50, 100, 200, 400$, and the number of bootstrap resamples is 199. Each experiment is based on 500 replications, and we use the significance level of 0.05. Simulation results are summarized in Table 1, where **ptest** refers to the method we proposed in this paper.

The numerical simulation results of Examples 1 – 3 show that **ptest** can control the Type I error very well when X and Y are identically distributed, while Examples 4 – 6 illustrate that **ptest** can identify the distribution differences of the paired variables more effectively than the classical t-test and Wilcoxon signed rank test.

Table 1 Proportion of rejections for Examples 1 − 6 at 5% level

model	method	$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
Example 1	t.test	0.054	0.030	0.052	0.030	0.040
	Wilcox.test	0.058	0.036	0.050	0.038	0.044
	ptest	**0.048**	**0.030**	**0.046**	**0.022**	**0.050**
Example 2	t.test	0.072	0.074	0.060	0.048	0.058
	Wilcox.test	0.072	0.068	0.060	0.048	0.062
	ptest	**0.062**	**0.066**	**0.052**	**0.044**	**0.058**
Example 3	t.test	0.080	0.068	0.070	0.048	0.050
	Wilcox.test	0.072	0.068	0.062	0.044	0.044
	ptest	**0.062**	**0.078**	**0.054**	**0.042**	**0.052**
Example 4	t.test	0.060	0.048	0.040	0.040	0.050
	Wilcox.test	0.048	0.038	0.050	0.034	0.042
	ptest	**0.574**	**0.884**	**1.000**	**1.000**	**1.000**
Example 5	t.test	0.046	0.050	0.038	0.042	0.042
	Wilcox.test	0.072	0.092	0.132	0.234	0.406
	ptest	**0.098**	**0.176**	**0.390**	**0.916**	**1.000**
Example 6	t.test	0.034	0.044	0.030	0.038	0.050
	Wilcox.test	0.032	0.040	0.022	0.038	0.056
	ptest	**0.994**	**1.000**	**1.000**	**1.000**	**1.000**

References

[1] SZÉKELY G J. E-statistics: the energy of statistical samples [R]. Technical Report No. 02-16, Bowling Green, Ohio: Bowling Green State University, Department of Mathematics and Statistics, 2002.

[2] RIZZO M L. A test of homogeneity for two multivariate populations [C] // *2002 Proceedings of the American Statistical Association, Physical and Engineering Sciences Section* [CD-ROM], Alexandria, VA: American Statistical Association, 2002.

[3] SZÉKELY G J, RIZZO M L. A new test for multivariate normality [J]. *J Multivariate Anal*, 2005, **93(1)**: 58–80.

[4] SZÉKELY G J, RIZZO M L. Hierarchical clustering via joint between-within distances: extending Ward's minimum variance method [J]. *J Classification*, 2005, **22(2)**: 151–183.

[5] SZÉKELY G J, RIZZO M L, BAKIROV N K. Measuring and testing dependence by correlation of distances [J]. *Ann Statist*, 2007, **35(6)**: 2769–2794.

[6] RIZZO M L, SZÉKELY G J. DISCO analysis: a nonparametric extension of analysis of variance [J]. *Ann Appl Stat*, 2010, **4(2)**: 1034–1055.

[7] SZÉKELY G J, RIZZO M L. Energy statistics: a class of statistics based on distances [J]. *J Statist Plann Inference*, 2013, **143(8)**: 1249–1272.

[8] LYONS R. Distance covariance in metric spaces [J]. *Ann Probab*, 2013, **41(5)**: 3284–3305.

[9] SZÉKELY G J, RIZZO M L. Brownian distance covariance [J]. *Ann Appl Stat*, 2009, **3(4)**: 1236–1265.

[10] SCHOENBERG I J. On certain metric spaces arising from Euclidean spaces by a change of metric and their imbedding in Hilbert space [J]. *Ann Math*, 1937, **38(4)**: 787–793.

[11] LEE A J. U-Statistics: Theory and Practice [M]. New York: Marcel Dekker, Inc., 1990.

[12] WANG F, CUI H J, JIN J. Symmetric center test by using U-statistics [J]. *J Beijing Normal Univ* (*Nat Sci*), 2009, **45(1)**: 17–21. (in Chinese)

关于能量距离的两点注记

陈 敏 琼

(中山大学新华学院, 广州, 510520)

摘 要: 本文讨论了能量距离的两个问题. 类似 Brownian 协方差的讨论提出了 Brownian 距离的定义, 并证明了 Brownian 距离与能量距离的一致性. 给出了配对变量的能量距离的表示, 并探讨了将能量距离用于配对样本同分布的检验问题时原假设下的渐近分布理论. 最后通过一个简单的数值模拟说明基于能量距离的配对样本的分布差异的检验方法比传统的 t 检验及 Wilcoxon 符号秩检验更有效.

关键词: 能量距离; Brownian 距离; 配对变量; 渐近分布; 数值模拟

中图分类号: O212.7