A NEW CLASS OF MEASURES FOR TESTING INDEPENDENCE

Xiangrong Yin

Department of Statistics, University of Kentucky,

725 Rose St. Lexington, Kentucky, 40536-0082, E-mail: yinxiangrong@uky.edu

Qingcong Yuan

Department of Statistics, Miami University,

Oxford, Ohio, 45056, E-mail: qingcong.yuan@miamioh.edu

Supplementary Material

This supplementary file provides additional materials related to the newly proposed index in Section 3 in the paper. It also includes proofs of propositions and theorems stated in the paper, and additional simulations that support our conclusion.

S1 Brownian Motion Approach

We use the discrepancy between the characteristic functions and a particular weight function to lead to our index (6). However, in this section, we show that a Brownian motion procedure also can derive our index (6).

Let W be a two-sided one-dimensional Brownian motion/Wiener process with expectation zero and covariance function $|s| + |t| - |s - t| = 2\min(s,t), s,t > 0$ (Székely and Rizzo (2009, 3.3)).

Definition 1. The Brownian conditional difference or the Wiener conditional difference of a real-valued random vector $\mathbf{X}$ given $\mathbf{Y}$ with finite second moments is a non-negative number defined by $\mathcal{D}^2_W(\mathbf{X}|\mathbf{Y}) = \mathrm{E}(\mathbf{X}_W\mathbf{X}'_W|\mathbf{Y})$, where W does not depend on $(\mathbf{X}, \mathbf{X}', \mathbf{Y})$.

With this definition, we then have the following result.

Proposition 1. *If* $\mathbf{X}$ *is an* $\mathbb{R}^p$ *valued random vector,* $\mathbf{Y}$ *is an* $\mathbb{R}^q$ *valued random vector, and* $\mathrm{E}[|\mathbf{X}|^2 + \mathrm{E}(|\mathbf{X}|^2|\mathbf{Y})] < \infty$*, then* $\mathrm{E}(\mathbf{X}_W\mathbf{X}'_W|\mathbf{Y})$ *is nonnegative and finite. Let* $\mathbf{X}$ *and* $\mathbf{X}'$ *be iid, and* $\mathbf{X}_\mathbf{Y}$ *and* $\mathbf{X}'_\mathbf{Y}$ *be iid; Expectations are taken over every random vector except conditioning on* $\mathbf{Y}$ *if it appears. Then, (6) holds. That is,*

$$\mathcal{C}^2(\mathbf{X}|\mathbf{Y}) = \mathrm{E}[\mathcal{D}^2_W(\mathbf{X}|\mathbf{Y})].$$

Proof.

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

which is nonnegative. Finiteness can be obtained as Székely and Rizzo (2009, page 1262). Note that $\mathcal{D}^2_W(\mathbf{X}|\mathbf{Y}) = \mathrm{E}[\mathrm{E}(\mathbf{X}_W\mathbf{X}'_W|\mathbf{Y}, \mathbf{X}, \mathbf{X}')|\mathbf{Y}]$. Now using the same argument on page 1263 of Székely and Rizzo (2009), we have that

$$\mathrm{E}(\mathbf{X}_W\mathbf{X}'_W|\mathbf{Y}, \mathbf{X}, \mathbf{X}') = \mathrm{E}'|\mathbf{X}_\mathbf{Y} - \mathbf{X}'| + \mathrm{E}|\mathbf{X}'_\mathbf{Y} - \mathbf{X}| - |\mathbf{X}_\mathbf{Y} - \mathbf{X}'_\mathbf{Y}| - \mathrm{E}|\mathbf{X} - \mathbf{X}'|,$$

where the first expectation E' is over $\mathbf{X}'$, the second expectation is over $\mathbf{X}$, and the last one is over both $\mathbf{X}$, and $\mathbf{X}'$. Thus, by using the fact that $\mathbf{X}$ and $\mathbf{X}'$ are iid, and $\mathbf{X}_{\mathbf{Y}}$ and $\mathbf{X}'_{\mathbf{Y}}$ are iid,

$$\mathcal{D}_W^2(\mathbf{X}|\mathbf{Y}) = \mathrm{E}[(\mathrm{E}'|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'|)|\mathbf{Y}] + \mathrm{E}[(\mathrm{E}|\mathbf{X}'_{\mathbf{Y}} - \mathbf{X}|)|\mathbf{Y}] - \mathrm{E}[(|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'_{\mathbf{Y}}|)|\mathbf{Y}] - \mathrm{E}|\mathbf{X} - \mathbf{X}'|.$$

By taking expectation over $\mathbf{Y}$, and the fact that the first term and the last term are equal, consequently, we have that $\mathcal{C}^2(\mathbf{X}|\mathbf{Y}) = \mathrm{E}[\mathcal{D}_W^2(\mathbf{X}|\mathbf{Y})]$. That is, again (6) holds. □

S2 Relations to DISCO

Our index does not require Y to be discrete. However, if Y is categorical, then it is much intuitive and clear that our estimation method provides a close link to ANOVA, MANOVA and, most recently DISCO (Rizzo and Székely (2010)).

To be more specific, we can define the following population within distance and sample within distance, total distance and its sample version, respectively. If we consider $e^{it^T\mathbf{X}_Y}$ as an observation, $\mathrm{E}(e^{it^T\mathbf{X}_Y})$ as the group mean and $\mathrm{E}(e^{it^T\mathbf{X}})$ as the overall mean in the typical ANOVA calculation, the resulting measures are exactly what are calculated in ANOVA.

Definition 2. The population within distance is defined as:

$$\mathcal{W}^2(\mathbf{X}|Y) = \mathrm{E}[\mathcal{W}_w^2(\mathbf{X}|Y)] = \mathrm{E}\int |e^{it^T\mathbf{X}_Y} - \mathrm{E}e^{it^T\mathbf{X}_Y}|^2 w(t)dt;$$

The sample within distance is defined as:

$\mathcal{W}_n^2(\mathbf{X}|Y) = \sum_{y=1}^{H} p_y ||e^{it^T\mathbf{X}_y} - f^n_{\mathbf{X}|y}(t)||^2.$

The population total distance is defined as:

$\mathcal{T}^2(\mathbf{X}|Y) = \mathrm{E}[\mathcal{T}_w^2(\mathbf{X}|Y)] = \mathrm{E}\int |e^{it^T\mathbf{X}_Y} - \mathrm{E}e^{it^T\mathbf{X}}|^2 w(t)dt;$

The sample total distance is defined as:

$\mathcal{T}_n^2(\mathbf{X}|Y) = \sum_{y=1}^{H} p_y ||e^{it^T\mathbf{X}_y} - f^n_{\mathbf{X}}(t)||^2.$

We can have their respective equivalent formulas, stated below.

Proposition 2. *The population within distance can be rewritten as:*

$\mathcal{W}^2(\mathbf{X}|Y) = \mathrm{E}[\mathcal{W}_w^2(\mathbf{X}|Y)] = \mathrm{E}|\mathbf{X}_Y - \mathbf{X}'_Y|;$

The sample within distance can be rewritten as:

$\mathcal{W}_n^2(\mathbf{X}|Y) = \frac{1}{n}\sum_{y=1}^{H} \frac{1}{n_y}\sum_{k_y,l_y=1}^{n_y,n_y} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}|.$

The population total distance can be rewritten as:

$\mathcal{T}^2(\mathbf{X}|Y) = \mathcal{C}^2(\mathbf{X}|\mathbf{X}) = \mathrm{E}|\mathbf{X} - \mathbf{X}'|;$

The sample total distance can be rewritten as:

$\mathcal{T}_n^2(\mathbf{X}|Y) = \frac{1}{n^2}\sum_{y,y'=1}^{H,H} \sum_{k_y,l_{y'}=1}^{n_y,n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}|.$

The following result is a straightforward calculation, thus we omitted its proof.

Proposition 3. *1.* $\mathcal{T}^2(\mathbf{X}|Y) = \mathcal{C}^2(\mathbf{X}|Y) + \mathcal{W}^2(\mathbf{X}|Y)$*;*

2. $\mathcal{T}_n^2(\mathbf{X}|Y) = \mathcal{C}_n^2(\mathbf{X}|Y) + \mathcal{W}_n^2(\mathbf{X}|Y)$*.*

Under the null hypothesis, by SLLN, as $n \to \infty$, $\mathcal{W}_n^2(\mathbf{X}|Y) \to \mathrm{E}|\mathbf{X} - \mathbf{X}'|$. Or note that $\mathrm{E}[\mathcal{T}_n^2(\mathbf{X}|Y)] = \mathrm{E}[\mathcal{W}_n^2(\mathbf{X}|Y)]$, thus analogous to ANOVA, we may use test statistic,

$$\frac{\mathcal{C}_n^2(\mathbf{X}|Y)/(H-1)}{\mathcal{W}_n^2(\mathbf{X}|Y)/(n-H)},$$

which is the ratio of between distance over within distance. Note that the previous test statistic in Section 4.1,

$$\frac{n\mathcal{C}_n^2(\mathbf{X}|Y)}{S_n} = \frac{\mathcal{C}_n^2(\mathbf{X}|Y)/(H-1)}{\mathcal{T}_n^2(\mathbf{X}|Y)/n} = \frac{n}{n-1}\frac{\mathcal{C}_n^2(\mathbf{X}|Y)/(H-1)}{\mathcal{T}_n^2(\mathbf{X}|Y)/(n-1)}.$$

With negligible factor $\frac{n}{n-1}$, this is the ratio of between distance over total distance. Note that $\frac{n\mathcal{C}_n^2(\mathbf{X}|Y)}{S_n}\frac{(H-1)}{n} = R_{c,n}^2$, an estimator of R_c^2.

In particular, one can show that for response with two categories, the energy distance of Rizzo and Székely (2010, page 1038) is proportion to $\mathcal{C}^2(\mathbf{X}|Y)$. Indeed, one also can show that $n\mathcal{C}_n^2(\mathbf{X}|Y) = 2S_\alpha$ and $n\mathcal{W}^2(\mathbf{X}|Y) = 2W_\alpha$, with $\alpha = 1$, where S_α and W_α are defined in Rizzo and Székely (2010).

Classical methods of ANOVA or MANVOA for multi-sample usually require normally distributed error (see, e.g., Cochran and Cox (1957); Hand and Taylor (1987); Mardia, Kent, and Bibby (1979)), especially for inference. When such condition fails, one may apply F statistics via permutation test procedure (Efron and Tibshirani (1998); Davison and Hinkley (1997)). Rich literature exists in beyond testing the mean differences but on distributions, for instance, Akritas and Arnold (1994) and Gower and Krzanowski (1999) for structured data, and Anderson (2001), McArdle and Anderson (2001), Excoffier, Smouse, and Quattro (1992) and Zapala and Schork (2006) with applications in ecology and genetics.

S3 The class of α-divergence

We also can extend our measure (6) to a one parameter family of measures indexed with a positive exponent α. Note that in our previous application $\alpha = 1$.

Suppose that $\mathrm{E}|\mathbf{X}_{\mathbf{Y}}|^{\alpha} < \infty$. Let $\mathcal{C}^{(\alpha)}(\mathbf{X}|\mathbf{Y})$ denote the $\alpha-$measure which is the nonnegative number defined by

$$\mathcal{C}^{2(\alpha)}(\mathbf{X}|\mathbf{Y}) = \mathrm{E}_{\mathbf{Y}}||f_{\mathbf{X}|\mathbf{Y}}(t) - f_{\mathbf{X}}(t)||_{\alpha}^{2} = \mathrm{E}_{\mathbf{Y}} \int_{\mathbb{R}^p} \frac{|f_{\mathbf{X}|\mathbf{Y}}(t) - f_{\mathbf{X}}(t)|^2}{\tilde{C}(p,\alpha)|t|^{\alpha+p}} dt.$$

The $\alpha-$measure statistics are defined by replacing the exponent 1 with exponent α in the respective formulas (6) and (8). That is, for instance, in (8) replace $|\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}|$ by $|\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}|^{\alpha}$. Lemma 4.2 can be generalized for $||\cdot||_{\alpha}-$norms, so that almost surely convergence of $\mathcal{C}_n^{2(\alpha)}(\mathbf{X}|Y) \to \mathcal{C}^{2(\alpha)}(\mathbf{X}|Y)$ follows if the $\alpha-$moments are finite. Similarly one can prove the weak convergence and statistical consistency for α exponents, $0 < \alpha < 2$, provided that α moments are finite. However, when $\alpha = 2$, it leads to $2\mathrm{E}(\mu_Y - \mu)^2$, where μ_Y is the mean for group Y and μ is the overall mean. Thus in such a case, $\mathcal{C}^{2(2)}(\mathbf{X}|Y) = 0$ iff $\mu_Y = \mu$ for all Y. Furthermore, for $0 < \alpha \leq 2$, $n\mathcal{C}_n^{2(\alpha)}(\mathbf{X}|Y) = 2S_{\alpha}$ and $n\mathcal{W}_n^{2(\alpha)}(\mathbf{X}|Y) = 2W_{\alpha}$, where S_{α} and W_{α} are defined in Rizzo and Székely (2010).

One can consider the Levy fractional Brownian motion $\{W_H^d(t), t \in \mathbb{R}^d\}$, with Hurst index $H \in (0,1)$, which is a centered Gaussian random process with covariance function (Herbin and Merzbach (2007)):

$$\mathrm{E}[W_H^d(t)W_H^d(s)] = |t|^{2H} + |s|^{2H} - |t-s|^{2H}, t, s \in \mathbb{R}^d.$$

Using Lemma 1 of Székely and Rizzo (2009), we can show that under $\mathrm{E}|\mathbf{X}|^{2h} < \infty$ and $\mathrm{E}|\mathbf{X}_\mathbf{Y}|^{2h} < \infty$, for Hurst parameters $0 < H \leq 1$, and $h = 2H$ $(0 < h \leq 2)$,

$$\mathcal{C}^2_{W_H^p}(\mathbf{X}|\mathbf{Y}) = \mathrm{E}_\mathbf{Y} \int_{\mathbb{R}^p} \frac{|f_{\mathbf{X}|\mathbf{Y}}(t) - f_\mathbf{X}(t)|^2}{\widetilde{C}(p,h)|t|^{h+p}} dt = \mathrm{E}|\mathbf{X} - \mathbf{X}'|^h - \mathrm{E}|\mathbf{X}_\mathbf{Y} - \mathbf{X}'_\mathbf{Y}|^h.$$

When $h = 1$, it is Theorem 3.1. Theory for $0 < \alpha < 2$ can be established similarly.

S4 Proofs of results in the paper

Proof of Lemma 2.1. If $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$, then $f_{\mathbf{X}|\mathbf{Y}}(t) = \mathrm{E}[e^{it^T\mathbf{X}}|\mathbf{Y}] = \mathrm{E}[e^{it^T\mathbf{X}}] = f_\mathbf{X}(t)$. Thus $\mathcal{C}^2_{w,\mathbf{Y}}(\mathbf{X}|\mathbf{Y}) = 0$, so does $\mathcal{C}^2(\mathbf{X}|\mathbf{Y})$. On the other hand, if $\mathcal{C}^2(\mathbf{X} \mid \mathbf{Y}) = 0$, then it implies that $\mathcal{C}^2_{w,\mathbf{Y}}(\mathbf{X} \mid \mathbf{Y}) = 0$ almost surely for $\mathbf{Y}$. Hence, $f_{\mathbf{X}|\mathbf{Y}}(t) = f_\mathbf{X}(t)$ almost surely for t. Let $s \in \mathbb{R}^q$, then $e^{is^T\mathbf{Y}} f_{\mathbf{X}|\mathbf{Y}}(t) = e^{is^T\mathbf{Y}} f_\mathbf{X}(t)$. Hence,

$$\mathrm{E}(e^{is^T\mathbf{Y}}\mathrm{E}[e^{it^T\mathbf{X}}|\mathbf{Y}]) = \mathrm{E}(e^{is^T\mathbf{Y}}\mathrm{E}[e^{it^T\mathbf{X}}])$$

$$\mathrm{E}[e^{is^T\mathbf{Y}}e^{it^T\mathbf{X}}] = \mathrm{E}(e^{is^T\mathbf{Y}})\mathrm{E}[e^{it^T\mathbf{X}}]$$

$$f_{\mathbf{X},\mathbf{Y}}(t,s) = f_\mathbf{X}(t)f_\mathbf{Y}(s)$$

That means, $\mathbf{X} \perp\!\!\!\perp \mathbf{Y}$. □

Proof of Theorem 2.1. 1. $\mathcal{C}^2(\mathbf{X}|\mathbf{X}) = 0$ iff $e^{it^T\mathbf{X}} = \mathrm{E}[e^{it^T\mathbf{X}}]$ almost surely for $\mathbf{X}, t$;

Note that the right hand side is constant with regards to $\mathbf{X}$. Hence, $\mathbf{X}$ must be a constant. And $\mathbf{X} = \mathrm{E}(\mathbf{X})$ almost surely. If $\mathbf{X} = \mathrm{E}(\mathbf{X})$ almost surely, the result is obvious.

2. For simplicity, in the following we omit the term $w(t)dt$ in the integrals. Note that by using the independence of $(\mathbf{W}_1, \mathbf{V}_1)$ and $(\mathbf{W}_2, \mathbf{V}_2)$, suppose $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^p$, $\mathbf{V}_1, \mathbf{V}_2 \in \mathbb{R}^q$, we have:

$$\mathcal{C}^2(\mathbf{W}_1 + \mathbf{W}_2 | \mathbf{V}_1 + \mathbf{V}_2) = \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |f_{\mathbf{W}_1+\mathbf{W}_2|\mathbf{V}_1+\mathbf{V}_2} - f_{\mathbf{W}_1+\mathbf{W}_2}|^2$$
$$= \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |\mathrm{E}[(\mathrm{E}e^{it^T\mathbf{W}_1+it^T\mathbf{W}_2}|\mathbf{V}_1, \mathbf{V}_2)|\mathbf{V}_1 + \mathbf{V}_2] - f_{\mathbf{W}_1} f_{\mathbf{W}_2}|^2.$$

Apply Propositions 4.6 and 4.5 of Cook (1998), then $\mathbf{W}_1 \perp\!\!\!\perp \mathbf{W}_2 | (\mathbf{V}_1, \mathbf{V}_2)$. Hence,

$$\cdots = \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |\mathrm{E}[(\mathrm{E}e^{it^T\mathbf{W}_1}|\mathbf{V}_1, \mathbf{V}_2)\mathrm{E}(e^{it^T\mathbf{W}_2}|\mathbf{V}_1, \mathbf{V}_2)|\mathbf{V}_1 + \mathbf{V}_2] - f_{\mathbf{W}_1} f_{\mathbf{W}_2}|^2.$$

Use $(\mathbf{W}_1, \mathbf{V}_1) \perp\!\!\!\perp \mathbf{V}_2$, we further have

$$\cdots = \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |\mathrm{E}[f_{\mathbf{W}_1|\mathbf{V}_1} f_{\mathbf{W}_2|\mathbf{V}_2} |\mathbf{V}_1 + \mathbf{V}_2] - f_{\mathbf{W}_1} f_{\mathbf{W}_2}|^2$$
$$= \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |\mathrm{E}[(f_{\mathbf{W}_1|\mathbf{V}_1} - f_{\mathbf{W}_1}) f_{\mathbf{W}_2|\mathbf{V}_2} + f_{\mathbf{W}_1} f_{\mathbf{W}_2|\mathbf{V}_2} |\mathbf{V}_1 + \mathbf{V}_2] - f_{\mathbf{W}_1} f_{\mathbf{W}_2}|^2$$
$$= \mathrm{E}_{\mathbf{V}_1+\mathbf{V}_2} \int |\mathrm{E}[(f_{\mathbf{W}_1|\mathbf{V}_1} - f_{\mathbf{W}_1}) f_{\mathbf{W}_2|\mathbf{V}_2} |\mathbf{V}_1 + \mathbf{V}_2] + f_{\mathbf{W}_1} \mathrm{E}[f_{\mathbf{W}_2|\mathbf{V}_2} - f_{\mathbf{W}_2} |\mathbf{V}_1 + \mathbf{V}_2]|^2$$

Let $a = \mathrm{E}[(f_{\mathbf{W}_1|\mathbf{V}_1} - f_{\mathbf{W}_1}) f_{\mathbf{W}_2|\mathbf{V}_2} |\mathbf{V}_1 + \mathbf{V}_2], b = f_{\mathbf{W}_1} \mathrm{E}[f_{\mathbf{W}_2|\mathbf{V}_2} - f_{\mathbf{W}_2} |\mathbf{V}_1 + \mathbf{V}_2]$,

$$\cdots = \mathrm{E} \int |a|^2 + 2\mathrm{E} \int |ab| + \mathrm{E} \int |b|^2.$$

By using Cauchy-Scwarz inequality twice $\mathrm{E}\int|ab| \leq (\mathrm{E}\int|a|^2\mathrm{E}\int|b|^2)^{1/2}$,

$$\cdots \leq ([\mathrm{E}\int|a|^2]^{1/2} + [\mathrm{E}\int|b|^2]^{1/2})^2.$$

That is,

$$\mathcal{C}(\mathbf{W}_1 + \mathbf{W}_2|\mathbf{V}_1 + \mathbf{V}_2) \leq [\mathrm{E}\int|a|^2]^{1/2} + [\mathrm{E}\int|b|^2]^{1/2}. \tag{S4.1}$$

By conditional Hölder's inequality, separately on a and b with power 2, then

$$\begin{aligned}
&\mathcal{C}(\mathbf{W}_1 + \mathbf{W}_2|\mathbf{V}_1 + \mathbf{V}_2) \\
&\leq [\mathrm{E}\int|f_{\mathbf{W}_1|\mathbf{V}_1} - f_{\mathbf{W}_1}|^2]^{1/2} + [\mathrm{E}\int|f_{\mathbf{W}_2|\mathbf{V}_2} - f_{\mathbf{W}_2}|^2]^{1/2} \\
&= \mathcal{C}(\mathbf{W}_1|\mathbf{V}_1) + \mathcal{C}(\mathbf{W}_2|\mathbf{V}_2).
\end{aligned} \tag{S4.2}$$

We can see that if (i) $\mathbf{W}_1$ and $\mathbf{V}_1$ are both constant, (ii) $\mathbf{W}_2$ and $\mathbf{V}_2$ are both constant, or (iii) $\mathbf{W}_1$, $\mathbf{V}_1$, $\mathbf{W}_2$ and $\mathbf{V}_2$ are mutually independent, then we have the equality. On the other hand, if we have the equality, then we must have equality in (S4.1) and (S4.2), which implies (i) or (ii) holds. If none of the (i) and (ii) conditions is satisfied, the equality holds only if $\mathbf{W}_1$ and $\mathbf{V}_1$, and $\mathbf{W}_2$ and $\mathbf{V}_2$ are independent, but $\mathbf{W}_1, \mathbf{V}_1$ and $\mathbf{W}_2, \mathbf{V}_2$ are already independent, so they must be mutually independent. We complete the proof.

3. This follows from item 2. above by choosing $\mathbf{W}_1 = \mathbf{V}_1 = \mathbf{X}$, and $\mathbf{W}_2 = \mathbf{V}_2 = \mathbf{Y}$. And the independence in item 2. means (i) $\mathbf{X}$ is constant; or (ii) $\mathbf{Y}$ is constant; or (iii) both of them are constant, because this is the only case when a random vector can be independent of itself.

4. Note that by definition,

$$
\begin{aligned}
\mathcal{C}^2(\mathbf{X}|\mathbf{Y}) &= \mathrm{E}_{\mathbf{Y}}[\int_{\mathbb{R}^p} |f_{\mathbf{X}|\mathbf{Y}}(t) - f_{\mathbf{X}}(t)|^2 w(t)dt] \\
&= \mathrm{E}_{\mathbf{Y}}[\int_{\mathbb{R}^p} (\mathrm{E}e^{it^T\mathbf{X}_\mathbf{Y}} - \mathrm{E}e^{it^T\mathbf{X}})(\mathrm{E}e^{-it^T\mathbf{X}_\mathbf{Y}} - \mathrm{E}e^{-it^T\mathbf{X}})w(t)dt] \\
&= \mathrm{E}_{\mathbf{y}}[\int_{\mathbb{R}^p} \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}=\mathbf{y},\mathbf{Y}'=\mathbf{y}) - \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}'=\mathbf{y}) \\
&- \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}=\mathbf{y}) + \mathrm{E}e^{it^T(\mathbf{X}-\mathbf{X}')})w(t)dt] \\
&= \mathrm{E}_{\mathbf{y}}[\int_{\mathbb{R}^p} -\{1 - \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}=\mathbf{y},\mathbf{Y}'=\mathbf{y})\} + \{1 - \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}'=\mathbf{y})\} \\
&+ \{1 - \mathrm{E}(e^{it^T(\mathbf{X}-\mathbf{X}')}|\mathbf{Y}=\mathbf{y})\} - \{1 - \mathrm{E}e^{it^T(\mathbf{X}-\mathbf{X}')}\}w(t)dt] \\
&= \mathrm{E}_{\mathbf{y}}[-\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')|(\mathbf{Y}=\mathbf{y},\mathbf{Y}'=\mathbf{y})]\}w(t)dt] \\
&+ \mathrm{E}_{\mathbf{y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')|\mathbf{Y}'=\mathbf{y}]\}w(t)dt] \\
&+ \mathrm{E}_{\mathbf{y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')|\mathbf{Y}=\mathbf{y}]\}w(t)dt] - \mathrm{E}_{\mathbf{Y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')]\}w(t)dt] \\
&\text{Note that the last three terms are equal} \\
&= \mathrm{E}_{\mathbf{Y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')]\}w(t)dt] \\
&- \mathrm{E}_{\mathbf{y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')|(\mathbf{Y}=\mathbf{y},\mathbf{Y}'=\mathbf{y})]\}w(t)dt] \\
&= \mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')]\}w(t)dt \\
&- \mathrm{E}_{\mathbf{y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')|(\mathbf{Y}=\mathbf{y},\mathbf{Y}'=\mathbf{y})]\}w(t)dt] \\
&\le \mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}-\mathbf{X}')]\}w(t)dt.
\end{aligned}
$$

However,

$$
\begin{aligned}
\mathcal{C}^2(\mathbf{X}|\mathbf{X}) &= \mathrm{E}[\mathcal{C}^2_{w,\mathbf{X}}(\mathbf{X}|\mathbf{X})] = \mathrm{E}\int |e^{it^T\mathbf{X}} - \mathrm{E}e^{it^T\mathbf{X}}|^2 w(t)dt \\
&= \mathrm{E}\int (1 - e^{it^T\mathbf{X}}\mathrm{E}e^{-it^T\mathbf{X}} - e^{-it^T\mathbf{X}}\mathrm{E}e^{it^T\mathbf{X}} + \mathrm{E}e^{it^T\mathbf{X}}\mathrm{E}e^{-it^T\mathbf{X}})w(t)dt \\
&= \int (1 - \mathrm{E}e^{it^T\mathbf{X}}\mathrm{E}e^{-it^T\mathbf{X}})w(t)dt = \mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X} - \mathbf{X}')]\}w(t)dt.
\end{aligned}
$$

Hence, conclusion follows. Consequently, $0 \le R_c \le 1$.

□

Proof of Theorem 3.1. By the proof in part 4 of Theorem 2.1 and Lemma 1 of Székely, Rizzo, and Bakirov (2007), we have

$$
\begin{aligned}
\mathcal{C}^2(\mathbf{X}|\mathbf{Y}) &= \mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X} - \mathbf{X}')]\}w(t)dt - \mathrm{E}_{\mathbf{Y}}[\mathrm{E}\int_{\mathbb{R}^p} \{1 - \cos[t^T(\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'_{\mathbf{Y}})]\}w(t)dt] \\
&= \mathrm{E}|\mathbf{X} - \mathbf{X}'| - \mathrm{E}|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'_{\mathbf{Y}}|.
\end{aligned}
$$

The last equality holds. Because $\mathrm{E}|\mathbf{X} - \mathbf{X}'| = \mathrm{E}_{\mathbf{Y}}\mathrm{E}[(|\mathbf{X} - \mathbf{X}'|)|\mathbf{Y}] = \mathrm{E}|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'|$, and hence, $\mathrm{E}|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'| = \mathrm{E}|\mathbf{X} - \mathbf{X}'|$, which immediately indicates that the first equality in (6) holds. Thus we complete the proof. □

Proof of Theorem 3.2. 1. This can be proved easily by plugging $\mathbf{X}$ for $\mathbf{Y}$ in the second formula of (6). Because, $\mathrm{E}|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'_{\mathbf{Y}}| = \mathrm{E}_{\mathbf{y}}\mathrm{E}[|\mathbf{X} - \mathbf{X}'||\mathbf{Y} = \mathbf{y}, \mathbf{Y}' = \mathbf{y}]$. If $\mathbf{X} = \mathbf{Y}$, then $\mathbf{X}' = \mathbf{Y}'$ and $\mathbf{X}' = \mathbf{Y}' = \mathbf{Y} = \mathbf{X}$. Hence, $\mathrm{E}|\mathbf{X}_{\mathbf{Y}} - \mathbf{X}'_{\mathbf{Y}}| = 0$. Or by the proof in part 4 of Theorem 2.1, and Lemma 1 in Székely, Rizzo, and Bakirov (2007) we have $\mathcal{C}^2(\mathbf{X}|\mathbf{X}) = \int (1 - \mathrm{E}e^{it^T\mathbf{X}}\mathrm{E}e^{-it^T\mathbf{X}})w(t)dt = \mathrm{E}|\mathbf{X} - \mathbf{X}'|$.

2. By using formula (6), and note that $\mathbf{B}^T\mathbf{B} = I_p$, we can prove it easily.

3. If $\mathbf{X} = \mathbf{g}(\mathbf{Y})$, for some function $\mathbf{g}$, then $\mathbf{X}_\mathbf{Y} = \mathbf{X}'_\mathbf{Y}$. Thus the second term in $\mathcal{C}^2(\mathbf{X}|\mathbf{Y})$ must be 0. Therefore, $\mathcal{C}^2(\mathbf{X}|\mathbf{Y}) = \mathcal{C}^2(\mathbf{X}|\mathbf{X})$, implying that $R_c = 1$. On the other hand, if $R_c = 1$, then the second term in $\mathcal{C}^2(\mathbf{X}|\mathbf{Y})$ must be 0, which means that almost surely for $\mathbf{Y}$, there is only one $\mathbf{X}$ corresponding to such a value of $\mathbf{Y}$. Thus, $\mathbf{X} = \mathbf{g}(\mathbf{Y})$.

□

Section 3.1: Conditional normal distribution.

$$
\begin{aligned}
\pi\mathcal{C}^2(X|Y) &= \int \mathrm{E}_Y|\mathrm{E}[e^{isX}|Y] - \mathrm{E}e^{isX}|^2 \frac{ds}{s^2} \\
&= \int \mathrm{E}_Y|e^{is\mu_y - s^2/2} - \mathrm{E}[\mathrm{E}(e^{isX}|Y)]|^2 \frac{ds}{s^2} \\
&= \int \mathrm{E}_Y|e^{is\mu_y - s^2/2} - p_0 e^{is\mu_0 - s^2/2} - p_1 e^{is\mu_1 - s^2/2}|^2 \frac{ds}{s^2} \\
&= \int \{p_0 p_1^2 |e^{is\mu_0 - s^2/2} - e^{is\mu_1 - s^2/2}|^2 + p_0^2 p_1 |e^{is\mu_1 - s^2/2} - e^{is\mu_0 - s^2/2}|^2\} \frac{ds}{s^2} \\
&= \int p_0 p_1 |e^{is\mu_0 - s^2/2} - e^{is\mu_1 - s^2/2}|^2 \frac{ds}{s^2} \\
&= \int p_0 p_1 |e^{is\mu_0} - e^{is\mu_1}|^2 e^{-s^2} \frac{ds}{s^2} \\
&= \int p_0 p_1 (2 - e^{is\Delta} - e^{-is\Delta}) e^{-s^2} \frac{ds}{s^2} \\
&= \int 2 p_0 p_1 (1 - \cos(s\Delta)) e^{-s^2} \frac{ds}{s^2} = 2 p_0 p_1 F(\Delta),
\end{aligned}
$$

where $F(\Delta) = \int (1-\cos(s\Delta))e^{-s^2}\frac{ds}{s^2}$. Note that $F(0) = 0$, and $F'(0) = 0$, but

$$F''(\Delta) = \int \cos(s\Delta)e^{-s^2}ds = \sqrt{\pi}e^{-\Delta^2/4}.$$

Thus

$$F'(Y) = \sqrt{\pi}\int_0^Y e^{-z^2/4}dz.$$

By using the function (error function, or Gaussian error function), $\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$, we have that $\int_0^Y e^{-z^2/4}dz = \sqrt{\pi}\mathrm{erf}(Y/2)$. Hence,

$$\begin{aligned} F(\Delta) &= \sqrt{\pi}\int_0^\Delta \int_0^y e^{-z^2/4}dzdy = \sqrt{\pi}\int_0^\Delta \sqrt{\pi}\mathrm{erf}(y/2)dy \\ &= \pi\int_0^\Delta \mathrm{erf}(y/2)dy = 2\pi\int_0^{\Delta/2}\mathrm{erf}(y)dy \\ &= 2\pi[\frac{\Delta}{2}\mathrm{erf}(\frac{\Delta}{2}) + \frac{e^{-\Delta^2/4}-1}{\sqrt{\pi}}], \end{aligned}$$

where, we have used the fact that $\int \mathrm{erf}(z)dz = z\mathrm{erf}(z) + \frac{e^{-z^2}}{\sqrt{\pi}}$.

Finally, $\mathcal{C}^2(X|Y) = 4p_0p_1[\frac{\Delta}{2}\mathrm{erf}(\frac{\Delta}{2}) + \frac{e^{-\Delta^2/4}-1}{\sqrt{\pi}}]$. □

Section 3.1: Bivariate normal distribution. Note that if $X \sim N(\mu_x, \sigma_x^2)$, then $\mathrm{E}(e^{isX}) = e^{is\mu_x - \frac{s^2}{2}\sigma_x^2}$, and $\mathrm{E}(e^{sX}) = e^{s\mu_x + \frac{s^2}{2}\sigma_x^2}$.

Hence, $\mathcal{C}^2(X|Y) = F(\rho)/\pi$, where

$$\begin{aligned} F(\rho) &= \int \mathrm{E}_Y|e^{is\rho Y - \frac{s^2}{2}(1-\rho^2)} - e^{-\frac{s^2}{2}}|^2\frac{ds}{s^2} = \int \mathrm{E}_Y|e^{is\rho Y + \frac{\rho^2 s^2}{2}} - 1|^2\frac{e^{-s^2}}{s^2}ds \\ &= \int \mathrm{E}_Y(e^{\rho^2 s^2} - e^{is\rho Y + \frac{\rho^2 s^2}{2}} - e^{-is\rho Y + \frac{\rho^2 s^2}{2}} + 1)\frac{e^{-s^2}}{s^2}ds \\ &= \int (e^{\rho^2 s^2} - 1)\frac{e^{-s^2}}{s^2}ds. \end{aligned}$$

By Taylor expansion, we have that $e^{\rho^2 s^2} - 1 = \sum_{n=1}^{\infty} \frac{(\rho^2 s^2)^n}{n!}$. Thus,

$$F(\rho) = \rho^2 \sum_{n=1}^{\infty} \frac{\rho^{2(n-1)}}{n!} \int s^{2(n-1)} e^{-s^2} ds = \rho^2 G(\rho).$$

Note that $G(\rho)$ is an increasing function, then

$$\pi \mathcal{C}^2(X|Y) = F(\rho) \leq F(1) = \pi \mathcal{C}^2(X|X).$$

In addition, $F(0) = 0, F'(0) = 0$. Simple calculation shows that $F'(\rho) = \frac{2\rho\sqrt{\pi}}{\sqrt{1-\rho^2}}$. Therefore, we have $F(\rho) = \int_0^{\rho} \frac{2z\sqrt{\pi}}{\sqrt{1-z^2}} dz = 2\sqrt{\pi}(1 - \sqrt{1-\rho^2})$, And we have:

$$\mathcal{C}^2(X|Y) = \frac{2}{\sqrt{\pi}}(1 - \sqrt{1-\rho^2}).$$

□

Section 3.1: Conditional binomial distribution. Note that if $X|Y \sim Bin(n, q_Y)$, where $Y \in \{0, 1\}$, then we have that

$$\mathcal{C}^2(X|Y) = \int \mathrm{E}_Y |\mathrm{E}[e^{itX}|Y] - \mathrm{E}e^{itX}|^2 w(t) dt$$

$$= p_0 p_1 \int |(q_0 e^{it} + 1 - q_0)^n - (q_1 e^{it} + 1 - q_1)^n|^2 w(t) dt$$

$$= p_0 p_1 \int |\sum_{k=0}^{n} c_n^k q_0^k e^{ikt} (1-q_0)^{n-k} - \sum_{k=0}^{n} c_n^k q_1^k e^{ikt} (1-q_1)^{n-k}|^2 w(t) dt$$

$$= p_0 p_1 \int |\sum_{k=0}^{n} c_n^k e^{ikt} [q_0^k (1-q_0)^{n-k} - q_1^k (1-q_1)^{n-k}]|^2 w(t) dt$$

$$= p_0 p_1 \int \{\sum_{k=0}^{n} c_n^k e^{ikt} [q_0^k (1-q_0)^{n-k} - q_1^k (1-q_1)^{n-k}]\}$$

$$\times \{\sum_{l=0}^{n} c_n^l e^{-ilt} [q_0^l (1-q_0)^{n-l} - q_1^l (1-q_1)^{n-l}]\} w(t) dt$$

$$= p_0p_1 \int \{ \sum_{k,l=0}^{n} c_n^k c_n^l [q_0^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}][q_0^l(1-q_0)^{n-l} - q_1^l(1-q_1)^{n-l}] e^{it(k-l)} \} w(t)dt$$

$$= p_0p_1 \int \{ \sum_{k,l=0}^{n} c_n^k c_n^l [q_0^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}][q_0^l(1-q_0)^{n-l} - q_1^l(1-q_1)^{n-l}]$$

$$\times [(e^{it(k-l)} - 1) + 1] \} w(t)dt$$

$$= -p_0p_1 \{ \sum_{k,l=0}^{n} c_n^k c_n^l [q_0^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}][q_0^l(1-q_0)^{n-l} - q_1^l(1-q_1)^{n-l}] |k-l| \} + 0$$

$$= -p_0p_1 \{ \sum_{k,l=0}^{n} c_n^k c_n^l [q_0^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}][q_0^l(1-q_0)^{n-l} - q_1^l(1-q_1)^{n-l}] |k-l| \}.$$

Now consider

$$q_0^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k} = (q_0 - q_1 + q_1)^k(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}$$

$$= \sum_{i=0}^{k} c_k^i (q_0 - q_1)^i q_1^{k-i}(1-q_0)^{n-k} - q_1^k(1-q_1)^{n-k}$$

$$= (q_0 - q_1) \sum_{i=1}^{k} c_k^i (q_0 - q_1)^{i-1} q_1^{k-i}(1-q_1)^{n-k} + q_1^k[(1-q_0)^{n-k} - (1-q_1)^{n-k}]$$

$$= (q_0 - q_1) \sum_{i=1}^{k} c_k^i (q_0 - q_1)^{i-1} q_1^{k-i}(1-q_1)^{n-k} + q_1^k(q_1 - q_0) \sum_{i=1}^{n-k} (1-q_0)^{n-k-i}(1-q_1)^{i-1}$$

$$= (q_0 - q_1)[\sum_{i=1}^{k} c_k^i (q_0 - q_1)^{i-1} q_1^{k-i}(1-q_1)^{n-k} - q_1^k \sum_{i=1}^{n-k} (1-q_0)^{n-k-i}(1-q_1)^{i-1}]$$

Therefore,

$$\mathcal{C}^2(X|Y)$$
$$= -p_0p_1(q_0-q_1)^2\{\sum_{k,l=0}^{n} c_n^k c_n^l [\sum_{i=1}^{k} c_k^i (q_0-q_1)^{i-1} q_1^{k-i}(1-q_1)^{n-k} - q_1^k \sum_{i=1}^{n-k}(1-q_0)^{n-k-i}(1-q_1)^{i-1}]$$
$$\times [\sum_{i=1}^{l} c_l^i (q_0-q_1)^{i-1} q_1^{l-i}(1-q_1)^{n-l} - q_1^l \sum_{i=1}^{n-l}(1-q_0)^{n-l-i}(1-q_1)^{i-1}]|k-l|\}.$$

When $n=1$, then $\mathcal{C}^2(X|Y) = 2p_0p_1(q_0-q_1)^2$; and when $n=2$, then $\mathcal{C}^2(X|Y) = 4p_0p_1(q_0-q_1)^2[1+(q_0+q_1-1)^2]$. □

Section 3.1: Conditional Cauchy distribution. Note that $q_0, q_1 > 0$, and without loss of generality we assume that $q_1 \geq q_0$. Define a function $E_i(x) = \int_{-\infty}^{x} \frac{e^s}{s} ds$, and integral is taken in the principal as ϵ to ϵ^{-1} when $\epsilon \to 0$. We then have,

$$\begin{aligned}\mathcal{C}^2(X|Y) &= \int \mathrm{E}_Y|\mathrm{E}[e^{itX}|Y] - \mathrm{E}e^{itX}|^2 w(t)dt \\ &= \frac{p_0p_1}{\pi}\int |e^{-q_0|t|} - e^{-q_1|t|}|^2 \frac{dt}{t^2} \\ &= \frac{2p_0p_1}{\pi}\int_0^{+\infty} [e^{-2q_0t} - 2e^{-(q_0+q_1)t} + e^{-2q_1t}]\frac{dt}{t^2}\end{aligned}$$

Look at

$$\mathcal{C}^2(X|Y;\epsilon) = \frac{2p_0p_1}{\pi}\int_{\epsilon}^{\epsilon^{-1}} [e^{-2q_0t} - 2e^{-(q_0+q_1)t} + e^{-2q_1t}]\frac{dt}{t^2}$$
$$= \frac{2p_0p_1}{\pi}\int_{\epsilon}^{\epsilon^{-1}} [e^{-2q_0t} - 2e^{-(q_0+q_1)t} + e^{-2q_1t}]\frac{dt}{t^2}$$

Now by using 1.3.2.20 and 1.3.2.12 of Prudnikov, Brychkov, and Marichev (1986), we have that

$$\mathcal{C}^2(X|Y;\epsilon) = \frac{2p_0p_1}{\pi}\int_{\epsilon}^{\epsilon^{-1}} [e^{-2q_0t} - 2e^{-(q_0+q_1)t} + e^{-2q_1t}]\frac{dt}{t^2}$$
$$= \frac{2p_0p_1}{\pi}[-\frac{e^{-2q_0t}}{t} - 2q_0E_i(-2q_0t) - 2(-\frac{e^{-(q_0+q_1)t}}{t} - (q_0+q_1)E_i(-(q_0+q_1)t))$$
$$- \frac{e^{-2q_1t}}{t} - 2q_1E_i(-2q_1t)]|_{\epsilon}^{\epsilon^{-1}}$$
$$= \frac{2p_0p_1}{\pi}[-\frac{e^{-2q_0t}}{t} + 2\frac{e^{-(q_0+q_1)t}}{t} - \frac{e^{-2q_1t}}{t}$$
$$- 2q_0E_i(-2q_0t) + 2(q_0+q_1)E_i(-(q_0+q_1)t) - 2q_1E_i(-2q_1t)]|_{\epsilon}^{\epsilon^{-1}}$$

But $[-\frac{e^{-2q_0t}}{t} + 2\frac{e^{-(q_0+q_1)t}}{t} - \frac{e^{-2q_1t}}{t}]|_{\epsilon}^{\epsilon^{-1}} \to 0$ as $\epsilon \to 0$. Thus, as $\epsilon \to 0$ we can have

$$\mathcal{C}^2(X|Y;\epsilon) = \frac{2p_0p_1}{\pi}[-2q_0E_i(-2q_0t) + 2(q_0+q_1)E_i(-(q_0+q_1)t) - 2q_1E_i(-2q_1t)]|_{\epsilon}^{\epsilon^{-1}}$$
$$= \frac{2p_0p_1}{\pi}[-2q_0\int_{-2q_0\epsilon}^{-2q_0\epsilon^{-1}} \frac{e^t}{t}dt + 2(q_0+q_1)\int_{-(q_0+q_1)\epsilon}^{-(q_0+q_1)\epsilon^{-1}} \frac{e^t}{t}dt - 2q_1\int_{-2q_1\epsilon}^{-2q_1\epsilon^{-1}} \frac{e^t}{t}dt]$$
$$= \frac{2p_0p_1}{\pi}[2q_0\int_{2q_0\epsilon^{-1}}^{(q_0+q_1)\epsilon^{-1}} \frac{e^{-t}}{t}dt - 2q_0\int_{2q_0\epsilon}^{(q_0+q_1)\epsilon} \frac{e^{-t}}{t}dt$$
$$- 2q_1\int_{(q_0+q_1)\epsilon^{-1}}^{2q_1\epsilon^{-1}} \frac{e^{-t}}{t}dt + 2q_1\int_{(q_0+q_1)\epsilon}^{2q_1\epsilon} \frac{e^{-t}}{t}dt]$$
$$= \frac{2p_0p_1}{\pi}[2q_0A_1 - 2q_0B_1 - 2q_1A_2 + 2q_1B_2].$$

But $A_1 = \int_{2q_0}^{q_0+q_1} e^{-y\epsilon^{-1}}y^{-1}dy \leq (2q_0)^{-1}(q_1 - q_01)e^{-2q_0\epsilon^{-1}} \to 0$ as $\epsilon \to 0$. Similarly, $A_2 \to 0$ as $\epsilon \to 0$. Now by using 1.3.2.13 of Prudnikov, Brychkov, and Marichev (1986), we have

$$\begin{aligned} B_1 &= \ln[(q_0+q_1)\epsilon] + \sum_{k=1}^{\infty} \frac{(-(q_0+q_1)\epsilon)^k}{k!k} - \ln(2q_0\epsilon) - \sum_{k=1}^{\infty} \frac{(-2q_0\epsilon)^k}{k!k} \\ &= \ln \frac{q_0+q_1}{2q_0} + \sum_{k=1}^{\infty} \frac{(-(q_0+q_1)\epsilon)^k - (-2q_0\epsilon)^k}{k!k} \\ &= \ln \frac{q_0+q_1}{2q_0} + \sum_{k=1}^{\infty} \frac{(-(q_0+q_1))^k - (-2q_0)^k}{k!k}\epsilon^k = \ln \frac{q_0+q_1}{2q_0} \text{ as } \epsilon \to 0. \end{aligned}$$

While by similar argument, we have $B_2 = \ln \frac{2q_1}{q_0+q_1}$ as $\epsilon \to 0$. Therefore,

$$\mathcal{C}^2(X|Y) = \lim_{\epsilon\to 0} \mathcal{C}^2(X|Y;\epsilon) = \frac{4p_0p_1}{\pi}(q_0 \ln \frac{2q_0}{q_0+q_1} + q_1 \ln \frac{q_1}{q_0+q_1}).$$

Note that $\mathcal{C}^2(X|Y) \geq 0$, and it is 0 if $q_1 = q_0$; However, $\mathcal{C}^2(X|Y) \geq 0$ increases as $q_1 > q_0$; decreases as $q_1 < q_0$. Thus $\mathcal{C}^2(X|Y) = 0$ iff $q_1 = q_0$. □

Proof of Theorem 4.1. Following Székely, Rizzo, and Bakirov (2007), we have that

$$\begin{aligned} f^n_{\mathbf{X}|y}(t)\overline{f^n_{\mathbf{X}|y}(t)} &= \frac{1}{n_y^2}\sum_{k_y,l_y=1}^{n_y,n_y} \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}) + v_1 \\ f^n_{\mathbf{X}|y}(t)\overline{f^n_{\mathbf{X}}(t)} &= \frac{1}{nn_y}\sum_{y'=1}^{H}\sum_{k_y,l_{y'}=1}^{n_y,n_{y'}} \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}) + v_2 \\ f^n_{\mathbf{X}}(t)\overline{f^n_{\mathbf{X}}(t)} &= \frac{1}{n^2}\sum_{y,y'=1}^{H,H}\sum_{k_y,l_{y'}=1}^{n_y,n_{y'}} \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}) + v_3, \end{aligned}$$

where v_1, v_2 and v_3 vanish when integral is evaluated. Since

$$\cos t^T(\mathbf{X}_k - \mathbf{X}_l) = 1-(1-\cos t^T(\mathbf{X}_k - \mathbf{X}_l)), \text{ and } \int [1-\cos t^T(\mathbf{X}_k - \mathbf{X}_l)]w(t)dt = |\mathbf{X}_k - \mathbf{X}_l|,$$

by choosing $k = y, k_y$ and $l = y, l_y$, we have

$$\cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}) = 1 - (1 - \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}))$$

$$\text{and} \quad \int [1 - \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y})] w(t) dt = |\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}|;$$

by choosing $k = y, k_y$ and $l = y', l_{y'}$, we have

$$\cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}) = 1 - (1 - \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}))$$
$$\text{and} \quad \int [1 - \cos t^T(\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}})] w(t) dt = |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}|.$$

We also have

$$|f^n_{\mathbf{X}|y}(t) - f^n_{\mathbf{X}}(t)|^2 = f^n_{\mathbf{X}|y}(t)\overline{f^n_{\mathbf{X}|y}(t)} - f^n_{\mathbf{X}|y}(t)\overline{f^n_{\mathbf{X}}(t)} - \overline{f^n_{\mathbf{X}|y}(t)} f^n_{\mathbf{X}}(t) + f^n_{\mathbf{X}}(t)\overline{f^n_{\mathbf{X}}(t)}.$$

Therefore,

$$\begin{aligned}
\mathcal{C}^2_{w,y,n}(\mathbf{X}|Y = y) &= ||f^n_{\mathbf{X}|y}(t) - f^n_{\mathbf{X}}(t)||^2 \\
&= \frac{2}{n n_y} \sum_{y'=1}^{H} \sum_{k_y, l_{y'}=1}^{n_y, n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}| - \frac{1}{n_y^2} \sum_{k_y, l_y=1}^{n_y, n_y} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}| \\
&- \frac{1}{n^2} \sum_{y,y'=1}^{H,H} \sum_{k_y, l_{y'}=1}^{n_y, n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}|.
\end{aligned}$$

And thus, we have

$$\begin{aligned}
\mathcal{C}^2_n(\mathbf{X}|Y) &= \sum_{y=1}^{H} p_y \mathcal{C}^2_{w,y,n}(\mathbf{X}|Y = y) \\
&= \frac{2}{n^2} \sum_{y=1}^{H} \sum_{y'=1}^{H} \sum_{k_y, l_{y'}=1}^{n_y, n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}| - \frac{1}{n} \sum_{y=1}^{H} \frac{1}{n_y} \sum_{k_y, l_y=1}^{n_y, n_y} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}| \\
&- \frac{1}{n^2} \sum_{y,y'=1}^{H,H} \sum_{k_y, l_{y'}=1}^{n_y, n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}| \\
&= \frac{1}{n^2} \sum_{y,y'=1}^{H,H} \sum_{k_y, l_{y'}=1}^{n_y, n_{y'}} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y',l_{y'}}| - \frac{1}{n} \sum_{y=1}^{H} \frac{1}{n_y} \sum_{k_y, l_y=1}^{n_y, n_y} |\mathbf{X}_{y,k_y} - \mathbf{X}_{y,l_y}|.
\end{aligned}$$

Note that the summation in the first and third term after the second equality sign are the same. We complete the proof. $\square$

Proof of Lemma 4.2. This can follow from Theorem 2 of Székely, Rizzo, and Bakirov (2007) and Theorem 3 of Shao and Zhang (2014). By applying SLLN of V-statistic to achieve the conclusion.

Note that let $\xi_{n,y}(t) = f^n_{\mathbf{X}|y}(t) - f^n_{\mathbf{X}}(t)$, then $\mathcal{C}^2_{w,y,n}(\mathbf{X}|y) = ||\xi_{n,y}(t)||^2$. Hence, by (7), we have $\mathcal{C}^2_n(\mathbf{X}|Y) = \mathrm{E}_Y \mathcal{C}^2_{w,y,n}(\mathbf{X}|Y) = \mathrm{E}_Y ||\xi_{n,Y}||^2 = \sum_{y=1}^H p_y ||f^n_{\mathbf{X}|y}(t) - f^n_{\mathbf{X}}(t)||^2$.

Define $\xi_y(t) = f_{\mathbf{X}|y}(t) - f_{\mathbf{X}}(t)$, and let $u_{y,k_y} = \exp(it^T\mathbf{X}_{y,k_y}) - f_{\mathbf{X}|y}(t)$ and $v_{y,k_y} = \exp(it^T\mathbf{X}_{y,k_y}) - f_{\mathbf{X}}(t)$. Then,

$$\xi_{n,y}(t) = \frac{1}{n_y}\sum_{k_y=1}^{n_y} u_{y,k_y} - \frac{1}{n}\sum_{y=1}^{H}\sum_{k_y=1}^{n_y} v_{y,k_y} + \xi_y(t).$$

In integrals, we can use the symbol $d\omega$, which is defined by $d\omega = w(t)dt$, where $w(t)$ is defined previously. Define the region $D(\delta) = \{t : \delta \le |t|_p \le 1/\delta\}$, for each $\delta > 0$, and the random variables

$$\mathcal{C}^2_{w,y,n,\delta}(\mathbf{X}|y) = \int_{D(\delta)} |\xi_{n,y}(t)|^2 d\omega.$$

For any fixed δ, the weight function $w(t)$ is bounded on $D(\delta)$. Hence, $\mathcal{C}^2_{w,y,n,\delta}(\mathbf{X}|y)$ is a combination of $V-$statistics with finite expectation. By the SLLN for $V-$statistics, it follows that almost surely

$$\lim_{n\to\infty} \mathcal{C}^2_{w,y,n,\delta}(\mathbf{X}|y) = \mathcal{C}^2_{w,y,\cdot,\delta}(\mathbf{X}|y) = \int_{D(\delta)} |\xi_y(t)|^2 d\omega.$$

Clearly $\mathcal{C}^2_{w,y,\cdot,\delta}(\mathbf{X}|y)$ converges to $\mathcal{C}^2_{w,y}(\mathbf{X}|y)$ as $\delta \to 0$. Therefore, it remains to prove that almost surely

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}|\mathcal{C}^2_{w,y,n}(\mathbf{X}|y) - \mathcal{C}^2_{w,y,n,\delta}(\mathbf{X}|y)| = 0.$$

For each $\delta > 0$,

$$|\mathcal{C}^2_{w,y,n}(\mathbf{X}|y) - \mathcal{C}^2_{w,y,n,\delta}(\mathbf{X}|y)| = \int_{|t|<\delta}|\xi_{n,y}(t)|^2 d\omega + \int_{|t|>\frac{1}{\delta}}|\xi_{n,y}(t)|^2 d\omega. \tag{S4.3}$$

For $z = (z_1, \cdots, z_p)^T \in \mathbb{R}^p$, define the function

$$G(s) = \int_{|z|<s}\frac{1-\cos z_1}{|z|^{1+p}}dz.$$

By Lemma 1 of Székely, Rizzo, and Bakirov (2007), clearly $G(s)$ is bounded by $\tilde{c}_p$ and $\lim_{s\to 0} G(s) = 0$. Using the inequality $|a+b+c|^2 \le 3(|a|^2+|b|^2+|c|^2)$, and applying Cauchy-Schwarz inequaltiy, we have that

$$\begin{aligned}
|\xi_{n,y}(t)|^2 &\le 3(|\frac{1}{n_y}\sum_{k_y=1}^{n_y}u_{y,k_y}|^2 + |\frac{1}{n}\sum_{y=1}^{H}\sum_{k_y=1}^{n_y}v_{y,k_y}|^2 + |\xi_y(t)|^2) \\
&\le 3(\frac{1}{n_y}\sum_{k_y=1}^{n_y}|u_{y,k_y}|^2 + \frac{1}{n}\sum_{y=1}^{H}\sum_{k_y=1}^{n_y}|v_{y,k_y}|^2 + |\xi_y(t)|^2).
\end{aligned} \tag{S4.4}$$

After a suitable change of variables, we have

$$\begin{aligned}
\int_{|t|<\delta}\frac{|u_{y,k_y}|^2}{\tilde{c}_p|t|^{1+p}}dt &\le 2\mathrm{E}_{\mathbf{X}|y}|\mathbf{X}-\mathbf{X}_{y,k_y}|G_{|y}(|\mathbf{X}-\mathbf{X}_{y,k_y}|\delta) \\
\int_{|t|<\delta}\frac{|v_{y,k_y}|^2}{\tilde{c}_p|t|^{1+p}}dt &\le 2\mathrm{E}_{\mathbf{X}}|\mathbf{X}-\mathbf{X}_{y,k_y}|G(|\mathbf{X}-\mathbf{X}_{y,k_y}|\delta)
\end{aligned}$$

Therefore, we have

$$\int_{|t|<\delta} |\xi_{n,y}(t)|^2 d\omega \leq \frac{6}{n_y} \sum_{k_y=1}^{n_y} \mathrm{E}_{\mathbf{X}|y} |\mathbf{X} - \mathbf{X}_{y,k_y}| G_{|y}(|\mathbf{X} - \mathbf{X}_{y,k_y}|\delta) + \frac{6}{n} \sum_{y=1}^{H} \sum_{k_y=1}^{n_y} \mathrm{E}_{\mathbf{X}} |\mathbf{X} - \mathbf{X}_{y,k_y}| G(|\mathbf{X} - \mathbf{X}_{y,k_y}|\delta) + 3 \int_{|t|<\delta} |\xi_y(t)|^2 d\omega$$

By the SLLN, then

$$\limsup_{n\to\infty} \int_{|t|<\delta} |\xi_{n,y}(t)|^2 d\omega \leq 6\mathrm{E}_{|y}(|\mathbf{X} - \mathbf{X}'|) G_{|y}(|\mathbf{X} - \mathbf{X}'|\delta) + 6\mathrm{E}(|\mathbf{X} - \mathbf{X}'|) G(|\mathbf{X} - \mathbf{X}'|\delta) + 3 \int_{|t|<\delta} |\xi_y(t)|^2 d\omega$$

By the Lebesgue Dominated Convergence theorem, we then have

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \int_{|t|<\delta} |\xi_{n,y}(t)|^2 d\omega = 0, \text{ almost surely.}$$

Now consider the second term in equation (S4.3), using the fact that $|u_{y,k_y}|^2, |v_{y,k_y}|^2, |\xi_y(t)|^2 \leq 4$ and the inequality (S4.4) implies that $|\xi_{n,y}(t)|^2 \leq 36$. Hence,

$$\int_{|t|>\frac{1}{\delta}} |\xi_{n,y}(t)|^2 d\omega \leq 36 \int_{|t|>\frac{1}{\delta}} \frac{1}{\tilde{c}_p |t|^{1+p}} dt = 36 h(\delta).$$

But $h(\delta)$ goes to zero as $\delta \to 0$. That means $\mathcal{C}^2_{w,y,n}(\mathbf{X}|y) \to \mathcal{C}^2_{w,y}(\mathbf{X}|y)$ almost surely, for any given y. And the conclusion then follows. □

Proof of Theorem 4.2. The argument is very similar to that presented in the proofs of Theorem 5 and Corollary 2 of Székely, Rizzo, and Bakirov (2007) and that of Theorem 4 of Shao and Zhang (2014). Let $\Gamma(\cdot)$ denote a complex-valued zero-mean Gaussian random process with covariance function $\mathrm{cov}_\Gamma(s, s_0) = [f_{\mathbf{X}}(s - s_0) - f_{\mathbf{X}}(s)\overline{f_{\mathbf{X}}(s_0)}]$, where

$s, s_0 \in \mathbb{R}^p$. Note that $f_{\mathbf{X}|Y}(s) = \mathrm{E}(e^{is\mathbf{X}}|Y), f_{\mathbf{X}}(s) = \mathrm{E}(e^{is\mathbf{X}})$, $n_y = [p_y n]$, where n_y is the number of observations in $Y \in y$, $y = 1, 2, ...H$ and $\sum_{y=1}^{H} n_y = n$. And $f_{\mathbf{X}}(s) = \mathrm{E}_Y f_{\mathbf{X}|Y}(s) = \sum_Y p_Y f_{\mathbf{X}|Y}(s)$, where $p_Y = P(y \in Y)$.

1. Define the empirical process

$$\Gamma_{n,y}(s) = \sqrt{n_y}[f^n_{\mathbf{X}|y}(s) - f^n_{\mathbf{X}}(s)].$$

Under independence hypothesis, $\mathrm{E}_{\mathbf{X}|y}[\Gamma_{n,y}(s)] = 0$ and $\mathrm{E}_{\mathbf{X}|y}[\Gamma_{n,y}(s)\overline{\Gamma_{n,y}(s_0)}] = (1 - \frac{n_y}{n})[f_{\mathbf{X}}(s - s_0) - f_{\mathbf{X}}(s)\overline{f_{\mathbf{X}}(s_0)}] = (1 - \frac{n_y}{n})\mathrm{cov}_{\Gamma}(s, s_0)$. In particular, $\mathrm{E}_{\mathbf{X}|y}|\Gamma_{n,y}(s)|^2 = (1 - \frac{n_y}{n})[1 - |f_{\mathbf{X}}(s)|^2] \le 1$.

Note that $n\mathcal{C}^2_n(\mathbf{X}|Y) = \sum_{y=1}^{H} ||\Gamma_{n,y}(s)||^2$.

For each $\delta > 0$, define the region $D(\delta) = \{s : \delta \le |s|_p < 1/\delta\}$. For each δ we construct a sequence of random variables $\{Q_n(\delta)\}$ such that

- (i) $Q_n(\delta) \xrightarrow{D} Q(\delta)$ for each $\delta > 0$;
- (ii) $\limsup_{n\to\infty} \mathrm{E}|Q_n(\delta) - ||\Gamma_n(s)||^2| \to 0$ as $\delta \to 0$; $\Gamma^T_n(s) = (\Gamma_{n,1}(s), \cdots, \Gamma_{n,H}(s))$.
- (iii) $\mathrm{E}|Q(\delta) - \mathcal{C}^2(\mathbf{X}|\mathbf{X})Q| \to 0$ as $\delta \to 0$.

Then the weak convergence follows from Theorem 8.6.2 of Resnick (1999). Therefore,

$$n\mathcal{C}^2_n(\mathbf{X}|Y) = \sum_{y=1}^{H} ||\Gamma_{n,y}(s)||^2 \Rightarrow \mathcal{C}^2(\mathbf{X}|\mathbf{X})Q.$$

Following the construction in Shao and Zhang (2014) and Székely, Rizzo, and Bakirov (2007), we define

$$Q_n(\delta) = \int_{D(\delta)} |\Gamma_n(s)|^2 d\omega \text{ and } Q(\delta) = \int_{D(\delta)} |A(s)|^2 d\omega,$$

where $A(s)$ is the limt of $\Gamma_n(s)$, and $|A(s)|^2 \sim (1 - f_{\mathbf{X}}(s)^2)Q$.

Given $\epsilon = 1/q > 0$, $q \in N$, choose a partition $\{D_k\}_{k=1}^N$ of $D(\delta)$ into $N = N(\epsilon)$ measurable sets with diameter at most ϵ. Then $Q_n(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma_n(s)|^2 d\omega$ and $Q(\delta) = \sum_{k=1}^N \int_{D_k} |A(s)|^2 d\omega$.

Define $Q_n^q(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma_n(s_0(k))|^2 d\omega$ and $Q^q(\delta) = \sum_{k=1}^N \int_{D_k} |A(s_0(k))|^2 d\omega$, where $\{s_0(k)\}_{k=1}^N$ is a set of distinct points such that $s_0(k) \in D_k$. By multivariate CLT and continuous mapping theorem, $Q_n^q(\delta) \xrightarrow{D} Q^q(\delta)$, for any $q \in N$. Note that under independence of slicing and under dependence of slicing, the two resulting $A(s)$ are different but distributions of $|A(s)|^2$ are the same (see **Remark 1** and **Remark 2** below). Thus based on Theorem 8.6.2 of Resnick (1999), (i) holds if we can show that

$$\limsup_{q\to\infty} \mathrm{E}|Q^q(\delta) - Q(\delta)| = 0, \tag{S4.5}$$

and

$$\limsup_{q\to\infty} \limsup_{n\to\infty} \mathrm{E}|Q_n^q(\delta) - Q_n(\delta)| = 0. \tag{S4.6}$$

Let $\beta_{n,y}(\epsilon) = \sup_{s,s_0} \mathrm{E}||\Gamma_{n,y}(s)|^2 - |\Gamma_{n,y}(s_0)|^2|$ and $\beta(\epsilon) = \sup_{s,s_0} \mathrm{E}||\Gamma(s)|^2 - |\Gamma(s_0)|^2|$, where the supremum is taken over all s and s_0, under the restrictions: $\delta <$

$|s|_p, |s_0|_p < 1/\delta$ and $|s - s_0|_p < \epsilon$.

$$\begin{aligned}
\beta(\epsilon) &= \sup_{s,s_0} \mathrm{E}||\Gamma(s)|^2 - |\Gamma(s_0)|^2| \\
&= \sup_{s,s_0} \mathrm{E}|(\Gamma(s) - \Gamma(s_0))\overline{\Gamma(s)} + \Gamma(s_0)(\overline{\Gamma(s)} - \overline{\Gamma(s_0)})| \\
&\le \sup_{s,s_0} \mathrm{E}^{1/2}|\Gamma(s) - \Gamma(s_0)|^2(\mathrm{E}^{1/2}|\Gamma(s)|^2 + \mathrm{E}^{1/2}|\Gamma(s_0)|^2) \\
&\le 2\sup_{s,s_0} \mathrm{E}^{1/2}|\Gamma(s) - \Gamma(s_0)|^2 \\
&= 2\sup_{s,s_0} |\mathrm{cov}_\Gamma(s,s) - \mathrm{cov}_\Gamma(s,s_0) - \mathrm{cov}_\Gamma(s_0,s) + \mathrm{cov}_\Gamma(s_0.s_0)|^{1/2}.
\end{aligned}$$

Since $f_{\mathbf{X}}(s)$ is uniform continuous in $s \in \mathbb{R}^p$, it is clear that $\beta(\epsilon) \to 0$ as $\epsilon \to 0$.

To show (S4.5), note that $A(s)$ is the limt of $\Gamma_n(s)$,

$$\begin{aligned}
\mathrm{E}|Q^q(\delta) - Q(\delta)| &= \mathrm{E}\left|\sum_{k=1}^{N}\int_{D_k}|A(s_0(k))|^2 d\omega - \int_{D(\delta)}|A(s)|^2 d\omega\right| \\
&= \mathrm{E}\left|\sum_{k=1}^{N}\int_{D_k}(|A(s_0(k))|^2 - |A(s)|^2)d\omega\right| \\
&\le 2\beta(1/q)\int_{D(\delta)} w(s)ds \to 0 \text{ as } q \to \infty.
\end{aligned}$$

By using the same argument, we can show that (S4.6) holds as well. Hence, (i) is true. To prove (ii), note that

$$\mathrm{E}|\int_{D(\delta)}|\Gamma_n(s)|^2 d\omega - \int_{\mathbb{R}^p}|\Gamma_n(s)|^2 d\omega| = \int_{|s|<\delta}\mathrm{E}|\Gamma_n(s)|^2 d\omega + \int_{|s|>1/\delta}\mathrm{E}|\Gamma_n(s)|^2 d\omega.$$

By noting that $\mathrm{E}_{\mathbf{X}|Y}|\Gamma_{n,y}(s)|^2 = (1 - \frac{n_y}{n})[1 - |f_{\mathbf{X}}(s)|^2]$ and following from the proof of Lemma 4.2, we have that

$$\int_{|s|<\delta} \mathrm{E}|\Gamma_{n,y}(s)|^2 d\omega \leq (1 - \frac{n_y}{n})\mathrm{E}|\mathbf{X} - \mathbf{X}'|G(|\mathbf{X} - \mathbf{X}'|\delta).$$

The fact $\mathrm{E}_{\mathbf{X}|Y}|\Gamma_{n,y}(s)|^2 \leq 1$ implies that

$$\int_{|s|>1/\delta} \mathrm{E}|\Gamma_{n,y}(s)|^2 d\omega \leq h(\delta),$$

where $h(\delta)$ is defined in Lemma 4.2 and goes to zero as $\delta \to 0$. Thus (ii) holds.

Applying a similar argument, (iii) holds. Thus we complete the proof of (1).

2. This can easily follow from Corollary 2 of Székely, Rizzo, and Bakirov (2007) and see Theorem 4 of Shao and Zhang (2014) as well.

   Based on (1), $n\mathcal{C}_n^2(\mathbf{X}|Y) \xrightarrow[n\to\infty]{D} \mathcal{C}^2(\mathbf{X}|\mathbf{X})Q$. By the SLLN for $V-$statistics, as $n \to \infty$, $\mathcal{C}_n^2(\mathbf{X}|\mathbf{X}) \to \mathcal{C}^2(\mathbf{X}|\mathbf{X})|$, almost surely. Therefore,

$$n\mathcal{C}_n^2(\mathbf{X}|Y)/\mathcal{C}_n^2(\mathbf{X}|\mathbf{X}) \xrightarrow[n\to\infty]{D} Q.$$

3. If $\mathbf{X}$ and Y are dependent, then $\mathcal{C}^2(\mathbf{X}|Y) > 0$. Lemma 4.2 implies that when for large n, $\mathcal{C}_n^2(\mathbf{X}|Y) > 0$, and thus $n\mathcal{C}_n^2(\mathbf{X}|Y) \to \infty$ as $n \to \infty$. By the SLLN, $\mathcal{C}_n^2(\mathbf{X}|\mathbf{X})$ converges to a constant and therefore, as $n \to \infty$, $n\mathcal{C}_n^2(\mathbf{X}|Y)/\mathcal{C}_n^2(\mathbf{X}|\mathbf{X}) \to \infty$.

□

Remark 1: Proof of Theorem 4.2 under independence of slicing. Since Y is sliced,

$$\begin{aligned}\Gamma_{n,y}(s) &= \sqrt{n_y}\left[f^n_{\mathbf{X}|y}(s) - f^n_{\mathbf{X}}(s)\right] \\ &= \sqrt{n_y}\left[\frac{1}{n_y}\sum_{k_y=1}^{n_y} e^{is^T\mathbf{X}_{y,k_y}} - \frac{1}{n}\sum_{y=1}^{H}\sum_{k_y=1}^{n_y} e^{is^T\mathbf{X}_{y,k_y}}\right].\end{aligned}$$

Then $n\mathcal{C}^2_n(\mathbf{X}|Y) = \sum_{y=1}^{H} ||\Gamma_{n,y}(s)||^2$. And note that

$$\begin{aligned}\begin{pmatrix}\Gamma_{n,1}(s)\\ \Gamma_{n,2}(s)\\ \vdots\\ \Gamma_{n,H}(s)\end{pmatrix} &= \begin{pmatrix}1-\frac{n_1}{n} & -\frac{\sqrt{n_1n_2}}{n} & \cdots & -\frac{\sqrt{n_1n_H}}{n}\\ -\frac{\sqrt{n_1n_2}}{n} & 1-\frac{n_2}{n} & \cdots & -\frac{\sqrt{n_2n_H}}{n}\\ \vdots & \vdots & & \vdots\\ -\frac{\sqrt{n_1n_H}}{n} & -\frac{\sqrt{n_2n_H}}{n} & \cdots & 1-\frac{n_H}{n}\end{pmatrix}\begin{pmatrix}\sqrt{n_1}\left(\frac{1}{n_1}\sum_{k_1=1}^{n_1} e^{is^TX_{1,k_1}} - f_{\mathbf{x}}(s)\right)\\ \sqrt{n_2}\left(\frac{1}{n_2}\sum_{k_2=1}^{n_2} e^{is^TX_{2,k_2}} - f_{\mathbf{x}}(s)\right)\\ \vdots\\ \sqrt{n_H}\left(\frac{1}{n_H}\sum_{k_H=1}^{n_H} e^{is^TX_{H,k_H}} - f_{\mathbf{x}}(s)\right)\end{pmatrix}\\ &\equiv \mathbf{A}_n\mathbf{U}_n.\end{aligned}$$

$\mathbf{A}_n \xrightarrow{P} \mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix}1-p_1 & -\sqrt{p_1p_2} & \cdots & \sqrt{p_1p_H}\\ -\sqrt{p_1p_2} & 1-p_2 & \cdots & -\sqrt{p_2p_H}\\ \vdots & \vdots & & \vdots\\ -\sqrt{p_1p_H} & -\sqrt{p_2p_H} & \cdots & 1-p_H\end{pmatrix}$$

$\mathbf{A}$ is an idempotent matrix with rank $H-1$. by central limit theorem, $\mathbf{U}_n \xrightarrow{D} \mathbf{U}$, where

$\mathbf{U} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ with $\sigma^2 = 1 - |f_{\mathbf{X}}(s)|^2$. Therefore,

$$\begin{pmatrix} \Gamma_{n,1}(s) \\ \Gamma_{n,2}(s) \\ \vdots \\ \Gamma_{n,H}(s) \end{pmatrix} \xrightarrow{D} \mathbf{AU} \sim N\left(\mathbf{0}, \begin{pmatrix} (1-p_1)(1-f^2_{\mathbf{X}}(s)) & & \\ & \ddots & \\ & & (1-p_H)(1-f^2_{\mathbf{X}}(s)) \end{pmatrix}\right),$$

$$\sum_{h=1}^{H} |\Gamma_{n,y}(s)|^2 = \mathbf{U}_n{}^T\mathbf{AU}_n \xrightarrow{D} \left[1 - |f_{\mathbf{X}}(s)|^2\right] Q,$$

follows continuous mapping theorem, where $A(s) = \mathbf{AU}$, and $Q \sim \chi^2_{H-1}$. □

Remark 2: Proof of Theorem 4.2 under dependence of slicing. Let $I_i^{(h)} \equiv I\{Y_i = y^{(h)}\}$, $S \equiv \{s_1, s_2, \cdots, s_H\}$, $S^{-h} \equiv S \setminus \{s_h\}$, $I_{s_h}^{(h)} = \left(\sum_{i=1}^n I_i^{(h)}\right)$, $M_S \equiv \prod_{h=1}^H I_{s_h}^{(h)}$, $M_S^{-h} \equiv M_S / I_{s_h}^{(h)}$, $p_n \equiv \prod_{h=1}^H \frac{n_h}{n}$ and $p \equiv \prod_{h=1}^H p_h$.

Define

$$\Gamma_{n,y}(s) = \sqrt{n_y}\left[\frac{1}{n_y}\sum_{t=1}^{n_y} e^{is^T\mathbf{X}_t^{(h)}} - \frac{1}{n}\sum_{t=1}^{n} e^{is^T\mathbf{X}_t}\right], h = 1, \cdots, H.$$

Then

$$n\mathcal{C}_n^2(\mathbf{X}|Y) = \int \sum_{h=1}^{H} |\Gamma_{n,h}(s)|^2 \, w(s) ds.$$

Note that

$$\begin{aligned}\Gamma_{n,h}(s) &= \sqrt{n_h}\left(\frac{\frac{1}{n}\sum_{t=1}^{n} e^{is^T\mathbf{X}_t} I_t^{(h)}}{\frac{1}{n}\sum_{t=1}^{n} I_t^{(h)}} - \frac{1}{n}\sum_{t=1}^{n} e^{is^T\mathbf{X}_t}\right) \\ &= \sqrt{n_h}\frac{\frac{1}{n^H}\sum_{t,S^{-h}} e^{is^T\mathbf{X}_t} I_t^{(h)} M_s^{-h} - \frac{1}{n^{H+1}}\sum_{t,S} e^{is^T\mathbf{X}_t} M_s}{\prod_{h=1}^{H}\left(\frac{1}{n}\sum_{t=1}^{n} I_t^{(h)}\right)} \\ &\equiv \sqrt{n_h}\frac{V_n^{(h)} - V_n^{(0)}}{p_n},\end{aligned}$$

where $V_n^{(h)}(h = 0, \cdots, H)$ are $V-$statistics. We denote the corresponding $U-$statistics by $U_n^{(h)}(h = 1, \cdots, H)$ with kernel $h^{(h)}(t, S^{-h}) \equiv e^{is^T\mathbf{X}_t} I_t^{(h)} M_S^{-h}$ $(h = 1, \cdots, H)$ and $U_n^{(0)}$ with kernel $h^{(0)}(t, S) \equiv e^{is^T\mathbf{X}_t} M_S$, respectively. There kernels are not symmetric in their arguments but can be fixed by averaging over permutations. Note that $Eh^{(h)} = pf_{\mathbf{X}}(s)$ for $h = 0, \cdots, H$. We consider the following projections of $U_n^{(h)}$: $\hat{U}_n^{(h)} \equiv \frac{H}{n}\sum_{t=1}^{n}\tilde{h}_1^{(h)}(\mathbf{X}_t) + Eh^{(h)}, h = 1, \cdots, H$ and $\hat{U}_n^{(0)} \equiv \frac{H+1}{n}\sum_{t=1}^{n}\tilde{h}_1^{(0)}(\mathbf{X}_t) + Eh^{(0)}$, where

$$\begin{aligned}\tilde{h}_1^{(h)}(\mathbf{X}_t) &\equiv E(h^{(h)}(t, S^{-h})|\mathbf{X}_t) - Eh^{(h)}(t, S^{-h}) \\ &= \frac{1}{H}\frac{p}{p_h} e^{is^T\mathbf{X}_t} I_t^{(h)} + \sum_{h' \neq h}\frac{1}{H}\frac{p}{p'_h} f_{\mathbf{X}}(s) I_t^{(h')} - pf_{\mathbf{X}}(s), h = 1, \cdots, H; \\ \tilde{0}_1^{(h)}(\mathbf{X}_t) &\equiv E(h^{(0)}(t, S^{-h})|\mathbf{X}_t) - Eh^{(0)}(t, S^{-h}) \\ &= \frac{1}{H+1} p e^{is^T\mathbf{X}_t} + \sum_{h=1}^{H}\frac{1}{H+1}\frac{p}{p_h} f_{\mathbf{X}}(s) I_t^{(h)} - pf_{\mathbf{X}}(s).\end{aligned}$$

Let $\tilde{\Gamma}_{n,h}(s) \equiv \sqrt{n_h}\frac{\hat{U}_n^{(h)} - \hat{U}_n^{(0)}}{p_n}$. Applying Theorem 5.3.2 in Serfling (2009), we have $E(\hat{U}_n^{(h)} - U_n^{(h)})^2 = O(n^{-2})$ as $E\left|h^{(h)}\right|^2 < \infty$ and hence, $\sqrt{n}(\hat{U}_n^{(h)} - U_n^{(h)}) \xrightarrow{P} 0$ by

Chebyshev's inequality. Similarly, $\sqrt{n}(V_n^{(h)} - U_n^{(h)}) \xrightarrow{P} 0$ by Lemma 5.7.3 in Serfling (2009) and Chevyshev's inequality. Therefore, $\Gamma_{n,h}(s) - \tilde{\Gamma}_{n,h}(s) \xrightarrow{P} 0$ for $h = 1, \cdots, H$. Note that

$$
\begin{aligned}
\tilde{\Gamma}_n(s) &= \begin{pmatrix} \tilde{\Gamma}_{n,1}(s) \\ \tilde{\Gamma}_{n,2}(s) \\ \vdots \\ \tilde{\Gamma}_{n,H}(s) \end{pmatrix} = p_n^{-1} \begin{pmatrix} \sqrt{\frac{n_1}{n}} & 0 & \cdots & 0 & -\sqrt{\frac{n_1}{n}} \\ 0 & \sqrt{\frac{n_2}{n}} & \cdots & 0 & -\sqrt{\frac{n_2}{n}} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\frac{n_H}{n}} & -\sqrt{\frac{n_H}{n}} \end{pmatrix} \begin{pmatrix} \sqrt{n}(\hat{U}_n^{(1)} - pf_{\mathbf{X}}(s)) \\ \sqrt{n}(\hat{U}_n^{(2)} - pf_{\mathbf{X}}(s)) \\ \vdots \\ \sqrt{n}(\hat{U}_n^{(H)} - pf_{\mathbf{X}}(s)) \\ \sqrt{n}(\hat{U}_n^{(0)} - pf_{\mathbf{X}}(s)) \end{pmatrix} \\
&\equiv p_n^{-1}\mathbf{A}_n\mathbf{U}_n.
\end{aligned}
$$

Applying multivariate Central Limit theorem,

$$
\hat{\mathbf{U}}_n = \sqrt{n}\frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} H\tilde{h}_1^{(1)}(\mathbf{X}_t) \\ \vdots \\ H\tilde{h}_1^{(H)}(\mathbf{X}_t) \\ (H+1)\tilde{h}_1^{(0)}(\mathbf{X}_t) \end{pmatrix} \equiv \sqrt{n}\frac{1}{n}\sum_{t=1}^{n} \tilde{h}_1(\mathbf{X}_t) \xrightarrow{D} \hat{\mathbf{U}},
$$

where $\hat{\mathbf{U}} \sim N(\mathbf{0}, \tilde{\Sigma})$ with $\tilde{\Sigma} = \mathrm{Cov}\left(\tilde{h}_1(\mathbf{X})\right)$. Also note that $p_n \xrightarrow{P} p$ and $\mathbf{A}_n \xrightarrow{P} \mathbf{A}$, where

$$\mathbf{A} = \begin{pmatrix} \sqrt{p_1} & 0 & \cdots & 0 & -\sqrt{p_1} \\ 0 & \sqrt{p_2} & \cdots & 0 & -\sqrt{p_2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \sqrt{p_H} & -\sqrt{p_H} \end{pmatrix}.$$

Eventually, we have $\tilde{\Gamma}_n(s) \xrightarrow{D} A(s)$ by Slutsky's theorem, where $A(s) \sim N(0, \sigma^2\Sigma)$ with $\sigma^2 = 1 - |f_{\mathbf{X}}(s)|^2$ and

$$\Sigma = \frac{1}{p^2\sigma^2}\mathbf{A}\tilde{\Sigma}\mathbf{A}^T = \begin{pmatrix} 1-p_1 & -\sqrt{p_1p_2} & \cdots & \sqrt{p_1p_H} \\ -\sqrt{p_1p_2} & 1-p_2 & \cdots & -\sqrt{p_2p_H} \\ \vdots & \vdots & & \vdots \\ -\sqrt{p_1p_H} & -\sqrt{p_2p_H} & \cdots & 1-p_H \end{pmatrix}.$$

Note that Σ is idempotent with rank $H-1$, then

$$\sum_{h=1}^{H} |\Gamma_{n,h}(s)|^2 \xrightarrow{D} \left[1 - |f_{\mathbf{X}}(s)|^2\right] Q,$$

where $Q \sim \chi^2_{H-1}$. □

Alternative proof of Theorem 4.2 using $U-$statistics method. Let $D_{ij} \equiv -|\mathbf{X}_i - \mathbf{X}_j|$, $I_i^{(h)} \equiv I\{Y_i = y^{(h)}\}$ for $h = 1, \cdots, H$, and $\tilde{D}_{ij} \equiv -|\mathbf{X}_i - \mathbf{X}_j| + \mathrm{E}_{\mathbf{X}}|\mathbf{X}_i - \mathbf{X}| + \mathrm{E}_{\mathbf{X}}|\mathbf{X} - \mathbf{X}_j| -$

$\mathrm{E}_{\mathbf{X},\mathbf{X}'}|\mathbf{X}-\mathbf{X}'|$. Then

$$\begin{aligned}\mathcal{C}_n^2(\mathbf{X}|Y) &= \sum_{h=1}^{H}\frac{n}{n_h}\frac{1}{n^2}\sum_{i,j=1}^{n}D_{ij}I_i^{(h)}I_j^{(h)} - \frac{1}{n^2}\sum_{i,j=1}^{n}D_{ij}\\ &= \sum_{h=1}^{H}\frac{n}{n_h}\frac{1}{n^2}\sum_{i,j=1}^{n}\tilde{D}_{ij}I_i^{(h)}I_j^{(h)} - \frac{1}{n^2}\sum_{i,j=1}^{n}\tilde{D}_{ij}\\ &\equiv \sum_{h=1}^{H}\frac{n}{n_h}V_n^{(h)} - V_n^{(0)},\end{aligned}$$

where $V_n^{(h)}(h=1,\cdots,H)$ and $V_n^{(0)}$ are $V-$statistics. We denote the corresponding $U-$statistics by $U_n^{(h)} \equiv \frac{1}{n(n-1)}\sum_{i\neq j}h_{ij}^{(h)}$ with kernel $h_{ij}^{(h)} \equiv \tilde{D}_{ij}I_i^{(h)}I_j^{(h)}$ for $h=1,\cdots,H$, and $U_n^{(0)} \equiv \frac{1}{n(n-1)}\sum_{i\neq j}h_{ij}^{(0)}$ with kernel $h_{ij}^{(0)} \equiv \tilde{D}_{ij}$, respectively. Let $\tilde{\mathcal{C}}_n^2(\mathbf{X}|Y) \equiv \sum_{h=1}^{H}\frac{n}{n_h}U_n^{(h)} - U_n^{(0)}$, then

$$\begin{aligned}n\mathcal{C}_n^2(\mathbf{X}|Y) - n\tilde{\mathcal{C}}_n^2(\mathbf{X}|Y) &= \sum_{h=1}^{H}\frac{n}{n_h}n(V_n^{(h)}-U_n^{(h)}) - n(V_n^{(0)}-U_n^{(0)})\\ &= \sum_{h=1}^{H}\frac{n}{n_h}(\frac{1}{n}\sum_{i=1}^{n}\tilde{D}_{ii}I_i^{(h)} - U_n^{(h)}) - (\frac{1}{n}\sum_{i=1}^{n}\tilde{D}_{ii} - U_n^{(0)})\\ &\xrightarrow{P} (H-1)\mathcal{C}^2(\mathbf{X}|\mathbf{X}).\end{aligned}$$

Thus, our objective is to show

$$n\tilde{\mathcal{C}}_n^2(\mathbf{X}|Y) \xrightarrow{D} \mathcal{C}^2(\mathbf{X}|\mathbf{X})[Q-(H-1)], \tag{S4.7}$$

where $Q \sim \chi^2_{H-1}$.

A special representation for $h_{ij}^{(h)}$ as in Serfling (2009, p.196) will be used. Let $\{\phi_m^{(h)}(\cdot)\}$ denote the orthonormal eigenfunctions corresponding to the eigenvalues $\{\lambda_m^{(h)}\}$ defined in connection with $h_{ij}^{(h)}$, i.e., $\{\phi_m^{(h)}(\cdot)\}$ satisfies the following for $h=0,\cdots,H$:

(i) $E_{(\mathbf{X}_j,Y_j)}[h_{ij}^{(h)}\phi_m^{(h)}(\mathbf{X}_j,Y_j)] = \lambda_m^{(h)}\phi_m^{(h)}(\mathbf{X}_i,Y_i)$

(ii) $E[\phi_{m_1}^{(h)}\phi_{m_2}^{(h)}] = \begin{cases} 1, & m_1 = m_2 \\ 0, & m_1 \neq m_2 \end{cases}$

(iii) $\lim_{M\to\infty} E[h_{ij}^{(h)} - \sum_{m=1}^{M}\lambda_m^{(h)}\phi_m^{(h)}(\mathbf{X}_i,Y_i)\phi_m^{(h)}(\mathbf{X}_j,Y_j)]^2 = 0.$

Then we write $h_{ij}^{(h)} = \sum_{m=1}^{\infty}\lambda_m^{(h)}\phi_m^{(h)}(\mathbf{X}_i,Y_i)\phi_m^{(h)}(\mathbf{X}_j,Y_j)$. In the same sense, we have $h_1^{(h)}(\mathbf{X}_i,Y_i) \equiv E_{(\mathbf{X}_j,Y_j)}h_{ij}^{(h)} = \sum_{m=1}^{\infty}\lambda_m^{(h)}\phi_m^{(h)}(\mathbf{X}_i,Y_i)E[\phi_m^{(h)}(\mathbf{X}_j,Y_j)]$. Not that $E_{\mathbf{X}_j}\tilde{D}_{ij} = 0$ so $h_i^{(h)}(\mathbf{X}_i,Y_i) = 0$. Therefore, $E(\phi_m^{(h)}) = 0$ since $\mathrm{Var}(h_i^{(h)}) = 0$, for all m.

Similarly, let $\{\phi_m(\cdot)\}$ denote orthonormal eigenfunctions corresponding to the eigenvalues $\{\lambda_m\}$ defined in connection with $h_{ij}^{(0)}$. Therefore, $E(\phi_m) = 0$ and we can deduce from (i) and (ii) that:

(a) $\phi_m^{(h)}(\mathbf{X}_i,Y_i) = \frac{1}{\sqrt{p_h}}\phi_m(\mathbf{X}_i)I_i^{(h)}$ for $h = 1,\cdots,H$

(b) $\phi_m^{(0)}(\mathbf{X}_i,Y_i) = \phi_m(\mathbf{X}_i)$

(c) $\frac{\lambda_m^{(h)}}{p_h} = \lambda_m^{(0)} = \lambda_m, h = 1,\cdots,H.$

The following are the explanation for (a) and (c), and the rest follows from the same logic. From (i),

$$\begin{aligned}\lambda_m^{(h)}\phi_m^{(h)}(\mathbf{X}_i,Y_i) &= E_{(\mathbf{X}_j,Y_j)}[h_{ij}^{(h)}\phi_m^{(h)}(\mathbf{X}_j,Y_j)] \\ &= \begin{cases} 0, & \text{if } \mathbf{Y}_i \neq \mathbf{y}^{(h)} \\ p_h E_{\mathbf{X}_j}[\tilde{D}_{ij}\phi_m^{(h)}(\mathbf{X}_j,\mathbf{y}^{(h)})], & \text{if } \mathbf{Y}_i = \mathbf{y}^{(h)} \end{cases}\end{aligned}$$

for $h = 1, \cdots, H$. Hence, $\phi_m^{(h)}(\mathbf{X}_i, Y_i) = c\phi_m(\mathbf{X}_i)I_i^{(h)}$ for some constant c and $\frac{\lambda_m^{(h)}}{p_h} = \lambda_m$, for $h = 1, \cdots, H$. Required by (ii), $c = \frac{1}{\sqrt{p_h}}$.

Let $\tilde{D} \equiv \sum_{m=1}^{\infty} \lambda_m [Q - (H-1)]$ and $\tilde{D}_M \equiv \sum_{m=1}^{M} \lambda_m [Q - (H-1)]$, where $Q \sim \chi^2_{H-1}$. putting $T_n^{(h)} \equiv \frac{1}{n}\sum_{i \neq j} h_{ij}^{(h)}$, we have $nU_n^{(h)} = \frac{n}{n-1} T_n^{(h)}$. In terms of the above presentation for $h_{ij}^{(h)}$, $T_n^{(h)} = \frac{1}{n}\sum_{i\neq j}\sum_{m=1}^{\infty} \lambda_m^{(h)} (\mathbf{X}_i, Y_i)\phi_m^{(h)}(\mathbf{X}_j, Y_j)$ and let $T_{n,M}^{(h)} = \frac{1}{n}\sum_{i\neq j}\sum_{m=1}^{M} \lambda_m^{(h)}(\mathbf{X}_i, Y_i)\phi_m^{(h)}(\mathbf{X}_j, Y_j)$ for $h = 0, \cdots, H$. Eventually, we will show that

$$n\tilde{\mathcal{C}}_n^2(\mathbf{X}|Y) = \frac{n}{n-1}\left[\sum_{l=1}^{L} \frac{n}{n_h} T_n^{(h)} - T_n^{(0)}\right] \xrightarrow{D} \tilde{D} \tag{S4.8}$$

by using characteristic functions. The proof is decomposed into three parts as follows.

(1) Given $\epsilon > 0$ and s, $\left| Ee^{is\left(\sum_{h=1}^{H} \frac{n}{n_h} T_n^{(h)} - T_n^{(0)}\right)} - Ee^{is\left(\sum_{h=1}^{H} \frac{n}{n_h} T_{n,M}^{(h)} - T_{n,M}^{(0)}\right)} \right| < \epsilon$ for M and n sufficientlly large. Using the inequality $|e^{iz} - 1| \le |z|$, we have

$$\begin{aligned} &\left| Ee^{is\left(\sum_{h=1}^{H} \frac{n}{n_h} T_n^{(h)} - T_n^{(0)}\right)} - Ee^{is\left(\sum_{h=1}^{H} \frac{n}{n_h} T_{n,M}^{(h)} - T_{n,M}^{(0)}\right)} \right| \\ &\le |s| E \left| \sum_{h=1}^{H} \frac{n}{n_h} (T_n^{(h)} - T_{n,M}^{(h)}) - (T_n^{(0)} - T_{n,M}^{(0)}) \right| \\ &\le |s| \left(\sum_{h=1}^{H} \frac{n}{n_h} E \left| T_n^{(h)} - T_{n,M}^{(h)} \right| + E \left| T_n^{(0)} - T_{n,M}^{(0)}) \right| \right) \\ &\le |s| \left\{ \sum_{h=1}^{H} \frac{n}{n_h} \left[E\left(T_n^{(h)} - T_{n,M}^{(h)} \right)^2 \right]^{1/2} + \left[E\left(T_n^{(0)} - T_{n,M}^{(0)} \right)^2 \right]^{1/2} \right\}. \end{aligned}$$

Similar to Serfling (2009, p.197-p.198), we can show that $\sum_{m=1}^{\infty} \left[\lambda_m^{(h)}\right]^2 = E\left[h_{ij}^{(h)}\right]^2 < \infty$ and $E\left(T_n^{(h)} - T_{n,M}^{(h)}\right)^2 \le 2\sum_{m=M+1}^{\infty} \left[\lambda_m^{(h)}\right]^2$ for $h = 0, \cdots, H$. Combining with the fact that $\frac{n}{n_h} \xrightarrow{a.s.} \frac{1}{p_h}$, the conclusion follows.

(2) $\sum_{h=1}^{H}\frac{n}{n_h}T_{n,M}^{(h)}-T_{n,M}^{(0)}\xrightarrow{D}\tilde{D}_M$. We may write

$$T_{n,M}^{(h)}=\sum_{m=1}^{M}\lambda_m^{(h)}\left[\left(W_{n,m}^{(h)}\right)^2-R_{n,m}^{(h)}\right],$$

where $W_{n,m}^{(h)}\equiv n^{-1/2}\sum_{i=1}^{n}\phi_m^{(h)}(\mathbf{X}_i,Y_i)$ and $R_{n,m}^{(h)}\equiv n^{-1}\sum_{i=1}^{n}\left[\phi_m^{(h)}(\mathbf{X}_i,Y_i)\right]^2$. From the foregoing consideration, it can be seen that

$$\mathbf{W}_{n,m}\equiv\begin{pmatrix}W_{n,m}^{(1)} & \cdots & W_{n,m}^{(H)} & W_{n,m}^{(0)}\end{pmatrix}^T\xrightarrow{D}\mathbf{W}_m,$$

where $\mathbf{W}_m\sim N(\mathbf{0},\Sigma)$ with

$$\Sigma=\begin{pmatrix}1 & \mathbf{0} & 0 & \sqrt{p_1}\\ \mathbf{0} & \ddots & \mathbf{0} & \vdots\\ 0 & \mathbf{0} & 1 & \sqrt{p_H}\\ \sqrt{p_1} & \cdots & \sqrt{p_H} & 1\end{pmatrix}$$

and $\mathrm{Cov}(\mathbf{W}_{n,m_1},\mathbf{W}_{n,m_2})=\mathbf{0}$ for $m_1\neq m_2$. Also, $R_{n,m}^{(h)}\xrightarrow{P}1$ for $h=0,\cdots,H$. Let $\mathbf{A}_n\equiv\mathrm{diag}\left(\sqrt{\frac{n}{n_1}p_1}\quad\cdots\quad\sqrt{\frac{n}{n_H}p_H}\ i\right)$, $i^2=-1$, then $\mathbf{A}_n\xrightarrow{P}\mathbf{A}\equiv\mathrm{diag}(1\quad\cdots\quad 1\ i)$ and $\mathbf{A}\Sigma\mathbf{A}^T$ has and only has non-zero eigenvalue 1 with multiplicity $H-1$. Therefore, $(\mathbf{A}\mathbf{W}_{n,m})^T\mathbf{A}\mathbf{W}_{n,m}\xrightarrow{D}Q$, where $Q\sim\chi_{H-1}^2$ and

$$\begin{aligned}\sum_{h=1}^{H}\frac{n}{n_h}T_{n,M}^{(h)}-T_{n,M}^{(0)}&=\sum_{m=1}^{M}\left[(\mathbf{A}\mathbf{W}_{n,m})^T\mathbf{A}\mathbf{W}_{n,m}-\left(\sum_{h=1}^{H}p_h\frac{n}{n_h}R_{n,m}^{(h)}-R_{n,m}^{(0)}\right)\right]\\&\xrightarrow{D}\tilde{D}_M.\end{aligned}$$

(3) Given $\epsilon>0$, $\left|Ee^{is\tilde{D}}-Ee^{is\tilde{D}_M}\right|<\epsilon$ for M sufficiently large. This can be seen by Serfling (2009, p.199). Combining (1) with (3), we can establish (S4.8). To

finish the proof of (S4.7), we need to show that $\sum_{m=1}^{\infty} \lambda_m = \mathcal{C}^2(\mathbf{X}|\mathbf{X})$. Indeed, $\mathcal{C}^2(\mathbf{X}|\mathbf{X}) = \sum_{m=1}^{\infty} \lambda_m E\phi_m^2(\mathbf{X}) = \sum_{m=1}^{\infty} \lambda_m$.

□

Proof of Theorem 4.3. Note that $\mathcal{C}_{n,k}^2(\mathbf{X}|\mathbf{Y}) = \frac{1}{n^2}\sum_{i,j} |\mathbf{X}_i - \mathbf{X}_j| - \hat{m}$, the first term is a V-statistic, which is root-n consistent to $\mathrm{E}|\mathbf{X} - \mathbf{X}'|$. For the second term,

$$
\begin{aligned}
\hat{m} - E(m(\mathbf{y})) &= \frac{1}{n}\sum_{i=1}^{n} \hat{m}(\mathbf{y}_i) - E(m(\mathbf{y})) \\
&= \frac{1}{n}\sum_{i=1}^{n} (\hat{m}(\mathbf{y}_i) - m(\mathbf{y}_i)) + \frac{1}{n}\sum_{i=1}^{n} m(\mathbf{y}_i) - E(m(\mathbf{y}))
\end{aligned}
$$

The first part tends to 0 based on Lemma 4.3 and the second part tends to 0 by LLN theory, Thus Theorem 4.3 holds. □

S5 Additional simulation studies

In this section, we report additional simulations related to the examples in the paper.

Example 1. This is an additional example, following Example 6.1. We try other dependency scenarios when p-value is far from 0. That is we simulate (X, Y) follows a standard bivariate normal distribution with covariance ρ and sample size n. Tables 1 and 2 show the mean and standard deviation of the p-values using dCov, kernel methods and slicing methods with different number of slices, under 100 replicates for $n = 200$ and 400. respectively.

Sample sizes affect all methods. In general, dCov is the best which is not surprising as the model is a "regression" type. For slicing methods, small number of slices are preferred, $H = 2, 5$ should be preferred. When $\rho = 0.05$ (signal is weak enough), or $\rho = 0.3$ (signal is strong enough), methods of dCov, R_c(gau), R_c(epa) and R_c(slice) for $H = 2, 5$ are quite consistent, especially for $n = 400$.

Table 1: Mean and standard deviation of p-value for $n = 200$ with 100 replicates

ρ	dCov	R_c(epa)	R_c(gau)	R_c(slice), H=2	R_c(slice), H=5
0.05	0.4691(0.2755)	0.4831(0.2951)	0.4727(0.2863)	0.43(0.2735)	0.5219(0.3013)
0.1	0.291(0.2743)	0.3589(0.2722)	0.3342(0.2787)	0.3512(0.2983)	0.3794(0.2889)
0.15	0.1469(0.1886)	0.2731(0.2514)	0.2579(0.2644)	0.2525(0.2659)	0.2422(0.2423)
0.2	0.0931(0.177)	0.1609(0.216)	0.1333(0.1953)	0.1469(0.2115)	0.1773(0.2376)
0.25	0.0165(0.0546)	0.0508(0.1336)	0.0354(0.0972)	0.036(0.0703)	0.0404(0.1007)
0.3	0.0064(0.0099)	0.013(0.0223)	0.0082(0.0116)	0.0279(0.1157)	0.0168(0.0348)
ρ	R_c(slice), H=10	R_c(slice), H=20	R_c(slice), H=40	R_c(slice), H=50	R_c(slice), H=100
0.05	0.4936(0.3011)	0.5(0.2952)	0.4879(0.297)	0.4982(0.2776)	0.5165(0.3034)
0.1	0.4072(0.2945)	0.4067(0.2825)	0.4082(0.298)	0.4328(0.2979)	0.482(0.2993)
0.15	0.2759(0.2467)	0.3356(0.2859)	0.3325(0.2622)	0.39(0.2799)	0.4421(0.2878)
0.2	0.2292(0.2645)	0.2878(0.2849)	0.3197(0.2682)	0.3751(0.2853)	0.4374(0.2801)
0.25	0.0583(0.1047)	0.124(0.1736)	0.2173(0.2387)	0.2394(0.2292)	0.3456(0.2741)
0.3	0.0348(0.0723)	0.0717(0.139)	0.1532(0.1907)	0.1487(0.1683)	0.28(0.2392)

Example 2. Following Example 6.2 in the paper, we construct models 6.2 (e)–(g), where the dimensions of $\mathbf{X}$ and Y are the same as the models 6.2 (a)–(d), except that each individual random variable is independently generated from t_2, t_3 and χ_2^2 distributions, respectively. The empirical type-I errors at the nominal level of 0.1 for models 6.2 (e)–(g) are shown in table 3, while at the nominal significance level of 0.05

Table 2: Mean and standard deviation of p-value for $n = 400$ with 100 replicates

ρ	dCov	R_c(epa)	R_c(gau)	R_c(slice), H=2	R_c(slice), H=5
0.05	0.4361(0.2778)	0.4167(0.2804)	0.4248(0.2615)	0.4839(0.2788)	0.481(0.3012)
0.1	0.1848(0.2579)	0.2558(0.2591)	0.2459(0.2698)	0.2161(0.2499)	0.2346(0.2579)
0.15	0.0679(0.1457)	0.1573(0.2004)	0.1443(0.2025)	0.1216(0.2132)	0.1292(0.1984)
0.2	0.0085(0.0108)	0.0372(0.0712)	0.0234(0.0435)	0.0322(0.0729)	0.0338(0.0678)
0.25	0.0061(0.0058)	0.0115(0.031)	0.0104(0.0369)	0.0106(0.0186)	0.0129(0.0311)
0.3	0.0047(0)	0.0049(0.0015)	0.0049(0.0013)	0.0048(8e-04)	0.0048(0.001)
ρ	R_c(slice), H=10	R_c(slice), H=20	R_c(slice), H=40	R_c(slice), H=50	R_c(slice), H=100
0.05	0.4795(0.2895)	0.5158(0.2951)	0.5516(0.3104)	0.5112(0.3205)	0.5389(0.3109)
0.1	0.2891(0.2697)	0.3392(0.2963)	0.3994(0.2952)	0.4154(0.3256)	0.4291(0.3047)
0.15	0.1992(0.2386)	0.2635(0.2503)	0.3267(0.2656)	0.3607(0.2798)	0.4389(0.3042)
0.2	0.0518(0.0833)	0.089(0.1257)	0.1641(0.1894)	0.193(0.2037)	0.2869(0.2448)
0.25	0.0228(0.0709)	0.0525(0.1293)	0.1019(0.1773)	0.1212(0.1802)	0.2036(0.2355)
0.3	0.0069(0.017)	0.0085(0.0144)	0.0198(0.038)	0.0272(0.0493)	0.0914(0.1334)

are shown in table 4 for models 6.2 (a)–(d), and in table 5 for models 6.2 (e)-(g). Again, we have the same conclusion as in the paper.

Example 3. These additional simulations follow from Example 6.5 in the paper, but with different combinations of a, p, σ_x^2 and σ^2. Figure 1 shows similar power changes as in the paper. Again, kernel methods are the best.

Example 4. This example reports the computing time of dCov and the proposed slicing and kernel methods. Huo and Székely (2016) discussed a fast computing algorithm of the distance covariance measure, which reduces the computational complexity from $O(n^2)$ to $O(n \log n)$. We believe it is similarly possible to reduce the calculation complexity of the proposed measure.

Table 3: Empirical type-I error rates for 10,000 tests at nominal significance level of 0.1, using B replicates for models (e)–(g)

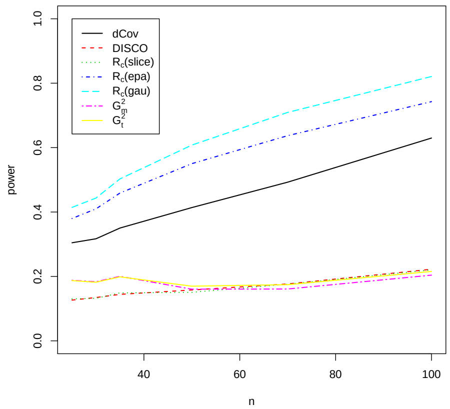
		(e) $t_2, p = 5, q = 1$					(f) $t_3, p = 5, q = 1$				
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)
25	400	0.105	0.103	0.105	0.101	0.102	0.101	0.102	0.101	0.105	0.100
30	366	0.097	0.096	0.096	0.093	0.099	0.101	0.099	0.101	0.096	0.099
35	342	0.105	0.103	0.102	0.097	0.105	0.098	0.102	0.102	0.100	0.096
50	300	0.095	0.096	0.095	0.101	0.102	0.096	0.097	0.097	0.102	0.099
70	271	0.100	0.103	0.103	0.100	0.101	0.098	0.096	0.096	0.097	0.097
100	250	0.098	0.095	0.097	0.098	0.100	0.099	0.098	0.099	0.102	0.102
		(g) $\chi_2^2, p = 5, q = 1$									
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)					
25	400	0.100	0.097	0.099	0.099	0.099					
30	366	0.099	0.097	0.098	0.096	0.097					
35	342	0.097	0.098	0.099	0.099	0.098					
50	300	0.102	0.102	0.103	0.103	0.104					
70	271	0.100	0.097	0.097	0.095	0.101					
100	250	0.100	0.100	0.099	0.100	0.096					

Table 4: Empirical type-I error rates for 10,000 tests at nominal significance level of 0.05, using B replicates for models (a)–(d)

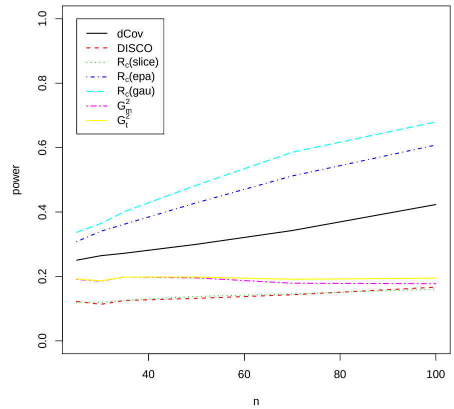
		(a) $N(0,1), p=5, q=1$					(b) $t_1, p=5, q=1$				
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)
25	400	0.051	0.054	0.054	0.050	0.051	0.047	0.046	0.048	0.050	0.050
30	366	0.049	0.055	0.055	0.050	0.049	0.050	0.053	0.051	0.049	0.052
35	342	0.049	0.050	0.051	0.049	0.053	0.051	0.047	0.046	0.049	0.050
50	300	0.049	0.051	0.051	0.054	0.054	0.048	0.048	0.048	0.050	0.051
70	271	0.050	0.048	0.048	0.047	0.048	0.045	0.046	0.047	0.051	0.049
100	250	0.047	0.049	0.051	0.049	0.043	0.044	0.046	0.047	0.045	0.046
		(c) $\chi^2_1, p=5, q=1$					(d) $\chi^2_3, p=5, q=1$				
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)
25	400	0.053	0.054	0.053	0.050	0.050	0.047	0.046	0.046	0.049	0.055
30	366	0.050	0.050	0.050	0.051	0.048	0.048	0.051	0.052	0.050	0.051
35	342	0.052	0.049	0.048	0.052	0.053	0.047	0.052	0.052	0.044	0.048
50	300	0.050	0.050	0.049	0.048	0.049	0.046	0.046	0.048	0.050	0.047
70	271	0.045	0.048	0.047	0.046	0.050	0.046	0.049	0.047	0.049	0.046
100	250	0.051	0.048	0.047	0.046	0.053	0.050	0.048	0.046	0.045	0.051

Table 5: Empirical type-I error rates for 10,000 tests at nominal significance level of 0.05, using B replicates for models (e)-(g)

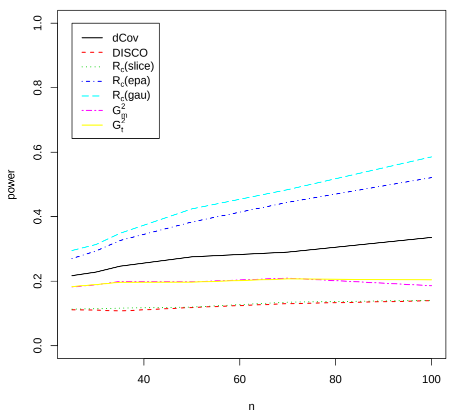
		(e) $t_2, p = 5, q = 1$					(f) $t_3, p = 5, q = 1$				
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)
25	400	0.051	0.050	0.050	0.054	0.053	0.051	0.050	0.051	0.052	0.049
30	366	0.050	0.049	0.048	0.050	0.051	0.049	0.046	0.045	0.050	0.047
35	342	0.050	0.050	0.049	0.051	0.047	0.050	0.048	0.048	0.055	0.051
50	300	0.052	0.050	0.049	0.049	0.051	0.050	0.050	0.051	0.050	0.048
70	271	0.045	0.047	0.048	0.048	0.045	0.044	0.045	0.046	0.047	0.046
100	250	0.047	0.046	0.045	0.045	0.049	0.046	0.047	0.047	0.047	0.047
		(g) $\chi^2_2, p = 5, q = 1$									
n	B	dCov	DISCO	R_c(slice)	R_c(epa)	R_c(gau)					
25	400	0.050	0.048	0.048	0.047	0.050					
30	366	0.051	0.052	0.051	0.048	0.050					
35	342	0.050	0.050	0.049	0.049	0.050					
50	300	0.046	0.050	0.050	0.050	0.048					
70	271	0.049	0.049	0.050	0.046	0.048					
100	250	0.051	0.052	0.050	0.045	0.049					



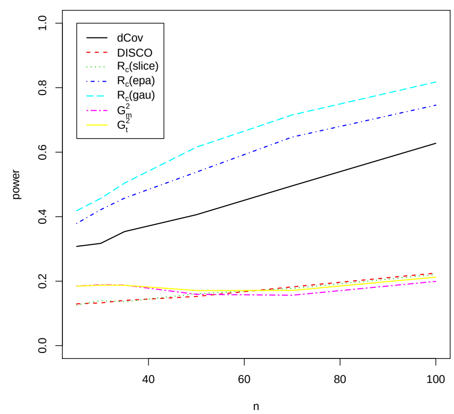
(a) $a = 0.5, p = 10, \sigma_x^2 = 1$ and $\sigma^2 = 1$.



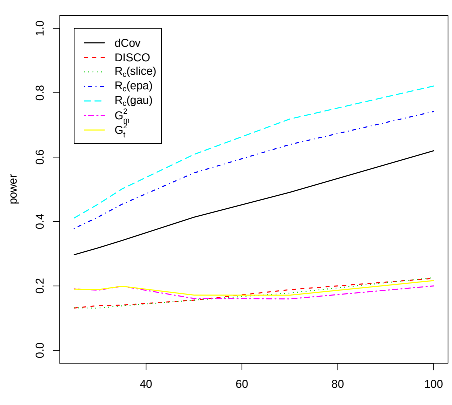
(b) $a = 0.3, p = 15, \sigma_x^2 = 1$ and $\sigma^2 = 1$.



(c) $a = 0.3, p = 20, \sigma_x^2 = 1$ and $\sigma^2 = 1$.



(d) $a = 0.3, p = 10, \sigma_x^2 = 1$ and $\sigma^2 = 0.25$.



(e) $a = 0.3, p = 10, \sigma_x^2 = 0.5$ and $\sigma^2 = 1$.

Figure 1: Empirical power with the change of sample size n for other different parameter combinations.

Since the fast computing algorithm is written in Matlab, it would not be fair to compare it with our methods which are written in R. In table 6, we compare the computation time (in sec) using the dCov method and the proposed slicing and kernel methods in R. We simulate X and Y independently from $U(0,1)$ with sample size n, n goes from 32 ($= 2^5$) to 2048 ($= 2^{11}$). Compute the average running time for the different implementations in R with 1000 replications at each sample size. We can see that slicing method is the fastest, while dCov is the second fastest. The two kernel methods are much slower as expected.

Table 6: Running time (in sec) for the direct dCov method, the slicing and kernel methods

Sample size	dCov	R_c(slice)	R_c(gau)	R_c(epa)
32	0.0007 (0.0007)	0.0014 (0.0005)	0.0014 (0.0026)	0.0055 (0.0011)
64	0.0021 (0.0035)	0.0023 (0.0028)	0.0034 (0.0028)	0.0195 (0.0050)
128	0.0059 (0.0062)	0.0047 (0.0041)	0.0157 (0.0077)	0.0671 (0.0126)
256	0.0287 (0.0194)	0.0135 (0.0106)	0.0868 (0.0184)	0.2983 (0.0233)
512	0.1347 (0.0117)	0.0856 (0.0194)	0.6008 (0.0301)	1.4453 (0.0577)
1024	0.4103 (0.0322)	0.1994 (0.0028)	4.8327 (0.1463)	8.1123 (0.1168)
2048	2.5765 (0.1683)	0.9775 (0.0391)	42.7910 (2.4216)	46.6297 (18.6964)

Bibliography

Akritas, M. G. and Arnold, S. F. (1994). Fully nonparametric hypotheses for factorial designs I: multivariate repeated measures designs. *J. Amer. Statist. Assoc.* **89**, 336-343.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology.* **26**, 32-46.

Chen, X., Cook, R. D. and Zou, C. (2015). Diagnostic studies in sufficient dimension reduction. *Biometrika.* **102**, 545-558.

Cochran, W. G. and Cox, G. M.(1957). *Experimental Designs.* 2nd ed. Wiley, New York.

Cook, R. D. (1998a). *Regression Graphics: Ideas for Studying Regressions Through Graphics.* Wiley, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Oxford.

Efron, B. and Tibshirani, R. J. (1998). *An Introduction to the Bootstrap.* Chapman and Hall/CRC, Boca Raton, Florida.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Gower, J. C. and Krzanowski, W. J. (1999). Analysis of distance for structured multivariate data and extensions to multivariate analysis of variance. *J. R. Statist. Soc.* C. **48**, 505-519.

Hand, D. J. and Taylor, C. C. (1987). *Multivariate Analysis of Variance and Repeated Measures*. Chapman and Hall, New York.

Herbin, E. and Merzbach, E. (2007). The multiparameter fractional Broanian motion. *In Math. Everwhere.* 93-101.

Huo, X. and Székely, G. (2016). Fast computing for distance covariance. *Technometrics.* **58**, 435-447.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press, San Diego.

McArdle, B. H. and Anderson, M. J. (2001). Fitting multivariate models to community data: a comment on distance-based redundancy analysis. *Ecology.* **82**, 290-297.

Prudnikov, A. P., Brychkov, A. and Marichev, O. I. (1986). *Integrals and Series.* Gordon and Breach Science Publishers, New York.

Resnick, S. I. (1999). *A Probability Path.* Birkhäuser, Boston.

Rizzo, M. L. and Székely, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Statist.* **4**, 1034-1055.

Shao, X. and Zhang, J. (2014). Martingale Difference correlation and its use in high-dimensional variable screening. *J. Amer. Statist. Assoc.* **109**, 1302-1318.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35**, 2769-2794.

Székely, G. J. and Rizzo. M. L. (2009). Brownian distance covariance. *Ann. Appl. Statist.* **3**, 1236-1265.

Zapala, M. A. and Schork, N. J. (2006). Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. *Proceedings of the National Academy of Sciences.* **103**, 19430-19435.